Svatopluk Kapounek, Hana Vránová (eds.)

38th International Conference on Mathematical Methods in Economics

September 9–11, 2020

Conference Proceedings

MENDELU
Faculty
of Business
and Economics

Mendel University in Brno
Faculty of Business and Economics

Svatopluk Kapounek
Hana Vránová (eds.)

38th INTERNATIONAL CONFERENCE ON MATHEMATICAL METHODS IN ECONOMICS (MME 2020)

Conference Proceedings

September 9–11, 2020
Mendel University in Brno
Czech Republic

International Conference on Mathematical Methods in Economics 2020 (MME 2020) was organized by the Faculty of Business and Economics, Mendel University in Brno.

Responsible Editors:

Svatopluk Kapounek
Hana Vránová

Programme Committee:

prof. Dr. Ing. Miroslav Plevný
President of the Czech Society for Operations Research
University of West Bohemia, Faculty of Economics

prof. Ing. Josef Jablonský, CSc.
Vice-president of the Czech Society for Operations Research
University of Economics in Prague, Faculty of Informatics and Statistics

prof. RNDr. Helena Brožová, CSc.
University of Life Science in Prague, Faculty of Economics and Management

prof. Ing. Mgr. Martin Dlouhý, Dr., MSc.
University of Economics in Prague, Faculty of Informatics and Statistics

doc. Ing. Jan Fábry, Ph.D.
University of Economics in Prague, Faculty of Informatics and Statistics

prof. RNDr. Ing. Petr Fiala, CSc., MBA
University of Economics in Prague, Faculty of Informatics and Statistics

Ing. Ludvík Friebel, Ph.D.
University of South Bohemia, Faculty of Economics

prof. Ing. Jana Hančlová, CSc.
Technical University of Ostrava, Faculty of Economics

doc. Ing. Svatopluk Kapounek, Ph.D.
Mendel University in Brno, Faculty of Business and Economics

doc. RNDr. Jana Klicnarová, Ph.D.
University of South Bohemia, Faculty of Economics

Ing. František Koblasa, Ph.D.
Technical University of Liberec, Faculty of Mechanical Engineering

doc. RNDr. Ing. Miloš Kopa, Ph.D.
Charles University, Faculty of Mathematics and Physics

doc. RNDr. Ing. Ladislav Lukáš, CSc.
University of West Bohemia, Faculty of Economics

prof. RNDr. Jan Pelikán, CSc.
University of Economics in Prague, Faculty of Informatics and Statistics

prof. RNDr. Jaroslav Ramík, CSc.
Silesian University in Opava, School of Business Administration in Karviná

Ing. Karel Sladký, CSc.
Academy of Sciences of the Czech Republic, Institute of Information Theory and Automation

Ing. Eva Šírová, Ph.D.
Technical University of Liberec, Faculty of Economics

doc. Ing. Tomáš Šubrt, Ph.D.
University of Life Science in Prague, Faculty of Economics and Management

doc. RNDr. Jana Talašová, CSc.
Palacký University in Olomouc, Faculty of Science

prof. RNDr. Milan Vlach, DrSc.
Charles University in Prague, Faculty of Mathematics and Physics
The Kyoto College of Graduate Studies for Informatics

prof. RNDr. Karel Zimmermann, DrSc.
Charles University in Prague, Faculty of Mathematics and Physics

prof. Ing. Miroslav Žižka, Ph.D.
Technical University of Liberec, Faculty of Economics

Organisation Committee:

Ing. Hana Vránová (chair)
Mendel University in Brno, Faculty of Business and Economics

Ing. Dominika Doubková
Mendel University in Brno, Faculty of Business and Economics

Ing. Barbora Šiklová
Mendel University in Brno, Faculty of Business and Economics

Ing. Andrea Prudilová
Mendel University in Brno, Faculty of Business and Economics

Technical Editors:

Ing. Jan Přichystal, Ph.D.
Mendel University in Brno, Faculty of Business and Economics

doc. Ing. Jiří Rybička, Dr.
Mendel University in Brno, Faculty of Business and Economics

ISBN 978-80-7509-734-7

Foreword

We are honoured to present you the Conference Proceedings devoted to the best selected contributions presented at the 38th international conference Mathematical Methods in Economics organized by Mendel University in Brno, Faculty of Business and Economics, under the auspice of the Czech Society of Operations Research, the Slovak Society for Operations Research, and the Czech Econometric Society. The conference was held in Brno, September 9–11, 2020.

This traditional meeting brings together academicians and professionals interested in the theory and applications of operations research and econometrics and it serves as a significant event in the field. We believe that the selected conference papers published in this proceedings will help you to develop new ideas to make the world a better place for life.

We welcomed more than 130 researchers who also served as discussants of the papers and helped to improve the quality of the research results presented during the conference days. Moreover, we hosted two distinguished well-known speakers who contributed to the conference programme with their speeches. Prof. Jesus Crespo Cuaresma (Vienna University of Economics and Business) gave a speech on a Model Uncertainty in Econometrics and Dr. Peter Molnár (University of Stavanger Business School in Norway) contributed to the discussion on the topic of Online attention at the Financial Markets.

In the presented Proceedings you find 103 papers which were selected based on the peer-review process. The contributions follow new trends in econometrics and operations research, and build bridges between researchers, academicians and practitioners in the industrial and institutional sectors sharing recent theoretical and applied results.

Finally, we would like to thank to all conference participants for their inspiring contributions. Furthermore, we are grateful to all the reviewers and the members of the scientific committee for their contribution to the organisation of this high-level scientific conference.

Let us also thank the members of the organising team for their support and hard work which contributed to the successful organisation of the conference.

Brno, September 2020 Svatopluk Kapounek

Table of Contents

Likelihood-Ratio Test and F-test for Two Exponential Means Equality: a Monte Carlo Power Exploration

Václav Adamec[1]

Abstract. In statistical practice, the exponential distribution is frequently evoked to explore problems of survival or utility times or in general, time to event scenarios. This study investigates power of statistical tests to verify equality of two exponential distributions: a traditional F-test and a likelihood-ratio test (LRT) in common and approximate variants. The tests were examined for power in both exact and bootstrapped forms. Two Monte Carlo simulations were set up to research the test power in response to the combined sample size, size of samples generated from the exponential distributions and mean ratio as a measure of the population means inequality. 15,000 MC runs were generated with additional 1,000 resampled data for the bootstrapped alternatives. The rejection rates generally increased with sample size, balanced samples and inequal means. Power was found inadequate for $n \leq 30$ in all examined combinations. A similar power was found, when samples were balanced, although it turned divergent with more unbalanced data. The bootstrapped tests generally showed increased empirical test size relative to nominal $\alpha = 0.05$ and superior power over the exact tests for simulated combinations with greater means occurring in large samples.

Keywords: Monte Carlo simulation, exponential distribution, F-test, likelihood--ratio test, bootstrap, power

JEL Classification: C12, C15, C41
AMS Classification: 65C05

1 Introduction

In statistics, the exponential distribution is often considered to model mathematically utility time, survival time or times between two or more successive events, which occur randomly and independently. For example, assuming individuals request service at the call center independently, the waiting times between the successive calls can be shown to be exponentially distributed.

The random variable X is said to follow the *exponential distribution*, i.e. $X \sim Exp(\delta)$, if its probability density function (pdf) is $f_X(x) = \delta e^{-\delta x}$ for $x \geq 0$, $\delta > 0$ or $f_X(x) = 0$, otherwise. Under the same conditions, the cumulative distribution function (cdf) is $F_X(x) = 1 - e^{-\delta x}$ and its inverse, the quantile function $F_X^{-1}(x) = -ln(1-p)/\delta$, where p denotes probability $p \in [0,1]$ associated with the quantile. For illustration, probability density and distribution functions for $X \sim Exp(3/4)$ are presented in Fig. 1. It is noticeable that with increasing rate parameter δ, the probability of X assuming small values increases and the long-run mean of the distribution becomes lower.

Moment-based characteristics of the exponential distribution depend on parameter δ, which could be interpreted as value of the pdf function corresponding to zero value of X, i.e. $\delta = f_X(x = 0)$. Specifically, the expected value (mean) of the distribution is

$$E(X) = \int_0^\infty x f_X(x) dx = \int_0^\infty x \delta e^{-\delta x} dx = 1/\delta, \tag{1}$$

and the variance

$$Var(X) = E(X^2) - E(X)^2 = 2/\delta^2 - 1/\delta^2 = 1/\delta^2, \tag{2}$$

since $E(X^2) = 2/\delta^2$. This implies that the expected value and standard deviation of the exponential distribution are both equal to $1/\delta$. The exponential distribution is known to have *memoryless* property, implying that

[1] Mendel University, School of Business and Economics, Department of Statistics and Operational Research, Zemědělská 1, 613 00 Brno, Czech Republic, vadamec@mendelu.cz

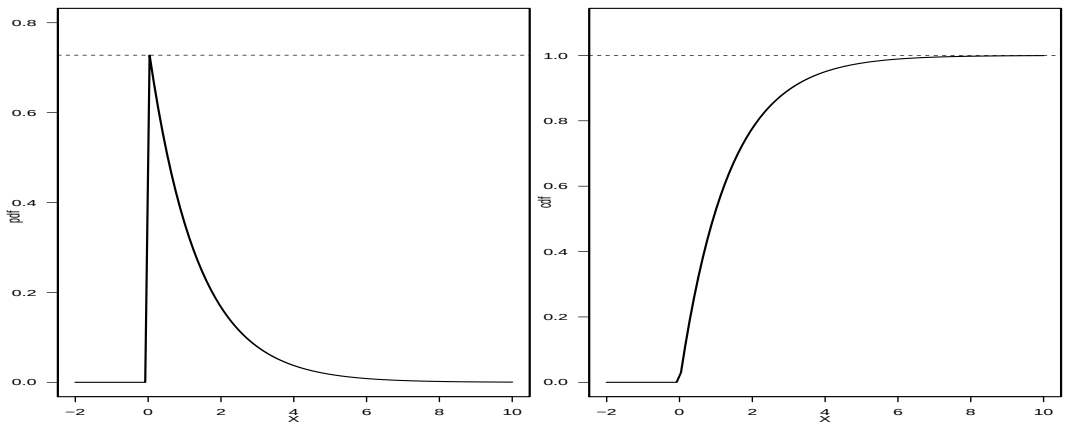


Figure 1 Probability density (left) and distribution function (right) of the exponential distribution with $\delta = 3/4$.

$P(X \geq t + w | X \geq w) = P(X \geq t)$. Proof thereof can be found in Chihara [3]. *Standardization* transformation $Y = X\delta$ yields $Y \sim Exp(1)$, where $E(Y) = Var(Y) = 1$.

Objective of this study is to simulate and investigate power of exact and bootstrapped two-sample statistical tests to verify equality in means of two exponential distributions via the simulation method of Monte Carlo (MC), which is one of the frequently applied methods. The power is expected to naturally oscilate with sample size, ratio of the means and sample size ratios. The statistical tests investigated include traditional exact F-test, exact likelihood ratio test, both regular and approximate, and boostrapped F-test and boostrapped likelihood-ratio test (F-test variant). Bootstrap modifications of the tests were also included in this study, since resampling could potentially increase power of statistical tests under specific scenarios. The empirical power curves are expected to assist the analyst with comparing competing statistical tests, selecting the best performing test for concrete data or refraining from use of insufficiently powerful test.

2 Material and Methods

2.1 Statistical Tests

F-test

Traditional F-test to verify equality of two exponential means is technically the same as for testing parameter equality of two Poisson distributions, as stated in Cox [4]. Specifically, the test hypotheses are defined $H_0 : \mu_x/\mu_y = 1$ vs. $H_1 : \mu_x/\mu_y \neq 1$. Assuming x_i, $i = 1, 2, \ldots, n_1$ and y_j, $j = 1, 2, \ldots, n_2$ are realized waiting times to event, which originated from two exponential distributions with rate parameters δ_1 and δ_2. The sample means of the distributions are estimated via $\hat{\mu}_x = n_1^{-1} \sum_{i=1}^{n_1} x_i$ and $\hat{\mu}_y = n_2^{-1} \sum_{j=1}^{n_2} y_i$. The test statistic is calculated as a proportion $F_{obs} = \hat{\mu}_x/\hat{\mu}_y$ following F-distribution with $df_1 = 2n_1$ and $df_2 = 2n_2$ degrees of freedom under H_0. The null hypothesis is rejected, if $F_{obs} \leq F_{\alpha/2;2n_1;2n_2}$ or $F_{obs} \geq F_{1-\alpha/2;2n_1;2n_2}$, as given in Han [6]. P-value is computed via $P = 2min(F_{2n_1;2n_2} \leq F_{obs}; F_{2n_1;2n_2} > F_{obs})$. Due to relationship between $E(X) = \mu$ and δ, the hypotheses verified by F-test relate directly to the parameters and thereby validate homogeneity of the exponential distributions.

$(1 - \alpha) \times 100\%$ two-tailed confidence interval for the ratio of the means can be constructed from the probabilistic statement (3) and used for testing statistical hypotheses

$$P\left(\frac{\hat{\mu}_x}{\hat{\mu}_y}\frac{1}{F_{1-\alpha/2;2n_1;2n_2}} \leq \frac{\mu_x}{\mu_y} \leq \frac{\hat{\mu}_x}{\hat{\mu}_y}\frac{1}{F_{\alpha/2;2n_1;2n_2}}\right) = 1 - \alpha. \tag{3}$$

Variant of the F-test based on resampling was implemented via applying principles of the nonparametric bootstrap. It assumed generating $B = 1,000$ bootstrap samples with replacement from the combined data x_i and y_i of size n_1, n_2, respectively and calculating ratios $F^* = \hat{\mu}_x^*/\hat{\mu}_y^*$ from the bootstrap samples, simulating thus the distribution of the test statistic under the asumption of mean equality. Provided $F_{obs} \geq 1$, p-value of the bootstrapped F-test is computed via $P = 2min(p_L, p_R)$, where $p_L = B^{-1} \sum_{i=1}^{B} I(F^* \leq F_{obs})$, $p_R = B^{-1} \sum_{i=1}^{B} I(F^* > F_{obs})$ and $I(.)$ indicates a binary indicator function. If $F_{obs} < 1$, then the respective left-

tailed and right-tailed probabilities were $p_L = B^{-1}\sum_{i=1}^{B} I(F^* \leq 1/F_{obs})$ and $p_R = B^{-1}\sum_{i=1}^{B} I(F^* > 1/F_{obs})$.

Likelihood Ratio Test

The likelihood ratio (LR) test is generally based on the -log likelihood ratio directly computed in (4) and identical to the formulas used in Lee [7]

$$-ln\lambda(F) = n_1 ln\left(\frac{n_1}{n_2} + \frac{1}{F_{obs}}\right) + n_2 ln\left(\frac{n_2}{n_1} + F_{obs}\right) + n_1 ln\left(\frac{n_2}{n_1+n_2}\right) + n_2 ln\left(\frac{n_1}{n_1+n_2}\right). \tag{4}$$

Under the assumption of mean equality, $-ln\lambda(F) = 0$. Provided $F_{obs} \geq 1$, a left-tailed F^L_{obs} must be identified, producing the same numerical value of the negative log likelihood ratio $-ln\lambda(F)$, as illustrated in Fig. 2. Consequently, a two-tailed p-value of the LR test can be easily computed as a sum $P = P_L + P_R$, where $P_L = P(F_{2n_1;2n_2} \leq F^L_{obs})$ and $P_R = P(F_{2n_1;2n_2} > F_{obs})$. Alternatively, an approximate p-value can be obtained via finding $P(\chi^2_1 \geq -2ln\lambda(F))$.

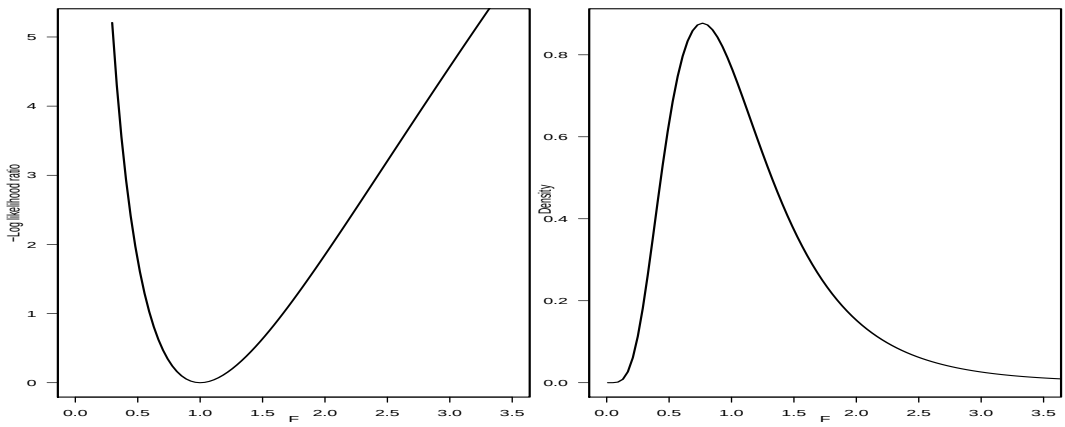


Figure 2 Negative log likelihood ratio (left) and probability density (right) for the ratio of the means; partially modified and adopted from https://stats.stackexchange.com.

Due to the fact that distribution of the mean ratio test statistic is strongly positively skewed, bootstrap variant of the likelihood ratio test was built based on a method of nonparametric bootstrap. $B = 1,000$ bootstrap samples with replacement from x_i, y_i of respective size n_1, n_2 were generated and F^* ratios were received. 95% confidence interval for the mean ratio was constructed from 0.025 and 0.975 percentiles of the resampled ratios F^*, as presented in Efron [5] or Ruppert [9]. Finally, the null hypothesis of mean equality was rejected, if the 95% confidence interval failed to include unity. This approach was facilitated by the one-to-one relationship between confidence intervals and hypothesis tests, as noted in Chernick [2].

2.2 Monte Carlo Simulations

To pursue objectives of this research, two separate Monte Carlo simulation schemes were set up. In the first MC scheme, two independent samples were simulated from the exponential distributions with parameters $\delta_1 = 1/\mu_x$ and $\delta_2 = 1/\mu_y$, where $\mu_y = k\mu_x$. Constant $k \in \{1.0, 1.3, 1.6, 1.9, 2.2, 2.5\}$ symbolizes the factor of mean increase in the second sample, relative to the first one and $\mu_x = 25$. Simulated sample size $n = n_1 + n_2$ of the combined data was $n \in \{30, 60, 90, 120\}$, where n_1 and n_2 were subsequently determined by spliting the total sample size n using sample size ratios $\kappa = n_1 : n_2$, $\kappa \in \{1:1, 3:2, 7:3, 4:1\}$. In overall, 96 combinations of parameters δ, sample size n and sample size ratios were produced.

For every simulated combination, we generated $R = 15,000$ MC runs, each producing two exponential samples of equal or unequal size. To every simulated sample pair, the researched two-sample statistical tests of equality in means of the exponential distribution were applied: exact F-test, exact LR test, exact LR test with chi-square approximation, bootstrapped F-test and bootstrapped LR test; the resulting binary outcomes of the named tests were stored.

Empirical rejection rates were calculated as a relative proportion of statistical test rejecting H_0 under uniformly applied significance level $\alpha = 0.05$. The empirical rejection rate under mean equality $\mu_x = \mu_y$ is generally regarded as an empirical size of the test. Simulated rejection rates were presented in tabular or graphical form. Program code of the tests, simulation schemes and plots were secured with R-software, v.

3.6.3. [8] and *lattice* package developed by Sarkar [10]. Simulation of a single MC scheme required approximately 21 hours on Windows 10 PC with Intel(R) Core(TM) i7-3770 processor 3.40 GHz.

In the second MC simulation scheme, we employed the identical setup as in the first scheme, however, the sample size ratios κ were inversed, i.e. $\kappa \in \{1:1, 2:3, 3:7, 1:4\}$. This scheme was implemented in response to possible confounding of the effects of the mean magnitude and the sample sizes n_1 and n_2 upon the simulated rejection rates.

3 Results and Discussion

3.1 Simulation 1

In Simulation 1, we observed that simulated rejection rates for all statistical tests were very similar, when applied to samples of equal size. The power curves indicate that the simulated power of the tests increased monotonically with the total sample size and mean ratio parameter k signifying greater differences between the means μ_x and μ_y. This finding is in conformity with the theoretical assumptions stated in Casella [1], that larger sample size and greater difference in means should lead to higher power of statistical test. For the smallest sample size combinations with $n \leq 30$, merely an inadequate power < 0.8 was attained due to lack of information in these samples, which makes it impossible for the tests to be efficiently applied on small samples in practice. Simulated rejection rates from Simulation 1 are presented in Fig. 3.

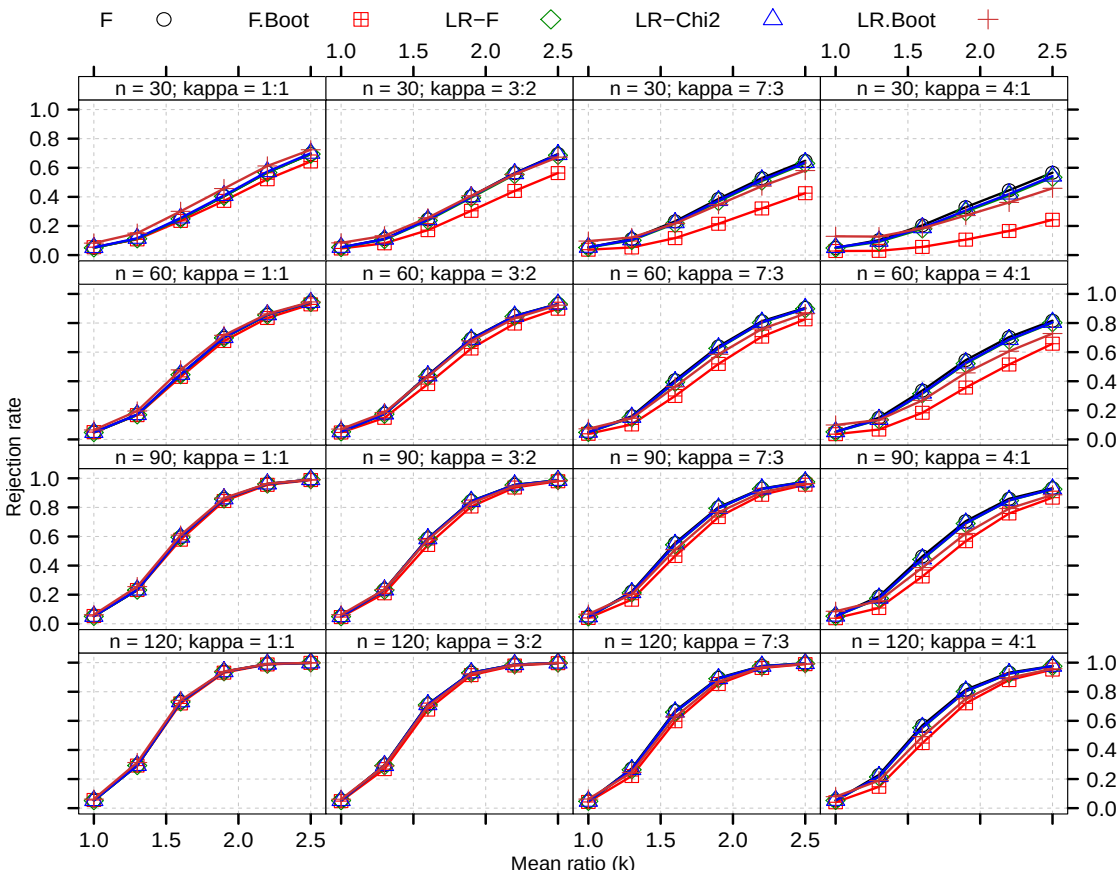


Figure 3 Rejection rates of exact and boostrapped statistical texts from Simulation 1 in response to sample size n, ratio of means k and sample size ratio κ.

In general, the power of the researched tests appeared the largest for the simulated variants with equal sample size n_1 and n_2. However, disparities in power among the tests visibly increased, when inequality in n_1 and n_2 became more prominent, as indicated by the rising value of κ ratios. The statistical tests in this study seemed to respond differentially to inequal amount of information stored in the simulated data: statistical tests applied on simulated data with larger differences in sample size (higher κ), clearly had reduced power, compared to more balanced data. Nonetheless, the power curves should be assessed in context of the simulation scheme, where data of lower sample size were drawn from a distribution with higher mean, in Simulation 1.

Bootstrapped tests, in this scheme, displayed generally inferior power, when compared to the exact tests. The best performing tests were F-test, LR χ^2-test, LR F-test, bootstrapped LR test and bootstrapped F-test, ranking from the best to the worst. Differences in power of the exact tests, however, could be viewed as of low importance. In the bootstrapped LR test, we nonetheless noticed elevated Type I error probability in small samples, compared to the other statistical tests, which is not an unusual discovery. Study of Han [6] indicates that two-sample F-test can be biased and show low power, which may cause erroneous inferences about the data. On the other hand, asymptotic LR test may require sufficiently large samples to maintain the α significance level, set forth by the analyst.

3.2 Simulation 2

In Simulation 2, the power curves should be viewed in context of the simulation design, where simulated data with higher sample size had increased mean. Simulated rejection rates from this scheme are presented in Fig. 4. We found out in Simulation 2, that empirical rejections rates were almost equal to the simulated power in Simulation 1 for all generated variants with sample size ratio $\kappa = 1$. This discovery nonetheless could be expected, since parameters of the simulation schemes 1 and 2 were in reality identical for this level of κ ratio.

Analogously to Simulation 1, the observed power was inadequate in all tests (< 0.8) for the smallest sample size combinations ($n = 30$). As a result, only random samples of the overall size $n > 30$ should be considered for applications of the researched two-sided tests with adequate power above 80%. Near identical finding of low power for the exact tests applied to small samples was reported in simulation studies of Stehlík and Wagner [11] and Střelec and Stehlík [12]. In addition, the current MC simulations confirmed that rejection rates increased monotonically with the mean-ratio parameter k, as anticipated by Casella [1]. The tests displayed acceptable power for majority of MC variants with $k > 2$. In the bootstrapped tests, increased simulated size of the tests was noticed in comparison to the nominal 5% observed mostly in the exact tests. The test size increase was primarily evident in the researched combinations with more unbalanced data.

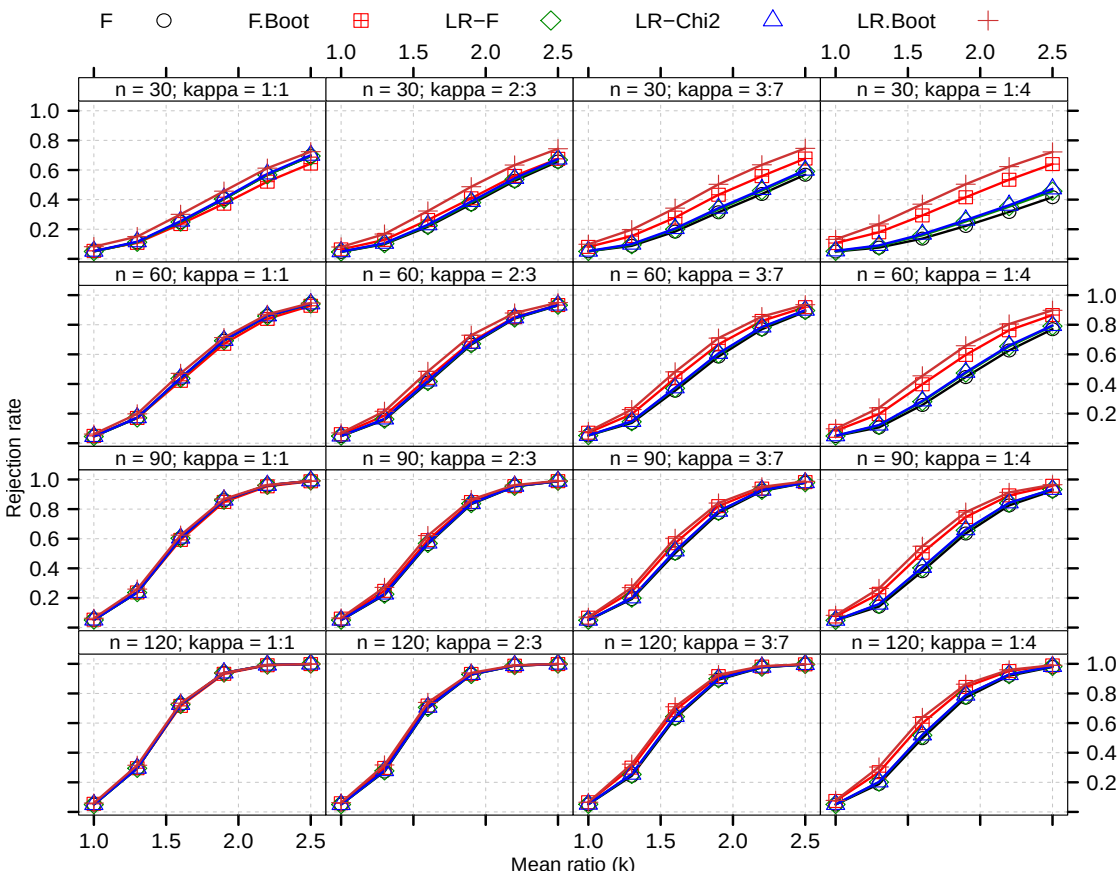


Figure 4 Rejection rates of exact and boostrapped statistical texts from Simulation 2 in response to sample size n, ratio of means k and sample size ratio κ.

For simulated variants with total sample size $n = 30$, the empirical rejection rates dropped with decreasing

levels of κ, implying greater differences in sample sizes n_1 and n_2 of the respective subsets. In general, the simulated rejections rates decreased primarily in the exact tests with lower κ, though a small drop in power was also seen in the bootstrapped tests. In terms of simulated rejection rates, the most promising tests were the bootstrapped LR test, bootstrapped F-test, LR χ^2-test, LR F-test and F-test, respectively, starting with the test possessing the largest power and ending with the test showing the lowest power. For simulated variants with large overall n, the differences in observed power withing the respective groups of the exact and bootstrapped tests were found unimportant. For variants with large n and more balanced samples, the differences in power among all tests dropped to zero.

Differences in power between the groups of exact and bootstrap tests, however, became more evident, as levels of κ ratio departed from unity. Increased rejection rates for the bootstrapped tests used on simulated data with $\kappa < 1$ and lower overall n could be attributed to the combination of data with higher mean occurring in samples of larger size. In the bootstrapped tests, this may have contributed to more precise simulation of the test statistic distribution under H_0, especially in the segments located close to the tails.

4 Conclusions

Simulated power curves via MC for three exact and two bootstrapped tests for equality of expected values of two exponential distributions were presented in this paper. The rejection rates increased monotonically with total sample size, more balanced data and ratio of the population means, as expected. Nonetheless, no test showed universal superiority over the remaining tests in terms of power. The exact and bootstrapped tests showed differential response in terms of power to increasingly unbalanced data: exact tests has greater power in Simulation 1, where data with higher population mean had lower sample size; on the contrary in Simulation 2, the bootstrapped tests showed greater rejection rates for unbalanced data variants, where data with higher population mean had higher sample size. The interaction of the bootstrapped and exact test variants with the size of the samples and means was thereby established.

For balanced samples and high sample size, the power performance of the tests was practically indistinguishable. With increasing sample size, the differences in power of the exact and bootstrapped tests approached zero in researched MC variants, as the simulated data became more unbalanced. The knowledge presented thereby could be valuable to the analyst when selecting a powerful statistical test for strongly unbalanced data with low or intermediate sample size.

Acknowledgements

The author wishes to express thanks to Dr. Luboš Střelec for careful reading of the manuscript and helpful comments.

References

[1] Casella, G. & Berger, R. L. (2002). *Statistical Inference*. Duxbury.

[2] Chernick, M. R. & LaBudde, R. A. (2011). *An Introduction to Bootstrap Methods with Applications to R*. New York: John Wiley & Sons, Inc.

[3] Chihara, L. & Hesteberg, T. (2011). *Mathematical Statistics with Resampling and R*. New York: John Wiley & Sons, Inc.

[4] Cox, D. R. (1953). Some Simple Approximate Tests for Poisson Variates. *Biometrika*, 40, 354–360.

[5] Efron, B. & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall, Inc.

[6] Han, G., Schell, M. J. & Kim, J. (2012). Comparing Two Exponential Distributions Using the Exact Likelihood Ratio Test. *Statistics in Biopharmaceutical Research*, 4, 348–356.

[7] Lee, E. T. & Wang, J. W. (1992). *Statistical Methods for Survival Data Analysis*. New York: John Wiley & Sons, Inc.

[8] R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Available from: http://www.r-project.org/.

[9] Ruppert, D. & Matteson, D. S. (2015). *Statistics and Data Analysis for Financial Engineering with R Examples*. New York: Springer.

[10] Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. New York: Springer.

[11] Stehlík, M. & Wagner, H. (2011). Exact Likelihood Ratio Testing for Homogeneity of the Exponential Distribution. *Communications in Statistics – Simulation and Computation*, 40, 663–684.

[12] Střelec, L. & Stehlík, M. (2012). Comparative Simulation Study of Likelihood Ratio Tests for Homogeneity of the Exponential Distribution. *Acta Univ. Agric. Silvic. Mendelianae Brun.*, 60, 307–314.

Optimal Transition Paths Toward Clean Technologies

Nastaran Ansari[1], Adriaan Van Zon[2], Olaf Sleijpen[3, 4]

Abstract. The climate change problem has encouraged many researchers to extend the Ramsey model to discuss the transition from carbon-intensive technologies to carbon-free technologies. The heterogeneity of capital stocks and the irreversibility of investment are intrinsic features of such a transition. However, in the current literature, few studies considered these features in the Ramsey model, and this problem is considered qualitatively with a focus on steady-state properties rather than on transitional dynamics. In this paper, we show that finding the optimal transition path to carbon-free technology in the Ramsey model with irreversible investment requires solving an optimal control problem with a changing set of constraints over time that defines a multi-stage optimal control problem. We develop a two-stage optimal control model which allows the quantitative analysis of transitional dynamics in the Ramsey model with two kinds of capital (clean and dirty) and irreversible investment.

Keywords: Ramsey model, optimal control theory, heterogeneity of capital stocks, irreversible investment, multi-stage optimal control

JEL Classification: P28, C610
AMS Classification: 49K04

1 Introduction

In the last two decades, many studies used the Ramsey model to discuss the interaction of economic growth, the environment, and technology. These studies are dealing with the transition from carbon-intensive technologies to carbon-free technologies [1]. Modelling such a transition entails the heterogeneity of capital stocks in that each kind of capital embodies a particular technology [1, 2]. The heterogeneity of capital stocks raises the need to distinguish between reversible and irreversible investment. The irreversibility of investment implies that once investment has taken place in a particular type of capital embodying a specific technology, that investment cannot be converted into another kind of capital embodying another production technology. Nor can it be used for consumption [3].

However, in the Ramsey models with environmental extensions, very little attention has been paid to the role of the irreversibility of investment. This lack of attention is rooted in the total absence of capital in the model as in [4] and [5] or the presence of generic capital as in [6] and [7]. Even without considering environmental aspects, in general, few studies discussed the irreversibility of investment in the Ramsey model. In this paper, we present a Ramsey model with heterogeneous capital stocks with irreversible investment to prepare a useful and extendable framework for further research.

Arrow and Kurtz [10] considered a Ramsey model with a homogeneous capital stock and no possibility to decumulate the capital stock directly and instantaneously to satisfy consumption. In addition, gross investment is non-negative hence irreversible. They based their discussion on two kinds of intervals: the free interval, during which the irreversibility constraint is not binding, and the blocked interval, during which the irreversibility constraint is binding and gross investment is zero. Hence, the Hamiltonian systems are different at each interval. Based on the Arrow and Kurtz [10] paper, if the initial amount of capital, k_0, is less than the amount of capital in the steady-state, k^*, then there is only one free interval and the transition path is similar to the reversible investment case, but if k^* is smaller than k_0 then first there is a blocked interval followed by a free interval.

Barro and Salai-i-Martin [3] consider a Ramsey model with two capital stocks, physical and human capital, with constant returns to the factor of production and irreversible investment. They show that the optimality conditions imply a fixed ratio of the two capital stocks. By means of a qualitative discussion of the optimal path and the steady-state, they show that when the optimal ratio of two capital stocks does not hold from

[1] Maastricht University, School of business and Economics, 6200 MD Maastricht, n.ansari@maastrichtuniversity.nl.
[2] Maastricht University, School of business and Economics, 6200 MD Maastricht, adriaan.vanzon@maastrichtuniversity.nl.
[3] Maastricht University, School of business and Economics, 6200 MD Maastricht, o.sleijpen@maastrichtuniversity.nl.
[4] De Nederlandsche Bank N.V., 1017 ZN Amsterdam, o.c.h.m.sleijpen@dnb.nl.

the beginning, there is a stage during which the capital-stock ratio adjusts to its optimal level through positive gross investment in the stock that is too small, and zero gross investment in the stock that is too large. In the next stage, both stocks grow in strict proportion thus maintaining their optimum ratio. Events during this stage are identical to the reversible case.

In a recent study, Rozenberg et al. [11] have used the Ramsey model to compare the effects of different climate policies on stranded assets in the transition from dirty to clean capital. Their model includes irreversible investment and a threshold for accumulated pollution. They propose that transition has two phases; in the first phase there is no investment in dirty capital, and in the second phase, there is an investment in both types of capital, while the second phase ends when the pollution stock attains the pollution threshold. They present a qualitative discussion of the problem based on optimal control theory, but they did not develop an analytical approach to finding the solution. Instead, they specified a discrete time version of their structural model and used the GAMS mathematical programming software as a black box solver.

With respect to the literature on the Ramsey model with irreversible investment, there is a lack of a systematic framework that allows analysing transitional dynamics and intertemporal utility. Because of the practical importance of irreversible investment for the timing of the carbon-transition, we consider a Ramsey model with two kinds of capital stock and irreversible investment in these two stocks, i.e. in dirty capital, which is carbon-intensive, and clean capital, which is carbon-free.

The structure of the problem in our study is similar to the one considered by Barro & Salai-i-Martin [3], where the optimal condition implies a fixed ratio of the two capital stocks. For quantitative analysis, the proposed method has to deal with the optimal timing of attaining the optimal ratio of two capital stocks, due to the different nature of the Hamiltonian problems before and after this point in time. From this perspective, we make a bridge between optimal timing problems and the irreversibility of investment in the context of the Ramsey model with heterogeneous capital stocks. To the best of our knowledge, no one has considered this problem in the Ramsey model before, but there are a number of related studies based on the AK setting such as [12], [8] and [9]. These studies consider the optimal timing of switches between two different kinds of technologies, and except [12], the studies mentioned above do not consider the irreversibility of investment.

2 Method

We consider a Ramsey model with two kinds of capital: dirty or polluting capital, k_p, which is carbon intensive; clean capital, k_c, which does not emit carbon. To keep the analysis as simple as possible, we use a Cobb-Douglas production function with constant returns to both type of capital:

$$y = A\, k_p^{1-\alpha c}\, k_c^{\alpha c}, \tag{1}$$

where y is total production. A represents total factor productivity, while αc represents the partial output elasticity of clean capital. Output can be used for consumption or investment in the two types of capital, so the budget constraint is:

$$y = c + I_c + I_p, \tag{2}$$

where c, I_c and I_p are consumption, gross investment in clean and dirty capital, respectively. The equations of motion for both capital stocks are given by:

$$\dot{k_c} = I_c - \delta . k_c, \tag{3}$$

$$\dot{k_p} = I_p - \delta . k_p, \tag{4}$$

where δ is the depreciation rate. The intertemporal utility function is given by:

$$U = \int_0^\infty u[c(t)].\, e^{-\rho t} \mathrm{d}t \tag{5}$$

where ρ is the rate of discount and t represents time. Barro and Salai-i-Martin [3] show that in the optimal path the ratio of two types of capital stock, i.e., ψ^* should be:

$$\psi^* = \frac{k_p}{k_c} = \frac{1-\alpha c}{\alpha c}. \tag{6}$$

This constraint (equation (6)) is required for the equality of the marginal productivity of the two types of capital, and it naturally arises in a setting with completely reversible investment in both types of capital. By substituting equation (6) in equation (1), we have:

$$y = \left(\frac{1-\alpha c}{\alpha c}\right)^{(1-\alpha c)} \cdot A \cdot k_p. \tag{7}$$

Equation (7) implies that in the case of a nonbinding irreversibility constraint, the production function effectively provides a linear AK setting. However, if the actual ratio, ψ, deviates from optimal one, ψ^*,and is bigger than ψ^* from the start, k_p is relatively abundant. Attaining ψ^* while investment is irreversible, implies positive gross investment only in k_c and zero gross investment in k_p causing k_p to decrease via depreciation [4]. What would the optimal transition path to ψ^* look like in this case? And when would ψ^* be achieved in order to maximize intertemporal utility, and what about the characteristic of the optimal path after attaining ψ^*?

In order to answer the above questions, we formulate a two-stage optimal control model. In the first stage ψ moves to ψ^* and in the second stage, while ψ remains constant at $\psi = \psi^* = \frac{1-\alpha c}{\alpha c}$, and, as equation (7) shows, we effectively have an AK setting. In the second stage, therefore, there are no transitional dynamics and k_p, k_c and consumption grow at a constant identical rates. Let us now assume that T signals the moment that the first stage ends and the second stage begins. We can write the discounted welfare function as follows:

$$U(c(t), T) = U_1(c(t), T) + U_2(c(t), T), \tag{8}$$

$$U_1(c(t), T) = \int_0^T u[c(t)].\, e^{-\rho t} \mathrm{d}t, \tag{9}$$

$$U_2(c(t), T) = \int_T^\infty u[c(t)].\, e^{-\rho t} \mathrm{d}t, \tag{10}$$

where

$$u(c(t)) = \log c(t). \tag{11}$$

To make our two-stage model, we first define the notion of the aggregate capital stocks k:[5]

$$k = k_p^{1-\alpha c} \cdot k_c^{\alpha c}. \tag{12}$$

So, total production is given by:

$$y = A \cdot k \tag{13}$$

This definition can represent the model in terms of ψ and aggregate k instead of k_p and k_c. In next sessions, at first, we consider the Hamiltonian and first order conditions in the first and second stage. Then we discuss the transversality conditions, which is required for total welfare maximization.

2.1 The second stage

In the second stage, ψ is constant and only k is changing. In that case the Hamiltonian and the first order conditions are given by[6]:

$$H^s = e^{-\rho\, t}.\log c^s + \lambda_k^s \cdot \dot{k^s}, \tag{14}$$

$$\frac{\partial H^s}{\partial c^s} = \frac{e^{-\rho \cdot t}}{c^s} - \frac{\lambda_k^s\, \psi^{*\,1-\alpha c}}{1+\psi^*} = 0 \Rightarrow\ c^s = e^{-\rho \cdot t}\frac{1+\psi^*}{\lambda_k^s\, \psi^{*\,1-\alpha c}}, \tag{15}$$

$$\frac{\partial H^s}{\partial k^s} = -\dot{\lambda_k^s} = \delta - \frac{A\, \psi^{*\,1-\alpha c}}{1+\psi^*}, \tag{16}$$

$$\frac{\partial H^s}{\partial \lambda_k^s} = \dot{k^s} = -k^s\delta - \frac{\mathrm{e}^{-t\rho}}{\lambda_k^s} + \frac{A\, k^s\, \psi^{*1-\alpha c}}{1+\psi^*}. \tag{17}$$

[5] For ease of exposition, we drop time-subscripts in the remainder of the paper.
[6] We add superscripts s to denote the second stage and f for the first stage.

2.2 The first stage

In the first stage we have two state variables, k^f and ψ.The Hamiltonian and the first order conditions are given by:

$$H^f = e^{-\rho\, t} * \log c^f + \lambda_k^f * \dot{k^f} + \lambda_\psi^f \dot{\psi}, \tag{18}$$

$$\frac{\partial H^f}{\partial c^f} = \frac{e^{-\rho \cdot t}}{c^f} - \frac{\psi^{\,1-\alpha c}\left(k^f\ \alpha c\, \lambda_k^f + \lambda_\psi^f \psi\right)}{k^f} = 0 \Rightarrow\ c^f = \frac{k^f e^{-\rho \cdot t}}{\psi^{\,1-\alpha c}\left(k^f\ \alpha c\, \lambda_k^f + \lambda_\psi^f \psi\right)}, \tag{19}$$

$$\frac{\partial H^f}{\partial k^f} = -\dot{\lambda_k^f} = \delta\, \lambda_k^f + \frac{\psi^{\,1-\alpha c}\left(A\, k^{f\,2}\ \alpha c\, \lambda_k^f + c^f\ \lambda_\psi^f \psi\right)}{(k^f)^2}, \tag{20}$$

$$\frac{\partial H^f}{\psi} = -\dot{\lambda_\psi^f} = \frac{(A\, k^f - c^f)\psi^{\,-\alpha c}(k^f(-1+\ \alpha c)\ \alpha c\, \lambda_k^f + (-2+\alpha c)\lambda_\psi^f \psi}{k^f}, \tag{21}$$

$$\frac{\partial H^f}{\lambda_k^f} = \dot{k^f} = -\delta\, k^f + A\, k^f\ \alpha c\, \psi^{\,1-\alpha c} - \frac{e^{-\rho \cdot t} \alpha c\, k^f}{k^f\ \alpha c\, \lambda_k^f + \lambda_\psi^f \psi}, \tag{22}$$

$$\frac{\partial H^f}{\lambda_\psi^f} = \dot{\psi} = A\, \psi^{\,2-\alpha c} - \frac{e^{-\rho \cdot t} \psi}{k^f\ \alpha c\, \lambda_k^f + \lambda_\psi^f \psi}. \tag{23}$$

2.3 Transversality conditions

The transversality conditions and the other optimality conditions in our two-stage model with infinite time horizon are as follows:

1. The standard infinite horizon transversality condition. In the second stage the standard infinite horizon transversality condition implies that the value of capital times its shadow price must approach zero as times goes to infinity in the second stage [3]. Therefore, we have:

$$\lim_{t\to\infty} k_t^s\ \cdot \lambda_{k,t}^s = 0, \tag{24}$$

where we have now added time-subscripts. The economic meaning of this condition is that the terminal stock at the end of planning horizon should have zero utility-value in present value terms. Note that the time path of λ_k^s is obtainable by integrating (16), as shown in (25). Using equation (25) and then by integrating (17) we can obtain k_t^s as shown in equation (26).

$$\lambda_{k,t}^s = e^{(t-T)(\delta - \frac{A\, \psi^{*\,1-\alpha c}}{1+\psi^*})} \cdot \lambda_{k,T}^s. \tag{25}$$

$$k_t^s = \frac{e^{-(t+T)\rho + (-t+T)(\delta - \frac{A\psi^{*\,1-\alpha c}}{1+\psi^*})} \cdot (-e^{t\rho} + e^{T\rho} + e^{(t+T)\rho}\ k_T^s \cdot\ \lambda_{k,T}^s \rho)}{\lambda_{k,T}^s \rho}. \tag{26}$$

Substituting equation (25) and (26) in (24), we find:

$$\lim_{t\to\infty} k_t^s\ \lambda_{k,t}^s = 0 \xRightarrow{\text{Equation (25) and (26)}} \lambda_{k,T}^s = \frac{\mathrm{e}^{-T\rho}}{k_T^s\ \rho}. \tag{27}$$

The other transversality conditions relate to the conditions that must be hold at the switching moment from the first stage to the second stage (see [13]). These transversality conditions and the necessary conditions for holding them are shown in equations (28)–(30).

2. Continuity condition: equations (28) and (29) show the required conditions for the optimality of the state variables at time T.

$$\lambda^f_{\psi,T} = \frac{\partial U_2(c^s,T)}{\partial \psi} \Rightarrow \lambda^f_{\psi,T}$$
$$= \frac{\mathrm{e}^{-T\rho}\psi^{*-1-\alpha c}(-1+\alpha c+\alpha c\,\psi^*)(-A\psi^*+\rho\psi^{*\alpha c}(1+\psi^*))}{\rho^2(1+\psi^*)^2} \xRightarrow{\psi^*=\frac{1-\alpha c}{\alpha c}} \lambda^f_{\psi,T} = 0^7. \tag{28}$$

$$\lambda^f_{k,T} = \frac{\partial U_2(c^s,T)}{\partial k} \Rightarrow \lambda^f_{k,T} = \frac{\mathrm{e}^{-T\,\rho}}{k^s_T\,\rho}. \tag{29}$$

3. The optimal length of the first stage follows from the requirement that $\frac{\partial U_1(c^f,T)}{\partial T} + \frac{\partial U_2(c^s,T)}{\partial T} = 0$, implying that $H^f_T + \frac{\partial U_2(c^s,T)}{\partial T} = 0$. This implies the following requirement:

$$\frac{\partial U_1(c^f,T)}{\partial T} + \frac{\partial U_2(c^s,T)}{\partial T} = 0 \Rightarrow H^f_T + \frac{\partial U_2(c^s,T)}{\partial T} = 0$$
$$\Rightarrow \frac{\mathrm{e}^{-T\rho}\psi^{*-\alpha c}(A\psi^*(-1+(1+\psi^*)(\alpha c+\mathrm{e}^{T\rho}\lambda\psi^*\rho\psi^*)) - \rho\psi^{*\alpha c}(1+\psi^*)(T\rho+\log\rho+\log(1+\psi^*)+\log(\frac{\mathrm{e}^{-T\rho}\alpha c}{\rho}+\lambda\psi^*\psi^*)))}{\rho(1+\psi^*)} \tag{30}$$
$$= 0 \xRightarrow{\psi^*=\frac{1-\alpha c}{\alpha c},\ \lambda^f_{\psi,T}=0} H^s_T + \frac{\partial U_2(c^s,T)}{\partial T} = 0.$$

This constraint is met for $\lambda^f_{\psi,T} = 0$ and $\psi^* = \frac{1-\alpha c}{\alpha c}$ as under point 2 above.

3 Numerical solution

The two systems of differential equations shown in equations (14)–(17) and (18)–(23) are nonlinear and cannot be solved analytically**.** However, it is possible to solve them numerically. Based on the information we have about co-state and state variables at the initial and terminal moment of the first stage, and based on the transversality conditions, it is possible to find a set of initial values to solve the differential equations. The initial values of the dirty and clean capital stocks are known, but the associated co-state variables are unknown. In the first stage the dirty capital stock, k_p, is depreciating at a fixed rate δ. So, we have:

$$k^f_{p,t} = e^{-\delta\cdot t}\cdot k_{p,0}. \tag{31}$$

Based on equations (12) and (31) we can obtain the aggregate capital stock at the end of the first stage, i.e. at $t = T$ where $\psi_T = \psi^*$, as follows:

$$k^f_t = k^f_{p,t}\psi^{-\alpha c} \xRightarrow{for\ t=T\ and\ \psi=\psi^*} k^f_t = k^s_T = e^{-\delta\cdot T}k_{p,0}\psi^{*-\alpha c}. \tag{32}$$

From equations (29) and (32) we can obtain $\lambda^f_{k,T}$ and from equation (28) we know $\lambda^f_{\psi,T}$ should be zero. Together with equations (6) and (32), and given the a set of initial values at time T, $(\lambda^f_{k,T}, \lambda^f_{\psi,T}, k^f_T, \psi^*)$, we have enough information to solve the differential equations pertaining to the first stage (equation (19)–(23)). The only remaining question is what is the value of T (the optimum length of the first stage) should be. However, for any given value of T, we can solve the differential equations system of the first stage backward in time, but there is only one T that will take us back to the given initial value of ψ at t=0, i.e. ψ_0. For any T, we define $\psi_e(T)$ as the 'estimated'[8] value of ψ_0, given our 'guess' T. The optimum value of T, i.e. T^* is then implicitly defined by $\psi_e(T) = \psi_0$. For the numerical simulations in Mathematica, shown in the next section, we used the following set of parameter values: $\alpha c = 0.4$ (implying $\psi^* = 1.5$), A=0.123, $\rho = 0.04$, $\delta = 0.05$, $k_p = 150$, $k_c = 10$, $\psi_0 = 15$. As shown in Figures 1 and 2, at $T^* = 18.773$ the error defined as $\psi_e(T) - \psi_0$ would be zero and total welfare would be maximized. Figure 3 and Figure 4 show the optimal time path for ψ and aggregate capital, *k,* respectively.

[7] This follows immediately from the fact that during the second stage we have that $\psi^* = \frac{1-\alpha c}{\alpha c}$. For this value of ψ, $\frac{\partial U2(c^*)}{\partial \psi_T} = 0$.

[8] Through backward integration, see also [3].

4 Conclusion

Our paper provides a structured analytical framework covering the Ramsey model with heterogeneous capital stocks and irreversible investment. As such it is a methodological contribution to the literature regarding the Ramsey model. We have used the proposed framework to discuss the optimal transition path from production using dirty capital to production using clean capital. We have shown that if the initial ratio of two types of capital is not equal to its optimum long-term value, the transition path has two stages with different Hamiltonian systems in both stages. We have devised a method that uses transversality conditions that can determine the optimal timing of the switch between both stages. Thus, we can analyse how parameter shocks or shocks in initial conditions would influence the transitional dynamics and welfare, both analytically to some degree and quantitatively. The method presented in this paper is a first attempt to introduce a Ramsey-type framework incorporating the heterogeneity of capital and the irreversibility of investment as well as environmental considerations. The next step(s) would be extending the method by relaxing the assumption of constant returns to the factor of production and adding a threshold for accumulated emission.

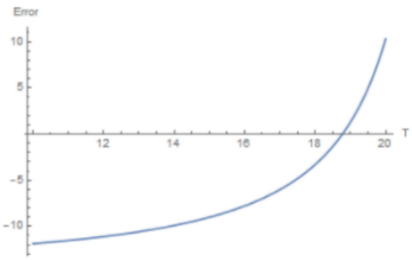


Figure 1 Error for different values of T

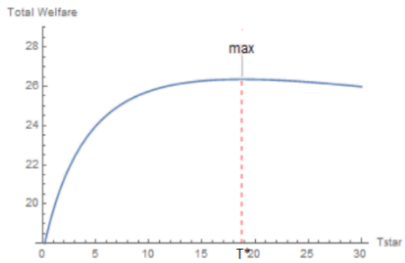


Figure 2 Total welfare for different values of T

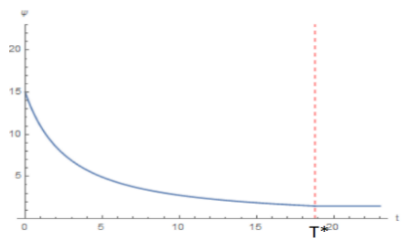


Figure 3 The ratio of dirty capital to clean capital, ψ, over time

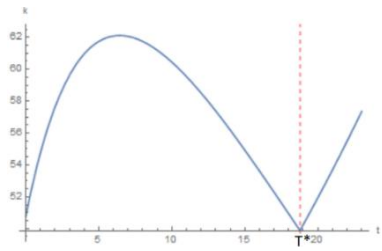


Figure 4 Aggregate capital, k, over time

5 References

[1] SMULDERS, S., TOMAN, M., & WITHAGEN, C. (2014) Growth theory and "green growth". *Oxford Review of Economic Policy*, 30(3), pp. 423–446. doi: 10.1093/oxrep/gru027.

[2] VAN ZON, A. (2015). *On the optimum timing of the global carbon-transition under conditions of extreme weather-related damages: further green paradoxical results*, *Unu-MERIT Working paper*. doi: 10.20955/r.85.67.

[3] BARRO, R. J. & SALA-I-MARTIN, X. (2004). *Economic Growth* (2nd ed.). The MIT Press.

[4] ACEMOGLU, D. ET AL. (2016). Transition to clean technology. *Political Economy*, 124(1), pp. 52–104.

[5] ACEMOGLU, D., AGHION, P., & HÉMOUS, D. (2014). The environment and directed technical change in a North-South model. *Oxford Review of Economic Policy*, 30(3), pp. 513–530.

[6] VAN DER PLOEG, F. & REZAI, A. (2016). Stranded Assets, the Social Cost of Carbon, and Directed Technical Change: Macroeconomic dynamics of optimal climate policy. *WU – Vienna University of Economics & Business,* Research Paper 176, 44(0).

[7] VAN DER PLOEG, F. & WITHAGEN, C. (2014). Growth, renewables, and the optimal carbon tax'. *International Economic Review*, 55(1), pp. 283–311.

[8] BOUCEKKINE, R., POMMERET, A., & PRIEUR, F. (2013). Optimal regime switching and threshold effects. *Journal of Economic Dynamics and Control*. Elsevier, 37(12), pp. 2979–2997.

[9] BOUCEKKINE, R., SAGLAM, C., & VALLÉE, T. (2004). Technology adoption under embodiment: A two-stage optimal control approach. *Macroeconomic Dynamics*, 8(2), pp. 250–271.
[10] ARROW, K. J. & KURTZ, K. (1970). Optimal Growth with Irreversible Investment in a Ramsey Model. *Econometrica*, 38(2), 331–344.
[11] ROZENBERG, J., ET AL. (2019). Instrument choice and stranded assets in the transition to clean capital, *Journal of Environmental Economics and Management*, https://doi.org/10.1016/j.jeem.2018.10.005.
[12] VAN ZON, A. & DAVID, P. (2012). Optimal multi-phase transition paths toward a global green economy. *Unu-MERIT Working paper*, (31), pp. 1–51.
[13] LEONARD, D. & VAN LONG, N. (1995). *Optimal control theory and static optimization in economics*. Cambridge university press.

Semantic Model of Project Management as Diagnostic tool in Project Stakeholder Management

Jan Bartoška[1], Tereza Jedlanová[2], Jan Rydval[3]

Abstract. The paper proposes the use of semantic networks and analytic network process (ANP) for quantification of the "soft" structure of a project in a corporate organization. The semantic project networks are based on the organization structure and on the life cycle of projects. Their subsequent quantification using the ANP creates the basis for the analysis of the project roles and analysis of their individual relationship to project documentation and stakeholders. Although the project roles or stakeholders during the management of projects are the keys to success, an approach to the quantification and analysis of their impact and influence on project organization structure has not been introduced yet. Semantic networks can be used to manage projects to illustrate and quantify links between internal and external objects of the project environment – project goals, project outputs, project documents, potential stakeholders, etc. The paper suggests a new approach for identifying and quantifying influence of project roles to project documentation and approach for estimation of power of potential stakeholders of project. The semantic model and its quantification can be used to construct a multi-criteria model for decision support in the project stakeholder management. The paper contains a case study of an application of the semantic model of project management in a commercial organization from the bank sector.

Keywords: analytic network process, limit supermatrix, project documentation, project management, project roles, project stakeholder management

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Project Management has several important parts, such as: project goals, outputs, scope statement, critical success factors, deliverables, schedule, budget, quality, human resources planning. However, communication and especially communication with all stakeholders in the project is the indispensable part. Often a project can occur in a complicated environment, then the management of communication with the stakeholder is even more important, because it means communicating with the right people in the right way at the right time.

According to the international project management standard PMBOK® Guide [10] and methodology PRINCE2 [1], stakeholder management is a process by which we identify and communicate with those people or groups who are interested in or influenced by the project outputs. The people or groups are actively involved in the project or their interests are influenced by the implementation of the project. These people or groups can influence the outputs and results of the project. Simply put, insufficient communication, especially with stakeholders, can have a negative effect on the project and it can bring the project into problems, or even cause project failure. On the other hand, the effective involvement of stakeholders will facilitate the day-to-day management of the project. Therefore, stakeholder management should never be underestimated, it is one of the most important points of project management with the aim to eliminate the negative impact on the project as much as possible states Bryson [6] and at the same time support the achievement of project goals. It is an activity that follows the analysis of stakeholders, and its integral part is communication with individual stakeholders.

To set up appropriate and effective communication with the stakeholders, it is necessary at first to define the project environment in a company. This can be conducted by creating a semantic model, as reported by Bartoška [2], and by El-Gohary, Osman, and El-Diraby [8], also according by Rydval, Bartoška and Brožová [11] semantic networks are suitable for displaying and expressing management structures and processes.

[1] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, bartoska@pef.czu.cz.
[2] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, jedlanova@pef.czu.cz.
[3] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, rydval@pef.czu.cz.

Using the semantic networks, the soft structure of a project in a corporate organization can be described. This is based on the Work Breakdown Structure, the oriented hierarchical decomposition of works and parts of the project management environment in the company, and the RASCI or RACI matrix which is mostly used for clarifying and defining roles and responsibilities in the company. Brennan [5] explains how to use the RACI matrix to obtain information about the key people for communication and how to split them into four groups: Responsible, Accountable, Consulted, and Informed.

After the project management environment is defined, it is necessary to conduct the analysis of the project roles and analysis of their individual relationship to project documentation and stakeholders. Then it will be possible to customize the communication plan to match the project environment and to ensure the success of the project. A properly designed semantic network can be converted into a network consisting of clusters and nodes for the analytic network process to determine the cardinal quantitative information about the main element of such a network. Saaty [13, 14] or Williams [15] describes how to conduct the analysis of particular network elements preferences using the analytic network process (ANP). Prioritization of the soft structure elements especially internal and external objects of the project environment – project goals, project outputs, project documents, potential stakeholders, etc., is then the key to successful project completion. Although the project roles or stakeholders during the management of projects are the keys to success, an approach to the quantification and analysis of their impact and influence on project organization structure has not been introduced yet.

The paper suggests a new approach for identifying and quantifying the influence of project roles on project documentation, and approach for estimation of the power of potential stakeholders on the project. The presented semantic model and its quantification using ANP can be used to construct a multi-criteria model for decision support in commercial practice in project stakeholder management. The paper contains a case study of the semantic model application of project management in a commercial organization from the bank sector.

2 Material and methods

2.1 Project Stakeholder Management

Project Stakeholder Management includes the processes required to identify the people, groups, or organizations that could impact or be impacted by the project [10]. Project Stakeholder Management analyze stakeholder expectations and their impact on the project, and to develop appropriate management strategies for effectively engaging stakeholders in project decisions and execution. The processes support the work of the project team to analyze stakeholder expectations, assess the degree to which they impact or are impacted by the project, and develop strategies to effectively engage stakeholders in support of project decisions and planning and execution of the work of the project. [10]

The management of a project's "stakeholders" means that the project is explicitly described in terms of the individuals and institutions who share a stake or an interest in the project. Thus, the project team members, subcontractors, suppliers, and customers are invariably relevant. The impact of project decisions must be considered in any rational approach to the management of a project. Management must also consider others who have an interest in the project and, by definition, are also stakeholders. These stakeholders are outside the authority of the project manager and often present serious management problems. [7]

2.2 Semantic model and Analytic network process (ANP)

The project management structure in the organization can be described in the form of a semantic network [2], both in the commercial and non-commercial spheres, as described in [3] or [9]. It is particularly useful to distinguish project roles, project documentation, project constraints, knowledge management areas, and interrelationships between elements or groups of elements, see Figure 1 as states [3].

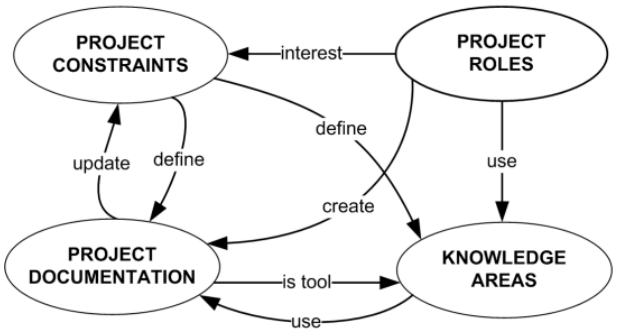


Figure 1 Semantic Model of Project Structure in the ANP model [3]

The project or project management semantic network defined in the organization can be simplified (without link interpretation) into the Analytic Network Process (ANP) model and then used to analyse the organization's environment, see Figure 2 as reported in [3].

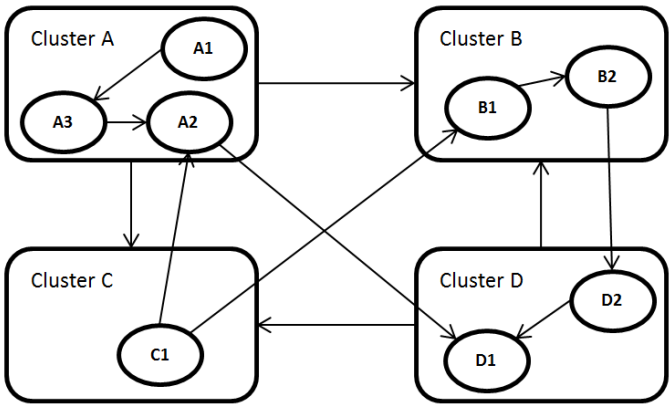


Figure 2 The structure of elements and clusters in ANP model [3]

The benefit of the ANP method is the ability to express different preferences of links between elements and clusters. To express preferences the method of pair comparison is used. Preferences always occur precisely in assessing the importance of the two elements in terms of the element which refers to them – there rises a question "which of the elements is more important, and by how much". The resulting values for the subclusters are then combined into a supermatrix where the normalization of columns is performed [13, 12]:

$$W = \begin{array}{c} \\ C_1 \\ C_2 \\ \vdots \\ C_N \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \dots & C_N \end{array} \\ \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \dots & W_{nn} \end{bmatrix} \end{array} \tag{1}$$

Where each block of the super matrix consists of:

$$W_{ij} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix} \tag{2}$$

Under condition:

$$\sum_{i}^{n} w_{ij} = 1, j \in \langle 1, n \rangle \tag{3}$$

For a weighted supermatrix (1) for which the relation (3) is valid, a calculation to obtain limit weights of the elements can be performed. The calculation is performed by squaring the weighted supermatrix to a sufficiently large number. Since the supermatrix has $N \times N$ size, the squaring is always feasible in a trivial manner (matrix multiplication). The result is the approximation of the weighted matrix to the limit matrix. Limit scales can be found in any column of the supermatrix. The limit weight of each element expresses the

strength of the effect on the overall structure of elements, i.e. it answers the question of how strongly an element affects the other elements [13, 12].

2.3 Case study: Organization project structure in bank sector

Currently, a multi-level project organizational structure has pushed through in the current project management practice in the corporate environment with the prevailing standard of "Guide to Project Management Body (PMBOK® Guide)" [10] and the methodology "Managing Successful Projects with PRINCE2" [1], which includes both project roles (PgM, PM), Project Leadership (PSC) and senior corporate units (PPM, PMO, BoD, EAB, PROB):

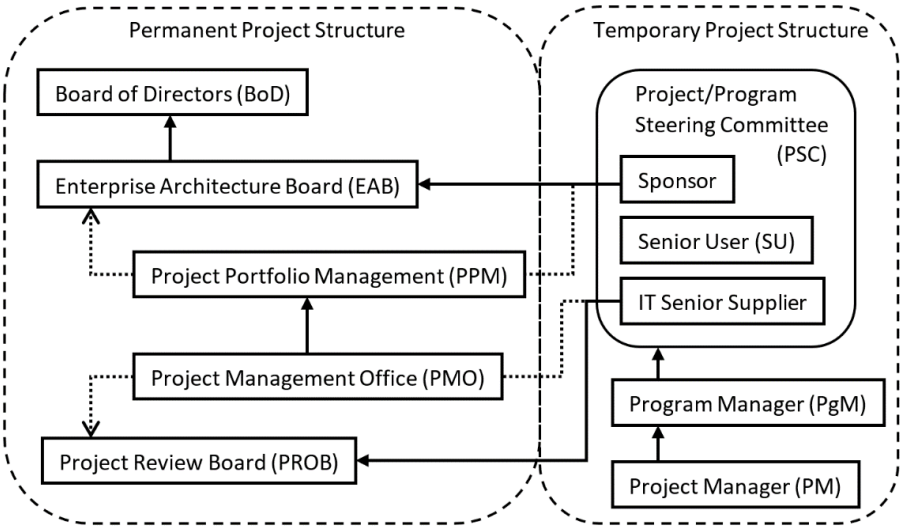


Figure 3 The organization project structure in bank sector [4]

Figure 3 shows the interaction between project roles (PgM, PM), the project's superior body (PSC), and corporate units (BoD, EAB, PPM, PMO, PROB – without any participation of the project roles). Project roles (PgM, PM) report on the progress and success of individual projects, while the superior PSC decides on changes and key life situations of individual projects (start, close, etc.). The project roles of PgM and PM and the parent PSC are temporary structures. The temporary project structure (Project roles) has full responsibility for the negotiation with Project stakeholders, whereas a part of the temporary project structure (PCS) includes also the internal project stakeholders (especially Sponsor, Senior User, and other senior managers in the bank corporation in the case).

The research in the chosen bank organization took place from 2016 to 2018. The organization is the international bank company with the extensive portfolio of the banking services for the corporate or the personal clients. The chosen bank company represents a typically corporate environment with developed project management. Within the research, a basic semantic model of project management was created as stated in [2] and further described and interpreted in [3] – the model includes a complete network of project roles, departments, project documentation, project restrictions, etc. The model was first quantified, i.e. the limit weights of the elements were determined, without project role preferences (Neutral model without preferences). This model was then repeatedly discussed with selected banking organization staff, with subjective preference recording, and always stored and quantified in the copy – thereby obtaining different limit weights for individual elements, incorporating individual respondent preferences. Two program managers (PgM 1, PgM 2) and four project managers (PM 1, PM 2, PM 3, PM 4) were selected among respondents. The semantic model quantification of the project structure organization can be performed by the ANP method. The advantage of this method is the possibility of bias preferences among elements. For example, with the help of the Saaty scale [13] or [14], the addressed project roles can differentiate their different attitude toward the project stakeholder. Creating the calculation of the ANP model can be performed e.g. in a software tool Super Decisions Software 2.1 (*http://www.superdecisions.com/*). By calculating Calculus Type, a supermatrix with limit weights can be obtained. The computed limit weights in the semantic model of project management can be used to evaluate the significance of the elements in the clusters of the project structure (project elements), i.e. for the importance of Project documentation and Project roles (clusters in Figure 1).

The product of the limit weights of the project elements can be called "*Neighbor-interaction*", i.e. a relationship where the elements of one category influence the elements of another category during the creation of the structure or during the using in the structure of management of the project. This interaction of the elements of project structure leads always to the specific output or the result during managing (especially Project documentation). The value of "*Neighbor-interaction*" should not lead to comparing the elements between themselves, but bring to assess the influence of the elements on the project structure and his managing. For computing of the value "*Neighbor-interaction*", it is possible to use the limit weights from supermatrix (results from the calculation of the ANP model). The value "Neighbor-interaction" is computed as the product of the limit weights of the project's elements (especially Project roles and Project documentation).

3 Results and Discussion

3.1 Neighbor-Interaction in Semantic model of Project Management

In the case study of the bank organization, it is possible to use the general model of categories (figure 1) to express Neighbor-interaction between the elements of the clusters. It was selected specific element of the project documentation: "*Deliveries/Products*". The chosen document represents the result of negotiation with project stakeholders for project delivery and product (notation of the project outputs with acceptances). The interactions between the project roles and the project documentation as follows (Table 1):

Project documentation	**Temporary project structure (Project roles)**			
	IT Delivery Manager	Project Manager	Senior User	Team Manager
Deliveries / Products	***accept***	***create /finalize***	***approve***	***responsible***

Table 1 The interactions between the project roles and the project documentation ("*Deliveries/Products*")

The active participation of the project stakeholders on development of the "*Deliveries/Products*" can be distinguished by their responsibility in the project and their role in the project-oriented organization (Table 1). Each type of stakeholder interaction (accept, create, finalize, approve, responsible) in the development of the "*Deliveries / Products*" can be expressed quantitatively by the value of Neighbor-interaction, i.e. by the product of the limit weights (clusters: Project documentation, Project roles). The values of Neighbor-interaction is possible to determine in the case study of the bank organization as follows (between the neutral model and the individual semantic models with preferences):

Limit weights	**Neutral model**	**PgM1**	**PgM2**	**PM1**	**PM2**	**PM3**	**PM4**
IT Delivery Manager	0.093359534	0.1370526	0.0591249	0.1386236	0.1679982	0.0933362	0.027358
Project Manager	0.85253138	0.793673	0.8803655	0.8019838	0.7678612	0.8524847	0.942425
Senior User	0.025381433	0.0589131	0.0544171	0.0195934	0.0405539	0.0254165	0.004083
Team Manager	0.028727653	0.0103613	0.0060925	0.0397992	0.0235867	0.0287626	0.026133

Table 2 The values of Neighbor-interaction between the elements (Project roles, Project documentation)

It is obvious (Table 2) that managers (PgM, PM) perceive expectations (interest) and influence (power) towards stakeholders differently – the managers' individual differences in preferences from the neutral model are as follows:

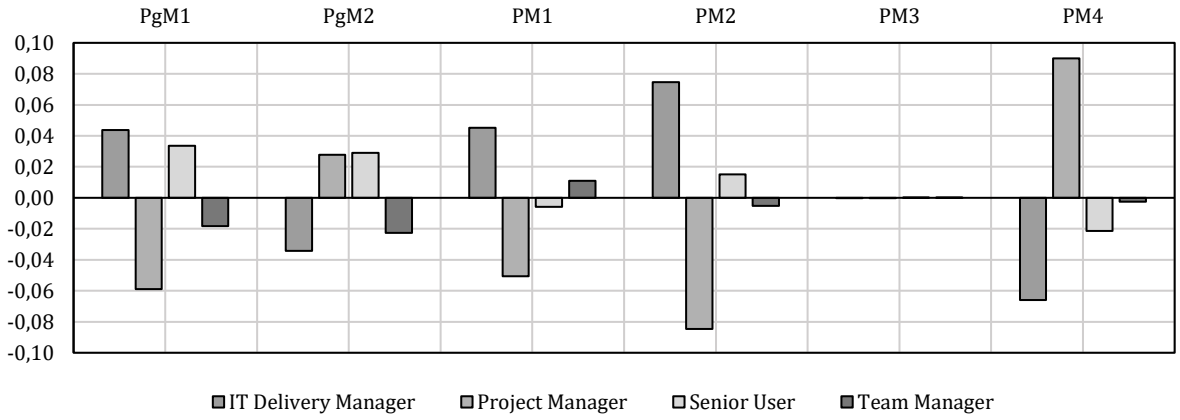


Figure 4 The differences between the neutral model and the models of individual preferences (PgM, PM)

The project document "*Deliveries/Products*" is based on the attendance of the project stakeholders differently. The biggest difference in perceiving of expectations is among project managers (especially between PM3 and PM4). Although the results of the research in the bank organization are rather illustrative, they demonstrate the applicability of the Semantic model of Project Management approach and applicability of the value of Neighbor-interaction.

The differences point out, how it would be possible to set the strategies in the Stakeholder Management of the project. Depending on the results achieved, it would be appropriate to inform the *Senior User* and *Team Manager*, while the *IT Delivery Manager* and *Project Manager* should become co-authors of the document "Deliveries / Products" (i.e. notation of the project outputs with acceptances).

4 Conclusion

The paper presents the use of semantic models of project management in the commercial sector, specifically in the banking sector. The presented results are derived from partial results of the authors' own research from 2016 to 2018. The results of the case study show that it is possible to quantify the individual attitude of the project roles to the potential stakeholders in the project organizational structure during the creating of the project documents (especially in case "*Deliveries/Products*").

The influence of stakeholders on the project is very significant and recognizing their strength is important to manage the projects. An equally important component of their interaction is the attitude of the program or project managers (PgM, PM) themselves when negotiating. The value of Neighbor-interaction can be used in the analysis of Project Stakeholder Management to determine the stakeholder power or interest.

5 Acknowledgements

This research is supported by the grant No. 2019A0015 "Ověření a rozvoj sémantického modelu řízení projektů" of the Internal Grant Agency of the University of Life Sciences Prague.

References

[1] AXELOS. (2017) *Managing Successful Projects with PRINCE2 2017th Edition*. London: The Stationery Office. ISBN 978-0113315338.

[2] BARTOŠKA, J. (2016). *A sematic model for the management of projects*. Habilitation work (thesis). DSE FEM CULS Prague.

[3] BARTOŠKA, J. (2016). Semantic Model of Organizational Project Structure. In: *Proceedings of the 34th International Conference Mathematical Methods in Economics*. Liberec: Technical University. ISBN 978-80-7494-296-9.

[4] BARTOŠKA, J., JEDLANOVÁ, T., & RYDVAL, J. (2019). Semantic Model of Project Management in Corporate practice. In: *Proceedings of the 37th International Conference Mathematical Methods in Economics*. České Budějovice: University of South Bohemia. ISBN 978-80-7394-760-6.

[5] BRENNAN, K. (2009). *A Guide to the Business Analysis Body of Knowledge (BABOK Guide)*. International Institute of Business Analysis. ISBN 0-9811292-1-8.

[6] BRYSON, J. M. (2004) What to do when Stakeholders matter, *Public Management Review*, 6(1), 21–53, DOI 10.1080/14719030410001675722.
[7] CLELAND, D. I. (1986). Project stakeholder management. *Project Management Journal*, 17(4), 36–44.
[8] EL-GOHARY, N. M., OSMAN, H., & EL-DIRABY, T. E. (2006) Stakeholder management for public private partnerships. *International Journal of Project Management*, *24*, 595–604.
[9] JEDLANOVÁ, T., BARTOŠKA, J., & VYSKOČILOVÁ, K. (2018). Semantic Model of Management in Student Projects. In: *Proceedings of the 36th International Conference Mathematical Methods in Economics*. Prague: MatfyzPress. ISBN 978-80-7378-372-3.
[10] PMI. (2017) *A Guide to the Project Management Body of Knowledge (PMBOK® Guide) — Sixth Edition and Agile Practice Guide*. Newtown Square (PA): Project Management Institute Inc. ISBN 978-1628253825.
[11] RYDVAL, J., BARTOŠKA, J., & BROŽOVÁ, H. (2014). Semantic Network in Information Processing for the Pork Market. *AGRIS on-line Papers in Economics and Informatics* 6, 59–67.
[12] SAATY, T. L. (1996). *Decision Making with Dependence and Feedback: The Analytic Network Process*. Pittsburgh: RWS Publications. ISBN 0-9620317-9-8.
[13] SAATY, T. L. (2001). *Decision Making with Dependence and Feedback: The Analytic Network Process: the Organization and Prioritization of Complexity*. Pittsburgh: RWS Publications. ISBN 978-0962031793.
[14] SAATY, T. L. & VARGAS, L. G. (2006) Decision Making with the Analytic Network Process: Economic, Political, Social and Technological Application with Benefits, Opportunities, Costs and Risks. New York: Springer. ISBN 978-0-387-33987-0.
[15] WILLIAMS, T. (2015) Identifying Success Factors in Construction Projects: A Case study. *Project Management Journal*, *47*, 97–112.

System Dynamic Conceptual Model for Landscape Fertility of Bees

Jan Bartoška[1], Tomáš Šubrt[2], Jan Rydval[3], Jan Kazda[4], Martina Stejskalová[5]

Abstract. The paper proposes the conceptual model and the partial mathematical model for displaying and capturing landscape fertility for bee breeding. The conceptual model is based on a causal loop diagram, while mathematical model is designed using analytical mathematical function. The authors' suggestions are based on long-term field research using hive weights as well as individual beekeeping observations. The proposals are based on causal relationships between beehive weight, humidity and temperature in the beehive, outdoor temperature, and et al. Landscape fertility for bees should be expressed as a number of colonies that can be effectively and sustainably fed for a given part of the landscape. It is conceived as a possible combination of weather, landscape, flowering and other variables. By designing a conceptual model, the authors define the dynamic behavior of real aspects of the environment (flowering, landscape type, fertility, etc.) and key aspects of bee swarm prosperity (honey clinker, vitality, activity of bees, etc.). The article will also outline the procedure how to estimate the landscape fertility in the long term, because due to frequent bee mortality, landscape fertility for bees has become one of the key topics in the field of nature conservation and agricultural production.

Keywords: system dynamic model; causal loop diagram; mathematical model; landscape fertility; bee breeding; honey stocks; temperature, weight and humidity of bee hive

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The concept of landscape fertility is not very widespread in the international academic and/or professional literature. Term is thus only vaguely described and similarly known as “trophy”; “fertility” or; “carrying capacity”. Models and landscape fertility functions are therefore not clearly defined in the literature and the topic can be encountered only in the form of isolated solutions. Original term “Landscape fertility” was mentioned by Smith [25] in connection with fertilization, biochemistry, urban management and the natural environment functions. In [12] is presented topics related to landscape fertility in the publication focused on the condition and chemical composition of the soil. However, commercial articles, some expert studies and, above all, beekeepers' discussion forums focus on the usefulness of the landscape fertility especially in Central European countries.

Due to the fact that this article focuses on the usability of the landscape fertility in relation to the behavior and performance of bees, it is also appropriate to mention a publication that compares these two aspects. In [5] is presented mathematical model for behavior of bee colonies monitoring. The research article focuses on a global problem known as the Colony Collapse Disorder (CCD) and assesses the possible causes (including e.g. pesticides, parasites, nutritional stress). Research paper explore and evaluates the effect of pollen on honey bee colony dynamics.

Anyway, it should be noted that this is a very important topic in the field of modern biology, which is closely related to biodiversity [6].

[1] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, bartoska@pef.czu.cz.
[2] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, subrt@pef.czu.cz.
[3] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, rydval@pef.czu.cz.
[4] Czech University of Life Sciences Prague, Department of Plant Protection, Kamýcká 129, Prague, kazda@af.czu.cz.
[5] Czech University of Life Sciences Prague, Department of Plant Protection, Kamýcká 129, Prague, volkovam@af.czu.cz.

1.1 Biodiversity as an indicator of landscape fertility

Biodiversity has been firstly defined by Bruce [7] and Wilcox [28]: "Biological diversity is the variety of life forms ... at all levels of biological systems (i.e., molecular, organismic, population, species and ecosystem)." Other definition was presented in 2004 by Gaston & Spicer's in their book [14]: "Biodiversity: an introduction" and says, that biodiversity is "variation of life at all levels of biological organization".

In [6] is states that the existence and life of bees has an effect on the existence and life of plants, just as the life of plants has an effect on the life of bees. Without bees, biodiversity would not reach the extent it reaches now.[6] Higher biodiversity means higher landscape fertility for bees. It is also necessary to mention that the importance of bees is absolutely essential for agriculture or fruit growing, where bees are an integral part of the whole system.

1.2 System dynamics models in living nature

In this section some models which are related to the behavior of bees are presented. In general, the use of system dynamics models in living nature is a quite new scientific discipline that develops especially due to computer technology since the 90s. Bruce [7] deals with the issue of modelling dynamic of biological systems and provides complete instructions for the application of models in this area. In the publication, Bruce [7] presents the use of biochemical models, genetic models and models of organisms, as well as population or catastrophic models.

As a result of the growing death of bee colonies, Stephen Russell [21] developed a dynamic model that seeks to identify the factors that have the greatest impact on the growth and survival of bee colonies. The model included an analysis of three years that covered significant fluctuations and simulated possible population growth / decline. In the conclusions the author states that bee colonies can be very sensitive to the composition of the food source or atypical fluctuations during changing seasons.

Other models of system dynamics that relate to the behavior or life of bees, were introduced by the following authors. Schmickland, T. a Crailsheim, K. [22] constructed a simple population model of bees using differential equations. Attached simulation scheme of stocks and flows shows a total of 5 flow quantities (number of young[7], number of adults, the amount of nectar, the amount of honey and the amount of pollen). Another approach was presented by David S. Khoury [18], who created a simulation model to determine the performance of the hive based on several reproduction scenarios. In [4] is proposed a dynamic model monitoring the production of food depending on the use of pesticides in crop fields.

2 Material and methods

2.1 Landscape fertility for bees and bee effect in landscape

Bee pollination of crops contributes to higher yields and thus to a better economic situation in the agricultural sector and the national economy [1]. The average number of hives per km^2 in 2019 was 10.2 hives, in 2018 was 9.42 hives, in 2017 was 9.39 hives and in 2016 was 9.69 hives in the Czech Republic. The implementation of the landscape in the Czech Republic is not uniform. Incorporation is from 0 to 149 beehives per km^2 [8]. The number of beehives in the landscape is limited by the amount of food that the bee colony can collect within its range.

Bees have a separate supply of energy food: nectar and pollen. Bees depend on the usability of the landscape both in terms of honey production and their health. Which also depends significantly on protein nutrition with pollen. The essentiality of amino acids for bees has been described by De Groot [16]. The positive effect of species diversity on pollen nutrition on bee immunity has been reported by [3]. Studies are describing the partial influence of individual characteristics of the landscape on its usability for bees. For example, studies report a positive correlation between the proportion of grasslands and deciduous forests in the range of flycatchers and the supply of pollen in the honeycomb [11] or the positive effect of catch crops and semi-natural habitats on workers' health in autumn [2].

In the Czech Republic the usefulness of the landscape for bees is diverse and currently highly variable – global climate change, extreme weather fluctuations, etc. The number of hives is increasing, the usefulness of the landscape for bees and other pollinators is declining. The Czech Republic has been struggling with

[6] Biodiversity is measured as the number of different species, plants and animals in a defined area. Biodiversity is highest in the tropical forest and lowest in the Arctic.
[7] This term includes the total number of eggs, larvae and pupae.

declining plant diversity since the 1980s. For example, the clover (the main food source for today's major pollinators) has lost its use in agriculture, and today is grown on just about 5% of arable land [9], [10] and [15]. On a summer day, one acre of the meadow can contain 3 million flowers that produce a kilogram of nectar sugar. That's enough to support 96,000 honey bees daily. In [17] is presented, that perennial meadows produced up to 20x more nectar and up to 6x more pollen than annual meadows, which in turn produced far more than amenity grassland controls. The average bee colony brings annually 200 kg of nectar and 30 kg of pollen. Another 20–150 kg of nectar is converted into 10–15 kg of honey, which can be collected by the beekeeper. The kilogram of honey is evidence of pollination of 3 million flowers and it contains up to 10 million pollen grains [27].

2.2 Internal research project *Včelstva Online* FEM CULS Prague

The web portal (https://vcelstva.czu.cz/) was design and started in 2017 at FEM CULS Prague as a volunteer project with the support of Czech commerzcial corporations (Česká spořitelna a.s., T-Mobile Czech Republic a.s., IBM Czech Republic, s.r.o.). In current time, the portal is a tool of citizen science within the internal research project on FEM CULS Prague.

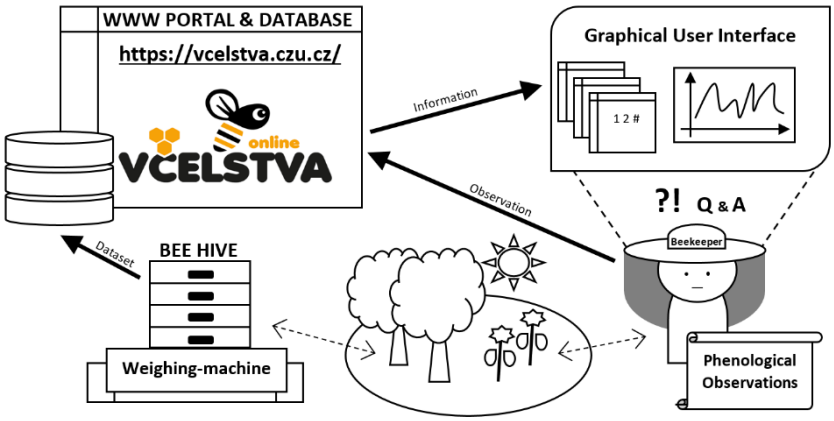


Figure 1 Schema of collecting records on the web portal Včelstva Online

The portal provides basic user functions for beekeepers: hive diary, records of locations with bee habitats, treatment records, etc. It also provides functions for beekeeping associations: evidence of beekeepers, treatments reports, etc. For citizen science offers functions (Figure 1): collecting phenological records, collecting data from bee hives (weight, inside temperature, outside temperature, humidity). The records from beehive weights and phenological records from beekeepers are collected together on the web portal – paired data are possible to use for monitor of the landscape fertility.

2.3 System Dynamics and Causal Loop Diagram

System Dynamics is a discipline which uses modelling and computer simulation to analyse, understand, and improve complex dynamic systems [13], [26] and [23]. The main idea is, that the system behaviour is determined mainly by its own structure, structure elements and by the interconnections between them [20], [23]. System dynamics methodology based on the feedback concepts of control theory [13], the principles of cognitive limitations, mental modelling [19] and bounded rationality [24] is an appropriate technique to handle complex systems to improve system thinking and system learning.

Symbol	Interpretation
	All else equal, if X increases (decreases), then Y increases (decreases) above (below) what it would have been.
	All else equal, if X increases (decreases), then Y decreases (increases) below (above) what it would have been.
	Delay mark
	Causal loop Reinforcing (+)

Symbol	Interpretation
B	Causal loop Balancing (–)

Table 1 Symbols of the causal loop diagram (based on [26])

To describe and define the system using system dynamics, a Causal Loop Diagram (CLD) is used firstly, and subsequently, a Stock-Flow Diagram (SFD) is created to enable mathematical modelling of the system. CLD is an important tool used for the definition and description of the complex system and it represents the structure of feedback systems. According to [26], CLD is very suitable for quickly capturing hypotheses about the causes of dynamics, capturing a mental model, and communicating the important feedbacks that seem to be responsible for a problem. CLD consists of variables which are connected by oriented causal links (arrows), these links represent the causal influences among the variables. The basic building blocks used in the CLD with icons and their interpretation are shown in Table 1. Each causal link is assigned a polarity, either positive (+) or negative (–) to show how the independent variable affects the dependent variable. Feedback loops are important for system behaviour. Feedbacks can be either positive ("+" sign or the letter R as Reinforcing) or negative ("–" sign or the letter B as Balancing). CLD does not enable mathematical modelling of the system.

3 Results and Discussion

3.1 System Dynamic Conceptual Model for Landscape Fertility of Bees

The CLD for landscape fertility for a bee hive during one season was compiled together with beekeeping experts using the basic building blocks of system dynamics (Table 1).

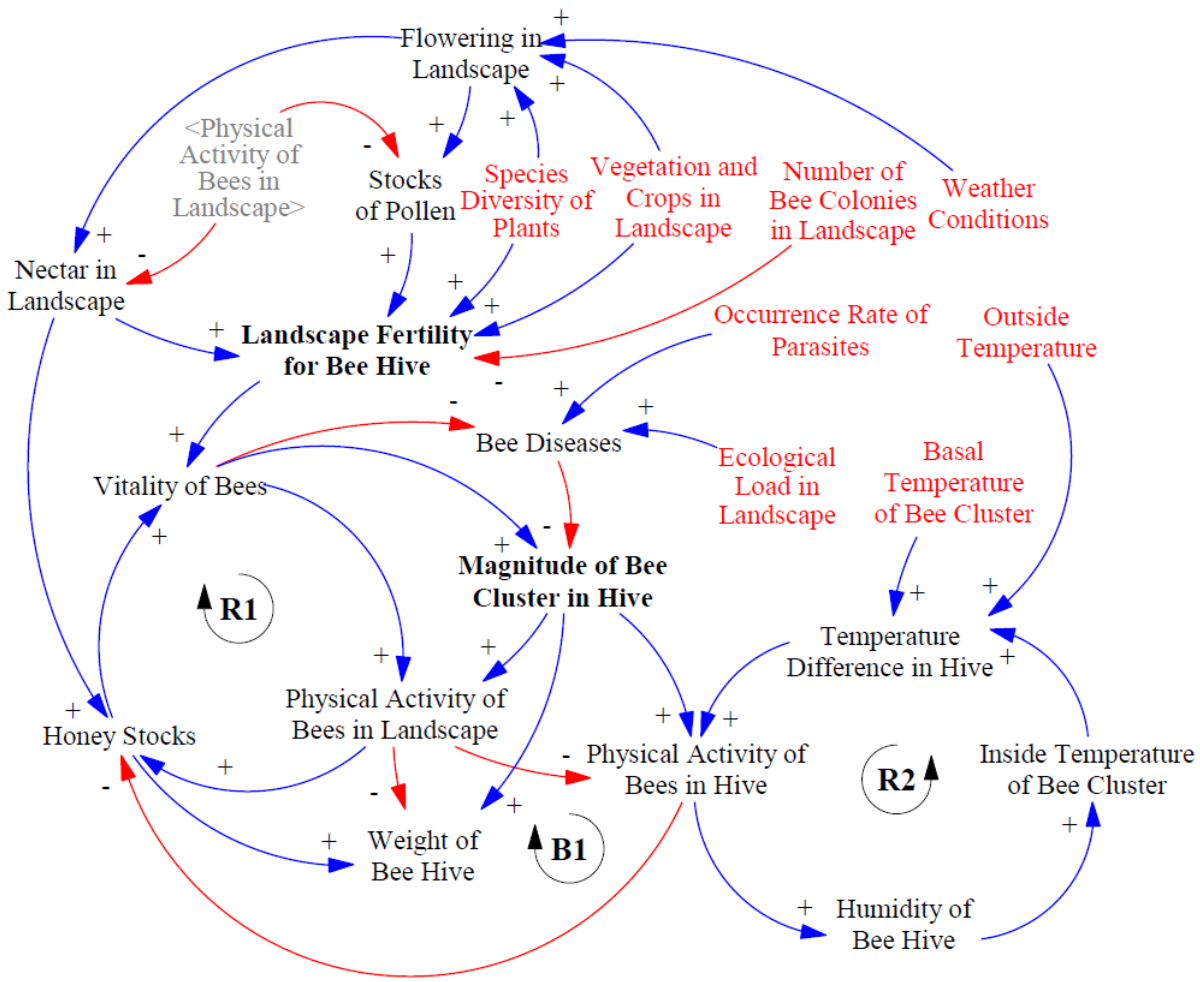


Figure 2 Causal Loop Diagram – Landscape Fertility for Bee Hive

The central point of this model (Figure 2) is the size of the bee cluster in the hive, which affects the amount of the physical activity of bees. The basic feedback loop (R1) describes the physical activity of the bees in the landscape. In simple terms, it can be assumed that with increased physical activity of bees in the landscape, the production of honey increases and therefore the honey stock increases as well. Honey stock contributes to keeping the high vitality of bees, which is necessary for the physical activity of bees. It can be seen from the diagram; this is a reinforcing feedback loop. However, the physical activity of bees is also inside the hive, where it can be caused, for example, by a high difference between the outside temperature and inside temperature of the hive. The high-temperature difference compared to the basal temperature of the bee cluster causes the physical activity of bees inside the hive (bees compensate too low temperature by consuming honey, too high temperature by increasing humidity of the hive). This physical activity of bees in hive subsequently affects the humidity and inside temperature of the hive (R2 – reinforcing effect). However, excessive physical activity of bees in the hive leads to the consumption of produced honey and that can lead to decreasing the vitality of bees due to honey deficiency (excessive honey consumption). And low bees' vitality can additionally reduce the size of the hive and thus the amount of physical activity of bees inside the hive (B1 – balance feedback). The model clearly shows that any physical activity of bees is influenced mainly by their vitality and, last but not least, the by honey stock. These two variables are determined by the landscape fertility for bees or nectar supply of the area respectively. The landscape fertility for bee hive is given by species diversity of plants, vegetation and corps in the landscape, stock of pollen, stock of nectar, and the number of bee colonies in the landscape. System Dynamic Conceptual Model for Landscape Fertility of Bees was designed on the ground of the results of research and practical knowledge (already mentioned above). The CLD (Figure 2) contains exogenous and endogenous variables that describe links and interactions among vegetation, crops and bee hive.

3.2 Mathematical model of Landscape Fertility for Bee Hive

The Landscape fertility for bee hive in our conceptual model is a key variable of the whole model and it can be designed using mathematical model with parametric logistic function (Figure 3):

$$LF = \frac{1 - SC^{-NP}}{BC + SC^{-NP}}$$

LF …	Landscape Fertility, $LF \in \langle 0; 1 \rangle$
SC …	Species Diversity of Plants (species per m^2)
	* Vegetation and Crops in Landscape (plants per m^2)
NP …	Nectar in Landscape (ten grams per m^2)
	* Stocks of Pollen (ten grams per m^2)
BC …	Number of Bee Colonies in Landscape (hives per km^2)

Figure 3 Mathematical model of Landscape Fertility

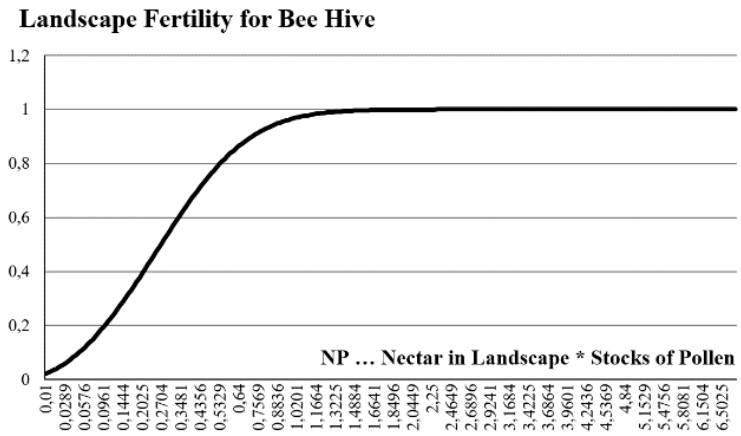


Figure 4 The course of the math. model of L.F.

The Landscape fertility (LF) for perimeter of 5 km^2 depends on exogenous variables Number of Bee Colonies in Landscape (BC), Species Diversity of Plants (first parameter in SC), Vegetation and Crops in Landscape (second parameter in SC), and endogenous variables Nectar in Landscape (first parameter in NP) and Stocks of Pollen (second parameter in NP). The Landscape fertility decreases by counts of bee colonies in Landscape (BC), increases by counts of plants and species in Landscape (SC), then mainly increases by abundance of nectar and pollen per m^2 (NP). In the next year of the research, authors plan to extend the model, perform to verify by the data collection and confront the results with praxis. The records from beehive weights and phenological records from beekeepers are now continuously collected and in next year will use for the extend and the verify of the model.

4 Conclusion

The paper suggests the first steps in the research of Landscape fertility for bees and its causal relations on various endogenous and exogenous factors. The conceptual model for Landscape fertility in the form of a Causal Loop Diagram was designed and the mathematical model for the key variable (Landscape fertility) was proposed. Suggested CLD and mathematical model based on logistic function are based on theoretical results of research and on practical experiences in the apiculture. Results of here published approach will serve practical use on the web portal *Včelstva Online*.

Although the concept of the Landscape fertility is not very widespread in the professional literature, the Landscape fertility for bees is nowadays a very popular topic in the agricultural praxis. Vitality and survival ability of bee cluster in Landscape decrease rapidly every year. Renewal of bee hives is very financially demanding, but agricultural production critically depends on bee pollination. The systematic monitoring of Landscape fertility for bees will ensure to the protection of bees and landscape as a whole too.

5 Acknowledgements

This research is supported by the grant No. 2019B0001 "*Monitoring a modelování trofické aktivity včel*" of the Internal Grant Agency of the Faculty of Economics and Management, University of Life Sciences Prague.

References

[1] AIZEN, M. A. & HARDER, L. D. (2009). The Global Stock of Domesticated Honey Bees Is Growing Slower Than Agricultural Demand for Pollination. *Current Biology*, *19*, 915–918.
[2] ALAUX, C., ALLIER, F., DECOURTYE, A., ODOUX, J. F., TAMIC, T., CHABIRAND, M., DELESTRA, E., DECUGIS, F., LE CONTE, Y. & HENRY, M. (2017). A 'Landscape physiology' approach for assessing bee health highlights the benefits of floral landscape enrichment and semi-natural habitats. *Scientific Reports*, *7*.
[3] ALAUX, C., DUCLOZ, F., CRAUSER, D. & LE CONTE, Y. (2010). Diet effects on honeybee immunecompetence. *Biology letters*, *6*.
[4] ATEHORTUA, F. L. & KRUIJTHOFF, K. (2014). Bees population decline as a threat to crops production. *32nd International Conference of the System Dynamics Society*, July 2014, Delft NL.
[5] BAGHERI, S. (2019). A mathematical model of honey bee colony dynamics to predict the effect of pollen on colony failure. *PLoS ONE*, 14.
[6] BRADBEAR, N. (2009). *Bees and their role in forest livelihoods*. Rome: FAO. ISBN: 978-92-5-106181-7.
[7] BRUCE, H. (2014). *Modeling Dynamic Biological Systems*. Springer International Publishing, Second Edition. ISBN 978-3-319-05614-2.
[8] COLOSS (2018). *Hustota zavčelení ČR 2015–2018*. [online] [cit. 2020-05-08] URL: https://colosscz.carto.com/builder/18c57198-e01d-4feb-8506-c3109913c972/embed.
[9] CZSO (2012). *Vývoj osevních ploch a první odhad sklizně*. [online] [cit. 2014-04-07] URL: http://www.czso.cz/csu/csu.nsf/1e01747a199f30f4c1256bd50038ab23/94912c113edf4287c1257a460031d437/$FILE/skl072612analyza.pdf.
[10] CZSO (2014). *Soupis ploch osevů 2014*. [online] [cit. 2014-07-23] URL: http://www.czso.cz/csu/2014edicniplan.nsf/p/270143-14.
[11] DONKERSLEY, P., RHODES, G., PICKUP, R., JONES, K., & WILSON, K. (2014). Honeybee nutrition is linked to landscape composition. *Ecology and Evolution*, *4*.
[12] DURHAM, R. E. (2001). *Principles of Home Landscape Fertilization*. University of Kentucky College of Agriculture.
[13] FORRESTER, J. W. (1968). Principles of system dynamics. MIT Press, Cambridge, MA.

[14] GASTON, K. J. & SPICER, J. I. (2004). *Biodiversity: An Introduction*. Second Edition. Oxford: Wiley-Blackwell Science. ISBN 9781405118576.
[15] GOULSON, D. (2010). *Bumblebees: Behaviour, Ecology, and Conservation (Oxford Biology)*. Second Edition. New York: Oxford University Press. ISBN 978-0199553075.
[16] GROOT DE, A. P. (1953). *Protein and amino acid requirements of the honeybee (Apis mellifica L.).* Academic theses. Den Haag: W. Junk.
[17] HICKS, D. M., OUVRARD, P., BALDOCK, K. C., BAUDE, M., GODDARD, M. A., KUNIN, W. E., MITSCHUNAS, N., MEMMOTT, J., MORSE, H., NIKOLITSI, M., OSGATHORPE, L. M., POTTS, S. G., ROBERTSON, K. M., SCOTT, A. V., SINCLAIR, F., WESTBURY, D. B., & STONE, G. N. (2016). Food for Pollinators: Quantifying the Nectar and Pollen Resources of Urban Flower Meadows. *PloS One*, *11*.
[18] KHOURY, S. D. (2014). Modelling Food and Population Dynamics in Honey Bee Colonies. *PLoS ONE*, 8.
[19] KIM, D. H. & SENGE, P. M. (1994). Putting systems thinking into practice, *Syst. Dyn. Rev*., 10 (2–3), 277–290.
[20] MEADOWS, D. H. (2008). *Thinking in Systems. A Primer*. Chelsea Green Publishing Company.
[21] RUSSELL, S. (2013). Dynamic modelling of honey bee (Apis mellifera) colony growthand failure. *Ecological Modelling*, 265, 158–169.
[22] SCHMICKLAND, T. & CRAILSHEIM, K. (2007). HoPoMo: A model of honeybee intracolonial population dynamics and resource management. *Ecological Modelling*, 204, 219–245.
[23] SENGE, P. M. (2006). Fifth Discipline: The Art and Practice of the Learning Organization. Random House Business Books.
[24] SIMON, H. A. (1979). Rational decision making in business organizations, *Am. Econ. Rev*. 69 (4), 493–513.
[25] SMITH, K. T. (2010). Biogeochemistry and landscape fertility. *The Landsculptor*, February 2010.
[26] STERMAN, J. D. (2000). Business Dynamics: System Thinking and Modelling for a Complex World. Irgwin/McGraw-Hill, Boston.
[27] ŠVAMBERK, V. (2017). *Základy chovu včelstev podle ročních období*. Kožlany: Spolek pro rozvoj včelařství MÁJA. ISBN 978-80-88045-04-5.
[28] WILCOX, B. A. (1984). In situ conservation of genetic resources: determinants of minimum area requirements. *Conference: National Parks, Conservation and Development, Proceedings of the World Congress on National Parks*. J. A. MCNEELY & K. R. MILLER, Smithsonian Institution Press. pp. 18–30.

Impact of Labour and Growth Shocks in Small Open Economy: DSGE Approach

Jakub Bechný[1]

Abstract. This paper aims to identify the impacts of the labour market shocks for economic growth, and the impact of the growth shocks for the labour market variables. We use a small open economy dynamic stochastic general equilibrium model with real wage rigidity and involuntary unemployment. The model is estimated by using a Bayesian approach and 16 observed variables from 2001Q1 to 2019Q4 for the Czech economy. The main results are as follows: (i) The labour supply shocks contributed to the economic growth and induced disinflationary pressure during 2010–2017 period, but are the barrier of growth from 2018 onwards; (ii) Productivity and domestic cost shocks are the dominant determinants of real wage growth; (iii) The effect of the domestic demand shock on the unemployment is twice as large as the foreign demand shock.

Keywords: Bayesian estimation, DSGE model, labour market, small open economy

JEL Classification: C32, E17
AMS Classification: 91B51

1 Introduction

The Czech labour market exhibits a remarkable development. Following the 2012–2013 recession that ended up with the exchange rate commitment of the Czech National Bank, the unemployment rate has steadily decreased from almost 8% to 2% and fluctuated close to this historical minimum since 2018. The labour market is tight, with the lowest unemployment rate in the whole European Union. According to the Czech Ministry of Labour and Social Affairs, there were 215 000 of involuntary unemployed and 340 000 of the unfilled job vacancies by the end of 2019.

Consequently, the goal of our short paper is: (i) To identify the impact of the labour market shocks for the growth of the Czech economy; (ii) To identify the impact of the growth shocks for the labour market variables. The analysis is done using a small open economy dynamic stochastic general equilibrium (DSGE) model with real wage rigidity and explicitly modelled involuntary unemployment. In particular, we employ the model that was in more detail presented in Bechný [4] and used for quantification of various output gap measures.

The research on the Czech labour market by using DSGE approach has been limited. The exceptions are Němec [11], Tonner et al. [13], and Železník [14]. We contribute to this literature by identification of the growth and labour market shocks within the balanced growth path model, and by quantification of the impacts of those shocks on the real economy. The remainder of this paper is organised as follows. Section 2 briefly describes our model and estimation procedure. Section 3 discusses the role of labour market shocks for the growth of the economy, Section 4 discusses the role of growth shocks for the labour market, and Section 5 concludes with the summary of the main findings.

2 Model and estimation

Our model combines a standard New Keynesian closed economy of Christiano et al. [7] with the small open economy structure of Adolfson et al. [1]. The model thus contains a standard set of nominal and real rigidities such as sticky prices, rigid wages, variable capital utilisation, investment adjustment costs or habit in consumption. The sticky prices of importers and exporters then allow for the incomplete exchange rate pass-through. Only domestic intermediate firms use capital and labour services in production. In total, there are five types of intermediate firms (and thus five types of New Keynesian Phillips curves): domestic, exporters, consumption good importers, investment good importers, and importers of re-exported goods. Aggregate consumption, investment and export goods are compiled as the constant elasticity of substitution

[1] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, Czech Republic, jakubbechny@mail.muni.cz

index of domestic production and imported goods. The model contains one type of representative household that attains utility from consumption, leisure, and real cash balances. The labour market with explicitly modelled involuntary unemployment and hiring costs then follows Blanchard & Galí [5] and combines the Nash-bargained wage with the real wage persistence of Hall [10]. The monetary policy in our model follows the inflation forecast targeting regime. The foreign economy represents the IS curve, Phillips curve, and monetary policy rule.[1]

Importantly, the model contains the balanced growth path steady-state and allows us to link the model variables with the observed growth rates of the data, without their pre-filtration that leads to loss of long-term frequency information as discussed by Andrle [2]. We use 16 observed variables from 2001Q1–2019Q4 for Bayesian estimation of the model, taken mostly from the Czech Statistical Office and Eurostat: the real output, consumption, investment, imports, exports, wages, CPI inflation, GDP deflator inflation, nominal interest rate (PRIBOR), inflation target, CZK/EUR exchange rate, unemployment rate, and job vacancies. The foreign economy is proxied by the euro area real output, CPI inflation, and nominal interest rate (EURIBOR). We use a standard Metropolis-Hastings algorithm and Kalman filter, as implemented in Matlab toolbox Dynare. In particular, we used two parallel chains for the Metropolis-Hastings, each chain with two million draws, of which 50% are dropped as a burn-in. We obtained an acceptance ratio around 30% and checked the convergence using the diagnostics of Brooks & Gelman [6].

Based on comments of an anonymous referee, we must also admit that our modelling approach does not allow for the permanent impacts of the transitory shocks, that is, for the hysteresis. This follows from the assumption of a unique balanced growth path steady-state, inherent to a vast majority of modern New Keynesian DSGE models. For an interesting analysis of a search and matching model that allows for the unemployment hysteresis see, e.g., Čížek [9]. Bechný [3] then provided some empirical evidence in favour of the unemployment hysteresis in the Czech economy.

3 Role of labour market for growth of economy

We will start with the analysis of the impacts of the labour market shocks on the growth of the economy. The model contains two labour market shocks – labour supply ζ_t^N and hiring cost ζ_t^x. Figure 1 shows that the Czech economy was hit by the sequence of a positive[2] labour supply shocks during the 2010Q1–2017Q4 period; the labour supply shocks have had a negative effect on the growth since 2018. The labour supply shocks were thus the major driving force which led to the decrease in the unemployment rate to its historically lowest levels, however, since 2018 the extremely tight labour market with symptoms of the labour shortage likely become the barrier of the further economic growth. The negative labour supply shocks were also identified for the 2008–2009 Great Recession period and around the year 2005 that was associated with the peak of structural unemployment after the 1997–1998 crisis. From the normal distribution point of view, the labour supply preference process ζ_t^N exhibited a remarkable development soon after the Great Recession and from the year 2014 onwards with the values of ζ_t^N outside the ± 1 standard deviation interval.

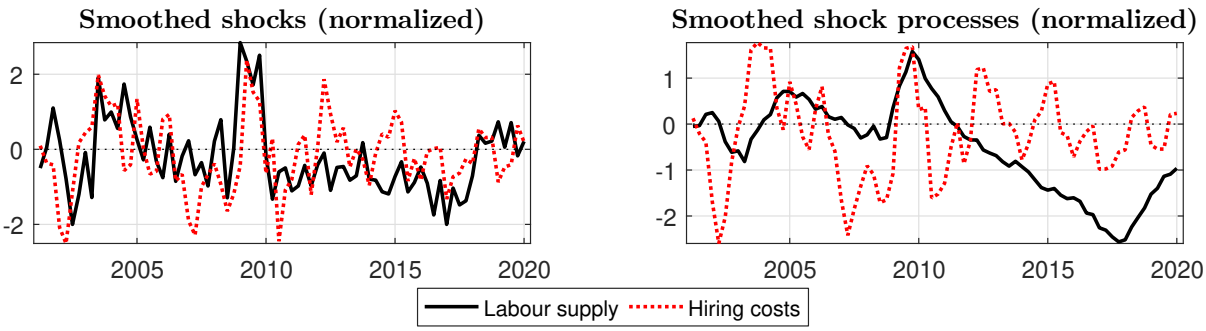


Figure 1 Smoothed labour market shocks (s.d.).
Notes: Smoothed shocks (innovations) and autoregressive processes are normalised by their standard deviations. Source: «Own computations»

[1] The log-linearised version of our model contains 84 equations. Therefore, it is above the scope of this short paper to present its structure in more detail here. For details, the readers are referred to Bechný [4].

[2] A positive labour supply shock leads to an increase in labour supply due to a decrease in households' disutility from work; that is, is associated with negative values of shock ζ_t^N.

Regarding the hiring cost shock process ζ_t^x, there are clear drops during 2002 and before the Great Recession, and peaks during the 2008–2009 Great Recession and around the year 2004. During the 2002–2004 period, the shock likely captures the ongoing structural changes in the Czech labour market. Dynamics of the hiring cost shock before and after the Great Recession likely capture a contribution of the domestic labour market to the overheating and subsequent slump of the economy. Importantly, there is no clear pattern for the period 2013 onwards. The hiring cost shock process ζ_t^x fluctuates within its ± 1 standard deviation ínterval, letting the labour supply shock ζ_t^N capture the recent phase of the labour market overheating.

Figure 2 compares the impacts of the one standard deviations labour market shocks for the growth of the Czech economy. Apparently, the hiring cost shock ζ_t^x has only a weak effect reducing the output growth Δy_t and the real wage growth Δw_t roughly by 0.2 p.p., a., and inducing a weak increase in inflation π_t^c of 0.2 p.p., a. The positive one standard deviation labour supply shock ζ_t^N has more significant effects on the growth. As can be seen in Figure 2, the increased labour supply due to the shock brings a drop in the unemployment rate $\widehat{U}_t$ of 0.15 p.p. and in the real wage growth Δw_t of 1.2 p.p., a. The increased labour supply induces increase in the growth rate of output Δy_t of 0.5 p.p., a., and reduction of the inflation π_t^c by 0.5 p.p., a.; the shock has a positive growth effect for all components of the domestic aggregate demand – the consumption Δc_t, investments Δi_t and exports Δex_t.

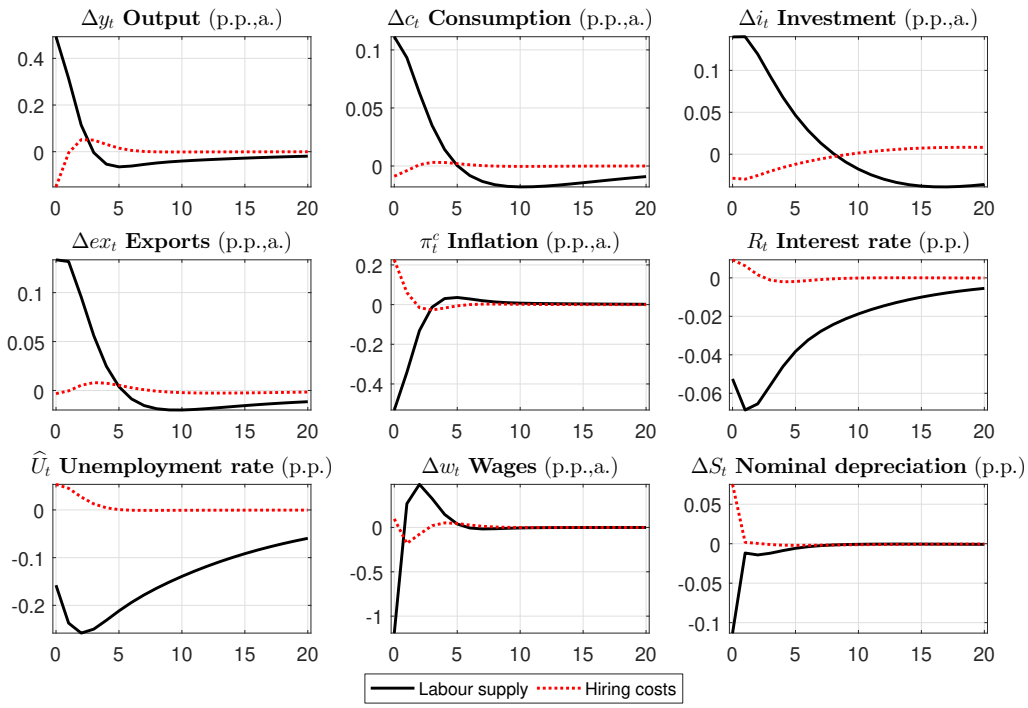


Figure 2 Impulse response functions to 1 s.d. labour market shocks

Notes: The variables are either in percentage points (p.p.), or in the percentage points and annualised (p.p.,a.). The unemployment rate is a deviation from the NAIRU. Source: «Own computations»

4 Impact of growth shocks on labour market

We will proceed with the analysis of the impacts of the key growth shocks on the labour market in the Czech economy. We will discuss in more detail impacts of the following shocks:

Permanent technology $\hat{\mu}_t^z$, due to its crucial role in the balanced growth path models in which this shock represents the only source of the long-term economic growth of all real variables, including the wages;

Foreign demand $\hat{\zeta}_t^{y^*}$ that captures the demand-driven dynamics of the output in the foreign economy block (that is, the euro area), and played an important role during the 2008–2009 Great Recession;

Consumption preference $\hat{\zeta}_t^c$ as the domestic demand shock that played an important role during the 2012–2013 recession;

Domestic markup $\hat{\lambda}_t^d$ as the domestic cost-push shock with quite a strong impact on the output growth and inflation.

Figure 3 shows the smoothed trajectories of the above-mentioned growth shocks and shock processes. We will again use the optic of the ± 1 standard deviation interval band to identify interesting realisations of the processes, respectively shocks.

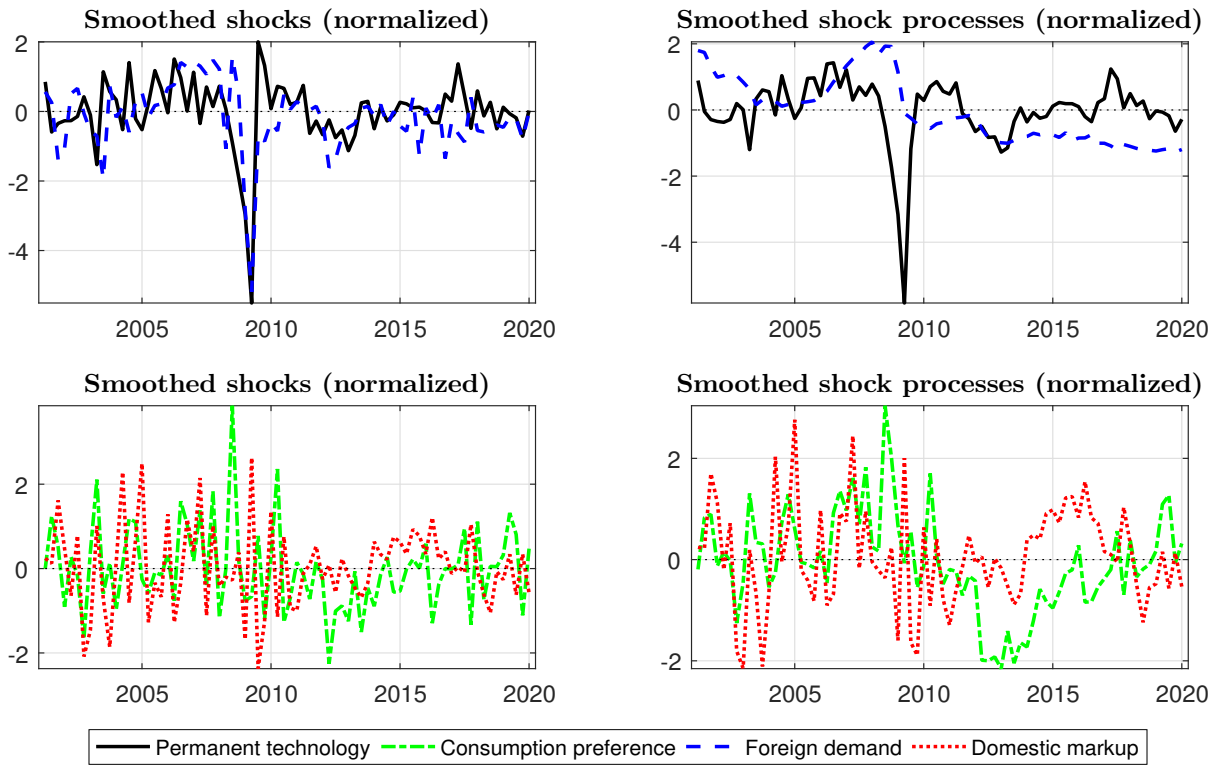


Figure 3 Smoothed growth shocks (s.d.)

Notes: Smoothed shocks (innovations) and autoregressive processes are normalised by their standard deviations. Source: «Own computations»

For most of the time, the permanent technology process $\hat{\mu}_t^z$ fluctuates within its ± 1 standard deviation interval band. It points to notable improvements in the technology during 2006–2007 and 2017. But most importantly, it captures the Great Recession through the shock of 5.5 standard deviations. The Great Recession was so severe that the model partially interprets it also as a drop in the potential output of the economy, due to its common-trend slump in almost all observed real variables (both domestic and foreign). This is not very intuitive (the recession was imported from abroad and relatively short-lived) and would require the incorporation of the expert judgment during the estimation. Regarding the current situation in the Czech economy, the permanent technology shock process $\hat{\mu}_t^z$ is identified as a negative from 2018 onwards – that is, recent growth is not driven by productivity improvement according to the model.

The foreign demand shock process $\hat{\zeta}_t^{y^*}$ shown in Figure 3 peaks just before the beginning of the Great Recession. The Great Recession itself is then associated with the negative foreign demand shock of 5 standard deviations. The model thus quite intuitively points to the crucial role of the development in the euro area before, during and after the Great Recession. The foreign demand shock process $\hat{\zeta}_t^{y^*}$ is persistently negative since 2013 onwards and quite intuitively captures the recent economic conditions in the euro area, associated with weak demand and deflationary tendencies.

The domestic demand shock played a crucial role during the 2012–2013 recession. As can be seen in Figure 3, the shock process $\hat{\zeta}_t^c$ falls to negative values of 2 standard deviations during the whole recession. It quite intuitively captures a weak domestic demand due to negative consumer sentiment and restrictive fiscal policy of the former government. The positive shocks before the Great Recession then capture the contribution of the domestic demand to the overheating, and significant positive shock in 2010Q1 is needed to explain the rise of the real consumption Δc_t during that period. Regarding the current situation in the Czech economy, the shock process $\hat{\zeta}_t^c$ peaked during the year 2019, capturing contribution of the domestic demand to the recent growth of the economy.

The process of the domestic cost-push shock $\hat{\lambda}_t^d$ in Figure 3 does not show any clear pattern until 2010. However, this shock is one of the key drivers of domestic inflation. The shock $\hat{\lambda}_t^d$ is persistently positive during the 2014–2017 period, capturing the inflationary domestic cost shocks. The impact of the cost shock $\hat{\lambda}_t^d$ is disinflationary from 2018 onwards.

Figure 4 compares the impacts of the selected one standard deviations growth shocks for the labour market variables of the Czech economy. We will start with the discussion of the effects of permanent technology shock $\hat{\mu}_t^z$. Similarly as in Sheen and Wang [12] the increased productivity drives up the output growth Δy_t and the real wages Δw_t that increase by 3.2 p.p., a. – the productivity is a dominant determinant of the real wage growth. The increase of the wages is partially offset by the drop in the hiring costs $\hat{g}_t$ (and thus also in labour market tightness $\hat{x}_t$, hiring rage ΔH_t, and vacancies ΔV_t). Domestic marginal costs $\widehat{mc}_t^d$ increase 0.2%, which induces the inflationary pressure of about the same magnitude. Similarly as in Christiano et al. [8], the productivity shock has no significant effect on unemployment.

As can be seen in Figure 4, the foreign and domestic (consumption) demand shocks have qualitatively but also quantitatively very similar impact on the labour market, although propagation of the consumption shock is a little bit stronger. Following the positive demand shock, the growth rate of output Δy_t and wages Δw_t increase. The unemployment $\hat{U}_t$ drops by 0.1 p.p. in response to the foreign shock $\hat{\zeta}_t^{y^*}$, and almost by 0.2 p.p. in response to the domestic demand shock $\hat{\zeta}_t^c$. In response to stronger demand the firms increase both posting of vacancies ΔV_t and hiring rage ΔH_t. The increase in the hiring and output is restricted by the rise in hiring costs $\widehat{g}_t$ caused by more tight labour market $\widehat{g}_t$, that is the result of the drop in unemployment. Both wages Δw_t and hiring costs $\widehat{g}_t$ also increase domestic marginal costs $\widehat{mc}_t^d$, that induce the inflation π_t^c of 0.5 p.p., a.

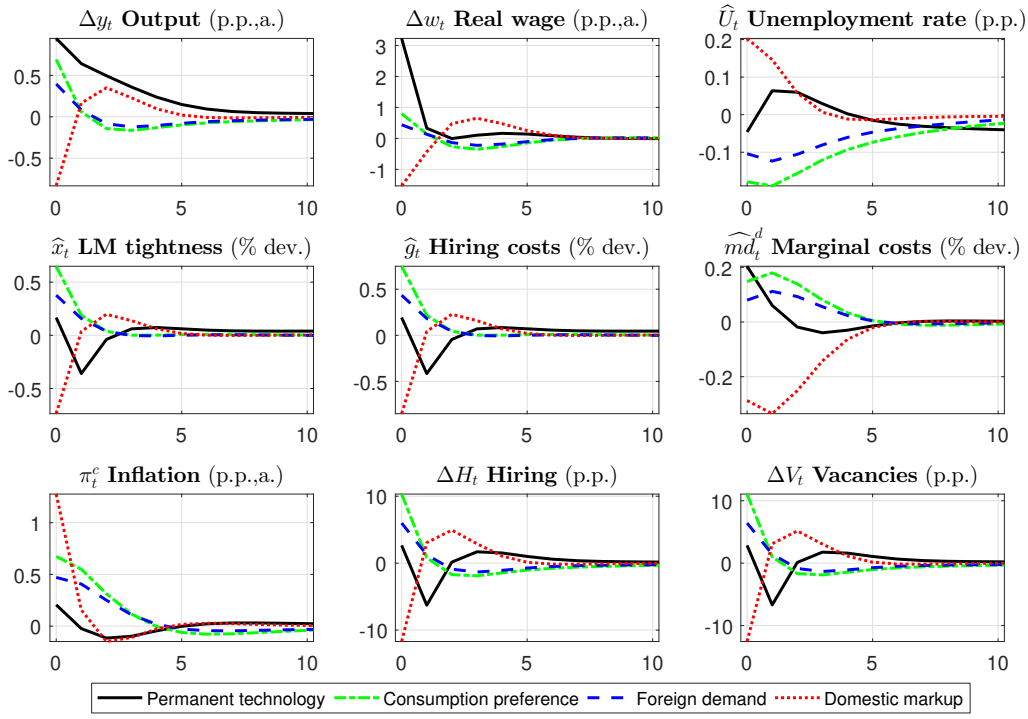


Figure 4 Impulse response functions to 1 s.d. growth shocks

Notes: The variables are either in percentage deviations (% dev.) from steady state, percentage points (p.p.), or in the percentage points and annualised (p.p.,a.). The unemployment rate is a deviation from the NAIRU. Source: «Own computations»

As can be also seen in Figure 4, the cost-push shock $\hat{\lambda}_t^d$ propagates with a remarkable magnitude. The cost shock increases the inflation π_t^c by 1.3 p.p., a. and shrinks the output growth Δy_t. The real wages Δw_t drop by 1.5 p.p., a., the unemployment $\hat{U}_t$ rises by 0.2 p.p. which reduces the labour market tightness $\hat{x}_t$, and thus the hiring costs $\hat{g}_t$. Nevertheless, the hiring rage ΔH_t and posting of vacancies ΔV_t drop by 12 p.p. due to reduced demand for labour. Note also that the inflation rises in response to the cost-push shock $\hat{\lambda}_t^d$, even

though the wages, hiring costs and thus also the domestic marginal costs $\widehat{mc}_t^d$ decrease – the similar result can also be found in Christiano et al. [8].

5 Summary and conclusions

The goal of this short paper was to identify the impact of the labour market shocks for the growth of the economy, and the impact of the growth shocks for the labour market variables. The analysis was done using a small open economy dynamic stochastic general equilibrium model with real wage rigidity and involuntary unemployment. The model was estimated using the Bayesian approach and data for the Czech economy from 2001Q1–2019Q4. We identified several remarkable realisations of the labour market and growth shocks and discussed the impacts of the shocks using the analysis of the impulse response functions.

We showed that positive one standard deviation labour supply shock reduces the unemployment rate by 0.15 p.p., increases the output growth by 0.5 p.p., a., and reduces the inflation rate by 0.5 p.p., a. The second labour market shock in our model, the hiring cost shock, has only a weak effect on the economy – the one standard deviation shock induces inflationary pressure of 0.2 p.p., a. The labour supply shocks contributed to the economic growth and induced disinflationary pressure during 2010–2017 period. On the other hand, the labour supply has been identified as the barrier of growth from 2018 onwards, and also during the Great Recession and year 2005.

We also showed that productivity and domestic cost shocks are the dominant determinants of real wage growth. Domestic cost shock also has a strong effect on the unemployment rate. The effect of the domestic demand on the unemployment is twice as large as of the foreign demand shock: the one standard deviation domestic consumption demand shock reduces the unemployment rate by 0.2 p.p., while the foreign demand shock only by 0.1 p.p. A significant productivity and foreign demand shocks were identified during the Great Recession. Adverse domestic demand shocks then played a crucial role during the 2012–2013 recession.And finally, our estimates show that the recent growth of the Czech economy is not driven by productivity improvement.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/1191/2019. This support is gratefully acknowledged.

References

[1] Adolfson, S., Laséen, J., Lindé, J. & Villani, M. (2007). Bayesian estimation of an open economy DSGE model with incomplete pass-through. *Journal of International Economics*, 72(2), 481–511.

[2] Andrle, M. (2008). *The Role of Trends and Detrending in DSGE Models*. MPRA Paper No. 13289.

[3] Bechný, J. (2019a). Unemployment Hysteresis in the Czech Republic. *Prague Economic Papers*, 28(5), 532–546.

[4] Bechný, J. (2019b). Output gap in the Czech economy: DSGE approach. *Review of Economic Perspectives*, 20(2), 137–156.

[5] Blanchard, O. J., J. & Galí, J. (2010). Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment. *American Economic Journal: Macroeconomics*, 2(2), 1–30.

[6] Brooks, S. P., Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4), 434–455.

[7] Christiano, L. J., Eichenbaum, M. & Evans, Ch. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1), 1–45.

[8] Christiano, L. J., Trabandt, M. & Walentin, K. (2011). Introducing financial frictions and unemployment into a small open economy model. *Journal of Economic Dynamics and Control*, 35(12), 1999–2041.

[9] Čížek, O. (2017). Unemployment and Aggregate Demand in Spain. *Journal of Economics*, 65(5), 459–481.

[10] Hall, R. E. (2005). Employment fluctuations with equilibrium wage stickiness. *American Economic Review*, 95(1), 50–65.

[11] Němec, D. (2013). Investigating Differences between the Czech and Slovak Labour Market Using a Small DSGE Model with Search and Matching Frictions. *Czech Economic Review*, 7(1), 21–42.

[12] Sheen, J. & Wang, B. Z. (2016). Assessing labor market frictions in a small open economy. *Journal of Macroeconomics*, 48, 231–251.

[13] Tonner, J., Tvrz, S., Vašíček, O. (2015). *Labour Market Modelling within a DSGE Approach*. Czech National Bank Working Paper Series, No. 2015/06.

[14] Železník, M. (2018). Labour Market in the Czech Republic: DSGE Approach. *Review of Economic Perspectives*, 18(3), 225–259.

Network Analysis of Intermediaries in Ecommerce

Ladislav Beranek[1], Radim Remes[2]

Abstract. E-commerce from the point of view of network science can be characterized as bipartite network. One part of such a network is made up of users and the other part is by electronic shops (e-shops). Transactions between these groups can be represented by user-to-e-shops edges. Edges can be weighted, when weight can express the size of the transaction, etc. Similarly, nodes can be assigned a similar parameter, node strength. It may mean reputation, the size of e-shop, ability to buy, etc. In this paper, we look at the situation where other entities enter such ecommerce bipartite network. These entities will occupy a central position between the aforementioned two parts of the originally bipartite network, affect relationships and change the topology of original bipartite network. Gaining a central position is beneficial due to the relative simplicity of the possibility to influence of other agents. Central position defines an information broker who accesses information and integrates it through social links. In this paper we study the properties of intermediaries and show its distribution characteristics in dependence of behavior of users and e-shops.

Keywords: ecommerce, evolving network, intermediary, price comparison web sites

JEL Classification: D85
AMS Classification: C63

1 Introduction

Online business systems on the Internet are similar to these ones in a physical environment where different interactions and relationships between customers and sellers also exist. However, in an online environment more than in a physical one, various intermediaries can influence these interactions. The result is that these systems have mostly complex arrangement and the different relationships between the elements, which affects their properties.

One example of network model is a bipartite network. It is widely applied in the modelling of various online platforms, such as online services where users view or purchase products [8, 10, 12], in biology [6], in medical science and in other areas [5, 13]. For example, Saavedra et al. have introduced two mechanisms (specialization and interaction) that lead to the exponential distribution of degrees for both parts of the bipartite network [11]. They found out that bipartite network can effectively characterize the structure of ecological and organizational networks. Further empirical analyzes of some models of bipartite networks have shown that user-level distributions truly follow a shifted power-law, while the distribution of an object's level always follows power law [15].

The network system is generally defined by patterns of different links between agents or various entities. Research on social networks has greatly advanced in understanding of how features such as links, paths, position, and other structural parameters determine the possibilities of a group of agents to influence other agents within the network [4]. In particular, taking a central position is generally considered to be beneficial because of the relative simplicity of control of the information distribution and a possibility of an influencing other agents. Like that, central position can form an information broker who accesses and integrates information through social links [4, 14].

In this paper, we deal with a situation where other entities enter into relationships modeled initially by the bipartite network (e.g., the already mentioned relationships in the online system customer (user) – seller (e-shop) in the B2C system) and will take a central position between the two parts of this bipartite network, will affect relationships and change network topology. An example of such an entity may be comparison-

[1] University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentska 13, 37005 Ceske Budejovice, Czech Republic, beranek@jcu.cz.
[2] University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentska 13, 37005 Ceske Budejovice, Czech Republic, inrem@jcu.cz.

shopping websites (hereinafter referred to only as price comparator). The general aim of this entering entity is to take the most significant position within the network and to have benefits from this position (see Figure 1).

Our contribution is based on our previous paper [1], which provides a detailed description of our model. In this paper, we describe the resulting network analysis of the behavior of intermediaries in e-commerce.

The rest of the paper is organized as follows. Basic principles of our model and corresponding mechanisms are described briefly in Chapter 2. In Chapter 3, we describe the experiments in short. In Chapter 4, we analyze the changes of nodes in strength over time and the strength distributions of the price comparison sites and objects. We will draw conclusions in Chapter 5.

2 Model

The aim of our model is to find out characteristics of a network originally bipartite after inserting new nodes (and edges) between both parties (users and e-shops) of the original bipartite network. These new nodes (price comparators) try to be an intermediary between the two parties. This means for example passing information about e-shops (for example on the prices of certain products in individual e-shops and the reliability of these e-shops) to users. We suppose that the evolution of network, the emergence of new edges is determined by the preferential connection mechanism. This assumption has its basis in the knowledge of user practice in online environment, when users select e-shops based on the popularity and ratings of other users.

Therefore, we assume that there are only users and objects at the beginning. Their relationship can be described by the bipartite graph $G(U, V, E)$, whose nodes can be divided into two disjunctive and independent sets U and V so that each edge joins a node u in U with one v in V. The set of edges is denoted as E, where $E \subseteq U \times V$, see Figure 1. In our concept, U denotes a set of users (buyers) and V denotes a set of objects (e-shops). At the beginning, this bipartite network is composed of u_0 nodes in the set U, v_0 of the node in the set V and e_0 edges in the set E. These initial sets can be called U_0, V_0 and E_0. Each edge between U and V is assigned an initial mass ω_0 ($\omega_0 > 0$). The mass between nodes i and j in time step t is denoted by $\omega_{ij}(t)$, we suppose that $\omega_{ij}(t) = \omega_{ji}(t)$.

For nodes, we use in our model a measure – strength of a node what is a generalization of a node degree (3aa). The strength of node i that is added in a time step t_i is denoted as $s_i(t)$ and is defined as $s_i(t) = \sum_{i \in \Upsilon(i)} \omega_{ij}$, where $\Upsilon(i)$ denotes a set of neighbors node i. If the node is also the starting node, $s_i(0) = k_i\omega_0$, where k_i is the number of neighbors. We assume that at the beginning $e_0\omega_0 = \sum_{i \in U_0} s_i(0) = \sum_{j \in V_0} s_j(0)$.

Now, a new entity (price comparator) r (R denotes a set of these entities) will enter these existing relationships. The original bipartite network will become tripartite with edges between sets U and V, U and R, R and V (see Figure 1).

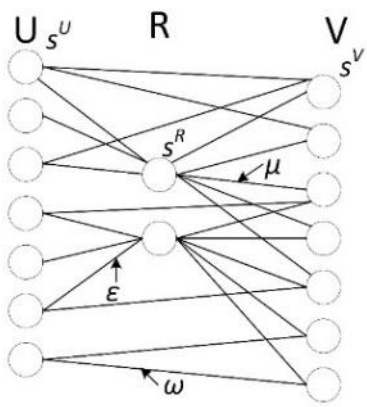


Figure 1 Marking of some model parameters

We will use the following marking (see Figure 1). The strength of a node u_i from the set U will be denoted as s_i^U, the strength of node v_i from the set V as s_i^V and the strength of node r_i from the set R as s_i^r. Furthermore,

we denote the weight of the edge between the node u_i of the set U and the node r_j of the set R as ε_{ij} and the weight of the edge between the node r_i of the set R and v_i of the set V as μ_{ij}.

The model is based on following mechanisms [1] with the probabilities p_1 to p_7 described within the text:

A) At time t, a certain number of already existing users will connect to the nodes already existing in V (e-shops) or in R (price comparators) using the following mechanisms (see also Figure 1):

i) to n_3 nodes already existing in V (e-shops) with the probability p_3 (n_3 new edges are created) based on preferential attachment:

$$\Pi(s_i^V) = \frac{s_i^V(t)}{\sum_{j\in V} s_j^V(t)} \tag{1}$$

This transaction will affect the weight ω_{ij} of the edge between the user and object (e-shop); we can express this change as following:

$$\omega_{ij}(t) = \omega_0 + \omega_{ij}(t-1) + \delta_3 \frac{\omega_{ij}(t-1)}{s_j^V(t-1)} + \delta_3 \frac{\omega_{ij}(t-1)}{s_i^U(t-1)}, \tag{2}$$

where ω_0 is an initial weight of the new edge.

(ii) users choose a price comparator with the probability p_4 based on preferential connection (n_4 edges are created) according to equation (1). This transaction will affect the edge weight between a user and a comparator by relationship:

$$\varepsilon_{ij}(t) = \varepsilon_0 + \delta_4 \frac{\varepsilon_{ij}(t-1)}{s_j^R(t-1)} + \delta_4 \frac{\varepsilon_{ij}(t-1)}{s_i^U(t-1)}. \tag{3}$$

where ε_0 is an initial weight of the new edge.

At the same time, selected price comparators will connect to m existing nodes in V (price comparator will show the possibilities of shopping in various e-shops to users) based on the preferential connection according to equation (1). This transaction will affect the weight of the edges (again, we only consider the local shifts of the weight), what we express by the equation:

$$\mu_{ij}(t) = \mu_0 + \delta \frac{\mu_{ij}(t-1)}{s_j^V(t-1)} + \delta \frac{\mu_{ij}(t-1)}{s_i^R(t-1)}. \tag{4}$$

B) The refusal of price comparator service. Large objects (e-shops) can have their own marketing activities. If they feel they are strong enough, they can reject with the probability of p_6 services of price comparator. The probability that refusal will realize the node i with the greatest force s_i^V can be expressed by equation (1). The strength of this object i is reduced by the weight value of the edge μ_{ij}, which disappears. We suppose that it also will influence other objects:

$$s_i^V(t) = s_i^V(t-1) - \mu_{ij}(t-1) - \delta \frac{\mu_{ij}(t-1)}{s_j^V(t-1)}. \tag{5}$$

This refusal will have an impact on price comparators whose service the object (e-shop) used. At the same time, the strength of this object will be reduced by a weight of the respective edge to the price comparator j that no longer exists. This reduction also affects the surroundings of the comparator j, thus reducing the weight of the node edges i, for which $i \in \Upsilon(j)$, by the relation:

$$s_j^R(t) = s_j^R(t-1) - \mu_{ij}(t-1) - \sum_{i\in\Upsilon(j)} \delta \frac{\mu_{ij}(t-1}{s_i^V(t-1)}. \tag{6}$$

C) Decreasing of the strength of a price comparator j with the probability p_7 using the anti-preferential probability:

$$\mathrm{P}(s_i^R) = \frac{1 - \frac{s_i^R(t)}{\sum_{j\in R} s_j^R(t)}}{\mu_0 + t \cdot p_7 - 1}. \tag{7}$$

If we select the node i, then the weight of the edge μ_{ij} is reduced for all nodes j with which i has a common edge, i.e. for $j \in \Upsilon(i)$, by the relation:

$$s_i^R(t) = (1 - \gamma_2)s_i^R(t-1). \tag{8}$$

The topology properties of the resulting tripartite developing network can be derived from strength distributions. The node strengths varying at time t are intermediate variables to obtain strength probability distributions for objects and comparators.

3 Model verification – data

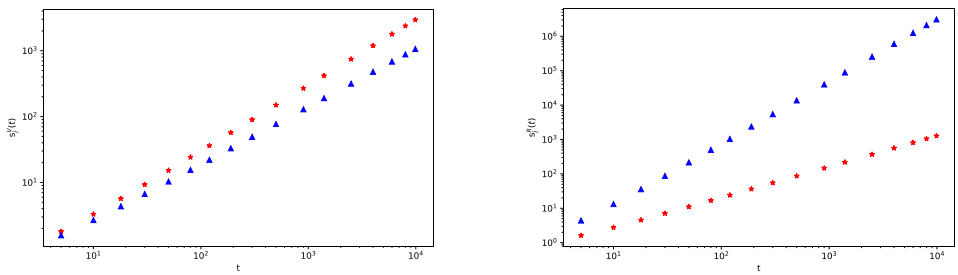
We used student participants to investigate user behavior when using price comparators. A total of 160 people who have had online shopping experience in the previous three months participated in this study. Overall, the participants were young (age average 21.2 years) with a good online shopping history. Participants received questionnaires and after two months they should indicate when they used a price comparator (most often they were price comparators Heureka.cz and Zbozi.cz) and when they went directly to the e-shop website, their shopping preferences, whether they notice the number of purchases listed in the e-shop information on the price comparator and other data. To find out the data concerning e-shops, we addressed 32 e-shops that used the services of price comparators. The questions we asked were about the number of user accesses, directly or through the price comparator, the consequences they would experience by refusing the price comparator service, how a position in the price comparator affects their performance and more. We have estimated some values based on publicly available statistics (https://www.czso.cz/, https://www.ceska-ecommerce.cz), literature (e.g., [9]) and our previous research [2, 3, 7]. This basic data was collected systematically between 1 March 2019 and the end of May 2019.

We processed the data and determined the basic values of the necessary parameters. For example, one of the conclusions stated that almost 50% of respondents regularly used price comparators for their purchases, 33% used them irregularly. Altogether 83% of respondents used price comparators because they helped them to navigate the extensive offers of e-shops. From these data we estimated p_3 and p_4. We proceeded similarly when estimating the values of other parameters. The δ_i parameters mean how the respective actors on the site interact. For example, if a new e-shop is created it also connects to the price comparator. However, this will also affect existing e-shops, must somehow respond, adjust the price of the product, shipping or otherwise respond. We therefore estimated these parameters based on data obtained from e-shops, their responses to their reactions to the emergence of new e-shops and the like.

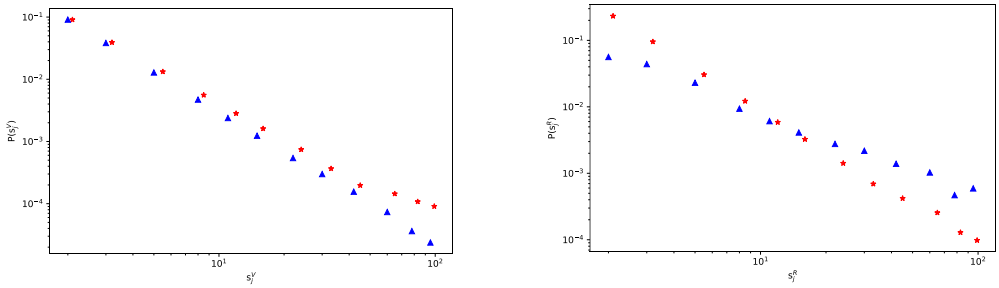
4 Results and discussion

We performed many simulations to analyze how different behaviors of users, e-shops and price comparators will affect different network characteristics. In this chapter, we describe the influence of the way the user gets into the e-shop, and the influence when e-shops reject the services of the price comparators.

4.1 Influence of the way the user gets to the e-shop on distribution characteristics



Figures 2a, 2b The log–log plot of the change of strength $s_j^V(t)$ (Fig. 2a, left) and log–log plot of the change of the strength $s_j^R(t)$ (Fig. 2b, right) with the time t = 2s for different values of p_3 and p_4 (other parameters are identical). Red line is for p_3= 0.63 and p_4= 0.309, blue line is for p_3 = 0.309 and p_4 = 0.63.



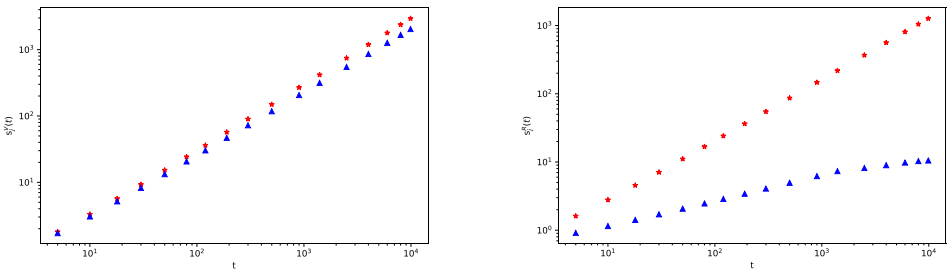
Figures 3a, 3b The log–log plot of objects (e-shops) strength distribution $P(s_j^V)$ changes over objects strength s_j^V (Fig. 3a, left) and the log–log plot of price comparator strength distribution $P(s_j^R)$ changes over price comparator strength s_j^R (Fig. 3b, right) for different values p2 and p4 (other parameters are identical). Red mark (∗) is for values p_3= 0.63, p_4= 0.309, blue mark (△) is for values p_3 = 0.309, p_4 = 0.63.

Figures 2 and 3 show the effect of the way users prefer to come to e-shops on nodes characteristics. Figure 2a shows how the strength of an object that was added at time t_j = 2 changes over time. Figure 2b shows similarly how the power of the price comparator, which was also added at time t_j = 2, changes over time. A red line indicates a greater probability a user will come to the object (e-shop) directly. Although the probability values are quite different in both cases, the impact of the way users come to some e-shop on the strength of this e-shop and its distribution function is not very large. This fact is shown in Figures 2a and 3a. Price comparators make it easier for users to search the Internet and compare product prices and their attributes in the real life. The result of our analyses and simulations show however that their functioning does not significantly affect the strength of individual objects. There are other factors that influence the strength of objects (marketing campaigns, influencers' marketing, recommendation and other). After all, the customer always will buy a product in some e-shop. A comparator can “only” lead on a customer to a specific e-shop.

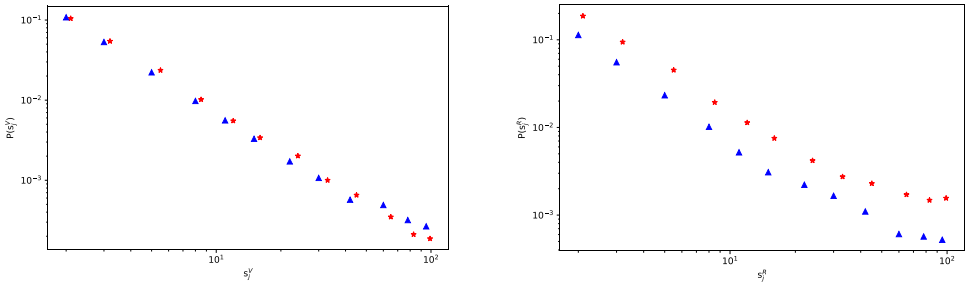
Figures 3a and 3b show the influence of the way in which users come to e-shops on the strength distribution characteristics of objects (Figure 3a) and the strength distribution characteristics of price comparators (Figure 2b). with a different probability of preferring the path to e-shops directly (red color) or through price comparators (blue color). From the figures, it is obvious that with a higher probability of using the comparators to access objects (e-shops), the power of the comparators is growing rapidly (Fig. 3b).

Of interest is especially Figure 3b, which shows the results of simulations of the distribution characteristic of comparators. The power of price comparators is rapidly increasing from a certain size. Price comparators added to the system later have little chance of gaining more power under given model conditions. This is given by the principle of preferential connection, especially for comparators, “richer becomes richer”.

4.2 The effect of the canceling of the use of a price comparator



Figures 4a, 4b The log–log plot of the change of strength $s_j^V(t)$ (Fig. 4a, left) and log–log plot of the change of the strength $s_j^R(t)$ (Fig. 4b, right) with the time t for different values of p6 (probability that the object (e-shop) will refuse to use the service of a comparator). Red mark (*) line is for value p_6= 0, blue mark (△) is for value p_6 = 0.01 (other parameters are identical).



Figures 5a, 5b The log–log plot of objects strength distribution $P(s_j^V)$ changes over objects strength s_j^V (Fig. 5a, left) and the log–log plot of comparator strength distribution $P(s_j^R)$ changes over comparator strength s_j^R (Fig. 5b, right) for different values p6 (probability that the object (e-shop) will refuse to use the service of a comparator). Red mark (*) line is for value p_6= 0, blue mark (△) is for value p_6 = 0.01 (other parameters are identical).

Figures 4 and 5 show the effect when an e-shop refuses to continue the use of a price comparator services. Users must only access this object (e-shop) directly. In the real situation, large e-shops may have their own policy of independence. If they are strong enough (have a lot of customers for example) they may, with some probability, to refuse to use a price comparator services. This means users will have to come then directly to such e-shop (the real situation is described in [5]). This decision can have a real basis, based on the situation shown in Figures 5a and 6a, when we stated that under given conditions the function of the comparator does not have a major influence on the strength of the object (e-shop). This is also reflected in Figures 4 and 5. Conversely, this decision of an e-shop can have a significant effect on a price comparator. It can be seen especially on the distribution characteristic curves without cancelation of comparator services (red mark) and with nonzero probability that some e-shop will cancel to use comparator services (blue mark) on Figure 5b. In a real case, a price comparator must attract all e-shops especially big ones.

5 Conclusion

Complex networks play an essential role in a wide range of disciplines such as social and behavioral sciences, biology, economics, industrial engineering, information technology and more. By analyzing the structure and behavior of a given network, we can draw a number of useful conclusions when studying various complex systems. In this paper we use a model of evolving weighted tripartite network with preferential growth mechanisms and different rules for changing the strength of nodes and edge weights. Our paper is closely related to our previous work [1], where we describe the model in detail. Our goal was to create and to analyze a model of a weighted tripartite network to gain an overview of the role of the network structure in e-commerce relationships involving three types of actors: user, e-shops and price comparators. Despite its simplicity, the model captures the essential characteristics of the modeled real network, the distribution of node strength then corresponds to the distribution characteristic of the power law under defined conditions. The results of numerical simulations correspond to performed theoretical analyzes. This weighted tripartite evolving network provides another certain ideas about the behavior of complex systems with more complicated behavior, such as the multi-actor e-commerce system.

In order to create a more realistic e-commerce model in the presence of a price comparator, we are considering further work in the future. First, simulation experiments with different parameter conditions can reveal more interesting e-commerce features in the presence of a price comparator. Secondly, consumer behavior in the online market is a more complex process than just price perception in the context of online shopping, depending on many circumstances, including for example perception of ease in assessing the quality of an online product. We will therefore try to further refine some parameters of our model expressing the behavior of individual actors in order to more accurately model the described system.

References

[1] BERANEK, L. & REMES, R. (2020). Distribution of Node Characteristics in Evolving Tripartite Network. *Entropy*, 22, 263.

[2] BERANEK, L. & HANZAL, P. (2017). Price comparison sites and their influence on e-commerce processes. In P. PRAŽÁK (Ed.) *Proceedings of the 35th International Conference Mathematical Methods in Economics* (MME), Hradec Kralove, Czech Republic: University of Hradec Králové, (pp. 48–53).
[3] BERANEK, L., TLUSTY, P., & REMES, R. (2010). The Usage of Belief Functions for an Online Auction Reputation Model. In: L. FRIEBEL, J. FRIEBELOVA (Eds.) *Proceedings of the 28th Conference on Mathematical Method in Economy 2010*. České Budejovice: University of South Bohemia, (pp. 49–54).
[4] BORGATTI, S. P. & HALGIN, D. S. (2011). On Network Theory. *Organization Science*, 22, 1168–1181.
[5] CERNY, A. (2016). E-shop war: Alza the biggest e-shop leaves biggest price comparator Heureka: (in Czech). *iDnes* [online]. 5-19-2016, Available from: https://www.idnes.cz/ekonomika/domaci/alza-opousti- heureku-a-vyzyva-k-tomu-i-ostatni.A160519_153822_ekonomika_rny (accessed on 14-02-2020).
[6] DACOSTA, L. F., RODRIGUES, F. A., & CRISTINO, A. S. (2008). Complex networks: The key to systems biology, *Genetics and Molecular Biology*, 31(3), 591–601.
[7] DRLIK, M. & BERANEK, L. (2015). The role of an online social network in informal learning. In M. HOUŠKA, I. KREJČÍ, M. FLÉGL, M. FEJFAROVÁ, H. URBANCOVÁ, J. HUSÁK (Eds.) *Proceedings of the 12th International Conference on Efficiency and Responsibility in Education* (ERIE), Prague: Czech University of Life Sciences, (pp. 81–88).
[8] HOU, R., WU, J. & DU, H. S. (2017). Customer social network affects marketing strategy: A simulation analysis based on competitive diffusion model. *Physica A: Statistical Mechanics and its Applications*, 469, 644–653.
[9] MICHAL, P., KLIMEK, P., JURICKOVA, E., & PALKA, P. (2016). Comparison shopping agents and Czech online customers' shopping behavior. *Int. J. of Entrepreneurial Knowledge*, 4(2), 62–69.
[10] NIKAKHTAR, N. & JIANZHENG, Y. (2013). Impact of network on E-commerce economy, Analysis on a price competition model. *Life Science Journal*, 10, 2801–2808.
[11] SAAVEDRA, S., REED-TSOCHAS, F., & UZZI B. (2009). A simple model of bipartite cooperation for ecological and organizational networks. *Nature*, 457, 463–466.
[12] WAN, CH., LI, T., GUAN, Z. H., WANG, Y. & LIU, X. (2017). Spreading dynamics of an e-commerce preferential information model on scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 467, 192–200.
[13] WANG, H., WANG, J., SMALL, M., & MOORE, J. M. (2019). Review mechanism promotes knowledge transmission in complex networks. *Applied Mathematics and Computation*, 340, 113–125.
[14] YAN, B., LIU, Y., LIU, J., CAI, Y., SU, H., & ZHENG, H. (2018). From the Periphery to the Center: Information Brokerage in an Evolving Network. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, Stockholm: International Joint Conferences on Artificial Intelligence, (pp. 3912–3918).
[15] ZHANG, CH. X., ZHANG, Z. K., & LIU, CH. (2013). An evolving model of online bipartite networks, *Physica A: Statistical Mechanics and its Applications*, 392, 6100–6106.

Efficiency of Social Transfers in Poverty Reduction: Development in the Czech Republic

Vladislav Bína[1], Jiří Přibil[2], Lucie Váchová[3]

Abstract. The Czech Republic is one of the EU member states with the lowest values of poverty rates for many years. The presented paper analyses the development of relative poverty in the Czech Republic in years from 2005 to 2018 together with the measures of efficiency of social transfers in the perspective of relative poverty reduction. All computations are based on the data from the European Union – Statistics on Income and Living Conditions sample survey.

This analysis is augmented by a regional view of situation in the year 2018. This perspective shows regional differences in poverty and efficiency of social transfers and presents the targeting of social transfers on the individuals and households with incomes (before social transfers) close to a poverty threshold in the year 2018.

Keywords: Relative poverty, social transfers, EU-SILC, efficiency

JEL Classification: D31, I32, I38
AMS Classification: 62P20

1 Relative Poverty and Social Transfers

During the 20th century, the modern welfare state developed gradually throughout the European states and addressed the issues of unemployment, poverty, economic crises and collapses of the financial system. Gradually, all European countries employed at least basic net of social transfers protecting most of population endangered by low incomes and insufficient living standards. Nowadays, the EU member states provide similar categories of social transfers and agreed on a common approach to poverty measurement which employs the concept of relative poverty based on OECD approach suitable for the assessment of developed countries [7].

In 2010 in reaction to the economic crisis and threat of unemployment and poverty European Commission formulated a communication named "EUROPE 2020: A strategy for smart, sustainable and inclusive growth" [4]. This strategy includes poverty alleviation as one of the five most important targets. This vision stated among its headline targets that by 2020 the poverty should be reduced by one third, i.e. 20 million fewer people should be at risk of poverty throughout EU states and member states were invited to incorporate this goal to their national targets. This effort is already claimed to be partially successful in most of the EU states but still in ten states (mostly in the states more affected by the preceding economic crisis) the poverty issue appears to be more significant [13].

This paper aims to analyse the development of poverty in the years 2005 to 2018 in the Czech Republic as a part of this picture. The analysis will be augmented by a perspective of efficiency of social transfers in poverty reduction. (Although it is apparent that the poverty alleviation is not the only aim of social transfers which includes other losses of income not necessarily leading to poverty like illness, disability, maternity, survival benefits and unemployment benefits.) In the Czech Republic it was shown already in times of Microcenzus surveys that the efficiency of transfers for poverty elimination is rather high, moreover, it is the highest across Europe [15]. For alternative approaches including the share of net social expenditures on GDP see, e.g. Caminada & Goudswaard [3] or for the effect of change in the public budget on social transfers see Leventi et al. [9].

EU-SILC data

Source of data for the above-sketched analyses are the EUROSTAT EU-SILC survey microdata files. EU-SILC (European Union – Statistics on Income and Living Conditions) is an annually performed survey held in randomly sampled households. The form of the survey is a four-year sample with a quarter of households replaced each year. Data concerning incomes and social condition are collected throughout Europe (with few exceptions) since 2005 and thanks to common methodology provide information comparable across the European states. Incomes of individuals and households are collected within the whole year and then

[1] University of Economics, F. of Management, Dept. of Exact Methods, Jarošovská 1117/II, 377 01, J. Hradec, vladislav.bina@vse.cz.
[2] University of Economics, F. of Management, Dept. of Exact Methods, Jarošovská 1117/II, 377 01, J. Hradec, jiri.pribil@vse.cz.
[3] University of Economics, F. of Management, Dept. of Exact Methods, Jarošovská 1117/II, 377 01, J. Hradec, lucie.vachova@vse.cz.

appear in a "next year" data set together with a large amount of individuals' background and household characteristic, i.e. 2018 EU-SILC data contain incomes from the year 2017 [5, 6]. The most important methodological issues and pitfalls are discussed, e.g., by Mysíkova [12].

1.1 Relative Poverty Measurement in EU

The OECD approach to the poverty measurement [7] was adjusted in means of weights in the calculation of equivalent household size and adopted by all EU states. The methodology began to unify when Atkinson proposed a common set of social indicators at the turn of millennia [2] and harmonized by the European Council in Laeken in 2001 [10]. Later from 2004, the unified output of national surveys resulted into the start of EU-SILC data and its coverage grows with the process of EU enlargement (with several non-member states who also adopted the methodology). In more detail see Jenkins [8].

In agreement with this methodology and according to the EUROSTAT [5] we define the following notions. The EU-SILC data contain among many other also variables concerning the income of households. In the context of this paper, we use variable $HY022$ **total disposable household income before social transfers** other than old-age and survivors benefits and $HY020$ **total disposable household income** (after social transfers).

The calculation of **equivalent household size** requires two auxiliary quantities, namely, HM^{14+} is a number of household members aged 14 and over, HM^{13-} is a number of household members aged 13 or less (both at the end of income reference period). The equalized household size is then given by the formula

$$HX050 = 1 + 0.5(HM^{14+} - 1) + 0.3HM^{13-} \tag{1}$$

Now, the computation of **equivalised household income** uses the income of households and in a standard variant of incomes after social transfers is defined as an income per (equivalised) household member

$$HX090 = \frac{HY020}{HX050}. \tag{2}$$

A national **poverty threshold** PT is then computed as a 60% of median (weighted by household weights) of equivalised household income $HX090$ in the respective state.

Household (and its members) are **at-risk-of-poverty** if their equivalised household income $HX090$ is below the poverty threshold (value of poverty indicator $HX080 = 1$). In the opposite case, the household poverty indicator takes value of $HX080 = 0$.

A **poverty gap** index measures the intensity of poverty calculating the average depth in the poverty as

$$PG = \frac{1}{N} \sum \frac{(PT - HX090) \cdot HX080}{PT} \tag{3}$$

where N is the total population and the implementation on EU-SILC data again employs the household cross-sectional weights [1].

1.2 Efficiency of Social Transfers in Poverty reduction

If we focus on the role of social transfers in poverty alleviation, it is rather straightforward that it can be measured in means of a reduction in the number of households which are taken above the poverty threshold by the social transfers or computing a proportion of transfers spent on the transition above the poverty threshold (for a profound analysis see [15]).

Groups of Social Transfers Covered by EU-SILC data

It is not within the focus of this paper to provide a complete summary of the Czech social system. The rest of the paper will take advantage of a simplified structure as covered by EU-SILC data. Let us also stress that pensions are considered as a primary income since they are related to life-cycle, not to a redistribution of resources.

The considered social transfers appearing as a difference between $HY020$ and $HY022$ income variables are the following (the first four concern individuals and therefore need to be summed up for all individuals in the household):

- unemployment benefits ($PY090G$);
- sickness benefits ($PY120G$);
- disability benefits ($PY130G$);
- education-related allowances ($PY140G$);
- family/children related allowances ($HY050G$);
- social exclusion not elsewhere classified ($HY060G$);

- housing allowances ($HY070G$).

Variants of Poverty Efficient Social Transfers

A basic overview of papers concerning the relations between social expenditures and poverty alleviation can be found in Miežienė & Krutulienė [11]. In this paper, we focus on the efficiency of above listed social transfers on poverty reduction in inspiration by Trbola & Sirovátka [15].

Therefore, we consider the following three views of efficiency in poverty reduction. All three possibilities consider the social transfers spend on households at-risk-of-poverty and using cross-sectional household weights aim to estimate the ratio of the sum of such social transfers and the sum of all social transfers in the Czech Republic in the analyzed year.

1. **poverty-reduction efficient social transfers**: estimate share of social transfers spent on households below the poverty threshold resulting in improvement of poverty status (i.e., social transfers pulling household above the poverty threshold);
2. **poverty-alleviation efficient social transfers**: estimate share of all social transfers spent on households below the poverty threshold;
3. **strictly efficient social transfers**: estimate share of social transfers spent on households below the poverty threshold in such a way that they are sufficient to pull household to the poverty threshold (no transfer exceeding the poverty threshold are considered).

	2005	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**
Poverty threshold (€)	2539.45	2878.06	3253.53	3640.54	4376.81	4235.01	4470.81
At-risk-of-pov. (after ST, %)	10.36	9.80	9.61	9.05	8.56	9.00	9.80
At-risk-of-pov. (before ST, %)	17.94	17.88	16.76	15.90	14.45	14.40	14.92
Poverty gap (after ST, %)	18.17	16.97	18.12	18.52	18.84	21.06	17.23
Poverty gap (before ST, %)	37.64	33.82	37.55	32.83	29.97	31.70	29.42
Sum of social transfers (mil. €)	4.80	5.52	6.08	7.04	7.62	7.50	7.44
Poverty r. efficient ST (1, %)	32.40	35.78	33.21	35.08	33.05	31.51	28.69
Poverty al. efficient ST (2, %)	53.33	53.79	52.02	48.55	46.54	44.94	42.69
Strictly efficient ST (3, %)	37.31	36.74	35.77	30.84	28.81	27.96	27.41

	2012	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Poverty threshold (€)	4674.79	4616.19	4573.40	4453.69	4702.78	4969.09	5452.97
At-risk-of-pov. (after ST, %)	9.62	8.60	9.72	9.75	9.68	9.15	9.59
At-risk-of-pov. (before ST, %)	14.45	13.77	14.52	14.10	13.96	13.20	13.03
Poverty gap (after ST, %)	19.10	16.61	17.96	19.19	19.47	16.60	14.98
Poverty gap (before ST, %)	30.93	27.66	31.49	33.11	33.22	27.56	24.79
Sum of social transfers (mil. €)	7.10	6.76	6.61	6.15	6.37	6.48	6.44
Poverty r. efficient ST (1, %)	30.94	33.66	30.69	30.76	28.76	29.77	27.85
Poverty al. efficient ST (2, %)	45.52	46.77	49.84	49.96	48.88	45.19	43.22
Strictly efficient ST (3, %)	29.40	29.68	33.75	33.85	33.82	30.15	28.86

Table 1 Trends of selected characteristics of poverty and measures of the efficiency of social transfers in the Czech Republic based on EU-SILC data.

2 Changes in the Czech Republic from 2005 to 2018

Throughout the analyzed period several significant socio-economic changes occur. After becoming a member of EU in 2004 the Czech Republic joined a Schengen Area at the end of 2007 and one year later the Great Recession occurred also in Central Europe and lasted approximately five years. Then years of prosperity arrived which lasted until the Coronavirus crisis of 2020. Unfortunately, the availability of EU-SILC data ends with the year 2018 and thus do not include this globally significant change in the presented picture.

2.1 Development of Basic Measures

Let us have a look at the summary of characteristics of poverty as mentioned in the previous section (see Table 1) and denote that both at-risk-of-poverty and poverty gap are computed as variants before and after social transfers.

It appears that both characteristics evaluated from incomes before social transfers decrease in the analyzed period with a small perturbation at the end of the Great Recession. The values of at-risk-of-poverty after social transfers decrease only slightly or are rather constant. On the other hand, the poverty gap based on incomes after social transfers appears to be lower in some years before the crisis and then in the last years (possibly in some connection with low unemployment rate).

Estimate sum of social transfers again show a significant growth in crisis years and later years the value decreases slightly and then remains constant. As a rather bad result, we can observe that all three measures of the efficiency of social transfers show a substantial decrease in the observed period with several disturbances in years of ending crisis from 2013 to 2016 (as already mentioned, income data are one year older than a year of the survey).

2.2 Regional Perspective

According to EUROSTAT [14], the capital Prague ranks among the regions with the seventh-highest regional GDP per capita in the EU in 2017. When analyzing the relative poverty in the Czech Republic, it appears to be important to analyze the situation in different regions of the Czech Republic. The EU-SILC data for the Czech Republic provide a structuring on NUTS-2 regions and thus the analysis is possible. The CZ-NUTS-2 regions are:

- **CZ01**: Prague;
- **CZ02**: Central Bohemia;
- **CZ03**: Southwest;
- **CZ04**: Northwest;
- **CZ05**: Northeast;
- **CZ06**: Southeast;
- **CZ07**: Central Moravia;
- **CZ08**: Moravian-Silesian.

Figure 1 shows eight pairs of graphs for NUTS-2 regions in 2018. The upper graph presents sums of social transfers in categories relative to the median of income before social transfers, the lower graph shows an average amount of social transfers spent in each category. The graphs show without any surprise that although a high amount of transfers is spent on households above the poverty threshold, still the distribution is concentrated in the surroundings of the poverty threshold. Higher amounts for households below the poverty threshold are spent in regions which are traditionally labelled as "poor", i.e. CZ04: Northwest, CZ07: Central Moravia and CZ08: Moravian-Silesian.

In all regions (with only insignificant differences) the average amount of social transfers decreases as the income before transfers increases towards the poverty threshold and then above it only much smaller amounts are paid (although the sum in the category can be quite high). The regional characteristics summarized in Table 2 again confirm the "rich" character of Prague (partially together with Central Bohemia) and show lower efficiency in poverty alleviation whereas the "poor" regions show the opposite behaviour.

	CZ01	CZ02	CZ03	CZ04	CZ05	CZ06	CZ07	CZ08
Poverty threshold (€)	5452.97	5452.97	5452.97	5452.97	5452.97	5452.97	5452.97	5452.97
At-risk-of-pov. (after ST, %)	10.54	8.71	8.13	11.90	9.84	8.44	8.34	9.33
At-risk-of-pov. (before ST, %)	12.05	11.43	11.21	15.35	13.82	11.20	11.74	13.51
Poverty gap (after ST, %)	18.93	17.41	14.09	18.90	12.14	15.66	13.70	16.29
Poverty gap (before ST, %)	22.97	27.40	16.18	33.83	26.19	24.96	22.78	39.46
Sum of social transfers (mil. €)	0.50	0.69	0.74	0.72	0.93	1.10	0.81	0.95
Poverty r. efficient ST (1, %)	18.18	19.48	25.24	32.32	29.10	24.62	35.83	33.48
Poverty al. efficient ST (2, %)	22.91	37.04	35.63	55.77	44.53	37.02	49.25	55.73
Strictly efficient ST (3, %)	13.01	29.28	23.02	35.81	30.89	24.80	29.12	38.81

Table 2 Selected characteristics of poverty and measures of the efficiency of social transfers in Czech NUTS-2 regions based on 2018 EU-SILC data.

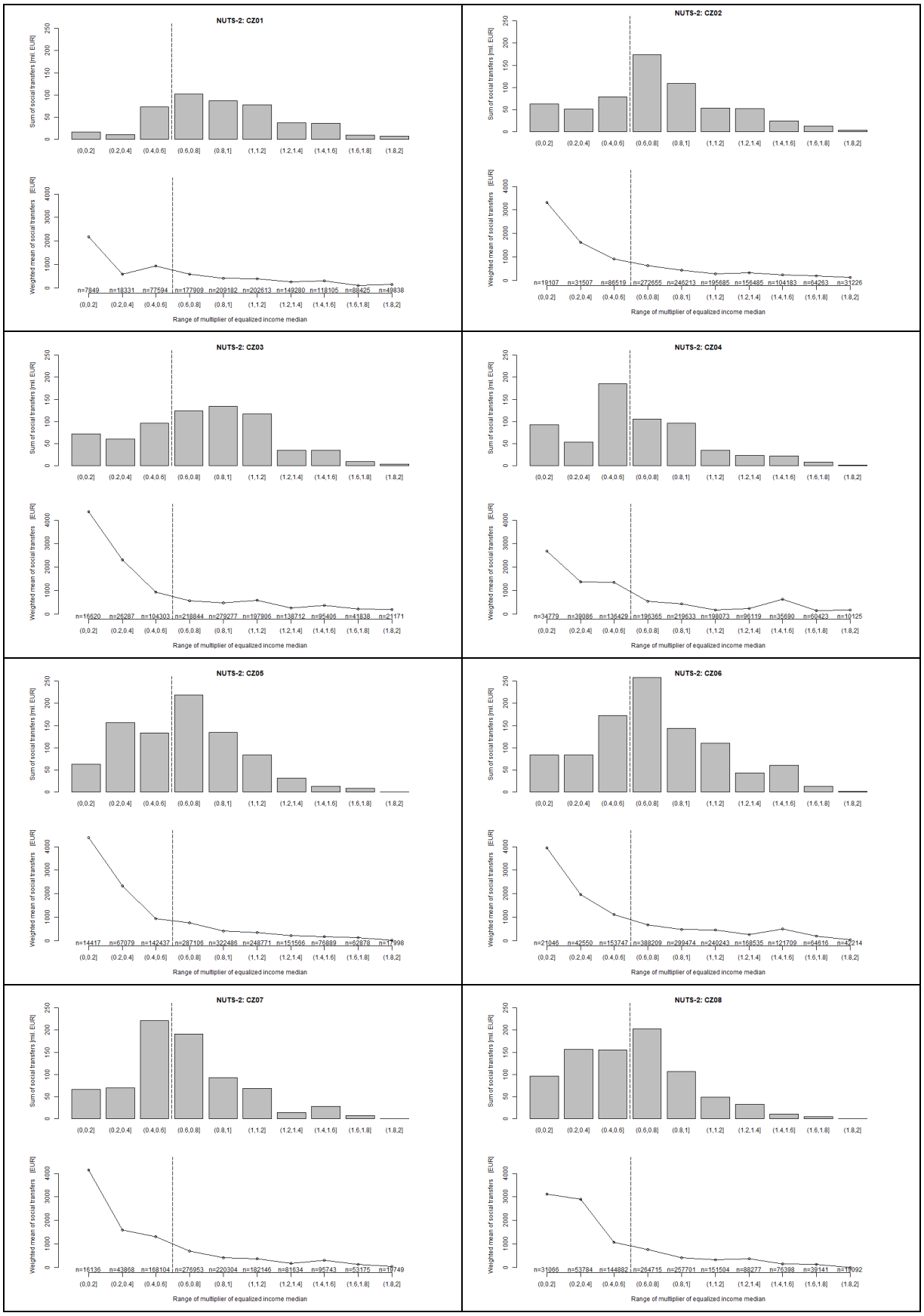


Figure 1 Eight pairs of graphs for Czech NUTS-2 regions in 2018. The upper graph presents sums of social transfers in categories relative to the median of income before social transfers, the lower graph shows an average amount of social transfers spent in each category. Dashed line visualizes a poverty threshold.

3 Conclusions

The presented analysis showed that in the observed period no substantial change in at-risk-of-poverty characteristic occurred and no clear trend can be observed in the case of poverty gap either. Therefore, within nearly the whole period of validity of EUROPE 2020 strategy, the extent of poverty in the Czech Republic remains more or less unchanged. Since the relative poverty is not so serious issue in the Czech Republic as in other EU countries, more interesting observation concerns the efficiency of social transfers with respect to the poverty alleviation. All three presented measures show a substantial decrease of the percentage of poverty efficient social transfers.

As the time span of EUROPE 2020 strategy ends it appears to be interesting to evaluate its results with respect to the poverty reduction, possibly together with the efficiency of European social transfer systems. The theme is even more interesting in the perspective of contemporary COVID-19 pandemic and consequent start of the economic crisis.

Acknowledgements

The first author was supported by the National Science Foundation of the Czech Republic (GAČR) under project no. 19-00015S (Identification of Poverty and Social Exclusion Temporal Patterns of Households Based on Multivariate Mixed Type Panel Data) and the third author was supported under EUROSTAT RPP 289/2018-EU-SILC.

References

[1] ALFONS, A. & TEMPL, M. (2012). Estimation of social exclusion indicators from complex surveys: The R package laeken. *KU Leuven, Faculty of Business and Economics Working Paper*.

[2] ATKINSON, T., CANTILLON, B., MARLIER, E., & NOLAN, B. (2002). *Social indicators: The EU and social inclusion*. OUP, Oxford.

[3] CAMINADA, K. & GOUDSWAARD, K. P. (2010). How well is social expenditure targeted to the poor? In: P. SAUNDERS & R. SAINSBURY (Eds.), Social Security, Poverty and Social Exclusion in Rich and Poorer Countries, *International Studies on Social Security, 16* (pp. 97–112). Mortsel, Belgium: Intersentia.

[4] EUROPEAN COMMISSION (EC). (2010). *EUROPE 2020: A strategy for smart, sustainable and inclusive growth*. Working paper {COM (2010) 2020}.

[5] EUROPEAN COMMISSION – EUROSTAT (2018a): Differences Between Original Database and Anonymised User Database. SILC Disclosure Control Rules.

[6] EUROPEAN COMMISSION – EUROSTAT (2018b): Methodological Guidelines and Description of EUSILC Target Variables. DocSILC065 (2018 operation).

[7] FORSTER, M., LEVY, H., D'ERCOLE, M. M., & RUIZ, N. (2013). *The OECD Approach to Measure and Monitor Income Poverty Across Countries*. Working Paper 17. np: United Nations Economic Commission for Europe: Conference of European Statisticians.

[8] JENKINS, S. P. (2020). Perspectives on Poverty in Europe. Following in Tony Atkinson's Footsteps. *Italian Economic Journal, 6*(1), 129–155.

[9] LEVENTI, C., SUTHERLAND, H., & TASSEVA, I. V. (2019). Improving poverty reduction in Europe: What works best where? *Journal of European Social Policy*, 29(1), 29–43.

[10] MARLIER, E., ATKINSON, T., ATKINSON, A. B., CANTILLON, B., & NOLAN, B. (2007). *The EU and social inclusion: Facing the challenges.* Policy Press.

[11] MIEŽIENĖ, R. & KRUTULIENĖ, S. (2019). The impact of Social transfers on poverty reduction in EU countries. *Baltic Journal of European Studies*, *9*(1), 157–175.

[12] MYSÍKOVÁ, M. (2011). EU-SILC a jeho metodologická úskalí: mezinárodní srovnatelnost a příjmové proměnné. *Data a výzkum – SDA Info*, *5*(2), 147–170.

[13] NEWSRELEASE, EUROSTAT (2017). Downward trend in the share of persons at risk of poverty or social exclusion in the EU.

[14] NEWSRELEASE, EUROSTAT (2019). GDP per capita in 281 EU regions: Regional GDP per capita ranged from 31% to 626% of the EU average in 2017.

[15] TRBOLA, R. & SIROVÁTKA, T. (2006). *Efektivnost sociálních transferů při eliminaci chudoby v České republice*. VÚPSV, výzkumné centrum Brno.

Portfolio Selection via Fuzzy Mean-Variance Model

Adam Borovička[1]

Abstract. Portfolio selection is still an interesting topic, as thousands of people around the world face this decision-making. Such a decision-making process may be nontrivial due to its potential complexity. There can be a number of effecting factors. The most important are undoubtedly return and risk. These characteristics can be reflected by a well-known mean-variance model using for a portfolio selection. However, return is usually instable over time. Even risk can also vary. This instability and associated uncertainty can be effectively quantified through the fuzzy set. Then return and risk are proposed as triangular fuzzy numbers. Model with fuzzy elements can also respect the investor's vague preferences. The fuzzified mean-variance model can be solvable by fuzzy mathematical programming techniques. Model respecting a typical uncertainty is then much closer to reality. Thus, a portfolio composition can be more representative and satisfactory. The application contribution of a developed fuzzy approach is demonstrated on selecting a portfolio from stocks traded on the Czech capital market. The results are analyzed and confronted with the output of a crisp mean-variance model.

Keywords: fuzzy, mean-variance, portfolio selection, stock

JEL Classification: C44, C61, G11
AMS Classification: 90B50, 90C30, 90C70

1 Introduction

This paper contributes into the still-living topic – portfolio selection problem. Many people around the world are making decisions about where to invest their free funds. Making such a decision is often not easy. The use of a supporting tool is so essential.

One of these instruments is a well-known mean-variance model developed by H. M. Markovitz [5, 6]. Over the last few decades, this model has undergone several modifications to increase a decision-making ability in a dynamic environment. One approach replaces a crisp return with its fuzzy form. (Triangular) fuzzy number then can reflect the uncertainty (instability) of return more satisfactory [3]. But what about risk represented by variance? The risk of the investment can also be unstable, especially over a longer period. In other words, the variance can change during the tracked time period, or investment horizon. Therefore, I propose to express the instable variance, or covariance, of returns as triangular fuzzy number which is not usual in the current concepts.

The main aim of this article is to develop the Markowitz model with fuzzy return and risk – fuzzy mean-variance model. The accompanying, and necessary aim is to design the complex methodological concept for a portfolio selection using the proposed fuzzy mean-variance concept. The application power of the developed methodological process is demonstrated on a real-life selecting a portfolio from the stocks traded on the Prague Stock Exchange. A comparison of the portfolios made by fuzzy and "standard" crisp mean-variance model clearly shows the importance of reflecting return and risk instability over time.

The structure of the article is as follows. After the introduction, mean-variance model is briefly described. Then its fuzzy modification is developed, and the methodological process of a portfolio selection is designed. Next section deals with real-life making a portfolio from the stocks traded on the Czech capital market. Finally, a contribution of the article is summarized and some ideas for further research are opened.

2 Portfolio selection process: Fuzzy mean-variance model

After a brief description of a well-known Markowitz mean-variance model, its fuzzy form is designed. Fuzzy portfolio characteristics are presented in detail. Finally, the complex methodological process of a portfolio selection using fuzzy mean-variance model is proposed.

[1] University of Economics, Department of Econometrics, W. Churchill Sq. 4, Prague, Czech Republic, adam.borovicka@vse.cz.

2.1 Mean-variance model

Mean-variance model is well-known optimization approach for a portfolio selection. It can be formulated as follows

$$\begin{aligned} &\min \quad \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m}\sigma_{ij}x_i x_j} \\ &\sum_{i=1}^{m} r_i x_i \geq R \\ &x_i \in X \qquad i = 1,2,\ldots,m, \end{aligned} \tag{1}$$

where $x_i, i = 1,2,\ldots,m$, or $x_j, j = 1,2,\ldots,m$, is the weight of the i-th, or j-th asset in the portfolio, $\sigma_{ij}, i,j = 1,2,\ldots,m$,represents the covariance of return of the i-th and j-th asset, $r_i, i = 1,2,\ldots,m$, denotes the expected return (mean) of the i-th asset, R is a minimum required level of the portfolio return. The set X reflects all other conditions, including "portfolio" constraint $\sum_{i=1}^{n} x_i = 1$ and non-negativity conditions for all variables.

2.2 Fuzzy mean-variance model

The significant specific, and also benefit, of the fuzzy form of the mean-variance model is an ability to taking into account a phenomenon of the capital market – uncertainty. The uncertainty is often connected with an unstable development of the asset return. However, instability can also be infiltrated into the risk of investment, which can thus change over time. Inclusion of an instability of the characteristics makes the optimization model closer to reality, which reinforces its applicability. A suitable means for expressing the element of uncertainty is the fuzzy set theory. Then a fuzzy (vague) return and risk can be expressed via triangular fuzzy number as can be seen in the following fuzzy formulation of a mean-variance model

$$\begin{aligned} &\min \quad \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m}\tilde{\sigma}_{ij}x_i x_j} \\ &\sum_{i=1}^{m} \tilde{r}_i x_i \geq \tilde{R} \\ &x_i \in X \qquad i = 1,2,\ldots,m, \end{aligned} \tag{2}$$

where $\tilde{\sigma}_{ij}, i,j = 1,2,\ldots,m$, is a fuzzy covariance of the i-th and j-th asset, $\tilde{r}_i, i = 1,2,\ldots,m$,represents a fuzzy expected return (mean) of the i-th asset, $\tilde{R}$ denotes a fuzzy minimum required level of portfolio return expressed by the triangular fuzzy number(r_p^l, r_p^m, r_p^u). Other symbols are interpreted the same way as in the model (1).

2.3 Fuzzy portfolio characteristics

As mentioned above, both characteristics are expressed as triangular fuzzy numbers. This often used type of fuzzy numbers usually satisfactorily represent an unstable (uncertain, vague) return and risk. Another benefit is that the basic computing operations are well-known and easy to implement. Now, let us introduce both vague elements of a portfolio selection process.

Fuzzy return of the i-th asset is represented using fuzzy mean expressed as a triangular fuzzy number denoted as $\tilde{r}_i = (r_i^l, r_i^m, r_i^u)$. Its lower, middle and upper parameter are proposed as follows

$$r_i^l = \min_{1\leq k\leq n}(r_{ik}), \quad r_i^m = \frac{1}{n}\sum_{k=1}^{n} r_{ik}, \quad r_i^u = \max_{1\leq k\leq n}(r_{ik}), \qquad i = 1,2,\ldots,m, \tag{3}$$

where $r_{ik}, i = 1,2,\dots,m, k = 1,2,\dots,n$, represents the k-th historical mean (calculated for a specified subperiod) of the i-th asset. Then fuzzy portfolio return (mean) $\tilde{r}_p$ can be also formulated as the following triangular fuzzy number

$$r_{ik}, i = 1,2,\dots,m, k = 1,2,\dots,n,. \tag{4}$$

Fuzzy risk is reflected through fuzzy variance, or covariance $\tilde{\sigma}_{ij}$ of i-th and j-th asset's return which can be formalized as $\tilde{\sigma}_{ij} = (\sigma_{ij}^l, \sigma_{ij}^m, \sigma_{ij}^u)$. Its lower, middle and upper parameter then are computed as follows

$$\sigma_{ij}^l = \min_{1\le k\le n}(\sigma_{ijk}), \quad \sigma_{ij}^m = \frac{1}{n}\sum_{k=1}^{n}\sigma_{ijk}, \quad \sigma_{ij}^u = \max_{1\le k\le n}(\sigma_{ijk}), \qquad i,j = 1,2,\dots,m, \tag{5}$$

where $\sigma_{ijk}, i,j = 1,2,\dots,m, k = 1,2,\dots,n$, represents the k-th historical covariance (calculated for a specified subperiod) between the return of i-th and j-th asset. Fuzzy portfolio risk (as a fuzzy variance of portfolio return) σ_p^2 then is presented as the triangular fuzzy number

$$\sigma_p^2 = \sum_{i=1}^{m}\sum_{j=1}^{m}(\sigma_{ij}^l, \sigma_{ij}^m, \sigma_{ij}^u)x_i x_j = \left(\sum_{i=1}^{m}\sum_{j=1}^{m}\sigma_{ij}^l x_i x_j, \sum_{i=1}^{m}\sum_{j=1}^{m}\sigma_{ij}^m x_i x_j, \sum_{i=1}^{m}\sum_{j=1}^{m}\sigma_{ij}^u x_i x_j\right). \tag{6}$$

2.4 Solving the fuzzy mean-variance model

To solve the problem of fuzzy (nonlinear) mathematical programming, the fuzzy model is standardly transformed to the strict form. Then the model (2) can be transformed through the multiple objective programming principle as follows [4]

$$\begin{aligned}
&\text{“min”} \quad \left\{f_{r_p}^l(\mathbf{x}), f_{r_p}^m(\mathbf{x}), f_{r_p}^u(\mathbf{x})\right\} \\
&\sum_{i=1}^{m} r_i^l x_i \ge r_p^l \qquad \sum_{i=1}^{m} r_i^m x_i \ge r_p^m \qquad \sum_{i=1}^{m} r_i^u x_i \ge r_p^u, \\
&x_i \in X \qquad i = 1,2,\dots,m
\end{aligned} \tag{7}$$

where $f_{r_p}^l(\mathbf{x}) = \sum_{i=1}^{m}\sum_{j=1}^{m}\sigma_{ij}^l x_i x_j$, $f_{r_p}^m(\mathbf{x}) = \sum_{i=1}^{m}\sum_{j=1}^{m}\sigma_{ij}^m x_i x_j$, $f_{r_p}^u(\mathbf{x}) = \sum_{i=1}^{m}\sum_{j=1}^{m}\sigma_{ij}^u x_i x_j$ with the vector of variables $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$. A comparison of triangular fuzzy numbers is performed by simple, however slightly simplifying (not complex), approach (more about simple comparison technique in [2]). As can be deduced from the model (7), for instance, fuzzy return $\tilde{r}_1 = (r_1^l, r_1^m, r_1^u)$ is not lower than $\tilde{r}_2 = (r_2^l, r_2^m, r_2^u)$, when the following holds $r_1^l \ge r_2^l \wedge r_1^m \ge r_2^m \wedge r_1^u \ge r_2^u$. This approach is obviously well applicable for a comparison of triangular fuzzy numbers within a mathematical model standardly operating with the apparatus of equations and inequalities.

In the spirit of the Bellman optimality principle using maxmin optimization approach [1], the multi-objective model (7) is solved via the following one-objective form

$$\begin{aligned}
&\max \alpha \\
&\frac{f_{r_p}^{lB} - f_{r_p}^l(x)}{f_{r_p}^{lB} - f_{r_p}^{lI}} \ge \alpha \qquad \frac{f_{r_p}^{mB} - f_{r_p}^m(x)}{f_{r_p}^{mB} - f_{r_p}^{mI}} \ge \alpha \qquad \frac{f_{r_p}^{uB} - f_{r_p}^u(x)}{f_{r_p}^{uB} - f_{r_p}^{uI}} \ge \alpha \\
&\sum_{i=1}^{m} r_i^l x_i \ge r_p^l \qquad \sum_{i=1}^{m} r_i^m x_i \ge r_p^m \qquad \sum_{i=1}^{m} r_i^u x_i \ge r_p^u \\
&x_i \in X \qquad i = 1,2,\dots,m \\
&0 \le \alpha \le 1,
\end{aligned} \tag{8}$$

where $f_{r_p}^{lB}, f_{r_p}^{mB}, f_{r_p}^{uB}$ or $f_{r_p}^{lI}, f_{r_p}^{mI}, f_{r_p}^{uI}$ are the worst, or the best possible values of the "lower", "middle" and "upper" portfolio risk achievable on the set X. These extreme values represent minimum and maximum of the "risk" functions $f_{r_p}^{l}(\mathbf{x})$, $f_{r_p}^{m}(\mathbf{x})$, $f_{r_p}^{u}(\mathbf{x})$ on the set X. They are therefore determined using one-objective models. Extreme levels of returns can be determined in a similar manner. These values can be used for a determination of the reference return values r_p^l, r_p^m and r_p^u.

2.5 Methodological process of a portfolio selection

In the last subsection, the complex methodological process for a portfolio selection is designed trough the following several steps.

Step 1: At first, the investment policy, or strategy is declared. Based on this, potential assets for investment are preselected.

Step 2: Further, the unnecessary data of the investment instruments are collected for a specified time period. This period is derived from the investment horizon. The prices tracked in the chosen historical period serve to a computation of the main investment characteristics – return and risk. To reflect an instability of these characteristics, this historical period is divided into shorter subperiod. In these subperiods, return (mean) and risk ((co)variances) are calculated. Thus, mean of the i-th asset in k-th subperiod is calculated

$$r_{ik} = \frac{\sum_{p=1}^{P_k} r_{ikp}}{P_k} \qquad i = 1,2,\dots,m, k = 1,2,\dots,n, \tag{9}$$

where $r_{ikp}, i = 1,2,\dots,m, k = 1,2,\dots,n, p = 1,2,\dots,P_k$, represents p-th return of the i-th asset in the k-th subperiod. P_k is the number of observations of return in the k-th subperiod. This number is differentiated according to the period because the number of observations can be different over all subperiods. For instance, every month has a specific number of trading days. The covariance of returns of the i-th and j-th asset in the k-th subperiod is calculated as follows

$$\sigma_{ijk} = \frac{\sum_{p=1}^{P_k}(r_{ikp} - r_{ik})(r_{jkp} - r_{jk})}{P_k} \qquad i,j = 1,2,\dots,m, k = 1,2,\dots,n, \tag{10}$$

where r_{ikp}, or r_{jkp} denotes the p-th return of the i-th, or j-th asset in the k-th subperiod, r_{ik}, or r_{jk} is a mean of the i-th, or j-th asset in k-th subperiod. These sub-characteristics participates in the determination of the fuzzy characteristics through the formulas (3)–(6).

Step 3: Now, the fuzzy model (2) and its strict form (8) can be formulated. The strict model is solved for a specified minimum required level of portfolio return. The parameters of the triangular fuzzy number representing a fuzzy reference return are determined according to the investor's preferences. This level can be also formed based on the worst, or the best possible values of the "lower", "middle" and "upper" portfolio return achievable on the set of feasible solution X. To make this process easier for the investor, s/he can only decide on the basis of middle parameter. Other two parameters are then proportionally computed. The optimal solution of this model represents an investment portfolio composition. For user friendliness, the fuzzy portfolio characteristics can be transformed to the strict form via simple or weighted average.

3 Stock portfolio selection via fuzzy mean-variance model

The last part of the article is devoted to a real-life portfolio selection problem from the environment of the Czech capital market. To select the most suitable portfolio, the designed methodological process using fuzzy mean-variance model is applied.

3.1 Step 1: Investment strategy specification

On the Czech capital market, there are many conservative investors who (continuously) "save" their (lower) free funds. To reflect many "smaller investor" cases on the capital market, this often frequented investment strategy is reflected in this work.

For the investment, the stocks issued by well-known companies traded on the Prime Market of the Prague Stock Exchange [8] are selected, see Table 1. Frequently traded stocks with a longer history were selected. The conservative strategy requires dividend-oriented portfolio to have a "certain" return (regardless of the

capital return). Then at least 50% of the portfolio will be made by the stocks of Philip Morris, Česká spořitelna and Komerční banka. These companies steadily pay solid dividends in relation to the market price of their shares, see more [7]. The willingness to take a higher risk for a higher return is low. Such an investor is afraid of capital loss of the investment. S/he is satisfied with the gradual, modest appreciation of his/her resources over a longer time period. The purpose of such saved funds is mostly in financial security in retirement age.

3.2 Step 2: Data collection

The preselected stocks are evaluated according to two evidently most important characteristics – return and risk. Both characteristics are tracked from January 2012 to January 2020. This longer time period can server for making a real expectation about a longer time development of portfolio return and risk. The daily prices are downloaded through the web site of the Prague Stock Exchange [8]. Characteristics below are calculated from these prices. The historical period is divided to the shorter parts, i.e. month. The daily mean (average return) is computed for every month. Then, fuzzy return of each stock can be determined from these 95 observations via the formulas (3). The (co)variances, involving in expression of the investment risk, for each couple of stocks are calculated from the daily historical return for each month. The fuzzy (co)variances then are computed through the formulas (5). The levels of the fuzzy covariances can be seen in the following table (Table 1) in the form of a triplet of the parameters of the triangular fuzzy numbers. The fuzzy returns (in the same form) are displayed in the last row.

Covar	**CETV**	**ČEZ**	**ERSTE**	**KB**	**O2 C.R.**	**PF**	**VIG**	**PM**
CETV	0.16,9.14,191.12	−3.58,0.43,8.78	−27.42,0.73,9.7	−1.28,0.64,7.73	−2.87,0.37,6.94	−0.86,0.42,4.47	−1.11,0.79,6.14	−1.35,0.21,2.24
ČEZ		0.19,1.57,6.75	−1.26,0.39,5.1	−0.84,0.38,2.99	−1.65,0.23,2.62	−0.8,0.07,2.32	−1.44,0.29,4.41	−0.93,0.09,1.39
ERSTE			0.71,3.53,19.98	−1.07,0.78,6.94	−1.41,0.31,2.71	−0.87,0.2,3.27	−0.61,1.16,6.44	−1.22,0.15,1.32
KB				0.2,1.75,6.14	−1.26,0.18,2.07	−1.03,0.11,1.89	−0.62,0.4,3.15	−0.85,0.12,1.68
O2 C.R.					0.07,2.41,13.29	−2.95,0.05,1.22	−1.23,0.17,2.27	−0.93,0.08,1.01
PF						0.15,1.01,7.63	−0.78,0.15,1.99	−0.44,0.06,1.41
VIG							0.17,1.94,13.96	−0.66,0.13,1.26
PM								0.07,1.07,4.12
Return	−1.64,0.05,2.47	−0.66,−0.01,0.64	−1.12,0.06,1.02	−0.62,0.02,0.6	−1.01,0.03,2.39	−0.59,0.03,0.6	−0.7,0,0.54	−0.68,0.02,0.41

Note: ERSTE ≈ Erste Group Bank, KB ≈ Komerční banka, PF ≈ PFNonwovens, PM ≈ Philip Morris

Table 1 Stock data

3.3 Step 3: Portfolio selection

To make a portfolio, a multi-objective model (7), specified from the original fuzzy model (2), is formulated. This model is transformed to the following strict form according to (8)

$$\begin{aligned}
&\max \alpha \\
&\frac{0.069 - f_{r_p}^{l}(x)}{0.069 - (-13.492)} \geq \alpha \qquad \frac{3.529 - f_{r_p}^{m}(x)}{3.529 - 0.367} \geq \alpha \qquad \frac{57.626 - f_{r_p}^{u}(x)}{57.626 - 2.606} \geq \alpha \\
&\sum_{i=1}^{8} r_i^l x_i \geq r_p^l \qquad \sum_{i=1}^{8} r_i^m x_i \geq r_p^m \qquad \sum_{i=1}^{8} r_i^u x_i \geq r_p^u \\
&x_i \in X \qquad i = 1,2,\ldots,8 \\
&0 \leq \alpha \leq 1,
\end{aligned} \tag{11}$$

where $x_i, i = 1,2,\ldots,8$, represents a share of the i-th stock in order from the Table 1 ($i = 1 \approx$ CETV, etc.). The set X includes, of course, the non-negativity conditions for the variables and additional condition for "conservative" requirement related to a minimum general share of the Erste, Komerční banka and Philip Morris stocks. The set X can be formally written as $X = \{x_3 + x_4 + x_8 \geq 0.5, x_i \geq 0, i = 1,2,\ldots,8,\}$. On this

set, the extreme values of the partial "return" and "risk" functions are found through one-objective models as mentioned above. All models are solved by LINGO optimization software,

The required minimum (fuzzy) return is derived from the investor preferences. Under the proven assumption of "higher return-higher risk", the conservative investor determines sober values due to the minimum possible fuzzy return $(-1.381, 0.001, 0.412)$. Then the reference return is determined approximately in the middle of the difference between minimum and maximum fuzzy return formalized as $\tilde{R} = (r_p^l, r_p^m, r_p^u) = (-0.914, 0.037, 1.210)$. The optimal solution of the model (11) representing the most suitable portfolio composition then has the following form: 8.53% CETV, 28.41% Erste Group Bank, 21.59% Komerční banka, 18.54% O2 C.R. and 22.93% PFNonwovens. Erste Group Bank stock has the biggest share from the dividend-oriented stocks which are required in the portfolio at the minimum level of 50% share. The main reason is a significant diversification ability of this stock, namely with the CETV stock. Stocks issued by ČEZ, VIG and Philip Morris do not participate in the portfolio due to their inconvenient "risk-return" profile. The comparison with the resulting portfolio of a crisp mean-variance becomes interesting. Let us formulate mean-variance model (1) working with mean from daily returns and their (co)variances. Under a comparable return reference level related to the fuzzy model, the portfolio composition is slightly different: 4.93% CETV, 31.27% Erste Group Bank, 17% O2 C.R. 28.07% PFNonwovens and 18.73% Philip Morris. The main difference is the exchange of dividend-oriented stock of Komerční banka for Philip Morris stock. The main reasons are as follows. First, the diversification ability of PM stock is higher than KB stock. On the other side, this ability is instable over time which is just reflected in the fuzzy model. The second aspect lies in return. The mean of daily return is one of the greatest for Philip Morris. No wonder, that such a stock is in the portfolio under the additional requirement for the presence of strong dividend-oriented stocks. On the other side, its fuzzy return shows its greater uncertainty then by Komerční banka. Then, the fuzzy model logically prefers the stock issued by Komerční banka. The shares of other stocks in both portfolios are similar. Finally, with a different minimum required return, portfolio differences may be even more significant.

4 Conclusion

The main effort of this article is to provide a methodological approach to the investor that would significantly support making a complex satisfactory investment decision. A key (technical) instrument in this process is a fuzzy mean-variance model. It is the modification of a well-known mean variance model that reflect a typical uncertainty (instability) in the environment of the capital market more satisfactory. This significant element is reflected via the instruments and principles of fuzzy set theory. Return and also risk of the investment are represented through the triangular fuzzy numbers. This unusual modification of the Markowitz model (unlike other concepts fuzzifying only return) is significantly proven in the practical part selecting the portfolio of the stocks traded on the Prague Stock Exchange. Comparison of the created portfolio with the portfolio made by a crisp mean-variance model shows their difference just caused by the instability of both characteristics. The result then is closer to the conditions of the capital market reality.

This fuzzy approach could be further studied. Another type of fuzzy number could be applied. A trapezoidal fuzzy number, which could specify an interval of return or risk values with the same grade of membership, would be considered. Another interesting topic is the possibility of applying another known optimization principle for solving a fuzzy mathematical programming model, and the related study of other techniques for fuzzy number comparison. These themes open the door to further research beneficial for practical use.

Acknowledgements

The research project was supported by Grant No. F4/66/2019 of the Internal Grant Agency, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] BELLMAN, R. E. & ZADEH, L. A. (1970). Decision making in a fuzzy environment. *Management Science*, *17*, 141–164.

[2] DOROHONCEANU, B. & MARIN, B. (2002). A simple method for comparing fuzzy numbers. *CiteSeerX Scientific Literature Digital Library and Search Engine*.

[3] HUANG, X. (2007). Portfolio selection with fuzzy returns. *Journal of Intelligent & Fuzzy Systems*, *18*, 383–390.

[4] LAI, Y. J. & HWANG, C. L. (1992). *Fuzzy mathematical programming*. Berlin: Springer-Verlag.

[5] MARKOWITZ, H. M. (1952). Portfolio selection. *Journal of Science*, *7*, 77–91.
[6] MARKOWITZ, H. M. (1959). *Portfolio selection: efficient diversification of investments*. New York: John Wiley & Sons, Inc.
[7] MIRAS LEBL. (2020). *Dividendy*. [online] Available at: https://www.miras.cz/akcie/dividendy.php, [Accessed 30 April 2020].
[8] PRAGUE STOCK EXCHANGE. (2020). *Prime market – stocks*. [online] Available at: https://www.pse.cz/udaje-o-trhu/akcie/prime-market, [Accessed 30 April 2020].

Efficient Amount of Selected Input Using DEA Approach

Helena Brožová[1], Ivana Boháčková[2]

Abstract. This contribution aims to use Data Envelopment Analysis models for estimation of a selected input amount for that the DMU will be efficient. The idea is based on the DEA model for optimal allocation of some dual-role factor. The suggested model supposes only factors behaving themselves as inputs and limited by lower and upper bound determined by amount of these factors for all DMU. If these bounds are the same for all units then the inputs amounts are found to ensure the efficiency of all units. However, it is not true when the bounds differ for different DMUs. The suggested model is used for analysing the amount of the Gross Fixed Capital Formation of the agricultural sector in the regions of the Czech Republic within the period between 2005 and 2015. These regions are characterized by three inputs (agricultural land, annual work units, and gross fixed capital formation) and by only output (gross value added). The model deals with estimating the efficient amount of the Gross Fixed Capital Formation.

Keywords: efficiency, dual role factor, efficient amount of input, CCR model, agricultural sector, gross fixed capital formation**.**

JEL Classification: C44, C61, D24
AMS Classification: 90B50, 90C05, 90C90

1 Introduction

The Data Envelopment Analysis models (DEA) measure the efficiency of the production process of Decision-Making Units (DMU) based on their own inputs and outputs. They were firstly developed by Charnes, Cooper and Rhodes [4] and Banker, Charnes and Cooper [1] applying the Farrel's approach [7]. A typical result of these models is represented by an efficiency index showing a necessary change in monitored DMU inputs (decreasing) or outputs (increasing) for achieving virtual units as an efficient image of this DMU. A different problem is to find the selected input amount during maintaining the other parameters of the DMU at which the unit will be efficient or to find this amount for all monitored DMUs to be efficient. Cook, Green and Zhu [6] dealt with the problem of the so-called dual-role factor or its reallocation leading to the efficient DMUs. Their approach (revisiting Beasley model [2]) assumed that some factors can play both roles– input from the point of view of one DMU and output from the point of view of another DMU – simultaneously. Accordingly, its amount needs to be reduced or increased. However, their approach does not always show specific changes and, in addition, may not lead to the efficiency of all units. This approach was extended and generalized by Chen [5]. Toloo et al. [8] incorporated the imprecise data into the models with dual role factors.

In this paper, we suggest the DEA model for estimation of selected input amount for each DMU to create all DMUs efficient. Then, this model is used for estimating necessary changes of the Gross Fixed Capital Formation (GFCF) of the agricultural sector in the Czech regions to be efficient in the period 2005–2015.

This paper is organized as follows. Section 2 defines the CCR DEA model and models including the dual-role factors in DEA in brief. Section 3 describes our suggested model for evaluation of selected input to be all DMUs efficient. Section 4 contains the model application for the agricultural sector of the Czech regions and necessary changes of GFCF in the years 2005–2015. Section 5 summarizes our findings.

2 DEA models

Charnes, Cooper and Rhodes [4] introduced the model CCR to evaluate the efficiency of a set of DMUs which transforms multiple inputs into multiple outputs under the assumption of the constant returns to scale

[1] Czech University of Life Sciences, Faculty of Economics, Department of Systems Engineering, 165 00 Praha 6 – Suchdol, e-mail: brozova@pef.czu.cz.
[2] Czech University of Life Sciences, Faculty of Economics, Department of Economics, 165 00 Praha 6 – Suchdol, e-mail: bohackiv@pef.czu.cz.

(CRS). The outputs and inputs can be of various characteristics and forms and their measuring can be also difficult.

The input oriented CCR DEA model [4] measures efficiency of a DMU in time t as the maximum of a ratio of weighted outputs and weighted inputs assuming that this ratio is less than or equal to 1 for all DMUs and the efficiency of the efficient DMU is equal to 1. The efficiency Φ_H^t explicitly shows the necessary of decreasing all inputs with the same amount of inputs of an/the inefficient DMU H in time t. Linearization of this model follows

$$\Phi_H^t(\boldsymbol{x}^t, \boldsymbol{y}^t) = \sum_{j=1}^{n} u_{jH} y_{jH}^t \to MAX \qquad \text{subject to} \qquad \begin{aligned} &\sum_{i=1}^{m} v_{iH} x_{iH}^t = 1 \\ &-\sum_{i=1}^{m} v_{iH} x_{ik}^t + \sum_{j=1}^{n} u_{jH} y_{jk}^t \le 0, \\ &k \in K = \{1, 2, \dots, p\} \\ &u_{jH} \ge 0, j = 1, 2, \dots, n \\ &v_{iH} \ge 0, i = 1, 2, \dots, m \end{aligned} \tag{1}$$

where variables u_{jH} and v_{iH} are weights of outputs and inputs, y_{jk}^t is the amount of jth output from unit k, and x_{ik}^t is the amount of ith input to k^{th} unit in time t, and H is an index of the evaluated DMU (this notation is used through the whole paper).

Small Example: *This model is generally used for evaluating efficiency and setting necessary changes on inputs (or outputs) for inefficient units to increase their efficiency. As an example, you can see the following four DMUs (Table 1). The dual-role factor is seen as an input. Graphical representation of real units and their virtual efficient units is on the picture (Table 1).*

	Real DMUs				Virtual DMUs		
	Input	**Dual**	**Output**	**CCR**	**Input***	**Dual***	**Output***
DMU 1	2	4	3	**0.6**	1.2	2.4	3
DMU 2	1	1.5	2,5	**1**			
DMU 3	4	2	5	**1**			
DMU 4	4	6	5	**0.5**	2	3	5

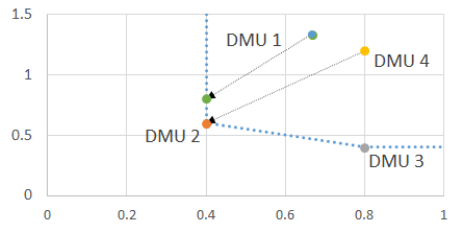


Table 1 Small example – CCR efficiency of DMUs and their virtual units

Cook, Green and Zhu [6] dealt with the evaluation of the DMUs where some factors can play an input or output role. They improve the idea of Beasley [2] that incorporated the dual-role factor into the DEA model two times – one time as input and once again as output. Again, they supposed the double incorporation of the dual role factor into criterion of the model and used the principle of complementarity in solution of the linear optimization model. The linear form of model follows:

$$\Phi^t(\boldsymbol{x}^t, \boldsymbol{y}^t) = \sum_{j=1}^{n} u_{jH} y_{jH}^t + \gamma w_H - \beta w_H \to \text{MAX} \qquad \text{subject to} \qquad \begin{aligned} &\sum_{i=1}^{m} v_{iH} x_{iH}^t = 1 \\ &-\sum_{i=1}^{m} v_{iH} x_{ik}^t + \gamma w_k - \beta w_k + \sum_{j=1}^{n} u_{jH} y_{jk}^t \\ &\le 0, \\ &k \in K \\ &u_{jH} \ge 0, j = 1, 2, \dots, n \\ &v_{iH} \ge 0, i = 1, 2, \dots, m \\ &\gamma \ge 0, \beta \ge 0 \end{aligned} \tag{2}$$

where variables γ and β are weights of dual-role factor and w_k is the amount of dual-role factor of unit k.

This model serves for dividing among the DMUs. The first group $K1$ of DMUs considers this factor as the input (if $\gamma - \beta \leq 0$), the second $K2$ as output (if $\gamma - \beta \geq 0$), and the third $K3$ considers it on equilibrium or optimal level (if $\gamma - \beta = 0$). $K = K1 + K2 + K3$.

Small Example: *Table 2 shows which of the DMUs from the small example above has to increase or decrease the amount of the dual-role factor or leave its initial amount. The results of model (2) show that the DMU 3 is found on its equilibrium due to the value of the dual-role factor, the DMU 1 and DMU 4 efficiency is improved by reducing the value of the dual role factor and the DMU 3 can use more of the dual-role factor (the last column of Table 2).*

	Efektivita	u1	beta	gama	v1	gama – beta	K1/K2/K3
DMU 1	**1**	0.5	0	0.18	09	0.18	**2**
DMU 2	**1**	1	0	0	0.4	0	**3**
DMU 3	**1**	0.25	0.5	0	0.4	–0.5	**1**
DMU 4	**0.77**	0.25	0	09	05	09	**2**

Table 2 Small example – solution of model (2)

Then, the following model [6] is used for optimal allocation of the dual-role factor under the condition that the whole amount of the factor has to be exploited. The nonlinear form of this model follows:

$$\begin{aligned}
\Phi^t(\boldsymbol{x}^t, \boldsymbol{y}^t) = \sum_{k \in K1} \left(\sum_{j=1}^{n} u_j y_{jk}^t + \alpha f_k \right) + \\
+ \sum_{k \in K2} \left(\sum_{j=1}^{n} u_j y_{jk}^t - \alpha f_k \right) + \\
+ \sum_{k \in K3} \sum_{j=1}^{n} u_j y_{jk}^t \to \text{MAX}
\end{aligned}
\quad \text{subject to} \quad
\begin{aligned}
\sum_{k \in K} \sum_{i=1}^{m} v_i x_{ik}^t = 1 \\
-\sum_{i=1}^{m} v_i x_{ik}^t + \alpha f_k + \sum_{j=1}^{n} u_j y_{jk}^t \leq 0, k \in K1 \\
-\sum_{i=1}^{m} v_i x_{ik}^t - \alpha f_k + \sum_{j=1}^{n} u_j y_{jk}^t \leq 0, k \in K2 \\
-\sum_{i=1}^{m} v_i x_{ik}^t + \sum_{j=1}^{n} u_j y_{jk}^t \leq 0, k \in K3 \\
\sum_{k \in K1 \cup K2} f_k = \sum_{k \in K1 \cup K2} w_k \\
w_k \leq f_k \leq w_k^U \\
w_k^L \leq f_k \leq w_k \\
u_j \geq 0, j = 1, 2, \dots, n \\
v_i \geq 0, i = 1, 2, \dots, m \\
\alpha \geq 0
\end{aligned}
\tag{3}$$

where variable α is weight of the dual-role factor, variable f_k is the amount of the dual-role factor of unit k and w_k^L, w_k^U are the lower and upper bound of the dual-role factor of unit k value (this notation is used through the whole paper). Because of product of variables αf_k, this model is not linear. It could be linearized using the change of variables $\alpha f_k = \delta_k$.

The results of this model show the new allocation of total amount of the dual-role factor into any given DMU resulting in maximum efficiency of all DMUs.

Small Example: *This small example model has optimal solution with zero values of variables* δ_k *and* α. *It means there is no reasonable re-allocation of the dual-role factor to increase the efficiency of all DMUs while maintaining the overall value of the dual-role factor. Unfortunately, such the solution of model (3) is quite common.*

3 Efficient value of selected input

We were inspired by the model (3) and suggested a method for finding an optimal amount of one selected input $f_k, k \in K$ so that all DMUs could be efficient. The proposed model follows:

$$\Phi^t(\boldsymbol{x}^t, \boldsymbol{y}^t) = \sum_{k\in K}\sum_{j=1}^{n} u_j y_{jk}^t \to \text{MAX} \qquad \text{subject to} \qquad \begin{aligned} &\sum_{k\in K}\left(\sum_{i=1}^{m} v_i x_{ik}^t + \alpha f_k\right) = 1 \\ &-\sum_{i=1}^{m} v_i x_{ik}^t - \alpha f_k + \sum_{j=1}^{n} u_j y_{jk}^t \le 0, k \in K \\ &w_k^L \le f_k \le w_k^U \\ &u_j \ge 0, j = 1, 2, \dots, n \\ &v_i \ge 0, i = 1, 2, \dots, m \\ &\alpha \ge 0, f_k \ge 0, k \in K \end{aligned} \tag{4}$$

The model is linearized using the new variables $\delta_k = \alpha f_k$ which has to be of the following form:

$$\Phi^t(\boldsymbol{x}^t, \boldsymbol{y}^t) = \sum_{k\in K}\sum_{j=1}^{n} u_j y_{jk}^t \to MAX \qquad \text{subject to} \qquad \begin{aligned} &\sum_{k\in K}\left(\sum_{i=1}^{m} v_i x_{ik}^t + \delta_k\right) = 1 \\ &-\sum_{i=1}^{m} v_i x_{ik}^t - \delta_k + \sum_{j=1}^{n} u_j y_{jk}^t \le 0, k \in K \\ &\alpha w_k^L \le \delta_k \le \alpha w_k^U \\ &u_j \ge 0, j = 1, 2, \dots, n \\ &v_i \ge 0, i = 1, 2, \dots, m \\ &\alpha \ge 0, \delta_k \ge 0, k \in K \end{aligned} \tag{5}$$

Lemma 1

Both the models (5) and (4) have always an optimal solution.

Proof. All constraints of the model (5) represent a bounded set of feasible solutions of the model (5) and, therefore, the linear objective function is on its maximum. For the model (4), input $f_k = {}^{\delta_k}/_{\alpha}$ is enough.

Lemma 2

There are such the lower and upper bounds w_k^L, w_k^U that the optimal solution of the model (5) has value $\Phi^t(\boldsymbol{x}^t, \boldsymbol{y}^t) = 1$.

Proof. If $\Phi^t(\boldsymbol{x}^t, \boldsymbol{y}^t) = 1$, then for all DMUs is $\sum_{i=1}^{m} v_i x_{iq}^t + \alpha f_q = \sum_{j=1}^{n} u_j y_{jq}^t$. Because $w_k^L \le f_k \le w_k^U$, it has to be $\alpha w_k^L \le \delta_k \le \alpha w_k^U$ and also $\alpha w_k^L \le -\sum_{i=1}^{m} v_i x_{ik}^t + \sum_{j=1}^{n} u_j y_{jk}^t \le \alpha w_k^U$.

After reformulation, we receive $w_k^L \le \frac{-\sum_{i=1}^{m} v_i x_{ik}^t + \sum_{j=1}^{n} u_j y_{jk}^t}{\alpha} \le w_k^U$ and, then, it is possible to find the values of the lower and upper limits for these inequalities.

Lemma 3

All DMUs are efficient only in the case that the optimal solution of the model (5) is $\Phi^t(\boldsymbol{x}^t, \boldsymbol{y}^t) = 1$.

Proof. A) For an efficient DMU, the following relation is valid: $\sum_{i=1}^{m} v_i x_{ik}^t + \alpha f_k = \sum_{j=1}^{n} u_j y_{jk}^t$.

From the constraint of aggregated inputs $\sum_{k\in K}(\sum_{i=1}^{m} v_i x_{ik}^t + \alpha f_k) = 1$ follows the following relation for aggregate outputs $\sum_{k\in K}\sum_{j=1}^{n} u_j y_{jk}^t = 1$.

B) Supposing the optimal aggregate efficiency (outputs) $\sum_{k\in K}\sum_{j=1}^{n} u_j y_{jk}^t = 1$ then, of course, the feasible solution the constraint $\sum_{k\in K}(\sum_{i=1}^{m} v_i x_{ik}^t + \alpha f_k) = 1$ is fulfilled.

Supposing the inefficient DMU p, then, the following relation is valid: $\sum_{i=1}^{m} v_i x_{ip}^t + \alpha f_p > \sum_{j=1}^{n} u_j y_{jp}^t$. In this case, there is some DMU q for which the following relation is valid: $\sum_{i=1}^{m} v_i x_{iq}^t + \alpha f_q < \sum_{j=1}^{n} u_j y_{jq}^t$. Such efficiency DMU would be greater than 1 and which is impossible.

The procedure for finding values of the selected input ensuring efficiency of all units consists of the following steps:

1. For the selected input k choose the values w_k^L, w_k^U arbitrary;
2. Create the model (5) for input k incorporating (??) all units;
3. Find an optimal solution;
4. If the optimal value $\Phi^t(\boldsymbol{x}^t, \boldsymbol{y}^t) < 1$,
   - then set some new higher upper bounds and return to Step 3,
   - otherwise, continue to Step 5.
5. Calculate optimal values of the input k as $f_k = {}^{\delta_k}/_{\alpha}$

Small Example: *This approach is shown on the small example for setting necessary values of the dual-role factor when it is considered as the input for all units. The lower bound was set as minimum of all values and upper bound as their maximum. The DMU 2 can use more than initial amount of an interesting factor while other units have to reduce use of this factor. The values of dual-role factor that all units are efficient are in the column Dual* in Table 3. Graphical representation is on the picture.*

	Real DMU			New DMU			
	Input	**Dual**	**Output**	**Input**	**Dual***	**Change**	**Output**
DMU 1	2	4	3	2	**3**	–25%	3
DMU 2	1	1.5	2.5	1	**6**	–300%	2.5
DMU 3	4	2	5	4	**1.5**	–25%	5
DMU 4	4	6	5	4	**1.5**	–75%	5

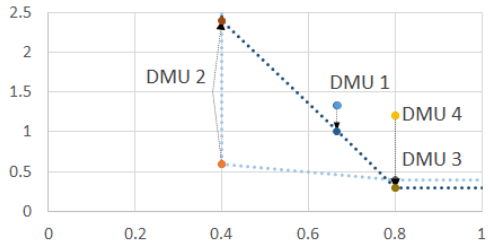


Table 3 Small example – efficient values of selected input

4 Efficient Amount of Gross Fixed Capital Formation of agricultural sector in regions of the Czech Republic

GFCF or investment is a very important factor of production that affects economic results and efficiency of production process. In agriculture, its value very often depends not only on the past economic results of companies but also on their ability to obtain subventions. Their investments, even considering their longer-term impact, often do not correspond to the results because the obtained subventions had to be spent.

Using the model (5) and procedure for obtaining the values of the selected input, we calculated the efficient values of GFCF for the agricultural sector in the CZ regions during the period 2005–2015. For this analysis, we used data from Eurostat databases and yearbooks of the Czech regions. Comparable basic factors of production, namely the area of agricultural land (AL), labour as annual work units (L), and gross fixed capital formation (GFCF) were chosen as inputs while gross value added (GVA) created in the agricultural sector was chosen as output (**Table 4**).

		2005	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
AL (ha)	Min	25797	25748	25662	25537	25458	25432	25349	25257	25138	25043	25020
	Median	450750	450451	450402	450342	450374	450283	450512	450383	450359	450536	450717
	Mean	493350	493111	492885	492651	492449	492211	492076	491879	491695	491554	491452
	Max	971984	971378	970986	970279	969861	969426	969084	968644	967946	967462	967208
L (AWU)	Min	3195	2862	1868	2004	2500	2715	3021	3382	3005	3088	3490
	Median	13477	12008	11721	11689	11819	11677	11970	11689	11597	11574	10674
	Mean	12660	12142	11852	11646	11471	11187	11576	11422	11473	11511	11094

		2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
	Max	22682	23604	22315	24263	23606	22124	23567	23064	23520	23794	23066
GFCF (CZK mil.)	Min	151	191	336	178	295	376	561	742	749	1090	811
	Median	1454	1483	1937	2011	1538	1892	1841	3060	2635	2281	2142
	Mean	1426	1529	2007	1855	1518	1761	2069	3143	2648	2666	2199
	Max	2494	4203	4248	4233	3063	3613	4290	5828	5258	5055	4101
GFA (CZK mil.)	Min	1673	1723	1671	1428	1511	1669	2214	2578	2528	2895	2654
	Median	5149	5448	5373	5264	4696	4005	5898	6279	6364	7298	7132
	Mean	5156	5273	5385	5557	4609	4301	6189	6730	6826	7623	7376
	Max	9118	9667	9404	10705	8321	8749	12522	14155	13725	15293	14837

Table 4 Basic information about data for all 14 regions

The results of the CCR model (1) show that the Prague region is rated by the model DEA significantly better than other regions in all periods. Therefore, the Prague region was excluded from further evaluation [3].

The value of the GFCF is significantly dependent on the size of subsidies and, from this point of view, this factor could be perceived as a factor of success (output). However, it is an input factor because future production undoubtedly depends on investments however not only in a given year but also in subsequent years.

Our aim is finding out the most appropriate level of investment (GFCF) in each region. Firstly, we tried to find the most appropriate reallocation of GFCF in each year between regions based on the models (2) and (3). For five years (2008, 2010–2012, 2015), however, the result of the model (2) suggests a way of change while the model (3) cannot find it out. The results, obviously indicate that in 31 cases DMU/year suppose increasing their efficiency by rising GFCF (pink) and only 16 ones reached it while in 100 cases by decreasing GFCF (yellow) and only 26 reached it. Balanced GFCF is shown in 12 cases.

Efficiency	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Min	0.79	0.72	0.79	0.54	0.80	0.84	0.72	0.73	0.73	0.84	0.82
Median	1	0.87	0.94	0.80	0.97	0.89	0.91	0.90	0.90	0.94	0.93
Average	0.94	0.88	0.93	0.81	0.93	0.91	0.88	0.90	0.89	0.94	0.92
Data changes (%)											
Středočeský	30	−12.8	0	0	−26.3	0	0	0	0	30	0
Jihočeský	30	0	0	0	0	0	0	0	0	−20.7	0
Plzeňský	0	0	0	0	0	0	0	0	0	−30	0
Karlovarský	30	30	0	0	0	0	0	0	0	0	0
Ústecký	30	30	30	0	30	0	0	0	0	−30	0
Liberecký	30	0	30	0	30	0	0	0	0	−30	0
Královéhradecký	−30	0	0	0	30	0	0	0	8.9	0	0
Pardubický	0	0	−30	0	0	0	0	0	0	30	0
Vysočina	−30	0	−13.9	0	0	0	0	0	0	−30	0
Jihomoravský	−15.8	0	30	0	0	0	0	0	0	0	0
Olomoucký	0	0	0	0	−30	0	0	0	0	30	0
Zlínský	30	0	−30	0	28	0	0	0	30	30	0
Moravskoslezský	−30	30	30	0	30	0	0	0	−30	30	0

Table 5: Efficiency and suggest changes of GFCF as dual-role factor for all regions and all years

Secondly, we used the proposed model (5) and proposed procedure to obtained results for all regions and all years. Correctness of the results was tested by different methods for solving the linear optimization model, Solver – simplex algorithm, Solver – gradient algorithm, and Open Solver. We followed the suggested steps.

Step 1. The lower and upper bounds of GFCF for each year were set as the minimum and maximum of the GFCF values of 13 analyzed regions (excluding Prague) for each year (**Table 6**).

Step 2 and 3. The model (5) was created and solved for each year.

Step 4. Because optimal values of objective function are equal to 1, we received immediately the optimal amount of GFCF for the agricultural sector in each region for each year excepting year 2014. For this year, we needed to increase the upper bound from 5055 to 5300 to ensure all DMUs would be efficient (**Table 6**).

		2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
GFCF (CZK mil.)	Min – w_k^L	293	221	364	353	295	376	561	742	749	1090	811
	Max – w_k^U	2494	4203	4248	4233	3063	3613	4290	5828	5258	5055/**5300**	4101

Table 6 Minimum and maximum of GFCF Except the Prague region)

Step 5. The optimal values of GFCF were calculated. **Table 7** shows the percentual change of amount of GFCF and the basic information of new values of GFCF (increasing (pink) or decreasing (yellow)). The agricultural sectors in all regions are efficient.

Change of GFCF (%)	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Středočeský	–6.5	–8.4	1.8	–37.5	–71.7	–3.5	–12.4	0	–17.7	3.7	–2.4
Jihočeský	13.7	92.2	19	–73.6	–52.5	–41.5	–17.3	–36.9	–44.6	11.6	5
Plzeňský	–39.4	43.4	26.3	–53.9	–55.3	–25.5	–42.3	–78.8	–72.4	–19.7	–34.1
Karlovarský	70.8	585.9	217.3	–24.2	–29.7	152.1	0	171.5	0	88	57.9
Ústecký	–52	–61.6	–18.5	36.3	74.3	–39.3	–14.7	–6.9	9	–9.6	–8.2
Liberecký	64.8	–38.7	0	8.7	0	–11.5	–26.1	43.3	–8.2	–19.6	–5
Královéhradecký	–20.8	134	57.7	–81.3	–41.5	–59.1	–37.2	–21.1	–2.8	50.6	49.4
Pardubický	61.9	53.6	35.6	–48.9	–25.8	–81.4	–28.5	–56.9	–71.7	–47.4	–57.7
Vysočina	–21	26.1	–4.8	–45.3	–90.4	–63.9	–53.2	–75.5	–62.7	–70.7	–73.7
Jihomoravský	7	76.8	54	60.6	30.7	29.4	0	–33.3	0	14.2	14.3
Olomoucký	30.3	–6.2	27.2	53.9	–47.4	–58	–16.8	–49	–8.4	23.8	–11.6
Zlínský	9.2	8.2	45.1	0	–21.4	–1	13.1	–5.9	14.8	42.9	–7.8
Moravsko–slezský	–78.2	202.5	49	69.2	2.6	32.5	24.5	40.9	22.8	63.3	47.8
Average change	3,1	85,2	39,2	–10,5	–25,2	–13,1	–16,2	–8,4	–18,6	10,1	–2,0
Efficient GFCF values (CZK mil.)											
Median	1542.56	2646.53	2827.88	1096.78	848.41	1037.78	1465.55	1913.16	2054.43	2370.76	1731.14
Mean	1447.63	2448.66	2701.72	1589.90	979.98	1395.72	1759.58	2342.28	2094.60	2812.31	2120.58
Average change	–76.06	817.12	566.57	–394.57	–619.79	–471.98	–410.80	–926.49	–660.79	50.38	–162.96

Table 7: Changes of GFCF for all regions and all years

Initially after EU accession, to the companies received more subsidies (by 85.2% in 2005, by 39.2% in 2006). After the crisis, companies learned how to draw and use subsidies but for their efficiency it would have been enough to get 10%, 25%, etc. less subsidies.

5 Conclusion

The efficient values of inputs and outputs represent a guide for an individual DMU to improve its production capabilities. The changes of some parameters are not often really possible so the calculation of a possible change of only selected parameter while maintaining others is important. This article shows the calculation for changing the value of the selected input with the so that all DMUs would be efficient. This model based

on an input-oriented DEA model can be extended on a problem with more than 1 selected input. Analogously, it is possible to use this model to change the selected output which will be based on an output-oriented DEA model.

References

[1] BANKER, R. D., CHARNES, R. F. & COOPER, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science* **30**, 1078–1092.

[2] BEASLEY, J. (1995). Determining teaching and research efficiencies. *Journal of the Operational Research Society*, **46**, 441–452.

[3] BROŽOVÁ, H., BOHÁČKOVÁ, I. (2016). Post-crisis development in the agricultural sector in the CR – Evidence by Malmquist index and its decomposition. *Mathematical Methods in Economics 2016*, TUL, Liberec.

[4] CHARNES, A., COPPER, W., & RHODES, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research* **2**, 429–444.

[5] CHEN, W.-CH. (2014). Revisiting dual-role factors in data envelopment analysis: derivation and implications, *IIE Transactions*, **46**, 653–663.

[6] COOK, W. D., GREEN, R. H., & ZHU, J. (2006). Dual-role factors in data envelopment analysis, *IIE Transactions*, **38**, 105–115.

[7] FARREL, M. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society, Series A* **120**, 253–281.

[8] TOOLO, M., KESHAVARZ, E., & HATAMI-MARBINI, A. (2017). Dual-Role Factors for Imprecise Data Envelopment Analysis, *Omega,* **77**, 15–31.

Unemployment and Job Vacancies

Andrea Čížků[1]

Abstract. The paper formulates, econometrically estimates and analyzes a model of unemployment and job vacancies in Spain and the United Kingdom of Great Britain and Northern Ireland before and after the current economic crisis. These two fundamentally different European economies were chosen in order to stress differences between advanced European economies and less developed countries in southern Europe. Job creation process in the formulated model depends on aggregate output demand and the goal of the paper is to assess the strength of this aggregate demand transmission mechanism in these two different European economies. Equilibrium analysis is also performed and a connection between the strength of aggregate demand transmission mechanism and a multiplicity of equilibrium unemployment rate is analyzed and discussed as well.

Keywords: unemployment rate, vacancies, aggregate output demand

JEL Classification: E24, J23, J64
AMS Classification: 91G70

1 Introduction

Stiglitz [17] pointed out that many economies are suffering from high unemployment and weak aggregate output demand. The hypothesis that aggregate output demand plays an important role in determining unemployment has a long tradition in post-Keynesian economics (Holt, Pressman [12], Lavoie [14], Diamond [6]). Michaillat a Saez [15] and Vujčić et al. [19] discuss a relationship between unemployment and aggregate output demand with a conclusion that unemployment is driven mainly by aggregate demand shocks. Importance of aggregate output demand on determining job vacancies has been analyzed and confirmed by both empirical and theoretical papers (Benigno, Fornaro [1], Cynamon, Fazzari [3], Čížek [5], Farmer [8], Heathcote, Perri [11]).

The goal of this paper is to contribute to this literature analyzing the link between unemployment and output demand, which is a topic highly discussed during the current economic crisis in daily economic news as well as in professional economic literature. Specifically, the paper formulates, estimates and analyzes empirically oriented model of the labor market aiming to incorporate aggregate output demand mechanism. This paper focuses on aggregate demand channel as an important factor of unemployment. Other factors influencing labor markets are discussed and analyzed by Bílková [2], Čabla, Malá [4], Flek & Mysíková [9], Večerník [18], Zdeněk & Střeleček [20].

The formulated model is applied to two fundamentally different economies – the Spain and the United Kingdom of Great Britain and Northern Ireland. The reason why these two fundamentally different economies were chosen is to point out differences between advanced European economies and less developed countries in southern Europe.

The structure of the paper is as follows. Firstly, the model is formulated in chapter 2. Then the data is described in chapter 3. Econometric estimation is discussed in section 4. This chapter discusses and analyzes the strength of the transmission mechanism describing the influence of aggregate output demand on creating job vacancies. Equilibrium analysis with the estimated model is performed in chapter 5. In this chapter, a connection between the strength of the aggregate output transmission mechanism and a multiplicity of equilibrium unemployment rate is analyzed. The final section 6 concludes.

2 Model

Shimer's [16] methodology is applied to (1) calculate job-finding and separation rates, (2) to describe dynamics of unemployment. Job-finding probability is calculated according to the relation

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 130 67, Prague 3, Czech Republic, andrea.cizku@vse.cz.

$$F_t = 1 - \frac{U_{t+1} - U_{t+1}^s}{U_t}, \tag{1}$$

where U_{t+1} is number of unemployed at the beginning of time period $t+1$, U_{t+1}^s is the number of short-term unemployed (duration of their of unemployment is shorter than one period) at the beginning of time period $t+1$.

The corresponding job-finding rate is $f_t = -\ln(1 - F_t)$. Dynamics of unemployment in Shimer's methodology is given by

$$U_{t+1} = \frac{s_t}{s_t + f_t} \cdot \left(1 - e^{-(s_t + f_t)}\right) \cdot L_t + e^{-(s_t + f_t)} \cdot U_t, \tag{2}$$

where E_t is number of employed, $L_t \equiv E_t + U_t$ represents labor force, s_t stands for separation rate, which is obtained by numerically solving the equation (2).

Matching vacancies with unemployed workers is described by Cobb-Douglas aggregate matching function

$$M_t \equiv M(V_t, U_t) = A \cdot U_t^{\alpha} \cdot V_t^{\beta}, \; A > 0, \; \alpha, \beta \in (0,1), \tag{3}$$

where U_t is number of unemployed, V_t represents number of vacancies, M_t is number of "matches" of unemployed workers with vacant jobs.

Probability of finding a job F_t is given by

$$F_t \equiv \frac{M_t}{U_t} = A \cdot U_t^{\alpha - 1} \cdot V_t^{\beta}. \tag{4}$$

Log-linearizing and adding i.i.d. random error ε_t^F for a purpose of econometric estimation leads to

$$\ln F_t = \ln A + (\alpha - 1) \cdot \ln U_t + \beta \cdot \ln V_t + \varepsilon_t^F. \tag{5}$$

Formulating a model for a creation of vacant jobs is motivated by Keynesian concept of aggregate output demand which can be summarized as follows

$$\boxed{\begin{array}{l} \uparrow \text{ unemployment } \rightarrow \downarrow \text{ purchasing power } \rightarrow \downarrow \text{ output demand} \\ \downarrow \text{ labor demand } \rightarrow \downarrow \text{ job vacancies} \end{array}}. \tag{6}$$

This transmission mechanism is approximated by a linear function of the form

$$V_t = a - b \cdot U_t + \varepsilon_t^V, \tag{7}$$

where $a, b > 0$ and ε_t^V is i.i.d. random error.

3 Data

Quarterly seasonally unadjusted data for Spain and the United Kingdom were obtained from the Eurostat database.[2] The data used for econometric estimation is the number of unemployed U_t (thousand persons), number of vacant jobs V_t (in thousands)[3] and a job-finding probability F_t. Probability of finding a job was calculated using the relation (1). Empirical counterpart to the number of short-term unemployed U_t^S is the number of unemployed workers whose spell of unemployment is shorter than one quarter (three months).[4]

4 Econometric Estimation

Econometric estimation was performed using seasonally unadjusted data as seasonal adjustments might lead to impure autocorrelation in the estimated regression equations. Seasonality is explicitly modeled by dummy binary variables Q_{2t}, Q_{3t} and Q_{4t} taking a value of 1 in a given quarter and zero otherwise.

2 http://ec.europa.eu/eurostat/data/database

3 The name of the corresponding time series in Eurostat database is 'Job vacancy statistics by NACE Rev. 2 activity – quarterly data (from 2001 onwards) [jvs_q_nace2]'. The item SIZECLASS was set to 'total' and the item NACE_R2 was set to 'Industry, construction and services (except activities of households as employers and extra-territorial organizations and bodies)'.

4 This time series can be found in the Eurostat database under the name 'Unemployment by sex, age and duration of unemployment [lfsq_ugad]'. The item DURATION was firstly set equal to 'less than one month', then it was set equal to 'from one to two months' and finally these two-time series were added together.

The Chow stability test confirmed structural break at time period 2009Q1 on 1% level of significance in both countries (Spain and United Kingdom). Estimated regression equations (5) and (7) were therefore modified as follows

$$\ln F_t = \left[\ln A + (\alpha - 1)\cdot \ln U_t + \beta \cdot \ln V_t\right] + \left[\ln A_2 \cdot P_{2t} + (\alpha - 1)_2 \cdot \ln U_t \cdot P_{2t} + \beta_2 \cdot \ln V_t \cdot P_{2t}\right] + \\ + \beta_2^F \cdot Q_{2t} + \beta_3^F \cdot Q_{3t} + \beta_4^F \cdot Q_{4t} + \varepsilon_t^F, \tag{8}$$

$$V_t = \left[a - b \cdot U_t\right] + \left[a_2 \cdot P_{2t} - b_2 \cdot U_t \cdot P_{2t}\right] + \left[\beta_2^V \cdot Q_{2t} + \beta_3^V \cdot Q_{3t} + \beta_4^V \cdot Q_{4t}\right] + \varepsilon_t^V, \tag{9}$$

where $P_{2t} = \begin{cases} 1, \text{ for } t \geq 2009\ Q1 \\ 0, \text{ otherwise} \end{cases}$, $Q_{it} = \begin{cases} 1, \text{ if } t \text{ corresponds to } i\text{ - th quarter} \\ 0, \text{ otherwise} \end{cases}$, $i = 2,3,4.$

Ramsey RESET test of the correct specification was also performed and the null hypothesis of correct specification was not rejected for all cases on standard 5% level of significance.

Regression equations (8)–(9) were estimated by ordinary least squares. Autocorrelation of random errors was confirmed on standard 5% level of significance using the Breusch-Godfrey Lagrange Multiplier test. Therefore, standard errors of estimated parameters were recalculated by Newey-West method (HAC – Heteroscedasticity and Autocorrelation Consistent Standard Errors). Results are summarized in the following tables.[5]

		Cobb-Douglas matching function									
country	**sample**	$\ln A$	$\ln A_2$	$\alpha - 1$	$(\alpha - 1)_2$	β	β_2	β_2^F	β_3^F	β_4^F	R^2
Spain	2001Q1–2019Q3	−0.21	7.34***	−0.63***	−0.31	0.91***	−1.04***	0.08**	−0.10***	−0.18***	0.88
United Kingdom	2001Q2–2019Q2	4.44*	−6.32***	−0.79***	0.57***	0.07	0.31	−0.01	0.16***	−0.03	0.87

Table 1 Econometric estimates of Cobb-Douglas matching function (8)

		Job-creation function							
country	**sample**	a	a_2	b	b_2	b_2^V	b_3^V	b_4^V	R^2
Spain	2001Q1–2019Q3	154.7***	–	0.03***	−0.02***	0.12	−0.35	−3.97	0.31
United Kingdom	2001Q2–2019Q2	730.9***	478.3***	0.09	0.22***	28.5***	63.8***	21.3***	0.92

Table 2 Econometric estimates of the job-creation function (9)

Estimated effectiveness of the matching process before the crisis is higher in United Kingdom $\hat{A}^{UK} = \exp(4.44) = 84.8$ than in Spain $\left(\hat{A}^{ESP} = \exp(-0.21) = 0.81\right)$. Nonetheless, this effectiveness rises in Spain and decreases in United Kingdom after the crisis on 1% level of significance.

Estimation of the parameter $1 - \alpha$ is also statistically significant on 1% level and the assumption $\alpha \in (0,1)$ is also satisfied in both countries. The coefficient $1 - \alpha$ has an interpretation of relative elasticity and gives us a percentage change in probability of finding a job F_t caused by an increase in number of unemployed U_t by 1%. For a periods before the crisis, the probability of finding a job F_t decreases by 0.63% in Spain and by 0.79% in United Kingdom in response to 1% increase in U_t. After the crisis, this elasticity decreases (in absolute value) in United Kingdom on 1% level of significance.

The influence of job vacancies V_t on probability of finding a job F_t is measured by the parameterβ, which turned out to be statistically significant on 1% level in Spain before the crisis. In this case, an increase in V_t by 1% causes an increase in F_t by 0.91%. After the crisis, this relative elasticity decreases dramatically in Spain, which is confirmed on 1% level of significance. In the United Kingdom, the Wald test refused the zero

[5] Symbols *, ** and *** represent statistical significance of parameters on 10%, 5% and 1%, respectively.

hypothesis $H_0: \beta^{UK} + \beta_2^{UK} = 0$, so this relative elasticity became statistically important on 1% level of significant after the crisis.

The parameter a_2 was not estimated in the case of Spain as it turned out statistically insignificant. Moreover, estimation of this parameter caused statistical insignificance of other parameters in the regression (9). The parameter b (and b_2) plays an important role in assessing the strength of the aggregate output transmission mechanism (6). In Spain, the estimated value of this parameter is rather low $\hat{b}^{ESP} = 0.03$. Nonetheless, it is statistically significant even on 1% level of significance. It follows from this estimate $\hat{b}^{ESP} = 0.03$ that the number of unemployed U_t would have to have to decrease roughly by $\frac{1}{0.33} \doteq 33$ in order to create 1 new vacant job. Moreover, the strength of the mechanism (6) decreased in Spain after the crisis, which is confirmed even on 1% level of significance. So the output demand mechanism (6) plays rather a minor role in Spain.

The mechanism (6) plays quantitatively more important role in the United Kingdom as the estimated parameters are $\hat{b}^{UK} = 0.09, \hat{b}_2^{UK} = 0.22$. The output demand mechanism became more relevant after the crises and the estimate $\hat{b}^{UK} + \hat{b}_2^{UK} = 0.31$ is statistically significant according to the Wald test even on 1% level of significance. This estimate also indicates that the number of unemployed U_t would have to decrease roughly by $\frac{1}{0.31} \doteq 3$ in order to create 1 new vacant job.

5 Equilibrium Analysis

Seasonal factors will now be neglected as they are not important for a purpose of equilibrium analysis. Lower bound $\underline{V}$ for the variable V_t is introduced, which is set as the minimum value in the observed data. Specifically, $\underline{V} = 54$ for Spain and $\underline{V} = 436$ for the United Kingdom. Econometric estimation of regression equations (8)–(9) under these modifications is given by

Spain:

$$F(U) = e^{-0.21} \cdot U^{-0.63} \cdot V(U)^{0.91} \cdot \left(e^{7.34}\right)^{P_2} \cdot \left(U^{-0.31}\right)^{P_2} \cdot \left(V(U)^{-1.04}\right)^{P_2} \tag{10}$$

$$V(U) = \max\{154.7 - 0.03 \cdot U + 0.02 \cdot U \cdot P_2; 54\} \tag{11}$$

United Kingdom:

$$F(U) = e^{4.44} \cdot U^{-0.79} \cdot V(U)^{0.07} \cdot \left(e^{-6.32}\right)^{P_2} \cdot \left(U^{0.57}\right)^{P_2} \cdot \left(V(U)^{0.31}\right)^{P_2} \tag{12}$$

$$V(U) = \max\{730.9 - 0.09 \cdot U + 478.3 \cdot P_2 - 0.22 \cdot U \cdot P_2; 436\} \tag{13}$$

Time index t was omitted in relations (10)–(13) as time is not relevant when analyzing equilibrium. A commonly used symbol of *hat* for fitted values is also omitted in these relations for a purpose of clarity of exposition. On the contrary, it is made explicit in relations (10)–(13) that the probability of finding a job $F(U)$ and also the number of job vacancies $V(U)$ are both functions of number of unemployed U.

Equilibrium unemployment rate satisfies $U \equiv U_t = U_{t+1}$. Using this condition in equation (2) yields

$$U = \frac{s}{s + f(U)} \cdot L, \tag{14}$$

where $f(U) = -\ln(1 - F(U))$ is the job-finding rate which corresponds to the job-finding probability *F(U)*

s represents the separation rate and L stands for the labor force – both are treated as constants within the analysis of equilibrium.[6]

The relation (14) can be expressed alternatively in terms of unemployment rate

[6] Arithmetic mean of separation rate and labor force is used for s and L for both countries. Performed analysis of equilibrium distinguishes between the period before the crisis (2001Q2–2008Q4) and after the crisis (2009Q1–2019Q2). The corresponding mean values of separation rate and labor force are:

Before the crisis: $\left(s_{ESP}^{2001-2008} = 0.05,\ s_{UK}^{2001-2008} = 0.03\right)$, $\left(L_{ESP}^{2001-2008} = 20656,\ L_{UK}^{2001-2008} = 29561\right)$

After the crisis: $\left(s_{ESP}^{2009-2019} = 0.07,\ s_{UK}^{2009-2019} = 0.03\right)$, $\left(L_{ESP}^{2009-2019} = 22916,\ L_{UK}^{2009-2019} = 31528\right)$

$$u = \frac{s}{s + f(u \cdot L)}, \tag{15}$$

where $u = \frac{U}{L}$ represents equilibrium unemployment rate.

Function $\frac{s}{s+f(u \cdot L)}$ is depicted together with 45° line representing the variable u on the left-hand side of the equation (15) on the following figure 1.

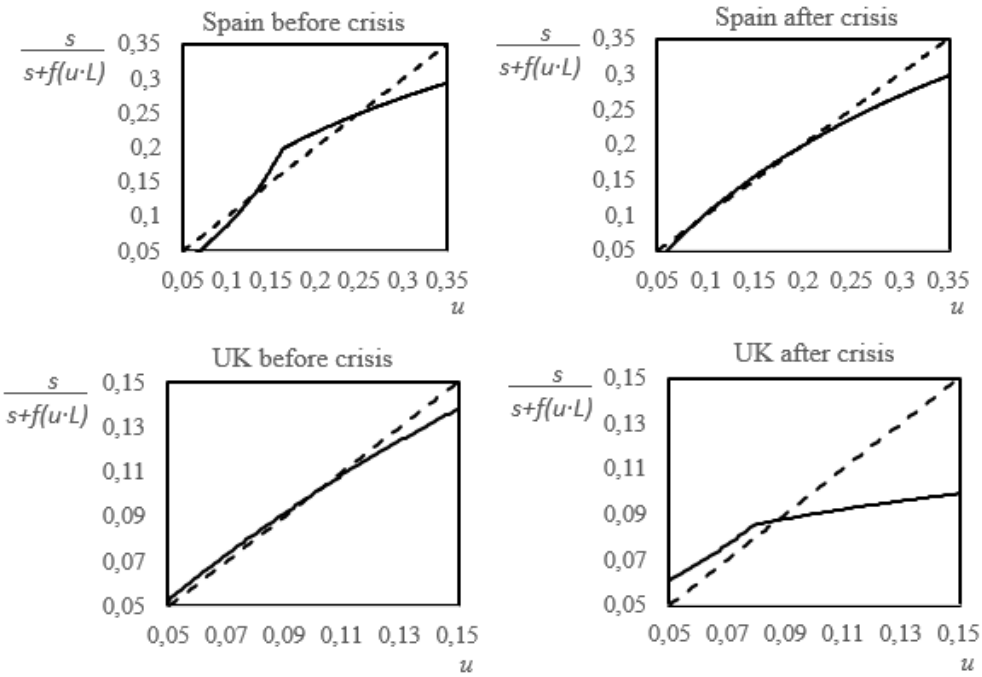


Figure 1 Equilibrium unemployment rates in Spain and United Kingdom before (2001Q2–2008Q4) and after (2009Q1–2019Q2) the crisis

The figure illustrates that there were two equilibrium unemployment rates in Spain before the crisis (13% and 24.5%). This finding is in line with conclusions of other empirical studies (Čížek [5]). After the crisis, there are still two equilibrium unemployment rates in Spain (9% and 19.5%). Nonetheless, the function values $\frac{s}{s+f(u \cdot L)}$ are so close to the 45° line that practically any value between 7% and 20% might be regarded as an equilibrium unemployment rate.

In United Kingdom, the equilibrium is unique both before and after the crisis despite the fact that the aggregate output demand mechanism (6) plays much more important role in the United Kingdom than in Spain. Therefore, the existence of multiple equilibrium unemployment rates is not so intimately connected to this demand mechanism as claimed by Čížek [5] who states that the reason for the existence of multiple equilibriums is the increased intensity of the output demand mechanism.

6 Conclusion

The paper formulated, econometrically estimated and analyzed the model of unemployment and job vacancies in which aggregate output demand mechanism and its effect on labor market is a central question of the analysis. Surprisingly, output demand mechanism turned out to be quantitatively more important in United Kingdom than in Spain. Despite this fact, labor market in Spain is characterized by a multiplicity of equilibriums which suggests that this multiplicity is not primarily caused by this aggregate demand mechanism as claimed by Čížek [5].

Influence of the aggregate output demand mechanism strongly increased after the crisis in the United Kingdom and slightly decreased in Spain. This mechanism turned out to be statistically significant in both countries either before or after the crisis, which is in line with the results of other papers (Kaplan, Menzio [13], Guerrazzi, Gelain [10], Eriksson, Stadin [7], Heathcote, Perri [11], Benigno, Fornaro [1], Cynamon, Fazzari [3]).

Acknowledgements

This work was supported by the Internal Grant Agency of University of Economics, Prague under Grant F4/53/2019.

References

[1] BENIGNO, G. & FORNARO, L. (2018). Stagnation Traps. Review of Economic Studies, 85(3), 1425–1470, https://doi.org/10.1093/restud/rdx063

[2] BÍLKOVÁ, D. (2015). Financial Position of Czech Employees at the Beginning of the 3rd Millennium according to Educational Attainment. Prague Economic Papers, 24(3), 307–331, https://doi.org/10.18267/j.pep.521

[3] CYNAMON, B. Z. & FAZZARI, S. M. (2015). Rising Inequality and Stagnation in the US Economy. European Journal of Economics and Economic Policies-Intervention, 12(2), 170–182, https://doi.org/10.4337/ejeep.2015.02.03

[4] ČABLA, A. & MALÁ, I. (2017). Modelling of Unemployment Duration in the Czech Republic. Prague Economic Papers, 26(4), 438–449, https://doi.org/10.18267/j.pep.620

[5] ČÍŽEK, O. (2017). Unemployment and Aggregate Demand in Spain. Ekonomický časopis, 65(5), 459–481.

[6] DIAMOND, P. A. (1982): Aggregate Demand Management in Search Equilibrium. Journal of Political Economy, 90(5), 881–894, https://doi.org/10.1086/261099

[7] ERIKSSON, S. & STADIN, K. (2017). What are the Determinants of Hiring? The Importance of Product Market Demand and Search Frictions. Applied Economics, 49(50), 5144–5165, https://doi.org/10.1080/00036846.2017.1302058

[8] FARMER, R. E. A. (2008). Aggregate Demand and Supply. International Journal of Economic Theory, 4(1), 77–93, https://doi.org/10.1111/j.1742-7363.2007.00069.x

[9] FLEK, V. & MYSÍKOVÁ, M. (2015). Unemployment Dynamics in Central Europe: A labour Flow Approach. Prague Economic Papers, 24(1), 73–87, https://doi.org/10.18267/j.pep.501

[10] GUERAZZI, M. & GELAIN, P. (2015). A Demand-Driven Search Model with Self-Fulfilling Expectations: The New 'Farmerian' framework under scrutiny. International Review of Applied Economics, 29(1), 81–104, https://doi.org/10.1080/02692171.2014.945993

[11] HEATHCOTE, J. & PERRI, F. (2018). Wealth and Volatility. Review of Economic Studies, 85(4), 2173–2213, https://doi.org/10.1093/restud/rdx074

[12] HOLT, R. P. F. & PRESSMAN, S. (eds.) (2001). A New Guide to Post Keynesian Economics. London and New York, Routledge, https://doi.org/10.4324/9780203466681

[13] KAPLAN, G. & MENZIO, G. (2015). Shopping Externalities and Self-Fulfilling Unemployment Fluctuations. Journal of Political Economy, 124(3), 771–825, https://doi.org/10.1086/685909

[14] LAVOIE, M. (2006). Introduction to Post-Keynesian Economics. Basingstoke, Palgrave Macmillan, https://doi.org/10.1057/9780230626300

[15] MICHAILLAT, P. & SAEZ, E. (2014). Aggregate Demand, Idle Time, and Unemployment. Working Paper no. 14–214. Kalamazoo, MI: W.E., Upjohn Institute for Employment Research, https://doi.org/10.17848/wp14–214

[16] SHIMER, R. (2012). Reassessing the Ins and Outs of Unemployment. Review of Economic Dynamics, 15(2), 127–148, https://doi.org/10.1016/j.red.2012.02.001

[17] STIGLITZ, J. E. (2013). Crisis: Principles and Policies: With an Application to the Eurozone Crisis. In: International Economic Association Roundtable on Debt Crises – How to prevent them, how manage them, how to ensure there is life after debt. Buenos Aires, August 13–14, 2012, https://doi.org/10.1057/9781137411488_2

[18] VEČERNÍK, J. (2007). The Czech Labour Market: Historical, Structural and Policy Perspectives. Prague Economic Papers, 16(3), 220–236, https://doi.org/10.18267/j.pep.306

[19] VUJČIĆ, B., DESKAR-ŠKRBIĆ, M., RATKOVSKI, Z., & ZRNC, J. (2014). Functional Distribution of Income and Economic Activity in Croatia: Post-Keynesian Approach. Zbornik Radova Ekonomskog Fakulteta u Rijeci-Proceedings of Rijeka Faculty of Economics, 32(1), 57–73.

[20] ZDENĚK, R. & STŘELEČEK, F. (2012). Evaluation of Development of Employment, Average Wage and Labour Productivity Using Shift-Share Analysis. E+M Ekonomie a Management, 15(3), 4–15.

Health System Efficiency and the COVID-19 Pandemic

Martin Dlouhý[1]

Abstract. The objective of the paper is to investigate the relationship between the health system efficiency and the performance of the health system in the time of the COVID-19 pandemic, which is measured by the case fatality rate. We used the data on 36 OECD member states. We evaluated the health system efficiency by the data envelopment analysis with three health system inputs (physicians, nurses, hospital beds) and two outputs (population, life expectancy). We run the logistic regression and have found that there is a piece of statistical evidence that health system efficiency, as defined in this study, increases the case fatality rate of COVID-19. Such finding has a substantial health policy impact. Clearly, the post-pandemic health systems will be different. Firstly, the aim of health system efficiency has to redefined. Secondly, some reserve capacity that was considered as inefficient in the past is needed. Thirdly, the health systems have to function with more flexibility, and health resources have to be mobilized in a short time in the case of unexpected demand.

Keywords: health system, efficiency, data envelopment analysis, COVID-19, case fatality rate

JEL Classification: I0, H51, C44
AMS Classification: 90C08

1 Introduction

The COVID-19 pandemic is causing many deaths and suffering in many countries of the world. In response to this pandemic, the national governments and public health authorities have adopted various public health measures, including social distancing, improved personal hygiene, early detection of new cases and others. The aim of these public health measures is not only reducing the number of the infected but also decreasing the pressure on health system capacities.

The case fatality rate (CFR) of COVID-19 indicates the severity of the disease and its significance as a public health problem, which helps public health authorities to determine the appropriate interventions. However, a precise estimate of the CFR of COVID-19 is impossible at present [1,2]. An approximate (and biased) estimate of the CFR can be obtained by dividing the number of deaths by the number of confirmed cases. With great caution, such calculations offer the public health authorities a piece of timely information. It is observed that the CFR significantly vary in time and across countries. The value of the CFR is affected by many factors such as the age structure of the population and by the availability of health system capacities.

The scientific community emphasizes that it is necessary to reduce peak health care demand in order to avoid a state of overwhelmed health systems, even in high-income countries [7, 8,10,11]. An unexpected pressure on health systems, especially on those with an existing shortage of physicians, nurses, and hospital beds is hardly manageable. The COVID-19 pandemic has shown the insufficiency of health system capacities in many countries. The numbers of hospital beds per capita and their occupancy rates vary greatly across the OECD member countries. The critical bottleneck in hospital bed capacity is intensive care unit beds, but the availability of comparable data is limited. The variation in intensive care beds capacity across selected OECD countries is ten-fold, ranging from 3.3 beds per 100,000 population in Mexico to 33.9 beds per 100,000 population in Germany [11].

Many governments pursue the goal of efficiency in health care. One way to achieve efficiency of the health system can be achieved by lowering numbers of health system capacities. However, lower health system capacities can be overwhelmed in the case of a sudden outbreak of infectious disease. So paradoxically, the health system efficiency in this particular context increases the probability of health system overload, and consequently, it increases the value of the CFR.

[1] University of Economics in Prague, Faculty of Informatics and Statistics, Department of Econometrics, Winston Churchill Sq. 4, 13067 Prague 3, Czech Republic, dlouhy@vse.cz.

The objective of this paper is to investigate the relationship between the health system efficiency and the performance of the national health system in the time of the COVID-19 pandemic, which is measured by the case fatality rate.

2 Methods

2.1 Logistic Regression

Since the CFR has a value between zero and one (it is the probability), we used the logistic regression to estimate the model. In the regression model, we include four independent variables. First, age is a significant risk factor [12], so the society with a higher proportion of the population aged 65 and over (POP65) can expect more deaths. Second, the high number of confirmed cases per capita (CASES) increases the probability of the health system overload. Hence, it may lead to decreased access to health services and their quality, and consequently to the increased CFR. Third, the gross domestic product per capita in USD purchasing power parity (GDP) expresses the power of the economy to mobilize resources, however, GDP also affects the quality of the national health system. Forth, the efficiency score representing the health system efficiency is included in the model. The efficiency score will be calculated by the data envelopment analysis that is used for efficiency evaluation in health care quite frequently [4, 5, 9, 13]. The observations are weighted by the number of confirmed cases.

2.2 Data Envelopment Analysis

For efficiency evaluation, we will use data envelopment analysis (DEA) that evaluates the technical efficiency of homogeneous production units and is able to deal with multiple inputs and outputs. DEA is based on the theory of mathematical programming and constructs the production frontier as the piecewise linear envelopment of the data [3, 6]. The production unit uses a number of inputs to produce outputs. The technical efficiency of the production unit is defined as the ratio of its total weighted output to its total weighted input.

In the DEA model, each production unit can choose its input and output weights to maximize its technical efficiency score. A technically efficient production unit is able to find such input and output weights that the production unit lies on the production frontier. The production frontier represents the maximum amounts of output that is produced by given amounts of input (the output maximization DEA model) or the minimum amounts of inputs required to produce the given amount of output (the input minimization DEA model).

Let us have n homogeneous production units that use m inputs to produce r outputs. The mathematical formulation of the input-oriented version of the constant returns-to-scale DEA model for production unit q is as follows:

$$\begin{aligned}
&\text{Maximize} \quad \varphi_q = \sum_{k=1}^{r} u_k y_{kq} \\
&\text{subject to} \quad \sum_{k=1}^{r} u_k y_{kj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, j = 1, 2, \ldots, n, \\
&\qquad \sum_{i=1}^{m} v_i x_{iq} = 1, \\
&\qquad u_k \geq \varepsilon, k = 1, 2, \ldots, r, \\
&\qquad v_i \geq \varepsilon, i = 1, 2, \ldots, m,
\end{aligned} \tag{1}$$

where φ_q is the technical efficiency score, x_{ij} is the amount of input i used by production unit j, y_{kj} is the amount of output k produced by production unit j, u_k and v_i are output and input weights, and ε is an infinitesimal constant. In the input-oriented DEA model, the technical efficiency score φ_q is one if the production unit q is technically efficient, and is lower than one if the unit is technically inefficient. The efficiency score calculates a required size of input reduction that makes production unit q technically efficient.

For the health system evaluation, we use three inputs and two outputs. The health system inputs are the number of physicians (head counts), the number of nurses (head counts), and the number of hospital beds.

Two outputs are the served population and life expectancy at birth. The input-oriented DEA model is considered.

3 Data

The sample includes 36 OECD member countries. The data on population (2018), life expectancy at birth (2017), GDP per capita in USD purchasing power parity (2018 or latest available), and three types of health system capacities (2017 or latest available) come from the OECD databases. The data are presented in Table 1. The cumulative number of COVID-19 cases and the cumulative number of deaths come from the European Centre for Disease Prevention and Control. The approximate value of CFR was calculated from the number of deaths and the number of confirmed cases on May 1, 2020.

4 Results

In the first step, the health system efficiency was calculated by the input-oriented constant returns-to-scale data envelopment analysis. The efficiency scores are presented in Table 1. Two health systems are technically efficient, Turkey and Mexico, which is obviously caused by the minimal numbers of health system capacities in these countries.

In the second step, the model was estimated by the logistic regression (Table 2). The percentage of deviance explained by the model was 53.2%. All independent variables were statistically significant ($p < 0.01$). The higher proportion of persons aged 65 years and over (POP65) and the higher number of cases per capita (CASES) increase the CFR. On the other hand, the higher GDP per capita decreases the CFR. Finally, we have found that there is a piece of preliminary statistical evidence that health system efficiency (EFF) increases the case fatality rate of COVID-19. Our hypothesis on the effect of health system efficiency was confirmed.

Country	Physicians per 1000	Nurses per 1000	Beds per 1000	Life expectancy	Population in 1000s	Efficiency Scores
Australia	3.68	11.68	3.84	82.6	25016	0.615
Austria	5.18	6.85	7.37	81.7	8822	0.386
Belgium	3.08	10.96	5.66	81.6	11399	0.634
Canada	2.65	9.96	2.52	82.0	36940	0.873
Chile	2.45	2.72	2.11	80.2	18552	0.983
Czech Republic	3.69	8.06	6.63	79.1	10610	0.513
Denmark	4.00	9.95	2.61	81.2	5781	0.634
Estonia	3.47	6.19	4.69	78.2	1319	0.563
Finland	3.21	14.26	3.28	81.7	5513	0.702
France	3.16	10.48	5.98	82.6	66926	0.626
Germany	4.25	12.93	8.00	81.1	82792	0.457
Greece	6.07	3.31	4.21	81.4	10741	0.672
Hungary	3.32	6.51	7.02	75.9	9778	0.563
Iceland	3.87	14.50	3.06	82.7	349	0.636
Ireland	3.07	12.16	2.96	82.2	4830	0.753
Israel	3.14	5.08	3.02	82.6	8873	0.740
Italy	3.99	5.80	3.18	83.0	60484	0.618
Japan	2.43	11.34	13.05	84.2	126177	0.830
Korea	2.34	6.91	12.27	82.7	51635	0.846
Latvia	3.21	4.57	5.57	74.8	1934	0.583
Lithuania	4.56	7.71	6.56	75.6	2809	0.417
Luxembourg	2.98	11.72	4.66	82.2	608	0.660
Mexico	2.43	2.90	1.38	75.4	123410	1.000

Country	Physicians per 1000	Nurses per 1000	Beds per 1000	Life expectancy	Population in 1000s	Efficiency Scores
Netherlands	3.58	10.88	3.32	81.8	17181	0.650
New Zealand	3.27	10.17	2.71	81.9	4865	0.736
Norway	4.66	17.67	3.60	82.7	5296	0.532
Poland	2.38	5.10	6.62	77.9	37977	0.786
Portugal	4.97	6.70	3.39	81.5	10291	0.507
Slovakia	3.42	5.65	5.82	77.3	5443	0.547
Slovenia	3.10	9.92	4.50	81.1	2067	0.635
Spain	3.88	5.74	2.97	83.4	46659	0.646
Sweden	4.12	10.90	2.22	82.5	10120	0.680
Switzerland	4.30	17.23	4.53	83.6	8484	0.531
Turkey	1.87	2.07	2.81	78.1	80811	1.000
UK	2.81	7.83	2.54	81.3	66274	0.829
United States	2.61	11.74	2.77	78.6	327892	0.821

Table 1 Data and Technical Efficiency

Parameter	Estimate	Standard Error	Odds Ratio
CONSTANT	−5.08734	0.0399018	
POP65	0.116193	0.00108592	1.12321
CASES	0.0336686	0.00038415	1.03424
GDP	−0.0000103661	2.30073E−7	0.99999
EFF	1.30983	0.0306328	3.70554

Table 2 Estimated Regression Model (Maximum Likelihood)

5 Conclusion

The finding the health system efficiency has a negative impact on the CFR has substantial health policy impact on long-term health resource planning. We must realize that the post-pandemic health systems will be different. Firstly, the aim of efficiency has to redefined. Secondly, some reserve capacity that was considered as inefficient in the past is needed. Thirdly, the health systems have to function with more flexibility, and health resources have been mobilized in a short time in the case of unexpected demand.

We are aware that this study has many limitations, above all the reliability of the data on COVID-19. However, we believe that such limitations do not change the main findings.

Acknowledgements

The study was supported by the project no. 19-08985S, funded by the Czech Science Foundation.

References

[1] BAUD, D., QI, X., NIELSEN-SAINES, K., MUSSO, D., POMAR, L., & FAVRE, G. (2020). Real estimates of mortality following COVID-19 infection. *Lancet Infectious Diseases*, published online March 12. https://doi.org/10.1016/S1473-3099(20)30195-X.

[2] BATTEGAY, M., KUEHL, R., TSCHUDIN-SUTTER, S., HIRSCH, H. H., WIDMER A. F., & NEHER, R. A. (2020). 2019 – Novel Coronavirus (2019-nCoV): estimating the case fatality rate – a word of caution. *Swiss Medical Weekly*, 150:w20203. https://doi.org/10.4414/smw.2020.20203.

[3] CHARNES, A., COOPER, W. W., & RHODES, E. (1978). Measuring the Inefficiency of Decision Making Units. *European Journal of Operational Research*, 2, 429–444.

[4] DLOUHÝ, M., JABLONSKÝ, J., & NOVOSÁDOVÁ, I. (2007). Využití analýzy obalu dat pro hodnocení efektivnosti českých nemocnic [Use of Data Envelopment Analysis for Efficiency Evaluation of Czech Hospitals, in Czech]. *Politická ekonomie*, 54, 60–71.
[5] DLOUHÝ, M. (2018). Measuring Geographic Inequalities: Dealing with Multiple Health Resources by Data Envelopment Analysis. *Frontiers in Public Health*, 6, article 53.
[6] DLOUHÝ, M., JABLONSKÝ, J., & ZÝKOVÁ, P. (2018). *Analýza obalu dat* [Data Envelopment Analysis, in Czech]. Prague: Professional Publishing. ISBN 978-80-88260-12-7.
[7] DLOUHÝ, M. (2020). *Health system capacities affect the case fatality rate of COVID-19*. Unpublished manuscript.
[8] FERGUSON, N. M., LAYDON, D., NEDJATI-GILANI, G., IMAI, N., AINSLIE, K., BAGUELIN, M., ET AL. (2020). *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. Imperial College London (16-03-2020) [accessed 2020, Apr 26]. https://doi.org/10.25561/77482.
[9] HOLLINGSWORTH, B. (2003). Non-Parametric and Parametric Applications Measuring Efficiency in Health Care. *Health Care Management Science*, 6, 203–218.
[10] LI, R., RIVERS, C., TAN, Q., MURRAY, M. B, TONER, E., & LIPSITCH, M. (2020). The demand for inpatient and ICU beds for COVID-19 in the US: lessons from Chinese cities. medRxiv. 2020.03.09.20033241. https://doi.org/10.1101/2020.03.09.20033241.
[11] OECD. (2020). *Beyond Containment: Health systems responses to COVID-19 in the OECD countries – Updated 16 April 2020* [accessed 2020 Apr 26]. https://oecd.dam-broadcast.com/pm_7379_119_119689-ud5comtf84.pdf.
[12] RUAN, S. (2020). Likelihood of survival of coronavirus disease 2019. *Lancet Infectious Diseases*, published March 30. https://doi.org/10.1016/S1473-3099(20)30257-7.
[13] VOTÁPKOVÁ, J. & ŠŤASTNÁ, L. (2013). Efficiency of Hospitals in the Czech Republic. *Prague Economic Papers*, 22, 524–541.

A New Approach to Investment Risk Assessment in HFT Systems

Michał Dominik Stasiak[1], Krzysztof Piasecki[2]

Abstract. Investment risk assessment in case of using HFT systems is a complicated and significant issue. The size of a single HFT transaction is decided based on the investor's capital and their preferences regarding possible risk. Techniques of measuring risk that are most popular among investors using HFT systems usually include analysing the return rate and the loss in capital (e.g. Calmar's indicator). Yet, those kind of techniques use only the results of an econometric analysis, realized based on historical data from a given time period. In the following article we propose a new approach, consisting in a proper extrapolation of capital increase function and the maximal drawdown increase function. This approach allows for an assessment of the change character of parameters over time, and, in consequence, for a more precise analysis of the risk undertaken by investors using HFT systems.

Keywords: technical analysis, investment decision support, algorithmic trade risk assessment, investment risk

JEL Classification: F31, G11,C49
AMS Classification: 91G70, 91B30

1 Introduction

The development in IT and telecommunication revolutionized the trade and methods of supporting investment decisions. Those changes can be seen most clearly on the biggest financial market, that is on the currency market. Thanks to introducing broker platforms about 10 years ago, such as Metatrader (which is the most popular), the investor obtained many new possibilities of action. Broker platforms allow for opening and closing orders from an arbitrary device in less than one second. Moreover, thanks to the MqL4 programming language, the platforms allow for automation of the transaction process [6, 7]. Because of those possibilities, systems automating the investment process were created and developed, i.e. High Frequency Trading systems (HFT) [2, 7]. Many of those systems use known methods of technical analysis [4, 10, 9], neural networks [8] or advanced modelling methods [11, 13]. They perform dozens, hundreds or even thousands of transactions in a short time. HFT systems are used by both individual and institutional investors.

HFT system design and software development requires, among others, an extensive IT and econometric knowledge. Because of this, many investors seek for ready-made solutions. On the other hand, designers offer their systems to investors, for a suitable price. In order to match those two groups, i.e. the designers and investors, so called social trading emerged and developed [7]. Dedicated services of the www.mql5.com or www.zulutrade.com type offer the possibility of copying investments of a given HFT system in exchange for a fixed subscription fee.

The growth of the HFT "market" requires potential investors to assess the risk incurred when investing on the basis of selected systems. Because of obvious reasons, the very algorithm of the HFT systems is not commonly known, contrary to the history of transactions made or testing results for historical data. So the question arises, how to assess the risk and choose a safe system? In the following article we present a concept of an advanced assessment of a HFT system regarding the investment risk.

2 HFT system risk assessment

Using an arbitrary HFT system, one should assess the bankruptcy risk, that is the risk of reaching a negative balance. A negative balance results in closing all open positions, and, in consequence, in losing all of the

[1] Poznan University of Economics and Business, Institute of Management, Department of Investment and Real Estate, Al. Niepodległości 10, 61-875 Poznań, michal.stasiak@ue.poznan.pl.
[2] Poznań University of Economics and Business, Institute of Management, Department of Investment and Real Estate, Al. Niepodległości 10, 61-875 Poznań, Poland e-mail: krzysztof.piasecki@ue.poznan.pl.

invested capital or even in having to repay the resultant negative balance. In this paper, the investment risk will be defined as a risk of losing all of the invested capital.

The most intuitive method of assessing the risk connected with investing based on HFT system indications, which is used by investors and recommended in many scientific publications, is the method based on using the analysis of maximal drawdown (MDD) [1, 2, 12] in the strategy testing period. Maximal drawdown is defined as a highest registered difference between the highest registered balance and the further drop. MDD can be calculated with the following formula:

$$\mathrm{MDD(T)} = \max_{t \in (0,T)} \left\{ \max_{s \in (0,t)} \{X(s) - X(t)\} \right\} \tag{1}$$

where X(t) is the account balance if the investor in the moment t, and T describes the time of HFT system performance. MDD parameter is the main parameter in evaluating any HFT system, that is calculated in the "strategy tester" of the MetaTrader platform and social trading services such as www.mql5.com. In order to effectively assess the investment risk, investor's balance increase analysis is also required. The increase in the account by the profits from successful investment decisions allow for an increase in the accepted MDD value. Coefficients describing the level of undertaken risk that are based on maximal drawdown and capital increase are usually applied in economics. Examples of such coefficients are the popular Calmar's coefficient [15] and its modifications, e.g. Sharp's coefficient [5].

Risk analysis with use of maximal drawdown and maximal capital increase, registered in a given time period, has some serious disadvantages, according to the Author. Maximal drawdown, as well as the maximal capital increase, are variable in time, so when analysing a shorter period of time, e.g. 6 months, we can obtain results vastly different from those obtained from a longer time period, e.g. 3 years. This stems from the fact that both values are maximal, registered in a given time period. By regarding only, the maximal values of those parameters one disregards the character of their changes over time. However, it is the change character that can be a valuable premise to assess the future dynamics of changes. To sum up, only based on the analysis of proper parameter changes over time one can deduce if the risk will drop or increase over time.

3 A new approach to investment risk assessment in HFT systems

In order to eliminate the disadvantages described in the previous section, in this paper we propose a new approach to HFT system investment risk assessment. This kind of approach is based on the analysis of the speed of maximal drawdown increase and the maximal capital increase in the tested time period. To do so, linear extrapolation of ensuing registered maximal drawdown and maximal capital increase values was performed, with use of least squares method [6]. The maximal drawdown function F_{MDD} can be thus defined as follows:

$$F_{MDD} = \alpha t + C_{MDD}, \tag{2}$$

where α is the directional coefficient, t is time and C_{MDD} is a constant that has no influence on the speed of changes and so its value is not analyzed. The directional coefficient of the maximal drawdown function (α) can be interpreted as the speed of the increase of this parameter over time. The higher the coefficient, the faster the maximal drawdown increase.

The maximal capital increase function F_{MAX} can be described the following way:

$$F_{MAX} = \beta t + C_{MAX}, \tag{3}$$

where β is the directional coefficient, t is time and C_{MAX} is a constant that has no influence on the speed of changes and so its value is not analyzed. The directional coefficient of this function (β) can be interpreted as the speed of capital increase (profit). In this paper we propose a new $\mathrm{Ind_{AR}}$ indicator as a measure of investment risk. It is defined as a ratio of directional coefficients of maximal capital increase and maximal drawdown functions:

$$Ind_{AR} = \frac{\beta}{\alpha}. \tag{4}$$

Thus, the $\mathrm{Ind_{AR}}$ indicator is, in fact, a directional coefficient of the increase function, calculated for the differences between the capital increase and the maximal drawdown increase. In presented considerations we assumed that $\alpha \neq 0$. $\alpha = 0$ would have suggested that the HFT system is characterized by 100% forecast accuracy, and such kind of system does not require any risk analysis.

Based on the Ind_{AR} parameter we can evaluate the HFT system regarding the risk taken by the investor.

- If $Ind_{AR} < 1$, then the increase in the maximal drawdown is higher than the increase in the capital – in the longer perspective the system will lead to a loss in capital (the increase in risk over time).
- If $Ind_{AR} > 1$, then the increase in the maximal drawdown is lower than the capital increase. This means that the difference between the forecasted investor's balance (the sum of initial balance and system performance profit) and the forecasted maximal drawdown will increase, and in consequence, the risk will drop.

The basic requirement for the HFT system is to allow for a risk decrease or at least to maintain a constant risk level. Therefore, one can formulate a condition for a HFT system to be suitable for application in investment decision support:

$$Ind_{AR} \geq 1. \tag{5}$$

The higher the Ind_{AR}, the safer the system, thus the maximization of Ind_{AR} should be used as a parameter when choosing a suitable trading system.

4 Risk assessment with use of Ind_{AR} indicator

Let us now consider using the indicator Ind_{AR} in assessing the investment risk of a HFT system. The investor considers the possibility of using a HFT system, which had made 50 transactions in a given time period in the transaction history (Figure 1). Often the historical results of investing with use of a given systems are replaced by simulation results performed on historical data (so called backtest, [2]). Based on the balance fluctuations, one can calculate the linear approximation of changes in the maximal capital drawdown (Figure 2). We solve this approximation task with use the Method of Least Square.

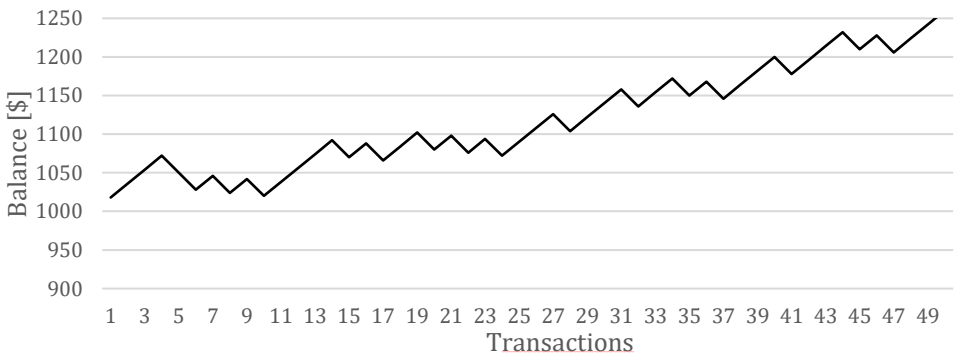


Figure 1 History of transactions made by the researched HFT system.

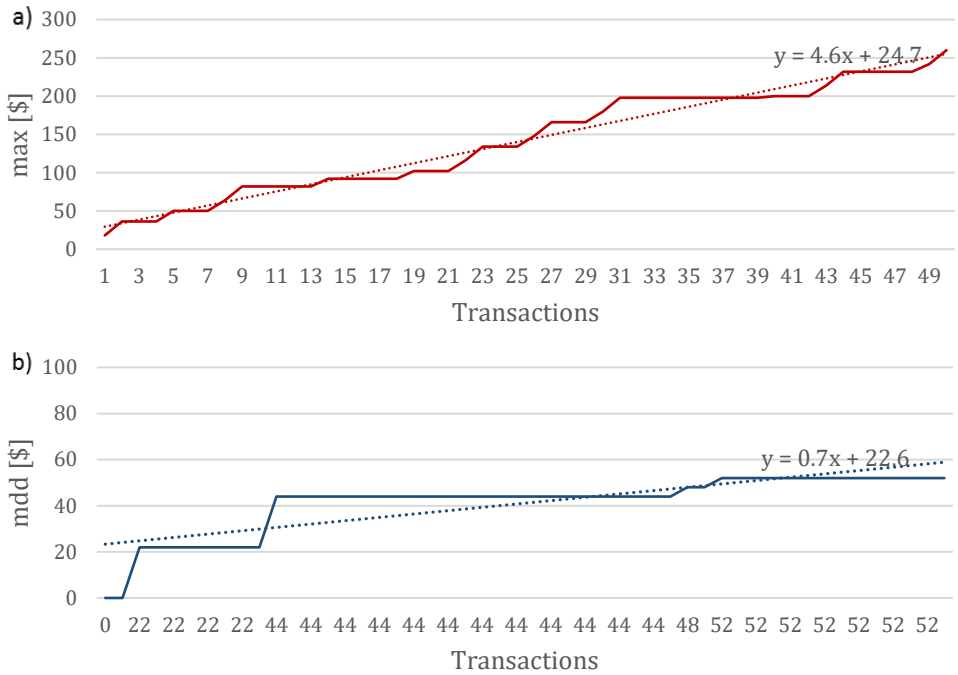


Figure 2 Linear function of the maximal capital increase (a) and changes in the maximal capital drawdown (b).

Figure 3 presents a forecast of changes in mentioned max and mdd values for the next 50 transactions.

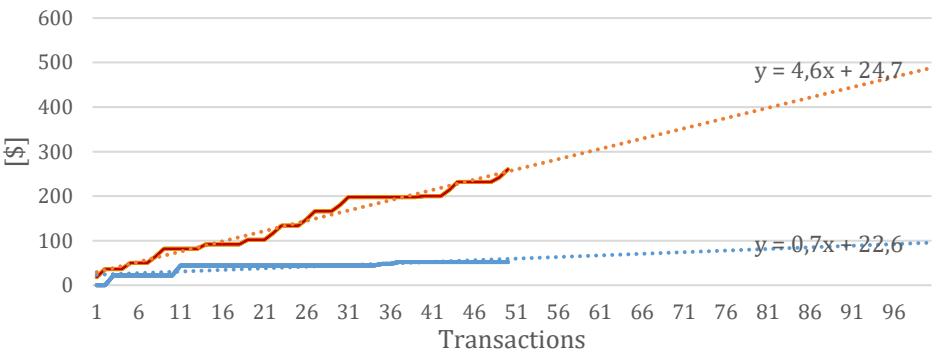


Figure 3 Forecast of the capital increase and the increase of maximal capital drawdown.

From the forecast we can deduce that the capital on the investor's account will increase its value definitely faster than the ensuing capital drawdowns occur. The fact can be interpreted most simply as a decreasing tendency in risk over time. Ind_{AR} indicator for the analysed strategy is 6.57. Therefore, according to (5), the researched system can be reccomended to the investor.

Let us now consider another HFT system, for which the transaction history is presented in the Figure 4. The system, despite the positive return rate, is characterised by risk increasing over time (Figure 5). The risk indicator Ind_{AR} equals 0.5.

Figure 4 History of transactions made by the second researched HFT system.

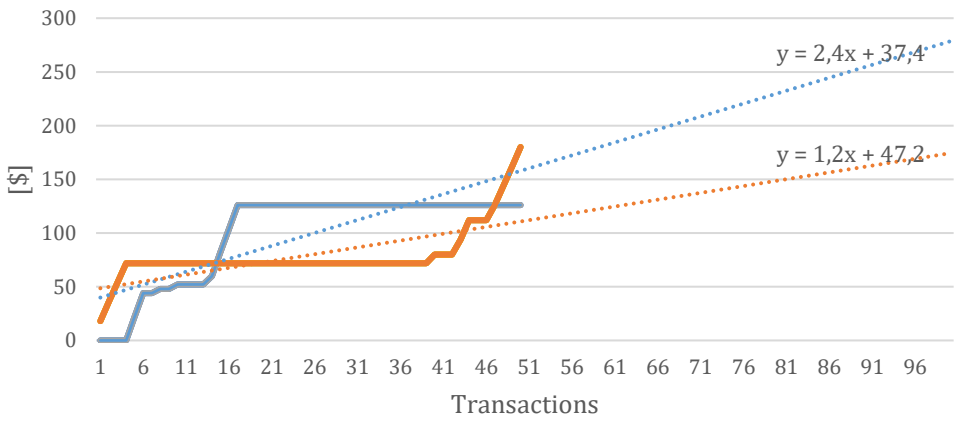


Figure 5 Forecast of the capital increase and the increase of maximal capital drawdown.

Many HFT systems make investment decisions based not on the technical analysis, but on the progression systems known from the mathematical analysis of hazard games (so called gambling strategies) [3, 14]. Those kind of systems are characterized by a rapid increase of risk over time. An example of such a strategy is the best-known martingale system [14]. It consists in randomly choosing the direction of the transaction of a given value (e.g. 1 Lot). In the event of a loss, the next transaction is concluded with twice the size (e.g. 2 Lots), etc. The use of such systems may lead to bankruptcy, yet due to the profits from subscription fees (for some indefinite period of time systems "allow" for some profit), they are unfortunately offered to investors. The analysis proposed in this paper, which is based on the use of Ind_{AR} risk indicator, allows to detect systems that use progression.

5 Summary

The investment risk analysis performed with use of Ind_{AR} indicator seems to be far more precise and credible than those carried out with existing tools. The cause of this is the possibility of analysing the change dynamics for parameters responsible for the risk undertaken by the investor. As contrary to the statistical risk indicators such as Calmar's or Sharp's, which do not encompass the character of the change, the proposed indicator allows for an assessment of potential direction of the change of risk in a given HFT system. This kind of approach allows to detect popular systems that are based on progression, which due to a drastic increase in maximal capital drawdown, lead in consequence to the bankruptcy of the investor.

References

[1] ACAR, E. & JAMES, S. (1997). Maximum loss and maximum drawdown in financial markets. In *Proceedings of International Conference on Forecasting Financial Markets*.

[2] ALDRIDGE, I. (2009) *High-frequency trading: a practical guide to algorithmic strategies and trading systems*. John Wiley and Sons.
[3] BALSARA, N. J. (1992). *Money management strategies for futures traders* (Vol. 4). John Wiley & Sons.
[4] BURGESS, G. (2010). *Trading and Investing in the Forex Markets Using Chart Techniques* (Vol. 543), John Wiley & Sons.
[5] BURKE, G. (1994). *A sharper Sharpe ratio*, The Computerized Trader.
[6] DAHLQUIST, G. & BJORCK, A. (2008). *Numerical Methods in Scientific Computing: Volume 1* (Vol. 103). Siam.
[7] GALLO, C. & FRATELLO, A. (2014). The Forex market in practice: a computing approach for automated trading strategies. *Int. J. Econ. Manag. Sci*, 3(169), 1–9.
[8] YAO, J. & TAN, C. L. (2000). A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing*, 34(1–4), 79–98.
[9] MURPHY, J. J. (2002). *Technical analysis of the futures markets: A comprehensive guide to trading methods and applications*, Prentice Hall Press.
[10] SCHLOSSBERG, B. (2006) *Technical Analysis of the Currency Market*. John Wiley & Sons.
[11] STASIAK, M. D. (2019). Trend analysis with use of binary representation. In: HOUDA, M., REMEŠ, R. (Eds.), *Proceedings of 37th International Conference Mathematical Methods in Economics MME 2019* (pp. 529–535) České Budějovice: University of South Bohemia.
[12] PARDO, R. (2011). *The evaluation and optimization of trading strategies*, John Wiley & Sons.
[13] PIASECKI, K. & STASIAK, M. D. (2019). The Forex Trading System for Speculation with Constant Magnitude of Unit Return. *Mathematics*, 7(7), 1–23.
[14] VINCE, R. (1992). *The mathematics of money management: risk analysis techniques for traders* (Vol. 18). John Wiley & Sons.
[15] YOUNG, T. W. (1991). Calmar ratio: A smoother tool. *Futures*, 20(1), 40.

GUHA Method Application in Behavioral Finance Focusing on the Length of Trade

Hana Dvořáčková[1]

Abstract. This paper focuses on application of the GUHA method in the field of behavioral finance. The aim is to use the GUHA method to automatic generation of hypotheses to specify characteristics of trades with respect to their length to explore new possible hypotheses within the experimental data set as a base for the further researching of the disposition effect. Experimental trading data used for the research were collected by Jochec and Dvořáčková, Dvorackova, Jochec, and Tichy [7]. The hypotheses given by GUHA were assessed according to their significance and statistically tested, while all of them were found statistically significant. The results support an assumption, that traders enjoying risk tend to make shorter trades, thus report a speculative behavior. Contrary students with no previous experience with speculation will most probably hold trades longer.

Keywords: behavioral finance, GUHA, financial markets, FX trading

JEL Classification: C12, G02

1 Introduction

Currently two main approaches to the financial decision-making and asset pricing are acknowledged. First of all the traditional neoclassical finance whose proponents treat sentiment as a minor determinant of market prices and assume that investors are mostly free of biases in their decision-making. They are focusing on the fundamental risk or time-varying risk aversion and seeking to maximize the expected utility. This is a rationality-based framework. Contrary, the behavioral approach accepts sentiment as the major determinant of market prices and proponents are very critical to the expected utility as the main descriptive theory. They assume that people usually do not behave exactly in accordance with the expected utility theory because their behavior contains several psychological biases.

The paper is denoted to behavioral finance, particularly to the disposition effect. The disposition effect has been described in the stock-investing context as a behavioral tendency of investors to hold on losing stocks for too long and sell winning stocks too early. This is an implication of the prospect theory to investment by Kahneman and Tversky [12]. According to Kahneman and Tversky, a person who has not made peace with his losses tends to accept risks and gamble which would not be otherwise acceptable. Later, Shefrin and Statman [16] published the first paper naming this behavioral effect as the disposition effect. According to the first large publication regarding to this topic by Odean [14], the most obvious explanations – explanations based on informed trading, re-balancing, or transaction costs – fail to capture important features of the data. For the purpose of the said paper, records of 10,000 accounts at a large discount brokerage house were analysed to prove the tendency of investors to hold losing investments for too long and sell winning investments too early.

Another studies focused on this topic were, for example, Shiller and Case [17], who tested the disposition effect at the real estate market. Weber and Camerer [21] confirmed the disposition effect within an experimental data set, collected from 103 students. Contrary to our experiments, their students could trade only with six selected stocks, which were modelled by the authors. Locke and Mann [13] were focused on the nature of trading discipline and investigated whether professional traders are able to avoid irrational behaviors confirmed in retail populations. They found that even professional traders succumbed to the disposition effect, however, they did not find any evidence of costs arising from this fact. Barberis and Xiong [4] investigated whether the prospect theory preferences can predict disposition effect, when considering two implementations of the prospect theory: firstly, preferences are defined over annual gains and losses; secondly, preferences are defined over realized gains and losses. According to their results, the annual gain/loss model often fails to predict a disposition effect, contrary the realized gain/loss model predicts it reliably. Cheng, Lee, and Lin [5] were focused on the effect of age and gender on the tendency to succumb to the disposition effect. According to their results obtained by comparing of the likelihood of realizing gains versus realizing losses, woman and mature traders showed stronger disposition effect. Eom [8] was focused on

[1] VŠB-TUO, Department of Finance, Sokolská 33, Ostrava, 702 00, Czech Republic, hana.dvorackova@vsb.cz

the trading behavior of investors, using data of transactions of the Korean stock index from 2003 to 2005. An intraday analysis, focused on the duration of trades was used, while he found an empirical evidence for the disposition effect, but also for the opposite of the disposition effect. Dvorackova, Jochec, and Tichy [7] analyzed the disposition effect in the experimental data set, while the disposition effect has been confirmed.

The aim of this paper is to use the GUHA method in context of behavioral finance to generate hypotheses specifying characteristics of trades with respect to their length to explore new possible hypotheses in the experimental data set as a base for further researching of the disposition effect. The knowledge of factors affecting the length of trades is important to understand besides others so called disposition effect. The motivation to use this method is to explore possible new hypotheses, which were not yet examined and look at the length of trade, and later also the disposition effect from very new points of view.

2 Data

The data set used in this research were collected by Dvorackova and Jochec [6] from 2009 to 2015 during lectures in several countries (for instance, the USA, New Zealand, Kazakhstan, etc.). Moreover, the data collection proceeded in 2016 at the Faculty of Economics of VŠB-TUO in the Czech Republic. Students were trading on the OANDA FX Trade Practice platform with currency pairs and CFDs.

Initially, students were given $100,000, meanwhile the trading period was standardized and took three months. As those students did not trade with real money, they were motivated to achieve as good result as possible by a financial reward together with extra points for the exam on account of the winner (student with the highest account balance at the end of the trading period). One of the learning objective was to experience the first-hand trading and use explained technical analysis techniques in practise. At the end of the given trading period students had to submit in detail recorded trading history together with their answers to a short questionnaire and demographic information. Also the student's login information and passwords of their official game account were collected, therefore it was not possible to change the account later, reset losses, use more accounts, etc. The advantage of the experimental setting is that traders do not self-select into roles and thus the sample is less biased: take, for example, the effect of gender on trading decisions. Moreover, the effect of the self-selection bias, examined in Adams and Funk [2] or survivorship bias, were declined. Another advantage to be a homogeneity of objectives of traders. Last but not least, the effect of traders' wealth was eliminated, as the money was not real. When using real trading data, each trader is affected by the absolute amount of his wealth and what portion of it uses for trading.

The counterargument might be that the "winner takes it all" reward scheme is problematic, there is no incentive for scoring second (third, etc.); similarly, scoring low does not bear any penalty. This and the fact that it is hard to predict currency rates even for professional traders means that students were encouraged taking higher-than-normal risk and engaged in "all or nothing" gamble. It was not possible to perfectly rule those problems out, however, there is no indication of more frequent occurrence of large bets on the last few days of the game which would point out a tendency towards a pure gambling. It can be assumed that students derived some benefits also from simply doing well, even if not the best. This could result from the long-term continuance of the experiment and the psychological benefit (cost) of favourable (unfavourable) comparison with the peers. It is possible to see a parable with gaming. People around the world are playing several online games which are for free and provide no reward for playing, just comparison with other gamers within a ranking. According to Stanley [18], the social media gaming has around 800 million monthly users around the world, thus 800 million of people are motivated to play a game every month to keep a peer pride among people who even do not know. They are even willing to pay for some extra features to achieve better position, Gainsbury et al. [9].

Based on collected information a unique data set was created, containing experimental trading data linked to the individual students/traders. In total 336 students were involved in the research during the collection time period, who made almost 20,000 trades with the total volume of over three billion units. Those data were cleaned from trades with final profit/loss below $1 and only active trades were taken into account. Active trade is a trade which is opened/closed by a person, not system through a command, for example, take profit, stop loss, etc., as this automatic trades do not bear the behavioral patterns examined in this research. After this filtration the number of students who made relevant trades was 311 and final number of trades, i.e. examined observations, was 10,548. From above-mentioned students, 46% (144 students) became profit makers. That is to say that their account balance at the end of trading period was higher than initial $100,000. Regarding the gender diversity of traders, 142 females and 169 males were involved after

the data cleaning. Overall, 30% of trades were made by females, 70% by males, who appeared to be more active traders, as proposed in Barber and Odean [3].

3 GUHA – General Unary Hypotheses Automaton

The General Unary Hypotheses Automaton Method (hereinafter the "GUHA") firstly described in Hájek, Havel and Chytil [10] is one of the oldest data mining methods used for generating of systematical hypotheses supported by the given empirical data. According to Turunen [19] the GUHA method is pursuant to the well-defined first order monadic logic which contains generalized quantifiers on finite models. A GUHA procedure brings about generates statements on association between complex Boolean attributes, which are built from the predicates corresponding to the columns of the a matrix.

This method is essentially suitable for an exploratory analysis (having no specific hypothesis to be tested, only a general problem) of large data, typically used, for instance, in pharmaceutical and medical research [1]. The author is not aware of previous application of this method in the field of behavioral finance or financial markets. The data form a rectangle matrix of zeros and ones, whereas rows corresponding to objects belonging to the sample and columns corresponding to the researched variable (attributes). Typically, the matrix has hundreds to thousands of rows and dozens of columns. As stated in Hájek, Holeňa and Rauch [11], let's assume to $P_1, \ldots, P_n$ be names of the attributes. For each P_i, $\neg P_i$ is it's negation. An elementary conjunction of length $k(1 \leq k \leq n)$ is a conjunction of k literals in which each predicate occurs at least once, for example, $\neg P_3;\ P_1 \wedge \neg P_3 \wedge P_7$. Similarly to this, an elementary disjunction is defined, for example, $P_1 \vee \neg P_3 \vee P_7$. An object satisfies an elementary conjunction in case of satisfaction of all its members and it satisfies an elementary disjunction if it satisfies at least one of its members.

Let $0 \leq P \leq A$; formula $A \Rightarrow S_p$, where A is an Antecedent, i.e. an elementary conjunction and S is a Succedent, i.e. an elementary disjunction, is true in given data set if at least 100p percent of objects satisfying A satisfies also S, i.e. $\frac{a}{r} \geq p$, where r is the number of objects satisfying A and a is the number of objects satisfying A together with S. The Antecedent A is *t-good* if at least t objects satisfy it.

The software used for data mining using the GUHA method is called the LISp-Miner [1] and it has been developed in 1966 at the Faculty of Informatics and Statistics of the University of Economics in Prague. The software provides six GUHA procedures. A GUHA procedure is a computer program which based on imputed data provides a simple definition of relevant, thus potentially interesting patterns. GUHA procedure generates particular relevant patterns and tests whether they are true within analyzed data. Afterwards the output contains all prime patterns which are true in analyzed data an do not immediately follow from the other, more simple output pattern.

When using this method it is necessary to understand terms *Support*, *Confidence*, *Lift*, and *Count*, also used in this paper. *Support* represents the percentage of rows from the overall data set satisfying the left side of hypotheses generated by GUHA, while *Count* represents this information in its absolute value. Variable *Confidence* shows the percentage of rows to which the right side of hypotheses is applicable. Variable *Lift* defines the increase of *Confidence* of data filtered based on the GUHA results comparing to the overall data set.

Hypotheses generated by GUHA need to be statistically tested for their significance, using generally known statistical methods, particularly the F-test and t-test.

4 Results

From the experimental data set, following variables were chosen to be explored by GUHA:

1. *Length of trade* (mostly the explained variable, presenting length of trades in days)
2. *Profitmaker/Lossmaker* (dichotomous variable showing the final account balance of a trader);
3. *P/L* (amount of profit or loss on particular trade);
4. *currency pair* (currency pair used for trading);
5. *units to median* (dichotomous variable telling us whether the volume of particular trade is below median (0) or over median (1));
6. *Gender* (dichotomous variable, only "male" and "female" possibility was taken into account);
7. *Country* (home country of the trader);
8. *Question Nr. 3* of the questionnaire (Have you ever had a feeling that you became "hooked-up" (addicted) to the game?);

[1] [22], [20], [15]

9. *Question Nr.5* of the questionnaire (Do you think that you are a person that enjoys taking risks?);
10. *Question Nr.7* of the questionnaire (Have you ever made money by speculation?);
11. *Question Nr.9* of the questionnaire (Have you ever done any adrenaline sports?).

Over ten thousand trades were divided according to their length (in days) into three intervals, (0, 0.351] (hereinafter the "*Interval A1*", the shortest), (0.351, 5.44] (hereinafter the "*Interval A2*"), and (5.44, 100] (hereinafter the "*Interval A3*", the longest). Hypotheses were specified for each interval separately. Selected hypotheses with the highest *Lift*, i.e. the highest improvement of each interval are presented in Table 1 in the column *Left Side of Hypotheses* (so called *Antecedent* labeled according to the relevant interval) defining the rule given by GUHA. As shown in that table, the length of trade belongs to the Interval A1 mostly if the trader meets the following characteristics given by the *Antecedent A1*, he/she:

1. is a loss maker at the end of the trading period,
2. is a male
3. did not feel addiction during the trading period,
4. enjoys taking risk,
5. has ever done an adrenaline sport,

whereas within current data this characteristics meet 64% of students. Trades belonging to the intervals A2 (based on the *Antecedent A2*) and A3 have one interesting similarity – students/traders answered "no" to the question related to their experience with money speculation and so they might treat the trade as a long-term investment opportunity rather than a speculation. Additionally, the trades of the *Interval B* are also defined by profit of the trade within the interval $(50; 200\,000]$, those rules given by *Antecedent A2* met 43% of students.

Left Side of Hypothesis	*Right Side of Hypothesis*	*Support*	*Confidence*	*Lift*	*Count*
Profit_maker/Loss_maker = 0 Gender = 0 Q_3 = 0 Q_5 = 1 Q_9 = 1	$\Rightarrow \{\text{length of trade} = (0, 0.351]\}$	8.6%	64.2%	1.28	908
P/L = $(50; 200\,000]$ Q_7 = 0	$\Rightarrow \{\text{length of trade} = (0.351, 5.44]\}$	9.5%	43.0%	1.28	1005
Q_7=0	$\Rightarrow \{\text{length of trade} = (5.44, 100]\}$	9.6%	19.7%	1.2	1015

Table 1 Hypotheses Generated by the GUHA Method

Hypotheses generated by GUHA shall be statistically tested for they statistical significance, using standard statistical tests. In this paper, the F-test and t-test relevant to the variances equality of a specific hypothesis were used. Two separate groups were created, while the first group contained overall data (length of trade in days of all trades), the second contained those data after filtration based on the particular left side of hypothesis resulting from the GUHA method. Initially, the first antecedent leading to the length of trade between 0 and 0.351 days, hence to the *Interval A1*, was chosen for filtration and the two sample F-test for equality of variances testing on the significance level $\alpha = 0.05$ was performed to compare variances of both groups. The F-test was performed assuming hypotheses $H_0 : \sigma_1 = \sigma_2$ and $H_1 : \sigma_1 \neq \sigma_2$, where σ_1 represents the variance of length of trades from the whole data set and σ_2 represents the variance of length of trades from the data set after filtration according to the particular hypothesis, resulting in $F = 1.638$ and $F_{crit} = 1.069$, thus $F > F_{crit}$. Hence, it is possible to reject H_0 on the significance level of 0.05.

Based on that, the two-tailed two-sample t-test was chosen to test the difference in means. Hypotheses established for this test were $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$, where μ_1 represents the mean length of trades of whole data set and μ_2 represents the mean length of trades of the data set after filtration. The resulting $t = 6.68$, while $t_{crit} = 1.65$, thus $t > t_{crit}$ and it is possible to reject H_0 and assume statistically significant difference in means of both data sets.

Consequently, the data filtered based on the antecedent, thus left side of hypothesis relevant to the *Interval A2* were statistically tested in the same way as the first group. The F-test based on the same hypotheses as in the previous case came out $F = 1.08$ and $F_{crit} = 1.06$. As $F > F_{crit}$, it is possible to reject H_0 and assume inequality of variances of tested samples. Accordingly, the two-sample two-tailed t-test for inequal

variances was chosen for further testing. As resulting $t = -2.16$ and $t_{crit} = 1.96$, $|t| > t_{crit}$, therefore, H_0 can be rejected and it is possible to suppose the statistically significant difference in means of both data sets.

Regarding the group obtained by data filtration based on the antecedent, thus left side of hypothesis relevant to the *Interval A3*, the result of the F-test for equality of variances gives $F = 0.71$ and $F_{crit} = 0.96$ so that $F < F_{crit}$. Hypotheses H_0 and H_1 were set as previously and H_0 cannot be rejected, since the difference in variances of both groups cannot be assumed. Pursuant to this assumption, in this case the two-sample two tailed t-test for equal variances was executed resulting in $t = -5.65$ and $t_{crit} = 1.96$, giving $|t| > t_{crit}$. Therefore, H_0 also in this case can be rejected and it is possible to assume the statistically significant difference in means of both data sets.

All Statistical tests were found statistically significant as $p < \alpha$, whereas $\alpha = 0.05$.

5 Conclusion

The paper was focused on the application of GUHA method in behavioral finance. The aim was to use the GUHA method to generate hypotheses specifying characteristics of trades with respect to their length to explore new possible hypotheses within the experimental data set as a base for further researching of the disposition effect. Over ten thousand trades were divided into three groups, intervals, according to their length. The knowledge of factors affecting the length of trades in crucial to explore so called disposition effect, a significant behavioral bias affecting most of traders[2].

The GUHA method provided several hypotheses, thus only those with the highest effect, measured by variable *Lift* were chosen to be statistically tested and presented hereabove. The broadest specification was related to the *Interval A*, whereas a male, loss maker, who did not feel addiction during trading, enjoys taking risk and has done an adrenaline sport has higher probability to trade within the shortest trading period.

Regarding the *Interval A2* and *A3*, the specification is more simple. Profit between \$50 and \$200,000 together with experience with speculation increase the probability of trades in the *Interval A2*. Trades which will most probably be in *Interval A3* are defined only by no experience with speculation of the trader. The results support an assumption that traders enjoying risk tend to make shorter trades, thus report a speculative behavior. Contrary students with no previous experience with speculation will most probably hold trades longer.

After exploring the hypotheses which affect the length of trades, further research will be focused on examination of difference within length of profitable trades and loss trades, taking into account results of GUHA method, to obtain information about potential disposition effect.

Acknowledgements

This work has been supported by the Czech Science Foundation (GAČR) under project 18-13951S and SP2020/11, an SGS research project of VSB – TU Ostrava, and within RRC/10/2017 support scheme of the Moravian-Silesian Region.

References

[1] Lisp-miner. online. URL `http://lispminer.vse.cz/index.html`.

[2] Adams, R. B., and Funk, P.: Beyond the Glass Ceiling: Does Gender Matter? *Management Science* **58** (2012), 219–235.

[3] Barber, B. M., and Odean, T.: Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics* **116** (2001), 261–292.

[4] Barberis, N., and Xiong, W.: What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation. *The Journal of Finance* **64** (2009), 751–784.

[5] Cheng, T. Y., Lee, C. I., and Lin, C. H.: An examination of the relationship between the disposition effect and gender, age, the traded security, and bull – bear market conditions. *Journal of Empirical Finance* **21** (2013), 195–213.

[6] Dvorackova, H., and Jochec, M.: Disposition effect in currency trading: An evidence from experimental student games. *9th International Scientific Conference Managing and Modelling of Financial Risks*. Ostrava: VSB–Technical University Ostrava.

[2] [16], [14], [17], [21], [13], [4], [5], [8], [7]

[7] Dvorackova, H., Jochec, M., and Tichy, T.: Disposition Effect in Currency Trading: An Evidence from Experimental Student Games. *Review of Research and Social Intervention* **64** (2019), 246–261.

[8] Eom, Y.: The opposite disposition effect: Evidence from the korean stock index futures market. *Finance Research Letters* **26** (2018), 261–265.

[9] Gainsbury, S. M., King, D. L., Russell, A. M., Delfabbro, P., and Hing, N.: Virtual addictions: An examination of problematic social casino game use among at-risk gamblers. *Addictive Behaviors* **64** (2017), 334–339.

[10] Hájek, P., Havel, I., and Chytil, M.: The guha method of automatic hypotheses determination. *Computing 1* **4** (1966), 293–308.

[11] Hájek, P., Holeňa, M., and Rauch, J.: The GUHA method and its meaning for data mining. *Journal of Computer and System Sciences* **76** (2010), 34–48.

[12] Kahneman, D., and Tversky, A.: Prospect theory: An analysis of decision under risk. *Econometrica* **46** (1979), 171–185.

[13] Locke, P. R., and Mann, S.: Professional trader discipline and trade disposition. *Journal of Financial Economics* **76** (2005), 401–444.

[14] Odean, T.: Are investors reluctant to realize their losses. *Journal of Finance* **53** (1998), 1775–1798.

[15] Renc, Z., Kubát, K., and Kouřim, J.: An application of the guha method in medicine. *International Journal of Man-Machine Studies* **10** (1978), 29–35.

[16] Shefrin, H., and Statman, M.: The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance* **40** (1985), 777–790.

[17] Shiller, R., and Case, K.: The behavior of home buyers in boom and post-boom markets, 1988. Cowles Foundation Discussion Papers.

[18] Stanley, M.: Social gambling: click here to play. *Morgan Stanley Research: Global* **23** (2012), 171–175.

[19] Turunen, E.: Interpreting guha data mining logic in paraconsistent fuzzy logic framework. *Lecture Notes in Computer Science* **5783** (2009), 284–293.

[20] Vlček, V.: An application of the guha method to chemical engineering. *International Journal of Man-Machine Studies* **15** (1981), 299–307.

[21] Weber, M., and Camerer, C. F.: The disposition effect in securities trading: an experimental analysis. *Journal of Economic Behavior & Organisation* **33** (1998), 167–184.

[22] Zvárová, J., Preiss, J., and Sochorová, A.: Analysis of data about epileptic patients using the guha method. *International Journal of Medical Informatics* **45** (1997), 59–64.

Efficiency of Credit Risk Management of Selected Commercial Banks in the Czech Republic

Xiaoshan Feng[1]

Abstract. This paper focuses on measuring the efficiency of credit risk management of banks in the Czech Republic. The main aim is to evaluate the performance of credit risk management reflected on global technical efficiency and pure technical efficiency and productivity change of selected commercial banks in the Czech Republic. We apply the Data Envelopment Analysis (DEA) on 12 commercial banks in the Czech Republic over the period from 2012 to 2018. We find that under different assumptions of returns to scale, the efficiency score of selected banks ranges from 0.28 to 0.43 under CCR model and ranges from 0.61 to 0.71 under BCC model. Moreover, strong evidence from the Malmquist Index showed that the Czech banking sector improved its efficiency during the past 7 years due to innovation in credit risk management. Furthermore, we employ logistic regression to find out that the likelihood of a bank being efficient increases with larger size and higher yearly change of CAR, while decrease with higher GDP growth rate with lag effect under CCR model. And increases with lower profitability under BCC model.

Keywords: data envelopment analysis, Malmquist index, credit risk management, the Czech Republic, logistic regression

JEL Classification: G21, C35, C67, C80, C61, C58
AMS Classification: 62M10, 91G40, 91G70

1 Introduction

As a robust growth rate of the volume of bank loans in recent years, commercial banks need to be more cautious about the quality of their assets. Therefore, to investigate the efficiency of credit risk management of commercial banks becomes one of the most important steps to measure the overall soundness of the banking sector.

As for the banking sector in the Czech Republic, there is abundant literature focus on the bank operational efficiency, and it is lack of timeliness. However, it is rare to investigate efficiency from the view of credit risk management of banks. Furthermore, scholars have investigated the determinants of bank efficiency through a dynamic panel data model in previous literature, while it is scarce research to figure out the determinants of credit risk efficiency from external and internal impacts. Therefore, research based on current credit risk management efficiency of the Czech banking sector and its determinants is the exhibited knowledge gap.

Motivated by this gap, the aim of this paper is to evaluate the performance of credit risk management reflected on global technical efficiency (CCR) and pure technical efficiency (BCC), and productivity change of selected commercial banks in the Czech Republic from 2012–2018. Furthermore, based on the efficiency results, we will get insight into the possible determinants of credit risk management efficiency.

We apply Data Envelopment Analysis and the Malmquist index (MI) to measure the efficiency of credit risk management and productivity change of 12 selected banks in the Czech Republic during the period from 2012–2018. Moreover, we employ a regression model to investigate the possible internal and external determinants of credit risk management efficiency. This paper is divided into five sections. The first section starts with the introduction and the last one ends with the conclusion. The second section includes the literature review. Section 3 presents a brief description of methodology and data collection. In the fourth section, the empirical results will be discussed.

[1] VSB – Technical University of Ostrava, Faculty of Economics, Department of Finance, Sokolská 33, 702 00 Ostrava

2 Literature Review

There is rich literature focused on bank efficiency with Data Envelopment Analysis (DEA). Řepková [12] finds that the efficiency scores from the BCC model have higher values than from the CCR model due to the elimination of deposit management inefficiency. To conduct a DEA estimation, the research employs labour and deposits as inputs, then loans and net interest income as two outputs. The study applied the Malmquist index to estimate the efficiency change in the Czech banks over time from 2001 to 2010. The negative growth on efficiency indicates the industry has lacked innovation or technological progress during the time. Moreover, the catch-up effect has a more accountable result than the frontier-shift effect.

To incorporate risk into bank efficiency has been raised among analysts due to the recent financial crisis. Tsolas and Charles [17] provide research on the efficiency profile of the Greek banking sector based on the DEA model, in which the financial risk is proxied by credit risk provisions. The study uses a satisficing DEA model that treats loan loss provision as a stochastic controllable input, which corresponding to the intermediation approach.

In real-life applications, undesirable outputs such as non-performing loans may present in the banking sector which needs to be minimized. There is an abundance of research that incorporates undesirable outputs into the analysis. Paradi and Zhu [9] have surveyed bank branch efficiency and performance research with DEA. The study mentioned three approaches when non-performing loans are incorporated in previous literature. The first is to leave NPL as an output but use the inverse value. The second method is to treat this undesirable output as input, which applied in other studies [11, 15]. The third one is to treat it as an undesirable output with an assumption of weak disposability, which requires that undesirable outputs can be reduced, but at a cost of fewer desirable outputs produced.

Partovi and Matousek [10] analyze technical and allocative efficiencies in Turkish banks from 2002 to 2017, under the assumption of constant returns to scale. The study applies a modified version of the DEA approach which employs a directional distance model to provide estimates of efficiency, with a focus on NPLs as an undesirable output.

Gaganis et al. [7] also examine the efficiency and productivity of a Greek bank's branches from 2000 to 2005. The finding shows that the inclusion of loan loss provisions as an input variable increases the efficiency score, then, fixed and random effects models were used to determine the impact of internal and external factors on the efficiency and productivity scores.

Notwithstanding there are abundant studies focused on the efficiency of the Czech banking sector, it is rare to investigate efficiency from the view of credit risk management of banks. Therefore, we incorporate non-performing loans as a proxy of credit risk to measure the efficiency of credit risk management in the banking sector.

3 Methodology and Data

Based on the previous literature, we apply Data Envelopment Analysis and the Malmquist index to measure the efficiency of credit risk management and productivity change of 12 selected banks in the Czech Republic during the period from 2012–2018. Moreover, we employ a regression model to investigate the possible internal and external determinants of probability of credit risk management efficiency. This section will briefly describe the essence behind each methodology and the collection of data for corresponding analysis.

3.1 Two Classic Models of Data Envelopment Analysis

DEA is a linear programming-based method, which introduced by Charnes, Cooper and Rhodes in 1978 [5]. DEA is used to evaluate the relative efficiency of a set of decision-making units (DMUs) with multiple inputs and multiple outputs. Then, Banker, Charnes and Cooper has proposed a model in 1984, named the BCC, which is an extended version of the CCR model. The main difference between these two models is different returns to scale. The CCR model assumes all DMUs are operating at an optimal scale, that is, constant returns to scale (CRS); While the BCC model assumes variable returns to scale (VRS).

In DEA models, we measure the efficiency of each DMU. One of the most frequently used methods to measure efficiency is by the ratio. Suppose we have n DMUs in the population, each DMU produces s outputs while consuming m inputs. Consider DMU_j, j represents n DMUs, x_{ri} and y_{ri} are the matrixes of inputs and outputs respectively. The efficiency rate of such a unit can be expressed as:

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \tag{1}$$

The efficiency rate is the ratio of the weighted sum of outputs to weighted sum of inputs. The DEA model assumed inputs and outputs should be non-negative. Let DMU_j to be evaluated on any trial be designated as DMU_o, where $o = (1,2,\dots,n)$.

A ratio of two linear functions can construct the linear-fractional programming model as follows:

$$\max_{v,u} \theta = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \tag{2}$$

$$\text{subject to } \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, (j = 1,2,\dots,n) \tag{3}$$

$$u_1, u_2, \dots, u_r \geq 0, (r = 1,2,\dots,s) \tag{4}$$

$$v_1, v_2, \dots, v_i \geq 0, (i = 1,2,\dots,m) \tag{5}$$

Where θ is the technical efficiency of DMU_o to be estimated, v_i $(i = 1,2,\dots,m)$ is the optimized weight of input and the output u_r $(r = 1,2,\dots,s)$. y_{rj} is observed amount of output of the r-th type for the j-th DMU, x_{ij} is observed amount of input of the i-th type for the j-th DMU.

Moreover, we will apply the Malmquist index to deal with our panel data, to evaluate the productivity change of a DMU between two time periods and is an example in comparative statistical analysis. Farrell developed the Malmquist index as a measurement of productive efficiency in 1957, then Fare decomposed MI into two terms in 1994, it can be defined as "Catch-up" and "Frontier-shift" terms. The catch-up term indicates the degree of a DMU improves or worsens its efficiency. The frontier-shift term is used to figure out the change in the efficient frontiers between two time periods. The Malmquist index is computed as the geometric mean of Catch-up and Frontier-shift.

$$MI = (\text{Catch-up}) \times (\text{Frontier-shift}) \tag{6}$$

When MI larger than 1, it means progress in the total factor productivity of the DMU_o from period 1 to period 2, while MI equals 1 means no change, and MI less than one indicates deterioration in the total factor productivity.

3.2 Data Selection for Measuring the Efficiency by DEA and MI

To make sure the results of applying DEA are accurate, the number of inputs and outputs, and DMUs must support the rule of thumb, which proposed firstly by Golany and Roll [8], then developed by Bowlin [4], that is, it should have three times the number of DMUs as there are input and output variables, if this condition will not be met, the results are not reliable [16].

In this paper, we have collected data of 12 commercial banks out of 23 commercial banks which include the large-size bank such as Česká spořitelna, ČSOB; Medium size such as MONETA money bank, Equa bank; Small size such as PPF bank, Air Bank. All data are collected from the annual report of each bank in a consolidated basis. In the data set, twelves banks are observed over seven years (2012–2018). In total, we have a balanced panel of 72 observations.

To select the inputs and outputs, there are basically three approaches when we apply DEA in the banking sector. The first one is production approach, in which, bank plays a role of producer, transfer some physical inputs such as labours into services [14]. The next one is intermediation approach, that is, bank acts as an intermediary, transfer funds or deposits from clients to loans and investments [3], which is more commonly used in analysing banking efficiency. Thirdly, the profitability approach measures how efficiently a bank maximizes its profits by decreasing expenses while increasing revenue [6].

This study aims to investigate the efficiency of credit risk management in the Czech banking sector. Therefore, we will apply the intermediation approach to measure the efficiency of credit risk management based on DEA model. To assess credit risk modelling in banking industry, Berg et al. [2] used non-performing loans as a proxy of credit risk in a nonparametric study of the bank production, Altunbas et al. [1] incorporate loan loss provisions to analyse the efficiency of Japanese bank.

Therefore, based on previous literature, we proxy credit risk by the ratio of non-performing loans to total gross loans, so-called NPL ratio. Then use loan loss provision ratio to represents the ratio of provision and

non-performing loan, which is primarily to reflect commercial banks' abilities to compensate for loan losses and to protect against credit risk. Generally, this paper developed two inputs and one output, with 12 DMUs which satisfy the rule of thumb. *Input* x_1 is loan loss provision, *Output* y loans and receivables as output, non-performing loan as undesirable output, based on the treatment of undesirable output from previous literature, NPL is transformed as *Input* x_2.

3.3 Logistic Regression Model

Furthermore, after we obtain efficiency score based on the CCR model and the BCC model, we can estimate the determinants of banking credit risk management efficiency using regression model, Since the efficiency can be measured as binary outcomes, we can model the conditional probabilities of the response outcome, rather than give a binary result. Therefore, we apply logistic regression model in this paper. The logistic model could be interpreted based on an underlying linear model, shows below:

$$Y_{i,t} = \beta_0 + X'_{i,t}\beta + \epsilon_{i,t}, i = 1, \dots, N, t = 1, \dots, T. \tag{7}$$

Where the subscripts i and t denote the cross-sectional (N) and time dimension (T) of the panel data, respectively. There is k $(k = 1, \dots, K)$ regressor in $X_{i,t}$, not including a constant term. $X_{i,t}$ is explanatory variable value for i-th section at t-th dimension; β_0 is the intercept; β is the slope coefficient of a $(k \times 1)$ vector. The variable $\epsilon_{i,t}$, can be called as the error term in the relationship, represents factors other than explanatory variables that affect dependent variables. Since we have a binary output variable $Y_{i,t}$, and we want to model the conditional probability $p(Y_{i,t} = 1|X'_{i,t} = x_{i,t})$ as a function of $x_{i,t}$:

$$\pi(x_{i,t}) = p(Y_{i,t} = 1|X'_{i,t} = x_{i,t}) \tag{8}$$

The logistic regression model can be constructed as follows:

$$\log\frac{\pi(x_{i,t})}{1 - \pi(x_{i,t})} = \beta_0 + X'_{i,t}\beta, i = 1, \dots, N, t = 1, \dots, T. \tag{9}$$

3.4 Data Selection for Logistic Regression Model

The starting point for model construction, we aim to investigate the determinants of efficiency of bank's credit risk management, therefore, our dependent variable is efficiency score obtained from previous mentioned DEA model. The independent variables are, respectively, size of the bank, which measured as the natural logarithm of the value of total assets in Czech koruna; Capital Adequacy Ratio (CAR), measured by dividing a bank's total capital by its risk-weighted assets; Return on average assets (ROAA), it is calculated as the ratio of net income to average total assets; GDP growth rate, which is the year-on-year annual GDP growth rate. The binary outcomes are measured by 1 and 0, which represent *DMU* is efficient and inefficient, respectively.

4 Empirical Results

Based on methodology described in Section 3.1, we can measure the efficiency under the two classic models, which are, the CCR model and the BCC model. To compute the empirical result, we used DEA SolverPro™. Table 1 presents the average scores of efficiencies under the CCR model and the BCC model. We find that under different assumptions of returns to scale, the average global technical efficiency of selected banks ranges from 0.28 to 0.43 and the average of pure technical efficiency ranges from 0.61 to 0.71.

	2012	2013	2014	2015	2016	2017	2018	Average
CCR	0.33	0.43	0.31	0.34	0.28	0.28	0.31	0.33
BCC	0.71	0.64	0.61	0.65	0.65	0.61	0.71	0.65

Table 1 Efficiency score under the CCR and the BCC (Industry average)

The average of efficiency scores under the CCR model is 0.33 for selected banks. It provides the results that if bank can produce its outputs on the efficient frontier, compared to the inputs which currently being used, only 33% of the inputs needed. Put differently, if selected banks in the Czech Republic adopt best practice technology, they can reduce inputs (i.e. more efficient credit risk management, and reduction in loan-loss

provisions) at least 69% and still produce the same level of outputs. However, the potential reduction in inputs from adopting best practices varies from bank to bank. The results of BCC model showed a higher efficiency score than results from the CCR model. However, ČSOB still performed well among selected banks.

Based on results, ČSOB has not only global technical efficiency but it also has pure technical efficiency, which forms the reference set for inefficient banks. For a specific DMU, if the technical efficiency scores differ in the CCR model and the BCC model, then the DMU presents an inefficiency of scale. EQUA bank has been one of the most inefficient banks under the CCR model, while it showed inverse results under the BCC model, which presents a large inefficiency of scale.

	CS	KB	CSOB	RB	CRED	AIR	PPF	FIO	MMB	UNI	JT	EQUA
CCR	0.89	0.77	1.00	0.23	0.01	0.03	0.15	0.04	0.08	0.50	0.13	0.08
BCC	0.96	0.86	1.00	0.57	0.41	0.46	0.52	0.53	0.34	0.68	0.56	0.97

Table 2 Efficiency score under the CCR and the BCC (Yearly average)

Next, we applied the Malmquist index (MI) to measure productivity change of a DMU between two time periods, Table 3 showed the overall technological productivity and its decomposition.

	Catch-up	Frontier	Malmquist
2012=>2013	1.79	1.04	1.85
2013=>2014	0.73	1.32	0.97
2014=>2015	1.08	1.08	1.17
2015=>2016	0.92	1.69	1.56
2016=>2017	1.06	1.62	1.72
2017=>2018	1.06	1.25	1.33
Mean	1.07	1.31	1.40

Table 3 Malmquist index and its decomposition

The measurement of the MI reflects the trend of overall productivity of the sample banks during periods. The Malmquist (total factor productivity) exceeded unity except period 2 (2013–2014), which shows that selected banks continuously improved their development in credit risk management and achieved an average 40% increase in the overall efficiency during past 6 periods, although slightly regress in second period.

Furthermore, we can discover the results from decomposition, it can be found that the frontier-shift effect was primarily accountable for the productivity growth rather than the catch-up effect. On average, the technical efficiency was growing at an annual rate of 7%, and the technological progress increased by 31% annually. Therefore, the technological progress (frontier-shift effect) was the main reason for the growth of total factor productivity in selected banks in the Czech Republic.

Under logistic regression, we estimated the likelihood of a DMU is efficient given the applicants SIZE, CAR, ROAA and GDPG. Table 4 presents the results of regression. We found only size and profitability were statistically significant in BCC model. Under CCR model, the probability of a bank being efficient increases with larger size and higher yearly change of CAR, while decrease with higher GDP growth rate with lag effect. And the probability increases with larger size and lower profitability under BCC model.

	CCR model		BCC model
Variables		Variables	
C	-132.3267***	C	-16.233***
	[39.2733]		[0.2365]
SIZE	9.500***	SIZE	1.4002***
	[2.706]		[0.4800]
LAGGDPG	-0.2766**	ROAA	-2.290*
	[0.1399]		[0.1536]
DCAR	0.416*		
	[0.238]		

Table 4 Regression results

5 Conclusion

In this paper, we aim to measure the efficiency of credit risk management of selected banks in the Czech Republic during the period from 2012–2018. Therefore, we apply two classic DEA models and the Malmquist index to measure efficiency and productivity change of 12 selected commercial banks in the Czech Republic. We find out ČSOB was the most efficient bank on credit risk management under the CCR model and the BCC model. Moreover, the Czech banking sector improved its efficiency during the investigated period mainly due to innovation (frontier-shift effect) in credit risk management. Furthermore, we employ logistic regression to get insight into the possible determinants of the likelihood of a bank managing credit risk efficiently. The probability of bank acting efficient increases with larger size and lower profitability under VRS, while it decreases with lower GDP growth rate with lagged effect in CCR model. Our findings provide strong evidence to explain how ČSOB was the most efficient DMU, that is, the second-largest bank in the Czech Republic and the lowest ROA among large-size banks. These findings also hold for bank efficiency in the Czech Republic within 2001–2012 [13].

The main contribution of this paper is that we provide a big picture of efficiency on credit risk management of the Czech banking sector and find the main drivers of productivity growth. Furthermore, we investigate the determinants of probability for efficiency on credit risk management to provide more specific insight into the Czech banking sector.

Acknowledgements

The author was supported through the Czech Science Foundation (GACR) under project 20-25660Y and moreover by SP2020/11, an SGS research project of VSB-TU Ostrava. The support is greatly acknowledged.

References

[1] ALTUNBAS, Y., LIU, M., MOLYNEUX, P., & SETH, R. (2000). Efficiency and risk in Japanese banking. *Journal of Banking & Finance*, Elsevier, vol. 24(10), pp. 1605–1628.

[2] BERG, S. A., FORSUND, F. R., & JANSEN, E. S. (1992). Malmquist Indexes of Productivity Growth During the Deregulation of Norwegian Banking, 1980–89. *The Scandinavian Journal of Economics* pp. S211–S228.

[3] BERG, S. A., FORSUND, F. R., HJALMARSSON, L., & SUOMINEN, M. (1993). Banking efficiency in the Nordic countries. *Journal of Banking and Finance*, 17: pp. 371–388.

[4] BOWLIN, W. F. (1998). Measuring Performance: An Introduction to Data Envelopment Analysis (DEA). *Journal of Cost Analysis* 7, pp. 3–27.

[5] CHARNES, A., COOPER, W. W., & RHODES, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, Vol. 2, No. 6, pp.429–444.

[6] EKEN, M., KALE, S., & SELIMLER, H. (2014). Analyzing the Efficiency of European Banks: A DEA-Based Risk and Profitability Approach. *Business Science Reference*, pp. 28–55.

[7] GAGANIS, C., LIADAKI, A., DOUMPOS, M., & ZOPOUNIDIS, C. (2009). Estimating and analyzing the efficiency and productivity of bank branches: Evidence from Greece. *Managerial Finance*, Vol. 35 No. 2, pp. 202–218. https://doi.org/10.1108/03074350910923518.

[8] GOLANY, B. & ROLL, Y. (1989). An Application Procedure for DEA. *Omega* 17, pp. 237–250.
[9] PARADI, J. C. & ZHU, H. (2013). A survey on bank branch efficiency and performance research with data envelopment analysis. *Omega*. Elsevier, 41(1), pp. 61–79. doi: 10.1016/j.omega.2011.08.010.
[10] PARTOVI, E. & MATOUSEK, R. (2019) Research in International Business and Finance Bank efficiency and non-performing loans: Evidence from Turkey. *Research in International Business and Finance*. Elsevier, 48 (March 2018), pp. 287–309. doi: 10.1016/j.ribaf.2018.12.011.
[11] PURI, J. & YADAV, S. P. (2014). Expert Systems with Applications A fuzzy DEA model with undesirable fuzzy outputs and its application to the banking sector in India. *Expert Systems With Applications*. Elsevier Ltd., 41(14), pp. 6419–6432. doi: 10.1016/j.eswa.2014.04.013.
[12] ŘEPKOVÁ, I. (2012). *Measuring the efficiency in the Czech banking industry: Data Envelopment Analysis and Malmquist index*, pp. 781–786.
[13] ŘEPKOVÁ, I. (2015). Banking Efficiency Determinants in the Czech Banking Sector. 23(October 2014), pp. 191–196. doi: 10.1016/S2212-5671(15)00367-6.
[14] SEALEY, C. W. & LINDLEY, J. T. (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions. *The Journal of Finance,* 32(4): pp.1251–1266.
[15] TOLOO, M., HANČLOVA, J. (2019). Multi-valued measures in DEA in the presence of undesirable outputs. *OMEGA*, https://doi.org/10.1016/j.omega.2019.01.010.
[16] TOLOO, M., TICHY, T. (2015). Two alternative approaches for selecting performance measures in data envelopment analysis. *Measurement*, Vol. 65, pp. 29–40.
[17] TSOLAS, I. E. & CHARLES, V. (2015). Expert Systems with Applications Incorporating risk into bank efficiency: A satisficing DEA approach to assess the Greek banking crisis. *Expert Systems With Applications*. Elsevier Ltd., 42(7), pp. 3491–3500.

Coordination of Supply Chains by Auctions

Petr Fiala[1], Renata Majovská[2]

Abstract. Supply chain is a decentralized system where material, financial and information flows connect economic agents. There is much inefficiency in supply chain behavior. Recently, considerable attention of researchers is drowned to provide some incentives to adjust the relationship of supply chain agents to coordinate the supply chain, i.e., the total profit of the decentralized supply chain is equal to that achieved under a centralized system. Auctions are important market mechanisms for the allocation of goods and services. A complex trading model between layers of the supply chain is proposed in the paper. The model is based on so called multidimensional auctions. There is possible to formulate multidimensional auctions as mathematical programming problems. Iterative methods are used to solve the problems.

Keywords: supply chain, coordination, auctions.

JEL Classification: L14, D44, C61
AMS Classification: 90C09

1 Introduction

Supply chain management is about matching supply and demand with inventory management. When one or more agents of the supply chain try to optimize their own profits, system performance may be hurt. Developing strategies to decrease the risk faced by the retailer is becoming more and more critical in a supply chain, especially in the global marketplace where firm-to-firm competition is being replaced by supply-chain-to-supply-chain competition. There is much inefficiency in supply chain behaviour. Recent years have seen a growing interest among researchers and practitioners in the field of supply chain management. Most of coordination mechanisms are based on game theory models (see [6]) and contracts between agents of the supply chain. However, little work has been done on using auctions for supply chain coordination. The paper proposes a complex trading model for coordination of agents in supply chain. The model is based on, so called, multidimensional auctions.

Auctions are important market mechanisms for the allocation of goods and services. Multidimensional auctions arise by extensions of standard auction models. Multi-item auctions can place bids on combinations of items, so called combinatorial auctions. The advantage of combinatorial auctions is that the bidder can more fully express his preferences. Multi-type auction model includes forward, reverse and double auctions. Multi-round, so called iterative, methods are used for analysis of combinatorial auctions and for negotiation process. Allowing bidders more fully to express preferences often leads to improved economic efficiency and greater auction profit. The multi-item model for multi-type auction is modelled and solved by multi-round approach. The proposed model illustrates the possibility to formulate and solve multidimensional auctions as mathematical programming problems. The model is based on a linear programming model and its extensions. A solution procedure is presented. The procedure is based on primal-dual algorithms.

The rest of the paper is organized as follows. Section 2 presents the supply chain coordination problem and the possibility to solve the problem. Section 3 summarizes the basics of auctions. In Section 4, a complex trading model based on multidimensional auctions is formulated. Multi-round iterative auctions as a solution approach are presented in Section 5. Finally, Section 6 provides conclusions.

2 Supply chain coordination

Supply chain is a decentralized system composed from layers of potential suppliers, producers, distributors, retailers and customers etc., where agents are interconnected by material, financial and information flows. A supply chain is the collection of steps that a system takes to transform raw components into the final

[1] University of Economics, Prague, department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, Czech Republic, pfiala@vse.cz.
[2] University of Finance and Administration, Prague, department of Computer Science and Mathematics, Estonska 500, 101 00 Praha 10, Czech Republic, renata.majovska@mail.vsfs.cz.

product. There is much inefficiency in supply chain behavior. When one or more agents of the supply chain try to optimize their own profits, system performance may be hurt. There are many concepts and strategies applied in designing and managing supply chains (see [9]). The expanding importance of supply chain integration presents a challenge to research to focus more attention on supply chain modelling (see [10, 11]).

Analysis and modelling of supply chains goes through the following phases: designing, managing, performance measurement, performance improvement. The most important part of managing phase is the coordination of individual activities to be optimal in terms of the whole system. Supply chains are decentralized systems. A centralized system can be taken as a benchmark. The question is: How to coordinate the decentralized supply chain to be efficient as the centralized one?

We made some experiments with evaluation of different supply network structures. The supplier rarely has complete information about customer's cost structure. However, the quantity the customer will purchase and therefore supplier's profit depends on that cost structure. Somehow, the supplier will have to take this information asymmetry into account. The numbers of suppliers and customers are denoted by m, n, respectively. The symbol S_i represents i-th seller while the symbol B_j represents j-th buyer. The seller-buyer relations in supply chain can be taken as decentralized or centralized (coordinator between suppliers and customers).

Most supply networks are composed of independent agents with individual preferences. It is expected that no single agent has the power to optimize the supply network. Each agent will attempt to optimize his own preference, knowing that all of the other agents will do the same. This competitive behaviour does not lead the agents to choose polices that optimize overall supply chain performance due to supply chain externalities. The agents can benefit from coordination and cooperation. The typical solution is for the agents to agree to a set of transfer payments that modifies their incentives, and hence modifies their behaviour. Many types of transfer payments are possible.

The problem of coordination in supply chains involves multiple agents with multiple goals. Coordination between suppliers and customers can be provided through information sharing. A seller S_i and a buyer B_j have information and analytical tools for their problem representations. A coordinator helps by information sharing and by formulation of a joint problem representation (see [5]).

3 Auctions

Auctions are important market mechanisms for the allocation of goods and services. Auctions are preferred often to other common processes because they are open, quite fair, and easy to understand by agents, and lead to economically efficient outcomes. Many modern markets are organized as auctions. Design of auctions is a multidisciplinary effort made of contributions from economics, operations research, informatics, and other disciplines. Auction theory has caught tremendous interest from both the economic side as well as the Internet industry. An auction is a competitive mechanism to allocate resources to buyers based on predefined rules. These rules define the bidding process, how the winner is determined, and the final agreement.

The auction mechanism is a process that transforms bids on allocation of objects to winners and determining the payments that must be paid by the buyer and the seller receives. An auction provides a mechanism for negotiation between buyers and sellers. Multidimensional auctions are a generalization of standard auctions. These auctions can be classified:

- multi-item auction,
- multi-type auction,
- multi-round auction.

Multi-item auctions can place bids on combinations of items, so called combinatorial auctions. Combinatorial auctions (see [3, 2]) are those auctions in which bidders can place bids on combinations of items, so called bundles. The advantage of combinatorial auctions is that the bidder can more fully express his preferences. This is particular important when items are complements. The auction designer also derives value from combinatorial auctions. Allowing bidders more fully to express preferences often leads to improved economic efficiency and greater auction revenues.

There are several types of auctions (forward, reverse, and double). Forward auctions are oriented to the sale, with one seller and multiple buyers. Reverse auctions are oriented to purchase, with only one buyer and multiple sellers. Double auctions combine the two previous types and mediate an exchange between

multiple sellers and multiple buyers. There is an effort to propose a general multi-type auction that covers all the types.

In the iterative approach, there are multiple rounds of bidding and allocation and the problem is solved in an iterative and incremental way. Iterative combinatorial auctions are attractive to bidders because they learn about their rivals' valuations through the bidding process, which could help them to adjust their own bids. There are possible combinations of the multidimensional characteristics. However, alongside their advantages, combinatorial auctions raise a host of questions and challenges.

4 Complex trading model

We propose a complex trading model based using of iterative process for combinatorial multi-type auctions.

4.1 Multi-item auctions

Many types of combinatorial auctions can be formulated as mathematical programming problems. From different types of combinatorial auctions, we present a forward auction of indivisible items with one seller and multiple buyers. Let us suppose that one seller S offers a set R of r items, j = 1, 2, ..., r, to n potential buyers $B_1, B_2, ..., B_n$.

Items are available in single units. A bid made by buyer B_i, i = 1, 2, ..., n, is defined as

$$b_i = \{C, p_i(C)\}, \tag{1}$$

where $C \subseteq R$, is a combination of items,

$p_i(C)$ is the offered price by buyer B_i for the combination of items C.

The objective is to maximize the revenue of the seller given the bids made by buyers. Constraints establish that no single item is allocated to more than one buyer. Binary variables are introduced for model formulation:

$x_i(C)$ is a binary variable specifying if the combination C is assigned to buyer B_i ($x_i(C) = 1$).

The forward auction can be formulated as follows

$$\begin{gathered} \sum_{i=1}^{n} \sum_{C \subseteq R} p_i(C) x_i(C) \to \max \\ \text{subject to} \sum_{i=1}^{n} \sum_{C \subseteq R} x_i(C) \leq 1, \quad \forall j \in R, \\ x_i(C) \in \{0,1\}, \quad \forall C \subseteq R, \forall i, i = 1,2, \dots, n. \end{gathered} \tag{2}$$

The objective function expresses the revenue. The constraints ensure that overlapping sets of items are never assigned. The problem (2) is called the winner determination problem.

4.2 Multi-type auctions

We present a reverse auction of indivisible items with one buyer and several sellers. This type of auction is important for supplier selection problem. Let us suppose that m potential sellers $S_1, S_2, ..., S_m$ offer a set R of r items, j = 1, 2, ..., r, to one buyer B. A bid made by seller S_h, h = 1, 2, ..., m, is defined as

$$b_h = \{C, c_h(C)\},$$

where $C \subseteq R$ is a combination of items,

$c_h(C)$ is the offered price by seller S_h for the combination of items C.

The objective is to minimize the cost of the buyer given the bids made by sellers. Constraints establish that the procurement provides at least set of all items.

Binary variables are introduced for model formulation:

$y_h(C)$ is a binary variable specifying if the combination C is bought from seller S_h ($y_h(C) = 1$).

The reverse auction can be formulated as follows

$$\sum_{h=1}^{m}\sum_{C\subseteq R} c_h(C)y_h(C) \to \min$$
$$\text{subject to} \sum_{h=1}^{m}\sum_{C\subseteq R} y_h(C) \geq 1, \quad \forall j \in R, \tag{3}$$
$$y_h(C) \in \{0,1\}, \quad \forall C \subseteq R, \forall h, h = 1,2,\dots,m.$$

The objective function expresses the cost. The constraints ensure that the procurement provides at least set of all items.

Double auctions (auctions with multiple buyers and multiple sellers) are becoming increasing popular in electronic commerce. For double auctions, the auctioneer is faced with the task of matching up a subset of the buyers with a subset of the sellers. The profit of the auctioneer (supply chain) is the difference between the prices paid by the buyers and the prices paid to the sellers. The objective is to maximize the profit of the auctioneer given the bids made by sellers and buyers. Constraints establish the same conditions as in single-sided auctions.

We present a double auction problem of indivisible items with multiple sellers and multiple buyers. Let us suppose that m potential sellers $S_1, S_2, \dots, S_m$ offer a set R of r items, $j = 1, 2, \dots, r$, to n potential buyers $B_1, B_2, \dots, B_n$.

A bid made by seller S_h, $h = 1, 2, \dots, m$, is defined as $b_h = \{C, c_h(C)\}$,

a bid made by buyer B_i, $i = 1, 2, \dots, n$, is defined as $b_i = \{C, p_i(C)\}$,

where $C \subseteq R$, is a combination of items,

$c_h(C)$, is the offered price by seller S_h for the combination of items C,

$p_i(C)$, is the offered price by buyer B_i for the combination of items C.

Binary variables are introduced for model formulation:

$x_i(C)$ is a binary variable specifying if the combination C is assigned to buyer B_i ($x_i(C) = 1$),

$y_h(C)$ is a binary variable specifying if the combination C is bought from seller S_h ($y_h(C) = 1$).

$$\sum_{i=1}^{n}\sum_{C\subseteq R} p_i(C)x_i(C) - \sum_{h=1}^{m}\sum_{C\subseteq R} c_h(C)y_h(C) \to \max$$
$$\text{subject to} \sum_{i=1}^{n}\sum_{C\subseteq R} x_i(C) \leq \sum_{h=1}^{m}\sum_{C\subseteq R} y_h(C), \quad \forall j \in R, \tag{4}$$
$$x_i(C) \in \{0,1\}, \quad \forall C \subseteq R, \forall i, i = 1,2,\dots,n,$$
$$y_h(C) \in \{0,1\}, \quad \forall C \subseteq R, \forall h, h = 1,2,\dots,m.$$

The objective function expresses the profit of the auctioneer. The constraints ensure for buyers to purchase a required item and that the item must be offered by sellers.

The formulated combinatorial double auction can be transformed to a combinatorial single-sided auction. Substituting $y_h(C)$, $h = 1, 2, \dots, m$, with $1 - x_i(C)$, $i = n+1, n+2, \dots, n+m$, and substituting $c_h(C)$, $h = 1, 2, \dots, m$, with $p_i(C)$, $i = n+1, n+2, \dots, n+m$, we get a model of a combinatorial single-sided auction.

$$\sum_{i=1}^{n+m}\sum_{C\subseteq R} p_i(C)x_i(C) - \sum_{i=n+1}^{n+m}\sum_{C\subseteq R} p_i(C) \to \max$$
$$\text{subject to} \sum_{i=1}^{n+m}\sum_{C\subseteq R} x_i(C) \leq m, \quad \forall j \in R, \tag{5}$$
$$x_i(C) \in \{0,1\}, \quad \forall C \subseteq R, \forall i, i = 1,2,\dots,n+m.$$

The model (5) can be solved by methods for single-sided combinatorial auctions. The specific forward (2) or reverse (3) auctions can be modeled as special cases of the model (5).

4.3 Multi-round auctions

The key challenge in the iterative combinatorial auctions design (see [7]) is to provide information feedback to the bidders after each iteration (see [8]). Pricing was adopted as the most intuitive mechanism of providing feedback. In contrast to the single-item single-unit auctions, pricing is not trivial for iterative combinatorial auctions. The main difference is the lack of the natural single-item prices. With bundle bids setting independent prices for individual items is not obvious and often even impossible. Different pricing schemes are introduced and discussed their impact on the auction outcome.

A set of prices $p_i(C)$, $i = 1, 2, \ldots, n$, $C \subseteq R$ is called:

- linear, if $\forall i, C: p_i(C) = \sum_{j \in S} p_i(j)$,
- anonymous, if $\forall k, l, C: p_k(C) = p_l(C)$.

Prices are linear if the price of a bundle is equal to the sum of the prices of its items, and anonymous if the prices of the same bundle are equal for every bidder. The simple pricing scheme with linear anonymous prices will be used. Linear anonymous prices are easily understandable and usually considered fair by the bidders. The communication costs are also minimized, because the amount of information to be transferred is linear in the number of items.

A set of prices $p_i(S)$ is called compatible with the allocation $x_i(C)$ and valuations $v_i(C)$, if

$$\forall i, C: x_i(C) = 0 \Leftrightarrow p_i(C) > v_i(C) \text{ and } x_i(C) = 1 \Leftrightarrow p_i(C) \leq v_i(C).$$

The set of prices is compatible with the given allocation at the given valuations if and only if all winning bids are higher than or equal to the prices and all loosing bids are lower than the prices (assuming the bidders bid at their valuations).

Compatible prices explain the winners why they won and the losers, why they lost. In fact, informing the bidders about the allocation $x_i(C)$ is superfluous, if compatible prices are communicated. However, not every set of compatible prices provides the bidder with meaningful information for improving bids in the next auction iteration. Another important observation is the fact that linear compatible prices are harder and often even impossible to construct, when the bidder valuations are super- or sub-additive.

A set of prices $p_i(C)$ is in competitive equilibrium with the allocation $x_i(C)$ and valuations $v_i(C)$, if

1. The prices $p_i(C)$ are compatible with the allocation $x_i(C)$ and valuations $v_i(C)$.
2. Given the prices $p_i(C)$, there exists no allocation with larger total revenue than the revenue of the allocation $x_i(C)$.

The idea behind this concept is to define prices characterizing the optimal allocation. The prices may not be too low to violate the compatibility condition 1, but they may not be too high to violate the condition 2. In general, one can show that the existence of competitive equilibrium prices implies optimality of the allocation.

5 Solving the trading model

Multi-round iterative auctions can be taken as a solution approach. There is a strong interrelationship between the iterative auctions and the primal-dual linear programming algorithms. One way of reducing some of the computational burden in solving the winner determination problem is to set up a fictitious market that will determine an allocation and prices in a decentralized way. In the iterative approach, there are multiple rounds of bidding and allocation and the problem is solved in an iterative and incremental way. Iterative combinatorial auctions are attractive to bidders because they learn about their rivals' valuations through the bidding process, which could help them to adjust their own bids.

There is a connection between efficient auctions for many items, and duality theory. The simplex algorithm can be taken as static approach to determining the outcome. Alternatively, the primal-dual algorithm can be taken as a decentralized and dynamic method to determine the pricing equilibrium. A primal-dual algorithm usually maintains a feasible dual solution and tries to compute a primal solution that is both feasible and satisfies the complementary slackness conditions. If such a solution is found, the algorithm terminates. Otherwise the dual solution is updated towards optimality and the algorithm continues with the next iteration. The fundamental work of Bikhchandani and Ostroy (see [1]) demonstrates a strong interrelationship between the iterative auctions and the primal-dual linear programming algorithms. A primal-dual linear

programming algorithm can be interpreted as an auction where the dual variables represent item prices. The algorithm maintains a feasible allocation and a price set, and it terminates as the efficient allocation and competitive equilibrium prices are found. The duality theory can be applied also for multi-criteria auctions (see [4]).

For the winner determination problem, we will formulate the LP relaxation and its dual. Consider the LP relaxation of the winner determination problem (2):

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{C\subseteq R} v_i(C)x_i(C) \to \max \\ &\text{subject to} \sum_{i=1}^{n}\sum_{C\subseteq R} x_i(C) \le 1, \quad \forall j \in R, \\ &x_i(C) \ge 0, \quad \forall C \subseteq R, \forall i, i = 1,2,\dots,n. \end{aligned} \tag{6}$$

The corresponding dual to problem (6)

$$\begin{aligned} &\sum_{i=1}^{n} p(i) + \sum_{j\in C} p(j) \to \min \\ &\text{subject to} \quad p(i) + \sum_{j\in C} p(j) \ge v_i(C) \quad \forall i, C, \\ &p(i), p(j) \ge 0, \quad \forall i, j, \end{aligned} \tag{7}$$

The dual variables $p(j)$ can be interpreted as anonymous linear prices of items, the term $\sum_{j\in C} p(j)$ is then the price of the bundle C and $p(i) = \max_{C}[v_i(C) - \sum_{j\in C} p(j)]$ is the maximal utility for the bidder i at the prices $p(j)$.

Several auction formats based on the primal-dual approach have been proposed in the literature. Though these auctions differ in several aspects, the general scheme can be outlined as follows:

1. Choose minimal initial prices.
2. Announce current prices and collect bids. Bids have to be higher or equal than the prices.
3. Compute the current dual solution by interpreting the prices as dual variables. Try to find a feasible allocation, an integer primal solution that satisfies the stopping rule. If such solution is found, stop and use it as the final allocation. Otherwise update prices and go back to 2.

6 Conclusions

The proposed trading model has some advantages in comparisons with other approaches. Auctions are the important subject of an intensive economic research. Auctions are very popular mechanisms in practice and it is not necessary to conclude contracts between agents. Restrictions are only rules of the auction process. Agents need to monitor the price only. Auctions can be made via Internet. The approach coordinates layers of agents in supply chain, not only individual agents as by contracts. Repeating the procedure of coordination between layers of agents makes it possible to coordinate the entire supply chain.

A possible flexible approach for modelling and solving such auctions is presented. The analysis of the simple cases gives recommendations for more complex real problem. The combination of such models can give more complex views on auctions. For example, the model can be extended by multiple objectives. Complex problems require consider multiple objectives, not just profit. Multi objective linear programming (MOLP) problem can be used for the extended model. The problem is possible to solve by interactive methods of MOLP problems.

Acknowledgements

This work was supported by the grant No. IGA F4/66/2019, Faculty of Informatics and Statistics, University of Economics, Prague. The paper is a result of an institutional research project no.7429/2020/02 “System approach to selected ICT trends” supported by University of Finance and Administration, Prague.

References

[1] BIKHCHANDANI, S. & OSTROY, J. M. (2002). The package assignment model. *Journal of Economic Theory*, *107*, 377–406.

[2] CRAMTON, P., SHOHAM, Y., & STEINBERG, R. (Eds.) (2006). *Combinatorial Auctions*. Cambridge: MIT Press.

[3] DE VRIES, S. & VOHRA, R. V. (2003). Combinatorial auctions: A survey. *INFORMS Journal of Computing*, *15*, 284–309.

[4] FIALA, P. (1981). Duality in linear vector optimization. *Ekonomicko-matematický obzor*, *17*, 251–266.

[5] FIALA, P. (2005). Information sharing in supply chains. *Omega*, *33*, 419–423.

[6] FIALA, P. (2016). Profit allocation games in supply chains. *Central European Journal of Operations Research*, *24*, 267–281.

[7] PARKES, D. C. (2005). Iterative Combinatorial Auctions. In P. CRAMTON, Y. SHOHAM, & R. STEINBERG (Eds.), *Combinatorial Auctions* (*pp*. 41–78). Cambridge: MIT Press.

[8] PIKOVSKY, A. & BICHLER, M. (2005). Information Feedback in Iterative Combinatorial Auctions. In *Conference proceedings Wirtschaftsinformatik.* Bamberg: Springer.

[9] SIMCHI-LEVI, D., KAMINSKY, P., & SIMCHI-LEVI, E. (1999). *Designing and Managing the Supply Chain: Concepts, Strategies and Case studies*. Boston: Irwin/ Mc Graw-Hill.

[10] SNYDER, L. V. & SHEN Z. J. M. (2011). *Fundamentals of Supply Chain Theory*. Hoboken: Wiley.

[11] TAYUR, S., GANESHAN, R., & MAGAZINE, M. (Eds.) (1999). *Quantitative models for supply chain management*. Boston: Kluwer.

Effect of Higher Moments in the Presence of Heterogeneous Financial Market Participants

Lukáš Frýd[1]

Abstract. The mainstream in financial economics follows the hypothesis of a random walk of the logarithm of an asset price with symmetric innovation and homogeneous investment horizons. We show, that different investment horizons, as well as higher conditional moments, play an important role in the asset's return dynamics. The results are presented for the VIX index during the financial crisis in 2008.

Keywords: wavelets, GAS, VIX

JEL Classification: C58, G00
AMS Classification: 62P20

1 Introduction

The most famous framework of the asset pricing model is based on the theory of the Efficient market hypothesis (EMH) and the random nature of asset price development. The first mathematical description of the random nature of asset prices was established by Samuelson [7]. Samuelson derived the EMH for the assumption that assets prices p_t are driven by random walk process:

$$\ln p_t = \ln p_{t-1} + \epsilon_t. \tag{1}$$

For the data generation process of ϵ_t Campbell et al. [8] define three types of the random walk:

- $\epsilon_{t+1}|\Omega_t \sim IID(0, \sigma^2)$, [1]
- $\epsilon_{t+1}|\Omega_t \sim INID$,[2]
- $\epsilon_{t+1}|\Omega_t \sim NID$. [3]

But EMH as well as the assumption that assets prices follow random walk face to the critique from the beginning of the theory. Furthermore, in the financial economic exists several alternatives to the EMH and so to the asset pricing. Nowadays the behavioral approach is very popular for example Baberis and Thaler [9]. Behavioral aspects accent Weron and Weron [10] too. The authors refuse the homogeneity of financial markets participants. They stress, that the market is made up of many individuals with different investment horizons. Hence, the price setting is based on the combination of many different investment horizons. The existence of different investment horizonts supports [1], [2], [3], [4].

The next part of the critique we can find in the econometrics framework. For example Pesaran [11] highlights these disputable conditions:

- Normal or symmetry distribution of returns,
- Distribution of returns constant over time,
- Statistically independent increment of return over time.

For example, in the empirical literature, we can find that predictability tends to rise during turmoils periods (Pesaran, [11]) and (or) the forecast accuracy rises with time horizons Malkiel [6], Lo and MacKinlay [5].

In this article, we focus on the two features of price setting. We stress the existence of different investment horizons. Next, we relax the assumption of time-invariant higher moments. For these reasons, we utilize wavelet transformation for analysis of different investment horizons and Generalized autoregressive score model (GAS) model for estimation of higher time-varying moments, more in section Methodology. In the empirical part, we analyze VIX index during the financial crisis in 2008. The last section contains a conclusion.

[1] University of Economics, Faculty of Informatics and Statistics, Sq. W. Churchill 1938/4; 130 67 Prague 3; Czech Republic, lukas.fryd@vse.cz

[1] most often $\mathcal{N}(0, \sigma^2)$
[2] INID-independent but not identically distributed
[3] NID-not independent but uncorrelated

2 Methodology

We utilize Maximal overlap discrete wavelet transformation (MODWT) by Percival and Walden [12] with LA (8) filter (Daubechies, [13]) to decompose the original time series r into the additive representations, of the same length.

$$\mathbf{X} = \sum_{j=1}^{J} \tilde{\mathcal{D}}_j + \mathbf{S}_J, \tag{2}$$

where $\mathbf{X}$ is matrix of size $N \times (J+1)$, when $\mathbf{x}_t = (r_{1,t}, r_{2,t}, ..., r_{J,t}, r_{J+1,t})$ for time scales $j = 1, 2, .., J$. $\tilde{\mathcal{D}}_j$ represents T dimensional column vector whose elements represents the changes in $\mathbf{X}$ at scale j and $\mathbf{S}_J$ is the smooth of $\mathbf{X}$ at scale J and represents the low frequencies or higher scales in the time series. More detailed explanation can be found for example in Frýd [15].

2.1 GAS model

Creal et al. [14] propose the observation driven time-varying parameter model named Generalized autoregressive score model (GAS). The model is based on the updating autoregressive function driven by the scaled score of the model's density function. The conditional distribution of return r_t is given as:

$$r_t \sim p(r_t|\mathbf{f}_t, \mathcal{F}_t; \boldsymbol{\Theta}), \tag{3}$$

where $\mathcal{F}_t = \{\mathbf{r}_{t-1}, \mathbf{f}_t\}$ for $t = 1, ..., n$ represent the available information set at time t, $\mathbf{f_t}$ is vector time-varying parameter and $\boldsymbol{\Theta}$ represents vector of static parameter. In this article we assume that conditional distribution from Equation 3 follows Skew–Student–t distribution with the four time varying parameters $f = (\mu, \sigma, \gamma, \nu)$. The main feature of the GAS model is given by the autoregressive updating equation for the vector of parameters $\mathbf{f}_t$:

$$\mathbf{f}_{t+1} = \boldsymbol{\omega} + \mathbf{A}\mathbf{s}_t + \mathbf{B}\mathbf{f}_t, \tag{4}$$

where $\boldsymbol{\omega}, \mathbf{A}, \mathbf{B}$ are vector and matrices of coefficient and are functions of $\boldsymbol{\Theta}$, $\mathbf{s}_t$ is the scaled score function $\mathbf{s}_t = s_t(r_t, \mathbf{f}_t, \mathbf{F}_t, \boldsymbol{\Theta})$. The mechanism of GAS model is based on the updating the parameters $\mathbf{f}_t$ for the next period $t+1$ under knowledge of realization of r_t at time t. The updating mechanism is given by function $\mathbf{s}_t$:

$$\mathbf{s}_t = \mathbf{S}_t(\mathbf{f}_t) \times \boldsymbol{\nabla}_t, \tag{5}$$

where $\mathbf{S}_t$ is positive definite scaling matrix and $\boldsymbol{\nabla}_t$ is score of 3 define as:

$$\boldsymbol{\nabla}_t = \frac{\partial \ln p(r_t|\mathbf{f}_t, \mathcal{F}_t; \boldsymbol{\Theta})}{\partial \mathbf{f}_t}, \tag{6}$$

The form of scaling matrix $\mathbf{S}_t$ plays an important role in the GAS model and different form of scaling matrix gives different GAS model. For example, Creal et al. [14] suggest that the scaling matrix depends on the variance of score 6:

$$\mathbf{S}_t(\mathbf{f}_t) = \mathcal{I}_{t|t-1}(\mathbf{f}_t)^{-\gamma}, \tag{7}$$

where

$$\mathcal{I}_{t|t-1}(\mathbf{f}_t) = E_{t-1}[\boldsymbol{\nabla}_t \boldsymbol{\nabla}_t^T]. \tag{8}$$

E_{t-1} is expectation with respect to distribution 3, parameter $\gamma \in \{0, 1/2, 1\}$. So, special case of scaling matrix for $\gamma = 0$ is identity matrix.

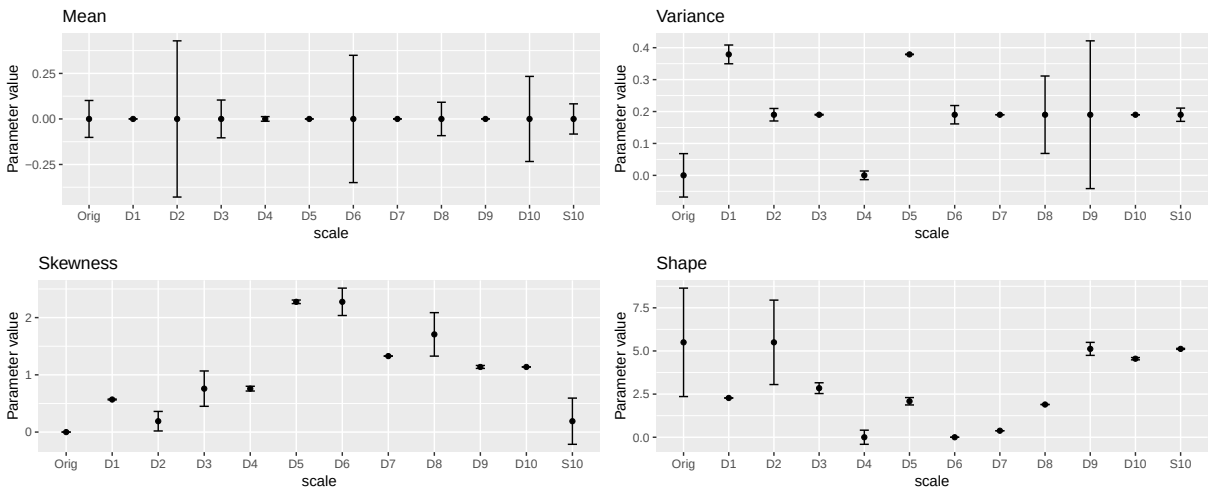


Figure 1 Innovation parameter a – VIX

3 Empirical results

3.1 Data

In this section, we analyze the daily returns of the VIX index during the financial crisis in 2008. The sample period span 1.1.2007–31.12.2009.

The estimation of GAS model parameters from the matrix **A** from Equation 4 are displayed in Figure 1 and parameters from the matrix **B** from Equation 4 are displayed in Figure 2. Every figure represents development of static parameter a/b across time scales where $a \in A$ and $b \in B$. Because we evaluate Skew–Student–t distribution with the four time-varying parameters figures are divided into the quadrant. The vertical line represents 95% confident interval. D_j, for $j = 1, 2, \ldots, 10$ and S_{10} represents scales (investment horizons). Parameter a represents the innovation in the dynamics of process 4 and parameter b represented autoregressive part (memory) in the trajectory of conditional moment. The first value in the figure is the parameter gained from the original time series (daily returns) and the next labels represent a specific scale from the MODWT.

The innovation parameters from matrix **A** have very different dynamics with respect to the original data. Only in case of mean, the estimated values are closed to zero. We can not reject the null hypothesis that the parameter is equal to zero. But for innovation parameter for variance, we can see scale-dependent $a_{2,j}$. Parameter $a_{2,j}$ fluctuates at lower scales and stabilize at higher scales (longer investment horizons). Almost all parameters $a_{2,j}$ are statistically significant. Even more interesting is the dynamics of a_3. The impact of the new information on the skewness dynamics is unstable. We can see revert U-shape. The interesting thing is in the instability of a_4 driven innovation for the time-varying shape parameter. Parameter a_4 is statistically significant for all scales.

Figure 2 provides interesting view into the process memory. For mean b_1, short investment horizons are connected with higher fluctuation of the parameter value. But from the scale D5 we can see parameter value stabilization at 0.9. Further, "memory" in variance (b_2) is very volatile. Furthermore, estimation variance fluctuates too. On the other hand, for all parameters, we can reject the null hypothesis that the true parameter is zero. Autoregressive parameter for skewness dynamics (b_3) fluctuates around the 0.9 up to the scale D8. But long-term investment horizons reveal the lower value of process memory. The last parameter b_4 driven the shape of return distribution is unstable from the scale D1 to the scale D6, but from the scale D7 to the scale S10 the autoregressive parameter is higher than 0.9 and statistically significant.

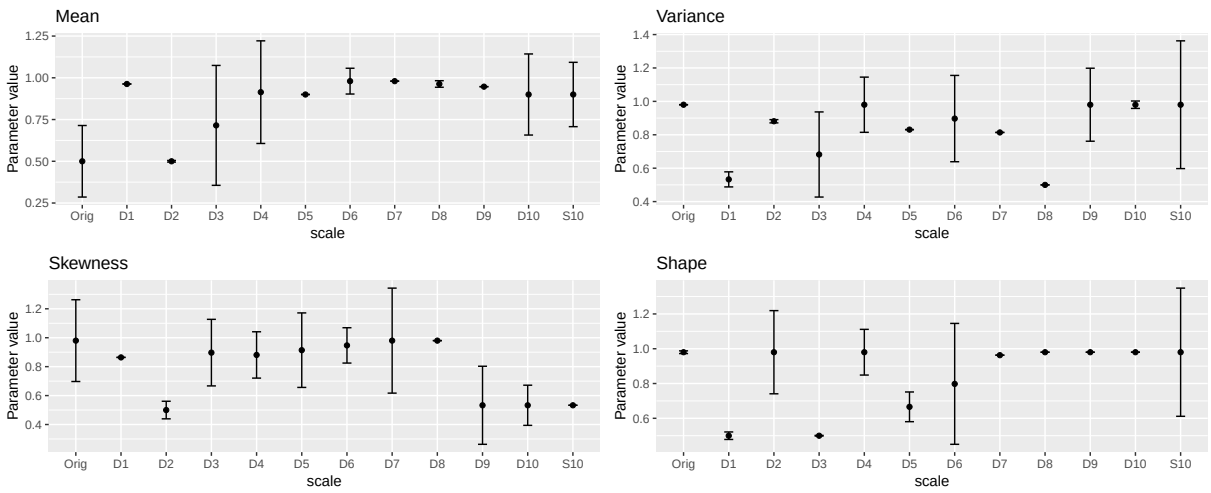


Figure 2 Autoregressive parameter b – VIX

4 Conclusion

We study the dynamics of conditional mean, variance, skewness, and shape parameter for the VIX index during the financial crisis in 2008 for different investment horizons. The disaggregation of daily returns is provided by wavelet transformation and estimation of conditional moments is done using the GAS model. We show that parameters driven conditional higher moments are investment horizons dependent. But, the distribution of parameters across the investment horizons does not follow a clear pattern.

Acknowledgements

We gratefully acknowledge the support of this project by the research grant VSE IGA F4/34/2020, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Kristoufek, L. (2013). Fractal markets hypothesis and the global financial crisis: Wavelet power evidence. *Scientific reports* 3, 2857.

[2] Vacha, L. & J. Barunik (2012). Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis. *Energy Economics* , 34(1), 241–247.

[3] Gençay, R., F. Selcuk, & B. Whitcher (2005). Multiscale systematic risk. *Journal of International Money and Finance*, 24(1), 55–70.

[4] Reboredo, J. C. & M. A. Rivera-Castro (2014). Wavelet-based evidence of the impact of oil prices on stock returns. *International Review of Economics & Finance* 29, 145–176.

[5] Lo, A. W. & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The review of financial studies*, 1(1), 41–66.

[6] Malkiel, B. G. & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2), 383–417.

[7] Samuelson, P. A. (2016). Proof that properly anticipated prices fluctuate randomly. In The world scientific handbook of futures markets (pp. 25–38).

[8] Campbell, J. Y., Champbell, J. J., Campbell, J. W., Lo, A. W., Lo, A. W. & MacKinlay, A. C. (1997). The econometrics of financial markets. princeton University press.

[9] Barberis, N. & Thaler, R. (2003). A survey of behavioral finance. Handbook of the Economics of Finance, 1, 1053–1128.

[10] Weron, A. & Weron, R. (2000). Fractal market hypothesis and two power-laws. *Chaos, Solitons & Fractals*, 11(1-3), 289–296.

[11] Pesaran, M. H. (2015). Time series and panel data econometrics. Oxford University Press.

[12] Percival, D. B. & Walden, A. T. (2000). Wavelet methods for time series analysis (Vol. 4). Cambridge university press.

[13] Daubechies, I. (1992). Ten lectures on wavelets (Vol. 61). *Siam*.

[14] Creal, D., Koopman, S. J. & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5), 777–795.

[15] Frýd, L. (2020). Volatility and Skewness Spillover Effects: Multiresolution Analysis.

R&D Spillover Effects: Impact of R&D Expenditure in the Business Sector on Innovation Activities

Andrea Furková[1]

Abstract. The paper examines the role of spatial interactions in the context of regional innovation activities across 220 European Union (EU) regions. The hypothesis was that there is a global level of spillover innovation effects among the EU regions. This hypothesis was verified based on the spatial autoregressive models and data comes from The Regional Innovation Scoreboard 2019. Patent Cooperation Treaty (PCT) applications were considered as an innovative output and three innovative inputs: scientific publications among the top-10% most cited publications worldwide, Research & Development (R&D) expenditure in the business sector and Small and Medium-Sized Enterprises (SME) introducing product or process innovations were used. We found out that all impacts associated with all innovative inputs were statically significant and thus innovation activities performed in one region do not have just local character. In addition, spatial partitioning of impacts related to the R&D expenditure in the business sector showed that even the regions corresponding to the fifth degree based on the spatial maximum likelihood estimation of the neighbourhood still have a statistically significant impact.

Keywords: R&D expenditure in the business sector, spatial spillover effects, spatial autoregressive model

JEL Classification: O31, R12
AMS Classification: 91B72

1 Introduction

The latest technologies, innovation and knowledge today play a huge role in the rapidly changing global economy. The measurement and evaluation of innovation activities especially on the regional level seems to be a significant challenge at academic and policy making levels. Empirical studies that analyse how knowledge is created, accumulated, and transferred make it possible to identify and explain the performance and productivity gaps between specific enterprises, activities, industries and even countries that have "knowledge potentials" – dynamic knowledge absorption capabilities [11]. So far, these research activities have led to three main conclusions: innovation is not uniformly distributed across regions, innovation tends to be spatially concentrated over time and even regions with similar innovation capacity have different economic growth patterns [5]. Studies, that support these conclusions include e.g., [9], [10], [7], [6], [2] or [4].

The lack of the regional innovation data makes it significantly more difficult to carry out analyses on regional innovation activities. The Regional Innovation Scoreboard (RIS) offers statistical facts on regions' innovation performance. At regional level, a composite indicator – the Regional Innovation Index (RII) is provides and it is constructed based on the regional data for 17 indicators. RIS 2019 provides a comparative assessment of performance of innovation systems across 238 regions of 23 European Union[2] (EU) member states, Norway, Serbia and Switzerland.

The capacity of firms to develop new products determines their competitive advantage. One indicator of the rate of new product innovation is the number of patents [5]. Thus Patent Cooperation Treaty (PCT) patent applications as an innovation output is an inevitable part of RII indicator. Figure 1 indicates strong geographical performance differences in PCT patent applications. Regions in Denmark, Finland, Germany, the Netherlands and Sweden are amongst the top one-third high performing group. On the other hand, most of the Eastern European regions, and regions in Portugal, Spain and the South of Italy perform relatively weak

[1] University of Economics in Bratislava. Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská 1/b, 852 35 Bratislava, andrea.furkova@euba.sk.
2 Cyprus, Estonia, Latvia, Luxembourg and Malta are included at the country level, as in these countries NUTS 1 and NUTS 2 levels are identical to the country territory.

on PCT patent applications. Evidently, innovation among the EU regions is not uniformly distributed and probably innovation tends to be spatially correlated.

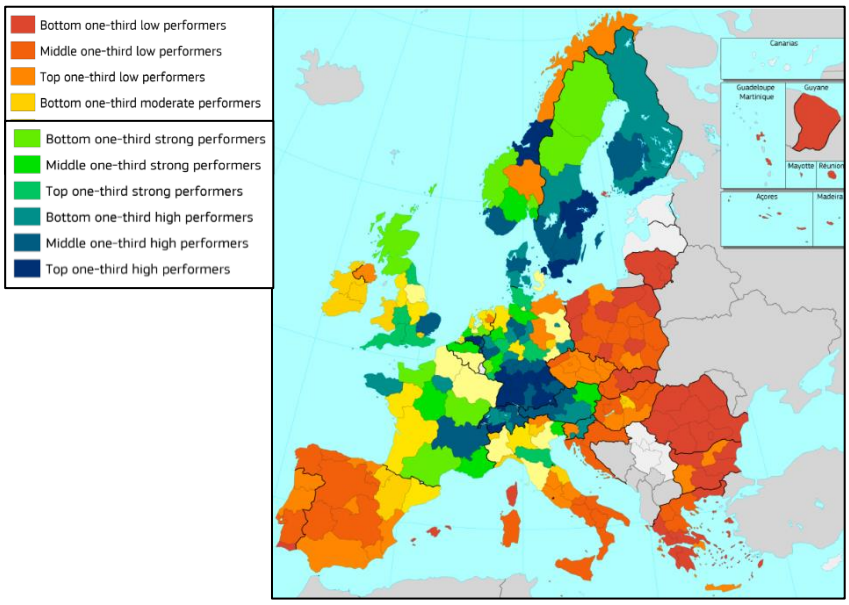


Figure 1 PCT patent applications per billion regional GDP
Source: author's elaboration based on the RIS 2019 [5]

Regions presented in Figure 1 constitute basis for our analysis of spatial dependences among the EU regions in terms of regional innovation problem. The innovation processes are often heavily localized into clusters of innovative companies, sometimes in close cooperation with public institutions such as research centres and universities. Geographic concentration of companies allows to exploit technological development, share experiences with similar technologies, knowledge, etc. Due to these facts, we suppose that there exist spatial spillover effects. As an innovation output we chose PCT applications and as innovative inputs we used Scientific publications among the top-10% most cited publications worldwide, R&D expenditure in the business sector and Small and Medium-Sized Enterprises (SME) introducing product or process innovations. We hypothesize that there is global level of spillover innovation effects among the EU regions, i.e., the changes in innovative inputs in the particular region will affect the number of patent applications not only the region itself but these changes will also have significant impacts on neighbouring regions with higher degree of neighbourhood.

Spatial econometric approach will be applied as the hypothesis validation tool. The models will be used to quantify and to test statistical significance of the direct, indirect and total impacts of the innovative inputs. We will pay a special attention to the R&D expenditure in the business sector and following spatial partitioning of its impacts we will try to answer the question what level of neighbourhood degree still matter.

The paper is structured as follows: section 2 provides brief theoretical backgrounds of the study, data and empirical results are presented and interpreted in section 3. Main concluding remarks contain section 4 and the paper closes with references.

2 Brief methodological backgrounds

Spatial econometric models account for the role that space plays in determining many of the variables that economists and other social scientists deal with. In spatial econometric models, the spatial interactions are typically accounted for by weighting matrix **W**. The (i,j)-th element of this matrix w_{ij} describes the "closeness" between unit i and unit j. Units are viewed as neighbours and interact in a meaningful way. This interaction could relate to geographic proximity issues, spillovers, externalities, industrial structure, similarity of markets, etc.

The most popular group of spatial econometric models explicitly allow for spatial dependence through so-called spatially lagged variables. The well–known SAR (Spatial Autoregressive) model assumes spatial spillover effects within the dependent variable and it can be written as follows:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad \mathbf{u} \sim N(\mathbf{0}, \sigma_u^2\mathbf{I}_N) \tag{1}$$

or a reduced form as follows:

$$\mathbf{y} = (\mathbf{I}_N - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_N - \rho\mathbf{W})^{-1}\mathbf{u} \quad \mathbf{u} \sim N(\mathbf{0}, \sigma_u^2\mathbf{I}_N) \tag{2}$$

where $\mathbf{y}$ is a $(N \times 1)$vector of a dependent variable (N is a size of the sample); $\mathbf{X}$ is a $\big(N \times (k+1)\big)$ matrix of explanatory variables (k is a number of explanatory variables); $\boldsymbol{\beta}$ is a $\big(1 \times (k+1)\big)$ vector of unknown parameters; $\mathbf{u}$ is a $(N \times 1)$ vector of errors terms; $\mathbf{I}_N$ is N dimensional identity matrix. The term $\mathbf{Wy}$ in (1) is a spatial lag component which reflects a relationship between the dependent variable in a given unit with the same variable in neighbouring units and spatial autoregressive parameter ρ indicates the direction and the strength of spatial dependence.

The estimation of spatial econometric models is based on familiar estimation econometric methods but they must be modified with respect to spatial aspects: Maximum Likelihood (ML) or two-stage least squares (2LS). Review of estimation methods and other specifications of spatial econometric models can be found in e.g., [1].

Spatial econometric models are characterized by a complicated structure of spatial dependencies which allows to accommodate modelling strategies that describe multi-regional interactions. Consequently, unlike classical linear regression model, a change in a single region associated with any given explanatory variable will affect not only the region itself (called a direct impact), but potentially affect all other regions indirectly (called an indirect impact). This means that the expected value of the dependent variable in the i-th location is no longer influenced only by exogenous location characteristics, but also by the exogenous characteristics of all other locations through a spatial multiplier $(\mathbf{I}_N - \rho\mathbf{W})^{-1}$ (see equation (2) and for more details [8]). In order to interpret parameter estimates correctly, LeSage and Pace [8] suggested summary impact measures: average total impact, average direct impact and average indirect impact.

3 Data and Empirical Results

The spatial spillover analysis of the EU regional innovation activities was performed based on the cross–sectional data set obtained from The Regional Innovation Scoreboard 2019 which covers 238 regions of 23 EU member states, Norway, Serbia[3] and Switzerland at different NUTS (Nomenclature of territorial units for statistics) levels. The components of RII indicator were used as model variables. Due to missing data and isolated observations, the data set reduction must have been done. Consequently, we dealt with 220 observations at NUTS 1 or NUTS 2 levels.

The basis of our analysis was spatially extended Regional Knowledge Production Function (RKPF) model (see e.g., [9]) following SAR specification in (1) or (2). As an innovative outcome, we consider PCT applications at the European Patent Office (EPO) (dependent variable y) expressed as the number of PCT applications at the EPO per billion regional GDP. The first innovative input is represented by scientific publications among the top 10% most cited publications worldwide (explanatory variable x_1) expressed as number of scientific publications among the top 10 % most cited publication worldwide per total number of scientific publications. This variable was chosen because it is considered as a measure for the efficiency of the research system as highly cited publications are assumed to be of higher quality. The second innovative input is represented by R&D expenditure in the business sector (explanatory variable x_2) expressed as all R&D expenditures in the business sector as percentage of regional GDP. This captures the formal creation of new knowledge within firms. It is particularly important in the science-based sector (pharmaceuticals, chemicals and some areas of electronics) where newest knowledge is created in or near R&D laboratories. Finally, small and medium-sized enterprises – SME introducing product or process innovations was used as the last innovative input (explanatory variable x_3). It is expressed as number of SMEs that introduced a new product or a new process per number of SMEs. Technological innovation as measured by the introduction of new products (goods or services) and processes is key to innovation in manufacturing activities. Higher shares of technological innovators should reflect a higher level of innovation activities. This was the reason for inclusion this variable into consideration. All model variables are in logarithmic form.

We start our econometric analysis with performing OLS (Ordinary Least Squares) estimation (model without spatial lag terms) and next we proceed with spatial modification of this model. Table 1 presents OLS and SAR model estimates based on a spatial weight matrix of queen case contiguity form (for more details see, e.g. [8] or [3]). The SAR model was estimated by spatial maximum likelihood (SML) procedure and also

[3] For Serbia, official NUTS codes are not yet available and therefor unofficial codes will be used (see [5]).

by spatial two-stage least squares (2SLS). Since the estimate for the spatial autoregressive parameter ρ according to both spatial estimations is significantly different from zero, OLS estimates are biased and inconsistent. Typically, non-spatial models tend to attribute variation in the dependent variable to the explanatory variables leading to larger (in absolute value terms) estimates, whereas the SAR model assigns this variation to the spatial lag of the dependent variable, producing smaller estimates (see Table 1). Also, spatial autocorrelation statistics based on the OLS estimation (Moran´s *I* and robust LM_{SAR}. LM_{SEM} tests) and *LR* test resulting from SML estimation provided the support for consideration of SAR model. Both spatial estimations provided very similar estimation results as for model parameters and none of these models indicate a problem of remaining spatial autocorrelation (see results of *LM* and *A–K* tests in Table 1).

	OLS Model OLS	**SAR Model** SML	**SAR Model** 2SLS
β_0	–0.064	0.057	0.053
$\beta_1(\ln x_1)$	0.658***	0.332***	0.342***
Direct impact	0.658	0.355***	0.364***
Indirect impact	0	0.283***	0.275***
Total impact	0.658	0.638***	0.639***
$\beta_2(\ln x_2)$	0.604***	0.440***	0.445***
Direct impact	0.604	0.471***	0.474***
Indirect impact	0	0.376***	0.359***
Total impact	0.604	0.847***	0.833***
$\gamma_1(\ln x_3)$	0.205***	0.097*	0.100*
Direct impact	0.205	0.104*	0.107*
Indirect impact	0	0.083*	0.081*
Total impact	0.205	0.186*	0.187*
ρ	–	0.480***	0.465***
R^2	0.700	–	–
pseudo R^2	–	0.796	0.795
AIC	270.863	201.792	–
ln*L*	–	–95.896	–
Jarque–Bera	29.388***	–	–
Moran´s *I* (residuals)	7.309 ***	–	–
Robust LM_{SAR}	26.195***	–	–
Robust LM_{SEM}	1.126	–	–
LM test	–	1.508	–
A–K test	–	–	0.126
LR test	–	71.071***	–

Table 1 Estimation results – OLS and SAR
Notes:
Symbols ***, **, * in all tables of the paper indicate the rejection of H_0 hypotheses at 1%, 5% and 10% level of significance, respectively.
AIC – Akaike information criterion; ln*L* – logarithm of log likelihood function; *A–K* – Anselin–Kelejian; *LR* – Likelihood Ratio; *LM* – Lagrange Multiplier
Source: own calculations in RStudio

As for SAR model estimates, we cannot ignore the fact that it is a model with global spillover effects and consequently, the model estimates cannot be interpreted as in the classical regression model fashion. To assess the signs and magnitudes of impacts arising from changes in the three explanatory variables – three

innovative inputs, we turn to the summary measures of direct, indirect and total impacts presented in Table 1, Table 2 and Table 3. All impacts associated with our three innovative inputs are statistically significant and have expected positive signs. Since both spatial models provided very similar results, we consider these estimates to be equivalent.

Now, for example, based on the SML estimation let us have a closer look at R&D expenditure in the business sector (variable x_2) and spillover effects associated with this variable. The average direct impact (0.471) does not match the estimate of the parameter β_2 (0.440) and this difference (0.031) is the amount of feedback effects among the regions. An average total impact equals 0.847 and since the variables in our models were expressed in logarithmic forms, average total impacts can be interpreted as elasticities. Based on the average total R&D expenditure (variable x_2) impact we can conclude that a 1% increase in total R&D expenditure will cause on average of 0.847 % increase in PCT applications, while approximately up to 56 % of total impact is attributed to indirect impact and only 44 % to direct impact (see Table 2).

	SML estimation	**2SLS estimation**
Direct impact	0.471***	0.474***
Indirect impact	0.376***	0.359***
Total impact	0.847***	0.833***
Indirect impact/Total impact	0.556	0.569
Direct impact/Total impact	0.444	0.431

Table 2 Cumulative impacts of R&D expenditure in the business sector
Source: own calculations in RStudio

It is clear that changes in explanatory variables will have a greater impact on regions with a lower degree of neighbourhood than on regions with higher degrees of neighbourhood. Table 3 provides the spatial partitioning of the direct, indirect and total R&D expenditure impacts on the basis of the SAR models. Spatially partitioned (marginal) direct, indirect, and total impacts are calculated for neighbourhood degrees 0 through 5. According to LeSage and Pace [8], it is possible to quantify the effect corresponding to each degree of neighbourhood. For instance, the direct impact corresponding to the neighbourhood $\mathbf{W}^0$degree equals to the parameter β and the indirect impact is zero because there is no spatial lag of explanatory variable in SAR model. From the spatial partitioning of the direct impact we can further notice that, until we reach the fifth degree of neighbourhood, we will "explain" a substantial part of the cumulative impacts (see Table 3). The statistical significance of marginal impacts corresponding to all degrees of neighbourhood suggests that even the regions corresponding to the fifth degree – SML estimation (fourth degree – 2SLS estimation) of the neighbourhood still have a statistically significant impact. Accordingly, R&D expenditure in the business sector does not have only have a local character but positive effects of R&D expenditure can also be spread beyond the boundaries of the analysed region.

	SML estimation2			**SLS estimation**		
W-order	**Direct**	**Indirect**	**Total**	**Direct**	**Indirect**	**Total**
$\mathbf{W}^0$	0.440***	0	0.440***	0.445***	0	0.445***
$\mathbf{W}^1$	0	0.211***	0.211***	0	0.207***	0.207***
$\mathbf{W}^2$	0.023***	0.078***	0.101***	0.022***	0.074***	0.096***
$\mathbf{W}^3$	0.003***	0.046***	0.049***	0.003**	0.042**	0.045**
$\mathbf{W}^4$	0.003**	0.021**	0.023**	0.003*	0.018*	0.021*
$\mathbf{W}^5$	0.001*	0.011*	0.011*	0.001	0.009	0.010
Σ	0.471	0.363	0.837	0.474	0.351	0.824

Table 3 Spatial partitioning of R&D expenditure in the business sector
Source: own calculations in RStudio

4 Conclusion

This paper dealt with spatial analysis of 220 European regions based on the components of RII 2019. We hypothesized that there is global level of spillover innovation effects among the EU regions. The spatial autoregressive models were used to quantify direct, indirect and total impacts of three innovative inputs. We paid a special attention to the R&D expenditure in the business sector. We found out that all impacts associated with all innovative inputs under the consideration were statically significant and the changes in innovative inputs in the particular region affect the number of patent applications not only the region itself but these changes also have significant impacts on neighbouring regions with higher degree of neighbourhood. Spatial partitioning of impacts related to the R&D expenditure in the business sector showed that even the regions corresponding to the fifth degree – SML estimation (fourth degree – 2SLS estimation) of the neighbourhood still have a statistically significant impact. In overall, innovation activities performed in one region do not have just local character but can spread to other regions. The diffusion of knowledge or positive effect of R&D expenditure increase is much more obvious in the case of the regions which create so called innovative clusters and they benefit from geographic proximity. In overall, we can conclude that the spatial interactions among the EU regions should not be neglected and if we omit these spatial dependencies we can come to truly misleading conclusions.

Acknowledgements

This work was supported by the Grant Agency of Slovak Republic – VEGA 1/0193/20 "Impact of spatial spillover effects on innovation activities and development of EU regions" and VEGA1/0294/18, "Short-run and long-run dynamic analysis of economic development of European post-communist countries and regions".

References

[1] ANSELIN L. & REY, S. J. (2014). *Modern Spatial Econometrics in Practice*. Chicago: GeoDa Press LLC.

[2] CHARLOT, S., CRESCENZI, R., & MUSOLESI, A. (2014). Econometric modelling of the regional knowledge production function in Europe. *Journal of Economic Geography*, *15*(6), 1227–1259.

[3] CHOCHOLATÁ, M. (2018). Spatial Analysis of the Tertiary Educational Attainment in European Union. In *Proceedings of the 15th International Conference Efficiency and Responsibility in Education* (pp.117–124). Praha.

[4] FURKOVÁ, A. (2019). Spatial spillovers and European Union regional innovation activities. *Central European Journal of Operations Research*, *27*(3), 815–834.

[5] HOLLANDERS, H., ES-SADKI, N., & MEKELBACH, I. (2019). Regional Innovation Scoreboard 2019. [online] Available at: https://ec.europa.eu/growth/sites/growth/files/ris2019.pdf [Accessed 1 Mar. 2020].

[6] KHAN, B. Z. (2012). Of Time and Space: A Spatial Analysis of Technological Spillovers among Patents and Unpatented Innovations in the Nineteenth Century. NBER Working Paper No. 20732. [online] Available at: http://www.nber.org/papers/w20732 [Accessed 11 Feb. 2017].

[7] KUMAR, I. (2008). Innovation Clusters: A Study of Patents and Citations. In *ESRI International User Conference*. [online] Available at: https://www.pcrd.purdue.edu/files/media/Innovation-Clusters-A-Study-of-Patents-and-Citations.pdf [Accessed 11 Jun 2017].

[8] LESAGE, J. & PACE, K. (2009). *Introduction to Spatial Econometrics*. Boca Raton: Chapman and Hall/CRC.

[9] MORENO, R., PACI, R., & USAI, S. (2005a). Spatial Spillovers and Innovation Activity in European Regions. *Environment and Planning A: Economy and Space*, *37*(10), 1793–1812.

[10] MORENO, R., PACI, R., & USAI, S. (2005b). Innovation clusters in the European regions. [online] Available at: http://crenosunicait/crenos/sites/default/files/wp/05-12pdf. [Accessed 2 Mar. 2017].

[11] TRACHUK, A. & LINDER, N. (2019). Kowledge Spillover Effects: Impacts of Export Learning Effects on Companies' Innovative Activities. [online] Available at: https://www.intechopen.com/books/current-issues-in-knowledge-management/knowledge-spillover-effects-impact-of-export-learning-effects-on-companies-innovative-activities. [Accessed 28 May. 2020].

,

Interval Transportation Problem: The Best and the Worst (Feasible) Scenario

Elif Garajová[1], Milan Hladík[2], Miroslav Rada[3]

Abstract. Interval programming presents a powerful mathematical tool for modeling optimization problems affected by uncertainty. We consider an interval linear programming model for the transportation problem with uncertain supply and demand varying within a priori known bounds. We address the problem of computing the optimal value range of an interval transportation problem, i.e. finding the best and the worst possible optimal value, and describing the corresponding scenarios of the problem.
Since the worst-case scenario in the traditional sense is often infeasible, thus leading to an infinite bound of the optimal value range, we consider the worst finite optimal value of the problem. We propose a decomposition method based on complementarity for computing the worst finite optimal value exactly. We also study the corresponding best and worst extremal scenarios for which the bounds of the finite optimal value range are attained. Moreover, we derive a description of the structure of the linear program corresponding to the best scenario.

Keywords: transportation problem, interval linear programming, optimal value range.

JEL Classification: C44, C61
AMS Classification: 90C70, 90C08

1 Introduction

Transportation problems [7] belong to the classical optimization models widely studied in operations research. The objective of the original problem is to find a transportation plan for shipping a given commodity from a set of supply centers (e.g. factories) to the destinations, while minimizing the transportation costs. Throughout the years, several types of transportation problems have emerged, considering additional constraints and requirements or different objective functions. Since the input data of real-world transportation problems are not always known exactly and in advance, attention has also been devoted to models reflecting such uncertainty or inexactness. Uncertain transportation problems have been studied in the literature by means of stochastic programming [16], fuzzy programming [2] or robust optimization [9]. We adopt the approach of interval linear programming and consider the interval transportation problem [3].

Interval linear programming [15] provides tools for modeling and solving uncertain linear programs, in which the uncertain coefficients can be independently perturbed within the given interval ranges. One of the first tasks considered in interval optimization was the problem of computing the *optimal value range*, which denotes the interval bounded by the best and the worst optimal value over all possible scenarios. For general interval linear programs, and even for some subclasses of programs with a fixed (non-interval) constraint matrix, computing the optimal value range is an NP-hard problem [14]. Since the traditional definition of the optimal value range is not restricted to feasible scenarios, one of its bounds may become infinite due to infeasibility. Recently, a modified optimal value range considering only finite optimal values has been proposed and studied [8].

The interval transportation problem (ITP) is represented by an interval linear programming model, in which the levels of supply and demand, as well as the transportation costs, can be uncertain and thus correspond to interval coefficients. Several contributions can be found in the literature regarding the optimal value range of an interval transportation problem. Chanas et al. [3] derived a transformation of ITPs with both equality and inequality constraints into classical linear programs to find the best optimal value. Liu [11]

[1] Charles University, Faculty of Mathematics and Physics, Dept. of Applied Mathematics, Malostranské nám. 25, Prague, Czech Republic,
University of Economics, Prague, Dept. of Econometrics, nám. W. Churchilla 4, Prague, Czech Republic, elif@kam.mff.cuni.cz
[2] Charles University, Faculty of Mathematics and Physics, Dept. of Applied Mathematics, Malostranské nám. 25, Prague, Czech Republic, hladik@kam.mff.cuni.cz
[3] University of Economics, Prague, Dept. of Econometrics & Dept. of Financial Accounting and Auditing, nám. W. Churchilla 4, Prague, Czech Republic, miroslav.rada@vse.cz

proposed a linearly constrained non-linear program based on strong duality to compute the worst optimal value. Juman and Hoque [10] designed a heuristic-based algorithm for approximating the optimal value range of an extended model with inventory costs. Xie et al. [17] formulated a bilevel programming model for the problem and proposed a heuristic method based on a genetic algorithm. Finally, Cerulli et al. [1] designed an iterative local search algorithm to find a lower bound on the worst optimal value. For a special class of transportation problems immune against the transportation paradox, an improved algorithm was proposed [4]. Furthermore, general methods for approximating the optimal value range of an interval linear program are also available (see e.g. [12]).

The paper is organized as follows: In Section 2, we formulate the interval model of the transportation problem and introduce the related notions of feasibility and optimality in interval linear programming. Section 3 is devoted to the worst-case scenario, for which we derive an exact decomposition method to compute the worst finite optimal value of the interval transportation problem. In Section 4, we address the best case and examine the structure of extremal scenario corresponding to the best optimal value. Finally, Section 5 concludes the paper and provides some directions for further research.

2 Interval Transportation Problem

Intervals. First, we introduce basic notation regarding intervals. For a given $\underline{v}, \overline{v} \in \mathbb{R}^n$ such that $\underline{v} \leq \overline{v}$, an *interval vector* $\mathbf{v} \in \mathbb{IR}^n$ is the set

$$\mathbf{v} = [\underline{v}, \overline{v}] = \{v \in \mathbb{R}^n : \underline{v} \leq v \leq \overline{v}\}.$$

Alternatively, we can also define an interval vector $\mathbf{v}$ by its *center* v^c and its *radius* v^Δ as $\mathbf{v} = [v^c - v^\Delta, v^c + v^\Delta]$, where

$$v^c = \frac{1}{2}(\overline{v} + \underline{v}), \qquad v^\Delta = \frac{1}{2}(\overline{v} - \underline{v}).$$

The symbol $\mathbb{IR}^n$ denotes the set of all n-dimensional interval vectors. Hereinafter, we denote intervals and interval vectors by bold letters.

Problem formulation. Let $I = \{1, \ldots, m\}$ be the set of m *sources* (supply centers) and let $J = \{1, \ldots, n\}$ be the set of n *destinations* (customers). Each of the sources $i \in I$ has a limited *supply* s_i, which is an uncertain quantity varying within the given non-negative interval $\mathbf{s}_i = [\underline{s}_i, \overline{s}_i]$. Similarly, each destination $j \in J$ is associated with an uncertain *demand* d_j to be satisfied, represented in the model by the interval $\mathbf{d}_j = [\underline{d}_j, \overline{d}_j]$. Finally, the interval $\mathbf{c}_{ij} = [\underline{c}_{ij}, \overline{c}_{ij}]$ denotes the uncertain *unit cost* of transporting goods from source i to destination j.

The objective of the problem is to find a minimal-cost transportation plan for shipping goods from the sources to the destinations such that the supply and demand requirements are satisfied. Formally, the *interval transportation problem* can be represented by the interval linear program

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in I} \sum_{j \in J} [\underline{c}_{ij}, \overline{c}_{ij}] x_{ij} \\ \text{subject to} \quad & \sum_{j \in J} x_{ij} = [\underline{s}_i, \overline{s}_i], && \forall i \in I, \\ & \sum_{i \in I} x_{ij} = [\underline{d}_j, \overline{d}_j], && \forall j \in J, \\ & x_{ij} \geq 0, && \forall i \in I, j \in J, \end{aligned} \tag{ITP}$$

where the non-negative variable x_{ij} corresponds to the amount of goods transported from source i to destination j.

The interval transportation problem can be understood as the set of all linear programs with coefficients lying in the corresponding intervals $\mathbf{c} \in \mathbb{IR}^{m \times n}$, $\mathbf{s} \in \mathbb{IR}^m$ and $\mathbf{d} \in \mathbb{IR}^n$. A particular linear program in the interval transportation problem, which is determined by the cost vector $c \in \mathbf{c}$, supply vector $s \in \mathbf{s}$ and demand vector $d \in \mathbf{d}$ is called a *scenario*.

Feasible and Optimal Solutions. Let us now introduce the notions of feasibility and optimality usually used in interval linear programming. A given vector $x \in \mathbb{R}^{m\times n}$ is said to be a *weakly feasible solution* of (ITP), if it is a feasible solution for some scenario of the problem. The set of all weakly feasible solutions (also known as the weakly feasible set) will be denoted by $\mathcal{F}(\mathbf{s}, \mathbf{d})$. Furthermore, x is said to be a *weakly optimal solution*, if it is optimal for some scenario of the problem. The set of all weakly optimal solutions of (ITP) will be denoted by $\mathcal{S}(\mathbf{c}, \mathbf{s}, \mathbf{d})$. In the context of interval programs, we will use the terms "feasible solution" and "optimal solution" to refer to the weakly feasible and optimal solutions, respectively.

Optimal Value Range. One of the essential questions in interval linear programming is to describe the set or the range of optimal values over all possible scenarios. Let $f(c, s, d)$ denote the *optimal value* of a given scenario of (ITP), i.e.

- $f(c, s, d) = c^T x^*$ for some optimal solution x^* (if an optimal solution exists),
- $f(c, s, d) = \infty$ if the program is infeasible.

Note that the program cannot be unbounded.

Then, we can define the *optimal value range* $[\underline{f}, \bar{f}]$ of (ITP) as the interval bounded by the *best optimal value* $\underline{f}$ and the *worst optimal value* $\bar{f}$, where

$$\underline{f} = \inf\{f(c, s, d) : c \in \mathbf{c}, s \in \mathbf{s}, d \in \mathbf{d}\},$$
$$\bar{f} = \sup\{f(c, s, d) : c \in \mathbf{c}, s \in \mathbf{s}, d \in \mathbf{d}\}.$$

Since by this definition we have $\bar{f} = \infty$ if there is at least one infeasible scenario, in many practical problems it may also be desirable to compute the *worst finite optimal value* $\bar{f}_{\text{fin}}$, which can be found as the worst optimal value over the feasible scenarios:

$$\bar{f}_{\text{fin}} = \max\{f(c, s, d) : c \in \mathbf{c}, s \in \mathbf{s}, d \in \mathbf{d} \text{ with } \mathcal{F}(s, d) \neq \emptyset\}.$$

3 The Worst (Feasible) Scenario

Worst Optimal Value. Computing the worst optimal value of a general interval linear program is an NP-hard problem. This holds even when we restrict the class of programs to those with a fixed (real) constraint matrix. For the special case of interval transportation problems, no polynomial-time algorithm for computing the worst optimal value is known either.

To compute the (possibly infinite) worst optimal value $\bar{f}$ of (ITP) exactly, we can apply the general decomposition method known for interval linear programs (see [15]). For a given vector $p \in \mathbb{R}^n$, let $\operatorname{diag}(p) \in \mathbb{R}^{n\times n}$ denote the diagonal matrix with elements of p on the diagonal. Using the decomposition, we have

$$\bar{f} = \sup\{f(\bar{c}, s^c + \operatorname{diag}(p)s^\Delta, d^c + \operatorname{diag}(q)d^\Delta) : (p, q) \in \{\pm 1\}^{m+n}\},$$

i.e. the worst optimal value $\bar{f}$ of the interval transportation problem can be found as the supremum over optimal values of all scenarios in the form

$$\begin{aligned}
\text{minimize} \quad & \sum_{i\in I}\sum_{j\in J} \bar{c}_{ij} x_{ij} \\
\text{subject to} \quad & \sum_{j\in J} x_{ij} = s_i^c + p_i s_i^\Delta, && \forall i \in I, \\
& \sum_{i\in I} x_{ij} = d_j^c + q_j d_j^\Delta, && \forall j \in J, \\
& x_{ij} \geq 0, && \forall i \in I, j \in J,
\end{aligned} \tag{1}$$

for all $p \in \{\pm 1\}^m, q \in \{\pm 1\}^n$. Therefore, at most 2^{m+n} linear programs are required to find the worst optimal value using the general decomposition approach. The tested scenarios are extremal, since the right-hand-side coefficients representing the supply and demand levels are set to their respective lower or upper bounds determined by the sign vectors p and q. Note that if any of the scenarios (1) is infeasible, we obtain an infinite bound $\bar{f} = \infty$.

The decomposition used to compute the worst optimal value also provides the corresponding (extremal) worst scenario of the interval program, for which the value $\bar{f}$ was attained. If $\bar{f}$ was attained for some vectors $p^* \in \{\pm 1\}^m$, $q^* \in \{\pm 1\}^n$, then the worst scenario $(c^{\mathrm{W}}, s^{\mathrm{W}}, d^{\mathrm{W}})$ is determined by setting

$$c_{ij}^{\mathrm{W}} = \bar{c}_{ij}, \qquad s_i^{\mathrm{W}} = s_i^c + p_i^* s_i^\Delta, \qquad d_j^{\mathrm{W}} = d_j^c + q_j^* d_j^\Delta. \tag{2}$$

Worst Finite Optimal Value. Obviously, the equation-constrained interval transportation problem will almost always have an infeasible scenario, since $\sum_i s_i \neq \sum_j d_j$ occurs for some $s \in \mathbf{s}$ and $d \in \mathbf{d}$ if at least one of the supply or demand levels is an interval of non-zero radius. Thus, the worst optimal value $\bar{f}$ as defined in the traditional way will usually be infinite.

Attention has therefore been devoted to computing the worst finite optimal value $\bar{f}_{\text{fin}}$, which can provide more information about the quality of the optimal transportation plan in the worst feasible scenario. Several algorithms for approximating the value $\bar{f}_{\text{fin}}$ have been proposed in the literature. Here, we design a method to compute the value exactly (albeit in exponential time).

The proposed method is based on a description of the optimal solution set of a general interval linear program with a fixed constraint matrix given in [6], which uses a decomposition by complementary slackness. The result states that the optimal solution set $\mathcal{S} \subseteq \mathbb{R}^n$ of an interval linear program in the form

$$\min \mathbf{c}^T x \text{ subject to } Ax = \mathbf{b}, x \geq 0$$

can be described as the union of feasible sets of the systems

$$\begin{aligned} Ax &= b, \\ (A^T u)_i &\leq c_i, \quad x_i = 0, \quad \forall i \in K \\ (A^T u)_j &= c_j, \quad x_j \geq 0, \quad \forall j \notin K \\ \underline{b} &\leq b \leq \overline{b}, \ \underline{c} \leq c \leq \overline{c}, \end{aligned}$$

for each $K \subseteq \{1, \ldots, n\}$.

Now, we can apply the decomposition by complementary slackness to the interval transportation problem and use the description of the optimal solution set to compute the worst finite optimal value. Similarly as in the former case, we can observe that the worst finite optimal value $\bar{f}_{\text{fin}}$ will be achieved for some scenario with $c = \overline{c}$. Furthermore, for a problem with a fixed objective vector, we can find $\bar{f}_{\text{fin}}$ simply by maximizing the objective over the set of all optimal solutions.

Specifically, for the interval transportation problem, we can thus compute $\bar{f}_{\text{fin}}$ as the maximum value of $\sum \overline{c}_{ij} x_{ij}$ over the optimal solution set $\mathcal{S}(\mathbf{c}, \mathbf{s}, \mathbf{d})$ described by

$$\begin{aligned} &\sum_{j \in J} x_{ij} = s_i, \quad \forall i \in I, \\ &\sum_{i \in I} x_{ij} = d_j, \quad \forall j \in J, \\ &u_i + v_j \leq \overline{c}_{ij}, \quad x_{ij} = 0, \quad \forall (i,j) \in K, \\ &u_i + v_j = \overline{c}_{ij}, \quad x_{ij} \geq 0, \quad \forall (i,j) \notin K, \\ &\underline{s} \leq s \leq \overline{s}, \quad \underline{d} \leq d \leq \overline{d} \end{aligned} \tag{3}$$

for all systems with $K \subseteq I \times J$.

Note that for the purpose of computing the worst finite optimal value, we can restrict the decomposition to systems corresponding to basic optimal solutions, slightly reducing the number of linear programs to be solved. The basis enumeration approach to computing $\bar{f}_{\text{fin}}$ was also further developed by Hladík [8].

Again, we can obtain the worst feasible scenario $(c^{\text{WF}}, s^{\text{WF}}, d^{\text{WF}})$ corresponding to the worst finite optimal value $\bar{f}_{\text{fin}}$ from the decomposition. In this case, the scenario is obtained directly from the instance of program (3), for which the maximal objective value was achieved. Namely, if the supply and demand levels in the worst program are s^* and d^*, then the worst feasible scenario is

$$c^{\text{WF}} = \overline{c}, \qquad s^{\text{WF}} = s^*, \qquad d^{\text{WF}} = d^*.$$

Outcome Range Problem. Note that the decomposition by complementary slackness can also be applied to a generalized problem proposed recently in the literature, known as the *outcome range problem* [13]. The goal of the outcome range problem is to determine the best value $\underline{g}$ and the worst value $\overline{g}$ of an additional linear function $g(x) = r^T x$ (called the *outcome function*) over the optimal solution set $\mathcal{S}$ of an interval linear program, thus we have the interval $[\underline{g}, \overline{g}]$ with bounds

$$\underline{g} = \min\{r^T x : x \in \mathcal{S}\}, \qquad \overline{g} = \max\{r^T x : x \in \mathcal{S}\}.$$

Regarding the transportation problem, where the main objective is to minimize the transportation costs, other outcome functions (such as environmental impact of the chosen transportation plan) can be modeled to aid the decision-makers in evaluating future consequences of optimal decision making.

4 The Best Scenario

Best Optimal Value. Unlike the worst optimal value, the best optimal value can be computed in polynomial time by solving a single linear program. The best optimal value $\underline{f}$ of (ITP) is obtained by optimizing the best objective $\underline{c}^T x$ over the weakly feasible set (see also [3, 15] for details regarding general interval linear programs), i.e.

$$\underline{f} = \inf\{\underline{c}^T x : x \in \mathcal{F}(\mathbf{s}, \mathbf{d})\}.$$

Namely, by expressing the feasible solution set of ITP, we can show that the value $\underline{f}$ can be computed by solving the following linear program:

$$\begin{aligned} \text{minimize} \quad & \sum_{i\in I}\sum_{j\in J} \underline{c}_{ij} x_{ij} \\ \text{subject to} \quad & \underline{s}_i \le \sum_{j\in J} x_{ij} \le \overline{s}_i, \qquad \forall i \in I, \\ & \underline{d}_j \le \sum_{i\in I} x_{ij} \le \overline{d}_j, \qquad \forall j \in J, \\ & x_{ij} \ge 0, \qquad \forall i \in I, j \in J. \end{aligned} \tag{4}$$

The corresponding best scenario can also be computed from the optimal solution x^* of program (4). Namely, the best scenario $(c^{\mathrm{B}}, s^{\mathrm{B}}, d^{\mathrm{B}})$ with $c^{\mathrm{B}} \in \mathbf{c}$, $s^{\mathrm{B}} \in \mathbf{s}$ and $d^{\mathrm{B}} \in \mathbf{d}$ can be found by setting

$$c^{\mathrm{B}}_{ij} = \underline{c}_{ij}, \qquad s^{\mathrm{B}}_i = \sum_{j\in J} x^*_{ij}, \qquad d^{\mathrm{B}}_j = \sum_{i\in I} x^*_{ij}. \tag{5}$$

Note that unlike the worst scenario (2), the obtained best scenario (5) is not necessarily extremal in the sense that coefficients in the supply and demand vectors s^B, d^B are not set to the bounds of their respective interval ranges. However, by studying the structure of the best scenario, we can show that a certain number of extremal values s_i, d_i can still be achieved.

Best Extremal Scenario. Let us now assume that the levels of supply and demand are positive, specifically we have $\underline{s}, \underline{d} > 0$. Since for a solution vector x to be feasible, we have to satisfy the supply and demand requirements at all supply centers and destinations, at least one edge from each supply center and to each destination needs to be used. Therefore, we can observe that $x_{ij} > 0$ holds for at least $\max\{m, n\}$ variables x_{ij}.

Furthermore, we have $2m + 2n$ inequality constraints in linear program (4). At most $2m + 2n - \max\{m, n\}$ slacks of the inequalities are non-zero in a (basic) solution, and therefore at least $\max\{m, n\}$ slacks are zero. Thus, we have at least $\max\{m, n\}$ extremal values s_i, d_i in the best scenario of the interval transportation problem.

5 Conclusion

We have addressed the problem of computing bounds of the optimal value range for the interval transportation problem, i.e. the transportation problem with uncertain transportation costs and uncertain levels of supply and demand modeled by interval coefficients. We have explored the traditional notions of the best and the worst optimal value used in interval linear programming and derived further properties of the optimal value range with respect to the interval transportation problem.

For the worst possible optimal value, which forms the upper bound of the range, we have reviewed the general approach to computing it by decomposing the interval problem into an exponential number of linear programs. Since the worst optimal value in the traditional sense is usually infinite in case of the transportation problem due to infeasibility, we have also considered the worst finite optimal value. Here, we have designed a method for computing the exact value by solving an optimization problem over the optimal solution

set obtained by using a decomposition based on complementary slackness. We have also applied the general approach for finding the best optimal value, which amounts to solving a single classical linear programming problem. For the best case, we have also further examined the most extremal scenario corresponding to the best optimal value.

The decomposition method used to compute the worst finite optimal value can also be applied to the outcome range problem, which is a generalization of the problem of computing the optimal value range. Conducting further research in this direction might be of interest, as well as extending the obtained results to a broader class of interval linear programming problems.

Acknowledgements

M. Rada and E. Garajová were supported by the Czech Science Foundation under Grant P403-20-17529S. M. Hladík was supported by the Czech Science Foundation under Grant P403-18-04735S. E. Garajová and M. Hladík were also supported by the Charles University project GA UK No. 180420.

References

[1] Cerulli, R., D'Ambrosio, C. & Gentili, M. (2017). Best and Worst Values of the Optimal Cost of the Interval Transportation Problem. In Sforza, A. & Sterle, C. (Eds.), *Optimization and Decision Science: Methodologies and Applications, ODS 2017.* (pp. 367–374). Springer, Cham.

[2] Chanas, S., Kołodziejczyk, W. & Machaj, A. (1984). A fuzzy approach to the transportation problem. *Fuzzy Sets and Systems*, 13(3), 211–221.

[3] Chanas, S., Delgado, M., Verdegay, J. L. & Vila, M. A. (1993). Interval and fuzzy extensions of classical transportation problems. *Transportation Planning and Technology*, 17(2), 203–218.

[4] D'Ambrosio, C., Gentili, M., Cerulli, R. (2019). The optimal value range problem for the Interval (immune) Transportation Problem. *Omega*, 102059 (in press).

[5] Garajová, E., Hladík, M. & Rada, M. (2017). On the Properties of Interval Linear Programs with a Fixed Coefficient Matrix. In Sforza, A. & Sterle, C. (Eds.), *Optimization and Decision Science: Methodologies and Applications, ODS 2017.* (pp. 393–401). Springer, Cham.

[6] Garajová, E. & Hladík, M. (2019). On the optimal solution set in interval linear programming. *Computational Optimization and Applications*, 72, 269–292.

[7] Hitchcock, F. L. (1941). The Distribution of a Product from Several Sources to Numerous Localities. *Journal of Mathematics and Physics*, 20, 224–230.

[8] Hladík, M. (2018). The worst case finite optimal value in interval linear programming. *Croatian Operational Research Review*, 9(2), 245–254.

[9] Hlavatý, R. & Brožová, H. (2017). Robust optimization approach in transportation problem. In Pražák, P. (Ed.), *The 35th International Conference Mathematical Methods in Economics, Conference Proceedings.* (pp. 225–230). Hradec Králové: Gaudeamus, University of Hradec Králové.

[10] Juman, Z. A. M. S. & Hoque, M. A. (2014). A heuristic solution technique to attain the minimal total cost bounds of transporting a homogeneous product with varying demands and supplies. *European Journal of Operational Research*, 239(1), 146–156.

[11] Liu, S.-T. (2003). The total cost bounds of the transportation problem with varying demand and supply. *Omega*, 31(4), 247–251.

[12] Mohammadi, M. & Gentili, M. (2019). Bounds on the worst optimal value in interval linear programming. *Soft Computing*, 23, 11055–11061.

[13] Mohammadi, M. & Gentili, M. (2019). The Outcome Range Problem. Preprint at arXiv:1910.05913 [math.OC].

[14] Rohn, J. (1995). NP-hardness results for some linear and quadratic problems. *Technical Report 619*, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague.

[15] Rohn, J. (2006). Interval linear programming. In *Linear Optimization Problems with Inexact Data* (pp. 79–100).

[16] Williams, A. C. (1963). A Stochastic Transportation Problem. *Operations Research*, 11(5), 759–770.

[17] Xie, F., Butt, M. M., Li, Z. & Zhu, L. (2017). An upper bound on the minimal total cost of the transportation problem with varying demands and supplies. *Omega*, 68, 105–118.

Hierarchical Forecasting of GDP

Jiří Georgiev[1]

Abstract. Precise forecasts of macroeconomic variables are essential for decision making of policymakers and other economic agents. In many cases, time series are organized in hierarchical structures and can be aggregated. For such series reconciliation methods and bottom-up approach have proved to provide coherent aggregations and may even be beneficial in terms of forecasting power. Whether this is a case for Gross Domestic Product is investigated in this contribution on data for 25 European countries. Components of Expenditure Approach represent base-level series and are forecasted by simple univariate models that can be aggregated and reconciled. Reconciliation improved top-level forecasts compared to the univariate model in 64% of countries and compared to the bottom-up approach in all but two countries. For bottom-level series, reconciliation improved forecasting precision in 62% of countries.

Keywords: forecasting, hierarchical time series, GDP, expenditure Approach

JEL Classification: C53
AMS Classification: 91B84

1 Introduction

Accurate forecasts of macroeconomic variables such as Gross Domestic Product (GDP), inflation and production are crucial for decisions of policymakers and other economic agents. Aggregation structures are commonly found in macroeconomics. Often these structures can be disaggregated by attributes of interest, geographical location or per industry sectors. For example, GDP in the Expenditure Approach is calculated as an aggregate of multiple series representing expenditure made on final goods and services throughout the economy. Each of these series can be split even further into series reflecting different areas, such as government expenditure, gross capital formation, net export, etc. This creates a hierarchical structure (can be seen in Figure 1) that is often omitted in the forecasting of these series even though some of these sub-series are commonly used as their predictors.

For example, Athanasopoulos [1] demonstrates improvement in forecast precision for Australian GDP by aggregation with reconciliation. On the other hand, Heinisch [3] finds only limited evidence that bottom-up approach is capable of improving predictive capabilities for German GDP. The goal of this contribution is to assess whether reconciliation or bottom-up approach is viable for predicting GDP for European countries and whether reconciliation is capable of improvements in the bottom-level series.

The structure of this paper is as follows. Used datasets, the methodology for assessing forecasting capabilities and hierarchical time series are described in Section 2. Results are presented in Section 3, and paper is concluded in Section 4.

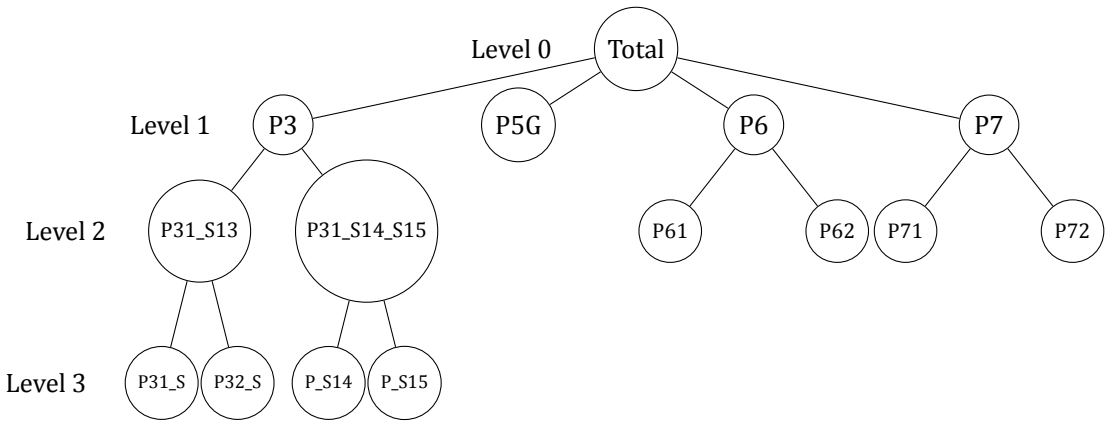


Figure 1 Three-level hierarchical structure of GDP by Expenditure Approach

[1] University of Economics, Prague, jiri.georgiev@vse.cz

2 Models, Datasets and Methodology

2.1 Hierarchical time series

Total series $y_{Tot,t}$, which is corresponding to level 0 of the hierarchy, is an aggregate of all series below this level. Simple three-level hierarchy is shown in Figure 1. Then relationships across hierarchical structure can be written as [5]

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t, \tag{1}$$

where $\boldsymbol{y}_t = (y_{Tot,t}, y_{A,t}, y_{B,t}, \dots)'$ is a vector of length n, which contains observations of all series of the hierarchy at time t. Then $\boldsymbol{b}_t$ contains m bottom-level series and $\boldsymbol{S}$ is summing matrix of size $n \times m$, which reflects linear aggregation constraints. One of the traditional approaches for forecasting hierarchical time series is a **bottom-up** approach, where only the most disaggregated bottom-level series are forecasted, and these forecasts are aggregated to obtain remaining series from the hierarchy. Another traditional approach is **top-down**, where in contrast to the bottom-up approach, only the top-level series is forecasted and disaggregated into lower levels of the hierarchy. Forecast h-step-ahead for bottom-up would be

$$\tilde{\boldsymbol{y}}^{BU}_{T+h|T} = \boldsymbol{S}\hat{\boldsymbol{b}}_{T+h|T}, \tag{2}$$

where $\hat{\boldsymbol{b}}_{T+h|T}$ are forecast of bottom-level series and top-down approaches would be

$$\tilde{\boldsymbol{y}}^{TD}_{T+h|T} = \boldsymbol{S}\boldsymbol{p}\hat{\boldsymbol{y}}_{Tot,T+h|T}, \tag{3}$$

where $\boldsymbol{p}$ is a vector of proportions used to disaggregated top-level forecast $\hat{\boldsymbol{y}}_{Tot,T+h|T}$. Both of these approaches are using only a single level, therefore ignoring correlations across the hierarchy. An alternative approach is **reconciliation** [9], where $\hat{\boldsymbol{y}}_{T+h|T}$ contains forecasts for each of the time series up to time T. First, all forecasts across the hierarchy are obtained by any method, and aggregation constraints are ignored. Sum of bottom-level forecast may not be equal to higher-level forecast, therefore needs to be adjusted as follows

$$\tilde{\boldsymbol{y}}_{T+h|T} = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{y}}_{T+h|T}, \tag{4}$$

where $\boldsymbol{P}$ is $m \times n$ matrix, when $\boldsymbol{P}$ is appropriately selected, forecasts will be coherent. One of the possible ways to determine $\boldsymbol{P}$ is finding a solution to the minimization of a trace of the covariance matrix of the h-step-ahead reconciled forecast errors

$$Var(\boldsymbol{y}_{T+h|T} - \tilde{\boldsymbol{y}}_{T+h|T}) = \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{S}^T\boldsymbol{P}^T, \tag{5}$$

where $\boldsymbol{W}_h$ is the covariance of h-step-ahead base forecast errors. This method is called MinT and was proposed by Hyndman [9].

2.2 Expenditure Approach for GDP

The Expenditure Approach is a method of measuring GDP by aggregating spending throughout the economy and can be defined as

$$GDP = C + I + G + NX, \tag{6}$$

where C is private final consumption expenditure, G is government spending measured as government final consumption expenditure, I is private investments as gross capital formation and, NX is net export [2]. Private final consumption expenditure is direct expenditure on goods and services for the satisfaction of individual needs. It can be divided into the consumption of households and consumption of non-profit institutions serving households (NPISH). Government spending can be divided into individual and collective consumption. The first relates to expenditure that is benefited by individuals such as health care, education, housing. The second reflects expenditures that benefit society and are often known as public goods and services, such as justice system, police, defence, etc. The net export is the difference between exports and imports of goods and services. Gross capital formation is the sum of gross fixed capital formation and the change in inventories (stocks). This aggregation creates a three-level hierarchical structure, where each of the nodes is a time series. Described structure can be seen in Table 1 and Figure 1. Bottom-level series are the most disaggregated and top-level (level 0) series, which corresponds to the Gross domestic product, is the most aggregated.

2.3 Datasets

Data used for this contribution were obtained from Eurostat[1] database `namq_10_gdp` with the use of R package `eurostat`. The analysis uses seasonally unadjusted quarterly GDP for 28 countries of the European Union spanning over the period from 2000 Q1 to 2019 Q3. For a large number of missing values, Poland, Malta and Croatia have been excluded. Components used as bottom-level series are listed in Table 1.

Item ID	Level 1	Level 2	Level 3	Description
				Level 3
P31_S13	1	1	1	Individual consumption expenditure of general government
P32_S13	1	1	2	Collective consumption expenditure of general government
P31_S14	1	2	3	Final consumption expenditure of households
P31_S15	1	2	4	Final consumption expenditure of NPISH
				Level 2
P3_S13	1	1		Final consumption expenditure of general government
P3_S14_S15	1	2		Household and NPISH final consumption expenditure
P61	3	3		Exports of goods
P62	3	4		Exports of services
P71	4	5		Imports of goods
P72	4	6		Imports of services
				Level 1
P3	1			Final consumption expenditure
P5G	2			Gross capital formation
P6	3			Exports of goods and services
P7	4			Imports of goods and services
				Level 0
B1GQ				Gross domestic product at market prices

Table 1 GDP and used components

2.4 Methodology

Data set is divided into training and test sets. The training set contains all observations from 2000 Q1 to 2015 Q4 and test set includes all remaining observations till 2019 Q3. Evaluation of forecast is carried out by using rolling forecast origin [4] with a forecasting horizon of 1, 2, 3 and 4. With each iteration training set is moved by one, forecasts are made for respective horizon and error is computed by RMSE and MASE as follows

$$RMSE_h = \sqrt{\frac{1}{h}\sum_{j=1}^{h}(e_j)^2}, \tag{7}$$

$$MASE_h = \frac{\frac{1}{h}\sum_{j=1}^{h}|e_j|}{\frac{1}{T-1}\sum_{t=2}^{T}|y_t - y_{t-1}|}, \tag{8}$$

where h is forecasting horizon and e_j is forecast error given as difference between forecasted value and actual value. MASE is preferable to other methods as it is scale-invariant and therefore allows comparison across countries [6]. For each of the forecasting horizons, the final error is computed as the average of error across all origins. This final error should provide a robust evaluation of forecasting capabilities and can be reliably used for assessment and comparison. For comparison in following the section, MASE is used.

[1] https://ec.europa.eu/eurostat/data/database

3 Results

The first step is to obtain base level forecasts for all series in the hierarchy. For Expenditure Approach 15 series are needed for reconciliation, as reconciliation demands series for all nodes of the structure to be estimated, and only 8 series for bottom-up approach, as only bottom-level series are needed. To estimate and forecast these series ARIMA and Exponential smoothing with errors, trend, seasonality components (ETS) were used. Lag order and final models were selected by a combining usage of information criterion AICc and unit-root testing for each of estimated series. As the benchmark model for top-level (GDP) series simple univariate ARIMA and ETS models were estimated for each country in a similar fashion to all other series.

Reconciliation on ARIMA and ETS models managed to improve the forecast accuracy of GDP compared to respective univariate models only in 44% and 64% countries respectively. Results for the top-level forecasts for a fraction of assessed countries are given in the Table 2, and all results in the text include all 25 analysed countries. ETS based reconciliation outperformed ARIMA based reconciliation in 60% of countries in one-step-ahead forecasts. Results for all evaluated horizons and countries are visualized in Figure 2. In countries, where precision was improved, forecasting error measured by MASE decreased on average by 12.6%, 12.3%, 14.2%, 14.3% for horizon 1, 2, 3, and 4 respectively. Overall bottom-up approach performed worse for ETS based models with forecasting horizon 1 in all countries with the exception of Italy and Ireland.

		AT			BG			CZ			DK			HU		
	h	Uni.	Adj.	BU	Uni.	Adj.	BU	Uni.	Adj.	BU	Uni.	Adj.	BU	Uni.	Adj.	BU
									ARIMA							
MASE	1	0.191	0.164	0.237	0.123	0.128	0.192	0.216	0.205	0.322	0.233	0.225	0.266	0.182	0.179	0.249
	2	0.207	0.194	0.272	0.121	0.118	0.185	0.250	0.243	0.328	0.289	0.268	0.302	0.199	0.195	0.266
	3	0.200	0.197	0.288	0.121	0.118	0.192	0.274	0.268	0.383	0.314	0.297	0.321	0.222	0.216	0.280
	4	0.187	0.210	0.305	0.127	0.122	0.191	0.298	0.297	0.423	0.326	0.309	0.316	0.251	0.248	0.290
RMSE	1	546.62	470.02	677.37	141.23	146.95	220.48	396.14	375.58	592.19	514.20	495.82	586.29	313.63	308.91	429.89
	2	664.99	615.62	843.17	157.54	150.23	227.64	504.87	488.04	653.82	697.60	649.19	734.35	373.70	360.63	480.53
	3	687.32	670.04	892.43	160.37	154.52	242.98	558.31	546.71	770.13	753.29	718.49	775.18	423.86	408.03	507.40
	4	661.47	710.04	948.10	170.60	161.38	245.71	607.91	600.48	852.18	787.07	742.09	764.25	480.01	466.93	525.40
									ETS							
MASE	1	0.214	0.187	0.227	0.125	0.137	0.223	0.278	0.243	0.430	0.275	0.274	0.340	0.153	0.151	0.189
	2	0.237	0.222	0.261	0.154	0.167	0.242	0.359	0.304	0.460	0.355	0.348	0.394	0.195	0.191	0.223
	3	0.237	0.233	0.286	0.188	0.199	0.273	0.366	0.298	0.459	0.390	0.374	0.383	0.244	0.229	0.228
	4	0.234	0.252	0.315	0.225	0.233	0.294	0.352	0.295	0.439	0.442	0.405	0.395	0.305	0.280	0.236
RMSE	1	614.184	536.501	649.411	143.824	157.190	257.148	510.662	445.548	789.686	605.817	605.037	749.258	263.569	260.492	326.723
	2	757.061	699.233	802.584	194.557	213.977	299.400	690.822	586.847	860.760	839.037	827.920	949.027	366.217	353.217	406.214
	3	784.352	775.451	910.102	241.452	255.263	341.702	721.660	595.346	875.156	949.546	907.526	938.441	468.585	434.464	436.970
	4	777.209	849.093	1011.097	290.377	300.140	373.242	706.522	590.259	852.019	1073.481	980.698	971.463	599.312	540.521	453.596

Table 2 Comparison of reconciliation (Ajd.), bottom-up approach (BU) and univariate model (Uni.) on forecasting error of top-level series measured by MASE, RMSE across selected countries and forecasting horizons *h*

	AT		BG		CZ		DK		HU		AT		BG		CZ		DK		HU	
NA items	Uni.	Adj.	Uni.	Adj.	Uni.	Adj.	Uni.	Adj.	Uni.	Adj.	Uni.	Adj.	Uni.	Adj.	Uni.	Adj.	Uni.	Adj.	Uni.	Adj.
							MASE										**RMSE**			
P31_S13	0.189	0.200	0.213	0.220	0.158	0.159	0.453	0.489	0.236	0.262	69.37	73.59	27.08	28.0 6	43.92	44.11	113.86	123.10	37.38	41.56
P31_S14	0.145	0.150	0.275	0.235	0.147	0.142	0.327	0.325	0.181	0.170	273.09	282.61	161.44	1 37.97	107.54	104.08	378.45	376.60	90.93	85.33
P31_S15	1.480	1.481	0.362	0.362	0.715	0.712	0.911	0.910	1.618	1.622	18.87	18.88	3.15	3.15	5.29	5.27	12.93	12.92	18.21	18.24
P32_S13	0.165	0.161	0.312	0.347	0.091	0.085	0.668	0.628	0.672	0.661	63.15	61.80	41.76	46.3 7	44.18	41.14	75.70	71.16	123.23	121.15
P5G	0.224	0.205	0.269	0.267	0.289	0.277	0.846	0.840	0.404	0.365	355.92	325.01	150.55	149.01	327.03	312.91	664.02	659.49	494.40	447.42
P61	1.060	0.815	0.740	0.638	1.013	0.908	1.297	1.360	1.420	1.522	932.47	716.95	190.35	163.96	913.19	818.16	733.70	769.31	820.47	879.44
P62	0.150	0.152	0.112	0.106	0.397	0.393	0.455	0.460	0.354	0.348	213.75	216.83	70.68	67.09	127.37	126.13	292.11	295.42	140.21	137.53
P71	1.109	1.158	0.366	0.309	0.473	0.550	1.327	1.225	1.022	1.035	675.68	705.38	117.79	99.66	482.59	560.45	715.66	660.41	642.23	650.48
P72	0.282	0.290	0.355	0.332	0.532	0.535	0.831	0.809	0.552	0.577	221.10	227.01	68.82	64.36	168.67	169.51	313.33	305.30	110.14	115.20

Table 3 Comparison of forecast errors of unadjusted base (Uni.) and adjusted by reconciliation (Adj.) forecasts with ARIMA (h= 1)

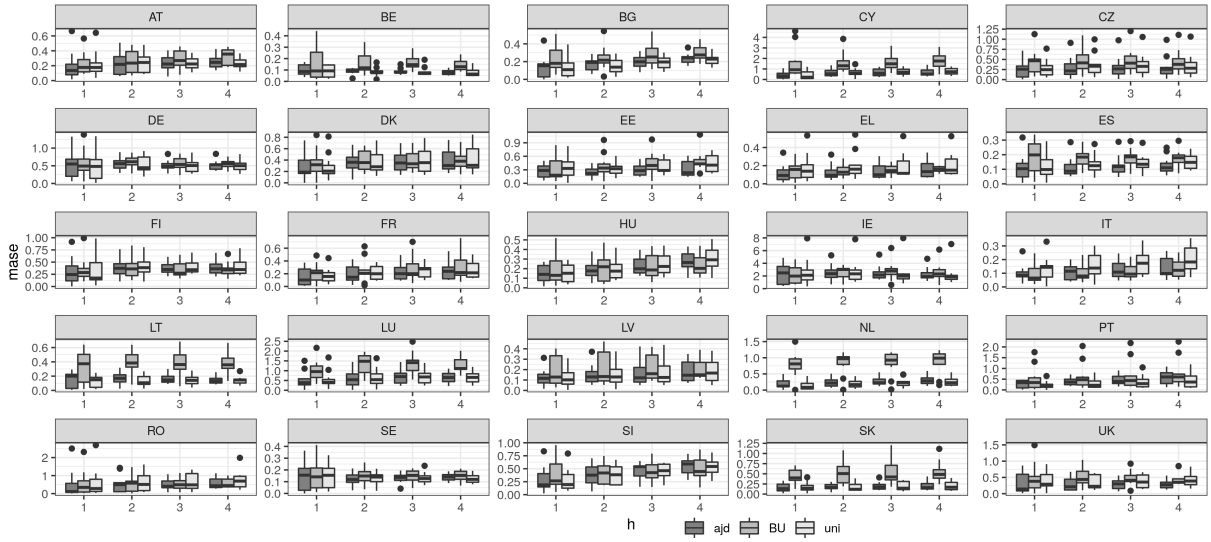


Figure 2 Top-level series forecasting errors for ETS based models across rolling origins, forecasting horizons and countries. Reconciliation (ajd.), bottom-up approach (BU) and univariate model (uni.)

Reconciliation, compared to the bottom-up approach, adjusts base forecasts in order to produce a coherent forecast for higher levels of the structure. This adjustment introduces further information, which may result in improvement in bottom-level forecasts. Results for bottom-level forecasts vary across different components and countries. ARIMA based reconciliation managed to improve precision over the base forecast in 59% of bottom-level series. Best results were for Gross capital formation, which was improved by reconciliation in 76% of countries. ETS based reconciliation achieved a similar result of 62% on the bottom-level series. Detailed results for selected countries can be seen in the Table 3. Similarly to top-level series, ETS provided better forecasts for bottom-level series than ARIMA in 51% of series.

In order to examine the reasons why in some countries, reconciliation with ARIMA performed poorly compared to the univariate model and reconciliation with ETS, time series features have been extracted. Features such as the strength of the trend, seasonality, spikiness (variance of the leave-one-out variance), lumpiness (variance of the variances), stability (variance of the means), linearity and curvature. For a detailed description of these features see [7] and [8]. The dimensionality of extracted features was reduced by principal component analysis and, the first six components managed to explain 99% of their variability. These components were utilized to visualize countries with similarities in the development of GDP, and countries with poor result were marked. In Figure 3 can be seen that these countries have high values for PC2 and PC3. These components mainly load features of linearity and curvature, which are calculated based on coefficients of an orthogonal quadratic regression of trend. For countries with poor performing reconciliation with ETS, no pattern have been found.

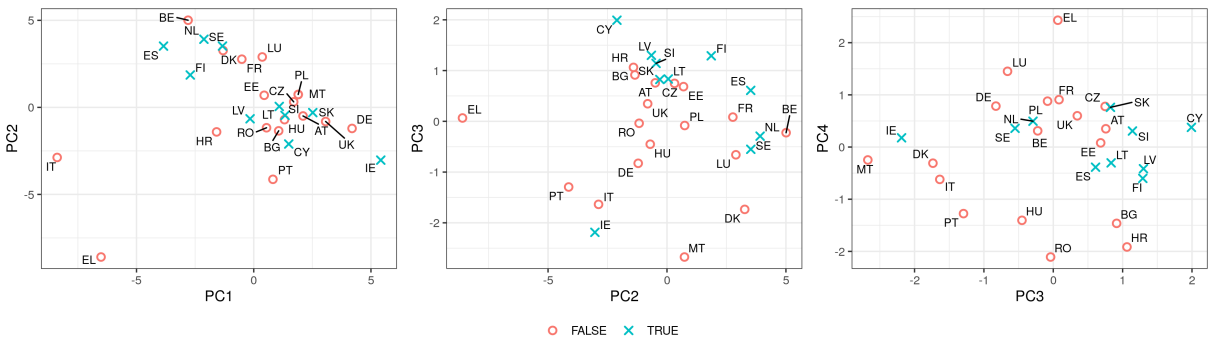


Figure 3 Principal components of time series features. Countries with poor performance are marked with a cross.

4 Conclusion

This contribution assessed forecasting improvement of estimating and forecasting disaggregated series of GDP components and their following aggregation and reconciliation on data for 25 European countries. Simple ARIMA model and ETS were used for modelling base-level series, which were aggregated with reconciliation and with the bottom-up approach to obtain a forecast for every component of Expenditure Approach. Forecasting error across all forecasted levels and different horizons was evaluated with rolling forecast origin against univariate models. Reconciliation with ETS proved to be a more viable option than the bottom-up approach and managed to improve top-level forecasts in 64% of countries, in these countries forecasting error for one-step-ahead forecasts was on average decreased by 12.6% compared to the univariate model. Similarly, forecasts for bottom-level series have been improved by reconciliation in 62% of cases. As reconciliation can be performed on forecasts obtained from any method, further research should focus on more complex methods for estimating individual series.

Acknowledgements

Supported by the grant No. IGA F4/27/2020, Faculty of Informatics and Statistics, University of Economics, Prague

References

[1] Athanasopoulos, G., Gamakumara, P., Panagiotelis, A., Hyndman, R. J. and Affan, M. (2020). Hierarchical Forecasting. In: *Macroeconomic Forecasting in the Era of Big Data* (Fuleky, P., ed.), volume 52 of *Advanced Studies in Theoretical and Applied Econometrics*. Springer International Publishing, 689–719. ISBN: 978-3-030-31149-0.

[2] Eurostat: Glossary: Expenditure approach – Statistics Explained. (2020). URL `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Expenditure_approach&oldid=347898`, (visited on 05/10/2020).

[3] Heinisch, K. and Scheufele, R. (2018). Bottom-up or direct? Forecasting German GDP in a data-rich environment, *Empirical Economics*; Heidelberg, **54**, 705–745. ISSN: 03777332.

[4] Hyndman, R. J. (2014). Measuring forecast accuracy .

[5] Hyndman, R. J., Ahmed, R. A. and Athanasopoulos, G. (2007). Optimal combination forecasts for hierarchical time series. Monash University, Department of Econometrics and Business Statistics, 9/07.

[6] Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, Oct. 1, 2006, ISSN: 0169-2070.

[7] Hyndman, R. J., Wang, E. and Laptev, N. (2015). Large-Scale Unusual Time Series Detection. *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, Atlantic City, NJ, USA: IEEE, Nov. 2015, pp. 1616–1619, ISBN: 978-1-4673-8493-3.

[8] Wang, X., Smith, K., and Hyndman, R. (2006). Characteristic-Based Clustering for Time Series Data, *Data Mining and Knowledge Discovery*, vol. 13, no. 3, pp. 335–364, Sep. 25, 2006, ISSN: 1384-5810, 1573-756X.

[9] Wickramasuriya, S. L., Athanasopoulos, G. and Hyndman, R. J. (2019). Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization, *Journal of the American Statistical Association*, vol. 114, no. 526, pp. 804–819, Apr. 3, 2019, ISSN: 0162-1459, 1537-274X.

Indicators for Efficiency Analysis of Health Systems with DEA

Mária Grausová[1], Miroslav Hužvár[2]

Abstract. The comparisons of productivity and efficiency of national healthcare systems attract a lot of attention since they may provide valuable information for decision makers at all levels and contribute to sharing good practices. Data Envelopment Analysis (DEA) represents a flexible and powerful instrument for such comparisons. In our paper we discuss a proper selection of inputs and outputs that represents one of the key steps in the analysis. The inputs in our approach cover the human health resources along with the performance of healthcare systems and the population behaviour. Some of the inputs are based on the composite indicators included in Euro Health Consumer Index and Euro Heart Index. We illustrate the application of selected indicators with the output oriented, slacks-based DEA model under variable returns to scale to assess the efficiency of European countries in supporting the health outcomes of their inhabitants by national healthcare systems. The results provide a benchmarking of the assessed countries and identify the potential for improvement of inefficient countries.

Keywords: health system efficiency, data envelopment analysis, slacks-based measure

JEL Classification: C61, I15
AMS Classification: 90C08

1 Introduction

The main function of a health system is to support the health conditions of the population which are traditionally measured by health outcomes such as life expectancy and mortality. Personal and capital resources of health systems are built to assure availability and provision of health services, and thus improve the health outcomes. From this perspective it is natural to study the efficiency of national health systems, i.e. to ask whether the health resources and provided services are efficiently used to reach desirable health outcomes of the population.

However, it is generally accepted that the impact of a health system on the health outcomes of inhabitants is limited and the differences among countries cannot be fully explained without taking into account a wider context, e.g. the life style of the population. This context includes a large variety of factors and it is hard to exactly evaluate the contribution of individual factors to improve or worsen the health outcomes in individual countries. That is why it may be better to rely on composite indicators designed by experts in the health sector that incorporate several important factors to provide a more complex characteristics of the national context.

Our approach to the assessment of health system efficiency is based on DEA methodology. The aim of the paper is to discuss the choice of proper input and output indicators for this purpose and suggest a solution that incorporates a larger number of factors into the model. The application of selected inputs and outputs is illustrated using a slack based DEA model (SBM) designed by Tone [9].

A few studies applying DEA to compare national health systems have been published in latest years. We address some drawbacks of the studies in this paper. In order to reduce the heterogeneity of socio-economic and geographic environment influencing the health conditions, we confine to health systems within a group of 30 European countries including EU members, Norway, Switzerland, and United Kingdom. These countries can be considered sufficiently homogeneous in socio-economic conditions. Moreover, the composite health indices [6, 7] describing the current state of health care provision are available for all of them. Thus,

[1] Matej Bel University, Faculty of Economics, Department of Quantitative Methods and Information Systems, Tajovského 10, 974 01 Banská Bystrica, Slovak Republic, maria.grausova@umb.sk.
[2] Matej Bel University, Faculty of Economics, Department of Quantitative Methods and Information Systems, Tajovského 10, 974 01 Banská Bystrica, Slovak Republic, miroslav.huzvar@umb.sk.

we can assume the selected countries apply equivalent measures to evaluate their healthcare delivery and are willing to share the best practices to improve their performance. The selected input and output indicators are discussed in Section 2. The results of analysis are discussed in Section 3. Conclusions and open issues are presented at the end of the paper.

2 Health Care Indicators for DEA

In accordance with the traditional view, we consider efficiency as a relative productivity of individual countries (acting as decision making units) in the use of health resources and services to reach the desired health outcomes. The resources and services consuming the costs of healthcare provision are inputs and the health outcomes are natural outputs in efficiency assessment.

In most of previous studies, labour and capital resources used in the health systems along with health expenditure per capita are considered as inputs while the health outcomes of the population being outputs [1]. Bhat [2] considered practising physicians, nurses, in-patient beds and pharmaceuticals as inputs to treat populations of various age groups and examine the efficiency of healthcare delivery systems in OECD countries. Hadad at al. [5] compared the health systems by two models, using discretionary inputs in the former and inputs beyond the control of health systems in the latter.

Hsu [8] used DEA to evaluate health expenditures to demonstrate how productivity had changed over time for selected countries in Europe and Central Asia. Grausová et al. [4] applied DEA to analyse the development of healthcare systems in the Visegrád group with regard to other European countries. Dlouhý [3] used a two-stage DEA approach to evaluate the efficiency of health systems in two steps: efficiency of health resources to produce health services and efficiency of health services to produce health of the population.

In our study we use input indicators that capture the human resources of health systems, the provision of health services, the access to healthcare and the prevention measures including the behaviour of the population. Output indicators characterize the healthy life length and mortality till the age of 65 as the outcomes that may be significantly influenced both by health system and the lifestyle of the population.

The inputs and outputs used in our analysis are briefly summarized in Table 1 and explained in detail below. Note that the indicators are independent of the size of a country. All data on the inputs and outputs used in the analysis refers to the year 2016 and is collected from the World Bank [10] and Health Consumer Powerhouse [6, 7].

	Indicator
Inputs	**Prevention** – based on EuroHealth Consumer Index, sub-discipline Prevention
	Procedures – based on Euro Heart Index, sub-discipline Procedures
	Access to care – based on Euro Heart Index, sub-discipline Access
	Health staff – number of physicians, nurses and midwives per 1,000 inhabitants
Outputs	**Survival rate to 65** of males
	Healthy life years at birth of males

Table 1 Indicators used in DEA efficiency assessment

In contrast with some of previous studies, we do not use the health expenditure as an input. Although data on health expenditure expressed in monetary units or as a percentage of GDP is widely available for a long time period and can be attractive for DEA as a unique measure covering all inputs, this indicator also has some essential drawbacks. Healthcare expenditures are difficult to interpret across countries as they are measured in different currencies with vastly different spending powers [2]. Due to the same reason we avoid using the indicators estimating the capital resources of healthcare systems, e.g. the hospital beds since their numbers may be also significantly influenced by different price levels in assessed countries.

The inputs in our analysis cover the human resources involved in providing healthcare and the scope and availability of healthcare services. The human resources are represented by a unique input – the number of **health staff** including doctors, nurses and midwives since the structure of these professionals strongly depends on their different roles and responsibilities in individual countries. The number of health staff per 1,000 inhabitants also varies widely across European countries, in 2016 it was lowest in Cyprus, Poland, and Latvia, and highest in Switzerland, Norway, and Denmark.

The level of healthcare services is expressed by three aggregated input indicators that are derived from the indicators Euro Health Consumer Index (EHCI) [7] and Euro Heart Index (EHI) [6]. Both EHCI and EHI include several dozens of indicators grouped into a few sub-disciplines. A country is assigned a value of 1, 2 or 3 for each indicator, depending on the level of compliance with the given criteria. A higher value represents better performance that is expected to have a positive impact on the health outcomes The value of our aggregated input for each country is calculated as the mean of values of all indicators in the corresponding sub-discipline of EHCI or EHI.

The inputs Procedures and Access to care are calculated based on corresponding sub-disciplines of EHI, reflecting the fact there has been a general advancement of therapies and procedures, contributing to reducing heart mortality in Europe and reducing cardiovascular disease from the position of the most prominent cause of death in numerous European countries. The achievement of good outcomes for people suffering from any cardiovascular disease is dependent on the provision of well-organised and coordinated services which draw on the knowledge and skills of health and social care professionals working across primary and secondary care [6].

- The input **Procedures** estimating overall health care performance is measured by the provision of cardiovascular disease services as an average value of indicators included in EHI 2016 sub-discipline – Procedures. This sub-discipline contains 11 indicators: time of the door to balloon delay for STEMI patients; the existence of health care and paramedical personnel certified for latest/appropriated cardiopulmonary resuscitation; the availability of thrombolysis as part of treatment given in ambulances or in primary care settings; emergency defibrillators widely available in public places; % of advised patients participating in rehabilitation / availability of rehabilitation; home care available for cardiac patients; total number of percutaneous coronary interventions per million population aged 50+; ratio of procedures: number of percutaneous coronary intervention (PCI) / coronary artery bypass (CABG); statin deployment (SU per capita 50+ SDR adjusted); clopidogrel deployment (SU per capita 50+ SDR adjusted); PCSK-9 inhibitor deployment (SU per capita 15+). The value of this input in 2016 was lowest in Bulgaria, Cyprus, Latvia, Romania, and Slovakia, and highest in France, Germany, and Netherlands.
- The input **Access to care** is measured by average value of indicators included in EHI 2016 sub-discipline Access to care. People with cardiovascular disease or who are at high cardiovascular risk need early detection and management using counselling and medicines, as appropriate. There are 6 indicators: average waiting time to echocardiography and diagnostics for suspected heart disease; average waiting time for non-acute revascularization (CABG, PCI) from time of catheterization; ratio: number of patients on waiting list / number of transplants per year; availability of support for families with children having congenital heart disease; access to free FH genetic testing; access to combination therapy to treat FH. The value of this input was lowest in Cyprus and Romania and highest (at the same level) in 5 countries: France, Luxembourg, Netherlands, Norway, and Sweden.
- The last input **Prevention** is calculated based on the corresponding sub-discipline of EHCI which include indicators capturing some important preventive measures applied within the health systems, but also indicators measuring some aspects of the behaviour or life style of the population with a significant positive or negative impact on the health outcomes. Of course, the values of latter indicators are not in direct control of the health system. However, they may be affected by appropriate governmental policies and human decision makers in a reasonably short time. The sub-discipline Prevention of EHCI 2016 contains 7 indicators that cover the prevention measures and population life style from different aspects: the infant 8-disease vaccination; the % of adult population registering high blood pressure; the actual cigarette sales per capita; the pure alcohol intake per capita; the hours of physical education in compulsory schools; the access to free HPV vaccination for teenage girls, and traffic deaths per 100,000 inhabitants. The values of the Prevention input were lowest in Romania, Bulgaria, Estonia, and Lithuania, and highest in Norway and United Kingdom.

As for outputs, life expectancy at births and infant mortality rate are frequently used in previous works. Instead, we choose the indicators of **healthy life years** (which incorporates the quality of population health as a main target of providing healthcare) and **survival rate to the age of 65** (which extends the infant mortality rate to cover a longer period of human life). The survival rate is defined as the ratio of survived persons to non-survived ones to the age of 65. The lowest numbers of healthy life years appear in Latvia and Estonia (both below 55) and the highest in Sweden, Spain, and Malta (over 70). The lowest values of survival rate are seen in Lithuania and Latvia (below 2 survived per 1 non-survived), the highest values in Switzerland, Sweden, Italy and Malta (over 8 survived per 1 non-survived). Basic descriptive statistics for inputs and outputs is given in Table 2.

	Prevention	Procedures	Access to care	Health staff	Survival rate to 65	HLY at birth males
Max	2.86	2.72	2.67	22.47	8.99	73.00
Min	1.15	1.00	1.01	7.20	1.83	52.30
Average	2.18	2.02	2.08	12.72	5.61	62.22
SD	0.39	0.46	0.47	3.91	2.17	4.95

Table 2 Descriptive statistics of inputs and outputs

3 Applied DEA model

In order to measure the efficiency of healthcare systems in European countries, we apply an output oriented non-radial (slacks-based) DEA model under variable returns to scale (VRS) [9]. The choice of the model is based on the following assumptions:

1. The primary target of the health system is to improve the health outcomes, i.e. the outputs of the model. Hence, the efficiency score used for ranking should measure the potential of a country for output improvements while the possible slacks observed on inputs are considered as additional information.
2. The position of some countries within the assessed group is rather different for the two outputs (see Figure 1). Thus, non-proportional recommendations for their improvement are naturally expected from the DEA.
3. The character of inputs and outputs as well as the mutual comparisons across countries advocate the use of VRS since at least the countries with better health outcomes can hardly reach additional increase of the outcomes that would be proportional to the increase in the inputs.

As can be seen in Table 3, all inputs and outputs are positively correlated, but the correlations within the inputs and within the outputs confirm that each indicator captures a specific characteristics of the healthcare and it is reasonable to consider all of them in the analysis. The inputs Prevention and Access to care exhibit the strongest correlation with the outputs. Both the outputs are naturally expected to be influenced by the quality of the healthcare as well as by the behaviour of the population measured by selected inputs.

	Prevention	Procedures	Access to care	Staff	Survival rate to 65	HLY at birth males
Prevention	1	0.75	0.73	0.54	0.75	0.48
Procedures		1	0.86	0.64	0.59	0.30
Access to care			1	0.72	0.66	0.34
Health staff				1	0.52	0.33
Survival rate to 65					1	0.71
HLY at birth males						1

Table 3 Linear correlation coefficients for the indicators

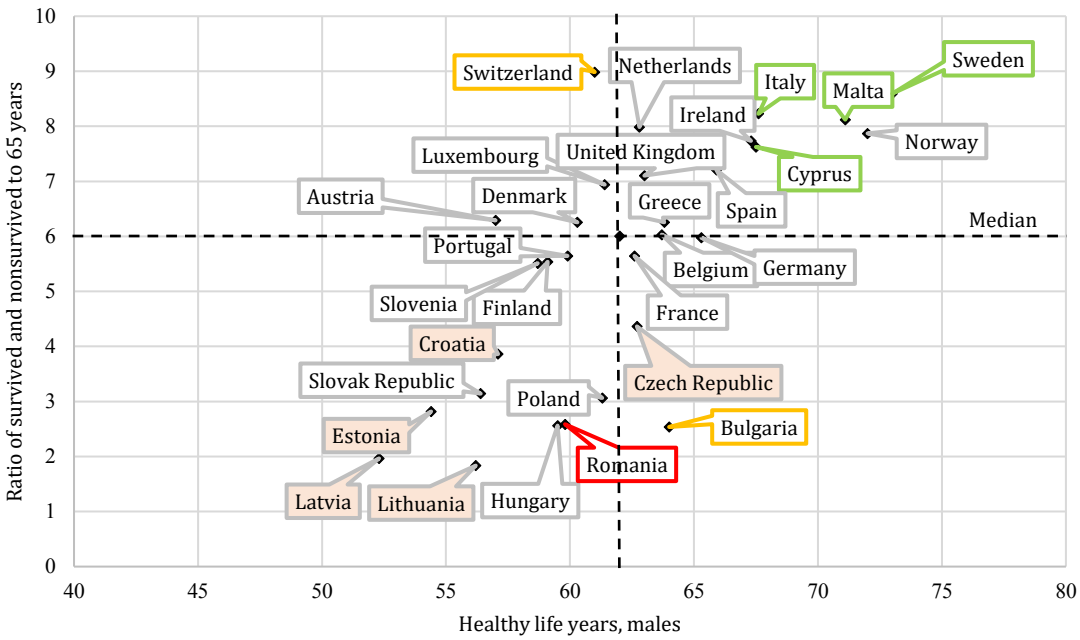


Figure 1 Comparison of European countries by health outcomes, the efficient countries are highlighted by red, orange and green frames

4 Results and Discussion

The benchmarking of national healthcare systems by the output oriented, slacks-based measure under VRS (SBM-O-V) is shown in Table 4. We can see regional differences in Europe that are mostly determined by the level of health outcomes. The countries in Western, Northern and Southern Europe reach significantly higher efficiency scores than those in Central and Eastern Europe. Thus, the comparison is much more valuable within the regions.

The model identifies seven fully efficient countries, namely Bulgaria, Cyprus, Italy, Malta, Romania, Sweden and Switzerland (see Table 4). However, not all of them really represent the best practices for inefficient countries. Switzerland and Bulgaria are in a special situation since they are good in one output, but poor in the other output (framed by orange colour in Figure 1). Fortunately, they are in the reference set only for themselves and do not have any impact on the efficiency of other countries. But Romania, that is efficient by default due to VRS with extremely low inputs and both outputs below median, is an efficient peer (together with Cyprus) for Croatia, the Czech Republic, Estonia, Latvia and Lithuania although the Czech Republic reaches significantly better values of both outputs, Croatia and Estonia are better in the ratio of survived to 65. Thus, Romania (framed red in Figure 1) can be hardly acceptable as a referential peer for countries with better health outcomes (shaded red in Figure 1).

Hence, really acceptable efficient peers seem those highlighted by green frame in Figure 1. Note that Cyprus that is referential for many inefficient countries is also a special case since it reaches very good outputs with very low two inputs (Access to care and Procedures).

A closer look on the results reveals some other drawbacks of the applied DEA measure. For some countries, there are great differences in the values of the weights that can be interpreted as the virtual costs and prices of inputs and outputs, respectively [9]. The disproportions in weights mean that the model does not consider all the inputs and outputs sufficiently important. We also observe many zero weights for inputs. Moreover, numerous slacks on inputs imply recommendations for their reduction that may be contradictory to the improvement of health care and health outcomes.

Country	Score	Rank	Country	Score	Rank
Bulgaria	1	1	Denmark	0.80	16
Cyprus	1	1	Germany	0.80	17

Country	Score	Rank	Country	Score	Rank
Italy	1	1	Slovenia	0.79	18
Malta	1	1	Austria	0.78	19
Romania	1	1	Portugal	0.78	20
Sweden	1	1	France	0.77	21
Switzerland	1	1	Czech Republic	0.77	22
Norway	0.95	8	Croatia	0.75	23
Ireland	0.94	9	Finland	0.74	24
Spain	0.93	10	Estonia	0.68	25
Netherlands	0.91	11	Slovak Republic	0.55	26
United Kingdom	0.89	12	Poland	0.55	27
Greece	0.88	13	Lithuania	0.52	28
Luxembourg	0.84	14	Hungary	0.47	29
Belgium	0.80	15	Latvia	0.43	30

Table 4 Efficiency scores and ranking by SBM-O-V model

5 Conclusion and further research

Finding a proper trade-off between an efficient use of health resources and services and the improvement of health outcomes that is critically dependent on the further development of the resources and services requires a special methodology for healthcare efficiency assessment to provide valuable knowledge for experts, policy makers and the public.

We discussed some drawbacks of the application of traditional DEA models to efficiency assessment of national healthcare systems. We showed how the drawbacks can be reduced by a more careful selection of assessed systems and input and output indicators covering a larger number of factors. Some restrictions of the suggested solution consist in the fact that the total volume of health services provided by healthcare systems were estimated by specialized procedures that cover only a part of healthcare provision (cardiovascular diseases). The results of the analysis may be also negatively influenced by missing data of EHI and EHCI indicators for some of assessed countries.

However, the illustrative example shows that traditional DEA models do not ensure a comparable importance of all indicators considered in the analysis since they prefer to minimize the role of weaker points of each health system. They also automatically provide recommendations for reducing the inputs that may be inevitable for providing a good healthcare. With proper weight restrictions it may be possible to modify the efficient frontier for DEA models so that all the inputs and outputs are taken into account with relevant weights and the analysis is focused on the improvement of health outcomes, health resources and services although some of them are treated as inputs and the others as outputs. The models can be adjusted with respect to the preferences of the health policy makers to follow good practices in efficient healthcare systems. This opens a space for further research.

Acknowledgements

This work was supported by the Grant Agency of Slovak Republic – VEGA 1/0843/18 “Methodological aspects of DEA application on efficiency assessment of production units”.

References

[1] AFONSO, A. & ST AUBYN, M. (2005). Non-parametric approaches to education and health efficiency in OECD countries. *Journal of Applied Economics, 8*, 227–246.

[2] BHAT, V. N. (2005). Institutional arrangements and efficiency of health care delivery systems. *The European Journal of Health Economics, 6*, 215–222.

[3] DLOUHÝ, M. (2015). Long-Term Trends in Efficiency of National Health Systems: An Evaluation by Two-Stage Data Envelopment Analysis. In *Proceedings of Mathematical Methods in Economics MME 2015: 33rd international conference* (pp. 127–132). Plzeň: University of West Bohemia.
[4] GRAUSOVÁ, M., HUŽVÁR, M., & ŠTRANGFELDOVÁ, J. (2014). Healthcare Systems Efficiency in the Visegrád Group. In *Applications of Mathematics and Statistics in Economics*. Conference Proceedings Full Text Papers, (pp. 104–113). Wroclaw: Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.
[5] HADAD, S., HADAD, Y., & SIMON-TUVAL, T. (2013). Determinants of healthcare system's efficiency in OECD countries. *The European Journal of Health Economics, 14 (2),* 253–265.
[6] HEALTH CONSUMER POWERHOUSE. (2016). Euro Heart Index 2016 Report. https://healthpower-house.com/media/EHI-2016/EHI-2016-report.pdf.
[7] HEALTH CONSUMER POWERHOUSE. (2017). Euro Health Consumer Index 2016 Report. https://health-powerhouse.com/media/EHCI-2016/EHCI-2016-report.pdf.
[8] HSU, Y.-CH. (2013). The efficiency of government spending on health: Evidence from Europe and Central Asia. *The Social Science Journal 50*, 665–673.
[9] TONE, K. (2001). A slacks-based measure of efficiency in data envelopment analysis, *The European Journal of Operational Research, 130*, 498–509.
[10] WORLD BANK DATA INDICATORS. (2016). Retrieved February 13, 2020, from http://data.worldbank.org/indicator.

Efficiency Measurement of OECD Insurance Industries with Data Envelopment Analysis

Biwei Guan[1]

Abstract. In recent years, the research of efficiency measurement has been widely concerned, especially in the insurance industry. As a financial institution, the insurance industry has the function of risk dispersion. How to maintain and improve the development speed, operating conditions, and social contribution of insurance companies need to be solved urgently. As a non-parametric method, DEA can be used to calculate the various efficiency values of the insurance industry. As an important organization, OECD is also worth studying. This paper selects 20 insurance markets from OECD member countries as research objects. The life insurance market and non-life insurance market are discussed respectively, with data from 2013 to 2017. By using the BBC model and the CCR model, the corresponding technical efficiency, pure technical efficiency, and scale efficiency are obtained. We found that during the sampling period, Germany's life insurance market and non-life insurance market achieved technical efficiency and scale efficiency. From the results, we also have a better understanding of the efficiency and environment of the OECD insurance market.

Keywords: data envelopment analysis, insurance industries, OECD

JEL Classification: C67, G15, G22
AMS Classification: 60H99

1 Introduction and Literature Review

In recent years, efficiency measures related to the insurance industry have been popular and attracted many regulators and investors. For the measurement of frontier efficiency, there are two main methods: Stochastic Frontier Analysis (SFA) and data envelopment analysis (DEA). Kaffash, Azizi, Huang, and Zhu [12] mentioned that recent papers have applied DEA rather than SFA to the evaluation of insurance companies, and more and more interested in evaluating the DEA efficiency of the insurance industry rather than other financial institutions. Therefore, this paper uses the DEA method to evaluate the frontier efficiency of insurance companies.

The purpose of this paper is to evaluate the technical efficiency, pure technical efficiency and scale efficiency of 20 OECD insurance industries through data envelopment analysis, and to understand the reasons and influencing factors of each market not reaching the effective state. This paper is divided into several parts. The first part is the introduction, the second part is the related previous research. The next section includes information about the selected samples, variables, and methods. The fourth part is the empirical results, which will show the results of technical efficiency, pure technical efficiency, and scale efficiency in each country. This part will introduce the findings of this paper, and the last part is the conclusion.

In previous studies, DEA and SFA were most commonly used to assess the efficiency of the insurance industry. Initially, SFA has advantages that DEA does not have. It can analyze influencing factors. However, after much research on improving and innovating DEA models such as Barros et al. [2], Kao and Hwang [13], and Yang and Pollitt [17] have been produced, this advantage has been greatly weakened. There is various research on the efficiency of insurance companies. Eling and Luhnen [7] mentioned that there are more than 90 studies on efficiency measurement in the insurance industry. Previous work focused on different documents. Eling and Schaper [8] and Vencappa et al. [16] observed changes in productivity and efficiency; Eling and Jia [6] and Greene and Segal [10] analyzed the relationship between efficiency and profitability; Kader et al. [11] studied the relationship between cost efficiency and board composition. What is missing from previous studies is that they only focus on one type of insurance market, or the samples are limited. Current research focuses on the U.S. market and European countries. In previous studies, many papers also chose OECD member countries as research objects. In Rai [15], more extensive cross-border research involving

[1] VŠB – Technical University of Ostrava, Department of Finance, Sokolská tř. 33, Ostrava 70200, Czech Republic, biwei.guan@vsb.cz

11 OECD countries; Donni and Fecher [5] studied the measurement of technical efficiency in 15 OECD insurance industries, etc.

In Farrell [9], a method was created to evaluate the efficiency of modern companies, which mentioned that the efficiency of manufacturers includes two parts: technical efficiency (TE) and allocative efficiency (AE), both of which are included in total cost efficiency (CE). Technical efficiency can be further divided into two parts: pure technical efficiency (PTE) and scale efficiency (SE). PTE reflects the production efficiency of DMU input with the best scale, and the scale return can be changed; SE reflects the gap between the actual scale and the best production scale. This paper uses the CCR model (Charnes et al. [3]) and the BBC model (Banker et al. [1]) to study the technical efficiency, pure technical efficiency, and scale efficiency of insurance companies.

2 Data and Methodology

This paper collected the data from OECD [14], 20 OECD insurance industries were included, and the data was from 2013 to 2017.

2.1 Selected Indicators

The input and output indicators used in this paper are shown in Table 1. In the past, the commonly used input indicators are divided into three categories: labour input, capital input, and other material input. (Irene and Jia [6], Irene and Luen [7], Irene and sharp [8]. When selecting input indicators, this paper refers to these indicators. When considering the output indicators, the non-insurance industry pays more attention to claims, while the life insurance industry pays more attention to income. Therefore, this paper selects different output indicators for them. It should be emphasized here that some previous studies have selected premium income as an output indicator. For example, Irene and Jia [6] use net premium income as one of the output indicators. In Kader et al. [11], the contribution rate of the total premium is chosen as the output indicator, but the actual situation is not ideal. Because premium income is the concept of output multiplied by price, the price here includes not only the component of risk prevention but also the investment factor, cost factor and profit factor of the company. Therefore, the premium income is not appropriate.

	Life insurance market	Non-life insurance market
Inputs indicators	1. Number of companies	1. Number of companies
	2. Debt capital	2. Debt capital
	3. Equity capital	3. Equity capital
Outputs indicators	1. Total investments	1. Total investments
	2. Net income + Technical provisions	2. Claims payments + Technical provisions

Table 1 Summarize of inputs and outputs indicators

2.2 Data Envelopment Analysis

Data envelopment analysis (DEA) is suitable for the evaluation of complex multi-output and multi-input problems. One of the characteristics of DEA is that it does not need any weight assumption before an analysis but uses the actual data input by DMU to obtain the best weight. This function enables the DEA to eliminate many subjective factors. Therefore, DEA has strong objectivity. The assumption of the CRR model is that in the production process, the scale return is fixed. When the input changes in proportion, the output should also change in proportion. For the input-oriented CCR model, there are the following constraints:

$$\begin{gathered} \min \theta \\ s.t. \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_0 \\ \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_0 \\ \lambda_j \geq 0, i = 1,2,\ldots,n; j = 1,2,3,\ldots,n; r = 1,2,\ldots,n \end{gathered} \tag{1}$$

where x_{ij} represents the *i-th* inputs on the *j-th* DMU, y_{rj} represents the *r-th* outputs on the *j-th* DMU, they are scalar vectors, here are three inputs, two outputs and 20 DMUs; λ_j is a scalar vector, and θ is an input radial measure of technical efficiency. Among them, the optimal solution is θ^*, $1-\theta^*$ represents the maximum input that can be reduced without reducing the output level at the current technical level. A larger θ^* means a smaller amount of input can be reduced, which means higher efficiency. When $\theta^* = 1$, it means that DMU is in a technical effective state currently.

The CCR model's assumptions apply when all manufacturers operate at an optimal scale. But in fact, due to financial constraints, incomplete competition, government regulations, and other factors, DMU is not in the best production state. Therefore, the BBC model is proposed to solve this problem. BBC model has almost the same constraints as the CCR model. The only difference is that in the BBC model, there is also a constraint on λ, which can basically ensure that manufacturers of similar size are compared with manufacturers that are not valid, rather than manufacturers with large gaps. The constraint is as follows:

$$\sum_{j=1}^{n} \lambda_j = 1 \tag{2}$$

Here, in the input-oriented BBC model, technical efficiency does not include scale efficiency, which is called pure technical efficiency. The calculation formula of scale benefit is:

$$SE = \frac{TE}{PTE} \tag{3}$$

3 Empirical Results

The results were calculated by DEA solver lv8. DEA solver lv8 is developed by Kaoru tone and can be applied to many different DEA models, such as BBC, CCR, IRS, DRS, etc. From this software, you can get the results like score, projection, weight, graph, etc. In the following analysis of each country, the time series is 5 years, this paper analyzes the data of each year separately and then selects their average value as the final result.

3.1 Life Insurance Industry

Inputs are the number of companies, debt capital, and equity capital. The outputs are the sum of total investment, technology supply, and net income. The results of DEA are shown in Table 2.

	TE	PTE	SE
Australia	0.4313	0.5099	0.7082
Belgium	0.9585	0.9625	0.9959
Denmark	0.9434	0.9467	0.9965
Finland	0.9726	0.9892	0.9833
Germany	1	1	1
Greece	0.7656	0.8206	0.9062
Hungary	0.9663	0.9697	0.9964
Iceland	0	1	0
Ireland	0.9968	0.999	0.9977
Italy	0.981	0.9854	0.9955
Luxembourg	0.975	0.9766	0.9984
Mexico	1	1	1
New Zealand	0.5277	0.5339	0.9889
Norway	1	1	1
Poland	0.9701	0.9718	0.9983
Portugal	0.9728	0.9739	0.9989
Spain	0.9294	0.962	0.9662

	TE	PTE	SE
Switzerland	0.996	0.9982	0.9978
Turkey	0.1712	0.2279	0.7708
United States	0.9718	1	0.9718

Table 2 Results of DEA for life insurance industry

From table 2, we can clearly find the countries that are in technical effective and scale effective, namely Germany, Mexico, and Norway. Among them, Iceland's technical efficiency is shown as "0", because the value of the original data from Iceland is only about 0.1%–0.01% of that of other countries, which leads to such results. Cummins and Weiss [4] mentioned that SFA's efficiency value is higher than DEA's, because DEA regards all deviations from the boundary as invalid, while SFA allows the use of random error terms. In Table 2, we can find that PTE from the BBC model is higher than the TE from the CCR model. The reason is that the BBC model ensures that invalid manufacturers are only compared to similar manufacturers. Whether it is for TE or PTE, Turkey is obviously the least efficient country, the value is only 0.17 and 0.23. This shows that Turkey's life insurance market can reduce its input by more than 77% without reducing its output level. Based on the information obtained from the BBC model, we can infer that for the Turkish life insurance market, if they want to achieve technical efficiency and effective scale, they should expand the scale. Interestingly, the United States, as the leader of insurance, has a long history, a large and mature market, but has not achieved technical efficiency.

3.2 Non-Life Insurance Industry

Input indicators are the same as for life insurance; outputs are the sum of total investment, technical reserves, and claim payments.

	TE	PTE	SE
Australia	0.8848	0.8905	0.9937
Belgium	0.7162	0.7195	0.9953
Denmark	0.9219	0.9234	0.9983
Finland	0.842	0.8515	0.9888
Germany	1	1	1
Greece	0.7713	0.8365	0.9016
Hungary	0	1	0
Iceland	0.7786	1	0.7786
Ireland	0.8753	0.8827	0.9904
Italy	0.8201	0.8222	0.9974
Luxembourg	0.7753	0.7824	0.9911
Mexico	0.7012	0.7219	0.9712
New Zealand	0.2878	0.2938	0.9798
Norway	0.7641	0.7677	0.9954
Poland	0.8762	0.8937	0.9806
Portugal	0.8959	0.9116	0.9827
Spain	0.865	0.8796	0.9841
Switzerland	0.8729	0.8755	0.997
Turkey	1	1	1
United States	1	1	1

Table 3 Results of DEA for non-life insurance industry

As can be seen from Table 3, for the non-life insurance industry, the efficiency value of Mexico and Norway is no longer "1". Germany still maintains the position of efficiency first, and the United States has achieved pure technical effectiveness and technical effectiveness. While Iceland and Hungary have achieved pure technical effectiveness. The most surprising is that Turkey has the worst performance in the life insurance market, with both technical and pure technical effectiveness in the non-life insurance market. As we mentioned earlier, the BBC model assumes that the scale gains can be changed in the production process, which means that when the input increases proportionally, the output will not necessarily increase proportionally, and the scale may increase or decrease. In the non-life insurance industry, the value of scale efficiency is more than 97% in most countries. Only Iceland, Hungary, and Greece have low scale efficiency values. Hungary has the same problems as Iceland's life insurance industry. Because the relevant data is much lower than in other countries, the technical efficiency value is calculated as "0", which results in the scale efficiency value of "0".

3.3 Summary

We are also interested in the comparison between overall life insurance and non-life insurance, which are composed of 20 OECD member countries. For the comparison between the life insurance industry and the non-life insurance industry, we select the average value of the 20 countries from each insurance industry from each year.

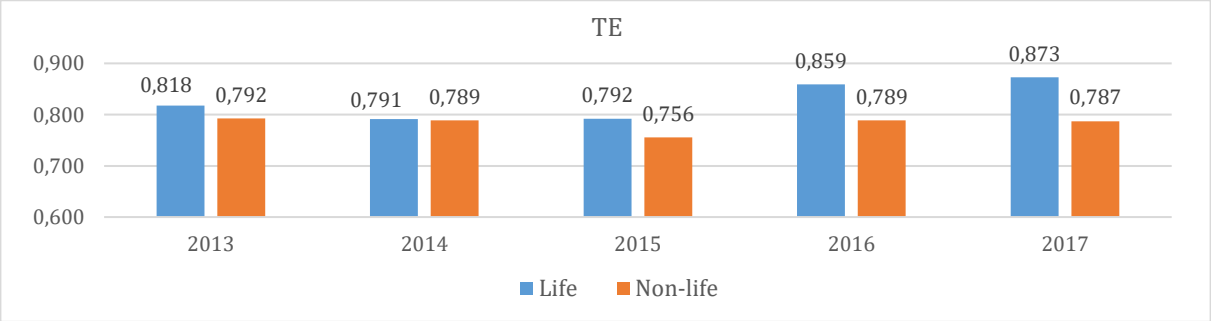


Figure 1 TE of life and non-life insurance industry from 2013 to 2017

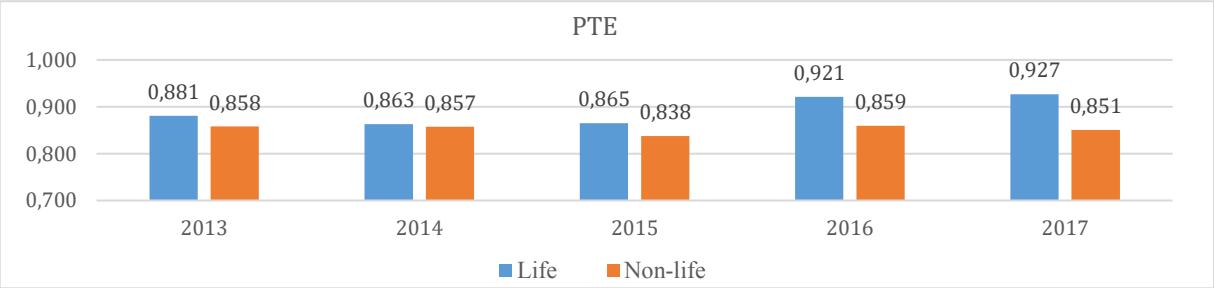


Figure 2 PTE of life and non-life insurance industry from 2013 to 2017

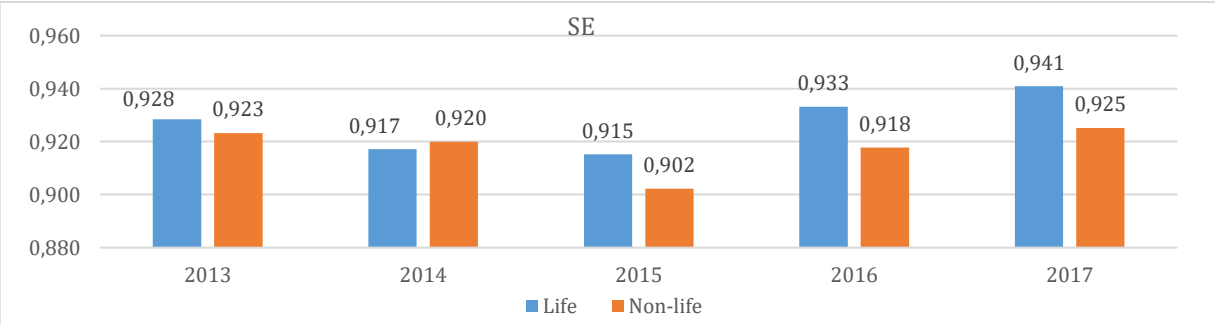


Figure 3 SE of life and non-life insurance industry from 2013 to 2017

For the life insurance industry, we can see from the three charts that the trend of TE, PTE, and SE began to decline in 2013 and began to rise after 2016, and the efficiency value in 2016 was higher than the previous

three years. In our opinion, the main reason is that in 2013, many countries/regions were inefficient and shrinking in scale, but in the second year, they did not make appropriate adjustments, resulting in reduced output and more input, thus making technology more efficient. For the non-life insurance industry, these three efficiency values change relatively little in five years. But in 2015, technical efficiency, pure technical efficiency, and scale efficiency are much smaller than other times. An important reason is that Greece's non-life insurance market experienced a cliff slide in 2015, especially in terms of technical efficiency. If we look at data from Greece over the past five years, we can see that in 2015, its technical efficiency was less than half of its average level.

From these three charts, we can draw a rough conclusion: the efficiency value of the life insurance market is usually higher than that of the non-life insurance market. The only exception in our sample is that in 2014, the scale efficiency of the life insurance market was lower than that of the non-life insurance market. In 2014, six countries in the life insurance market achieved economies of scale. They are Germany, Ireland, Mexico, Norway, Switzerland, and the United States. In the non-life insurance market, only Germany, Ireland, Turkey, and the United States have achieved an effective scale. So why is the scale efficiency value of the non-life insurance market still higher than that of the life insurance market? The reason is that although more countries in the life insurance market have reached an effective scale, in the non-life insurance market, all countries except Iceland have similar economies of scale, all of which are more than 95%. In the life insurance market, Australia, Greece, and Turkey are relatively inefficient in scale, which reduces the efficiency value of the whole market.

4 Conclusion

In this paper, the CCR model and BBC model are used to calculate the technical efficiency and pure technical efficiency respectively, so as to get the scale efficiency value. Through these data, we understand the efficiency status of each market and the general market environment.

The sample data are all from the life insurance market and the non-life insurance market of OECD member countries. Considering the lack and inaccuracy of some data, we have selected 20 countries from OECD as the analysis object, not all member countries. The innovation of this paper is that it analyses the life insurance and non-life insurance market at the same time and compares them.

Through this study, we confirm that the pure technical efficiency calculated by the BBC model is indeed higher than that calculated by the CCR model. Because the former eliminates the efficiency of the scale. Among our research objects, Germany is the best in both markets and has reached an effective state. Turkey has also achieved effective status in the non-life insurance market, but its efficiency value in the life insurance market is at the lowest point. We also have a general understanding of how to adjust a market to an effective position. In general, most countries have reached an effective position in the life insurance market.

Acknowledgements

The author was supported through the Czech Science Foundation (GACR) under project 20-25660Y and moreover by SP2020/11, an SGS research project of VSB-TU Ostrava. The support is greatly acknowledged.

References

[1] BANKER, R. D., CHARNES, A., & COOPER, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*. 30(9). 1031–1142.

[2] BARROS, C. P., NEKTARIOS, M., & ASSAF, A. (2010). Efficiency in the Greek insurance industry. European *Journal of Operational Research*. 205 (2). 431–436.

[3] CHARNES, A., COOPER, W. W., & RHODES, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*. 2 (6). 429–444.

[4] CUMMINS, J. D. & WEISS, M. A. (2000). Analyzing firm performance in the insurance industry using frontier efficiency methods. In: Dionne, G. (Ed.) *Handbook of Insurance Economics.* Kluwer Academic Publishers, Boston. 767–829.

[5] DONNI, O. & FECHER, F. (1997). Efficiency and productivity of the insurance industry in the OECD countries. *Geneva Papers on Risk and Insurance*. 22 (84). 523–535.

[6] ELING, M. & JIA, R. (2019). Efficiency and Profitability in the global insurance industry. *Pacific-Basin Finance Journal*. 57 (2019), 101190.

[7] ELING, M. & LUHNEN, M. (2010). Efficiency in the international insurance industry: A cross-country comparison. *Journal of Banking & Finance*, 34 (7), 1497–1509.
[8] ELING, M. & SCHAPER, P. (2017). Under pressure: how the business environment and efficiency of European Life insurance companies. *European Journal of Operational Research*, 258(2017), 1082–1094.
[9] FARRELL, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*: Series A (General). 120 (3). 253–281.
[10] GREENE, W. H. & SEGAL, D. (2004). Profitability and efficiency in the U.S. life insurance industry. *Journal of Productivity Analysis*. 21 (3), 229–247.
[11] Kader, H. A., Adams, M., Hardwick, P., & Kwon, W. J. (2014). COST EFFICIENCY AND BOARD COMPOSITION UNDER different takaful insurance business models. *International Review of Financial Analysis*. 32 (2014), 60–70.
[12] KAFFASH, S., AZIZI, R., HUANG, Y., & ZHU, J. (2019). A survey of data envelopment analysis applications in the insurance industry 1993–2018. *European Journal of Operational Research*. Available online 22 July 2019.
[13] KAO, C. & HWANG, S.-N. (2014). Multi-period efficiency and Malmquist productivity index in two-stage production systems. *European Journal of Operational Research*. 232 (3). 512–521.
[14] OECD. (2014–2018). OECD Insurance Statistics. Database version 2014, 2015, 2016, 2017 and 2018.
[15] RAI, A. (1996). Cost efficiency of international insurance firms. *Journal of Financial Services Research*. 10 (3), 213–233.
[16] VENCAPPA, D., FENN, P., & DIACON, S. (2013). Productivity growth in the European insurance industry: Evidence from life and non-life companies. *International Journal of the Economics of Business*, 20 (2), 281–305.
[17] YANG, H. & POLLITT, M. (2009). Incorporating both undesirable outputs and uncontrollable variables into DEA: The performance of Chinese coal-fired power plants. *European Journal of Operational Research*. 197 (3). 1095–1105.

Actual Revision of Cost and Revenue Functions for Reforestation System in Drahanska Highlands

David Hampel[1], Lenka Viskotová[2]

Abstract. This paper deals with simulation-based reforestation system described by the revenue and the cost functions estimated in 2011–2013. Actually, these functions are reestimated with data from period 2017–2019. The use of simulation techniques is necessary due to the fact that the available data exist only in the form of unit costs or revenues. Underlying regressions are statistically significant and coefficients of determination correspond to volatility of timber prices of particular tree species. Although the simulation procedure has not captured certain phenomena such as drought and bark beetle attack, relations among the estimated revenue and cost functions correspond to today´s reality. Profitability of spruce forests is the worst of all tree species included in our study; on the other hand, profitability of oak planting sustains the best. We can conclude that the used simulation procedure is robust enough to capture the current development in forestry; however, in the further work it is advisable to consider its extension, for example, in the direction of mixed forest cultivation.

Keywords: biodiversity, cost function, reforestation, revenue function, simulation

JEL Classification: C53
AMS Classification: 91B74

1 Introduction

We have been dealing with an optimal reforestation strategy from the point of view of the forest owners for a long time. We focus on the Drahanska Highlands with 80 800 ha of forest land, where the five most common tree species, i.e. beech, oak, pine, larch and spruce, occupy an area of 72 122 ha. In Regional forest development plan for Drahanska highlands area during 2000–2020 we can read about historical distribution of these species in the Drahanska highlands area, see Table 1.

Tree species	Beech	Oak	Pine	Larch	Spruce
Year 1732 [%]	16	44	11	0	0
Year 1830 [%]	18	8	10	0	2
Year 2010 [%]	15	8	12	8	46

Table 1 Historical and actual tree species distribution in Drahanska highlands area

It is visible that shares of beech and pine are historically almost the same. Further important information is the share of fir: 29 % in 1732 and 52 % in 1830. This relates with rapid oak elimination from 1732 to 1830. Fir was typical species in the Czech republic, specially in Drahanska highlands area it sustained till the first half of the twentieth century. Nevertheless, fir suffers from various harmful influences, e.g. air pollution, and it was gradually replaced by spruce planted mostly in monocultures from 1880 and much faster from 1910. Spruce has become dominant tree species over time. Since the end of the twentieth century voices which warn about this share has grown louder. The combination of drought and bark beetle, which has been intensively attacking and drastically devastating forests since 2015, has confirmed these warnings. Current situation is illustrated in Figure 1. In the left graph it is visible, that salvage felling has become almost 100 % of total felling. In the right graph we can see the reasons: bark beetle attack followed by a destructive weakening of spruce forests by drought and wind.

Desired changes of forest species structure were publicly discussed many years ago. Beside biological aspects and related risks it is necessary to respect the economic viewpoint including economic risks arising

[1] Department of Statistics and Operational Analysis, Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, qqhampel@mendelu.cz
[2] Department of Statistics and Operational Analysis, Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, lenka.viskotova@mendelu.cz

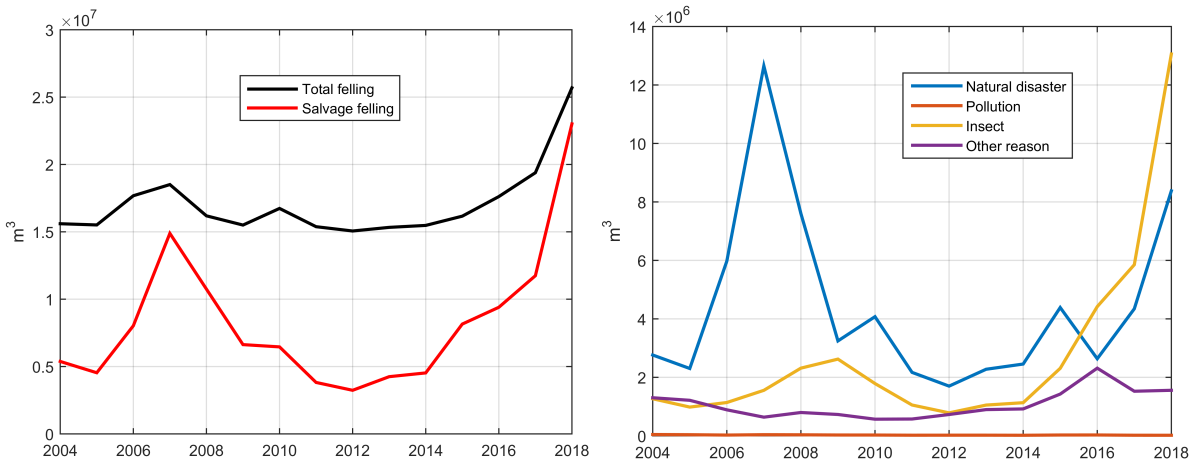


Figure 1 Total felling in the Czech Republic with the share of salvage felling (left graph), share of salvage felling caused by specific reason (right graph)

from the cultivation of spruce monoculture. In [7] the dynamic optimization technique is used to identify the optimal long run reforesting strategy for the forest owners in the Czech Republic and the optimal control was employed for finding the optimal time path for the state variables – the areas forested by the particular forest types. Optimal control techniques have been steadily employed in the forestry management research problems for analyzing the potential optimal strategies, see for example [1], [3] or [4]. Results of [7] are based on solving infinite horizon problem with current value profit maximization criterion. The profit function consists of the revenue and the cost functions, which are necessary to estimate. This is done using simulation technique. In the field of forestry, simulation techniques are used very often – especially for investigating growth or another production characteristics of a forest, see for example [9], [10] and [11].

The aim of this paper is to estimate the reforestation revenue and cost functions analogously to [7], but based on simulated data appropriate to actual situation in forestry. Reestimated revenue and cost function will be compared to old ones and observed differences will be discussed. The paper is organized as follows: section 2 describes motivation of our research in exact terms, data and simulation procedure, section 3 comprises results and the last section concludes.

2 Material and methods

In this section, we present the cost and the revenue function, the estimation of which is our goal, and the brief motivation of our work. Further, we discuss available real data and their sources. Finally, we describe simulating procedure needed for estimation of the cost and the revenue functions for particular species: beech, oak, pine, larch and spruce. For all calculations, i.e. simulation and regression, we use computational system MATLAB R2019b.

2.1 Cost and revenue functions

Results of this work are related to the decision problem described by infinite horizon optimal control problem with free terminal points, which can be schematically described as

$$\max V = \int_0^\infty \Pi(t) e^{-\rho t} \mathrm{d}t, \tag{1}$$

where

$$\Pi(t) = \sum_{i=1}^{n} R_i(x_i(t), u_i(t)) - K_i(x_i(t), u_i(t)), \tag{2}$$

under appropriate constrains. Here, n denotes the number of tree species appropriate for the given region, $x_i(t)$ are the state variables representing the area of land forested by species i at time t, ρ is discount rate. Further we denote total area reforested at time t by species i (in hectare per year) $u_i(t)$, the current profit $\Pi(t)$, total revenues from selling the timber and subsidies $R_i(t)$ and total costs of growing and logging at time t as $K_i(t)$. We do not consider forestation of non-forest land in our model. Further we assume that

$$R_i(t) = G_i(x_i(t)) + \sigma_i u_i(t), \tag{3}$$

where G_i represents a given function of revenues from logging and $\sigma_i u_i(t)$ is the subsidy (constant over time) derived from the increase in area of species i. Only the broadleaves species are considered to be supported by subsidies. Solution of this optimal control problem and further details are stated in [7] and some related discussion about timber prices stability is presented in [2]. For the estimation of the functions $G_i(x_i)$ and $K_i(x_i, u_i)$ we adapt the assumptions made in [3] and set

$$G_i = g_{i0} + g_{i1}x_i + \frac{1}{2}g_{i2}x_i^2, \tag{4}$$

$$K_i = k_i + a_i x_i + b_{i1}u_i + \frac{1}{2}b_{i2}u_i^2, \tag{5}$$

where the total cost function (5) is composed from the logging costs and reforestation costs. We assume $b_{i1} = b_{i2} = 0$ for the spruce type forest, which is supposed to be reduced. It means that decreasing the area of the forest type (after logging the current stand) is costless. Further details related to the estimation of the cost and the revenue functions parameters are stated in [6]. To proceed with the solution of optimal control problem (1) it is necessary to estimate parameters of the functions (4) and (5).

2.2 Data

There is limited availability of real data related to wood processing economics. We have collected the original data on the local forest structure, land characteristics and technological-economical characteristics of forest growing, thinning and the timber selling.

The data were acquired from the state agencies (Czech Statistic Office, The Forest Management Institute), from the Mendel University forest company (managing 10 000 ha of forest land in Morava region of the Czech Republic) and from the Costs of the usual measures of the Ministry of the Environment. This material serves as a basis for the evaluation of projects and measures within the subsidy programs of the Ministry focusing on nature and landscape protection. The costs within the document are expressed in terms of the current prices of activities that are usually implemented within the given type of measure. Data have the form of unit cost or revenue tabulation mainly, which omits possible changes given by the size of stumped area.

We use two datasets as input for simulation process. Actual situation is described by timber prices data from 2017 to 2019, and other prices (seedlings, human work, etc.) are related to the beginning of 2019. We wish to compare actual situation with the period 2011–2013 where no problems with spruce planting were visible for the general public. Analogously, timber price from 2011 to 2013 and other costs from beginning of 2013 are taken into simulation.

2.3 Simulation study

According to datasets insufficient for immediate regression analysis, a simulation procedures were suggested to generate the missing data using the information available. For generating new samples we use approach developed in [8]. Simulation is provided simultaneously for five tree species (spruce, oak, beech, larch, pine), where costs and revenues related to logging and reforesting can be established using known data. Detailed description of the simulation process is given in [5]. Generally, the process of costs simulation was expressed symbolically as

$$\text{costs} \sim f(\text{area, soil quality, terrain profile, working power, reforestation})$$

and the process of revenues simulation as

$$\text{revenues} \sim f(\text{area, soil quality, prices of different quality timber}).$$

Currently it is clear that two important factors were omitted in [5]: drought effects and bark beetle damage. We provide schematic illustration of both the costs and the revenues simulation including drought effects and bark beetle damage in Figure 2. Simulation scheme used in [5] and then in [7] is still usable for our purposes, because huge portion of drought and bark beetle effects is projected into prices of timber and working power prices.

We focus the simulation on forests of an area from 40 ha to 80 ha. For evaluation we assume huge number of these relatively small forest units, and more, we expect the same economics conditions for all these forest units (for example, the same curve of supply of working power or the same possibilities when buying seedlings). We select 70 years as the time horizon for simulations. Change of this horizon affects only absolute numbers, not relations among particular species. Administrative and similar costs are omitted and assumed to be the same for all forest units; again, it affects for example absolute costs per hectare, but not relations among costs for particular species.

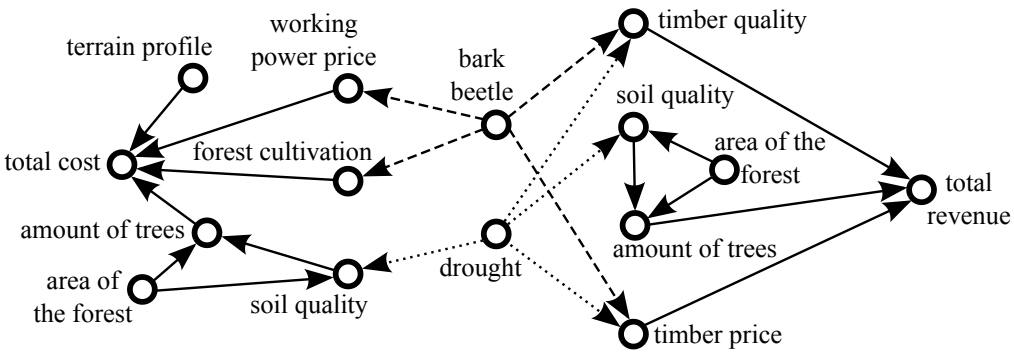


Figure 2 Scheme of costs and revenues simulation

3 Results and Discussion

On the basis of 100 000 simulated cost and profit values we estimate coefficients of the cost and the revenue functions (4) and (5), see Tables 2 and 3. All parameters except constants are statistically significant on the significance level $\alpha = 0.05$. Coefficients of determination for the cost function are always higher than 0.95, which points out the high quality of particular regressions. For the revenue functions we can see differences: the variability of simulated revenues for beech and oak was visibly lower than for other tree species, what results to higher coefficients of determination. This is in accordance with volatility of timber prices for particular tree species.

i	Species	k_i	a_i	b_{i1}	b_{i2}	R^2
1	Beech	1.7036	0.3340	0.3313	−0.0091	0.9670
2	Oak	−3.4534	0.3997	0.2642	0.3620	0.9701
3	Pine	0.9182	0.3547	0.2468	0.0675	0.9809
4	Larch	0.6229	0.4044	0.2775	0.0377	0.9779
5	Spruce	−1.7017	0.5657	×	×	0.9942

Table 2 Estimated parameters of the cost function for particular species; all values are in 10^6

i	Species	g_{i0}	g_{i1}	g_{i2}	R^2
1	Beech	7.6636	0.5378	−0.0024	0.8042
2	Oak	13.8751	0.7816	−0.0037	0.6653
3	Pine	3.4975	0.6931	−0.0034	0.3940
4	Larch	1.1305	1.0005	−0.0057	0.4787
5	Spruce	1.4971	0.9036	−0.0050	0.5458

Table 3 Estimated parameters of the revenue function for particular species; all values are in 10^6

It is possible to compare estimated parameters with those estimated based on data from 2011–2013, see [7]. Better insight is provided by graphics presentation given in Figure 3. Here we can see increase of the cost function for all species, especially for spruce. Difference between spruce and the other species can be explained as follows: at first, the prescribed cost function (5) underestimates the costs for the range from 70 to 80 ha; at second, the yield for spruce timber per hectare is higher than for other species, what means higher working power costs. Usually the higher yield per hectare means an advantage resulting in higher revenues, but actually it is not true. Huge supply of spruce timber, in most cases bark beetle affected – i.e. of low quality, cause fall in spruce timber prices. Effect of this is visible in Figure 3 (bottom left graph). Analogous, but not so deep fall of revenues is visible for pine. The revenue function of beech seems to be stable over time, and for larch and oak we can see an increase of the revenue function.

Finally, the profit function for particular tree species can be calculated, see Figure 4. In period 2011–2013 it is visible, that the most profitable tree species are oak and spruce. Pine and larch create a group with lower profitability than spruce, and beech exhibits the lowest profitability. This distribution was true roughly from

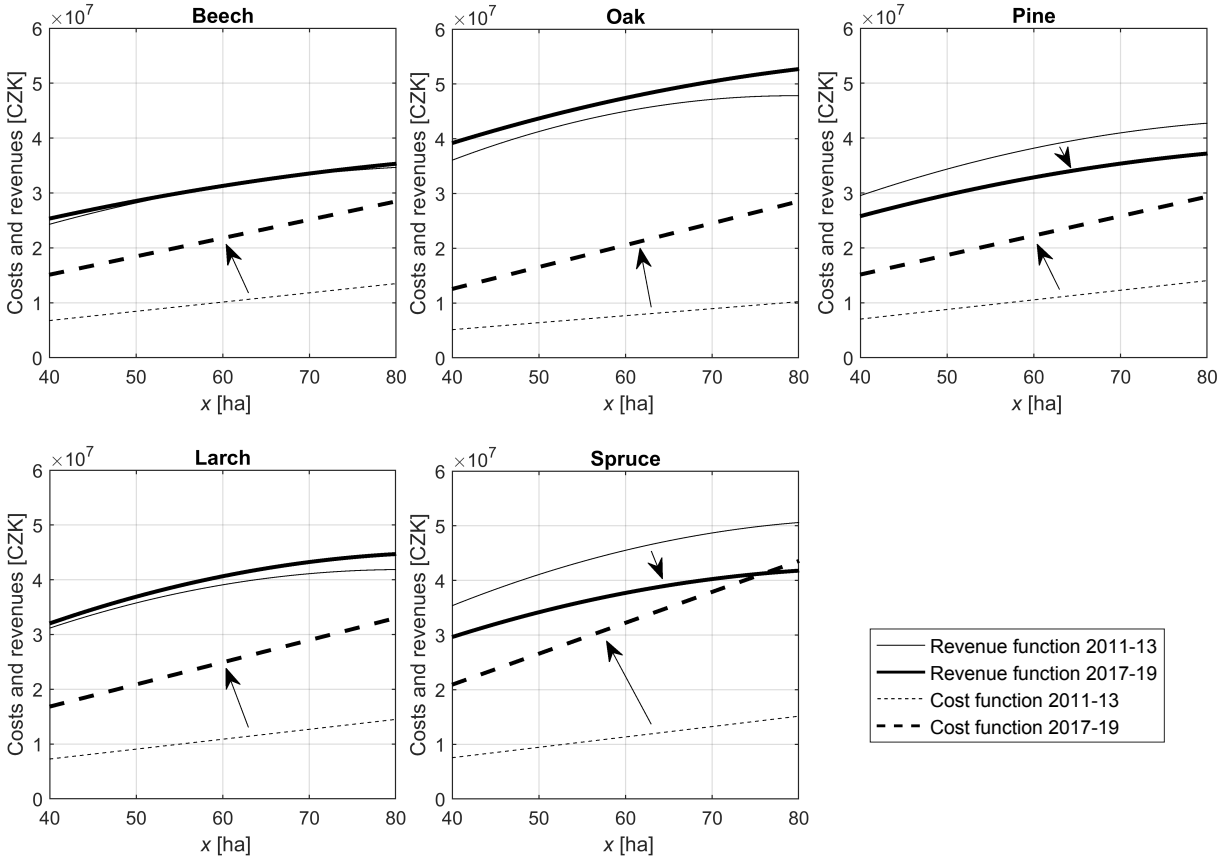


Figure 3 Estimated revenue and cost functions for particular species based on data from the years 2011–2013 (thin curves) and 2017–2019 (bold curves)

2000 to 2015. Briefly, it was caused by strong demand after spruce timber and by the closure of production in the hardwood processing industry in the Czech Republic. Oak was understood as a premium wood, where mainly valuable trees are chosen for felling; but beech, which has wood with similar properties, was used more for heating. Actually, the situation has changed dramatically, see Figure 4 (right graph). Oak exhibits the highest profitability again, but profitability of spruce is the worst. Larch seems to be more profitable than pine and beech. Note that the drop of profitability in the range of 70–80 ha should be for all species analogous to spruce (it is given by insufficient cost function expression, see above), but only for spruce an economical loss is reached. This can be supported by the fact, that many spruce forest owners wish to sell or even give the forest for free because of currently holding of forest is not profitable (the owners are forced to remove the bark beetle-infested timber in a short time).

4 Conclusions

We deal with the simulation of reforestation system presented in [7], which has lead to estimate the revenue and the cost functions and has been based on data from the years 2011–2013. We employ the simulation procedure with actual data from the period 2017–2019, reestimate the revenue and the cost functions and calculate the actual profit functions. The simulation seems to be robust although it does not capture certain phenomena such as drought and bark beetle attack. Relations among estimated functions correspond to the real situation from biological as well as economical aspects. There is no doubt that current situation is not stable. Intensive reforestation must be carried out in a smart way: it is necessary to take measures against drought (including local ones as planting birch as a tree to protect subsequently planted seedlings) and to plant broadleaves species. It is not desirable to completely avoid the cultivation of spruce, as spruce timber is very useful in many fields of human activity. However, it is necessary to carefully consider which localities are suitable for planting spruce and whether it is not more advantageous to grow spruce in a mixture with other tree species. These aspects we would like to consider in our future work.

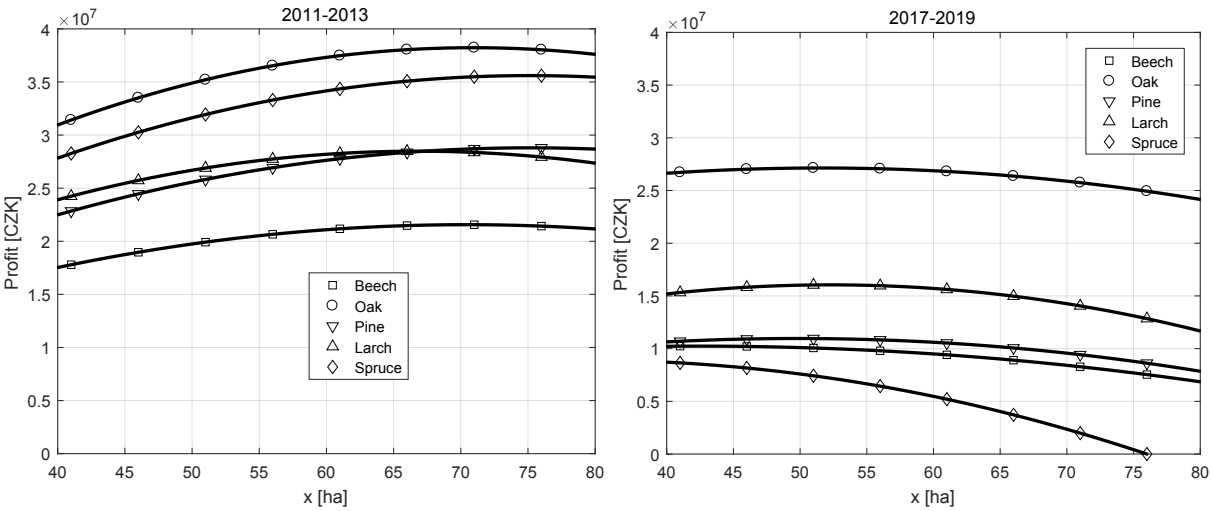


Figure 4 Estimated profit functions for particular species based on data from the years 2011–2013 (left graph) and 2017–2019 (right graph)

Acknowledgements

This paper presents the results of the research supported by Czech Science Foundation grant GA18-08078S.

References

[1] Bach, C. F. (1999). Economic incentives for sustainable management: a small optimal control model for tropical forestry, *Ecological Economics*, 30(2), 251–265.

[2] Beneš, O., Blašková, V. & Střelec, L. (2015). Verifying Timber Price Stability for the Reforestation Optimal Control Model. In *Proceedings of the International Conference of Numerical Analysis and Applied Mathematics 2017*. Melville: American Institute of Physics (AIP).

[3] Caparros, A., Cedra, E., Ovando, P. & Campos, P. (2013). Carbon sequestration with reforestations and biodiversity-scenic Values, *Environmental Resource Economics*, 45(1), 49–72.

[4] Cerda, E. & Martin-Barroso, D. (2013). Optimal Control for Forest Management and Conservation Analysis in Dehesa Ecosystems, *European Journal of Operational Research*, 227(3), 515–526

[5] Hampel, D. & Janová, J. (2014). Simulation of data for reforestation system. In *Conference Proceedings of 32nd International Conference Mathematical Methods in Economics* (pp. 251–256). Olomouc: Univerzita Palackého v Olomouci.

[6] Hampel, D., Janová, J. & Kadlec, J. (2015). Estimation of Cost and Revenue Functions for Reforestation System in Drahanska Highlands. In *Proceedings of the International Conference on Numerical Analysis and Applied Mathematics 2014*. Melville: American Institute of Physics (AIP).

[7] Janová, J. & Hampel, D. (2016). Optimal managing of forest structure using data simulated optimal control. *Central European Journal of Operations Research*, 24(2), 297–307.

[8] Janová, J., Hampel, D. & Nerudová, D. (2019). Design and validation of a tax sustainability index. *European Journal of Operational Research*, 278(3), 916–926.

[9] Knoke, T. (2008). Mixed forests and finance – Methodological approaches, *Ecological Economics*, 65(3), 590–601.

[10] Lohmander, P. (1992). The multi species forest stand, stochastic price process and adaptive selective thinning, *Systems Analysis – Modeling – Simulation*, 9, 229–250.

[11] Valkonen, S. & Valsta, L. (2001). Productivity and economics of mixed two-storied spruce and birch stands in Southern Finland simulated with empirical models, *Forest Ecology and Management*, 140(2–3), 133–149.

Estimation of Long-Term Projects in Terms of Uncertainty of Inputs

Simona Hašková[1]

Abstract. The state of "uncertainty" is regarded as a situation where only a "sufficiently trusted" set or interval of occurrence of possible numerical values of considered "known unknown" is known and no relevant reason exists to prefer one value over another. In such a situation a manager often appears. The tool used here is a three-stage fuzzy system of multicriteria evaluation of alternatives, the inputs of which can be various uncertain data of economic, ecological, aesthetic or other nature, and the output of which is the subjectively expected value of the social benefit index of the assessed alternative. The process of transformation of uncertain inputs to output is an analogy of the process taking place when solving similar problems in the human mind. This is in the theoretical part exactly formulated by the mathematical apparatus of interval algebra, fuzzy logic and fuzzy set theory, which is utilized within the fuzzy approach to multicriteria evaluation. In the practical part, the fuzzy system is applied to the problem of selecting the optimal type of production equipment.

Keywords: Fuzzy logic, uncertainty, uncertain inputs, interval algebra.

JEL Classification: C51, C58
AMS Classification: 90B50, 90B90

1 Introduction

Standard mathematical approaches to multicriteria evaluation and managerial decision-making that is based on it, more or less ignore the specifics of the role that the human factor plays in evaluation and in decision-making activities in terms of individual knowledge, skills and experience of a manager [3].

In contrast, alternative approach called the fuzzy approach, implemented in the form of a three-stage fuzzy system of multicriteria evaluation in rough outlines in [4], provides space for the application of the human factor. It is based on the thesis that the key elements of human thinking are not numbers, but intuitive concepts, the contents of which can be modelled by fuzzy sets [6].

Fuzzy logic systems, figuratively speaking, concentrate on the human brain "software" emulating fuzzy and symbolic reasoning. Neural networks are another system for imitating the operation of human brain while concentrating on the structure of human brain, i.e., on the "hardware" emulating its basic functions [10]. The core of this paper is to show the processing of "human factor" within the computation based on the three-stage fuzzy system in case of ex ante evaluation.

2 Elements of fuzzy set theory

Let the set U be a field of reasoning or discussion (universe). Let $\mu_{\underline{A}}: U \to \langle 0,1 \rangle$ be a membership function and let $\underline{A} = \{(x, \mu_{\underline{A}}(x)): y \in U\}$ be the set of all pairs $(x, \mu_{\underline{A}}(x))$, in which the numbers $0 \le \mu_{\underline{A}}(x) \le 1$ state to the given $x \in U$ the degree of membership to the set $\underline{A}$. Then $\underline{A}$ is a *fuzzy subset* on the universe U. The significant characteristic of the fuzzy subset $\underline{A}$ is its support $U_{\underline{A}} = \{x: 0 < \mu_{\underline{A}}(x) \le 1, x \in U\} \subset U$. In terms of fuzzy logic $\mu_{\underline{A}}(x) = |x \in U_{\underline{A}}|$, where $|x \in U_{\underline{A}}|$, denotes the degree of veracity of the proposition that x is the element of the support of the fuzzy set $\underline{A}$. The element $x \in U$ with $\mu_{\underline{A}}(x) = 0.5$ is called the *crossover point* in $\underline{A}$. At values greater than 0.5, the element x rather belongs to $U_{\underline{A}}$, at the values smaller it rather does not.

The fuzzy subset contains pairs $(x, \mu(x))$ with all x from the universe of discussion. Therefore, it is more general than the fuzzy set used to formalize intuitive concepts containing only $(x, \mu(x))$ with $\mu(x) > 0$. The fuzzy subset $\underline{A}$ (whose support $U_{\underline{A}} \subset U \subset \mathcal{R}$, where $\mathcal{R}$ is the set of real numbers, and its function $\mu_{\underline{A}}$ is gifted

[1]The Institute of Technology and Business/The School of Expertness and Valuation, Okružní 517/10, České Budějovice, Czech Republic, e-mail: haskovas@post.cz.

by the property of normality and convexity, i.e., at least, in the case of one element $x \in U_{\underline{A}}$ it applies $\mu_{\underline{A}}(x) = 1$, and $\mu_{\underline{A}}(x') \geq \min\{\mu_{\underline{A}}(x_1), \mu_{\underline{A}}(x_2)\}$ for all $x' \in \langle x_1, x_2 \rangle \subset U_{\underline{A}}$) is called the fuzzy number. Theoretically, there may be different shapes of membership function $\mu_{\underline{A}}$ of fuzzy numbers: triangular, trapezoidal, bell, sinusoidal, cosinusoidal (see e.g. [5], [12]). Fuzzy numbers are also formal models of linguistic terms (i.e., expressed in natural language) of variables occurring in managerial decision-making tasks [2]. Linguistic variables thus acquire their values at two levels: at the linguistic level and at the numerical level. At the linguistic level, these are usually terms (fuzzy numbers) of the low value ($\underline{L}$), common value ($\underline{M}$) and high value ($\underline{H}$) type; at the numerical level they perform the real numbers from the interval $U = \langle 0, 100 \rangle$.

The relationship between the two mentioned levels of values of the linguistic variable is evident from the following fuzzification table defining the projection of μ: $\{\underline{L}, \underline{M}, \underline{H}\} \times U \rightarrow \langle 0,1 \rangle$ in the form $\mu(\underline{T}, u) = \mu_{\underline{T}}(u)$, in which constants $a, b, c, d \in \langle 0, 100 \rangle$, $0 \leq a \leq b \leq c \leq d \leq 100$ are given by an expert:

Interval	$u < a$	$a \leq u < b$	$b \leq u < c$	$c \leq u < d$	$u \geq d$
$\underline{L}$	1	$(b - u) / (b - a)$	0	0	0
$\underline{M}$	0	$(u - a) / (b - a)$	1	$(d - u) / (d - c)$	0
$\underline{H}$	0	0	0	$(u - c) / (d - c)$	1

Table 1 Fuzzification table with $a, b, c, d \in \langle 0, 100 \rangle$, $0 \leq a \leq b \leq c \leq d \leq 100$

The inner nonzero fields of the fuzzification table define for each $u \in U$ an one-element or two-element subset $h(u) = \{\underline{T}: \underline{T} \in \{\underline{L}, \underline{M}, \underline{H}\}, \mu(\underline{T}, u) > 0\}$ of the set $\{\underline{L}, \underline{M}, \underline{H}\}$.

3 Three-stage fuzzy system of multicriteria evaluation of alternatives

The basic functions of all elements of the fuzzy system structure diagram in Fig. 1 have already been described verbally and in exact form in detail in [4], where a similar task was solved in the case of ex post evaluation, i.e., under conditions of certainty on the inputs side. Herein, Fig. 1 captures ex ante evaluation within the fuzzy system in the form of the analogy to the fuzzy system formulated in [4].

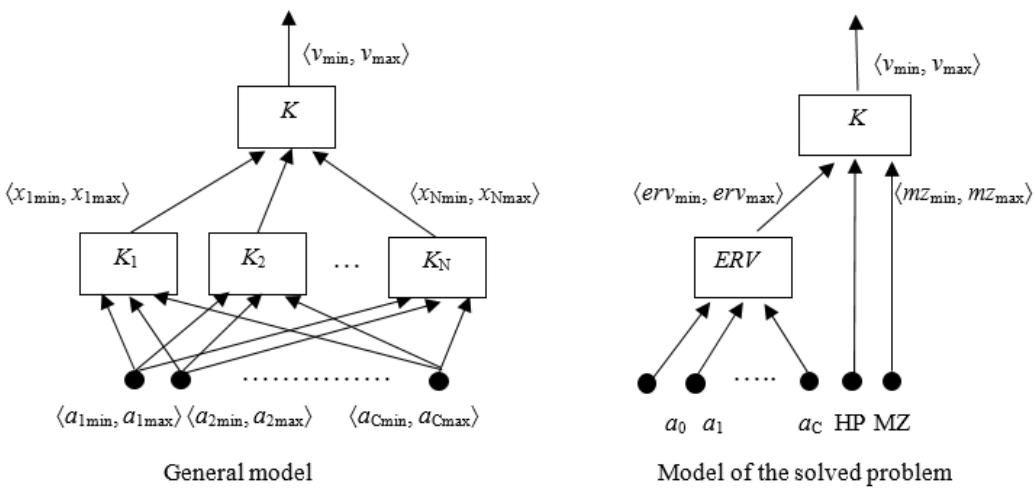


Figure 1 Structure of a three-stage fuzzy system of multicriteria evaluation of alternatives

In the case of ex ante evaluation, the fuzzy system appears in conditions of uncertainty of input data. In its first stage, only uncertain data characterized by intervals of values of their possible occurrences (see intervals $\langle a_{min}, a_{max} \rangle$ to $\langle mz_{min}, mz_{max} \rangle$ in the models in Fig. 1) or mixes of uncertain and numerical data enter the problem solving process. This external uncertainty is then transferred by the criterion functions to the internal uncertainty of the inputs to block K (see intervals $\langle x_{min}, x_{max} \rangle$ to $\langle erv_{min}, erv_{max} \rangle$ at the second stage of the models), as well as to its output (see interval $\langle v_{min}, v_{max} \rangle$). The right part of Fig. 1 shows a case where some external uncertain data enter block K directly. It is usually a subjective quantification of qualitative data by means of intervals on the value scales from 0 to 10 or from 0 to 100.

The first step of the solution is the conversion of external uncertainty to internal uncertainty. For this purpose, the relations defining the individual criterion functions are translated into the language of interval algebra, defined and discussed in detail in [1] and [11], and by applying its operations, the input intervals are converted to output intervals. In the second stage of the general model, N intervals are created, which can be represented by at most two-element sets of their extreme values (in the case of a mix of uncertain and numerical data, the numerical value is considered as an interval with two identical extreme values). Their Cartesian product can result in up to 2^N N-dimensional numerical vectors entering block *K*. By gradual processing of each of them by the computational fuzzy algorithm, whose block diagram is in Fig. 2, a set of partial calculation results (numbers *v*) is created. The arithmetic mean $y = (v_{min} + v_{max}) / 2$ of the minimum and maximum of this set is then the subjectively expected numerical value of the output linguistic variable ISU (social benefit index) of the evaluated alternative.

4 Computational fuzzy algorithm

The procedures by which the fuzzy system (block *K*) processes its numerical inputs are analogous to the procedures by which the human mind processes visual, auditory, tactile and other stimuli and generates corresponding responses. The N-dimensional vector $(x_1,\ldots,x_N)$, entering the block *K*, is transformed into a vector $(u_1,\ldots,u_N)$ by converting the coordinates x_i into a scale in the range 0 to 100. Its coordinates u_i, i = 1,…,N are included in the adequate input fuzzy sets, together they select suitable inference rules for the manipulation with them, define their "strength" and generate the membership function μ_{agg} on the field $V = \langle 0, 100\rangle$ of the output linguistic variable ISU. The horizontal coordinate of the centre of gravity of the surface below its course $\mu_{agg}(v)$ is the result of the calculation.

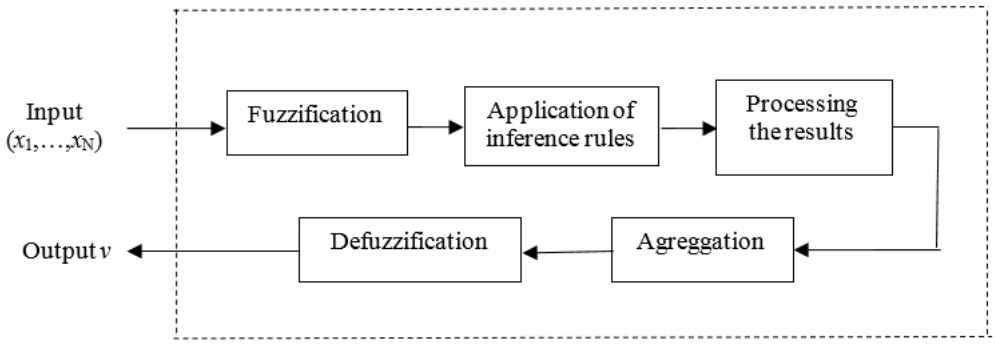


Figure 2 Phase continuity diagram of the computational algorithm of the fuzzy approach

In the fuzzification phase, the expected influence of the x_i coordinate of the vector $(x_1,\ldots,x_N)$ on the output *v* is first taken into account. If positive, x_i is converted to the *internal* value u_i used in Tab. 1 in section 2 according to the formula $u_i = 100\cdot(x_i - Xi_{min}) / x_{iref}$, $x_{iref} = Xi_{max} - Xi_{min}$, where $\langle Xi_{max}, Xi_{min}\rangle$ interval is a domain of numerical values of a linguistic variable on the i-*th* input. If the effect is negative, the formula $u_i = 100 - 100\cdot(x_i - Xi_{min}) / x_{iref}$ is used for the conversion. If the positive effects prevail in the given task, the resulting social benefit has the character of a contribution, if the negative effects prevail, it has the character of a harm (cost). Each i-*th* input, i = 1 to N, has its own fuzzification table, which generates its set $h(u_i) = \{\underline{T}_i: \underline{T}_i \in \{\underline{L}_i, \underline{M}_i, \underline{H}_i\}, \mu(\underline{T}_i, u_i) > 0$. The Cartesian product $\mathcal{H} = h(u_1)\times\ldots\times h(u_N) = \{(\underline{T}_1,\ldots,\underline{T}_N): \underline{T}_1\in h(u_1),\ldots,\underline{T}_N\in h(u_N)\}$ with 2^α elements is formed from them, where α, $0 \le \alpha \le N$ is the number of two-element sets $h(u_i)$ in the Cartesian product $\mathcal{H}$.

The interference rule is the element of projection p: $\{\underline{L}_1, \underline{M}_1, \underline{H}_1\}\times\ldots\times\{\underline{L}_N, \underline{M}_N, \underline{H}_N\} \to \{\underline{L}, \underline{M}, \underline{H}\}$, where $\underline{L}$, $\underline{M}$ and $\underline{H}$ are the terms of the output linguistic variable ISU (social benefit index) with a domain of numerical values $V = \langle 0, 100\rangle$. The set of inference rules thus consists of a total 3^N pairs of the type $((\underline{T}_1,\ldots,\underline{T}_N), T)$ given by a knowledgeable expert. In the application phase of inference rules, three classes are created

$$\mathcal{H}(\underline{T}) = \{(\underline{T}_1,\ldots,\underline{T}_N): (\underline{T}_1,\ldots,\underline{T}_N) \in \mathcal{H} \cap p^{-1}(\underline{T})\},\ \underline{T} = \underline{L}, \underline{M}, \underline{H},$$

of decomposition of the set $\mathcal{H}$ according to the terms of the output linguistic variable ISU, where the relation p^{-1} is the inversion of projection *p*. Each resulting class $\mathcal{H}(\underline{T})$ is then assigned by its characteristic number $\mathcal{M}_{\underline{T}} \in \langle 0, 1\rangle$:

- if the class is empty ($\mathcal{H}(\underline{T}) = \emptyset$), then $\mathcal{M}_{\underline{T}} = 0$;

- if $\mathcal{H}(\underline{T}) \neq \emptyset$, then $\mathcal{M}_{\underline{T}} = \max\{\{\min\{\mu(\underline{T}_1, u_1),\ldots, \mu(\underline{T}_N, u_N)\}: (\underline{T}_1,\ldots,\underline{T}_N) \in \mathcal{H}(\underline{T})\}\}$.

In the result processing phase, the fuzzy algorithm by means of the numbers $\mathcal{M}_{\underline{L}}$, $\mathcal{M}_{\underline{M}}$ and $\mathcal{M}_{\underline{H}}$ restricts (cuts off) the course of the functions $\mu_{\underline{L}}(v)$, $\mu_{\underline{M}}(v)$ and $\mu_{\underline{H}}(v)$ of the terms $\underline{L}$, $\underline{M}$ and $\underline{H}$ of the output linguistic variable ISU on the domain of its numerical values $V = \langle 0, 100\rangle$, see Fig. 3 left. Then, in the aggregation phase, the fuzzy algorithm fuzzy logically adds their torso, thus aggregating them into the resulting function $\mu_{agg}(v)$, which is their wrapper, see Fig. 3 on the right.

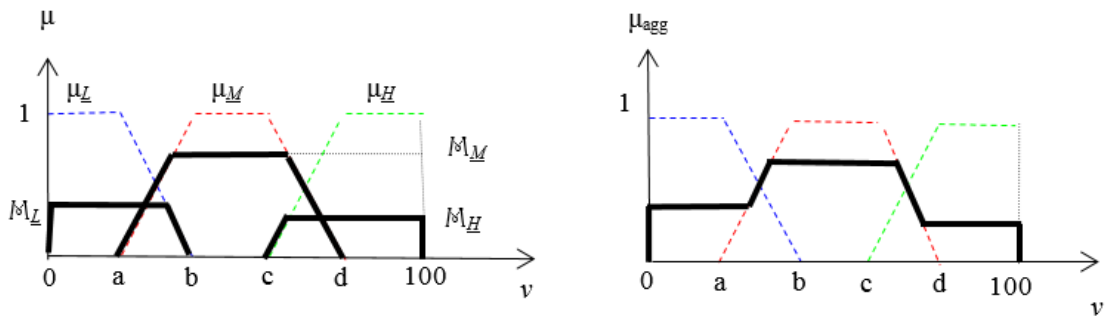


Figure 3 Example of cutting off the functions $\mu_{\underline{L}}(v)$, $\mu_{\underline{M}}(v)$, $\mu_{\underline{H}}(v)$ and their aggregation into the resulting $\mu_{agg}(v)$

5 Example of application of fuzzy algorithm in ex ante managerial evaluation

The manager responsible for investing in the company's production facilities is considering investing in one of two equally powerful machines, Machine X and Machine Y, necessary to implement a long-term production plan of the given product (in detail in [7] and [8]). The revenues from the expected annual sales volumes of the product do not depend on which machine is going to be used in the production [9], therefore, the profitability of machines should be assessed according to the following three criteria: annual expenditure equivalent (ERV), traffic noise (HP) and environmental pollution (MZ). The ERV criterion is an economic criterion, the HP and MZ criteria are ecological criteria. In all cases they perform social costs for which the lowest possible value is desirable. Only the purchase prices of machines and their economic life are exactly known in advance. Other relevant data entering the task are uncertain and only estimates of the intervals in which their values may be located are available – see the following section.

5.1 Specification of the task parameters

Everything needed for making decision, which machine to prefer, is listed in the following table:

Year	0	1	2	3	4	5	6	HP	MZ
Machine X	31,0	⟨7, 9⟩	⟨11, 14⟩	⟨16, 20⟩	⟨24, 30⟩			⟨20, 40⟩	⟨50, 70⟩
Machine Y	42,5	⟨5, 6⟩	⟨8, 10⟩	⟨12, 15⟩	⟨16, 20⟩	⟨20, 25⟩	⟨24, 30⟩	⟨60, 70⟩	⟨60, 80⟩

Table 2 Intervals of possible values of annual expenses of machines X and Y, intervals of HP and MZ estimates (the noise and pollution zone)

The zero column states the acquisition costs of both machines in thousands of EUR. Machine X has a four-year and machine Y a six-year economic life. In the next four, respectively, six columns, the intervals of possible values of annual expenses (also in thousands of EUR) associated with the operation of machine X, respectively Y are placed. These intervals, including the numeric data in the zero column, are used for calculating the interval $\langle erv_{min}, erv_{max}\rangle$ of the ERV criterion values. In Fig. 1 on the right, they are indicated as the symbols a_0 to a_C (for machine X $a_C = a_4$, for machine Y $a_C = a_6$). For calculating the values of *erv*, of which the upper acceptability is limited by 50 thousand EUR, it is necessary to know the considered value of the cost of capital of the project (an annual discount rate of the expenditure flow). This is assumed to be 10 %.

The values of intervals in the columns HP and MZ have the completely different character. These are subjective estimates of the noise or pollution zone, projected by a knowledgeable expert to a scale from 0 to 100.

5.2 Solution

From the table of entered parameters, it is obvious that X machine is better in terms of ecological criteria. If it was better according to the economic criteria, there would be nothing to deal with any more. The following applies for machine X:

- $erv_{\min} = (31 + 7 / 1{,}1 + 11 / 1{,}1^2 + 16 / 1{,}1^3 + 24 / 1{,}1^4) / (1 / 1{,}1 + 1 / 1{,}1^2 + 1 / 1{,}1^3 + 1 / 1{,}1^4) = 23{,}61$,
- $erv_{\max} = (31 + 9 / 1{,}1 + 14 / 1{,}1^2 + 20 / 1{,}1^3 + 30 / 1{,}1^4) / (1 / 1{,}1 + 1 / 1{,}1^2 + 1 / 1{,}1^3 + 1 / 1{,}1^4) = 27{,}14$.

For machine Y it applies:

- $erv_{\min} = (42{,}5 + 5 / 1{,}1 + 8 / 1{,}1^2 + 12 / 1{,}1^3 + 16 / 1{,}1^4 + 20 / 1{,}1^5 + 24 / 1{,}1^6) / (1 / 1{,}1 + 1 / 1{,}1^2 + 1 / 1{,}1^3 + 1 / 1{,}1^4 + 1 / 1{,}1^5 + 1 / 1{,}1^6) = 22{,}86$,
- $erv_{\max} = (42{,}5 + 6 / 1{,}1 + 10 / 1{,}1^2 + 15 / 1{,}1^3 + 20 / 1{,}1^4 + 25 / 1{,}1^5 + 30 / 1{,}1^6) / (1 / 1{,}1 + 1 / 1{,}1^2 + 1 / 1{,}1^3 + 1 / 1{,}1^4 + 1 / 1{,}1^5 + 1 / 1{,}1^6) = 26{,}09$.

As can be seen, according to the economic criterion machine Y is better. The problem therefore has to be solved in block *K* of the diagram in Fig. 1 on the right.

It includes numerical vectors (erv, hp, mz), $erv \in \langle 0, 50\rangle$, $hp \in \langle 0, 100\rangle$, $mz \in \langle 0, 100\rangle$, of the triple of linguistic variables ERV, HP and MZ with terms $\underline{L}_i$, $\underline{M}_i$ a $\underline{H}_i$, i = erv, hp, mz. For simplicity, let us suppose that their fuzzification tables are symmetric and except for the identical indices (i.e., for all i = erv, hp, mz for which it applies: $a_i = 20$, $b_i = 40$, $c_i = 60$, $d_i = 80$). From the block *K* the number $v \in \langle 0, 100\rangle$ exits of the linguistic variable ISU, having in this case the character of social cost, with the terms $\underline{L}$, $\underline{M}$ and $\underline{H}$. The set of inference rules of the type $((\underline{T}_{erv}, \underline{T}_{hp}, \underline{T}_{mz}), \underline{T})$, $\underline{T}_i \in \{\underline{L}_i, \underline{M}_i, \underline{H}_i\}$, $\underline{T} \in \{\underline{L}, \underline{M}, \underline{H}\}$ has 27 elements formed by the strategies of the predominant element. It assigns to the given left side the very term $\underline{T}$, which predominates on the left side. If there is no such term, it assigns the term $\underline{M}$.

For the recalculated vector (u_1, u_2, u_3) the following applies for both machines: $u_1 = 100 \cdot erv / 50 = 2 \cdot erv$, $u_2 = hp$ and $u_3 = mz$. In the general case, to obtain the outputs $v_{\min}$ and $v_{\max}$, it would be necessary for both machines X and Y to calculate the values of v to the eight recalculated vectors of all combinations of the limits of intervals of uncertain values. It is easy to see that in this case we can only use a combination $(u_{1\min}, u_{2\min}, u_{3\min})$ for $v_{\min}$ and $(u_{1\max}, u_{2\max}, u_{3\max})$ for $v_{\max}$.

Therefore for machine X it applies: $h(u_{1\min}) = h(47{,}22) = \{\underline{M}_{erv}\}$, $h(u_{2\min}) = h(20) = \{\underline{L}_{hp}\}$, $h(u_{3\min}) = h(50) = \{\underline{M}_{mz}\}$, $\mathcal{H} = \{(\underline{M}_{erv}, \underline{L}_{hp}, \underline{M}_{mz})\}$, $\mathcal{M}_{\underline{L}} = \mathcal{M}_{\underline{H}} = 0$, $\mathcal{M}_{\underline{M}} = \max\{\{\min\{1, 1, 1\}\} = 1$, $\mu = \mu_{VI}$, $\boldsymbol{v_{\min} = 2000 / 40 = 50}$.

$h(u_{1\max}) = h(54{,}28) = \{\underline{M}_{erv}\}$, $h(u_{2\max}) = h(40) = \{\underline{M}_{hp}\}$, $h(u_{3\max}) = h(70) = \{\underline{M}_{mz}, \underline{H}_{mz}\}$, $\mathcal{H} = \{(\underline{M}_{erv}, \underline{M}_{hp}, \underline{M}_{mz}), (\underline{M}_{erv}, \underline{M}_{hp}, \underline{H}_{mz})\}$, $\mathcal{M}_{\underline{L}} = \mathcal{M}_{\underline{H}} = 0$, $\mathcal{M}_{\underline{M}} = \max\{\{\min\{1, 1, 0{,}5\}, \{\min\{1, 1, 0{,}5\}\} = \max\{0{,}5, 0{,}5\} = 0{,}5$, $\mu = \mu_{VI}$, $\boldsymbol{v_{\max} = 1250 / 25 = 50}$.

$y_x = (50 + 50) / 2 = 50$.

For machine Y it applies: $h(u_{1\min}) = h(45{,}72) = \{\underline{M}_{erv}\}$, $h(u_{2\min}) = h(60) = \{\underline{M}_{hp}\}$, $h(u_{3\min}) = h(60) = \{\underline{M}_{mz}\}$, $\mathcal{H} = \{(\underline{M}_{erv}, \underline{M}_{hp}, \underline{M}_{mz})\}$, $\mathcal{M}_{\underline{L}} = \mathcal{M}_{\underline{H}} = 0$, $\mathcal{M}_{\underline{M}} = \max\{\{\min\{1, 1, 1\}\} = 1$, $\mu = \mu_{VI}$, $\boldsymbol{v_{\min} = 2000 / 40 = 50}$. $h(u_{1\max}) = h(52{,}18) = \{\underline{M}_{erv}\}$, $h(u_{2\max}) = h(70) = \{\underline{M}_{hp}, \underline{H}_{hp}\}$, $h(u_{3\max}) = h(80) = \{\underline{H}_{mz}\}$, $\mathcal{H} = \{(\underline{M}_{erv}, \underline{M}_{hp}, \underline{H}_{mz}), (\underline{M}_{erv}, \underline{H}_{hp}, \underline{H}_{mz})\}$, $\mathcal{M}_{\underline{L}} = 0$, $\mathcal{M}_{\underline{M}} = \max\{\{\min\{1, 0{,}5, 1\}\} = \max\{0{,}5\} = 0{,}5$, $\mathcal{M}_{\underline{H}} = \max\{\min\{1, 0{,}5, 1\}\} = \max\{0{,}5\} = 0{,}5$, $\mu = \mu_{VI}$, $\boldsymbol{v_{\max} = 2341{,}66 / 37{,}5 = 62{,}44}$.

$y_y = (50 + 62{,}44) / 2 = 56{,}22$.

Since $y_x < y_y$, machine X is better.

6 Summary and conclusion

The topic of the paper is the issue of multicriteria evaluation of alternatives in terms of uncertainty of input data. Data uncertainty is generally understood as a situation where only a sufficiently trusted set or interval

of occurrence of possible numerical values of considered "known unknown" is known and there is no relevant reason to prefer one value to another. A three-stage fuzzy system is presented, which sophisticatedly projects the results of partial criteria into a range of values of the social benefit index, by which it comprehensively evaluates the given alternatives.

At its first stage, each uncertain data entering the partial criteria is entered by the extreme points of the interval of its values. For them, the partial criteria, expressed in the language of interval algebra, generate on the second stage of the fuzzy system the sets of their corresponding evaluation results. The elements of the Cartesian product of these sets are then gradually processed in the third stage by the computational fuzzy algorithm. The procedure of processing each element gradually goes through a five-phase. The arithmetic mean of the maximum and minimum of the algorithm results above all elements of the mentioned Cartesian product is *subjectively expected value* of the social benefit index of the assessed alternative. The subjectively expected value is derived from the basic parameters of the fuzzy system (the shape of μ functions of relevant terms and a set of inference rules), which are generally given by a knowledgeable expert based on their knowledge, skills and experience, thus influencing the function of the fuzzy system by what is called a "human factor".

In the application part, the course of calculations is demonstrated on the solution of the example of choosing a more advantageous investment in one of two equally powerful machines with different lengths of economic life.

References

[1] BĚHOUNEK, L. & CINTULA, P. (2006). From fuzzy logic to fuzzy mathematics: A methodological manifesto. *Fuzzy Sets and Systems*, ***157***(5), 642–646.

[2] DUBOIS, D. & PRADE, H. (1996). What are fuzzy rules and how to use them. *Fuzzy sets and systems*, 1996, ***84***(2), 169–185.

[3] FIALA, P. & BOROVIČKA, A. (2015). Investment decision-making by a two-step multi-criteria procedure. *International Journal of Multicriteria Decision Making*, ***5***(3), 274–293.

[4] HAŠKOVÁ, S. (2018). A Contribution to the Application of the Fuzzy Approach to the Economic Analyses. In VACHOVA & KRATOCHVIL. *36th International Conference Mathematical Methods in Economics* (*MME*). 1. vyd. Jindřichův Hradec, ČR: Univerzity of Economics, Prague, 127–132.

[5] HAŠKOVÁ, S. & FIALA, P. (2019). A fuzzy approach for the estimation of foreign investment risk based on values of rating indices. *Risk Management*, ***21***(3), 1–17.

[6] HERRERA, F. & HERRERA-VIEDMA, E. (2000). Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and systems*, ***115***(1), 67–82.

[7] KARSAK, E. E. & TOLGA, E. (2001). Fuzzy multi-criteria decision-making procedure for evaluating advanced manufacturing system investments. *International journal of production economics*, ***69***(1), 49–64.

[8] KENLEY, R. & WILSON, O. D. (1989). A construction project net cash flow model. *Construction Management and Economics*, ***7***(1), 3–18.

[9] MEREDITH, J. R. & SURESH, N. C. (1986). Justification techniques for advanced manufacturing technologies. *International Journal of Production Research*, ***24***(5), 1043–1057.

[10] WU, D., LIN, C. T., HUANG, J., & ZENG, Z. (2019). On the functional equivalence of TSK fuzzy systems to neural networks, mixture of experts, CART, and stacking ensemble regression. *IEEE Transactions on Fuzzy Systems*.

[11] ZADEH, L. A. (1965). Fuzzy sets. *Information and control*, ***8***(3), 338–353.

[12] ZADEH, L. A. (1975). Fuzzy logic and approximate reasoning. *Synthese*, ***30***(3–4), 407–428.

Redundancy in Interval Linear Systems

Milan Hladík[1]

Abstract. In a system of linear equations and inequalities, one constraint is redundant if it can be dropped from the system without affecting the solution set. Redundancy can be effectively checked by linear programming. However, if the coefficients are uncertain, the problem becomes more cumbersome. In this paper, we assume that the coefficients come from some given compact intervals and no other information is given. We discuss two concepts of redundancy in this interval case, the weak and the strong redundancy. This former refers to redundancy for at least one realization of interval coefficients, while the latter means redundancy for every realization. We characterize both kinds of redundancies for various types of linear systems; in some cases the problem is polynomial, but certain cases are computationally intractable. As an open problem, we leave weak redundancy of equations. Herein, a characterization is known only for certain special cases, but for a general case a complete characterization is still unknown.

Keywords: interval analysis, interval system, redundancy, linear programming

JEL Classification: C44, C61
AMS Classification: 15A39, 65G40, 90C05

1 Introduction

Detecting and removing redundant constraints is a fundamental problem in optimization and important for variable elimination, convex hull algorithms, verification of hybrid systems and many others [1, 9]. A constraint is *redundant* in a system of constraints if it can be omitted and the solution set remains the same. For a real system of linear equations or inequalities, redundancy can be easily checked by means of linear programming (for efficient algorithms see [4, 16]), but computationally it is not easier.

If the coefficients are inexact, redundancy is more difficult and, indeed, not well defined. For the sake of this paper, we assume that coefficients of the system lie in specified compact intervals; no other information is given, so the intervals are considered as deterministic. We discuss two concepts of redundancy in this context. Notice that redundancy of interval linear inequalities was pioneered by Lodwick [11]. This paper extends his results and presents an overview on redundancy of different types of interval systems.

Interval notation. By an interval matrix we understand the set of matrices

$$\boldsymbol{A} = \{A \in \mathbb{R}^{m\times n};\ \underline{A} \le A \le \overline{A}\},$$

where $\underline{A}, \overline{A} \in \mathbb{R}^{m\times n}$ are given matrices and the inequality is meant entrywise. Interval vectors are defined analogously. The midpoint A_c and the radius A_Δ of $\boldsymbol{A}$ respectively read

$$A_c = \frac{1}{2}(\underline{A} + \overline{A}), \quad A_\Delta = \frac{1}{2}(\overline{A} - \underline{A}).$$

Given $u \in [-1,1]^m$ and $v \in [-1,1]^n$, we define a matrix $A_{u,v} = A_c - \operatorname{diag}(u) A_\Delta \operatorname{diag}(v) \in \boldsymbol{A}$, where $\operatorname{diag}(u)$ stands for the diagonal matrix with entries $u_1, \dots, u_m$.

Problem formulation. By an interval linear system in a general form we understand the family of linear systems

$$Ax + By = a,\ Cx + Dy \le b,\ x \ge 0, \quad (A, B, C, D, a, b) \in (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{a}, \boldsymbol{b}),$$

where $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ are interval matrices and $\boldsymbol{a}, \boldsymbol{b}$ interval vectors of appropriate size. Variables x are nonnegative, whereas variables y are free. The interval system is usually written in short as

$$\boldsymbol{A}x + \boldsymbol{B}y = \boldsymbol{a},\ \boldsymbol{C}x + \boldsymbol{D}y \le \boldsymbol{b},\ x \ge 0.$$

[1] Charles University, Department of Applied Mathematics, Malostranské nám. 25, 118 00, Prague, Czech Republic, hladik@kam.mff.cuni.cz

A *weak solution* is a solution of at least one realization, whereas a *strong solution* is a solution that fulfills any realization. In a similar way we introduce two types of redundancy.

Definition 1. Let an interval system of constraints be given. A selected constraint is *strongly redundant* if it is redundant for every realization of interval coefficients. A selected constraint is *weakly redundant* if it is redundant for at least one realization.

Strong redundancy means that omitting the constraint makes no harm since it is redundant whatever is the true value of the coefficients. In contrast, if a constraint is not weakly redundant, then it is not redundant regardless what the true coefficients are.

In the following section we characterize weak and strong redundancy of a general interval system as well as special types of interval systems that appear in various problems.

2 Redundancy of inequalities

Let $\mathcal{M}(A,b)$ be a convex polyhedron with constraint matrix A and the right-hand side vector b, and let $c^T x \le d$ an inequality in question. Then $c^T x \le d$ is redundant for $\mathcal{M}(A,b)$ if and only if

$$d \ge \max_{x \in \mathcal{M}(A,b)} c^T x. \tag{1}$$

Therefore, redundancy for real linear systems is effectively decidable by means of linear programming.

Let interval data $\boldsymbol{A} \in \mathbb{IR}^{m\times n}$, $\boldsymbol{b} \in \mathbb{IR}^m$ and $\boldsymbol{c} \in \mathbb{IR}^n$ be given. Let $\boldsymbol{f} = [\underline{f}, \overline{f}]$ be the optimal value range corresponding to the interval LP problem $\max_{x\in\mathcal{M}(\boldsymbol{A},\boldsymbol{b})} \boldsymbol{c}^T x$ defined as [3, 6, 7, 12]

$$\underline{f} := \inf_{A\in\boldsymbol{A},\, b\in\boldsymbol{b},\, c\in\boldsymbol{c}} \max_{x\in\mathcal{M}(A,b)} c^T x,$$
$$\overline{f} := \sup_{A\in\boldsymbol{A},\, b\in\boldsymbol{b},\, c\in\boldsymbol{c}} \max_{x\in\mathcal{M}(A,b)} c^T x.$$

Condition (1) for testing redundancy is extended to the interval case in the following way.

Proposition 1. *The inequality $\boldsymbol{c}^T x \le \boldsymbol{d}$ is*
1. *strongly redundant if and only if $\overline{f} \le \underline{d}$,*
2. *weakly redundant if and only if $\underline{f} \le \overline{d}$.*

Redundancy in inequality systems with nonnegative variables is easy to detect directly due to inclusion monotonicity w.r.t. the coefficients.

Proposition 2. *The inequality $\boldsymbol{c}^T x \le \boldsymbol{d}$ is strongly redundant for*

$$\boldsymbol{A}x \le \boldsymbol{b},\ x \ge 0$$

if and only if $\overline{c}^T x \le \underline{d}$ is redundant for

$$\underline{A}x \le \overline{b},\ x \ge 0.$$

Proposition 3. *The inequality $\boldsymbol{c}^T x \le \boldsymbol{d}$ is weakly redundant for*

$$\boldsymbol{A}x \le \boldsymbol{b},\ x \ge 0$$

if and only if $\underline{c}^T x \le \overline{d}$ is redundant for

$$\overline{A}x \le \underline{b},\ x \ge 0.$$

If we lose nonnegativity of variables, strong redundancy becomes NP-hard to check due to [5]. The strong and weak redundancy characterizations below are based on the optimal value range descriptions from [2, 3]. In the following we denote by e the vector of ones of suitable dimension. So, for example, in the statement below we have $A_{e,s} = A_c - \operatorname{diag}(e)A_\Delta \operatorname{diag}(s) = A_c - A_\Delta \operatorname{diag}(s)$ or $c_{-s,1} = c_c + \operatorname{diag}(s)c_\Delta$.

Proposition 4. *The inequality $\boldsymbol{c}^T x \le \boldsymbol{d}$ is strongly redundant for*

$$\boldsymbol{A}x \le \boldsymbol{b}$$

if and only if the inequality $c_{-s,1}^T x \le \underline{d}$ is redundant for $A_{e,s}x \le \overline{b}$ for every $s \in \{\pm 1\}^n$, that is,

$$\underline{d} \ge \max\ c_{-s,1}^T x \ \textit{ subject to } \ A_{e,s}x \le \overline{b}$$

for every $s \in \{\pm 1\}^n$.

Proposition 5. *The inequality $\boldsymbol{c}^T x \le \boldsymbol{d}$ is weakly redundant for*

$$\boldsymbol{A}x \le \boldsymbol{b}$$

if and only if the inequality $\underline{c}^T x^1 - \overline{c}^T x^2 \le \overline{d}$ is redundant for $\overline{A}x^1 - \underline{A}x^2 \le \underline{d},\ x^1, x^2 \ge 0$, that is,

$$\overline{d} \ge \max\ \underline{c}^T x^1 - \overline{c}^T x^2 \ \text{ subject to } \ \overline{A}x^1 - \underline{A}x^2 \le \underline{d},\ x^1, x^2 \ge 0.$$

Proposition 6. *The inequality $\boldsymbol{c}^T x + \boldsymbol{c}'^T y \le \boldsymbol{d}$ is strongly redundant for*

$$\boldsymbol{A}x + \boldsymbol{B}y = \boldsymbol{a},\ \boldsymbol{C}x + \boldsymbol{D}y \le \boldsymbol{b},\ x \ge 0$$

if and only if the inequality $\overline{c}^T x + c'^T_{-s,1} y \le \underline{d}$ is redundant to

$$\underline{A}x + B_{e,s}y \le \overline{a},\ \ -\overline{A}x - B_{-e,s}y \le -\underline{a},\ \ \underline{C}x + D_{e,s}y \le \overline{b},\ x \ge 0$$

for every $s \in \{\pm 1\}^k$, where k denotes the size of y.

Proof. By Proposition 1, the inequality is strongly redundant if and only if $\underline{d} \ge \overline{f}$, which in view of [8] takes the form

$$\begin{aligned} \underline{d} \ge \max_{s \in \{\pm 1\}^m}\ \max\ & \overline{c}^T x + c'^T_{-s,1} y \\ \text{subject to}\quad & \underline{A}x + B_{e,s}y \le \overline{a}, \\ & -\overline{A}x - B_{-e,s}y \le -\underline{a}, \\ & \underline{C}x + D_{e,s}y \le \overline{b},\ x \ge 0. \end{aligned}$$

This proves the statement. □

Proposition 7. *The inequality $\boldsymbol{c}^T x + \boldsymbol{c}'^T y \le \boldsymbol{d}$ is weakly redundant for*

$$\boldsymbol{A}x + \boldsymbol{B}y = \boldsymbol{a},\ \boldsymbol{C}x + \boldsymbol{D}y \le \boldsymbol{b},\ x \ge 0$$

if and only if the inequality $\underline{c}^T x + \underline{c}'^T y^1 - \overline{c}'^T y^2 \le \overline{d}$ is redundant to

$$A_{-s,e}x + B_{-s,e}y^1 - B_{s,e}y^2 = a_{s,1},\ \ \overline{C}x + \overline{D}y^1 - \underline{D}y^2 \le \underline{b},\ \ x, y^1, y^2 \ge 0$$

for some $s \in \{\pm 1\}^m$, where m denotes the number of equations.

Proof. By Proposition 1, the inequality is strongly redundant if and only if $\overline{d} \ge \underline{f}$, which in view of [8] takes the form

$$\begin{aligned} \overline{d} \ge \min_{s \in \{\pm 1\}^k}\ \max\ & \underline{c}^T x + \underline{c}'^T y^1 - \overline{c}'^T y^2 \\ \text{subject to}\ & A_{-s,e}x + B_{-s,e}y^1 - B_{s,e}y^2 = a_{s,1}, \\ & \overline{C}x + \overline{D}y^1 - \underline{D}y^2 \le \underline{b}, \\ & x, y^1, y^2 \ge 0. \end{aligned}$$

This proves the statement. □

3 Redundancy of equations

Let $\mathcal{M}(A, b)$ be a convex polyhedron. An equation $c^T x = d$ is redundant for $\mathcal{M}(A, b)$ if and only of

$$d = \min_{x \in \mathcal{M}(A,b)}\ c^T x = \max_{x \in \mathcal{M}(A,b)}\ c^T x. \tag{2}$$

Redundancy of equations is thus reduced to solving two LP problems.

Let interval data $\boldsymbol{A} \in \mathbb{IR}^{m \times n}$, $\boldsymbol{b} \in \mathbb{IR}^m$ and $\boldsymbol{c} \in \mathbb{IR}^n$ be given. Let $\boldsymbol{f}^1$ and $\boldsymbol{f}^2$ be the optimal value ranges corresponding to the interval LP problems $\min_{x \in \mathcal{M}(\boldsymbol{A},\boldsymbol{b})}\ \boldsymbol{c}^T x$ and $\max_{x \in \mathcal{M}(\boldsymbol{A},\boldsymbol{b})}\ \boldsymbol{c}^T x$, respectively. Strong redundancy can be easily characterized via these optimal value ranges, but for weak redundancy no such a simple condition is known.

Proposition 8. *The following statements are equivalent for strong redundancy of $\boldsymbol{c}^T x = \boldsymbol{d}$:*

1. $\boldsymbol{f}^1, \boldsymbol{f}^2$ *and* $\boldsymbol{d}$ *are real and* $f^1 = f^2 = d$,
2. $\overline{f}^2 \leq \underline{d}$ *and* $\underline{f}^1 \geq \overline{d}$,
3. $\bigcup_{A \in \boldsymbol{A},\, b \in \boldsymbol{b}} \mathcal{M}(A, b) \subseteq \bigcap_{c \in \boldsymbol{c},\, d \in \boldsymbol{d}} \{x;\ c^T x = d\}$.

Proof. The first condition follows from (2).

The second condition implies that for every realization and the optimal values $f^1 := \min_{x \in \mathcal{M}(A,b)} c^T x$ and $f^2 := \max_{x \in \mathcal{M}(A,b)} c^T x$ we have

$$\overline{d} \leq \underline{f}^1 \leq f^1 \leq f^2 \leq \overline{f}^2 \leq \underline{d}.$$

Thus all inequalities hold as equations, from which the statement follows.

The third condition is obvious. □

The second condition implies that strong redundancy of equations can be reduced to redundancy of inequalities.

Corollary 1. The equation $\boldsymbol{c}^T x = \boldsymbol{d}$ is strongly redundant if and only both inequalities $\boldsymbol{c}^T x \leq \boldsymbol{d}$ and $\boldsymbol{c}^T x \geq \boldsymbol{d}$ are strongly redundant.

The third condition in Proposition 8 says that the weak solution set of the interval system must be contained in the strong solution set of the interval equation $\boldsymbol{c}^T x = \boldsymbol{d}$. Weak solutions of interval systems were thoroughly dealt with in [3, 8]. In particular, strong solutions of $\boldsymbol{c}^T x = \boldsymbol{d}$ are described by linear equations

$$c_c^T x = d, \ \ x_i = 0 \ \forall i : (c_\Delta)_i > 0,$$

where $\boldsymbol{d} = d$ must be real.

By Corollary 1 and Proposition 2, we have:

Proposition 9. *The equation $\boldsymbol{c}^T x = \boldsymbol{d}$ is strongly redundant for*

$$\boldsymbol{A} x \leq \boldsymbol{b},\ x \geq 0$$

if and only if two inequalities $\overline{c}^T x \leq \underline{d}$ and $\underline{c}^T x \geq \overline{d}$ are redundant for

$$\underline{A} x \leq \overline{b},\ x \geq 0.$$

Similarly, by Corollary 1 and the Oettli–Prager characterization of weak solutions of systems of interval linear equations [3, 13], we have:

Proposition 10. *The equation $\boldsymbol{c}^T x = \boldsymbol{d}$ is strongly redundant for*

$$\boldsymbol{A} x = \boldsymbol{b},\ x \geq 0$$

if and only if two inequalities $\overline{c}^T x \leq \underline{d}$ and $\underline{c}^T x \geq \overline{d}$ are redundant for

$$\underline{A} x \leq \overline{b},\ \overline{A} x \leq \underline{b},\ x \geq 0.$$

Proposition 11. *The equation $\boldsymbol{c}^T x = \boldsymbol{d}$ is strongly redundant for*

$$\boldsymbol{A} x \leq \boldsymbol{b}$$

if and only if

$$\underline{d} \geq \max\ c_{-s,1}^T x \ \ \textit{subject to}\ \ A_{e,s} x \leq \overline{b}$$

and

$$\overline{d} \leq \min\ c_{s,1}^T x \ \ \textit{subject to}\ \ A_{e,s} x \leq \overline{b}$$

for every $s \in \{\pm 1\}^n$.

The above characterization has exponential complexity, which is justified by the following observation.

Proposition 12. *Checking strong redundancy of $\boldsymbol{c}^T x = \boldsymbol{d}$ for system $\boldsymbol{A} x \leq \boldsymbol{b}$ is a co-NP-hard problem.*

Proof. It is known that checking weak solvability of $\boldsymbol{A}x \leq \boldsymbol{b}$ is an NP-hard problem [3, 14]. Consider the interval equation $\boldsymbol{c}^T x = \boldsymbol{d}$ in the form $e^T x = K$, where $K > 0$ is sufficiently large constant (its value of polynomial size can be a priori determined by means of [15]). Now, if $\boldsymbol{A}x \leq \boldsymbol{b}$ is weakly solvable, then $e^T x = K$ is not strongly redundant, and if $\boldsymbol{A}x \leq \boldsymbol{b}$ is not weakly solvable, then $e^T x = K$ is strongly redundant. □

For weak redundancy, the situation is worse. One of the few cases easy to characterize are interval equations. Notice, however, that this type of redundancy is NP-hard to check since weak solvability of interval linear equations is intractable [3, 10].

Proposition 13. *The equation $\boldsymbol{c}^T x = \boldsymbol{d}$ is weakly redundant for $\boldsymbol{A}x = \boldsymbol{b}$ if and only if the interval system*

$$\boldsymbol{A}^T z = \boldsymbol{c},\ \boldsymbol{b}^T z = \boldsymbol{d}$$

is weakly solvable.

Proof. A real equation $c^T x = d$ is redundant for $Ax = b$ if and only if it is linearly dependent on the equation. In other words, $A^T z = c$, $b^T z = d$ is solvable. Weak redundancy is thus characterized by weak solvability of the interval case. □

For interval systems of linear inequalities with nonnegative variables, we can see by direct inspection that weak redundancy reduces to redundancy of a certain real instance of the interval system. Together with Proposition 13, we then derive a complete characterization.

Proposition 14. *The equation $\boldsymbol{c}^T x = \boldsymbol{d}$ is weakly redundant for*

$$\boldsymbol{A}x \leq \boldsymbol{b},\ x \geq 0$$

if and only if it is weakly redundant for

$$\overline{A}x \leq \underline{b},\ x \geq 0. \tag{3}$$

Recall that the affine hull of a set $\mathcal{S}$ is the smallest (w.r.t. inclusion) affine space containing $\mathcal{S}$, and each affine space can be described by a system of linear equations.

Proposition 15. *The equation $\boldsymbol{c}^T x = \boldsymbol{d}$ is weakly redundant for*

$$\boldsymbol{A}x \leq \boldsymbol{b},\ x \geq 0$$

if and only if the interval system

$$\tilde{A}^T z = \boldsymbol{c},\ \tilde{b}^T z = \boldsymbol{d} \tag{4}$$

is weakly solvable, where $\tilde{A}x = \tilde{b}$ is the affine hull of (3).

Proof. By Proposition 14, in order that $\boldsymbol{c}^T x = \boldsymbol{d}$ is weakly redundant, the set described by (3) has to lie in some realization of $\boldsymbol{c}^T x = \boldsymbol{d}$, that is, in some hyperplane $c^T x = d$. Therefore also its affine hull must lie in the hyperplane, and the rest follows from Proposition 13. □

Notice that solvability of the interval system (4) is easily recognized since intervals are situated in the right-hand side only. So the weak solution set of the interval system (4) is equivalent to solution set of the real linear system

$$\underline{c} \leq \tilde{A}^T z \leq \overline{c},\ \ \underline{d} \leq \tilde{b}^T z \leq \overline{d}.$$

For other types of interval systems, only sufficient or necessary conditions are known.

Proposition 16. *The equation $\boldsymbol{c}^T x = \boldsymbol{d}$ is weakly redundant for*

$$\boldsymbol{A}x = \boldsymbol{b},\ x \geq 0$$

if the interval system

$$\boldsymbol{A}^T z = \boldsymbol{c},\ \boldsymbol{b}^T z = \boldsymbol{d}$$

is weakly solvable.

Proof. By Proposition 13, weak solvability of the system implies that $\boldsymbol{c}^T x = \boldsymbol{d}$ is weakly redundant for $\boldsymbol{A}x = \boldsymbol{b}$, so it must be weakly redundant for $\boldsymbol{A}x = \boldsymbol{b}, x \geq 0$ as well. □

The condition is only sufficient for weak redundancy and not necessary. This is shown by the example of the (real) system $x_1 + x_2 = 0, x_1, x_2 \geq 0$, for which $x_1 + 2x_2 = 0$ is redundant, but the condition is not satisfied.

4 Conclusion

We considered redundancy of a general interval system of linear equations and inequalities as well as its special form cases. We discussed two concepts of redundancy – the weak ("for some") and the strong ("for every"). We characterized them by a reduction to one or more real cases. In some situations, the problem is polynomial, but for certain systems it is NP-hard. What remains as an open problem is a complete characterization of weak redundancy of an interval equation; here, we state only a sufficient condition or a characterization for special form systems.

Acknowledgements

The author was supported by the Czech Science Foundation Grant P403-18-04735S.

References

[1] Caron, R. J., Boneh, A., and Boneh, S.: Redundancy. In: *Advances in Sensitivity Analysis and Parametic Programming* (Gal, T., and Greenberg, H. J., eds.). Springer, Boston, MA, 1997, 449–489.

[2] Chinneck, J. W., and Ramadan, K.: Linear programming with interval coefficients. *J. Oper. Res. Soc.* **51** (2000), 209–220.

[3] Fiedler, M., Nedoma, J., Ramík, J., Rohn, J., and Zimmermann, K.: *Linear Optimization Problems with Inexact Data*. Springer, New York, 2006.

[4] Fukuda, K., and Szedlák, M.: Redundancies in linear systems with two variables per inequality, 2016. Preprint, `https://arxiv.org/abs/1610.02820`.

[5] Gabrel, V., Murat, C., and Remli, N.: Linear programming with interval right hand sides. *Int. Trans. Oper. Res.* **17** (2010), 397–408.

[6] Garajová, E., Hladík, M., and Rada, M.: Interval linear programming under transformations: optimal solutions and optimal value range. *Cent. Eur. J. Oper. Res.* **27** (2019), 601–614.

[7] Hladík, M.: Interval linear programming: A survey. In: *Linear Programming – New Frontiers in Theory and Applications* (Mann, Z. A., ed.), chapter 2. Nova Science Publishers, New York, 2012, 85–120.

[8] Hladík, M.: Weak and strong solvability of interval linear systems of equations and inequalities. *Linear Algebra Appl.* **438** (2013), 4156–4165.

[9] Karwan, M. H., Lotfi, V., Telgen, J., and Zionts, S., eds.: *Redundancy in mathematical programming. A state-of-the-art survey*, volume 206 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Cham, 1983.

[10] Lakeev, A. V., and Noskov, S. I.: On the solution set of a linear equation with the right-hand side and operator given by intervals. *Sib. Math. J.* **35** (1994), 957–966.

[11] Lodwick, W. A.: Analysis of structure in fuzzy linear programs. *Fuzzy Sets Syst.* **38** (1990), 15–26.

[12] Novotná, J., Hladík, M., and Masařík, T.: Duality gap in interval linear programming. *J. Optim. Theory Appl.* **184** (2020), 565–580.

[13] Oettli, W., and Prager, W.: Compatibility of approximate solution of linear equations with given error bounds for coefficients and right-hand sides. *Numer. Math.* **6** (1964), 405–409.

[14] Rohn, J.: *NP-hardness results for some linear and quadratic problems*. Technical Report 619, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, 1995. URL `http://hdl.handle.net/11104/0122691`.

[15] Schrijver, A.: *Theory of Linear and Integer Programming. Repr.* Wiley, Chichester, 1998.

[16] Szedlák, M. K.: *Redundancy in Linear Systems: Combinatorics, Algorithms and Analysis*. Ph.D. thesis, ETH Zurich, 2017. URL `https://doi.org/10.3929/ethz-b-000167108`.

Optimizing Production in Small-Scale Beef Cattle Farm

Robert Hlavatý[1], Igor Krejčí[2]

Abstract. The focus of our contribution is optimization of planning and breeding in a small-scale beef cattle farm in a long-term horizon. We present a vast linear programming model capturing all categories of cattle. The overall balance of cattle categories is influenced in a positive sense by the reproduction cycle and purchasing of heavily pregnant heifers in a monthly sense. The decrease in the overall state of the herd is caused by herd turnover and sales in the meat sector. Our intention is to design a mathematical model and find through linear programming such balance that generates the maximum and stable profit in a long-term horizon. The constraints are determined by the availability of space in cattle stables and herd turnover. We maximize the profit over sales of meat bulls and meat cows. The life cycle of all categories is captured by a tree graph. On its basis, the mathematical model of linear optimization is presented in detail. We optimize according to various scenarios of purchases and sales subjected to changing prices, subsidies and limits of farm capacity. It occurs that the subsidies play a key role in maintaining of prosperous farm under the current circumstances in the market.

Keywords: beef cattle farm, linear programming, optimization, subsidies

JEL Classification: C61, Q12
AMS Classification: 90C05, 90C90

1 Introduction

Small farms are very specific subjects in the agriculture entrepreneurship and its sustainability is an important part of Common agricultural policy. It is a usual phenomenon that subsidies for farmers are the main factor of farmer's business survival. The small beef cattle farms are especially vulnerable to ineffective decision making since the overall profitability is low and there are large delays between the decisions and their impact caused by the biological characteristic of the species. We seek to design a detailed optimization model that would be able to capture real-farm characteristics and consequently help a farmer to find the optimal patterns of behaviour in terms of treating the beef cattle herd. First, we introduce a brief review of relevant sources, our own modelling approach follows afterwards and it is demonstrated in the end how to possibly take advantage of the calculations in the favour of the farmer.

The methods of OR have been proved an efficient tool for various sectors of agriculture and related decision making, as recently shown by Krejčí and Houška [14], Marušák et al. [15], Kašpar et al. [13], Behzadi et al. [3]. It is often the mathematical optimization that is used as a tool to maximize profit under the given economic and organization constraints which was demonstrated in the last decade by Ahumada and Villalobos [1] or Tong et al. [21]. In this contribution, we especially focus on linear programming (LP) and therefore we introduce some important papers of past two decades that are focused on using LP in the agriculture sector, especially in relation to beef herd management.

Already in 1974, Wilton et al. [24] introduced quite a complex LP model for cattle farming, which included not only breeding factors but was also meant to optimize towards cropping, feeding and nutritional requirements. Many case studies were developed since then, which is apparent from the review study of Stygar and Makulska [18], wherein it is summarized how the various authors used different techniques to treat the beef herd management. It turns out that optimization through LP is the most frequent tool along with the dynamic programming and simulations. All LP models described differ in the modelling purposes. Some of them are oriented to economic aspects and policies while the others are concerned with environmental issues. It is presented by Veysset et al. [22] how to set the production activities in an optimal manner in reaction to EU policy changes. Similar goals are pursued by Crosson et al. [5] but under the conditions in Ireland whereas the previous takes places in France. Our model gets inspired by the cooperation with Czech farmers and it is not the first attempt to use LP for the purposes of beef herd optimization. Havlík et al. [11]

[1] CZU Prague, Department of Systems Engineering, Kamýcká 129, 165 00 Prague, hlavaty@pef.czu.cz
[2] CZU Prague, Department of Systems Engineering, Kamýcká 129, 165 00 Prague, krejcii@pef.czu.cz

introduce an LP model, which focused mainly on the biodiversity and its relationship with the production of beef. Unlike this approach, we rather focus on internal processes within the beef herd itself. There is a strong bond between beef farming and environmental sustainability and the matters of sustainable management in this sector are discussed by Costa and Rehman [4] or Gallegos Rivero and Daim [9].

Nevertheless, using mathematical programming in the agriculture sector is not necessarily reduced to beef cattle only. Rodríguez-Sánchez et al. [17] use the LP for complex modelling of processes in sow farms. Another study is dedicated to dairy sheep farms by Villalba et al. [23] who optimize by the means of the multi-objective point of view with the help of stochastic simulation. After all, it is not always the profit maximization being a goal of optimization in the farms. This is demonstrated by Groot et al. [10] or Anastasidis and Chukova [2] who take advantage of mathematical optimization to deal with the problems of organic matter balance or water quality in the latter one. The specifics of our modelling approach are the conditions of small-scale beef farm, which, unlike its larger competitors, is dependent mainly on beef production as such. Our methodology is designed to capture detailed herd structure with respect to aging of each category of cattle. It takes into account the dynamic nature of the problem, which causes the model to grow considerably large. In this sense, we take advantage of P-class complexity of LP with the support of powerful optimization solver.

2 Materials and methods

2.1 Herd lifecycle and variables

In this section, the model of beef herd is described in detail. First, we introduce a tree graph to capture the beef herd lifecycle (Figure 1) and at the same time introduce the variables used in our model further (Table 1).

VARIABLE	DESCRIPTION
C_i	Set of all calves in generation i of age $\in (0,7]$ months
C_i^H	Set of all heifer calves in generation i of age $\in (0,7]$ months
F_HEI_i	Set of all fattening heifers in generation i of age $\in (7,24]$ months
B_HEI_i	Set of all breeding heifers in generation i of age $\in (7,24]$ months
$P\_HEI\_A_{ij}$	Set of all pregnant heifers in generation i, j-th pregnancy ≤ 5 months, $j = 1$
$P\_HEI\_B_{ij}$	Set of all pregnant heifers in generation i, j-th pregnancy > 5 months, $j = 1$
P_COW_{ij}	Set of all cows in generation i, j-th pregnancy
$C_{(i+1)j}$	Set of all calves born from generation i from j-th pregnancy
C_i^B	Set of all bull calves in generation i of age $\in (0,7]$ months
F_BUL_i	Set of all fattening bulls in generation i of age $\in (7,24]$ months
HP_HEI_l	Number of heavily pregnant heifers bought in the year $l, l = 1, \dots, 8$

Table 1 Variables and beef herd categories

Note that the variable indices are not specifically defined in their dimension. The generation means the set of all calves born in the same month and its older equivalents. The dimeson of index i denotes the number of all generations in the model. The index $j = 1, \dots, n$ denotes the number of pregnancies that can occur for a single heifer or consequently cow. The number n depends on the individual generation i.

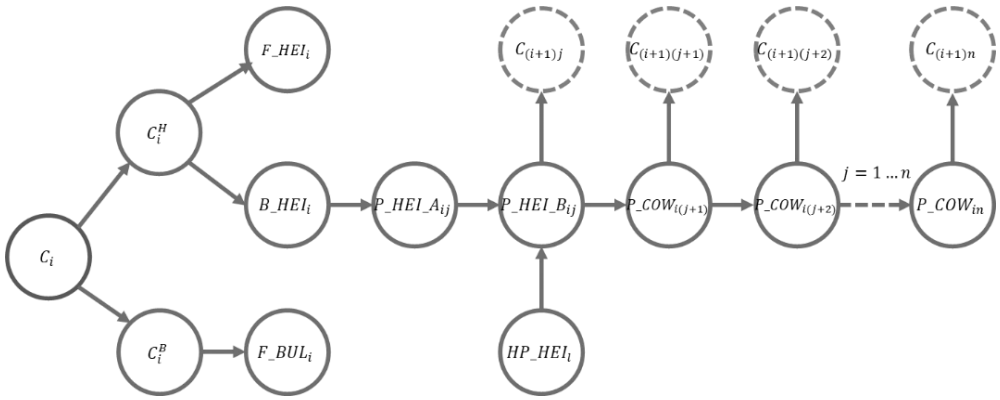


Figure 1 Tree graph of beef herd lifecycle

The Figure 1 depicts the tree with the relationships between all beef herd categories. Each vertex represents a category and the oriented edges show how a category generates another one. There is a moment at the start of each generation tree when a calf C_i is born (or a set of calves of the same generation, in fact, however, for the sake of simplicity, the rest of the graph will be described for a single animal). The calf C_i can be either female (heifer calf) C_i^H or male (bull calf) C_i^B. The bull calf C_i^B grows later into mature fattening bull F_BUL_i which is consequently terminated by slaughter.

The heifer calf C_i^H becomes either fattening heifer F_HEI_i and is terminated by slaughter between the age of 18–24 months. Alternatively, the heifer calf C_i^H becomes breeding heifer B_HEI_i (which is at most 50%). After a period of time, the breeding heifer B_HEI_i becomes pregnant and consequently $P\_HEI\_A_{ij}$ which is a first part of the heifer's pregnancy, $j = 1$. After 5 months, it turns to $P\_HEI\_B_{ij}$ (heavily pregnant heifer), which is the second part of the heifer pregnancy resulting in calf $C_{(i+1)j}$ birth. Distinguishing between two stages of heifer pregnancy is important due to different cost coefficients for both categories. The heavily pregnant heifer category is also the moment when a farmer mostly makes an acquisition of new heads to extend the herd. After the first birth, a new pregnancy occurs after a mandatory service period and it shifts to category $P\_COW_{i(j+1)}$ whereas it becomes a mature cow being pregnant for $(j + 1)$th time. The pregnancy leads to another birth of calf $C_{(i+1)(j+1)}$ and so the process of pregnancy iterates until n-th pregnancy occurs.

Whenever a calf $C_{(i+1)j}, \forall i, j = 1, \dots, n$ is born, it creates a new tree of the very same structure. We want to avoid excessive notation and assume that in the new tree, the index of the generation is $\hat{\imath} = (i + 1)j$ and thus becomes one digit larger than the previous generation. This way, the generation index allows tracking the "distance" of each generation from the initial generation based on the number of digits.

2.2 Linear programming model

We construct the optimization problem based on the relationships described in the tree graph (Figure 1) and we take into consideration the time aspects and the constraints described further. Profit maximization is the objective of optimization and is subjected to internal and external constraints. The general description of the optimization problem is given as:

$$\begin{gathered} \text{maximize } (P) \\ \text{s. t.} \\ H_k \in \mathrm{I} \cap \mathrm{O}, \forall k \\ H_k \subset \mathbb{R}^+ \end{gathered} \tag{1}$$

P is an aggregate variable expressing the total profit. H_k is a set of all variables that occur in month k. I and O are polyhedral sets of all internal and external constraints. All variables are allowed to attain real non-negative values. A reader should note that we do not impose any integrality constraints on our problem. Considering the size of our model in terms of variables and constraints, this would be a senseless trade-off between increased precision and computational complexity. Prior to the very description of the constraints, it is necessary to introduce a couple of more parameters and cost coefficients, as presented in Table 2 and Table 3.

PARAMETER	DESCRIPTION	RANGE	DEFAULT
ω	Calves' mortality	3–7%	3%
ψ	Heifer and cow culling rate	10–20%	15%
N	Normal initial size of the herd	100	
Γ	Farm capacity (with regards to the normal size of the herd)	140–180%	150%
π	Limit to a single purchase of HP_HEI_l as a share of normal herd size	0–10%	10%

Table 2 Beef herd parameters

COEFFICIENT	DESCRIPTION	DEFAULT[5]
s^{FAT}	Monthly costs per cattle for fattening[1]	45.75 EUR
s^{COW}	Monthly costs per suckler cow including the calve[1]	69.78 EUR
s^{HEI_A}	Monthly costs per heifers till 5 months of gestation[1]	36.54 EUR
s^{HEI_B}	Monthly costs per heavily pregnant heifers[1]	42.63 EUR
r^{SUB}	Average monthly subsidies per large livestock unit including the land subsidies[2]	44.61 EUR (Range: ±50%)
r^{FAT}	Fattening bull unit sale price[3]	998.74 EUR
r^{COW}	Fattening heifer / suckler cow unit sale price[3]	739.81 EUR
s^{HP}	Price of heavily pregnant heifer[3]	1,294.67 EUR
N/A	Hectare per livestock unit used for the subsidies calculation[3,4]	1 ha/1 LSU

Table 3 Costs, revenues and subsidies (Sources: [1][12], [2][19], [3][16], [4][6], [5]Exchange rate 27.034 CZK/EUR in 2016 [8])

According to Eurostat [7], the small-scale farmers mostly depend on their family member's workforce and thus the costs are already lowered by the possible labour costs. All cost coefficients are related to the year 2016. Using the coefficients from Table 3, we construct the objective function which as an aggregate of costs, revenues and subsidies of individual beef cattle categories in our model. The objective stands as:

$$\begin{aligned} P = \sum_i \sum_{j=1}^{n} \Big(& (6 * 0.4r^{SUB} + 0.6r^{SUB})C_i^H + (14 * 0.6r^{SUB} - 14s^{FAT} + r^{COW})(F\_HEI_i + B\_HEI_i) \\ & + (5r^{SUB} - 5s^{HEI\_A})P\_HEI\_A_{i1} + (8r^{SUB} - 8s^{HEI\_B} + \psi r^{COW})P\_HEI\_B_{i1} \\ & + (13r^{SUB} - 13s^{COW} + \psi r^{COW})P\_COW_{j+1}\Big) \\ & + \sum_i \Big((3 * 0.6r^{SUB})C_i^B + (14 * 0.6r^{SUB} - 14 * 0.6s^{FAT} + r^{FAT})F\_BUL_i\Big) \\ & + \sum_{l=1}^{8} s^{HP} * HP\_HEI_l \end{aligned} \tag{2}$$

Three sums in formula (2) complete the entire objective. The first sum is related to heifers and suckler cows, the second sum is related to bulls and the final sum is related to purchasing of heavily pregnant heifers. All parameters are expressed in monthly manner and it is necessary to modify them by the multipliers accordingly. The integer multipliers simply express the number of months the individual category exists and the decimal multipliers are used to lower the coefficients in case of young categories, which have different cost/revenue numbers compared to the mature livestock categories. The set of constraints is divided into two subsets I and O to distinguish between internal and external factors of the herd lifecycle. While the internal constraints represent relationships between categories, the external constraints affect the herd lifecycle. The set I is formed as the intersection of the following equations and inequalities (3–11)

$$(1 - \omega)C_i = C_i^H + C_i^B, \forall i \tag{3}$$

$$C_i^H = C_i^B, \forall i \tag{4}$$

$$C_i^H = F\_HEI_i + B\_HEI_i, \forall i \tag{5}$$

$$F\_HEI_i \geq 0.5\, C_i^H, \forall i \tag{6}$$

$$B\_HEI_i \leq 0.5\, C_i^H, \forall i \tag{7}$$

$$B\_HEI_i = P\_HEI\_A_{ij} = P\_HEI\_B_{ij}, \forall i;\ j = 1 \tag{8}$$

$$P\_COW_{i(j+1)} = (1-\psi) P\_HEI\_B_{ij}, \forall i\ ;\ j = 1 \tag{9}$$

$$P\_COW_{i(j+1)} = (1-\psi) P\_COW_{i(j+2)}, \forall i\ ; j = 1 \ldots n-2 \tag{10}$$

$$C_i^B = F\_BUL_i \quad \forall i \tag{11}$$

(3) expresses how all new-born calves are split into heifer calves C_i^H and bull calves C_i^B with the given mortality rate ω. The distribution of male and female calves is even (4). Each heifer calf C_i^H can later become either fattening heifer F_HEI_i or breeding heifer B_HEI_i (5). It is assured by 6 and 7 that the decision how to split heifers in 5 is made individually for each generation i. This is influenced by the current state of the herd capacity in the given month. If the farm capacity appears to nearly full then the heifer calves C_i^H are sent to become fattening heifers F_HEI_i and consecutively leave the farm as they are sold for slaughter and thus free the capacity and increase the profit. If the current farm capacity offers enough space for new breeding heifers, then it is more likely that C_i^H become B_HEI_i and further breed, thus filling the farm capacity. (8) merely shows that every breeding heifer B_HEI_i becomes pregnant heifer $P\_HEI\_A_{ij}$ (and $P\_HEI\_B_{ij}$ later on) without any losses in numbers. Whenever a birth occurs and a heifer (or suckler cow) is under service period, a culling is done afterwards by the coefficient ψ as shown in (9). The same culling process is applied in (10) as the suckler cows grow older. Bull calves C_i^B only grow older and become F_BUL_i as we do not assume any bull calves to be sent to test of their genetic performance and become breeding bulls (11).

The set of external constraints O is described by the following set of inequalities:

$$\begin{gathered} C_i^H + F\_HEI_i + B\_HEI_i + P\_HEI\_A_{ij} + P\_HEI\_B_{ij} + P\_COW_{ij} + C_i^B + F\_BUL_i + HP\_HEI_l \leq \Gamma\mathrm{N}, \\ \left\{C_i^H, F\_HEI_i, B\_HEI_i, P\_HEI\_A_{ij}, P\_HEI\_B_{ij}, P\_COW_{ij}, C_i^B, F\_BUL_i, HP\_HEI_l\right\} \in H_k, \forall i, j, k, l \end{gathered} \tag{12}$$

$$HP\_HEI_l \leq \pi\mathrm{N}, \quad l = 1, \ldots, 8 \tag{13}$$

(12) expresses that the overall capacity of the farm can be no more than the number of ΓN and this restriction is to be satisfied in each month k. All categories from various generations cannot exceed the capacity ΓN in every single month k as stated in the second part of constraint (12). Constraint (13) imposes a restriction that the number of heavily pregnant heifers HP_HEI_l both in the month l can be no more than a part of the entire herd. The reason is the farmer's effort to avoid big leaps in capacity and there are legal restrictions as well, regarding the genetic structure and the age of the herd, which are obviously influenced by the new purchases.

3 Results and discussion

In this section, we would like to present some of the findings that we have reached upon running the various scenarios of the LP model described in the previous part of the chapter. The scenarios differ in two aspects. First, it is the initial structure of the herd that the farm starts with. Note that the constraints (3–11) only describe the relationships between individual categories but the initial values of starting generation must be fixed to determine the evolution of the herd in the following generations to be born. As a simple illustration, a reader can imagine that if there is a bigger portion of females than males in the beginning, this is going to lead to a larger amount of births in next generations and due to limited capacity of the farm, this will strongly influence the sales and purchases, consequently influencing the farm revenue. The second aspect of scenarios is changing individual parameters within a given range (Table 2), one at a time ceteris paribus. Combining different initial herd structures and different parameter values generate a vast amount of scenarios and due to the limited space offered by the conference paper, we present here only some of our findings related to the amount of subsidies that the farmer possibly acquires.

The *OpenSolver* environment was used for model creation and *Gurobi*© linear solver was consequently called to optimize the problem. The automated creation of different scenario inputs was carried out by our own-made VBA macro. The model consisting of 1,248 variables and 45,245 constraints was solved for all

scenarios. Each run took approximately 200 seconds. Our computations covered 8-year period evolvement of the initial herd.

The calculation shows how significant is the role of subsidies in the beef cattle farming sector. For the 100-heads initial herd, the 8-year optimal profit ranges from 67 852 EUR to 676 756 EUR if the subsidies are changed within –50% and +50% compared to normal. These changes are of course hypothetical but it is an evidence of its role and importance in the farmer's budget. We made a threshold analysis to find out what is the critical level of subsidies that determines the profitability of the entire business. By the sensitivity analysis, it turns out that a decrease of 28%, compared to normal subsidies, causes the business to cease being profitable. This also leads to continuous decrease of the heard size and causes the absolute preference of fattening and slaughtering before breeding. These findings are in accordance with Syrůček et al. [19] and Syrůček et al. [20] showing that beef cattle farming profitability is not achievable without subsidies under the current circumstances.

4 Conclusion

We presented a modelling approach to capture detailed behaviour of the beef herd in the long run. Usage of linear programming is advantageous, considering the size of the model while it does not spoil the information value of the results where the over-time herd proportions between individual beef categories are important, not the exact integer numbers. Being limited by the paper extent, we focused on the demonstration of how the proposed model can be used for determining the importance of individual model parameters on the objective. The attention was paid to changes in subsidies and its impact on the overall profit. Further on, we plan to investigate how the farm capacity extension contributes to competitiveness and determine the proper moments for the acquisition of heavily pregnant heifers to make the herd extension true.

Acknowledgements

This paper was supported by Czech Science Foundation (GAČR) project No. P403-18-04735S and Technology Agency of the Czech Republic (TAČR), project No. TJ01000068.

References

[1] AHUMADA, O. & VILLALOBOS, J. R. (2011). A tactical model for planning the production and distribution of fresh produce. *Annals of Operations Research*, *190 (1)*, 339–358.

[2] ANASTASIADIS, S. & CHUKOVA, S. (2016). An inertia model for the adoption of new farming practices. *International Transactions in Operational Research*, *26 (2)*, 667–685.

[3] BEHZADI, G., O'SULLIVAN, M. J., OLSEN, T. L., & ZHANG, A. (2018). Agribusiness supply chain risk management: A review of quantitative decision models. *Omega (United Kingdom)*, *79*, 21–42.

[4] COSTA, F. P. & REHMAN, T. (2005). Unravelling the rationale of 'overgrazing' and stocking rates in the beef production systems of Central Brazil using a bi-criteria compromise programming model. *Agricultural Systems*, *83 (3)*, 277–295.

[5] CROSSON, P., O'KIELY, P., O'MARA, F. P., & WALLACE, M. (2006). The development of a mathematical model to investigate Irish beef production systems. *Agricultural Systems*, *89 (2–3)*, 349–370.

[6] EUROSTAT (2013). Glossary:Livestock unit (LSU) [online]. Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Livestock_unit_(LSU) [Accessed 10 May 2019].

[7] EUROSTAT (2018). Small and large farms in the EU – statistics from the farm structure survey [online]. Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php/Small_and_large_farms_in_the_EU_-_statistics_from_the_farm_structure_survey [Accessed 27 Feb 2020].

[8] EUROSTAT (2020). ECU/EUR exchange rates versus national currencies [online]. Available from: https://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tec00033&plugin=1 [Accessed 29 Mar 2020].

[9] GALLEGOS RIVERO, A. R. & DAIM, T. (2017). Technology roadmap: Cattle farming sustainability in Germany. *Journal of Cleaner Production*, *142*, 4310–4326.

[10] GROOT, J. C. J., OOMEN, G. J. M., & ROSSING, W. A. H. (2012). Multi-objective optimization and design of farming systems. *Agricultural Systems*, *110*, 63–77.

[11] HAVLÍK, P., JACQUET, F., BOISSON, J.-M., HEJDUK, S., & VESELÝ, P. (2006). Mathematical programming models for agri-environmental policy analysis: A case study from the White Carpathians. *Agricultural Economics (Zemědělská ekonomika)*, *52 (2)*, 51–66.
[12] INSTITUTE OF AGRICULTURAL ECONOMICS AND INFORMATION. (2018). Costs of Agricultural Products [online]. Available from: http://www.iaei.cz/costs-of-agricultural-products/ [Accessed 5 Oct 2018].
[13] KAŠPAR, J., HLAVATÝ, R., KUŽELKA, K., & MARUŠÁK, R. (2017). The Impact of Assumed Uncertainty on Long-Term Decisions in Forest Spatial Harvest Scheduling as a Part of Sustainable Development. *Forests*, *8 (9)*, 335.
[14] KREJČÍ, I. & HOUŠKA, M. (2014). Discussion with the paper 'Project costs planning in the conditions of uncertainty' by H. Stikova, *60 (11)*, 517–523.
[15] MARUŠÁK, R., KAŠPAR, J., HLAVATÝ, R., KOTEK, V., KUŽELKA, K., & VOPĚNKA, P. (2015). Alternative Modelling Approach to Spatial Harvest Scheduling with Respect to Fragmentation of Forest Ecosystem. *Environmental Management*, *56 (5)*, 1134–1147.
[16] POLÁKOVÁ, J., MOULIS, P., KOLÁČKOVÁ, G., & TICHÁ, I. (2016). Determinants of the Business Model Change – A Case Study of a Farm Applying Diversification Strategy. *Procedia – Social and Behavioral Sciences*, *220*, 338–345.
[17] RODRÍGUEZ-SÁNCHEZ, S. V., PLÀ-ARAGONÉS, L. M., & ALBORNOZ, V. M. (2012). Modeling tactical planning decisions through a linear optimization model in sow farms. *Livestock Science*, *143 (2–3)*, 162–171.
[18] STYGAR, A. & MAKULSKA, J. (2010). Application of mathematical modelling in beef herd management – a review. *Annals of Animal Science*, *10 (4)*, 333–348.
[19] SYRŮČEK, J., KRPÁLKOVÁ, L., KVAPILÍK, J., & VACEK, M. (2017). *Kalkulace Ekonomických Ukazatelů v Chovu Skotu*. Prague: Institute of Animal Science.
[20] SYRŮČEK, J., KVAPILÍK, J., BARTOŇ, L., VACEK, M., & STÁDNÍK, L. (2017). Economic efficiency of suckler cow herds in the Czech Republic. *Agricultural Economics (Zemědělská ekonomika)*, *63 (1)*, 34–43.
[21] TONG, D., REN, F., & MACK, J. (2012). Locating farmers' markets with an incorporation of spatio-temporal variation. *Socio-Economic Planning Sciences*, *46 (2)*, 149–156.
[22] VEYSSET, P., BEBIN, D., & LHERM, M. (2005). Adaptation to Agenda 2000 (CAP reform) and optimisation of the farming system of French suckler cattle farms in the Charolais area: A model-based study. *Agricultural Systems*, *83 (2)*, 179–202.
[23] VILLALBA, D., DÍEZ-UNQUERA, B., CARRASCAL, A., BERNUÉS, A., & RUIZ, R. (2019). Multi-objective simulation and optimisation of dairy sheep farms: Exploring trade-offs between economic and environmental outcomes. *Agricultural Systems*, *173*, 107–118.
[24] WILTON, J. W., MORRIS, C. A., LEIGH, A. O., JENSON, E. A., & PFEIFFER, W. C. (1974). A Linear Programming Model for Beef Cattle Production. *Canadian Journal of Animal Science*, *54 (4)*, 693–707.

Impact of the Parametrization and the Scaling Function in Dynamic Score-Driven Models: The Case of the Negative Binomial Distribution

Vladimír Holý[1]

Abstract. Generalized autoregressive score (GAS) models offer a unified framework for modeling of time series based on any underlying distribution with any time-varying parameters. The extensive GAS literature demonstrates that it is quite effective and natural to use the score of the conditional density function to drive the time-varying parameters. There are, however, two aspects in GAS modeling which seem rather arbitrary or based purely on convenience. The first is the choice of the parametrization of the underlying distribution, especially regarding the parameters linked to the GAS dynamics. The second is the choice of the function scaling the score term in the GAS dynamics. In the paper, we investigate the effects of the parametrization and the scaling function in the case of dynamic count models based on the negative binomial distribution with time-varying location parameter. In two empirical examples of E. coli cases and robbery charges, we demonstrate that the choice of the parametrization and the scaling function does not dramatically impact the performance of the GAS models.

Keywords: generalized autoregressive score model, parametrization, scaling function, model misspecification, negative binomial distribution.

JEL Classification: C22, C46

1 Introduction

A modern approach for modeling of time series based on any underlying probability distribution with any time-varying parameters is the *generalized autoregressive score (GAS) model* of [4], also known as *dynamic conditional score (DCS) model* by [9]. The GAS model is an observation-driven model capturing dynamics of time-varying parameters by the autoregressive term and the scaled score of the conditional density function (or the conditional probability mass function). For a comprehensive list of GAS papers, see [14].

We focus on the negative binomial distribution. In the literature, the score-driven models based on the negative binomial distribution are utilized in [8] for offensive conduct reports and in [1] for trade durations with frequent split transactions. For the specifications of various count models with GAS dynamics, see [11]. For the general methodology of count data analysis, see [16], [10] and [3].

In the paper, we investigate the impact of two rather arbitrary aspects in GAS modeling – the parametrization of the underlying distribution and the function scaling the score term. Using publicly available time series of E. coli cases and robbery charges, we find that the choice of these two building blocks does not distinctly affect the performance of the GAS model.

2 Generalized Autoregressive Score Model

Let us model non-negative random variables $Y_t \in \mathbb{N}_0$, $t = 1, \dots, n$ and assume Y_t follow the negative binomial distribution with time-varying location parameter f_t and static scale parameter g. In the *generalized autoregressive score (GAS)* dynamics of [4], the time-varying parameter follows recursion

$$f_{t+1} = c + bf_t + aS(f_t, g)\nabla(y_t; f_t, g),$$

where c is the constant parameter, b is the autoregressive parameter, a is the score parameter, $S(f_{t-i+1}, g)$ is the *scaling function* for the score and $\nabla(y_t, f_t, g)$ is the *score* given by

$$\nabla(y_t; f_t, g) = \frac{\partial \log P[Y_t = y_t | f_t, g]}{\partial f_t}.$$

[1] University of Economics, Prague, vladimir.holy@vse.cz

The score has zero mean and (under some regularity conditions) variance equal to the *Fisher information* given by

$$\mathcal{I}(f_t, g) = \mathrm{E}\left[\nabla(y_t; f_t, g)\nabla(y_t; f_t, g)' | f_t, g\right].$$

The scaling function is typically choosen to be equal to one (the *unit scaling*), the square root of the inverse of the Fisher information $\mathcal{I}(f_t, g)^{-\frac{1}{2}}$ (the *square-root scaling*) or the inverse of the Fisher information $\mathcal{I}(f_t, g)^{-1}$ (the *inverse* scaling). The model can be straightforwardly estimated by the *maximum likelihood method*.

In the GAS framework, when parameters of the distribution β are required to be positive or to belong in a given interval, they are transformed by the *link function* $\tilde{\beta} = h(\beta)$ in order to make them unrestricted. Let $\dot{h} = \partial h(\beta)/\partial \beta'$. Note that the reparametrization affects the score and the Fisher information as

$$\tilde{\nabla}\left(y; \tilde{\beta}\right) = \left(\dot{h}'\right)^{-1} \nabla\left(y; \tilde{\beta}\right), \qquad \tilde{\mathcal{I}}\left(\tilde{\beta}\right) = \left(\dot{h}'\right)^{-1} \mathcal{I}\left(\tilde{\beta}\right)\left(\dot{h}\right)^{-1}.$$

For a parameter $\beta > 0$, the *logarithmic transform* $\tilde{\beta} = \ln \beta$ is utilized. For a parameter β in $(0, 1)$, the *logistic transform* $\tilde{\beta} = \ln(\beta/(1-\beta))$ is utilized.

3 Negative Binomial Distribution

The *negative binomial distribution* is a two-parameter distribution for count data exhibiting overdispersion (i.e. variance equal to the mean). It contains the Poisson distribution with equidispersion as a special case. The Poisson and negative binomial distributions are the most common distributions used for the analysis of count data.

It can be defined as the distribution of the number of failures in a sequence of i.i.d. Bernoulli trials with probability of success $\pi \in (0, 1)$ before $\nu > 0$ successes occur. Note that ν should be integer to make the definition meaningful but can be straightforwardly extended to postive real number. We denote this parametrization as the *waiting time (WT) parametrization*. We refer to π as the location parameter and ν as the scale parameter. The probability mass function is given by

$$\mathrm{P}\left[Y = y | \pi, \nu\right] = \frac{\Gamma(y+\nu)}{\Gamma(y+1)\Gamma(\nu)}(1-\pi)^{\nu}\pi^{y}, \qquad \text{for } y = 0, 1, \ldots,$$

where $\Gamma(\cdot)$ denotes the gamma function. The expected value and variance are respectively given by

$$\mathrm{E}[Y] = \frac{\pi\nu}{1-\pi}, \qquad \mathrm{var}[Y] = \frac{\pi\nu}{(1-\pi)^2}.$$

The score is a vector with 2 elements given by

$$\begin{aligned}
\nabla_1(y; \pi, \nu) &= \frac{\nu}{\pi - 1} + \frac{y}{\pi}, \\
\nabla_2(y; \pi, \nu) &= \ln(1-\pi) - \psi_0(\nu) + \psi_0(y+\nu),
\end{aligned}$$

where $\psi_0(\cdot)$ denotes the digamma function. The Fisher information is a 2×2 matrix with elemets given by

$$\begin{aligned}
\mathcal{I}_{11}(\pi, \nu)) &= \frac{\nu}{\pi(1-\pi)^2}, \\
\mathcal{I}_{12}(\pi, \nu)) = \mathcal{I}_{21}(\pi, \nu)) &= \frac{1}{1-\pi}, \\
\mathcal{I}_{22}(\pi, \nu)) &= \psi_1(\nu) - \mathrm{E}\left[\psi_1(y+\nu)\right].
\end{aligned}$$

where $\psi_1(\cdot)$ denotes the trigamma function. Note that the Fisher information for ν is not available in a closed form.

An alternative parametrization denoted as the *NB2 parametrization* is proposed in [2]. It is more suitable for modeling of count data as it has location parameter $\mu > 0$ equal to the expected value and scale parameter $\alpha > 0$ representing the degree of overdispersion. The probability mass function is given by

$$\mathrm{P}\left[Y = y | \mu, \alpha\right] = \frac{\Gamma\left(y + \frac{1}{\alpha}\right)}{\Gamma\left(y+1\right)\Gamma\left(\frac{1}{\alpha}\right)}\left(\frac{1}{1+\alpha\mu}\right)^{\frac{1}{\alpha}}\left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{y} \qquad \text{for } y = 0, 1, \ldots$$

The expected value and variance are respectively given by

$$\mathrm{E}[Y] = \mu, \qquad \mathrm{var}[Y] = \mu(1+\alpha\mu).$$

The score is a vector with 2 elements given by

$$\begin{aligned}
\nabla_1(y;\mu,\alpha) &= \frac{y-\mu}{\mu(1+\alpha\mu)}, \\
\nabla_2(y;\mu,\alpha) &= \frac{1}{\alpha^2}\ln(1+\alpha\mu) + \frac{y-\mu}{\alpha(1+\alpha\mu)} + \frac{1}{\alpha^2}\psi_0\left(\frac{1}{\alpha}\right) - \frac{1}{\alpha^2}\psi_0\left(y+\frac{1}{\alpha}\right).
\end{aligned}$$

The Fisher information is a 2×2 matrix with elemets given by

$$\begin{aligned}
\mathcal{I}_{11}(\mu,\alpha) &= \frac{1}{\mu(1+\alpha\mu)}, \\
\mathcal{I}_{12}(\mu,\alpha) &= \mathcal{I}_{21}(\mu,\alpha) = 0, \\
\mathcal{I}_{22}(\mu,\alpha) &= \frac{2}{\alpha^3}\ln(1+\alpha\mu) - \frac{\mu}{\alpha^2(1+\alpha\mu)} + \frac{2}{\alpha^3}\psi_0(\alpha^{-1}) + \frac{1}{\alpha^4}\psi_1(\alpha^{-1}) - \mathrm{E}\left[\frac{2}{\alpha^3}\psi_0(y+\alpha^{-1}) + \frac{1}{\alpha^4}\psi_1(y+\alpha^{-1})\right].
\end{aligned}$$

Both parametrizations are related as

$$\begin{aligned}
\pi &= \frac{\alpha\mu}{1+\alpha\mu}, & \nu &= \frac{1}{\alpha}, \\
\mu &= \frac{\pi\nu}{1-\pi}, & \alpha &= \frac{1}{\nu}.
\end{aligned}$$

4 Empirical Example of E. Coli Cases

As the first illustration, we utilize a time series of E. coli cases. Specifically, the time series represents the *weekly number of reported disease cases caused by Escherichia coli in the state of North Rhine-Westphalia (Germany) from January 2001 to May 2013, excluding cases of EHEC and HUS* [13]. The source of the data is [15]. The dataset is available in R package `tscount` as `ecoli` [12, 13].

The number of observations is 646, the unconditional mean is 20.33, the unconditional variance is 88.75 and the whole unconditional probability mass function is illustrated in Figure 1. The negative binomial distribution is clearly favored over the Poisson distribution due to significant overdispersion in data. The dependence on past values is illustrated in Figure 2. The estimated GAS model with various specification is presented in Table 1. The maximum likelihood is obtained among the candidates for the square-root scaling with either the WT or NB2 parametrization. However, the other specifications have almost the same log-likelihood and can be utilized as well.

Furthermore, we conduct a simulation study in order to investigate the effects of model misspecification. We simulate 1000 observations using the models given by Table 1 and estimate models with various parametrizations and scaling functions. We repeat this 30,000 times and report the mean differences in log-likehoods between the estimated models and the true model in Table 2. Differences between the two parametrizations are extremely small. Differences between the scaling functions are a bit higher but still small. Naturally, largest differences in log-likelihoods are in the case of the true inverse scaling when assuming unit scaling and vice versa. The square-root scaling offers a middle way.

5 Empirical Example of Robbery Charges

As the second illustration, we utilize a time series of robbery charges. Specifically, the time series represents the *monthly counts of charges laid in Local Courts in armed robbery in New South Wales from 1995–2007* [5]. The source of the data is [7]. The dataset is available in R package `glarma` as `RobberyConvict` [6, 5].

The number of observations is 150, the unconditional mean is 18.12, the unconditional variance is 81.41 and the whole unconditional probability mass function is illustrated in Figure 3. Again, the negative binomial distribution is appropriate for the data. The dependence on past values is illustrated in Figure 4. Table 3 contains the estimated coefficients with log-likehoods for various model specifications. The best model is based on the NB2 parametrization with the unit scaling while the others perform quite comparably.

We conduct a similar simulation study as in the previous section with differences in log-likelihoods reported in Table 4. The process is simulated 90,000 times. The results have the same implications as in the previous

empirical example. The differences in log-likelihoods are, however, more pronounced as the autocorrelation structure is stronger in this case.

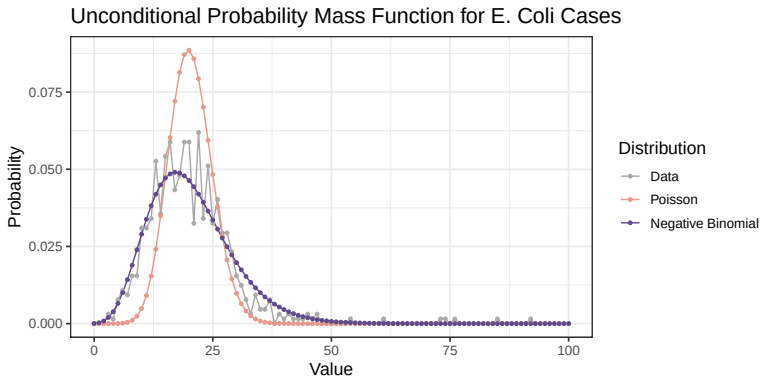


Figure 1 The unconditional probability mass function for the E. coli cases data.

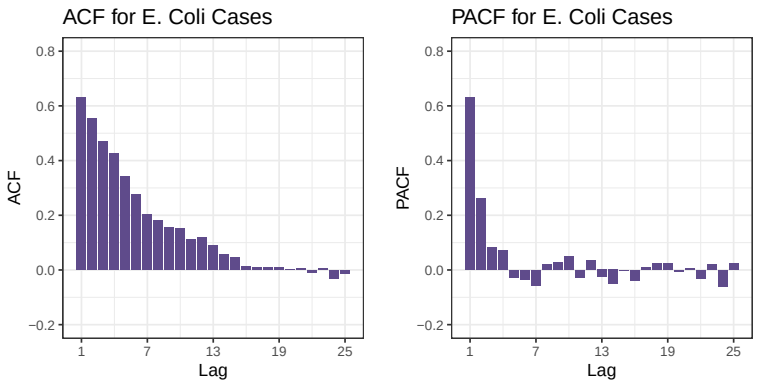


Figure 2 The autocorrelation and partial autocorrelation functions for the E. coli cases data.

Model	c	b	a	g	Log-Lik
WT/Unit	0.01	0.89	0.04	2.88	−2112.21
WT/Sqrt	0.01	0.89	0.11	2.87	−2112.14
WT/Inv	0.01	0.90	0.32	2.88	−2113.19
NB2/Unit	0.33	0.89	0.04	−2.86	−2112.49
NB2/Sqrt	0.31	0.89	0.11	−2.87	−2112.14
NB2/Inv	0.29	0.90	0.34	−2.87	−2112.27

Table 1 The estimated coefficients with the log-likelihoods for the E. coli cases data.

	Estimated Model					
True Model	WT/Unit	WT/Sqrt	WT/Inv	NB2/Unit	NB2/Sqrt	NB2/Inv
WT/Unit	0.00	−0.31	−1.17	−0.01	−0.31	−1.13
WT/Sqrt	−0.28	0.00	−0.26	−0.22	0.00	−0.31
WT/Inv	−0.97	−0.23	0.00	−0.92	−0.23	−0.07
NB2/Unit	−0.07	−0.28	−1.05	0.00	−0.28	−1.08
NB2/Sqrt	−0.28	0.00	−0.27	−0.21	0.00	−0.32
NB2/Inv	−1.11	−0.39	−0.16	−1.12	−0.22	0.00

Table 2 The differences in the log-likelihoods for various models based on the E. coli cases data.

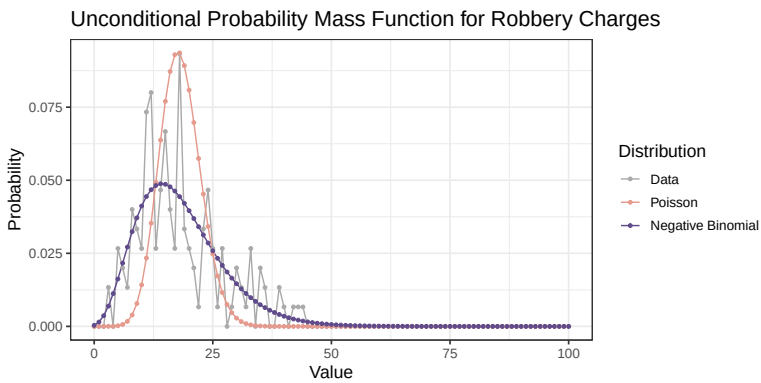


Figure 3 The unconditional probability mass function for the robbery charges data.

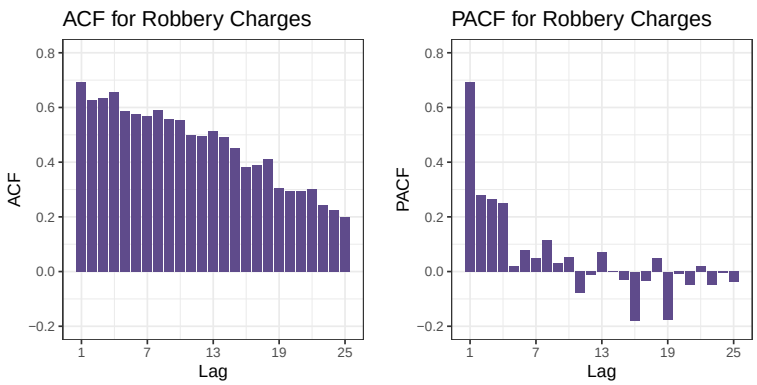


Figure 4 The autocorrelation and partial autocorrelation functions for the robbery charges data.

Model	c	b	a	g	Log-Lik
WT/Unit	−0.01	0.98	0.03	3.15	−465.76
WT/Sqrt	−0.01	0.98	0.08	3.16	−465.74
WT/Inv	−0.01	0.98	0.25	3.18	−465.75
NB2/Unit	0.05	0.98	0.02	−3.15	−465.68
NB2/Sqrt	0.05	0.98	0.08	−3.16	−465.74
NB2/Inv	0.04	0.98	0.26	−3.15	−465.88

Table 3 The estimated coefficients with the log-likelihoods for the robbery charges data.

	Estimated Model					
True Model	WT/Unit	WT/Sqrt	WT/Inv	NB2/Unit	NB2/Sqrt	NB2/Inv
WT/Unit	0.00	−1.57	−5.51	−1.81	−1.65	−2.38
WT/Sqrt	−0.84	0.00	−0.77	−0.04	−0.03	−0.18
WT/Inv	−2.39	−0.57	0.00	−0.84	−0.58	−0.48
NB2/Unit	−0.83	−0.10	−0.96	0.00	−0.13	−0.37
NB2/Sqrt	−0.82	0.00	−0.75	−0.02	0.00	−0.15
NB2/Inv	−1.09	0.00	−0.69	−0.37	−0.02	0.00

Table 4 The differences in the log-likelihoods for various models based on the robbery charges data.

6 Discussion and Conclusion

We compare the performance of various specifications of the score-driven models based on the negative binomial distribution with time-varying location parameter. The results are not surprising and in line with the GAS literature. The differences between the models based on the waiting time parametrization and the NB2 parametrization with the commonly used scaling functions are quite negligible in terms of the model fit. This allows to choose the parametrization with respect to its interpretability (e.g. the parametrization mirroring the moments of the distribution) and to choose the scaling function with respect to its theoretical properties (e.g. the square-root scaling for the unit variance of the scaled score) or computational performance (e.g. the unit scaling when the Fisher information is not available in a closed form and must be numerically approximated).

Acknowledgements

The work on this paper was supported by the grant No. F4/27/2020 of the Internal Grant Agency of University of Economics, Prague. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

References

[1] Blasques, F., Holý, V., and Tomanová, P.: Zero-Inflated Autoregressive Conditional Duration Model for Discrete Trade Durations with Excessive Zeros, 2020.

[2] Cameron, A. C., and Trivedi, P. K.: Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests. *Journal of Applied Econometrics* **1** (1986), 29–53.

[3] Cameron, A. C., and Trivedi, P. K.: *Regression Analysis of Count Data*. Second edition. Cambridge University Press, New York, 2013.

[4] Creal, D., Koopman, S. J., and Lucas, A.: Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics* **28** (2013), 777–795.

[5] Dunsmuir, W. T. M., Li, C., and Scott, D. J.: Package 'glarma', 2018. URL `https://cran.r-project.org/package=glarma`.

[6] Dunsmuir, W. T. M., and Scott, D. J.: The glarma Package for Observation-Driven Time Series Regression of Counts. *Journal of Statistical Software* **67** (2015), 1–36.

[7] Dunsmuir, W. T. M., and Weatherburn, D.: *Assessing the Impact of Mandatory DNA Testing of Prison Inmates in NSW on Clearance, Charge and Conviction Rates for Selected Crime Categories*. Technical report, 2008. URL `https://www.bocsar.nsw.gov.au/documents/l17.pdf`.

[8] Gorgi, P.: Integer-Valued Autoregressive Models with Survival Probability Driven By a Stochastic Recurrence Equation. *Journal of Time Series Analysis* **39** (2018), 150–171.

[9] Harvey, A. C.: *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. First edition. Cambridge University Press, New York, 2013.

[10] Hilbe, J. M.: *Negative Binomial Regression*. Second edition. Cambridge University Press, New York, 2011.

[11] Holý, V.: Score-Driven Count Time Series. In: *Proceedings of the 15th International Symposium on*

Operations Research in Slovenia. Slovenian Society Informatika, Bled, 399–404. URL http://fgg-web.fgg.uni-lj.si/ /sdrobne/sor/SOR'19 - Proceedings.pdf.

[12] Liboschik, T., Fokianos, K., and Fried, R.: tscount: An R Package for Analysis of Count Time Series Following Generalized Linear Models. *Journal of Statistical Software* **82** (2017), 1–51.

[13] Liboschik, T., Fried, R., Fokianos, K., Probst, P., and Rathjens, J.: Package 'tscount', 2020. URL https://cran.r-project.org/package=tscount.

[14] Lucas, A.: Generalized Autoregressive Score Models, 2020. URL http://www.gasmodel.com.

[15] Robert Koch Institute: SurvStat@RK, 2020. URL https://survstat.rki.de.

[16] Winkelmann, R.: *Econometric Analysis of Count Data*. First edition. Springer, Berlin, Heidelberg, 2008.

Use of the BCC and Range Directional DEA Models within an Efficiency Evaluation

Michal Houda[1]

Abstract. The contribution deals with two data envelopment analysis (DEA) models, in particular the BCC model (radial DEA model with variable returns to scale), and the range directional model. The mathematical description of the models are provided and several properties reported. A numerical comparison of the two models on real industrial data is provided with discussion about possible drawbacks of simplifying modeling procedures.

Keywords: Data Envelopment Analysis, BCC Model, Range Directional Model

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Question of measuring productive efficiency is classically based on the use of the Pareto efficiency notion to define the production function. Traditionally, the inefficiencies in input/output usage are neglected and total production growth is represented by shift in technologies. The pioneering works of Shephard [10] and Farrell [9] in measuring the productive efficiency using all inputs and outputs in order to prevent the index number problem (inadequacy of separate indices for labor and capital productivity), and in introducing the conceptual use of various types of efficiencies, were followed by a successful attempt to compute the productivity efficiency using a linear optimization model by Charnes, Cooper, and Rhodes [5]. Their non-parametric approach to estimate the production function as the efficiency frontier made up as the boundary of the convex hull of data points adopted the name of *Data Envelopment Analysis* (DEA) and spread around the scientific world quickly.

The classical CCR DEA model implicitly assumes constant returns-to-scale and continuous linear production possibility set. Many extensions to this original model were adopted, e. g. the widely used variable returns-to-scale model [1], discrete and continuous additive models [2], [4], slack-based measure models [11], or stochastic extensions [6]. We refer the reader to the monographs [8] and [7] for further information.

In our contribution, we want to point out that the classical DEA models are not able to work with negative data. We want emphasize, in particular, that units with negative data cannot be simply dropped from investigation as the results will become distorted. In Section 2, we provide definitions of basic notions and several DEA models used for comparison. Section 3.2 present some insights to numerical results based on real dataset, discussed and concluded in Section 4.

2 DEA Models

2.1 Production Possibility Set

Consider K decision-making units denoted $DMU_k,\ k = 1, \dots, K$. Each unit k is characterised by a collection of m inputs $x_{ik},\ i = 1, \dots, m$ and outputs $y_{jk},\ j = 1, \dots, n$; the input and output matrix are then denoted $X = (x_{ik})$ and $Y = (y_{jk})$, respectively. For the sake of convenience, let also denote $x_k = (x_{1k}, \dots, x_{mk})^T$ and $y_k = (y_{1k}, \dots, y_{nk})^T$ the input and output vectors of DMU_k. A unit under actual investigation (for which the efficiency is evaluated) is denoted DMU_0 thorough the paper.

The set of all combinations of allowed input and output vectors are known as *production possibility set* (*PPS*) and its correct specification plays an elementary role in DEA analysis. In general, *PPS* is defined through

$$PPS = \{(x, y) \mid y \text{ may be produced from } x\}\,. \tag{1}$$

A unit may be then characterized as efficient in Pareto–Koopmans dominance sense with respect to such defined production possibility set:

[1] University of South Bohemia in České Budějovice, Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, České Budějovice, and Czech Academy of Sciences, Institute of Information Theory, Pod Vodárenskou věží 4, Praha 8, Czech Republic, houda@ef.jcu.cz

Definition 1. *DMU_1 dominates DMU_2* if $x_1 \leq x_2, y_1 \geq y_2$, and at least one (one-dimensional) inequality is strict.

Definition 2. *DMU_0* is *efficient* with respect to *PPS* if there is no (real or virtual) unit with $(x,y) \in PPS$ dominating DMU_0.

2.2 BCC Model

Giving an example, Banker, Charnes, and Cooper's (BCC) model [1] assuming variable returns to scale is related to the following *continuous convex PPS*

$$PPS_C = \left\{ (x,y) \mid x \leq X\lambda,\ y \geq Y\lambda,\ \sum_k \lambda_k = 1,\ \lambda \geq 0 \right\} \tag{2}$$

It is not hard to see that PPS_C represents the convex hull of all available input-output data points. Let s^- and s^+ be the slack (surplus) for the inequalities $X\lambda \leq \theta x_0$ and $Y\lambda \geq y_0$ with some $\theta \in \mathbb{R}$. The efficiency of DMU_0 with respect to PPS_C may verified solving the *BCC input oriented* model (in envelopment form)

$$\begin{aligned} \min \theta + \epsilon \left(\sum_i s_i^+ + \sum_j s_j^+ \right) \text{ subject to} \\ X\lambda + s^- = \theta x_0 \\ Y\lambda - s^+ = y_0 \\ \sum_k \lambda_k = 1, \lambda \geq 0, s^-, s^+ \geq 0,\ \theta \text{ unconstrained}, \end{aligned} \tag{3}$$

where ϵ is so-called non-Archimedean infinitesimal (a positive number smaller than any other positive number). DMU_0 is efficient with respect to PPS_C if the optimal solution to (3) has $\theta^* = 1$ and $s^{-*} = s^{+*} = 0$, implying that DMU_0 is lying on the boundary of PPS_C and is (in fact) an extreme point of it.

Remark 1. Although the model (3) is not linear in principle due to the infinitesimal part, it may be solved by two-stage procedure with two linear optimization problems, see e. g. [8]. The infinitesimal part is present to ensure that some boundary points with $\theta^* = 1$ but non-zero slacks (also known as weakly efficient) are excluded from the final optimal solution.

2.3 Directional Distance Model

A single factor θ in (3) ensures that all the inputs are improved proportionally when projecting onto the efficient frontier. This is the property that characterizes so-called *radial* DEA models: the actual efficiency of the unit is furthermore a proportion of input values of DMU_0 and its peer unit. This implies a main drawback of the presented model: it may work well only if the data—the matrices X and Y—contain only positive elements. With negative data, the efficient measure is not well defined and the model cannot give appropriate results.

To generalize the radial feature and overcome the inability to work with negative data, Chambers, Chung, and Färe [3] proposed a so-called *directional distance model*:

$$\begin{aligned} \max \beta \text{ subject to} \\ X\lambda \leq x_0 - \beta g^x \\ Y\lambda \geq y_0 + \beta g^y \\ \sum_k \lambda_k = 1, \lambda \geq 0,\ \beta \geq 0 \end{aligned} \tag{4}$$

where g^x and g^y are pre-specified vectors of improvement directions and β is called *directional distance measure*. DMU_0 is efficient with respect to PPS_C if $\beta^* = 0$ in (4). The input oriented BCC model is then seen as a special case of (4) with $g^x = x_0, g^y = 0$ and $\theta = 1 - \beta$.

2.4 Range Directional Model

The free choice of possible improvement directions g^x and g^y gives the decision maker a great flexibility to represent many particular kinds of preference strengths inside input and output vectors. On the other

hand, this may represent a shortcoming too, especially in the case of missing or limited information about the nature of the data (the input/output improvement directions should be given in advance). This potentially unwanted freedom may be solved by choosing a particular reference point I (sometimes called "ideal point") with respect to which the improvements are considered. A frequent choice for such point is $I = (\min_k x_{\cdot k}; \max_k y_{\cdot k})$ (the minimums of inputs and maximums of outputs are taken component-wise in the above notation). The improvement directions are then defined by

$$g^x = x_0 - \min_k x_{\cdot k},$$
$$g^y = \max_k y_{\cdot k} - y_0,$$

and the corresponding directional distance model, called the *range directional model*, takes the form

$$\begin{aligned} &\max \beta \text{ subject to} \\ &\quad X\lambda \leq (1-\beta)x_0 + \beta \min_k x_{\cdot k} \\ &\quad Y\lambda \geq (1-\beta)y_0 + \beta \max_k y_{\cdot k} \\ &\sum_k \lambda_k = 1, \lambda \geq 0, \beta \geq 0 \end{aligned} \tag{5}$$

Again, DMU_0 is efficient with respect to *PPS* if the optimal solution of (5) is $\beta^* = 0$. (Adaptation of the two-stage technique to deal with non-efficient boundary points is straightforward and we will not provide additional details here.) The range directional model is still radial but the reference point is now I and not the origin as in the BCC model. Furthermore, the particular choice of I ensures that the improvement directions are meaningfully defined for all kind of the data and there is no need to restrict the DEA investigation to nonnegative inputs and outputs only.

3 Numerical Illustration

In this section we will compare two particular DEA models in order to point out the danger of bad model specification when working with real economic data and problems.

3.1 Example Setting

We have considered annual accounts of 380 Czech companies from the food industry (NACE C.10) from the year 2014. For the purpose of this paper, wee have chosen to evaluate the companies using

1. the input oriented model with variable returns to scale (BCC model), and
2. the range directional model (RD model).

The analysis was based on four inputs: *SPMAAEN* (material and energy consumption), *ON* (personnel costs), *STALAA* (fixed assets), and *POSN* (percentage of personnel costs); and two outputs *VYKONY* (business performance), and *ROA* (return on assets). Among 380 companies, 89 reported no material and energy costs and were removed from investigation. Further, 70 companies have negative *ROA* and cannot be analysed using BCC model. This resulted into 244 feasible observations for the BCC model and 291 feasible observation for the RD model.

3.2 Numerical Results

Among the 244 companies, 22 of them (9%) were evaluated as BCC efficient; additional three companies have the efficiency score higher than 95%. The alternative RD model evaluated only 10 companies (3.4%) to be efficient. The histograms of efficiency scores for both models are given in Figure 1 (BCC model is on the left hand side, RD model on the right hand side of the figure).

Furthermore, only four from nine NACE subgroups comprise an RD efficient company, while the distribution of BCC efficient companies is more widespread. Figure 2 provides an additional insight into the relationship between the subgroup size and the relative number of efficient companies in the overall model.

Another interesting comparison of the two models was made for selected input and output of both models. Figure 3 provides the empirical distribution plot for the personnel cost percentage (variable *POSN*). Note, in particular, the noticeable difference in distribution of this input for efficient (cyan colored) units – units with small percentages were marked more often as efficient in BCC model. Similar plot is provided as Figure 4 for

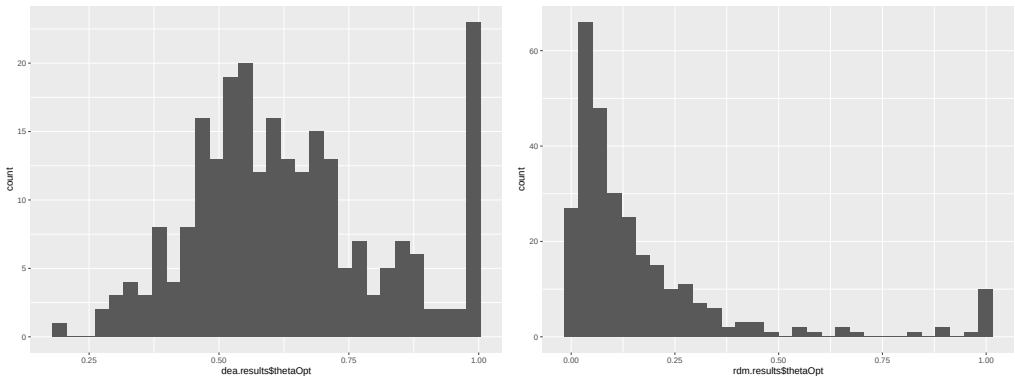


Figure 1 Efficiency Scores for BCC and RD Models

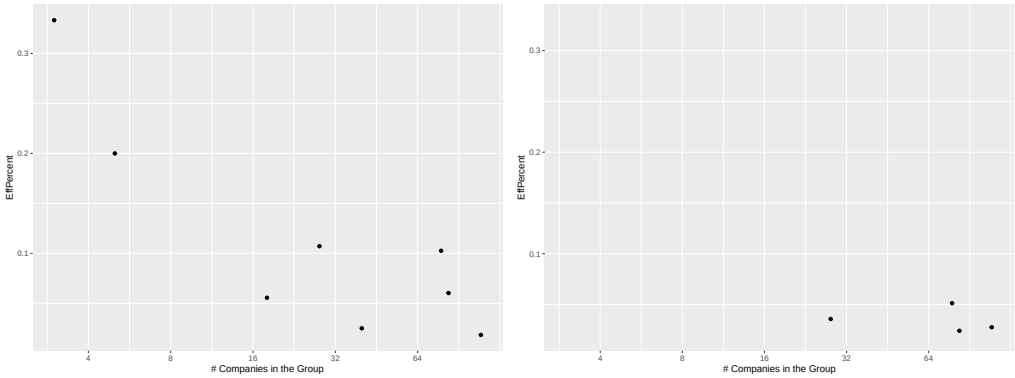


Figure 2 Relationship between the Size of the Group and the Relative Number of Efficient Units

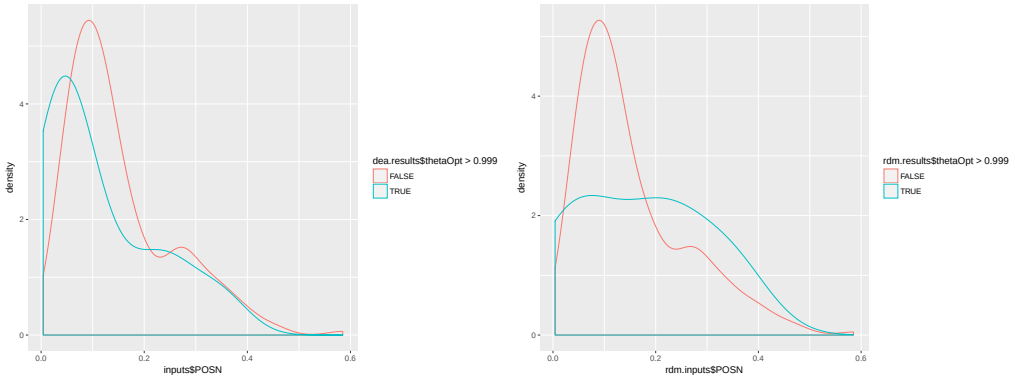


Figure 3 Empirical Distribution of Personnel Cost Percentage

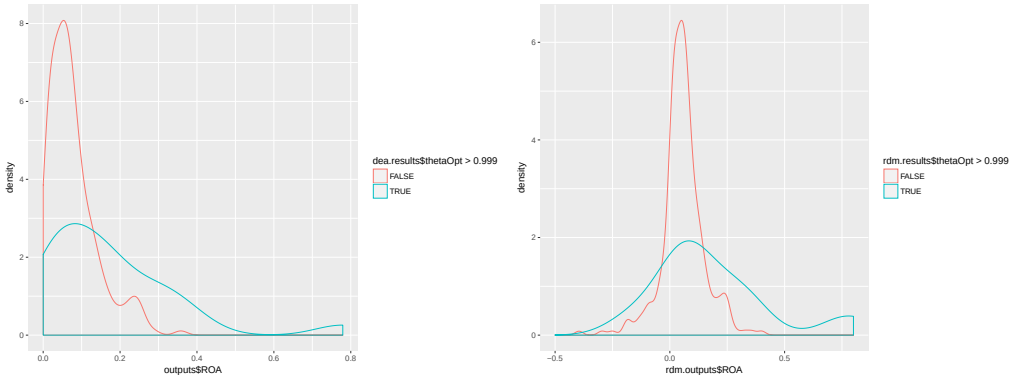


Figure 4 Empirical Distribution of Return on Assets

the return on assets (variable *ROA*). For the BCC model, all units with negative *ROA* must be excluded from the computations.

A scatter plot of a (partial) relationship of the input and output mentioned above is provided in Figure 5. Interestingly, one unit with negative *ROA* and average *POSN* was made efficient by RD model. Again, only the upper part of the right hand side plot (units with *ROA* above zero) is comparable with the left hand side plot.

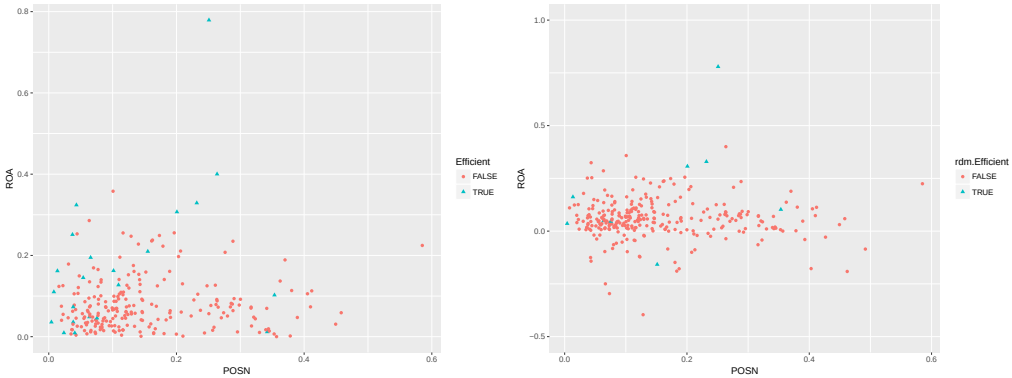


Figure 5 Relationship between the Personnel Cost Percentage and Return on Assets

4 Discussion and Conclusion

The comparison made in Section 3.2 clearly demonstrates that choice of a good model in production analysis is crucial. For example, the distribution of the optimal efficiency θ^* among the investigated units shows that BCC model (probably inadequately) overvalued efficiencies of the remaining units (with positive *ROA*). Notice that only a small part of the units exceeded the efficiency of 0.50 for the RD model. Furthermore, analysing Figure 5 one may notice a great number of BCC efficient units with relatively small *ROA* which were not marked as efficient by the RD model.

To conclude: simple deleting the observations which do not conform to the assumption of the chosen model (as done, for example, by the BCC case) ineluctably leads to false conclusions concerning efficiency of the remaining units and any subsequent results.

Acknowledgements

This paper was partially supported by the Czech Science Foundation This research was partially supported by the Czech Science Foundation under Grant No. GA18-02739S (Stochastic Optimization in Economic Processes).

References

[1] Banker, R. D., Charnes, A., and Cooper, W. W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* **30** (1984), 1078–1092.

[2] Bowlin, W. F., Brennan, J., Cooper, W. W., and Sueyoshi, T. *DEA models for evaluating efficiency dominance*. Research report, The University of Texas, Center for Cybernetic Studies, Austin, Texas, 1984.

[3] Chambers, R. G., Chung, Y. H., and Färe, R. Benefit and distance functions. *Journal of Economic Theory* **70** (1996), 407–419.

[4] Charnes, A., Cooper, W. W., Golany, B., Seiford, L. M., and Stutz, J. D. Foundations of data envelopment analysis for pareto-koopmans efficient empirical production functions. *Journal of Econometrics* **30** (1985), 91–107.

[5] Charnes, A., Cooper, W. W., and Rhodes, E. Measuring the efficiency of decision making units. *European Journal of Operational Research* **2** (1978), 429–444.

[6] Cooper, W. W., Deng, H., Huang, Z., and Li, S. X. Chance constrained programming approaches to technical efficiencies and inefficiencies in stochastic data envelopment analysis. *Journal of the Operational Research Society* **53** (2002), 1476–9360.

[7] Cooper, W. W., Seiford, L. M., and Joe, Z., eds. *Handbook on Data Envelopment Analysis*, volume 164 of *International Series in Operations Research & Management Science*. 2 edition. Springer, New York, 2011.

[8] Cooper, W. W., Seiford, L. M., and Tone, K. *Data Envelopment Analysis*. 2 edition. Springer, New York, 2007.

[9] Farrell, M. J. The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)* **120** (1957), 253–290.

[10] Shephard, R. W. *Cost and Production Functions*. Princeton University Press, Princeton, 1953.

[11] Tone, K. A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research* **130** (2001), 498–509.

Numerical Pricing of American Lookback Options with Continuous Sampling of the Maximum

Jiří Hozman[1], Tomáš Tichý[2]

Abstract. Exotic options whose payoff depends on the extrema of the underlying asset over a certain period of time form a class of the lookback options. Moreover, the American constraint admits early exercise and thus these contingent claims have become increasingly popular hedging and speculation instrument over recent years.
In this paper we restrict ourselves to the floating and fixed strike contracts with the continuously observed maximum only. Since no analytic formulae exist for this case, we follow a PDE approach. The corresponding American lookback option pricing problem leads to the parabolic partial differential inequality subject to a constraint, which can be handled by penalty techniques. As a result, we obtain an option pricing equation of the Black-Scholes type, where the path-dependent variable appears as a parameter only in the initial and boundary conditions.
The numerical approach proposed is based on the modification of the discontinuous Galerkin method incorporating a penalty term that handles the early-exercise constraint. The capabilities of the numerical scheme are demonstrated within a simple empirical study on the reference experiments.

Keywords: option pricing, discontinuous Galerkin method, Black–Scholes equation, American feature, lookback option, continuous sampling, numerical solution

JEL Classification: C44, G13
AMS Classification: 65M60, 35Q91, 91G60

1 Introduction

Decision-making of various market participants, such as households and firms, meets different constraints and needs. Some of them focus first of all on risk elimination, others look for specific return opportunities. The diversity of exotic options allows anyone to design a unique combination of risk and return. However, an inevitable condition for efficient usage of options is the knowledge of reliable valuation techniques.

It is thus not surprising that development of new pricing techniques as well as deep analysis of currently known approaches is in the center of interest of many researchers since the seminal paper on modern option pricing [2]. For example, as concerns lookback options, a special type of path-dependent options, payoff of which depends on the maximum or minimum asset price observed over some prespecified period of the option life, we can name [3], who provided explicit formulae (in specific cases) and bounds for option values, and [1], who focused on binomial approximation schemes; see also [14] with finite difference/element approaches.

In this short contribution we extend our previous results on the topic of numerical valuation of various option contracts, see [6], [7], [8] and [9]; and focus on developing of numerical scheme for valuation of American lookback options with continuous sampling of the maximum using discontinuous Galerkin (DG) method. We proceed as follows – in Section 2 lookback options with continuous sampling of the maximum are discussed, while in Section 3 a numerical valuation procedure is presented. Finally, in Section 4 a numerical experiments with call and put options are provided.

2 Lookback Options with Continuous Sampling of the Maximum

In this paper we concentrate on valuing a lookback option (with maturity time T) that depends on the maximum value M of the underlying asset S obtained by the continuous measurement on the whole time interval

[1] Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic, jiri.hozman@tul.cz
[2] Department of Finance, VSB-TU Ostrava, Sokolska 33, 701 21, Ostrava, Czech Republic, tomas.tichy@vsb.cz

$[0,t] \subset [0,T]$, i.e., $M(t) = \max_{0\le\tilde{t}\le t} S(\tilde{t})$, where t is the actual time. One can easily observe that for continuous sampling the asset price is necessarily less or equal to the maximum, i.e., $0 \le S \le M$. It is not true in case of the discrete measurement where the asset price S takes also values greater then M, cf. [14]. This is a very important difference between continuously and discretely sampled lookback options which is reflected on the (S,M)-domain on which the problem is posed.

Therefore, in general the option value $V = V(S,M,t)$ is a function of three variables S, M and t. The value V at maturity T is simply given by a payoff function. According to the way in which the variable M is incorporated into the payoff function, we distinguish a floating strike lookback put with payoff: $\max(M(T) - S(T), 0)$, and a fixed strike lookback call with payoff: $\max(M(T) - \mathcal{K}, 0)$, where $\mathcal{K}$ stands for a strike price. The first contract gives the holder the right to sell the underlying asset at the highest price observed over the lookback period. The second one is a call option on the maximum price with agreed price $\mathcal{K}$. For the remaining subtypes of lookback options (e.g., discrete monitoring or sampling of the minimum) we refer interested reader to the book [4].

In order to determine the option value V at arbitrary time instants $0 \le t < T$, we follow the theory of semimartingales [11] and this value is characterized as a solution of a deterministic governing equation or inequality (according to the European or American exercise features). Next, we introduce the Black-Scholes framework modified for the specified lookback case. Suppose that the price process $S(t)$ evolves over time according to the following stochastic differential equation:

$$\mathrm{d}S(t) = \mu S(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W(t), \quad S(0) > 0, \tag{1}$$

where $\mu S(t)\mathrm{d}t$ is a drift term with a constant rate $\mu \ge 0$, $W(t)$ is a standard Brownian motion and $\sigma > 0$ is the volatility of the asset price. Unfortunately, the measurement of maximum M is not differentiable and thus has to be approximated by a path-dependent quantity:

$$M_n(t) = \left(\int_0^t S(\tilde{t})^n \mathrm{d}\tilde{t}\right)^{1/n}, \quad \text{and} \quad M(t) = \lim_{n\to\infty} M_n(t) = \max_{0\le\tilde{t}\le t} S(\tilde{t}). \tag{2}$$

Then, the derivative of (2) satisfies

$$\mathrm{d}M_n(t) = \frac{1}{n}\frac{S(t)^n}{M_n(t)^{n-1}}\,\mathrm{d}t, \quad \text{and} \quad \lim_{n\to\infty} \mathrm{d}M_n(t) = 0. \tag{3}$$

Since $S(t) \le M(t)$ for all $t \in [0,T]$, the derivative (3) tends to zero as $n \to \infty$. In what follows we describe both situations for European- and American-style options.

European Case

Within the European feature the exercise of options is permitted only at maturity time T. We use the standard technique consisting of a construction of a hedged portfolio, an elimination of the stochastic components and a comparison of the portfolio dynamics by virtue of multidimensional Itô's lemma. Following these arguments together with (3), the price function of European lookback option V is the unique solution of the following backward partial differential equation

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0, \quad 0 < S < M < +\infty,\ 0 < t \le T. \tag{4}$$

where $r \ge 0$ is the risk-free interest rate. In fact, equation (4) is the classical Black-Scholes equation in the standard variables S and t. The second spatial variable M enters here only as a parameter, but it also features in the terminal condition

$$V(S,M,T) = V_T(S,M) := \begin{cases} \max(M-S,0), & \text{for a floating strike put,} \\ \max(M-\mathcal{K},0), & \text{for a fixed strike call.} \end{cases} \tag{5}$$

American Case

In contrast to the European-style option an American-style option can be exercised before the expiry of the contract. In this case, we have to encompass the additional constraint to the problem (4)–(5) that $V(S,M,t) \ge V_T(S,M)$ at any time $t \in [0,T]$. This American feature leads to a moving-boundary problem where, apart from solving the governing equation, it is also necessary to determine two regions separated by a free boundary

$\mathcal{E}$ driven by the optimal exercise price $S^*(M,t)$. Let $\Omega_{\mathrm{E}} \subset \{[S,M] \in (\mathbb{R}^+ \times \mathbb{R}^+) : S < M\}$ denote the exercise region. Since it is optimal to exercise the option early for $[S,M] \in \Omega_{\mathrm{E}}$, we solve the problem

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV < 0, \quad [S,M] \in \Omega_{\mathrm{E}},\ 0 < t \le T, \quad \text{with } V(S,M,t) = V_T(S,M). \tag{6}$$

While in the continuation region it is not optimal to exercise early, we solve the following problem

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0, \quad [S,M] \in (\mathbb{R}^+ \times \mathbb{R}^+) \backslash \overline{\Omega_{\mathrm{E}}},\ S < M,\ 0 < t \le T, \quad \text{with } V(S,M,t) > V_T(S,M). \tag{7}$$

Note that to guarantee the well-posedness of (6)–(7), it is enforced a continuity of the option value V and partial derivatives $\partial V/\partial S$ and $\partial V/\partial M$ on the free boundary $\mathcal{E}$, see [9].

There are several approaches how to handle the early-exercise feature, among the widely used ones let us cite the linear complementarity problem with penalty techniques [15] or operator splitting methods [10]. In this paper we follow the penalty approach and reformulate both problems (6) and (7) into one equation valid everywhere in both regions, i.e.,

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV + q = 0, \tag{8}$$

where q is defined to ensure American constraint $V(S,M,t) \ge V_T(S,M)$ and satisfy the conditions:

$$q(S,M,t) = 0, \text{ if } V(S,M,t) > V_T(S,M), \qquad q(S,M,t) > 0, \text{ if } V(S,M,t) = V_T(S,M). \tag{9}$$

This new quantity can be viewed as an additional nonlinear source term q in the governing equation and its essential role is to guarantee that the value of an American option cannot fall below its payoff function at any time t. The specific choice of q is discussed in Section 3.

3 Numerical Approach

Since there is no analytical option pricing formula for finite maturity American options under BS framework, the valuation should rely on numerical approaches. In our study, we employ the DG method, successfully used also in the field of financial engineering (see, e.g., [8] and [9]), that improves the valuation process for options. We proceed in three steps. At first, we change the time running and localize the option pricing problem to a bounded spatial domain. Next, we recall the variational form of the penalty term for the American constraint. Finally, we mention the standard discretization steps and present the numerical scheme.

3.1 Initial-boundary value problem

Since the governing equation (8) is accompanied by the particular payoff (5) prescribed at maturity T, it is suitable to use the forward time running from the numerical point of view. Setting $\hat{t} = T - t$ the time to maturity, we get $u(S,M,\hat{t}) = V(S,M,t)$ and $\hat{q}(S,M,\hat{t}) = q(S,M,t)$ as a new option price function and a new penalty term, respectively.

Further, the numerical approach, related to solving of the pricing equation (8), requires the restriction of the Cauchy problem to a bounded (S,M)-domain Ω. For this purpose let $S_{\max} > S^*$ and $M_{\max}$ denote the maximal sufficient value of the underlying asset and maximal possible value of its maximum, respectively. Without loss of generality $M_{\max} = S_{\max}$, i.e., we consider the upper triangular domain $\Omega := \{[S,M] \in (\mathbb{R}^+ \times \mathbb{R}^+) : S < M \wedge M < M_{\max}\}$. Consequently, the transformed governing equation with the initial condition, localized on the bounded domain, can be rewritten as

$$\frac{\partial u}{\partial \hat{t}} - \frac{\partial}{\partial S}\left(\frac{1}{2}\sigma^2 S^2 \frac{\partial u}{\partial S}\right) + \frac{\partial}{\partial S}\left((\sigma^2 - r)Su\right) + (2r - \sigma^2)u = \hat{q} \quad \text{in } \Omega \times (0,T), \tag{10}$$

$$u(S,M,0) = u_0(S,M) := V_T(S,M) \quad \text{in } \Omega. \tag{11}$$

Since the problem (10)–(11) is defined on the bounded domain Ω, we have to impose values of u on appropriate parts of boundary $\partial\Omega$. The prescribed values are chosen to be compatible with the payoff function, in accordance with the vector field determined by the characteristics of (10) and by the American constraints. Since the variable M is not present in the differential operator in (10), the convection does not propagate in the M-direction and thus no boundary condition has to be imposed on the boundary parallel to the S-axis

(i.e., $M = M_{\max}$). On line $S = 0$ the price of a floating put is given by American constraint $u(0, M, \hat{t}) = M$. In the case of a fixed call the American constraint $u(0, M, \hat{t}) = M - \mathcal{K}$ is enforced only for $M > \mathcal{K}$ and the remaining part $\{0\} \times (0, \mathcal{K})$ is considered as an outflow boundary. For the particular situation $S = M$, we can argue that the value of the lookback option for both cases should be insensitive to infinitesimal changes in M, i.e., $\frac{\partial u}{\partial M}(M, M, \hat{t}) = 0$, see [12].

The rigorous treatment of these boundary conditions plays an important role to achieve highly accurate solutions. A special interest here is due to the presence of a derivative boundary condition with respect to the parameter M. In contrast to [12], we propose here the weak enforcement of this homogeneous boundary condition that is incorporated in the discrete formulation by the following term

$$\frac{\sigma^2}{2} \int_\Gamma S^2 \left(\frac{\partial u}{\partial S} + \alpha \frac{\partial u}{\partial M} \right) n_S \, v \mathrm{d}s, \tag{12}$$

where n_S is the first component of outer unit vector to boundary $\Gamma = \{[S, M] \in (\mathbb{R}^+ \times \mathbb{R}^+) : S = M \wedge M < M_{\max}\}$ and $\alpha > 0$ is a suitably defined large number that represents a weight with which this boundary condition is enforced. The term (12) with $\alpha = 0$ naturally arises from the application of Green's theorem in variational formulation. Without any specifications, we assume that functions u and v are sufficiently regular to ensure well-posedness of (12).

Finally, let us mention that relations (10)–(11) accompanied with proper boundary conditions pose the initial-boundary value problem, which is closely related to the class of convection-diffusion problems. Although the governing equation (10) has no explicit dependency on the variable M, it is still present in the initial-boundary value problem due to initial and boundary conditions. Therefore, the proposed numerical schemes for solving such problems have to take these properties into account.

3.2 Penalty method

In order to handle the American early exercise feature and force the solution of (10) to be equal to the payoff in the exercise region Ω_{E}, we were inspired by [15] and introduce, for a sufficiently regular function v, the variational form of penalty term $\hat{q}$ as

$$\left(\hat{q}(\hat{t}), v\right) = c_p \int_\Omega \chi_{\mathrm{exe}}(\hat{t}) \left(u_0 - u(\hat{t})\right) v \, \mathrm{d}S \, \mathrm{d}M = \underbrace{c_p \int_\Omega \chi_{\mathrm{exe}}(\hat{t}) u_0 v \, \mathrm{d}S \, \mathrm{d}M}_{\mathcal{Q}_{\mathrm{R}}(v)} - \underbrace{c_p \int_\Omega \chi_{\mathrm{exe}}(\hat{t}) u(\hat{t}) v \, \mathrm{d}S \, \mathrm{d}M}_{\mathcal{Q}_{\mathrm{L}}(u,v)}, \tag{13}$$

where $(\cdot, \cdot)$ denotes the inner product in $L^2(\Omega)$. The function $\chi_{\mathrm{exe}}(\hat{t})$ in (13) is defined as an indicator function of the region Ω_{E} at time instant $\hat{t}$ and $c_p > 0$ represents a weight to enforce the early exercise. In line with [8], we set c_p proportional to $1/\tau$, where τ is the time step introduced in (16). The form (13) can be split into linear functional $\mathcal{Q}_{\mathrm{R}}$ and bilinear form $\mathcal{Q}_{\mathrm{L}}$, and we place them on opposite sides of the variational formulation of (10), see (14).

3.3 Discontinuous Galerkin scheme

We present a numerical scheme based on a simple modification of the DG method (see [13] for a complete overview) that extends the lookback option pricing approach from [7] to numerical pricing of American style options using penalty technique. Our aim is to construct solution $u_h = u_h(\hat{t})$ from the finite dimensional space S_h^p consisting from piecewise polynomial, generally discontinuous, functions of the p-th order defined over the partition $\mathcal{T}_h$ of the domain Ω with the assigned mesh size h.

This DG discretization in spatial coordinates leads to a system of the ordinary differential equations for unknown price function u_h, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}\hat{t}} (u_h, v_h) + \mathcal{A}_h(u_h, v_h) + \mathcal{Q}_h(u_h, v_h) = \ell_h(v_h)(\hat{t}) + q_h(v_h)(\hat{t}) \quad \forall v_h \in S_h^p, \forall \hat{t} \in (0, T), \tag{14}$$

where the initial condition $u_h(0)$ is given by (11), the bilinear forms $\mathcal{A}_h(\cdot, \cdot)$ and $\mathcal{Q}_h(\cdot, \cdot)$ stand for the discrete variants of the spatial partial differential operator from (10) and form $\mathcal{Q}_{\mathrm{L}}$ from (13), respectively. Further, the term $\ell_h(\cdot)(\hat{t})$ arises from boundary conditions and $q_h(\cdot)(\hat{t})$ is given by $\mathcal{Q}_{\mathrm{R}}$ from (13). For the detailed derivation of the above-mentioned forms we refer the interested reader to [9].

Next, we introduce the temporal discretization of (14) that is realized by implicit Euler method over the uniform partition of the interval $[0, T]$ with time step τ. Denote $u_h^m \in S_h^p$ the approximation of the solution

$u_h(\hat{t})$ at time level $\hat{t}_m \in [0, T]$. Moreover, for practical purpose, to evaluate forms $\mathcal{Q}_h$ and q_h we use an element-wise approximation of the early exercise region as

$$\chi_{\text{exe}}(\hat{t}_m)\big|_K \approx \widetilde{\chi_{\text{exe}}}(\hat{t}_m)\big|_K := \begin{cases} 1, & \text{if } u_h^{m-1}(S_c, M_c) < u_0(S_c, M_c), \\ 0, & \text{if } u_h^{m-1}(S_c, M_c) \geq u_0(S_c, M_c), \end{cases}, \quad \hat{t}_m \in [0, T],\ K \in \mathcal{T}_h, \tag{15}$$

where $[S_c, M_c]$ denotes a barycenter of the element K. Let $u_h^0 \approx u_0$ be the initial state, then the discrete solutions u_h^m, $m \geq 1$, are computed within the DG framework by the recurrence scheme

$$\left(u_h^{m+1}, v_h\right) + \tau \mathcal{A}_h(u_h^{m+1}, v_h) + \tau \mathcal{Q}_h(u_h^{m+1}, v_h) = (u_h^m, v_h) + \tau \ell_h(v_h)(\hat{t}_{m+1}) + \tau q_h(v_h)(\hat{t}_{m+1}) \quad \forall\, v_h \in S_h^p. \tag{16}$$

Finally, note that the equation (16) results into a sequence of systems of linear algebraic equations with sparse matrices. The solvability of a such system is proven in [8]. Since the system matrix is non-symmetric the restarted GMRES solver is incorporated into the numerical procedure.

4 Reference Numerical Experiments

In this section, we present numerical experiments on two benchmarks widely referred in the literature in order to verify the validity of the proposed numerical scheme and illustrate its capabilities. All computations are carried out with an algorithm implemented in the solver Freefem++; for more details to a mesh generation/adaptation, the spatial and temporal discretization, assembly of a linear algebraic problem, and its solving, see [5].

Within the first experiment we examine a newly issued half-year American lookback put option with floating strike. As in [1] and [3] we consider the following model parameters: $T = 0.5, r = 0.1, \sigma = 0.2, S_{\text{ref}} = M_{\text{ref}} = 100, S_{\max} = M_{\max} = 2S_{\text{ref}}$, where S_{ref} and M_{ref} determine the initial price and the current maximum.

For the purpose of a broader illustration of convergence properties and enforcement of the boundary condition on line $S = M$, we compute the piecewise linear solutions on a sequence of the adaptively generated grids with the prescribed fixed refinement $n_D \in \{125, 250, 500, 1000, 2000, 3000\}$ on line $S = M$. In parallel with this, we take values $\alpha \in \{10\sqrt{2}, 50\sqrt{2}, 100\sqrt{2}, 500\sqrt{2}, 1000\sqrt{2}\}$. The factor $\sqrt{2}$ in α is only artificial to compensate component sizes of a normal vector in (12), because $n_S = 1/\sqrt{2}$ on line $S = M$. For all scenarios, we assume that there are 360 days in a year, take time step proportional to a half-day and the American early exercise feature is handled with $c_p = 10/\tau$ in (13).

The comparative results evaluated at a given reference node $[S_{\text{ref}}, M_{\text{ref}}, T]$ are presented in Table 1. We also appends values of European-style options under the same market conditions (see values in brackets). Without surprise, one can easily observe that the proposed approach gives promising results that asymptotically tend (as n_D and α increase) to reference values from [1] obtained by the binomial method. Moreover, Table 1 illustrates the typical findings of American-style options resulting from the given constraint, i.e., American options cost more than their European counterparts.

α \ n_D	125	250	500	1000	2000	3000	ref. val. [1]
$10\sqrt{2}$	11.1981 (10.1210)	10.5054 (9.5204)	10.1798 (9.2893)	10.0105 (9.1482)	9.9253 (9.0775)	9.8982 (9.0537)	
$50\sqrt{2}$	11.5772 (10.4493)	10.8039 (9.7858)	10.4414 (9.5321)	10.2567 (9.3770)	10.1604 (9.2997)	10.1304 (9.2735)	
$100\sqrt{2}$	11.6272 (10.4941)	10.8447 (9.8219)	20.4769 (9.5651)	10.2916 (9.4082)	10.1923 (9.3299)	10.1619 (9.3034)	10.17 (9.29)
$500\sqrt{2}$	11.6801 (10.5283)	10.8781 (9.8513)	10.5059 (9.5903)	10.3238 (9.4319)	10.2365 (9.3543)	10.1881 (9.3331)	
$1000\sqrt{2}$	11.7047 (10.5353)	10.8834 (9.8550)	10.5096 (9.5954)	10.3281 (9.4367)	10.2211 (9.3599)	10.1908 (9.3354)	

Table 1 Floating strike lookback put: Comparison between American options (upper values in cells) and European options (bottom values in brackets) evaluated at reference node $[S_{\text{ref}}, M_{\text{ref}}, T]$.

Secondly, we investigate the behaviour of three-month American lookback call option with fixed strike. The experiment data come from [3], where the lower and upper bounds for option values were derived. We

consider $T = 0.25$, $r = 0.1$, $\sigma = 0.2$ and $\mathcal{K} \in \{95, 100, 105\}$. The reference node coordinates and sizes of computational domain Ω are the same as in the preceding experiment.

For fixed $\alpha = 1000\sqrt{2}$ (taking larger values does not improve the results significantly), we compute piecewise linear approximation of option prices on a sequence of consecutively refined meshes with a quarter-day time step. In Table 2 we compare obtained results (evaluated at $[S_{\text{ref}}, M_{\text{ref}}, T]$) with bounds from [3] for three scenarios with different strike prices. One can again observe that the obtained results are of higher accuracy and match better the reference bounds as the computational grid is finer. From this point of view, thus the presented DG approach shows one part of its promising potential.

$\mathcal{K}$ \ n_{D}	125	250	500	1000	2000	3000	bounds [3]
95	15.7917	15.0386	14.6603	14.4929	14.4053	14.3756	14.32 – 14.68
100	11.0066	10.0720	9.7593	9.6021	9.5206	9.4911	9.45 – 9.69
105	6.3448	5.8193	5.5734	5.4519	5.3851	5.3697	5.34 – 5.47

Table 2 Fixed strike lookback call: Comparison of American option values at reference node $[S_{\text{ref}}, M_{\text{ref}}, T]$ for various consecutively refined grids and strike prices.

5 Conclusion

Pricing of options is very challenging and no less important part of financial engineering. In this paper we have presented numerical scheme based on the DG approach for pricing of continuously sampled American lookback options with floating as well as fixed strike. Experimental results show quite good similarity to selected benchmarks. Moreover, when the procedure is combined with our previous results presented in [7], one can easily price either discretely or continuously sampled lookback options, depending on actual needs – while in some cases, discrete monitoring is proper, in others continuous sampling might be more relevant. Obviously, the DG approximations presented in the paper can be relatively easily extended to other classes of path-dependent options with different complexity of payoff functions.

Acknowledgements

The first author acknowledges support provided within the project No. PURE-2020-4003 financed by TU Liberec. The second author was supported through the Czech Science Foundation (GACR) under project 18-13951S and moreover by SP2020/11, an SGS research project of VSB-TU Ostrava.

References

[1] Babbs, S. (2000). Binomial valuation of lookback options. *Journal of Economics Dynamics & Control*, 24, 1499–1525.

[2] Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81, 637–659.

[3] Conze, A. & Viswanathan, R. (1991). Path-dependent options: The case of lookback options. *Journal of Finance*, 46, 1893–1907.

[4] Haug, E.G. (2007). *The Complete Guide to Option Pricing Formulas*. New York: McGraw-Hill.

[5] Hecht, F. (2012). New development in FreeFem++. *Journal of Numerical Mathematics*, 20, 251–265.

[6] Hozman, J., et al. (2018). *Robust Numerical Schemes for Pricing of Selected Options*. SAEI, Vol. 57., Ostrava: VŠB-TU Ostrava.

[7] Hozman, J. & Tichý, T. (2017). The Valuation of Discretely Sampled European Lookback Options: A DG Approach. In P. Pražák (Eds.), *Proceedings of the 35th International Conference Mathematical Methods in Economics* (pp. 242–247). Hradec Králové: University of Hradec Králové.

[8] Hozman, J. & Tichý, T. (2018). DG framework for pricing European options under one-factor stochastic volatility models. *Journal of Computational and Applied Mathematics*, 344, 585–600.

[9] Hozman, J. & Tichý, T. (2020). The discontinuous Galerkin method for discretely observed Asian options. *Mathematical Methods in the Applied Sciences*, doi: 10.1002/mma.6160.

[10] Ikonen, S. & Toivanen, J. (2004). Operator splitting methods for American option pricing. *Applied Mathematics Letters*, 17, 809–814.

[11] Jacod, J. & Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes*. Berlin: Springer.

[12] Kwok, Y.-K. (2008). *Mathematical Models of Financial Derivatives*. Berlin: Springer.

[13] Rivière, B. (2008). *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*. SIAM, Philadelphia.

[14] Willmott, P. & Dewynne, J. & Howison, J. (1993). *Option pricing: mathematical models and computations*. Oxford: Oxford financial press.

[15] Zvan, R. & Forsyth, P.A. & Vetzal, K.R. (1998). Penalty methods for American options with stochastic volatility. *Journal of Computational and Applied Mathematics*, 91, 199–218.

Spatial Analysis of the Flat Market in Prague

Petr Hrobař[1], Vladimír Holý[2]

Abstract. Our study aims to examine the effect of spatial dependency in the Prague flat market. Following the Tobler's first law of geography: near things are more related than distant things, we allow the flat's price to be not only the function of its own characteristics but also function of its neighbourhood unit characteristics. We also evaluate different spatial dependence matrices for analysis of estimated results. The findings are as follows. First, based on the positive parameter of spatial autocorrelation, we confirm spatial dependency in our dataset, i.e. prices of flats tend to form spatial clusters. Second, once controlling for spatial dependency, we are able to evaluate flat's prices more accurately. Third, based on the residual distribution across space, we identify "grandiose" clusters in which the price of estate can be multiple times higher than outside the cluster simply due to the location factor.

Keywords: spatial econometrics, spatial autocorrelation, Moran test, real estate

JEL Classification: C33, C38, R31

1 Introduction

Prices of flats or generally any estates tend to form spatial clusters. As a result of this process, the location characteristics are key determinants to accurate evaluation of estates. When some form of spatial autocorrelation occurs in data generating process (DGP), using ordinary least square regression (OLS) can lead to inaccurate results and estimation bias.

Spatial econometric framework allows us to account for various spatial processes and provide much more accurate analysis and estate evaluation. Even though usual spatial analyses are applied to regional macroeconomic data, e.g., regional unemployment rates, GDPs, etc., applications to non-regional data can also be utilized. Many studies contributing to hedonic price modeling of flats have been conducted, e.g. [1]. To our best knowledge, latest study that contributes to spatial analysis of flat market in Prague, using spatial econometrics framework, is study [2] from 2016. We follow and expand the very study by applying different set of regressors, by using contemporary data and by evaluating spatial stability and estimation robustness for multiple structures of spatial dependency.

Our study is structured as follows: Section 2 describes all models and methods used in this study as well as fundamental literature sources, Section 3 discusses the dataset and the selected methodology and provides an illustrative examples of spatial methods applications. Section 4 shows illustrative evaluation of spatial stability. Lastly, Section 5 concludes our study.

2 Methods for Cross-Sectional Data

In this section we briefly describe non-spatial and spatial models used in our study. We provide fundamental description of methods and used models as well as references to corresponding literature.

2.1 Linear Regression

First model used in our study is simple linear regression given as

$$y = X\beta + \varepsilon, \tag{1}$$

where y is a $(n \times 1)$ vector of dependent variable, X represents a $(n \times k+1)$ matrix of exogenous regressors, β is a $(k+1 \times 1)$ vector of regression coefficients to be estimated and ε represents vector of error elements. The most used estimation method is the ordinary least square (OLS) method. For a quick recapitulation of the OLS method as well as all statistical assumptions for errors ε, see eg. [5].

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, hrop00@vse.cz

[2] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, vladimir.holy@vse.cz

Model 1 does not take spatial dependency into account. In order to partially allow for spatial dependency, we propose simple feature based on statistical learning clustering algorithms. Since we are working with spatial data, we have coordinates available although we do not recommend including coordinates into linear model directly.

However, various types of clustering methods can be applied. We use widely known *k-means* clustering method (see e.g. [3]). Then we include dummy variables representing each cluster into linear regression model. Assuming that all/some dummy variable/s representing cluster are statistically significant, we can conclude that there are some spatial dissimiarities present in the DGP. Since we are using both variants of the linear regression model – with and without dummy cluster variables, we shall be able to evaluate efficiency of both models.

2.2 Spatial Models

As mentioned in [5] the standard approach is to start by using no spatial models at all, then considering various types of spatial interactions. Generally, three main different types of spatial interactions can arise:

1. Dependent variable y of unit i affects/is affected by dependent variable y of unit j.
2. Dependent variable y of unit i affects/is affected by independent variable x of unit j.
3. Error term ϵ_i of unit i affects/is affected by ϵ_j of unit j.

Using a model to cover all interaction effects above, we get model that takes form

$$\begin{aligned} y &= \rho Wy + X\beta + WX\theta + u, \\ u &= \lambda Wu + \epsilon, \end{aligned} \tag{2}$$

where y is $(n \times 1)$ vector of dependent variable, W is a $(n \times n)$ weight matrix that represent a spatial dependency among observed units, X is a $(n \times k + 1)$ matrix of independent variables, β is a $(k + 1 \times 1)$ vector of regression coefficient to be estimated and ρ, θ, λ are spatial autocorrelation parametrs to be estimated. Wy denotes the endogenous interactions among dependent variable. Similarly, WX represents interaction effect among independent variables. Lastly, Wu stands for interaction effect among error terms.

For the purpose of our analysis we shall not use entire model as described above but rather use two derived specifications described down below.

Spatial Lag Model

Spatial lag model can be obtained from 2 when θ and λ are both equal to 0. Under this assumption derived model allows for accounting for spatial dependency in dependent variable of neighbourhood units. Therefore, model can be written as

$$y = \rho Wy + X\beta + \varepsilon. \tag{3}$$

Under the assumption of existing inverse of $(I_N - \rho W)$, we can obtain a reduced form given as

$$y = (I_N - \rho W)^{-1}(X\beta + \varepsilon), \tag{4}$$

where 4 is usually described as a DGP equation for y as emphasised by [4]. To obtain regression parameters β and ρ, the maximum likelihood approach is commonly used. Paper [4] offers an overview of spatial econometric models as well as detailed maximum likelihood (ML) estimation for model 3 and 5.

Spatial Error model

Second spatial model used in our study is spatial error model. Given the model 2 we can obtain spatial error model form when ρ and θ are both equal to 0. Hence, using this model specification we can account for spatial interactions among the error terms. Formal model form can be described as

$$\begin{aligned} y &= X\beta + u, \\ u &= \lambda Wu + \varepsilon. \end{aligned} \tag{5}$$

Selection for model 5 can imply that some (spatially distributed) independent variable was not included in the model. Once again, parameters β, λ are estimated by the ML, see e.g. [4]. Unlike the spatial lag model, coefficients from spatial error model can be directly interpreted as marginal effects.

2.3 Moran *I* test

Most importantly, the use of spatial model should always be confirmed, i.e. formal statistical test should be implemented. In Section 2.1 we propose simple clustering approach to (moderately) account for spatial dependency. However, more sophisticated tests are at disposal. The most common test is the Moran *I* test, see e.g. [4].

The formal testing statistic takes form given by

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \times \frac{\varepsilon' W \varepsilon}{\varepsilon' \varepsilon}, \tag{6}$$

where N is the number of (spatial) observations, w_{ij} are elements of spatial weight matrix W and ε is a $(n \times 1)$ vector of residuals obtained from the OLS model. Generally, the Moran I statistic value is between $[-1; 1]$. However, values outside the interval can also arise. Under the null of spatial independence, statistic take form as

$$E(I) = \frac{-1}{N-1}. \tag{7}$$

Not only is the Moran the most broadly used test for spatial dependency but also simulations showed that moran test is the most accurate test for spatial dependency [5].

Spatial dependency structure

Since W matrix is needed to perform tests for spatial dependency, it is necessary to correctly specify spatial W matrix and spatial dependency structure in general. Main obstacle in spatial framework is the fact that spatial dependency structure is not estimated but predefined and thus one needs to evaluate spatial dependency thoroughly.

Spatial Weight matrix W is derived from neighbor structure. As far as neighborhood dependency goes multiple approaches are common. Since we are not working with data on regional level we shall not consider methods based on polygons such as *queen and rook* methods.

Since we are evaluating flat estates we assume that estates that are more distant tend to effect prices slightly, but on the other hand estates closer to each other tend to influence one another more significantly. Thus using neighborhood dependency based on the number of neighborhoods units seems as a good starting point. Another approach considered in our study is method based on the maximum distance of units. However, this approach would be very computationally demanding and time consuming. Therefore we create neighborhood dependency based on the number of nearest neighbor units. This dependency is then transformed in W matrix in order to perform Moran test and estimate spatial models described in Section 2.2.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price	2,984	8,495,653.000	5,696,973.000	80,000	4,990,000	9,990,000	79,000,000
Rooms	2,984	2.804	1.128	1	2	4	6
Meters	2,984	79.634	40.968	15	52	96	435
KK	2,984	0.792	0.406	0	1	1	1
Maisonette	2,984	0.027	0.163	0	0	0	1
Panel	2,984	0.132	0.338	0	0	0	1
Balcony/Terrace	2,984	0.514	0.500	0	0	1	1
New Estate	2,984	0.291	0.454	0	0	1	1

Table 1 Descriptive statistics.

3 Dataset and Results

Flats estates are retrived from Czech estates site Sreality.cz which contains various estates that are available to rent or buy. We suppose that Prague flats listed here are credible representation of the real flats market of the city, i.e. the listed estates follow the same DGP as the estates not listed and thus not included in the dataset.

In our study, more than 4 000 flats profiles were collected data were collected over a one month period starting on March 20, 2020. The data collection process was done in PYTHON programming language using

webscrapping approach. After the data collection, a filtering processes were implemented to obtain credible dataset. Data are cross-sectional hence no individual effects nor time effects need to taken into account. First step was to select relevant variables for analysis. Thus, variables preserved for analysis are as follows: *Price* – price of estate, *Rooms* – number of rooms of flat, *Meters* – number of square meters of flat, *Maisonette* – an apartment occupying multiple floors *KK* – kitchenette. Also variables describing estate building type were retained, i.e. *Panel* – if an apartment is in a high-rise building, *Balcony/Terrace* – if an estate has balcony or terrace, and lastly, *New Estate* – if a flat is new-build property.

Additionally, since we are interested in spatial modeling, *coordinates*, i.e., *longitude and latitude*, were also kept. Last steps taken in data cleaning process were to delete all observations that do not have information about variables described above available and therefore were removed. Eventually, we have dataset that has 2 984 observations which should allow us for proper estimation and authentic results. Descriptive statistics are in Table 1.

Lastly, we investigate correlation between all independent variables. Particularly, we were concerned about correlation between *Meters* and *Rooms*. However, this correlations turns out not to be excessively high and we shall keep both variables in all models. We also considered whether variable *Rooms* should be dummy or not. We decided not to use dummy since we would be adding another five variables in model. However, we are fully aware that since variable *Rooms* is discrete we should not interpret associated coefficient, nevertheless coefficient sign and statistical significance can be interpreted. The final model is selected as

$$\begin{aligned}\ln(Price) = \beta_0 + \beta_1 Rooms + \beta_2 \ln(Meters) + \beta_3 Maisonette + \beta_4 KK + \beta_5 Panel \\ + \beta_6 Balcony/Terrace + \beta_7 NewEstate + \varepsilon.\end{aligned} \tag{8}$$

After the OLS model estimation we calculated Moran *I* statistic to identify whether spatial autocorrelation is present in flats prices. The results of Moran test are as follows: *I-statistic = 0.43, p-value = 0.000.*

Based on the significant results given by Moran test we shall apply spatial models. Interestingly enough, the result also shows that flats prices are not randomly distributed in space, i.e., prices of flats tend to form spatial clusters. Estimations of all presented models are summarized in Table 2. Model comparison metrics are in Table 3.

	Dependent variable:			
	log(price)			
		kmeans	Spatial lag	Spatial Error
	(1)	(2)	(3)	(4)
Rooms	0.061***	0.077***	0.084***	0.085***
	(0.010)	(0.009)	(0.008)	(0.007)
log(Meters)	0.808***	0.749***	0.685***	0.691***
	(0.021)	(0.019)	(0.017)	(0.017)
Maisonette	−0.003	−0.016	−0.062**	−0.065***
	(0.031)	(0.028)	(0.025)	(0.023)
KK	0.117***	0.155***	0.171***	0.178***
	(0.016)	(0.014)	(0.012)	(0.012)
Panel	−0.324***	−0.255***	−0.127***	−0.118***
	(0.016)	(0.015)	(0.013)	(0.014)
Balcony/Terrace	−0.007	0.039***	0.055***	0.075***
	(0.011)	(0.010)	(0.009)	(0.009)
New Estate	−0.011	0.016	0.060***	0.077***
	(0.011)	(0.011)	(0.009)	(0.010)
$Kmean_2$		0.344***		
		(0.019)		
Constant	12.145***	12.097***	2.432***	12.412***
	(0.065)	(0.062)	(0.223)	(0.075)
ρ			0.636***	
λ				0.933***
Observations	2,984	2,984	2,984	2,984

Note: *p<0.1; **p<0.05; ***p<0.01
Number of cluster (5) selected based on AIC.
We do not show additional 3 dummy coefficients for cluster.

Table 2 Estimated models.

	OLS	*OLS Kmeans*	*Spatial Error*	*Spatial Lag*
AIC	583.827	75.441	−931.300	−688.538
Log-like.	−282.913	−24.720	475.650	354.269
R^2	0.748	0.788	0.854	0.837

Table 3 Models metrics.

Model selection is be done using the metrics in Table 3. We also include pseudo R^2 calculated as $corr(y, \hat{y})^2$. Firstly, when comparing baseline OLS model and OLS model with clustering approach we can conclude that once (partially) controlling for spatial heterogeneity estimated results tend to be more steady. Additionally, model with clustering still allows for nice interpretability. For example, we can identify cluster in which prices of states are (on average) 34% higher (supposedly) due to unobserved location factor.

However, spatial models particularly spatial error model seems to be the best for DGP rendering. Therefore, spatial error model shall be used for statistical inference. Lastly, we tested error model residuals for homoscedasticity using the modified Breusch–Pagan test.

4 Model Specification Robustness

An important step is to evaluate model stability for various types of spatial dependency. Study describing importance of such step is e.g. [6]. Also, since we are using neighborhood dependency based on the number of nearest neighbor units we do not know the optimal number that should be used. Therefore, evaluation of spatial stability using information criteria should serve as a fair indicator. Results of this "spatial cross validation" are in Figure 1.

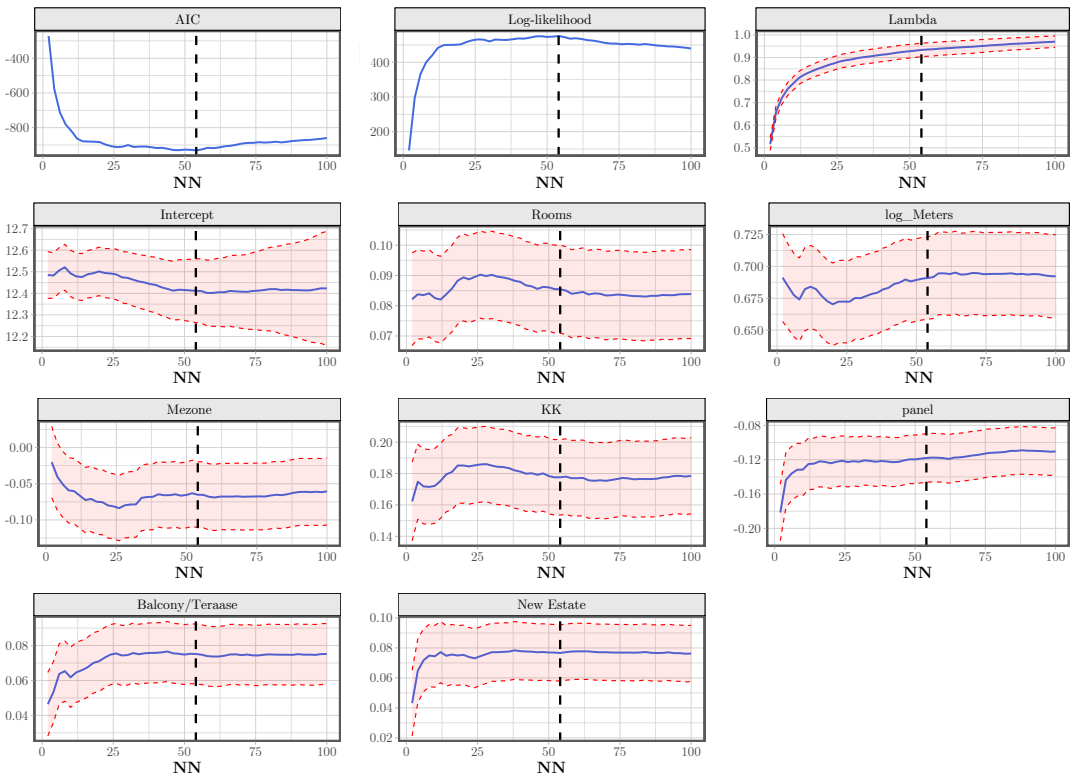


Figure 1 Estimated model parameters and the *AIC* statistics for different numbers of neighbor units.

Spatial dependency structure and derived spatial weight matrix W that minimizes *AIC* is for 54 nearest neighbors. Models in Table 2 are using such spatial dependency structure. Lastly, we take OLS model residuals and investigate their distribution across space. Since we are using the logarithmic transformation we need to use correct transformation of residuals rather than using the simple exponential transformation. After this step we discretize residual values and inspect residuals distribution in Figure 2.

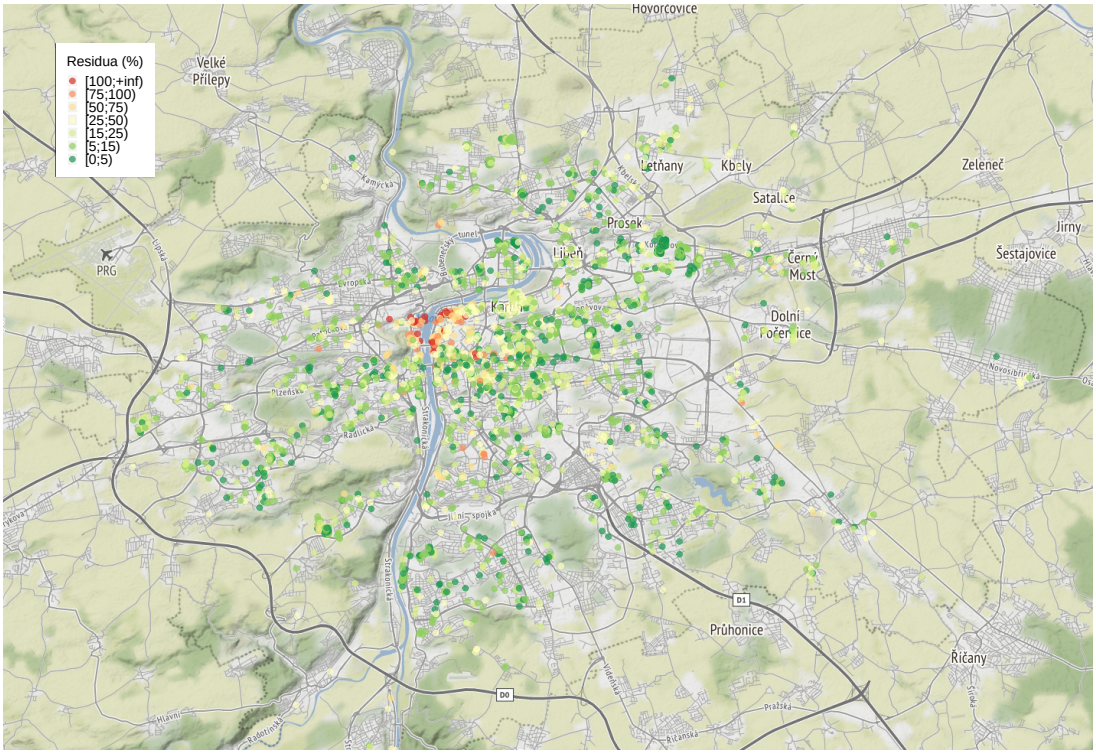


Figure 2 Residual distribution in space.

Clearly, we can identify cluster with high prices in the historical part of Prague. Those findings are in compliance with [2], where after applications of different sets of variables we still see significant clustering. This allows us to believe that residual noise is (presumably) entirely due to the factor of location.

5 Conclusions

Spatial econometrics framework was utilized to investigate spatial dependency of flat market. To account for spatial heterogeneity we used simple clustering approach. After confirming spatial dependency using moran I test, we used two straightforward models, i.e., *spatial lag* and *spatial error*. Based on the models fit metrics we found that spatial models are more suitable for flats analysis. Lastly, we investigate residuals from OLS model which do no account for spatial dependency and investigate their distribution in space. Using this approach we are able to identify "grandiose" clusters witch tends to increase estate price simply because of location factor.

Acknowledgements

The work on this paper was supported by the grant No. F4/27/2020 of the Internal Grant Agency of University of Economics, Prague.

References

[1] Dubin, R. A. (1992). *Spatial autocorrelation and neighborhood quality*. Regional science and urban economics.

[2] Lipán, M. (2016). *Spatial approaches to hedonic modelling of housing market: Prague case.* Bachelor's thesis, Charles University, Faculty of Social Sciences, Institute of Economic Studies.

[3] Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

[4] LeSage, J. P. & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton: CRC Press, Taylor & Francis Group.

[5] Anselin, L. *Spatial econometrics: methods and models. Vol. 4.* Springer Science & Business Media, 2013.

[6] Formánek, T. & Hušek, R. (2016). *On the stability of spatial econometric models: Application to the Czech Republic and its neighbors.* Mathematical Methods in Economics: 213–218.

The Long–term Relationship between Military Expenditure and Economic Growth the NATO Countries

Filip Hron[1], Lukáš Frýd[2]

Abstract. This paper aims to detect long term equilibrium between economic growth and military expenditure for 22 NATO members during the years 1988 and 2017. This study is used advanced panel techniques, that assumes the existence of cross-section dependence. We find that among GDP, military expenditure and capital formation exist cointegration relationship. Moreover, the impact of military expenditures is relatively similar to capital formation impact.

Keywords: cross-section dependence, military, economic growth , panel cointegration

JEL Classification: C50, O40
AMS Classification: 62P20

1 Introduction

Military expenditures represent a significant source of fiscal expenditures in the economy. For example, members of NATO follow the rule that the budget of the ministry of defense should be equal to 2% of GDP. However, there are only seven countries, which have already met the two percentage target, based on NATO report from 2019. This leads us to the question, whether long term relationship exists during the investigated period. The crucial question is also if military spendings represent consumption or capital formation. If these spendings may produce a positive spillover effect on the economy, or represent only consumption driven demand.

For example, Benoit [2][3] supports the idea that military expenditures have a positive effect on economic growth. Another supporting fact is Keynesian militarism, which assumes that increasing government military spending supports the growth of aggregate demand thanks to the Keynesian multiplicative effect. In addition thesis by Yildirim, Sezgin a Öcal [16] supports existence of positive effect in Turkey and Middle Eastern countries based on Feder model approach. Another comprehensive study supporting this relationship is publicated by Lee and Chen [17]. They investigated 27 OECD and 62 non-OECD countries for 1988–2003 period with panel data approach and they also tested existence of cointegration. Authors drawn a conclusion, there was a significant influence of military expenditure to the gross domestic product. It is appropriate to mention, that the effect was positive in OECD members. On the other hand, the effect was demonstrably negative for non-OECD countries. In the most cases, this fact was explained as non-OECD country relied on military supply from OECD members, which led to positive spin–off effect for OECD countries. In the recent publication by Cavatorta and Smith [6] is investigated relationship of demand for military expenditure and economic growth. They provided evidence of long term equilibrium. However, demand was not able to react on speed of the economic growth during short-term period. They also used similar methodology as in this thesis.

In addition, there are several thesis, where is used estimators, which are not robust against existence of cross section dependence and it could leads to bias and inconsistent standard errors. The cross section dependence may occur in the form of economic shocks, changes in oil prices and etc. These factors are also related to country security and therefore it is appropriate to include it, thereby this paper differs from most of the existing literature. Within this fact, it is necessary to use robust approach, in our case named Common Correlated effect estimator (*CCE*), which is more described in Pesaran [4]. We equally important provide testing cointegration and unit roots in cross-sectional dependent sample.

The main goal of this study is to find if the similar impact of military expenditures on economic growth may be found in the NATO countries, with the stress of the presence of cross–section dependence.

For this reason, we utilized advanced panel data methods in our analysis. Hence, the utilized econometric framework represents the novelty in our study.

[1] Department of Econometrics, University of Economics in Prague, Winston Churchill Square 4, 13067 Prague, Czech Republic, hrof01@vse.cz
[2] 1Department of Econometrics, University of Economics in Prague, Winston Churchill Square 4, 13067 Prague, Czech Republic, lukas.fryd@vse.cz

2 Model and methodology

2.1 Model

In our study, we do not work with all NATO members. Due to the irrelevance of a few countries, we remove them from the analysis. It concerns especially members, without any records about military spending as Iceland. Furthermore, countries with weak quality of available data are removed too. The study is established on annual data set with a total of twenty-two NATO members ($N = 22$) during the period from 1988 to 2017 ($T = 30$). Similar to the d'Agostino, Dunne and Pieroni [5] the general model is determined based on Cobb–Douglas production function and we follow typical approach:

$$\ln GDP_{it} = \alpha_1 + \beta_1 \ln M_{it} + \beta_2 \ln C_{it} + \epsilon_{it}, \tag{1}$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$. GDP_{it} is gross domestic product, M_{it} represents military expenditures and C_{it} represents gross fixed capital formation. All three variables represent per capita ratio and are accounted in current prices of *USD*. The crucial part of equation is random error ϵ_{it}. The most famous panel estimator, Fixed effect estimator, see Green [13]. assumes that $E(\epsilon_{it}\epsilon_{jt}) = 0\ \forall i, j$, when $i \neq j$. Pesaran stress, that in macroeconomic panel data, this assumption is too restrictive. In reality, the unobserved common factors (shocks) have influenced all cross section units. These factors are the source of cross-section dependence. Following Pesaran [4] we assume that:

$$\begin{aligned} \epsilon_{u,it} &= \alpha_{u,i} + \boldsymbol{\lambda}_i \mathbf{f}_t + e_{it}, \\ \mathbf{x}_{it} &= \boldsymbol{\alpha}_{x,i} + \boldsymbol{\lambda}_{x,i}^T \mathbf{f}_t + \mathbf{u}_{x,it}, \end{aligned} \tag{2}$$

where $\mathbf{x}_{it} = (\ln M_{it}, \ln C_{it})$ f_t is vector of unobserved common factors ($m \times 1$), $\alpha_{,i}$ represents time independent individual effects for each country i, e_{it} is random variable and $\mathbf{u}_{x,it}$ represents the vector of random variables. We assume that random variables are independent and follow *iid* processes. It is obvious that unobserved factors $\mathbf{f}_t$ are common for random error $\epsilon_{u,it}$ as well as for regressors $\mathbf{x}_{it}$. Hence, traditional estimators like a Fixed effect estimator will produce inconsistent estimates. Pesaran suggests to approximate the unobserved common factors by arithmetic averages of each variable from Equation 1 across $i = 1, \ldots, N$ expressed as:

$$\begin{aligned} \overline{\lambda} = N^{-1} \sum_{i=1}^{N} \lambda_i, \quad \overline{\epsilon}_{,t} = N^{-1} \sum_{i=1}^{N} \epsilon_{,it}, \\ \overline{\mathbf{x}}_t = N^{-1} \sum_{i=1}^{N} \mathbf{x}_{it}, \quad \overline{GDP}_t = N^{-1} \sum_{i=1}^{N} GDP_{it}. \end{aligned} \tag{3}$$

Then auxiliary equation which follows assumptions from Pesaran [4] with mentioned averages is written as:

$$\overline{GDP}_t = \overline{\alpha} + \overline{\boldsymbol{\beta}}^T \overline{\mathbf{x}}_t + \overline{\lambda} \mathbf{f}_t + \overline{\epsilon}_t. \tag{4}$$

If parameter $\overline{\lambda}$ is not equal to zero, then averages of unobserved common factors $\overline{\lambda} f_t$ are expressed as:

$$\mathbf{f}_t = \overline{\lambda}^{-1} [\overline{G}_t - (\overline{\alpha} + \overline{\boldsymbol{\beta}}^T \overline{\mathbf{x}}_t + \overline{\epsilon}_t)]. \tag{5}$$

Hence, we can rewrite Equation 1:

$$G_{it} = \alpha_0 + \beta_1^C C_{it} + \beta_2^M M_{it} + \delta_1^C \overline{C}_t + \delta_2^M \overline{M}_t + \eta_1 \overline{G}_t + e_{g,it}. \tag{6}$$

$e_{g,it}$ represents idiosyncratic and approximation error of unobserved common factors. Equation 6 can be estimated by Ordinary Least Squares and the estimator is named Common Correlated Effect (CCE). Furthermore, estimation is robust contra heteroskedasticity, autocorrelation and non-stacionarity of common unobserved factors f_t Kapetanios at al.[8].

Specifically Peseran presents two estimation methods for CCE, which names are Pooled (CCEP) and Mean group (CCEMG). Pooled CCE estimator assumes homogeneity of parameters $\beta_i = \beta$. But, we use CCEMG estimator due to allowed heterogeneity in β_i coefficients from Equation 6, which is expressed as arithmetic

average of slope coefficients $\hat{\beta}_{CCEMG} = N^{-1} \sum_{i=1}^{N} \hat{\beta}_i$. This helps us also to produce a cointegration test robust to the presence of cross–section dependence, which is presented in Subsection 2.4.

2.2 Cross sectional dependence test

In the first place, we test cross–section dependence. We follow Peseran´s Cross–section dependence (CD) test [7]. The test statistic has the following form:

$$CD = \sqrt{\frac{2T}{N(N-1)}} (\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}) \xrightarrow{d} N(0,1), \tag{7}$$

where $\hat{\rho}_{ij}$ are estimation of residual pair correlation coefficients based on individual *OLS* regressions. According to Pesaran [7], *CD* test is applicable to both small and large samples. It is also usable on unbalanced panel data.

2.3 Unit root test

Panel unit root tests which relax the assumption of cross–section dependence are labeled as the First generation unit root tests. These tests may produce misleading results due to the existence of cross–section dependence [9]. Hence, we utilize a second generation panel unit root test. Specifically, we use the CIPS test by Pesaran [10]. This test is based on a panel version of Augmented Dickey–Fuller unit root test [15], extended of approximation of unobserved common factors. Pesaran [10] utilizes Common Correlated Estimation framework. Hence, we get Common Correlated Augmented Dickey–Fuller regression (CADF):

$$\Delta y_{it} = \alpha_{CIPS,i} + d_i t + \beta_{CIPS,i} y_{it-1} + \psi \bar{y}_{t-1} + \sum_{j=1}^{p} \gamma_{ij} \Delta y_{it-j} + \sum_{j=0}^{p} \gamma_{ij} \Delta \bar{y}_{t-j} + u_{p,it}, \tag{8}$$

where $\bar{y}_t = N^{-1} \sum_{i=1}^{N} y_{it}$, Δy_{it} represents first difference and deterministic trend is contained in $d_i t$. Then null hypothesis is $H_0 : \beta_{CIPS,i} = 0$ for $\forall i = 1, ..., N$ and alternative $H_1 : \beta_{CIPS,i} < 0$. Pesaran provide test statistic, where t_i is t-statistic from regression 8 and with null hypothesis non–stationary, expressed as:

$$CIPS(N,T) = N^{-1} \sum_{i=1}^{N} t_i(N,T) \xrightarrow{d} N(0,1) \tag{9}$$

Critical values are tabulated by Pesaran [10]. Optimal number of lags p is based on rule $p^* = [4(T/100)^{1/4}]$.

2.4 Cointegration

The cointegration testing follow Engle and Granger [14] methodology for panel version. In the first stage we get residuals $\hat{e}_{it}$ from regression presented in Equation 6 from CCEMG estimation.

In the second stage, we test the presence of unit root in residuals $\hat{e}_{it}$ with CIPS test. In the case the test rejects the null hypothesis about the stationary process of residuals, we can highlight long term equilibrium between economic growth, military spending, and capital formation.

3 Empirical results

3.1 Cross sectional dependence

Based on Table 1, we can reject the null hypothesis about the existence of non–cross sectional dependence.

Variable	CD test	p-value
GDP	75.27	0.000
M	61.68	0.000
C	64.33	0.000

Table 1 Cross-section dependence test

3.2 Stationarity

For testing stationarity, we use the CIPS test. Following Pesaran [10] the optimal number of lags is chosen as $p^* = 2.960 \doteq 3$. In the Table 2 are displayed results from CIPS test with $lags = 3$. According to this result, we can not reject the null hypothesis that all variables are non–stationarity. We also provide a test with and without deterministic trend $d_i t$ from the Equation 8. In the next step, we test the first differences. Based on results we are able to reject the null hypothesis. Hence, we assume that all variables from Equation 1 are integrated of order one $I(1)$.

Specification		const		trend	
Variables	lags	zt bar	p-value	zt bar	p-value
GDP	3	−1.425	0.077	−1.572	0.058
M	3	−0.129	0.449	1.441	0.925
C	3	−0.699	0.242	0.604	0.727

Table 2 Cross-sectionally augmented IPS test results

3.3 Cointegration

Following Engle and Granger framework, in the first instance, we estimate model from the Equation 1 using CCEMG estimator. The results are presented in Table 3.

Variable	coef.	p-value	St. error
M	0.331	0.000	0.035
C	0.382	0.000	0.046

Table 3 Estimation of long-term relationship – CCEMG approach

In the next step, we test residuals for the presence of unit root by CIPS test. The results are shown in Table 4. Hence, we can reject the null hypothesis. Furthermore, this result indicates long–term relationship among our three variables. Military expenditures have a positive and significant effect on economic growth.

Specification		const		trend	
	lags	zt bar	p-value	zt bar	p-value
Residuals	3	−5.011	0.000	−1.699	0.045

Table 4 Cointegration testing

4 Conclusion

In this paper, we use advanced panel data methods to examine common relationship primary between gross domestic product and military expenditure in 22 NATO countries during 1988 and 2017. We handle with the existence of cross–sectional dependence, which we positively tested. Further, we utilize panel unit root test and cointegration test robust to the presence of cross–sectional dependence. We confirm, that all variables follow $I(1)$ processes. Next, we find the presence of cointegration vector among the variables. We conclude that military expenditure positively influences gross domestic product.

Acknowledgements

We gratefully acknowledge the support of this project by the research grant VSE IGA F4/34/2020, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Bilingsley, P. (1968). *Convergence of Probability Measures*. New York: John Wiley & Sons Inc.

[2] Benoit, E. (1973). *Defense and economic growth in developing countries.*. Mass., Lexington Books

[3] Benoit, E. (1978). Growth and defense in developing countries. *Economic development and cultural change*, 26, 271–280.

[4] Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4), 967–1012.

[5] d'Agostino, G., Dunne, J. P. & Pieroni, L. (2010). Assessing the effects of military expenditure on growth. *Oxford handbook of the economics of peace and conflict*. Oxford University Press, Oxford.

[6] Cavatorta, E. & Smith, R. P. (2017). Factor models in panels with cross-sectional dependence: an application to the extended SIPRI military expenditure data. *Defence and Peace Economics*, 28(4), 437–456.

[7] Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels.

[8] Kapetanios, G., Pesaran, M. H. & Yamagata, T. (2011). Panels with non-stationary multifactor error structures. *Journal of Econometrics*, 160(2), 326–348.

[9] Breitung, J. & Das, S. (2005). Panel unit root tests under cross-sectional dependence. *Statistica Neerlandica*, 59(4), 414–433.

[10] Pesaran, M. H., Smith, L. V. & Yamagata, T. (2013). Panel unit root tests in the presence of a multifactor error structure. *Journal of Econometrics*, 175(2), 94–115.

[11] Everaert, G. & De Groote, T. (2016). Common correlated effects estimation of dynamic panels with cross-sectional dependence. *Econometric Reviews*, 35(3), 428–463.

[12] Pesaran, M. H., Shin, Y. & Smith, R. P. (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the american statistical association*, 94(446), 621–634.

[13] Greene, W. H. (2003). Econometric analysis. Pearson Education India.

[14] Engle, R. F. & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251–276.

[15] Cheung, Y. W. & Lai, K. S. (1995). Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business & Economic Statistics*, 13(3), 277–280.

[16] Yildirim, J., Sezgin, S. & Öcal, N. (2005). Military expenditure and economic growth in Middle Eastern countries: A dynamic panel data analysis. *Defence and Peace Economics*, 16(4), 283–295.

[17] Lee, C. C. & Chen, S. T. (2007). Do defence expenditures spur GDP? A panel analysis from OECD and non-OECD countries. *Defence and Peace Economics*, 18(3), 265–280.

EOQ as a Tool for Increased Transport Efficiency

Michal Husinec[1], Tomáš Šubrt[1], Jiří Fejfar[1]

Abstract. Modern logistics is affected by a high degree of inefficiency in the optimum order size caused by the large number of smaller shipments in the distribution network. The developed transport network and advanced transport technologies enable end-users and distribution centers to reduce the amount of money in stock if they order only the amount needed to satisfy the consumer, regardless of minimum insurance stocks. The use of the EOQ principle would define the optimal order and thus the efficient utilization of the vehicle in relation to the total transport costs. The analyzed problems are linked to the lack of information in the sector concerned, the limited details and the closed environment for freely available vehicle capacities. This article introduces selected economic and mathematical methods in relation to the application for the purpose of determining possible economies of scale and savings from specialization. The essence of the application of economic and mathematical methods is the efficient use of resources and mitigating the negative impacts of road transport.

Keywords: economic order quantity, transport costs, return to scale, return to specialization

JEL Classification: C44
AMS Classification: 90B50, 91B06

1 Introduction

The objective of modern transport today is to accelerate the service and ensure the quality of the service. But from the economic point of view, this does not constitute an efficient use of resources, in particular as concerns the utilization of vehicles and their overall economic efficiency. The problem of unutilized vehicles points to the inability of transport firms to correctly define transport costs and specify the composition of their fleets, depending on the requirements of the customers purchasing their goods. This document formulates the EOQ model of an economic nature, which is frequently used for increasing the reliability of the entire system, thanks to its simplicity. Even though all orders may be placed randomly, the overall result always factors in annual demand as being constant. Taking into account that this is a stochastic process, for example, due to unreliable transport, it may make demand during the time of its implementation even riskier. Overall, decisions on cost savings should therefore not be made in advance, but rather, they must be the result of the minimization of all appropriate costs. Strategic measures that reduce transport time or uncertainty in transport help companies reduce their safety stock, while maintaining or increasing their present level of service. Transport requirements grow with the needs of modern society, and today purchasing trends are shifted into the sphere of online purchases, with customers usually expecting delivery to their home [11]. As a standard, delivery is expected within the shortest time possible. With increasing demand for transport capacity, the adverse influence of road transport on the environment is increasing, in particular pollutants and increased noise level in urban agglomerations [3]. The objective of cities is an overall reduction of these adverse influences, which is why low-emission zones are being set up today. Both of these adverse influences involve traditional vehicles with conventional drives. That is why an increasing number of transport companies look for alternatives in terms of sustainable development of transport and distribution of goods using less environmentally demanding vehicles. One of the options is to use electric cars or hybrid drive vehicles [5]. The initial investment, which his higher compared to traditional cars, leads to thoughts about efficiency of use. Modern logistic systems provide nearly unlimited options of delivery of goods and the end consumer keeps only minimal stocks, as it expects delivery within a short period of time, usually within 24 hours from ordering. That means that transport companies are forced to invest in more vehicles with differing capacity. The final volume of investment in vehicles for distribution will also result in a change in consumer behaviors, because carriers will have to reflect the higher acquisition costs in the final price of the goods delivered.

[1] Czech University of life Sciences in Prague, Department of Systems Engineering, Kamýcká 129, Prague, Czech Republic, e-mail: husinecm@pef.czu.cz

In order to achieve better consumer conduct, and hence greater efficiency of delivery costs, we can apply the EOQ (Economic order quantity) principle. It is one of the most widely discussed models in production logistics [12]. Probably the most widely known adaptation of the principle is the EPQ method (Economic Production Quantity), which reflects daily demand for product production. Another well-known modification of the EOQ method is "Japanese production", a practical approach focused on the size of an order and the time required for the production or preparation of the order, [6], with the introduction of principles such as Just-in-time (JIT). An approach that is usually based on the "Pull Production" principle [1] and where production time is adapted to customer demand. Final products are distributed to customers immediately. But what happens if production or preparation of an order cannot respond to demand from customers, due to longer delivery periods or potential production outages? Then delivery processes should be planned in advance, for example, with the use of "Push Systems" [2]. In this case, however, operative processes will increase, as well as stocks of goods and overall logistics costs. Slower or unbalanced orders then introduce into the system increased costs of the realization of a repeated order [4, 8, 9, 10, 14, 15]. This contribution expands the possibilities of further use of the EOQ principle in determining the optimal order, with a view to cost minimization and increasing efficiency in the distribution of goods. Furthermore, attention must also be paid to stock optimization at a time of uncertainty, including tracking of carriage time or of the preparation of an order as a part of the overall process. [15] In such cases, there is usually a higher likelihood that stocks will run out, which means that it will not be possible to deliver goods from the warehouse. In order to prevent the exhaustion of stocks, a so-called insurance stock is made, which allows a company to optimally distribute its stock supplies or set up an interim store for its products. A small interim store may make for smooth supplies of goods, whereas a large interim store will have a significant impact on overall logistics costs. Uncertainty with respect to stock supplies means that, on top of minimization of logistical costs, the costs of an insurance supply and costs of stocks must be added. It is evident that optimal stock supplies and costs of a warehouse increase uncertainty. On the other hand, unreliable transport results in the creation of excessive safety stocks and higher costs of stocks. We assume that the objective of optimization using EOQ is to minimize the economic impact of the model. That is, to minimize all logistics costs, such as the costs of transport, non-transport costs of an order, costs of the storage of stocks, including costs of the storage of the safety stocks, in order to prevent uncertain supplies during the delivery period – and the costs of stocks. A cost and benefit analysis of the EOQ method uses concepts of economic theory to improve the efficiency of distribution and reduce the adverse impact of modern modes of transport [6].

2 Material and Methods

Economic order quantity is the optimal order quantity that minimizes the total operations logistics costs. Model EOQ is a one of the oldest models. The EOQ model was developer by Ford W. Harris in 1913 [7]. The general EOQ formula only applies if demand is constant throughout the year and we expect every future order to be delivered in full. Fixed costs must be entered for each order. The price of the goods purchased is known. The objective of the EOQ model is to minimize aggregate costs of inventory stock and determine the optimal size of an order [13]. In order to be able to apply to EOQ model in a distribution system, we have adapted the general formulation as follows:

$$EOQ^* = \sqrt{\frac{2PD}{C}} \tag{1}$$

Where: EOQ^* recommended optimal amount for one delivery

P costs per order (shipping and costs of assembly)

D annual demand quantity

C actual storage costs

2.1 Elements of logistic costs

To get a general understanding of the nature of optimization models for cost reduction, we must first define the constituent parts of the process and define individual logistical costs related to EOQ optimization. The fundamental issue is to always minimize overall costs, including individual transport costs, with a view to the size of an order, considering the fact that the size of the order must not exceed the size of the vehicle. We must therefore know in advance that it will be feasible and annual transport capacity must cover minimum annual demand.

Transport costs – these are the total costs related to the delivery of goods. That means not only costs incurred between the warehouse and the client, but also the costs of combined transport. In the case of combined transport, the tariff rate is assumed to be the total cost of transport (TC, Total Costs) per one unit of the commodity carried. We also expect that the frequency is sufficient for the total quantity, in order that the size of order need not be reduced. Transport costs are influenced by an effort to make maximum use of the fleet for the distribution of orders and by the variability of the vehicles used: in order to be able to define total costs of transport, we need to define the following model parameters.

$$TC = (T_n P_t) + (T_p P_t) + (L_{ij} P_r) + (L_0 P_r) \tag{2}$$

Where: T_n *unproductive time (waiting to loading, unloading)*

T_p *productive time (loading, unloading)*

L_{ij} *distance (km)*

L_o *distance without goods*

P_r *rate CZK/km*

P_t *rate / time*

Handling costs – the costs related to the carriage of goods, that is the costs related to the administration of an order. The costs of order assembly (OC) and the costs of packaging of the order and its preparation for shipment. Handling costs also include costs related to the receipt of the goods from production and the costs involved in the placement of the goods in store. In terms of cost classification, these are variable costs that may have a major impact on operating management. Individual parameters of the model are defined:

$$OC = (T_o A_c) + (T_o P_m) + (T_m A_c) + (T_0 P_m) + O_c \tag{3}$$

Where : T_o *time for receipt and storage*

T_m *time for handling shipments*

A_c *administration costc, rate per hours*

P_m *rate per hours – handling*

O_c *other costs*

Warehousing costs – costs related to the storage of the goods, i.e., the costs related to the rental of premises, an investment in a shelving system, utilities, and maintenance costs (inspection of the goods stored). Warehousing costs (WC) are related both to the placement of goods in a shelving system and to their storage on the warehouse floor. Warehousing costs also include the costs of handling areas used for inspections, preparation, and handling of goods. The costs of shipping areas can also be included in other handling areas. In terms of cost classification, these are fixed costs which we can influence significantly by optimizing the way items are stowed and by changing the structure of the shelving system. Warehousing costs also include the costs of handling equipment. Individual parameters of the model are defined:

$$WC = (T_w P_m) + (Q_{rs} P_c) + (Q_{fp} P_f) + O_c \tag{4}$$

Where : T_w *time for removal from storage*

T_m *time for handling shipments*

P_c *rate for storage racks*

P_m *rate for spacestorage (freearea)*

Q_{rs} *goods in storage rack*

Q_{fp} *goods in spacestorage (freearea)*

P_m *rate per hours – handling*

O_c *other costs*

Insurance stock – purposefully created stock designed to cover extraordinary fluctuations in demand. It is a reserve that is used if a supplier does not delivery its supply. It helps level out fluctuations in demand or extensions in delivery times, when supplies drop below the level when orders are placed. Individual

parameters of the insurance stock (IS) model are defined for insurance stock and a standard deviation of sales.

$$\sigma c = \sqrt{\bar{R}(\sigma S^2) + \overline{S^2}(\sigma R^2)} \tag{5}$$

$$\sigma S = \sqrt{\frac{\sum f d^2}{n-1}} \tag{6}$$

Where : σc *insurance stock*

$\bar{R}$ *average stock replenishment cycle*

σR *standard deviation of the replenishment cycle*

$\bar{S}$ *average daily sales*

σS *standard deviation of average daily sales*

f *frequency of the case of the same daily sale*

d d*eviation of cases from the mean value*

n *number of observation*

3 Principle of the EOQ model in a distribution system – conditions of uncertainty

In order to set real-state conditions, i.e., a certain level of risk, the EQO method was further modified for stock management in conditions of uncertainty. The creation of an insurance stock will be added to the model design to cover for outages in deliveries of goods. The use of the insurance stocks will make it possible for the company to better respond to an uncertain situation on the market and to better optimise its transport costs that can arise if a quantity less than the optimal EOQ is ordered. For the calculation, data from companies that used greater delivery frequencies and whose daily sales manifest a greater level of variability. The selected set manifests different quantities of goods ordered with a repeating frequency *fd =1*. The distribution area of the Czech Republic was chosen for the application of the EOQ calculation. The data presented in this contribution is real data provided by a logistical company for the purpose of optimising the efficiency of distribution direction. The comprehensive data is an output from the internal system Logenius. The lack of constancy in transport cost is caused by the utilisation of vehicles of differing capacity. The costs of assembly and warehousing are constant, and the company bills them to customers in their actual amounts. Some companies report a zero deviation $\bar{R}$ (Company2). These are cases when restocking takes place once a month.

Company	$\bar{R}$	σS	$\bar{S}$	σR	σC
Company2	44,5	17,7	30,00	0,00	118
Company4	22,5	4,8	30,00	0,00	23
Company5	34,6	21,9	7,50	6,12	137
Company6	61,8	44,9	30,00	0,00	352
Company9	66,9	46,4	7,50	5,59	382
Company12	68,5	92,4	7,50	6,61	767
Company13	50,5	21,9	30,00	0,00	156
Company15	110,3	83,9	7,50	5,59	882
Company16	8,7	7,2	30,00	0,00	21
Company21	27,0	17,4	7,50	5,59	100
Company23	18,0	9,5	15,00	8,66	136
Company24	5,5	3,5	30,00	0,00	8
Company30	10,0	2,8	30,00	0,00	9
Company31	26,3	37,0	11,00	2,24	191
Company32	4,0	4,2	30,00	0,00	8
Company36	1722,5	857,7	4,50	2,24	35598
Company40	57,3	57,0	10,00	0,00	432
Company41	13,0	2,8	30,00	0,00	10
Company43	38,3	18,4	10,00	0,00	114
Company45	35,0	3,6	30,00	0,00	21
Company46	45,0	7,1	30,00	0,00	47

Figure 1 EOQ – condition of uncertainty

4 Result

The first results of the use of the EOQ method indicate a significant cost saving and hence also a greater level of utilization of trucks. With an optimization of its existing system, the logistics company could reduce its costs and could also reduce the total costs to the customer, in terms of the preparation and delivery of the goods. The total savings in the distribution area are displayed in the enclosed table. It is evident that a 52.5% cost saving has been noted. A secondary objective is CO2 reduction, as an adverse impact of road transport on the environment. EOQ optimization also discovered that some companies already optimize their costs, with a view to the quantity of goods ordered. Furthermore, an overall reduction in the number of transport runs was achieved, so an overall benefit can be noted in the sphere of freight transport and its adverse impact on the environment.

	Actual	Optimatized
Number of deliveries	129	61
Number of cartons / deliveries (Average)	8927	5451
Transport costc	124 113,7 CZK	58 967,1 CZK
Saving costs (CZK)		65 146,7 CZK
Saving costs (%)		52,5%

Figure 2 The result of optimization

In terms of insurance stock creation by certain companies, warehousing costs increased due to the higher amount, as is evident from the table below, by a total of 48.3%

Company	Actual Stocks	Insurance Stock	New Stocks	Actual storage cost	New storage costs	Incerase Costs
Company2	89	118	207	7 058 CZK	16 409 CZK	9 351 CZK
Company4	90	23	113	7 137 CZK	8 941 CZK	1 804 CZK
Company5	381	137	518	30 213 CZK	41 050 CZK	10 837 CZK
Company6	247	352	599	19 587 CZK	47 536 CZK	27 949 CZK
Company9	1071	382	1453	84 930 CZK	115 237 CZK	30 307 CZK
Company12	754	767	1521	59 792 CZK	120 587 CZK	60 795 CZK
Company13	101	156	257	8 009 CZK	20 362 CZK	12 353 CZK
Company15	421	882	1303	33 385 CZK	103 312 CZK	69 927 CZK
Company16	26	21	47	2 062 CZK	3 751 CZK	1 689 CZK
Company21	162	100	262	12 847 CZK	20 767 CZK	7 921 CZK
Company23	169	136	305	13 402 CZK	24 185 CZK	10 784 CZK
Company24	11	8	19	872 CZK	1 530 CZK	658 CZK
Company30	20	9	29	1 586 CZK	2 295 CZK	709 CZK
Company31	86	191	277	6 820 CZK	21 993 CZK	15 173 CZK
Company32	8	8	16	634 CZK	1 307 CZK	673 CZK
Company40	172	432	604	13 640 CZK	47 872 CZK	34 233 CZK
Company41	26	10	36	2 062 CZK	2 871 CZK	809 CZK
Company43	153	114	267	12 133 CZK	21 144 CZK	9 011 CZK
Company45	105	21	126	8 327 CZK	10 018 CZK	1 692 CZK
Company46	90	47	137	7 137 CZK	10 899 CZK	3 762 CZK

Figure 3 Update of stock costs – insurance stock

5 Discussion

For greater effectiveness, the EOQ model formulated here combines two sets of logistical literature. Depending on the point of view, there are two areas where transport costs can translate into the overall model. One of the angles could be to merge all costs in the total sales price. The second model defines each cost separately. From the point of view of overall optimization, it is more suitable to use non-cumulated costs. With a view to market uncertainty and development of modern technologies, among other things, it is suitable to monitor each cost separately, and thus better optimize individual components of logistical costs. The above mentioned model does not factor in the volume size of a delivery, but given the great variability of the vehicles used, its implementation on the basis of the EOQ model executed can be assumed. It does not fully reflect uncertain demand and delivery periods as a consequence of creating insurance or safety stocks, while maintaining maximum level of customer service. A derived problem reflects minimization of costs and can be resolved analytically in all cost areas. Its use in Excel indicates its easy application in data and their fast processing. Furthermore, for maximum utilization of the model, it is limited by the entering of all

variables, i.e., the greatest problem will probably be the correct entering of all costs. Defining all costs provides suitable material for a study of the economic efficiency of costs, economies of scale, or savings due to specialization in specific cases. The approach selected can also be used for analyzing the costs versus the benefits of freight transport, factoring in the actual proliferation of vehicles with an alternative drive. Recently, road transport has been struggling with a lack of professional drivers and hence also the question whether optimization of load size, and hence a reduction in the total number of freight vehicles in a distribution system, may not help address the situation. In the future, a greater level of cooperation among logistics operators, owners, and clients purchasing goods can be expected, in an attempt to optimize costs as much as possible and to use transport capacity efficiently. And hence, also the use of the EOQ model can be expected, and its potential modifications for optimization of the issue at hand. On the basis of the optimization, distribution of goods at least once a month would be ensured, and the possibility of warehousing on the client's premises would not be assessed.

6 Conclusion

Of essence for the right definition of the issue – construction of the EOQ model and its subsequent modification for use in road transport – is the set-up of the system, such that it would suit all parties that take part in the logistical chain. Therefore, we would like to examine greater variability of the EOQ model in the future, which would include the client's costs before final consumption. And also concern ourselves with the application of modifications of the EOQ model in the Czech Republic.

References

[1] ABDULMALEK, F. A. & RAJGOPAL, J. (2007). Analyzing the Benefits of Lean Manufacturing and Value Stream Mapping via Simulation: A Process Sector Case Study. *International Journal of Production Economics*, vol. 107, pp. 223–236.

[2] AIGBEDO, H. (2000). Some Structural Properties for the Just-in-Time Level Schedule Problem. *Production Planning & Control*, pp. 357–362.

[3] CATTARUZZA, D. ET AL. (2017). Vehicle routing problems for city logistics. *EURO Journal on Transportation and Logistics*. Berlin Heidelberg: Springer, vol. 6, pp. 51–79. DOI: 10.1007/s13676-014-0074-0.

[4] GRAUERS, A., SARASINI, S., & KARLSTRÖM, M. (2013). Why Electromobility and What Is It? In *Systems Perspectives on Electromobility*, pp. 10–21, ISBN 978-91-980973-1-3. Available at: http://publications.lib.chalmers.se/records/fulltext/211430/local_211430.pdf.

[5] GRAHAM, D. J. & GIBBONS, S. (2019). Quantifying Wider Economic Impacts of agglomeration for transport appraisal: *Existing evidence and future directions*. DOI: 10.1016/j.ecotra.2019.100121.

[6] HALL, R. W. (1981). *Driving the Productivity Machine: Production Planning and Control in Japan*. Las Vegas, American Production, Inventory Control Society.

[7] HARRIS, F. W. (1915). What quantity to make at once. *The Library of Factory Management*, vol. V. Operation and A. W. Shaw Company, Chicago, pp. 47–52.

[8] CHIARINI, A. (2013). Relationships between Total Quality Management and Six Sigma inside European Manufacturing Companies: A Dedicated Survey. *International Journal of Productivity and Quality Management*, vol. 2, pp. 179–194.

[9] JUŘIČKOVÁ, E., PILÍK, M., & KVARTENG, A., M. (2017). Economics and Management of Enterprises. On-line shopping behaviour in the Czech Republic under the digital transformation of economy. *Economic Annals*–XXI. 165(165), pp. 119–1235. DOI: 10.21003/ea.V165-24.

[10] KHAN, M., JABER, M. Y., GUIFFRIDA, A. L., & ZOLFAGHARI, S. (2011). A Review of the Extensions of a Modified EOQ Model for Imperfect Quality Items. *International Journal of Production Economics*, vol. 132 (1), pp. 1–12.

[11] KUNDU, K. & STAUDACHER, A. P. (2016). *Investigation of Effectiveness of Order Review and Release Models in Make to Order Supply Chain*. DOI: 10.1051/0.

[12] MONDEN, Y. (2011). *Toyota Production System: An Integrated Approach to Just-in-Time*. New York. CRC Press.

[13] SHINGO, S. (1988). *Non-Stock Production: The Shingo System of Continuous Improvement*. Cambridge, MA: Productivity Press.

[14] SPEARMAN, M. L. & ZAZANIS, M. A. (1992). Push and Pull Production Systems: Issues and Comparisons. *Operations Research*, vol. 40 (3), pp. 521–532.

[15] ULUTAS, B. (2011). An Application of SMED Methodology. World Academy of Science, *Engineering and Technology*, vol. 5(7), pp. 63–66.

Spatial Variations in Unemployment Across the Slovak Districts: A GWR Approach

Michaela Chocholatá[1]

Abstract. This paper uses the geographically weighted regression (GWR) approach to analyse the spatial variations in unemployment across the 79 Slovak districts. The spatial character of unemployment as well as of its considered determinants (average nominal monthly wage, index of economic dependence of young people and crude rate of net migration) together with the substantial spatial heterogeneity across the geographical space were confirmed. Both the traditional global regression model and the local GWR model were estimated. The results proved that the local GWR model describes the data set better than the global regression model. Furthermore, the estimation of the local GWR parameters enabled to assess the diverse impact of analysed unemployment determinants in individual districts.

Keywords: unemployment, district, geographically weighted regression

JEL Classification: J64, C21, R23
AMS Classification: 91B72, 62P20

1 Introduction

Unemployment is one of the most frequently discussed socio-economic issues usually characterized by huge disparities both from the national and the regional perspective [4, 5]. The Slovak Republic has to face the phenomenon of substantial spatial variations in unemployment rates. The National Employment Strategy of the Slovak Republic until 2020 [14] states that the huge regional unemployment differences are associated with social, political and economic changes and identifies that the highest unemployment rates are typical for regions located in the southern part of middle Slovakia and eastern part of Slovakia.

The analysis of regional unemployment disparities based on various theoretical and empirical approaches has attracted increasing interest of researchers. Elhorst [5] provides an overview of 41 empirical studies dealing with the regional unemployment. Besides the traditional econometric approaches, consideration both of the geographical location and of the spatial effects has become very popular especially during the last two decades. Niebuhr [16] confirmed the significance of spatial dependence with respect to regional unemployment in selected European countries during 1986–2000. Using the measures of spatial autocorrelation and the spatial econometric methods, she suggests some implications for regional policy. Based on spatial econometric models, Cracolici, Cuffaro and Nijkamp [3] analysed the geographical distribution of unemployment across Italian provinces and confirmed significant degree of spatial dependence among provincial labour markets. Formánek and Hušek [6] investigated the regional unemployment dynamics in 7 EU member states based on semiparametric spatial models and underlined the importance to consider the spatial dependence in regional modelling of unemployment. Mariš and Marišová [12] used various statistical methods to assess the labour specialization and its impact on spatial patterns of unemployment across the Slovak regions. Netrdová, Nosek and Hurbánek [15] studied the spatial patterns of unemployment in seven Central European countries (including Slovakia) using different measures. To capture the spatial heterogeneity, the geographically weighted regression (GWR) approach is being widely used. The study of Salvati [17] analysed the Okun's relationship between district income and unemployment rate in Italy. Based on the GWR approach the study was able to reflect the different intensity in the modelled relationship. Lewandowska-Gwarda [11] used the GWR in the analysis of regional unemployment in Poland. She confirmed that the use of the GWR approach enabled to capture the diversity of unemployment determinants across the geographical space.

The aim of this paper is to investigate the spatial variations in unemployment rate in 2019 for the cross-sectional data from 79 Slovak districts. Besides the traditional global regression approach, the GWR approach taking into account the spatial heterogeneity of the analysed regions, will be applied as well. The

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1, 852 35 Bratislava, michaela.chocholata@euba.sk.

average nominal monthly wage, index of economic dependence of young people and crude rate of net migration are used as independent variables.

The structure of the paper is as follows: the introductory section is followed by section 2 which provides the methodology concerning the global regression model and the local GWR model. Data and empirical results are gathered in section 3, the conclusion of the paper can be found in section 4.

2 Methodology

The first step in modelling of the unemployment rate in this paper is based on the estimation of the traditional global regression model using the ordinary least squares (OLS) method, i.e.

$$y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k \, x_{ik} + \varepsilon_i \tag{1}$$

where index i = 1, ..., n, denotes the i-th region, y_i is the value of dependent variable at region i, x_{ik} denotes the values of the k-th independent variable at region i, β_0 is the intercept, β_k is the regression parameter for the k-th independent variable, p is the number of regression terms, and ε_i denotes the error term at region i. The OLS method supposes that the relationship between the dependent variable and independent variables is uniform over the whole area under consideration. However, when using the geographical data, one needs to pay attention to the spatial effects. In general, two types of spatial effects can be distinguished: spatial autocorrelation and spatial non-stationarity (heterogeneity). Within the cross-sectional data it is quite problematic to distinguish between these two effects [1]. Fotheringham [8] pointed out that sometimes spatial autocorrelation can be caused by estimation of a global model for a spatially non-stationary process and suggests in such cases to use the local approach. The local approach, i.e., the GWR approach, incorporates both the spatial non-stationarity and spatial autocorrelation [2] and enables local variations in the parameter estimates considering the coordinates (u_i, v_i) of individual regions. For each region i = 1, ..., n, the local GWR model is thus formulated as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p-1} \beta_k \, (u_i, v_i) x_{ik} + \varepsilon_i \tag{2}$$

where $(y_i, x_{i1}, x_{i2}, \dots, x_{ip-1})$ are the values of dependent variable and independent variables at location (u_i, v_i) of the i-th region, $\beta_j(u_i, v_i)$ for j = 0, 1, ..., p-1 are local regression parameters and ε_i for i = 1, ..., n are the error terms. The local regression parameters of model (2) are estimated for each location (u_i, v_i) using the weighted least squares method [7], [9], [13]:

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = [\boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}(u_i, v_i) \boldsymbol{X}]^{-1} \boldsymbol{X}^{\boldsymbol{T}} \boldsymbol{W}(u_i, v_i) \boldsymbol{y} \tag{3}$$

The spatial weights $w_j(u_i, v_i)$ $j = 1,2, \dots, n$ at each location (u_i, v_i) are defined as a distance function from the location (u_i, v_i) to the other locations. Symbol $\boldsymbol{W}(u_i, v_i)$ denotes the n x n diagonal local weight matrix at location (u_i, v_i) the off-diagonal elements of which are zero and diagonal elements are calculated based on a spatial kernel function giving higher weight to the locations closer to the location (u_i, v_i) in comparison to more distant locations ([7], [18]):

$$\boldsymbol{W}(u_i, v_i) = diag[w_1(u_i, v_i), w_2(u_i, v_i), \dots, w_n(u_i, v_i)\,] \tag{4}$$

The parameter estimates are thus dependent on the choice of the kernel function for geographical weighting and its bandwidth size. Either the fixed kernel function or the adaptive kernel function can be used (for more information see e.g. [7], [18]). From the various bandwidth selection methods, the most frequently used are the minimization of the corrected Akaike Information Criterion – *AICc* and minimization of the cross validation score – *CV*, respectively. The GWR provides the set of estimated parameters for each location (u_i, v_i). The local coefficients of determination R_i^2 as well as the parameter estimates for each location (u_i, v_i) can be mapped in order to gain a useful insight into the non-stationarity of the modelled relationships. Leung, Mei and Zhang [10] present some approaches for testing the goodness of fit of the local GWR model and propose the procedures for testing the variation of the GWR model parameters.

3 Data and Empirical Results

The data set comprises the regional data for the 79 Slovak regions (districts). The dependent variable is the registered unemployment rate in 2019 (in %), the considered independent variables are: the average nominal monthly wage in 2018 (in Euro), index of economic dependence of young people (i.e. the number of persons 0 – 14 years old per 100 persons aged 15 to 64) in 2019 and crude rate of net migration (i.e. ((immigration – emigration) / average population).1000) in 2019 (in ‰). The data were downloaded from the DATAcube database of the Statistical Office of the Slovak Republic [20]. The spatial analysis was performed in software GeoDa and the GWR fit was done in the MGWR 2.1 software environment. The shape file for the Slovak districts was retrieved from [19].

Natural breaks maps (for more information, see e.g., [21]) of variables under consideration are depicted in Figure 1 and document the considerable spatial diversification. The values of unemployment rate (*un_t19*) varying between 1.93% (Trenčín) and 15.14% (Rimavská Sobota) sharply differ across the Slovak districts. Especially districts located in the southern part of middle Slovakia and eastern Slovakia suffered from the high unemployment. Since the average nominal monthly wage (*w18*) was lower than 1022 Euro in more than a half of Slovak districts (located mainly in the southern, middle and eastern part of Slovakia), there were 8 districts with extreme values exceeding 1308 Euro. Figure 1 furthermore enables to reveal that while in the majority of Slovak districts the amount of children (0–14 years old) per 100 persons aged 15 to 64 was lower than 23.17, there were 8 districts with considerably higher ratios of children to the population in productive age with a maximum value of 35.99 in Kežmarok district. Mapping of the last independent variable, the crude rate of net migration (*ms19*), indicates the extremely high values for the 3 districts. Based on maps in Figure 1 we are able to identify some spatial patterns and outliers.

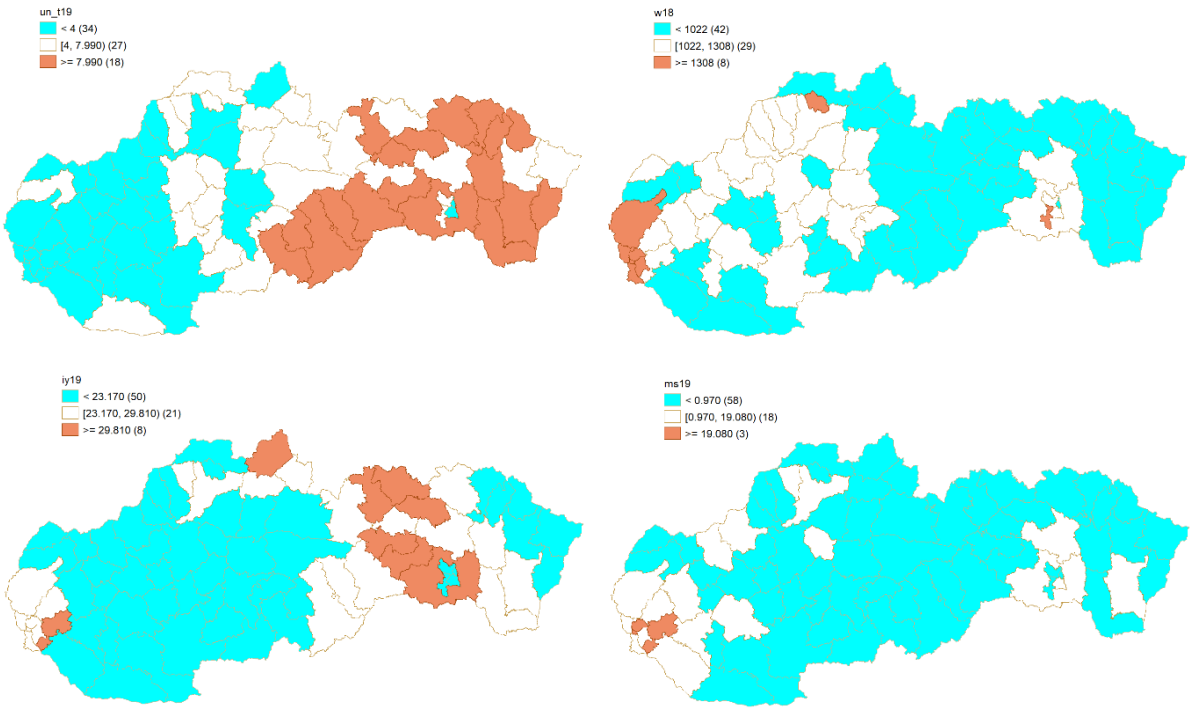


Figure 1 Natural breaks maps for unemployment rate (*un_t19*), average nominal monthly wage (*w18*), index of economic dependence of young people (*iy19*) and crude rate of net migration (*ms19*)
Source: author's calculations in GeoDa.

The values of global Moran's *I* statistics calculated for individual variables (Table 1) confirming the statistically significant positive spatial autocorrelation clearly indicate that there exist clusters of districts with similar values of individual variables. The location of a district in the geographical space thus does matter and should be considered in further econometric analysis (global Moran's *I* statistics were calculated based on contiguity weights – consideration of the queen contiguity and rook contiguity, respectively delivers the same results).

un_t19	*w18*	*iy19*	*ms19*
0.590***	0.569***	0.514***	0.499***

Source: author's calculations in GeoDa.

Notes: Symbol *** indicates statistical significance proved by the randomization approach based on 999 permutations.

Table 1 Moran's *I* values of unemployment rate (*un_t19*), average nominal monthly wage (*w18*), index of economic dependence of young people (*iy19*) and crude rate of net migration (*ms19*)

The traditional econometric approach starts with estimation of the global regression model (1) supposing the homogenous relationship between unemployment rate (*un_t19*) – as a dependent variable and average nominal monthly wage (*w18*), index of economic dependence of young people (*iy19*) and crude rate of net migration (*ms19*) – as independent variables. The OLS estimates of model (1) are gathered in Table 2 (column: Global model) indicating statistical significance of all estimated parameters with exception of intercept. The signs of estimated parameters indicate the negative impact of wages and of crude rate of net migration and positive impact of amount of children (0 – 14 years old) per 100 persons aged 15 to 64. The residuals from the global regression model showed statistically significant positive spatial autocorrelation with the Moran's *I* value of 0.362. To capture the spatial non-stationarity across analysed regions, i.e. spatially varied effects of individual independent variables on the unemployment rates, the estimation of the global regression model (1) was followed by estimation of the local GWR model (2). The GWR estimation results (minimum, median and maximum) are presented in Table 2 (columns: Local model). The adaptive bi-square kernel based on 53 nearest neighbours (calculated using the AICc minimisation) was used for the geographical weighting in the estimated local GWR model. The bottom part of the Table 2 contains the diagnostic statistics both for the global and local model fits. The values of the adjusted coefficients of determination R^2 and of AICc for global and local model, respectively indicate that the local model has better fit of the analysed data. Furthermore, in accordance with [7], Moran's *I* for the residuals from the GWR fit of 0.133 indicates considerable reduction in the degree of the spatial autocorrelation in comparison to the global regression model.

Model	Global model (OLS)	Local model (GWR)		
		Minimum	Median	Maximum
β_0	3.791	−3.108	6.405	16.831
β_1 (*w18*)	−0.007***	−0.014	−0.008	−0.001
β_2 (*iy19*)	0.389***	0.000	0.194	0.812
β_3 (*ms19*)	−0.124**	−0.317	−0.021	0.158
AICc	372.159		349.836	
R^2	0.460		0.684	
Adjusted R^2	0.438		0.628	

Source: author's calculations in GeoDa and MGWR 2.1

Notes: Symbols ***, ** indicate rejection of H_0 hypotheses at 1% and 5 % level of significance, respectively.

Table 2 Estimation results of global OLS regression and of GWR

Since the local R_i^2 were above the 0.415 in districts of middle and eastern part of Slovakia, the goodness-of-fit in western districts was substantially lower (results are available upon request). Monte Carlo test confirmed the spatial variability of all independent variables. Colour-coded maps of estimated local parameters from the GWR fit together with the corresponding *p*-values gathered in Figure 2 clearly document the different model performance across individual districts. Although the impact of average nominal monthly wage was negative across all districts, the statistically significant impact was proved only for regions located in the middle and eastern part of Slovakia. No changing positive signs in local parameter values were recorded for the second independent variable, the index of economic dependence of young people [4], [5]. Similarly, as in case of wages, the statistically significant impact of this variable was detected in almost the same

districts in middle and eastern part of Slovakia. The signs of the last set of estimated parameters corresponding to the crude rate of net migration were changing across the Slovak districts. By assessing the statistical significance of these parameters it can be concluded that the results proved statistical significance only in 3 districts (Levice, Krupina and Banská Štiavnica).

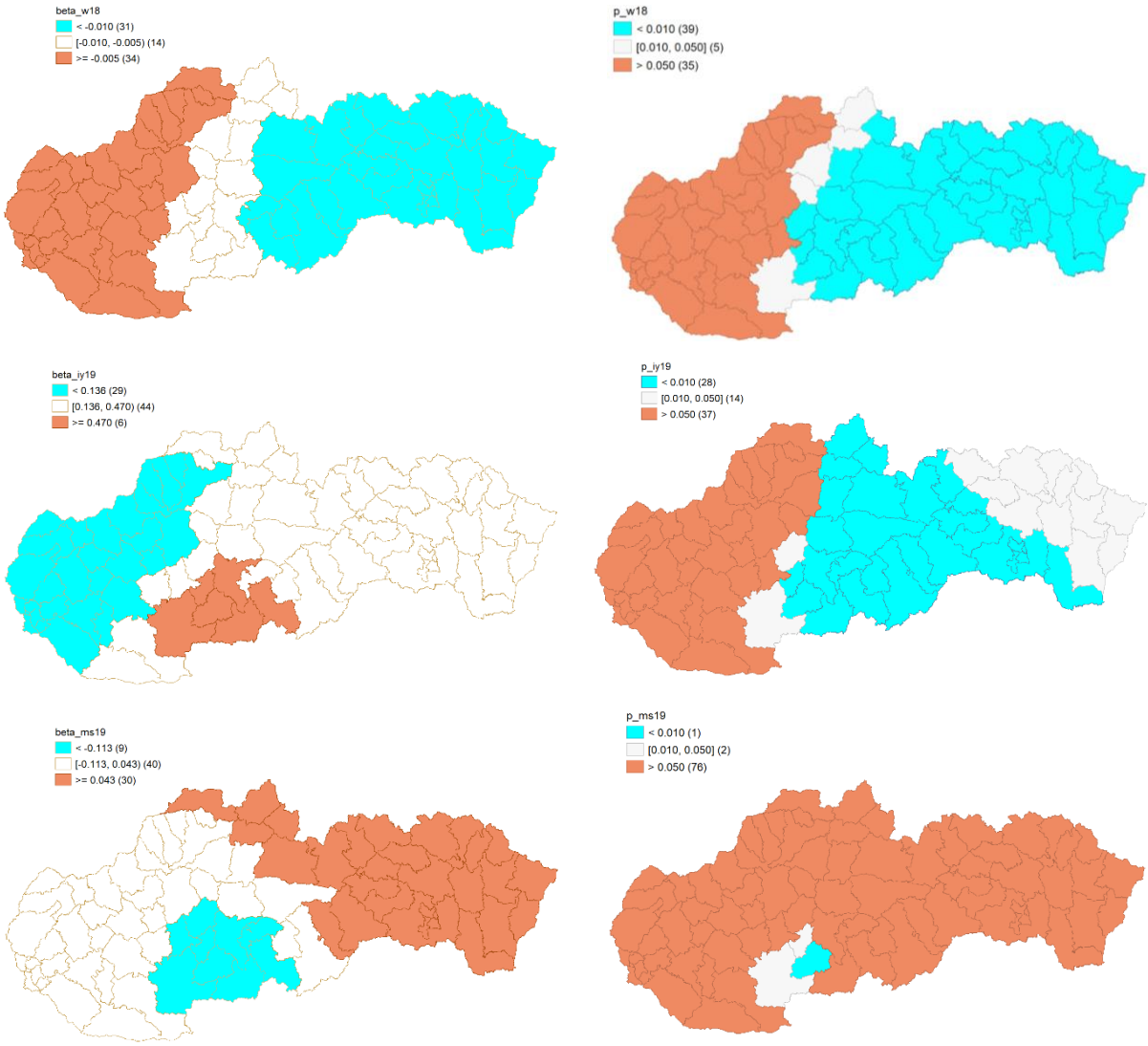


Figure 2 Local parameters and corresponding *p*-values (*GWR fit)*
Source: author's calculations in GeoDa and MGWR 2.1

4 Conclusion

This paper analysed the spatial variations in unemployment rate across the Slovak districts. The traditional global regression approach was extended by the GWR approach to allow for local parameters to be estimated. The average nominal monthly wage, the crude rate of net migration and the demographic variable – index of economic dependence of young people were used as independent variables. The preliminary spatial analysis revealed huge spatial disparities across individual districts for all variables under consideration. The results implied negative impact of wages on the unemployment rate both in the global and local model. Furthermore, similar as pointed out by e.g., [4], [5], it was confirmed, that higher proportion of children increases the unemployment rate (in both models). The impact of net migration was ambiguous, since the global model parameter indicated positive impact of this variable on the unemployment rate, the local parameter estimates had changing signs. The results presented in this paper clearly confirmed the spatial heterogeneity of Slovak districts and enabled to assess the diverse impact of analysed unemployment determinants in individual districts.

Acknowledgements

This work was supported by the Grant Agency of Slovak Republic – VEGA grant no. 1/0193/20 "Impact of spatial spillover effects on innovation activities and development of EU regions" and VEGA grant no. 1/0294/18 "Short-run and long-run dynamic analysis of economic development of European post-communist countries and regions".

References

[1] ANSELIN, L. (1999). The future of spatial analysis in the social sciences. *Geographic Information Sciences,* ***5***(2), 67–76.

[2] BLAINEY, S. & MULLEY, C. (2013). Using Geographically Weighted Regression to forecast rail demand in the Sydney Region. In *Australasian Transport Research Forum 2013 Proceedings (pp. 1–16).* Available online at: http://www.patrec.org/atrf.aspx [cited 2018-08-28].

[3] CRACOLICI, M. F., CUFFARO, M., & NIJKAMP, P. (2009). A spatial analysis on Italian unemployment differences. *Statistical Methods and Applications*, ***18***(2), 275–291.

[4] DEMIDOVA, O. & MOKHOVA, V. (2018). Spatial econometric modeling of regional unemployment in Russia: comparison of resource-rich and resource-deficient regions. Available online at: http://conference.iza.org/conference_files/Transition_2018/demidova_o26942.pdf [cited 2019-05-28].

[5] ELHORST, P. (2003). The mystery of regional unemployment differentials: Theoretical and empirical explanations. *Journal of Economic Surveys*, ***17***(5), 709–748.

[6] FORMÁNEK, T. & HUŠEK, R. (2019). Semi-parametric spatial model of unemployment dynamics based on Moran's eigenvectors. In *Proceedings of the 37th International Conference on Mathematical Methods in Economics 2019* (pp. 73–78). České Budějovice: University of South Bohemia in České Budějovice, Faculty of Economics.

[7] FOTHERINGHAM, A. S., BRUNSDON, C. M., & CHARLTON, M. E. (2002). Geographically Weighted Regression. The Analysis of Spatial Varying Relationships. Chichester: Wiley.

[8] FOTHERINGHAM, A. S. (2009). The problem of spatial autocorrelation and local spatial statistics. *Geographical Analysis*, ***41***(4), 398–403.

[9] FURKOVÁ, A. (2018). Spatial Heterogeneity of Regional Innovation Processes: Geographically Weighted Regression Approach. In *Proceedings of the International Scientific Conference Quantitative Methods in Economics: Multiple Criteria Decision Making XIX* (pp. 127–134). Bratislava: Letra Edu.

[10] LEUNG, Y., MEI, C.-L., & ZHANG, W.-X. (2000). Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model. *Environment and Planning A*, ***32***(1), 9–32.

[11] LEWANDOWSKA-GWARDA, K. (2018). Geographically Weighted Regression in the Analysis of Unemployment in Poland. *International Journal of Geo-Information*, ***17***(7), 1–16.

[12] MARIŠ, M. & MARIŠOVÁ, E. (2015). Labor specialization and its impact on spatial patterns of unemployment. Journal of Economy & Society, 1–2., 41–57.

[13] MEI, C.-L. (2006). Geographically Weighted Regression Technique for Spatial Data Analysis. Working Paper of School of Science, Xi'an Jiaotong University. Available online at: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=26930A9D9616E9D4C3A7DD808375F313?doi=10.1.1.550.1777&rep=rep1&type=pdf [cited 2020-03-28].

[14] National employment strategy of the Slovak Republic until 2020. Available online at: https://www.employment.gov.sk/files/slovensky/praca-zamestnanost/podpora-zamestnanosti/national-employment-strategy-slovak-republic-until-2020.pdf [cited 2020-05-01].

[15] NETRDOVÁ, P., NOSEK, V., & HURBÁNEK, P. (2020). Using Areal Interpolation to Deal with Differing Regional Structures in International Research. *International Journal of Geo-Information*, ***9*****,** 126, 1–14.

[16] NIEBUHR, A. (2003). Spatial interaction and regional unemployment in Europe. *European Journal of Spatial Development*, ***5***.

[17] SALVATI, L. (2015). Space matters: Reconstructing a Local-scale Okun's Law for Italy. *International Journal of Latest Trends in Finance and Economic Sciences*, ***5***(1), 833–840.

[18] WHEELER, D. C., & PÁEZ, A. (2010). Geographically Weighted Regression. In Handbook of Applied Spatial Analysis. Software Tools, Methods and Applications (pp. 461–486). Berlin Heidelberg: Springer-Verlag.

[19] https://www.geoportal.sk/sk/zbgis_smd/na-stiahnutie/ [cited 2020-04-28].

[20] http://datacube.statistics.sk/ [cited 2019-02-07].

[21] https://geodacenter.github.io/workbook/3a_mapping/lab3a.html#natural-breaks-map [cited 2020-05-07].

Using to Data Envelopment Analysis in Panel Regression to Forecast Health Care Performance Over Time in Europe

Lucie Chytilova[1], Jana Hanclova[2]

Abstract. Nowadays, health care is one of the most discussed topic all over the world. Also forecasting is an important tool used by governments to plan and evaluate. One of the most commonly used techniques for forecasting is regression analysis. Often forecasts are produced based upon a set of comparable units which could be individuals, groups or companies that perform similar activities such as a set of countries. This paper applies a methodology that includes a new independent variable, the comparable unit's DEA relative health efficiency of each country, into the panel regression analysis. The results of applying this methodology to the performance of health care in Europe over time is be presented.

Keywords: Data Envelopment Analysis, panel regression, health care

JEL Classification: C23
AMS Classification: 90C08

1 Intoduction

In this paper is used the extend work of Klimberg et al. where they had introduced a methodology that incorporates into the regression forecasting analysis a new variable that captures the unique weighting of each comparable unit [7] – [10]. This new variable is the technical efficiency of each comparable unit. It is generated by a non-parametric technique called data envelopment analysis (DEA). The DEA efficiency variable is a nonlinear variable that takes into account a set of weighted inputs and outputs. In their case always the inclusion of this multivariate variable has improved the regression forecasting model. The main objective of this paper is to use this method for panel regression of health care in Europe and see if the inclusion of the new variable is helpful.

Health care in Europe is provided through a wide range of different systems run at individual national levels. Most European countries have a system of tightly regulated, competing private health insurance companies, with government subsidies available for citizens who cannot afford coverage. The World Health Organization has listed 53 countries as comprising the European region. health outcomes vary greatly by country in the Europe. Countries in western Europe have had a significant increase in life expectancy since World War II, while most of eastern European and the formerly Soviet countries have experienced a fall in life expectancy. There are many factors which are causing preventable cause of death in Europe – tobacco, obesity and so on. All these info are important to detect for forecasting and to define how the health care should be in future.

Health care is important topic for all governments and countries all around the world. This topic can be viewed from many sides and angles. There are a lot of articles and analysis. For example, Bedir [2] has been examining the relationship between income and health expenditures. He found out a two way causality for the Czech Republic and Russian Federation. The health view over the income view is seen in Greece, Poland, the United Arab Emirates, China, Indonesia, and the Korean Republic supports the income view over the health view. The empirical results have indicated that income is an important factor in explaining the difference in healthcare expenditures among countries. Health care and bad habits has been also analysed by panel regression, for example, by Lightwood and Glantz [11]. In their paper, authors estimates changes in health care expenditure attributable to changes in aggregate measures of smoking in the USA. Also DEA is widely used, for example, authors of article [6], examined countries that joined the EU in 2004. In their conclusion, they classified countries which were the best performers in terms of balanced longevity increase followed by health expenditure growth. All these articles and many others analysis shows that this topic and used methods are good and important for future research.

[1] VSB – TU Ostrava, Faculty of Economics, Department of Systems Engineering, Sokolska tr. 33, Ostrava 2, 702 00, Czech Republic, lucie.chytilova@vsb.cz
[2] VSB – TU Ostrava, Faculty of Economics, Department of Systems Engineering, Sokolska tr. 33, Ostrava 2, 702 00, Czech Republic, jana.hanclova@vsb.cz

The rest of the paper has the following structure: section Methodology provides brief information about DEA models and panel regression. In Section 3 – Data, all variables are defined and information about them are given. Section 4 – Analysis and application – focuses on efficiency health analysis of countries in Europe and application in panel regression. The discussions about the results are provided in this section as well. Section 5 – Conclusion gives some conclusions and remarks.

2 Methodology

2.1 Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a non-parametric approach. It is widely used for measuring relative efficiency of decision making units (DMUs) with multiple inputs and outputs. Assume, there is a set of T DMUs (DMU_k for $k = 1, ..., T$), let input and output variables data be $X = \{x_{ik}, i = 1, ...R; k = 1, ...T\}$ and $Y = \{y_{jk}, j = 1, ...S; k = 1, ...T\}$, respectively. Also, u_i for $i = 1, ...R$ and v_j for $j = 1, ..., S$ be the weights of the ith input variable and the jth output variable, respectively. Mathematically, the relative efficiency score of DMU_k can be defined as:

$$e_k = \frac{\sum_{j=1}^{S} v_j y_{jk}}{\sum_{i=1}^{R} u_i x_{ik}}, \text{ for } k = 1, ..., T. \tag{1}$$

Charnes et al. [3] have proposed the following CCR model to measure the efficiency score of the under evaluation unit, DMU_Q where $Q \in \{1, ..., T\}$:

$$\begin{aligned} & \max e_Q = \tfrac{\sum_{j=1}^{S} v_j y_{jQ}}{\sum_{i=1}^{R} u_i x_{iQ}}, \\ \text{s.t.} \quad & \textstyle\sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} \leq 0, && k = 1, ..., T, \\ & u_i \geq 0, && i = 1, ..., R, \\ & v_j \geq 0, && j = 1, ..., S. \end{aligned} \tag{2}$$

The model (2) is non-linear. It is the model of linear-fractional programming. The model (2) could be transferred by Charmes-Cooper transformation to the standard linear programming problem:

$$\begin{aligned} & \max e_Q = \textstyle\sum_{j=1}^{S} v_j y_{jQ}, \\ \text{s.t.} \quad & \textstyle\sum_{i=1}^{R} u_i x_{iQ} = 1, \\ & \textstyle\sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} \leq 0, && k = 1, ..., T, \\ & u_i \geq 0, && i = 1, ..., R, \\ & v_j \geq 0, && j = 1, ..., S, \end{aligned} \tag{3}$$

where $Q \in \{1, ..., T\}$. DMU_Q is CCR-efficient if and only if $e^* = 1$ and if there exists at least one optimal solution $(\mathbf{u}^*, \mathbf{v}^*)$ with $\mathbf{u}^* > 0$ and $\mathbf{v}^* > 0$ for the set $Q \in \{1, ..., T\}$. The inefficient units have a degree of relative efficiency that belongs to interval $[0, 1)$. Note: The model must be solved for each DMU separately.

The model (3) is called a multiplier form of the input-orient-CCR model. However, for computing and data interpretation, it is preferable to work with model that is dual associated to model (3). The model is referred as envelopment form of input-oriented CCR model, see [3]. There also exists a multiplier form and envelopment form of output-oriented CCR model. Both models give the same results, see [3].

Banker et al. [1] have extended CCR model. The extended model is called BCC model and considers variable returns to scale assumption. The model has convex envelope of data which leads to more efficient DMUs. The mathematical model of dual multiplier form of input-oriented BCC model is:

$$\begin{aligned} & \max e_Q = \textstyle\sum_{j=1}^{S} v_j y_{jQ} - v_0, \\ \text{s.t.} \quad & \textstyle\sum_{i=1}^{R} u_i x_{iQ} = 1, \\ & \textstyle\sum_{j=1}^{S} v_j y_{jk} - \sum_{i=1}^{R} u_i x_{ik} - v_0 \leq 0, && k = 1, ..., T, \\ & u_i \geq 0, && i = 1, ..., R, \\ & v_j \geq 0, && j = 1, ..., S, \\ & v_0 \in (-\infty, \infty), \end{aligned} \tag{4}$$

where v_0 is the dual variable assigned to the convexity condition $\mathbf{e}^{\mathbf{T}}\lambda = 1$ of envelopment form of input-oriented BCC model. Note: The BCC model can be rewritten into the envelopment form or changed into the output orientation.

The input-oriented BCC model will continue into the next section.

2.2 Panel regression

Panel regression is a modeling method adapted to panel data, also called longitudinal data or cross-sectional data. In this paper the panel regression with fixed effects is used. Fixed-effects (FE) regression is a method that is especially useful in the context of causal inference [4].

Panel data are typically set up in long format – observations of each subject are ordered chronologically, and the time series (panels) of subjects are stacked below each other. This is used when estimating an effect of some causal variable using data from an annual panel survey conducted in years $t = 1, ..., T$ for individuals $i = 1, ..., N$. It is assumed that the outcome variable Y is continuous, while the K regressors $X_1, ..., X_K$ may be measured on any scale.

FE estimation builds on the error components model,

$$y_{it} = \mathbf{x}_{it}\beta + \alpha_i + \epsilon_{it}. \tag{5}$$

Here, y_{it} denotes the observed outcome of individual i at time t, $\mathbf{x}_{it}$ is the $(1 \times K)$ vector of covariates of this individual measured contemporaneously, and β is the corresponding $(K \times 1)$ vector of parameters to be estimated. The error term of this model is split into two components. The α_i are stable individual-specific characteristics which not only are often unobserved by the researcher, but also are very often related to the covariates. Hence, the α_i are unobserved effects capturing time-constant individual heterogeneity. The second component ϵ_{it} is an idiosyncratic error that varies across subjects and over time. The intercept α that is standard in regression models is dropped here due to collinearity with the person-specific errors α_i. The α_i is define as the dummy variable. This variable expresses the intensity of the differences in the level constant of the country "country i" ($i \neq 1$) compared to the base country "country 1". Also, depends on the type of model.

3 Data

The important task for efficiency measurements and also for the panel regression is to identify the right and relevant variables for the calculation. There is no defined approach for evaluation of health care in Europe. The model had been set up with 2 inputs which should be minimised and 3 outputs which should be maximised. The variables are as follows:

- Input 1 – Alcohol consumption among adults (litr);
- Input 2 – Death by neoplasms (Years lost, /100 000 population, aged 75 years old; Potential years of life lost);
- Output 1 – Health expenditure as a share of GDP (%);
- Output 2 – Life expectancy (years);
- Output 3 – Percentage of people if good health (%).

Inputs are variables which are the risk variables and may be changed – restriction or the preventive inspections masy be supported more. The outputs are variable which cannot be changed easily. The main output factor is life expectancy which can be used as the first detection that the country has good health care (but long life does not mean good life, that is why the Output 3 is in the model). The BCC input model is chosen according to the type of variable. The variable return to scale is chosen because the different type of health care financing over the European countries.

Inputs and outputs had been chosen according to the analysis of Health at a Glance: Europe 2018[1]. Health expenditure as a share of GDP should be according to their analysis maximised. This is the reason why Health expenditure as a share of GDP is define as the output. The rest of the variables are chosen according to the general logic. In future analysis, the change may be done and Health expenditure as a share of GDP may be define as the input.

The required data set for input and output variables have been collected from OECD data[2]. There are 27 countries in the analysis: Austria (country 1), Belgium (country 2), Czech Republic (country 3), Denmark (country 4), Estonia (country 5), Finland (country 6), France (country 7), Germany (country 8), Greece (country 9), Hungary (country 10), Iceland (country 11), Ireland (country 12), Italy (country 13), Latvia

[1] https://www.oecd-ilibrary.org/sites/health_glance_eur-2018-en/index.html?itemId=/content/publication/health_glance_eur-2018-en

[2] https://stats.oecd.org/index.aspx?queryid=30127

(country 14), Lithuania (country 15), Luxembourg (country 16), Netherlands (country 17), Norway (country 18), Poland (country 19), Portugal (country 20), Slovak Republic (country 21), Slovenia (country 22), Spain (country 23), Sweden (country 24), Switzerland (country 25), Turkey (country 26) and United Kingdom (country 27). The data are taken for the time period from 2010 to 2017.[3]

4 Analysis and application

Using the data for these 27 European countries, the DEA model (BCC-I) had been run for the time period 2010 to 2017. Table 1 below lists some descriptive statistics of the DEA efficiency scores for these years. As shown in Table 1, the number of efficient countries in health care is decreasing. On the other hand, the average efficiency is slightly increasing. Ireland, Norway, Sweden, Switzerland and Turkey re efficient during the whole time period. First four countries are probably not surprising, but in case on Turkey. This is debatable why this is the case and whether other methodologies or skewed analyses are used.

	2010	2011	2012	2013	2014	2015	2016	2017
min	0.4521	0.4452	0.4539	0.4526	0.4659	0.4723	0.4215	0.4681
max	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
average	0.8377	0.7953	0.8018	0.8105	0.8100	0.8027	0.7466	0.7835
median	0.8934	0.7818	0.8078	0.7980	0.8051	0.7804	0.7093	0.7670
# efficient countries	10	7	7	9	9	7	7	5

Table 1 Descriptive statistics of the DEA efficiency scores for each year (source: own calculation in GAMS)

Using the DEA efficiency scores as an input and life expectancy as the primary output variable, the regression model is ran. The panel regression equation is[4]:

$$y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + \beta_4 x_{4it} + \beta_5 x_{5it} + \epsilon_{it}, \tag{6}$$

where $t = 2010, \ldots, 2017$, i are 27 European countries and

- y_{it} = Output 2 – Life expectancy (years),
- x_{1it} = Input 1 – Alcohol consumption among adults (litr),
- x_{2it} = Input 2 – Death by neoplasms (Years lost, /100 000 population, aged 75 years old; Potential years of life lost),
- x_{3it} = Output 1 – Health expenditure as a share of GDP (%),
- x_{4it} = Output 3 – Percentage of people if good health (%),
- x_{5it} = DEA efficiency.

To decide between fixed or random effects a Hausman test has been run [5]. As expected, the model with fixed effects been detected as the good.

Table 2 summarise the panel regression models results. There is a model without x_{5it}. This model is called N/DEA model. The second model, the model with x_{5it} is called W/DEA model. The α_i is still dummy variable which expresses the intensity of the differences in the level constant of the country "country i" – Belgium, ..., United Kingdom ($i \neq 1$) compared to the base country – Austria. The intensity of the differences is also seen in Table 2. The legend for the Table 2 is following: * p<0.05; ** p<0.01; *** p<0.001.

For both types of models there had been done three calculations:

- fixed effects using n entity-specific intercepts (xtreg) – fixed;
- fixed effects using least squares dummy variable model (LSDV) – ols;
- fixed effects using another n entity-specific intercepts (areg) – areg.

The calculation for both models are closed for each other, but always the model W/DEA has little higher R^2 and R^2_{adj}[5]. The reason it that the special new variable x_{5it} is not statistically significant at lower levels so it does not bring a lot of info for the whole model. Generally speaking, the calculations ols and areg were giving the similar results and high R^2 around 97%. The calculation fixed had R^2 around 42% in both case. This big difference may be reason to try different type of model – change the variables.

[3] We wanted to have longer time period, but there had been missing data. Also we wanted to add another input – Daily smoking among adults (Grammes per capita), but there had not been data for half of the countries.

[4] Also, depends on the type of model.

[5] R^2 shows the amount of variance of Y explained by X

variable	fixed	ols	areg	variable	fixed	ols	areg
Alcohol consumption	-0.2810^{***}	-0.2810^{***}	-0.2810^{***}	Alcohol consumption	-0.2685^{***}	-0.2685^{***}	-0.2685^{***}
Death by neoplasms	-0.0040^{***}	-0.0040^{***}	-0.0040^{***}	Death byneoplasms	-0.0039^{***}	-0.0039^{***}	-0.0039^{***}
Health expenditure	-0.1589^{*}	-0.1589^{*}	-0.1589^{*}	Health expenditure	-0.1589^{*}	-0.1589^{*}	-0.1589^{*}
Good health	0.0087	0.0087	0.0087	Good health	0.0093	0.0093	0.0093
Belgium		-0.5820		DEA efficiency	-0.8635	-0.8635	-0.8635
Czech Republic		-2.2716^{***}		Belgium		-0.5879	
Denmark		-1.2030^{***}		Czech Republic		-2.3973^{***}	
Estonia		-3.6187^{***}		Denmark		-1.2014^{***}	
Finland		-2.3422^{***}		Estonia		-3.7508^{***}	
France		2.0048^{***}		Finland		-2.1195^{***}	
Germany		-0.3101		France		2.0750^{***}	
Greece		-1.5898^{**}		Germany		-0.2981	
Hungary		-1.6121^{*}		Greece		-1.5118^{**}	
Iceland		-1.3634^{**}		Hungary		-1.9777^{**}	
Ireland		-0.4282		Iceland		-1.1248^{*}	
Italy		-0.1416		Ireland		-0.2458	
Latvia		-5.4416^{***}		Italy		-0.0766	
Lithuania		-4.2952^{***}		Latvia		-5.6676^{***}	
Luxembourg		-0.7599		Lithuania		-4.5722^{***}	
Netherlands		-0.2393		Luxembourg		-0.6713	
Norway		-1.8617^{***}		Netherlands		-0.2070	
Poland		-2.7474^{***}		Norway		-1.5982^{**}	
Portugal		0.0691		Poland		-2.9454^{***}	
Slovak Republic		-3.5132^{***}		Portugal		-0.0021	
Slovenia		-0.2164		Slovak Republic		-3.7060^{***}	
Spain		0.7237		Slovenia		-0.3225	
Sweden		-1.5675^{**}		Spain		0.8946^{*}	
Switzerland		0.2805		Sweden		-1.3114^{*}	
Turkey		-8.6249^{***}		Switzerland		0.5037	
United Kingdom		-1.0248^{**}		Turkey		-8.3078^{***}	
constant	89.4682^{***}	91.0488^{***}	89.4682^{***}	United Kingdom		-1.0021^{**}	
				constant	89.8824^{***}	91.4372^{***}	89.8824^{***}
N	216	216	216	N	216	216	216
R^2	0.4182	0.9741	0.9741	R^2	0.4229	0.9744	0.9744
R^2_{adj}	0.3238	0.9699	0.9699	R^2_{adj}	0.3257	0.9700	0.9700

Table 2 Panel regression for N/DEA model and W/DEA model (source: own calculation in STATA)

Based on calculation "ols", Table 2 shows that not all countries has a significant influence on dependent – Life expectancy. For example, Belgium, Germany, Ireland, Italy, Luxembourg, Netherlands, Portugal, Slovenia, Spain and Switzerland did not have a significant influence on Life expectancy. This is very interesting in the case of Ireland and Switzerland, given that these countries have been shown to be effective in the DEA analysis. Also, the influence of x_{1it} – x_{3it} is interesting. More precisely, the first two variables were identified as inputs, so the negative effect is logical, but the effect of the third variable is surprising. x_{3it} – health expenditure as a share of GDP (%) should be look after.

5 Conclusion

In this paper, a new panel regression forecasting methodology to forecast comparable units is used. The approach included in the panel regression analysis a surrogates measure of the unique weight of variables and of performance. This new variable is the relative efficiency of each comparable unit that is generated by DEA. The results of applying this new regression forecasting methodology including a DEA efficiency variable to a data set demonstrated that this does not provides an enhanced approach to forecasting comparable units every time as it was in work before. The study case was done fot health care for European countries. As the results are not constant, there is plan to perform further testing with other data sets from same industry, with more comparable units, more variables, different definitions of input and output and possibly more years of data. Also, the DEA for missing date may be used as the problem is the missing data in important factor – tobacco consumption.

Acknowledgements

This article was supported by the Czech Science Foundation within the project GA19-13946S.

References

[1] Banker, R. D., Charnes, A. and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, **30**, 1078–1092.

[2] Bedir, S. (2016). Healthcare Expenditure and Economic Growth in Developing Countries. *Advances in Economics and Business*, **4**, 76–86.

[3] Charnes, A., Cooper, W. W. and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, **2**, 429–444.

[4] Gangl, M. (2020). Causal inference in sociological research. *Annual Review of Sociology*, **36**, 21–47.

[5] Greene, W. H. (2008). *Econometric analysis*. 6th ed., Upper Saddle River, N.J. : Prentice Hall.

[6] Jakovljevic, M. B., Vukovic, M. and Fontanesi, J. (2016). Life expectancy and health expenditure evolution in Eastern Europe-DiD and DEA analysis. *Expert Review of Pharmacoeconomics & Outcomes Research*, **4**, 537–546.

[7] Klimberg, R. K. et al. (2009). Using regression and Data Envelopment Analysis (DEA) to forecast bank performance over time. *Applications of Management Science*, **13**, 133–142.

[8] Klimberg, R. K., Lawrence, K. D. and Lawrence, S. M. (2008). Improved Performance Evaluation of Comparable Units with Data Envelopment Analysis (DEA). *Advances in Business and Management Forecasting*, **5**, 65–75.

[9] Klimberg, R. K., Lawrence , K. D. and Lawrence, S. M. (2005). Forecasting Sales of Comparable Units with Data Envelopment Analysis (DEA). *Advances in Business and Management Forecasting*, **4**. 201–214,

[10] Klimberg, R. K., Lawrence, K. D. and T. Lal. (2008). *Using Data Envelopment Analysis (DEA) to Forecast Bank Performance*. NEDSI, New York, NY.

[11] Lightwood, J., and Glantz S. A. (2016). Smoking Behavior and Healthcare Expenditure in the United States, 1992–2009: Panel Data Estimates. *PLOS Medicine*, **13**.

An Aggregation-Based Procedure for Ranking of Efficient Units in DEA Models

Josef Jablonský[1]

Abstract. Data envelopment analysis (DEA) models evaluate a set of decision-making units (DMUs) according to the efficiency of transformation of multiple inputs into multiple outputs. In DEA models, the DMUs obtain efficiency scores that allow their ranking for inefficient ones only, because the efficient DMUs have the same maximum efficiency score. In this paper, we propose a procedure for ranking of efficient units based on the aggregation of individual rankings obtained by partial efficiency measures defined as all possible ratios an output divided by input. A mixed-integer linear optimization model that aggregates individual rankings is introduced and illustrated on an example. The same approach can be applied for aggregation of rankings of alternatives obtained by various multiple criteria decision-making (MCDM) methods.

Keywords: data envelopment analysis, multiple criteria decision-making, ranking, aggregation, efficiency

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

DEA models are widely applied for efficiency and performance evaluation of a set of DMUs. Since 1978, when the first model of this category was introduced by Charnes et al. [3], many modifications have been formulated by various researchers. In current studies, among the most used DEA models standard radial models – [2, 3] – and slacks-based measure (SBM) models, formulated by Tone [8], belong. The envelopment form of the CCR (derived from Charnes, Cooper and Rhodes) input-oriented model for evaluation of DMU_q is formulated as follows:

$$
\begin{aligned}
\text{Minimize} \quad & \theta_q \\
\text{subject to} \quad & \sum_{i=1}^{n} x_{ij}\lambda_i \le \theta_q x_{qj}, && j = 1,\dots,m, \\
& \sum_{i=1}^{n} y_{ik}\lambda_i \ge y_{qk}, && k = 1,\dots,r, \\
& \lambda_i \ge 0, && i = 1,\dots,n,
\end{aligned}
\tag{1}
$$

where n is the number of DMUs, m is the number of inputs with input values x_{ij}, $i = 1, \dots, n$, $j = 1, \dots, m$, r is the number of outputs with output values y_{ik}, $i = 1, \dots, n$, $k = 1, \dots, r$, and λ_i, $i = 1, \dots, n$, are the weights of DMUs. Model (1) looks for efficient target inputs $\sum_{i=1}^{n} x_{ij}\lambda_i$ and outputs $\sum_{i=1}^{n} y_{ik}\lambda_i$ and minimizes the scalar variable θ_q that reaches its maximum value $\theta_q = 1$ in case that the unit under evaluation is on the efficient frontier. i.e. it is at least weakly efficient. $\theta_q < 1$ indicates inefficiency. It is a radial measure of efficiency because all inputs of the model have to be reduced by θ_q to reach the efficient frontier. Another option of constructing

[1] University of Economics, Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, Praha 3, Czech Republic, e-mail: jablon@vse.cz

the efficient frontier is based on minimization of slacks (SBM models). A typical representative of this group of models is Tone's SBM model. Its general form is as follows:

Minimize

$$\rho_q = \frac{1 - \frac{1}{m}\sum_{j=1}^{m}\left(s_j^- / x_{qj}\right)}{1 + \frac{1}{r}\sum_{k=1}^{r}\left(s_k^+ / y_{qk}\right)},$$

subject to

$$\sum_{i=1}^{n} x_{ij}\lambda_i + s_j^- = x_{qj}, \qquad j = 1, \ldots, m, \tag{2}$$

$$\sum_{i=1}^{n} y_{ik}\lambda_i - s_k^+ = y_{qk}, \qquad k = 1, \ldots, r,$$

$$\lambda_i \geq 0, \qquad i = 1, \ldots, n$$

Model (2) is not linear in its objective function but can be linearized by Charnes-Cooper transformation easily. The unit under evaluation is efficient if $\rho_q = 1$, lower values indicate inefficiency.

Both typical representatives of most applied DEA models are not able distinguishing among efficient units. That is why their super-efficiency modifications have been formulated. In super-efficiency versions, both models (1) and (2) are extended by a new constraint $\lambda_q = 0$. This extension causes the optimal objective function of super-efficiency models is greater than one, which allows the ranking of initially efficient DMUs. More about (not only) ranking models in DEA can be found in [4].

The idea of this paper was partly motivated by Mohammadi and Rezaei [6]. The authors present an interesting approach for aggregation of individual rankings based on half-quadratic theory. For the same purpose, we introduce an optimization model that results in a unique aggregated ranking. This approach is applied for ranking of efficient units in DEA based on partial efficiency measures. Section 2 contains the formulation optimization models for aggregation of rankings and discusses their use for various purposes. Section 3 illustrates the application of the models for ranking of efficient units and compares the results with other ranking models. The final section discusses further research and concludes the paper.

2 Aggregation of individual rankings

Let us suppose n individual rankings of m units (DMUs, alternatives) where r_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$, is the rank of the i-th unit in the j-th ranking. The aim of the following optimization model is to find an aggregated ranking of all units, i.e. a vector $\mathbf{x} = (x_1, \ldots, x_m)^T$ with the elements that express an aggregated rank of the i-th unit. The model that minimizes the weighted sum of deviations of the aggregated ranking from all individual rankings is as follows:

Minimize

$$\sum_{i=1}^{m}\sum_{j=1}^{n} w_j (d_{ij}^- + d_{ij}^+) \tag{3}$$

subject to

$$
\begin{aligned}
& r_{ij} + d_{ij}^{-} - d_{ij}^{+} = x_i, && i = 1,\dots,m, j = 1,\dots,n, \\
& \sum_{j=1}^{n} y_{ij} = 1, && i = 1,\dots,m, \\
& \sum_{i=1}^{m} y_{ij} = 1, && j = 1,\dots,n, \\
& x_i = \sum_{j=1}^{n} j.y_{ij}, && i = 1,\dots,m, \\
& y_{ij} - binary, &&
\end{aligned}
\tag{4}
$$

where d_{ij}^{-}, d_{ij}^{+} are negative and positive deviations of the elements of the aggregated ranking from the elements of all individual rankings, w_j, $j = 1, \dots, n$, is the weights of the j-th individual ranking (typically, all rankings have the same weights). Binary variables y_{ij}, $i = 1, \dots, m, j = 1, \dots, n$, are used for ensuring the uniqueness of the final ranking (each rank is contained in the aggregated ranking just once). Their elements $y_{ij} = 1$ express that the i-th unit is on the j-th rank in the final ranking.

Let us denote D_j, $j = 1, \dots, n$, the maximum deviation of the elements of the aggregated ranking from the elements of the j-th individual ranking. Then, it is possible to derive the aggregated ranking by minimization of the sum (or the weighted sum) of maximum deviations. The objective function of this model is the following:

Minimize

$$
\sum_{j=1}^{n} w_j D_j \tag{5}
$$

In this model, the set of constraints (4) must be extended by the additional set of constraints

$$
d_{ij}^{-} + d_{ij}^{+} \le D_j, \qquad i = 1,\dots,m, j = 1,\dots,n, \tag{6}
$$

ensuring that the absolute value of all deviations is less than or equal than the variable D_j. This second model often has multiple optimum solutions. That is why, it is reasonable to apply a lexicographic approach – in the first step to optimize the objective function (5), and in the second step to minimize the sum of deviations (3) with the additional constraint

$$
\sum_{j=1}^{n} w_j D_j = D^*
$$

where D^* is the optimum value of the objective function obtained in the first step.

The proposed models may be applied in all cases where it is necessary for adopting a final unique decision based on individual preferences among several units. The following 3 cases demonstrate possible applications of the presented procedures:

1. In the introductory section of the paper, we have discussed that traditional DEA models are not able to distinguish among efficient unit because of their identical maximum efficiency score. Many models have been proposed in the past to rank efficient units in DEA models. Those models are based on various principles and usually generate different results. In general, the rankings of efficient DMUs produced by various models are not identical. Our procedure can be applied for aggregation of those individual rankings into one final ranking.
2. The above optimization procedures may be used for definition of an own measure that allows ranking of efficient (or even all) DMUs. Considering a DEA model with m inputs and r outputs, it is possible to define $m \cdot r$ partial efficiency measures as all ratios an output divided by input. It is natural ranking all efficient DMUs under evaluation according to those partial efficiency measures. Then, those partial rankings may be aggregated into one final ranking using the proposed procedures.
3. The proposed procedures may be applied for aggregation of individual rankings in traditional MCDM problems – evaluation of the finite set of alternatives by multiple criteria. There are many MCDM methods for evaluation of alternatives (TOPSIS, ELECTRE and PROMETHEE class methods,

AHP, VIKOR, and others). In general, their application often leads to different rankings. They can be aggregated to one final ranking using the models introduced above.

3 Numerical illustration

For an illustration of both models, an example taken from [5] is considered. The source dataset contains characteristics of 194 DMUs – bank branches of one of the Czech commercial banks. Evaluation of the efficiency of the branches by CCR model (1) leads to the identification of 12 efficient units. The data set for the efficient units can be found in [5]. The following inputs and outputs have been considered:

- Total operational costs in thousands of CZK per year (I_1).
- The number of inhabitants within the region of the branch (I_2).
- The number of employees (I_3),
- Value of credits in millions of CZK (O_1),
- The total number of accounts (O_2).

Table 1 contains all six ratios of output/input calculated from the source dataset for all efficient DMUs.

DMU	O1/I1	O1/I2	O1/I3	O2/I1	O2/I2	O2/I3
26	0.90	2.01	18.39	158.38	0.35	3227.43
28	0.81	1.74	29.72	143.34	0.31	5242.80
37	0.94	0.42	33.33	133.90	0.06	4739.50
71	0.58	3.56	16.89	133.26	0.82	3896.29
79	0.52	5.72	22.00	95.10	1.05	4021.85
82	0.46	9.47	14.40	101.48	2.08	3170.80
83	0.55	5.72	14.50	123.70	1.29	3282.80
105	0.81	2.32	18.47	148.93	0.43	3390.00
133	0.68	6.52	18.98	115.96	1.11	3241.33
147	0.38	40.32	10.77	73.80	7.89	2107.33
182	0.80	4.63	28.77	116.56	0.67	4189.71
184	0.41	7.34	10.70	118.34	2.12	3091.00

Table 1 Partial efficiency measures

The results of the optimization models (3)–(6) are presented in Table 2. The first six columns contain rankings according to the partial efficiency scores. The last two columns show the aggregated ranking derived using the minimization of the sum of deviations (SUM) and the minimization of the sum of maximum deviations (MM). The last two rows of Table 2 present the values of both criteria for both models. In the first model, the minimum sum of deviations is 194, and the sum of the maximum deviations is 42. In the second model, the optimum objective function is 34, and the corresponding sum of deviations is 214.

DMU	O_1/I_1	O_1/I_2	O_1/I_3	O_2/I_1	O_2/I_2	O_2/I_3	SUM	MM
26	2	10	7	1	10	9	9	5
28	3	11		3	11	1	2	1
37	1	12	1	4	12	2	1	2
71	7	8	8	5	7	5	7	8
79	9	5	4	11	6	4	4	6
82	10	2	10	10	3	10	10	10
83	8	6	9	6		7	8	9
105	4	9	6	2	9	6	6	7

133	6	4	5	9	5	8	5	4
147	12	1	11	12	1	12	12	11
182	5	7	3	8	8	3	3	3
184	11	3	12	7	2	11	11	12
The total (weighted) sum of deviations							194	214
The total sum of (weighted) maximum deviations							42	34

Table 2 Ranking of DMUs according to their partial efficiency measures

The results presented in Table 2 are compared with rankings of the same units obtained by various models for ranking of efficient DMUs in DEA models as presented in [5]. The following ranking models have been applied in this study:

- AP – Andersen and Petersen super-efficiency model with the assumption of constant returns to scale technology [1].
- SBMT – Tone's SBM general super-efficiency model [9].
- OPT/PES – Optimistic and pessimistic super-efficiency measure [10].
- CROSS – Cross-efficiency approach [7].
- SBMG – Goal programming super-efficiency model [5].
- AHP – AHP based super-efficiency measure [5].

Ranking of the DMUs according to those six models is shown in Table 3. This table also contains an aggregated ranking obtained by two models presented in this paper.

Table 3 shows that the ranking obtained by AP, SBMT, SBMG, and partly AHP, are very close to each other. On the contrary, results given by OPT/PEC concept and cross-efficiency approach are somewhat different. Due to a higher number of the group of models, both aggregation procedures lead to the rankings that are similar to traditional models (AP, SBMT). Comparison of aggregated results in Tables 2 and 3 demonstrates a high level of differences in both cases. Especially, DMU #147 is ranked as the best by all traditional ranking models, and also among the best DMUs in the aggregation of six rankings. The same unit is the worse or nearly worse by aggregation of partial efficiency measures (Table 2). This aggregation result is an expected outcome because this unit is the worse or almost worse by all partial measures except O_1/I_2 and O_2/I_2. This unit has a very low input I2 in comparison with other DMUs which leads to an extremely high level of two partial efficiency ratios. Due to this level, traditional super-efficiency models identify DMU #147 as extremely efficient, but the models based on the aggregation of partial rankings rank this unit as one of the worse.

DMU	AP	SBMT	OPT/PES	CROSS	SBMG	AHP	SUM	MM
26	6	9	7	7	9	7	8	8
28	2	3	2	6	3	3	3	1
37	3	4	1	12	4	8	4	4
71	9	10	8	3	10	11	10	10
79	7	7	4	8	7	6	7	7
82	8	6	10	9	6	2	6	6
83	11	11	9	4	11	10	11	11
105	12	12	6	5	12	12	12	12
133	10	8	5	2	8	9	9	9
147	1	1	12	11	1	1	1	3
182	4	2	3	1	2	5	2	2
184	5	5	11	10	5	4	5	5
The total (weighted) sum of deviations							134	146

The total sum of (weighted) maximum deviations	29	27

Table 3 Ranking of DMUs based on various models and its aggregation

4 Conclusions

The paper is a contribution to the approaches that allow ranking of efficient units in DEA models. The presented procedures are based on the aggregation of rankings of the DMUs according to their partial efficiency measures that are defined as all ratios of outputs and inputs. The procedures can be used not only for ranking of efficient units. They can find their applications in all cases where it is necessary to obtain a unique ranking of any units that are characterized by several partial rankings.

The proposed models can be extended in several directions which can be a good starting point for future research. In the current version of the models, the partial rankings are strictly generated by the original data, i.e. even very tiny differences in source data lead to different rankings. A possible solution to avoid this unreasonable situation may consist in considering indifference thresholds of a certain level. The DMUs with the differences lower than those thresholds may be considered as identical and will have the same rank. Some other extension can improve or enrich the current models.

Acknowledgements

The research is supported by the Czech Science Foundation, project no. 19-08985S – Models for efficiency and performance evaluation in non-homogeneous economic environment.

References

[1] ANDERSEN, P. & PETERSEN, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science,* 39**(10)**, 1261–1264.

[2] BANKER, R. D., CHARNES, A., & COOPER, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science,* 30**(9)**, 1078–1092.

[3] CHARNES, A., COOPER, W. W., & RHODES, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2**(6)**, 429–444.

[4] DLOUHÝ, M., JABLONSKÝ, J., & ZÝKOVÁ, P. (2018). *Data Envelopment Analysis* (in Czech). Praha, Professional Publishing.

[5] JABLONSKÝ, J. (2012). Multicriteria approaches for ranking of efficient units in DEA models. *Central European Journal of Operations Research,* 20**(3)**, 435–449.

[6] MOHAMMADI, M. & REZAEI, J. (2020). Ensemble ranking: Aggregation of rankings produced by different multiple-criteria decision-making methods. *Omega,* In press.

[7] SEXTON, T. R., SILKMAN, R. H., & HOGAN, A. J. (1986). Data envelopment analysis: Critique and extensions. In: SILKMAN, R. H. (ed.) *Measuring efficiency: An assessment of data envelopment analysis*. CA: Jossey-Bass, San Francisco.

[8] TONE, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research,* 130**(3)**, 498–509.

[9] TONE, K. (2003). A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research,* 143**(1)**, 32–41.

[10] WANG, Y. M., CHIN, K. S., & YANG, J. B. (2007). Measuring the performances of decision making units using geometric average efficiency. *Journal of the Operational Research Society,* 58**(7)**, 929–937.

Adaptive Path-Relinking Method for Public Service System Design

Jaroslav Janáček[1], Marek Kvet[2]

Abstract. Public service system design problem can be modelled as a kind of the weighted *p*-median problem. To solve real-sized instances of the *p*-median problem in acceptable computational time, a family of heuristics and metaheuristics have been implemented and tested. Surprisingly, simple incrementing heuristics as the swap algorithm or path-relinking method proved to be able to reach a near-to-optimal solution in relatively short time, even if they were applied to the weighted *p*-median problem with the generalized objective min-sum function. Having analyzed performance of the mentioned heuristics, we found that the path-relinking method wan as concerns accuracy, but it paid for better accuracy by bigger computational time caused by inefficient search along the shortest paths. In this contribution, we focus on accelerating the path-relinking method by an adjustment of the searching process. The adjustment consists in application of adaptive approach, which enables to avoid the inspection of unprofitable parts of the inspected paths and thus it reduces the computational time necessary to reach the final solution.

Keywords: public service system design, path-relinking method, adaptive approach

JEL Classification: C44
AMS Classification: 90C06, 90C10, 90C27

1 Introduction

Public service systems are usually designed to provide the system users with more secure life or to guarantee necessary service in case of emergency. Since the resources, from which the service is provided, are limited, the associated mathematical model often takes the form of a weighted *p*-median problem, which belongs to the family of so-called location problems with wide spectrum of applications in various areas of life [2, 3, 7, 13, 14].

To make the system more flexible and realistic, the concept of generalized disutility has been introduced [3, 8, 11, 15]. It follows from the idea that the service does not have to be necessarily provided only from the nearest located service center, but from such nearest center, that is currently available. This way, the stochastic behavior of the real system can be partially incorporated into the model. Mentioned adjustment can make the model more complicated for effective solution.

The necessity of solving large instances has led to the development of many exact [1, 5, 11] and heuristic approaches [4, 16]. Currently, the attention is paid to metaheuristic methods, i.e. genetic algorithms, scatter search, path-relinking method and many others with the goal to obtain a good solution in a short time [6, 17].

In this paper, we focus on special kinds of the path-relinking method, in which we try to accelerate the computational process of the method. We suggest and compare three schemes of the adaptive approach to the generalized weighted *p*-median problem solving, in which we study the result accuracy and computational time. Our suggestions are experimentally verified using real sized instances obtained from the road network of Slovakia.

[1] University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovakia, jaroslav.janacek@fri.uniza.sk
[2] University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovakia, marek.kvet@fri.uniza.sk

2 Path-Relinking Search Based Approaches

2.1 Generalized *p*-Median Problem and Uniformly Deployed Sets

The generalized weighted p-median problem is formulated as a task to service n users so that an average distance traveled due to service be minimal. A user j=1 … n has frequency b_j of demands for the service. In the problem, m possible facility locations are considered and exactly p of them are to be chosen for service center location. Standard p-median problem assumes that a user is serviced by the nearest facility, but the generalized problem formulations consider r nearest facility locations for each user and assume that the user is serviced by the k-th nearest facility with probability q_k. If the symbol d_{ij} denotes the distance between locations i and j and operation $\min_k\{\}$ denotes the k-th minimal value of the list of values in the brackets, and having described a solution by list P of p locations chosen from m possible facility locations, the problem can be described by (1).

$$\min\left\{\sum_{j=1}^{n} b_j \sum_{k=1}^{r} q_k \min_k\left\{d_{ij}: i \in P\right\}, P \subset \{1,2,\dots,m\}, |P| = p\right\} \tag{1}$$

The original version of the path-relinking method searches one of the shortest paths, which connects two given input solutions and returns the best feasible solution, which is met during the search [6]. The inspected path is considered in the graph, which vertices correspond to vertices of a unit m-dimensional hypercube and a graph edge is represented by an edge which connect two neighboring hypercube vertices on the hypercube surface. Obviously, the neighboring hypercube vertices represented by zero-one m-dimensional vectors differs only in one component, which takes the value one in one solution and the value of zero in the other one. This way, each even vertex in the shortest path inspected by the original version of the path-relinking method does not belong to the set of feasible solutions of (1), because the cardinality of the associated set P s either $p-1$ or $p+1$. The suggested version of path-relinking method below inspects the shortest path in the graph, which vertices correspond only to feasible solution of (1) and the which edges connect only the vertices, which Hamming distance equals to two. Efficiency of an algorithm using any version of the path-relinking methods is influenced by an initial set of solutions, from which the pairs of solutions are chosen, to inspect the connecting shortest path. In our contribution, we used uniformly deployed set of solutions instead randomly generated starting set of solutions.

The uniformly deployed set of p-location problem solutions can be defined as a subset S of all feasible solutions that Hamming distance of each pair of solutions of S is greater than or equal to a given threshold h and have cardinality exceeding a given limit. As formal construction of a maximal uniformly deployed set does not depend on the specific instance of the p-location problem, but it depends only on the numbers p, m and h, the once constructed set can be used for a broad class of the p-location problems. We make use of the results published in [9, 16], which yield near-to maximal uniformly distance sets with suitable cardinality for used benchmarks.

2.2 Mixed Truncated Path-Relinking Method

The path-relinking method searches one of the shortest paths, which connects two given input solutions in general and returns the best feasible solution, which is met during the search. The suggested version inspects the path, which consists only of feasible solutions of the p-median problem. To describe the suggested version, we will use the following denotations. The objective function value $f(P)$ is described by (2).

$$f(P) = \sum_{j=1}^{n} b_j \sum_{k=1}^{r} q_k \min_k\left\{d_{ij}: i \in P\right\} \tag{2}$$

Operation $argmin\{f(P), f(Q)\}$ returns the argument P or Q, which has the least function value and operation $argmin\{f((P-\{i\})\cup\{g\}): i\in U, g\in V\}$ returns the pair (i^*, g^*), for which the expression $f((P-\{i\})\cup\{g\})$ takes the least value. In addition, we introduce two sets U and V of locations for the input solutions P and Q. Set U contains the selected locations of P, which are not contained in Q and set V contains the selected locations of Q, which are not contained in P. Briefly $U = P - Q$ and $V = Q - P$. The newly suggested *Path-relinking* method performs according to the steps described below.

Algorithm *Path-relinkingTruncated(P, Q, α)*

0. Initialize P^* by $argmin\{f(P), f(Q)\}$, determine $U = P - Q$ and $V = Q - P$ and set $e=\lceil \alpha|V| \rceil$.

1. While *Cont1*(U, e) or *Cont2*(V, e) perform the following steps otherwise terminate and return P^*.

2. If *Cont1*(*U, e*) then perform

2.1 Determine locations i^* and g^* by $(i*, g*) = argmin\{f((P - \{i\}) \cup \{g\}): i \in U, g \in V\}$.

2.2 Update *P*, *U*, *V*, and P^* according to $P = (P - \{i*\}) \cup \{g*\}$,
$U = U - \{i*\}, V = V - \{g*\}, P* = argmin\{f(P), f(P*)\}$.

3. If *Cont2*(*V, e*) then perform

3.1 Determine locations i^* and g^* by $(i*, g*) = argmin\{f((Q - \{i\}) \cup \{g\}): i \in V, g \in U\}$.

3.2 Update *Q*, *U*, *V*, and P^* according to $Q = (Q - \{i*\}) \cup \{g*\}$,
$U = U - \{g*\}, V = V - \{i*\}, P* = argmin\{f(Q), f(P*)\}$.

The control Boolean functions *Cont1*(*U, e*) and *Cont2*(*V, e*) are defined in the following way.

Cont1(*U, e*): If $|U| > e$, then return true, else return false.

Cont2(*V, e*): If $|V| > e$, then return true, else return false.

These two functions allow the algorithm to inspect a part of the shortest path from the input solution *P* to the input solution *Q*. The inspected part of the shortest path consists of two sections and each section is adjacent to one of the input solutions. Let us express the length of the path by a number of solutions forming the path. Thus, the uninspected part of the shortest path is located around its center and the length of the uninspected part is *e*. If the parameter *e* equals to one, then each solution of the shortest path is inspected.

2.3 Adaptive Breaking Rule

As noted above, the mixed path-relinking method inspects the shortest path starting from its both input ends and stops the inspection, when a given threshold is reached. The suggested adaptive threshold breaking rule allows continuing with the inspecting, when the inspection of the current end of the partial path improved the best-found solution. Inspection of the promising partial path continues, whilst inspection of the other partial path is terminated. The inspection of the promising partial path continues until its current end inspection brings improvement. To modify the algorithm *Path-relinkingTruncated* to *Path-relinkingTruncatedBroken*, the Boolean variables *improv1* and *improv2* are introduced and their initialization and updating procedures are added. Then, the algorithm is described below.

Algorithm *Path-relinkingTruncatedBroken(P, Q, α)*

0. Initialize P^* by *argmin*{*f*(*P*), *f*(*Q*)}, determine $U = P - Q$ and $V = Q - P$ and set $e = \lceil \alpha |V| \rceil$.
Perform *InitializeB*(*improv1, improv2*).

1. While *ContB1*(*U, e, improv1*) or *ContB2*(*V, e, improv2*) perform the following steps otherwise terminate and return P^*.

2. If *ContB1*(*U, e, improv1*) then perform

2.1 Determine locations i^* and g^* by $(i^*, g^*) = argmin\{f((P-\{i\}) \cup \{g\}): i \in U, g \in V\}$.

2.2 Update *P*, *U*, *V*, and P^* according to $P=(P-\{i^*\}) \cup \{g^*\}$,
$U = U - \{i*\}, V = V - \{g*\}, P^* = argmin\{f(P), f(P^*)\}$.
If $f(P) > f(P*)$ then *improv1=false*, else *improv1=true*.

3. If *ContB2*(*V,e, improv2*) then perform

3.1 Determine locations i^* and g^* by $(i^*, g^*) = argmin\{f((Q-\{i\}) \cup \{g\}): i \in V, g \in U\}$.

3.2 Update *Q*, *U*, *V*, and P^* according to $Q=(Q-\{i^*\}) \cup \{g^*\}$,
$U = U - \{g*\}, V = V - \{i*\}, P^* = argmin\{f(Q), f(P^*)\}$.
If $f(Q) > f(P*)$ then *improv2=false*, else *improv2=true*.

The initialization procedure is defined as: InitializeB(improv1, improv2): improv1=false, improv2=false.

The control Boolean functions *ContB1*(*U, e, improv1*) and *ContB2*(*V, e, improv2*) are defined in the form:

ContB1(*U, e, improv1*): If ($|U| > e$) or *improv1*, then return true, else return false.

ContB2(*V, e, improv2*): If ($|V| > e$) or *improv2*, then return true, else return false.

2.4 Truncated Adaptive Path-Relinking Method

The algorithm below omits the compulsory inspection of the partial paths. It tracks improvements and fails in updating of the best-found solutions from the beginnings of the partial paths inspection and stops the inspection at the path, when a given threshold is reached or the inspection of the current end of the partial path does not improve the best-found solution. To modify the algorithm *Path-relinkingTruncatedBroken* to *Path-relinkingTruncatedAdaptive*, the Boolean functions *ContB1*(*U, e, improv1*) and *ContB2*(*V, e, improv2*) are redefined to *ContA1*(*U, e, improv1*) and *ContA2*(*V, e, improv2*) respectively and initialization and updating procedures are modified. Furthermore the procedure *InitializeB*(*improv1, improv2*) is replaced with *InitializeA*(*improv1, improv2*). The initialization procedure is defined as: *InitializeA*(*improv1, improv2*): *improv1=true*, *improv2=true*. The control Boolean functions *ContA1*(*U, e, improv1*) and *ContA2*(*V, e, improv2*) are defined in the following way. *ContA1*(*U, e, improv1*): If ($|U| > e$) and *improv1*, then return true, else return false.
ContA2(*V, e, improv2*): If ($|V| > e$) and *improv2*, then return true, else return false.

2.5 Incrementing Heuristic Based on the Path-Relinking Method

The suggested heuristic makes use of a special starting set of p-location problem solutions called the uniformly deployed set described above. If a uniformly deployed set S of the p-location problem solutions is at disposal, then arbitrary kind of the *Path-relinking()* algorithm can be used according to the following scheme. Let us denote general function *Path-Relinking*(P, Q) to represent an arbitrary path-relinking algorithm described in the previous sub-sections. The function returns the resulting solution P^* obtained by partly inspection of the shortest path connecting the input solutions P and Q. We assume that the solutions of the uniformly deployed set $S = \{P^1, \dots, P^{|S|}\}$ are subscripted and ordered by permutation O so that $f(P^{O(1)}) \leq f(P^{O(2)}) \leq \dots \leq f(P^{O(|S|)})$ holds.

Then, the search performs according to the following steps:

0. Initialize P^* by $P^{O(1)}$.

1. For k=2 … $|S|$ perform subsequently P^*=*Path-Relinking*(P^*, $P^{O(k)}$).

2. Terminate and return P^*.

The algorithm performs $|S|$ times the path-relinking method, which inspects the shortest path in maximum $h/2$ steps. At each step the objective function f is enumerated at most $h/2$ times. Thus, the complexity of the above algorithm is $O(|S|*(h)^2)$.

3 Numerical Experiments

The goal of performed computational study is to study efficiency of the suggested path-relinking applications in combination with a uniformly deployed set. We compare results of the adaptive path-relinking modifications.

The used benchmarks were derived from the Slovak self-governing regions. The individual instances are denoted by the names of capitals of the individual regions, which are reported by abbreviations of the region denotations. The list of instances consists of Banská Bystrica (BB), Košice (KE), Nitra (NR), Prešov (PO), Trenčín (TN), Trnava (TT) and Žilina (ZA). The sizes of the individual benchmarks are m and p, where m stands for the number of possible center locations and p is the number of service centers to be located. The constant B reported in Table 1 denotes the total number of service demands computed as a sum of coefficients b_j. The generalized disutility objective function value was computed for r = 3. The associated coefficients q_k for k=1 … r were set in percentage in the following way: $q_1 = 77.063$, $q_2 = 16.476$ and $q_3 = 100 - q_1 - q_2$. These values were obtained from a simulation model of existing EMS system in Slovakia as described in [12]. The uniformly deployed sets were obtained from [9] and their cardinalities $|S|$ and minimal Hamming distance h are reported in Table 1.

To obtain enough robust comparison of the suggested optimization approaches, we made use of the above-mentioned uniformly deployed set property that an arbitrary permutation of m subscripts of locations generates different uniformly deployed set with the same parameters. We generated randomly ten different uniformly deployed sets for each benchmark. Table 1 contains evaluation of the sets as concerns generalized objective function values. We report the following characteristics:

F^{tmin} – the total minimum of objective function of all solutions in ten uniformly deployed sets.

F^{tmax} – the total maximum of objective function of all solutions in ten uniformly deployed sets.

F^{avgmin} – the average of minimal objective function values, where minimum is computed for each set.

F^{avgmax} – the average of maximal objective function values, where maximum is computed for each set.

F^{avg} – the average objective function value of all solutions of ten uniformly deployed sets.

Region	m	p	B	$\lvert S \rvert$	h	F^{tmin}	F^{tmax}	F^{avgmin}	F^{avgmax}	F^{avg}
BB	515	36	6609	172	66	69249	277492	72193	139936	90787
KE	460	32	7929	60	60	66219	217298	71712	136884	90564
NR	350	27	6900	83	50	71327	211015	73547	126075	90185
PO	664	32	8183	232	60	88632	528030	94338	218427	122070
TN	276	21	5942	137	38	53163	245018	54629	111993	71026
TT	249	18	5563	212	32	56017	312195	57487	124472	71270
ZA	315	29	6911	112	52	60361	287821	63631	124394	80913

Table 1 Benchmarks derived from the self-governing regions of Slovakia and characteristics of uniformly deployed sets for individual benchmarks

The experiments were run on a PC equipped with the Intel® Core™ i7 3610QM 2.3 GHz processor and 8 GB RAM. The algorithms were implemented and run in Java language using NetBeans IDE 8.2 environment.

The experiments with the optimization approaches were organized so that three algorithms were used to solve each instance of each problem benchmark, i.e. ten instances of each of seven self-governing regions. To study an impact of the parameter α on the computational time of the specific algorithm, each instance was solved for each setting of α from a predetermined range. The range of α was limited from above by the value of 0.97, which corresponds to the case, when each inspected part of the shortest path contains at least one solution different from the input solutions. Minimal value of α was determined so that it allows to inspect whole shortest path. The modifications of the path-relinking method showed tendencies, which did not change with used benchmarks.

The following Table 2 contains the optimal solutions of studied benchmarks in the part denoted by ***Exact***. The objective function is denoted by *OptSol* and the computational time in seconds is reported in column denoted by *CT*. The objective function values were taken from [10]. The denotation ***PatR*** is used for the simple path-relinking method and ***PatA*** denotes the *Path-relinkingTruncated* with α around zero. The objective function values are reported in the columns denoted by *ObjF* and the computational times in seconds are given in the column denoted by *CT*. The right part of the table contains the results of three suggested path-relinking algorithms. Let ***Alg1*** denote the *AdaptiveBroken* method, in which threshold breaking was forbidden. Similarly, ***Alg2*** denotes the *TruncatedBroken* approach, in which threshold breaking was permitted and finally, ***Alg3*** denotes the *TruncatedAdaptive* algorithm. Parameter α was set to the value 0.7.

Region	***Exact***		***PatR***		***PatA***		***Alg1***		***Alg2***		***Alg3***	
	OptSol	*CT*	*ObjF*	*CT*	*ObjF*	*CT*	*ObjF*	*CT*	*ObjF*	*CT*	*ObjF*	*CT*
BB	44752	10.57	44752	25.94	44752	25.94	44752	15.77	44752	15.90	44780	4.93
KE	45588	7.58	45667	5.60	45667	5.59	45671	3.48	45667	3.54	45673	1.44
NR	48940	19.21	48985	3.72	48985	3.77	48966	2.21	48973	2.23	48985	0.96
PO	56704	76.53	56740	34.71	56740	34.79	56746	21.37	56740	21.48	56777	6.91
TN	35275	4.04	35287	2.33	35287	2.36	35291	1.43	35287	1.44	35291	0.68
TT	41338	2.79	41370	2.23	41370	2.24	41370	1.30	41370	1.32	41488	0.70
ZA	42110	2.70	42118	5.63	42118	5.60	42118	3.13	42119	3.16	42140	1.23

Table 2 Comparison of the results obtained by the path-relinking method to the optimal solutions

The impact of parameter α on studied characteristics (result accuracy and computational time) is reported in Table 3. Here, only the benchmark Žilina (ZA) is used.

α	Alg1		Alg2		Alg3	
	ObjF	CT	ObjF	CT	ObjF	CT
0.50	42119	4.36	42119	4.39	42139	1.24
0.60	42117	3.79	42119	3.84	42140	1.24
0.70	42118	3.13	42119	3.16	42140	1.23
0.80	42127	2.29	42119	2.32	42139	1.20
0.85	42119	1.72	42120	1.82	42139	1.17
0.87	42119	1.48	42120	1.61	42132	1.16
0.90	42135	1.07	42139	1.25	42132	1.10
0.95	42135	0.56	42139	0.75	42132	0.58

Table 3 The impact of parameter α on studied characteristics using the benchmark Žilina (ZA)

As demonstrated by selected results plotted in Tables 2 and 3, the following findings can be stated. When the parameter α is changing in the mentioned range, the final results are very near to the optimal solutions and their negligible fluctuation under 0.5 percent of the best-found solution is small and it is caused probably by random data coincidence during the computational process. Significant differences appeared, when computational times of the individual modifications were compared and their dependence on the parameter α was studied. In all cases, the times monotonously decreased with increasing value of the mentioned parameter α. This effect can be explained by relation between the parameter α and the length of the shortest path part, which is inspected. The coefficient α gives the length of the part excluded from the inspection. As concerns the algorithms *Path-relinkingTruncated* and *Path-relinkingTruncatedBroken*, the decreasing courses of computational time depending on α were almost identical and they decreased more than five times. The adaptive breaking rule applied after inspection of the prescribed part of the shortest path proved to be inactive. Contrary to these partial results, the adaptive approach limited by the threshold α performed by algorithm *Path-relinkingTruncatedAdaptive* behaved in a different way. The adaptive rule stopped the shortest path inspection far from reaching the limit and considerably reduced the computational time, when the parameter α is lower than 0.85.

4 Conclusions

The paper is devoted to the research of various schemes of adaptive approach applications to the generalized weighted *p*-median problem solving, where the core of the solving tool is based on the truncated path-relinking method. Three schemes were studied and their results were compared from the point of result accuracy and computational time. We verified that truncation itself may lead to considerable reduction of computational time necessary for reaching a near-to-optimal solution of the studied problem. In addition, we showed that the adaptive approach could lead to the time reduction, when applied within the limit given by the truncation rule. A further research in this area of path-relinking method usage may be aimed at some other adaptive rules and at possible adaptively controlled reduction of the input uniformly deployed set of solutions.

Acknowledgements

This work was supported by the research grants VEGA 1/0342/18 “Optimal dimensioning of service systems”, VEGA1/0089/19 “Data analysis methods and decisions support tools for service systems supporting electric vehicles”, and VEGA 1/0689/19 “Optimal design and economically efficient charging infrastructure deployment for electric buses in public transportation of smart cities”. This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-19-0441.

References

[1] AVELLA, P., SASSANO, A., & VASIĽEV, I. (2007). Computational study of large scale p-median problems. *Mathematical Programming*, 109, pp. 89–114.

[2] BROTCORNE, L, LAPORTE, G., & SEMET, F. (2003). Ambulance location and relocation models. *European Journal of Operational Research*, 147, pp. 451–463.
[3] CURRENT, J., DASKIN, M., & SCHILLING, D. (2002). Discrete network location models, Drezner Z. et al. (ed) *Facility location: Applications and theory*, Springer, pp. 81–118.
[4] DOERNER, K. F., GUTJAHR, W. J., HARTL, R. F., KARALL, M., & REIMANN, M. (2005). Heuristic Solution of an Extended Double-Coverage Ambulance Location Problem for Austria. *Central European Journal of Operations Research*, 13(4), pp. 325–340.
[5] GARCÍA, S., LABBÉ, M., & MARÍN, A. (2011). Solving large p-median problems with a radius formulation. *INFORMS Journal on Computing*, 23(4), pp. 546–556.
[6] GENDREAU, M. & POTVIN, J. (2010). *Handbook of Metaheuristics*, Springer Science & Business Media.
[7] CHANTA, S., MAYORGA, M. E., & MCLAY, L. A. (2014). Improving emergency service in rural areas: a bi-objective covering location model for EMS systems. *Annals of Operations Research*, 221, pp. 133–159.
[8] JANÁČEK, J. (2019). Smart Search of Neighborhood of p-Location Problem Solution for Generalized Disutility. In: *Zborník zo seminára využitie kvantitatívnych metód vo vedeckovýskumnej činnosti a v praxi XIII*, pp. 59–64.
[9] JANÁČEK, J. & KVET, M. (2019). Uniform Deployment of the p-Location Problem Solutions. *Operations Research Proceedings 2019*, Springer, in print.
[10] JANÁČEK, J. & KVET, M. (2019). Usage of Uniformly Deployed Set for p-Location Min-Sum Problem with Generalized Disutility. In: *SOR 2019 proceedings*, pp. 494–499.
[11] JANÁČEK, J. & KVET, M. (2016). Min-max Optimization and the Radial Approach to the Public Service System Design with Generalized Utility. *Croatian Operational Research Review*, 7(1), pp. 49–61.
[12] JANKOVIČ, P. (2016). Calculating Reduction Coefficients for Optimization of Emergency Service System Using Microscopic Simulation Model. In: *17th International Symposium on Computational Intelligence and Informatics*, pp. 163–167.
[13] JÁNOŠÍKOVÁ, Ľ. & ŽARNAY, M. (2014). Location of emergency stations as the capacitated p-median problem. In: *Quantitative Methods in Economics (Multiple Criteria Decision Making XVII)*, pp. 117–123.
[14] KARATAS, M. & YAKICIA, E. (2019). An analysis of p-median location problem: Effects of backup service level and demand assignment policy. *European Journal of Operational Research*, 272(1), pp. 207–218.
[15] KVET, M. (2014). Computational Study of Radial Approach to Public Service System Design with Generalized Utility. In *Digital Technologies 2014*, pp. 198–208.
[16] KVET, M. & JANÁČEK, J. (2019). Usage of Uniform Deployment for Heuristic Design of Emergency System. *Operations Research Proceedings 2019*, Springer, in print.
[17] RYBIČKOVÁ, A., BURKETOVÁ, A., & MOCKOVÁ, D. (2016). Solution to the locating – routing problem using a genetic algorithm. In: *SmaRTT Cities Symposium Prague (SCSP)*, pp. 1–6.

Verification and Validation of the Dynamic Optimization Model for Forest Conversion Management

Jitka Janová[1]

Abstract. There is a longstanding need for structure change in the spruce dominated forests in the Czech Republic. While the authorities and experts agree that the change must take place, the willingness of forest managers to make substantial changes has been very low so far. As the information on the economic consequences of the conversion in the Czech Republic is missing, we introduce a dynamic optimization model of forest conversion to supply the lacking knowledge and support the decision process.
The model based on optimal control delivers the time path for forest structure under objective of profit maximization and when considering uncertainties in future timber prices and quality. The aim of this contribution is particularly to perform the verification and validation for the model in a way that its results may be considered an appropriate help for policy making and forest management.

Keywords: forest structure, land use, optimal control, decision support, mixed forest, spruce

JEL Classification: C44
AMS Classification: 90C20, 90B90

1 Introduction

Although out of its environmental conditions, spruce dominates forest structure in the Central Europe. Since the early 19th century the pure coniferous stands have been artificially extended up to the current range which is far beyond natural limits [1]. Conversion of pure spruce stands into mixed forests is an effective strategy to improve ecological environmental situation of forests in Central Europe [2]. The conversion of forest stands is indeed a fundamental policy change, therefore taking stakeholders into account is important for the process to be successful [3] [4]. While the authorities and experts agree that the change must take place, the willingness of forest managers to make substantial changes has been very low so far. The problem stems mainly from economic dimension: the pure coniferous stands are economically advantageous and even though mixed forests are supposed to be economically comparable when including ecological stability [5], the long term conversion process has unknown economic impacts.

In this contribution we introduce the optimal control model for forest conversion in the Czech Republic with maximizing profit objective and ecologic-legislative constraints aiming at delivering an optimal long run policy for conversion of coniferous dominated forest into mixed forest with higher ratios of ameliorative and soil improving species. The particular aim is to perform thorough verification and validation discussion which not only approves the appropriateness and technical correctness of the model but also delivers understanding of the results application in the decision making process.

2 The problem

The problem is to find an optimal conversion strategy for representative forest area of 72.2 ha, approximatively $1/1000$ of total forest area in the Drahanska highlands. The initial forest structure is given by area distribution of 5 species in the Drahanska highlands: beech, oak, pine, larch and spruce. The objective is to maximize total current value from future profits of the forests, while the area of particular species may increase or decrease in time. The profit is given by difference between total revenues $R_i(t)$ from selling the timber and subsidies and total costs $K_i(t)$ of growing and logging at time t.

2.1 The optimization model

The problem has been formulated in [6] as infinite horizon optimal control problem with free terminal points:

[1] Mendel University in Brno, Department of statistics and operational analysis, Address, janova@mendelu.cz

$$
\begin{aligned}
\max \quad & V = \int_0^\infty \Pi(t) e^{-\rho t} \mathrm{d}t && (1)\\
& \Pi(t) = \sum_{i=1}^{n} R_i(x_i(t), u_i(t)) - && (2)\\
& -K_i(x_i(t), u_i(t)) && (3)\\
\text{s.t.} & && (4)\\
& u_i(t) = \dot{x}_i(t), && (5)\\
& \sum_{i=1}^{n} x_i(t) = A, && (6)\\
& x_i(t) \geq 0, && (7)\\
& u_j(t) \geq 0, \ 1 \leq j \leq n-1 && (8)\\
& -\frac{1}{\tau_5} x_5(t) \leq u_5(t) \leq 0, && (9)\\
& x_i(0) \text{ given}, \ 1 \leq i \leq n, && (10)
\end{aligned}
$$

where n denotes the number of tree species considered in the given region, $x_i(t)$ are the state variables representing the area of land forested by species i in time t, $u_i(t)$ are control variables representing total area reforested at time t by species i (in ha per year), ρ is discount rate, $\Pi(t)$ the profit in time t. The control variables $u_i(t)$ are non negative for all species but spruce (u_5) for which enlarging of its total area is not allowed, i.e. the total area forested by spruce may only remain at the initial level or decrease in our model. Constraint (9) defines upper limit of annual spruce cutting by the area of spruce in rotation age (τ_5).

For the purpose of further estimation of the functions $R_i(x_i)$, $K_i(x_i, u_i)$, we have set

$$
\begin{aligned}
R_i &= g_{i0} + g_{i1} x_i + \frac{1}{2} g_{i2} x_i^2 + \sigma_i u_i, && (11)\\
K_i &= k_i + a_i x_i + b_{i1} u_i + \frac{1}{2} b_{i2} u_i^2, && (12)
\end{aligned}
$$

where σ_i is the subsidy from increasing the area of species i by 1 ha. Note that only the ammeliorative and soil improving species are considered to be supported by subsidies.

2.2 Results and Discussion

The optimal control problem may be analyzed and steady state solution may be found analytically for the special case when we allow only one species to enlarge its area at the expense of spruce as shown in [6]. Let us allow only beech to spread in the forest. Beech is the most important natural species in the area, historically it covered 77%[1] of the forest land in the Drahanska highland. Following the calculation presented in [6] we obtain the steady state area for beech 18.8 ha.

To obtain the time path solution for our dynamic problem we must solve the complete optimal control problem. We have transformed the problem into discrete finite time horizon problem with planning period 100 years for which the decision variables are areas of particular species in the given year and control variables are the annual changes in these total areas. Note that in this representation we have neglected the terminal value function in the objective functional (compare [7]). We assume that this may be done because the terminal value of the forest will not depend on particular forest land distribution among species, the forest will be established and no further conversion will be expected. Thanks to quadratic function in objective and linear constraints, the discrete equivalent of the control problem may be transformed into quadratic programming problem for which solutions may be found more easily. Solving the quadratic programming problem we obtain the time path for forest land use distribution visualized in Figure 1. The terminal state of land use distribution is given by 17.2 ha for beech, 14.4 ha for oak, 9.9 ha for pine, 8.4 ha for larch and 22.15 ha for spruce.

[1] http://www.uhul.cz/images/ke_stazeni/oprl_oblasti/OPRL-LO30-Drahanska_vrchovina.pdf

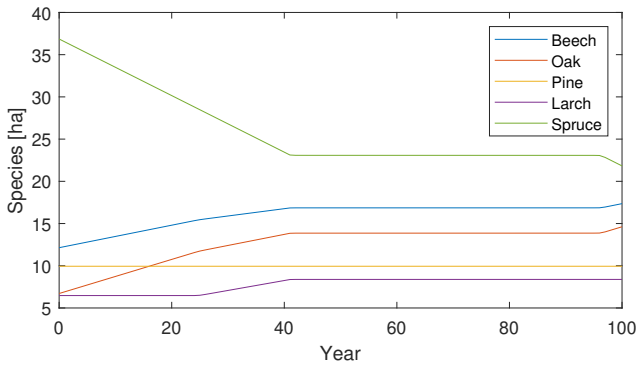


Figure 1 Forest land use time path, planning period 100 years, $\rho = 0.03$

3 Verification and Validation

The forest management optimization models focus mainly on the harvests scheduling, optimal rotation, optimal thinning, supply chain management, regeneration and land use optimization for which predominantly linear programming is selected as means of optimization [8], [9]. The inherent part of forest management are risks, uncertainties and dynamic nature of the problem due to the long term planning (the planning period often spans several hundred years). While risks are often covered mainly by portfolio optimization approach (e.g. [10], [11]), dynamic optimization techniques are seldom employed (from dozen recent articles on forest management optimization only a couple of papers deals with dynamic programming [12],[13], [7], [14]). Theoretically, the dynamic approach appears appropriate when searching for long term forest strategies. However, being less applied so far there is not enough practice when producing and interpreting the available results. In consequence, one must carefully handle the dynamic model and thoroughly understand the assumptions and possible shortcomings that might arise when the results would be incorrectly interpreted. To this end, thorough verification and validation (V&V) is of atmost importance when developing dynamic optimization model for forest management. The V&V phenomena, although theoretically being inherent part of decision support system (DSS) development, are not uniquely defined in operations research ([15], [16], [17], [18]). We follow the general procedure of building the DSS [15] and identify the particular steps according to [19]:

1. Real world problem statement (Forest conversion problem)
2. Conceptual model development (Optimal control transformed into Quadratic programming model)
3. Building appropriate computerized model (Matlab implementation of Quadratic programming model)
4. Solution (Time path).

In our contribution we follow the V&V definitions presented in [19]:

"Verification of DSS evaluates the formal appropriateness and correctness of the conceptual model and it ensures technical correctness of the subsequent computerized model."

"Validation means determining whether the conceptual model and the solutions provided are in accurate representation of DSS intended use."

The V&V can't be applied generally, but must be carefully tailormade for the particular DSS. Applying the V&V framework formulated in [19] we define the verification and validation processes for forest conversion dynamic optimization problem.

3.1 Verification

As structured in [19], the verification may be split in two parts: (i) conceptual model validation and (ii) computerized model verification. Verification is often not explicitly mentioned when developing standard optimization model, such as linear or quadratic programming [20]. Our problem is formulated initially as optimal control problem that was transformed into quadratic programming problem which is a common well established optimization tool. Hence one must just make sure that the quadratic programming model is correctly formulated with respect to original optimal control problem (conceptual model validation) and technically accordingly implemented in the solver (Matlab) (computerized model verification).

The objective function. The objective in quadratic programming has the meaning of current value of future profits from forest stand throughout 100 years planning period. This goal function is common in op-

timization models although there has been steady debate on effects of current value objective onto decision making, recently e.g. [22]. If we accept for this model the current value approach, then only the cost K_i and revenue functions R_i need verification. The cost K_i and revenue functions R_i in objective function were estimated in [6], [23], [21] as annual average money flows in the representative area. Simulations were used to incorporate the uncertainty about future timber prices, soil quality and timber quality. Hence, the functions are appropriate only when used in long term planning, which is our case in the optimal conversion problem. Further, the functions do not intend to cover the revenue and costs precisely and to full scope, but to cover the main aspects of the variable components of profit and costs and enable relative comparisons among species. This also suits to the optimization model, as the relative relations matter in optimization. The revenue and cost functions provide annual values that are in good correspondence to real annual revenues and costs from comparable forest stands. Therefore also the values of goal function are realistic and may be used for interpreting the results and estimation of future profit flows.

Constraints and decision variables. We have verified the constraint in quadratic programming representing the equation of motion of the optimal control (5) to cover appropriately the dynamics of the problem. Further, we have verified technical transformation from optimal control with $2n$ dynamic variables $x_i(t)$, $u_i(t)$, $1 \leq i \leq n$ to quadratic programming with $2nT$ variables x_{it}, u_{it}, $1 \leq t \leq T$. This was done by common check of correctness when developing the quadratic programming model and formulating first the small scale example and crosschecking with the original optimal control problem. The remaining constraints are straithforward and do not need extra verification, so only correct rewriting into computerized model in matlab was ensured.

Results. To ensure compliance of the computerized model and the original optimal control problem we have compared their results. After formulating the complete quadratic programming problem and finding its solution we compared the results for steady state optimal control solution for beech (18.8 ha) with the terminal state value for time path of beech obtained from quadratic programming (17.2 ha) (see 2.2). As the values are comparable (though having on mind limitation of steady state approximation to finite time problem), we take this a supportive element in verifying the technical accordance of quadratic programming model and its implementation in Matlab with original optimal control problem.

3.2 Validation

Validation means determining whether the conceptual model and the solutions provided are accurate representation of the DSS's intended use [19]. But how to acknowledge the validity is an open question, no universal approach exists. The key issue is to formulate the intended use of the optimization model, which in our case is "to provide the decision maker initial information on optimal 100 years conversion of forest in the Drahanska highlands under pure economic criteria of current profit maximization".

The optimal control or quadratic programming with the objective of current value profit maximization may be considered appropriate methods for the desired goal and verification ensures that the results obtained are technically correct with respect to the conceptual model (which is quadratic programming model (QPM) in our case). The validation must ensure that the quadratic programming model is doing the right thing, i.e. we need to understand the assumptions and limitations of the quadratic programming model:

- QPM is based on average cost and revenues functions based on aggregate simulated data on environmental, technical and economical scenarios in the Drahanska highlands, therefore it is a sketch model (see [19]) and its results represent directions of decision making rather than particular solution for the particular forest stand.
- Via the cost and revenue functions QPM incorporates basic uncertainties and risks (uncertainty in timber quality and timber prices). The functions do not incorporate the risks from severe weather conditions or biotic risks, such as bark beetle calamity. However, the changing climate and risk of drought is considered to some extent as the model is converting the spruce monocultures into mixed forest.
- Up to constraint (9), the constraints do not bound explicitly the forest conversion. Hence, if the conversion takes place in the model solution, it is a result of economic optimization only.

After summarizing the main assumptions and limitations we may turn to the QPM results (see Figure 1). We apply the following validation techniques [19]:

1. **Face validation.** We review the QPM behavior with respect to our knowledge of the real world problem. The results in Figure 1 show the forest structure conversion up to certain equilibrium state. The area of spruce is decreased and ratios of beech together with oak are increasing, which is what experts theoretically presume: both broadleaf species are natural in the area, are reasonably profitable and are expected to improve the forest resistance against drought and climate changes. The conversion from

spruce to other species happens only thanks to incorporating risks from climate change into the decision model. If these risks were not incorporated, the economic advantageousness of spruce would dominate and the conversion would not begin at all, which corresponds to the historical experience from Czech forest management and explains why we do not observe forest conversion in large extent in the Czech Republic. However, the model is relevant, as there are studies confirming the rise of willingness to plant mixed forests in the Czech Republic.

2. Comparison to other models is not currently possible as we are not aware of any similar study.
3. **Sensitivity analysis.** We have solved QPM for various parameters setting. Important from the validation viewpoint are the solutions with different subsidies. When increasing subsidies for ammeliorative and soil improving species (ASIS) the results change as one would expect: the convergence into forest with higher rations of ASIS is accelerating. This may be considered another element confirming the compliance of the model solution with real world situation.

4 Conclusion

Under selected assumptions, we have verified and validated an optimal conversion strategy model of a forest stand aiming to serve as a forest management decision support tool.

Despite the possible shortcomings of the model with respect to representing the original real world problem [15], the validation delivers good understanding of the optimization model capabilities and prevents the decision maker from inappropriate use of the model results.

The finite planning horizon model will be further elaborated with respect to precise expression of the terminal value of the established forest. Currently, the bark beetle calamity escalates in the Czech Republic which instantly affects forest stands with critical impacts to forest planning. The willingness of forest managers to consider mixed stands has increased rapidly which intensifies the need for appropriate decision support under changing situation. In the following research the model is to be further elaborated in this direction.

Acknowledgements

This paper presents the results of the research supported by Czech Science Foundation grant GA18-08078S.

References

[1] Knoke, T., Stimm, B., Ammer, C. and Moog, M. (2005). Mixed forests reconsidered: A forest economics contribution on an ecological concept. *Forest Ecology and Management*, 213(1-3):102–116, July 18 2005.

[2] Seppelt, R., Dormann, C. F., Eppink, F. V., Lautenbach, S. and Schmidt, S. (2011). A quantitative review of ecosystem service studies: approaches, shortcomings and the road ahead. *Journal of Applied Ecology*, 48(3):630–636, June 2011.

[3] Paletto, A., Giacovelli, G., Grilli, G., Balest, J., and De Meo, I. (2014). Stakeholders' preferences and the assessment of forest ecosystem services: A comparative analysis in italy. *Journal of Forest Science*, 60(11):472–483, 2014. cited By 12.

[4] Paletto, A., Cantiani, M. G. and De Meo, I. (2015). Public Participation in Forest Landscape Management Planning (FLMP) in Italy. *Journal of Sustainable Forestry*, 34(5):465–482, 2015.

[5] Knoke, T. (2008). Mixed forests and finance – Methodological approaches. *Ecological Economics*, 65(3):590–601, April 15 2008.

[6] Janova, J. and Hampel, D. (2016). Optimal managing of forest structure using data simulated optimal control. *Central European Journal of Operations Research*, 24(2, SI):297–307, June 2016.

[7] Cerda, E. and Martin-Barroso, D. (2013). Optimal control for forest management and conservation analysis in dehesa ecosystems. *European Journal of Operational Research*, 227(3):515–526, June 16 2013.

[8] Caglayan, I., Yesil, A., Cinar, D. and Cieszewski, C. (2018). Taxonomy for the optimization in forest management: a review and assessment. *Forestist*, 68(2):122–135, 2018.

[9] Martell, D. L., Gunn, E. A. and Weintraub, A. (1998). Forest management challenges for operational researchers. *European Journal of Operational Research*, 104(1):1–17, January 1 1998.

[10] Neuner, S., Albrecht, A., Cullmann, D., Engels, F., Griess, V. C., Hahn, W. A., Hanewinkel, M., Haertl, F., Koelling, C. (2015). Kai Staupendahl, and Thomas Knoke. Survival of Norway spruce remains higher

in mixed stands under a dryer and warmer climate. *Global Change Biology*, 21(2):935–946, February 2015.

[11] Dragicevic, A., Lobianco, A., and Leblois, A. (2016). Forest planning and productivity-risk trade-off through the Markowitz mean-variance model. *Forest Policy and Economics*, 64:25–34, March 2016.

[12] Ferreira, L., Constantino, M. F., Borges, J. G. and Garcia-Gonzalo, J. (2015). Addressing Wildfire Risk in a Landscape-Level Scheduling Model: An Application in Portugal. *Forest Science*, 61(2):266–277, Aprin 2015.

[13] Ferreira, L., Constantino, M., Borges, J. G., Garcia-Gonzalo, J. and Barreiro, S. (2016). A climate change adaptive dynamic programming approach to optimize eucalypt stand management scheduling: a Portuguese application. *Canadian Journal of Forest Research*, 46(8):1000–1008, August 2016.

[14] Alvarez, P. P. and Vera, J. R. (2014). Application of Robust Optimization to the Sawmill Planning Problem. *Annals of Operations Research*, 219(1):457–475, August 2014.

[15] Gass, S. I. (1993). Model Accreditation – a Rationale and Process for Determining a Numerical Rating. *European Journal of Operational Research*, 66(2):250–258, April 16 1993.

[16] Dery, R. Landry, M. and Banville, C. (1993). Revisiting the Issue of Model Validation in or – an Epistemological View. *European Journal of Operational Research*, 66(2):168–183, April 16 1993.

[17] Miser, H. J. (1993). A Foundational Concept of Science Appropriate for Validation in Operational-research. *European Journal of Operational Research*, 66(2):204–215, April 16 1993.

[18] Janova, J. and Hampel, D. (2016). alfaDRA: A program for automatic elimination of variety self-proximities in alpha-design. *Computers and Electronics in Agriculture*, 122:156–160, March 2016.

[19] Janova, J., Hampel, D. and Nerudova, D. (2019). Design and validation of a tax sustainability index. *European Journal of Operational Research*, 278(3):916–926, November 1 2019.

[20] Janova, J. (2012). Crop planning optimization model: the validation and verification processes. *Central European Journal of Operations Research*, 20(3, SI):451–462, September 2012.

[21] Hampel, D. and Janová, J. (2014). Simulation of data for reforestation system. In *Conference Proceedings of 32nd International Conference Mathematical Methods in Economics* (pp. 251–256). Olomouc: Univerzita Palackého v Olomouci.

[22] 10.1016/j.ecolecon.2020.106664

[23] Hampel, D., Janová, J. and Kadlec, J. (2015). Estimation of Cost and Revenue Functions for Reforestation System in Drahanska Highlands. In: *Proceedings of the International Conference on Numerical Analysis and Applied Mathematics 2014*. Melville: American Institute of Physics (AIP), 1648.

Evaluation of an AFT Model for Misclassified, Interval-censored Data in an Economical Setting

Tomáš Kala[1], Arnošt Komárek[2]

Abstract. We adapt an Accelerated Failure Time (AFT) model designed for the analysis of misclassified, interval-censored medical study data, into an economical setting. The model performance is evaluated on a simulated dataset, and a centering reparameterization of its event times part is applied to ensure faster convergence of MCMC algorithms.

Keywords: interval-censored data, misclassified data, bayesian approach

JEL Classification: C11, C34
AMS Classification: 62N01

1 Data and research problems

The motivation for our study is the analysis of a number of low-income households, some of whose members currently receive welfare benefits from the state. At the same time, they are encouraged to find a job occupation as soon as possible to alleviate the taxpayers. As such, they are periodically checked for whether they have found a posting since the previous checkup. The interest is to determine the dependence of the unemployment time on given set of covariates. The problem is that the unemployment time can only be determined to lie between two consecutive examinations, giving rise to interval-censored data.

Another complication is that even the determination of the employment status can be subject to error. We assume employment to be an occupation which ensures a regular monthly income of at least 14,600 CZK (the minimum monthly wage). As it is often the case that the subjects find undeclared work, they may fail to disclose their employment status to a government officer. Another reason for not confessing their employment status is to simply keep collecting welfare. To conclude, not only is the exact unemployment time interval-censored, but even its occurrence within an interval can be misclassified.

We describe how an AFT model proposed in García-Zattera et al. [1] to analyze misclassified interval-censored data can be adapted to this situation. Although originally developed in the context of a medical study, it is well suited for household data as well.

The following section gives an overview of the model formulation. Discussed are the the regression part describing the dependence of the event (unemployment) time on a given set of covariates, and the misclassification part linking together the event times and the (potentially) error-corrupted event indicators. Finally, we describe how the model can be adapted to household data and perform a simulation study to evaluate the model performance.

2 AFT model for misclassified, interval-censored data

We assume a dataset consisting of N independent subjects (households in our case), each consisting of J units (household members). Let $T_{(i,j)} \in \mathbb{R}_+$ be the time-to-event (employment) of the jth member of the ith household. The employment status of all members of the ith household is assessed at evaluation times $v_{(i,0)} \equiv 0 < v_{(i,1)} < \ldots < v_{(i,K_i)} < \infty \equiv v_{(i,K_i+1)}$ by one of Q government officers. The exact event time $T_{(i,j)}$ is not directly observable. If there is no misclassification of the event status, the event time $T_{(i,j)}$ is exactly known to lie in an interval $\left(v_{(i,l_{(i,j)}-1)}, v_{(i,l_{(i,j)})} \right]$, $l_{(i,j)} \in \{1, \ldots, K_i\}$, where $v_{(i,l_{(i,j)})}$ is the first examination where the employment event was observed. Nevertheless, the event occurrences can be misclassified as described in Section 1. The observed data then consist of the (potentially) corrupted event occurrence indicators $Y_{(i,j,k)} \in \{0, 1\}$, $k = 1, \ldots, K_i$ where a value of one indicates that the event occurrence was (perhaps wrongly) indicated by an officer who performed the examination at time $v_{(i,k)}$.

Finally, we assume that for each subject i and unit j, a p-dimensional covariate vector $\mathbf{x}_{(i,j)}$ is available. The goal is to infer the dependence of the unobserved event times $T_{(i,j)}$ on $\mathbf{x}_{(i,j)}$, where $T_{(i,j)}$ are only observed

[1] Charles University, Faculty of Mathematics and Physics, Dept. of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic, kalat@karlin.mff.cuni.cz
[2] Charles University, Faculty of Mathematics and Physics, Dept. of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic, komarek@karlin.mff.cuni.cz

through sequences of misclassified binary indicators $Y_{(i,j,k)}$ of the event occurrence. The proposed model consists of two parts: the misclassification model linking the observed indicators $Y_{(i,j,k)}$ to the unobserved event times $T_{(i,j)}$, and the event times model specifying the dependence of $T_{(i,j)}$ on $\mathbf{x}_{(i,j)}$.

For the misclassification model, let $\xi_{(i,k)} \in \{1, \ldots, Q\}$ denote the government officer present at the kth examination of the ith household. Further, denote the specificity and sensitivity of the qth examiner by $\eta_{(q)}$ and $\alpha_{(q)}$, respectively. These parameters are assumed to lie in a restricted space $\left\{\left(\eta_{(q)}, \alpha_{(q)}\right) \in [0,1]^2 : \eta_{(q)} + \alpha_{(q)} > 1\right\}$ to avoid identification problems. By definition of sensitivity and specificity, we can write the probabilities of event classification as follows:

$$\mathbb{P}\left(Y_{(i,j,k)} = 1 \mid T_{(i,j)} \in (0, v_{(i,k)}]\right) = \alpha_{\left(\xi_{(i,k)}\right)},$$
$$\mathbb{P}\left(Y_{(i,j,k)} = 0 \mid T_{(i,j)} \in (v_{(i,k)}, \infty)\right) = \eta_{\left(\xi_{(i,k)}\right)}.$$

Next, we describe the event times model assumed by García-Zattera et al. [1]. Its structure follows an AFT model with random effects for the time-to-event data:

$$\log T_{(i,j)} = \mathbf{x}_{(i,j)}^{\intercal}\boldsymbol{\beta} + b_i + \varepsilon_{(i,j)}, \tag{1}$$

where $b_i \overset{\text{iid}}{\sim} g_b$ and $\varepsilon_{(i,j)} \mid \sigma_\varepsilon^2 \overset{\text{iid}}{\sim} \mathrm{N}\left(0, \sigma_\varepsilon^2\right)$. Here, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a fixed effect vector, $b_i \in \mathbb{R}$ is a subject-specific random effect with density g_b and $\mathbb{R} \ni \varepsilon_{(i,j)} \perp b_i$ is an error term.

To parameterize the random effects in a flexible way, a penalized Gaussian mixture model (Komárek and Lesaffre [2]) is assumed, which takes the form

$$g_b(\cdot) = \sum_{l=-M}^{M} w_l \frac{1}{\tau}\phi\left(\frac{\cdot - \mu}{\tau}\middle|\kappa_l, \zeta^2\right),$$

where $\phi\left(x\middle|\mu, \sigma^2\right)$ denotes the density of $\mathrm{N}\left(\mu, \sigma^2\right)$ evaluated at x. The vector $(\kappa_{-M}, \ldots, \kappa_0, \ldots, \kappa_M)^{\intercal}$ contains a fixed grid of knots centered at $\kappa_0 = 0$, ζ^2 is a fixed basis variance and $\mu \in \mathbb{R}$, $\tau > 0$ are unknown location and scale parameters. The mixture weights are given by $w_l = \frac{\exp\{a_l\}}{\sum_{m=-M}^{M}\exp\{a_m\}}$ and parameterized by the vector $\mathbf{a} = (a_{-M}, \ldots, a_M)^{\intercal} \in \mathbb{R}^{2M+1}$. The mixture model is characterized by $\boldsymbol{\theta} = (\mu, \mathbf{a}^{\intercal}, \tau)^{\intercal}$, and the penalized Gaussian mixture distribution is abbreviated as PGM $(\boldsymbol{\theta})$.

As the aim is to fit the model through Bayesian methods, we need to specify prior distribution for its parameters. For the regression model, we assume $\boldsymbol{\beta} \sim \mathrm{N}(m_{\boldsymbol{\beta}}, V_{\boldsymbol{\beta}})$, $\sigma_\varepsilon^{-2} \sim \Gamma(\nu_{\varepsilon_1}, \nu_{\varepsilon_2})$, $\mu \sim \mathrm{N}\left(m_\mu, s_\mu^2\right)$, $\tau^{-2} \sim \Gamma(\nu_{\tau_1}, \nu_{\tau_2})$. The transformed mixture weights a are assumed to follow an intrinsic Gaussian Markov random field with a precision matrix $\lambda P_o^{\intercal} P_o$, P_o denoting the finite difference operator matrix of order o and $\lambda \sim \Gamma(\nu_{\lambda_1}, \nu_{\lambda_2})$. Order $o = 3$ is used. Finally, the sensitivities and specifities are assumed to follow a constrained beta distribution, $(\alpha_q, \eta_q) \overset{\text{iid}}{\sim} \mathrm{B}\left(a_q^{(\alpha,0)}, a_q^{(\alpha,1)}\right) \times \mathrm{B}\left(a_q^{(\eta,0)}, a_q^{(\eta,1)}\right) \times \mathbb{I}\left\{\alpha_{(q)} + \eta_{(q)} > 1\right\}$.

3 Evaluation

The model is evaluated on a simulated dataset. The event times $T_{(i,j)}$ are generated according to the model

$$\log T_{(i,j)} = \beta_0 + \beta_1 x_{(i,j,1)} + \beta_2 x_{(i,j,2)} + b_i + \varepsilon_{(i,j)},$$

where $i = 1, \ldots, N$, $j = 1, \ldots, 4$, $x_{(i,j,1)} \overset{\text{iid}}{\sim} \mathrm{U}(2,3)$, $x_{(i,j,2)} \overset{\text{iid}}{\sim} \mathrm{Alt}(0.6)$ and $\varepsilon_{(i,j)} \overset{\text{iid}}{\sim} \mathrm{N}(0, 0.08)$. The variable interpration is as follows. The variable $x_{(i,j,1)}$ represents (in months) how long the jth member of the ith household collects welfare, and $x_{(i,j,2)}$ corresponds to the gender of the jth member of the ith household encoded as a binary indicator. The true coefficient values are $\beta_0 = 0.8$, $\beta_1 = 0.2$, $\beta_2 = -0.1$. The random effects b_i are sampled from the Gumbel distribution with mean 0 and variance $\sigma_b^2 = 0.02$.

The interval censoring was produced by simulating $K_i = 10$ examination times of each household i. The time of the first examination was sampled from $\mathrm{N}(0.8, 0.01)$, while the times between two consecutive examinations were sampled from $\mathrm{N}(1, 0.01)$.

The households were examined by $Q = 5$ government officers, whose sensitivities and specificities were set differently in three scenarios. The first scenario ("low precision") assumes the sensitivities to be linearly

Precision	N	β_0	β_1	β_2	σ_ε^2	σ_b^2
	500	−0.2168	0.0759	0.0299	0.0161	−0.0009
Low	1000	0.1307	−0.0461	−0.0317	−0.0095	−0.0012
	2000	0.0733	−0.0297	0.0049	−0.0056	0.0003
	500	−0.1219	0.0441	−0.0041	0.0029	−0.0009
Medium	1000	−0.0191	−0.0044	0.0257	0.0049	−0.0031
	2000	−0.1210	0.0547	−0.0200	0.0026	0.0019
	500	0.0103	−0.0030	−0.0041	0.0025	−0.0005
High	1000	0.0240	−0.0039	−0.0154	−0.0004	−0.0018
	2000	−0.0054	−0.0006	0.0003	−0.0000	0.0005

Table 1 Bias of the posterior means of the estimated regression parameters.

Precision	N	β_0	β_1	β_2	σ_ε^2	σ_b^2
	500	0.1662	0.0661	0.0377	0.0138	0.0012
Low	1000	0.1040	0.0407	0.0240	0.0071	0.0009
	2000	0.0769	0.0299	0.0174	0.0054	0.0007
	500	0.1002	0.0394	0.0229	0.0062	0.0012
Medium	1000	0.0717	0.0278	0.0166	0.0047	0.0008
	2000	0.0512	0.0200	0.0116	0.0031	0.0007
	500	0.0602	0.0237	0.0137	0.0022	0.0012
High	1000	0.0420	0.0165	0.0096	0.0015	0.0011
	2000	0.0296	0.0116	0.0067	0.0011	0.0007

Table 2 Standard deviation of the posterior means of the estimated regression parameters.

spaced between 0.6 and 0.8, and the specificities between 0.5 and 0.7. The second scenario ("medium precision") assumes the sensitivities and specificities to be between 0.7 and 0.9, and 0.6 and 0.8, respectively. The third scenario ("high precision") assumes both the sensitivities and specificities to be perfect and equal to one.

The prior hyperparameters of the regression model were set to $m_{\boldsymbol{\beta}} = 0_3$, $V_{\boldsymbol{\beta}} = 1000 \times \mathbf{I}_3$, $\nu_{\varepsilon_1} = \nu_{\lambda_1} = \nu_{\tau_1} = 1$, $\nu_{\varepsilon_2} = \nu_{\lambda_2} = \nu_{\tau_2} = 0.005$. The hyperparameters for the misclassification model were set to identical vectors of ones, leading to uniform prior distribution over sensitivities and specificities.

In each setting, increasing number of households was sampled and N set progressively to 500, 1000 and 2000. This way, we can assess how the estimates improve with increasing number of data even under misclassification. The Markov chains were run for 440,000 iterations in total, with a burn-in period of 40,000 and thinning equal to 25.

In Tables 1 and 2, the biases and standard deviations, respectively, of the posterior means of the regression parameters are shown. As can be expected, both parameters decrease with increasing number of households N and precision (sensitivities and specificities) of the examiners. Even under lowest examiner precision, the estimates concentrate around the true values with increasing N, meaning that the model is applicable even to situations with a high degree of misclassification.

Tables 3 and 4 show the same results for the misclassification parameters. We observe similar results as in the case of the regression parameters, i.e., both the bias and standard deviation decreases with increasing N even under low classification precision.

In addition, having assumed uniform priors over the sensitivities and specificities, we conclude that no external knowledge of the examiner behavior is necessary, as the estimates concentrate around the true values even under no prior knowledge.

Precision	N	α_1	α_2	α_3	α_4	α_5
Low	500	0.3047	0.3018	0.3049	0.3065	0.3015
	1000	0.3055	0.2985	0.3071	0.3049	0.3068
	2000	0.2959	0.2981	0.2972	0.2989	0.2989
Medium	500	0.0047	−0.0010	0.0010	0.0051	0.0091
	1000	0.0064	−0.0048	−0.0037	0.0008	0.0019
	2000	0.0077	−0.0048	−0.0011	−0.0014	0.0053
High	500	−0.0005	−0.0004	−0.0004	−0.0005	−0.0004
	1000	−0.0002	−0.0002	−0.0002	−0.0002	−0.0002
	2000	−0.0001	−0.0001	−0.0001	−0.0001	−0.0001

Precision	N	η_1	η_2	η_3	η_4	η_5
Low	500	0.0957	0.0787	0.1091	0.1002	0.0952
	1000	0.1041	0.0850	0.1144	0.1077	0.1045
	2000	0.0960	0.0999	0.0985	0.1121	0.0962
Medium	500	−0.0090	−0.0171	0.0033	0.0017	0.0092
	1000	0.0073	−0.0046	−0.0055	0.0115	0.0125
	2000	0.0088	−0.0090	−0.0149	0.0007	0.0010
High	500	−0.0011	−0.0011	−0.0010	−0.0011	−0.0010
	1000	−0.0005	−0.0005	−0.0005	−0.0005	−0.0005
	2000	−0.0002	−0.0003	−0.0003	−0.0002	−0.0003

Table 3 Bias of the posterior means of the estimated misclassification parameters.

Precision	N	α_1	α_2	α_3	α_4	α_5
Low	500	0.0097	0.0099	0.0098	0.0098	0.0091
	1000	0.0071	0.0069	0.0068	0.0067	0.0061
	2000	0.0050	0.0050	0.0050	0.0046	0.0044
Medium	500	0.0091	0.0091	0.0084	0.0079	0.0062
	1000	0.0068	0.0063	0.0060	0.0053	0.0047
	2000	0.0047	0.0045	0.0043	0.0039	0.0033
High	500	0.0004	0.0004	0.0004	0.0005	0.0004
	1000	0.0002	0.0002	0.0002	0.0002	0.0002
	2000	0.0001	0.0001	0.0001	0.0001	0.0001

Precision	N	η_1	η_2	η_3	η_4	η_5
Low	500	0.0145	0.0158	0.0164	0.0155	0.0150
	1000	0.0112	0.0108	0.0109	0.0103	0.0104
	2000	0.0074	0.0076	0.0078	0.0074	0.0078
Medium	500	0.0141	0.0152	0.0151	0.0138	0.0129
	1000	0.0105	0.0109	0.0104	0.0095	0.0093
	2000	0.0075	0.0072	0.0073	0.0069	0.0066
High	500	0.0011	0.0011	0.0010	0.0011	0.0010
	1000	0.0005	0.0005	0.0005	0.0005	0.0005
	2000	0.0002	0.0003	0.0003	0.0002	0.0003

Table 4 Standard deviation of the posterior means of the estimated misclassification parameters.

4 Modification of the event times distribution structure

Our contribution lies in a modification of Equation 1 in a way that ensures a faster convergence of the MCMC algorithms. The formulation also allows the random effects to involve more variables than just the random intercept.

Let the coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^p$ be divided into $\boldsymbol{\beta} = (\boldsymbol{\beta}_F^\intercal, \boldsymbol{\beta}_R^\intercal)^\intercal$, where $\boldsymbol{\beta}_F \in \mathbb{R}^d$ is a fixed effects vector, and $\boldsymbol{\beta}_R \in \mathbb{R}^{p-d}$ is the random effects mean vector. That is, hierarchically centered random effects will be assumed, as recommended by Gelfand et al. [3].

Specifically, We rewrite Equation 1 into

$$\log T_{(i,j)} = \mathbf{x}_{(i,j)}^\intercal \boldsymbol{\beta}_F + \mathbf{z}_{(i,j)}^\intercal \mathbf{b}_i + \varepsilon_{(i,j)},$$

where $\mathbf{z}_{(i,j)} \in \mathbb{R}^{p-d}$ is a covariate vector corresponding to the random effects, and the meaning of the other symbols remains the same as in Equation 1.

As for the distributional assumptions, we let $\mathbf{b}_i \sim \mathrm{N}_{p-d}(\boldsymbol{\beta}_R, \boldsymbol{\Sigma})$ and $\varepsilon_{(i,j)} \overset{\text{iid}}{\sim} \mathrm{PGM}(\mu, \mathbf{a}, \tau)$. That is, we move the PGM model to the error terms to retain the flexibility this distribution brings, and specify the multivariate normal distribution for the random effects, as is the case in standard mixed effect models.

The model specification is completed by assuming the following multivariate normal and inverse Wishart prior distributions, respectively: $\boldsymbol{\beta}_R \sim \mathrm{N}_{p-d}(m_R, S_R)$ and $\Sigma_b^{-1} \sim \mathrm{W}_{p-d}(V_\Sigma, \nu_\Sigma)$. The remaining priors remain the same as in the original formulation.

The results of our simulation study indicate comparable results of the estimated parameters bias and standard deviation, which will therefore not be repeated. As discussed by Gelfand et al. [3], the convergence rate of the underlying Markov chain is improved by the centered parameterization. Since it leads to a model of equivalent precision, it can be used as a simple mean to improve the model performance.

5 Conclusion

We discussed how an AFT model designed for the analysis of misclassified, interval-censored data coming from a medical study can be applied in an economical setting. The model was evaluated on a simulated dataset and found to estimate the main parameters of interest sufficiently well. In addition, a simple reparameterization of the event times part of the model was adapted, leading to a faster convergence of the MCMC algorithms.

Acknowledgements

This research was supported by the Czech Science Foundation (GAČR) grant 19-00015S.

References

[1] García-Zattera, M. J., Jara, A. and Komárek, A. (2016). A flexible AFT model for misclassified clustered interval-censored data. *Biometrics*, **72**(2), 473–483.

[2] Komárek, A. and Lesaffre, E. (2008). Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *Journal of the American Statistical Association*, **103**(482), 523–533.

[3] Gelfand, A. E., Sahu, S. K. and Carlin, B. P. (1995). Efficient parameterisations for normal linear mixed models. *Biometrika*, **82**(3), 479–488.

A Note on Stochastic Optimization Problems with Nonlinear Dependence on a Probability Measure

Vlasta Kaňková[1]

Abstract. Nonlinear dependence on a probability measure begins to appear (last time) in a stochastic optimization rather often. Namely, the corresponding type of problems corresponds to many situations in applications. The nonlinear dependence can appear as in the objective functions so in a constraints set. We plan to consider the case of static (one-objective) problems in which nonlinear dependence appears in the objective function with a few types of constraints sets. In details we consider constraints sets "deterministic", depending nonlinearly on the probability measure, constraints set determined by second order stochastic dominance and the sets given by mean-risk problems. The last case means that the constraints set corresponds to solutions those guarantee an acceptable value in both criteria. To introduce corresponding assertions we employ the stability results based on the Wasserstein metric and $\mathcal{L}_1$ norm. Moreover, we try to deal also with the case when all results have to be obtained (estimated) on the data base.

Keywords: stochastic optimization problem, nonlinear dependence, static problems, second order stochastic dominance, mean–risk model, stability, empirical estimates

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Optimization problems with nonlinear dependence on a probability measure begin to appear (in a stochastic literature) (see, e.g., [2], [3], [6], [7]). This type of problems corresponds to many situations from practice.

To introduce the above mentioned type of the problems let $(\Omega, \mathcal{S}, P)$ be a probability space; $\xi\,(:= \xi(\omega) = (\xi_1(\omega), \ldots, \xi_s(\omega))$ an s–dimensional random vector defined on $(\Omega, \mathcal{S}, P)$; $F(:= F_\xi(z), z \in R^s)$ the distribution function of ξ; P_F, Z_F the probability measure and a support corresponding to F; $X_F \subset X \subset R^n$ a nonempty set generally depending on F; $X \subset R^n$ a nonempty "deterministic" set. If $\bar{g}_0(:= \bar{g}_0(x, z, y))$ is a real–valued function defined on $R^n \times R^s \times R^{m_1}$; $h(:= h(x, z)) = (h_1(x, z), \ldots, h_{m_1}(x, z))$ is an m_1–dimensional vector function defined on $R^n \times R^s$, then stochastic (static) optimization problem with a nonlinear dependence on the probability measure can be introduced in the form:

Find

$$\bar{\varphi}(F, X_F) = \inf\{\mathsf{E}_F \bar{g}_0(x, \xi, \mathsf{E}_F h(x, \xi)) | x \in X_F\}. \tag{1}$$

Evidently a nonlinear dependence can appear as in the objective function so in the constraints set. We consider a few types of X_F :

$$\begin{aligned} a. \qquad X_F &:= X, \\ b. \qquad X_F &:= \{x \in X : \mathsf{E}_F \bar{g}_i(x, \xi, \mathsf{E}_F h(x, \xi)) \le 0, \, i = 1, \ldots, m\}, \\ &\qquad \text{where} \quad \bar{g}_i(x, z, y), \, i = 1, \ldots, m \quad \text{are defined on} \quad R^n \times R^s \times R^{m_1}. \end{aligned} \tag{2}$$

Of course it is assumed that all finite mathematical expectation in the relations (1), (2) exist.

Problem (1) covers a classical problem with linear dependence in the form:

Find

$$\varphi(F, X_F) = \inf\{\mathsf{E}_F g_0(x, \xi) | x \in X_F\}, \tag{3}$$

[1] The Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, kankova@utia.cas.cz

with $\bar{g}_0(x, z, y) := g_0(x, z)$, $X_F = X$; $g_0(:= g_0(x, z))$ a real-valued function defined on $R^n \times R^s$.

To introduce the next type of X_F we have first to recall a notion of second order stochastic dominance. If $V := V(\xi)$, $Y := Y(\xi)$ are random values for which there exist finite $\mathsf{E}_F V(\xi)$, $\mathsf{E}_F Y(\xi)$ and if

$$F^2_{Y(\xi)}(u) = \int_{-\infty}^{u} F_{Y(\xi)}(y)dy, \quad F^2_{V(\xi)}(u) = \int_{-\infty}^{u} F_{V(\xi)}(y)dy,$$

then $V(\xi)$ dominates in second order $Y(\xi)$ $(V(\xi) \succeq_2 Y(\xi))$ if

$$F^2_{V(\xi)}(u) \leq F^2_{Y(\xi)}(u) \quad \text{for every} \quad u \in R^1.$$

To define second order stochastic dominance constraints set X_F, let $g(x, \xi)$ (defined on $R^n \times R^s$) be for every $x \in X$ a random variable with distribution function $F_{g(x,\xi)}$. Let, moreover, for every $x \in X$ there exists finite $\mathsf{E}_F g(x, \xi)$, $\mathsf{E}_F Y(\xi)$ and

$$F^2_{g(x,\xi)}(u) = \int_{-\infty}^{u} F_{g(x,\xi)}(y)dy, \quad F^2_{Y(\xi)}(u) = \int_{-\infty}^{u} F_{Y(\xi)}(y)dy, \quad u \in R^1,$$

then rather general second order stochastic dominance constraints set X_F can be defined by

c.
$$X_F = \{x \in X : F^2_{g(x,\xi)}(u) \leq F^2_{Y(\xi)}(u) \quad \text{for every} \quad u \in R^1\}, \tag{4}$$

equivalently by

$$X_F = \{x \in X : \mathsf{E}_F(u - g(x, \xi))^+ \leq \mathsf{E}_F(u - Y(\xi))^+ \quad \text{for every} \quad u \in R^1\}. \tag{5}$$

(The equivalence of the constraints sets (4) and (5) can be found in [10]. For definitions of the stochastic dominance of others orders see, e.g., [9].)

To introduce the last considered type of the set X_F we start with classical mean-risk problem:

Find

$$\max \mathsf{E}_F g_0(x, \xi), \quad \min \rho_F(g_0(x, \xi)) \quad \text{s.t.} \quad x \in X; \quad \rho(\cdot) := \rho_F(\cdot) \quad \text{denotes a risk measure.} \tag{6}$$

Evidently to optimize simultaneously both objectives is mostly impossible. However it can happen that there exist two real-valued acceptable constants ν_2, ν_1 and $x_0 \in X$ such that

$$\mathsf{E}_F g_0(x_0, \xi) \geq \nu_2, \quad \rho_F(g_0(x_0, \xi)) \leq \nu_1. \tag{7}$$

If furthermore the function g_0 and risk measure ρ_F follow the following definition:

Definition. [4] The mean–risk model (6) is called consistent with the second order stochastic dominance if for every $x \in X$ and $x' \in X$,

$$g_0(x, \xi) \succeq_2 g_0(x', \xi) \quad \Longrightarrow \quad \mathsf{E}_F g_0(x, \xi) \geq \mathsf{E}_F g_0(x', \xi) \quad \text{and} \quad \rho_F(g_0(x, \xi)) \leq \rho_F(g_0(x', \xi)); \tag{8}$$

then we can define the set $X_F(x_0)$ by

$$X_F(x_0) = \{x \in X, x \neq x_0 : \mathsf{E}_F(u - g_0(x, \xi))^+ \leq \mathsf{E}_F(u - g_0(x_0, \xi))^+ \quad \text{for every} \quad u \in R^1\}. \tag{9}$$

In the case when $X_F(x_0)$ is a nonempty set, then according to the relation (8) we can see that

$$x \in X_F(x_o) \quad \Longrightarrow \quad \mathsf{E}_F g_0(x, \xi) \geq \mathsf{E}_F g_0(x_0, \xi) \quad \text{and simultaneously} \quad \rho_F(g_0(x, \xi)) \leq \rho_F(g_0(x_0, \xi)).$$

Evidently, we can set

d. $X_F := X_F(x_0)$.

2 Some Definitions and Auxiliary Assertions

In this section we recall some auxiliary assertions suitable for stability and empirical estimates of the problems (1), (3) with constraints sets a., b., c., d. To this end, first, let $\mathcal{P}(R^s)$ denote the set of all (Borel) probability measures on R^s and let the system $\mathcal{M}_1^1(R^s)$ be defined by the relation:

$$\mathcal{M}_1^1(R^s) \;=\; \{\nu \in \mathcal{P}(R^s) : \int\limits_{R^s} \|z\|_1 d\nu(z) < \infty\}, \quad \|\cdot\|_1 \quad \text{denotes} \quad \mathcal{L}_1 \quad \text{norm in} \quad R^s.$$

We introduce the system of the assumptions:

A.1

1. $g_0(x,z)$ is for $x \in X$ a Lipschitz function of $z \in R^s$ with the Lipschitz constant L (corresponding to the $\mathcal{L}_1$ norm) not depending on x,
2. $g_0(x,z)$ is either a uniformly continuous function on $X \times R^s$ or there exists $\varepsilon > 0$ such that $g_o(x,\, z)$ is a convex bounded function on $X(\varepsilon)$; ($X(\varepsilon)$ denotes the ε–neighborhood of the set$X$),

B.1 For $P_F,\, P_G \in \mathcal{M}_1^1(R^s)$, there exist $\varepsilon > 0$ such that

1. $\bar{g}_0(x,\, z,\, y)$ is for $x \in X(\varepsilon)$, $z \in R^s$ a Lipschitz function of $y \in Y(\varepsilon)$ with a Lips. constant L_y; $Y(\varepsilon) = \{y \in R^{m_1} : y = h(x,\, z) \quad \text{for some} \quad x \in X(\varepsilon),\, z \in R^s\}$, $\mathsf{E}_F h(x,\, \xi)$, $\mathsf{E}_G h(x,\, \xi) \in Y(\varepsilon)$,
2. for every $x \in X(\varepsilon)$, $y \in Y(\varepsilon)$ there exist finite mathematical expectations, $\mathsf{E}_F \bar{g}_0(x,\, \xi,\, \mathsf{E}_F h(x,\, \xi))$, $\mathsf{E}_F \bar{g}_0(x,\, \xi,\, \mathsf{E}_G h(x,\, \xi))$, $\mathsf{E}_G \bar{g}_0(x,\, \xi,\, \mathsf{E}_F h(x,\, \xi))$, $\mathsf{E}_G \bar{g}_0(x,\, \xi,\, \mathsf{E}_G h(x,\, \xi))$,
3. $h_i(x,\, z)$, $i = 1,\, \ldots,\, m_1$ are for every $x \in X(\varepsilon)$ Lipschitz functions of z with the Lipschitz constants L_h^i (corresponding to $\mathcal{L}_1$),
4. $\bar{g}_0(x,\, z,\, y)$ is for every $x \in X(\varepsilon)$, $y \in Y(\varepsilon)$ a Lipschitz function of $z \in R^s$ with the Lipschitz constant L_z (corresponding to $\mathcal{L}_1$ norm),

B.2 $\mathsf{E}_F \bar{g}_0(x, \xi, \mathsf{E}_F h(x, \xi))$, $\mathsf{E}_G \bar{g}_0(x, \xi, \mathsf{E}_G h(x, \xi))$ are continuous functions on X.

If $F_i(z_i), G_i(z_i),\, i = 1,\, \ldots,\, s$ are one dimensional marginal distributions corresponding to $F,\, G$, then we can recall

Proposition 1. [7]. Let $P_F, P_G \in \mathcal{M}_1^1(R^s)$ and let X be a compact set. If

1. Assumption A.1 1 is fulfilled, then for $x \in X$ it holds

$$|\mathsf{E}_F g_0(x,\xi) - \mathsf{E}_F g_0(x,\xi)| \;\le L \sum_{i=1}^{s} \int\limits_{-\infty}^{+\infty} F_i(z_i) - G_i(z_i)| dz_i. \tag{10}$$

   If moreover A.1 2 is fulfilled, then also

$$|\varphi(F,\, X) - \varphi(G,\, X)| \le L \sum_{i=1}^{s} \int\limits_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| dz_i, \tag{11}$$

2. Assumptions B.1 are fulfilled, then there exist $\hat{C} > 0$ such that for $x \in X$

$$|\mathsf{E}_F \bar{g}_0(x, \xi, \mathsf{E}_F h(x, \xi)) \;-\; \mathsf{E}_G \bar{g}_0(x, \xi, \mathsf{E}_G h(x, \xi))| \le \hat{C} \sum_{i=1}^{s} \int\limits_{-\infty}^{\infty} |F_i(z_i) - G_i(z_i)| dz_i. \tag{12}$$

   If moreover B.2 is fulfilled, then also

$$|\bar{\varphi}(F,\, X) - \bar{\varphi}(G,\, X)| \le \hat{C} \sum_{i=1}^{s} \int\limits_{-\infty}^{\infty} |F_i(z_i) - G_i(z_i)| dz_i. \tag{13}$$

To complete auxiliary assertions we recall a very useful inequalities. To this end let $P_F,\ P_G \in \mathcal{M}_1^1(R^s)$ and let problems (1), (3) be well defined, then on employing the triangular inequality we can obtain

$$\begin{aligned} |\varphi(F,\, X_F) - \varphi(G,\, X_G)| &\leq |\varphi(F,\, X_F) - \varphi(G,\, X_F)| + |\varphi(G,\, X_F) - \varphi(G,\, X_G)|, \\ |\bar{\varphi}(F,\, X_F) - \bar{\varphi}(G,\, X_G)| &\leq |\bar{\varphi}(F,\, X_F) - \bar{\varphi}(G,\, X_F)| + |\bar{\varphi}(G,\, X_F) - \bar{\varphi}(G,\, X_G)|. \end{aligned} \tag{14}$$

Evidently, employing the assertion of Proposition 1 and the relations (14) we can bounded the gaps between optimal values of the problems (1), (3) with different distributions F and G. However to this end it is reasonable, first, to define for $\varepsilon \in R^1$ the sets X_F^ε by

- $$X_F^\varepsilon = \{x \in X:\ \mathsf{E}_F \bar{g}_1(x,\, \xi,\, \mathsf{E}_F h(x,\, \xi)) \leq \varepsilon\} \quad \text{in the case of constraints set } \ b. \ \text{ with } \ m = 1, \tag{15}$$
- $$\begin{aligned} X_F^\varepsilon \quad = \quad & \{x \in X:\ \mathsf{E}_F(u - g(x,\, \xi))^+ - \mathsf{E}_F(u - Y(\xi))^+ \leq \varepsilon \quad \text{for every} \quad u \in R^1\} \\ & \text{in the case of constraint set} \quad c. \end{aligned} \tag{16}$$

Employing the last relations, assumptions of Proposition 1 and the approach of the work [8] we can obtain:

1. in the case of constraint b. with $m = 1$ and the function $\bar{g}_1$ fulfilling the assumption B.1 that
$$X_G^{\delta-\varepsilon} \subset X_F^\delta \subset X_G^{\delta+\varepsilon} \quad \text{with} \quad \delta \in R^1, \quad \varepsilon = \hat{C} \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| dz_i, \tag{17}$$
2. in the case of constraint set c; $g(x,\, z),\ Y(z)$ to be for every $x \in X$ Lipschitz functions of $z \in R^s$ with the Lipschitz constant L_g not depending on $x \in X$, that
$$X_G^{\delta-\varepsilon} \subset X_F^\delta \subset X_G^{\delta+\varepsilon} \quad \text{with} \quad \delta \in R^1, \quad \varepsilon = 2L_g \sum_{i=1}^{s} \int_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| dz_i. \tag{18}$$

Evidently, the analysis and the results of this section can be employed in the case when the distribution function G is replaced by empirical one.

3 Empirical Estimates

First, in this section, we introduce a new system of the assumptions.

A.2
- $\{\xi^i\}_{i=1}^\infty$ is an independent random sequence corresponding to F,
- F^N is an empirical distribution function determined by $\{\xi^i\}_{i=1}^N, N = 1, 2, \ldots,$

A.3 $P_{F_i}, i = 1, \ldots, s$ are absolutely continuous w. r. t. the Lebesgue measure on R^1.

Empirical problem, corresponding to the "underlying" problem (1), can be introduced in the form:

Find
$$\bar{\varphi}(F^N,\, X_{F^N}) = \inf\{\mathsf{E}_{F^N} \bar{g}_0(x, \xi, \mathsf{E}_{F^N} h(x, \xi)) | x \in X_{F^N}\}. \tag{19}$$

Employing the idea of the paper [8], we can obtain.

Proposition 2. Let X be a nonempty compact set, $P_F \in \mathcal{M}_1^1(R^s)$, X_F be defined by the relation (5); $X_F,\ X_{F^N},\ N = 1,\ \ldots$ be nonempty compact sets. Let, moreover, $g(x,\, z),\ Y(z)$ be for every $x \in X$ Lipschitz functions of $z \in Z_F$ with the Lipschitz constant L_g not depending on $x \in X$. If

1. Assumptions B.1, B.2 , A.2 are fulfilled,
2. $\mathsf{E}_F\bar{g}_0(x,\,\xi,\,\mathsf{E}_F h(x,\,\xi))$ is a Lipschitz function on X,
3. there exists $\varepsilon_0 > 0$ such that X_F^ε (defined by the relation (15)) are nonempty compact sets for every $\varepsilon \in \langle -\varepsilon_0,\,\varepsilon_0\rangle$ and, moreover, there exists a constant $\hat{C} > 0$ such that

$$\Delta_n[X_F^\varepsilon,\; X_F^{\varepsilon'}] \;\le \hat{C}|\varepsilon - \varepsilon'| \quad \text{for} \quad \varepsilon, \varepsilon' \in \langle -\varepsilon_0,\,\varepsilon_0\rangle,$$

then

$$P\{\omega :\; |\bar{\varphi}(F,\,X_F) - \bar{\varphi}(F^N,\,X_{F^N})| \longrightarrow_{N\longrightarrow\infty} 0\} = 1. \tag{20}$$

($\Delta[\cdot,\,\cdot] := \Delta_n[\cdot,\,\cdot]$ denotes the Hausdorff distance in the subsets of n-dimensional Euclidean space; for definition see, e.g., [11].)

Proposition 3. Let X be a nonempty compact set, $P_F \in \mathcal{M}_1^1(R^s)$, X_F be defined by the relation (2); $X_F,\; X_{F^N},\; N = 1,\,\ldots$ be nonempty compact sets.

1. functions $\bar{g}_0,\,\bar{g}_1$ fulfil Assumptions B.1, B.2; Assumption A.2 is fulfilled,
2. $\mathsf{E}_F\bar{g}_0(x,\,\xi,\,\mathsf{E}_F h(x,\,\xi))$ is a Lipschitz function on X,
3. there exists $\varepsilon_0 > 0$ such that X_F^ε (defined by the relation (15)) are nonempty compact sets for every $\varepsilon \in \langle -\varepsilon_0,\,\varepsilon_0\rangle$ and, moreover, there exists a constant $\hat{C} > 0$ such that

$$\Delta_n[X_F^\varepsilon,\; X_F^{\varepsilon'}] \;\le \hat{C}|\varepsilon - \varepsilon'| \quad \text{for} \quad \varepsilon, \varepsilon' \in \langle -\varepsilon_0,\,\varepsilon_0\rangle,$$

then

$$P\{\omega :\; |\bar{\varphi}(F,\,X_F) - \bar{\varphi}(F^N,\,X_{F^N})| \longrightarrow_{N\longrightarrow\infty} 0\} = 1. \tag{21}$$

Remark. In the both cases (Proposition 2, Proposition 3) $\bar{\varphi}(F^N,\,X_{F^N})$ is a consistent estimate of $\bar{\varphi}(F,\,X_F)$. Evidently, it is possible also to prove results about the rate of convergence for this estimates. However to present the corresponding assertion is beyond the scope of this contribution.

It remains to deal with an analysis of the constraints set X_F corresponding to the case d. If we can assume that constants $\nu_2,\,\nu_1$ fulfill condition (7) with some $x_0 \in X$, then we can replace distribution function F by F^N and try to find x_0^N and $X_{F^N}(x_0^N)^N$ such that for $N = 1, 2, \ldots$

$$\mathsf{E}_F g_0(x_0^N,\,\xi) \ge \nu_2 - \frac{1}{N}, \quad \rho_F(g_0(x_0^N,\,\xi)) \le \nu_1 + \frac{1}{N}.$$

$$X_{F^N}(x_{0^N}) = \{x \in X, x \ne x_0^N : \mathsf{E}_F(u - g_0(x,\xi))^+ \le \mathsf{E}_F(u - g_0(x_0^N,\xi))^+ \quad \text{for every } u \in R^1\} \tag{22}$$

Let X be a compact set and let there exist a constant $C^1 > 0$ such that

$$|\rho_F(g_0(x,\,\xi)) \;-\; \rho_{F^N}(g_0(x,\,\xi))| \;\le\; C^1 \sum_{i=1}^{s} \int_{-\infty}^{\infty} |F_i(z_i) - F_i^N(z_i)| dz_i \quad \text{for every} \quad x \in X. \tag{23}$$

If we can assume that $\rho_F(g_0(x,\,\xi))$ is a uniformly continuous on X, assumptions A.1 are fullfilled, then we can obtain

$$\begin{aligned} P\{\omega : |\rho_F(g_0(x_0,\,\xi)) \;-\; \rho_{F^N}(g_0(x_0^N,\,\xi))| \longrightarrow_{N\longrightarrow\infty} 0\,\} &= 1, \\ P\{\omega : |\mathsf{E}_F g_0(x_0,\,\xi) \;-\; \mathsf{E}_{F^N} g_0(x_0^N,\,\xi)| \longrightarrow_{N\longrightarrow\infty} 0\,\} &= 1. \end{aligned} \tag{24}$$

Consequently, we have proven the convergence $\mathsf{E}_{F^N} g_0(x_0^N,\,\xi)$ to $\mathsf{E}_F g_0(x_0,\,\xi)$ and $\rho_{F^N}(g_0(x_0^N,\,\xi))$ to $\rho_F(g_0(x_0,\,\xi))$ (a.s.). This result is important for us. However, our aim is to find out the assumptions under which

$$P\{\omega : |\bar{\varphi}(F, X_F(x_0)) - \bar{\varphi}(F^N, X_{F^N}(x_0^N)))| \longrightarrow_{N\longrightarrow\infty} 0\} \;= 1. \tag{25}$$

Evidently if we can assume that

$$P\{\omega : \Delta[X_F^0(x_0),\, X_{F^N}^0(x_0^N)] \longrightarrow_{N\longrightarrow\infty} 0\} \;= 1 \tag{26}$$

and if we add to all above mentioned assumptions:

- $P_F \in \mathcal{M}_1^1(R^1), X, X_F(x_0), X_{F^N}(x_0^N),\ N = 1, 2, \ldots$ are nonempty compact sets,
- $\mathsf{E}_F g_0(x_0,\ \xi,\ \mathsf{E}_F h(x, \xi))$ is a Lipschitz function on X,
- Assumptions B.1, B.2, A.2 are fulfilled,

then the assertion (25) is valid.

4 Conclusion

The contribution is focused on a special type of the stochastic optimization problems in which dependence on the probability measure is not linear. This type of problems corresponds to real-life situations rather often. A risk given by variance (in the mean-risk problem) is well known example of this class. The aim of this contribution is to show that many properties of these problems (under acceptable assumptions) are similar to them in the "classical" case. However the detailed analysis is beyond the scope of this contribution.

Acknowledgements

This work was supported by the Czech Science Foundation under grant 18-02739S.

References

[1] Birge, J. R. & Louveaux, F. (1999). *Introduction in Stochastic Programming.* Springer: Berlin.

[2] Dentcheva, D., Penev, S. & Ruszczynski, A. (2014). *Statistical Estimation of Composite Risk Functionals and Risk Optimization Problem.* Conwelt University Library 2014.

[3] Ermoliev, Yu. & Norkin, V. (2013). Sample average approximation method for compound stochastic optimization problems. *SIAM J. Optimization*, 23 (4), 2231–2263.

[4] Gutjahr, W. J. & Pichler, A. (2016). Stochastic multi–objective optimization: a survey on non–scalarizing methods. *Ann. Oper. Res.*, 236, 475–499.

[5] Kaňková, V. & Houda, M. (2006). Empirical processes in stochastic programming. In M. Hušková and M. Janžura (Eds.) *Proceedings of Prague Stochastics 2006* (pp. 426–436). Prague: MATFYZPRESS.

[6] V. Kaňková (2013). Risk measures in optimization problems via empirical estimates. *Czech Economic Review,* 7(3), 162–177.

[7] Kaňková, V. & Houda, M. (2015). Thin and heavy tails in stochastic programming. *Kybernetika*, 51(3), 433–456.

[8] Kaňková, V. & Omelčenko, V. (2018). Stochastic optimization problems with second order stochastic dominance constraints via Wasserstein metric. *Kybernetika*, 54 (6), 1231–1246.

[9] Levy, H. (2006). *Stochastic Dominance (Investment, Decision Making under Uncertainty), 2nd edition.* Springer: New York.

[10] Ogryczak, W. & Rusczynski, A. (1999). From the stochastic dominance to mean–risk models: semidevations as risk measure. *European J. Oper. Res.*, 116, 33–50.

[11] Rockafellar, R. & Wets, R. J. B. (1983). *Variational Analysis.* Berlin: Springer.

[12] Schulz, R. (1996). Rates of convergence in stochastic programs with complete integer recourse. *SIAM J. Optimization*, 6 (4), 1138–1152.

[13] Shapiro, A., Dentcheva, D. & Ruszczynski, A. (2009). *Lectures on Stochastic Programming – Modeling and Theory.* Society for Industrial and Applied Mathematics and Mathematical Programming Society: Philadelphia.

[14] Shorack G. R. & Wellner J. A . (1986). *Empirical Processes and Applications to Statistics*. Wiley: New York.

Remarks on Economic Characteristics of Rectifying AOQL Plans by Variables

Nikola Kaspříková[1]

Abstract. The average outgoing quality limit acceptance sampling plans were originally designed by Dodge and Romig to minimize the mean inspection cost per lot of the process average quality. The paper addresses the sampling plans for the inspection by variables and considers the effects of wrong guess of the process average quality value on the economic performance of the plans.
The rectifying AOQL sampling plans are calculated and evaluated using an R software extension package.

Keywords: acceptance sampling, inspection cost, AOQL

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The sampling inspection plans as the tools for statistical quality control include the decision procedures using the inspection by attributes and the decision procedures using the inspection by variables. Dodge and Romig (see e.g. [2]) have designed the rectifying average outgoing quality limit (AOQL) sampling plans minimizing the mean inspection cost per lot of process average quality when the remainder of rejected lots is inspected. The plans were originally designed by Dodge and Romig for the inspection by attributes. Plans for the inspection by variables and for the inspection by variables and attributes (all items from the sample are inspected by variables, the remainder of rejected lots is inspected by attributes) were then proposed and it was shown that these plans are in many situations more economical than the corresponding Dodge-Romig attribute sampling plans. The AOQL plans for inspection by variables and attributes have been introduced in [8], using approximate calculation of the plans. Exact operating characteristic, using non-central t distribution, has been later implemented for the calculation of the plans in the LTPDvar package [5]. The operating characteristics used for these plans are discussed by Jennett and Welch in [3] and by Johnson and Welch in [4]. It has been shown that these plans are in many situations better than the original attribute sampling plans and similar results have been obtained for the LTPD plans, the analysis is provided in [7]. The recent development of acceptance sampling plans includes the design of plans which are using the exponentially weighted moving average (EWMA) statistic in the inspection procedure. Using the EWMA statistic enables some savings in the cost of inspection in comparison with the plans without memory, as it allows using information on the quality in the previous lots. The EWMA-based average outgoing quality limit plans for the unknown standard deviation case are introduced in [6]. A simple economic model has been used for the evaluation of economic efficiency of the plans.

This paper considers the plans proposed in [6] and shows the economic characteristics of these plans measured by the mean cost of inspection per lot of the process average quality. It considers the effects of wrong guess of the process average quality value on the economic performance of the plans.

The structure of this paper is as follows: first, the design of the original AOQL sampling plans for the inspection by attributes, as introduced by Dodge and Romig (see [2]), is recalled. Then we recall the design of the the AOQL variables sampling plans based on the usage of the EWMA statistic in the inspection procedure and also recall the simple economic model used to asses the comparative economic efficiency of the plans. The optimal acceptance sampling plan for the unknown standard deviation case is calculated in a short case study and the effect of the wrong supposed value of the process average proportion defective on the mean inspection cost per lot of the process average quality is shown.

The calculation and economic evaluation of the plans is done using the free [5] package which has been published on the Comprehensive R Archive Network.

[1] University of Economics in Prague, Department of Mathematics,
Nám. W. Churchilla 4, Praha, Czech Republic, nikola.kasprikova@vse.cz

2 AOQL attributes inspection plans

For the inspection procedures in which each inspected item is classified as either good or defective (the acceptance sampling by attributes), Dodge and Romig (see [2]) consider sampling plans $(n,\ c)$ which minimize the mean number of items inspected per lot of process average quality I_s, assuming that the remainder of the rejected lots is inspected

$$I_s = N - (N - n) \cdot L(\bar{p}; n, c) \tag{1}$$

under the condition

$$\max_{0<p<1} AOQ(p) = p_L. \tag{2}$$

The notation in equations (1) and (2) is as follows:
N is the number of items in the lot (the given parameter),
$\bar{p}$ is the process average fraction defective (the given parameter),
p_L is the average outgoing quality limit (the given parameter, denoted AOQL),
n is the number of items in the sample ($n < N$),
c is the acceptance number (the lot is rejected when the number of defective items in the sample is greater than c),
$L(p)$ is the operating characteristic (the probability of accepting a submitted lot with the fraction defective p).

The function AOQ is the *average outgoing quality, AOQ*(p) is the mean fraction defective after inspection when the fraction defective before inspection was p. The AOQ function is continuous in [0,1] and reaches minimum value 0 for $p = 0$ and for $p = 1$. For a more detailed discussion of the properties of the AOQ function, see [2]. The average outgoing quality (where all defective items found are replaced by good ones) is approximately

$$AOQ(p) = \left(1 - \frac{n}{N}\right) \cdot p \cdot L(p; n, c). \tag{3}$$

So the condition (2) can be rewritten as

$$\max_{0<p<1} \left(1 - \frac{n}{N}\right) \cdot p \cdot L(p; n, c) = p_L. \tag{4}$$

The condition (2) protects the consumer against having an average outgoing quality higher than p_L (the chosen value), regardless of what the fraction defective p is before inspection. The equation (4) can be solved using numerical methods (interval halving or Fibonacci search respectively).

3 AOQL sampling plans for the inspection by variables

As an alternative to the sampling plans for the inspection by attributes, the AOQL plans for the inspection by variables and attributes based on the exponentially weighted moving average statistic for the case of the unknown standard deviation were designed. The operating characteristic used in [6] for the design of the AOQL plans is applied here. The AOQL plans were designed under the following assumptions: The measurements of a single quality characteristic X are independent and identically distributed normal random variables with parameters μ and σ^2. We consider the unknown σ case. For the quality characteristic X, either an upper specification limit U (the item is defective if its measurement exceeds U), or a lower specification limit S (the item is defective if its measurement is smaller than S), is given.

For the rectifying AOQL plans minimizing the mean inspection cost per lot of the process average quality we shall use a procedure based on the EWMA statistic. The procedure is as follows: draw a random sample of n items from the lot and compute the sample mean $\bar{x}$, sample standard deviation s and the statistic Z at time t (lot order $t = 1, 2, \ldots$; the starting value for Z is $\bar{x}$) as

$$Z_t = \lambda \bar{x} + (1 - \lambda) Z_{t-1}, \tag{5}$$

where λ is a smoothing constant (between 0 and 1).

Accept the lot if

$$\frac{U - Z_t}{\sigma} \geq k \quad \text{or} \quad \frac{Z_t - L}{\sigma} \geq k. \tag{6}$$

The operating characteristic is (see [1])

$$L(p) = \Phi(u_{1-p} c_4 - k) \sqrt{\frac{1}{\frac{\lambda}{n(2-\lambda)} + k^2 (1 - {c_4}^2)}}, \tag{7}$$

where

$$c_4 = \sqrt{(2/(n-1))}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}. \tag{8}$$

The symbols used in (7) and (8) are Γ for the gamma function and Φ for the cumulative distribution function of the standard normal distribution.

The plan parameters (n, k) are determined so that the plan has optimal economic characteristics and satisfies the requirement (2), when (7) is used as the operating characteristic.

Regarding the economic efficiency, Klufa in [8] uses the economic model, used for a more detailed evaluation of the sampling plans in [7] and searches for the acceptance plan (n, k), minimizing the mean inspection cost per lot of the process average quality C_{ms} under the condition (2). We use this model for the economic evaluation of the plans. The inspection cost per lot, assuming that the remainder of the rejected lots is inspected by attributes (the inspection by variables and attributes), is $n\,c_m^*$, with the probability $L(p; n, k)$, and $[n\,c_m^* + (N-n)\,c_s^*]$ with the probability $[1-L(p; n, k)]$, where c_s^* is the cost of the inspection of one item by attributes, and c_m^* is the cost of the inspection of one item by variables. The mean inspection cost per lot of the process average quality is then

$$C_{ms} = n \cdot c_m^* + (N-n) \cdot c_s^* \cdot [1 - L(\bar{p}; n, k)]. \tag{9}$$

Let us denote

$$c_m = \frac{c_m^*}{c_s^*}. \tag{10}$$

Instead of C_{ms} we will look for the acceptance plan (n, k) minimizing

$$I_{ms} = n \cdot c_m + (N-n) \cdot [1 - L(\bar{p}; n, k)] \tag{11}$$

(both functions C_{ms} and I_{ms} have a minimum for the same acceptance plan ($C_{ms} = I_{ms} \cdot c_s^*$)) under the condition (2).

The I_{ms} function is the mean inspection cost of the plan per lot of the process average quality, expressed in the c_s^* units.

4 Economic evaluation of the plans

Let's calculate the AOQL acceptance sampling plan for sampling inspection by variables when the remainder of rejected lots is inspected by attributes and if the standard deviation of the quality characteristic is unknown in a case study below. The EWMA-based statistic will be used in the inspection procedure. The economic performance of the plan will be evaluated with the mean inspection cost per lot of the process average quality.

Example 1. A lot of N = 5000 items is considered in the acceptance procedure. The average outgoing quality limit p_L = 0.015. It is known that average process quality is $\bar{p}$ = 0.01. A cost of inspecting an item by variables is known to be by 80 percent times higher than the cost of inspecting an item by attributes, so c_m parameter equals 1.8. Find the AOQL acceptance sampling plan for sampling inspection by variables when the remainder of rejected lots is inspected by attributes, using the EWMA statistic with smoothing constant 0.9.

The plan can be calculated using the LTPDvar package [5]. for the R software [9]. The solution is n = 80, k = 1.929489. The mean inspection cost per lot of the process average quality for this plan is 213.36.

The values of the input parameters influence the resulting sampling plan and its economic characteristics. The AOQL acceptance sampling plans are optimized with respect to mean inspection cost per lot of the process average quality $\bar{p}$. If the supposed value of process average quality (let us denote such value p_a) is different from the true value of $\bar{p}$, the resulting acceptance sampling plan will still satisfy the condition (2), but the value of I_{ms} may not be optimal. The Table 1 shows the mean inspection cost per lot of the process average quality I_{ms} for plans calculated for various values of the supposed process average quality p_a, keeping the other parameters from our example unchanged. The values of the mean inspection cost (see also Figure 1) are increasing if the guess value p_a becomes farther from the true process average quality value. Underestimating the true values shows more significant increase in the mean inspection cost than overestimating $\bar{p}$.

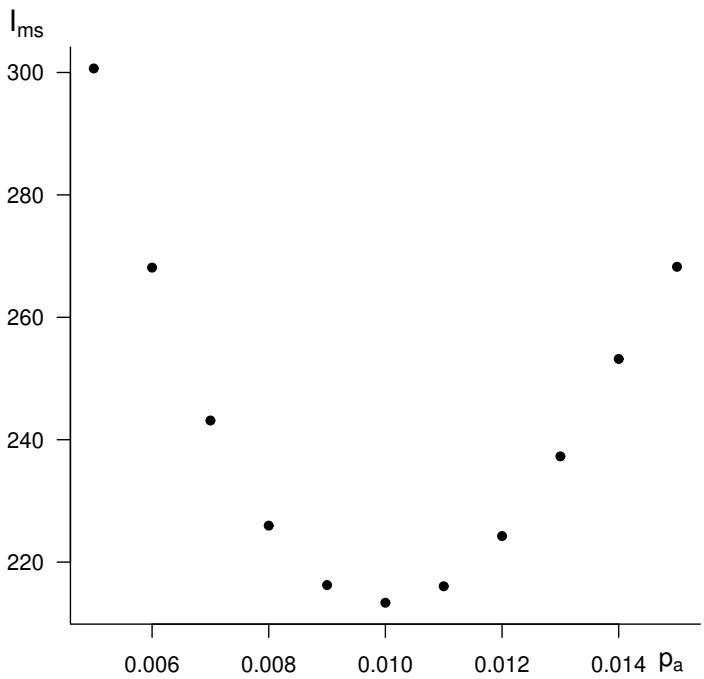


Figure 1 Inspection cost for plans constructed for various p_a, real $\bar{p} = 0.01$

Example 2. Let us change some of the input parameter values from Example 1 and consider now the lot size $N = 3500$, $\bar{p} = 0.005$, $c_m = 2$.

The Table 2 and Figure 2 show the situation after parameter update in Example 2. The values of the mean inspection cost are again increasing if the guess value p_a becomes farther from the true process average quality value, underestimating the true values shows more significant cost increase.

p_a	n	k	I_{ms}
0.005	40	1.903871	300.66
0.006	46	1.907847	268.12
0.007	53	1.912658	243.14
0.008	61	1.918028	225.96
0.009	70	1.923705	216.25
0.01	80	1.929489	213.36
0.011	90	1.934730	216.06
0.012	102	1.940369	224.26
0.013	115	1.945773	237.28
0.014	128	1.950551	253.20
0.015	139	1.954172	268.25

Table 1 AOQL plans for process average quality guess between 0.005 and 0.015, real $\bar{p} = 0.01$

5 Conclusion

The AOQL sampling plans for the inspection by variables when the statistic based on the exponentially weighted moving average is used and the remainder of the rejected lots is inspected by attributes were

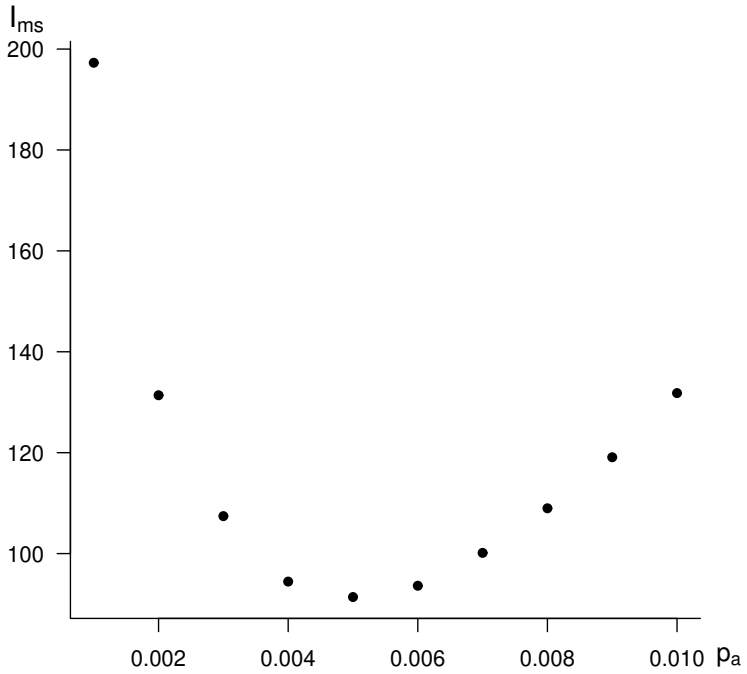


Figure 2 Inspection cost for plans constructed for various p_a, real $\bar{p} = 0.005$

p_a	n	k	I_{ms}
0.001	16	1.916008	197.29
0.002	21	1.901807	131.39
0.003	25	1.898272	107.43
0.004	30	1.897935	94.45
0.005	35	1.899709	91.39
0.006	40	1.902432	93.62
0.007	46	1.906207	100.14
0.008	52	1.91013	108.98
0.009	58	1.913999	119.1
0.01	65	1.918322	131.81

Table 2 AOQL plans for process average quality guess between 0.001 and 0.01, real $\bar{p} = 0.005$

addressed under the assumption that the standard deviation of the quality characteristic is unknown. The mean inspection cost per lot of the process average quality has been used as the economic characteristic of the plans. The effects of the supposed value of the process average proportion defective on the mean inspection cost per lot of the process average quality eas shown and it has been observed that the mean inspection cost is increasing when the guess differs from true process average quality value in both directions, underestimating the true values showed more significant increase in the mean inspection cost than overestimating.

Acknowledgements

This paper has been produced with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Aslam, M., Azam, M., and Jun, C. (2015). A new lot inspection procedure based on exponentially weighted moving average, *International Journal of Systems Science*, 46, 1392–1400.

[2] Dodge, H., F., and Romig, H. G. (1998). *Sampling Inspection Tables: Single and Double Sampling*. New York: John Wiley.

[3] Jennett, W. J., and Welch, B. L.(1939). The Control of Proportion Defective as Judged by a Single Quality Characteristic Varying on a Continuous Scale, *Supplement to the Journal of the Royal Statistical Society*, 6, 80–88.

[4] Johnson, N. L., and Welch, B. L. (1940). Applications of the Non-central t distribution, *Biometrika*, 38, 362–389.

[5] Kasprikova, N. (2020). *LTPDvar: LTPD and AOQL plans for acceptance sampling inspection by variables.* R package version 1.2. http://CRAN.R-project.org/package=LTPDvar.

[6] Kasprikova, N. (2018). Economic Efficiency of Rectifying AOQL Plans Based on the EWMA Statistic. *Journal of Testing and Evaluation*, 47, DOI: 10.1520/JTE20170312.

[7] Kasprikova, N., and Klufa, J. (2015). AOQL sampling plans for inspection by variables and attributes versus the plans for inspection by attributes, *Quality technology and quantitative management*, 6, 133–142.

[8] Klufa, J. (1997). Dodge-Romig AOQL single sampling plans for inspection by variables, *Statistical Papers*, 38, 111–119.

[9] R Core Team. (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Czech Local Action Groups Technical Efficiency

Renata Klufová[1], Jana Klicnarová[2]

Abstract. The rural development policy and CLLD are widely discussed topics. Hence, it is necessary to evaluate the effects of provided financial means. A negative impact of the subsidies on lowering of the motivation of beneficiaries and worsening of their operation is often discussed in the literature (besides other aspects). The aim of the contribution is to calculate a technical efficiency of the beneficiaries and evaluate the influence of the subsidies on it. We focus on the financial means from the Integrated Regional Operational Program (IROP – Priority Axis 4) on the period of 2014–2020, but also considers the previous programming period. The Data Envelopment Analysis (DEA) method was used for the LAGs assessment.

Keywords: LEADER/CLLD, LAGs, DEA, Integrated Regional Operation Program

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The LEADER programme supports the endogenous development of rural regions through a bottom-up and integrated approach implemented by the Local Action Groups (LAGs). LAGs are expected to induce added values not achieved by the other sources such as better targeting of funds by addressing local needs by decision-making at the ground and evoking synergies through cross-sectoral collaboration and sharing of resources between partners as well as improving the quality of life and furthering local identity [2] and governance skills. The LAGs contribute to strengthening the social capital, which is an important factor for the development of the locality, and the LEADER with LAGs is an approached built upon social capital [9].

However, in many cases, LAGs have become synonymous with inefficiency, a lack of control and transparency, and the removal of private shareholder responsibility, due in part to the standard on the basis of which their intervention programs are fully funded by EU resources[15].

Regional checks on LAG activities are often formal – expenditure is checked to ensure that it complies with the procedures defined in the contracts. No long-term assessment of their initiatives is carried out, nor are checks performed to verify the functioning of their bodies regarding compliance with legal requirements on the part of administrators, the disclosure of documents, the balance of powers between bodies, the regularity of budgets, and the actual payment of shares by private shareholders.

The scientific community has already discussed the evaluation of LAG´s effectiveness in achieving the programme´s objectives. One reason for the discussion is that the effectiveness of LAGs also depends on intangible capacities such as social relations whose evaluation still appears to be difficult [12].

LEADER was administered originally only in the Rural Development Program, since 2014 has been extended to another three operation programs as one of three integrated tools under the name Community-led Local Development (CLLD) – others are Integrated Territorial Investments (ITI) and Integrated Territorial Development Plans (ITDP). In the programming period 2014–2020 priority axis 4 of IROP provides options for the development of CLLD. This priority axis contributes to the achievement of thematic objectives [6] – TO 5 Promoting climate change adaptation, risk prevention and management, TO 6 Preserving and protecting the environment and promoting resource efficiency, TO 7 Promoting sustainable transport and removing bottlenecks in key network infrastructures, TO 9 Promoting social inclusion, combating poverty and discrimination, TO 10 Investing in education, training and vocational training for skills and lifelong learning and TO 11 Increasing institutional capacity of public authorities.

1.1 Approaches to LAGs evaluation

The history of LEADER and its evaluation is extensive and has grown progressively over the past twenty-five years. LEADER began as a pilot initiative in 1991 (LEADER I). It has evolved with numerous iterations

[1] University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, klufova@ef.jcu.cz
[2] University of South Bohemia/Faculty of Economics, Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, janaklic@ef.jcu.cz

(LEADER II, LEADER +) into a widespread; mainstream policy instrument used to support the development of rural areas across the European Union. The findings and recommendations of various LEADER evaluations prove that this policy tool has played a critical role in improving and developing the LEADER approach's conceptual foundation. The path of LEADER evaluations conveys a particular story, one of institutional and societal learning.

In the programming period 2014–2020, LEADER is implemented as part of Community-Led Local Development (CLLD). The CLLD strategies prepared and implemented through the LEADER method include the description of specific monitoring and evaluation arrangements. These arrangements are the basis for Local Action Groups (LAGs) to carry out the CLLD strategy's monitoring and evaluation activities.

The evaluation of the impact of the LEADER program has been subject to broad discussion. Especially the assessment of possible intangible outputs and added values is no easy task. Therefore first analyses and evaluations of the activity of LAGs undertaken in 2007 were mostly descriptive [13]. According to Osti [11] it is "Impossible to construct a standardized model for a general application, ... any local the territory must determine its own development parameters." As we used publicly available data of the LAGs' activities, and decided to evaluate the LAGs' effectiveness by using a DEA method, we focus on similar analyses.

Vrabková & Šaradín [17] evaluated LAGs in terms of technical efficiency based on DEA model, Pechrová, A. & K. Boukalová [14] created a typology of the Czech LAGs according to their individual features and its organizational background. A similar approach also used Lopolito et al. [8]; however, no demographics parameters were included in the model.

2 Material and Methods

Our aim is also to include demographic parameters in the evaluation. In [7], we studied correlation coefficients among some demographic parameters (resp. their changes) and amount of subsidies; and regression analysis with these factors. In this study, we focus on the possibility of identification of effective LAGs, where the efficiency could be based on demographics factors, too. Publicly available data [16] has been used for the LAGs evaluation. Table 1 shows variables used for a particular programming period.

variable	**2007–2013**	**2014–2020**
UNEM1/2	share of unemployed persons registered (%)	share of unemployed persons registered (%)
lTr0814/1317	linear trend of net migration 2008–2014	linear trend of net migration 2013–2017
EMP	no. of employees of the LAG administration	no. of employees of the LAG administration
IROPPP	subsidies from IROP per capita	subsidies from IROP per capita
ClCall		number of closed calls
PCFR		percent. of funds reimbursed to applicants (%)

Table 1 Analyzed characteristics of LAGs – programming periods

Because the calculation of the unemployment rate changed [18] in 2013, this indicator – for both programming periods – has been recalculated for individual LAGs according to the new methodology. The unemployment is thus calculated as the share of unemployed people on the population aged 15–64 years. Let us remark that the correlation coefficient of the unemployment rate before and after the change (for individual LaG's) was not too high (about 0.3).

The aim of the LAGs evaluation in the Czech Republic would be to identify factors that affect their success rate in obtaining subsidies. However, as we mentioned above, we know (see [7]) that this success is too strongly correlated with the number of inhabitants. We do not see any other parameter important for LAG efficiency, which could cause this dependence; hence it seems that the amount of obtained subsidies is not significantly related to any other factor indicating LAG's efficiency. More precisely, we know (see [7]) that for the first period, the correlation coefficient between the number of inhabitants and the amount of IROP subsidies is higher than 0.995. Therefore, the amount of subsidies per capita is close to a constant. For the second period, the dependence is similar; the correlation coefficient is even higher. For this reason, we do not see any sense in such a study. We decided to identify effective units according to other criteria and then (in future research) to study the relevancy of such results in comparison with the allocation of subsidies. For such evaluation, we decided to apply Data Envelopment Analyses.

In our analysis, we are looking for an answer to the following question:

- Which demographic parameters can be used for the identification of LAG's efficiency?

The criteria for LAG's efficiency we can split into two main groups:

1. LAG's parameters,
2. demographics parameters.

How to use parameters from the first section is clear and applied in other studies, see for example [8], [15]. In our study, we use a number of employees (EMP), IROP subsidies per capita (IROPPP), number of closed calls (ClCall); and percentage of funds reimbursed to applicants (PCFR).

On the other hand, there is a problem with the choice of demographic parameters. It is well-known that the aim of LAG is to help with employment in the region, to help with social life, and so on. Hence, we decided to use a share of unemployed and net migration, which many researchers also used. The absolute values of the unemployment rate and migration balance depend mainly on the location of the LAG (not on LAG activities).

So, we decided to identify successful LAGs to search for LAGs with the smallest number of employees, the best progression of the unemployment rate and the migration during the programming period, and the highest value of subsidies per capita. To evaluate the progress in migration balance, we first suppose to use an index which would be a quotient of migration balances at the beginning and the end of the programming period. However, the correlation coefficient between these indexes for the first and second programming periods was non-significant (it was less than 0.02), so such index is not a correct measure of migration development in individual LAGs. Hence, it was necessary to apply a more complicated measure of this development – for any LAG, and each period we estimated a linear trend of migration balance, we set that the best LAG has the highest slope. The same idea we applied for the unemployment rate. Since there was a strong trend in the unemployment rate in the whole country, we used a ratio of the unemployment rate for given LAG and year to the average unemployment rate in the year. And we studied behaviour of these ratios. We suppose that this choice of parameters could be right because if we check the correlation coefficient between these parameters for individual LAGs in the first and second periods, it is 0.85. Therefore, it seems to be a stable parameter. The situation with a measure of the change in the unemployment rate is more complicated. It is probably caused by a difference in the trend of the unemployment rate – it grew up to 2014, and then it has been declining.

3 Czech LAGs performance evaluation – results and discussion

For the evaluation of LAG's efficiency, we applied the Data Envelopment Analysis. For the first programming period, we used the following criteria – the number of employees, IROP subsidies per capita, the trend in migration balance, and the trend of ratios of the individual unemployment rate to an average one. For the second period, we used the same variables and added the number of closed calls and the percentage of funds reimbursed to applicants. The following LAGs were not included in the analyses due to the missing data – Dolnobřežansko, Rozvoj Kladenska a Prahy-Západ, Svitava, and Otevřené zahrady Jičínska for both periods; and Lužnicko, Na cestě k prosperitě, and Polabí for the second period. The LAG Jihozápad must be excluded too because it was set later, hence it is not correct to use some parameters.

Figures 1 and 2 show LAGs efficiencies and the development of the share of unemployed for both of the programming periods. LAGs Achát, Brdy, Cínovecko, Šluknovsko, Frýdlanstko, Slezská brána, Uničovsko, Jemnicko, and Kelečsko-Lešensko-Starojicko seem to be effective in the period 2007–2013. From some point of view, these LAGs have the best combination (each in a different way) of analyzed parameters. For other LAGs, there exist so-called peer LAGs, which have a better combination of parameters.

As we can see from figure 1, LAGs with high values of efficiency in 2007–2013 shows mainly negative trend in the development of unemployment (northern and western Bohemia, part of the Vysočina District and South Moravian region), i. e. LEADER activities probably helped to reduce unemployment in these LAGs. For LAGs in border areas, the question arises as to how much these trends are affected by cross-border cooperation and commuting abroad. Consequently, there is a need to include other variables that appropriately express these processes.

On the opposite, a growing trend in unemployment is visible in Central and South Bohemia and the northern part of Olomoucký district (Jesenicko), which is problematic in the long run. This growing trend correlates with lower LAGs efficiency. In this context, it is possible to ask whether the above facts are influenced by the structure of activities of individual LAGs (e.g., the predominant focus on infrastructure support and technical projects, with less emphasis on the development of human and social capital, etc.). This is one of the other issues that need to be addressed in further research.

Comparing the effectiveness of the LAG in the period 2007–2013 with the following programming period,

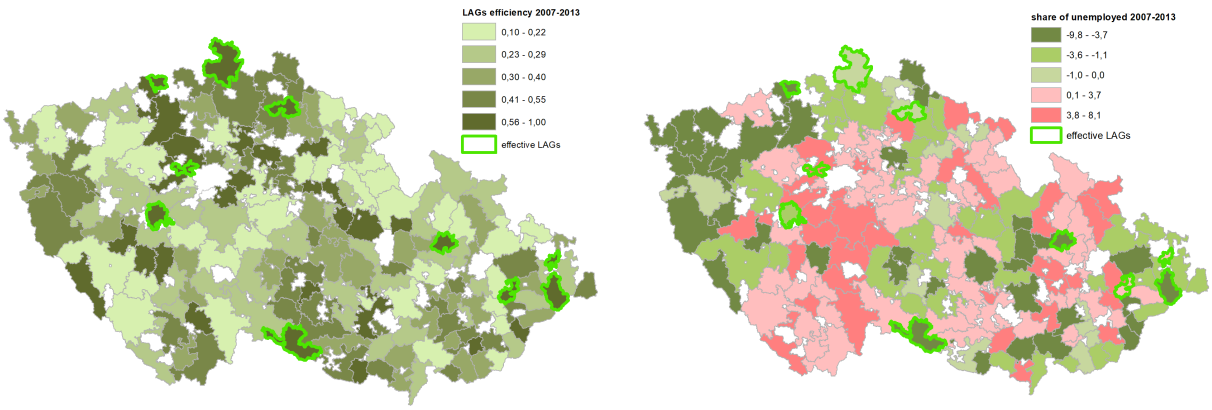


Figure 1 LAGs efficiency score and share of unemployed 2007–2013

the figure shows a similar geographical structure. Number of effective LAGs in 2014–2020 has arisen by 9 to 13 (more parameters have been included). Some LAGs north of Prague continue to show high efficiency values, but efficiency has improved in the southern part of the state and in some peripheral areas. This may be related to the growing experience of LAG staff in designing and implementing individual activities, specific tools of state authorities for regional development focused on peripheral areas (whether external or internal peripheries), but also, for example, other processes occurring in recent decades (counter urbanization).

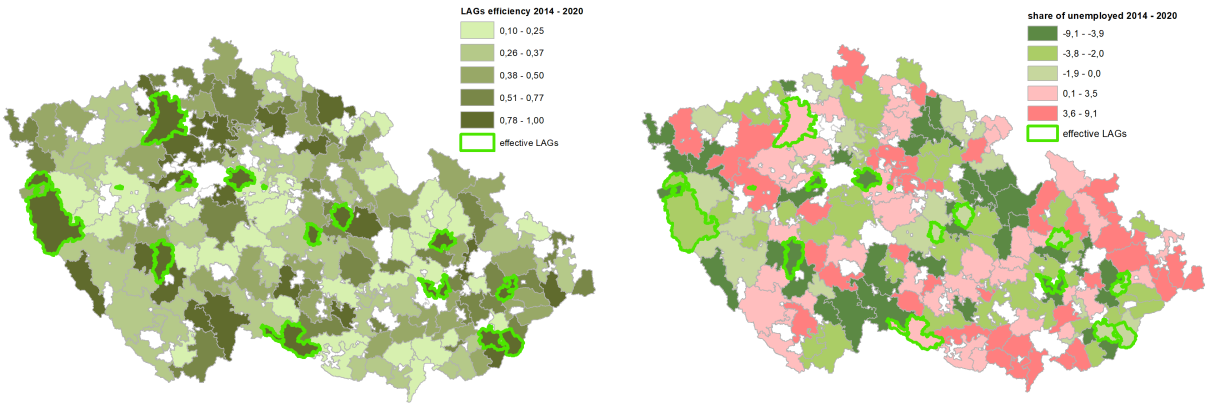


Figure 2 LAGs efficiency score and share of unemployed 2014–2020

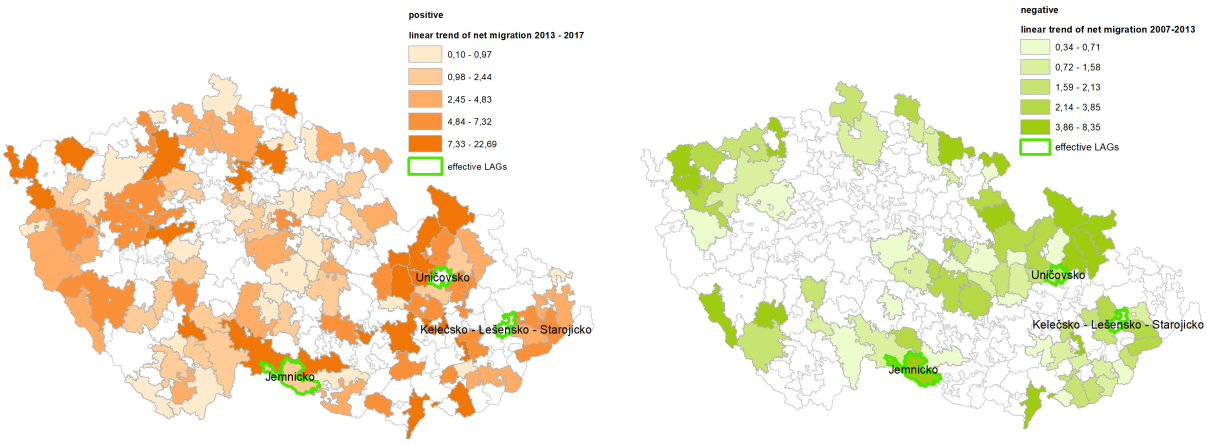


Figure 3 Linear trend of net migration 2013–2017

The growing attractiveness of non-urban areas, which can be supported by bottom-up activities, can be seen, among other things, in the development of the net migration (see figure 3). It is closely connected with the

offer of job opportunities, suitable housing, civic amenities, leisure activities, etc.

LAG	EMP	Migration – lin. trend	ratio unemployment – lin. tend	IROPPP
Uničovsko	1	−1.624	−0.082	1.748
Jemnicko	1	−5.010	−0.045	1.929
Kelečsko-Lešensko-Starojicko	1	6.273	0.012	2.245

Table 2 LAGs effective in both programming periods – 2007–2013

LAGs Jemnicko, Uničovsko a Kelečsko-Lešensko-Starojicko seem to be effective in both programming periods (tables 2 and 3). They are depicted in figure 2. Two of them are situated in the Moravian part of the Czech Republic, LAG Jemnicko consists of 47 municipalities from three districts (Vysočina, Jihočeský, and Jihomoravský). The cooperation of a relatively large number of municipalities of this LAG evidently helps to overcome the disadvantages of the peripheral location of the area on the border with Austria, but partly also the so-called inner periphery, which is represented by municipalities at the regional border (see e.g. [1]).

LAG	EMP	Migration – lin. trend	ratio unemployment – lin. tend	IROPPP	ClCall	PFCR
Uničovsko	1	−2.020	−0.013	46.730	4	27.45
Jemnicko	1	−5.494	0.027	49.991	8	21.14
Kel.-Leš.-Star.	1	6.055	0.0166	48.020	5	26.33

Table 3 LAGs effective in both programming periods – 2014–2020

All the effective LAGs have only one full-time employee, which was influential in the evaluation. In the first period, each of them was successful in one demographic parameter and not in the other (the aim is to have a positive trend in migration balance and a negative trend in the unemployment rate). LAG Kelečsko-Lešensko-Starojicko obtained significantly higher subsidies from IROP per capita in the first period, as well as Jemnicko. LAG Jemnicko has been successful in subsidies per capita also in the second period. In the second period, all of these LAGs have quite a high number of closed calls (especially if we take into account that they have only one full-time employee).

Despite the use of different variables in a given programming period, the quintile cartograms of DEA efficiency in both periods show a similar spatial pattern. Certain common features can be identified in both:

1. Effective LAGs and the ones with a low share of unemployment and positive migration gains perform the highest values of DEA efficiency score in our definition and vice versa. It can, therefore, be assumed that the support of local activity through the LAG leads, among other things, to the creation of job opportunities and the reduction of unemployment. This leads to the increasing attractiveness of the LAG for migration. However, without other detailed analyses, we can not make such conclusion. It is necessary to study the trends of chosen demographic factors if they are affected by LAG's activities, or are affected by some external factors.
2. High DEA efficiencies seem to be correlated with convenient geographical location with good connections to highways and expressways or railways visible when displaying appropriately selected data layers in a GIS environment. Therefore, in further analysis, it would be appropriate to include a suitably chosen and expressed the geographical location factor. The role of geographical location for regional development and efficiency is widely discussed in many works (e. g. [4], [5]).

4 Conclusions

When evaluating the LAG, we tried to answer the question of which demographic factors should be used to evaluate the effectiveness of the LAG. Based on the results discussed in the article above, it was confirmed that the basic factors that need to be monitored include the level of unemployment and its development (it would be appropriate to supplement the data on labor migration and commuting) and net migration (as an indicator of permanent change of residence to some extent expressing the attractiveness of the area). However, it will be appropriate to include variables expressing the geographical location and structure of the LAG's activities in the evaluation of the effectiveness of the LEADER program.

The evaluation of the effectiveness of the LAG was carried out based on quantitative publicly available data. However, as already mentioned in the introduction, it is relatively challenging to include soft factors (social capital, successful involvement in various networks, regional identity, the existence of quality strategic documents, etc.) to the LAG's evaluation. It is also necessary to check that LAGs have an impact on analyzing

demographics factors. Therefore, these topics are a major challenge for further work on this topic, provided, of course, that appropriate data are available. The factors mentioned above are becoming more frequent in the various evaluation and evaluation methodologies, but mostly on a descriptive level, but there is no universally applicable method of evaluating effectiveness that would use a quantitative apparatus. See the study of Nieminen et al.[10].

From the point of view of the success of drawing subsidies, it is not only possible to consider their quantity, but above all also the (increasingly discussed) effectiveness of their spending for various purposes. So this is another topic for further activity. The question is also whether it is appropriate to recommend increasing the size of subsidies. In view of the discussions concerning a certain degree of harmfulness of subsidy titles and recent events, it would be appropriate, in view of future developments, to consider strengthening local economies and reducing dependence on subsidy titles. Nevertheless, CLLD and the associated search for a suitable model for evaluating its effectiveness need to be supported, especially as bottom-up activities that support local actors' activity, regional authenticity, and specificity.

Acknowledgements

This paper was supported by the International Visegrad Fund under the project ID 22010034.

References

[1] Bernard, J. & Šimon, M. (2017). Vnitřní periferie v Česku: Multidimenzionalita sociálního vyloučení ve venkovských oblastech. *Sociologický časopis/Czech Sociological Review*, 53, 1, 1–26.

[2] Delín, M. & Pospěch, P. (2016). Komodifikace venkova a utváření identity regionu. Sociologický vćasopis/Czech Sociological Review, 52 (2), 209–235.

[3] Grigsby, W., J. (2001). Community vitality: Some conceptual considerations. *Rural Development Paper no. 6*, The Northeast Regional Center for Rural Development.

[4] Hampl, M., & Gardavský, V., & Kühnl, K. (1987). *Regionální struktura a vývoj systému osídlení ČSR*. Praha: Univerzita Karlova v Praze, Přírodovědecká fakulta.

[5] Hampl, M. (2005). *Geografická organizace společnosti v České republice: transformační procesy a jejich obecný kontext*. Praha: Univerzita Karlova v Praze, Přírodovědecká fakulta.

[6] Integrated Regional Operational Programme for Period 2014–2020 (2020, July 29). Retrieved from http://www.dotaceeu.cz/getmedia/a1500e9a-34af-4a8e-b6cb-8e0b119f26d4/Programming-document-of-IROP.pdf.

[7] Klufová, R., & Klicnarová, J. (2019). Local Action Groups in the Czech Republic and their assessment. In *International Conference on Mathematical Methods in Economics 2019* (pp. 470–475). České Budějovice: University of South Bohemia, Faculty of Economics.

[8] Lopolito, A., Nardone, G. & R. Sisto (2011). Towards a comprehensive evaluation of Local Action Groups in LEADER programmes. *New Medit*, 1, 43–49.

[9] Marquardt, D. & Möllers, J. & Buchenrieder, G. (2012). Social Networks and Rural Development: LEADER in Romania. *Sociologia Ruralis*, 52, 398–431.

[10] Nieminen, T. Martellin, T. Koskinen, S. Simpura, J. Alanen, E. Härkänen, T. & Aromaa, A. (2008). Measurement and socio-demograhic variation of social capital in a large population-based survey. *Soc Indic Res*, 85, 405–423.

[11] Osti, G. (2000). LEADER and partnerships: the case of Italy. *Sociologia Ruralis*, 40, 172–180.

[12] Pappalardo, G. Marquardt, D. & Pecorino, B. (2013). Assessing Social Relationship within Local Action Groups: a Worthwhile Tool for Measuring Lag´s Effectiveness. Experiences from two Sicilian Leader Lags. In. B. C. de Gennaro & G. Nardone (Eds.), *Sustainability of the agri-food system: Strategies and Performances.*, Proceedings of the 50th SIDEA Conference, Lecce, Chiostro dei Domenicani, 26–28 September, (pp. 277–290). Prague: Union of Czech Mathematicians and Physicists.

[13] Pavlíková, G. (2007). Program LEADER ČR – nástroj pro rozvoj venkova. Praha: ČZU.

[14] Pechrová, A. & K. Boukalová (2015). Differences among Czech Local Action Groups in using selected principles of LEADER. *Scientia Agriculturae Bohemica*, 46 (1), 41–48.

[15] Pechrová, M. (2015). Vliv dotací z Programu rozvoje venkova na technickou efektivnost příjemců. Praha: Česká zemědělská univerzita v Praze, PEF.

[16] Přehledy projektů a výzev (2020, July 29). Retrieved from https://irop.mmr.cz/cs/statistiky-a-analyzy/prehledy-projektu-a-vyzev.

[17] Vrabková, I. & P. Šaradín (2017). The technical efficiency of Local Action Groups: A Czech republic case study. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 65 (3), 1065–1074.

[18] Změna metodiky ukazatele registrované nezaměstnanosti (2020, July 29). Retrieved from https://www.mpsv.cz/upozorneni-na-zmenu-metodiky.

Selection Strategies in Evolution Algorithms and Biased Selection with Incest Control

František Koblasa[1], Miroslav Vavroušek[2], František Manlig[3]

Abstract. Evolution algorithms are suitable to solve the large field of NP problems. The main difficulty of applying an evolutionary algorithm is to determine an appropriate set of parameter values of evolution operators. One of the most difficult operators to set is the selection of appropriate parents to be reproduced into new solutions. This parameter influences not only the quality of future solutions directly by parent chromosome, but also indirectly by keeping suitable diversity of the whole population.

The main motivation of this article is to discuss the ability to prevent premature convergence of selection procedures to make the base for further adaptive scheme to select parents. The presented research is then focused on suggesting a new mechanism of incest control in selection strategies schemes by considering the relationship between selected parents.

The new incest control mechanism in biased selection is introduced and experimentally compared with well-known selections. The control mechanism showed promising results in both premature convergence and the objective function.

Keywords: evolution algorithm, incest, lineage, selection strategies, scheduling

JEL Classification: C63, C65
AMS Classification: 90C59

1 Introduction

From the beginning of the first Evolution Algorithm (EA) proposal [17] as a tool to optimize complex decision making problems was selection together with defining representation-population size-crossover-mutation and elimination one of the key operators which setting up heavily influences all measurable optimization parameters. Selection as same as elimination is influencing quality on a macroscopic level by selecting individuals who will mate and who survive. The common concern was always so-called premature convergence of population, which cannot be effectively neglected by other than previously mentioned two operators. That is the reason this research is focused on the selection procedure as a way how to improve the performance of selection strategies.

The article is organized as follows. The second chapter is mapping and analysing the main mechanisms of selection operators in EA. The third chapter is introducing incest control mechanism based on lineage analysis, as a way how to improve existing strategies. The fourth chapter is describing the optimized problem and experiment setup leading to make results analysis.

2 Selection strategies in Evolution Algorithms

Selection strategies were analysed not only by the ability to prevent the whole population to converge to one local optimum but also by analysing time complexity [12], selection probabilities [3], takeover time, selection intensity and loss of diversity [4]. The most well-known reviews incorporate theoretical background together with experimental evaluation [2, 19, 25, 26,28]. However, this research is based on one of the most cited theoretical paper of Sivaraj [31] which provides a structured classification of most used selection strategies.

Following chapters are presenting most used selection strategies beginning with simple once as uniform or tournament, through classical roulette wheel fitness proportional based selection, ending with independent

Technical university of Liberec, Department of manufacturing systems and automation, Studentská 2, Liberec 1, Czech Republic,
[1] frantisek.koblasa@tul.cz.
[2] miroslav.vavrousek@tul.cz.
[3] frantisek.manlig1@tul.cz.

scaling selection whose do not take in account quality of individuals. This chapter has a goal to map types of selections to select once to be compared by testing together with the proposed incest control strategy.

2.1 Uniform selections

The first mentioning of randomly selecting parents to create children was from the early days of evolutionary optimization by Bäck [2] referring to earlier research in 1989 [30] by Schwefel. Although this selection is the least sophisticated it is still used for optimization technique experimental comparison as a typical for SGA (Simple Genetic Algorithm) [32]. Uniform types of selections are also popular to develop new operators which are trying to limit selection pressure leading to premature convergence. Hutter [20] introduced two-stage uniform selection where individuals are selected with uniform distribution, but not based on sequence ID (population is still sorted by an objective function), but base on fitness. This research is considering for comparison the two of the most used uniform selections – uniformly random and threshold selection. The uniformly random selection selects every parent randomly with uniform probability [2]. Threshold or Truckian selection first selects a fraction of best individual's t and then are copied deterministically [33] as often as needed until the mating pool size or select randomly among them.

2.2 Tournament selections

Tournament based selections are considered as one of the most popular because of their simple implementation and in most cases acceptable results [12]. There are several strategies, [22], however main principle remains the same. Selected groups of individual solutions are competing usually base on fitness to become parents. The simplest strategy Binary Tournament [1] is to randomly select two individuals and the one with better fitness function succeed to become a parent. Elite tournament selection [23] is selecting first parents of mating pair sequentially from the best solution group to depth defined by the size of the mating pool, the second parent of this pair comes from the classical binary tournament. There are several modifications or similarities to other tournament selection as modified elite selection [18] or biased tournament selection [25] whose are considering also worst fit individuals. Size of the tournament is also considered [25] as well as probability to be selected in Botzman tournament selection [11].

2.3 Proportional selections

Proportional selections are bounding selection probability proportionally to objective or fitness function or number of potential parents (size of the population). A typical example of purely selecting parents base on quality of objective function is classical roulette wheel selection which was an original selection strategy used in the first EA [17]. There are various other selections which main to consider are fitness scaling, biased, Boltzmann [24] or sampling selection.

Roulette wheel

The roulette wheel selections (RW) is the most used among a proportional selection Its simple mechanism defines probability to be selected based on the percentage of individual objective function from the sum of objective functions of all individuals. RW suites to maximization problems, however, for minimisation problem objective function has to be transformed into fitness function to not to put selection pressure on to least fit individuals. There are several approaches as simply reverting sequence of probabilities, or considering the highest and lowest values in the population. This research is using probability taking into account minimal objective function (1) as follows:

$$P(x_i) = \frac{O_{bw}(x_i)}{\sum_{j=1}^{N} O_{bw}(x_j)} \quad \text{where } O_{bw}(x_i) = \frac{f(x_i)}{f_{min}} \tag{1}$$

Fitness scaling selections

Fitness scaling selections transform objective function by scale to fitness function to minimize the effect of uniform randomness in case of large populations or great selection pressure in case of very strong individuals. There are various strategies as relative fitness scaling [15] or sigma scaling [16] which is still popular in various applications [29]. Sigma scaling exists in several variations and it is based on incorporating information about the average objective function and standard deviation of population fitnesses to rescale objective function into fitness (2):

$$O(x_i) = 1 + \frac{f(x_i) - \bar{f}}{2\sigma} \tag{2}$$

Procedure than continues with standard roulette wheel selection using fitness $O(x_i)$.

Biased selections

Biased selections (BI) are used to remove selection pressure from the best individuals. Classical biased selection uses RW in first of pair selection and the second one is selected randomly base on uniform distribution [13]. Biased selection can be also dividing groups in low-fit and best-fit groups where the roulette wheel is applied to an objective function or probability to select the second parent.

Sampling selections

A special group of proportional selections are sampling selections. There is a group which uses division of objective function and population average to calculate kind of decision number as deterministic sampling, Stochastic Remainder or its replacement modifications [10,31].

The most used sampling algorithm according to a number of publications is Stochastic Universal Sampling (US) developed by Baker [5]. US is using a focus of the pointer to select N parents with given F/N distance (where F represents the sum of all objective function values) e.g. minimum spread. Individual selection probability is then based on initial selection and base on defined distance. The population is shuffled randomly and a single random number (pointer) in the range [0, F/N] is generated (see Figure 1).

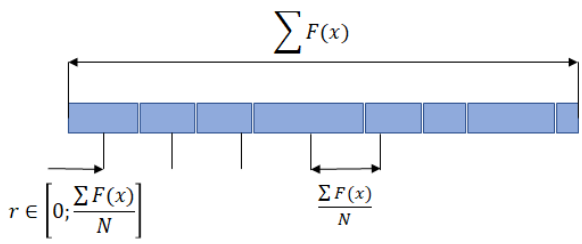


Figure 1 Stochastic Universal Sampling

Selection is popular for its zero bias and used in wide filed of applications beginning with generalized Nets [27], process optimization [6] ending with cluster horticulture [21]

2.4 Ranking selections

Ranking selections are working with sequence and number of individuals rather than with objective function. During the process, there is a calculated fitness base on defined function as Linear (RL), Exponential (RE), Non-linear [10] etc.

Linear scaling (ranking) was one of the first tries to limit premature convergence [14]. This strategy at first sorts individuals by their objective function and assign rank R=N (where N is the size of the population) to the best individual and rank 1 to the worst. The probability to be selected is then linearly assigned to the individuals by (3)

$$P(x_i) = \frac{1}{R}\left(n^- + (n^+ - n^-)\frac{i-1}{R-1}\right) \tag{3}$$

where $k_1 = \frac{n^-}{R}$ is the probability of selecting the worst individual and $k_2 = \frac{n^+}{R}$ the best. However, for minimisation problem probability has to be transformed by assigning $R = 1$ to the best individual and then probability to select an individual (4) is:

$$P(x_i) = \frac{O(x_i)}{\sum_{j=1}^{N} O(x_j)} \quad \text{where } O(x_i) = \frac{1}{R}\left(n^+ + (n^- - n^+)\frac{i-1}{N-1}\right) \tag{4}$$

This scaling uses objective function only to sort individual from the best to worst and probability depends on rank based on population size. The biggest difficulty is then set appropriate rank for the best/worst individual.

Exponential scaling works by the same system as linear only with different distribution along with the population. The base of this method is the exponent $0 < c < 1$. Probability of i-th individual is then (5)

$$P(x_i) = \frac{c^{N-i}}{\sum_{j=1}^{N} c^{N-j}}; i \in \{1, \dots, N\}. \tag{5}$$

3 The incest control mechanism in selection procedures

The newly proposed control mechanism is trying to improve existing selection strategies to prevent mating which will corrupt chromosome by sharing genes in the sequence which is very similar to each other. This problem was first addressed as incest problem by Eshelman [9]. The proposed solution was comparing mating individuals by measuring Hamming distance and when it passes threshold one of parents were replaced. This is still the most used principle to measure and prevent similarity of individuals [7, 34]. However, measuring Hamming distance in the case of large problems and big population sizes can be very time-consuming. This problem is solved by considering so-called lineage [8] which is keeping information about predecessors. Base on the assumption that in most GA, sharing chromosome has genetically strong (root) parent (child is sharing a bigger portion of the chromosome). So-called Lineage selection is then made by grouping solutions based on their root parent (initial solution). The tournament is done between individuals from different lineages and each lineage has than the same probability to be selected.

This research is using both of these approaches with the goal to generalize lineage for different kinds of selections. It is using lineage assumption and is generating lineage (see Fig. 2) in defined deepness R_{ld} (not just an initial parent) and is checking by vector comparison of common predecessors.

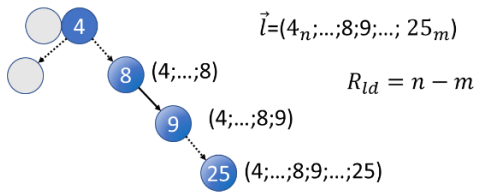


Figure 2 Lineage and relationship deepness

Incest control is in this research applied to Biased selection. During the selection of second parent lineage vectors $\vec{l}$ are compared and when they cross selection tries to randomly find non-relative for a number of defined times I.

4 Experiment and results

Job shop Scheduling as the tested problem is the combinatorial and in the most cases NP-Hard problem, in which multiple jobs n are processed on several machines m. Each job$\{J_i\}_{1 \le i \le j}$ consists of a sequence of tasks O_{jr} with processing time P_{jr}, which must be performed in a given order, and each task must be processed on a specific machine $\{M_r\}_{1 \le r \le m}$. The classical scheduling $n \times m$ problems are usually minimizing the objective function of makespan, e.g. time required to complete all the jobs. Solution together with fitness function is calculated by constructive Giffler and Thompson active schedule generation algorithm. There is used FT10 problem which is considered as one of the hardest because of its strong local extremes with theoretical 3.95 10^{65} size of (semiactive schedules) solution space.

Simple genetic algorithm (SGA) for testing is used with population size μ=100 and 200, random key representation, uniform crossover with crossing probability P_c=85%, λ= 0.5μ new individuals (selections) per generation, no mutation and with simple elitist elimination strategy(μ + λ). Each experiment was repeated 10 times to find the effect of each selection method to premature convergence (generation in which will population converge) and also an objective function of makespan to find the ability to not only keep diverse population but also to find good results.

We tested methods as Uniform ranking (UR), Roulette wheel (RW), Uniform scaling (SC), Sigma scaling (6S), Biased (B1) and its incest modification (IB1 with R_{ld}=6 and I=3), Linear (RL with k_1=0,2) and Exponential ranking (RE with c=0,95), Binary (TB) and Elite Tournament (Te) and finally Threshold (Th with t=0.5 μ). All selections are set as best to fit its original definitions in referenced literature in this paper.

Table 1 shows results from a mentioned experiment where b*f*(x), a*f*(x) and w*f*(x) represents best, the average and the worst result of objective function makespan during 250 repetitions together with best, average and worst (bCvg, aCvg, wCvg) result of generation in which total convergence of population happens or there is no improvement for 100 generations and with *at* representing average optimization timespan of one generation in seconds.

		N=100							N=200						
N	Strategy	b*f*(x)	a*f*(x)	w*f*(x)	bCvg(g)	aCvg(g)	wCvg(g)	at(g)	b*f*(x)	a*f*(x)	w*f*(x)	bCvg(g)	aCvg(g)	wCvg(g)	at(g)
1	UR	959	1010.90	1067	281	106.62	49	0.13	951	992.90	1057	349	150.40	70	0.27
2	RW	970	1013.80	1067	280	109.70	46	0.14	937	992.10	1047	313	146.60	55	0.26
3	6S	967	1012.90	1078	219	112.85	54	0.13	945	992.70	1035	350	153.80	64	0.26
4	US	985	1041.80	1137	237	74.93	12	0.13	971	1027.30	1092	284	89.00	19	0.26
5	B1	958	1012.50	1085	234	108.60	53	0.13	950	991.00	1042	261	146.70	63	0.26
6	RL	961	1025.10	1110	182	66.72	28	0.13	956	1001.80	1052	241	90.60	38	0.26
7	RE	997	1057.40	1141	89	23.51	7	0.13	980	1046.80	1105	104	16.90	7	0.26
8	B1I	951	988.20	1097	409	214.68	79	0.13	937	982.00	1078	593	295.40	148	0.24
9	TB	970	1033.90	1099	162	52.46	17	0.13	959	1006.70	1070	173	73.20	38	0.26
10	TE	972	1042.70	1134	148	37.02	14	0.13	956	1013.40	1063	144	46.80	21	0.25
11	Th	977	1033.10	1098	171	50.78	23	0.13	962	1007.60	1060	238	68.00	27	0.26

Table 1 JSP FT10 problem with 930 makespan optimum

Results shove great difference between each selection methods and can be divided into main three groups. The first one performs well in all terms (green in Table1). This includes surprisingly simplest UR strategy which has good results on average without concerning population size. This group also includes 6S, RW and B1. The best in this group is Biased (B1I) using defined incest control mechanism which outperformed all of the tested selections in both objective function and premature generation convergence. The second group (yellow), which includes RL, TB and Th, shows that groups can't be easily divided by the main principle of selection to group them by results. Those selections can perform acceptably, but with bigger population sizes which cause longer optimization time. The last (orange) group require revision as its results are not acceptable. That can be caused by setting its parameters. However, in some cases it can neglect nature of selection as setting RE with setting *c* to values closer to 1 makes it RL.

Results of objective function from point of ability to reach optimum of 930 can't be compared with usual experiments as its setup were not focused to find general optimum, but the influence of selection methods to the objective function and premature convergence. It would be necessary to implement different elimination method and mutation to prevent convergence and find an optimal solution.

5 Conclusion

Developed incest control mechanism shoves that it can be a significant mechanism to improve selection strategy. It can improve premature convergence and also objective function in a matter of optimization time. The main predicted disadvantage of increased optimization timespan by the mechanism of comparing lineage vectors was not confirmed as timespan of generation calculation has not significantly changed.

Future research will focus on comparing these results on other problems and on implementing designed incest control mechanism in other selections to look for possible synergy. The main focus will be put on implementing the results of selection procedure research to be implemented in adaptive algorithms to solve various kinds of combinatorial problems from operation research.

Acknowledgements

This publication was written at the Technical University of Liberec as part of the project "(21278) – Optimization of manufacturing systems, 3D technologies and automation" with the support of the Specific University Research Grant, as provided by the Ministry of Education, Youth and Sports of the Czech Republic in the year 2019

References

[1] AGRAWAL, R. B., DEB, K., & AGRAWALC, R. (1995). Simulated binary crossover for continuous search space. *Complex systems*, vol. 9, no. 2, pp. 115–148.

[2] BÄCK, T., & HOFFMEISTER, F. (1991). *Extended selection mechanisms in genetic algorithms*. Morgan Kaufmann.

[3] BÄCK, T., FOGEL, D. B., & MICHALEWICZ, Z. (1997). *Handbook of evolutionary computation*. CRC Press.

[4] BÄCK, T., FOGEL, D. B., & MICHALEWICZ, Z. (Eds.). (2018). *Evolutionary computation 1: Basic algorithms and operators*. CRC press.

[5] BAKER, J. E. (1987, July). Reducing bias and inefficiency in the selection algorithm. *In Proceedings of the second international conference on genetic algorithms* (Vol. 206, pp. 14–21).

[6] BALDOVINO, R. G., BAUTISTA, M. G. A. C., AQUINO, A. U., CALILUNG, E. J., SYBINGCO, E., & DADIOS, E. P. (2017). GA optimization of coconut sugar cooking process: A preliminary study using stochastic universal sampling (SUS) technique. *In Proceedings of the 9th International Conference on Computer and Automation Engineering* (pp. 346–349).

[7] BIDAR, M. & MOUHOUB, M. (2019). Solving Weighted Constraint Satisfaction Problems Using a new Self-Adaptive Discrete Firefly Algorithm. In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) (pp. 2198–2205). IEEE.

[8] BURKE, E. K., GUSTAFSON, S., KENDALL, G., & KRASNOGOR, N. (2003). Is increased diversity in genetic programming beneficial? An analysis of lineage selection. In The 2003 Congress on Evolutionary Computation, 2003. CEC'03. (Vol. 2, pp. 1398–1405). IEEE.

[9] ESHELMAN, L. J. (1991). Preventing premature convergence in genetic algorithms by preventing incest. *In Proc. of the 4th. Conf. on GA*, July, 1991 (pp. 115–122).

[10] GEETHA, T. & MUTHUKUMARAN, K. (2013). An observational analysis of genetic operators. International *Journal of Computer Applications*, 63(18), 24–34.

[11] GOLDBERG, D. E. (1990). A note on Boltzmann tournament selection for genetic algorithms and population-oriented simulated annealing. *Complex Systems*, 4(4), 445–460.

[12] GOLDBERG, D. E. & DEB, K. (1991). A comparative analysis of selection schemes used in genetic algorithms. *In Foundations of genetic algorithms* (Vol. 1, pp. 69–93). Elsevier.

[13] GONÇALVES, J. F. & RESENDE, M. G. (2011). Biased random-key genetic algorithms for combinatorial optimization. *Journal of Heuristics*, 17(5), 487–525.

[14] GREFFENSTETTE, J. J. & BAKER, J. E. (1989). How genetic algorithms work: A critical look at implicit parallelism. *In Proceedings of the 3rd International Conference on Genetic algorithms* (pp. 20–27).

[15] GUPTA, S. (2009). Relative fitness scaling for improving efficiency of proportionate selection in genetic algorithms. *In Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference*: Late Breaking Papers (pp. 2741–2744).

[16] HANCOCK, P. J. (1994). An empirical comparison of selection methods in evolutionary algorithms. *In AISB Workshop on Evolutionary Computing* (pp. 80–94). Springer, Berlin, Heidelberg.

[17] HOLLAND, J. H. (1975) *Adaption in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, MIT

[18] HUANG, S. C., JIAU, M. K., & LIN, C. H. (2014). Optimization of the carpool service problem via a fuzzy-controlled genetic algorithm. *IEEE Transactions on Fuzzy Systems*, 23(5), 1698–1712.

[19] HUSSAIN, A. & MUHAMMAD, Y. S. (2019). Trade-off between exploration and exploitation with genetic algorithm using a novel selection operator. *Complex & Intelligent Systems*, 1–14.

[20] HUTTER, M. (2002). Fitness uniform selection to preserve genetic diversity. *In Proceedings of the 2002 Congress on Evolutionary Computation*. CEC'02 (Cat. No. 02TH8600) (Vol. 1, pp. 783–788). IEEE.

[21] IRFIANTI, A. D., WARDOYO, R., HARTATI, S., & SULISTYONINGSIH, E. (2016). Determination of Selection Method in Genetic Algorithm for Land Suitability. In *MATEC Web of Conferences* (Vol. 58, p. 03002). EDP Sciences.

[22] LAVINAS, Y., ARANHA, C., SAKURAI, T., & LADEIRA, M. (2018). Experimental Analysis of the Tournament Size on Genetic Algorithms. *In 2018 IEEE International Conference on Systems, Man, and Cybernetics* (SMC) (pp. 3647–3653). IEEE.

[23] MATOUŠEK, R. (2008). GAHC: improved genetic algorithm. In Nature Inspired Cooperative Strategies for Optimization (NICSO 2007) (pp. 507–520). Springer, Berlin, Heidelberg.

[24] MAZA, M. D. L. & TIDOR, B. (1993, June). An analysis of selection procedures with particular attention paid to proportional and Boltzmann selection. *In Proceedings of the 5th International Conference on Genetic Algorithms* (pp. 124–131).

[25] ORAL, M. & GENÇAL, M. (2018). Harmony Between the Best and the Worst Individuals in Tournament Selection. *Majlesi Journal of Mechatronic Systems*, 6(3). Retrieved from http://journals.iaumaj-lesi.ac.ir/ms/index/index.php/ms/article/view/329.
[26] PANDEY, H. M. (2016). Performance evaluation of selection methods of genetic algorithm and network security concerns. *Physics Procedia*, 78, 13–18.
[27] PENCHEVA, T., ATANASSOV, K., & SHANNON, A. (2009, December). Modelling of a stochastic universal sampling selection operator in genetic algorithms using generalized nets. *In Proceedings of the Tenth International Workshop on Generalized Nets*, Sofia (pp. 1–7).
[28] RASKA, P. & ULRYCH, Z. (2018). Testing different evolution strategy selection strategies. *MM Science Journal*., 2290–2299 DOI: 10.17973/MMSJ.2018_03_2017110. ISSN 18031269.
[29] RAVINDRAN, S., JAMBEK, A. B., MUTHUSAMY, H., & NEOH, S. C. (2015). A novel clinical decision support system using improved adaptive genetic algorithm for the assessment of fetal well-being. *Computational and mathematical methods in medicine*, 2015.
[30] SCHWEFEL, H. P. (1981). *Numerical optimization of computer models*. John Wiley & Sons, Inc.
[31] SIVARAJ, R. & RAVICHANDRAN, T. (2011). A review of selection methods in genetic algorithm. *International journal of engineering science and technology*, 3(5), 3792–3797.
[32] TREJOS-ZELAYA, J., AMAYA-BRICEÑO, L. E., JIMÉNEZ-ROMERO, A., MURILLO-FERNÁNDEZ, A., PIZA-VOLIO, E., & VILLALOBOS-ARIAS, M. (2020). Clustering Binary Data by Application of Combinatorial Optimization Heuristics. *arXiv preprint arXiv:2001.01809*.
[33] WEISE, T. (2009). *Global optimization algorithms-theory and application*. Self-Published Thomas Weise.
[34] ZHU, H., WANG, J., SUN, Y., & WU, X. L. (2020). An Improved Genetic Algorithm for Optimizing EBG Structure with Ultra-Wideband SSN Suppression Performance of Mixed Signal Systems. IEEE Access, 8, 26129–26138.

A Modified Kaldor–Kalecki Model and its Impulse Response Analysis

Jan Kodera[1], Quang Van Tran[2]

Abstract.
The goal of our research in this paper is to analyze the dynamics of a new modification of the Kaldor–Kalecki (KK) model. It is a model based on macroeconomic principles and is represented by a system of delay differential equations. Though it has been studied in the past, the model has often been introduced in a simpler version with single delay. We present a novel specification which embeds the gestation period of investments into the model with multiple delays. We conduct impulse – response analysis which under our framework is to define conditions that diverge system from equilibrium, its restoration dynamics, and time necessary for its return to equilibrium. First, we investigate a KK model with a simple delay. Then, we analyze the dynamics of a KK model with distributed delays for gradual implementation of investments. We find that in both models the dynamics is complex. The trajectories of the models does not converge to an equilibrium point but to a steady state mode of nonlinear oscillations. The simulation results show that spreading the delay leads to a more complexity of model dynamics.

Keywords: Kaldor–Kalecki model, impulse response analysis, differential difference equation, steady state motion region.

JEL Classification: E11, E17
AMS Classification: 37M10

1 Introduction

Deterministic macroeconomic models have not been in the central focus of economists for a long time. One on the reasons was the lack of proper numerical tools in the past. However, new computational achievements have made a huge advance forward recently. This progress allows us to investigate models of this category in a more complex way than before. One type of these models is the Kaldor–Kalecki (KK) model. The model is built on macroeconomic principles and it is represented by a system of differential difference equations. In this paper, unlike previous studies, for example in [7], [2], the KK model is specified with multiple delays. This representation of the model allows us to spread Tinbergen's investment gestation period [10] over several periods of time. In our opinion, this specification can better capture the fact that investments often need more time to be finished. The model is first numerically solved. We also conduct impulse-response analysis which is to define conditions which diverge system from equilibrium under our framework and its restoration dynamics as well as time necessary for its return to equilibrium. To get a full comparison, we begin with a KK model with a simple delay. Then, we analyze the dynamics of a KK model with delays spread over time for gradual installation of investments. Further, in both cases, the impact of parameters of the model on its dynamics and geometric representation of the trajectories is also investigated.

2 Kaldor–Kalecki model

Kaldor [3] proposed in 1940 a macroeconomic business cycle model in which product $Y(t)$ a capital stock $K(t)$ intereact by a system of two differential equations as follows

$$\begin{aligned} \dot{Y}(t) &= \alpha[I(Y(t),K(t)) - S(Y(t))] \\ \dot{K}(t) &= I(Y(t),K(t)) - \delta K(t), \end{aligned} \tag{1}$$

where $\dot{Y}(t), \dot{K}(t)$ are derivatives of product and capital stock with respect to time, $I(.), S(.)$ are the investment and savings functions and α and δ are adjustment parameters, respectively. At first glance, the system of

[1] University of Economics in Prague, Faculty of Finances and Accounting, náměstí W. Churchilla 4, Praha 3, 130 00 – Czech Republic, kodera@vse.cz
[2] University of Economics in Prague, Faculty of Finances and Accounting, náměstí W. Churchilla 4, Praha 3, 130 00 – Czech Republic, tran@vse.cz

two equations constructs a simple two-dimension non-linear dynamic system. However, the dynamics of the system can be complex when investment and savings functions are sufficiently sophisticatedly specified. There have been many experiments with the specification of the Kaldor model. One of them can be found in [6].

Prior to Kaldor, Kalecki [4] furtherly developed Tinbergen's [10] idea of industrial investment cycles based on the ship-building model. According to him, investment decisions are derived from the expected profits in the future. Then, from a decision to invest to putting the investment to operation has to go through a gestation period. Hence, an investment process consists of three phases separated with time delays:

- the decision to invest,
- production of capital goods,
- actual investment – installation of capital goods.

This idea can be embodied in a delay-differential equation

$$\dot{J}(t) = aJ(t) - bJ(t-\theta), \tag{2}$$

where $J(t) = I(t) - U$ is the deviation of investment $I(t)$ from a constant consumption of capital U, θ is the time delay, and a, b are non-negative parameters[1]. The solution of equation (2) will be $J(t)$ as a function of t, which may display the endogenous cyclical fluctuations in an economic system. As the investment function is a function of product and capital stock, incorporating the gestation period principle into the Kaldor model in (1) and with adequate rearrangement we get a new model called Kaldor–Kalecki's model.

$$\begin{aligned} \dot{Y}(t) &= \alpha[I(Y(t), K(t)) - S(Y(t))] \\ \dot{K}(t) &= I(Y(t-\theta), K(t-\theta)) - \delta K(t). \end{aligned} \tag{3}$$

System (3) is a system of two difference differential equations which may potentially generate much richer dynamics than system (1).

3 An innovative modification of the original Kaldor–Kalecki model

In system (3) investment and savings functions are once again not specified. Hence, they become the subject of many experiments. For example, Krawiec and Szydlowski ([7], [8]) have proposed the following specification conditions: $I(Y,K) = I_1(Y) + I_2(K)$, $\frac{\partial I_1}{\partial Y} > 0$, $I_2 = \beta K, \beta < 0$. They showed that the system under these conditions may generate either the occurrence of a bifurcation or a limit cycle depending on the values of parameters of the model. Hu et al. [2] specified system (3) as follows

$$\begin{aligned} \dot{Y}(t) &= \tanh(Y(t)/2) - 0.5625Y(t) - 0.8K(t) \\ \dot{K}(t) &= \tanh(Y(t-\theta_1)/2) - 0.8K(t-\theta_2) - 0.9K(t), \end{aligned} \tag{4}$$

where tanh is the hyperbolic tangent function. With this specification, the system can produce stable and unstable dynamics depending on the values of the delays. Other specifications can be found in [11], [12].

In our contribution, we introduce two innovations into the specification of model (3). First, inspired by the work of Kydland and Prescott [9], we assume that multiple successive periods of time are needed to realize an investment project. Let q be the number of time periods needed to complete an investment project, let I_{jt} be the investment value of a project which needs j periods to be completed, for $j = 1, ..., q-1$. It implies that a new investment project starting at time t is I_{qt}. Then, the dynamics of capital accumulation will be

$$\dot{K}(t) = I_{1t} - \delta K(t) \tag{5}$$

$$I_{j,t+1} = I_{j+1,t}, \tag{6}$$

where δ is the depreciation rate. Let γ_j is the fraction of an investment project realized in the j-th period from the last one, then the total investment implemented in period t is

$$I_t = \sum_{j=1}^{q} \gamma_j I_{jt}, \tag{7}$$

The second innovation is the one which has been implemented in our previous work [5]. In system (3) the investment function is specified in a particular form. We suppose that $i(Y,K)$ is an investment-product ratio

[1] For more information, see the original Kalecki's article [4].

which is often called the propensity to invest. It is an increasing function of productivity of capital, hence $i(Y,K) = i(Y/K)$. In our model, we further assume the logarithmic form for both Y and K, therefore the original function $i(Y,K)$ can be also written as

$$i(Y/K) = i(e^{y-k}) = i(y-k), \tag{8}$$

where $y = \log Y, k = \log K$. We propose the functional form for the propensity to invest $i(t)$ as a logistic function, hence it can be specified as follows:

$$i(y(t)-k(t)) = \frac{x_1}{x_2 + x_3 e^{-a(y(t)-k(t))}}, \tag{9}$$

where x_1, x_2, x_3, a are real parameters. Then the original investment function can be rewritten as

$$I(Y(t),K(t)) = i(y(t)-k(t))Y(t) = \frac{x_1}{x_2 + x_3 e^{-a(y(t)-k(t))}} Y(t). \tag{10}$$

Similarly, we define the real savings function as a product of the propensity to save and the production, where the propensity to save is increasing in logarithm of production. This assumption leads to fact that the propensity does not grow as fast as production. Formally, the propensity to save function can be expressed as

$$s(\log Y(t)) = s_0 + s_1 \log Y(t) = s_0 + s_1 y(t), \tag{11}$$

where s_0, s_1 are real parameters. Then, the original savings function is

$$S(Y(t)) = (s_0 + s_1 y(t))Y(t). \tag{12}$$

Implementing these two innovations into system (3), we get this innovative specification for Kaldor–Kalecki model

$$\begin{aligned} \dot{y}(t) &= \alpha \left[\sum_{j=0}^{q} \gamma_j \frac{x_1}{x_2 + x_3 e^{-a(y(t-\theta_j)-k(t-\theta_j))}} - (s_0 + s_1 y(t)) \right] \\ \dot{k}(t) &= \frac{x_1}{x_2 + x_3 e^{-a(y(t-\theta_0)-k(t-\theta_0))}} e^{y(t)-k(t)} - \delta, \end{aligned} \tag{13}$$

where θ_j is the number of time periods to completion of a investment project. Non-linear dynamic system (13) of difference differential equations[2] (DDE) has nine parameters $\alpha, x_1, x_2, x_3, a, s_0, s_1, \delta, q$ and $q+1$ weight parameters $\gamma_j, j = 0, 1, ..., q$. They will be defined and the system will be solved in the next section.

4 Solving the model and impulse response analysis

System of difference differential equations (13) is a DDE system with constant delays. There is no way to find the solution of the system analytically and the only way to deal with this problem is the numerical solution. Numerically solving DDEs is similar to solving ordinary differential equations (ODE), but more challenging. A brief description of a method is provided here to understand how we proceed. A DDE in the following form

$$\dot{x}(t) = \mathrm{f}(t, x(t), x(t-\theta)), x(t) = u \text{ for } t \leq 0, u \in \mathbb{R} \tag{14}$$

can be transformed to this problem

$$\dot{x}(t) = \mathrm{f}(t, x(t), \varphi(t)), \varphi(t) = x(t-\theta), x(t) = u \text{ for } t \leq 0, u \in \mathbb{R}. \tag{15}$$

One can look at (15) as an ODE equation and proceed analogically. One of the often used methods for this problem is the well-known Runge–Kutta method. A solution of (15) obtained with s-th order Runge–Kutta method (RKS) can be expressed as follows

$$x(t_{j+1}) = x(t_j) + h \sum_{i=1}^{s} c_i k_i, \tag{16}$$

where

$$k_i = \mathrm{f}(t_j + a_i h, h \sum_{q=1}^{i-1} b_{iq} k_q, \varphi(t_j + a_i h)), \tag{17}$$

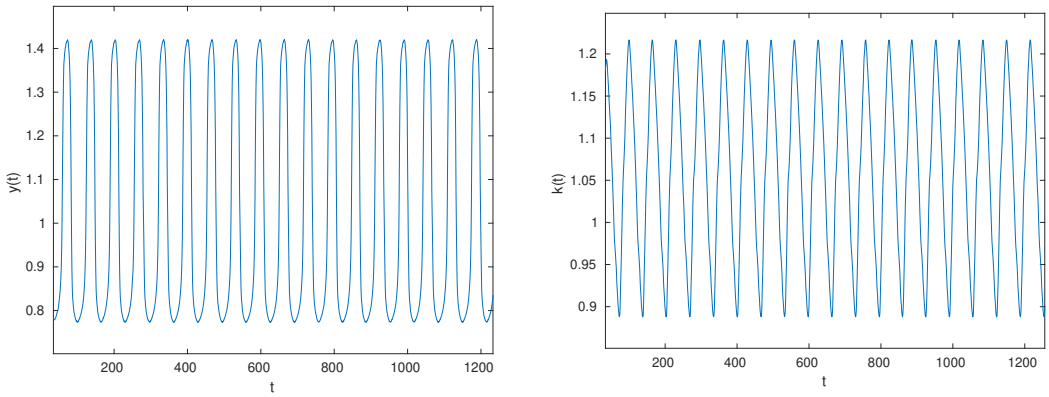


Figure 1 Time evolution of production and capital in the model with a simple delay

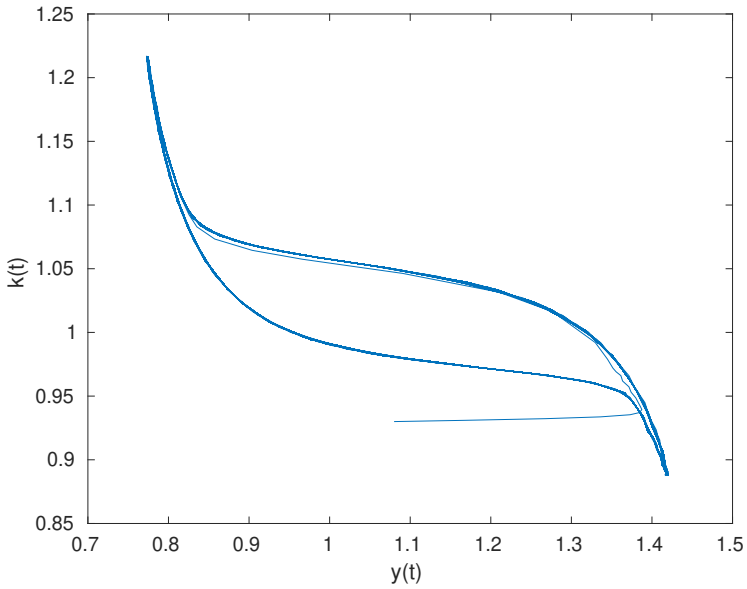


Figure 2 Phase portrait of Kaldor–Kalecki model with simple delay

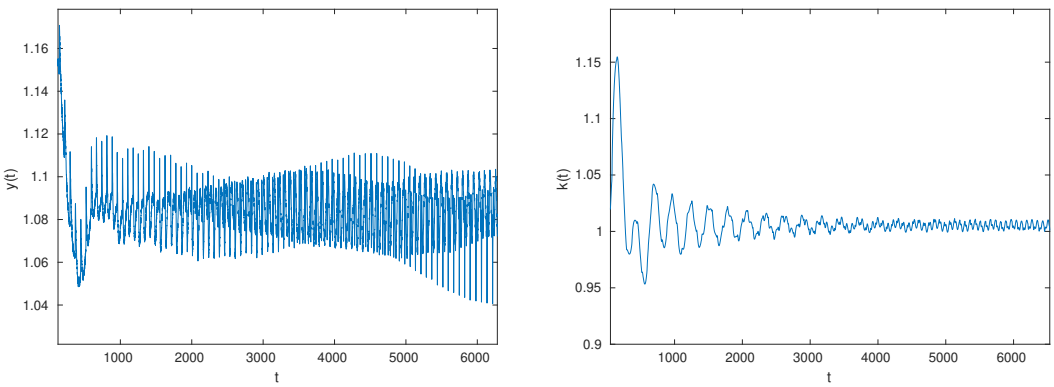


Figure 3 Time evolution of production and capital in the model with distributed delay

h is a fixed step length, $t_{j+1} = t_j + h$, c_i, a_i are the so called weights and nodes, and b_{iq} is an element of Runge–Kutta matrix. As values of $\varphi(t_j + a_i h)$ are often not known in advance, they are derived by interpolation from already known values which is the reason of the complexity of solving DDEs and it is also the source of errors.

For more information on solution of DDEs, see [1].

We calibrate model (13) with these values for its parameters: $\alpha = 120, x_1 = 0.1, x_2 = 1, x_3 = 1.8, a = 20$, $s_0 = -0.2, s_1 = 0.27, \delta = 0.1, q = 4$. This implies the values of weights are $\gamma_0 = \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0.20$. We use MATLAB to solve system (13) with this vector of values for parameters in the model. The computation is highly extensive and often time consuming if higher precision is required. First, we compute the steady state value of the system. Then, we solve them to find series $y(t)$ and $k(t)$ In order to evaluate the role of delays, we also solve system (13) with one delay. In this case $q = 1$ and $\gamma = 1/2$. The solution of $y(t)$ and $k(t)$ is displayed in Figure 1 and their phase portrait is shown in Figure 2. For the system with four delays the results are shown in Figures 3 and 4, respectively.

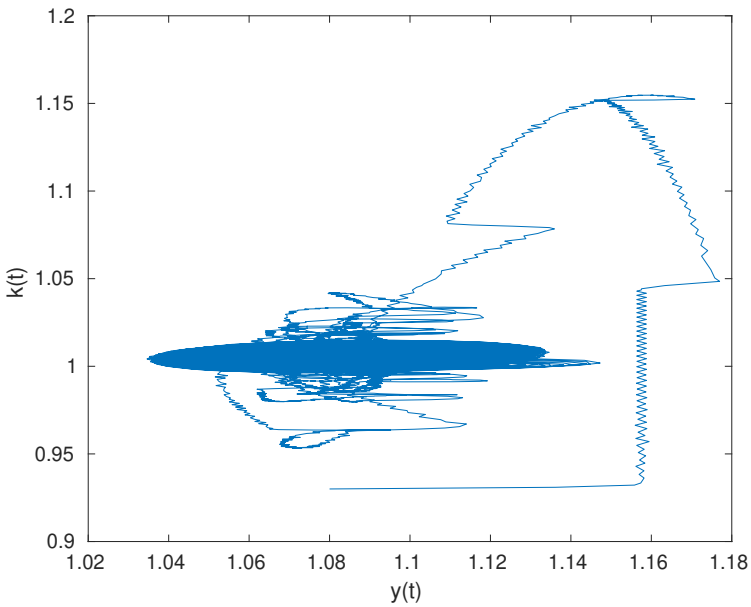


Figure 4 Phase portrait of Kaldor–Kalecki model with distributed delay

In Figure 1 one can observe that the modified Kaldor–Kalecki model with simple delay exhibits a singular periodical pattern of production and capital. The geometric representation of the trajectories of the Kaldor–Kalecki model in a plane (phase portrait) is shown in Figure 2. It indicates the existence of a limit cycle. The shape of the limit cycle is a double nonidentical rotated S-shape curve implying the opposite nonlinear dynamics of each variable inside the system. However, the specific dynamics in Figures 1 and 2 should be seen as a numerical example without any exact economic conclusions because this result is hardly economically interpretable for a relatively large amplitude of oscillations. In Figure 3 the dynamics of production and capital in the modified Kaldor–Kalecki model with distributed delays is shown, the one for production on the left panel, the one for capital on the right panel. it seems that in this mode the production displays a multiple periodical behavior while the capital stock exhibits a dampen oscillation toward a stable band. And a small fluctuations indefinitely continue inside this band. In Figure 4 the phase portrait of the two variables in the modified Kaldor–Kalecki model with distributed delays is shown. First, one can observe in the figure a complicated dynamics toward a shaded area, which can be called as steady state region. Then, there is a region, narrow in direction of capital and wider in direction of production, in which the two variables regularly oscillate. As both models are calibrated with the same values for parameters (wherever possible), the difference in the dynamics of the two models is distinctive. This dissimilarity originates from the spreading out investment installation over many consecutive periods of time, which we showed during the construction of the model. The behavior of the modified Kaldor–Kalecki model with simple delay is shown in figures 1 and 2. Unlike the relatively simple and unrealistic dynamics of the modified model with a simple delay, the nature of oscillations, especially multiperiodicity in the evolution of production and capital of the model with multiple delay, better corresponds to economic experience. In our opinion, the multiple-delay model is considered as a usable analytic tool for economic analysis. It also can be used as a starting block for building more comprehensive models of this type.

[2] Sometimes they are also called as delay differential equations and the two terms are interchangeable.

We also perform an analysis of impact of model parameter values on the dynamics of variables in the model. As in dynamic deterministic systems the fluctuations of variables are endogenous, changing parameter values of the model may cause changes in the dynamics, the investigation is called impulse-response analysis as the term is often used in DSGE modeling. The results of our analysis show that values of parameters of the model can actually distort the dynamic of variables as well as the limit cycles if they exist.

5 Conclusion

We have presented a new version of the classical Kaldor–Kalecki model. In our specification, unlike in previous versions, the implementation of an investment project is spread out over several periods before completion. Then the model has become a system of differential difference equations with multiple constant delays. First, we calibrate the model with a suitable set of values for its parameters and then we solve the system using higher order Runge-Kutta method for DDE. After obtaining the solution for a principle version, we conduct impulse – response analysis which under our framework is to define conditions which diverge system from equilibrium and its restoration dynamics as well as time necessary for its return to equilibrium. The dynamics of the version of KK model with multiple delays is compared with those of the version with of the KK model with a simple delay. We find that in both models the dynamics is complex. However, with multiple delays, the dynamics is much richer. Further, in this model, the equilibrium means a steady-state motion region with non-linear periodic oscillations. We also detect the impact of parameters on the dynamics in both versions. By definition, the model established and analyzed in this paper differs from models of main DSGE stream. However, in our opinion, it can provide more convenient dynamics which fits better real evolution of actual economy. The problem of DSGE models is that they usually have simple dynamics no matter they are in a linearized form or in a non-linear one. Their diminishing oscillations converge to their equilibrium points. The model presented here has different and more complex dynamics. Though it has steady state point, its dynamics does not converge to one-point steady state and it rather persists in periodic or non-periodic oscillations around a steady state region.

Acknowledgements

This research was supported by the University of Economics, Prague under the framework of Institutional Research Support no. IP 100040.

References

[1] Bellen, A., Zennaro, M. (2003). *Numerical Methods for Delay Differential Equations*. Oxford: Oxford University Press.

[2] Hu, W., Zhao, H., Dong, T. (2018) Dynamic Analysis for a Kaldor–Kalecki Model of Business Cycle with Time Delay and Diffusion Effect. *Complexity* 2018.

[3] Kaldor, N. (1940). A Model of the Trade Cycle. *Economic Journal* 50 (1940), 78–92.

[4] Kalecki, M. (1935). A Macro–Economic Theory of Business Cycles. *Econometrica* 3 (1935), 327–344.

[5] Kodera, J., Radová, J., Tran, Q. V. (2012). A modification of Kaldor–Kalecki model and its analysis. In: Ramík, J., Stavárek, D. (editors): Proceedings of 30th International Conference on Mathematical Methods in Economics 2012, Silesian University in Opava, School of Business Administration in Karviná, 420–425.

[6] Kodera, J., Vošvrda, M. (2007). Production Capital Stock, and Price Level Dynamic in the Light of Kaldorian Model. *Acta Oeconomica Pragensia* 15 (2007), 79–87.

[7] Krawiec, A., Szydlowski, M. (1999). The Kaldor–Kalecki business cycle model. *Annals of Operations Research* 89 (1999), 89–100.

[8] Krawiec, A., Szydlowski, M. (2001). The Kaldor–Kalecki Model of Business Cycle as a Two-Dimensional Dynamical System. *Journal of Nonlinear Mathematical Physics* 8 (2001), 266–271.

[9] Kydland, F. E., Prescott, E. C. (1982). Time to Build and Aggregate Fluctuations. *Econometrica* 50 (1982), 1345–1370.

[10] Tinbergen, J. (1959). Ein Schiffsbauzyklus? Weltwirtschaftliches Archiv. Reprinted in Klaasen L.H. et al. Selected Papers. Amsterdam.

[11] Wang, L., Wu, X. (2009). Bifurcation analysis of a Kaldor–Kalecki model of business-cycle with time delay. *Journal of Qualitative Theory of Differential Equations*, 1(27), 1–20.

[12] Wu, X. P. (2011). Codimension-2 bifurcations of the Kaldor model of business cycle. *Chaos, Solitons & Fractals* 44 (2011), 28–42.

Time to Build and Aggregate Fluctuations with Sticky Prices and Wages

Jan Kodera[1], Quang Van Tran[2]

Abstract. Kydland and Prescott (1982) in their seminal work introduced the notion gestation time, which is the time needed to introduce new capital into production process, into a dynamic stochastic general equilibrium (DSGE) model. This notion was first expressed in 1930s. We implement this condition into a DSGE model with sticky prices and wages, which is more common in DSGE models today. In this contribution, first we derive a simplified new-keyesian model with sticky prices and wages. The, using a set of sensible values for the parameters of the model, we performed the calibration. The calibration results show that taking time to build the productive capital into account, the dynamics of variables display interesting features. The model can be easily estimated from data by Bayesian technique or extended into larger size to obtain more variability in model variables whose smoother dynamics often is a deficiency of DSGE models.

Keywords: DSGE model, sticky prices and wages, gestation period, impulse response analysis

JEL Classification: E11, E17
AMS Classification: 37M10

1 Introduction

From experience observed in ship building industry, Tinbergen [8] assumed the need of gestation period from an investment decision to its full operationality. This idea was later further formalized by Kalecki [4]. After that it was long forgotten until Kydland and Prescott [5] revived it again. In their famous work, this idea was renamed as time to build and it was incorporated in the first real business cycle model to generate interesting fluctuations in variable dynamics. We renew this approach and introduce the gestation period or time to build into a more appropriate environment of today's macroeconomic modeling: a new keynesian model with sticky prices and wages. For this purpose, we derive a simplified version of Smets and Wouters' model ([7],[6]) for a closed economy with investments needing time to build. The model can be easily expanded into its original shape for many additional studies. After its derivation, the model will be calibrated and the statistical properties of its endogenous variables will be investigated as well as the impulse response analysis will be performed with respect to exogenous representative shock. To evaluate the impact of time to build, we compare the results with those of a similar model without time for investment installation.

2 A sticky price and waged model with time to build productive capital

In this section a sketch of derivation of the model is provided due to the limited space of a conference paper. A more detailed derivation procedure can be found in [9].

2.1 Households and utility maximization problem

There is a continuum of households in the economy and a typical household indexed by $j \in [0,1]$ tries to maximize

$$\mathbb{E}\sum_{0}^{\infty}\beta^{t}\mathrm{U}(C_t(j), L_t(j)), \tag{1}$$

subject to its budget constraint in the following form

$$P_tC_t(j) + P_tI_t(j) + Q_tB_t(j) = B_{t-1}(j) + R_tK_t(j) + W_tL_t(j) + \Pi_t(j), \tag{2}$$

[1] University of Economics in Prague, Faculty of Finances and Accounting, náměstí W. Churchilla 4, Praha 3, 130 00 – Czech Republic, kodera@vse.cz
[2] University of Economics in Prague, Faculty of Finances and Accounting, náměstí W. Churchilla 4, Praha 3, 130 00 – Czech Republic, tranvqua@fjfi.cvut.cz

where $\Pi_t(j)$ is the profits resulting from firms ownership of the household, other symbols should be well-known notations for those familiar with DSGE models. As the model embraces capital stock whose quantity depends on investment and capital consumption, we introduce investment as it was in the seminal paper of Kydland and Prescott [5]. They embody Tinbergen's [8] idea of gestation time into investment process. When investment decisions have been made, it takes time to build the productive capacity. Generally, it requires multiple periods of time to complete an investment decision. Let $S_{qt}(j)$ be an investment decision of the household with q periods from completion, $q = 1, 2, ..., Q-1$ where Q is the number of periods needed to finish new productive capital. Then the dynamics of capital in the economy can be described as

$$\begin{aligned} K_{t+1}(j) &= (1 - \delta)K_t(j) + S_{1t}(j) && (3) \\ S_{q,t+1}(j) &= S_{q+1,t}(j), q = 1, 2, ..., Q-1, && (4) \end{aligned}$$

where δ is the depreciation rate. Hence, $S_{Qt}(j)$ is a decision variable for period t. Let φ_q be the fraction of an investment decision installed in the q-th period from the final one. Then the total volume of investment installed in one period is

$$I_t(j) = \sum_{q=1}^{Q} \varphi_q S_{qt}(j). \tag{5}$$

Let us assume that the representative household has the preferences which can be expressed by this standard utility function

$$\mathrm{U}(C_t, L_t) = \frac{C_t^{1-\sigma}}{1-\sigma} - \frac{L_t^{1+\phi}}{1+\phi}. \tag{6}$$

Problem (1) with respect to constraints (2), (3), (4) and (5) with utility function (6) can be solved through the Lagrange multiplier method. After constructing the adequate Lagrangian and taking the derivative with respect to decision variables, the resulting first order conditions for $Q = 4$ are

$$\begin{aligned} \lambda_t &= C_t^{-\sigma}/P_t \\ \lambda_t &= \beta(1 + ir_t)\mathbb{E}\lambda_{t+1} \\ \lambda_{1t} &= \beta\mathbb{E}\lambda_{t+1} q * R_{t+1} + \beta(1-\delta)\mathbb{E}\lambda_{1,t+1} \\ \lambda_{1t} &= P_t\varphi_1\lambda_t + \lambda_{4,t-1}/\beta \\ \lambda_{4t} &= P_t\varphi_2\lambda_t + \lambda_{3,t-1}/\beta \\ \lambda_{3t} &= P_t\varphi_3\lambda_t + \lambda_{2,t-1}/\beta \\ \lambda_{2t} &= P_t\varphi_4\lambda_t, \end{aligned} \tag{7}$$

where $\lambda_t, \lambda_{1t}, \lambda_{2t}, \lambda_{3t}$ and λ_{4t} are Lagrangian multipliers corresponding to each constraint of the problem. As all households face the same problem, index j is left out without a loss of accuracy.

2.2 Household optimal wage setting problem

There is a continuum of households and each of them has monopoly on its type of labor $L_t(j)$. The aggregate demand of all types of labor L_t in the economy is

$$L_t = \left[\int_0^1 L_t(j)^{1-\frac{1}{\epsilon_w}} \mathrm{d}j\right]^{\frac{\epsilon_w}{\epsilon_w - 1}}, \tag{8}$$

where index $j \in [0, 1]$ denotes the type of labor and ϵ_w is the elasticity of substitution among labor types. If the labor of type j is priced at time t as $W_t(j)$, then the optimal demand for this labor type in the economy is

$$L_t(j) = \left(\frac{W_t(j)}{W_t}\right)^{-\epsilon_w} L_t \tag{9}$$

and the aggregate wage index W_t in the economy is

$$W_t = \left[\int_0^1 W_t(j)^{1-\epsilon_w} \mathrm{d}j\right]^{\frac{1}{1-\epsilon_w}}. \tag{10}$$

The wage stickiness is introduced into the model through the well-known Calvo-pricing-mechanism-like mode [2]. At each period, a fraction $1 - \theta_w$ of households in the economy reoptimize their wages by setting

them at new level W_t^* and the remaining fraction θ_w of households leave their wages unchanged at level W_{t-1} and θ_w hence expresses the stickiness of labor market. According to (9) the demand for this type of labor L_t^* will be

$$L_t^* = \left(\frac{W_t^*}{W_t}\right)^{-\epsilon_w} L_t. \tag{11}$$

As a monopolist of its type of labor, taking the demand for its labor supply, it has to solve the utility maximization problem with respect to labor supply by optimally setting the new wage claim W_t^*. The optimal solution of this problem is

$$W_t^{*\phi\epsilon_w+1} = \frac{\epsilon_w}{\epsilon_w - 1}\frac{\Gamma_{1,t}}{\Gamma_{2,t}}, \tag{12}$$

where

$$\Gamma_{1,t} = W_t^{(1+\phi)\epsilon_w} L_t^{1+\phi} + \beta\theta_w \mathbb{E}\Gamma_{1,t+1} \tag{13}$$

$$\Gamma_{2,t} = \frac{C_t^{-\sigma}}{P_t} W_t^{\epsilon_w} L_t + \beta\theta_w \mathbb{E}\Gamma_{2,t+1}, \tag{14}$$

where all symbols retain their standard economic meaning in DSGE modeling. Again, as all households in the fraction set the same wage, index j is hence redundant. The adequate aggregate wage index in the economy will be

$$W_t = \left[(1-\theta_w)W_t^{*1-\epsilon_w} + \theta_w W_{t-1}^{1-\epsilon_w}\right]^{\frac{1}{1-\epsilon_w}}. \tag{15}$$

Equations (12), (13), (14), and (15) implicitly construct the so called wage Phillips curve. Its explicit form can be obtained by log-linearization and adequate rearrangement.

2.3 Profits maximization and price setting problem of firms

There is a continuum of firms in the economy producing differentiated goods with a common technology expressed by the Cobb-Douglas production function. Each of them is indexed by $i \in [0, 1]$, hence

$$Y_t(i) = A_t K_t^{\alpha}(i) L_t^{1-\alpha}(i), \tag{16}$$

where all symbols are already known. To reach the maximum level of profits, each firm first tries to minimize the production costs embodied in the following costs function

$$\min_{K_t(i), L_t(i)} W_t L_t(i) + R_t K_t(i). \tag{17}$$

Solving problem (17) with respect to the production level in (16) at price level P_t gives us these first order conditions

$$R_t = \alpha MC_t P_t A_t K_t^{\alpha-1} L_t^{1-\alpha} \tag{18}$$

$$W_t = (1-\alpha) MC_t P_t A_t K_t^{\alpha} L_t^{-\alpha}. \tag{19}$$

As all firms solve the same problem, index i can be dropped. From conditions (18) and (19) we can derive the marginal costs MC_t as

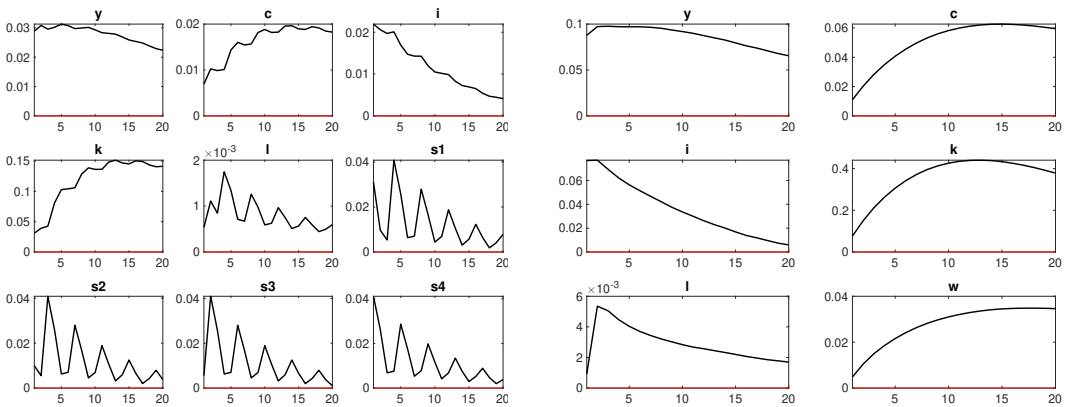
$$MC_t = \frac{1}{A_t}\left(\frac{R_t}{\alpha}\right)^{\alpha}\left(\frac{W_t}{1-\alpha}\right)^{1-\alpha}. \tag{20}$$

The marginal costs are also identical for all firms. The second task a firm in the economy has to deal with is the optimal price setting problem. As firms have monopoly on their differentiated products, they can set the price at will. The price stickiness is implemented through the already mentioned Calvo pricing mechanism. At a given period, each firm can independently reset its price with probability $1-\theta$, which also implies that at each period a fraction $1-\theta$ of firms in the economy reoptimize their prices by setting them at P_t^*. The remaining fraction θ of firms keep their prices unchanged at P_{t-1}. As such, θ is interpreted as price stickiness. At price P_t^* the demand for production with this price at any time after the change is

$$Y_{t+k}^* = \left(\frac{P_t^*}{P_{t+k}}\right)^{-\epsilon} Y_{t+k}. \tag{21}$$

The profits function of these firms Π are

$$\Pi = \max_{P_t^*} \mathbb{E}\sum_{k=0}^{\infty} \theta_p^k Q_{t,t+k}(P_t^* - MC_{t+k}) Y_{t+k}^*. \tag{22}$$



with time to build

without time to build

Figure 1 Impulse response analysis results

Solving problem (22) with respect to P_t^* under condition (21) gives us

$$P_t^* = \frac{\epsilon}{\epsilon - 1} \frac{\Phi_{1,t}}{\Phi_{2,t}}, \tag{23}$$

where

$$\Phi_{1,t} = MC_t Y_t + \theta_p \mathbb{E}\Phi_{1,t+1} \tag{24}$$

$$\Phi_{2,t} = Y_t + \theta_p \mathbb{E}\Phi_{2,t+1}. \tag{25}$$

And the resulting aggregate price level in the economy is

$$P_t = \left[(1-\theta_p)P_t^{*1-\epsilon} + \theta_p P_{t-1}^{1-\epsilon}\right]^{\frac{1}{1-\epsilon}} \tag{26}$$

and the aggregate profit level of all firms in the economy is

$$\Pi_t = (P_t - MC_t)Y_t. \tag{27}$$

Again, equations (23), (24), (25), and (26) implicitly construct the so called price Phillips curve. Its explicit form can be obtained by log-linearization and adequate rearrangement. It is obvious that this condition must hold in equilibrium in the goods market

$$Y_t = C_t + I_t. \tag{28}$$

The whole system of equations of the model with one exogenous shock of productivity can be found inside the dynare code in the next section.

3 Calibration and impulse response analysis

We calibrate the model derived in the previous section with a set of parameters described in the first part of the dynare code. The values of parameters are those often used for a DSGE model. The calibration then is performed in Dynare [1] and the code in Dynare for calibration and impulse-response analysis in shown below for those with interest of further extension. For a comparison purpose, the same task is carried out for a model without time to build. The moments of simulated variables of both models are calculated and the are displayed in Tables 1 and 2. The results of impulse response analysis with respect to technology shock are shown in Figure 1.

```
// Variables and Parameters Declaration
var y c i a k l p ps Phi1 Phi2 w ws Ga1 Ga2 r ir mc s1 s2 s3 s4 Pi la la1 la2 la3 la4;
varexo e;
parameters alpha beta sigma phi delta thetap epsp thetaw epsw rho Phipi Phiw
```

VARIABLE	MEAN	STD. DEV.	VARIANCE	SKEWNESS	KURTOSIS
y	2.634414	0.086367	0.007459	−0.040531	−0.903653
c	1.963509	0.063466	0.004028	0.095316	−1.587227
i	0.670904	0.053079	0.002817	0.263299	−0.056627
k	13.393548	0.379899	0.144323	−0.200812	−0.880443
l	0.922819	0.003522	0.000012	0.050038	−0.307364
w	1.419715	0.143959	0.020724	0.150481	−1.712851

Table 1 Moments of simulated variables, without case

```
Phiy phi1 phi2 phi3 phi4 q; alpha = 0.4; beta = 0.99; sigma = 0.6;
    delta = 0.05; phi = 3; rho = 0.95; thetap = 1/3; thetaw = 0.25; epsp = 5; epsw = 6;
    Phiw = 0.5; Phipi = 1.5; Phiy = 0.5; phi1 = 1/4; phi2 = 1/4; phi3 = 1/4; phi4 = 1/4;
    q = phi1 + phi2/beta + phi3/beta^2 + phi4/beta^3;
model;
    la = c^(-sigma)/p; la = beta*la(1)*(1 + ir);
    ws^(phi*epsw + 1) = epsw/(epsw - 1)*Ga1/Ga2; Ga2 = la*w^epsw*l + beta*thetaw*Ga2(1);
    Ga1 = w^((1+phi)*epsw)*l^(1+phi) + beta*thetaw*Ga1(1);
    w^(1-epsw) = (1 - thetaw)*ws^(1-epsw) + thetaw*w(-1)^(1-epsw);
    la1 = beta*la(1)*q*r(1) + beta*la1(1)*(1 - delta);
    mc = (q*r/alpha)^alpha*(w/(1-alpha))^(1-alpha)/a;
    y = a*k(-1)^alpha*l^(1-alpha); Phi2 = y + thetap*Phi2(1);
    w = p*(1 - alpha)*mc*y/l; i = phi1*s1 + phi2*s2 + phi3*s3 + phi4*s4;
    k = (1 - delta)*k(-1) + s1; s2 = s1(1); s3 = s2(1); s4 = s3(1);
    ps = epsp/(epsp - 1)*Phi1/Phi2; Phi1 = mc*y + thetap*Phi1(1);
    p^(1-epsp) = (1 - thetap)*ps^(1-epsp) + thetap*p(-1)^(1-epsp);
    Pi = (p - mc)*y; y = c + i; a = a(-1)^rho*exp(e);
    la1 = p*phi1*la + la4(-1)/beta; la4 = p*phi2*la + la3(-1)/beta;
    la3 = p*phi3*la + la2(-1)/beta; la2 = p*phi4*la;
    ir = 1/beta - 1 + Phipi*log(p/p(-1)) + Phiw*log(w/w(-1)) + Phiy*log(y/y(-1));
end;
steady_state_model;
    a = 1; p = 1; ps = 1; mc = (epsp - 1)/epsp; ir = 1/beta - 1; r = 1/beta - 1 + delta;
    w = (1 - alpha)*(mc*(alpha/(q*r))^alpha)^(1/(1-alpha)); ws = w;
    pom = (epsw - 1)/epsw; pom1 = q*r/mc/alpha;
    l = (w*pom*pom1^(-sigma/(alpha-1))*(pom1-delta)^(-sigma))^(1/(sigma+phi));
    k = pom1^(1/(alpha-1))*l; i = delta*k; Pi = (p - mc)*y;
    y = k^alpha*l^(1-alpha); c = y - i; s1 = i; s2 = i; s3 = i; s4 = i;
    Ga1 = 1/(1 - beta*thetaw)*w^(epsw*(1+phi))*l^(1+phi);
    Ga2 = 1/(1 - beta*thetaw)*c^(-sigma)*w^epsw*l;
    Phi1 = 1/(1 - thetap)*mc*y; Phi2 = 1/(1 - thetap)*y;
    la = c^(-sigma); la2 = phi4*la; la3 = phi3*la + la2/beta;
    la1 = beta*q*r*la/(1 - beta*(1 - delta)); la4 = phi2*la + la3/beta;
end;
// Diagnostics
    steady(maxit=1000); resid(1); check;
// Shock specification
    shocks; var e; stderr 0.03; end;
// Simulation
stoch_simul(irf=20,periods=200) y c i k l w s1 s2 s3; %s4;
```

In terms of moments of simulated endogenous variables, their computed values are quite similar which is expected. However, the first and the second moments of variables from the model without time to build are slightly higher than those of the model with time to build, except the case of capital. This means that the spread out of implementation of investment projects has a weak smoothing effect due to lower values of their second moment. Regarding higher moment, there is no substantial difference between the two groups which is consistent with the lack of reasoning for any existence of possible differences. As far as the results

of impulse response analysis are concerned, there is a visually distinctive deviation of response of variables of both models on productivity shock. While without delay in investment implementation the response of all variables is relatively smooth, the response in the case of gradual investment installation is somewhat saw-toothed reflecting the fact that previous investment decisions have to be fulfilled. However, the saw-toothed response seems to die out over the course of time.

VARIABLE	MEAN	STD. DEV.	VARIANCE	SKEWNESS	KURTOSIS
y	2.628587	0.085608	0.007329	−0.010141	−1.106601
c	1.963083	0.061604	0.003795	0.135723	−1.604694
i	0.665504	0.044467	0.001977	0.180989	−0.014053
k	13.357055	0.403139	0.162521	−0.039108	−1.355678
l	0.920778	0.002927	0.000009	0.186886	−0.006978
w	1.418555	0.116064	0.013471	0.137509	−1.697539

Table 2 Moments of simulated variables, with time to build case

4 Conclusion

We have introduced a gestation time of multiple periods into a more Smets-and-Wouters-like new-keynesian model for a closed economy. The adequate model has been derived and the calibration and impulse-response analysis have followed. In order to make a comparison of the impact of the gestation time, a similar procedure has been applied to a model without the time to build. As, by definition, these two models must have identical steady state, the magnitude of simulated endogenous variables are close to each other. However, the response of endogenous variables in the models are different, especially of investment decisions, workhours, and, to some extent, also consumption. These variables do not react to productivity shock smoothly, but in a saw-teeth-like shape. As the introduction of time to build approximates the investment more realistically and the dynamics of responses to exogenous shock is interesting, it would be worth a more comprehensive investigation of the impact of gestation time in macroeconomic models, estimation effort included. The model derived in this paper can be easily expanded further and it can usefully serve as a starting framework for many additional studies.

Acknowledgements

This research was supported by the University of Economics, Prague under the framework of Institutional Research Support no. IP 100040.

References

[1] Adjemian, S., Bastani, H., Juillard, M., Mihoubi, F., Perendia, G., Ratto, M., Villemot, S. (2011). Dynare: Reference manual, version 4. *Dynare Working Papers*, 1, CEPREMAP.

[2] Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics* 12(3), 383–398.

[3] Galí, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton: Princeton University Press.

[4] Kalecki, M. (1935). A Macro-Economic Theory of Business Cycles. *Econometrica* 3 (1935), 327–344.

[5] Kydland, F. E., Prescott, E. C. (1982). Time to Build and Aggregate Fluctuations. *Econometrica* 50, 1345–1370.

[6] Smets, F., Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review* 97(3), 586–606.

[7] Smets, F., Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European economic association*, 1(5), 1123–1175.

[8] Tinbergen, J. (1931). Ein Schiffsbauzyklus? Weltwirtschaftliches Archiv. Reprinted in Klaasen L.H. et al. Selected Papers. Amsterdam, 1959.

[9] Tran, V. Q. (2019). *Macroekonomické modely pro měnovou analýzu*. Praha: Oeconomica.

Comparison of Customer Satisfaction with Catering Facilities across Four Independent Studies

Tomáš Konderla, Jaromír Běláček[1]

Abstract. We compare four customer satisfaction studies with catering facilities, two conducted in Brno (for McDonalds and fast food chains) and the other two in Vyškov and in Vsetín. Available data are presented as the means derived from 1–5 satisfaction scales for 7 satisfaction attributes that were jointly interviewed in the studies, including their standard deviations with a total sample size of $n = 747$ respondents. We decided to subject these data to formal meta–analysis, where we compare 21 possible pairs of attributes for two basic model schemes. We found the most significant statistical difference between the best rated quality of food and the lowest nominal price of provided catering services ($p < 0.001$) for the model with random effects. The remaining five attributes were placed among them as follows: restaurant staff – services, cleanliness, offer – advertising, environment – appearance of restaurant facilities, availability. In the same order, a consistent series of statistical significances of paired tests was found, but here only in the model with a fixed effect. These results hold according to the model assumption of internal correlation $r = 0.5$ within individual studies.

Keywords: catering facilities, customer satisfaction, meta-analysis, pair comparison of the means, Cohen's d, fixed- and random-effect size models

JEL Classification: M30, M31, J13
AMS Classification: 62F10, 62J12

1 Introduction

Evaluation of consumer satisfaction in restaurants is an important indicator of the quality of provided catering services. In the market we can find different types of catering facilities that are visited by different consumer segments. Satisfaction ratings can therefore vary according to the type of catering establishment (chains, regions), consumer groups (broken down by age, gender, their eating habits, for example) and in term of different aspects services provided (evaluation attributes). Within the available data sources – see Hálová [4], Halvelandová [5], Ingrová [6], and Trtíková [7] – we decided to compare four independent studies on customer satisfaction in four regional destinations: two studies concern chains of McDonalds and fast foods chains in Brno, the third study covers several restaurants in Vyškov and the fourth in Vsetín.

In addition to the data published above, we would like to get a more universal view about consumer ratings, which are not from individual studies so transparent. We used the methodology of Borenstein et al. [2], MS Office tools and practical experience with similar marketing researches to achieve this goal. At last we were interested in possibility to assess the hierarchy of relationships between general categories of customer satisfaction attributes of catering facilities in Moravia region.

2 Data and methodology

From available variables we separate for our goals the following attributes: 1 – availability, 2 – offer/advertising, 3 – food quality, 4 – prices, 5 – services, 6 – cleanliness and 7 – appearance/environment. In each of the four studies we have the results of surveys containing of the mentioned attributes of visitor satisfaction in five to seven catering establishments ($K = 6$ for *I* – McDonalds, $K = 5$ for *II* – other fast foods, $K = 6$ for *III* – Vyškov and $K = 7$ for *IV* – Vsetín). The data presented at the scale of 1–5 (1 – *very satisfied*; 2 – *rather satisfied*; 3 – *I don't know/I don't care*; 4 – *rather dissatisfied*; 5 – *very dissatisfied*) we transformed to the level of trios: n_{ijk} – the number of respondents, m_{ijk} – an average (mean score) of values 1 – 5 and s_{ijk}^2 –

[1] Mendel University, Faculty of Business and Economics, Department of Statistics and Operations Research, Zemědělská 1, 613 00 Brno, Czech Republic, tomas.konderla@mendelu.cz, jaromir.belacek@mendelu.cz

individual variance – for study i ($i = I, \dots, IV$), satisfaction attribute j ($j = 1, \dots, 7$) and catering facility k ($k = 1, \dots, K$).

At the level of each study and a chosen satisfaction attribute we work with sample sizes

$$n_{ij} = \frac{1}{K}\sum_{k=1}^{K} n_{ijk}, i = I, \dots, IV, j = 1, \dots, 7 \tag{1}$$

and with Generalized Least Squared, so called Aitken's GLS, estimators – see Cipra [3], p. 233 –

$$m_{ij} = \left(J'V_{ij}^{-1}J\right)^{-1}\left(J'V_{ij}^{-1}M_{ij}\right), \quad i = I, \dots, IV, \quad j = 1, \dots, 7, \tag{1a}$$

where $J'=(1,\dots,1)$ indicates the K-vector of ones (apostrophe here means transposition), $V_{ij}=(s_{ijk}.r_{km}.s_{ijm})$ are KxK variance matrices for assumed model correlations r_{km} for within studies attributes (r_{km}=1 for k=m, r_{mk}= r > 0 for $k \neq m$) and $M'_{ij}=(m_{ij1},\dots, m_{ijK})$ are K-vectors of mean scores in individual restaurants; equivalents of individual scattering parameters s^2_{ijk} are estimated by

$$s^2_{ij} = (J'V_{ij}^{-1}J)^{-1}, \quad i = I, \dots, IV, j = 1, \dots, 7. \tag{1b}$$

For summary level for the j-th attribute we can make estimates (see Anděl [1], pp. 132–134)

$$n_{+j} = \sum_{i=I}^{IV} n_{ij}, \quad j = 1, \dots, 7, \tag{2}$$

$$m_{+j} = v_{+j}(s_{ij})\sum_{i=I}^{IV}\frac{m_{ij}}{s^2_{ij}}, \quad j = 1, \dots, 7 \tag{2a}$$

$$v_{+j}(s_{ij}) = \left(\sum_{i=I}^{IV}\frac{1}{s^2_{ij}}\right)^{-1}, \quad j = 1, \dots, 7, \tag{2b}$$

which are analogous to (1), (1a) and (1b) for four *independent* groups/studies.

The primary data material modified by formulas (1)–(2b) is presented in Tables 1–3 – the trios n_{ijk}, m_{ijk} and s^2_{ijk} are located in parallel cells. To fully complete the data, we substituted the three missing attributes in two studies (see Note below Table 3) by simple row averages. The values from Table 2 are drawn in Figure 1–2 to provide a visual idea of the differentiation of customer satisfaction attribute values across the studies. (The results are partially interpreted in Part 3.) The average satisfaction scores in the figures are supplemented by 95% Confidence Limits (95% CI), the width of which is a function of significantly different numbers of respondents inside individual studies (see Table 1). Figure 2 shows two alternatives of generalized GLS estimates of summarized main effects, both based on formula (2a) and (2b) (for definition see legend below Figure 2).

Following the methodology presented in Borenstein (2011) we performed 21 pairwise comparisons for fixed and random effect-size models. For the i-th study and the selected pair of attributes, say with indices j_1=1 and j_2=2 for the first couple, we followed formula 4.18 and 4.15 in Chapter 4 (in slightly modified connotation):

$$d_{i12} = \frac{m_{i1} - m_{i2}}{s_{d_{i12}}}, \quad i = I, \dots, IV, \tag{3a}$$

where

$$s^2_{d_{i12}} = \frac{n_{i2}.s^2_{i1} + n_{i1}.s^2_{i2}}{n_{i1} + n_{i2}} - s^2_{i1}.s^2_{i2}.r, \quad i = I, \dots, IV, \tag{3b}$$

which are position and variance parameters for Cohen's differences at the level of individual studies with correlation r in matched pairs. For each of the 21 pairs created from seven satisfaction attributes, we then use the variance approximation for d_{i12} expressed by the formula (see 4.20 in Borenstein (2011)):

$$V_{d_{i12}} = \frac{n_{i2} + n_{i1}}{n_{i2}.n_{i1}} + \frac{d_{i12}}{2(n_{i1} + n_{i2})}, i = I, \dots, IV, \tag{4}$$

Based on the same principles as above, we get the summary GLS estimates for bilateral differences and for their summary variances for fixed-effect models if we replace the values of $V_{d_{i12}}$ instead of s_{ij} in formulas (2a) and (2b). We further performed all necessary additional calculations in accordance with Chapters 12–14 in Borenstein [2], especially Working Examples – Part 1, and also with reference to the random effect models for Cohen's d and its Hedges corrections. We apply those computations to all 21 couples of attributes in our own user application created in MS Excel, including automatic marking of the statistical significance of the results and the necessary graphic utilities. The main results are given in the following chapter.

	1 – General availability	2 – Offer – advertising:	3 – Food quality:	4 – Prices:	5 – Services:	6 – Clean-liness:	7 – Appearance – environment:
I – McDonalds	156.8	156.7[1)]	155.7	156.7[1)]	157.8	157,0	156.2
II – Fast foods	382.2	382.0	382.2	382.4	381.8	382.1[1)]	382.2
III – Vyškov	36.7	36.7	36.7	36.7	36.7	36.7	36.7
IV – Vsetín	189.1	191.6	191.6	191.6	191.6	191.6	191.6
Summary	764.8	766.9	766.1	767.3	767.9	767.4	766.6

Table 1 Mean number of respondents across four studies (*I–IV*) and the interviewed satisfaction attributes – see formula (1) or (2) respectively

choice: r=0.5	1 – General availability	2 – Offer – advertising:	3 – Food quality:	4 – Prices:	5 – Services:	6 – Clean-liness:	7 – Appearance – environment:
I – McDonalds	2.534	2.630[1)]	2.536	2.630[1)]	2.652	2.592	2.835
II – Fast foods	2.542	2.210	2.278	2.438	2.300	2.338[1)]	2.259
III – Vyškov	1.453	2.289	1.465	1.829	1.321	1.703	1.514
IV – Vsetín	2.408	2.523	2.335	2.517	2.345	2.219	2.665
Summary	1.993	2.392	1.932	2.243	1.921	2.164	2.090

Table 2 Mean scores of customer satisfaction across four studies (*I–IV*) – see formula (1a) or (2a) respectively

choice: r=0.5	1 – General availability	2 – Offer – advertising:	3 – Food quality:	4 – Prices:	5 – Services:	6 – Clean-liness:	7 – Appearance – environment:
I – McDonalds	0.853	0.831[1)]	0.772	0.831[1)]	0.850	0.786	0.888
II – Fast foods	0.832	0.749	0.779	0.818	0.732	0.771[1)]	0.711
III – Vyškov	0.516	0.726	0.470	0.593	0.517	0.662	0.533
IV – Vsetín	0.922	0.870	0.873	0.910	0.893	0.904	0.875
Summary	0.359	0.394	0.330	0.378	0.348	0.383	0.352

Note: [1)] estimated due to missing data

Table 3 Individual standard deviations of satisfaction attributes across four studies (*I–IV*) – see formula (1b) or (2b) respectively

3 Results

The results based on 21 pairwise comparisons at the level of individual studies correspond with those presented through Figure 1. Especially at the level of fixed-effect size models, we verified a number of statistically significance specifications: for the study with potentially the most accurate results (*II* – other fast foods with about 382 respondents), the worst ratings for 1 – availability (m_{21} = 2.542) and the second worst for 4 – prices (m_{24} = 2.438) differ statistically significantly from each other ($p = 0.014$) and also from others ($p < 0.001$). The other attributes in the arrangement according to their improving values no longer differ so significantly in pairs. Similarly, in the study for *I* – McDonalds and independently among evaluated restaurants in *IV* – Vsetín, the worst ratings for the 7 – appearance/environmental of catering facilities (m_{17} = 2.835; m_{47} = 2.665) differ statistically significantly from the others (at least for $p = 0.025$). The rest attributes of these

two studies with a comparable number of respondents (about 157 and 191, respectively) are inside their ranges of satisfaction scores 2.534–2.630 and 2.219–2.523, insufficiently distinguishable.

The largest range of values (scores 1.321–2.289) is characterized by the means of satisfaction in the study *III*, monitoring catering situation in Vyškov: In this study, the lowest rated 2 – offer/advertisement (m_{32} = 2.289) also differs statistically significantly from all others ($p < 0.001$); the second and third worst rated (4 – prices and 6 – cleanliness of restaurant facilities with value of m_{34} = 1.829 and m_{36} = 1.703) in Vyškov are also divided from the rest by higher significance. Certainly it is related to the incomparably low sample size equal to 37, because with such a small number of respondents we must assume a more or less selective assessment of satisfaction with catering services. This hypothesis is supported also by the visual comparison of mean profiles on Figure 1 itself, because with the exception of the already mentioned 2 – offer / advertisement, all other attributes evaluated by respondents from Vyškov have incomparably better scores than in other studies. However, the 2 – offer / advertisement itself is the only one that differs statistically significantly from all other satisfaction attributes in Vyškov even for random-effect models (!).

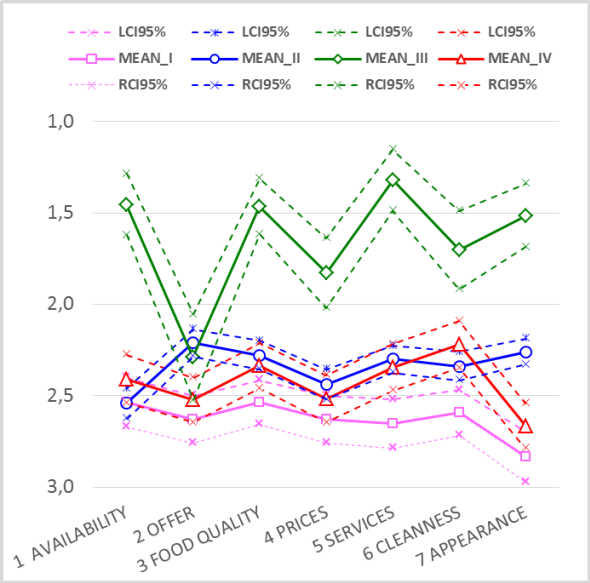


Figure 1: Input mean scores for seven attributes of consumer satisfaction in four studies with lowest and highest confidence limits (95% CI)
Legend: For the *i*-th study, we present data from Tables 1–3: values of mean satisfaction scores m_{ij} computed by (1a) ± 1.96-times their standard deviations computed by $s_{ij} \cdot n_{ij}^{-1/2}$ with s_{ij} from (1b) and n_{ij} from (1); formal links between satisfaction attributes at the level of individual studies suggest relatively significant structural differences; the widths of the confidence intervals reflect significantly different sample sizes at the level of individual studies

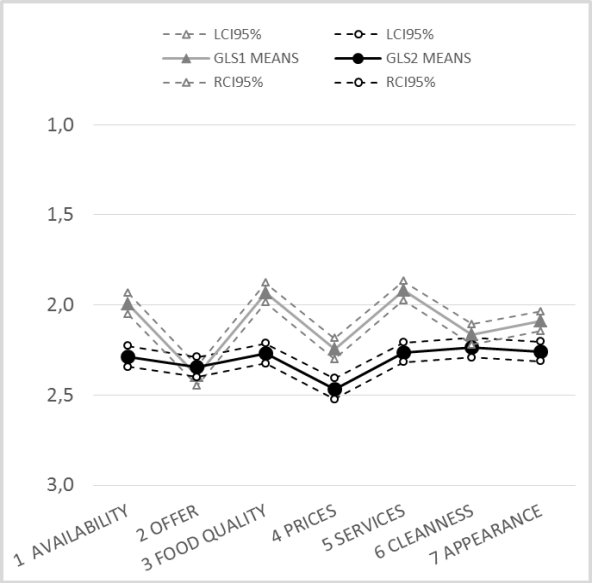


Figure 2: Summary GLS mean scores for attributes of consumer satisfaction by two computational methods with lowest and highest confidence limits (95% CI)
Legend: For summary values from the last rows of Table 1–3: the GLS1 MEANS m_{+j} were computed by (2a) ± 1.96-times standard deviations by formula (2b), where all s_{ij} were replaced by $s_{ij} \cdot n_{ij}^{-1/2}$; the alternative GLS2 MEANS we obtained due to (2a–b) for all s_{ij} replaced by $s_{ij} \cdot n_{ij}^{-1/2}$; despite of the same, but higher formal accuracy of both "summary" GLS estimators (measured by confidence limits), the observed mean lines suggest the possibility of an ambiguous assessments

Because random-effect size models also eliminate possible heterogeneities at the level of between or within factors, we would like to achieve the similar goal finally at the level of pairwise comparisons of summary GLS estimators. But among the 21 options we managed to identify only one statistically significant pair: the difference between the overall lowest rating of the 4 – prices on services against the overall best rating of 3 – quality of the meal ($p < 0.001$). This result to be the most significant one – not only in terms of their factual good interpretation.

Because random-effect size models also eliminate possible heterogeneities at the level of between or within factors, we would like to achieve the similar goal finally at the level of pairwise comparisons of summary GLS estimators. But among the 21 options we managed to identify only one statistically significant pair: the difference between the overall lowest rating of the 4 – prices on services against the overall best rating of 3 – quality of the meal ($p < 0.001$). This result to be the most significant one – not only in terms of their factual good interpretation.

For the level of the fixed-effect size model, the elimination of the influence of possible heterogeneous disturbances can no longer be generally assumed. Nevertheless, based on the performed calculations, and especially with regard to the high overall range of selection (about 767 in this case), we could consider any statistically significant results to be relatively valid. For all 21 couples created from satisfaction attributes, the results are presented in Table 4. This table contains summary GLS estimates based on Cohen's d (in higher triangle) computed by (3a), (3b) and (4) for the choice of $r = 0.5$, the corresponding GLS estimates of standard deviations (in lower triangle of table) and the marks indicating their statistical significance to zero (the null hypothesis). The satisfaction attributes in the table were arranged after the preliminary analysis in such way so that only positive GLS values of the position estimates remained in the upper triangle. This gives us a clear assessment for all 7 attributes of satisfaction, from the best-lowest hypothetical value to the worst-highest: 3 – food quality, 5 – services, 6 – cleanliness, 2 – offer-advertising, 7 – appearance – environment of catering facility, 1 – general availability, and 4 – price.

Cohen's *d* vs Std. dev.	**3 – Food quality**	**5 – Services**	**6 – Clean-liness**	**2 – Offer – advertising**	**7 – Appearance – environment**	**1 – Availability**	**4 – Prices**
3 – Food quality	×	0.046	0.056	0.110*	0.220***	0.255***	0.289***
5 – Services	0.05112	×	0.004	0.044	0.170***	0.216***	0.237***

Cohen's *d* vs Std. dev.	3 – Food quality	5 – Services	6 – Clean-liness	2 – Offer – advertising	7 – Appearance – environment	1 – Availability	4 – Prices
6 – Cleanliness	0.05117	0.0512	×	0.063	0.156**	0.204***	0.230***
2 – Offer	0.05156	0.05161	0.05152	×	0.114*	0.149**	0.158**
7 – Appearance	0.05147	0.05137	0.05162	0.05153	×	0.047	0.072
1 – Availability	0.05146	0.05144	0.05142	0.05207	0.05181	×	0.027
4 – Prices	0.05141	0.05141	0.05129	0.05146	0.05144	0.05131	×

Legend: *, ** or *** respectively denote statistical significance on level of $p = 0.05$, $p = 0.01$ or $p = 0.001$ for bilateral differences; the levels of significance are based on standardized Cohen's d differences (on higher triangle by formula (3a)) divided by their standard deviations (in lower triangle by (4)) in comparison to critical values of normal (Gaussian) distribution

Table 4 Summary GLS estimates for Cohen's d differences with corresponding GLS estimates of their standard deviations for 21 fixed-effect sample size pair models

If we add to the above arrangement the statistical significance given in Table 4 for models with a fixed effect, we find also through p-values the consistent ranges from which the satisfaction attributes differ statistically significantly from each other and where they no longer: Above all, the subgroup of three generally worst perceived attributes appears as the more compact (e.g. 7 – appearance differs from 2 – offer only with the meaning of $p = 0.027$, but 2 – offer toward 1 – availability and further to 4 – prices differs pairwise already with $p < 0.01$). On the opposite side of the spectrum of average ratings we also find at the level of significance $p = 0.05$ an indistinguishable group of best rated attributes: 3 – food quality, 5 – services, 6 – cleanliness – however 2 – offer/advertising is statistically significantly closer to this group (only the best rated 3 – food quality differs statistically significantly from it due to $p = 0.032$).

The last results are valid for a range of experimentally chosen values of r approximately from 0.4 to 0.6. For lower values of r ($0 \leq r < 0.4$), the final arrangement of attributes based on bilateral differences partially changes, but the statistical significance of these differences is no longer as consistent with this new arrangement. For higher values of r ($r > 0.6$) the basic arrangement of attributes does not change so much, but formally increasing statistical significance between attributes leads to gradual atomization of all seven attributes; thus, the force of the larger aggregate sample size for too high r ($r \approx 1.0$) gradually decreases.

4 Discussion

The advantage of using the principles of meta-analysis is the overall increase of the sample to a sample size corresponding to the sum of all studies. For the task of general evaluation of satisfaction attributes, the final arrangement of attributes presented in section 3 might seem sufficient. Unfortunately, the final arrangement of satisfaction attributes presented by us does not correspond faithfully to either of the two estimates for summary GLS estimators that we consider natural for a given task (see in Figure 2). This problem is well-known starting with the standard One-Way fixed-effect ANOVA methodology, when problem with more than two levels of the main factor is preliminary solved by testing of simultaneous (zero) hypothesis of equality of all mean values in groups (in our case with seven dependent attributes). Only after rejecting this basic hypothesis, the pairwise comparisons go on at a level not exceeding the specified level of significance. The opposite procedure is not considered right.

In this case, we feel justified to make it, as we found at least one statistically significant pair already among the models with random effects (for objectively the most distant attributes of satisfaction – 3 – food quality vs. 4 – prices). Based on the same values for input mean scores, standard deviances and sample sizes at the level of all studies, the fixed-effect and the random-effect size models differ only by larger confidence intervals around the values of the bilateral differences for the second model. In Figures 3–4, which correspond to the standard outputs of the meta-analysis illustrated in Borenstein (2011), we see that the impact of aggregate models must generally increase significance levels if the couples of attributes behave consistently in most studies. If the pair is in most studies inconsistent (bilateral differences are placed more common on both sides of the zero axis or their standard deviations are too large), the summary bilateral differences do not have a chance to be located far enough from the zero axis including 95% confidence interval as the whole.

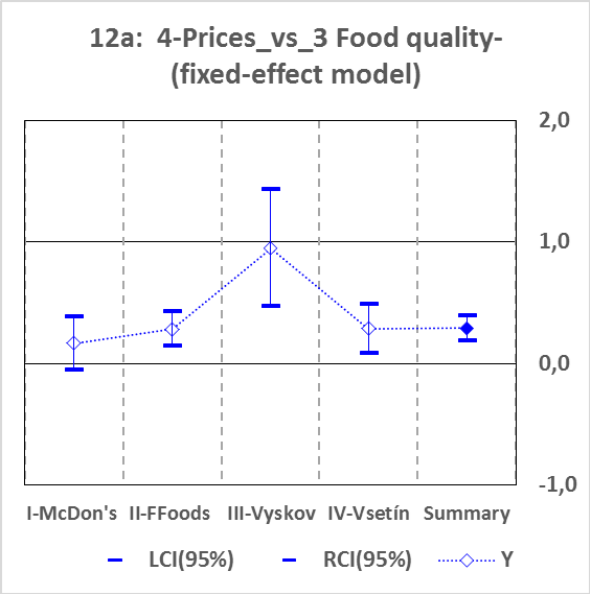


Figure 3: Impact of intervention (fixed-effect model) for the most significant mean difference between 3 – quality and 4 – prices found among 21 couples based on 7 attributes of satisfaction across four studies and summary effect
Legend: For fixed-effect size models we compare first the Cohen's standardized differences among the four studies, which are aggregated to summary mean effect with less confidence interval; scale estimators are based on assumption of the same (homogeneous) theoretical expected value between the studies

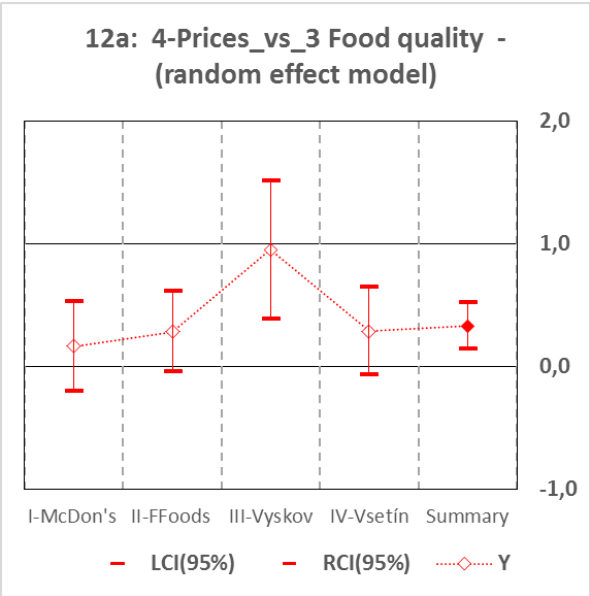


Figure 4: Impact of intervention (random-effect model) for only one significant mean difference between 3 – quality and 4 – prices found among 21 couples based on 7 attributes of satisfaction across four studies and summary effect
Legend: For random-effect size models we compare the Cohen's standardized differences with assumption of (maybe) different (nonhomogeneous) expected values across the studies; hence confidence intervals for input studies and also aggregated summary GLS estimators have larger range.

In order to standardize the problem, we solve, we would need to work with pairwise estimates, which would operate within each tested pair with universal variance estimates uniform for all possible pairs at the same time (similarly as in Scheffé or Tukey's method). Of course, such estimates can still be made. The available methodology is more suitable for situations where only one of the two compared groups is experimental and the other is the control one. However, the methodology presented in Borenstein [2] was created right for such tasks, which are more typical, for example, for biomedical applications.

5 Conclusions

The tests based on synthetic meta-analysis show statistically significant generally worse evaluation of prices (with summary mean score 2.367) and the best evaluation for food quality (with mean score 2.046) – this difference is statistically significant for fixed-effect and also random-effect size models (by $p < 0.001$). Other attributes of satisfaction with catering services, which we found located "somewhere between them", we managed to arrange in parts only by using the significance of pair differences in fixed-effect size models. Our results hold even when based direct on the size of effect d (according to Cohen) or based on the correction of deviation of such effect (according to Hedges). Such result is implied by the fact that the sample sizes used in our four studies are large enough, so there is no need to do the correction. What is more, the results hold under assumption that the average correlation r between the attributes of satisfaction is in the range of approximately from 0.4 to 0.6. For the other choices of r ($r > 0$), the hierarchy of the mutual arrangement of attributes changes partially; the statistical significance of all pairwise differences changes by a similar way.

References

[1] ANDĚL, J. (1985). *Matematická statistika*. SNTL, Praha.

[2] BORENSTEIN, M., HEDGES, L. V., HIGGINS, J. P., & ROTHSTEIN, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

[3] CIPRA, T. (1986). *Analýza časových řad s aplikacemi v ekonomii*. SNTL/Alfa, Praha.

[4] HÁLOVÁ, M. (2018). *Vyhodnocení kvality restauračních zařízení ve Vyškově*. Bachelor thesis.

[5] HALVELANDOVÁ, V. (2018). *Hodnotenie spokojnosti zákazníkov s fastfoodovými reťazcami spoločnosti McDonald's Česká Republika v Brne*. Bachelor thesis.

[6] INGROVÁ, A. (2017). *Hodnocení spokojenosti zákazníků s fastfoodovými řetězci v Brně a okolí*. Bachelor thesis.

[7] TRTÍKOVÁ, J. (2018). *Hodnocení spokojenosti zákazníků s restauračními zařízeními ve městě Vsetín.* Bachelor thesis.

Application of Network Analysis Methods to the Process of Preparing the Operation of a New Scheduled Flight Connection for Passenger Transport

Lenka Kontriková[1], Michal Dorda[2], Jiří Hořínka[3], Martin Kubáň[4]

Abstract. In the current competitive environment, the main activity of air carriers operating scheduled air passenger transport is the search for new business opportunities. Establishing a new route is a long-term and complicated decision-making process composed of many series-parallel sub-activities which follows each other in objective and time meaning. The preparation of the new routing operation can, therefore, be understood as a project. In the past we have dealt with the application of network analysis methods to manage such a project. The article deals with the application of the PERT method in the preparation of the scheduled flight connection between Leos Janacek Ostrava Airport – Warsaw Chopin Airport.

Keywords: network analysis methods, flight connection, passenger transport

JEL Classification: C44
AMS Classification: 90B10

1 Motivation to solve

In an evolving competitive environment, each airline operating scheduled flights must continuously update its network. This update is performed on the basis of the analysis of the operated flights connections and the evaluation of the potentials of individual flight connections. Based on this analysis, some regular flights whose potential has not been proven may be cancelled and new flights that promise commercial potential may be launched. The creation of each new scheduled flight requires long-term preparations, which can take up to two years before the airport and the carrier agree on the operation of the new flight connections.

The current situation on the air transport market is characterized by the fact that the demand of mainly regional airports for the start of operation of new flights connections significantly exceeds the capacity of air carriers. The start of operation of a new scheduled flight is a complicated process, business negotiations between the airport and the air carrier on the start of operation of the new flight will not be successful and the start of operations on the new flight connection will not take place. The reasons can be business and procedural. Therefore, the whole process of launching a new flight connection needs to be carefully planned so that its possible delays do not unnecessarily complicate the process leading to the start of the operation of the new flight connecting. Two parties are involved in the launch process – the air carrier and the airport. The project will be conceived from the perspective of an air carrier that is interested in starting operations on the scheduled flight connection. This means that the issues of the effective arrangement of partial steps are solved, which follow each other objectively and temporally.

The network analysis methods are a suitable tool for managing the process of preparations for the start of operations on a new flight connecting. The methods of network analysis will make it possible to time the implementation of individual partial activities leading to the start of operation of a new flight connecting so that the start of operation of the flight takes place without unnecessary time delays (i.e. in a minimum of time). As suitable methods for planning the course of the creation of a new flight connecting, we can choose the CPM method or the PERT method – see for example [1]. The PERT method is more appropriate at a time

[1] VSB-Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic, lenka.kontrikova@vsb.cz
[2] VSB-Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic, michal.dorda@vsb.cz
[3] VSB-Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic, horinka@centrum.cz
[4] VSB-Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic, MartinKuban@seznam.cz

when the duration of individual activities is more complicated to estimate. Some examples of using the CPM and PERT methods in air transport can be seen in [2].

2 Process of preparing a new scheduled flight connection

The first important moment of the whole decision-making process is the moment of the start of preparations for the operation of the new flight connecting. Such a decision is made several months in advance and, for long-distance lines, up to one or two years, depending on the company's internal procedures and circumstances. Underestimation of the preparation may be the cause of the subsequent non-implementation of the plan due to, for example, its impracticability at a time when it is already offered on the market, or a new flight connecting is launched on the market when it is not yet ready for it. A number of proposals, after analysing all the information, maybe rejected, given the small potential, or the proposal is postponed for a later period, e.g. due to operational or aeropolitical impracticability.

We have already dealt with the preparation of a new scheduled flight connection using the network graphs in conference papers [4] and [5]. The conference paper [4] dealt with using mathematical methods for planning the establishment of a new flight connection. In this article, the CPM method was used, the conference paper [5] dealt with the modelling process of launching a new scheduled flight connection where was used the CPM method.

The PERT method will be used to manage the project of introducing a new connection flight from Ostrava to Warsaw, as the duration of individual activities in both stages is subject to random influences. Thus, they cannot be considered as constants, but as random variables.

The process of preparing a scheduled flight connection includes activities that could be divided into two groups – activities that are always carried out (mandatory activities) and activities that are carried out beyond mandatory activities (other activities). The scope of other activities depends on the specific air carrier and its operational requirements, which it plans to comply with when operating the flight connecting. As the flight from Ostrava to Warsaw will be operated in the hub & spoke system, the project will include mandatory other activities as well, as this is characteristic of the flights connection operated in the hub & spoke system.

Mandatory activities in the case of the Ostrava – Warsaw flight will be: verification of the possibility of air restrictions, identification of the clientele, communication of the request to airports, time at participating airports, aircraft selection in terms of the proposed flight, times and capacity, adjustment of requirements according to airport dispositions, financial analysis of planned flight, closing ancillary contracts on operating conditions, preparation of a preliminary draft timetable, the conclusion of commercial contracts, phase of ensuring the sale of tickets in a new market, assessment of the benefits of using the services of a general agent in ticket distribution, pricing, operating components, quality and safety (part 1 and part 2), preparation of a marketing plan for the introduction of the flight, marketing, schedules flight, business contracts, complete provision of entry into the sales system.

Other activities can be expected to be arrangements with other air carriers in case of participation in the airline alliance, preparation and implementation of presentations for contractors, the conclusion of contracts in relation to other partners, implementation of a marketing plan, which includes marketing activities in relation to partners and auction prices, preparation of a package that includes, for example, contacts with hotels, the printing of the tour operator's catalogue, distribution of the tour operator's catalogue.

Compared to the general procedure reported in the literature [6], the following activities are not considered: verification of the existence of restricted airspace, preparation of commercial agreements concerning the formal initiation of the inter-state airspace contracting process and the time required for bilateral negotiations between States leading to the signing of trade agreements, the conclusion of trade agreements in relation to the contractual provision of operations to partners and the necessary actions for the approval of the commencement of operations by government authorities. These activities are not considered because they are not relevant for the project.

3 Application of the PERT method

The whole process leading to the launch of a new scheduled connection flight from Ostrava to Warsaw can be divided into two stages. Stage 1 includes the preparation of documents for the approval of the business

plan and the elaboration of recommendations for management, and Stage 2 includes the actual implementation of the approved business plan. Stage 2 can be further divided into two parts. The first part of Stage 2 begins with the approval of the business plan to start operations on the new flight connection and ends with the issuance of a final decision on whether or not to start operations. The start of operations is conditioned by an acceptable risk that the business plan will not be fulfilled. The second part of Stage 2 begins with the issuance of the final decision to start operation and ends on the day of the start of the first flight.

The initial step in providing input to the PERT method application is to create a default activity table. The table 1 contains the names of individual activities and their semantic content. Some of the activities listed in the table are divided into two parts (e.g. 3a, 3b). The division of activity into several phases is used in a situation where a certain phase of activity is conditioned by the termination of another activity.

Activity number	**Activity name –** The semantic content of the activity
1	**Verifying the existence of restricted airspace –** in the case of the regular flight Ostrava – Warsaw, this is a preliminary investigation into the existence of free slots at Warsaw Airport during the day.
2	**Identification of clientele –** the offers for any of the target groups, e.g. for premium passengers or business travellers, tourists, students, for whom the pricing and flight schedules differ.
3a	**Passing the intention to the airport authorities –** survey and pre-selection of a handling agency at Warsaw Airport, specification of the type and scope of services provided, and their price evaluation.
3b	**Response time of participating airports –** confirmation or modification of requested services by the airport.
4	**Selection of the aircraft in terms of proposed route, times and capacity** – the selection is done either from existing company fleet, or by acquiring the new type of aircraft.
5	**Adjustment of requirements according to airport dispositions** – we expect some time to assess any comments and take them into account for the preparation of the plan.
6	**Financial analysis of the upcoming flight –** preliminary cost-benefit analysis based on the model of clientele type distribution and dependent pricing.
7	**Preparation of documents for the management's final decision on the operation of the line** – completion and sorting of all collected documents.
8	**Examine the codeshare options on the route –** it is expected to take some time to assess whether our alliance partners or companies outside the alliance are interested to cooperate on this route.
9	**Support and operating conditions contract conclusion** – there will be the negotiations about conditions of operation between the MS region or another entity that could be interested in the flight introduction and support it (Hyundai, etc.)
10	**Creating the draft of the new flight schedule** – incorporation of new route into the existing flight schedule of the company to be optimal in terms of connection options in the destination, in terms of the our fleet utilization and includes the restrictions given in point 1.
11	**Conclusion of business contracts –** contractual provision of operations in relation to a specific draft flight schedule, with WAW airport, with a handling agency and with distribution companies.
12	**Phase to establish the ticket sales in a new market** – own distribution channels and possibly external sales in cooperation with distribution companies.
13	**Benefits assessment of using the general agent services in the tickets distribution** – selection and implementation of a general sales agent.
14	**Pricing, trade** – calculations of the air tickets price for tour operators on the domestic market and abroad

Activity number	**Activity name –** The semantic content of the activity
15a	**Operational components, quality, and safety (part 1)** – technical feasibility of flight and checking
15b	**Operational components, quality, and safety (2nd part)** – in case of the technical feasibility deficiencies, the proposal of technical feasibility solution
16	**Preparation of marketing plan for line introduction**
17	**Marketing** – preparation and implementation of presentations for contractors
18	**Flight schedule** – specification of preliminary flight schedule
19	**Business contracts** – other contractual reinsurance in relation to partners
20	**Complete access to the BSP** (Billing and Settlement Plan) **sales system**
21	**Pricing, trade, distribution of prices to tour operators**
22a	**Pricing, revenue management (part 1)** – preparation of prices for individual travellers including published tariffs, revenue management settings, distribution of prices for individual travellers.
22b	**Pricing, revenue management (part 2)** – pricing arrangements with other carriers.
23a	**Operational components, quality, and safety (part 3)** – general phase of preparation of operational-technical provision of clearance and flight.
23b	**Operational components, quality, and safety (Part 4)** – incorporation of contractual specifics, requirements, and possibilities of airports, technical handling, etc. into the preparation of operational-technical ensuring of handling and flight.
24	**Marketing** – addressing agencies and business partners (e.g. presentations).
25a	**Marketing** – preparation of the marketing plan implementation for the introduction of a line that includes marketing activities in relation to partners
25b	**Marketing** – concrete design of marketing plan including action prices
26a	**Tour operator (part 1)** – preparation of a package that includes e.g. contacts with hotels
26b	**Tour operator (2nd part)** – specific preparation package according to the price
27	**Print tour operator catalogue**
28	**Distribution of tour operator catalogue**
29	**Preparation of the product sale according to the catalogue**

Table 1 Significant content of individual activities

Due to its size, the network graph representing the given project is decomposed into several parts shown in Figures 1–3. The interconnection of individual parts of the network graph is performed via end/start nodes in the given figures. Figures 1–3 show the earliest possible and latest permissible start and end times of the individual activities, calculated on the basis of the mean durations of the individual activities $E(T_{ij})$. The figures also highlight the activities that make up the critical path.

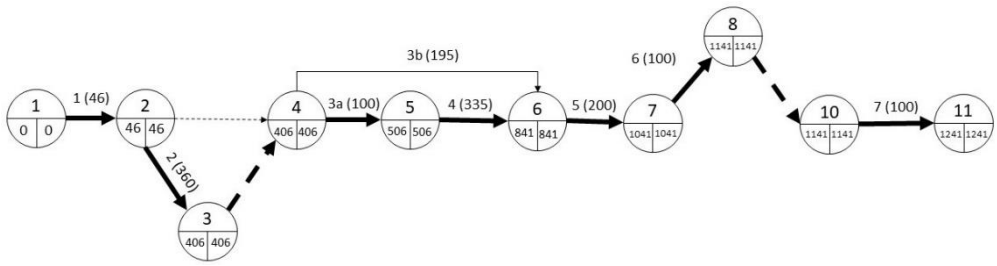


Figure 1 Network graph Stage 1

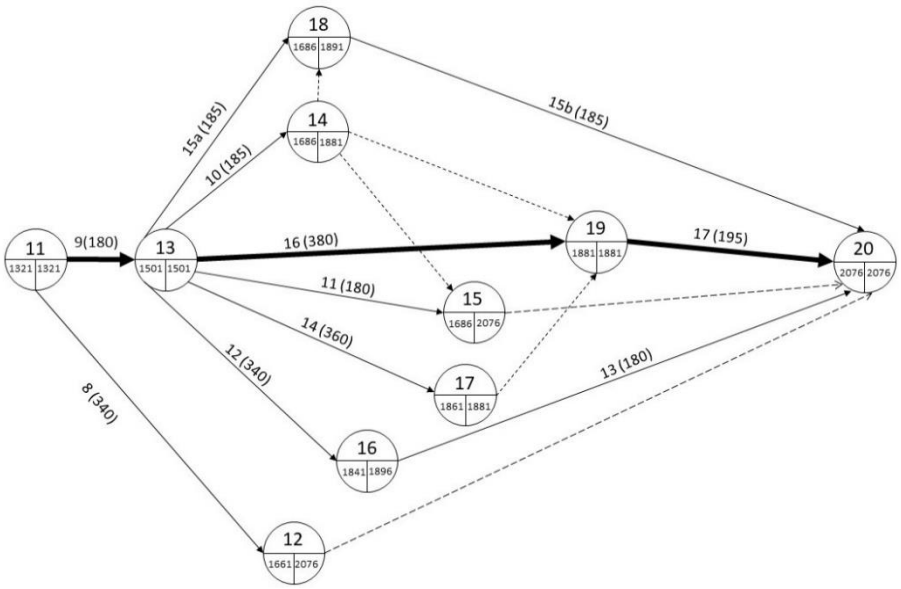


Figure 2 Network graph Stage 2, part 1

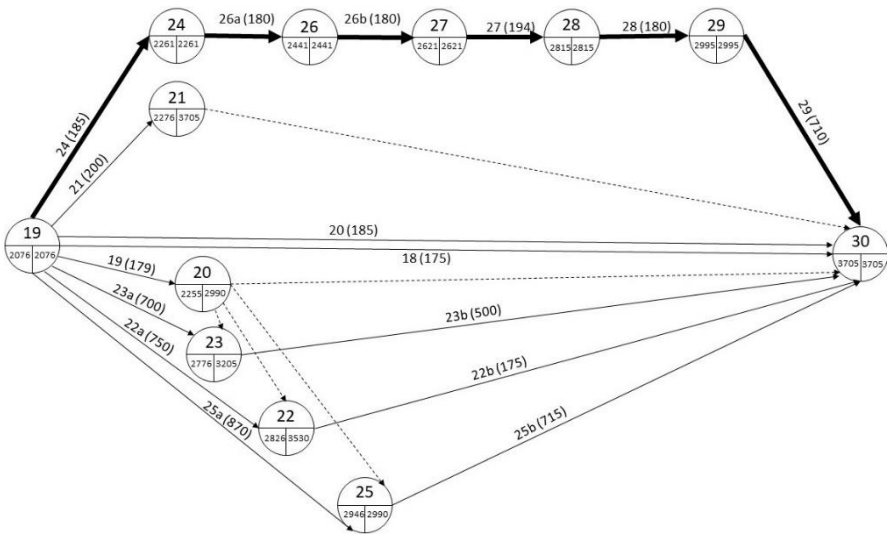


Figure 3 Network graph Stage 2, part 2

It has been stated above that the process of preparing a new scheduled flight connection can be divided into two stages. The first stage is defined by the vertex of a graph 1 and the vertex of a graph 11 in Figure 1, when the business plan for the start of operation of the new flight connection is submitted to the management of the air carrier. When the business plan is approved by the air carrier's management, Stage 1 ends, and Stage 2 can be started. Another important decision-making situation occurs after the completion of activities 13, 15b, and 17 in Figure 2, when the risks of possible further continuation and a final decision are issued on whether or not to start a new flight connection. If it is definitely decided to start the operation of a new flight connection, the second part of Stage 2 will follow, see Figure 3.

The basis for the creation of network graphs was the data presented in the Table 2, where the optimistic calculation a_{ij} and the pessimistic calculation b_{ij} are expert estimates. These data were determined by one of the authors who work in the field. The most likely calculation m_{ij} are the values given in the literature [6].

Activity	Edge $[i,j]$	Duration of activity (days) a_{ij}	b_{ij}	m_{ij}	Previous activity	Following activity	$E(T_{ij})$ (days)	$\sigma(T_{ij})$ (days)
1	[1; 2]	4	14	7	×	2	46/6	10/6
2	[2; 3]	30	90	60	1	3a, 3b	360/6	60/6
3a	[6; 7]	10	30	15	2	4	100/6	20/6
3b	[6; 8]	15	60	30	2	5	195/6	45/6
4	[7; 8]	5	90	60	3a	5	335/6	85/6
5	[8; 9]	20	60	30	3b, 4	6	200/6	40/6
6	[10; 11]	10	30	15	5	7	100/6	20/6
7	[11; 12]	15	45	30	6	8	180/6	30/6
8	[6; 10]	10	90	60	7	×	340/6	80/6
9	[12; 13]	15	45	30	7	10, 11, 12, 14, 15a, 16	180/6	30/6
10	[13; 14]	20	45	30	9	17	185/6	25/6
11	[13; 16]	15	45	30	9	×	180/6	30/6
12	[13; 17]	25	75	60	9	13	340/6	50/6
13	[17; 21]	20	40	30	12	×	180/6	20/6
14	[13; 18]	40	80	60	9	17	360/6	40/6
15a	[13; 19]	15	50	30	9	15b	185/6	35/6
15b	[19; 21]	15	50	30	15a	18, 19, 20, 21, 22a, 23a, 24, 25a	185/6	35/6
16	[13; 20]	45	95	60	9	17	380/6	50/6
17	[20; 21]	15	60	30	16	18, 19, 20, 21, 22a, 23a, 24, 25a	195/6	45/6
18	[21; 33]	10	45	30	8, 11, 13, 15b, 17	×	175/6	35/6
19	[21; 22]	14	45	30	8, 11, 13, 15b, 17	×	179/6	31/6
20	[21; 33]	20	45	30	8, 11, 13, 15b, 17	×	185/6	25/6
21	[21; 24]	20	60	30	8, 11, 13, 15b, 17	×	200/6	40/6
22a	[21; 25]	90	180	120	8, 11, 13, 15b, 17	22b	750/6	90/6
22b	[25; 33]	15	40	30	22a	×	175/6	25/6
23a	[21; 26]	60	160	120	8, 11, 13, 15b, 17	23b	700/6	100/6
23b	[26; 33]	30	110	90	22a	×	500/6	80/6
24	[21; 27]	25	40	30	8, 11, 13, 15b, 17	26a	185/6	15/6
25a	[21; 28]	90	180	150	8, 11, 13, 15b, 17	25b	870/6	90/6
25b	[28; 33]	100	135	120	25a	×	715/6	35/6
26a	[27; 29]	25	35	30	24	26b	180/6	10/6
26b	[29; 30]	25	35	30	26a	27	180/6	10/6
27	[30; 31]	14	60	30	36b	28	194/6	46/6
28	[31; 32]	15	45	30	27	29	180/6	30/6
29	[32; 33]	90	140	120	28	×	710/6	50/6
∑	—	—	—	—	—	—	617.5	91.83

Table 2 Initial list of project activities

Based on the network graphs, a critical path has been found, consisting of 16 activities (1, 2, 3a, 4, 5, 6, 7, 9, 16, 17, 24, 26a, 26b, 27, 28, 29). Tab. 2 shows the calculated mean times of individual activities $E(T_{ij})$ and calculated standard deviations σ_{ij}. Based on the data given in the total line in Tab. 2, is the mean duration of the whole project $E(T) = 617.5$ days and the standard deviation $\sigma(T) = 91.83$ days.

$$\sigma(T) = \sqrt{8432.7489} = 91.83 \text{ days.}$$

While respecting the normality of the input data and assuming that the number of 16 activities of which the critical path is composed is sufficient and assuming that the sum of random variables representing the duration of activities on the critical path will also follow the normal distribution, we can calculate the probability that the project will take place within two years of its launch.

We calculate this probability from the relation: [3]

$$P(T \leq T_p) = \Phi\left[\frac{T_p - T}{\sigma(T)}\right] = \Phi\left[\frac{730 - 617.5}{91.83}\right] = \Phi[1.2250] \doteq \Phi[1.22] = 0.88877$$

where, T_p is the planned duration of the project for two years, T is the expected value of the whole project and $\sigma(T)$ is the standard deviation. The probability of starting operation within two years from the start of Stage 1 is therefore with a probability of 0.88877.

T_p [years]	2	2,1	2,2	2,5
$P(T \leq T_p)$	0.88877	0.95154	0.97670	0.99931

If the air carrier requests, that the project should be successful with a probability of 0.95, then it must allow for a period of 2.1 years.

4 Conclusion

The preparation of a new flight connection from Leos Janacek Ostrava Airport to Warsaw Chopin Airport (the project) is a very complicated process for the parties involved. During this process, the air carrier must take into account a number of factors and must also gather a lot of information that are important for the preparation. The article deals with the application of network analysis methods (specifically PERT methods) to a given project. When applying the PERT method, it has been found that the whole project can be completed in an average of 617.5 days, with a standard deviation of 91.83. If the air carrier sets a maximum of two years for the preparation of a new flight connection, we can say that this requirement will be met with a probability of 0.88877. With a very short extension of the time to prepare for the start of operation, the probability of successful implementation of the project increases significantly. E.g. when the time for preparations for the start of operation of the new flight connection is extended by 0.1 year (from 2 to 2.1 years), the probability increases by approx. 0.07 (from 0.88 to 0.95), etc.

Acknowledgements

The paper was supported by internal project from the Faculty of Mechanical Engineering Faculty of VSB – Technical University of Ostrava, SP 2020/90 Research in the field of transport and process systems and technologies.

References

[1] ANDERSON, D. R., SWEENEY, D. J., WILLIAMS, T. A., & CAMM JEFFREY, D. J. J. (2016). *Quantitative Methods for Busines.* Boston: Cengage Learning. ISBN 978-1-285-86631-4.

[2] FLOURIS, T. G. & LOCK D. (2009). *Managing Aviation Project from Concept to Completion*. Ashgate Publishing. ISBN 978-0-7546-7615-7.

[3] JABLONSKÝ, J. (2007). *Operational research.* Praha: Professional Publishing. ISBN 978-80-86-946-44-3, 323 p.

[4] KONTRIKOVÁ, L., TEICHMANN, D., HOŘÍNKA, J., & OLIVKOVÁ, I. (2020). Using mathematical methods for planning the establishment of a new flight connection. In: *19th Conference on Applied Mathematics 2020*. Bratislava: Institute of Mathematics and Physics, Faculty of Mechanical Engineering, pp. 692–700.

[5] KONTRIKOVÁ, L., HOŘÍNKA, J., TEICHMANN, D., & KUBÁŇ, M. (2020). The modeling process of launching a new scheduled flight connection. In: *20th International scientific conference 2020.* Púchov: Quantitative Methods in Economics – Multiple Criteria Decision Making (in print).

[6] PRŮŠA, J., BRANDÝSKÝ, M., HLINOVSKÝ, L., HORNÍK, J., PAZOUREK, M., SLABÝ, F., TŘEŠŇÁK, M., & ŽEŽULA, J. (2015). *The world of air transport II*. Extended edition. Praha: Galileo Training, 2015. ISBN 978-80-260-8309-2, 647 p.

How Tight is the Necessary Condition for the Second-order Stochastic Dominance?

Miloš Kopa[1]

Abstract. The paper deals with the second-order stochastic dominance relation and its necessary conditions. Assuming a discrete distribution with many atoms, the second-order stochastic dominance implementation in portfolio optimization is extremely computationally costly, especially if atoms are not equiprobable. On the other hand, there exist necessary conditions for this relation (based on comparisons of mean and minimal realization) which are easy to apply. Therefore, this paper explores the strength of these necessary conditions when considering financial data. In particular, daily returns of 49 industry portfolios from the Kenneth French library from the last 50 years are analysed. First, the number of pairs of portfolios fulfilling the necessary condition is calculated. Second, the number of pairs obeying also the second-order stochastic dominance is identified. Finally, the dependence between the strenght of the necessary condition and correlation (average mean return) of portfolios is investigated. The whole empirical analysis is performed using annual moving (non-overlapping) window approach.

Keywords: second order stochastic dominance, portfolio returns, necessary condition, correlation

JEL Classification: D81, G11
AMS Classification: 91B16, 91B30

1 Introduction

The basics of stochastic dominance theory go back to 1960th, see [25], [8], [9] or [28]. Since then the most commonly used stochastic dominance relations in portfolio optimization are the first-order and the second-order stochastic dominance which have clear and easy interpretation. They allow comparing random variables, in financial applications random returns or losses of assets (portfolios). In general, we say that a random variable X dominates a random variable Y with respect to the N-th order stochastic dominance, $N = 1, 2$ if

$$\mathbb{E}u(X) \geq \mathbb{E}u(Y) \text{ for all } u \in U_N$$

where the set of admissible utility functions for the N-th order stochastic dominance is defined as follows:

$$U_N = \{u(x) \in D_N : (-1)^k u^{(k)} \leq 0, \ \forall x, \ k = 1, ..., N\}$$

where D_N is the class of N-times differentiable functions and $u^{(k)}$ stands for the k-th derivative of function u. See [18] for more details. In the last decade, a substantial development of the stochastic dominance applications in finance was observed. In particular:

- necessary and sufficient conditions for portfolio efficiency with respect to stochastic dominance criteria were derived and applied, see e.g. [23], [7], [15], [22]
- portfolio enhancement using stochastic dominance rules proved to be a promising way comparing to classical mean-risk models, see e.g. [27], [10], [24], [14], [17], [29].
- Special Data Envelopment Analysis models wich are equivalent to portfolio efficiency tests with respect to stochastic dominance were presented in [1], [2], [3].
- more robust version of stochastic dominance relations and stochastic dominance efficiency were proposed, see e.g. [4], [12], [5] or [6].
- more general stochastic dominance (ordering) rules were introduced, analyzed and applied, see e.g. [20], [19] and references therein, [21], [13] or [11].

In this paper, we focus on the second-order stochastic dominance under assumption of empirical distribution of random returns given by N equiprobable realizations. In this case, the second-order stochastic dominance relation could be verified by comparison of cumulative returns (realizations). This, however,

[1] Charles University, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, kopa@karlin.mff.cuni.cz

could be computationally intractable in portfolio optimization problems with the second-order stochastic dominance constraints where we can not pre-order the returns. Therefore, we can relax this condition by the necessary condition based on comparison of mean returns and minimal returns. The goal of this paper is to analyze how tight is this necessary condition when applied to returns of 49 US representative industry portfolios. Moreover, the analysis is enriched by correlations of the returns (average mean returns) because we expect that higher the correlation (smaller the average mean) is, stronger the necessary condition should be. Finally, we repeat the analysis for 50 non-overlapping time periods (50 years).

The remainder of this paper is structured as follows. Section 2 presents a notation and basic properties of the second-order stochastic dominance relation including the necessary condition. It is followed by an empirical study in Section 3 and the paper is concluded in Section 4.

2 Second-order Stochastic Dominance Relation

We say that a random variable X dominates a random variable Y with respect to the second-order stochastic dominance (SSD), if

$$Eu(X) \geq Eu(Y) \quad for \quad all \quad u \in U_2$$

where the set of admissible utility functions U_2 for the second-order stochastic dominance is defined as follows:

$$U_2 = \{u(x) : (-1)^k u^{(k)} \leq 0, \quad \forall x, \quad k = 1, 2\}.$$

See Levy (2016) and references therein for more details. In the rest of the paper, the random variables will represent daily returns of the US representative industry portfolios. And we say that one portfolio dominates the other one if stochastic dominance between their daily returns holds true.

2.1 Stochastic dominance for empirical distributions

Let r_i be the random return of the i-th asset (portfolio) with distribution function $F_i(x)$. Then i-th portfolio dominates j-th portfolio with respect to SSD if and only if

$$\int_{-\infty}^{z} F_i(x)\, dx \leq \int_{-\infty}^{z} F_j(x)\, dx \;\; \forall z \in \mathbb{R}$$

If r_i takes value $r_{i,t}$, t = 1,...T with the same probabilities, such that

$$r_{i,1} \leq r_{i,2} \leq \ldots \leq r_{i,T}$$

then i-th portfolio dominates j-th portfolio with respect to SSD if and only if

$$\sum_{s=1}^{t} r_{i,s} \geq \sum_{s=1}^{t} r_{j,s} \quad t = 1, \ldots, T.$$

Since this necessary and sufficient condition needs to order the returns first, it is not applicable in the portfolio selection problem with the second-order stochastic dominance constraints. Therefore, some necessary conditions may be useful as a relaxation. In this paper, we focus on the following necessary condition for the second-order stochastic dominance relation: If i-th portfolio dominates j-th portfolio with respect to SSD then:

$$r_{i,1} \geq r_{j,1} \wedge \sum_{s=1}^{T} r_{i,s} \geq \sum_{s=1}^{T} r_{j,s}. \tag{1}$$

3 Empirical study

In this section we analyze the strength of the necessary condition (1) for the daily returns of US 49 representative industry portfolios using moving annual window analysis of 50 years history. First, we compute NC_i – the number of pairs (out of all 49*48 pairs) which satisfy the necessary condition (1) in i-th year. Then we check every such pair whether it obeys SSD or not. The number of pairs obeying SSD in year i is denoted by SSD_i, $i = 1, 2, \ldots, 50$ This is done for all 50 considered annual periods. The results are summarized in Figure 1.

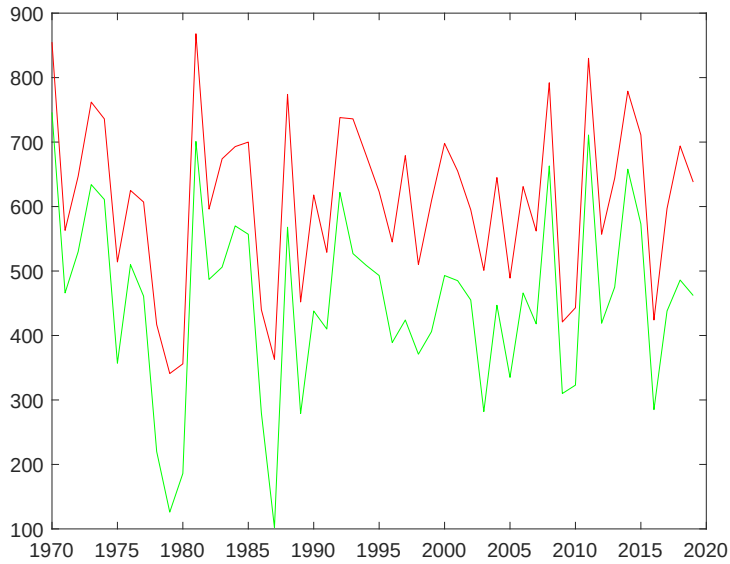


Figure 1 Number of pairs of portfolios satisfying the necessary condition (red) and SSD relation (green)

Figure 1 shows coincidence and high correlation (0.96) between number of pairs satisfying the necessary condition and number of pairs obeying SSD. In average, the necessary condition is also a sufficient condition for SSD in 72% cases, that is:

$$\frac{1}{50}\sum_{i=1}^{50}\frac{SSD_i}{NC_i} = 0.72$$

So, in average, only 28% of pairs of portfolios satisfying the necessary condition do not obey SSD. The ratio:

$$S_i = \frac{SSD_i}{NC_i}$$

is called the strength of the necessary condition. If the portfolios are perfectly correlated (positively or negatively) then $NC_i = SSD_i$, i.e. $S_i = 1$. So we may expect that S_i positively depends on the absolute value of average correlation over all pairs of portfolios satisfying the necessary condition (AC_i^{NC}) or SSD (AC_i^{SSD}). Figure 2 depicts the results of S_i, AC_i^{NC} and AC_i^{SSD}.

From Figure 2, we can conclude:

- except of two outliers (1979 and 1987), the stregth of the necessary condition ranges between 0.5 and 0.85. We will exclude the outliers in the rest of the analysis.
- the average correlation of portfolios satisfying the necessary condition do not differ substantially from the average correlation of portfolios obeying SSD relation, so we will consider only the average correlation of portfolios satisfying the necessary condition in the rest of the paper.
- the dependence of the strength of the necessary condition on the average correlation was not observed.

To confirm the last conclusion with excluding the two outliers, Figure 3 shows no significant dependence between the strength of the necessary condition and the average correlation.

Therefore, we proceed with the analysis considering an average mean return over all 49 portfolios in i-th year. We may expect that if the market is going down the strength of the necessary condition is increasing. And Figure 4 confirms this expectation. For the three years when the average mean return is smaller than -10% the strength of the necessary condition is very high (0.82–0.84) and as the average mean return increases to +10% the strength falls down to (0.5–0.8) and vary much more.

4 Conclusions

In this paper we analyzed the strength of the necessary condition for the second-order stochastic dominance when applied to daily returns of 49 US industry representative portfolios. We provided a moving window analysis over last 50 years. For every year, we computed the strength of the necessary condition as the ratio of number of portfolio pairs satisfying the necessary condition and of the portfolio pairs obeying SSD

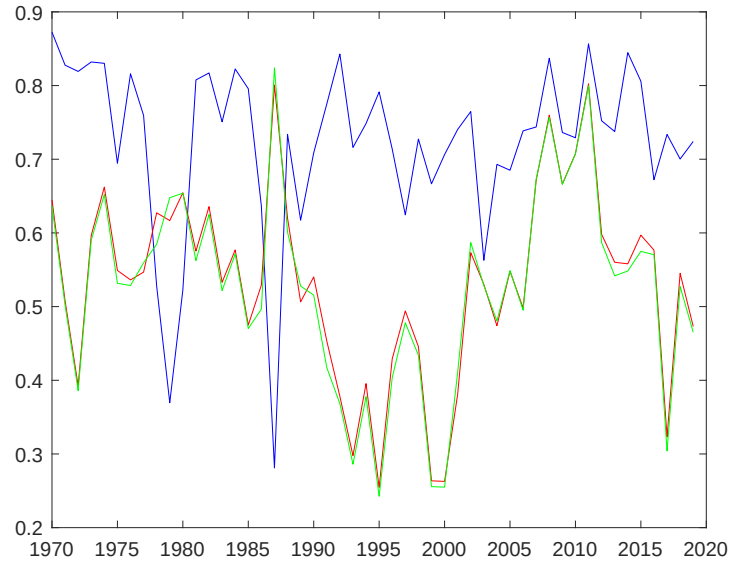


Figure 2 The strength of the necessary condition (blue), average correlation for portfolios satisfying the necessary condition (red) and SSD relation (green) is presented.

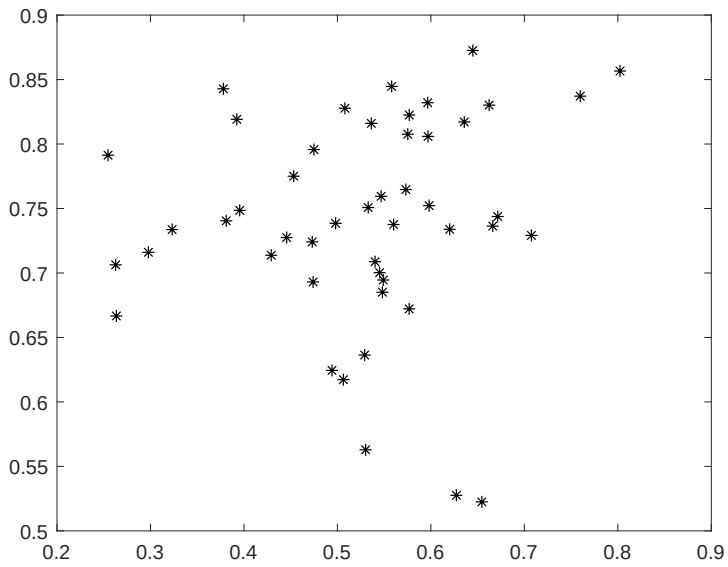


Figure 3 The strength of the necessary condition in dependence on the average correlation

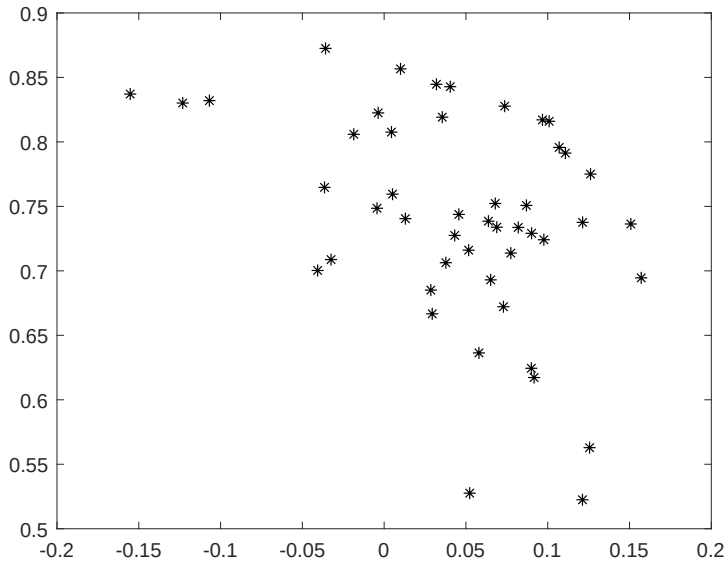


Figure 4 The strength of the necessary condition in dependence on the average mean return

relation. We assumed empirical distributions of returns. Then we compared the evolution of the strength with the evolution of the average pair correlations. Perhaps surprisingly, we did not find any significant dependence between these two variables. Therefore, we enriched our empirical study with analysis of the dependence between the strength and average (over all portfolios) mean returns. We found an interesting conclusion that the strength negatively depends on the average mean returns, that is, when market is going down the strength is the highest while when the market is going up the strength is very volatile and substantially smaller.

These findings could be of interest in a complicated optimization problem when the SSD constraints have to be relaxed somehow. It may typically occur in some multistage stochastic portfolio optimization problems (e.g. [17], [14] or [29]) or in the problems with endogenous randomness (e.g. [16])

For the future research, this study can be improved in various ways. For example, one can consider more robust stochastic dominance relations following [12], [5] or [6], using contamination techniques and the worst-case approach. Alternatively, one can compare these results also with the case when the first or the third order stochastic dominance is used [24].

Acknowledgements

The paper was supported by the grant No. 19-28231X of the Czech Science Foundation.

References

[1] Branda, M., and Kopa, M.: DEA-risk Efficiency and Stochastic Dominance Efficiency of Stock Indices. *Czech Journal of Economics and Finance* **62** (2012), 106–124.

[2] Branda, M., and Kopa, M.: On relations between DEA-risk models and stochastic dominance efficiency tests. *Central European Journal of Operations Research* **22** (2014), 13–35.

[3] Branda, M., and Kopa, M.: DEA models equivalent to general *N*-th order stochastic dominance efficiency tests. *Operations Research Letters* **44** (2016), 285–289.

[4] Dentcheva, D., and Ruszczyński, A.: Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming, Series B* **123** (2010), 85–100.

[5] Dupačová, J., and Kopa, M.: Robustness in stochastic programs with risk constraints. *Annals of Operations Research* **200** (2012), 55–74.

[6] Dupačová, J., and Kopa, M.: Robustness of optimal portfolios under risk and stochastic dominance constraints. *European Journal of Operational Research* **234** (2014), 434–441.

[7] Grechuk, B.: A simple SSD-efficiency test. *Optimization Letters* **8** (2014), 2135–2143.

[8] Hadar, J., and Russell, W. R.: Rules for Ordering Uncertain Prospects. *American Economic Review* **59** (1969), 25–34.

[9] Hanoch, G., and Levy, H.: The efficient analysis of choices involving risk. *Review of Economic Studies* **36** (1969), 335–346.

[10] Hodder, E., Jackwerth, J. C., and Kolokolova, O.: Improved Portfolio Choice Using Second-Order Stochastic Dominance. *Review of Finance* **19** (2015), 1623–1647.

[11] Kabasinskas, A.; Sutiene, K.; Kopa, M.; Luksys, K.; Bagdonas, K. Dominance-Based Decision Rules for Pension Fund Selection under Different Distributional Assumptions. *Mathematics* **8** 2020, n.719.

[12] Kopa, M.: Measuring of second-order stochastic dominance portfolio efficiency. *Kybernetika* **46** (2010), 488–500.

[13] Kopa, M., Kabasinskas, A., Sutiene, K.: Preference selection of Lithuanian second pillar pension funds by stochastic dominance rules, *IMA Journal of Management Mathematics*, to appear.

[14] Kopa, M., Moriggia, V. and Vitali, S.: Individual optimal pension allocation under stochastic dominance constraints. *Annals of Operations Research* **260** (2018), 255–291.

[15] Kopa, M., and Post, T.: A general test for SSD portfolio efficiency. *OR Spectrum* **37** (2015), 703–734.

[16] Kopa, M., and Rusý, T.: A decision-dependent randomness stochastic program for asset–liability management model with a pricing decision. *Annals of Operations Research*, https://doi.org/10.1007/s10479-020-03583-y.

[17] Moriggia, V., Kopa, M. and Vitali, S.: Pension fund management with hedging derivatives, stochastic dominance and nodal contamination. *Omega* **87** (2019), 127–141.

[18] Levy, H.: *Stochastic dominance: Investment decision making under uncertainty*. Second edition, Springer, New York, 2006.

[19] Ortobelli, S., Lando, T., Petronio, F., and Tichy, T.: Asymptotic stochastic dominance rules for sums of i.i.d. random variables. *Journal of computational and applied mathematics* **300** (2016), 432–448.

[20] Ortobelli, S., Shalit, H., and Fabozzi, F.: Portfolio selection problems consistent with given preference orderings. *International Journal of Theoretical and Applied Finance* **16** (2013), 38p.

[21] Post, T.: Standard Stochastic Dominance. *European Journal of Operational Research* **248** (2016), 1009–1020.

[22] Post, T., Fang, Y., and Kopa, M.: Linear Tests for DARA Stochastic Dominance. *Management Science* **61** (2015), 1615–1629.

[23] Post, T., and Kopa, M.: General Linear Formulations of Stochastic Dominance Criteria. *European Journal of Operational Research* **230** (2013), 321–332.

[24] Post, T., and Kopa, M.: Portfolio Choice based on Third-degree Stochastic Dominance. *Management Science* **63** (2017), 3381–3392 .

[25] Quirk, J. P., and Saposnik, R.: Admissibility and measurable utility functions. *Review of Economic Studies* **29** (1962), 140–146.

[26] Rockafellar, R. T. and Uryasev, S.: Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking and Finance* **26** (2002), 1443–1471.

[27] Roman, D., Mitra, G., and Zverovich, V.: Enhanced indexation based on second-order stochastic dominance. *European Journal of Operational Research* **228** (2013), 273–281.

[28] Rothschild, M., and Stiglitz, J. E.: Increasing risk: I. A definition. *Journal of Economic Theory* **2** (1970), 225–243.

[29] Rusý, T. and Kopa, M.: An asset liability management stochastic program of a leasing company. *Kybernetika* **54** (2018), 1247–1263.

Using Machine Learning to Predict Optimal Parameters in Portfolio Optimization Problems

Karel Kozmík[1]

Abstract. We use machine learning methods in portfolio optimization problems. Most portfolio optimization problems require selection of one or more parameters and we create machine learning model to predict the optimal values of such parameter with respect to out-of-sample performance. In this paper we use mean-CVaR portfolio optimization model and xgboost machine learning model. Extensive simulations were performed to create the dataset with the optimal choice of the desired parameter. We explore the dependencies of the optimal choice of minimal in-sample mean on input data, like number of stocks or number of scenarios. Predictor importance and prediction evaluation is presented, showing that the model gives reasonable predictions for parameter that is otherwise very hard to select.

Keywords: portfolio optimization, mean-CVaR, machine learning

JEL Classification: G11
AMS Classification: 91G10

1 Introduction

The problem of portfolio optimization is a typical problem in economics and finance. We want to invest in predefined assets and the goal is to determine the percentage share of our wealth to invest in each of the assets. The classical approach is based on models considering the trade-off between mean and risk. The mean-risk models try to maximize mean and minimize risk. Given that the models have 2 criteria, we usually put one criterion into the constraints of a optimization program, then the mean-risk models try to find the optimal portfolios that maximize mean given a maximum risk or minimize risk given a minimum mean. This means we bound the mean from below or the risk from above by a constant. The main problem is the selection of the constant. We want to select this constant in an optimal way with respect to the performance of our portfolio.

We focus on commonly used mean-CVaR model (Rockafellar & Uryasev [5], [6]), where the mean criterion is placed into the constraints, bounded from below by parameter r_0:

$$\begin{aligned} &\min_{x_i} CVaR_\alpha\left(-\sum_{i=1}^{n} x_i R_i\right) \\ &\text{subject to } \mathbb{E}\left(\sum_{i=1}^{n} x_i R_i\right) \geq r_0 \\ &\sum_{i=1}^{n} x_i = 1,\ x_i \geq 0, i = 1, \dots, n, \end{aligned} \tag{1}$$

where R_i denotes random return of the $i-th$ asset, x_i denotes the share to invest in the $i-th$ asset (weights).

Usually we use historically observed values of returns (scenarios) as the true distribution. For S equiprobable scenarios we get a linear programming formulation. One can imagine the returns r_{si} as a matrix that has stocks as columns and scenarios as rows. So r_{si} denotes returns of stock i in scenario s:

[1] Charles University, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8, kozmikk@karlin.mff.cuni.cz

$$
\begin{aligned}
&\min_{\xi, x_i, u_s} \xi + \frac{1}{(1-\alpha)S} \sum_{s=1}^{S} u_s \\
&\text{subject to } u_s \geq -\sum_{i=1}^{n} x_i r_{si} - \xi, \quad s = 1, \dots, S \\
&\sum_{i=1}^{n} x_i \bar{R}_i \geq r_0 \\
&\sum_{i=1}^{n} x_i = 1, \; x_i \geq 0, \quad i = 1, \dots, n, \\
&\xi \in \mathbb{R}, \; u_s \geq 0, \quad s = 1, \dots, S,
\end{aligned}
\tag{2}
$$

where $\bar{R}_i = \frac{1}{S} \sum_{s=1}^{S} r_{si}$.

Parameter r_0 represents minimal in-sample (mean) return and this inequality is usually fulfilled as an equality for the optimal solution. When selecting the parameter r_0 for our mean-CVaR model, the only values that make sense are values between $(\min_i \bar{R}_i)$ and $(\max_i \bar{R}_i)$.

The rest of the paper is structured as follows. The second section presents general approach, simulation setup and dataset creation. It is followed by creation of an XGBoost model in Section 3 and Section 4 concludes the paper.

2 General approach

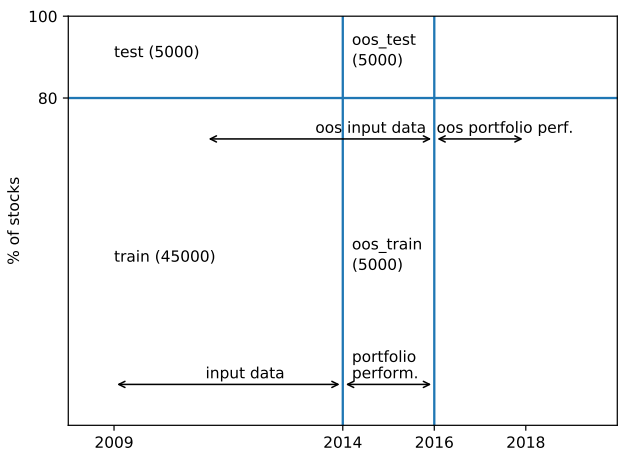


Figure 1 Simulation schema

The work-flow is briefly captured by Figure 1 and will be described throughout this paper. We divide the stocks into 2 groups: train and test, and use 2 periods: training and out-of-sample. Both train and out-of-sample periods consist of data-collecting period and target evaluation period. We want to predict the optimal value of the parameter r_0 with respect to out-of-sample performance. Firstly, we need to define how the performance is measured. We use a combination of return (yearly return denoted ret_y) and maximum drawdown (MDD) in the next 2 years (out-of-sample performance):

$$
(1-\lambda) ret_y - \lambda \cdot MDD, \tag{3}
$$

where λ is a parameter of risk aversion. For $\lambda = 0$ the investor cares only about the return (mean) and for $\lambda = 1$ the investor cares only about the maximum drawdown (representing risk).

We use a different criterion for the out-of-sample performance than the one that is used in the mean-CVaR model. We want the approach to be general and independent of the chosen portfolio optimization model. One can choose any method that uses stocks returns data and its output are portfolio weights.

Now we need a dataset containing:

- historical returns r_{si},
- λ, α,
- optimal value of r_0.

When we have this dataset, we hope that for similar values of historical returns and values of λ and α, the optimal value of parameter r_0 would be also similar. On the whole, we would like to create a model, where values of historical returns and values of λ and α are predictors and parameter r_0 is the target.

2.1 Data preparation

We use data capturing daily prices (adjusted closing prices) of a large number of American stocks downloaded from Kaggle [3]. We compute monthly returns and take only stocks which have complete data from January 2005. This results with more than 2600 American stocks with complete observations from 2005 to 2018 in our dataset.

One of the tickers – VST (Vistra Energy Corp.) – was removed from the dataset, because it disturbed the computation. There was very high volatility or some inaccuracy in the input data, when the price jumped from 2000 USD to 2 USD and back in a matter of days. However the same volatility was found on finance.yahoo [4].

2.2 Creating the learning set

Creation of the learning set corresponds to the bottom part of Figure 1, we get input data $\rightarrow$ solve the linear programming problem and get weights $x_i \rightarrow$ evaluate out-of-sample performance of this portfolio.

Most of the features are randomly generated:

- 12–60 scenarios,
- 10–100 stocks,
- α: 0.9, 0.91, ..., 0.99,
- λ: 0, 0.1, ..., 0.9, 1.

January 2014 is a fixed point, the number of scenarios determines the number of scenarios from this point to the past, which jointly with the selected stocks defines the matrix of returns r_{si}. Now that we have everything prepared, the mean-CVaR model is solved on a grid of values of parameter r_0 to determine its optimal value. Using this procedure, we generate 45000 observation of a pair (input data, target) – training dataset.

3 additional datasets containing 5000 observations each are also generated. The original train dataset was generated using only 80% of all stocks (we want the training dataset as large as possible, but we still keep enough for testing). Now we use the rest 20% to generate test dataset, so we can test our models on stocks that the model did not learn on. The other 2 datasets are generated by shifting the fixed date to January 2016 for out-of-sample evaluation, again for both groups of stocks, this corresponds to the top part of Figure 1.

Remark 1. All of the computation was done in Python.

2.3 Exploration of the dataset

Before we construct a prediction model, we investigate the dependence of optimal parameter r_0 on basic features of the data like number of stocks, number of scenarios and other parameters λ and α.

In Figure 2, we can see the dependencies. On the x axis, there is always the feature we change and want to explore its effect on optimal value of parameter r_0. On the y axis, median of all optimal r_0 for the specific value of feature on x axis is plotted. We also tried to use mean, but the same plots for mean were much more volatile.

We can see that when we have a small number of available scenarios, we should be more greedy in terms of selection the minimal in-sample return. Then with increasing number of scenarios, we should choose smaller r_0 and from about 30 scenarios (corresponding to 2.5 years of data), there is no significant change.

For the parameter α, which states the confidence level of CVaR, there seems to be no significant impact. One must check the scale on the left side, the values are almost constant. We generated α only in interval (0.9, 0.99), because this is the common practice, it might be the case that using broader range of values would unravel some dependence.

For the parameter of risk aversion λ, the results are very intuitive, the more risk averse an investor is, the lower in-sample return they should select.

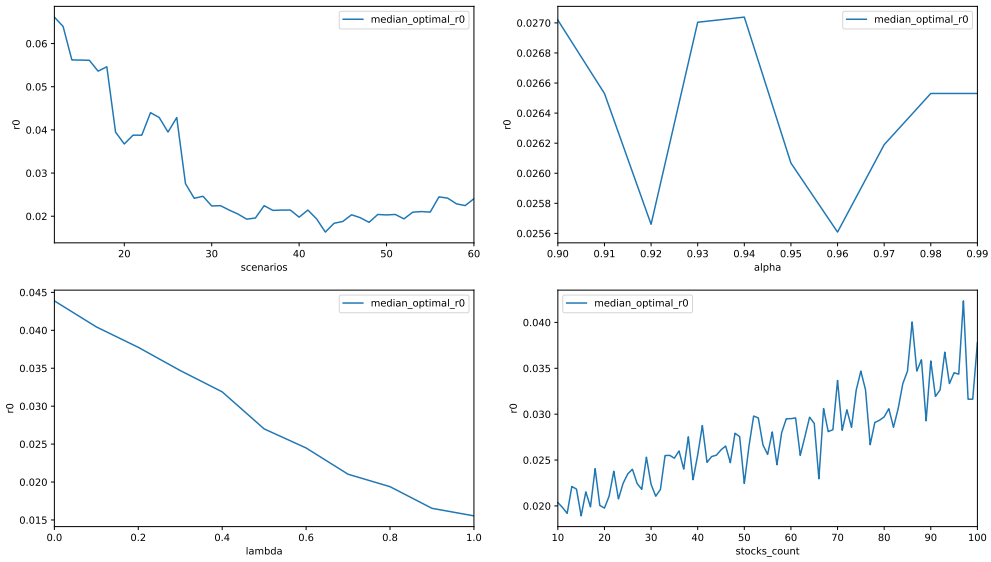


Figure 2 Exploration of basic dependencies

For the stocks count, there is clearly an increasing trend, the more stocks I am willing to invest in, the greedier I should be. It seems that with more investment opportunities, more of them have high in-sample performance and then some of them also have high out-of-sample performance.

3 Creating machine learning model

3.1 Definition of predictors

We need to define predictors that aggregate the information in observations r_{si}. We cannot simply use all the values, because we need to be able to get the same amount of predictors for different number of stocks and different number of scenarios.

The general approach is to use quantiles of different features of the data. Firstly, we use mean and standard deviation of return for each stock. From these values, we take mean and the following quantiles: 0, 0.25, 0.5, 0.75, 1. This means we take minimum, median, maximum and some other returns based on stocks. We do similar procedure using the other dimension – scenarios. We compute mean and standard deviation for each scenario and then we take 0, 0.25, 0.5, 0.75, 1 quantiles. In the end, we also take 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 1 quantiles from the combined values r_{si}.

We also include number of stocks, number of scenarios and parameters λ and α as predictors.

3.2 Definition of the target

Since the only values of r_0 that make sense are between $(\min_i \bar{R}_i)$ and $(\max_i \bar{R}_i)$, the interval is different for each observation. This makes prediction of the optimal values harder to generalize to out-of-sample observations. Noting this, we create 2 models, one when we predict directly the value of r_0 and another one, where we predict scaled (relative) value of r_0:

$$r_0^{rel} = (r_0 - \min_i \bar{R}_i)/(\max_i \bar{R}_i - \min_i \bar{R}_i). \tag{4}$$

This scaling gives us values between 0 and 1. When we do the same data exploration as we did with the absolute values, the distribution of the optimal values is not so skewed now, the graphs are almost the same for both mean and median, in Figure 3 we present medians so it is comparable.

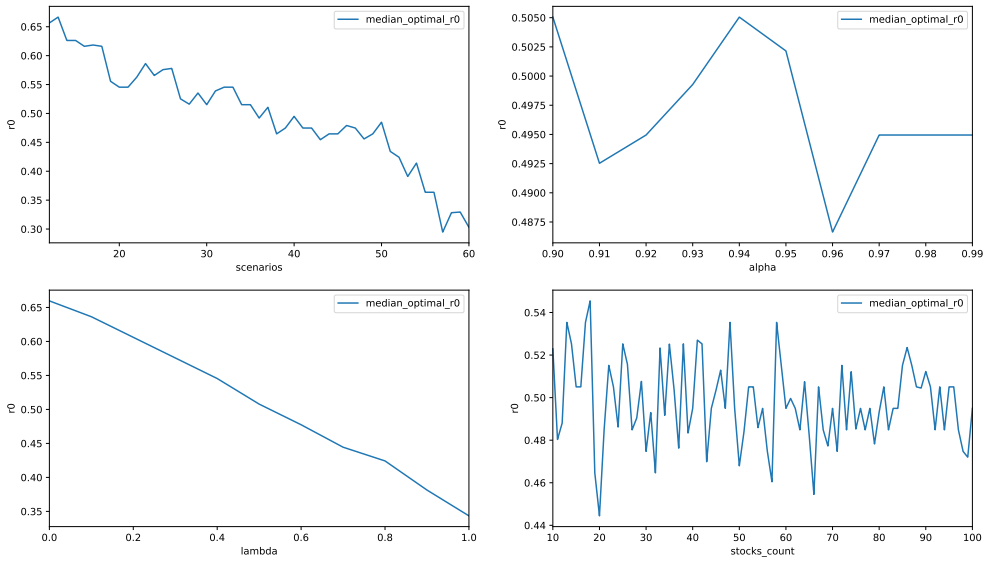


Figure 3 Exploration of basic dependencies for normalized values

3.3 Creating the model

To create the prediction model, we selected xgboost [1] package for Python, which is an implementation of a gradient boosting model, first proposed by Friedman [2]. The hyper-parameters were set expertly, using common recommended values like $depth = 5$ or $eta = 0.3$. To slightly increase the performance, one could optimize the hyper-parameters on out-of-sample performance, but the gain should not be very significant. Since we have a continuous target, we used $objective = reg : squarederror$ and $eval\_metric = rmse$, which corresponds to usual linear regression.

Both the models actually gave very similar results, but we prefer the model predicting relative optimal r_0, which is then transformed back using inverse transformation:

$$r_0 = r_0^{rel} \cdot (\max_i \bar{R}_i - \min_i \bar{R}_i) + \min_i \bar{R}_i. \tag{5}$$

Let us explore predictor importance graph, which is in figure 4 for the model predicting the relative r_0. The number after underscore denotes the quantile, when it is computed for scenarios, there is "scenario_" and "all_" denotes all the values r_{si} combined. We can see that the risk aversion parameter λ is the most important and then minimal and maximal values of standard deviations and means are important.

Remark 2. Predictor importance denotes how much the variable affected the objective value (root mean square error), i.e. how it was important for explaining the target (r_0^{rel}).

3.4 Prediction evaluation

One must keep in mind that we are in some sense predicting out-of-sample performance of stocks, which is very hard task and do not have very high expectations.

We evaluate the model separately on each of the prepared sets separately and on the train set, we use only the validation part of the data, the observations that were actually used to train the model are not evaluated as every machine learning model over-fits training data.

To evaluate the prediction, we used some standard prediction metrics for the values of r_0, but the interpretation is not straightforward. Even though the actual prediction of r_0 might not be very accurate, the final portfolio may not suffer from this, especially for higher values of r_0, which generate high error, the difference in portfolios is very small. We present prediction metrics regarding weights: absolute value of difference between the optimal portfolio (portfolio achieved using the optimal parameter r_0) and portfolio achieved by

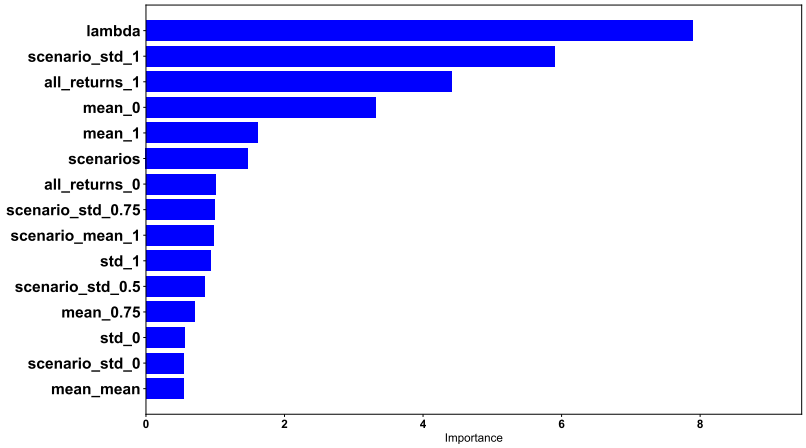


Figure 4 Predictor importance

using the predicted parameter r_0 – denoted wd (weight difference). We take absolute absolute value of the difference of the weights for each stock and sum the absolute values. Weight difference is always between 0 and 2 and we can see that for the out-of-sample, it is around 1, which means half of our capital is invested in an optimal way (when we use mean-CVaR model with the same input data). We find this result having the best interpretation. For comparison, we created a benchmark prediction (bwd = benchmark weight difference), which always takes the median value of optimal r_0 in the training dataset (0.0265). We can see that the by using the model in out-of-sample, we optimally invest around 2% more of our capital. When mean of optimal r_0 in the training data set (0.071) is taken as the benchmark, the difference is 20% in favour of using our model.

	mean_wd	median_wd	mean_bwd	median_bwd
train	0.824	0.768	0.976	0.976
test	0.900	0.889	0.977	0.970
oos_train	0.997	1.007	1.038	1.046
oos_test	1.003	1.029	1.043	1.070

Table 1 Prediction metrics

4 Conclusion

We used a large dataset containing monthly returns of more than 2600 American stocks to create a learning dataset for our machine learning model. We were able to achieve a prediction model for the parameter of minimal expected return in mean-CVaR model, for which no clear rules for the selection were available so far. Using this simple portfolio optimization problem, we wanted to show the problem and present the methods. Future research will concern more complicated models than mean-CVaR and ways to predict multiple parameters at once, for example using neural networks. Apart from methodological problems, generalizations of the problem face the issue of extensive computational power requirements, because not only that we want to use more complex problems, but we also want to predict multiple parameters, whose optimal value was now found by a grid search.

Acknowledgements

The paper was financially supported by grant EXPRO 19-28231X and SVV No. 260580.

References

[1] Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. CoRR abs/1603.02754 (2016). *arXiv preprint arXiv*:1603.02754.

[2] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

[3] Daily Historical Stock Prices (1970–2018). (2018). *https://www.kaggle.com/ehallmar/daily-historical-stock-prices-1970-2018/version/1*. Accessed 25.3.2019.

[4] Vistra Energy Corporation. (2020). *https://finance.yahoo.com/quote/VST/*. Accessed 25.10.2019.

[5] Rockafellar, R. T. & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2, 21–42.

[6] Rockafellar, R. T. & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7), 1443–1471.

Economic-Environmental and Technological Assessment of ETICS

Michal Kraus[1], Kateřina Žáková[2], Jaroslav Žák[3]

Abstract. The current societal trend is to reduce energy consumption and emissions of substances harmful to human health and threatening the climate and biodiversity. Buildings account for 40% of total energy consumption in the EU, and the sector is growing steadily. Buildings also account for 36% of greenhouse gas emissions. It is therefore clear that the energy status of buildings plays a significant role in both the economy and people's lives. ETICS (External Thermal Insulation Composite System) of buildings is particularly beneficial in energy savings. However, the impact on human health and the environment is not investigated enough, both during construction and during the lifetime and demolition of the structure. Financial demands and thermal insulation properties are usually main criteria for the selection of the best insulation. However, the impact on the environment and public health should also be taken into account. The aim of the paper is to find the most suitable variants of thermal insulation using appropriate statistical methods considering all aspects. This means considering not only variables involving thermal, technical, operational, economic, but also health and environmental aspects. Products that are friendly to the health of the population and the environment also bring economic benefits, which are sometimes difficult to quantify.

Keywords: ETICS, economic assessment, energy savings, environmental impacts

JEL Classification: C89, Q51
AMS Classification: 65-00

1 Introduction

External thermal insulation compound systems (ETICS) of buildings are already well researched in terms of technology, thermal insulation and economics. The impacts of different types of systems on the environment and the health of the population have already been studied less thoroughly [1]. There is almost no sophisticated decision-making process for selecting the specific variant for the particular building, considering all the above-mentioned aspects. There is a complete lack of quantification of the perception of the general public of different types of ETICS. The article will describe the decision-making process for finding the best (effective) solutions with consideration of thermal-technical, economic, health and environmental aspects.

Thermal insulation of buildings brings benefits to society as a whole, especially in energy savings. However, the risk of producing adverse effects on the environment and the health of the population is not negligible. The correct choice of a suitable variant and product should include an assessment not only of the thermal-technical, operational, economic but also of the health and environmental aspects.

When assessing environmental and health impacts, often only CO and sulphur emissions are assessed, and only quantitatively. The influence of other very dangerous chemical substances that are released during the production of ETICS components and subsequently also during the operation of the construction is investigated only very marginally. Maximum attention will be paid to this aspect and the individual variants will be carefully examined from this point of view as well.

[1] The Institute of Technology and Businesses in České Budějovice, Department of Civil Engineering, Okružní 517/10, 370 01 České Budějovice, Czech Republic, krausm@mail.vstecb.cz.
[2] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, myskova@mendelu.cz.
[3] The Institute of Technology and Businesses in České Budějovice, Department of Civil Engineering, Okružní 517/10, 370 01 České Budějovice, Czech Republic, zak@mail.vstecb.cz.

The general public usually does not have enough quality data on the various available variants of ETICS. The chosen method of insulation should be perceived positively by the user, which will contribute to well-being and a high standard of living.

2 Materials

The thermal-technical parameters of ETICS are provided by a layer of thermal insulation, which in thermal insulation systems can be formed by available insulation materials on the market. This is usually polystyrene foam, mineral fibre boards or lamellas, or materials based on polyurethane, polyisocyanurate or phenolic foam [13].

2.1 Styrofoam

Styrofoam is a light and rigid foam that is widely used in European construction, mainly as thermal insulation. It is also an excellent packaging material because its impact properties combined with its low weight make it a material suitable for countless uses. Expanded polystyrene is produced in two basic types, differing in production technology and properties: EPS (Expanded Polystyrene) and XPS (Extruded Polystyrene). Consumption of polystyrene in the Czech construction industry has been growing for a long time and in 2017 it repeatedly exceeded 60 thousand tons (compared to 1997, when the production of polystyrene for the construction industry was 10,000 tons). By 2030, according to EU legislation on circular economics, all plastics must be fully recyclable, including expanded polystyrene (EPS). Cut and pure polystyrene is fully recyclable. The cut polystyrene foam is then crushed and, depending on the degree of contamination, used further. Clean material without unwanted impurities can be repeatedly used for the production of new packaging or insulation. Contaminated polystyrene foam can be used for aerated concrete, aerated plasters, backfills or in the production of aerated bricks [5].

2.2 Mineral and glass wool

Mineral wool is actually a fleece made by pressing fibres made from minerals. Stone thermal insulation is produced at high temperatures by pulping basalt or gabbro basalt in a kiln and forming these fibres into mats or boards. Stone wool is non-flammable, so it is used in constructions with increased requirements for fire safety – fire dividing strips in contact thermal insulation systems, constructions with higher fire resistance, etc.

Glass wool is produced from new glass or by recycling and pulping container glass. The molten glass is blown into fibres and formed into plates or mats. The use of glass wool is similar to stone wool. In the case of mineral wool, phenol-formaldehyde resins, which are carcinogenic, have previously been widely used in the manufacture. Compared to natural materials, non-natural materials have worse disposal options. Mineral fibre recycling is very difficult. This is due to the specific production of this building material. Recycling is only possible during production from waste material. This recycling can be divided into three types: reuse without modification, regeneration to the same product or use in the production of building materials.

2.3 Polyurethane

Polyurethane is a polymer that is produced by the polyaddition of diisocyanates and di- or polyhydric alcohols to form a carbamate (urethane) bond. Polyurethane and polyisocianurate foam (PUR) are used in addition to casting and spraying directly on the construction site and for the production of board materials. The boards can be produced by cutting from blocks formed by free foaming or in moulds. A big problem for the environment is expanded gases – CFCs (HCKW, HCFC), which were used in the past. The expanding gases bound in the pores slowly escape, for their half-lives the time is about 100 years. A large part of CFCs, which can still get into the air, falls on foams for thermal insulation in construction. Therefore, polyurethanes must be disposed of in a special way. In modern household waste incinerators, CFCs are practically completely destroyed. Polyurethane is an insulating material with the most complex and energy-intensive production process. Environmental stress and the risk of poisoning occur during production. Polyurethane is a typical product of so-called hard chemistry. Inputs for the production of polyurethanes are obtained from crude oil.

2.4 Phenolic foam

Phenolic foam is a material obtained by foaming phenol formaldehyde resins. The foamed blocks are then cut to the required dimensions. The boards are provided with an aluminium layer to improve the thermal-

technical parameters of the material. The disadvantage of this insulation is the high price and is not recommended for insulating damp substrates (soaked masonry, damp concrete).

2.5 The blown cellulose

The basic component in the production of thermal insulation material blown cellulose is recycled paper. Using special technology, it is divided into cellulose fibres, which are then enriched with other ingredients.

2.6 Long shavings of wood

Long shavings of wood are a 100% natural material without chemicals and dyes. The fibres are usually 2 mm thin, 30 cm long and do not contain dust or chips. Before they are rolled into balls and wrapped, the dirt is shaken off with a special machine.

3 Impacts on human health, the environment and the availability of natural resources

Climate changes include systematic changes in the global atmosphere, which is inextricably linked to the hydrosphere, pedosphere and biosphere. The Earth maintains its thermal balance thanks to the delicate balance between the incident shortwave solar radiation and the emitted infrared, longwave radiation that escapes from the Earth's atmosphere. Gases present in the atmosphere, such as water vapour, carbon dioxide, methane, and others, allow sunlight to pass to the Earth's surface, but reflect the infrared radiation emitted by the Earth's surface partially back to Earth. The term “greenhouse effect” has been used for this phenomenon. The most important greenhouse gas is water vapour. The current concentration of CO_2 is probably the highest in the last 20 million years, the concentration of CH_4 is probably the highest in the last 420 thousand years and the current concentration of N_2O has not been exceeded for at least the last thousand years [6]. The impact of CO_2 on climate change is currently being intensively studied. In particular, the quantification of its impact and the quantification of anthropogenic climate change remain unclear [9].

The following environmental aspects were taken into account in the present study: Primary Energy Input (PEI), Global Warming Potential (GWP), Acidification Potential (AP), Eutrophication (EP), Depletion of Stratospheric Ozone (ODP) and Photochemical Oxidation Potential (POCP) [12].

3.1 Depletion of stratospheric ozone

The depletion of stratospheric ozone (ODP) allows greater penetration of solar UV radiation on the earth's surface, which negatively affects human health, the quality of the natural environment, natural resources and human creations. The release of chlorinated and brominated organic compounds into the atmosphere, especially CFCs, CFCs, HCFCs, carbon tetrachloride, 1,1,1-trichloroethane, halons and methyl bromide, leads to increasing concentrations of chlorine and bromine atoms in the stratosphere. A common feature of these substances is their relative stability. Through the transport mechanisms of the atmosphere, these substances enter the stratosphere, where they are decomposed by higher intensities of UV radiation and release chlorine and bromine atoms. In addition to halogenated hydrocarbons, N2O and methane emissions also contribute to the decomposition of ozone. Ozone decomposition occurs by chemical reactions on solid particles in the stratosphere. The extremely low temperature over Antarctica leads to the condensation of water and nitric acid to form polar stratospheric clouds. In the presence of such particles, reactions leading to ozone decomposition are accelerated [3].

3.2 Human toxicity

Human toxicity expresses the adverse effects of substances emitted into the environment on human health. There are a large number of different adverse effects of substances, from acute toxicity to mutagenicity, teratogenicity, nephrotoxicity, etc. These are effects that are of different chemical and biological nature. The category indicator is the Acceptable Daily Intake (ADI) [3, 11, 12].

Human activities release a significant amount of inorganic respiratory substances into the air: SO_2 coal combustion, CO_2 combustion processes in general, NO_x transport, CO combustion processes, HCl chemical production, HF chemical production or dust particles from combustion processes. Due to atmospheric flow and the resistance of these substances to decomposition, these substances have also been found in places on Earth where they have never been produced or applied [6].

3.3 Ionizing radiation

Ionizing radiation or radiation (radioactivity) includes the adverse effects of the release of radioactive substances into the environment as well as direct exposure to radiation, for example in building materials. Two types of emission fluxes play a role in the category of impact of ionizing radiation. Firstly, there are emissions of radioactive materials into the environment and secondly, it is direct radiation into the environment, not linked to the specific release of radionuclides (building materials).

So far, environmental noise pollution has been relatively rarely mentioned. However, increased noise levels have been shown to have an adverse effect on the health of the human population. Adverse effects of noise on animals have also been observed. More information on the development of noise pollution can be found on the European Environment Agency portal.

The main cause of photooxidants is the increased concentration of ozone, nitrogen oxides and Volatile Organic Compounds (VOCs) in the atmosphere, especially in places with lower air exchange, such as cities or valleys. In addition to ozone, we consider peroxoacetyl nitrate (PAN) to be the main toxic photooxidants, as well as hydrogen peroxide, the hydrogen peroxide radical, and other radicals which are formed as intermediates in oxidation reactions [3].

3.4 Acidification

Acidification is the process of acidification of the soil or water environment caused by an increase in the concentration of hydrogen cations that have entered the environment by atmospheric precipitation of emissions of sulphur dioxide, nitrogen oxides and ammonia. Substances causing acidification are SO_2, NO_x and NH_x, acids (HCl), H_2S. Lowering the pH of water and soil has far-reaching consequences. These are mainly the death of mountain forests, acidic surface water without fish, groundwater with a high content of toxic metals released from soils and rocks [3, 10, 12].

3.5 Eutrophication

Eutrophication (usability) is the process of increasing the content of nutrients in surface waters and soils. It is a natural phenomenon which, as a result of human activities, has exceeded an acceptable limit in the affected ecosystems. A visible consequence of eutrophication is the overgrowth of surface freshwater and seawater with aquatic cyanobacteria and algae. The primary consequences of eutrophication are disruption of the oxygen and light regime of water bodies, disturbance of the ecological balance by supporting fast-growing organisms at the expense of slow-growing ones, production of cyanobacterial toxins and loss of drinking water resources [3, 8, 10].

3.6 The ecotoxicity

The ecotoxicity impact category includes the toxic effects of emitted substances on water, soil and sediment ecosystems. These are not only toxic effects on individual organisms, but on ecosystems as a whole. The direct consequence is not only a reduction in natural wealth, but also a reduction in the quality and abundance of natural raw material resources.

Groundwater resources are depleted faster than their yield, resulting in salt water penetration in coastal areas or local pollution intrusions. The limited availability of raw material resources may lead to an increase in international tensions in the future [3]. Technologies requiring less water should be used as a matter of priority.

4 Economic aspects

The price of the examined materials was determined as the average price offered in the region of South Moravia. The prices of individual materials vary by region and are, of course, variable over time. Prices can also be affected by special prices and free shipping offers. When searching for the best variants for a specific building in a given place and at a given time, it is necessary to update the prices. Economic losses caused by damage to human health and negative effects on the environment are not quantified. They are included in the analysis through environmental variables.

5 Data and analysis

As mentioned above, in addition to economic and technical criteria, social and environmental aspects were also considered. Price was taken as an economic variable; the technical properties are described by thermal resistance. The social aspect is described by the quantity SI (Social Impact), which we obtained by our own survey (questioning). The last six variables represent the environmental aspects. Environmental data were retrieved from Czech database Envimat [4, 7].

	Outputs			Inputs					
Thermal insulation	**Price** Kč/m²	**Thermal resistance** m² K/W	**SI** Range 0–10	**PEI** MJ/(m².a)	**GWP** kg CO_2 eq/ (m² a)	**AP** g SO_2 eq/ (m² a)	**EP** g $(PO_4)^{3-}$eq/ (m² a)	**ODP** g R-11 eq/ (m² a)	**POCP** g C_2H_4 eq/ (m² a)
EPS	450	7,50	9	37,826	1,516	5,364	0,918	4,75E–05	2,432
XPS	690	7,50	8	28,954	1,146	4,018	0,904	2,65E–05	0,461
Rock wool	868	7,75	8	8,012	0,450	3,317	0,726	2,20E–05	0,177
Glass wool	770	8,75	7	25,499	0,838	3,902	1,481	1,35E–04	0,312
PUR	870	10,00	5	23,824	1,163	4,867	1,314	5,74E–06	0,226
Blown celulose	558	7,75	6	4,429	0,228	1,801	0,396	2,51E–05	0,076
Shavings of wood	897	13,80	9	1,422	0,077	0,388	0,206	5,98E–06	0,020

Table 1 Inputs table

To determine which of the studied materials are effective with respect to all considered variables, we used DEA analysis, specifically the BCC model (see [2] for details). We used environmental variables as inputs because we want to minimize them. We used technical and economic variables as outputs, which we want to maximize. We therefore considered the price with a negative sign. The results are shown in the following table.

Thermal insulation	**EPS**	**XPS**	**Rock wool**	**Glass wool**	**PUR**	**Blown cellulose**	**Shavings of wood**
Efficiency	1,000	0,474	0,306	0,235	0,168	1,000	1,000

Table 2 Table of results

Three types of thermal insulation are effective, namely EPS, blown cellulose and wood wool. The other four are ineffective. XPS, glass wool and polyurethane foam work ineffectively mainly due to high values of ecological characteristics. Glass wool has worse results compared to XPS because it has a lower SI rating and is more expensive. Polyurethane foam is the worst, although one environmental indicator has the lowest, so in other ecological indicators it approaches the maximum values, it is the worst evaluated in terms of SI and, moreover, it is relatively expensive. Stone wool does not come out efficiently mainly due to the price, in other respects it is roughly on average.

6 Conclusion

DEA proved to be a suitable mathematical tool for finding optimal materials for ETICS with consideration of all aspects, not only technical-economic, but especially environmental. In practical use, it is necessary to update the prices of materials. For the final decision, it is also necessary to take into account the specific climatic conditions, the detailed composition (layers) of the ETICS and the construction to which the EICS will be applied.

Acknowledgements

The authors gratefully acknowledge the financial support received from the Institute of Technology and Business, under project SVV202006.

References

[1] ABDALLAH, M. A., SHARKEY, M., BERRESHEIM, H., & HARRAD, S. (2018). Hexabromocyclododecane in polystyrene packaging: A downside of recycling? *Chemosphere [online], 199*, 612–616. DOI: 10.1016/j.chemosphere.2018.02.084.

[2] BANKER, R., CHARNES, A., & COOPER, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30, 1078–1092.

[3] DENSLEY TINGLEY, D., HATHWAY, A., & DAVIDSON, B. (2015). An environmental impact comparison of external wall insulation types. *Building and Environment [online], 85*, 182–189. DOI: 10.1016/j.buildenv.2014.11.021.

[4] ENVIMAT. The first Czech catalog of construction products and the impacts of their production on the environment, Prague, Czech Technical University in Prague. Retrieved from http://www.envimat.cz/

[5] GREEN SOLUTION S.R.O. (2017). *Průvodce předcházením vzniku stavebních odpadů*. Ministerstvo životního prostředí. Praha, ISBN 978-80-7212-613-9.

[6] KOČÍ, V. (2006). Hodnocení environmentálních dopadů technologií: Posuzování životního cyklu – LCA. Praha: Vysoká škola chemicko-technologická v Praze.

[7] KRAUS, M., NOVÁKOVÁ, P., & ŽÁK, J. (2018). Environmental Perspective of Thermal Insulation. In *18th International Multidisciplinary Scientific GeoConference SGEM 2018 Vienna GREEN Conference. vol. 18, issue 6.4.* (pp. 693–699). Vienna (Austria): STEF92 Technology.

[8] KUTZ, M. (2018). *Handbook of Environmental Engineering.* John Wiley & Sons.

[9] DE SCHEPPER, M., VAN DEN HEEDE, P., VAN DRISSCHE, I., & DE BELIE, N. (2014). Life Cycle Assessment of Completely Recyclable Concrete. *Materials [online], 7(8)*, 6010–6027. DOI: 10.3390/ma7086010.

[10] SEPPALA, J., POSCH, M., JOHANSSON, M., & HETTELINGH, J. P. (2006). Country-dependent characterisation factors for acidification and terrestrial eutrophication based on accumulated exceedance as an impact category indicator. *Int J LCA, 6,* 403–416.

[11] ROSENBAUM, R. K., HUIJBREGTS, M. A. J., HENDERSON, A. D., MARGNI, M., MCKONE, T. E., VAN DE MEENT, D. ET AL. (2011). USEtox human exposure and toxicity factors for comparative assessment of toxic emissions in life cycle analysis: sensitivity to key chemical properties. *Int J Life Cycle Assess, 16*, 710–727.

[12] VONKA, M. (2011). *SBToolCZ methodology*. Prague: Czech Technical University in Prague.

[13] ZATEPLOVACÍ SYSTÉMY ETICS. *TZB-info* [online]. [cit. 2018-10-22]. ISSN 1801-4399. Dostupné z: https://stavba.tzb-info.cz/zateplovaci-systemy/303-zateplovaci-systemy-etics

Default Rate in the Sector Information and Communication Activities in the Czech Republic

Radmila Krkošková[1]

Abstract. The main goal of this article is to analyze the relationship between the macroeconomic indicators and the default rate in the sector information and communication activities in the Czech Republic in the both long and short term. The vector error correction model was used for this purpose to determine both long-term and short-term causal relationships. To create the resulting model, the econometric methodology was used, namely unit root tests, Granger causality for the determination of statistically significant relationships, and the Johansen co-integration test. The results confirm the existence of relationships between macroeconomic variables and the probability of default in the sector information and communication activities. The model is based on the time series of the share of outstanding loans and the total amount of loans, and on selected macroeconomic indicators. The empirical results could be influenced by a period of a currency crisis. The data used have the character of quarterly time series in the period from 2005Q1 to 2019Q4. EViews software version 9 was used for the calculations.

Keywords: ADF test, co-integration test, default rate, macroeconomic factors, sector IC activities, VECM

JEL Classification: G21, G28
AMS Classification: 91B84, 91B64

1 Introduction

Given the growing financial markets, banks need to have certain criteria to help them both in their own decisions and in assessing the impact of individual decisions. The risk of failure rather than success is considered the most important criterion. Successful risk identification and management is a key tool for increasing an institution's profitability. From the point of view of the bank or other financial institution, the most important risk is credit risk, which is the possibility of a loss due to the debtor's inability to fulfill its obligations, as Foglia [10] states. The key parameter of credit risk is the probability of default (PD). Other parameters include default loss and default exposure, as reported by Ferrari at al., [9]. The probability of default is not only an important parameter of credit risk, but it is also an important dynamic component in banks 'stress testing, where banks' resilience, as well as the overall financial system are tested under various macroeconomic scenarios [22].

The section Information and Communication activities includes the production and distribution of information and cultural products, the provision of the means to transmit or distribute these products, as well as data or communications, information technology activities and the processing of data and other information service activities [20]. The main components of this section are publishing activities, including software publishing, motion picture and sound recording activities, radio and TV broadcasting and programming activities, telecommunications activities, information technology activities and other information service activities. Publishing includes the acquisition of copyrights for content (information products) and making this content available to the general public by engaging in (or arranging for) the reproduction and distribution of this content in various forms. All the feasible forms of publishing (in print, electronic or audio form, on the Internet, as multimedia products such as CD-ROM reference books etc.) are included in this section.

The core of this paper is the analysis of two hypotheses concerning the relationship between the macroeconomic indicators and the default rate in the sector information and communication activities in the Czech Republic: H1: There is a long-term relationship between the macroeconomic indicators and the default rate in the sector information and communication activities in the Czech Republic, H2: There is a short-term

[1] School of Business Administration in Karviná, Silesian University in Opava, Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 733 40 Karviná, Czech Republic, e-mail: krkoskova@opf.slu.cz

relationship between the macroeconomic indicators and the default rate in the sector information and communication activities in the Czech Republic.

The present text has the following structure. The second part of this paper presents theoretical background and data. There are introduced methods which are used in the empiric part of this paper. The third part contains results of econometric model describing the default rate in the sector information and communication activities in the Czech Republic. The results are summarized in the conclusion.

2 Theoretical background and data

2.1 Theoretical background

Macroeconomic models are motivated by observed assumptions that default rates for different entities rise during the recession. This fact led to the implementation of econometric models aimed at explaining the default rate using macroeconomic indicators. Estimating the default rate is at the forefront of both the professional and academic public, as shows Stoklasová [21].

Three approaches are used to model macroeconomic credit risk. The first approach is based on the search for an empirical relationship between some of the default probability indicators as a dependent variable and key macroeconomic indicators as independent variables, [4]. The second approach is based on a model which, as the debtor's default, raises the situation where the value of the return on his assets falls below a certain threshold as shows Jakubík [14]. The third approach is based on models that treat default as a random factor, as reported by Jarrow and Turnbull [16].

This paper deals with the first approach and the aim of this work is to find the influence of macroeconomic variables on the default rate within the sector information and communication activities. To this end, long-term relationships between the default rate and some macroeconomic variables are investigated by means of a co-integration analysis, Johansen [18]. A vector autoregression model is used to determine specific relationships between the monitored variables.

VAR model was also used for modelling the probability of default by, for example, Alessandri et al. [2] or Hamerle et al. [11]. Blaschke et al. [3] describe models are based on the assumption of credit quality sensitivity to changes in the economic cycle. It is important to use macroeconomic and financial variables. Variables such as economic growth, unemployment and interest rates can affect credit risk. Drehmann [6] addresses the selection of suitable model variables. Marcucci & Quagliariello [19] applied the VAR approach for credit risk modelling and four macroeconomic variables: GDP, inflation, interest rates and exchange rates. Hamerle et al. [11] used classic variables such as GDP growth, unemployment, inflation, and stock market variables. Jakubík a Reininger [15] in their paper analyse the influence of individual macroeconomic variables on the probability of default in the case of Central and Eastern Europe.

2.2 Data and methods

Quarterly data for the period from 2005/Q1 to 2019/Q4 were used for the calculations. The ARAD database of the Czech National Bank [5] was the primary data source. The description of individual variables is shown in Table 1. All these variables were seasonally adjusted, in addition by logarithmic transformation. EViews software version 9 was used for the calculations. The selection of variables was done according to Allen and Saunders [1]. Both econometric (the existence of direct links between dependent and independent variables) and the information contents of the variables play a role in the selection of suitable variables.

Variable designation	Description of variable
Y	The default rate is defined as the proportion of newly created 'bad' loans to the total volume of loans.
GDP	Gross domestic product at current prices in millions of CZK. Lower GDP growth means lower sales growth and it is more difficult for businesses to generate profits. For this reason, the default rate is expected to be higher with lower GDP.
CPI	Consumer Price Index (2005=100). It is assumed that the rise in price indices (inflation growth) will cause an increase in the default rate.
M2	Monetary aggregate M2 in millions of CZK. In the short term, the money supply decreases the interest rate and a reduction in the default rate can be expected. In the long run, a positive relationship between money supply and the default rate is expected due to increases in price.

ER	CZK/EUR. The exchange rate depreciation does not have a clear effect on the default of firms. For import firms, the growth of the exchange rate has the effect of increasing their costs (the growth of the exchange rate leads to an increase in the probability of default). The opposite is true for export companies.
BRENT	Oil price index (2005=100). Growth in oil prices is reflected in the growth of fuel, the price of which is entering most of the production. We can expect an increase in default with the increase in oil prices.
UN	Unemployment rate. A positive long-term and short-term relationship is assumed.

Table 1 Description of variables

The first step of this research is stationarity test. One of the most popular stationary tests is called the augmented Dickey-Fuller (ADF) test. Most time series in macroeconomics are non-stationary or integrated with order I(1), as stated in Engle and Granger [8], Enders [7]. The ADF test allows you to test the presence of a unit root based on three models A, B, C. Model A represents a random walk model, Model B contains a constant (μ), and Model C contains a constant (μ) and a trend component (t).

Test models are defined as follows:

Model A:
$$\triangle y_t = \gamma y_{t-1} + \sum_{i=1}^{K} \rho_i \triangle y_{t-1} + \varepsilon_t \tag{1}$$

Model B:
$$\triangle y_t = \mu + \gamma y_{t-1} + \sum_{i=1}^{K} \rho_i \triangle y_{t-1} + \varepsilon_t \tag{2}$$

Model C:
$$\triangle y_t = \mu + \beta t + \gamma y_{t-1} + \sum_{i=1}^{K} \rho_i \triangle y_{t-1} + \varepsilon_t \tag{3}$$

The determination of the order of integration of the individual time series is based on the zero hypothesis: $H_0: \gamma = 0$, which states that a time series contains a unit root, i.e. that the non-systematic component of time series is type of I(1). An alternative hypothesis is placed against the zero hypothesis: $H_1: \gamma < 0$, which states that a time series is stationary.

If the levels values of the series are stationary we should use vector autoregressive model and if the differences of the variable are stationary we test co-integration. With regard to the results we use the vector autoregressive model (VAR model) or vector error correction model (VECM model) and determine the Granger causality connections. The VECM belongs to the group of VAR models and is a special form that can take into account co-integration relationships between variables. The VECM model is based on a co-integration approach that models non-stationary time series the long-term relation of which is expressed through the error correction mechanism.

Impulse-response analysis allows analysis of both the short-term and long-term relations between the analysed variables based on the derived model. It is stated that the impulse-response analysis is related to the question of what reaction in one-time series will be caused by an impulse in another time series within a system that contains multiple time series. This is the study of the relation between two one-dimensional time series in a multidimensional system.

3 Econometric model

3.1 The probability of default

In this paper it used a model that suggested Wilson [23] and the credit risk is modelled through a probability of default (PD) and macroeconomic variables (it is assumed that PD is affected by the business cycle).

PD is defined as follows:

$$p_t = \frac{1}{1 + e^{-y_t}}, \tag{4}$$

where p_t is the probability of the default ICT sector at time t expressed by the default rate y_t. This transformation captures non-linear relationships between the default rate and macroeconomic variables. By solving equation (4) we obtain:

$$y_t = ln\left(\frac{p_t}{1-p_t}\right). \tag{5}$$

Index y_t is defined as follows:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \varepsilon_t, \tag{6}$$

where $x_t = (x_{1t}, x_{2t}, \dots, x_{kt})$ is a vector of exogenous macroeconomic variables in time t, $\varepsilon_t \sim N(0, \sigma_\varepsilon)$.

3.2 VAR/VECM model

This part of article deals with the estimation and testing of the default rate model depending on the selected economic indicators listed in Table 1. First, a stationarity of VAR model variables is tested and the order of the VAR model is determined. As the second step, the co-integration relationships for the VAR(p) model are tested using the Johansen's method and a number of co-integration relationships is determined. There is to estimate a VECM(p) model. The third part deals with model diagnostics. The methods which are presented below we can find in the articles Hendry and Juselius [12, 13].

The preparatory phase of estimating the VAR model is testing the stationarity of variables included in the model or their first differences. The test results for all variables are provided in Table 2. The Dickey-Fuller test (ADF) was used to test the stationarity. The second column provides information on the model type of testing the unit root (n = no trend and level constants /c = constant /c+t = level constant and trend), the third column contains the calculated T-statistics; the following column contains the corresponding level of statistical significance. The last column includes the result of testing: N = non-stationary (H_0 not rejected), S = stationary (H_0 rejected). The testing was performed at level 0.05.

Variable	n/c/c+t	T-stat	Signif.	Result	Variable	n/c/c+t	T-stat	Signif.	Result
Y	c+t	–0.001	0.995	N	*D(Y)*	c	–6.337	0.000	S
GDP	c+t	–2.317	0.418	N	*D(GDP)*	c	–3.435	0.014	S
CPI	c+t	–2.019	0.577	N	*D(CPI)*	c	–3.184	0.026	S
M2	c+t	–3.156	0.104	N	*D(M2)*	n	–1.272	0.185	N
ER	c+t	–2.605	0.279	N	*D(ER)*	c	–7.707	0.000	S
BRENT	c	–2.517	0.116	N	*D(BRENT)*	c	–7.329	0.000	S
UN	c	–0.928	0.945	N	*D(UN)*	c	–3.105	0.032	S

Table 2 Testing the unit root of the variables in levels and their first differences

If the time series are non-stationary and co-integrated, the Vector Error Correction Model (VECM) can be used to examine the relationship of variables. VECM is the VAR model in the first differences complemented by the co-integration residue vector.

The variables (*Y*, *GDP, CPI, ER, BRENT, UN*) for the VAR model exhibit the properties of first-order non-stationarity, i.e. I(1). All variables except *M2* contain a unit root, therefore the long-run co-integration relationships may exist between these variables. Therefore, a Johansen co-integration test with a null hypothesis was performed. This test assumes that there is no co-integration relationship between the variables tested. This test works with a test statistic based on a co-integration matrix track and a test statistic based on the eigenvalue value of the same matrix. Prior to determining the number of co-integration relationships, causal relationships with all used macroeconomic variables with different delay length settings were tested for the *Y* index. Variables with statistically insignificant coefficients in relation to default were not included in the resulting model.

The length of the delay is equal to 2. A statistically significant relationship was found between *Y* and the variables *GDP*, *CPI*, *BRENT*. One co-integration vector was determined by Johansen's test. The estimation led to a co-integration equation (7), where all coefficients are statistically significant at the significance level of 0.01.

$$Y = -12.17 \cdot GDP + 15.51 \cdot CPI - 1.87 \cdot BRENT \tag{7}$$

Growth of the *GDP* and *BRENT* variables reduces the probability of the default, and the growth of the *CPI* variable increases the probability of the default. The effect of the exchange rate is not statistically significant. In long-term relationships, one-point *GDP* growth will cause *Y* to fall by 12 points, one-point growth in oil prices will cause *Y* to fall by 2 points. CPI growth will increase *Y* by 15 points. The following formula applies to the complex shape of the VEC model:

$$\begin{aligned} D(Y) = {} & -0.11^{*} \cdot [Y(-1) + 12.17^{***} \cdot GDP(-1) - 15.51^{***} \cdot CPI(-1) + 1.87^{***} \cdot BRENT(-1)\,] + \\ & + 0.21^{*} \cdot D\big(Y(-2)\big) - 3.31^{**} \cdot D\big(GDP(-2)\big), \end{aligned} \tag{8}$$

where the statistical significance at the 0.01 level (***), at the 0.05 level (**), at the 0.1 level (*).

Based on the resulting shape of the adjustment vector, it can be argued that the complete elimination of short-term imbalances will take about 9 quarters. *BRENT* growth causes decrease in the default rate in the long run, which is in contradiction with the stated assumption. In the case of short-term effects, GDP growth is acting to reduce the default, and this effect will take effect after two quarters.

This model meets the assumptions: the residual component is not correlated; residual component heteroscedasticity and residual component non-normality were not demonstrated. These results are described in Table 3.

	Autocorrelation	**Heteroscedasticity**	**Normality**
Null hypothesis	H_0: absence of autocorrelation	H_0: absence of heteroscedasticity	H_0: normality of residues
Test	Ljung – Box	ARCH – LM	Doornik – Hansen
Significance	0.623	0.527	0.392

Table 3 The assumptions of the model

4 Conclusion

The aim of this article was to find out the influence of macroeconomic variables on the default rate in the ICT sector and to find out which variables are in long-term respectively short-term relationship. In the model, the default rate was considered as a dependent variable, defined as the ratio of outstanding and total loans in the sector over the period from 2005 to 2019 at quarterly frequency. The default rate was transformed into a macroeconomic index by means of a logit transformation. The macroeconomic index gains a larger range of values compared to the default rate. Furthermore, the Vector Error Correction model was used to capture both short and long term relationships.

The co-integration equation shows that the default rate is negatively affected in the long term by *GDP* and the variables *BRENT*. *BRENT* growth causes decrease in the default rate in the long run, which is in contradiction with the stated assumption. This result may be due to the fact that higher fuel prices discourage consumers from visiting foreign destinations and households invest, for example, in ICT equipment and thus support businesses in the sector information and communication activities in the Czech Republic. The positive relationship between the default rate and variable *CPI* is in line with the stated assumption. In the case of short-term effects, *GDP* growth is acting to reduce the default, and this effect will take effect after two quarters.

H1: There is a long-term relationship between the macroeconomic indicators and the default rate in the sector information and communication activities in the Czech Republic. The results confirmed that the long-term relationship is between the default rate and *GDP* and *BRENT.* H2: There is a short-term relationship between the macroeconomic indicators and the default rate in the sector information and communication activities in the Czech Republic. The short-term relationship was confirmed between the default rate and *GDP*.

By comparing the results with the article [17], we can see the results of the influence of individual macroeconomic variables on the probability of default in the sector accommodation and food service activities and in the sector transportation and storage. The results confirm the existence relationships between macroeconomic variables and the probability of defaults within both sectors. For the sector accommodation and

food service activities the long-term relationship was confirmed between the default rate and *GD*, *BRENT*. For the sector transportation and storage, the long-term relationship was confirmed between the default rate and *GDP*, *ER* and *CPI*.

Acknowledgements

This research was supported by the Ministry of Education, Youth and Sports Czech Republic within the Institutional Support for Long-term Development of a Research Organization in 2020.

References

[1] ALLEN, L. & SAUNDERS, A. (2003). A Survey of Cyclical Effects in Credit Risk Measurement Models. *BIS Working Papers,* 126.
[2] ALLESSANDRI, P. & GAI, P. (2009). Towards a Framework for Quantifying Systemic Stability. In S. KAPADIA, N. MORA & C. PUHR (Eds.), *International Journal of Central Banking*, 5(3), 47–81.
[3] BLASCHKE, W. & JONES, M. T. (2001). Stress Testing of Financial Systems: An Overview of Issues, Methodologies and FSAP Experiences. In G. MAJNONI & S. MARTINEZ PERIA (Eds.), *IMF Working Paper*, 6(88).
[4] CHAN-LAU, J. A. (2006). Fundamentals-Based Estimation of Default Probabilities: A Survey. *IMF Working Paper*, 6(149).
[5] CNB. (2020). ARAD – systém časových řad. Retrieved March 1, 2020, from http://www.cnb.cz/docs/ARADY/HTML/index.htm.
[6] DREHMANN, M. & QUAGLIARIELLO, M. (2009). Macroeconomic stress testing banks: A survey of methodologies. *Stress testing the banking system: Methodologies and applications*. Cambridge: Cambridge University Press.
[7] ENDERS, W. (2014). *Applied econometric time series*. 4th Edition. John Wiley, New York.
[8] ENGLE, R. F., & GRANGER, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55 (2), 251–276.
[9] FERRARI, S. & ROY, P. V. (2011). Stress Testing Credit Risk: Modelling Issues. In C. VESPRO (Eds.), *Financial Stability Review*, 9(1), s. 105–120.
[10] FOGLIA, A. (2009). Stress Testing Credit Risk: A Survey of Authorities Approaches. *International Journal of Central Banking*, 1(5), pp. 9–45.
[11] HAMERLE, A. & DARTSCH, A. (2011). Integrating macroeconomic risk factors into credit portfolio models. In R. JOBST & P. KILIAN (Eds.), *The Journal of Risk Model Validation*, 5(2), 3–24.
[12] HENDRY, D. & JUSELIUS, K. (2000). Explaining Cointegration Analysis: Part I. *The Energy Journal,* 21(1), 1–42.
[13] HENDRY, D. & JUSELIUS, K. (2001). Explaining Cointegration Analysis: Part II. *The Energy Journal,* 22(1), 75–120.
[14] JAKUBÍK, P. (2007). Macroeconomic Environment Credit Risk. *Czech Journal of Economics and Finance*, 57 (1–2), 60–78.
[15] JAKUBÍK, P. & REININGER, T. (2013). Determinants of Nonperforming Loans in Central, Eastern and Southeastern Europe, *Focus on European Economic Integration*, (3), 48–66.
[16] JARROW, R. & TURNBULL, S. (1995). Pricing Options on Financial Securities Subject to Default Risk. *Journal of Finance*, 50(1), s. 53–86.
[17] JEŘÁBEK, T. & MARTINOVIČOVÁ, D. (2017). Influence of Macroeconomic Variables Development on Corporate Default Risk in Selected Sectors. In V. ŠEFČÍK (Eds.), *Socioekonomické a humanitní studie,* 1(7), 59–78.
[18] JOHANSEN, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
[19] MARCUCCI, J., & QUAGLIARIELLO, M. (2008). Is Bank Portfolio Riskiness Procyclical? Evidence from Italy Using a Vector Autoregression. *Journal of International Financial Markets, Institutions and Money,* 18(1), 46–63.
[20] NACE. (2020). Retrieved March 1, 2020, from http://www.nace.cz/J-informacni-komunikacni-cinnosti.
[21] STOKLASOVÁ, R. (2018). Default rate in the Czech Republic depending on selected macroeconomic indicators. *E&M Economics and Management*, 21(2), 69–82.
[22] VIROLAINEN, K. (2004). Macro Stress Testing with a Macroeconomic Credit Risk Model for Finland. *Bank of Finland Discussion Paper,* 18/2004, 1–48.
[23] WILSON, T. C. (1997). Portfolio Credit Risk (II). *Risk Magazine*, 10(10), 56–61.

Modeling FX Rate with a Novel Heavy Tail Distribution

Jaromír Kukal[1], Quang Van Tran[2]

Abstract. An appropriate model of the distribution of returns of exchange rates is important pricing, risk management decisions, and portfolio allocations. Several models have been applied to them so far. However, the results have not been totally satisfactory. To amend the current state, we propose the use of a novel distribution for modeling the returns of exchange rates. This distribution originates from generalized gamma distribution which is regularized and symmetrized to make it fit for return modeling. After these adjustment, the distribution is smooth, differentiable and defined in the whole real domain. It can also capture the well-known heavy tail property of financial asset returns. The suitability of this model is verified on the returns of USD/EUR exchange rate. Verification is performed in an innovative way so that the results can be usable in practical financial engineering. The results of the testing show a promising applicability perspective of this novel distribution.

Keywords: generalized gamma distribution, regularization, symmetrization, USD/EUR exchange rate.

JEL Classification: C13, C46
AMS Classification: 37M10

1 Introduction

The returns of final assets often display a common property: leptokurticity of their distribution. Hence, the gaussian distribution is not suitable for their modeling. There have been several attempts to capture them with generalized hyperbolic distribution [6] or skewed generalized t-distribution [10] or others [9], but so far without sufficient success. As a proper model of their distribution is important for pricing, risk management decisions, and portfolio allocations, we propose a new alternative for this purpose. The novel distribution for modeling the returns of exchange rates comes from generalized gamma distribution. Since generalized gamma distribution is defined only for positive values, it must be is regularized and symmetrized so that it can be used for modeling returns of financial assets. Being modified, the distribution is smooth, differentiable and defined in the whole real domain and it it also is able to capture the well-known heavy tail property of financial asset returns. To test the usability of this novel distribution, we use a relatively long series of daily returns of USD/EUR exchange rate from 1/1999 to 6/2019. First, we use data to estimate parameters of the novel distribution. Unlike in many other studies which often use the whole dataset for estimation, we divide dataset into many smaller equidistant subintervals with window length ranging from roughly a half of a year to two years. The aim is to find a suitable length of history for proper estimation procedure as well as the possible extension of history into the future. After estimating parameters of the distribution these values are used to test for whether other intervals have the same distribution by Kolmogorov–Smirnov test. This approach would allow more practical usage of the novel distribution in financial engineering.

2 A Novel Model for Heavy Tail Distribution

We derive our novel distribution from the generalized gamma distribution [13], which has a wide range of applications in various areas [11]. The standardized probability density function (PDF) of this distribution is

$$\mathrm{f}(x,a,b) = \frac{bx^{a-1}\exp(-x^b)}{\Gamma(a/b)} \tag{1}$$

and its cumulative distribution function (CDF) is

$$\mathrm{F}(x,a,b) = 1 - \frac{\Gamma(a/b, x^b)}{\Gamma(a/b)}, \tag{2}$$

[1] Czech Technical University in Prague, Faculty of Nuclear Sciences and Physical Engineering, Trojanova 13, Praha 2, 120 00, Czech Republic, jaromir.kukal@fjfi.cvut.cz
[2] Czech Technical University/Faculty of Nuclear Sciences and Physical Engineering, Trojanova 13, Praha 2, Czech Republic, tranvqua@fjfi.cvut.cz

where

$$\Gamma(p,\xi) = \int_{\xi}^{\infty} t^{p-1}\exp(-t)\mathrm{d}t, \quad \Gamma(p) = \Gamma(p,0), \tag{3}$$

is the so called incomplete gamma function [2].

The moments of the generalized gamma distribution[1] are finite

$$\mathrm{E}X^k = \frac{\Gamma((a+k)/b)}{\Gamma(a/b)} < +\infty \tag{4}$$

for $k \in \mathbb{N}_+$. The second useful property is the occurrence of heavy tail for $0 < b < 1$ when for all $\lambda > 0$ holds

$$\lim_{x\to+\infty} \exp(\lambda x)(1-\mathrm{F}(x,a,b)) = +\infty. \tag{5}$$

This property makes generalized gamma distribution a heavy tail distribution[2]. The generalized gamma distribution also has an unwanted property:

- $\lim\limits_{x\to 0^+} \mathrm{f}(x,a,b) < +\infty$ only for $a \geq 1$,
- $\lim\limits_{x\to 0^+} \dfrac{\partial \mathrm{f}(x,a,b)}{\partial x} = 0$ only for either $a > 2$ or $a = 1 \wedge b > 1$,

This property makes the distribution for a certain set of values of its parameters either undefined or unsmooth when it is symmetrized for the whole domain. We eliminate it by regularizing the generalized gamma distribution. The regularization is completed by replacing part of the standardized distribution with a parabolic polynomial in the vicinity of zero.

Let $s > 0$ be the regularization parameter satisfying $a < 3 + bs^b$. The PDF of the regularized distribution is:

$$\mathrm{f}(x,a,b,s) = \begin{cases} (P+Qx^2)/R & \text{for} \quad 0 \leq x \leq s, \\ x^{a-1}\exp(-x^b)/R & \text{for} \quad x > s, \end{cases} \tag{6}$$

where

$$\begin{aligned} \mathcal{D} &= \{(a,b,s) \in \mathbb{R}_+^3 : a < 3 + bs^b\}, \\ P &= (3-a+bs^b)s^{a-1}\exp(-s^b)/2, \\ Q &= (a-1-bs^b)s^{a-3}\exp(-s^b)/2, \\ R &= Ps + Qs^3/3 + \Gamma(a/b, s^b)/b. \end{aligned} \tag{7}$$

Condition (7) ensures that the PDF is continuous and smooth. The resulting CDF of the novel distribution is

$$\mathrm{F}(x,a,b,s) = \begin{cases} (Px+Qx^3/3)/R & \text{for} \quad 0 \leq x \leq s, \\ 1-\Gamma(a/b, x^b)/(bR) & \text{for} \quad x > s. \end{cases} \tag{8}$$

The moments of random variable Y distributed by (6) are

$$\mathrm{E}X^k = \frac{Ps^{k+1}/(k+1) + Qs^{k+3}/(k+3) + \Gamma((a+k)/b, s^b)/b}{Ps + Qs^3/3 + \Gamma(a/b, s^b)/b} < +\infty \tag{9}$$

for $k \in \mathbb{N}_+$. The heavy tail defined by of (5) is preserved for $0 < b < 1$. The mode $\hat{X} = 0$ not only for $0 < a \leq 1$ but also for $a > 1 \wedge s \geq \left(\dfrac{a-1}{b}\right)^{1/b}$. Also, $\lim\limits_{x\to 0^+} \mathrm{f}(x,a,b,s) = \mathrm{f}(0,a,b,s) = P \in [0,+\infty)$ and $\lim\limits_{x\to 0^+} \dfrac{\partial \mathrm{f}(x,a,b,s)}{\partial x} = 0$.

The symmetrization of one-sided regularized generalized gamma distribution is performed as follows. Let $\mathcal{D}$ be a convex open set and $\mu \in \mathbb{R}, \sigma > 0, (s,a,b) \in \mathcal{D}$ be the mean value, the scaling factor, and the shape parameters of a random variable X following the two-sided regularized generalized gamma distribution, respectively. Then, its PDF can be expressed for $x \in \mathbb{R}$ as

$$\mathrm{g}(x,\mu,\sigma,a,b,s) = \frac{\mathrm{f}\left(\frac{|x-\mu|}{\sigma}, a, b, s\right)}{2\sigma}, \tag{10}$$

[1] The generalized gamma distribution is also known as Amoroso distribution [1].

[2] The proof of proposition (5) can be intuitively derived from the asymptotic property of incomplete gamma function and its logarithmic transform.

where $\mathrm{f}(x, a, b, s)$ is the PDF of random variable distributed according to (6).
The corresponding CDF of X is

$$\mathrm{G}(x, \mu, \sigma, a, b, s) = \frac{1 + \mathrm{sign}(x - \mu) \cdot \mathrm{F}\left(\frac{|x-\mu|}{\sigma}, a, b, s\right)}{2}, \tag{11}$$

where $\mathrm{F}(x, a, b, s)$ is the CDF defined in (8).

The first moment and the median of a random variable distributed according to (10) is μ. Their absolute central moments are

$$\mathrm{E}|X - \mu|^k = \sigma^k \cdot \frac{Ps^{k+1}/(k+1) + Qs^{k+3}/(k+3) + \Gamma((a+k)/b, s^b)/b}{Ps + Qs^3/3 + \Gamma(a/b, s^b)/b} < +\infty \tag{12}$$

for $k \in \mathbb{N}_+$ and even. However, if k is odd, the central moments are equal to zero. Therefore, its skewness is zero, its variance is finite as well as its kurtosis which approaches $+\infty$ when $b \to 0_+$. In Figure 1 we show three examples of symmetrization and regularization. The upper left panel is the case of $a = 0.5$ and $b = 2$. The upper right panel is the case of $a = 1$ and $b = 1.5$ and the bottom panel is the case $a = 2.5$ and $b = 1.5$.

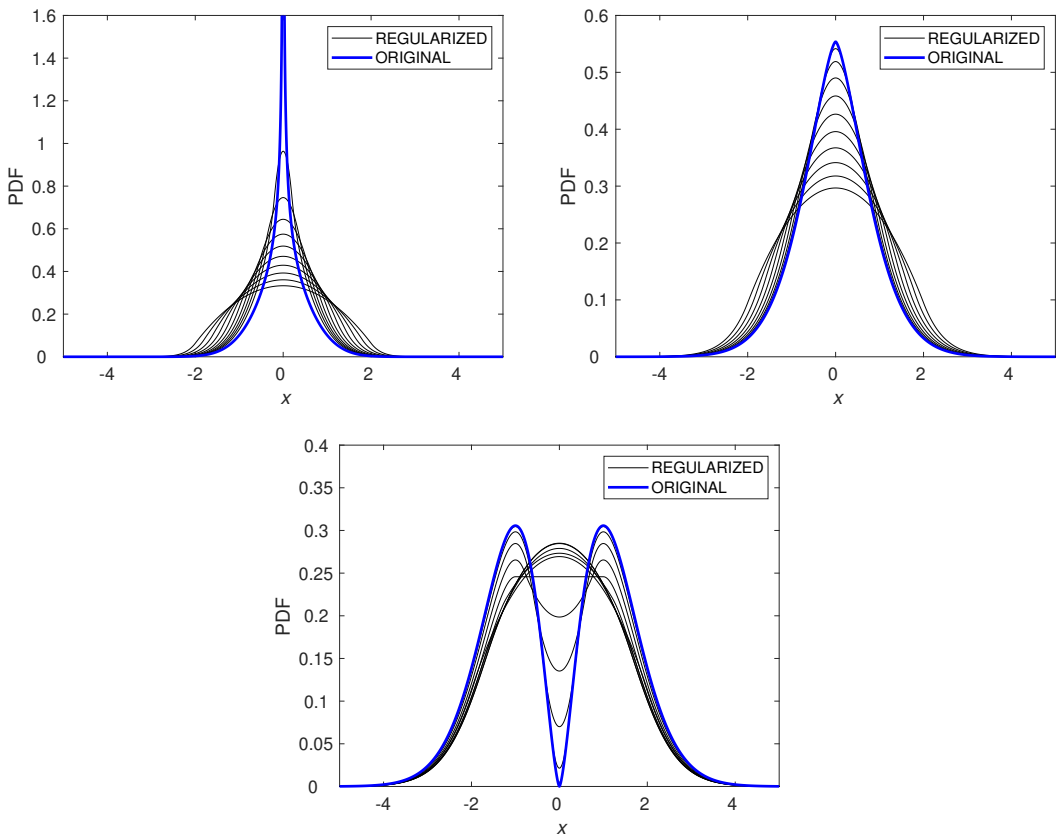


Figure 1 Regularization and Symmetrization in Examples

The novel distribution has many special cases when its parameters attain specific values, see [5] and their family tree could be approximated according to [4]. As our primary focus is the general case, the studies of special cases is out of touch of our interest in this research.

3 Experimental Verification

In this part of the research we verify the applicability of the novel distribution on return series of EUR/USD exchange rate. Data for verification is a series of daily close EUR/USD exchange rate from period 1/1999 to 6/2019. The original exchange rate series is converted into a series of logarithmic returns. The descriptive statistics of both series are shown in Table 1. The statistics show that logarithmic returns of EUR/USD exchange rate exhibit higher fourth moment than the one of a corresponding normal distribution.

Characteristic	Original series	Return series
Mean	1.2050	$-7.23 \cdot 10^{-3}$
Median	1.2198	$7.33 \cdot 10^{-3}$
Mode	1.1235	0
Minimum	0.8255	-0.0294
Maximum	1.1590	0.0394
Std deviation	0.1671	0.0059
Skewness	-0.2324	0.1452
Kurtosis	2.2465	5.1804
Obs.	5341	5340

Table 1 Descriptive Statistics of Exchange Rate and Logarithmic Return Series

The relatively long history of exchange rate return series is divided into equi-distant intervals with length $W = 125, 250,$ and 500, which roughly correspond to a history of six months, one year, and two years, respectively. We use F-test to verify the possibility that all intervals come from the same distribution with similar variance [7]. As it is a multiple hypothesis testing problem, we control the false discovery rate (FDR) by using the procedure proposed by Benjamini and Hochberg [3]. The results are displayed in Table 2. In the table column 2 shows the number of intervals passing the FDR test while column 3 shows the number of intervals passing F-test with $\alpha \geq 0.05$. The results imply that not all subperiods originate from the same distribution.

W	Total number	number of stat	number with $\alpha \geq 0.05$	α_{FDR}
125	42	38	34	$9.91 \cdot 10^{-4}$
250	21	17	13	0.0013
500	10	4	4	$7.54 \cdot 10^{-4}$

Table 2 Analysis of Variance Results for Sub-periods at Different Length

Parameter	Value	Std. error	z-score
a	0.8810	0.1040	-1.1442
b	1.6278	0.2380	-1.5639
$\log_{10} s$	-6.3548	1.5192	–
μ	$5.88 \cdot 10^{-4}$	$6.46 \cdot 10^{-8}$	9102.2
$\log_{10} \sigma$	-1.9425	0.0822	–

Table 3 Estimated Values of Parameters of the New Distribution, $W = 125$

As the novel distribution is smooth and differentiable, its parameters can be estimated from data using maximum likelihood estimation (MLE) method. For each window length, a subinterval from the middle of the dataset is selected and parameters of the distribution are estimated for the chosen subsample. As regularization parameter s and scale parameter σ are required to be positive and they can be small in their magnitude, the $\log_{10} s$ and $\log_{10} \sigma$ are estimated instead. This substitution secures their positiveness without imposing any additional condition on these two parameters. The numerical procedure for estimation is performed in Matlab. It is relatively stable and the stability increases with the growing size of the window. The estimation results with statistics resulting from with their asymptotic properties [15] are shown in Tables 3–5.

In tables 3–5 the estimated values of a and b are tested for the null hypotheses $a = 1$ and $b = 2$, which are parameter values of special cases of the novel distribution, while μ is tested for the null hypothesis $\mu = 0$[3]. The estimated values of parameters of the distribution for all window lengths are relatively close to each other. The reason may be the fact that the shorter interval is a subinterval of the longer one. According to the test result, the estimated values for two shorter periods may belong to a special case of the general form with $a = 1$ and $b = 2$. However, with more observations, $W = 500$, it is not the case. We tend to opt for the

[3] There is no reason to test $\log_{10} s$ and $\log_{10} \sigma$ against any specific value, hence the result is left blank.

Parameter	Value	Std. error	z-score
a	0.8761	0.1344	-0.9219
b	1.7010	0.3858	-0.7750
$\log_{10} s$	-5.5040	0.5311	–
μ	$-7.09 \cdot 10^{-4}$	$2.71 \cdot 10^{-8}$	$2.62 \cdot 10^{4}$
$\log_{10} \sigma$	-2.1432	0.0960	–

Table 4 Estimated Values of Parameters of the New Distribution, $W = 250$

Parameter	Value	Std. error	z-score
a	0.8632	0.0616	-2.2208
b	1.7217	0.1131	-2.4607
$\log_{10} s$	-6.0602	0.9746	–
μ	$3.27 \cdot 10^{-4}$	$9,81 \cdot 10^{-8}$	3333.3
$\log_{10} \sigma$	-2.2038	0.0558	–

Table 5 Estimated Values of Parameters of the New Distribution, $W = 500$

latter. Also, the estimation results show that as $b > 1$, the distribution of EUR/USD exchange rate returns may not have so heavy tails. It is similar to the case displayed in upper left panel in Figure 1.

We use the estimation results to examine whether other subintervals of the dataset also come from the same distribution as the one chosen for parameter estimation. For this purpose, we use the nonparametric Kolmogorov–Smirnov goodness of fit test [8]. The test statistic is

$$S = \max_{0 \le i \le N} \left(F(x_i) - \frac{i-1}{N}, \frac{i}{N} - F(x_i) \right), \tag{13}$$

where x_i is the i-th order observation, N is the number of observations and $F(x_i)$ is the value of theoretical CDF at x_i. Basically, the test statistic measures the maximum distance between theoretical CDF and its empirical counterpart. In order to perform the test, not only do we have to compute the test statistic, but we also must find the critical value of the test. To achieve it, we generate ten thousand samples with parameters identical with the estimated values and the window lengths as the interval accordingly. Then, the test statistics of these samples are computed and they are sorted into a ascending order and the 95 percentile value is extracted for critical value at level $\alpha = 0.05$.

W	Critical value	Number of not rejected
125	0.1193	14
250	0.0850	7
500	0.0605	4

Table 6 The Kolmogorov–Smirnov Goodness of Fit Test Results

The test results are displayed in Table 6. One can observe that the lower the number of observations is, the higher is the critical value which is intuitively reasonable. Further, if the number of observations in a subinterval is low, e.i., the history is short [12], then even though F-test indicates that all intervals are from the same distribution, the Kolmogorov–Smirnov test results show that only fewer than half of the cases are confirmed. It implies that passing F-test does not ensure that the interval will have the same distribution as the benchmark interval and the chosen history lengths may be insufficient for being a benchmark window. However, if the length of the window is sufficiently long, which is the case $W = 500$, then the Kolmogorov–Smirnov test confirms the results of the F-test [14]. The number of subintervals with the same distribution by Kolmogorov–Smirnov test is identical as the number of subperiods passing F-test. This means that our novel distribution can model the logarithmic returns of EUR/USD exchange rate and with a two-year history. The distribution then can be used to model other periods of the data series, that is the distribution can be applied for out-of-sample prediction if no dramatic changes are expected.

4 Conclusion

We propose a novel alternative to already known distributions with heavy tails for modeling returns of financial assets. This distribution is a result of modification of the generalized gamma distribution. Since the generalized gamma distribution is a one-sided distribution for positive values and it also has specific property at the left end, the generalization is the regularization and symmetrization of the original distribution. After modification the distribution is smooth, differentiable and defined in the whole real domain. It also can capture the well-known heavy tail property of financial asset returns. The verify the applicability of this distribution on real data, we choose daily a series of returns of USD/EUR exchange rate from 1/1999 to 6/2019. First, we use data to estimate parameters of the novel distribution. We divide dataset into many smaller equidistant subintervals with window length ranging from roughly a half of a year to two years. For one chosen subperiod with a given length, the parameters of the distribution are estimated by MLE. The remaining periods are tested for whether other intervals have the same distribution by Kolmogorov–Smirnov test. The results we obtained show that our distribution can be used to model returns of exchange rate and the optimal length of history used for estimation and out-of-sample verification should be two years for daily data, both for numerical stability and distributional stability. If no out-of-magnitude changes occur, the nature of distribution will not change in the future.

Acknowledgements

This paper is financially supported by grant SGS20/190/OHK4/3T/14.

References

[1] Amoroso, L. Ricerche intorno alla curva dei redditi. *Annali di Matematica Pura ed Applicata* 1925, 2, 123–159 (1925).

[2] Askey, R., Roy, R. Gamma function. In Olver, F. W. J., Lozier, D. M., Boisvert, R. F.; Clark, C. W. (eds.): NIST Handbook of Mathematical Functions. Cambridge: Cambridge University Press (2010).

[3] Benjamini, Y., Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, Series B. 57 (1): 289–300 (1995).

[4] Di Battista, G., Tamassia, R. Algorithms for plane representation of acyclic digraphs. *Theoretical Computer Science*, 61 (2–3), 175–178 (1988).

[5] Crooks, G. E. The Amoroso Distribution. *arXiv* 2010, arXiv:1005.3274 (2010).

[6] Foss, S., Konstantopoulos, T., Zachary, S. Discrete and Continuous Time Modulated Random Walks with Heavy-Tailed Increments. *Journal of Theoretical Probability* 20(3): 581–612 (2007).

[7] Huber-Carol, C., Balakrishnan, N., Nikulin, M. S., Mesbah, M. Goodness-of-Fit Tests and Model Validity. *Springer Springer Science & Business Media*, (2012).

[8] Kleiber, C., Kotz, S. *Statistical size distributions in economics and actuarial sciences* Vol. 470. John Wiley & Sons, (2003).

[9] Kukal, J., Tran, Q. V. Voigt Distribution and Its Heavy-tail Modeling Ability for Cryptocurrencies. In: Houda, M., Remes, R. (editors): Proceedings of 37th International Conference on Mathematical Methods in Economics 2019, University of South Bohemia in České Budějovice, 139–144 (2019).

[10] Kyriazis, N. A., Daskalou, K., Arampatzis, M., Prassa, P., Papaioannou, E. Estimating the volatility of cryptocurrencies during bearish markets by employing GARCH models. *Heliyon*, 5(8), e02239 (2019).

[11] Pham, T., Almanha, J. The Generalized Gamma Distribution: Its Hazard Rate and Stress-Strength Model. *IEEE transactionson reliability* 44(3), (1995)

[12] Rossi, R. J. *Mathematical Statistics: An Introduction to Likelihood Based Inference*. New York: John Wiley & Sons (2018).

[13] Stacy, E. W. A Generalization of the Gamma Distribution. *The Annals of Mathematical Statistics*, 33(3), 1187–1192 (1962).

[14] Wilks, S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9, 60–62 (1938).

[15] Woodridge, J. M. *Econometric Analysis of Cross Section and Panel Data* 2e. Cambridge: MIT Press (2010).

Influence of the Different Transformation of the Minimization Criteria on the Result – the Case of WSA, TOPSIS and ARAS Methods

Martina Kuncová[1], Jana Sekničková[2]

Abstract. Several multi-criteria evaluation of alternatives methods can be used to find the best alternative or the order of alternatives according to the utility maximization (such as WSA or ARAS methods) or the minimization of the distance from ideal solution (such as TOPSIS method). These methods differ in the calculation steps but also in the formulas of data normalization or in the rules for changing the minimization criteria into maximization. For some of them all criteria must be of the maximization type before analyses and calculations. But the transformation of the minimization criteria to the maximization ones can affect the resulting order of alternatives. This article describes the influence of different rules of minimization criteria transformation on the results. Three multi-criteria evaluation of alternatives methods are analysed: WSA, TOPSIS and ARAS. As an example, the evaluation of the Czech regions according to eight economic criteria in the years 2017 and 2018 is used.

Keywords: multi-criteria evaluation of alternatives, WSA, TOPSIS, ARAS, minimization criteria

JEL Classification: C44, C61
AMS Classification: 90B50, 90C29

1 Introduction

Multi-criteria decision making (MCDM) as a part of operations research is used in a situation when it is necessary to find the best or the compromise solution or to understand reality better. MCDM usually covers the multi-criteria evaluation of alternatives methods, sometimes called multi-criteria decision analysis (MCDA), but it can also cover multi-criteria linear programming or other decision-making theories and concepts such as group decision-making or data envelopment analysis (DEA). The main aim of the MCDA as the discrete decision-making problem could be to find the good (or the best) alternatives according to several selected criteria [4]. Based on the type of the problem or the aim of the decision, several different multi-criteria decision-making methods can be used – a lot of them are described in [1], [4], [6] and their usage in [10]. In this article we analyse WSA, TOPSIS and ARAS methods. WSA and TOPSIS were used, for example, for the comparison of economic activity of the EU countries [3], or the evaluation of the Czech regions [9] and in many other comparisons. ARAS method is not so common as it belongs to the newer ones [14]. It was used for the ranking of companies [7], determining the location of new student admission [12] or for the measuring of the quality of websites [13]. In this paper we compare results taken from WSA, TOPSIS and ARAS methods when different rules for the transformation of the minimization criterion into maximization are used. As an example, the economic criteria for 14 Czech regions for the years 2017 and 2018 are used. This problem was also analysed in [5] with the TOPSIS method. It is obvious that any changes in data can influence the MCDA results more or less, but this article is aimed only at the problem of "min to max" transformations. There are several procedures for this transformation, but we focus on the most common ones used in the 3 mentioned methods. The aim is primarily to analyse the impact of the transformation on the order of alternatives in different methods. As no "best" generally possible order of alternatives can be determined, the correlation of orders will be monitored in particular.

[1] University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 13067, Prague 3, kuncovam@vse.cz
[2] University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 13067, Prague 3, jana.seknickova@vse.cz

2 Methods and data

In this part, the basic description of WSA, TOPSIS and ARAS methods as representatives of MCDA is presented. Data for the analysis were taken from the Czech statistical office [2]. They describe the selected economic characteristics of the 14 Czech regions – similar data were used in the comparison in [8].

MCDA methods have been developed to help the decision-maker to find the best alternative or the ranking of alternatives [6]. The main areas of the methods' usage can be found at [10]. As these methods use different principles and different information, including subjective information, they do not have the same results. To solve this kind of model, it is necessary to know the preferences of the decision maker when all the alternatives ($a_1, a_2, ..., a_p$) and criteria ($f_1, f_2, ..., f_k$) are known. These preferences can be described by aspiration levels (or requirements), criteria order, or by the weights of the criteria (usually stated as v_j for $j = 1, \ldots, k$) [4]. In this article, the methods with criteria weights are analyzed.

2.1 WSA method

The Weighted Sum Approach method (WSA) belongs to the utility maximization type of methods. Criteria can be minimized or maximized as for WSA two formulas could be applied for the data normalization – formula (1) for maximization type and (2) for minimization criteria type. For each criterion f_j symbol A_j^+ denotes the highest value of this criterion, $A_j^+ = \max_i y_{ij}$, and A_j^- denotes the lowest value of this criterion, $A_j^- = \min_i y_{ij}$. According to the data y_{ij} for each alternative a_i and each criterion f_j the normalized values r_{ij} are calculated [1]:

$$r_{ij} = \frac{y_{ij} - A_j^-}{A_j^+ - A_j^-} \tag{1}$$

$$r_{ij} = \frac{A_j^+ - y_{ij}}{A_j^+ - A_j^-} \tag{2}$$

The final ranking is based on the utility – the higher is the better: $u(a_i) = \sum_{j=1}^{k} v_j r_{ij}, \ \forall\, i = 1, \cdots, p$.

2.2 TOPSIS method

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method ranks the alternatives using the relative index of distance of the alternatives from the ideal and non-ideal (basal) alternative [5]. The required input information includes decision matrix **Y** with the information about all selected alternatives $a_1, \ldots, a_p$ according to all criteria $f_1, \ldots, f_k$ and weight vector **v** of these criteria. The steps of this method based on [5] are:

- normalize the decision matrix (types of all criteria are maximization) according to Euclidean metric:

$$r_{ij} = \frac{y_{ij}}{\sqrt{\sum_{i=1}^{p} y_{ij}^2}}, \quad \forall\ i = 1,\ \ldots,\ p, \quad j = 1,\ \ldots,\ k, \tag{3}$$

- calculate the weighted decision matrix $\mathbf{W} = (w_{ij}) = v_j \cdot r_{ij}$, and from the weighted decision matrix **W** identify vectors of the hypothetical ideal A_j^+ and basal A_j^- alternatives over each criterion

$$A_j^+ = \max_i\ w_{ij}, \quad \forall\ j = 1,\ \ldots,\ k, \quad \text{and} \quad A_j^- = \min_i\ w_{ij}, \quad \forall\ j = 1,\ \ldots,\ k, \tag{4}$$

- measure the Euclidean distance of every alternative a_i to the ideal and to the basal alternatives over each attribute:

$$d_i^+ = \sqrt{\sum_{j=1}^{k} (w_{ij} - A_j^+)^2} \quad \text{and} \quad d_i^- = \sqrt{\sum_{j=1}^{k} (w_{ij} - A_j^-)^2}, \quad \forall\ i = 1,\ \ldots,\ p, \tag{5}$$

- for all alternatives determine the relative ratio of its distance to the basal alternative:

$$c_i = \frac{d_i^-}{d_i^+ + d_i^-}, \quad \forall\ i = 1,\ \ldots,\ p, \tag{6}$$

- rank order alternatives by maximizing ratio c_i.

Usually if some criteria are of the minimization (cost) type the values are changed into maximization (benefit) by calculation the differences from the worst value. But this transformation can influence the results of the TOPSIS method [5]. That is why the multiplication by -1 or the inverse values should be better – as analyzed in this article.

2.3 ARAS method

The Additive Ratio Assessment (ARAS) method belongs to the set of methods of the utility maximization. This method was first introduced by Zavadskas and Turskis in 2010 [14]. It is also mentioned in the book of Alinezhad and Khalili [1], but in this book the authors have a mistake as they forgot to mention the first step – changing the minimization criteria type into maximization. That is why we describe the steps according to Zavadskas and Turskis [14], using the symbols described above. The first step is the transformation of the minimization criteria into maximization:

$$y'_{ij} = \frac{1}{y_{ij}}, \tag{7}$$

where y_{ij} denotes values of original minimization criterion and y'_{ij} transformed values of corresponding maximization criterion. Afterwards the decision matrix is extended by a new ideal alternative a_0 for which all the values are taken as maximum values from each criterion, $y_{0j} = A_j^+, \forall\ j = 1,\ \dots,\ k$. The next step is the normalization:

$$r_{ij} = \frac{y_{ij}}{\sum_{i=0}^{p} y_{ij}}, \quad \forall\ i = 0,\ \dots,\ p, \quad j = 1,\ \dots,\ k. \tag{8}$$

Given the weights v_j of criteria f_j the weighted normalized values of each alternative are obtained $w_{ij} = r_{ij} \cdot v_j$. The following step is to determine the value S_i of the optimality function of i-th alternative:

$$S_i = \sum_{j=1}^{k} w_{ij}, \forall\ i = 0, \dots, p. \tag{9}$$

Finally, the utility for the alternative a_i is calculated (10) and based on it the order is created (the higher the better).

$$u(a_i) = \frac{S_i}{S_0}. \tag{10}$$

2.4 Data used for the analysis

In our study we compared 14 Czech regions with respect to available data from 2017 and 2018 using MCDA methods (no. 1 – Prague, the Capital City, no. 2 – Central Bohemian Region, no. 3 – South Bohemian Region, no. 4 – Plzeň Region, no. 5 – Karlovy Vary Region, no. 6 – Ústí nad Labem Region, no. 7 – Liberec Region, no. 8 – Hradec Králové Region, no. 9 – Pardubice Region, no. 10 – Vysočina Region, no. 11 – South Moravian Region, no. 12 – Olomouc Region, no. 13 – Moravian-Silesian Region and no. 14 – Zlín Region). The models have 8 criteria, 3 should be minimized (unemployment rate, average age and free workplaces per capita) and 5 should be maximized (economic activity, average wage, income per capita, consumption per capita and investments per capita). This data was described in [9], another comparison in the years 2012–2018 via PROMETHEE method is in [8].

2.5 Ways of data transformation from minimization to maximization

In the multi-criteria decision problem two types of criteria exist: maximization type (sometimes called benefit) where the higher values are preferred and minimization type (cost type) with the preference of lower values. Some MCDM methods require all the criteria to be maximized before the calculation starts. In general, there are several ways how to transform the minimization criterion type into maximization. The methods mentioned above use 3 of them. The first (T1) converts the minimization criterion f to maximization f' using the relation $f' = -f$. According to [6] this way is more suitable in TOPSIS method than the second way (T2) which assumes that negative values could cause problems in some methods. Therefore, it is necessary to add a suitable constant to the given criterion. A suitable constant can be, for example, the maximum (i.e. worst) value of the criteria f and f' then can be interpreted as savings over the worst alternative.

The transformation way T2 uses the formula $f' = \max - f$. This formula is commonly used for the WSA or TOPSIS method [6]. The third option is using of inverted values of criterion f and transforming the minimization criterion via the formula T3: $f' = \frac{1}{f}$. In the ARAS method, T3 transformation type is used [14]. At first glance, it may seem that the type of transformation relationship may affect the results obtained by the respective MCDM method, and this article aims to confirm this hypothesis. Due to the comparability of the criteria, most methods normalize the criterion values, i.e. transformation of the original criterion values into the $\langle 0, 1\rangle$ scale. The first way (N1) is the normalization of the ideal and non-ideal (basal, nadir) alternative that is applied in WSA – formula (1). When the T1 transformation for the minimization criteria is used, the formula (1) is only changed to formula (2) but afterwards the situation is the same – the best (minimal) value has the normalized value equal to 1 and the worst (highest) value is changed into 0. The same is valid for the transformation type T2. In the transformation T3 where $y'_{ij} = \frac{1}{y_{ij}}$, $A'^{+}_j = \frac{1}{A^-_j}$, $A'^{-}_j = \frac{1}{A^+_j}$ is $r_{ij} = \frac{y'_{ij} - A'^{-}_j}{A'^{+}_j - A'^{-}_j} = \frac{\frac{1}{y_{ij}} - \frac{1}{A^+_j}}{\frac{1}{A^-_j} - \frac{1}{A^+_j}}$. Again, it holds that for the ideal value A^-_j is $r_{ij} = 1$ and for the non-ideal value A^+_j is $r_{ij} = 0$.

The second way (N2) is the normalization against the sum of the criterion values – with respect to the distances in the linear metric. This principle is used, for example, by the ARAS method with the normalization function (7). In case of the transformation T1 usage the normalized values should be calculated as $r_{ij} = \frac{y'_{ij}}{\sum_{i=1}^{p} y'_{ij}} = \frac{y_{ij}}{\sum_{i=1}^{p} y_{ij}}$. In this case, however, the higher value of y_{ij} (for the minimization criterion, the worse value) would have a higher normalized value (meaning that it is better). Such normalization is unacceptable and is necessary at least in the case of minimization the formula (7) change as follows: $r_{ij} = \frac{y'_{ij}}{\sum_{i=1}^{p} |y'_{ij}|} = \frac{-y_{ij}}{\sum_{i=1}^{p} |y_{ij}|}$. For the transformations T2 and T3 (which is taken as basic for the ARAS method) the formula (7) can be used and it is still true that the best value will have the highest normalized value and the worst value will be normalized as worst.

The third way of normalization (N3) is used in TOPSIS method as described in formula (3). When the T1 transformation is used, the values $r_{ij} = \frac{y'_{ij}}{\sqrt{\sum_{i=1}^{p} {y'_{ij}}^2}} = \frac{-y_{ij}}{\sqrt{\sum_{i=1}^{p} (-y_{ij})^2}} = \frac{-y_{ij}}{\sqrt{\sum_{i=1}^{p} {y_{ij}}^2}}$. It is true that these values are negative or non-positive (so not from the interval $\langle 0, 1\rangle$) but still the lower y_{ij} means the higher r_{ij}. The same is true for the T2 transformation where the r_{ij} values will be in the interval $\langle 0, 1\rangle$ as $r_{ij} = \frac{y'_{ij}}{\sqrt{\sum_{i=1}^{p} {y'_{ij}}^2}} = \frac{A^+_j - y_{ij}}{\sqrt{\sum_{i=1}^{p} (A^+_j - y_{ij})^2}}$. In the T3 transformation the normalized values r_{ij} are decreasing with increasing of y_{ij} as $r_{ij} = \frac{y'_{ij}}{\sqrt{\sum_{i=1}^{p} {y'_{ij}}^2}} = \frac{\frac{1}{y_{ij}}}{\sqrt{\sum_{i=1}^{p} (\frac{1}{y_{ij}})^2}} = \frac{\frac{1}{y_{ij}}}{\sqrt{\sum_{i=1}^{p} \frac{1}{{y_{ij}}^2}}} = \frac{1}{y_{ij} \cdot \sqrt{\sum_{i=1}^{p} \frac{1}{{y_{ij}}^2}}}$.

3 Results

In our analysis, we used three MCDA methods (WSA, TOPSIS and ARAS) and applied three different approaches (T1, T2, and T3) of data transformation from minimization to maximization for each of them. These nine approaches were applied to regional data from 2017 and 2018. The results for the year 2017 are shown in Table 1.

The obtained data were used for a statistical test of the hypothesis that the method of converting minimization criteria to maximization ones can significantly affect the obtained results (null hypothesis: independence of results, type of data transformation is significant; against alternative hypothesis: dependence of results, type of data transformation is insignificant). This hypothesis was tested by Pearson correlation test [11] assuming normality distribution. Pearson's correlation coefficient measures the linear relationship between two variables. In our analysis we study the correlation between score (utility or minimal distance) as well as between the ranking of alternatives. All correlation coefficients were positive and so we tested the positive relationship between approaches.

2017	*WSA T1, T2*	*WSA T3*	*TOPSIS T1*	*TOPSIS T2*	*TOPSIS T3*	*ARAS T1*	*ARAS T2*	*ARAS T3*
Prague	0.818 (1)	0.799 (1)	0.674 (1)	0.628 (2)	0.671 (1)	0.757 (1)	0.802 (1)	0.899 (1)
Central Bohemian	0.556 (2)	0.499 (2)	0.441 (6)	0.646 (1)	0.354 (7)	0.418 (4)	0.765 (2)	0.710 (4)
South Bohemian	0.398 (5)	0.338 (6)	0.462 (4)	0.425 (5)	0.373 (4)	0.413 (5)	0.605 (4)	0.705 (5)
Plzeň	0.382 (6)	0.357 (4)	0.380 (11)	0.366 (10)	0.359 (6)	0.306 (8)	0.547 (10)	0.695 (6)
Karlovy Vary	0.302 (9)	0.237 (9)	0.360 (12)	0.300 (14)	0.224 (12)	0.264 (10)	0.507 (11)	0.638 (12)
Ústí nad Labem	0.267 (11)	0.226 (13)	0.409 (9)	0.490 (3)	0.325 (11)	0.264 (11)	0.605 (5)	0.648 (11)
Liberec	0.287 (10)	0.228 (12)	0.351 (13)	0.433 (4)	0.217 (13)	0.238 (12)	0.577 (7)	0.630 (13)
Hradec Králové	0.367 (7)	0.314 (7)	0.478 (3)	0.381 (9)	0.384 (3)	0.426 (3)	0.562 (9)	0.711 (3)
Pardubice	0.260 (12)	0.229 (11)	0.272 (14)	0.334 (12)	0.216 (14)	0.141 (14)	0.505 (12)	0.623 (14)
Vysočina	0.424 (3)	0.369 (3)	0.499 (2)	0.422 (6)	0.418 (2)	0.474 (2)	0.614 (3)	0.731 (2)
South Moravian	0.399 (4)	0.349 (5)	0.435 (7)	0.420 (7)	0.345 (9)	0.371 (6)	0.598 (6)	0.690 (7)
Olomouc	0.317 (8)	0.268 (8)	0.455 (5)	0.383 (8)	0.366 (5)	0.370 (7)	0.564 (8)	0.690 (8)
Moravian-Silesian	0.235 (14)	0.234 (10)	0.387 (10)	0.339 (11)	0.352 (8)	0.214 (13)	0.496 (13)	0.655 (10)
Zlín	0.255 (13)	0.213 (14)	0.413 (8)	0.317 (13)	0.326 (10)	0.297 (9)	0.494 (14)	0.661 (9)

Table 1 Results of all three methods and all three transformations (final utility or score and rank)

3.1 The results of methods

In principle, WSA does not require any data transformation, but we analysed it too, concerning the other methods. The results obtained by WSA were identically the same for T1 a T2 transformation (correlation coefficient 100%), as we assumed. Correlation between results of these two transformations and the third one (T3) is 99.32% for score (utility) and 92.97% for rank in 2017 (resp. 98.94% and 95.60% in 2018). We can conclude the type of transformation does not significantly affect the results obtained by the WSA method (level of significance 1%).

We obtained similar results for the ARAS method, although, for this method, the data transformation is necessary. The correlation coefficients for the utility (score) are high. At the 5% level of significance, we can conclude the type of transformation is not significant in any case. Note that at the 1% level of significance, the resulting ranks obtained by ARAS using T2 in 2018 are not statistically correlated with the results gained with T1 or T3 (correlation coefficients are about 55%, resp. 51%). Other results are correlated at the 1% level.

The different results were found for TOPSIS method, as we expected. At the 5% level of significance, the results (score, minimal distance) are independent on type of transformation but at 1% level of significance, the correlation between T2 and T3 is too low (60% in 2017 and 47% in 2018) to reject the independence hypothesis. The correlation coefficients for the ranking are much more remarkable. Classical TOPSIS method (with T2 transformation) provides statistically different ranking than the TOPSIS with T3 transformation (correlation about 36% in 2017 and 26% in 2018) at 10% level and at 1% level of significance the results in 2018 are uncorrelated also with T1 transformation (62% in 2017 and 47% in 2018). In contrast, the correlations between T1 and T3 (90.77% in 2017 and 84.62% in 2018) are statistically significant, even at the 1% level. These results support the theory in [5] that, especially with the TOPSIS method, the results can be significantly dependent on the transformation used.

3.2 The results of transformations

In our paper we also focus on the hypothesis that results (score or ranking) are dependent only on the type of transformation rather than on the method used. Firstly, we compare the results obtained after T1 transformation. All three methods gave results (score or ranking) that did not differ statistically significantly at the 5% level. Also, at the 1% level of significance, the hypothesis of independence is rejected (correlation greater than 63%), except in the case of the ranking in 2018 where the results of WSA and TOPSIS differ (correlation about 53%).

The second transformation (T2) provides similar results. All three methods gave results (score or ranking) that did not differ statistically significantly at the 5% level. Also, at the 1% level of significance, the hypothesis of independence is rejected (correlation greater than 62%), except in the case of the ranking in 2018 where the results of WSA and TOPSIS differ (correlation about 55%).

Similar results are given after the third transformation (T3). All three methods gave results (score or ranking) that did not differ statistically significantly at the 5% level. Also, at the 1% level of significance, the hypothesis of independence is rejected (correlation greater than 68%), except in the case of the ranking in 2018 where the results of WSA and TOPSIS differ (correlation about 45%).

From the results, we can accept the hypothesis that the results depend on the transformation rather than on the method used. If the same transformation is used, results obtained by different methods are statistically correlated.

3.3 The impact of transformations on the Czech region rankings

In this paper we showed that the type of transformation could significantly affect the rank of the alternative. See, for example, Ústí nad Labem Region in Table 1. The WSA method (with classical transformation T2) and the ARAS method (with classical transformation T3) place this region at 11th place, while by the TOPSIS method (with classical transformation T2) this region is placed at the 3rd place in 2017. Similarly, this region in 2018 is by WSA the 14th (the last one), by ARAS the 13th but by TOPSIS is the 5th. To be the fifth or the last is a very significant difference, as well as the 3rd or 11th. We can analyse original data for Ústí nad Labem region (for example in 2018). In 6 criteria it is very bad (less than 12% of ideal value; 4 of them less than 2.5%), in one, it is around average (56% of ideal value) and only in one, it is good (93% of ideal value). This only one value moves the region by TOPSIS to the forefront. The fifth place is probably unfair in this case. The rank of the region may appear to be determined by the method. But if we change transformation, the rank of region will not be so variable. In 2017 using T1 transformation the ranks are 11, 9, 11 and using T3 the ranks are 13, 11, 11. In 2018 using T1 the ranks are 14, 11, 13 and using T3 the ranks are 14, 9, 13.

Note also that inappropriate transformation T2 in TOPSIS method can cause not only improvement (as in the case of Ústí nad Labem Region) but also a significant decrease of position – see Hradec Králové Region in 2018. It is the 7th by WSA and the 6th by ARAS but the 12th by TOPSIS. After using T1 transformation, it will be 7th by WSA, 6th by TOPSIS and 4th by ARAS and after T3 it will be 8th, 8th and 6th. The ranking after the same transformation used is more similar. So the ranking of alternatives is influenced not only by original data and method used but also by the type of criteria transformation. Let's focus once again on the results of the classical TOPSIS method. In 2017, Prague, the Capital City did not finish as the winner in only one case – in the TOPSIS with T2. Transformations T1 and T3 place half of the alternatives in the same place and the order according to T2 differs significantly in these cases (in the Liberec Region it is by 9 places better, in the Hradec Králové Region by 6 places worse). In these cases, we use the same weights, the same method and only difference is in the type of transformation.

4 Conclusions

From the results and notes above, it is visible that the type of transformation of minimizing criteria to maximizing is more significant than the method used. The difference between TOPSIS and other MCDA methods (e.g. WSA and ARAS) seems to be given not only by the principle of a method (especially type of normalization) but also by the type of transformation the minimal criteria types into the maximal ones. There are numbers of papers presented that the type of transformation in classical TOPSIS method (T2) is not suitable and the results after T1 transformation are better [5] and we came to the same conclusion in the analysis. T2 transformation is not proper also for the ARAS method. Although only one example was tested, we showed that via different ways of transforming of the minimization criteria, it was possible to improve the position of an alternative. While the WSA method is not very sensitive to the way of transformation, TOPSIS method is really sensitive, especially concerning the resulting order. To conclude: if an analysis is done using more MCDM methods with the aim to compare the results of the methods, it makes sense to transform the minimization criteria to maximization in the same way.

Acknowledgements

The research was supported by the grant No. 19-08985S of the Czech Science Foundation and also by the Internal Grant Agency of the University of Economics in Prague IGA F4/66/2019.

References

[1] ALINEZHAD, A. & KHALILI, J. (2019). *New Methods and Applications in Multiple Attribute Decision Making (MADM)*. Springer Nature Switzerland AG.

[2] CZECH STATISTICAL OFFICE. (2020) [online], available at https://www.czso.cz/csu/czso/ [Accessed 17 Jan 2020].
[3] DINÇER, S. E. (2011). Multi-criteria Analysis of Economic Activity for European Union Member States and Candidate Countries: TOPSIS and WSA Applications. *European Journal of Social Sciences*, 21(4), 563–572.
[4] FIGUEIRA, J., GRECO, S., & EHRGOTT M. (2005). *Multiple Criteria Decision Analysis – State of the Art Surveys.* New York : Springer Science + Business Media Inc.
[5] HOUSKA, M., DOMEOVA, L., & BERANKOVA, M. (2012). The conversion of cost and benefit criteria in the TOPSIS method. *International Journal of Applied Decision Sciences*, 5 (4), 329–341.
[6] ISHIZAKA, A. & NEMERY, P. (2013). *Multi-Criteria Decision Analysis*. UK: John Wiley & Sons, Ltd.
[7] KARABASEVIC, D., PAUNKOVIC, J., & STANUJKIC, D. (2016). Ranking of Companies According to the Indicators of Corporate Social Responsibility Based on SWARA and ARAS Methods. *Serbian Journal of Management* 11 (1), 43–53.
[8] KUNCOVÁ, M. & SEKNIČKOVÁ, J. (2020). Visual PROMETHEE Usage for the Multicriteria Evaluation of the Czech Regions from the Selected Economic Activity Aspects Point of View. *Quantitative Methods in Economics* (Multiple Criteria Decision Making XXI), Púchov. Fortcoming.
[9] KUNCOVA, M. & SEKNICKOVA, J. (2018). Multicriteria Evaluation of the Czech Regions from the Selected Economic Activity Aspects Point of View. *Proceedings of the International Scientific Conference Quantitative Methods in Economics* (Multiple Criteria Decision Making XIX) 2018 (pp. 193–200), Tren. Teplice.
[10] MARDANI, A., JUSOH, A., NOR, K. M. D., KHALIFAH, Z., ZAKWAN, N., & VALIPOUR, A. (2015). Multiple criteria decision-making techniques and their applications – a review of the literature from 2000 to 2014. *Economic Research – Ekonomska Istraživanja*, 28(1), 516–571.
[11] PEARSON, K. (1895). Notes on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London*, 58, 240–242.
[12] PUNGKASANTI, P. T., HENNY INDRIYAWATI, H., & SUSANTO (2019). Application of the ARAS Method. In: Problem Completion of Determining The Location of New Student Admission. *Int. Journal of Information Technology and Business*, 2 (1), 16–20.
[13] STANUJKIC, D. & JOVANOVIC, R. (2012). Measuring a Quality of Faculty Website Using ARAS Method. *Contemporary Issues in Business, Management and Education*'12, 545–554.
[14] ZAVADSKAS, E. K. & TURSKIS, Z. (2010). A new additive ratio assessment (ARAS) method in multicriteria decision-making, *Technological and Economic Development of Economy*, 16(2), 159–172.

Financial Health Assessment of Pig-breeding Companies in Slovakia – Altman's Z-score and TOPSIS Results Comparison

Martina Kuncová[1], Roman Fiala[2], Veronika Hedija[3]

Abstract. This paper aims to investigate the financial health of firms belonging to the pig-breeding sector in Slovakia using two different methods which are suitable for financial health assessment and to compare the results of these methods. The research sample contained 32 pig-breeding companies (group 01.4.6 according to Statistical classification of economic activities in the European Community – NACE). These firms were assessed for the year 2017. Given the procedures used to evaluate financial health, 11 criteria were selected to focus on profitability, liquidity, activity and leverage. For measuring financial health Altman Z´ score model and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) were used. The data were obtained from Bisnode Albertina database. It was revealed that about half of the companies in the research group are financially healthy; the other half are not. Both methods applied for the analysis provide similar results.

Keywords: multi-criteria comparison, TOPSIS, Altman Z-score, pig-breeding companies, bankruptcy models

JEL Classification: C44, G33, Q14
AMS Classification: 90B50, 91B28

1 Introduction

The pig-breeding sector belongs to the traditional and very significant sector of agricultural animal production in Slovakia. Consumption of pork in Slovakia in the last ten years has been at the level of 30–32 kg per citizen. It is more or less stable consumption, but it must also be said that in the not so distant past, about 20–25 years ago, its consumption was over 40 kg. At the beginning of the 1990s, Slovakia was 100% self-sufficient; in 2017, pig-breeding companies in Slovakia were able to cover 36.6% of consumption [5]. The consumption of pork in Slovakia in 2016 was 35.9 kilograms per person. In 2017, this value was 35.4 kilograms [5].

This fact is certainly related to the integration of European economies and deepening globalization. This industry, like other sectors, faces increasing competition. However, the sector has still its specificities, thanks to the common agricultural policy of the European Union associated with agricultural subsidy policy and to its high dependency on the natural environment. Nevertheless, the managerial and economic logic and the importance of strategic management are comparable to other businesses [8], [14].

Competitive pressure in the European market has increased even more after the introduction of Russian sanctions in August 2014. Of the Visegrad group countries, Polish and Hungarian pig-breeding companies were most affected by Russian sanctions. The economic recession, which started in 2008, affected most of the Visegrad Group pig-breeding countries (mainly the Czech Republic, Slovakia and Hungary) [6].

There are studies which focused on the Altman model application in Slovakia or for all Visegrad group countries. Rybárová et al. [18] applied the model when tested the improving condition of the construction industry in Slovakia. Boďa and Úradníček [1] was investigated the usability of Altman Z Score model in the Slovak economic environment. All countries of the Visegrad Group were mentioned in articles of the authors Režňáková and Karas [17] or Misankova et al. [15]. I addition to Altman's Z-Score model, Misankova et al. [15] investigated Czech IN05, Polish Poznanski's model, Hungarian Virag and Hajda's model, Canadian Springate's model, and the UK Taffler's model.

[1] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 586 01 Jihlava, Czech Republic, kuncova@vspj.cz
[2] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 586 01 Jihlava, Czech Republic, roman.fiala@vspj.cz
[3] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 586 01 Jihlava, Czech Republic, veronika.hedija@vspj.cz

Altman Z-score model was investigated to estimate the financial efficiency of companies [2]. It usually uses four or five different criteria. From the mathematical point of view, it is a kind of multi-criteria evaluation where the final score is taken as a kind of weighted sum (but in this case weights are not taken from the 0–1 scale) with the aim of maximization of the score (the higher score means the better position of the company). According to this fact, we decided to use TOPSIS method as a typical representative of the multi-criteria evaluation of alternatives methods [7] to see whether it is possible to find the same or similar categories as from the Altman Z-score model. This method was used to compare the financial health of the Czech tour operators [9] or the financial health of the pig-breeding companies in the Czech Republic in 2013 [13] and that is why the same methodology is applied here for the pig-breeding companies' financial health in Slovakia.

There are many technics which are used to evaluate firm performance. Yang et al. [23] summarized several research techniques for performance measurement, including graphical tools), integrated performance indices (for instance analytic hierarchy process – AHP), statistical methods, or data envelopment analysis. Other authors used for measurement of organizational effectiveness multiple-criteria evaluation of alternatives methods as TOPSIS or VIKOR. In [18], the authors focused on the evaluation of the financial performance of manufacturing firms in Turkey with the help of TOPSIS and VIKOR (VlseKriterijumska Optimizacija I Kompromisno. Resenje) methods. Companies in the Taiwan Stock Market were evaluated with the help of four financial ratios: the inventory turnover, net income ratio, earnings per share and current ratio using TOPSIS method [20].

The aim of this article is to examine the financial health of firms belonging to the pig-breeding sector in Slovakia using two different methods which are suitable for financial health assessment and to compare the results of these methods.

2 Data and methodology

2.1 Data

Data from the Bisnode Albertina database [5] are used to evaluate the financial health of selected companies. The Albertina database contains selected data of all firms to which an identification number has been assigned. Bisnode Albertina database currently contains information of more than 2.7 million companies. Data for 2017 were selected, namely those companies which as their primary activity according to the classification of economic activities (CZNACE) reported activities of Raising of swine/pigs (NACE 0146).

The research sample was further narrowed down to companies which achieved any sales and sales during for the year 2017 and which have all the necessary data for this year. Required information to assess financial health are included in the financial reports of companies. Database Albertina does not contain data from the financial statements of all companies, but only the companies that are registered in the public register and are thus obliged by law to establish financial statements to the collection of documents, as well as companies that voluntarily entered this data into the collection of documents. The final sample includes data from 32 companies.

2.2 Financial ratios

The financial health of firms is assessed using a set of financial indicators that are used to assess the financial health of the farm in the financial analysis as in [12] or **Chyba! Nenalezen zdroj odkazů.**. Eleven ratios are traditionally used in the financial analysis of the company from these four basic categories of indicators: liquidity ratios, profitability ratios, activity ratios and leverage ratios. Selected indicators are described in Table 1.

These indicators enter as criteria for multi-criteria evaluation with the designation Ki. Subsequently, the total benefits or relative distances from the ideal alternative are assigned, which will allow to rank companies and evaluate their financial health. The interpretation of individual authors differs regarding the recommended values of financial analysis indicators, especially for liquidity indicators and leverage. The determination of "optimal values" is based on the recommendations in [12] and **Chyba! Nenalezen zdroj odkazů.** (see Table 1). Before the calculations, it is necessary to change the data inputs for TOPSIS method as all criteria should be maximized.

For some criteria, it was necessary to modify the input data to be applicable to the TOPSIS method. Criteria K_1, K_2 and K_3 are maximizing, but to avoid negative values, all values were increased by the absolute value of the most negative indicator (i.e. the minimum gets to zero). Values of criterion K_4 are preferably higher

than one, so the maximization is correct (and we decided not to change the data lower than one as they are still taken as bad values). The main changes were necessary for criteria K_5 and K_6, which are not maximized nor minimized. According to the recommended values, we decided to change the input data as follows: if the criterion value is in the recommended area, the value is transformed to one. If the real value is lower than the recommended minimum, the y_{ij} is equal to the ratio of the real value and the minimal threshold. If the value is higher than the recommended maximum, then the y_{ij} is equal to the ration of the maximal threshold and the real value. Criteria K_7 and K_8 were multiplied by -1 to change the minimum into maximum. Criterion K_9 is maximizing, but again there are some companies with negative values, so to have all positive values, we added the constant of the most negative value to all. For the criterion K_{10} good values are those that are lower than 0.6. That is why the real values were changed – the y_{i10} is equal to 0.6/real value. The last K_{11} criterion is essentially a maximization type, but again as for the negative values a constant value (max. negative in absolute value) was added to each value of this criterion. These adjustments are not the only possible ones, it would certainly be interesting to test the effect of changes on the result, but this is not part of the analyses in this article.

Assessed area	Marking indicators	Financial indicators	Recommended values
Profitability	K_1	Return on Equity (ROE) = EAT / total equity	Maximum
	K_2	Return on Assets = earnings before interest and taxes (EBIT) / total assets	Maximum
	K_3	Return on Sales = EBIT / sales	Maximum
Liquidity	K_4	Current Ratio = Current Assets / Current Liabilities	Higher than 1
	K_5	Quick Ratio = Cash and Cash Equivalents / Current Liabilities	0,4 – 1,5
	K_6	Cash Position Ratio = Cash and Cash Equivalents / Current liabilities	0,2 – 0,5
Activity	K_7	Total Assets Turnover Period = total assets / (sales / 360)	Minimum
	K_8	Average Collection Period = short-term trade receivables / (sales / 360)	Minimum
	K_9	Creditors Payment Period – Average Collection Period	Maximum
Leverage	K_{10}	Total Debt to Total Assets = Total debt / Total assets	Lower than 0.6
	K_{11}	Times Interest Earned Ratio = EBIT / Total interest charges	Higher than 3

Table 1 Financial indicators (Source: [9], [12], **Chyba! Nenalezen zdroj odkazů.**)

For the TOPSIS usage, it is necessary to set the weights of the criteria. Based on the previous research we set the equal weights for each of the four parts (profitability, activity, liquidity, leverage) which means, that the weight for the criteria K_1–K_9 is equal (0.0833) and the weights for the criteria K_{10} and K_{11} are 0.125 per each.

2.3 TOPSIS

To evaluate the economic performance of companies, the method of multiple criteria evaluation of alternatives can be used. These methods are usually applied on the situation where a comparison of more alternatives according to several criteria is necessary in order to find out the best alternative or to create the order of them [11]. The preferences can be described by aspiration levels, criteria order or by the weight of the criteria [7].

The model of multi-criteria evaluation of alternatives contains a list of alternatives $A = \{a_1, a_2, \dots, a_p\}$, a list of criteria $F = \{f_1, f_2, \dots, f_k\}$ and an evaluation of the alternatives by each criterion [11].

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method is able to rank the alternatives using the relative index of distance of the alternatives from the ideal and basal alternative. Higher relative index of distance means better alternative. The user must supply only the information about the weights of criteria [7]. The steps of this method can be described as follows [11]:

- normalize the decision matrix according to Euclidean metric:

$$r_{ij} = \frac{y_{ij}}{\sqrt{\sum_{i=1}^{p} y_{ij}^2}}, \quad \forall\ i = 1,\ \dots,\ p, \quad j = 1,\ \dots,\ k, \tag{1}$$

- calculate the weighted decision matrix W = (w_{ij}) = $v_j \cdot r_{ij}$, and from the weighted decision matrix W identify vectors of the hypothetical ideal H and non-ideal (basal) D alternatives over each criterion, where $H_j = \max_i w_{ij}$ and $D_j = \min_i w_{ij}$.
- measure the Euclidean distance of every alternative to the ideal and to the basal alternatives over each attribute:

$$d_i^+ = \sqrt{\sum_{j=1}^{n} (w_{ij} - H_j)^2} \quad and \quad d_i^- = \sqrt{\sum_{j=1}^{n} (w_{ij} - D_j)^2}, \quad \forall\ i = 1,\ \dots,\ p, \tag{2}$$

- for all alternatives determine the relative ratio of its distance to the basal alternative:

$$c_i = \frac{d_i^-}{d_i^+ + d_i^-}, \quad \forall\ i = 1,\ \dots,\ p, \tag{3}$$

- rank order alternatives by maximizing ratio c_i.

All these steps assume that all criteria are of the maximization type. If not, it is necessary to transform the minimization criterion into maximization. Usually the difference from the worst case is used for transformation [11] but it was proved that the multiplication by -1 is better for TOPSIS method [10], that is why we use this kind of transformation.

2.4 Altman Z-Score model

From several Z-score models, we chose one that is suitable for both manufacturing and non-manufacturing firms and publicly listed and privately held firms [10]. This model is described as follows [10]:

$$Z'' = 6.56.X_1 + 3.26.X_2 + 6.72.X_3 + 1.05.X_4, \tag{4}$$

where X_1 is working capital/total assets, X_2 denotes retained earnings/total assets, X_3 is earnings before interest and taxes/total assets, X_4 represents book value of equity/book value of total liabilities. Here the firms are also classified into three zones based on the value of Z´´-score: Z´´ ≤ 1.1 is Distress zone; 1.1 < Z´´ ≤ 2.6 is Grey zone and Z´´ > 2.6 denotes Safe zone. For the purposes of this study, we will use the Z´´ model for the above-mentioned reasons.

Financial ratios (see Table 1) are divided into four areas – profitability, liquidity, activity and leverage. X_1 expresses the ratio of liquidity to total assets, and X_2 expresses cumulative profitability. We classify X_3 indicator in the area of profitability and X_4 represents leverage area. Z´´-score model therefore includes three of the four areas used in the TOPSIS method. Thus, both approaches do not work with diametrically different indicators, but they are different.

3 Results and discussion

For evaluation of financial health pig-breeding companies, Alman's Z-score (formula 4) model is first calculated. Table 2 contains the number of companies that are classified into one of the zones according to the model (distress zone, grey zone, safe zone).

	Z´´ score	
Distress zone (Z´´ ≤ 1.1)	15	(46.875%)
Grey zone (1.1 < Z´´ ≤ 2.6)	3	(9.375%)
Safe zone (Z´´ > 2.6)	14	(43.75%)
Sum	32	(100%)

Table 2 Results of Z´´-score model (Source: own computation)

It can be seen from Table 2 that the companies are divided into two equally large parts according to the Z-score results. The first part of the 15 companies is located in the safe zone and the companies can be described as financially sound. On the contrary, according to the results, 14 companies are located in the distress zone and the Z-score model evaluated them as financially unhealthy.

The TOPSIS method and, in general, the methods of multi-criteria evaluation of alternatives do not divide companies into different areas or zones. However, these methods can be used to determine the order of companies and rank them from best to worst according to the score obtained, which in this case evaluates financial health. The resulting score of the TOPSIS method can be in the interval ⟨0; 1⟩. The closer the value is to the number 1, the closer it is to the ideal variant and the more in this case the company is rated as financially healthier. Based on the calculation of the TOPSIS method, the point interval of 32 companies ranged from ⟨0.293; 0.686⟩. The median score is 0.489 and average score is 0.483. Scores higher than 0.5 were achieved by 11 companies, which is about 35% of the research sample. Both approaches identified the same best 5 companies (see Table 3), but the next order differed for some of them.

When we normalize the results into the same scale, it means into the interval $\langle 0, 1\rangle$ from the $\langle 0.293, 0.686\rangle$ for TOPSIS and $\langle -6.108, 19.718\rangle$ for the Altman Z-score, we see that the results are similar (Figure 1) but the Altman Z-score moves the financial unhealthy companies more to the worse position.

If we compare the results with the study [9], then the results of pig-breeding firms are much better in comparison with travel agencies operating in the Czech Republic. In this study [9], the evaluated companies achieved a median of only 0.095 and the average value was 0.103. Scores higher than 0.5 was achieved only 1 firm (0.43 % of research sample). On the other hand the results are worse than in the study [13] where from the 42 pig-breeding companies from the Czech Republic only 23% were under 0.5 normalized score in TOPSIS while in this study it is 65% of companies (but the analysed years differ so it is not possible to conclude that the Slovak companies are worse than the Czech ones).

Company name	TOPSIS score	rank	Altman Z-score	rank
JK Gabčíkovo, s.r.o.	0,686	1	19,718	1
Poľnohospodárske družstvo Trenčín-Soblahov	0,622	2	10,155	4
Agrovýkrm Rybany, s.r.o.	0,604	3	8,508	5
Zväz chovateľov ošípaných na Slovensku – družstvo	0,590	4	14,760	2
Roľnícke družstvo Turá Lúka v Myjave	0,537	5	11,287	3

Table 3 Best 5 companies according to TOPSIS and Altman (Source: own computation)

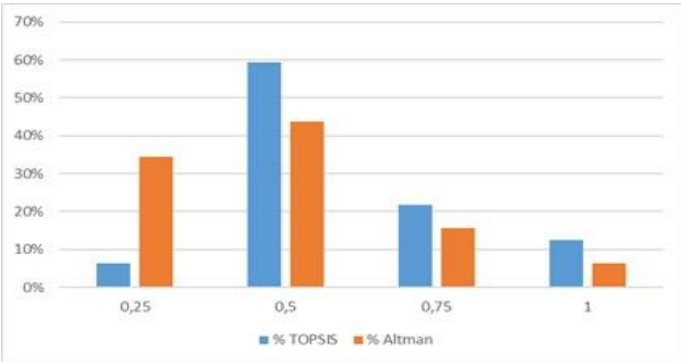


Figure 1 Normalised results of 32 companies in the TOPSIS and Altman Z-score methods (Source: own computation)

4 Conclusion

The aim of the paper was to investigate the financial health of pig-breeding firms in Slovakia using the Altman Z´´-score model and TOPSIS method and to compare the results of both methods.

Firstly, we examined the financial health of 32 firms which are suitable for financial health using Altman Z´´-score model. This model includes four financial ratios. The methodology of the Altman Z´´-score model divides companies into three areas – distress, grey and safe zone. According to this Altman Z´´-score model,

we can say that from the research sample of companies, about half of the companies are in a good financial situation and half of the firms are located in the distress zone, i.e. their financial situation is not good. There are only 3 companies in the grey zone. Then we evaluated the economic performance using TOPSIS method, which is able to rank the alternatives using the relative index of distance of the alternatives from the ideal and basal alternative. For the purpose of this method, we selected eleven indicators, which formed the input data of the multiple criteria evaluation of alternatives method. Both methods allowed each of the firm to assign one aggregate value, which made it possible to assess its financial health and determine the ranking within the research sample.

We found that both used method identified (with some exceptions) similar results, which shows that both methods identified the same top five companies. One of the topics for the further research may be an examination of relationship between a level of enterprises' innovativeness) and the level of enterprises' financial liquidity (see e. g. [5]) or to compare results of pig-breeding companies located in the Czech Republic and Slovakia.

Acknowledgements

The paper was supported by the contribution of long term institutional support of research activities by the College of Polytechnics Jihlava and by the Internal Grant of the College of Polytechnics Jihlava No. 1170/4/199.

References

[1] ALTMAN, E. I., IWANICZ-DROZDOWSKA, M., LAITINEN, E. K., & SUVAS. (2017). Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journal of International Financial Management & Accounting*, *28*, 131–171.

[2] ALTMAN, E. I. (1983). *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy*. New York: Wiley.

[3] BISNODE. [disk]. Data obtained from Bisnode company from Bisnode Albertina database. More information at https://www.bisnode.cz/produkty/albertina/

[4] BOĎA, M. & ÚRADNÍČEK, V. (2016). The portability of Altman's Z-score model to predicting corporate financial distress of Slovak companies. *Technological and Economic Development of Economy, **22***, 532–553.

[5] DOLEŠOVÁ, P. (2017). Aktuálna situácia v chove ošípaných v SR. Retrieved from https://www.rno.sk/aktualna-situacia-v-chove-osipanych-v-sr/

[6] FIALA, R., HEDIJA, V., DVOŘÁK, J., & JÁNSKÝ, J. (2020). Are profitable firms also financially healthy? Empirical evidence for pig-breeding sector. *Custos e Agronegocio On Line, **16***, 173–201.

[7] FIGUEIRA, J., GRECO, S., & EHRGOTT, M. (2005). *Multiple Criteria Decision Analysis – State of the Art Surveys*. New York: Springer Science + Business Media Inc.

[8] GUIMARÃES, V. P., TEDESCHI, L. O., & RODRIGUES, M. T. (2009). Development of a mathematical model to study the impacts of production and management policies on the herd dynamics and profitability of dairy goats. *Agricultural Systems, **101***, 186–196.

[9] HEDIJA, V. & KUNCOVÁ, M. (2018). Assessment of the financial health of the Czech tour operators. *Scientific Papers of the University of Pardubice, Series D, **42***, 53–66.

[10] HOUSKA, M., DOMEOVA, L., & BERANKOVA, M. (2012). The conversion of cost and benefit criteria in the TOPSIS method. *International Journal of Applied Decision Sciences*, 5 (4), 329–341.

[11] ISHIZAKA, A. & NEMERY, P. (2013). *Multi-Criteria Decision Analysis*. UK: John Wiley & Sons, Ltd.

[12] KNÁPKOVÁ, A., PAVELKOVÁ, D., REMEŠ, D., & ŠTEKER, D. (2017). *Finanční analýza: Komplexní průvodce s příklady*. Praha: Grada Publishing.

[13] KUNCOVÁ, M., FIALA, R., & HEDIJA, V. (2016). A Comparison of Specialised Agricultural Companies Performance. In M. REIFF & P. GEŽÍK (Eds.), *Proceedings of the International Conference: Quantitative Methods in Economics: Multiple Criteria Decision Making XVIII*. (pp. 209–2015). Vrátná: The Slovak Society for Operations Research, Department of Operations Research and Econometrics, Faculty of Economic Informatics, University of Economics in Bratislava.

[14] LOSOSOVÁ, J. & ZDĚNĚK, R. (2014) Key Factors Affecting the Profitability of Farms in the Czech Republic. *Agris On-line Papers in Economics and Informatics*, 6, 21–36.

[15] MISANKOVA, M., ZVARIKOVA, K., & KLIESTIKOVA, J. (2017). Bankruptcy Practice in Countries of Visegrad Four. *Economics and Culture, **14***, 108–118.

[16] OKRĘGLICKA, M. (2016). Internal Innovativeness and Management of Current Finances of Enterprises in Poland. In M. H. BILGIN, H. DANIS, E., DEMIR, E., & CAN, U. (Eds.), *Proceedings of the 14th Eurasia Business and Economics Society Conference* (pp. 225–237). Barcelona: Springer.
[17] REŽŇÁKOVÁ, M. & KARAS, M. (2015). The Prediction Capabilities of Bankruptcy Models in a Different Environment: An example of the Altman Model under the Conditions in the Visegrad Group Countries. *Ekonomický časopis, **63***, 617–633.
[18] RYBÁROVÁ, D., BRAUNOVÁ, M., & JANTOŠOVÁ, L. (2016). Analysis of the Construction Industry in the Slovak Republic by Bankruptcy Model. *Procedia – Social and Behavioral Sciences, **230***, 298–306.
[19] STATISTICAL OFFICE OF THE SLOVAK REPUBLIC (2018). Food consumption in the SR in 2017. Retrieved from https://www7.statistics.sk/PortalTraffic/fileServlet?Dokument=60b895b1-1a10-4eb4-85a8-6bb2829f335d
[20] WANG, T-C. & HSU, J. C. (2004). Evaluation of the Business Operation Performance of the Listing Companies by Applying TOPSIS Method. *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics*. (pp. 1286–1291). The Hague: IEEE.
[21] WEETMAN, P. (2006). Management Accounting. Harlow: Pearson Education Limited.
[22] YALCIN, N., BAYRAKDAROGLU, A., & KAHRAMAN, C. (2012). Application of fuzzy multi-criteria decision making methods for financial performance evaluation of Turkish manufacturing industries. *Expert Systems with Applications, **39***, 350–364.
[23] YANG, H., YEUNG, J. F. Y., CHAN, A. P. C., CHIANG, Y. H., & CHAN D. W. M. (2010). A critical review of performance measurement in construction. *Journal of Facilities Management, **8***, 269–284.

Volatility of American Stock Indexes During the Corona Crisis Period

Radoslav Lacko[1]

Abstract. During the post-crisis period the stock indexes beat their maximum several times. The huge and unexpected fall of index prices came with the coronavirus. This article investigated the volatility fluctuation during the market fall in March 2020. The correction after fall became only several weeks after the March bottom. This partial recovery on the stock markets was the result of massive injection by Federal Reserve System. Before the stock prices bounced off the bottom, volatility significantly lowered.

Keywords: American Stock Market, Coronavirus, Volatility

JEL Classification: C50
AMS Classification: 91B28

1 Introduction

The financial and then economic crisis in 2008 brought the biggest falls of American stock market in the modern history. The trigger was fall of the Lehman Brothers investment bank, which resulted in the bankruptcy of many well-known financial and insurance institutions. Economists could not find the right answers for questions about the future of the financial and stock market.

The central banks supported the economics with billions of fiat money – by quantitative easing or asset purchasing programs. This resulted in the longest period of continuous growth in the history, which was stopped by the Coronavirus pandemic. The initial reaction of markets was massive assets sale. Hence, governments and central banks promised the rescue worth trillion of dollars.

Situations described above are investigated and analysed in this article with the use of quantitative econometric methods. Firstly, the particular data are described. Followed by the theoretical background examination. Then the estimation of the best possible model will be done. We expect it will be the ARIMA methods, the basic technique for time series research. Then the residuals of the best estimated models will be excluded and used for further estimation with more sophisticated models, the GARCH family. Finally, our results are analysed and concluded. The aim of the article is to estimate volatility and evaluate its changes as a response to actions of central banks.

2 Data

As the topic says, the American stock indexes daily data are employed for the volatility investigation. We use three well-known indexes, namely data of NASDAQ Composite Index, S&P 500 and Dow Jones Industrial Index.

The examined period of one year starts on 2019/04/30 and ends on 2020/05/01. The daily data are obtained from Yahoo Finance website. The Figure 1 shows how indexes moved during the selected period. The line of all indexes have increasing trend up to February 2020, the Covid-19 pandemic have outbroken, when the fall came.

The Figure 2 presents daily returns of index prices as the result of the first difference of logarithm. This step also makes the dataset stationary and thus suitable for further processing.

3 Theoretical Framework

Our data features and the aim of this article encourage us to make the best possible estimation on our data and then save the residuals for ARCH family models estimation. According to our investigation, there are several studies that investigated the relationship between stock returns and the unexpected volatility. The first study we included in our theoretical framework was published in 1987 by the French et al. (1987) [4].

[1] University of Economics, Prague, Faculty of Informatics and Statistics, lacr02@vse.cz

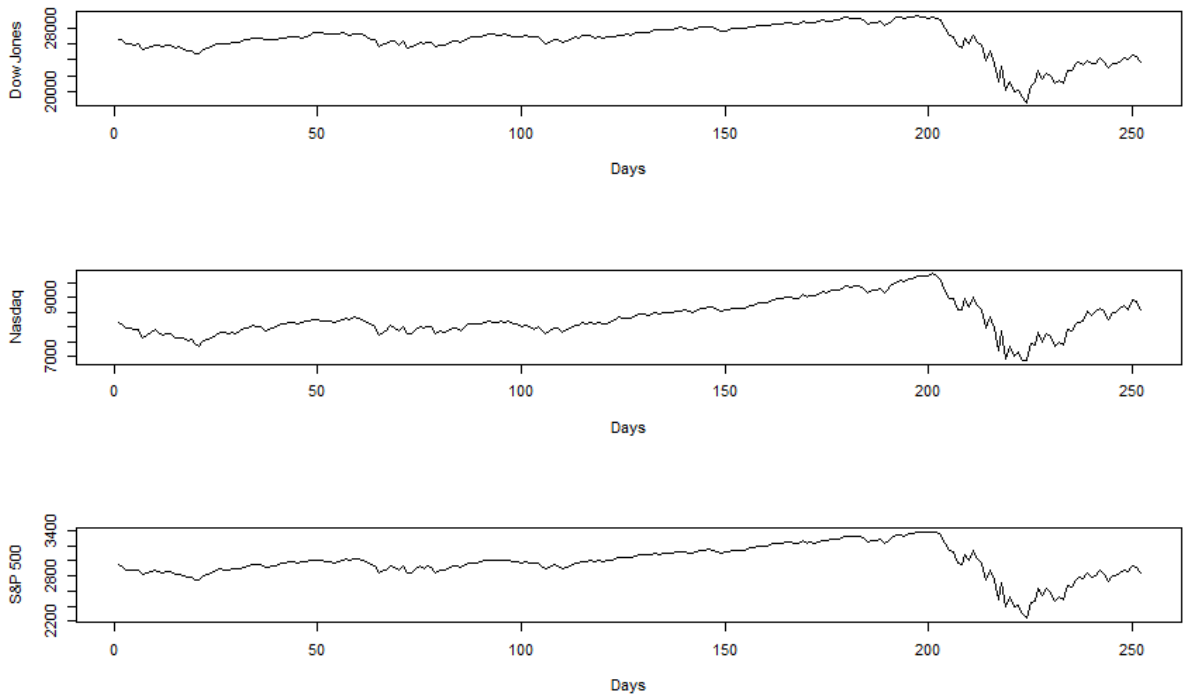


Figure 1 The figure shows American Stock Indexes, in order Dow Jones, NASDAQ, S&P 500, during the observed period.

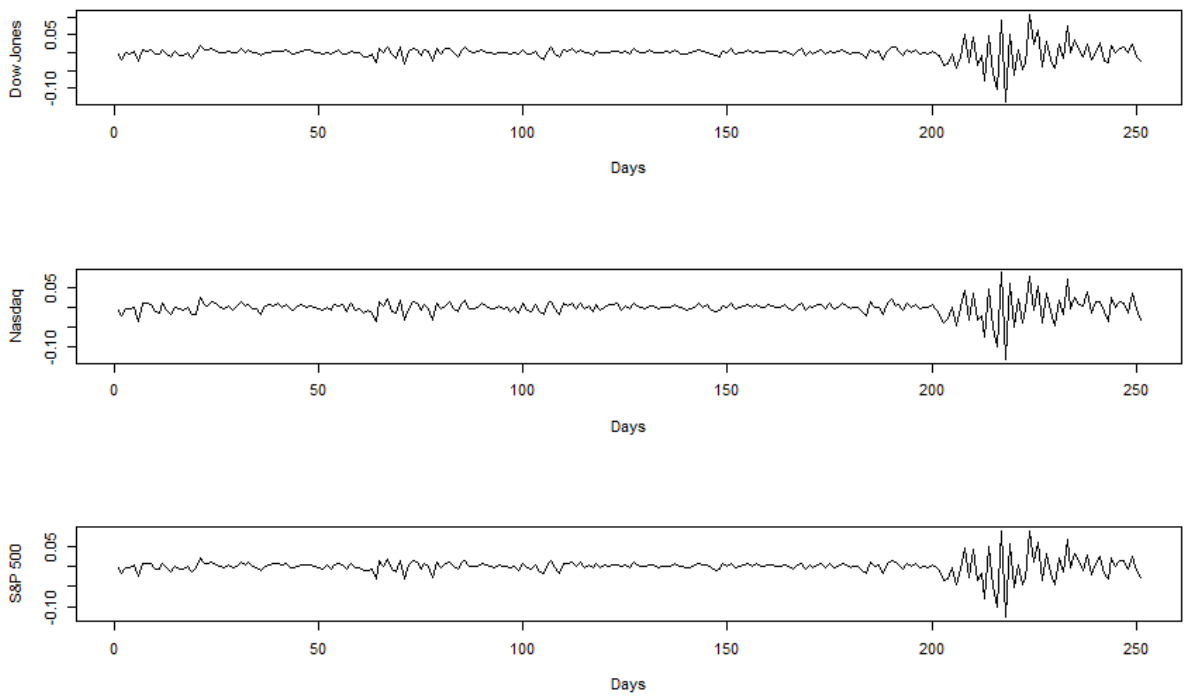


Figure 2 The figure shows the relative daily returns of American Stock Indexes, in order Dow Jones, NASDAQ, S&P 500, during the observed period.

There was found the evidence between the expected market risk premium and the predictable volatility

of stock return, exactly "the indirect evidence of a positive relation between expected risk premiums and volatility". The estimates were based on the data of the New York Stock Exchange (NYSE) in the period from 1928 to 1984. There has been substantial variation between estimated volatility and the expected risk premiums found almost over the 60 years. The regressions were based on the ARIMA model and its residuals, which were used for further estimation by ARCH family models.

The Baillie and DeGennaro (1990) [2] concluded the same relationship in their results, where the excessed kurtosis of the GARCH were monitored. Many other studies research the transmissions of volatility between markets. Karolyi et al. (1995) [5] examined the short-run transmission of volatility between NYSE and Toronto Stock Exchange (TSE) for the period from 1981 and 1989. The study of Adrian and Rosenberg (2008) [1] found that shocks to systematic volatility have higher impact to equity returns that was determined in the studies before. This study has been done on both the short-run and the long-run volatility. All results of the mentioned studies employed the ARCH family models for their estimations.

4 Methodology

The main source of information presented in the methodology part is the book Introductory Econometrics for Finance by Chris Brook (2019)[3]. First of all there is necessary to test for ARCH effects and then build volatility (GARCH type) models. The test employed by Brook can be formulated in several steps.

1. Run any postulated linear regression and save the residuals u_t.

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t \tag{1}$$

2. Square the residual, and regress them on q own lags to test for ARCH of order q and obtain the R^2.

$$\hat{u}_t^2 = \gamma_0 + \gamma_1 \hat{u}_{t-1}^2 + \cdots + \gamma_q \hat{u}_{t-q}^2 + v_t \tag{2}$$

3. The test statistic is defined as TR^2, which is the number of observations multiplied by the coefficient of multiple correlation, from the last regression.
4. The null and alternative hypotheses are

$$H_0 : \gamma_1 = 0 \, and \, \gamma_2 = 0 \, and \, \gamma_3 = 0 \, and \, \ldots \, and \, \gamma_q = 0$$

$$H_1 : \gamma_1 \neq 0 \, and \, \gamma_2 \neq 0 \, and \, \gamma_3 \neq 0 \, and \, \ldots \, and \, \gamma_q \neq 0$$

Based on the hypothesis results we may build the model. The most common ARCH family model for volatility modeling is Generalized ARCH (GARCH) model that allows the conditional variance to be dependent upon previous own lags, with the simpliest version GARCH(1,1):

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{3}$$

Where $\alpha_1 u_{t-1}^2$ is information about volatility during the previous period and $\beta \sigma_{t-1}^2$ is the fitted variance from the model during the previous period.

5 Results

First of all, it is necessary to mention that the ARCH-LM test rejected the null hypothesis. Hence, the conditional heteroskedasticity is presented in residuals. Moreover, we confirmed our findings by running the Ljung-Box test.

Thus, the GARCH model was estimated and the most suitable model (with respect to p and q) was chosen according to best information criteria (Akaike, Bayes, Shibata, Hannan-Quinn) results. According to our calculation, all of investigated American Stock Indexes were most performed at the GARCH(1,1). The variables α_1 and β_1 was significant in all models, but the constant was not, according to p-value.

The Figure 3 shows estimated volatility for American Stock Index. We can conclude that shape of all indexes is very similar. Before the Corona crisis the volatility was stable with values up to 2%. The volatility significantly were rising when the crisis started with the peaks in the middle and the late March 2020. The highest estimated volatility was in March 16. The difference between selected indexes is in the peaks. While the observed volatility reached almost 8% for S&P 500, Dow Jones Industrial Index have not reached 7% and NASDAQ highest point observed about 6.2%.

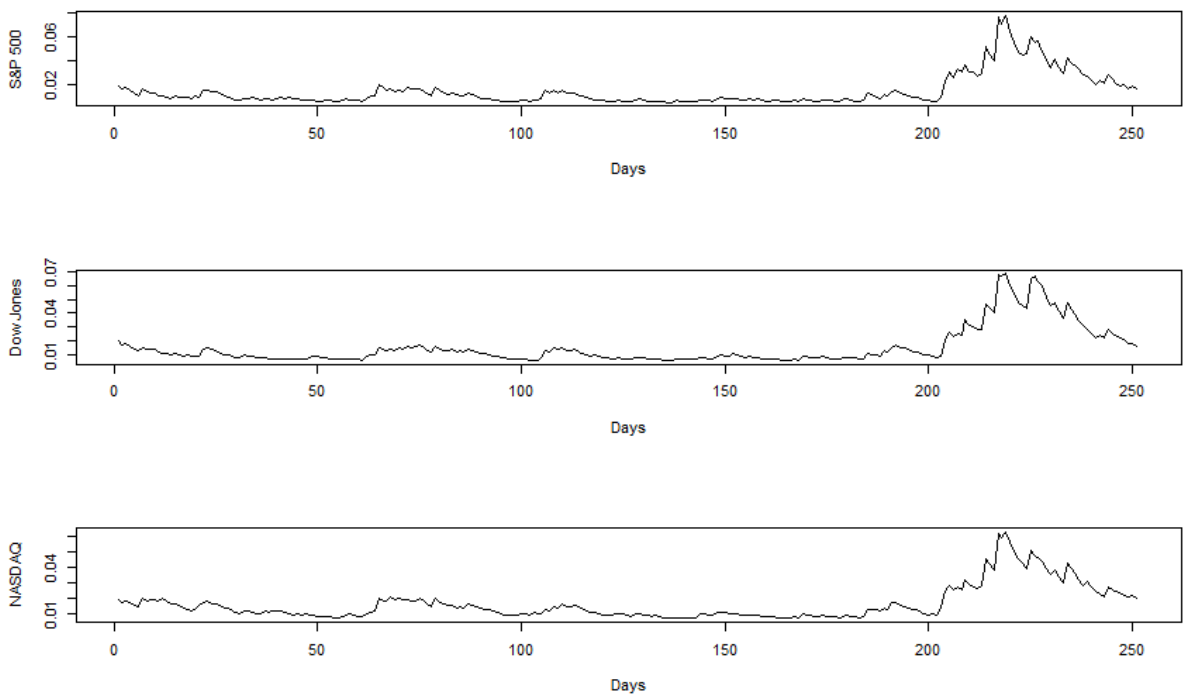


Figure 3 The figure shows the estimated volatility of American Stock Indexes, during the Corona crisis period.

6 Conclusions

The main point of this article was to provide estimation of volatility during the COVID-19 crisis. As we mentioned in the results, the peak of the volatility were observed on March 16, 2020 and then continuously lowered with several spikes. In our opinion this decrease of volatility might be the response to the first massive injection of Federal Reserve System (FED) announced on March 12, 2020. There were several other announcements about economic help from FED and government in next few weeks. Although the prices of stock went down, the volatility was much lower compared to its peak. At the beginning of April the stock bounced of the bottom. We can conclude that secure programs announced by governments and central banks calmed the situation on the market.

Acknowledgements

The work on this paper was supported by the grant No. F4/27/2020 of the Internal Grant Agency of University of Economics, Prague.

References

[1] Adrian T. & Rosenberg J. (2006). Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk. *SSRN Electronic Journal*.

[2] Baillie R. & DeGennaro R. (1996). Stock Returns and Volatility. *The Journal of Financial and Quantitative Analysis 25.2.*

[3] Brooks Ch. (2019). Introductory Econometrics for Finance. Cambridge University Press.

[4] French K., Schwert G. & Robert F. Stambaugh. (1987). Expected stock returns and volatility. *Journal of Financial Economics 19.1.*

[5] Karolyi A. (1995). A Multivariate GARCH Model of International Transmissions of Stock Returns and Volatility: The Case of the United States and Canada. *Journal of Business & Economic Statistics 13.1*

Threshold Analysis of Tourism Indicators of Selected Countries of EU

Eva Litavcová[1], Mária Popovičová[2], Igor Petruška[3]

Abstract. The total number of overnight stays in accommodation facilities is one of the key indicators of tourism. The aim of the paper is to analyze this indicator using the panel threshold model with fixed effect. Threshold models are widely used in econometric analyzes for non-linear time series modeling. They describe jump or structural breaks in the relationship between variables. Panel analysis allows models with one or more levels of fixed effects to be applied to multilevel longitudinal data. The panel threshold model combines these two approaches. In this article, it was applied to data on the number of overnight stays in seven European countries on a monthly basis for the period 2002–2019. We found such a combination of explanatory variables in which the difference of specific constants in the panel threshold model was proved and was the reason for its use. The paper shows the suitability of using these procedures for the analysis of tourism indicators. Statistical tests have shown that we can talk about different time series regimes when analyzing transformed data from the number of nights spent in tourist facilities.

Keywords: tourism, panel threshold model, SETAR model, threshold variable, regional variable

JEL Classification: C32, L83, Z32
AMS Classification: 62H20, 91B82

1 Introduction

The aim of this paper is to analyze one of the performance indicators in tourism, namely 'The total number of overnight stays in accommodation facilities' in selected EU countries. It deals with empirical research based on the panel threshold model. Tourism has an important territorial dimension, with uneven spatial distribution between and within countries, and delivering localized impacts [2]. The growth of tourism in one region has a positive effect on tourism in neighboring regions [12]. Countries with lower development (importance) of the tourism sector within a country tend to be more interconnected with international tourism demand of other countries [9]. The number of overnight stays in tourist facilities is one of the basic indicators of volume. A typical feature for tourism volume indicators is their seasonality, which is in [10] characterized for EU countries in more details using spectral analysis. The forecasts of the development of the number of overnight stays using linear SARIMA models for these countries can be found in [8].

According to Tsay [13], threshold models with a threshold variable (TAR) as well as with a threshold variable containing the delay of the analyzed variable (SETAR – k-regime Self Exciting Threshold Autoregresive Model) are widely used in econometric analyzes in nonlinear time series. It is applicable to nonlinear characteristics such as jump character or structural breaks in the relationship between variables. The TAR model uses piecewise linear models to obtains a better approximation of the conditional mean equation. It uses threshold space to improve linear approximation. The SETAR model is a piecewise linear AR model in the threshold space. It is nonlinear provided that number of regimes k is at least 2 [13, 6].

Panel analysis, explained in detail by Baltagi [1] makes it possible to apply models with one or more levels of fixed effects or random effects to multilevel longitudinal data. Panel data are measurements of the considered variable for the same set of N cases (entities, individuals, …) in several time points T, allow the identification and control of individual effects and dynamics. A limitation of the linear model is the assumption that a monotone change in a predictor variable leads to a monotone change in the dependent variable. If we assume a change in the behavior of the model after exceeding a certain threshold value of a variable, then linear panel regression is not appropriate. A suitable tool then is the panel threshold model with fixed

[1] University of Presov, Faculty of Management, Konstantinova 16, 080 01 Presov, Slovakia, eva.litavcova@unipo.sk.

[2] University of Presov, Faculty of Management, Konstantinova 16, 080 01 Presov, Slovakia, maria.popovicova@unipo.sk.

[3] University of Presov, Faculty of Management, Konstantinova 16, 080 01 Presov, Slovakia, igor.petruska@unipo.sk.

effects, which combines these two approaches. The core idea of the nonlinear threshold model of panel data is to examine whether the marginal impact of explanatory variables on explained variable will change significantly before or after a threshold value of the threshold variable [15].

This article is arranged as follows. In section 2, we briefly introduce the panel threshold model with fixed effects [5] and [6]. Section 3 presents the data and variables (tourism indicators) used in our empirical application. According to Wang [14], we apply a panel threshold model to selected variables using STATA 15 software in section 4. In section 5, we conclude the article.

2 Data and Methodology

The aim of this study is to analyze the development of the number of nights spent in tourist facilities of selected EU countries using models with changeable regimes. The year-on-year values of the growth rate were included in the analysis, ie $tpr = 100 \cdot (\mathrm{y_t} - \mathrm{y_{t-12}})/\mathrm{y_{t-12}}$. Descriptive statistics of the selected variables are given in Table 5.

The analyzed sample contains data from 8 countries, Slovakia, the Czech Republic, Hungary, Austria, Estonia, Latvia and Poland for 191 months (2004m1–2019m11) and the United Kingdom with 190 months. One extreme value (2018M1) for the United Kingdom was corrected by the average value of neighboring values. Two panels are analyzed. The first panel consists of the first 6 countries and the data of Poland, which serve as an added exogenous variable. The second panel consists of the first 7 countries (without 2019m11) and the added exogenous variable contains data for the United Kingdom. The panel data is strongly balanced. The data, namely "the total number of overnight stays in accommodation facilities", come from the Eurostat database [4].

Panel models are widely used in macroeconomics and financial analysis. Panel data are combined cross-sectional and time data. This type of data is advantageous due to easily controllable individual variability of the investigated entities, possibility of identification and measurability of effects, which are impossible to verify solely by cross-sectional data and time series and they are also more appropriate for the study of dynamics adjustment [7, 11]. A panel data regression in form of one-way error component model for the disturbances according to Baltagi [1] is

$$y_{it} = \alpha + X'_{it}\beta + \mu_i + \nu_{it}\,, \qquad t = 1, \dots, T, \qquad i = 1, \dots, N \tag{1}$$

with i denoting individuals (cross-section dimension), t denoting time (time-series dimension), α is a scalar, β is $K \times 1$ and X_{it} is the i-th observation on K explanatory variables. The character μ_i denotes the unobservable time-invariant individual-specific effect, ν_{it} denotes the remainder disturbance, y_{it} denotes it-th value of dependent variable. The assumption is independency of X_{it} and ν_{it}, disturbances are independent of each other. Random errors ν_{it} are independent random variables with $IID(0, \sigma_\nu^2)$ with constant variance. For model with fixed effects (FE) the assumption of a correlation between the entity's error term (individual-specific) and the prediction variables is expected to be met. FE removes the effect of those time-invariant characteristics. For model with random effects (RE) have to be fulfilled the assumptions: $\mu_i \sim IID(0, \sigma_\mu^2)$, $v_i \sim IID(0, \sigma_v^2)$, they do not depend on each other, nor with prediction variables.

Threshold regression methods are developed for non-dynamic panels with individual specific fixed-effects. Hansen [5] proposed least squares estimation of the threshold and regression slopes using fixed-effects transformations. Wang [14] introduced a command (xthreg) for implementing this model in Stata. He presents the equation of a panel threshold model with fixed effects with one threshold by

$$y_{it} = \mu + X_{it} I(q_{it} < \gamma)\beta_1 + X_{it} I(q_{it} \geq \gamma)\beta_2 + u_i + e_{it}, \tag{2}$$

where the threshold parameter γ of a threshold variable q_{it} divides regression into two regimes with coefficients β_1 and β_2. The function $I(\bullet)$ is an indicative one. By analogy for the threshold model with two thresholds by

$$y_{it} = \mu + X_{it} I(q_{it} < \gamma_1)\beta_1 + X_{it} I(\gamma_1 \leq q_{it} < \gamma_2)\beta_2 + X_{it} I(q_{it} \geq \gamma_2)\beta_3 + u_i + e_{it} \tag{3}$$

two threshold parameters γ_1 and γ_2 divide regression into three regimes. The parameter u_i is the individual effect, while e_{it} is the disturbance.

Caner and Hansen [3] developed consistent estimators for the threshold in a model with endogenous variables and an exogenous threshold variable. The estimator for the threshold is a 2SLS estimator, and the estimator of the slope parameters is a GMM estimator.

3 Results

Several models were considered in the analysis of the rate of increase in the number of nights spent in tourist facilities. Based on the results of the works [10] and [8], there is an obvious trend in the time course of the indicator of the number of nights spent and the periodicity of time series is also obvious. The analyses used in the SARIMA models and the spectral analysis show that the wave lengths of 12 and 6 months dominate over the others. For this reason, the following were selected as regional variables:

- lagged values of variable *tpr*, the year-on-year values of the growth rate: *L.tpr* (Lag) one month, *L6.tpr* six months, *L12.tpr* twelve months backward,
- time *t* was examined too.

The aim was to find such a combination of panel threshold model parameters that would be correctly applied to the data. The resulting models were selected from a number of tested models. Motivation to select data from Poland as a threshold variable (*L.tprPoland*) was the question of the influence of the largest country on the other countries of the panel. The motivation for choosing the United Kingdom (threshold variable *L.tprUK*) was Brexit.

For the first panel with the countries Slovakia, the Czech Republic, Hungary, Austria, Estonia and Latvia and with the threshold variable *L.tprPoland*, 4 regional explanatory variables, time *t* and the above-mentioned delayed values of the variable *tpr* are used. The resulting optimal model found by the Stata program contains one threshold value. The resulting coefficients of the first panel threshold model with fixed effects are shown in Table 1. Almost all coefficients, except coefficient of time *t* in second regime, are significant. Estimated threshold value of the threshold variable *L.tprPoland* is $\gamma = -0.576$, $p < 0.001$ from the threshold effect test with bootstrap. The panel regression *F* test, $F(8, 1060) = 104.92$, $p < 0.001$, gives one reason for its use. Also the *F* test that all specific effects are 0, $F(5, 1060) = 3.42$, $p = 0.005$, rejects the null hypothesis and confirms that they are not. An insight into the 95% confidence intervals of the model coefficients listed in Table 1 (95% LCL and UCL) can be used to estimate the significance of the coefficient differences in the different regimes. It is clear from this that there are significant differences between the regimes in the coefficients of the variables *t*, *L.tpr*, *L12.tpr*.

variable	regime	Coef.	Std. Err.	t	P>|t|	95% LCL	95% UCL
t	0	−0.0583615	0.0141297	−4.13	0.000	−0.0860869	−0.0306361
	1	0.0007049	0.0038533	0.18	0.855	−0.006856	0.0082658
L.tpr	0	0.5823369	0.0688816	8.45	0.000	0.4471771	0.7174967
	1	0.3433440	0.0289042	11.88	0.000	0.2866281	0.4000599
L6.tpr	0	0.2931719	0.0857987	3.42	0.001	0.1248173	0.4615265
	1	0.2694900	0.0278976	9.66	0.000	0.2147491	0.3242309
L12.tpr	0	−0.9304969	0.1048821	−8.87	0.000	−1.136297	−0.7246968
	1	−0.2410301	0.0242764	−9.93	0.000	−0.2886654	−0.1933948
const	c	3.1282440	0.4796958	6.52	0.000	2.186983	4.0695050

Table 1 Coefficients of the panel threshold model with fixed effect with threshold variable *L.tprPoland*

In the second panel, in which 7 countries are together with Poland and the threshold variable used is *L.tprUK*, the 3 delayed values of the variable *tpr* mentioned above are used as regional variables. The time variable *t* was not significant. As for the explanatory variables, the coefficients were found by optimization in three regression regions, which are given by two threshold values. The estimated two threshold values of the variable *L.tprUK* are $\gamma_1 = -12.430$, $p < 0.001$ and $\gamma_2 = 7.719$, $p = 0.027$. Figure 1 shows the corresponding values of LR statistics according to Wang [15]. The dashed line denotes the critical value (7.35) at the 95% confidence level.The result of the panel regression F test is $F(9,1230) = 98.46$, $p < 0.001$ and, the result of the *F* test that all specific effects are 0 is $F(6, 1230) = 4.13$, $p < 0.001$. The resulting coefficients of the second panel threshold model with fixed effects are shown in Table 2. All of them are significant. In the same way as in the previous case, it is possible to determine the significance of the inter-regime differences in the coefficients of the explanatory variables. In the case of *L.tpr* all three coefficients

are significantly different. In the case of *L6.tpr* and also in the case of *L12.tpr*, there are two pairs of significantly different coefficients.

For comparison, panel regression was performed with the same variables without the use of thresholds for both investigated panels. Breusch and Pagan Lagrangian multiplier test did not confirm the preference for a random effect against the null hypothesis of the pooled model. The Hausman test confirmed that the preferred model is a fixed effect in both models for both panels. Time variable *t* was not significant in both panels.

variable	regime	Coef.	Std. Err.	t	P>|t|	95% LCL	95% UCL
L.tpr	0	0.7474622	0.0626048	11.94	0.000	0.6246382	0.8702862
	1	0.4606080	0.0323420	14.24	0.000	0.3971564	0.5240596
	2	0.3148810	0.0404793	7.78	0.000	0.2354650	0.3942971
L6.tpr	0	0.5609915	0.0867462	6.47	0.000	0.3908045	0.7311785
	1	0.1775785	0.0320786	5.54	0.000	0.1146436	0.2405134
	2	0.4576132	0.0451928	10.13	0.000	0.3689496	0.5462768
L12.tpr	0	−0.7619161	0.0668132	−11.40	0.000	−0.8929967	−0.6308356
	1	−0.2045605	0.0287468	−7.12	0.000	−0.2609586	−0.1481623
	2	−0.2540637	0.0361114	−7.04	0.000	−0.3249104	−0.1832171
const		2.6681250	0.2303283	11.58	0.000	2.2162460	3.1200050

Table 2 Coefficients of the panel threshold model with fixed effect with threshold variable *L.tprUK*

Threshold models applied to the countries studied individually were also made. In each country, the same explanatory regional variables were used, ie *L.tprCountry*, *L6.tprCountry* and *L12.tprCountry*, where *Country* is understood as the identifier of a specific country. The results of the six threshold models with the *L.tprPoland* threshold variable are shown in Table 3 and with the *L.tprUK* threshold variable in Table 4. The significance of the coefficients is indicated by an index in the tables where "a", "b", "c", "d" denotes 10%, 5%, 1%, and 0.1% significance level. A significant inter-regime difference in the coefficients is indicated in bold. In addition to the coefficients, the tables also contain the estimated threshold value and values of the information criteria. The stated values of the information criteria AIC, BIC and HQIC with the same number of parameters and measured values allow a comparison of the quality of the model. In terms of model quality, the model for Slovakia has the best results in both tables. Then the Czech Republic, Estonia, Hungary, Austria and the last one is Latvia.

variable	regime	Slovakia	Czechia	Hungary	Austria	Estonia	Latvia
t	0	0.03784^{c}	0.01466	0.12844	−0.00319	**0.18621^{b}**	−0.03318
	1	0.01438	0.01746	0.00728	0.03239	**−0.01334^{a}**	−0.03107^{b}
L.tprCountry	0	**0.82055^{d}**	**0.61379^{d}**	0.49616	**−0.38842^{d}**	0.32900^{a}	**0.90441^{d}**
	1	**0.55869^{d}**	**0.21968^{b}**	0.21435^{c}	**−0.07687**	0.52167^{d}	**0.49102^{d}**
L6.tprCountry	0	**0.06034**	0.27025^{d}	0.13838	0.04999	0.10760	**−0.22317**
	1	**0.27754^{c}**	0.20249	0.28555^{d}	−0.01362	0.18796^{b}	**0.14994^{b}**
L12.tprCountry	0	−0.32355^{d}	**−0.34444^{d}**	−0.64950^{b}	**−0.53381^{d}**	**−0.56017^{c}**	−0.09641
	1	−0.24981^{d}	**−0.12612**	−0.23922^{d}	**−0.12597**	**−0.14960^{c}**	−0.21357^{d}
const	0	−1.9002	−0.13316	−11.1377	4.0857^{d}	−16.2645^{c}	−0.60759
	1	0.31307	2.3944	3.1844^{c}	2.12700	3.7138^{c}	8.2355^{d}
L.tprPoland	threshold	7.5235	9.2826	1.1998	13.9466	1.1998	1.1998
	AIC	543.7677	547.2026	585.4481	661.2726	565.3412	731.8525
	BIC	575.6415	579.0764	617.3220	693.1464	597.2151	763.7264
	HQIC	556.6923	560.1272	598.3727	674.1972	578.2658	744.7771

Table 3 Results of the threshold model with fixed effect with threshold variable *L.tprPoland*, $n = 179$

As can be seen from Table 3 a significant inter-regime differeces in the coefficients are indicated for all countries except Hungary.The countries Hungary, Estonia and Latvia have the same treshold values, which differ from other countries. The most common are different regional coefficients for the variable *L.tprCountry.* A significant inter-regime differences in the coefficients are indicated for all countries in Table 4. Czechia, Hungary, Austria and Estonia have very similar treshold values. The absolute values of thresholds are almost the same in the case of Slovakia and Latvia, but they have the opposite sign. The most common are different regional coefficients for the variable *L12.tprCountry.*

The criteria – Akaike information criterion (AIC), Bayesian information criterion (BIC) and Hannan–Quinn information criterion (HQIC) were used to compare the models. The information criteria acquire the lowest values in both cases *(L.tprPoland and L.tprUK)* for Slovakia and the Czech Republic.

variable	regime	Slovakia	Czechia	Hungary	Austria	Estonia	Latvia
t	0	0.04174[c]	0.01672	0.01639	0.01063	0.00587	**0.00363**
	1	0.01990[b]	0.01496[b]	0.01569[a]	–0.00118	–0.00684	**–0.04353[b]**
L.tprCountry	0	0.69659[d]	**0.74024[d]**	0.44279[b]	–0.27481	0.76209[d]	**0.77683[d]**
	1	0.70094[d]	**0.462793[d]**	0.30801[d]	–0.37401[d]	0.53068[d]	**0.34663[c]**
L6.tprCountry	0	0.03774	0.42148[a]	**1.4254[c]**	0.17735	**0.90412[c]**	**0.01853**
	1	0.17604[b]	0.24270[d]	**0.30700[d]**	0.03022	**0.21895[c]**	**0.50816[d]**
L12.tprCountry	0	**–0.47434[d]**	**–0.85464[d]**	**–0.66181[d]**	**–0.86451[d]**	–0.47167[a]	–0.13149[b]
	1	**–0.21436[d]**	**–0.23461[d]**	**–0.24150[c]**	**–0.32943[d]**	–0.15976[c]	–0.29054[c]
const	0	–2.6257	1.1897	–3.6835	3.4702	–1.3738	1.8454
	1	–0.35363	0.45434	1.3218	3.6540[c]	2.6108[b]	8.3934[b]
L.tprUK	threshold	–6.3486	–12.4304	–12.4304	–12.1414	–12.0708	6.3138
	AIC	547.3962	550.1930	603.5652	656.3380	578.0744	744.5521
	BIC	579.2700	582.0668	635.4391	688.2119	609.9483	776.4260
	HQIC	560.3208	563.1176	616.4898	669.2626	590.9990	757.4767

Table 4 Results of the threshold model with fixed effect with threshold variable *L.tprUK*, $n = 179$

We can view in Figure 1 the threshold confidence intervals of panel threshold model with fixed effect with threshold variable *L.tprUK* by plotting the LR statistics. The dashed line denotes the critical value (7.35) at 95% confidence level.

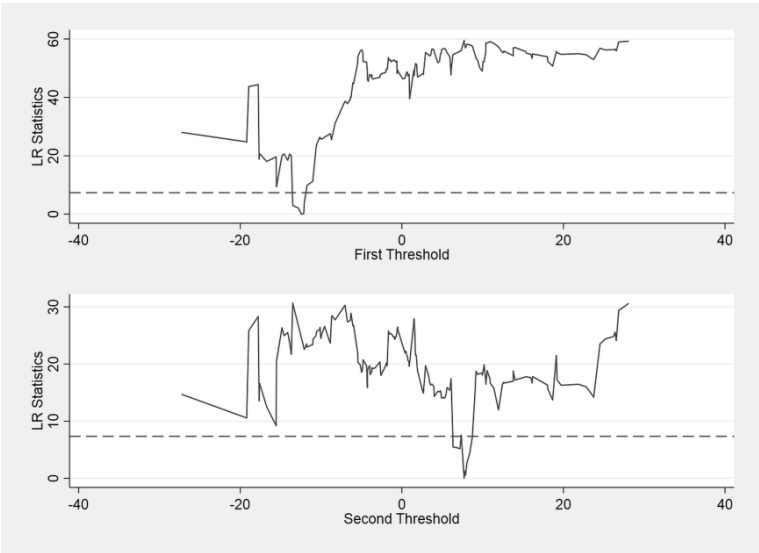


Figure 1 The LR statistics from model with threshold variable *L.tprUK*

Variable	Mean	Median	Min	Max	Std.D.	Variable	Mean	Median	Min	Max	Std.D.
tprAustria	2,05	1,70	–25,12	43,00	8,27	*tprLatvia*	7,09	6,80	–31,36	38,93	12,47
tprCzechia	3,91	3,85	–17,99	20,91	6,17	*tprPoland*	8,97	8,07	–13,10	36,45	7,33
tprEstonia	5,54	4,16	–19,97	51,80	9,60	*tprSlovakia*	3,37	3,56	–23,46	28,58	9,32
tprHungary	4,38	4,57	–17,85	23,45	6,42	*tprUK*	1,57	0,67	–28,17	39,74	11,92

Table 5 Descriptive statistics of selected variables, except $tprUK\ n = 191$, for $tprUK\ n = 190$.

4 Conclusion

A nonlinear panel threshold model was used to analyze the number of nights spent in tourist facilities (in form of year-to-year growth rate) in selected European countries. Slovakia, the Czech Republic, Hungary, Austria, Estonia, Latvia were in the first panel and also with Poland in the second panel. The threshold variable was vector of first lagged values of Poland in the first panel and first lagged values of United Kingdom in the second one. Our goal was to find such a combination of panel threshold model parameters that would be correctly applied to the examined data. Using the threshold regression on the panel, interesting properties of the investigated relationships were shown. It is known that after exceeding the threshold value, the direction of the dependence can change significantly, which has also been demonstrated here in one case. When using linear panel regression, the significance of time variable was not demonstrated on both panels. In the panel threshold model, it was demonstrated in one regime on the first panel.

It would be too simplistic to interpret the impact of the values of the two countries selected as threshold variables on the values of the other countries in the panel. In any case, these findings may be a reason to investigate the other causes of this impact. These results suggest the suitability of using panel threshold models in similar contexts for appropriately transformed data. The presented models can be extended by adding other exogenous variables.

Acknowledgements

This contribution was supported by research grant VEGA No. 1/0470/18 “Economic activity of tourism in the European area”. This work was supported by the Slovak Research and Development Agency under the contract no. APVV-17-0166.

References

[1] BALTAGI, B. H. (2005). *Econometric Analysis of Panel data*. Wst Sussex: John Wiley &Sons.
[2] BATISTA E SILVA, F., HERRERA, M. A. M., ROSINA, K., BARRANCO, R. R., FREIRE, S., & SCHIAVINA, M. (2018). Analyzing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management*, 68, 101–115.
[3] CANER, M., & HANSEN, B. E. (2004). Instrumental variable estimation of a threshold model. *Economic Theory*, 20, 813–843.
[4] EUROSTAT. (2020). Nights spent at tourist accommodation establishments – monthly data. Retrieved from https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=tour_occ_nim&lang=en, 2. 2. 2020.
[5] HANSEN, B. E. (1999). Threshold effects in non-dynamic panels: estimation, testing, and inference. *Econom*, 93, 345–368.
[6] HANSEN, B. E. (2011). Threshold autoregression in Economics. *Statistics and Its Interface*, 4, 123–127.
[7] KREMER, S., BICK, A., & NAUTZ, D. (2013). Inflation and growth: new evidence from a dynamic panel threshold analysis. *Empir Econ*, 44, 861–878.
[8] LITAVCOVÁ, E. & VAŠANIČOVÁ, P. (2019). On modelling of development of tourism in selected European Countries. In T. LÖSTER & T. PAVELKA (Eds.), *13th International Days of Statistics and Economics*, Praha, Czech Republic, September 5th–7th 2019 Conference proceedings. (pp. 950–959). Prague: Melandrium.
[9] LYÓCSA, Š., VAŠANIČOVÁ, P., & LITAVCOVÁ, E. (2019). Interconnectedness of international tourism demand in Europe: A cross-quantilogram network approach. *Physics A: Statistical Mechanics and its Applications*, 526, 1–13.
[10] PETRUŠKA, I. (2018). Spektrálna analýza ukazovateľov cestovného ruchu európskych krajín (Spectral analysis indicators of tourism in European Countries). In S. JENČOVÁ & E. LITAVCOVÁ (Eds.),

Ekonomická aktivita turizmu (Economic Activity of Tourism): Conference proceedings (pp. 120–128). Prešov: Faculty of Management.
[11] PETRUŠKA, I. (2019). Comparison of sources healthcare financing in European selected countries through panel data analysis. *Journal of Management and Business: Research and Practice*, 11(1), 12–19.
[12] ROMAO, J., GUERREIRO, J., & RODRIGUES, P. M. M. (2017). Territory and sustainable tourism development: A space-time analysis on European regions. *The Region*, 4(3), 1.
[13] TSAY, R. (2010). *Analysis of Financial Time Series*. New Jersey: John Wiley & Sons.
[14] WANG, Q. (2015). Fixed-effect panel threshold model using Stata. *The Stata Journal*, 15(1), 121–134.
[15] YANLONG, D. (2018). Fiscal Expenditure Structure and Environmental Pollution: Empirical Research Based on the Panel Threshold Model. *Meteorological and Environmental Research*, 9(1): 32–37.

Simulation of an Invetory Model by Stochastic Differential Equation Using sw Mathematica

Ladislav Lukáš[1]

Abstract. The paper deals with stochastic dynamic inventory models which are solved numerically. An area of inventory theory research that has recently been receiving considerable attention involves situations in which the demand rate is dependent on the level of inventory. Inventory for an item is replenished by a time dependent deterministic process, in the simples case, with constant rate one. Simultaneously, the product under inventory is depleted by stochastic demand which rate depends upon the current inventory level. Stochastic differential equation of Itô type for the inventory level is formulated and used for numerical computations. All simulated paths of inventory levels are generated by our Mathematica code. Sample mean path is calculated thereof and further filtered in order to get some inventory management important values. Two main Mathematica commands for generation stochastic processes is presented, too.

Keywords: stochastic dynamic inventory model, replenishment process, stochastic demand, stochastic differential equation, sample mean process, numerical simulation in Mathematica

JEL Classification: C630, L230
AMS Classification: 60H10, 60G40, 65C30

1 Introduction

There is a lot of literature devoted to inventory control and management methods. For general reference, we select two books, only. First, a classical book [1], and our one [5]. There is well-known, that in general framework of dynamic inventory models two functions of time $s(t)$ and $d(t)$ are introduced, which indicate replenishment process and demand process of goods, respectively. Usually, they are assumed non-negative and continuous functions of time depending upon various parameters, but which are not further specified here. Dimensions of both these functions are [goods unit/time unit], as usual. In [1] Chapter 3, and in [5] Chapter 2, we can read dynamic inventory models in deterministic framework.

Let $z(t)$, denote an instantaneous level of goods at inventory at time t, called an inventory level, in short hereafter. The elementary balance equation gives infinitesimal inventory level increment $\mathrm{d}z(t)$ during time period $\mathrm{d}t$, in following form

$$\mathrm{d}z(t) = z(t + \mathrm{d}t) - z(t) = s(t)\mathrm{d}t - d(t)\mathrm{d}t.$$

From this equation, one gets simply a rate of inventory level change in time

$$\frac{\mathrm{d}z(t)}{\mathrm{d}t} = s(t) - d(t),$$

which is an ordinary differential equation for inventory level $z(t)$, a well-known basic differential equation in inventory theory. To solve it, one needs an initial condition $z(0) = z_0$, expressing an initial inventory level at time $t = 0$. It is an initial value problem, which provides a solution by integration over period $(0, t)$, taking the form

$$z(t) = z_0 + \int_0^t s(\tau)\mathrm{d}\tau - \int_0^t d(\tau)\mathrm{d}\tau. \tag{1}$$

In the paper, we are focused on demand processes, since the replenishment ones take usually simple forms, e.g.

$$s(t) = \sum_{i=0}^{n} q_i \delta(t - t_i),$$

which actually represents a discrete time process at instants $t_i \in [0, T]$ defined using the Dirac function $\delta(.)$, with T giving the inventory global period, $q_i, i = 0, \ldots, n$ are inventory repenishment quantities, where q_0

[1] University of West Bohemia, Faculty of Economics, Dept. of Economics and Quantitative Methods, Univerzitní 22, 306 14 Pilsen, lukasl@kem.zcu.cz

gives the intial inventory level $z(0) = q_0$, in the case of $d(0) = 0$, which is an usual assumption. The index n is determined simply by relation $n = \max_j(t_j \leq T),\ j = 1, 2, \ldots$

Following [9] and [10], but other literature, too, found simply by searching the internet with relevant search profiles, we know that inventory management researchers and practioners have long recognized the demand of many retail items/goods is proportional to the amount of inventory displayed. The idea behind that is that a large stock for certain products may generate a higher demand, frequently. However, it is not an only reason, there are also the other ones, for example, reflecting fashion waves, seasonalities, price discount actions, and immense number of other externalities mainly with marketing viewpoints and expressing consumers' moods, too. In [10], one can find a classification of demand functions which are inventory-level-dependent, into two distinct types. The first type, denoted Type I, assumes that demand rate is a function of the initial inventory level at each inventory cycle, more precisely, the order-up-to level S. The second type, denoted Type II, assumes the demand rate to be dependent on the instantaneous inventory level. Denoting demand rate $r(t) = \dot{d}(t) = \mathrm{d}d(t)/\mathrm{d}t$, we can list the most usual demand rates of Type I, in following forms

- constant function: $r_S = \alpha,\ \alpha > 0,$ i.e. it is invariant upon S,
- linear function: $r_S = \alpha + \beta S,\ \alpha, \beta > 0,$
- simple power function: $r_S = \alpha S^\beta,\ \alpha > 0,\ \beta \in (0, 1),$
- posynomial function: $r_S = \gamma + \alpha S^\beta,\ \gamma > -\alpha S^\beta,\ \beta < 1,\ \mathrm{sign}(\alpha) = \mathrm{sign}(\beta),$
- exponential function: $r_S = \alpha + \beta \exp(-S),\ \alpha > 0, \beta < 0.$

with α, β, γ, given constants, Here, r_S, points out that the demand rate does not depend on t, actually, but on S, only, which also means that the demand rate is constant over each inventory cycle. As well-known, the posynomial functions is a subclass of power functions with fractional exponents, which was introduced in field of geometrical programming, in particular.

To get demand rates of Type II, one may use the same functions as for the Type I just replacing S with $z(t)$. It looks quite simple in principle, but such an approach is rather involved since it contains a hidden influence of replenishment process $s(t)$ on demand rate via $z(t)$. To eliminate such difficulty while keeping a level-dependence property, one may replace S either with t thus getting a time-dependent demand rate $r(t)$, in explicit form, or with $d(t)$ which yields a demand-dependent $r(t)$, in form of an ordinary differential equation for demand process $d(t)$, just keeping in mind that $r(t) = \dot{d}(t) = \mathrm{d}d(t)/\mathrm{d}t$, and adding an usual initial condition $d(0) = 0$. For example, using the linear and power functions of Type I, one gets the following demand rates of Type II

- $r(t) = \alpha + \beta t,\ \alpha, \beta > 0,$ explicit time-dependent demand rate within an inventory cycle,
- $r(t) = \dot{d}(t) = \alpha + \beta d(t),\ \alpha, \beta > 0,$ instantaneous demand-level-dependent demand rate applied in an inventory cycle; an ordinary differential equation of the first order, with $d(0)$, to be given for proper setting of the IVP (initial value problem),
- $r(z(t)) = \alpha z(t)^\beta,$ for $z(t) \geq S_0,$ and $= d,$ for $0 \leq z(t) \leq S_0,\ d = \alpha S_0^\beta,$ where $\alpha > 0,\ 0 < \beta < 1,$ and S_0 denotes a lower level of inventory which separates two different regimes of $r(z)$, i.e. constant if $z(t)$ falls bellow S_0, and level-dependent one on the opposite.

In [2], we already discussed several other types of demand processes based upon time-dependent demand rates of Type II, $r(t)$. The proposed processes were constructed within a framework of the generalized EOQ models, where coefficients $a_i, i = 0, \ldots, 3$, are determined by various interpolation conditions applied on $z(t)$

$$r(t) = \alpha_1 t + \alpha_2 t^2, \quad \text{yielding } z(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3.$$

In Fig. 1, various trajectories of $z(t)$, are depicted at the inventory cycle with demand rates of Type I, and Type II, too. Till now, we have discussed demand process $d(t)$ in deterministic framework, only. However, no doubt, a more proper framework is a stochastic one thus reflecting more realistic situations in practice.

2 Stochastic demand processes

2.1 Stochastic differential equations – basics

In order to emphasize a difference between deterministic approach and stochastic one, we change notation of demand process, formally. Instead of $d(t)$, let $x(t)$, denote a stochastic demand process precisely, called simply a demand process hereafter, as well. Framework of stochastic differential equations is a proper one for description of stochastic demand processes, in general. In [5] Chap. 4, we have formulated a stochastic inventory control model with the demand process taking a simple form, with $\alpha, \sigma > 0$, given constants, and

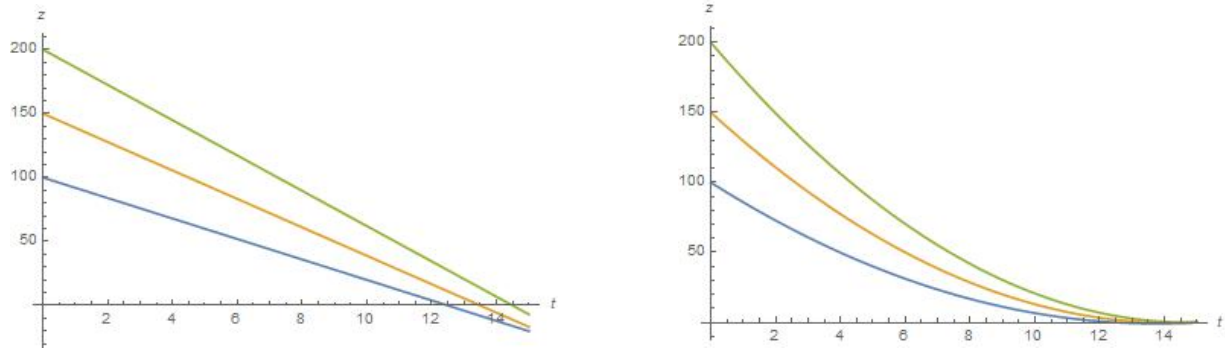


Figure 1 $z(t), S = 100$ (blue), 150 (red), 200 (green), with demand rate $r(t)$: Left: Type I, Right: Type II.

$w(t)$ is the Wiener process

$$\mathrm{d}x(t) = \alpha \mathrm{d}t + \sigma \mathrm{d}w(t). \tag{2}$$

In [7] Chap. 5, and [3] Chap. 5, a general introduction to stochastic differential equations is presented. Here, we restrict ourselves to one-dimensional case, and following [3] Chap. 5, we say that an $\mathbb{R}$ -valued stochastic process $y(t)$ is a *solution* of the *Itô stochastic differential equation*, called SDE in short, for $0 \leq t \leq T$, provided

$$\mathrm{d}y(t) = b(y(t), t)\mathrm{d}t + B(y(t), t)\mathrm{d}w(t), \tag{3}$$

(i) y(.) is progressively measurable with respect to $\mathcal{F}(.)$,
(ii) $b(y(t), t) \in L(0, T)$, a real-valued function given, called *drift term*,
(iii)$B(y(t), t) \in L^2(0, T)$, a real-valued function given, called *diffusion term*,
(iv)$y(t) = y_0 + \int_0^t b(y(s), s)\mathrm{d}s + \int_0^t B(y(s), s)\mathrm{d}w(s)$, a.s. $\forall t \in [0, T]$, a stochastic process, called *Itô process*.

where $y_0 = y(0)$ is a random variable which is independent of $w(.)$, and $\mathcal{F}(t)$ denotes the σ -algebra, denoted $\mathcal{S}$ generally, being generated by y_0 and the history of the Wiener process up to (and including) time t on some probability space $(\Omega, \mathcal{S}, \mathcal{P})$; it takes the following form

$$\mathcal{F}(t) = \mathcal{S}(y_0, w(s)(0 \leq s \leq t)), \ t \geq 0.$$

An intuitive interpretation of the SDE is that in a small time interval of length Δ the $y(t)$ changes its value by an amount that is normally distributed with expectation $b(y(t), t)\Delta$, and variance $B(y(t), t)^2\Delta$, and is independent of the whole past history of $y(.)$. Well-known example of linear SDE with given real-valued functions $f(t), g(t)$, which has the unique solution $y(t)$, is following one

$$\begin{aligned} \mathrm{d}y(t) &= f(t)y(t)\mathrm{d}t + g(t)y(t)\mathrm{d}w(t), \ \ y(0) = 1, \\ y(t) &= \exp\Big(\int_0^t (f(s) - \frac{1}{2}g(s)^2)\mathrm{d}s + \int_0^t g(s)\mathrm{d}w(s)\Big), \ \forall t \in [0, T], \end{aligned} \tag{4}$$

where the first integral is a classical Riemann (or Lebesgue) integral, and the second one is an Itô integral. In the case $f(t) = \mu, g(t) = \sigma, \mu, \sigma > 0$, given, denoting drift and diffusion coefficient, respectively, the (4) describes so called geometric Brownian motion yielding an explicit solution

$$\begin{aligned} \mathrm{d}y(t) &= \mu y(t)\mathrm{d}t + \sigma y(t)\mathrm{d}w(t), \ \ y(0) = y_0, \\ y(t) &= y_0 \exp\big((\mu - \frac{\sigma^2}{2})t + \sigma w(t)\big), \ \forall t \in [0, T]. \end{aligned} \tag{5}$$

As well-known, the Wiener process (standard Brownian motion) is the continuous time stochastic process having following properties

(i) $t_1 \neq t_2 \implies w(t_1), w(t_2)$ are independent random variables,
(ii) $\{w(t)\}$ is stationary process, i.e. the joint distribution of $\{w(t_1 + t), w(t_2 + t), \ldots, w(t_k + t)\}$ does not depend on t for arbitrary set of different times t_j, $j = 1, \ldots, k$,
(iii)$\mathrm{E}[w(t)] = 0$, $\forall t \in \mathbb{R}$,
(iv)$w(t)$ has continuous paths, $\forall t \in \mathbb{R}$.

2.2 Stochastic inventory models – demand processes

To apply scientific inventory control it is important to have, as much as possible, an estimate of demand process. Usually, a simple EOQ model is extended to cases of random demand realized by Poisson process, and also a special consideration is given to models assuming normally distributed demand. However, the normal assumption of a stationary demand process is not always realized in practice. Demand levels are subject to fluctuations over time. In recent paper [8], which has served as our main motivation for the subject, the inventory-level-dependent stochastic demand process of Type II is discussed thoroughly; it takes the following form

$$\mathrm{d}x(t) = (d_1 z(t) + d_2 z(t)^2)\mathrm{d}t - \sigma \mathrm{d}w(t), \quad d(0) = 0, \tag{6}$$

where d_1, d_2, are constants given, $x(0) = 0$, is a natural initial condition, and σ is diffusion coefficient measuring the intensity of the stochastic disturbance. This example also shows, that to get a reasonable formulation of an inventory problem, one needs to consider a replenishment (supply) process, too. The proposed SDE for inventory level $z(t)$, has the following form

$$\mathrm{d}z(t) = \big(u - (d_1 z(t) + d_2 z(t)^2)\big)\mathrm{d}t + \sigma z(t)\mathrm{d}w(t), \quad z(0) = z_0, \tag{7}$$

where the replenishment (production) rate $u > 0$, is constant given, thus yielding a replenishment process $s(t) = s(0) + ut$, within an inventory cycle; now, setting $s(0) = 0$, as usual, we get a simple linear production function, and finally, $z(0) = z_0$, gives an initial level of inventory at $t = 0$. Another interesting model of $z(t)$, is defined by the SDE, with given d_1, d_2, u, σ, which takes the following form

$$\mathrm{d}z(t) = \Big(d_1 z(t) - \frac{d_2}{\exp(\frac{\sigma^2 t}{2} - \sigma w(t))} z(t)^2\Big)\mathrm{d}t - u \exp\Big(\frac{\sigma^2 t}{2} - \sigma w(t)\Big)\mathrm{d}w(t), \quad z(0) = z_0. \tag{8}$$

Finally, the SDE which serves us to simulate paths of both $x(t)$, and $z(t)$, with given μ, σ, has following form

$$\mathrm{d}x(t) = -\mu x(t)\mathrm{d}t + \sigma\sqrt{(1 + x(t)^2)}\mathrm{d}w(t), \quad x(0) = x_0. \tag{9}$$

3 Numerical results – simulations

Basic functions for generation stochastic processes in sw Mathematica®, Wolfram Research Inc., are following:

- ItoProcess[*sdeqns*,*expr*,$x, t, w \approx$ *dproc*], which represents an Itô process specified by a SDE *sdeqns* driven by *w* following the process *dproc*, e.g. WienerProcess[ ], output expression *expr*, with state *x*, and time *t*.

- RandomFunction[*proc*, $\{t_{min}, t_{max}, dt\}, n$, Method], which generates an ensamble of *n* pseudorandom functions from t_{min} to t_{max} in steps of *dt* from the process *proc* using specified method. Let ξ_t, denote a generated path, and $\{\xi_{t,k}, k = 1, \ldots, n\}$, a sample of *n* paths.

In [4], and [6], the numerical methods for solution of SDE are presented, and in particular, the numerical approach for detecting a *first-passage-time*, called FPT is discussed, too.

This quantity playes an important role in inventory control and management, because of determining safety stock, in particular. Following [4], the FPT of a stochastic process $y(t)$ reaching the barrier b is defined by

$$h_b = \inf\{t > 0 : y(t) = b\}. \tag{10}$$

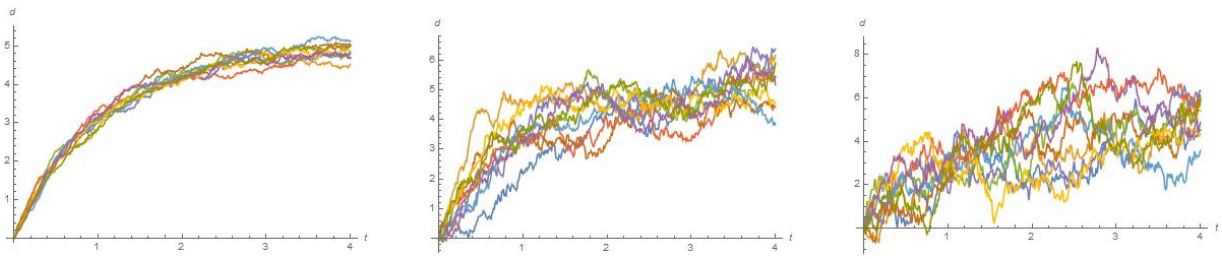


Figure 2 Sample paths of model (5), $\mu = 1$, diffusion part $\sigma \mathrm{d}w(t)$: Left: $\sigma = .25$, Middle: $\sigma = 1$, Right: $\sigma = 2$.

In the case of Wiener process $w(t)$, the FPT can be directly calculated by probabilities for given $b > 0$, using the up-down symmetry of Wiener increments. Let $t > 0$, then

$$\Phi(t,b) = \mathcal{P}[h_b < t] = \mathcal{P}[h_b < t, w(t) < b] + \mathcal{P}[h_b < t, w(t) > b] = \sqrt{(\frac{2}{\pi t})} \int_b^\infty e^{-\frac{s^2}{2t}} \mathrm{d}s = 1 - \mathrm{erf}(\frac{b}{\sqrt{2t}}),$$

and the corresponding density is obtained by differentiating $\Phi(t,b)$with repect to t and integration per-partes

$$\varphi(t,b) = \frac{\mathrm{d}\Phi(t,b)}{\mathrm{d}t} = \frac{|b|}{\sqrt{(2\pi t^3)}} e^{-\frac{b^2}{2t}}. \tag{11}$$

We adopted a method of moving particle along a path of stochastic process which was proposed in [4]. Our algoritms has following steps:

1. prepare data: set sample of paths, set value of barrier b,
2. convert set of paths into a numeric matrix apathVals $(q_{i,k})$: a column corresponds to individual path, a row element stores an actual value of the path,
   Mma command: $q_{i,k}$ = apath["SliceData",k],{i,*length*}, $k \ldots$ path index, *length* $\ldots$ number of values in the path,
3. loop over all paths in the sample: select k, set initial value of output variable $\kappa = 0$,
4. loop over all values in the k-th path $(i = 0, length, i++)$: **if** $(q_{i,k} \leq b \,\&\, q_{i+1,k} \geq b)$ **then** $(\kappa = i, goto\ End)$ **else** (set: $\kappa = i + 1$),
5. *End*: return κ.

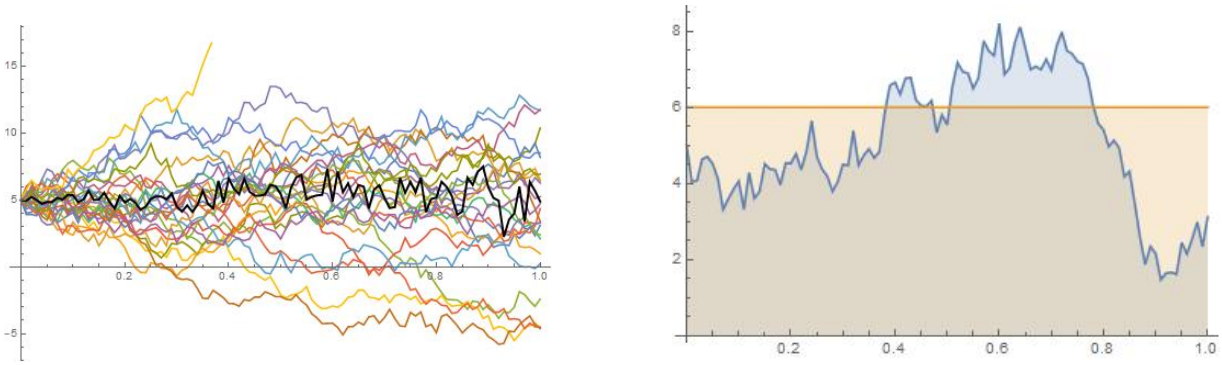


Figure 3 Sample paths of model (8): Left: 25 paths with sample mean (black), Right: a path with barrier $b = 6$.

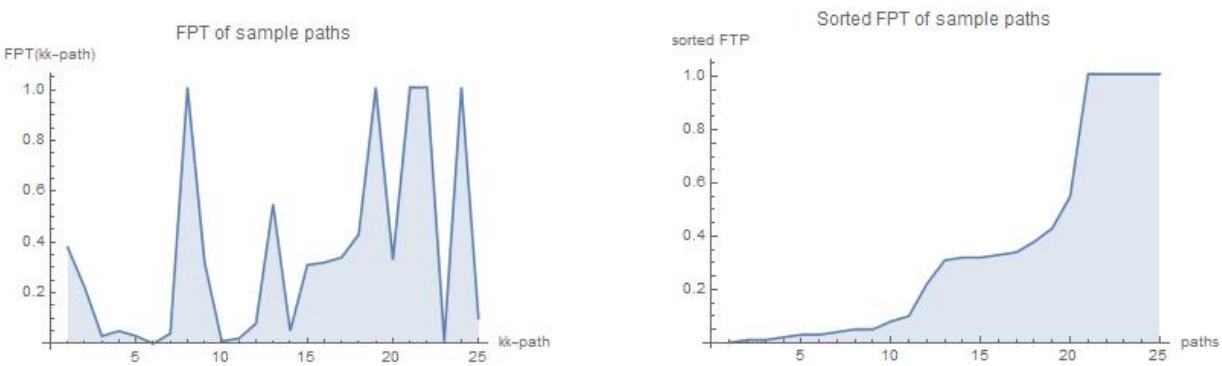


Figure 4 Calculated FPT κ, for $b = 6$, from sample paths of model (8): Left: unsorted κ, Right: sorted κ.

Data for model (8): $d_1 = 0.2, d_2 = 0.005, u = 5, \sigma = 0.1, z_0 = 5$. In Fig. 3, the simulated paths $\{\xi_{t,k}\}$, using the model (8) are displayed – on the left: 25 paths and sample mean $\mu_{t,n} = \mathrm{E}[\{\xi_{t,k}, k = 1, \ldots, n\}]$; on the right: a path ξ_t, with depicted barrier $b = 6$, which suggests a FPT value κ. In Fig. 4, the calculated FPT values from sample paths $\{\xi_{t,k}\}$, of model (8) are displayed – on the left: calulated values $\kappa_k, k = 1, \ldots, n$;

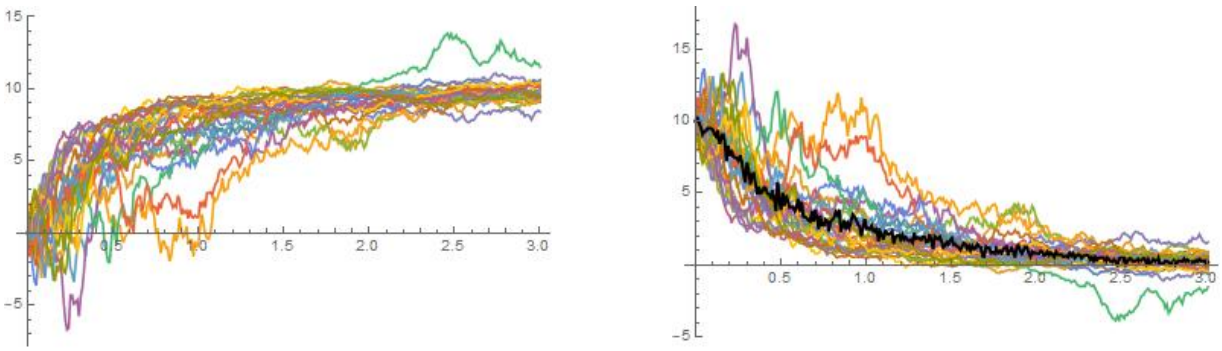


Figure 5 Sample paths of model (9): Left: 25 paths of $x(t)$, Right: 25 paths of $z(t)$, with sample mean (black).

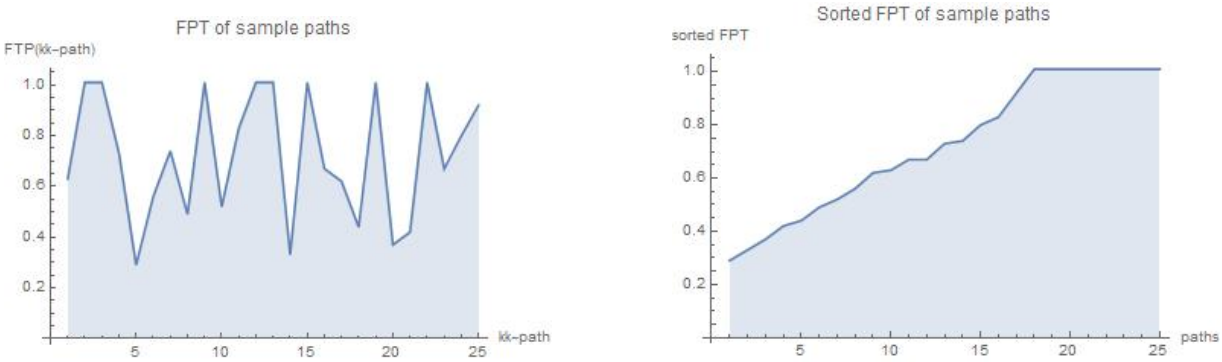


Figure 6 Calculated FPT κ, for $b = 7.5$, from sample paths of model (9): Left: unsorted κ, Right: sorted κ.

on the right: sorted values of κ_k. If $\kappa_k > t_{max} - t_{min}$, then it means that a pacticle moving along a k-the path $\xi_{t,k}$, during the given time span $t_{max} - t_{mix}$, does not meet the barrier b, at all. Data for model (9): $\mu = 1.25, \sigma = 0.75, x_0 = 0$. In Fig. 5, sample paths from process of model (9) are displayed – on the left: paths of $x(t)$; on the right: path of $z(t)$, with sample mean $\mathrm{E}[\{\xi_{t,k}, k = 1, \ldots, n\}]$. In Fig. 6, the calculated FPT values from sample paths $\{\xi_{t,k}\}$, of model (9) are displayed – on the left: calulated values $\kappa_k, k = 1, \ldots, k_n$; on the right: sorted values of κ_k.

4 Conclusions

We have discussed stochastic dynamic inventory models, and stochastic demand processes, in particular. We also construct a numerical procedure for detecting first-passage-time when a stochastic process meets the given barrier.

Acknowledgements

In particular, we acknowledge the financial support of our research by the grant SGS2017-004 "Finance and sustainable development from viewpoint of theory and practice", registered in and provided by the Internal Grant System of the University of West Bohemia in Pilsen, Czech Republic.

References

[1] Axsaeter, S. (2006). *Inventory Control*. 2. ed. New York: Springer.

[2] Hofman, J. & Lukáš, L. (2015). Generalized EOQ Type Inventory Models with Time Dependent both Demand Rate and Holding Cost Rate. In: *Conf. Proceedings, 33-th Int. Conf. Mathematical Methods in Economics 2015*, (Martinčík, D., Ircingová, J., Janeček, P. Eds.). Univ. of West Bohemia in Pilsen, Cheb, 2015, 249–255.

[3] Evans, L. C. (2006). *An Introduction to Stochastic Differential Equations*. ver. 1.2. Berkeley: UC Berkeley,

Dept. of Mathematics, http://ft-sipil.unila.ac.id/dbooks/AN INTRODUCTION TO STOCHASTIC DIFFERENTIAL EQUATIONS VERSION 1.2.pdf.

[4] Jansons, K. & Lythe, G. D. (2000). Efficient Numerical Solution of Stochastic Differential Equations Using Exponential Timestepping. *Journal of Statistical Physics*, 100(5/6), 1097–1109.

[5] Lukáš, L. (2012). *Probabilistic models in management – Theory of inventory and statistical description of demand*. (in czech: *Pravděpodobnostní modely v managementu – Teorie zásob a statistický popis poptávky*). Praha: Academia.

[6] Nyberg, M., Ambjörnsson, T. & Lizana, L. (2016). A simple method fo calculate first-passage-time densities with arbitrary initial conditions. *New Journal of Physics*, 18: 063019, http://dx.doi.org/10.1088/1367-2630/18/6/063019.

[7] Øksendal, B. (2003). *Stochastic Differential Equations An Introduction with Applications*. 5-th ed. Heidelberg New York: Springer-Verlag.

[8] Tsoularis, A. (2019). A Stochastic Differential Equation Inventory Model. *Int. J. Appl. Comput. Math.*, (2019) 5:8, https://doi.org/10.1007/s40819-018-0594-7.

[9] Urban, T. L. (1992). An Inventory Model with an Inventory-Level-Dependent Demand Rate and Relaxed Terminal Conditions. *J. Opl Res. Soc.*, 43 (7), 721–724.

[10] Urban, T. L. (2005). Inventory models with inventory-level-dependent demand: a comprehensive review and unifying theory. *Eur.J. Oper. Res.*, 162 (3), 792–804.

Resource-Workflow Petri Nets with Time Stamps and Their Using in Process Management

Ivo Martiník[1]

Abstract. Resource-workflow Petri nets with time stamps (RWPNTS) are the newly introduced class of low-level process Petri nets in this article. RWPNTS are specifically designed for their use in the process management area for modeling processes consisting of individual time-dependent activities and for determining their time-optimal critical paths. The critical path is defined as the (temporally) longest possible path from the process start point to the process end point and the Critical Path Method (CPM) is used by default for its determination. The new RWPNTS class of process Petri nets allows not only modeling processes composed of individual time-dependent activities and determining their critical paths, but, in addition to standard CPM method, it enables the analysis of other process properties with using of well-known methods of Petri nets theory and it also has the possibility to determine the time-optimal critical path of the given process whose activities share a finite set of different resources that are necessary to successfully complete any of them. These properties of RWPNTS class are demonstrated on the simple process example in this article.

Keywords: resource-workflow Petri nets with time stamps, critical path method, discrete time, process management, resource sharing

JEL Classification: C610
AMS Classification: 90C15

1 Introduction

The number of classes of *Petri nets* [2] are formally defined and studied for the modeling of generally parallel systems of all types at the present time. Modeling of time variables associated with individual system events, time duration of activities, time history of modeled system and many other time characteristics plays an important role in the study of many types of economic systems. Time Petri nets and timed Petri nets [2, 11] are the two most important classes of low-level Petri nets that use the concept of discrete time in their definition. It can be stated for these and many other classes of discrete-time Petri nets [3, 4, 8] that most of them use only relative time variables usually related to the specific labeling of the given Petri net. This can then cause problems, for example, when modeling complex time-synchronized systems in which their individual components must be synchronized with the given external time source.

Workflow Petri nets (WPN) [1] were defined as the subclass of low-level process Petri nets [5] for their using primarily in the area of workflow management. WPN is a connected Petri net that includes within the set of all its places the unique input place with no input arcs and the unique output place with no output arcs. *Resource-workflow Petri nets with time stamps* (RWPNTS) are the newly introduced class of low-level process Petri nets whose definition and properties are the main topics of this article. The class of RWPNTS is based on the definition of the *Workflow Petri nets with time stamps* (WPNTS) [7] class of low-level process Petri nets and they are specifically designed for their use in the process management area for modeling processes consisting of individual time-dependent activities and for determining their critical paths. The critical path is defined as the (temporally) longest possible path from the process start point to the process end point. Each critical path consists of the list of activities that the project manager should focus most on to ensure timely completion of the given process. The completion time of the last mission on the critical path is also the process completion time. The Critical Path Method (CPM) [9, 10] is used by default to determine the critical path of the given process. The new RWPNTS class allows not only modeling processes composed of individual time-dependent activities and determining their critical paths, but, in addition to standard CPM method, it also enables the analysis of other process properties with using of standard methods of Petri nets analysis and it also has the possibility to determine the time-optimal critical path of the given process whose activities share a finite set of different resources [6] that are necessary to successfully

[1] VSB-Technical University of Ostrava/Faculty of Economics, Department of Applied Informatics, 17. listopadu 2172/15, 708 00 Ostrava-Poruba, ivo.martinik@vsb.cz.

complete any of them. These properties of RWPNTS class are then demonstrated on the simple process example and its time-optimal critical path specification in this article.

2 Resource-Workflow Petri nets with time stamps and their properties

Let $\boldsymbol{N}$ denotes the set of all *natural numbers*, $\boldsymbol{N} := \{1, 2, \ldots\}$; $\boldsymbol{N}_0$ denotes the set of all *non-negative integer numbers*, $\boldsymbol{N}_0 := \{0, 1, 2, \ldots\}$; $\emptyset$ denotes the *empty set*; $\boldsymbol{N}_\emptyset := \boldsymbol{N}_0 \cup \{\emptyset\}$, i.e., $\boldsymbol{N}_\emptyset := \{\emptyset, 0, 1, 2, \ldots\}$; $|A|$ denotes the *cardinality* of the given set A; $A \times B := \{(x, y) \mid (x \in A) \wedge (y \in B)\}$ denotes the *Cartesian product* of the sets A and B; $\ulcorner$ denotes the *logical negation* operator; $f: A \rightarrow B$ denotes the *function* f with the *domain* A and the *codomain* B; by the (non-empty finite) *sequence* σ over the non-empty set A we understand the function σ: $\{1, 2, \ldots, n\} \rightarrow A$, where $n \in \boldsymbol{N}$, that is denoted by $\sigma := \langle a_1, a_2, \ldots, a_n \rangle$, where $a_i = \sigma(i)$ for $1 \le i \le n$; if $\sigma := \langle a_1, a_2, \ldots, a_n \rangle$ is the sequence, then the set $\boldsymbol{ELEMS}_\sigma := \{a \mid \exists i, 1 \le i \le n: a = \sigma(i)\}$. Let $(A \subset \boldsymbol{N}_0) \wedge (\exists n \in \boldsymbol{N}: |A| = n) \wedge (A \neq \emptyset)$; then $max(A) := x$, where $(x \in A) \wedge (\forall y \in A: x \ge y)$.

Definition 1. *Net NET* is the ordered triple $NET := (P, T, A)$, where P is finite non-empty set of the *places*; T is finite set of the *transitions*, $P \cap T = \emptyset$; and A is finite set of the *arcs*, $A \subseteq (P \times T) \cup (T \times P)$. □

Some commonly used notations for the nets are $\bullet y = \{x \mid (x, y) \in A\}$ for the *preset*, $\bullet\bullet y = \{x \mid (x \in \bullet z) \wedge (z \in \bullet y)\}$ for the *prepreset* and $y\bullet = \{x \mid (y, x) \in A\}$ for the *postset* of the net node y (i.e., place or transition). The *path* leading from the node x_1 to the node x_k of the net is the non-empty sequence $\langle x_1, x_2, \ldots, x_k \rangle$ of net nodes, where $k \in \boldsymbol{N}$, which satisfies $(x_1, x_2), (x_2, x_3), \ldots, (x_{k-1}, x_k) \in A$. We will denote the set of all such paths of the given net *NET* by $\boldsymbol{PATHS}_{NET}(x_1, x_k)$. The path of the net leading from its node x to its node y is the *circuit* if $(y, x) \in A$. We will denote the set of all the circuits of the given net *NET* by $\boldsymbol{CIRCUITS}_{NET}$.

Definition 2. *Resource-workflow net with time stamps* (RWNTS) *RWNET* is the ordered tuple $RWNET := (P, T, A, TI, IP, OP, RP)$, where:

1. (P, T, A) is the *net*;
2. $TI: P \rightarrow \boldsymbol{N}_0$ is the *time interval function*;
3. IP is the *input place*, $(IP \in P) \wedge (\bullet IP = \emptyset) \wedge (|IP\bullet| \le 1) \wedge (TI(IP) = 0)$;
4. OP is the *output place*, $(OP \in P) \wedge (OP\bullet = \emptyset) \wedge (|\bullet OP| \le 1) \wedge (TI(OP) = 0)$;
5. RP is finite set of the *resource places*, $(RP \subseteq (P \setminus \{IP, OP\})), \forall rp \in RP: (TI(rp) = 0) \wedge (|\bullet rp| = |rp\bullet|)$;
6. $\forall p \in \underline{P}: (|\bullet p| = 1) \wedge (|p\bullet| = 1)$, where $\underline{P} := P \setminus (RP \cup \{IP, OP\})$;
7. $\forall rp \in RP\ \forall t \in rp\bullet\ \exists p \in \underline{P}\ \exists s \in \bullet rp: (t\bullet = \{p\}) \wedge (\bullet s = \{p\})$;
8. $\forall x \in (P \cup T)\ \exists \sigma \in \boldsymbol{PATHS}_{RWNET}(IP, OP): x \in \boldsymbol{ELEMS}_\sigma$;
9. $\forall \sigma \in \boldsymbol{CIRCUITS}_{RWNET}\ \exists rp \in RP: rp \in \boldsymbol{ELEMS}_\sigma$. □

The RWNTS *RWNET* consists of the net (P, T, A); the *time interval function TI* then assigns with each place p the non-negative integer number d (with the default value of 0 in the net diagram) that expresses the minimum time interval during which the token has to remain in the place p instead of being able to participate in the next firing of some transition; the *input place IP* is the only one place of RWNTS *RWNET* with no input arc(s) and with no more than one output arc; the *output place OP* that is the only one place of RWNTS *RWNET* with no output arc(s) and with no more than one input arc; the finite set *RP* of the *resource places* is used for expressing the conditions of the modeled process containing some initial resources and we use circles with the double line for their representation; the number of all the input arcs of every resource place rp is the same as the number of all the output arcs of this resource place rp. It must also be fulfilled for the RWNTS *RWNET* that every of its non-input, non-output and non-resource places has exactly one input arc and one output arc; every input arc and every output arc of the given resource place rp must fulfill the condition 7 of Definition 2; every node x (i.e., place or transition) must be the element of some path σ from the input place *IP* to the output place *OP* (i.e., every RWNTS is the connected net); and every circuit of the given RWNTS *RWNET* must contain at least one resource place rp.

Definition 3. Let $RWNET := (P, T, A, TI, IP, OP, RP)$ be the RWNTS. Then:

1. *marking M* of the RWNTS *RWNET* is the function $M: P \rightarrow \boldsymbol{N}_\emptyset$;
2. variable $\tau \in \boldsymbol{N}_0$ is the *net time* of the RWNTS *RWNET*;
3. *state S* of the RWNTS *RWNET* is the ordered pair $S := (M, \tau)$;
4. transition $t \in T$ is *enabled* in the state $S := (M, \tau)$ of the RWNTS *RWNET* that is denoted by t ***en*** S, if $\forall p \in \bullet t: (M(p) \neq \emptyset) \wedge (M(p) \le \tau)$;

5. *firing of the transition* $t \in T$ results in changing the state $S := (M, \tau)$ of the RWNTS *RWNET* into its state $S' := (M', \tau)$ that is denoted by $S\ [t\rangle\ S'$, where $\forall p \in \bullet t$: $M'(p) := \emptyset$; $\forall p \in t\bullet$: $M'(p) := \tau + TI(p)$; $\forall p \in P \setminus (\bullet t \cup t\bullet)$: $M'(p) := M(p)$; we denote the net time value τ at which the transition t fired by $FT(t)$;
6. *elapsing of time interval* $\delta \in \boldsymbol{N}$ results in changing the state $S := (M, \tau)$ of the RWNTS *RWNET*, where $\forall t \in T$: $\ulcorner(t\ \boldsymbol{en}\ S)$, into its state $S' := (M, \tau + \delta)$ that is denoted by $S\ [\delta\rangle\ S'$, so that:
$$(\forall t \in T\ \forall n \in \boldsymbol{N}, 1 \le n < \delta: \ulcorner(t\ \boldsymbol{en}\ (M, \tau + n))) \wedge (\exists t \in T: t\ \boldsymbol{en}\ (M, \tau + \delta));$$
7. if the mutually different transitions $t_1, t_2, \ldots, t_n \in T$ are enabled in the state $S := (M, \tau)$ of the RWNTS *RWNET* (i.e., $(t_1\ \boldsymbol{en}\ S) \wedge (t_2\ \boldsymbol{en}\ S) \wedge \ldots \wedge (t_n\ \boldsymbol{en}\ S)$) we say that these transitions are *enabled in parallel* in the state S that is denoted by the statement $\{t_1, t_2, \ldots, t_n\}\ \boldsymbol{en}\ S$;
8. finite non-empty sequence $\sigma := \langle t_1, t_2, \ldots, t_n\rangle$ of the transitions $t_1, t_2, \ldots, t_n \in T$ for which the following is valid in the state $S_1 := (M_1, \tau_1)$ of the RWNTS *RWNET*:
   a) $(M_1, \tau_1)\ [t_1\rangle\ (M_2, \tau_1)\ [t_2\rangle \ldots [t_n\rangle\ (M_{n+1}, \tau_1)$,
   b) $\forall t \in T$: $\ulcorner(t\ \boldsymbol{en}\ (M_{n+1}, \tau_1))$,
   is called *step* σ in the given state S_1 of the RWNTS *RWNET* and it is denoted by $(M_1, \tau_1)\ [\sigma\rangle\ (M_{n+1}, \tau_1)$;
9. finite non-empty sequence ρ of steps and time intervals elapsing that represents the state changes
$$(M_1, \tau_1)\ [\sigma_1\rangle\ (M_2, \tau_1)\ [\delta_1\rangle \ldots (M_{n+1}, \tau_n)\ [\delta_n\rangle\ (M_{n+1}, \tau_{n+1})$$
   of the RWNTS *RWNET* is the sequence $\rho := \langle\sigma_1, \delta_1, \sigma_2, \delta_2, \ldots, \sigma_n, \delta_n\rangle$ of the steps $\sigma_1, \sigma_2, \ldots, \sigma_n$ and the time intervals elapsing $\delta_1, \delta_2, \ldots, \delta_n$;
10. we say that the state S' of the RWNTS *RWNET* is reachable from its state S if there exists the finite sequence $\rho := \langle\sigma_1, \delta_1, \sigma_2, \delta_2, \ldots, \sigma_n, \delta_n\rangle$ of the steps $\sigma_1, \sigma_2, \ldots, \sigma_n$ and the time intervals elapsing $\delta_1, \delta_2, \ldots, \delta_n$ such that $S\ [\sigma_1\ \delta_1\ \sigma_2\ \delta_2 \ldots \sigma_n\ \delta_n\rangle\ S'$;
11. *input state* $S_i := (M_i, 0)$ of the RWNTS *RWNET* is such state where $(M_i(IP) = 0) \wedge (\forall p \in (P \setminus (RP \cup \{IP\})): M_i(p) = \emptyset) \wedge (\forall rp \in RP: M_i(rp) = 0)$;
12. *otput state* $S_o := (M_o, \tau_o)$ of the RWNTS *RWNET* that is reachable from its input state $S_i := (M_i, 0)$ is every of its states S_o where $(S_o \in [S_i\rangle) \wedge (M_o(OP) \neq \emptyset) \wedge (\forall p \in (P \setminus (RP \cup \{OP\})): (M_o(p) = \emptyset)) \wedge (\forall p \in RP: M_o(p) \neq \emptyset)$.
13. the set of all the output states $S_o := (M_o, \tau_o)$ of the RWNTS *RWNET* that are reachable from its input state $S_i := (M_i, 0)$ is denoted by $[S_i\rangle_o$. □

Definition 4. *Resource-Workflow Petri net with time stamps* (RWPNTS) *RWPNET* is the ordered couple *RWPNET* := (*RWNET*, S_i), where *RWNET* := (P, T, A, TI, IP, OP, RP) is the RWNTS and $S_i := (M_i, 0)$ is the input state of the RWNTS *RWNET*. □

3 RWPNTS and their applications in process management

Critical Path Method (CPM) [9, 10] is one of the basic deterministic methods of the network analysis. It is used for straightforward processes where time durations of their activities can be estimated with a high degree of precision. Its aim is to determine the time duration of the given process on the basis of the length of the so-called critical path, which is the sequence of interdependent activities with the least time reserve. The critical path is then defined as the temporally longest possible path from the process start point to the process end point. The completion time of the last mission on the critical path is also the process completion time. The theory of RWPNTS can be then successfully applied especially to modeling processes and determining their critical paths. We will present that RWPNTS can also be used to generalize the original process model that is used by the CPM method, further considering that each activity may require a finite set of so-called resources (e.g., energy, financial, material, etc.) to be successfully completed and every of these resources can be also used by other activities. The simple resource-sharing model in which each of these resources is represented by the separate token 0 that is located at the appropriate resource place in the input state S_i of the given RWPNTS will be studied in this article (i.e., each of these resources is immediately available to the appropriate activity for its completion at the input state S_i of RWPNTS and also for its further use by another activity once the previous activity has been successfully completed).

The simple example of such process whose activities need to share selected resources for their successful completion is described in the following table of its activities (see Table 1). The RWPNTS *PROC* = (P, T, A, TI, IP, OP, RP) that represents the process comprising the activities with the shared resources listed in Table 1 can be seen in Figure 1, where (for instance) P := {**IP**, **A1**, **A**, **A2**, **B1**, **B**, **B2**, **C1**, **C**, **C2**, **D**, **E**, **F**, **F2**, **G**, **R1**, **R2**, **R3**, **OP**}; T := {**T1**, **TA1**, **TA2**, **TB1**, **TB2**, **TC1**, **TC2**, **TF1**, **TF2**, **T2**, **T3**, **T4**}; TI := {(**IP**, 0), (**A1**, 0), (**A**, 6), (**A2**, 0), (**B1**, 0), (**B**, 3), (**B2**, 0), (**C1**, 0), (**C**, 2), (**C2**, 0), (**D**, 5), (**E**, 7), (**F**, 4), (**F2**, 0), (**G**, 8), (**R1**, 0), (**R2**, 0), (**R3**, 0),

(**OP**, 0)}; IP := **IP**; OP := **OP**, RP := {**R1**, **R2**, **R3**}. RWPNTS *PROC* is in its input state S_i := (M_i, 0), where marking M_i := (M_i(**IP**), M_i(**A1**), M_i(**A**), M_i(**A2**), M_i(**B1**), M_i(**B**), M_i(**B2**), M_i(**C1**), M_i(**C**), M_i(**C2**), M_i(**D**), M_i(**E**), M_i(**F**), M_i(**F2**), M_i(**G**), M_i(**R1**), M_i(**R2**), M_i(**R3**), M_i(**OP**)) = (0, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, 0, 0, 0, ∅), net time τ = 0 (i.e., $\tau_i = \tau$). Marking M of any RWPNTS expresses the current time state of the modeled system using the time stamp associated with each of its token that occurs in any of its places in its state S. Individual values of the time stamps associated with the tokens, informally said, represent the values of the net time τ at which the respective token can first participate in the firing of the selected enabled transition t of the RWPNTS. The places **A**, **B**, **C**, **D**, **E**, **F** and **G** represent individual activities of the studied process, the appropriate values of the time interval function *TI* then express the time durations of individual activities (i.e., for instance, the activity A is represented by the place **A**, its time duration is given by the value of TI(**A**) = 6 and it has no previous activities with non-zero time duration because TI(**IP**) = 0 and TI(**A1**) = 0). Each of the resource places **R1**, resp. **R2**, resp. **R3**, then include the token 0 that models the shared resource R1, resp. R2, resp. R3. All of the resources can be used immediately after the starting of the whole process (because every of them is modeled by the token 0) and they all will be also immediately available for its further use by another activity once the previous activity has been successfully completed (because TI(**R1**) = 0, TI(**R2**) = 0 and TI(**R3**) = 0).

The transition **T1** is enabled in the input state S_i of the RWPNTS *PROC* because $\forall p \in \bullet$**T1**: ($M(p) \neq \emptyset$) ∧ ($M(p) \leq 0$), i.e., ($\bullet$**T1** = {**IP**}) ∧ (M_e(**IP**) = 0 ≤ 0 = τ) (see 4 of Definition 3). Firing of the transition **T1** changes the input state S_i := (M_i, 0) of the RWNTS *PROC* into its state S_1 := (M_1, 0) = ((∅, 0, ∅, ∅, 0, ∅, ∅, 0, ∅, ∅, ∅, ∅, ∅, ∅, ∅, 0, 0, 0, ∅), 0) (i.e., S_i [**T1**⟩ S_1), where (see 5 of Definition 3) M_1(**IP**) := ∅; M_1(**A1**) := τ_i + TI(**A1**) = 0 + 0 = 0; M_1(**B1**) := τ_i + TI(**B1**) = 0 + 0 = 0; M_1(**C**) := τ_i + TI(**C1**) = 0 + 0 = 0. The transitions **TA1**, **TB1** and **TC1** are enabled in the state S_1 and so called *conflict* occurs at their enabling, i.e., both transitions **TA1** and **TB1** have at least one common place **R2** in their presets, each of the these transitions is individually enabled in the state S_1, but these transitions are not enabled in parallel in this state (see 7 of Definition 3) and enabling of one of them will prevent enabling of the other (i.e., ($\bullet$**TA1** ∩ $\bullet$**TB1** ≠ ∅) ∧ (**TA1** ***en*** S_1) ∧ (**TB1** ***en*** S_1) ∧ ⌐({**TA1**, **TB1**} ***en*** S_1)). An analogous statement then also applies to the pair of the transitions **TB1** and **TC1.** Since no explicit rule has been defined which of the conflicting transitions **TA1**, **TB1** and **TC1** to fire in this case, we will randomly select the **TA1** and **TC1** transitions to fire. Firing of the transition **TA1** changes the state S_1 := (M_1, 0) into the state S_2 := (M_2, 0) = (((∅, ∅, 6, ∅, 0, ∅, ∅, 0, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, 0, ∅), 0), 0) (i.e., S_1 [**TA1**⟩ S_2) and then firing of the transition **TC1** changes the state S_2 := (M_2, 0) into the state S_3 := (M_3, 0) = (((∅, ∅, 6, ∅, 0, ∅, ∅, ∅, 2, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅), 0), 0) (i.e., S_2 [**TC1**⟩ S_3) (see Figure 1). There is no enabled transition in the state S_3 := (M_3, 0) because the token in the place **A** can first participate in firing of the transition **TA2** if the net time τ = 6 and the token in the place **C** can first participate in firing of the transition **TC2** if the net time τ = 2. So it is necessary to perform the time interval elapsing with the value of δ = 2 to enable the firing of the transition **TC2**. This will change the state S_3 := (M_3, 0) into the state S_4 := (M_3, 2), i.e., S_3 [2⟩ S_4) and the transition **TC2** can be then fired in this state. We can then write that (see 8 of Definition 3) S_i [**T1 TA1 TC1**⟩ S_3 [2⟩ S_4.

Activity	Time duration	Shared resources	Previous activities
A	6	R1, R2	—
B	3	R2, R3	—
C	2	R3	—
D	5	—	C
E	7	—	C
F	4	R1	A
G	8	—	B, D

Table 1 Table of activities, their time durations, dependencies and shared resources

Classic methods used for the analysis of Petri nets based on the study of the set of all the reachable states of the given RWPNTS can be used for determining the time-optimal critical path of the process represented by given RWPNTS. If $[S_i\rangle_o = \{S_o := (M_o, \tau_o) \mid \exists\rho = <\sigma_1, \delta_1, \sigma_2, \delta_2, \ldots, \sigma_n, \delta_n>: S_i\,[\sigma_1\,\delta_1\,\sigma_2\,\delta_2 \ldots \sigma_n\,\delta_n\rangle\, S_o, n \in \mathbf{N}_0\}$, it is then necessary to determine such sequence $\rho_{min} = <\sigma_1, \delta_1, \sigma_2, \delta_2, \ldots, \sigma_n, \delta_n>$ for which the output state S_o := (M_o, τ_o)

has the minimal value of the net time τ_o, i.e., we must find such output state $S_o := (M_o, \tau_{min})$ where $\tau_{min} = min(\{\tau_o \mid (M_o, \tau_o) \in [S_i\rangle_o\}$. We will demonstrate these concepts for the case of RWPNTS *PROC* (see Figure 1).

It can be easily determined that there are two different finite sequences ρ_1 and ρ_2 of the steps and the time intervals elapsing that represent the following state changes of RWPNTS *PROC*:

1. $\rho_1 = (M_i, 0)$ [**T1 TA1 TC1**⟩ $(M_3, 0)$ [2⟩ $(M_3, 2)$ [**TC2 T3**⟩ $(M_5, 2)$ [4⟩ $(M_5, 6)$ [**TA2 TB1 TF1**⟩ $(M_8, 6)$ [3⟩ $(M_8, 9)$ [**TB2 T2**⟩ $(M_{10}, 9)$ [1⟩ $(M_{10}, 10)$ [**TF2**⟩ $(M_{11}, 10)$ [7⟩ $(M_{11}, 17)$ [**T4**⟩ $(M_o, 17)$,
   where $M_3 = (\varnothing, \varnothing, 6, \varnothing, 0, \varnothing, \varnothing, \varnothing, 2, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing)$, $M_5 = (\varnothing, \varnothing, 6, \varnothing, 0, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 2, 2, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 2, \varnothing)$, $M_8 = (\varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 9, \varnothing, \varnothing, \varnothing, \varnothing, 2, 2, 10, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing)$, $M_{10} = (\varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 2, 10, \varnothing, 17, 9, 9, \varnothing, \varnothing)$, $M_{11} = (\varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 2, \varnothing, 10, 17, 9, 9, 10, \varnothing)$, $M_o = (\varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 9, 9, 10, 17)$;
2. $\rho_2 = (M_i, 0)$ [**T1 TB1**⟩ $(M_2, 0)$ [3⟩ $(M_2, 3)$ [**TB2 TA1 TC1**⟩ $(M_5, 3)$ [2⟩ $(M_5, 5)$ [**TC2 T3**⟩ $(M_7, 5)$ [4⟩ $(M_7, 9)$ [**TA2 TF1**⟩ $(M_9, 9)$ [1⟩ $(M_9, 10)$ [**T2**⟩ $(M_{10}, 10)$ [3⟩ $(M_{10}, 13)$ [**TF2**⟩ $(M_{11}, 13)$ [5⟩ $(M_{11}, 18)$ [**T4**⟩ $(M_o, 18)$,
   where $M_2 = (\varnothing, 0, \varnothing, \varnothing, \varnothing, 3, \varnothing, 0, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 0, \varnothing, \varnothing, \varnothing)$, $M_5 = (\varnothing, \varnothing, 9, \varnothing, \varnothing, \varnothing, 3, \varnothing, 5, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing)$, $M_7 = (\varnothing, \varnothing, 9, \varnothing, \varnothing, \varnothing, 3, \varnothing, \varnothing, \varnothing, 10, 12, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 5, \varnothing)$, $M_9 = (\varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 3, \varnothing, \varnothing, \varnothing, 10, 12, 13, \varnothing, \varnothing, 9, 9, 5, \varnothing)$, $M_{10} = (\varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 12, 13, \varnothing, 18, 9, 9, 5, \varnothing)$, $M_{11} = (\varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 12, \varnothing, 13, 18, 9, 9, 5, \varnothing)$, $M_o = (\varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, \varnothing, 9, 9, 5, 18)$.

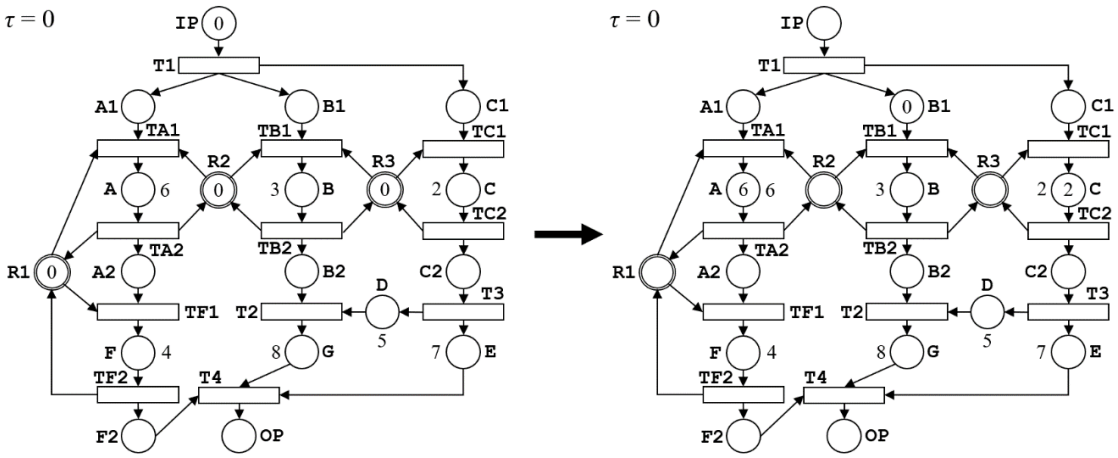


Figure 1 RWPNTS *PROC* in its input state S_i and in the state S_3 (S_i [**T1 TA1 TC1**⟩ S_3)

The above implies that if the sequence ρ_1 is selected then the process represented by the RWPNTS *PROC* will be terminated with the network time value $\tau = 17$ and if the sequence ρ_2 is selected, the process will be terminated with the network time value $\tau = 18$ (i.e., $[S_i\rangle_o = \{(M_o, 17), (M_o, 18)\}$, $\tau_{min} = 17$, $\rho_{min} = \rho_1$). The sequence ρ_1 is then the time-optimal sequence ρ_{min} minimizing overall process time duration.

To apply the algorithm for determining the time-optimal critical path of the given process represented by the relevant RWPNTS it is then necessary first to associate with each transition t of this RWPNTS the value of the net time τ at which the given transition t will be fired (i.e., the value $FT(t)$) within the application of the time-optimal sequence ρ_{min} – see 5 of Definition 3). An example of such an association of the value $FT(t)$ with each transition t for the case of RWPNTS *PROC* and its time-optimal sequence ρ_1 is shown in Figure 2, where RWPNTS *PROC* is in its output state S_o after the application of the time-optimal sequence ρ_1 and the individual value $FT(t)$ associated with each transition t is then shown in bold and underlined type (i.e., FT(**TA1**) = 0, FT(**TB1**) = 6, etc.).

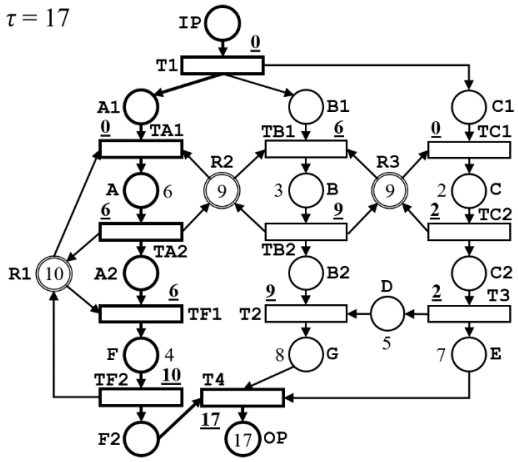


Figure 2 RWPNTS *PROC* in its output state S_o

The procedure for finding the set of nodes of the given RWPNTS that forms the process time-optimal critical path is then obvious. The (nondeterministic) algorithm for finding the sequence **CriticalPath** of the transitions that belong to the time-optimal critical path of the process represented by the given RWPNTS is based on the original algorithm for finding the critical path within the application of the classical CPM method and it is informally expressed as follows (individual steps of the algorithm are demonstrated on the example of RWPNTS *PROC*):

1. **CriticalPath** := <•*OP*> (i.e., **CriticalPath** := <**T4**> in RWPNTS *PROC*);
2. select such a transition *t* from the set of transitions forming the *prepreset* of the first element of the **CriticalPath** sequence which is associated with the maximum value of the net time τ at which the selected transition *t* will be fired (i.e., with the value $FT(t)$), and then put this selected transition *t* as the first element of the **CriticalPath** sequence (i.e., ••**T4** = {**T2, TF2**}; $max(\{FT(\mathtt{T2}), FT(\mathtt{TF2})\}) = FT(\mathtt{TF2}) = 10$; **CriticalPath** := <**TF2**, **T4**> in RWPNTS *PROC*);
3. repeat the step 2 until the first element of the sequence **CriticalPath** is the transition *IP*• (i.e., the transition **T1** in RWPNTS *PROC*).

From the above, it is also clear that the given RWPNTS may contain two or more critical paths with the same total time duration. The set of the transitions that belong to the time-optimal critical path of the process represented by RWPNTS *PROC* is after applying of this algorithm equal to the sequence **CriticalPath** := <**T1**, **TA1**, **TA2**, **TF1**, **TF2**, **T4**> and the whole time-optimal critical path is then represented by the sequence of the places and the transitions <**IP**, **T1**, **A1**, **TA1**, **A**, **TA2**, **A2**, **TF1**, **F**, **TF2**, **F2**, **T4**, **OP**> (see Figure 2).

4 Conclusions

The definition of RWPNTS presented in this article can be further generalized and their modeling capabilities for the project management applications can be enhanced, especially in the areas of modeling the temporary unavailability of individual shared resources, modeling the lower- and upper-time duration of each activity, representing the duration of individual activities in the form of fuzzy numbers, etc. Finding the time-optimal critical path of such a process as well as verifying the properties of the given RWPNTS that models such a process is generally non-trivial problem and the use of Petri nets theory plays a crucial role here.

Acknowledgements

This paper was supported by the Student Grant Competition Project No. SP2020/125 by the Ministry of Education, Science, Research and Sport of the Czech Republic.

References

[1] VAN DER ALST, W. & VAN HEE, K. (2004). *Workflow Management: Models, Methods and Systems. 1st ed.* Massachusetts: MIT Press.

[2] DIAZ, M. (2009). *Petri Nets: Fundamental Models, Verification and Applications. 1st ed*. London: John Willey & Sons.
[3] FURIA, C. A., MANDRIOLI, D., MORZENTI, A., & ROSSI, M. (2012). *Modeling Time in Computing. 1st ed*. Heidelberg: Springer-Verlag.
[4] VAN HEE, K. & SIDOROVA, N. (2013). The Right Timing: Reflections on the Modeling and Analysis of Time. In: COLOM, J. & DESEL, J. (Eds.) *Application and Theory of Petri Nets and Concurrency 2013*, LNCS, vol. 7927. (pp. 1–20). Heidelberg: Springer.
[5] HUANG, H., JIAO, L., CHEUNG, T., & MAK, W., M. (2012). *Property-Preserving Petri Net Process Algebra. 1st ed*. Singapore: World Scientific Publishing.
[6] LIN, CH.-P. & DAI, H.-L. (2014). Applying Petri Nets on Project Management. *Universal Journal of Mechanical Engineering*, 2(8), 249–255.
[7] MARTINIK, I. (2018). Workflow Petri Nets with Time Stamps and Their Using in Project Management. In: SAEED, K. & HOMENDA, W. (Eds.) *Computer Information Systems and Industrial Management, CISIM 2018*, LNCS, vol. 11127. (pp. 193–206). Heidelberg: Springer.
[8] MARTOS-SALGADO, M. & ROSA-VELARDO, F. (2014). Dynamic Networks of Timed Petri Nets. In: CIARDO, G. & KINDLER, E. (Eds.) *Application and Theory of Petri Nets and Concurrency 2014*, LNCS, vol. 8489. (pp. 294–313). Heidelberg: Springer.
[9] NASSAR, A. H. (2018). *Introduction to Project Management Using Critical Path Method. 1st ed*. New York: LAP LAMBERT Academic Publishing.
[10] O'BRIEN, J. J. & PLOTNICK, F. L. (2016). *CPM in Construction Management.* New York: McGraw-Hill.
[11] POPOVA-ZEUGMANN, L. (2013). *Time and Petri Nets. 1st ed*. Heidelberg: Springer-Verlag.

Occurrence and Violation of Transitivity of Preferences in Pairwise Comparisons

Jiří Mazurek[1], Zuzana Neničková[2]

Abstract. Pairwise comparisons are an important tool of modern (multiple criteria) decision making and form a fundamental part of the analytic hierarchy/network process (AHP/ANP), ELECTRE, PROMETHEE and other sophisticated methods. One of the most important features of logically consistent preferences is their transitivity. The aim of this paper is to compare occurrence and violation of transitivity of preferences in pairwise comparisons for pairwise comparisons matrices (PCMs) provided by human experts and PCMs generated randomly (Monte Carlo method) with the use of Saaty's fundamental scale. Our results suggest that the occurrence of transitivity is more frequent for human-made PCMs, while transitivity violation is more frequent for random PCMs. This indicates that transitivity analysis can be used for discrimination between human-made and random PCMs.

Keywords: pairwise comparisons, pairwise comparisons matrix, preferences, transitivity, Monte Carlo method

JEL Classification: C44, C63
AMS Classification:

1 Introduction

The core of pairwise comparisons is that objects are pair-compared to each other and the preferred (better, more eligible) one is determined. It is rather easy to compare only two (and not more) entities at the same time. Results of pairwise comparisons are taken into pairwise comparison matrix. This matrix has some typical properties: it is square and reciprocal. Special kinds of pairwise comparisons (multiplicative, additive and others) give rise to other special attributes of pairwise comparison matrices, e.g. irreducibility, or non-negativity. More about pairwise comparison matrices properties in the next chapter.

Transitivity of preferences should be characteristic to every rational human being. The axiom for transitivity of preferences is part of rational choice theory [2]: given three alternatives X, Y and Z, if an individual prefers X to Y and Y to Z, then X should be preferred to Z. Gisin [4] introduces a general notion of a strict preference relation associated with a preference relation valued in a semi-lattice with difference and shows that if a preference relation is valued in a linear scale then the associated strict preference relation is always transitive. In [15] Switalski investigates violation of transitivity in human preferences represented by fuzzy relations and in [16] he defines general transitivity conditions for fuzzy reciprocal preference relations. Pekala et al. [11] discusses the transitivity problem of Atanassov's intuitionistic fuzzy relations. Gong et al. [5] develops three determination theorems and the corresponding algorithms to judge the weak transitivity of an intuitionistic fuzzy preference relations from different angles. Wang [17] deals with relationships among max-min transitivity, restricted max-min transitivity, quasi-transitivity, weak transitivity, consistency and acyclicity are investigated. Haddenhorst et al. [6] compares three frameworks of generalized transitivity (g-stochastic transitivity, T-transitivity, and cycle-transitivity) and introduces E-transitivity as an even more general notion. Dias et al. [3] analyses the propagation of transitivity from reciprocal preferences on an ordinal scale to the associated crisp strict preference and indifference relations and conversely. Regenvetter et al. [12] reconsiders data from more than 20 studies of intransitive human or animal decision makers and challenges the standard operationalizations of transitive preferences and discusses pervasive methodological problems in the collection, modeling, and analysis of relevant empirical data.

However, the difference between transitivity of human-made preferences and randomly generated preferences were never studied before. Therefore, the aim of this paper is to compare occurrence and violation of transitivity of preferences in pairwise comparisons for pairwise comparisons matrices provided by human

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 733 40 Karviná, mazurek@opf.slu.cz.
[2] Silesian University in Opava, School of Business Administration in Karviná, Department of Informatics and Mathematics, Univerzitní náměstí 1934/3, 733 40 Karviná, nenickova@opf.slu.cz.

experts and pairwise comparison matrices generated randomly (Monte Carlo method) with the use of Saaty's scale.

Paper is organized as follows: Section 2, deals with necessary preliminaries, Section 3 provides the data and the method, in Section 4 results of the presented, and Section 5 provides an application of our approach. Conclusions end the article.

2 Preliminaries

Let G be a nonempty set and let R be a binary relation on R. We say that relation R over G is transitive if and only if aRb and bRc, then aRc for all $a, b, c \in G$. Transitive property is required for other more complex relations such as equivalence relation, preorder, partial order or total ordering.

For simplicity, but without loss of generality, let us consider pairwise comparisons of alternatives. Let X be a given set of n alternatives to be compared. Let a_{ij} denote the decision maker's preference for the i-th alternative over the j-th alternative. Also, we set $a_{ij} > 0; \forall i, j \in \{1,2,\dots,n\}$.

Pairwise comparisons are called reciprocal, if the following property is satisfied:

$$a_{ij} = \frac{1}{a_{ji}}, \forall i, j \in \{1,2,\dots,n\} \tag{1}$$

Property (1) is usually strictly required for pairwise comparisons, hence we will assume it is always satisfied. All pairwise comparisons can be arranged into a square $n \times n$ matrix, $A(a_{ij})$, called a pairwise comparison matrix (PCM):

$$A_{n\times n}(a_{ij}) = \begin{pmatrix} 1 & a_{12} & \dots & a_{1n} \\ a_{21} & 1 & \dots & a_{2n} \\ \dots & \dots & 1 & \dots \\ a_{n1} & a_{n2} & \dots & 1 \end{pmatrix}.$$

Further, pairwise comparisons (a pairwise comparison matrix) are called consistent, if the following property is satisfied:

$$a_{ij} \cdot a_{jk} = a_{ik}; \forall i, j, k \tag{2}$$

The matrix A is consistent, if and only if a priority vector (a vector of weights of all alternatives) $w = (w_1, \dots, w_n)$satisfies the following relation:

$$a_{ij} = \frac{w_i}{w_j}, \forall i, j.$$

However, in real decision-making situations, PCMs are rarely fully consistent.

A priority vector w can be derived via Saaty's eigenvalue method (EM) ([13, 14]):

$$Aw = \lambda_{\max} w \tag{3}$$

where $\lambda_{\max}$is the largest (positive) eigenvalue of A. The existence of the largest (positive) eigenvalue $\lambda_{\max}$of the matrix A is guaranteed by the Perron-Frobenius theorem, see [13, 14]. Usually, the vector w is normalized so that$\|w\| = 1$.

Also, the geometric mean (GM) method (the least squares method) was proposed by Crawford [1] for the derivation of a priority vector w:

$$w_i = \frac{\left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}}{\sum_{i=1}^{n} \left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}} \tag{4}$$

Both methods yield the same result when a pairwise comparison matrix A is consistent. Otherwise, the priority vectors differ slightly, see the comparative study of Ishizaka and Lusti [7].

For further considerations, it should be noted that if the matrix A is reciprocal and consistent, then $\lambda_{\max} = n$ and rank(A) = 1, which means that all of the rows (columns) of A differ only by a multiplicative constant.

Pairwise comparisons (preferences) are called transitive, if the following property is satisfied:

$$\text{If } a_{ij} > 1 \text{ and } a_{jk} > 1 \text{ Then } a_{ik} > 1, \forall i, j, k \tag{5a}$$

$$\text{If } a_{ij} < 1 \text{ and } a_{jk} < 1 \text{ Then } a_{ik} < 1, \forall i, j, k \tag{5b}$$

Relation (5a) has the following meaning: if an object A is preferred to an object B, and an object B is preferred to C, then an object A is preferred to C as well. The meaning of relation (5b) is similar, but the preferences are reversed. Since we assume the reciprocity condition (1) is preserved, it suffices to examine entries above the main diagonal of a pairwise comparisons matrix.

In the case of n objects, the number of triads (triples of a_{ij} from relation (5a) or (5b) such that $1 \leq i < j < k \leq n$ is given as $\binom{n}{3}$.

3 Data and the Method

We randomly generated 10,000 pairwise comparison matrices of the order n = 4 in MS Excel with entries drawn from Saaty's fundamental scale $S_F = \{1/9, 1/8, \ldots, 1, \ldots 8, 9\}$, and for each matrix two properties were observed:

- transitivity occurrence (how often transitivity occurs),
- transitivity violation (how often is transitivity violated).

For human expert matrices we used matrices provided by 85 respondents (university undergraduate students) who compared an area of four geometric figures in an experiment published in [9], and from the study [8]. Again, Saaty's fundamental scale was applied for comparisons. Respondents were not informed about the actual objective of the experiment.

Since n = 4, there were four triads that had to be examined with respect to relations (5a, b). Therefore, for each pairwise comparisons matrix, the values of transitivity occurrence and transitivity violation attained values from the set $V = \{0,1,2,3,4\}$.

After data collection, both samples were statistically processed and analyzed.

Two null research hypotheses were formulated:

- H_{01}: Frequency of occurrence of transitivity for random matrices and human-made matrices is the same.
- H_{02}: Frequency of violation of transitivity for random matrices and human-made matrices is the same.

Both hypotheses were tested by the two-sample t-test with unequal variances.

4 Results

Mean values and standard deviations of transitivity occurrence and transitivity violation for random and human-made pairwise comparisons matrices are provided in Table 1 and Figure 1 a–d. It is clear that transitivity occurrence is more frequent for human-made matrices, while transitivity violation is more frequent for random matrices.

The biggest difference between the two categories of PC matrices were found in preservation/violation of transitivity. While for human-made matrices only 18.8% violated (once of more times) transitivity, for random matrices this value was 60.4%.

The hypothesis H_{01} was tested by the two-sample t-test with unequal variances in MS Excel. Since reported$p < 10^{-10}$, the null hypothesis H_{01} was rejected.

The hypothesis H_{02} was tested by the two-sample t-test with unequal variances in MS Excel. Because reported$p < 10^{-11}$, the null hypothesis H_{02} was rejected.

Sample ($n = 4$)	Transitivity occurrence Mean (std. dev.)	Transitivity violation Mean (std. dev.)
Random matrices	0.830 (0.825)	0.929 (0.852)
Human-made matrices	1.494 (0.815)	0.356 (0.727)

Table 1 Mean values for human-made and random PC matrices.

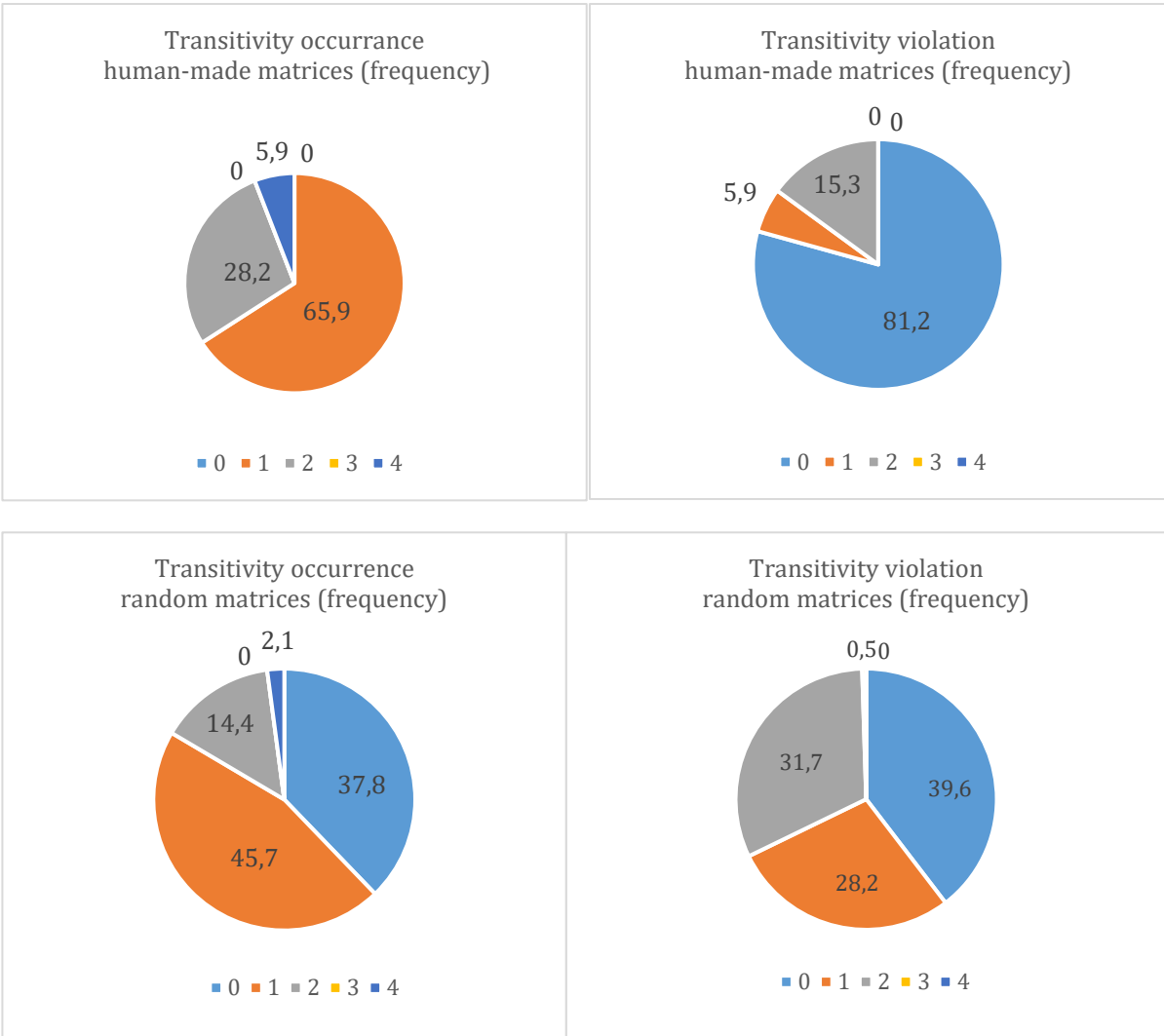


Figure 1a–d Transitivity occurrence and violation for random and human-made matrices (in %)

5 Application

It is estimated that 37.9% of the Internet traffic is made up by bots, and 20.4% by bad bots [10]. Bad bots are responsible for scraping and stealing data, publishing fake content or reviews, or skewing advertising and visitor metrics. The most targeted area by bad bots is financing followed by ticketing, education and IT [10].

Therefore, differentiating between human-made preferences and preferences collected randomly or created by bots might be of great importance. Next example demonstrates how the transitivity can be used to achieve this goal.

Example 1. Consider the set of pairwise preferences about four products collected from one Internet source Z in the form of 30 PC matrices. Out of these 30 PCMs, 18 violated transitivity at least once, while the rest preserved transitivity. The question arises: Were these preferences collected from real human consumers, or were they generated (randomly) by Internet bots?

To solve this problem, we state the following null hypothesis: "The preferences expressed by the Internet source Z were formed by humans."

Secondly, we use the data collected on human-made PC matrices of the order n = 4 from 85 respondents described in the previous sections and form a 2 times 2 contingency table:

PCMs Source/ Transitivity preserved	Yes	No
Human-made	69	16
Unknown	12	18

We evaluate the null hypothesis by chi-square test, the critical value t_{crit} ($alpha$ = 0.05, df = 1) = 3.8. The test value G = 18.05. Since $G > t_{crit.}$, the null hypothesis is rejected. The preferences provided by the Internet source Z were not (with 95% probability) created by humans (consumers).

6 Conclusions

In this paper we examined and compared transitivity occurrence and transitivity violation of human-made and random pairwise comparisons matrices of the order $n = 4$. Potential application of our study is as follows: consider a situation where a decision maker is given preferences in a form of pairwise comparisons, but the source of the comparisons is unknown. It could be a human expert, but also these preferences could be randomly collected from an Internet, or randomly generated by an Internet bot.

Since we showed that human-made and random pairwise comparisons (preferences) statistically significantly differ, 95% (99%) confidence intervals for both cases can be derived and based on transitivity of a given PC matrix, this matrix can be classified (at least theoretically) into one of the two categories.

In the future research we are going to examine transitivity for cases with $n > 4$, and also we plan to add another category to random and human-made PCs: pairwise comparisons provided (intentionally) by an artificial intelligence.

7 Acknowledgments

The paper was partially supported by the Czech Grant Agency grant number 18-01246S.

References

[1] CRAWFORD, G. B. (1987). The geometric mean procedure for estimating the scale of a judgement matrix, *Mathematical Modelling*, 9(3), 327–334.

[2] DEATON, A. & MUELLBAUER, J. (1999). *Economics and consumer behavior*. Cambridge: Cambridge University Press.

[3] DIAZ, S., GARCIA-LAPRESTA, J. L., & MONTES, S. (2008). Consistent models of transitivity for reciprocal preferences on a finite ordinal scale. *Information Sciences*, ***178***, 2832–2848.

[4] GISIN, V. B. (1994). On Transitivity of Strict Preference Relations. *Fuzzy Sets and Systems,* ***67***, 293–301.

[5] GONG, Z., LI, L., CAO, J., & ZHOU, F. (2010). On additive consistent properties of the intuitionistic fuzzy preference relation. *International Journal of Information Technology & Decision Making,* ***9***, 1009–1025.

[6] HADDENHORST, B., HUELLERMEIER, E., & KOLB, M. (2020). Generalized transitivity: A systematic comparison of concepts with an application to preferences in the Babington Smith model. *International Journal of Approximate Reasoning*, ***119***, 373–407.

[7] ISHIZAKA, A., LUSTI, M. (2006). How to derive priorities in the AHP: A comparative study. *Central European Journal of Operations Research*, 14(4), 387–400.

[8] MAZUREK, J., PERZINA, R. (2017). On the inconsistency of Pairwise Comparisons: An Experimental Study. *Scientific Papers of the University of Pardubice*, 24(3), 102–109.

[9] MAZUREK, J., RAMÍK, J. (2020). Which pairwise comparisons method is the best? *DEMSME*, Karvina, Czechia.

[10] OSBORNE, C. (2019). Bad bots now make up 20 percent of web traffic. ZDNet. Available from: https://www.zdnet.com/article/bad-bots-focus-on-financial-targets-make-up-20-percent-of-web-traffic/.
[11] PEKALA, B., SZMIDT, E., & KACPRZYK, J. (2018). Group Decision Support under Intuitionistic Fuzzy Relations: The Role of Weak Transitivity and Consistency. *International Journal of Intelligent Systems*, ***33***, 2078– 095.
[12] REGENWETTER, M., DANA, J., & DAVIS-STOBER, C. P. (2011). Transitivity of preferences. *Psychological Review,* ***118***, 42–56.
[13] SAATY, T. L. (1977) A Scaling Method for Priorities in Hierarchical Structures. *Journal of Mathematical Psychology*, 15, 234–281.
[14] SAATY, T. L. (1980). *Analytic Hierarchy Process*. McGraw-Hill, New York.
[15] SWITALSKI, Z. (2001). Transitivity of fuzzy preference relations – an empirical study. *Fuzzy Sets and Systems,* ***118***, 503–508.
[16] SWITALSKI, Z. (2003). General transitivity conditions for fuzzy reciprocal preference matrices. *Fuzzy Sets and Systems,* ***137***, 85–100.
[17] WANG, X. (1997). An investigation into relations between some transitivity-related concepts. *Fuzzy Sets and Systems,* ***89***, 257–262.

Periodicity of Convex and Concave Monge Matrices in Max-min Algebra

Monika Molnárová[1]

Abstract. Periodic properties of both convex and concave Monge matrices over max-min algebra are studied. Structure of the corresponding threshold digraphs in regard to matrix period is described. The period of a convex Monge matrix is proved to be equal to one or two. Moreover, equivalent conditions for both cases are presented. The period of a concave Monge matrix is shown to be equal to one.

Keywords: (max, min) algebra, period, Monge matrix

JEL Classification: C02
AMS Classification: 08A72, 90B35, 90C47

1 Introduction

In many optimization problems regarding discrete dynamic systems (DDS), graph theory, knowledge engineering, cluster analysis or fuzzy logic programs maximum and minimum operations are involved. Both of these operations have the property that they create no new elements. An extremal algebra is defined as an algebra in which at least one of the operations possesses the above-mentioned property. The max-min algebra (called also fuzzy algebra) with operations maximum and minimum and max-plus algebra (called also max algebra) with operations maximum and addition are most frequented. How the concept of max-plus algebra can be used in modelling discrete event systems was presented in [9]. Fuzzy sets and their applications were studied in [10].

The algorithms for matrix computations can be more efficient in case of special classes of matrices, e.g., Toeplitz, Circulant or Monge matrices. We have studied Monge matrices, their structural properties and algorithms solving many problems related to Monge matrices in [1], [2]. Periodic properties of matrices in max-plus algebra were presented in [5]. Periodicity in max-min algebra was studied in [3]. The formula for computing matrix period in max-min algebra was proved in [3], namely the period of a fuzzy matrix is equal to the least common multiple of the periods of all non-trivial strongly connected components from all threshold digraphs of the matrix.

The aim of this paper is to describe the structure of threshold digraphs of convex and concave Monge fuzzy matrices with regard to matrix periodicity and to prove equivalent conditions for the periodicity of convex and concave Monge fuzzy matrices.

We briefly outline the content and main results of the paper. Section 2 provides the necessary preliminaries on max-min algebra and periodicity in max-min algebra. In Section 3, the notion of a convex Monge matrix and a concave Monge is introduced. Important properties in regard of digraph structure of Monge matrices are presented and the differences between convex and concave Monge matrices are pointed out. In Section 4, results concerning periodicity are presented. The key results of this section are Theorem 2, which determines the period of a convex Monge matrix, and the Theorem 4, which formulates a necessary and sufficient condition for the period to be equal to two.

2 Background of the problem

The fuzzy algebra $\mathcal{B}$ is a triple $(B, \oplus, \otimes)$, where $(B, \leq)$ is a bounded linearly ordered set with binary operations *maximum* and *minimum*, denoted by $\oplus, \otimes$. The least element in B will be denoted by O, the greatest one by I. By $\mathbb{N}$ we denote the set of all natural numbers. The greatest common divisor of a set $S \subseteq \mathbb{N}$ is denoted by $\gcd S$, the least common multiple of the set S is denoted by $\operatorname{lcm} S$. For a given natural $n \in \mathbb{N}$, we use the notation N for the set of all smaller or equal positive natural numbers, i.e., $N = \{1, 2, \ldots, n\}$.

For any $m, n \in \mathbb{N}$, $B(m, n)$ denotes the set of all matrices of type $m \times n$ and $B(n)$ the set of all n-dimensional column vectors over $\mathcal{B}$. The matrix operations over $\mathcal{B}$ are defined formally in the same manner (with respect

[1] Technical University of Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, Monika.Molnarova@tuke.sk

to $\oplus$, $\otimes$) as matrix operations over any field. The rth power of a matrix $A \in B(n,n)$ is denoted by A^r, with elements a_{ij}^r. For $A,\ C \in B(n,n)$ we write $A \le C$ if $a_{ij} \le c_{ij}$ holds for all $i,\ j \in N$.

A *digraph* is a pair $G = (V,E)$, where V, the so–called vertex set, is a finite set, and E, the so–called edge set, is a subset of $V \times V$. A digraph $G' = (V',E')$ is a subdigraph of the digraph G (for brevity $G' \subseteq G$), if $V' \subseteq V$ and $E' \subseteq E$. A path in the digraph $G = (V,E)$ is a sequence of vertices $p = (i_1,\ \dots,\ i_{k+1})$ such that $(i_j, i_{j+1}) \in E$ for $j = 1,\ \dots,\ k$. The number k is the length of the path p and is denoted by $\ell(p)$. If $i_1 = i_{k+1}$, then p is called a cycle. For a given matrix $A \in B(n,n)$ the symbol $G(A) = (N,E)$ stands for the complete, edge–weighted digraph associated with A, i.e., the vertex set of $G(A)$ is N, and the capacity of any edge $(i,j) \in E$ is a_{ij}. In addition, for given $h \in B$, the *threshold digraph* $G(A,h)$ is the digraph $G = (N,E')$ with the vertex set N and the edge set $E' = \{(i,j);\ i,j \in N,\ a_{ij} \ge h\}$.

By a *strongly connected component* of a digraph $G(A,h) = (N,E)$ we mean a subdigraph $\mathcal{K} = (N_{\mathcal{K}}, E_{\mathcal{K}})$ generated by a non-empty subset $N_{\mathcal{K}} \subseteq N$ such that any two distinct vertices $i,j \in N_{\mathcal{K}}$ are contained in a common cycle, $E_{\mathcal{K}} = E \cap (N_{\mathcal{K}} \times N_{\mathcal{K}})$ and $N_{\mathcal{K}}$ is the maximal subset with this property. A strongly connected component $\mathcal{K}$ of a digraph is called non-trivial, if there is a cycle of positive length in $\mathcal{K}$. For any non-trivial strongly connected component $\mathcal{K}$ is the *period* of $\mathcal{K}$ defined as $\operatorname{per}\mathcal{K} = \gcd\{\,\ell(c);\ c \text{ is a cycle in } \mathcal{K},\ \ell(c) > 0\,\}$. If $\mathcal{K}$ is trivial, then $\operatorname{per}\mathcal{K} = 1$. By $\mathrm{SCC}^\star(G)$ we denote the set of all non-trivial strongly connected components of G.

Let $\lambda \in B$. A matrix $A \in B(n,n)$ is *ultimately λ–periodic* if there are natural numbers p and R such that the following holds: $A^{k+p} = \lambda \otimes A^k$ for all $k \ge R$. The smallest natural number p with above property is called the period of A, denoted by $\operatorname{per}(A,\lambda)$. In case $\lambda = I$ we denote $\operatorname{per}(A,I)$ by abbreviation $\operatorname{per} A$.

According to [3] we define

$$\mathrm{SCC}^\star(A) \quad = \quad \cup\{\,\mathrm{SCC}^\star(G(A,h));\ h \in \{a_{ij}; i,j \in N\}\,\}$$

Theorem 1. [3] *Let $A \in B(n,n)$. Then*

$$\operatorname{per} A = \operatorname{lcm}\{\operatorname{per}\mathcal{K};\ \mathcal{K} \in \mathrm{SCC}^\star(A)\}.$$

3 Structure of threshold digraphs of convex and concave Monge matrices

In this section we introduce convex and concave Monge matrices. We describe some of the properties of the corresponding threshold digraphs in regard to matrix period. Moreover, we point out the differences between the convex and concave Monge matrices with respect to the structure of the threshold digraphs.

Definition 1. We say, that a matrix $A = (a_{ij}) \in B(m,n)$ is a convex Monge matrix (concave Monge matrix) if and only if

$$a_{ij} \otimes a_{kl} \le a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l$$
$$\big(\, a_{ij} \otimes a_{kl} \ge a_{il} \otimes a_{kj} \qquad \text{for all} \quad i < k,\ j < l\,\big).$$

Let us denote by $h^{(1)},\ h^{(2)},\ \dots,\ h^{(r)}$ the elements of the set $H = \{a_{ij};\ i,j \in N\}$ ordered into a strictly decreasing sequence, i.e.,

$$h^{(1)} > h^{(2)} > \dots > h^{(r)}. \tag{1}$$

The number r is equal to the number of different values in the matrix A.

Lemma 1. [6] Let $A \in B(n,n)$. Then the sequence of threshold digraphs corresponding to the sequence (1) is ordered by inclusion

$$G(A,h^{(1)}) \subseteq G(A,h^{(2)}) \subseteq \ldots \subseteq G(A,h^{(r)}).$$

Lemma 2. [4] Let $c = (i_1,\ i_2,\ \dots,\ i_k,\ i_1)$ be a cycle of length $k \ge 3$. Then there are arcs $(i_j,\ i_{j+1})$ and $(i_l,\ i_{l+1})$ in c such that

$$i_j < i_l \quad \text{and} \quad i_{j+1} < i_{l+1}.$$

The period of a fuzzy matrix depends on the periods of all non-trivial strongly connected components of all threshold digraphs (Theorem 1). The period of a non-trivial strongly connected component depends on the lengths of all cycles in the component. Hence the structure of the cycles in a component is of crucial importance investigating the periodicity of a fuzzy matrix.

The convex Monge property of a fuzzy matrix guaranties the decomposition of every cycle of length at least three in a non-trivial strongly connected component to cycles of length one and two. In addition, all loops in a threshold digraph lie in the same non-trivial strongly connected component.

Lemma 3. Let $A \in B(n,n)$ be a convex Monge matrix. Let $h \in H$. Let $\mathcal{K} \in \mathrm{SCC}^{\star}(G(A,h))$. Let c be a cycle of length $\ell(c) \geq 3$ in $\mathcal{K}$. Then c can split in $\mathcal{K}$ into finite number of cycles of length one or two.

Proof. Let $h \in H$. Let $\mathcal{K} \in \mathrm{SCC}^{\star}(G(A,h))$. Let $c = (i_1, i_2, \ldots, i_k, i_1)$ be a cycle of length $\ell(c) \geq 3$ in $\mathcal{K}$. By Lemma 2 there are arcs (i_j, i_{j+1}) and (i_l, i_{l+1}) in c such that $i_j < i_l$ and $i_{j+1} < i_{l+1}$. By the Monge property

$$h \leq a_{i_j i_{j+1}} \otimes a_{i_l i_{l+1}} \leq a_{i_l i_{j+1}} \otimes a_{i_j i_{l+1}}.$$

Thus $h \leq a_{i_l i_{j+1}}$ and $h \leq a_{i_j i_{l+1}}$. Consequently the component $\mathcal{K}$ contains arcs (i_l, i_{j+1}) and (i_j, i_{l+1}) as well. Hence the cycle c splits into cycles $c_1 = (i_1, i_2, \ldots, i_j, i_{l+1}, \ldots, i_1)$ and $c_2 = (i_{j+1}, \ldots, i_l, i_{j+1})$ with $\ell(c) = \ell(c_1) + \ell(c_2)$. If $\ell(c_1) < 3$ and $\ell(c_2) < 3$ the assertion follows. If $\ell(c_1) \geq 3$ we repeat the procedure and split the cycle c_1 into two cycles again. The procedure of splitting the cycle will be repeated with both new arose cycles until we get cycles of length one and two. The case when $\ell(c_2) \geq 3$ we handle by analogy with above case. □

Lemma 4. Let $A \in B(n,n)$ be a convex Monge matrix. Let $h \in H$. Let $\mathcal{K} \in \mathrm{SCC}^{\star}(G(A,h))$. Let c be a cycle of odd length $\ell(c) \geq 3$ in $\mathcal{K}$. Then there is a node in c with a loop.

Proof. The assertion follows by Lemma 3 □

Lemma 5. [6] Let $A \in B(n,n)$ be a convex Monge matrix. Let $h \in H$. Let for $i, k \in N$ be the loops (i,i) and (k,k) in the digraph $G(A,h)$. Then the nodes i and k are in the same non-trivial strongly connected component $\mathcal{K}$ of $G(A,h)$.

Example 1. Let us check the structure of the threshold digraphs of the given convex Monge matrix $A \in B(6,6)$ for $B = [0,2]$

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 2 & 2 & 1 & 0 \\ 0 & 2 & 2 & 2 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

According to Lemma 3 every cycle in all non-trivial strongly connected components is a concatenation of cycles of length one and two (see Figure 1). The threshold digraph $G(A,1)$ consists of one non-trivial strongly connected component with two loops. The threshold digraph $G(A,2)$ contains two trivial and two non-trivial strongly connected components. The non-trivial strongly connected component $\mathcal{K}_1$ generated by the node set $N_{\mathcal{K}_1} = \{2,5\}$ has no loop, whereas the non-trivial strongly connected component $\mathcal{K}_2$ generated by the node set $N_{\mathcal{K}_2} = \{3,4\}$ contains according to Lemma 5 all loops from $G(A,2)$, i.e. two loops.

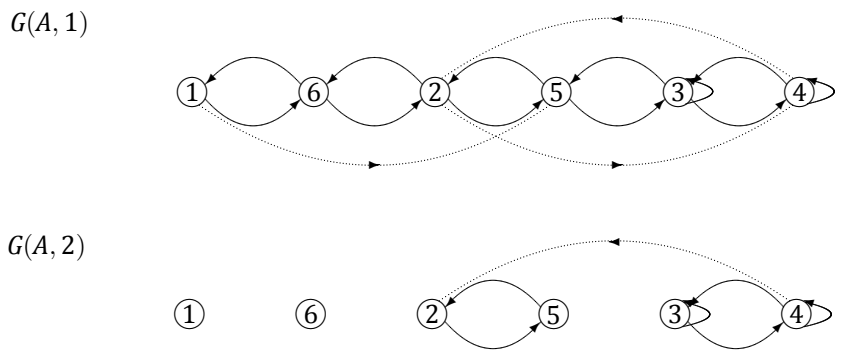


Figure 1 Structure of components in threshold digraphs of a convex Monge matrix

An analogy with Lemma 3 is for concave Monge matrices the following Lemma. However, every cycle can split now into cycles of length one, i.e. loops on every node of the cycle. Moreover, in contrast to the convex Monge matrices two loops can lie in different strongly connected components.

Lemma 6. [7] Let $A \in B(n,n)$ be a concave Monge matrix. Let c be a cycle in $G(A,h)$ for $h \in H$. Then there is a loop on every node of the cycle c.

Lemma 7. [7] Let $A \in B(n,n)$ be a concave Monge matrix. Let $h \in H$. Let $\mathcal{K} \in \mathrm{SCC}^{\star}(G(A,h))$. Then there is a loop on every node of the component $\mathcal{K}$.

Lemma 8. [7] Let $A \in B(n,n)$ be a concave Monge matrix. Element $a_{kk} < h$ represents a trivial strongly connected component of $G(A,h)$.

In contrast to convex Monge matrices there is a kind of ordering among the nodes from different strongly connected components for concave Monge matrices, namely the nodes of a component are numbered by a sequence of consecutive natural numbers.

Lemma 9. [7] Let $A \in B(n,n)$ be a concave Monge matrix. Let $c_1 = (i_0, i_1, \ldots, i_k)$ with $i_0 = i_k$ and $c_2 = (j_0, j_1, \ldots, j_l)$ with $j_0 = j_l$ be cycles in different non-trivial strongly connected components in $G(A,h)$ for $h \in H$. Let $i_s = \min\{i_0, i_1, \ldots, i_{k-1}\}$ and $i_t = \max\{i_0, i_1, \ldots, i_{k-1}\}$. Then exactly one of the conditions holds
(i) $j_m < i_s$ for all $m \in \{0, 1, \ldots, l\}$,
(ii) $j_m > i_t$ for all $m \in \{0, 1, \ldots, l\}$.

Consequently a concave Monge matrix has for every $h \in H$ a block form, where the diagonal blocks represent the strongly connected components.

Example 2. Let us check the structure of the threshold digraphs $G(A,h)$ of the given concave Monge matrix $A \in B(8,8)$ for $B = [0,2]$

$$A = \left(\begin{array}{c|ccc||cccc}
2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
1 & 2 & 2 & 2 & 0 & 0 & 0 & 0 \\
1 & 2 & 2 & 2 & 0 & 0 & 0 & 0 \\ \hline
0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ \hline\hline
0 & 0 & 0 & 0 & 2 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & 2 & 2 & 1 \\
0 & 0 & 0 & 0 & 0 & 2 & 2 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 2 & 2
\end{array}\right).$$

According to Lemma 6 every cycle in all non-trivial strongly connected components is a concatenation of cycles of length one (see Figure 2). The threshold digraph $G(A,1)$ contains two non-trivial strongly connected components defined by diagonal blocks bounded by double lines, whereas the threshold digraph $G(A,2)$ contains three non-trivial strongly connected components and one trivial strongly connected component (defined by $a_{44} < h = 2$) represented by diagonal blocks bounded by single lines. However, according to Lemma 7 there is a loop on every node in all non-trivial strongly connected components.

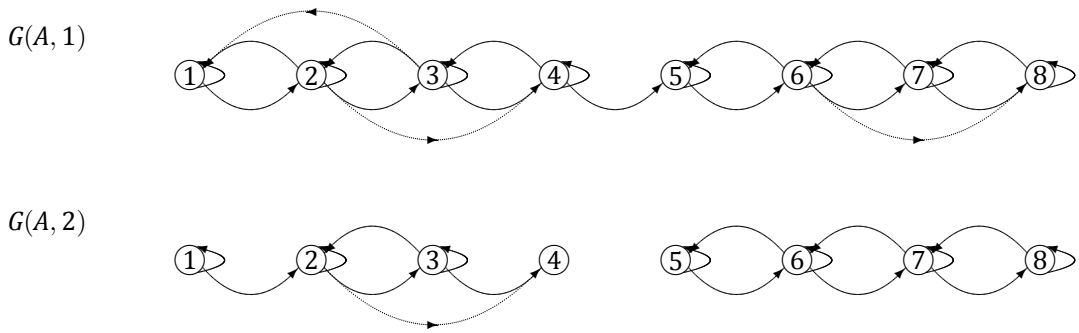


Figure 2 Structure of components in threshold digraphs of a concave Monge matrix

4 Periodicity of convex and concave Monge matrices

Results considering the structure of threshold digraphs introduced in previous section are helpful to study the period of Monge matrices. In this section we show that the period of a Monge matrix can not exceed the value two. Anyway, there are again differences between convex Monge matrices on one side and concave Monge matrices on other side.

Theorem 2. *Let $A \in B(n,n)$ be a convex Monge matrix. Then one of the statements holds*
(i) $\operatorname{per} A = 1$,
(ii) $\operatorname{per} A = 2$.

Proof. By Lemma 3 a decomposition of every cycle of length at least three in $G(A,h)$ of a convex Monge matrix results in cycles of length one and two. Hence the period $\operatorname{per} A$ equals to one or two using Theorem 1 and the definition of the period of a strongly connected component. □

Since robustness of a matrix means in fact that the period of the matrix equals one, we can formulate the equivalent condition for a convex Monge matrix to have the period equal to one.

Theorem 3. [6] *Let $A \in B(n,n)$ be a convex Monge matrix. Then* $\operatorname{per} A = 1$ *if and only if for each $h \in H$ the digraph $G(A,h)$ contains at most one non-trivial strongly connected component and this has a loop.*

Theorem 4. *Let $A \in B(n,n)$ be a convex Monge matrix. Then* $\operatorname{per} A = 2$ *if and only if there is $h \in H$ such that the digraph $G(A,h)$ contains a non-trivial strongly connected component without a loop.*

Proof. We prove the sufficient condition. Let $\operatorname{per} A = 2$. By Theorem 1 there exists $h \in H$ such that there is a non-trivial strongly connected component $\mathcal{K} \in \mathrm{SCC}^\star(G(A,h))$ with $\operatorname{per} \mathcal{K} = 2$. Since $\operatorname{per} \mathcal{K} = \gcd\{\, \ell(c);\ c \text{ is a cycle in } \mathcal{K},\ \ell(c) > 0 \,\}$ there is no loop in $\mathcal{K}$.
For the converse implication let us assume that there is $h \in H$ such that the digraph $G(A,h)$ contains a non-trivial strongly connected component $\mathcal{K}$ without a loop. Using Lemma 3 every cycle in $\mathcal{K}$ of length at least three can split into cycles of lengths one and two. Since $\mathcal{K}$ contains no loop by Lemma 4 there is no cycle of odd length in $\mathcal{K}$. Hence the length of every cycle is even and the decomposition of the cycle contains cycles of length two exclusively. Consequently $\operatorname{per} \mathcal{K} = 2$. Existence of a non-trivial strongly connected component with a loop does not influence the period of A, since the period of a component with a loop equals one. Using Theorem 1 is $\operatorname{per} A = 2$. □

Example 3. Let us compute the period of the convex Monge matrix from Example 1.
Since $G(A,2)$ contains a non-trivial strongly connected component $\mathcal{K}$ generated by the node set $N_{\mathcal{K}_1} = \{2,5\}$ (see Figure 1) due to Theorem 4 is $\operatorname{per} A = 2$.

The matrix in the next example is a slight modification of the above matrix (modified values are highlighted by bold characters) but the period equals one.

Example 4. Let us compute the period of the convex Monge matrix $A \in B(6,6)$ for $B = [0,2]$

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 2 & 2 & 2 & 0 \\ 0 & 2 & 2 & 2 & 0 & 0 \\ 0 & 2 & 2 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

In contrast to the previous example the digraph $G(A,h)$ contains at most one non-trivial strongly connected component and this with a loop for each $h \in H = \{0,1,2\}$ (see Figure 3). Due to Theorem 3 the period of the considered matrix $\operatorname{per} A = 1$.

Due to Lemma 7 the period of every non-trivial strongly connected component of a threshold digraph of a concave Monge matrix is equal to one. Consequently is the period of a concave Monge matrix one.

Theorem 5. [7] *Let $A \in B(n,n)$ be a concave Monge matrix. Then* $\operatorname{per} A = 1$.

We can illustrate the period of a concave Monge matrix on Example 2. There is no non-trivial strongly connected component without loops on every node. Hence the period of every component equals one and consequently the period of the matrix is equal to one as well.

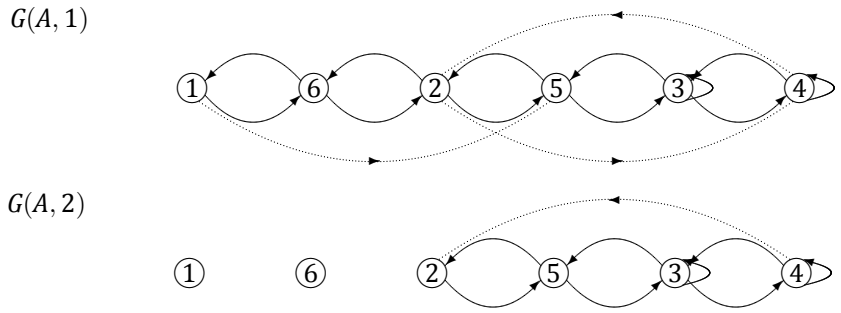


Figure 3 Threshold digraphs of a convex Monge matrix with period one

References

[1] Burkard, R. E. Monge properties, discrete convexity and applications, *EJOR*, Volume **176** (2007), 1–14.

[2] Burkard, R. E., Klinz, B. and Rudolf, R. Perspectives of Monge properties in optimization, *DAM*, Volume **70** (1996), 95–161.

[3] Gavalec, M. Computing matrix period in max-min algebra. *DAM*, Volume **75** (1997), 63–70.

[4] Gavalec, M. and Plavka, J. An $O(n^2)$ algorithm for maximum cycle mean of Monge matrices in max algebra, *DAM*, Volume **127** (2003), 651–656.

[5] Molnárová, M. and Pribiš, J. Matrix period in max-algebra, *DAM*, Volume **103** (2000), 167–175.

[6] Molnárová, M. Robustness of Monge matrices in fuzzy algebra, In: *Proceedings of* 32nd *Int. Conference Mathematical Methods in Economics 2014*, Olomouc, 2014, 679–684.

[7] Molnárová, M. Convex and concave Monge matrices in fuzzy algebra – comparison with respect to robustness, (preprint).

[8] Molnárová, M., Myšková, H. and Plavka, J. The robustness of interval fuzzy matrices, *DAM*, Volume **438** (2013), 3350–3364.

[9] De Schutter, B., van den Boom, T., Xu, J. and Farahani, S. S. Analysis and control of max-plus linear discrete-event systems: An introduction. *Discrete Event Dyn. Syst.* (2020), Volume **30**, 25–54.

[10] Zimmermann, H. J. Fuzzy Set Theory And Its Applications; Springer Science and Business Media, Berlin, Germany, 2011.

Minimum Norm Solution of the Markowitz Mean-variance Portfolio Optimization Model

Hossein Moosaei[1], Milan Hladík[2]

Abstract. In finance, Markowitz' model, which is a portfolio optimization model, assists in the selection of the most efficient portfolio by analyzing various possible portfolios of the given securities. This model was considered in many different aspects by researchers. In this paper, we study an extended version of the classical Markowitz' mean-variance portfolio optimization model when this problem has multiple solutions. In this case the natural and in some sense the best choice is finding the solution with minimum norm. We focus on this problem and find the minimum-norm solution of the extended Markowitz's model. To achieve this goal, we characterize the solution set of the model, and by using a standard method and an augmented Lagrangian method we obtain the minimum norm solution of the mentioned problem. The numerical results show that the proposed method is efficient and works well even for large scale problems.

Keywords: Markowitz' model, solution set of convex problems, Newton method, augmented Lagrangian method

JEL Classification: C61, G00
AMS Classification: 90C20, 90C30

1 Introduction

In this paper, we consider a practical problem arising in finance. The standard portfolio optimization problem model known as the Markowitz' Mean-Variance portfolio optimization model can be formulated in an extended version as follows:

$$\begin{aligned} \min_x \quad & x^T Q x, \\ s.t. \quad & e^T x = 1, \\ & \mu^T x \geq r, \\ & Ax \leq b, \\ & x \geq 0, \end{aligned} \tag{1}$$

where e is a vector of ones and a vector of expected returns μ and the covariance matrix of the returns of the asset Q are known.

When the covariance matrix is positive semidefinite and rank deficient, the problem (1) has multiple optimal solutions, and in this case the natural choice is finding the minimum norm solution. This situation, i.e., the covariance matrix is rank deficient, can be acquired in many times because the covariance matrix is estimated from the past trading price data, and when the number of sampled periods is smaller than the number of assets, the covariance matrix is rank deficient.

In this paper, we consider the problem (1) that has multiple optimal solutions, and we suggest a novel method for finding the minimum norm solution based on an augmented Lagrangian method. We will compare our method with a standard method. The numerical results clearly show that the proposed method is efficient enough and also provides better results than the other methods.

As for our notations, a few points need to be made. We denote n-dimensional real space by R^n, and symbols A^T and $\|.\|$ stand for the transpose of matrix A and the Euclidean norm, respectively. Next, a_+ replaces negative components of the vector a by zeros.

The paper is organized as follows. A characterization of solution sets and the minimum-norm solution are described in Section 2. In Section 3, numerical experiments are reported to illustrate the efficiency of the proposed method, and finally Section 5 concludes the paper.

[1] University of Bojnord, Department of Mathematics, Bojnord, Iran.
Charles University, Department of Applied Mathematics, Prague, Czech Republic, hmoosaei@kam.mff.cuni.cz, moosaei@ub.ac.ir
[2] Charles University, Department of Applied Mathematics, Prague, Czech Republic, hladik@kam.mff.cuni.cz

2 Minimum-norm solution

As we mentioned in the introduction, the problem (1) might have multiple optimal solutions, and the related problem for finding its minimum norm solution can be written as follows

$$\min \ \frac{1}{2}\|x\|^2 \ \ s.t. \ \ x \in X^*, \tag{2}$$

where X^* is the optimal solution set of problem (1).

In this section, we first characterize the optimal solution set of the problem (1). Then, we introduce a novel and efficient method for finding its minimum norm solution.

At first, we briefly characterize the solution set of a convex problem. Let $S \subseteq \mathbb{R}^n$ be an open convex subset, $f\colon S \to \mathbb{R}$ a convex differentiable function, and $X \subseteq S$ any convex subset. The following theorem gives a precise characterization of solution sets of convex programs with twice continuously differentiable convex objective functions.

Theorem 1. *[4, 5] Consider the following convex problem:*

$$\min_{x \in X} f(x),$$

and suppose that its optimal solution set X^* is nonempty. Then

$$X^* = \{x \in X : \nabla f(x^*)^T x^* = \nabla f(x^*)^T x, \ \nabla f(x^*) = \nabla f(x)\},$$

where $x^* \in X^*$.

According to this theorem, we have the following theorem.

Theorem 2. *Let X^* be the optimal solution set for the problem (1) and assume $x^* \in X^*$. Then $x \in X^*$ if and only if x satisfies*

$$\begin{aligned} Qx &= Qx^*, \\ e^T x &= 1, \\ \mu^T x &\geq r, \\ Ax &\leq b, \\ x &\geq 0. \end{aligned} \tag{3}$$

Proof. By Theorem 1, we have $x \in X^*$ if and only if x is feasible and $\nabla f(x^*)^T x^* = \nabla f(x^*)^T x$, $\nabla f(x^*) = \nabla f(x)$. This yields $Qx = Qx^*$. Therfore $x \in X^*$ if and only if x satisfies (3). □

The regarding to this theorem, we should solve the following problem:

$$\begin{aligned} \min \ & \frac{1}{2}\|x\|^2, \\ s.t. \ & Qx = Qx^*, \\ & e^T x = 1, \\ & \mu^T x \geq r, \\ & Ax \leq b, \\ & x \geq 0. \end{aligned}$$

Or equivalently:

$$\begin{aligned} \min \ & \frac{1}{2}\|x\|^2, \\ s.t. \ & A_{eq} x = b_{eq}, \\ & A_{iq} x \leq b_{iq}, \\ & x \geq 0, \end{aligned} \tag{4}$$

where

$$A_{eq} = \begin{bmatrix} Q \\ e^T \end{bmatrix}, \ b_{eq} = \begin{bmatrix} Qx^* \\ 1 \end{bmatrix}, \ A_{iq} = \begin{bmatrix} -\mu^T \\ A \end{bmatrix}, \ b_{iq} = \begin{bmatrix} -r \\ b \end{bmatrix}.$$

Now, for solving the problem (4), we have two approaches. The first approach is solving this problem by a standard method, and the second approach is solving this problem by an augmented Lagrangian method, which is addressed in the remainder of the paper.

The augmented Lagrangian (AL) method is also known as the method of multipliers and solves the constrained optimization problem. The general idea behind the AL method is to incorporate some or all of the constraints into the objective function [2, 7]. In the AL method, we add a quadratic penalty to the Lagrangian function. For all $\rho, \lambda_2, \lambda_3 > 0$ and λ_1, the AL function of (4) can be defined as follows:

$$AL_\rho(x, \lambda_1, \lambda_2, \lambda_3) = \tfrac{1}{2}\|x\|^2 + \tfrac{\rho}{2}\left\|A_{eq}x - b_{eq} + \tfrac{\lambda_1}{\rho}\right\|^2 + \tfrac{\rho}{2}\left\|\left(A_{iq}x - b_{iq} + \tfrac{\lambda_2}{\rho}\right)_+\right\|^2 + \tfrac{\rho}{2}\left\|\left(-x + \tfrac{\lambda_3}{\rho}\right)_+\right\|^2. \quad (5)$$

We use the following formula to update the Lagrangian multipliers:

$$\lambda_1^{k+1} = (\lambda_1^k + \rho(A_{eq}x - b_{eq})), \quad \lambda_2^{k+1} = (\lambda_2^k + \rho(A_{iq}x - b_{iq}))_+, \quad \lambda_3^{k+1} = (\lambda_2^k + \rho(-x))_+. \quad (6)$$

The following algorithm, which is an augmented Lagrangian algorithm for solving (4), proceeds by minimizing the augmented Lagrangian function at each iteration and updating Lagrange multipliers and penalty parameters between iterations.

Algorithm 1 Augmented Lagrangian algorithm for finding the minimum norm solution of problem (1)

Input: $\rho^1, \lambda_2^1, \lambda_3^1 > 0$, λ_1^1, and $k = 1$.

1: Find x^k as an approximate solution of the following problem:

$$\min_x \ AL_{\rho^k}(x, \lambda_1^k, \lambda_2^k, \lambda_3^k).$$

2: Update the Lagrange multipliers by using:

$$\lambda_1^{k+1} = (\lambda_1^k + \rho^k(A_{eq}x^k - b_{eq}), \quad \lambda_2^{k+1} = (\lambda_2^k + \rho^k(A_{iq}x^k - b_{iq})_+, \quad \lambda_3^{k+1} = (\lambda_2^k + \rho^k(-x^k))_+.$$

3: Choose new penalty parameter $\rho^{k+1} \geq \rho^k$.

4: Set $k \leftarrow k + 1$ and go to step 1.

5: **return** An approximate solution x^* of the problem (4).

The essential part of this algorithm is solving the following problem:

$$\min_x \ AL_\rho(x, \lambda_1, \lambda_2, \lambda_3). \quad (7)$$

The objective function of the problem (7) is convex, piecewise quadratic, and differentiable, but not twice differentiable. Indeed, the gradient

$$\nabla_x AL_\rho(x, \lambda_1, \lambda_2, \lambda_3) = x + \rho A_{eq}^T\left(A_{eq}x - b_{eq} + \frac{\lambda_1}{\rho}\right) + \rho A_{iq}^T\left(A_{iq}x - b_{iq} + \frac{\lambda_2}{\rho}\right)_+ - \rho\left(-x + \frac{\lambda_3}{\rho}\right)_+$$

of $AL_\rho(x, \lambda_1, \lambda_2, \lambda_3)$ with respect to x is not differentiable because the subplus function (i.e., the positive part) is not differentiable. However, we can define the generalized symmetric positive semidefinite Hessian matrix [3, 8] as follows:

$$\partial^2 AL_\rho(x, \lambda_1, \lambda_2, \lambda_3) = I + \rho A_{eq}^T A_{eq} + \rho A_{iq}^T D(z_1) A_{iq} + \rho D(z_2).$$

Here, $D(z_1)$ denotes a diagonal matrix whose ith diagonal entry is equal to one if the ith entry of $(A_{iq}x - b_{iq} + \frac{\lambda_2}{\rho}) > 0$, and to zero if $(A_{iq}x - b_{iq} + \frac{\lambda_2}{\rho}) \leq 0$, and similarly $D(z_2)$ denotes a diagonal matrix whose ith diagonal entry of it is equal to one if the ith entry of $\left(-x + \frac{\lambda_3}{\rho}\right) > 0$, and to zero if $\left(-x + \frac{\lambda_3}{\rho}\right) \leq 0$.

Since the generalized Hessian matrix can be singular, the following modified Newton direction is used [6].

$$-(\partial^2 AL_\rho(x, \lambda_1, \lambda_2, \lambda_3) + \delta I)^{-1} \nabla AL_\rho(x, \lambda_1, \lambda_2, \lambda_3),$$

where δ is a small positive number (in our calculations, $\delta = 10^{-4}$), and I is the identity matrix of appropriate order. In this case, the modified Newton method has the form

$$x_{n+1} = x_n - (\partial^2 AL_\rho(x_n, \lambda_1, \lambda_2, \lambda_3) + \delta I)^{-1} \nabla AL_\rho(x_n, \lambda_1, \lambda_2, \lambda_3).$$

In addition, our stopping criterion for this method is as follows: $\|x_{n+1} - x_n\| \leq tol$ (in our calculations, $tol = 10^{-9}$).

3 Numerical Experiments

In this section, we present numerical results on various instances of the extended Markowitz' portfolio optimization model. To further analyse our suggested method, we will compare the suggested method with an interior-point algorithm for convex problems, which is a standard optimization method for solving these kinds of problems. Finally, we illustrate the application of the proposed method on a real problem. The computational results were performed in MATLAB R2019b on a *Core™2 Quad CPU Q9300 @ 2.50GHz* $\times$ *4* with *8 GB* RAM computer.

3.1 Random artificial problems

In this subsection, to show that the proposed method is efficient, some various random problems are generated. Recall that we characterized the solution set of problem (1) as a system (3), and for finding the minimum norm solution of the problem (1), the following convex quadratic programming problem must be solved,

$$\min_x \frac{1}{2}\|x\|^2 \ \ s.t. \ \ x \text{ satisfies } (3). \tag{8}$$

We solve this problem directly with an interior-point method from the optimization toolbox in Matlab, and we name this method as QM. The proposed method based on augmented Lagrangian method is named as ALM.

In order to compare the QM and ALM and also in order to assess the computational behavior of our methods, a random program was written to generate sets of problems and we find the minimum norm solution to each problem. The following random code generates problems for which the optimal value of the objective function of the problem (1) is 0 and the norm of the minimum norm optimal solution of this problem is 0.7071.

```
    % Generate extended Markowitz' portfolio optimization model with
        minimum norm 0.7071
Q=zeros(n,n);
for i=1:n-2
    Q(i,i)=10;
end
e= ones(n,1);
mu=10*ones(n,1);
r=1;
A=rand(m,n);b=10*ones(m,1);
x1=zeros(n,1);x1(n)=1; % x1 is one of the optimal solutions
```

Table 1 reports the following information for each test problem:

- Algorithm: the name of the algorithm
- m, n: the size of problem
- Obj: optimal value of the objective function of problem (1)
- $\|\widetilde{x}^*\|$: the norm of the minimum norm solution of x
- cpu_s: CPU time in seconds.

Algorithm	n	m	Obj	$\|\widetilde{x}^*\|$	cpu_s
ALM	10	500	9.9668e-19	0.7071	0.092
QM	10	500	0	0.7071	1.715
ALM	15	1000	8.5546e-19	0.7071	0.166
QM	15	1000	0	0.7071	4.286
ALM	20	2000	6.5695e-18	0.7071	0.222
QM	20	2000	0	0.7071	30.042
ALM	50	5000	2.7574e-16	0.7071	0.547
QM	50	5000	0	0.7071	231.840
ALM	10	50000	9.9668e-19	0.7071	1.250
QM	10	50000	Out of memory		
ALM	20	100000	6.2295e-18	0.7071	4.398
QM	20	100000	Out of memory		
ALM	15	500000	2.7384e-18	0.7071	20.461
QM	15	500000	Out of memory		
ALM	30	1000000	2.423e-17	0.7071	82.179
QM	30	100000	Out of memory		
ALM	50	2000000	2.0244e-16	0.7071	268.278
QM	50	2000000	Out of memory		

Table 1 Comparison of the numerical results of the *ALM*, and *QM*

Table 1 shows that the norm of the minimum norm solution of all test problems is close to 0.7071 and the objective function is zero. This indicates that we have successfully obtained the minimum norm solution for all test problems.

In order to more analyse these results, we note that ALM and QM work well for small and medium problems, and also we see that ALM is faster than the alternative method. The another advantage of the proposed method is finding minimum norm solutions of large-scale problems in an appropriate time while the second method is not able to find the minimum norm solution because the memory cannot support the requirements for the algorithms in large-scale problems (we report "Out of memory" in table 1).

3.2 Standard Markowitz' Mean-Variance portfolio optimization model

In this section, we apply our theory on a practical problem arising in finance. A reduced version of the problem (1) that is a standard Markowitz' Mean-Variance portfolio optimization model can be formulated as follows:

$$\begin{aligned} \min_{x} \quad & x^T \Sigma x, \\ s.t. \quad & e^T x = 1, \\ & \mu^T x \geq r, \\ & x \geq 0, \end{aligned} \tag{9}$$

where e is a vector of ones. Vector of expected returns μ and the covariance matrix of the returns of the asset Σ are known.

We consider a specific example of portfolio optimization model. The real data for problem (9), which are the data between the years 1974 and 1977, come from `https://vanderbei.princeton.edu/ampl/nlmodels/markowitz/` and can also be found in [1]. The data have 8 types of assets ($N = 8$): US 3 month treasury bills, US government long bonds, SP 500, Wilshire 500, NASDAQ composite, corporate bond index, EAFE and Gold during the years between 1974–1977.

R is the 8×4 matrix containing the assets' returns for each of the 4 years. The expected returns vector μ and sample covariance matrix were computed via the following known formulas:

$$\mu := \frac{1}{T}R\mathbf{1}, \qquad \Sigma := \frac{1}{T-1}R(I - \frac{1}{T}\mathbf{1}\mathbf{1}^{\boldsymbol{T}})R^T,$$

where $\mathbf{1}$ denotes a conformable vector of ones, I a conformable identity matrix and T number of periods (here $T = 4$). We suppose that $r = 1.05$.

Thus the estimate of Σ and μ can be obtained as follows:

$$\Sigma = \begin{bmatrix} 0.0002 & -0.0005 & -0.0028 & -0.0032 & -0.0039 & -0.0007 & -0.0024 & 0.0048 \\ -0.0005 & 0.0061 & 0.0132 & 0.0136 & 0.0126 & 0.0049 & -0.0003 & -0.0154 \\ -0.0028 & 0.0132 & 0.0837 & 0.0866 & 0.0810 & 0.0196 & 0.0544 & -0.1159 \\ -0.0032 & 0.0136 & 0.0866 & 0.0904 & 0.0868 & 0.0203 & 0.0587 & -0.1227 \\ -0.0039 & 0.0126 & 0.0810 & 0.0868 & 0.0904 & 0.0192 & 0.0620 & -0.1232 \\ -0.0007 & 0.0049 & 0.0196 & 0.0203 & 0.0192 & 0.0054 & 0.0090 & -0.0261 \\ -0.0024 & -0.0003 & 0.0544 & 0.0587 & 0.0620 & 0.0090 & 0.0619 & -0.0900 \\ 0.0048 & -0.0154 & -0.1159 & -0.1227 & -0.1232 & -0.0261 & -0.0900 & 0.1725 \end{bmatrix}, \ \mu = \begin{bmatrix} 1.0630 \\ 1.0633 \\ 1.0670 \\ 1.0853 \\ 1.0882 \\ 1.0778 \\ 1.0820 \\ 1.1605 \end{bmatrix}.$$

The rank of matrix Σ is at most equal to the rank of the matrix $I - \frac{1}{T}\mathbf{1}\mathbf{1}^{\boldsymbol{T}}$ which is $T - 1$. In our example, the rank of matrix Σ is 3, and so it is rank-deficient. Therefore the portfolio optimization problem (9) has multiple optimal solutions. We solve the problem (9) by using function "quadprog.m" in MATLAB and the norm of optimal solution is $\|x\| = 0.4089$ and the minimum norm solution through our proposed ALM method is $\|x\| = 0.3934$.

4 Conclusion

In this paper, we were concentrated on finding the minimum norm solution of the Markowitz' Mean-Variance portfolio optimization model when it has multiple solutions. At first we characterized the optimal solution set of the model and then introduced an augmented Lagrangian method for finding the minimum norm solution of this model. In our algorithm we faced with an once differentiable unconstrained optimization problem and for solving it a modification of Newton's method was proposed. We introduced an augmented Lagrangian method for finding the minimum norm solution of this model. We compared our method with a standard method, and we also applied our method on a real problem. The results show that the proposed method is efficient and can be used even for large scale problems.

Acknowledgements

The authors were supported by the Czech Science Foundation Grant P403-18-04735S.

References

[1] Beck, A. & Sabach, S. (2014). A first order method for finding minimal norm-like solutions of convex optimization problems, *Mathematical Programming*, 147, 25–46.

[2] Birgin, E. G., & Martínez, J. M. (2014). *Practical augmented Lagrangian methods for constrained optimization*. Philadelphia, PA: SIAM.

[3] Hiriart-Urruty, J., Strodiot, J. & Nguyen, V. H. (1984). Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data, *Applied Mathematics and Optimization*, 11, 43–56.

[4] Ketabchi, S. & Ansari-Piri, E. (2007). On the solution set of convex problems and its numerical application, *Journal of Computational and Applied Mathematics*, 206, 288–292.

[5] Mangasarian, O. (1988). A simple characterization of solution sets of convex programs, *Operations Research Letters*, 7, 21–26.

[6] Mangasarian, O. (2004). A Newton method for linear programming, *Journal of Optimization Theory and Applications*, 121, 1–18.

[7] Nocedal, J. & Wright, S. (2006). *Numerical Optimization*. New York: Springer.

[8] Pardalos, P. M., Ketabchi, S. & Moosaei, H. (2014). Minimum norm solution to the positive semidefinite linear complementarity problem. *Optimization*, 63, 359–369.

Comparison of Selected Portfolio Approaches with Benchmark

David Neděla[1]

Abstract. Over the years, a popular topic amongst scientific economists has been the exploration and validation of modeling used to achieve an optimal portfolio composition that meets the financial goals of the investor. The portfolio optimization model offered tools for finding weights to align an investor's risk attitude with potential portfolio return. In this paper, the comparison between the backtesting of several portfolio optimization models, and the benchmark is solved. The portfolio models used for back testing were the most widely known Markowitz model, the Conditional Value at Risk (CVaR) model, and the Bayesian optimization model. The comparison criteria used were; final wealth, mean return and mean standard deviation together with Sharpe ratio, Rachev ratio and Value-at-Risk portfolio performance measures. The stock market used were the US, the Chinese, and the UK markets, with stock data drawn at boom and crisis time periods to allow comprehensive comparison across world markets and time periods.

Keywords: portfolio optimization, Markowitz model, Bayesian model, CVaR model.

JEL Classification: C61, G11
AMS Classification: 91G10,

1 Introduction

Over the years, a significant number of papers focused on portfolio selection problems have been published. One of the pioneers in portfolio optimization was Harry Markowitz who proposed the mean-variance approach that minimizes the variance of a portfolio whilst achieving the expected return, see [12]. Thanks to this paper, this approach has become the most widely known form of portfolio selection and the cornerstone of modern portfolio theory. To implement the Markowitz model, the mean returns and asset's dependence need to be estimated. In this model, the dependence between financial returns is explained by the covariance coefficients and the efficient portfolios are the conventional mean-variance optimization model.

As already mentioned, the possible optimization tasks for this approach is to find the portfolio with minimal risk at the predefined level of expected return. Several alternatives may be chosen as an indicator of portfolio risk, not just the variance or standard deviation. Similarly, several approaches to estimating input variables can be used.

Give the above information, the goal of this paper is to compare selected portfolio optimization models and strategies with the benchmarks, both in different world markets and differing time periods. For simplification in empirical analysis, only stocks are used and US, Chinese and UK market are considered. The topic and the analysis are motivated by the effort to find a suitable and simple model for ordinary investors, whose decision-making is made under different economic conditions.

The whole paper is divided into 5 sections. The introduction and the structure of the paper are described in section 1. In section 2, a brief literature review of the related research is made, and the descriptions of portfolio models are introduced in section 3. To verify the efficiency of the applied approaches, the empirical analysis is provided in section 4, and in section 5, the whole paper is summarized and concluded.

2 Literature Review

In his path-breaking work on a portfolio optimization problem, Markowitz in [12] considers how investors can maximize an expected return for a given risk level, or equivalently, minimize the risk for a given expected return. In original Markowitz's formulation and later specifications, e.g. including the CAPM of Sharpe in [15], the measure of risk is considered for a variance of returns in an investor's overall portfolio.

[1] VŠB – Technical University of Ostrava, Department of Finance, Sokolská tř. 33, Ostrava 70200, Czech Republic, david.nedela@vsb.cz.

By this meaning, the mean return vector and covariance matrix of individual asset returns provide us everything necessary to carry out a portfolio selection. In [5] proposed the use of approximation to variance instead of using variance as a measure of risk. They concluded that the linear model for portfolio optimization is computationally efficient and the results are not different from the Mean-Variance model.

In other approaches, different ways of capturing portfolio risk can be found. For example, optimization problems involving VaR can be found also in many papers, e.g. [9] or [10]. In [14] providing a linear formulation for the CVaR model that extends the traditional VaR approach and focuses only on the downside risk. CVaR is an alternative measure of risk due to better properties than VaR, see [2].

Although CVaR was not standard in the financial sector, it was primarily applied in the insurance sector. Uryasev in [16] provides a simple description of the approach for minimization of CVaR and also optimization problems with CVaR. To find methodology provided for simultaneous calculating VaR and optimizing CVaR see [8]. In [11] authors focused on a linear programming extension of the CVaR model mainly based on simple combinations of CVaR measures (weighted CVaR) and two different types of weight settings, being wide CVaR and tail CVaR. Their research concluded that by the tail CVaR is generated diversified portfolios than wide CVaR.

In the matter of portfolio optimization, the risk of parameter estimation and subjective views can be studied with the Bayesian portfolio optimization framework, e.g. [3], [4]. All the studies assume that asset returns are identically and independently distributed (i.i.d.) through time. In [1], they account for the possibility that returns of assets are predictable by macroeconomics variables. The various Bayesian models commonly define optimal asset weights in the portfolio, by using expected utility maximization (mean-variance optimization), see e.g. [13].

3 Portfolio Optimization Approaches

In this section, the chosen approaches of resolving portfolio optimization problems are described. In this paper, the investor's general utility function (1) is applied with parameter k, denoting the investor's attitude to risk.

$$U = k \cdot v_P - (1 - k) \cdot E(R_P) \tag{1}$$

Where $E(R_p)$ is expected return of a portfolio and v_p is risk indicator of a portfolio. If $k = 0$, then the investor is risk neutral and if $k > 0$, the investor is risk averse. If $k = 1$, then the investor is absolute risk averse.

3.1 Markowitz Model

The Markowitz model belongs to the group of mean-variance models based on the conversion of factors into two parameters: expected return (mean) and risk expressed by variance, see [12]. The reality is simplified by the following main assumptions: 1. single-period model, where the investor decides only for one time period; 2. the investor is rational and risk-averse with concave and increasing utility function; 3. markets are (informatively) effective and transaction costs and taxes are neglected; 4. assets can be infinitely divided.

Before using this model, it is necessary to calculate following variables that are inserted into the model. If $E(R_i)$ is assumed as the expected return of i-th asset, x_i as the weight of i-th asset in a portfolio, then the calculation of expected return of a portfolio $E(R_p)$ is as following formulation:

$$E(R_P) = \sum_{i=1}^{N} x_i \cdot E(R_i) \tag{2}$$

As is previous mentioned, in the Markowitz model, it is worked with variance or standard deviation, expressing the risk measure. The calculation of variance σ_p^2 and standard deviation σ_p of a portfolio are in equation (3), (4).

$$\sigma_P^2 = \sum_i \sum_j x_i \cdot x_j \cdot \sigma_{ij} \tag{3}$$

where the standard deviation is calculated as the square root of variance:

$$\sigma_P = \sqrt{\sigma_P^2} = \sqrt{\sum_i \sum_j x_i \cdot x_j \cdot \sigma_{ij}} \tag{4}$$

The covariance is defined as statistical dependency between *i*-th and *j*-th assets in a portfolio. The calculation of covariance is achieved by using following formula,

$$\sigma_{ij} = \frac{1}{N} \cdot \sum_t [R_{it} - E(R_i)] \cdot [R_{jt} - E(R_j)] \tag{5}$$

To find the optimal portfolio, it is necessary to know the efficient frontier from the allowable frontier. The efficient frontier is made up of points representing the expected mean value of return and variance, whilst noting it is not possible to improve one parameter without deteriorating the other. A portfolio with a minimum of risk can be defined as the following optimization problem:

$$\begin{gathered} \min \sigma_P^2 \\ \sum_i^n x_i = 1 \\ x_i \geq 0, \text{for } i = 1,2, \dots, N, \end{gathered} \tag{6}$$

3.2 Bayesian Optimization Model

In the Bayesian approach we consider the subjective assumption of estimating the parameter of probability distribution and the covariance matrix. The probability distribution is characterized by a combination of subjective distribution and the probability distribution of selection. For these differences, the Bayes-Stein portfolio strategy is applied in this paper. Due to the instability of the classical mean-variance optimum, many studies have been written on how to reduce the mean-variance error of the optimal portfolio. As stated in [7], according to the Bayes-Stein strategy, the expected return on assets should be estimated using equations (7) and (8),

$$\mu_t^{BS} = (1 - \delta_t) \cdot E(R_{i,t}) + \delta_t \cdot \bar{\mu}_t \tag{7}$$

$$\delta_t = \frac{n + 2}{n + 2 + m \cdot \left(E(R_{i,t}) - \bar{\mu}_t\right)^T \cdot C_t^{-1} \cdot \left(E(R_{i,t}) - \bar{\mu}_t\right)} \tag{8}$$

where μ_t^{BS} in equation (7) represents the expected return while using the Bayesian approach, $E(R_{i,t})$ is the expected return on the *i*-th asset in time t, $\bar{\mu}_t$ is the average expected return on all assets, δ_t is an estimation variable used for calculation μ_t^{BS}. In the second equation (8), the variable n represents the amount of invested assets, m is the number of historical observations and C_t is the covariance matrix.

The Bayesian strategy not only shrinks the expected return, but also shrinks the covariance and the amount of shrinkage depends on the number of assets and the number of observations. The process for the calculation of the covariance matrix of the Bayesian strategy C_t^{BS} is described as:

$$C_t^{BS} = C_t \cdot \left(1 + \frac{1}{m + \tau_t}\right) + \frac{\tau_t}{m \cdot (m + 1 + \tau_t)} \cdot \frac{1_n \cdot (1_n)^T}{1_n \cdot C_t^{-1} (1_n)^T} \tag{9}$$

where $\tau_t = \frac{m \cdot \delta_t}{1 - \delta_t}$, 1_n is n unit vector, τ_t is the information accuracy before μ.

The newly obtained expected return values together with the covariance matrix can then be used for the optimized utility function (1). The weights of individual assets in the portfolio are calculated as follows:

$$x_t^{BS} = \frac{(C_t^{BS})^{-1} \cdot \mu_t^{BS}}{(1_n)^T \cdot (C_t^{BS})^{-1} \cdot \mu_t^{BS}} \tag{10}$$

3.3 CVaR Model

Another approach to optimizing a portfolio of financial instruments reduce risk again but expressed by Conditional Value-at-Risk rather than variance or Value-at-Risk (VaR). CVaR is a popular tool for managing risk and it approximately (or exactly) equals the average of some percentage of the worst-case loss scenarios.

Portfolios with low CVaR necessarily have a low value of VaR as well, see [14]. The general equation of $\mathrm{CVaR}_{X,\alpha}$ for variable X is formulated as:

$$\mathrm{CVaR}_{X,\alpha} = -\mathrm{E}[X|X < -\mathrm{VaR}_{X,\alpha}] \tag{11}$$

where $\mathrm{CVaR}_{X,\alpha}$ is defined as

$$\mathrm{VaR}_{X,\alpha} = -\min\{x|\ \Pr(X \leq x) \geq \alpha\} \tag{12}$$

The $\mathrm{CVaR}_{X,\alpha}$ value of vector X_i expressing asset returns at significance level α for length of vector n can be calculated as following equation,

$$\mathrm{CVaR}_{X,\alpha} = -\frac{1}{\alpha}\left(\frac{1}{n}\sum_{i=1}^{[\alpha n]-1} X_i + \left(\alpha - \frac{[\alpha n]-1}{n}\right) X_{[\alpha n]}\right) \tag{13}$$

If $\mathrm{CVaR}_{X,\alpha}$ is considered as a measure of risk, and being implement in the optimization task, minimizing CVaR for utility function (1) where CVaR replace v_p, is then defined as

$$\begin{gathered} \min \mathrm{CVaR} \\ \sum_{i}^{n} x_i = 1 \\ x_i \geq 0, for\ i = 1,2,\dots,N \end{gathered} \tag{14}$$

4 Empirical Analysis and Data Description

4.1 Data

For the empirical analysis, daily adjusted close prices of stocks included in indices NASDAQ-100, Hang Seng, FTSE 100 traded on the US, Chinese, and UK stock market are used. These indices represent the benchmarks for comparison and since the comparison is performed at different time periods, two periods are selected. These contain both the crisis period (2005–2010) and the boom period (2014–2019). The prices of individual stocks were obtained from the Yahoo Finance[2] website. The risk-free rate used for the calculation of performance indicators, was the US 10 Year Treasury return, available from the CNBC[3] website. In all indices, there are several stock time series that are not included in the analysis due to the incomplete data in the analysis period.

4.2 Applications of Optimization Models

In this subsection, selected portfolio models are applied to described data and compared with the benchmark. In all models it is assumed that at the beginning of each month the portfolio is re-optimized. In other words, the monthly back-testing procedure with the one-year rolling window is applied. For the selected six-year time periods, the daily returns of the first year are used for optimizing the created portfolio, where the investment itself begins on the first trading day of the following year. The investment will last until the end of 6 years, i.e. 5 years with monthly re-optimization. In relation to the utility function, different investor attitudes to risk, expressed by the value of k, are applied and where the interval is from 0 to 1, in increments of 0.2.

According to the daily returns of the stocks, together with the calculated weights using the optimization function in equations (6) and (14), the final wealth W for all portfolios and portfolio performance measures are calculated. For simplification, in all portfolio investments the amount of initial wealth W_0 is equal to 1. The expected return, standard deviation, Sharpe ratio (SR), Rachev ratio (RR) and Value-at-Risk are all based on daily results, and the (VaR) values of $\alpha = 0.05$ and $\beta = 0.05$ are used. The results of portfolio wealth values and portfolio performance measures are in the Tables 1–6 for both markets and time periods. If $k = 0$ the results are identical for all models.

Throughout the time periods, it is not possible to conclusively define which model is the most appropriate to use for a particular market, or for a particular economic condition. Obviously, it should mainly depend

[2] Available from: https://finance.yahoo.com/
[3] Available from: https://www.cnbc.com/quotes/?symbol=US10Y

on the investor's behavior and rationality. Compared to the other selected models, it is visible that in all markets the CVaR model is the one with least risk. Generally, the lower values of W or $E(R_p)$ are mostly associated to risk. However, in some cases, e.g. in Table 1 or in Table 5, the CVaR model would be a preferred choice for individual types of investors, having the highest W and related SR or RR, expressing a relation of potential gain to potential loss.

	Markowitz Model						Bayesian Optimization Model					CVaR Model				
k	0	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
W	0.5431	0.5690	0.9597	1.0787	1.2224	1.2224	0.9377	1.0541	1.2489	1.2077	1.2164	1.5134	1.3498	1.3274	1.4091	1.3918
$E(R_P)$	0.0002	0.0002	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003	0.0002	0.0004	0.0003	0.0003	0.0003	0.0003
σ_P	0.0361	0.0343	0.0294	0.0248	0.0199	0.0199	0.0309	0.0258	0.0215	0.0164	0.0109	0.0130	0.0118	0.0115	0.0115	0.0114
SR	0.0036	0.0030	0.0112	0.0119	0.0142	0.0142	0.0117	0.0117	0.0155	0.0128	0.0128	0.0260	0.0197	0.0187	0.0229	0.0221
RR	0.9340	0.9337	0.9837	0.9923	0.9487	0.9487	0.9807	0.9940	0.9803	0.9119	0.9188	0.9369	0.9967	0.9925	1.0001	0.9921
$VaR_{0.05}$	0.0503	0.0469	0.0449	0.0408	0.0338	0.0338	0.0456	0.0418	0.0350	0.0282	0.0157	0.0198	0.0167	0.0167	0.0164	0.0170

Table 1 Portfolio performance indicators of NASDAQ-100 stocks in the investing period 2006–2010

	Markowitz Model						Bayesian Optimization Model					CVaR Model				
k	0	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
W	1.6409	2.2441	3.2204	3.4172	4.9781	1.8198	2.7817	2.9653	3.6734	4.7593	1.8207	3.3871	2.5303	2.1240	1.9141	1.8877
$E(R_P)$	0.0010	0.0012	0.0014	0.0013	0.0015	0.0006	0.0013	0.0012	0.0013	0.0014	0.0006	0.0011	0.0008	0.0007	0.0006	0.0006
σ_P	0.0355	0.0323	0.0291	0.0255	0.0202	0.0120	0.0302	0.0262	0.0217	0.0179	0.0120	0.0141	0.0127	0.0124	0.0124	0.0125
SR	0.0269	0.0341	0.0445	0.0489	0.0709	0.0401	0.0400	0.0438	0.0561	0.0755	0.0402	0.0718	0.0597	0.0495	0.0426	0.0416
RR	1.1062	1.0703	1.0712	1.0421	1.0370	0.9502	1.0813	1.0513	1.0592	1.0092	0.9504	1.0326	1.0188	0.9906	0.9821	0.9731
$VaR_{0.05}$	0.0556	0.0498	0.0434	0.0395	0.0316	0.0155	0.0457	0.0396	0.0320	0.0279	0.0155	0.0205	0.0172	0.0173	0.0175	0.0176

Table 2 Performance indicators of Hang Seng stock portfolio in the investing period 2006–2010

	Markowitz Model						Bayesian Optimization Model					CVaR Model				
k	0	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
W	1.4357	1.2325	0.9268	1.0109	0.9092	1.3181	0.9906	1.0537	0.9109	1.1291	1.3181	1.4770	1.2468	1.2132	0.9642	1.0771
$E(R_P)$	0.0008	0.0006	0.0003	0.0002	0.0001	0.9875	0.0003	0.0003	0.0001	0.0002	0.0003	0.0004	0.0002	0.0002	0.0001	0.0001
σ_P	0.0322	0.0298	0.0262	0.0219	0.0166	0.0111	0.0263	0.0217	0.0173	0.0136	0.0111	0.0118	0.0116	0.0117	0.0141	0.0115
SR	0.0230	0.0182	0.0080	0.0080	−0.001	0.0187	0.0101	0.0094	0.0000	0.0084	0.0187	0.0260	0.0145	0.0126	−0.000	0.0043
RR	0.9875	0.9472	0.8817	0.8829	0.8858	0.9375	0.8894	0.8966	0.8920	0.8686	0.9376	0.9217	0.9201	0.9136	0.9990	0.9390
$VaR_{0.05}$	0.0426	0.0400	0.0390	0.0325	0.0238	0.0172	0.0387	0.0320	0.0243	0.0208	0.0172	0.0169	0.0169	0.0173	0.0236	0.0172

Table 3 Performance indicators of FTSE 100 stock portfolio in the investing period 2006–2010

	Markowitz Model						Bayesian Optimization Model					CVaR Model				
k	0	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
W	2.0737	2.9036	3.2647	3.6192	3.1033	1.5621	3.1574	3.5404	3.4202	2.4591	1.5614	1.9490	1.8214	1.7753	1.7699	1.7699
$E(R_P)$	0.0011	0.0013	0.0013	0.0013	0.0011	0.0004	0.0013	0.0013	0.0012	0.0008	0.0004	0.0006	0.0005	0.0005	0.0005	0.0005
σ_P	0.0304	0.0290	0.0259	0.0225	0.0171	0.0071	0.0271	0.0232	0.0192	0.0134	0.0071	0.0091	0.0078	0.0076	0.0075	0.0075
SR	0.0320	0.0414	0.0465	0.0534	0.0569	0.0427	0.0448	0.0518	0.0566	0.0544	0.0426	0.0547	0.0554	0.0541	0.0543	0.0543
RR	1.0200	1.0313	0.9942	0.9595	0.8834	0.8909	0.9980	0.9625	0.9185	0.8235	0.8907	0.8691	0.8536	0.8458	0.8713	0.8713
$VaR_{0.05}$	0.0463	0.0438	0.0401	0.0337	0.0257	0.0123	0.0418	0.0352	0.0289	0.0224	0.0123	0.0143	0.0130	0.0130	0.0131	0.0127

Table 4 Portfolio performance indicators of NASDAQ-100 stocks in the investing period from 2015–2019

	Markowitz Model						Bayesian Optimization Model					CVaR Model				
k	0	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
W	2.2248	2.2938	1.4497	1.4452	1.9108	1.8840	1.6119	1.3814	1.8942	2.0610	1.9201	2.3893	2.2737	2.1734	2.1323	2.1409
$E(R_P)$	0.0010	0.0010	0.0006	0.0005	0.0007	0.0005	0.0007	0.0005	0.0007	0.0007	0.0006	0.0008	0.0007	0.0007	0.0007	0.0007
σ_P	0.0276	0.0265	0.0251	0.0219	0.0169	0.0074	0.0255	0.0223	0.0183	0.0135	0.0074	0.0102	0.0091	0.0088	0.0087	0.0086
SR	0.0345	0.0358	0.0215	0.0211	0.0350	0.0633	0.0250	0.0196	0.0334	0.0447	0.0656	0.0672	0.0694	0.0670	0.0661	0.0673
RR	1.0527	1.0533	0.9893	0.9896	0.9856	0.8915	1.0039	0.9933	0.9928	0.9934	0.8914	0.9452	0.9550	0.9415	0.9336	0.9293

	Markowitz Model						Bayesian Optimization Model					CVaR Model				
k	0	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$VaR_{0.05}$	0.0400	0.0381	0.0381	0.0334	0.0265	0.0121	0.0383	0.0333	0.0285	0.0207	0.0121	0.0146	0.0124	0.0122	0.0122	0.0120

Table 5 Performance indicators of Hang Seng stock portfolio in the investing period 2015–2019

	Markowitz Model						Bayesian Optimization Model					CVaR Model				
k	0	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
W	5.3794	5.6703	5.9450	5.3828	4.1550	1.4686	6.7624	5.7492	4.6151	3.0602	1.4684	3.0095	2.1660	2.3406	2.1348	2.5475
$E(R_P)$	0.0016	0.0016	0.0016	0.0015	0.0012	0.0003	0.0018	0.0016	0.0013	0.0010	0.0003	0.0009	0.0007	0.0007	0.0006	0.0008
σ_P	0.0238	0.0231	0.0211	0.0181	0.0148	0.0077	0.0227	0.0188	0.0158	0.0117	0.0077	0.0095	0.0093	0.0082	0.0080	0.0082
SR	0.0642	0.0673	0.0735	0.0782	0.0780	0.0334	0.0741	0.0787	0.0793	0.0747	0.0334	0.0881	0.0626	0.0768	0.0692	0.0847
RR	1.2414	1.2604	1.3025	0.0780	1.1989	0.9574	1.3484	1.2619	1.2011	1.1235	0.9574	1.2037	1.0054	1.1292	1.1335	1.2009
$VaR_{0.05}$	0.0317	0.0303	0.0285	0.0245	0.0200	0.0119	0.0290	0.0250	0.0216	0.0176	0.0119	0.0133	0.0124	0.0117	0.0120	0.0118

Table 6 Performance indicators of FTSE 100 stock portfolio in the investing period 2015–2019

t	2006–2010						2015–2019					
	W	$E(R_P)$	σ_P	SR	RR	$VaR_{0.05}$	W	$E(R_P)$	σ_P	SR	RR	$VaR_{0.05}$
NASDAQ-100	1.2029	0.0003	0.0164	0.0125	0.9823	0.0261	1.8889	0.0006	0.0102	0.0473	0.8396	0.0167
Hang Seng	1.5202	0.0005	0.0198	0.0231	1.0188	0.0306	1.1103	0.0001	0.0109	0.0063	0.9291	0.0177
FTSE 100	1.0500	0.0001	0.0147	0.0048	0.9642	0.0238	1.1487	0.0001	0.0087	0.0083	0.9740	0.0137

Table 7 Benchmarks performance indicators

In some periods through the Markowitz model or Bayesian model, W is rapidly higher than in the CVaR model. If the Markowitz model and the Bayesian model are compared, the W for individual k is similar because the optimization is based on an equivalent function. If the results of one model for individual k exceed the other model, either for k = 0.2 or k = 0.8, the W is opposite and indicates a different slope, either at the beginning or the end of the efficient frontier curve. However, in almost all periods and based on risk indicators, the Bayesian model is less risky than the Markowitz model. It is also more profitable due to SR for individual k.

The comparison with the benchmark (Table 7) is influenced most significantly by the absence of transaction costs for the portfolio models. It can be observed that in some cases, the return of investment for the benchmark surpasses the return of investment by the portfolio model. Surprisingly, a higher $E(R_p)$ and W could be achieved in the financial crisis period in the Chinese market but also with a considerably higher σ_p and $VaR_{0.05}$. The advantage of the investment in this crisis period is confirmed by the SR and RR values and it could be concluded that in this crisis, the Chinese market was less affected than the US and UK markets.

5 Conclusion

Due to the possibility of using many portfolio models, the comparison of the selected general models was provided in this paper. The goal of this paper was defined as a comparison of the portfolio models in differing periods and markets, with benchmarks. The back-testing approach of the Markowitz model, the Bayesian model, and the CVaR model was applied for the dataset, and then the performance of each model measured. Subsequently, the same comparable benchmark indicators were calculated for comparison, both for individual markets and periods.

For the empirical results, it is not possible to unequivocally determine whether any of the selected models are more suitable in a crisis time period, or a more favorable economic one. However, from the results of the risk indicators it was obvious that investors can achieve the lowest risk of investment when using the CVaR model for creating a portfolio. The lower risk is offset by lower daily return and final wealth but, in some cases, both were below the related benchmark. Another identified conclusion was that the Chinese market was more profitable in the financial crisis period than the US and UK markets, which may represent a more favorable investment location.

Acknowledgements

The author acknowledges a support provided by the Czech Science Foundation (GACR) under project 20-16764S and VSB-TU Ostrava under the SGS project SP2020/11.

References

[1] AVRAMOV, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64(3), 423–458.
[2] ARTZNER, P., DELBAEN, F. E., & HEATH, M. D. (1997). ''Thinking Coherently". *Risk,* 10, pp. 68–71.
[3] JORION, P. (1986). Bayes Stein estimation for portfolio analysis. *Journal of Finance and Quantitative Analysis*, **21**(3), 279–292.
[4] BLACK, F. & LITTERMAN, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5), 28–43.
[5] KONNO, H. & YAMAZAKI, H. (1991). Mean-Absolute deviation portfolio optimization model and its applications to Tokyo stock-market. *Management Science*, 37(5), 591–531.
[6] KOUAISSAH, N., ORTOBELLI, S., & TICHÝ, T. (2018). *Portfolio Theory and Conditional Expectations: Selected Models and Applications.* SAEI, vol. 59, VŠB-TU Ostrava. ISBN 978-80-248-4343-8.
[7] KRESTA, A. (2016). *Kvantitativní metody investování s aplikacemi v prostředí Matlab*. SAEI, vol. 43, VŠB-TU Ostrava. ISBN 978-80-248-3898-4.
[8] KROKHMAL, P., PALMQUIST, J., & URYASEV, S. (2002). Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, 4, 43–68.
[9] LITTERMAN, R. (1997). "Hot spots and hedges (I)". *Risk,* 10(3), 42–45.
[10] LUCAS, A. & KLAASSEN, P. (1998). Extreme returns, downside risk, and optimal asset allocation. *The Journal of Portfolio Management*, 25(1), 71–79.
[11] MANSINI, R., OGRYCZAK, W., & SPERANZA, M. G. (2007). Conditional value at risk and related linear programming models for portfolio optimization. *Annals of Operations Research*, 152(1), 227–256.
[12] MARKOWITZ, H. (1952). Portfolio Selection. *The Journal of Finance,* 7(1), 77–91.
[13] POLSON, N. G. & TEW, B. V. (2000). Bayesian portfolio selection: An empirical analysis of the S&P 500 index 1970–1996. *Journal of Business & Economic Statistics,* 18(2), 164–173.
[14] ROCKAFELLAR, R. T. & URYASEV, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk,* 2(3), 21–41.
[15] SHARPE, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425–442.
[16] URYASEV, S. (2000). Conditional value-at-risk: optimization algorithms and applications. *Financial Engineering News*, 14, February 2000.

Parametric and Non-parametric Estimates of Military Expenditure Probability Distribution

Jiří Neubauer[1], Martin Tejkal[2], Jakub Odehnal[3], Tereza Ambler[4]

Abstract. The contribution focuses on modeling of probability distribution of military expenditure. Analysis of the link between military expenditure and other macroeconomic variables is a widely discussed issue in the defense economic literature. The probability distribution of analyzed variables can significantly influence the performance of estimated models. The aim of this paper is to find an appropriate probability distribution for data describing military expenditure as a percentage of gross domestic product in the years 1993–2017. The statistical analysis involves data of 162 countries. The authors apply three parametric models of probability distribution (lognormal, gamma and Weibull distribution) and compare these models with non-parametric models based on the kernel estimation. The quality of a distribution fit is based on comparison of an empirical and a theoretical cumulative distribution function and is tested by Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling test. Based on the estimated models, quantiles describing the distribution of military expenditures for small and large values were calculated.

Keywords: distribution fitting, military expenditure, kernel estimates

JEL Classification: C12, E69
AMS Classification: 62F99

1 Introduction

Military expenditure represents financial means intended for financing defense as a public good. Many recently published articles focus on military expenditure definition, manners of measurement, quantification of determinants, and analyses of the possible multiplier effect of military expenditure, see [11], [6], [9] or [10]. The need for keeping international records and comparing military expenditure arises inter alia from the requirements for military expenditure analyses indicating and showing potential armament of armies and increasing security threats and risks. The article features an approach to the military expenditure analysis from the point of view of probability distribution modeling and is a follow-up article on the results [15].

The authors [11], [6], [9] and [10] analyze military expenditure from the point of view of its absolute level, or as the share of military expenditure of the respective country's gross domestic product. This article analyses military expenditure (military expenditure share of gross domestic product), perceived as an indicator of the military burden on the economy, in 162 selected countries from 1993 through 2017. NATO member states use this indicator as an optimum indicator of the recommended military expenditure level which should represent 2% of gross domestic product based on the conclusions of the Prague NATO summit and subsequently also the Wales NATO summit.

Countries carrying the greatest military burden measured as the military expenditure share of gross domestic product within the analyzed period are Oman (average level 12.4%) and Saudi Arabia (9.8%). The geographical division of 162 countries into groups, see [14] (Africa, America, Asia and Oceania, Europe, Middle East), and a comparison of these groups show that Eritrea stands out on the African continent; however, its average level (22.6%) is affected by a large number of missing observations when military expenditure data of Eritrea is available only until 2004. According to [14], the quality of data files can be influenced by different data reporting and checking systems in developed and developing countries, or by deliberate efforts of governments to understate actual military expenditure levels.

The United States dominate the American continent in a long-term perspective; the U.S.A. are simultaneously one of few NATO member states fulfilling the recommended obligation to spend 2% of gross domestic

[1] University of Defence in Brno, Department of Quantitative Methods, Kounicova 65, Brno, Czech Republic, jiri.neubauer@unob.cz
[2] University of Defence in Brno, Department of Quantitative Methods, Kounicova 65, Brno, Czech Republic, martin.tejkal@unob.cz
[3] University of Defence in Brno, Department of Economics, Kounicova 65, Brno, Czech Republic, jakub.odehnal@unob.cz
[4] University of Defence in Brno, Department of Economics, Kounicova 65, Brno, Czech Republic, intedo@seznam.cz

product on defense. The military burden on the economy significantly increased after 2001 when the military expenditure level was rising dramatically in consequence of the 2001 terrorist attacks and the resulting wars in Iraq and Afghanistan. The military burden on the U.S. economy kept increasing even during the economic crisis; however, this increase was primarily caused by a decrease in gross domestic product in consequence of the crisis in the U.S.A. Among Asian countries, it is Pakistan that shows the greatest average military burden on the economy. As regards European countries, Russia clearly dominates. Other countries having great military expenditure measured as a share of gross domestic product are Ukraine, Georgia, Armenia, Azerbaijan, and Croatia. Greece, the United Kingdom, and France are the only European NATO member states meeting the above political obligation to spend 2% of gross domestic product on defense. Oman and Saudi Arabia dominate the region of the Middle East. Significantly higher levels of military expenditure in comparison with other regions can be seen in Israel, Kuwait, Syria, the United Arab Emirates, and Yemen.

The following part of the article analyzes military expenditure (share of gross domestic product) of 162 countries by means of probability distribution. We apply selected parametric models (lognormal, gamma, and Weilbull distribution) for the purpose of modeling military expenditure. We compare these models with non-parametric models based on the kernel approach.

2 Materials and Methods

The main goal of this article is to analyze and model the distribution of military expenditure. The necessary data was obtained from the SIPRI (Stockholm International Peace Research Institute) database. This article follows the research results presented in [15] where authors applied selected parametric probability distributions to describe military expenditure per capita in years 1993–2016. We focus now on military expenditure, expressed as a percentage of gross domestic product (GDP), in years 1993–2017. We analyze data from $n = 162$ countries. In some years, the data were not available for all of the 162 countries. Denote m_j the number of countries for which the data are absent in a given year $j \in \{1993, \dots, 2017\}$. Henceforth, the random sample of military expenditure as percentage of GDP in a year $j \in \{1993, \dots, 2017\}$ will be denoted $\boldsymbol{X}_j = (X_{j1}, X_{j2}, \dots, X_{jn-m_j})^T$.

2.1 Parametric Methods

In the parametric approach, it was assumed that the random sample $\boldsymbol{X}_j$ for each year $j \in \{1993, \dots, 2017\}$ is drawn from an unknown distribution with cumulative distribution function (CDF) $F_{\boldsymbol{X}_j}$. The following probability distributions were considered: the lognormal, gamma, and the two-parametric Weibull distribution. The normal distribution was not considered in this paper. Based on the previous research [15] it was determined that the normal distribution is not suitable for modeling military expenditure. The unknown parameters μ_j and σ_j^2 of the log-normal distribution, α_j and β_j of the gamma distribution, and k_j and λ_j of the Weibull distribution were estimated from $\boldsymbol{X}_j$ for each year $j \in \{1993, \dots, 2017\}$. The maximum goodness of fit method as described in [7] was used to obtain the estimates. The respective parameter estimates will be denoted with the "hat" symbol. I. e. $\widehat{\mu}_j$ is the estimate of μ_j and so on.

In the next step the goodness-of-fit tests based on the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling metric were carried out for each $j \in \{1993, \dots, 2017\}$ to test the following hypothesis:

$$H_0 : F_{\boldsymbol{X}_j} = \widehat{F}, \tag{1}$$

where $\widehat{F}$ is either the CDF of the log-normal distribution $F_{LN}(\widehat{\mu}_j, \widehat{\sigma}_j^2)$, the gamma distribution $F_G(\widehat{\alpha}_j, \widehat{\beta}_j)$, or the Weibull distribution $F_W(\widehat{k}_j, \widehat{\lambda}_j)$, and the parameters are the estimates obtained from $\boldsymbol{X}_j$. For more details about the tests see [3].

2.2 Non-parametric Methods

Let $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ be a random sample drawn from some distribution with an unknown density f. Its kernel density estimator is [12, 16]

$$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{2}$$

where K is the *kernel* (non-negative function) and $h > 0$ is a smoothing parameter called the *bandwidth*. Commonly used kernels are: uniform, triangular, biweight, triweight, Epanechnikov, normal, and others. In

our computation, we applied the normal kernel. For smooth densities and the normal kernel, we used the bandwidth

$$h = \frac{1.06\widehat{\sigma}}{n^{1/5}}, \tag{3}$$

where $\widehat{\sigma} = \min\left\{s, \frac{IQR}{1.34}\right\}$, s is the sample standard deviation and IQR is the interquartile range.

Military expenditure described as a percentage of GDP is non-negative in general. The kernel estimator (2) is defined for all real values x. Therefore, for negative values near zero a non-zero estimates of the density are produced. We want to obtain density estimator for non-negative data which is non-zero only for $x > 0$. In order to achieve that, we decided to use the following two methods: a reflected density and a logarithmic transformation. The density estimator based on the reflected density method is [12, 2, 13]

$$\widehat{f}_h^{\text{refl}}(x) = \frac{1}{nh}\sum_{i=1}^{n}\chi_{[0,\infty]}(x)\left[K\left(\frac{x-X_i}{h}\right) + K\left(\frac{x+X_i}{h}\right)\right], \tag{4}$$

where

$$\chi_{[0,\infty]}(x) = \begin{cases} 1 & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases} \tag{5}$$

The logarithmic transformation offers another way to estimate the density. Let $Y = \log X$, $Y_i = \log X_i$ and $f_{h,Y}(y)$ be the density of Y. The density estimator is given by the formula [12, 1, 17, 8]

$$\widehat{f}_h^{\text{log}}(x) = x^{-1}\widehat{f}_{h,Y}(\log x) = \frac{1}{nh}\sum_{i=1}^{n} x^{-1}K\left(\frac{\log x - \log X_i}{h}\right). \tag{6}$$

It should be noted that the estimator $\widehat{f}_h^{\text{log}}(x)$ is "unstable" for values of x near zero (the estimates are not smooth). We solved this problem by estimating the density (6) for given $x_0 > 0$ and and then using a linear interpolation on the interval $(0, x_0]$.

3 Results

The density estimates were computed in statistical software R. For parametric models, we use the functions from the package *fitdistrplus* [4]. The kernel density estimates were computed with packages *stats*, *GoFKernel* and *logKDE* [8]. Figure 1 shows density estimates, in the left graph are the parametric models' density estimates, in the right graph are the kernel density estimates.

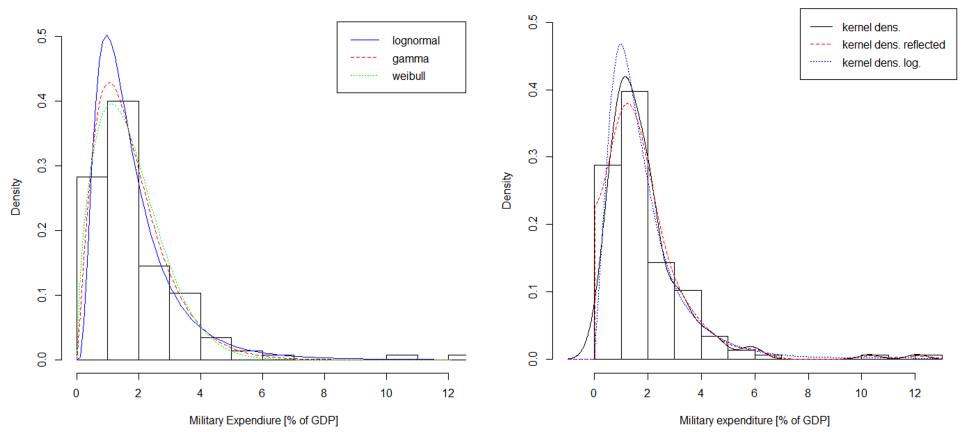


Figure 1 Parametric and non-parametric estimates of military expenditure density in 2017

The quality of the fit can be assessed by QQ-plots. Figure 2 shows QQ-plot corresponding to parametric (left) and non-parametric (right) density estimates of military expenditure in 2017. One can see that the parametric models describe the distribution of military expenditure satisfactorily for small values. However, for greater values of military expenditure, the fit of the parametric models is not very good. Non-parametric

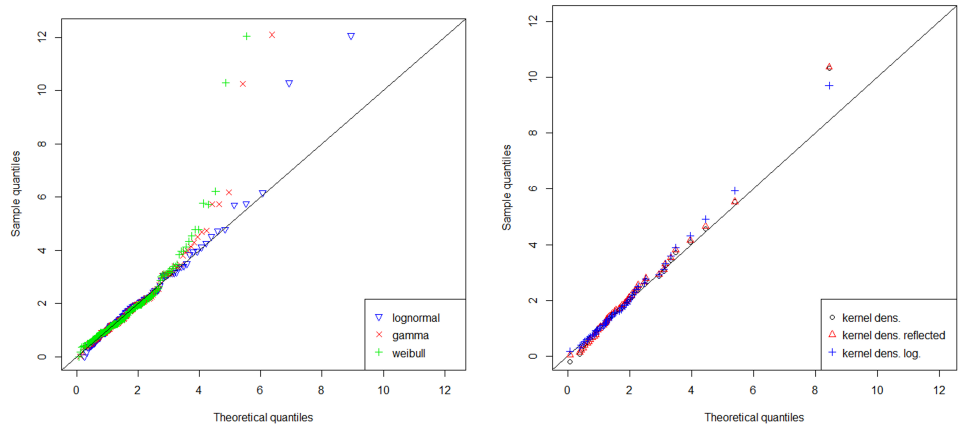


Figure 2 QQ-plots of kernel density estimates – year 2017

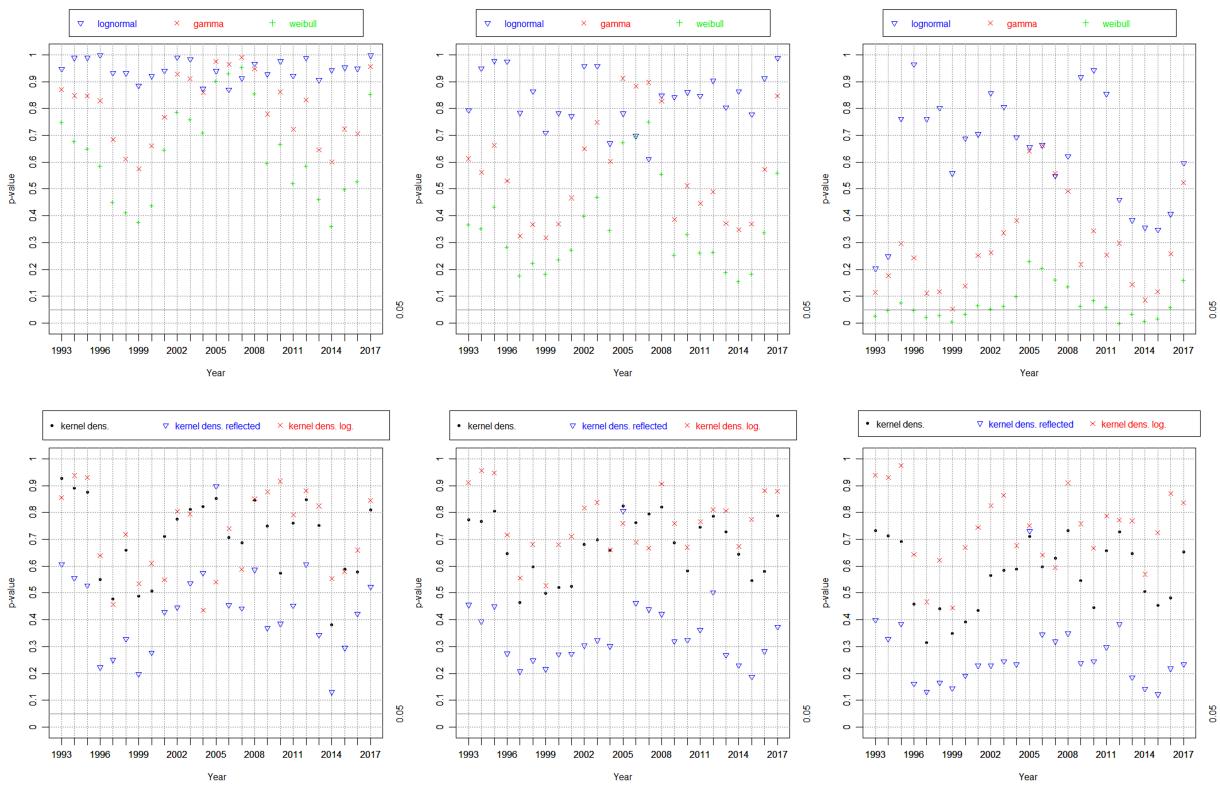


Figure 3 *p*-values of Kolmogorov-Smirnov (left), Cramer-von Mises (middle) and Anderson-Darling test (right) for parametric (top) and non-parametric (bottom) density estimates

method based on kernel estimates fit these values better, on the other hand, they suffer from the so-called boundary effects, which can be observed at small values of the military expenditure.

We applied the previously mentioned methods to describe the distribution of military expenditures in each year in the range from 1993 to 2017. The quality of the fit was tested by Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling test [5]. Figure 3 shows the p-values of these tests for each year and method. Based on the Figure 3, it is evident that the p-values of the Anderson-Darling test are mostly smaller than the p-values of the other two tests. This fact is more pronounced for the parametric models. In terms of p-values, the lognormal distribution offers the best fit. On the other hand, Weibull distribution is rejected by Anderson-Darling test in years 1993, 1994, 1996, 1997, 1998, 1999, 2000, 2012, 2013, 2014 and 2015. For all of the nonparametric methods used in the analysis, all p-values are higher than 0.05. The method based on the logarithmic transformation seems to offer the best fit, p-values of the reflected density estimates are smallest.

The motivation for modeling the probability distribution of military expenditures was, among other things, an attempt to describe the probabilities of extreme values, i.e. especially large values of military expenditures. These large values may indicate a potential safety risk. We computed 1%, 5%, 95% and 99% quantiles from estimated densities and compared them with the empirical quantiles. The results are summarized in Figure 4. The quantiles computed from lognormal densities are not too far from the empirical values. Only in the case of 99% quantile, the values are underestimated. This is probably due to the presence of a few extreme values of military expenditure. Non-parametric methods estimate 95% and 99% quantiles better. On the other hand, the estimates of 1% and 5% quantiles are not good. The classical kernel estimator underestimates the empirical values and sometimes gives negative values. The estimator using the logarithmic transformation offers the best estimates from the non-parametric methods. Recall, however, that the instability issue of the estimator, that manifested itself for small values of the expenditure, was treated by substituing the unstable segment by a linear approximation. This artificially improves the estimates. The boundary effect has significantly affected the estimates.

If we look for example at the data from 2017 in more detail (see Figure 1, 4 and SIPRI database), the results showed that Oman is an example of country with extreme military spending (higher than 99% quantile, measured as military spending as a share of GDP). Oman, Israel, Jordan, Kuwait, Saudi Arabia, Congo and Algeria are examples of countries with military spending higher than 95% quantile. On the other end of the scale are countries like Mauritius, Guatemala, Haiti, Panama, Ghana, and Ireland with military expenditure less than 5% quantile.

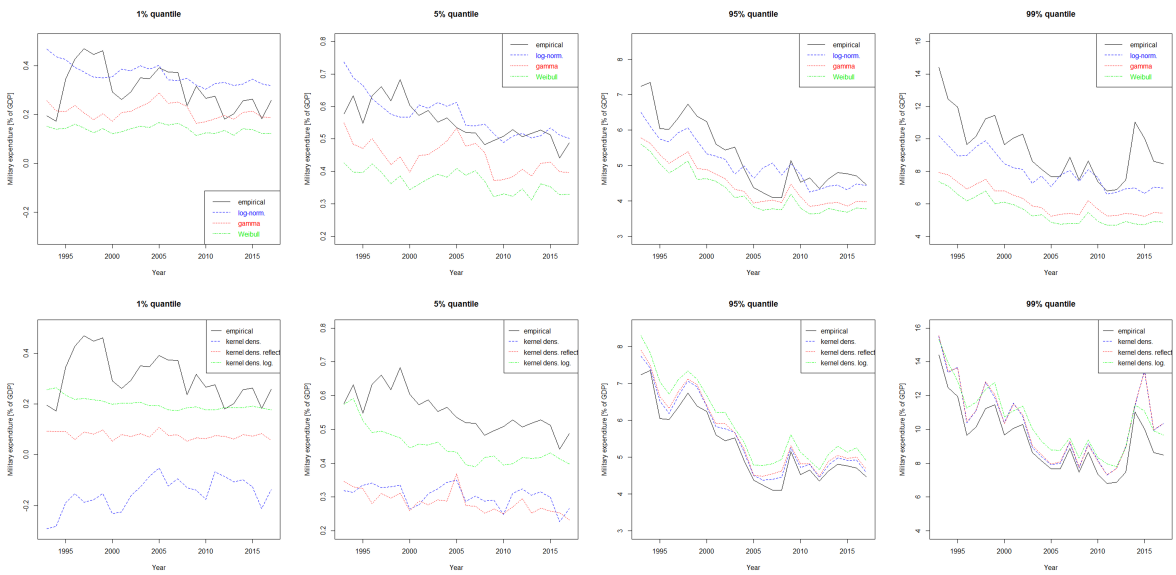


Figure 4 Empirical and estimated quantiles of military expenditure

4 Conclusion

The goal of this contribution was to find a suitable probability distribution of military expenditure expressed as a share of GDP from 1993 to 2017. We can conclude that the lognormal distribution is a suitable parametric model for the military expenditure. This distribution offers better fit than the gamma and Weibull distribution. In addition to the parametric models, the authors also used a non-parametric approach to modeling of the probability distribution of military expenditures. This approach is based on kernel density estimates. The non-parametric methods yield better results than the parametric ones for large values of military expenditure. This is crucial, since the large values of the military expenditure may signalize a potential security risk. On the other hand, the kernel estimates suffers from boundary effects for small values of military expenditure. Normality of variables is a common and frequent assumption of many widely used statistical models and methods. However, the results collected in this contribution further support the fact that normal distribution is not suitable for the modeling of military expenditure. Out of the methods compared in this paper, the lognormal model and the normal kernel estimator used together with the logarithmic transformation yields satisfactory results.

Acknowledgements

The paper was supported by FML Development Projects AERO.

References

[1] Charpentier, A. & Flachaire, E. (2015). Log-Transform Kernel Density Estimation of Income Distribution. *L'Actualité économique*, 91(1–2), 141–159.

[2] Cline, D. B. H. & Hart, J. D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics: A Journal of Theoretical and Applied Statistics*, 22(1), 69–84.

[3] D'Agostino, R. B.& Stephens, M. A. (1986). *Goodness-of-Fit Techniques.* New York: Marcel Dekker.

[4] Delignette-Muller, M. L. & Christophe Dutang (2015). fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, 64(4), 1–34. Available from: http://www.jstatsoft.org/v64/i04/

[5] Faraway, J., Marsaglia, G, Marsaglia J. & Adrian Baddeley (2017). goftest: Classical Goodness-of-Fit Tests for Univariate Distributions. R package version 1.1-1. Available from: https://CRAN.R-project.org/package=goftest

[6] Kollias, C., Manolas, T. & Paleologou, S. Z. (2004) Defence Expenditure and Economic Growth in the European Union: A Causality Analysis. *Journal of Policy Modeling*, 26(5), 553–569.

[7] Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics & Data Analysis*, 51(2), 904–917.

[8] Nguyen, H. D., Jones, A. T. & McLachlan, G. J. (2019). Positive Data Kernel Density Estimation via the LogKDE Package for R. Islam, R., Koh, I. S., Zhao, Y., Warwick, G., Stirling, D.,Li, Ch-T., Islam, Z., eds. *Data mining: 16th Australasian Conference, AusDM 2018, Bahrurst, NSW, Australia, November 28–30, 2018, Revised Selected Papers.* Singapore: Springer.

[9] Odehnal, J. & Neubauer, J. (2015). Ekonomické determinanty vojenských výdajů – kauzální analýza. *Ekonomický časopis*, 63(10), 1019–1032.

[10] Odehnal, J. & Neubauer, J. (2018). Economic, Security, and Political Determinants of Military Spending in NATO Countries. *Defence and Peace Economics.* [online] Available from: https://doi.org/10.1080/10242694.2018.1544440

[11] Sezgin, S. & Yildirim, J. (2002). The Demand for Turkish Defence Expenditure. *Defence and Peace Economics*, 13(2), 121–128

[12] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

[13] Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics – Theory and Methods*, 14(5), 1123–1136.

[14] SIPRI Yearbook 2013. (2013). [online] Avalaible from: http://www.sipri.org/yearbook/2013/

[15] Tejkal, M., Mikulášková, B. & Neubauer, J. (2018). Probability Distribution of Military Expenditure. In L. Váchová & V. Kratochvíl (Eds.), *36th International Conference Mathematical Methods in Economics* (pp. 585–590). Praha: MatfyzPress.

[16] Wand, M. P. & Jones, M.C. (1994). *Kernel Smoothing.* CRC Press.

[17] Wand, M. P., Marron, J. S. & Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association*, 86(414), 343–353.

Risk Measures Prediction and Its Sensitivity to the Refit Step: A Score-Driven Approach

Kateřina Nováková[1], Petra Tomanová[2]

Abstract. The aim of this paper is an assessment of the refit step impact on financial risk measures prediction – Value at Risk (VaR) and Expected Shortfall (ES) – for four world market price indices. Generalized Autoregressive Score (GAS) models assuming the Student's *t*, skew-Student's *t*, and Gaussian distributions are analyzed and compared to the t-GARCH model. VaR and ES predictions are backtested using rolling windows while considering various refit steps. Three different performance measures for predictions are utilized: dynamic quantile test, quantile loss function, and Fissler and Ziegel loss function. The results show that the choice of the refit step does not significantly influence VaR and ES predictions based on GAS models with the Student's *t* and skew-Student's *t* distribution. However, VaR and ES predictions based on Gaussian distribution react extensively in the periods of price shocks.

Keywords: expected shortfall, generalized autoregressive score model, prediction, value at risk

JEL Classification: C22
AMS Classification: 91G70

1 Introduction

Risk measures evaluate the risks that a financial institution goes through. The two leading risk measures are Value at Risk (VaR) and Expected Shortfall (ES). The VaR measures the largest expected portfolio loss over a particular time horizon at a given probability level assuming normal market conditions. It can also be comprehended as an estimate of the largest loss that could occur with $100\alpha\%$ probability based on already known losses within a certain period of time. Despite being widely used by all banks and regulators, VaR does not fulfill one of the axioms of coherence [1]. These axioms strive to distinguish *good* and *bad* risk measures. Breaking some of them can lead to paradoxical results. Specifically, the VaR violates subadditivity since a sum of portfolios might exhibit a higher risk (VaR) than sub-portfolios. The ES is introduced by Rockafellar et al. [6] and it measures the expected loss in the 100α % worst cases where usually $\alpha \in \{0.01, 0.05\}$ and therefore it takes into consideration the shape of the distribution tail. However, it does not consider only the worst case that can occur but the average of the worst cases. The ES is proven to be a coherent indicator [5].

The risk measure estimation requires an accurate estimate of the conditional distribution of future returns. Then, the VaR and ES at time *t* for a risk level α can be computed as

$$VaR_t(\alpha) \equiv F^{-1}(\alpha; \boldsymbol{\theta}_t, \boldsymbol{\xi}), \qquad ES_t(\alpha) \equiv \frac{1}{\alpha}\int_{-\infty}^{VaR_t} z dF(z, \boldsymbol{\theta}_t, \boldsymbol{\xi}),$$

where F^{-1} denotes the inverse of the continuous cumulative density function, $\boldsymbol{\theta}_t$ is a vector of time-varying parameters and $\boldsymbol{\xi}$ is a vector of additional static parameters. Thus, the VaR is simply the 100α % quantile of the return distribution at time *t* and the ES is the average of the 100α % worst cases.

In this paper we utilize the Generalized Autoregressive Score (GAS) framework for the time-varying parameter estimation. First, the GAS models are defined in Section 2. Second, the GAS estimates are compared to the well-known GARCH models on an empirical study of four stock market indices in Section 3. Third, in Section 4, the VaR and ES estimates are backtested using rolling windows and the impact of the length of the refit step is analyzed. Section 5 concludes.

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czech Republic, katerina.novakova1@seznam.cz
[2] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czech Republic, petra.tomanova@vse.cz

2 GAS Models

Generalized autoregressive score (GAS) models proposed by Creal et al. [4] belong to the class of observation-driven models that utilize a scaled score of the likelihood function as the driving mechanism. The possibility to let some parameters vary in time is necessary for capturing the dynamic behavior of time series. A huge benefit of GAS models is their ability to take advantage of the complex density structure rather than only consider means and higher moments. Moreover, the likelihood evaluation is straightforward.

In accordance with a notation of [3], let $\boldsymbol{y}_t$ be an N-dimensional random vector of the dependent variables at time t and $\boldsymbol{\theta}_t$ be a vector of time-varying parameters. Then $\boldsymbol{y}_t$ follows conditional observation density $p(\cdot)$

$$\boldsymbol{y}_t | \boldsymbol{y}_{1:t-1} \sim p(\boldsymbol{y}_t; \boldsymbol{\theta}_t)$$

for $t = 1, \ldots, T$, where $\boldsymbol{y}_{1:t-1}$ is a matrix which contains the past values of $\boldsymbol{y}_t$ up to time $t-1$. The vector of time-varying parameters $\boldsymbol{\theta}_t$ depends on $\boldsymbol{y}_{1:t-1}$ and a set of additional static parameters $\boldsymbol{\xi}$, $\boldsymbol{\theta}_t \equiv \boldsymbol{\theta}(\boldsymbol{y}_{1:t-1}, \boldsymbol{\xi})$.

The GAS updating mechanism for the time-varying parameter $\boldsymbol{\theta}_t$ is

$$\boldsymbol{\theta}_{t+1} \equiv \boldsymbol{\kappa} + \boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{B}\boldsymbol{\theta}_t,$$

where $\boldsymbol{\kappa}$ is a vector of constants measuring the level of the process, $\boldsymbol{B}$ is a diagonal matrix of autoregressive coefficients controlling for the persistence of the process and $\boldsymbol{A}$ is a diagonal matrix of parameters indicating the step of the update. $\boldsymbol{\kappa}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ are collected in the set $\boldsymbol{\xi}$. $\boldsymbol{s}_t$ is the scaled score, which depends on the past observations and the time-varying parameters

$$\boldsymbol{s}_t \equiv \boldsymbol{S}_t(\boldsymbol{\theta}_t)\boldsymbol{\nabla}_t(\boldsymbol{y}_t, \boldsymbol{\theta}_t),$$

where $\boldsymbol{S}_t$ is the scaling function and $\boldsymbol{\nabla}_t$ is the score

$$\boldsymbol{\nabla}_t(\boldsymbol{y}_t, \boldsymbol{\theta}_t) \equiv \frac{\partial \log p(\boldsymbol{y}_t; \boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t}, \qquad \boldsymbol{S}_t(\boldsymbol{\theta}_t) \equiv \mathcal{I}_t(\boldsymbol{\theta}_t)^{-\gamma}.$$

Creal et al. [4] suggest to set the scaling matrix to the γ-th power of the Fisher information matrix

$$\mathcal{I}_t(\boldsymbol{\theta}_t) \equiv \mathsf{E}_{t-1}\left[\boldsymbol{\nabla}_t(\boldsymbol{y}_t, \boldsymbol{\theta}_t)\boldsymbol{\nabla}_t(\boldsymbol{y}_t, \boldsymbol{\theta}_t)'\right].$$

The vector of static parameters $\boldsymbol{\xi} \equiv (\boldsymbol{\kappa}, \boldsymbol{A}, \boldsymbol{B})$ can be estimated by maximizing the log-likelihood function.

3 Empirical Study

Four major world stock market indices are analyzed in the empirical study. The first two indices, DJIA and S&P 500, are related to the U.S. stock market. The FTSE 100 assesses the market in Great Britain and TOPIX covers the Japanese market. Two time periods are analyzed: (i) January 3, 2000 – December 31, 2010, which covers 2,767 days, (ii) January 4, 2010 – March 15, 2019, which covers 2,315 days. The first period contains the global financial crisis and the second represents recent years. Each of the chosen periods evinces diverse shapes of the return distribution which allows to overview of each model and distribution reaction. The dataset is downloaded from Thomson Reuters Datastream.

3.1 Comparison of GAS Models

GAS models can utilize a wide range of possible conditional distributions. However, since the price returns are often fat-tailed and possibly skewed, the most common distributions are Student's t and skew-Student's t. This property is often demonstrated by comparing the results with the Gaussian distribution which is symmetric and very sensitive to extreme values and changes in return variance. The scale parameter is treated as time-varying which follows the properties of price returns. The skewness and kurtosis parameters of the related distribution are tested whether the parameters vary over time. Dynamics for both of them are not statistically significant across various time periods, thus skewness and kurtosis are treated as constant and the scale is the only dynamic parameter. Since different scaling functions have no significant effect on results, it is set to identity.

GAS models are estimated for each price index over each time period and evaluated based on the Akaike information criterion (AIC). The results for both periods are shown in Table 1. The values of AIC are the

	2000–2010				2010–2019			
AIC	**DJIA**	**S&P 500**	**FTSE 100**	**TOPIX**	**DJIA**	**S&P 500**	**FTSE 100**	**TOPIX**
Student's t	7.966	8.275	8.177	8.962	5.267	5.457	5.819	6.825
Skew-Student's t	7.968	8.276	8.177	8.966	5.269	5.459	5.820	6.826
Gaussian	8.073	8.354	8.199	9.005	5.418	5.629	5.885	6.948

Table 1 AIC values for GAS models in 2000–2010 and 2010–2019

lowest for the model utilizing the conditional Student's t distribution, however, the differences between its skewed version are negligible. Models based on normal distribution perform a lot worse as expected.

Figure 1 compares individually estimated ES series for GAS models with Student's t and Gaussian distributions for two indices S&P 500 and TOPIX in the periods of 2000–2010 and 2010–2019. The green and blue lines correspond to the Gaussian and the Student's t distribution respectively. Both periods are characterized by a different behavior. While the first one (2000–2010) exhibits higher returns fluctuations due to the financial crisis, the second one (2010–2019) is more tranquil with occasional jumps. These properties result in different estimates of risk measures as well as the sensitivity of the index itself.

The S&P 500 fluctuates less and therefore does not exhibit too many sudden drops. On the other hand, TOPIX is more sensitive and occasional jumps result in significant drops in estimated risk measures. In the period of 2010–2019, the drops are more severe, e.g. the estimated ES of TOPIX drops to -20.158 in 2011. It confirms that the Gaussian distribution results in undue sensitivity to extreme values since it does not treat the return heavy-tails properly.

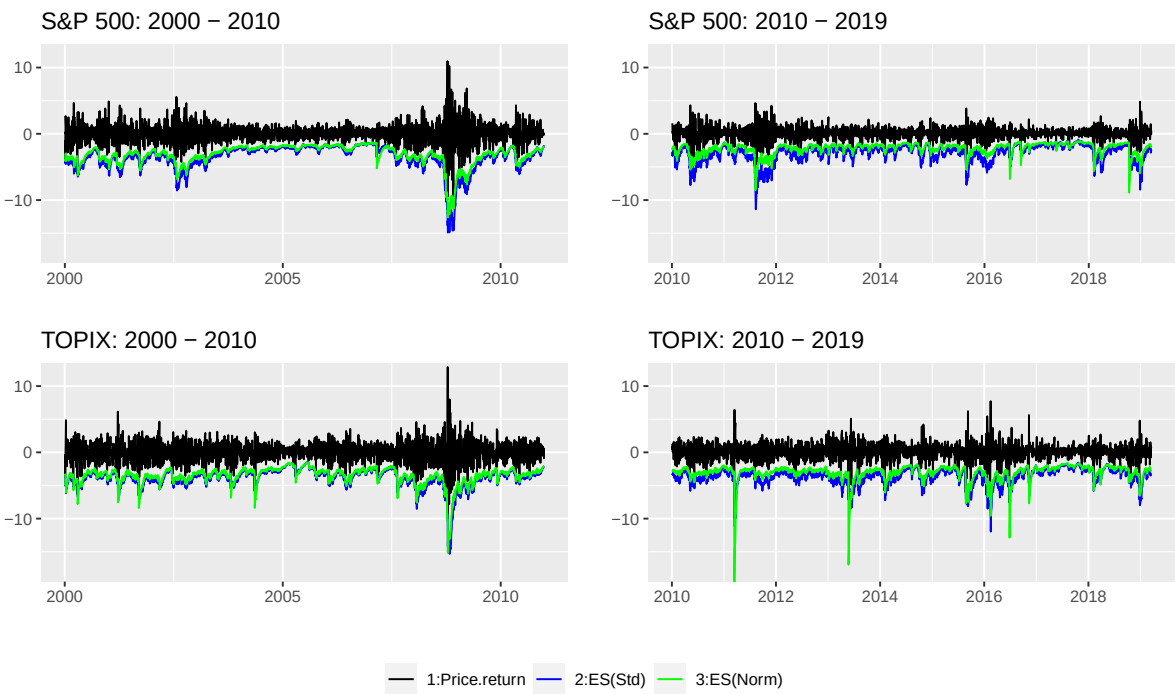


Figure 1 The estimated ES for S&P 500 and TOPIX based on Student's t and Gaussian distributions

3.2 Comparison of GAS and GARCH Models

GARCH models are so far one of the most often applied volatility models. Both GARCH and GAS models belong to the class of observation-driven models. Moreover, the original GARCH model is a special case of the GAS model, specifically, they coincide when the normal distribution and the inverse of Fisher information are utilized. However, it does not apply to the t distribution since t-GARCH and t-GAS updating mechanisms differ. Generally, GAS models have the advantage of using the complex density structure rather than only means and higher moments.

We compare t-GARCH and t-GAS on an example of TOPIX and S&P 500 in the periods of 2000–2010 and 2010–2019. Results in Table 2 show that while the t-GARCH performs better for S&P 500 in terms of the AIC, it is the opposite case for the more volatile TOPIX where the t-GAS is superior. The estimated VaR alongside the estimated volatility is plotted in Figure 2 where the blue and purple lines represent the estimated volatility and VaR from t-GARCH model respectively, the red and green lines refer to the estimated volatility and VaR from t-GAS model respectively, and the returns are black colored. For the TOPIX, the differences in estimated VaR are rather small. However, VaRs for S&P 500 exhibit noticeable differences. As expected, the score of the t distribution in GAS dynamics avoids the volatility to react too extensively to large values of returns. The idea is that such large values might be caused by the fat-tailed nature of the data and thus, should not be fully attributed to increases in the variance.

	2000–2010		**2010–2019**	
AIC	**t-GARCH**	**t-GAS**	**t-GARCH**	**t-GAS**
S&P 500	8.266	8.275	5.440	5.457
TOPIX	8.963	8.962	6.833	6.825

Table 2 The AIC values for t-GARCH and t-GAS

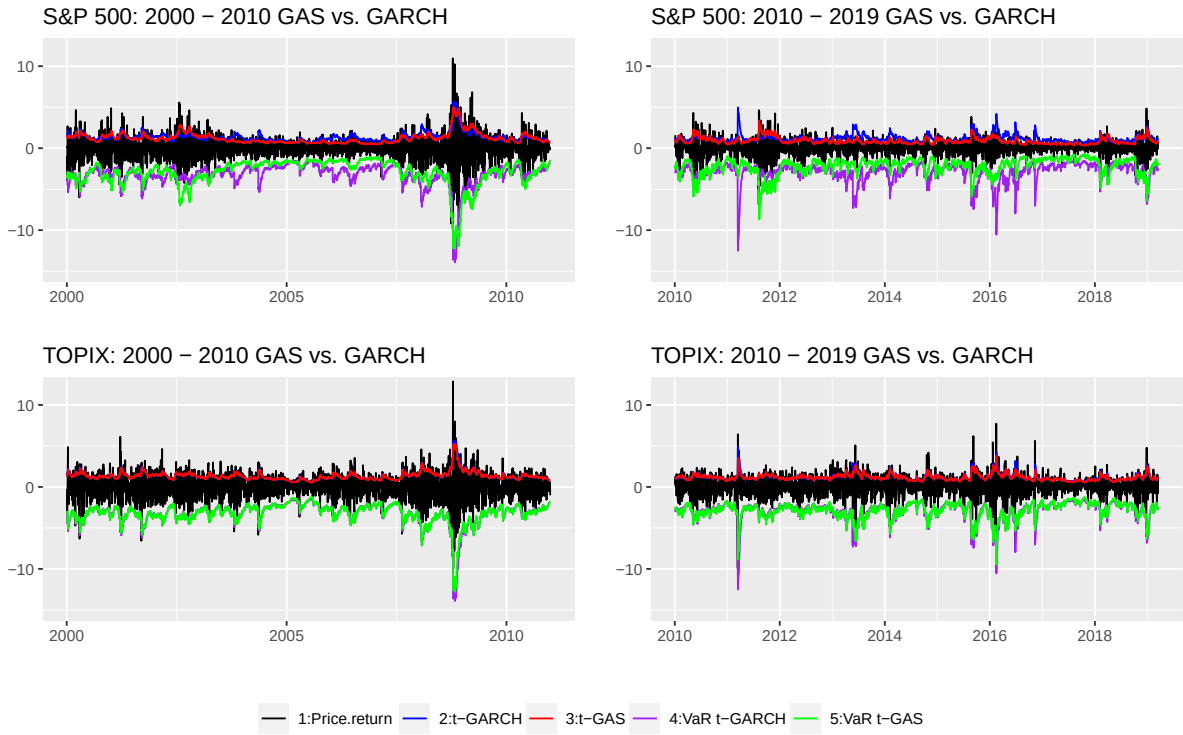


Figure 2 The estimated 1% VaR for S&P 500 and TOPIX based on t-GAS and t-GARCH models

4 Backtesting VaR and ES for GAS Models

Backtesting verifies the precision of VaR and ES predictions. The sample of length T is divided into two parts: in-sample of the length m and out-of-sample of the length $T-m$, and the approach of rolling windows is utilized. We analyze different refit steps for rolling windows and their impact on predictions. The output of the rolling windows are predicted values of the length $T-m$ and they are used to calculate VaR and ES. Then, the models are assessed by the dynamic quantile (DQ) test and the Fissler and Ziegel loss (FZL) function for the joint VaR and ES evaluation.

4.1 Sensitivity of the Refit Step for Rolling Windows

There are two parameters that in rolling windows that are required to be set: the forecast length $T - m$ and the length of the refit step. The forecast length is set to 1000 which represents approximately one-third of the whole sample. The length of the refit step can vary from 1 to T and there is no rule how to set it. The natural choice seems to be the length of 1 and 5 for the daily data [2] and the length of 4 for quarterly data [3]. Therefore, the sensitivity of the refit step is analyzed for the daily returns by comparing the estimated VaR and ES based on GAS models. Considered lengths are 1, 5, and 30.

For the GAS model utilizing the Student's t distribution, the lengths of 1 and 5 result in almost identical predictions, i.e. the estimated VaR and ES basically copy each other. The differences between the lengths of 1 and 30 are negligible as well, however, slight departures can be observed during the financial crisis in the end of 2008. On the other hand, the VaR predictions from models with Gaussian distribution differ substantially from their fitted values. The Gaussian predictions exhibit high sensitivity to the sudden changes in the index prices causing the estimated VaR to drop rapidly. This holds for shocks more than for slight changes but the effect is present for both. Moreover, the predictions based on the longer refit step are even more sensitive.

4.2 Quantile Loss Function and Quantile Dynamic Test

Quantile losses are averaged over the forecasting periods and the preferred model is the one with the lowest average value. The quantile loss (QL) function for time t at risk level α

$$QL_t(\alpha) \equiv (\alpha - d_t)(y_t - VaR_t(\alpha)), \qquad d_t \equiv I\left\{y_t < VaR_t(\alpha)\right\},$$

where $I\{\cdot\}$ is an indicator function. Series $d_t, t = 1, \ldots, T$, is called the hitting series and if the model is correctly conditionally covered, d_t should be independently distributed. This is tested by the dynamic quantile (DQ) test which is based on the joint hypothesis that (i) the hitting series are independently distributed, and (ii) the expected proportion of exceedance is equal to the risk level. The null hypothesis of the DQ test can be interpreted as the correct unconditional and conditional coverage and not rejecting the null hypothesis is desired.

The averaged QL functions are calculated for 1% and 5% VaR and the results show that the lowest values belong to the GAS models with the Student's t or skew-Student's t distribution. Their differences are negligible. On the other hand, the gap between these and the Gaussian model is more profound.

The results show that the QL does not change with the refit step, i.e. if the model is the best-performing one using the refit step of 1 then it also performs best when refit step of 5 or 30 is utilized. However, it does not apply for various confidence levels α, e.g. for FTSE 100 in 2000–2010 period, the model with the Gaussian distribution exhibits a better fit for 5% VaR while the skew-Student's t distribution fits the 1% VaR better than the Gaussian one.

Based on the DQ test and the 5% significance level, we cannot reject the null hypothesis of the correct specification for 5% VaR – this applies for all GAS models and indices in each period with the refit step of 1. However, the results for 1% VaR vary. Generally, the DQ test rejects the correct GAS model specification for models with the Gaussian distribution rather than for models with the Student's t distribution or skew-Student's t distribution. Moreover, for a given refit step in a given time period, the null hypothesis tends to be rejected either for all considered distributions or none of them which is a common and well-documented issue of the DQ test.

4.3 FZ loss function

Despite the QL function for the VaR, there is no such loss function for the ES since it is not an elicitable risk measure. However, the VaR and ES are jointly elicitable using test Fissler and Ziegel loss (FZL) function. Let's assume that VaR and ES are strictly negative and the generated loss differences are homogeneous of degree zero. Then the associated joint loss function FZL for time t at risk level α is formulated as

$$FZL_t^\alpha \equiv \frac{1}{\alpha ES_t^\alpha} d_t(y_t - VaR_t^\alpha) + \frac{VaR^\alpha}{ES_t^\alpha} + \log(-ES_t^\alpha) - 1$$

for the case when $ES_t^\alpha \leq VaR_t^\alpha < 0$. FZL functions are also averaged over the forecasting period and the preferred models are those with lower average values.

FZL values are calculated for all considered indices, periods, refit steps, distributions, and for both 1% and 5% risk levels. The results almost copy the QL function results. Generally, the GAS models with Student's t

or skew-Student's *t* distribution have the lowest FZL values and the differences between them are negligible. The gap is more profound between the Gaussian and Student's *t* or the Gaussian and the skew-Student's *t*. Furthermore, the gaps between Gaussian and *t* distributions are much higher for 1% VaR than 5% VaR as expected. Overall, the choice of the refit step does not influence the prediction performance of models in terms of the average FZL value. There are only a few exceptions – the differences are negligible and occur usually between the Student's *t* and skew-Student's *t* distribution.

5 Conclusion

In this paper, two risk measures – the Value at Risk and Expected Shortfall – are modeled and predicted by GAS models for four major world stock market indices. The considered GAS models are assessed from different perspectives. First, our results show that GAS models with conditional Student's *t* and skew-Student's *t* distributions perform similarly. Differences in terms of AIC values are negligible and the models are considered to be equally good in all assessed cases for all four indices. The estimated volatility and values of both risk measures differ very slightly. On the contrary, the GAS model utilizing the Gaussian distribution performs worse and leads to extreme values of the estimated volatility and consequently extreme values of the VAR and ES in days of price shocks. Additionally, the t-GAS model is compared to the well-known t-GARCH model. The results are ambiguous in terms of AIC values since examined periods result in no dominance of any model. However, the estimated volatility from t-GARCH reacts more extensively to large values of returns than volatility from the t-GAS model. This is caused by the score of the *t* distribution in GAS dynamics since such large values are considered to be caused by the fat-tailed nature of the data and thus, they are not be fully attributed to increases in the variance. Hence, GAS models might be of a better choice.

The choice of the refit step for rolling windows has a negligible impact on predictions based on GAS models assuming the Student's *t* or skew-Student's *t* distributions. Consequently, the estimated values of VaR and ES are almost indifferent. The choice of the refit step has a higher impact on predictions from the GAS model assuming the Gaussian distribution where the predicted values of VaR and ES tend to be significantly lower than their fitted values in days with price shocks. There is even a noticeable difference among predicted values considering different refit steps.

GAS models with the Student's *t* or skew-Student's *t* distribution outperform the GAS models with the Gaussian distribution in terms of the prediction power. Three different performance measures for predictions are considered: dynamic quantile test, quantile loss function, and Fissler and Ziegel loss function. Our results show that the model ranking is not sensitive to the choice of the refit step, i.e. if the model is the best-performing one when the refit step of 1 is used then it also performs the best when the refit step of 5 or 30 is used. On the other hand, it does not apply for various VaR levels ($\alpha = \{1\%, 5\%\}$), i.e. if the model is the best-performing one when $\alpha = 1\%$ is used, it might not perform the best for $\alpha = 5\%$. However, generally, the differences are negligible between models utilizing the Student's *t* and skew-Student's *t* distributions which both perform better than the Gaussian one.

Acknowledgements

This work was supported by the Internal Grant Agency of University of Economics, Prague under Grant F4/53/2019.

References

[1] Acerbi, C. & Tasche, D. (2002). Expected shortfall: a natural coherent alternative to Value at Risk. *Economic notes*, 31, 379–388.

[2] Ardia, D. & Boudt, K. & Catania, L. (2018). Downside risk evaluation with the R package GAS. *R Journal*, 10, 410–421.

[3] Ardia, D. & Boudt, K.& Catania, L. (2019). Generalized autoregressive score models in R: The GAS package. *Journal of Statistical Software*, 88, 1–28.

[4] Creal, D. & Koopman, S. J. & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5), 777–795.

[5] Delbaen, F. (2002). *Coherent risk measures on general probability spaces*. Springer: Advances in finance and stochastics.

[6] Rockafellar, R. T. & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of banking & finance, Elsevier*, 26, 1443–1471.

Financial Frictions in a Small Open Economy: The Case of Czech Republic

Tomáš Oravec[1], Osvald Vašíček[2]

Abstract. In the period during the Great Recession, it has been clear that financial factors played important role in economic downturn. Hence, abstracting from them in the business cycle modelling is currently no longer a valid simplification. In the submitted paper, we interpret structure and behaviour of a bayesian small open economy DSGE model with financial frictions. Originally, this is a Swedish New Keynesian model estimated and analysed on Czech data. Czech economy has been currently overheating and is subject to turbulent development in terms of key macroeconomic variables. The aim is to contribute in understanding financial frictions in the Czech economy and their impact on business cycle. We provide results of the economy in the period of recession and overheating using impulse responses framework and shock decomposition techniques. Where possible, financial frictions specification is compared to a baseline (frictionless) model.

Keywords: bayesian estimation, DSGE model, small open economy, financial frictions

JEL Classification: C53, E32, E37
AMS Classification: 91B50, 91B64

1 Introduction

The Czech economy has been undergoing a turbulent development. After declines during the 2008–2009 financial crisis and the 2012–2013 recession, ending with the Czech National Bank's unilateral exchange rate commitment regime, the economy has gradually recovered. Even in a connection with the development on the labour market, it can be argued that the Czech economy is currently overheating. While unemployment rate reached almost 8 % during the crisis, the labour market is currently extremely tense with unemployment rate slightly above the threshold of 2 %. In the coming years, we can expect a slowdown in the economy due to a decline in exports (mainly stemming from Brexit and constantly growing trade barriers) and a reduction in the inflow of funds from the European Union. In addition, the effects of these shocks on the economy may be strengthened by financial frictions, which are particularly the most evident in the real estate market.

The main research question is therefore to use dynamic stochastic general equilibrium (DSGE) models with frictions on the financial market to examine the development of the Czech economy in the context of actual overheating and the expected slowdown in growth. To achieve that objective, we use the structure of a New Keynesian small open economy model with financial frictions proposed by Christiano *et.al.* [5] (henceforth CTW) and estimate it on the Czech economy data. We continue in work done by Ryšánek *et.al.* [8] with the comparison based solely on model with and without financial frictions. In general, there is not yet known and widely-accepted approach of how to include financial frictions into DSGE model framework as Duncan and Nolan [7] point out in their discussion paper. CTW model concept allows us to involve simultaneously labour market frictions and assess their particular impact on model's performance. Nevertheless, modelling frictions on labour market is behind the scope of this particular contribution.

The rest of the paper is structured as follows. Section 2 sketches baseline model structure. For the reason that the model is really complex and has complicated structure, we aim to provide economical insight of causal relationships rather than employ any technical details. Section 3 expands original (frictionless) model into model embracing frictions on the financial market. Similarly, we abstract from exact form of how the financial frictions are implemented, and hence only transmission channels are intuitively described. Empirical results of analysed model are the content of Section 4. Finally, Section 5 concludes.

[1] Masaryk University, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, tomas.oravec@mail.muni.cz
[2] Masaryk University, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, osvald@econ.muni.cz

2 Baseline Model

This section introduces basic features of the Czech New Keynesian small open economy model. The baseline version (without introducing financial frictions) follows standard New Keynesian model microeconomic foundations. Basic building blocks rely on models analysed in e.g. Adlfson *et.al.* [1], Christiano *et.al.* [4] or Smets and Wouters [9]. The microeconomic background in the aforementioned models is deemed to be standard and has been widely accepted by most modern New Keynesian DSGE models. CTW model was developed to study Swedish economy, nevertheless, it has been shown that the model is generally suitable also for studying other small open economies.

In a domestic economy, a homogeneous domestic good is produced from intermediate goods by using Dixit-Stiglitz [6] production function with constant elasticity of substitution. Domestic producers are competitive representative firms taking price of inputs and price of output as given.

Each intermediate good producer has its production function, where he rents and reallocates capital and labour services using certain technology. Specific form of this production function provides two sources of economic growth. First, positive drift in a technology shock causing better reallocation of labour services in the production and second, positive drift in an investment-specific technology shock triggering downgrade of fixed production costs. At the same time, each firm is a monopolist in producing its good. Hence, each firm has its own price setting which is subject to Calvo frictions.

Final consumption goods are produced by a representative competitive firm using linear homogeneous technology and are purchased exclusively by households. In general, when producing a final consumption good, two inputs are used. First, homogeneous domestic good is transferred as one-for-one, and second, homogeneous composite of specialized consumption import goods serves as alternative input. Profit maximization leads to optimal schedules for both types of goods given their price.

In the context of this model, investment is defined as the sum of investment goods used in accumulation of physical capital and in capital maintenance. Similarly to consumption good sector, representative investment goods producers take all prices as given. Yet, the split to homogeneous investment good and homogeneous composite of specialized investment import goods is in place.

Import and export activities involve Calvo pricing frictions and Dixit-Stiglitz type of introducing a range of specialized goods. Exports are represented by a continuum of exporters. Each exporter is a monopolist producing specialized export good. Basically, each export good is produced using a homogeneous domestically manufactured good and a homogeneous good derived from import. Created specialized export goods are then sold to competitive foreign retailers creating a homogeneous foreign good subsequently sold to foreign citizens. On the other hand, specialized domestic importers turn a purchased homogeneous foreign good into specialized input sold to domestic retailers. Domestic retailers use specialized import good to create a homogeneous good used as an input into production of either specialized exports, investment, or consumption goods. These imported goods are combined with domestic inputs prior to creating a final good. There are pricing frictions present at each stage of the production.

Household preferences are given by utility of consumption and disutility of labour. Specialized labour is supplied by households and is combined with labour contractors into a homogeneous labour service. Again, households are subject to Calvo frictions and in choosing optimal wage rate, they maximize their discounted expected utilities.

Domestic economy's saving is done by the households. Saving in period t occurs by acquisition of the net foreign assets A^*_{t+1} and a domestic asset. The domestic asset is used for financing firms' working capital requirements. This asset provides a nominally non-state contingent return R_t from period t to $t+1$. The model assumes the same tax treatment of domestic agents' earnings on foreign bonds as agents' earnings on domestic bonds. The first order condition connected with the asset A^*_{t+1} paying R^*_t in terms of foreign currency is

$$v_t S_t = \beta \mathbb{E}_t v_{t+1} \left[S_{t+1} R^*_t \Phi_t - \tau^b (S_{t+1} R^*_t \Phi_t - \frac{S_t}{P_t} P_{t+1}) \right] \tag{1}$$

Here S_t denotes domestic currency price of a unit of foreign currency (CZK/EUR). Left side of the equation (1) represents the cost of acquiring a unit of foreign asset – the currency cost is S_t and the conversion into utility terms is ensured by multiplying it by the Lagrange multiplier on the household's budget constraint v_t. The term in the square brackets represents the after-tax payoff of the foreign asset converted into domestic currency units. The pre-tax interest payoff on A^*_{t+1} at period $t+1$ is $S_{t+1} R^*_t \Phi_t$. In this expression, R^*_t denotes foreign nominal interest rate that is actually risk-free in foreign currency units. The term Φ_t represents a

relative risk adjustment of the foreign asset return, so that a unit of foreign asset acquired in time t pays off $R_t^* \phi_t$ unit of foreign currency in period $t+1$. The remaining term in brackets pertains to the impact of taxation on the return on foreign assets.

The risk adjustment term has following form

$$\Phi_t = \Phi(a_t, R_t^* - R_t, \tilde{\phi}_t) = \exp\left[-\tilde{\phi}_a(a_t - \overline{a}) - \tilde{\phi}_s(R_t^* - R_t - (R^* - R)) + \tilde{\phi}_t\right], \tag{2}$$

where

$$a_t = \frac{S_t A_{t+1}^*}{P_t z_t^+},$$

country risk premium shock is denoted by $\tilde{\phi}_t$, and $\tilde{\phi}_a > 0$ and $\tilde{\phi}_s > 0$ are parameters. Particular calibration of $\tilde{\phi}_s = 0$ in case of Buss [3] model for Latvia ensures nominal interest rate peg regime. This calibration is not valid in the case of economy with the independent monetary policy (interest rate) regime. Dependence of ϕ_t on relative level of interest rates $R_t^* - R_t$ is designed to reproduce uncovered interest parity (UIP), although in a bit complicated manner than usual.

Foreign economy is modelled as a structural VAR(1) process built under an assumption that the source of growth in foreign economy is considered to be the same as in the domestic one.

3 Financial Frictions

Incorporation of financial frictions block is carried out by accumulation and management of capital. This approach to extend the baseline model by financial frictions is built on the financial accelerator mechanism similarly as in Bernake *et.al.* [2]. Connections between the real economy and financial markets provide framework for economic growth in the expansion and shrinkage in the recession. The background reflects that borrowers and lenders are different agents and there is asymmetry in the information available to these parties.

Financial frictions introduce new agents – entrepreneurs – owning and managing capital stock. Financing of the stock is performed by internal and borrowed funds. Households deposit money in banks for nominally non state-contingent interest rate. Banks act as non-risky intermediaries of debt contracts to entrepreneurs, hence are not modelled explicitly. The amount to lend under a debt contract is a function of entrepreneur's net worth. Entrepreneurs face shocks to their assets (return on capital) which influence their ability to borrow. Hence, negative shock translates into reduction in investment and eventually into economic slowdown. In this model setting, stock market value is a proxy for net worth.

At the end of each period, entrepreneur is required to pay gross interest rate on the bank loan, if possible. After each purchase of a capital, idiosyncratic productivity shock ω to newly acquired capital occurs. This shock is modelled as a unit-mean log-normally and independently distributed across entrepreneurs. However, there exists a threshold value of $\omega = \overline{\omega}$ such that the entrepreneur left with the only resources to cover the interest. Entrepreneurs for which $\omega < \overline{\omega}$ are bankrupt and have to turn over all their assets. Survivors return just the gross interest.

We assume competition and free entry across banks. Derivation of optimal contract (zero profit condition) implies that entrepreneur's loan is proportional to his net worth. One of the outcomes of this theoretical analysis is the fact, interest rate is the same across entrepreneurs regardless their net worth.

4 Empirical Results

As mentioned earlier, we have estimated two model specifications – one without financial frictions (baseline) and second with financial frictions – using Bayesian techniques. Hard-to-estimate parameters have been calibrated based on stylized facts about the Czech economy and have been reflected with experts from Czech National Bank. Priors have been based on mainly on CTW [5] and Buss [3]. As a data base has been used quarterly measured Czech economy data, same as in CTW [5] in a period of 2001Q1 – 2019Q4. Estimates are based on two independent Random Walk Metropolis Hastings chains, both consisting of 1 million random draws with a burn-in ratio of 2/3.

In what follows, impulse response functions (IRFs) are discussed. In order to quantify impact of financial frictions, we have plot the IRFs for the same fixed endogenous variable for different model specifications. Red

line represents the baseline model described in Section 2 while blue line stands for model including financial frictions carried out in Section 3. The results presented in this contribution do not take into consideration effects of labour market frictions on chosen variables, discussed in CTW [5] extensively. Horizontal axis represents time in terms of years under an assumption that certain shock occurs in the inception. On the contrary, units on vertical axis are either percentage deviations from steady state (% dev.), annualized basis points (ABP), or level deviation (Lev. dev.). Worth noticing is the fact that financial frictions model specification allows us to inspect responses of other endogenous variables not present in the baseline version of the model.

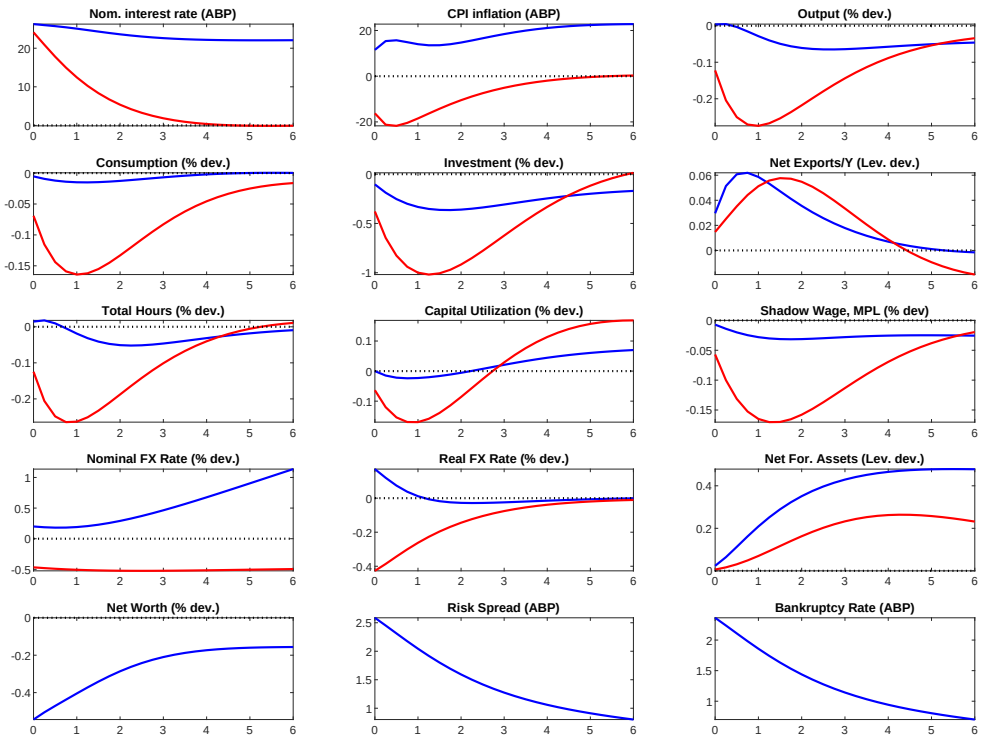


Figure 1 Impulse Response Functions to Monetary Policy Shock (red line represents baseline model, blue one stands for financial frictions specification)

Figure 1 depicts impulse responses of key macroeconomic variables to monetary policy shock. A temporary increase of the nominal interest rate causes hump-shaped response of CPI inflation, however the reduction or increase depends on the model specification. For the baseline model we experience cutdown effect, similarly as for the Swedish economy, but financial friction specification show some discrepancies. The reduction of consumption, investment and output stemming from domestic nominal interest rate accrual is as expected. Entrepreneurial net worth decrease effect can be explained by falling price of capital. On account of default (bankruptcy) risk acceleration, interest rate risk spread increases by 2.5 basis points. This risk-compensation effect is evident in the development of investment and output over time – investment strengthen its aggregated force to output movement. Depreciation of domestic currency then creates pressure on exports segment of the economy, hence together with contracted domestic demand produces gain in net exports. This effect is more pronounced with the financial frictions.

Impulse response functions to stationary neutral technology shock are shown on Figure 2. The responses of the estimated model comply with theoretical economic expectations. The most pronounced and sudden impact experiences CPI inflation with progressive recovery afterwards. Reduction of nominal interest rate persistently supports consumption, investment and finally translates also into GDP. Labour supply and capital utilization substantially fall at the moment when shock occurs below their steady state level, however during relative short time span they tend to rebound. Real exchange rate fluctuating response transfers into unclear long-term of net exports. Dampened investment response creates downward pressures for en-

trepreneurial net worth emerging higher default risk. Naturally, investors bearing this amplified risk are rewarded by higher risk spread.

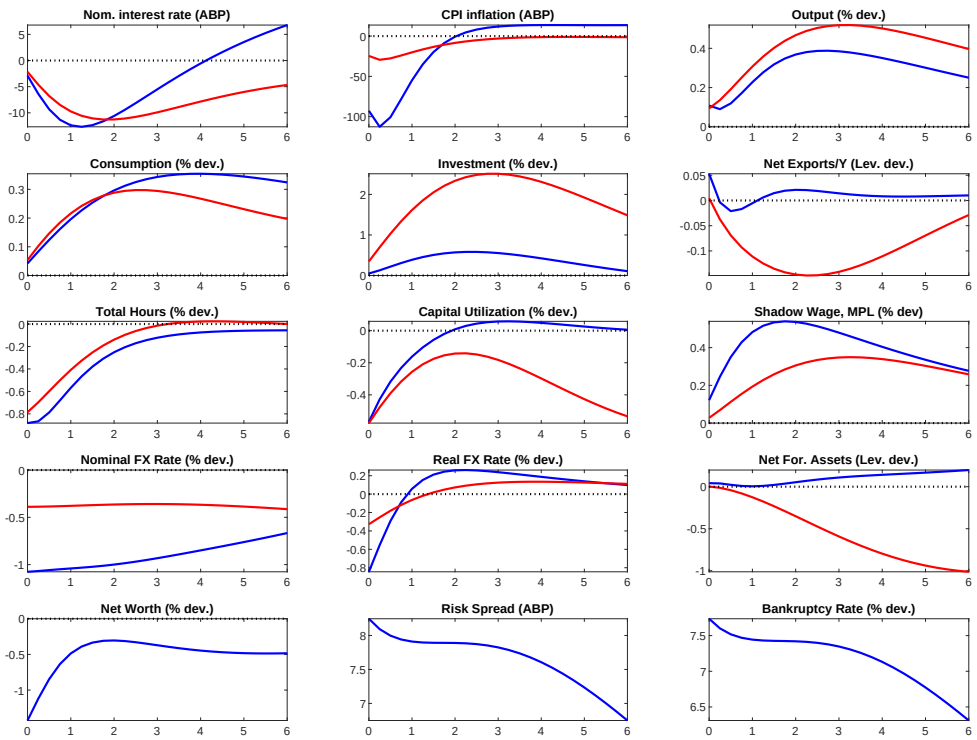


Figure 2 Impulse Response Functions to Stationary Neutral Technology Shock (red line represents baseline model, blue one stands for financial frictions specification)

Using structural historical shock decomposition technique to the most important figure of economic growth is presented on Figure 3. Empirical quarter-over-quarter decomposition of domestic GDP growth is decomposed into contribution of financial frictions model shocks. Because the model is pretty complex, economical grouping of shock was performed prior to plotting the final figure. In this framework, it is apparent that financial frictions together with technological shocks play cumulatively crucial role in explaining variations in the GDP growth. From the figure it is pivotal that financial frictions are providing leading indicator prior to recessions in 2008–2009 and 2012–2013.

5 Conclusion

We have shown main model differences between standard DSGE modelling and modelling with financial frictions. Based on the provided results, we have definitely confirmed that abstracting from financial frictions in modern business cycle modelling is not a valid option any more. Nevertheless, there is still respectable uncertainty connected with the results and modelling itself. Experiencing current COVID-19 pandemic is definitely event that could not be captured by the aforementioned model properly and it is completely evident, that this will have strong negative impact on economies worldwide.

Acknowledgements

This work is supported by funding of specific research at Faculty of Economics and Administration, project MUNI/A/1191/2019. This support is gratefully acknowledged.

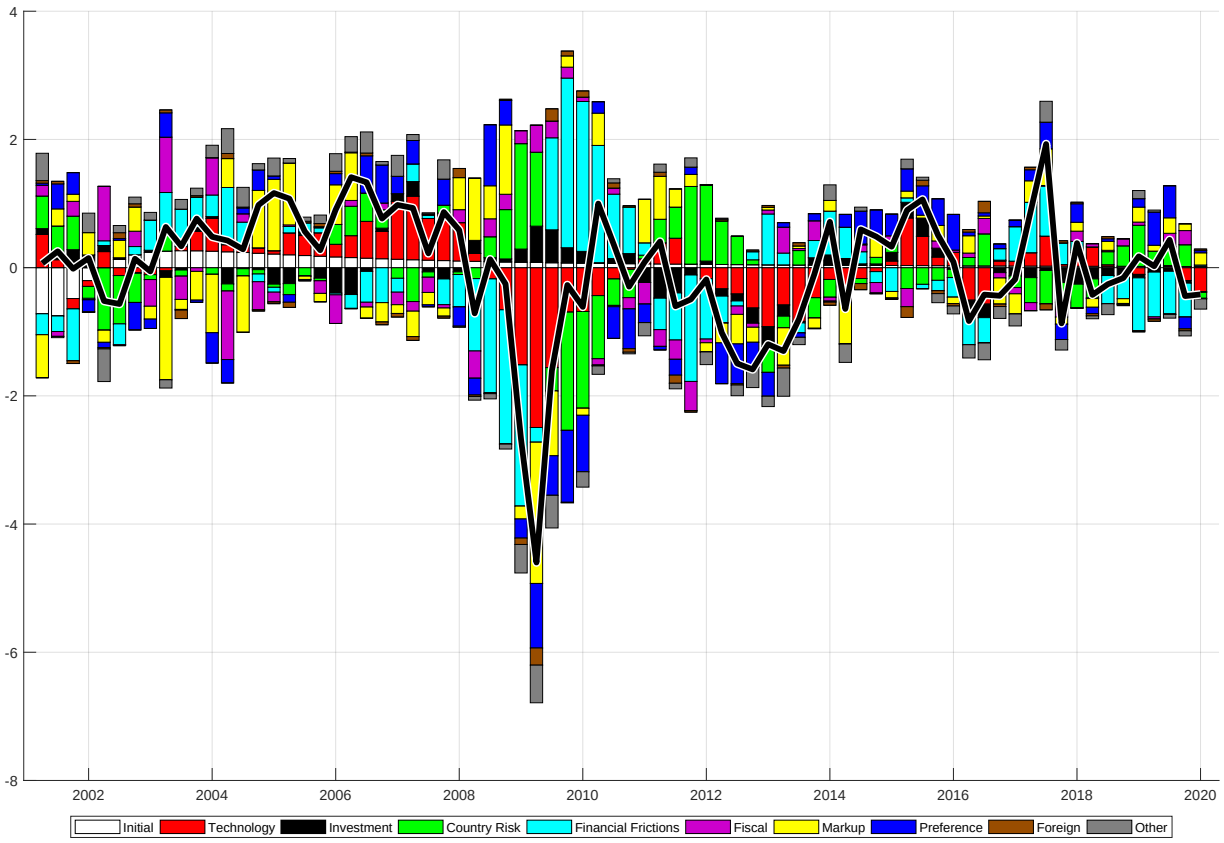


Figure 3 Historical Shock Decomposition of GDP Growth

References

[1] Adolfson, M., Laséen, S., Lindé, J. & Villani, M. (2008). Evaluating an Estimated New Keynesian Small Open Economy Model. *Journal of Economic Dynamics and Control*, vol. 32, issue 8, August, pp. 2690–2721.

[2] Bernake, B., Gertler, M. & Gilchrist, S. (1999). *The Financial Accelerator in a Quantitative Business Cycle Framework*. Handbook of Macroeconomics, vol. 1. Ed. by John B. Taylor and Michael Woodford. Elsevier Science, pp. 1341–1393.

[3] Bušs, G. (2014). *Financial Frictions in a DSGE model for Latvia*. Dynare Working Papers 42.

[4] Christiano, L. J., Eichenbaum, M. & Evans, C. L. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy*, vol. 113, No. 1, pp. 1–45.

[5] Christiano, L. J., Trabandt, M. & Walentin, K. (2011). *Introducing Financial Frictions and Unemployment into a Small Open Economy Model*. Riksbank Research Paper No. 214.

[6] Dixit, A., & Stiglitz, J. (1977): Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, Vol. 67, No. 3, pp. 297–308.

[7] Duncan, A. & Nolan, Ch. (2017). *Financial frictions in macroeconomic models*. School of Economics Discussion Papers, No. 1719, University of Kent

[8] Ryšánek, J., Tonner, J. & Vašíček, O. (2011). *Monetary Policy Implications of Financial Frictions in the Czech Republic*. Working Papers 2011/12, Czech National Bank.

[9] Smets, F. & Wouters R. (2003): An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association*, Vol. 1, pp. 1123–1175.

Capital Mobility in V4 Economies

Václava Pánková[1]

Abstract. Capital is one of the most important productive inputs. Under globalization, it is characterized by a certain mobility degree, which can be a significant factor that influences economic growth. In general, a more open capital account implies a higher productive performance than economies with restricted capital mobility. However, this relation is not straightforward, and there is empirical evidence that for weaker economies a high degree of capital mobility is undesirable. To measure the degree of capital mobility, the Feldstein–Horioka hypothesis is usually applied: Under perfect capital mobility, domestic savings and investment rates should be uncorrelated. In econometric models, a possible long-run relation between savings and investment is studied by the help of the concept of cointegration. Recently, the importance of including foreign direct investment into the cointegration equation has been documented.

The relevant econometric analysis of capital mobility in V4 economies is presented treating the countries as a panel. The four countries sharing a similar economic history have a high demand for investment from abroad but the domestic capital only rarely is exported. This might be why in aggregation the results tend to be insignificant. In the long-run, a low capital mobility is stated.

Keywords: capital mobility, econometric models, unit root, cointegration, panel data

JEL Classification: E20, C22, C51
AMS Classification: 62J02

1 Introduction

Capital is one of the most important productive inputs. Under globalization, it is characterized by a certain mobility degree, which can be a significant factor that influences economic growth. On the other hand, inflows that are large relative to the system may then cause sudden fluctuations in important economic indicators. In particular, developing economies are constrained by limited savings but they have strong demand for investment. Contemporaneously, they tend to have weaker regulatory and institutional frameworks. Even in the Eurozone, there is an apparent imbalance in foreign direct investment (FDI) flows, by which small countries are disadvantaged for different economic reasons. Moreover, rapid capital movements across national borders substantially amplify the impact on the domestic unemployment rate in large economies. Capital mobility, seen in a context of the macroeconomic trilemma, is a factor that is usually not in the focus of government policy. The concept of reindustrialisation emerged recently in the EU and became a topic of treatment under European official structures, which calls today's economic orthodoxy of international capital mobility into question.

The above reasons emphasize the importance of relevant measurements. To measure the degree of capital mobility, the Feldstein–Horioka hypothesis [7] is usually applied: Under perfect capital mobility, domestic savings and investment rates should be uncorrelated. In econometric models, a possible long-run relation between savings and investment is studied by the help of the concept of cointegration. Not cointegrated variables report high capital mobility and vice versa. Recently, the importance of including FDI into the equation has been documented. After subtracting FDI inflows from investment, the estimated coefficient of the savings rate reflects more precisely the extent to which domestic savings are used to finance domestic investment.

Econometric models based on the Feldstein–Horioka approach are applied to four central European economies, the Czech Republic, Hungary, Poland and Slovakia, together grouped as Visegrád (V4). All these countries profit from foreign investment, but they also share some less favourable features of capital mobility.

[1] University of Economics, Department of Econometrics, Winston Churchill sq. 4, 130 00 Praha, pankova@vse.cz

Capital mobility is defined as the ability of private funds to move across national boundaries in the pursuit of the efficient allocation of resources and higher returns. Capital mobility increases national savings and investment with an impact on capital accumulation and economic growth. This mobility depends on the absence of currency restrictions on the inflows and outflows of capital. In the historical survey opening his text, Taylor [17] reminds readers of the fact that late-nineteenth century capital markets were relatively well integrated under the classical gold standard. Disintegration and imperfect capital mobility relate to the interwar period, especially after 1929; the postwar period showed gradually increasing capital market integration. There is strong evidence (e.g. [9]) in favour of a higher degree of capital mobility for countries in the euro currency and in the eurozone. If capital is mobile, then it is easier to attract FDI into a country. Efficiency improved by technology transfer, global production chains and trade and financial development are a common part of the process. It will also increase investment opportunities abroad. As stated in [5], economies with a more open capital account show a higher productive performance than economies with restricted capital mobility. However, a positive influence of an open capital account emerges only after a country has achieved a certain degree of economic development. Common experience shows that capital mobility is appreciated in different ways according to the economic development of a country. Developing countries tend to be acceptors of capital and the conditions are supported by their governments that attract investment. Such economies are constrained by limited savings but they have strong demand for investment.

In [14], similar financial integration with the world market is documented for countries that share similar characteristics, such as geographical region and/or level of per capita income. On the other hand, Rademacher [13] shows that even in the Eurozone, there is an apparent imbalance in FDI flows, by which small countries are disadvantaged for different economic reasons.

In some literature (e.g. [11]), international capital mobility is described as today's economic orthodoxy. The contemporary view of capital controls is that they are relevant only for developing countries where a sudden outflow of capital in times of uncertainty may occur. Nevertheless, even the International Monetary Fund (see [11]) has been unable to find clear empirical support for the claim that capital mobility increases growth.

Capital mobility cannot be seen without a context of the macroeconomic trilemma (see [9]). Having typically desirable, yet contradictory, objectives:

1. To stabilize the exchange rate;
2. To enjoy free international capital mobility;
3. To engage in a monetary policy oriented toward domestic goals,

Only two out of the three objectives can be mutually consistent. That is why policymakers must decide which one to give up. Because policymakers usually insist on controlling the exchange rate and interest rate, they have to drop point 2.

In [11], different arguments against capital mobility are analysed. He deduces that a flexible exchange rate and capital openness can lead to large inflows that appreciate the exchange rate and cause a contraction of investment spending and net exports. That in turn reduces output and employment. In economies more sensitive to exchange rates, it also may change the share of exports and imports in GDP. Further problems are described at a microeconomic level and are specified as particularly acute in developing economies, which tend to have weaker regulatory and institutional frameworks. Inflows that are large relative to the system may then cause sudden fluctuations in important economic indicators. Palley [11] also mentions the hot money flows and short-term speculating investors who create negative externalities. His conclusions lead to a proposal to replace the above trilemma, also known as the impossible trinity, by the possible trinity of coordinated monetary policy, managed exchange rates and managed capital flows.

In the recent literature, attention has been paid to the broader economic context. Rapid capital movements across national borders, such as those experienced by developed nations in the past 20 years, substantially amplify the impact on the domestic unemployment rate. Capital flows increase the riskiness of labour income, as shown by Azariadis and Pissarides [2], who are not the only authors interested in the relation between capital mobility and the labour market. What are the wage effects of migration in the immigration country with international capital adjustment is the question of Ruist and Bigsten [15].

The argument is that FDI and outsourcing can be good for companies but may be bad for national income and wages. Hence, there are good economic reasons for restoring capital controls as a standard part of the policy arsenal. The concept of reindustrialisation emerged recently in the EU and became a topic of the treatment of European official structures not only from a technological but also from a social point of view.

In "Draft report on reindustrialising Europe" [3], the following two points are made:

- Europe's future industrial strength lies in a Renaissance of Industry for a Sustainable Europe (RISE) strategy that pursues technological, business and social innovation towards a third industrial revolution including a low-carbon modernisation offensive. RISE will create new markets, business models and creative entrepreneurs, new jobs and decent work, bringing industrial renewal with economic dynamism, confidence and competitiveness. Moreover, energy and resource efficiency are the key pillars of such a strategy
- RISE could repatriate manufacturing to the EU, paying attention to supply chain management and specific regional manufacturing cultures

V4 economies profit from foreign investment in their countries but they also share some less favourable features of capital mobility. That is why a measurement of the degree of capital mobility in this region may be of interest.

2 Measurement of Capital Mobility

To measure the degree of capital mobility, different approaches appear in the literature. In [5] and [16], two alternative measures of the extent of capital mobility are presented and explained, both constructed as indices during the 1990s.

An econometric treatment is usually based on studying the savings–investment correlation, which is known as the Feldstein–Horioka [7] hypothesis. The authors argue that if there is perfect capital mobility, we should observe low correlation between domestic investment share in output and domestic savings shares, which can be formalized by

$$\left(\frac{I}{GDP}\right)_t = \alpha + \beta\left(\frac{S}{GDP}\right)_t + u_t \tag{1}$$

with I for investment, S savings and GDP. Finding that parameter β is close to unity means that changes in domestic savings passed through almost fully into domestic investment, which implies imperfect international capital mobility. Under perfect capital mobility, domestic savings and investment rates should be uncorrelated. Under savings–investment dynamics, a simple correlation of both variables and the regression equation (1) represent a temporary short-run phenomenon. Long-run relations between time series can be seen through the theory of cointegration (e.g. [8]). Cointegration, in fact, is a long-run relation that can be represented by an error-correction model with long-run and a short-run parts; the validation of a long-run relation between investment and savings thus means low capital mobility, and vice versa.

Therefore, instead of relation (1), a formulation is proposed in [10] as

$$\Delta\left(\frac{I}{GDP}\right)_t = \alpha + \beta\Delta\left(\frac{S}{GDP}\right)_t + \gamma ecm_{t-1} + \delta\Delta\left(\frac{S}{GDP}\right)_{t-1} + u_t \tag{2}$$

with $ecm_t = \left(\frac{I}{GDP}\right)_t - \vartheta_0 - \vartheta_1\left(\frac{S}{GDP}\right)_t$. In the case of an open economy, low values of β and γ are expected. On the other hand, capital immobility should be reflected in the results, as high values of β, γ, α and δ being near to zero.

Nevertheless, the weak points of the above approach are criticized (e.g. [18]). The author sees that one problem with the conventional way of gauging capital mobility based on the correlation between domestic savings and domestic investment lies with the inclusion of FDI in the latter. FDI is not financed by the savings of residents. After subtracting FDI inflows from investment, the estimated coefficient of the savings rate reflects more precisely the extent to which domestic savings are used to finance domestic investment.

His findings suggest that capital is remarkably more mobile in both developed as well as developing countries when FDI is excluded from the domestic investment of the recipient country. Instead of (1), it is more precisely formulated as

$$\left(\frac{I - FDI}{GDP}\right)_t = \alpha + \beta\left(\frac{S}{GDP}\right)_t + u_t \tag{3}$$

and a long-run counterpart to (2) is formulated on the basis of (4) as

$$\Delta\left(\frac{I - FDI}{GDP}\right)_t = \alpha + \beta\Delta\left(\frac{S}{GDP}\right)_t + \gamma ecm_{t-1} + \delta\Delta\left(\frac{S}{GDP}\right)_{t-1} + u_t \tag{4}$$

with $ecm_t = \left(\frac{I-FDI}{GDP}\right)_t - \vartheta_0 - \vartheta_1 \left(\frac{S}{GDP}\right)_t$. Here again, the lowest values of β and especially γ parameters speak in favour of the highest degree of capital mobility.

Annual data provided by Eurostat covering years 1996–2018 are used.

The correlation of national investment and savings is given in the following survey:

	96–18	96–10	10–18
Czech Republic = CZ	0.734	0.871	0.291
Hungary = HU	0.743	0.961	0.306
Poland = PL	0.944	0.975	0.223
Slovakia = SK	0.678	0.694	0.063

Decreasing correlations in the last decade give a rough information about a raising capital mobility.

The preliminary analysis was performed dealing the countries separately. Let us denote $Y^f = (I - FDI)/GDP$, $X = S/GDP$ and $Y = I/GDP$. According to [1], KPSS test was preferred to individual time series the length of which is 23. Null hypothesis is stationarity, the critical value (trend and intercept) is 0.1460 at the 5 % level.

country	**KPSS**		
	Y^f	Y	X
CR	0.1290	0.1195	0.0648
HU	0.0594	0.1242	0.1439
PL	0.0672	0.0584	0.1171
SK	0.1228	0.0687	0.1139

Table 1 Unit root test (*source: own computation*)

In all cases null is not rejected, the series are stationary and that is why an ECM model cannot be justified. As for the dependence of Y^f, respective Y on X, it is evidently none; all the Probabilities are greater than 0.05.

country	**Y^f**		**Y**	
	Coeff.	***Prob.***	**Coeff.**	***Prob.***
CR	0.6402	0.0787	0.3881	0.1223
HU	–0.0581	0.9743	0.0350	0.8800
PL	0.6507	0.0738	0.4604	0.0580
SK	–0.0751	0.9081	0.2961	0.4905

Table 2 Short – run information (*source: own computation*)

The results seem to speak in favour of a high capital mobility but the coefficients show only the marginal propensity, no long-run information is comprised.

The panel structure hopefully will provide us with more information.

When dealing with small sample data, as in our case, one has to know the tests offered by a software, as the alternatives usually do not have the same power. The power of a test is its probability of rejecting the null hypothesis when it is false and the null is the unit root. Individual unit root tests (Im, Pesaran and Shin) have limited power in comparison with a common unit root process (Levin, Lin and Chu). Unfortunately, the power of the Levin, Lin and Chu test strongly depends on the time dimension being large. Using both alternatives (intercept plus trend), Y has no unit root, Y^f and X are found to be $I(0)$.

		Y^f		DY^f		X		DX	
Method	**Null hypothesis**	**Statistic**	***Prob***	**Statistic**	***Prob***	**Statistic**	***Prob***	**Statistic**	***Prob***
Levin, Lin and Chu t*	Unit root	−0.709	0.239	−2.053	0.020	−0.814	0.207	−2.112	0.017
Im, Pesaran and Shin W-stat	Unit root	1.289	0.099	−3.733	0.000	−0.822	0.205	−2.907	0.002

Table 3 Panel unit root tests (*source: own computation*)

Model (4) is estimated using Y^f, X and alternatively Y, X. The letter alternative is questionable; the stationarity of Y does not correspond with the assumptions of the Representation theorem [6]. The situations in which Representation theorem is used though the assumptions are not completely fulfilled often occur, unfortunately not with a proof of their correctness. Applying the methodologies of fixed effects, random effects and pooled regression, we can see that the results are very similar. According to the tests, fixed effects are redundant, random effects are not rejected by the Hausman test but no random effects are preferred by the Breusch–Pagan test.

	Y^f				Y			
	α	β	γ	R^2	α	β	γ	R^2
fixed	−0.001 (0.010)	0.343 (0.656)	−0.914 (0.113)	0.446	−0.001 (0.001)	0.271 (0.113)	−0.165 (0.064)	0.157
random	−0.001 (0.010)	0.354 (0.650)	−0.915 (0.113)	0.445	−0.001 (0.001)	0.279 (0.112)	−0.164 (0.064)	0.153
pooled	−0.001 (0.010)	0.354 (0.638)	−0.915 (0.111)	0.445	−0.001 (0.001)	0.279 (0.110)	−0.164 (0.063)	0.153

Table 4 Panel estimations, standard errors in parentheses (*source: own computation*)

In the Y version, parameters β and γ are significant what shows a low capital mobility. Comprising FDI in consideration (Y^f version), β becomes insignificant but γ is significantly non zero. The former version can be called into question as for the theoretical grounds, the letter as for the interpretation of the results. There is no evidence allowing us to give a clear statement about capital mobility.

Going back to the basic idea that cointegration means low capital mobility and vice versa we have to admit that ECM models are not the only way how to study cointegration. Applying Pedroni residual cointegration test based on Engle – Granger with H_0: no cointegration, we have main results given in Table 5.

	Panel rho-statistic	***Prob***	**Weighted panel rho-statistic**	***Prob***
Y^f, X	−3.705	0.000	−2.842	0.002
Y, X	−1.610	0.054	−0.709	0.239

Table 5 Cointegration testing (*source: own computation*)

Pedroni [12] studied the properties of small samples ($T = 20$) with the conclusion that trading off size and power, the panel rho statistic appears to be the most consistently reliable statistic. Following his recommendation, we reject H_0 for Y^f and do not reject for Y. Here again, the attempt to compare $I(0)$ variable Y with $I(1)$ variable X as for a possible cointegration is often discussed with prevailing opinion that no cointegration can occur.

Relating savings and investment only, an un-appropriate treatment was performed not respecting theoretical assumptions. The results are contradictory: ECM model shows low capital mobility, Pedroni test speaks in favour of high mobility. Conclusion: ignoring exact assumptions one can prove everything. Such a non-

correct handling with facts unfortunately occurs in literature and gave inspiration to Deng at al. [4] to show its pitfalls.

Following the idea of Younas [19], the technique was applied correctly. The results do not support a notion of a high capital mobility in long – run. As the incoming FDI evince the increasing trend, the outcomming FDI are not strong enough to make the investment – savings relation independent.

3 Conclusions

The responses to demand for the measurement of capital mobility are econometric models based on investment–savings relations. As a technique, the possible existence of cointegration is verified. If confirmed, a long-run relation between the indicators exists, which implies capital immobility. If no cointegration occurs, the importance of domestic savings to finance domestic investment is low, which speaks in favour of capital mobility.

Relevant models and methods were applied to the V4 economies. By handling the region as a homogeneous structure, a rather low capital mobility in the long-run should be stated though a raising trend seems to be a very probable eventuality. For that matter, there is the permanent high demand for foreign investment in this economies. On the other hand, the V4 savings only rarely are invested in abroad.

Acknowledgements

The financial support of IGA F4/34/2020 is gratefully acknowledged.

References

[1] ARLTOVÁ, M., FEDOROVÁ, D. (2016). Selection of Unit Root Test on the Basis of Length of the Time Series, *Statistika* 96(3) (pp. 47–64).
[2] AZARIADIS, C., PISSARIDES, C. A. (2004). Unemployment Dynamics with International.
[3] Capital Mobility [online]. Available at: http://personal.lse.ac.uk/pissarid/papers/AP04.pdf.
[4] Committee on Industry, Research and Energy, EUROPEAN PARLIAMENT, (2013). *DRAFT REPORT on reindustrialising Europe to promote competitiveness and sustainability.*
[5] DENG, Y., HU, Y., LIN, J. H., WANG, M. X. (2013). Engle-Granger two-step method in panel data. In *Xitong Gongcheng Lilun yu Shijian/System Engineering Theory and Practice* 33(12) (pp. 3112–3119).
[6] EDWARDS, S. (2001). *Capital Mobility and Economic Performance: Are Emerging Economies Different?,* NBER Working Paper No. w. 8076, [online]. Available at:http://www.cedeplar.ufmg.br/economia/disciplinas/ecn933a/crocco/ Liberaliza cao_regulamentacao contas_capitais/EDWAR~18.PDF.
[7] ENGLE, R. F., GRANGER, W. J. (1991). *Long – Run Economic Relationships*, Oxford University Press.
[8] FELDSTEIN, M.; HORIOKA, C. (1980). Domestic Saving and International Capital Flows, *Economic Journal* 90 (358) (pp. 314–329).
[9] HENDRY, D. (1995). *Dynamic Econometrics*, Oxford University Press.
[10] MASUD, A. M., RAFIQUL, I. M. (2010). *Revisiting the Feldstein-Horioka Hypothesis of savings, investment and capital mobility* [online]. Available at: http://mpra.ub.uni-muenchen.de/39383/.
[11] OBSTFELD, M., SHAMBAUGH, J. C., TAYLOR, A. M. (2004). *The Trilemma in History: Tradeoffs among Exchange Rates, Monetary Policies, and Capital Mobility*, NBER Working Paper No. 10396.
[12] PALLEY, T. (2009). Rethinking the economics of capital mobility and capital controls, *Brazilian Journal of Political Economy*, vol. 29, no. 3 (pp. 15–34).
[13] PEDRONI, P. (1997). Panel cointegration [online], Indiana University. Available athttps://web.williams.edu/Economics/pedroni/pc-revc.pdf.
[14] RADEMACHER, I. (2013). *Tax Competition in the Eurozone–Capital mobility, Agglomeration, and the Small Country Disadvantage*, Max-Planck-Institut für Gesellschaftsordnung, Discussion paper 13/13.
[15] RAKOVSKI, T. (2009). *Consumption Smoothing Current Account and International Capital Mobility–Evidence from Time Series and Panel Data*, [online]. Available at: http://digitool.library.mcgill.ca/ webclient/ StreamGate? folder_id=0&dvs=1385703336096~593.
[16] RUIST, J., BIGSTEN, A. (2013). Wage Effects of Labour Migration with International Capital Mobility, *The World Economy*, vol. 36, issue 1 (pp. 31–47).
[17] SCHNEIDER, B. (1999). *Saving. Investment Correlation and Capital Mobility in Developing Countries*, Working Paper No. 48, Indian Council for Research in International Economic Relations.

[18] TAYLOR, A. M. (1996). *International Capital Mobility in History: The Saving-Investment Relationship*, NBER Working Paper No. 5743, Cambridge.
[19] YOUNAS, J. (2013). Role of Foreign Direct Investment in Estimating Capital Mobility: A Reappraisal of Feldstein-Horioka Puzzle, [online]. Available at: http://www.freit.org/WorkingPapers/Papers/ForeignInvestment/FREIT106.pdf.

Analysis of the Empirical Distribution of Daily Returns of DJIA Index and Possibilities of Using Selected Distributions

Juraj Pekár[1], Mário Pčolár[2]

Abstract. In the past, a large number of authors pointed to several alternative distributions that would better capture the characteristics of daily returns than is the case with a normal distribution. Empirical observations point to the presence of thicker tails, more spiked center, and also the possible presence of skewness in return distributions. This may indicate the possibility of using alternative distributions in order to capture such externalities compared to the initial assumption of a normal distribution. The paper deals with the analysis of the empirical distribution of daily returns of selected stocks and the comparison of selected alternative distributions with the normal distribution in the ability to model the behavior of daily returns during the interval of 20 years as well as during partial intervals. The aim of this paper is to examine the goodness of fit of selected distributions for the needs of modeling the daily returns of DJIA shares in comparison with the normal distribution. The assumption that the data is not normally distributed has been confirmed, the best goodness of fit from the selected distributions provides the scalable Student's t-distribution.

Keywords: fitting distributions, Student's t-distribution, normal distribution, Laplace distribution, logistic distribution, Cauchy distribution

JEL Classification: C13, G19
AMS Classification: 62E17, 91B84

1 Introduction

Accepted fact in the financial literature is that the empirical distribution of stocks is leptokurtic, i.e. the distribution is more spiked in the center and the distribution is thicker in the tails compared to the normal distribution. The aim of this paper is to analyze the empirical distribution of daily returns of shares contained in the DJIA index. Also examine the possibility of using other alternative distributions compared to the normal distribution. In the analysis, we will consider the following distributions: scalable Student's t-distribution, Laplace distribution, logistic distribution, power exponential distribution and Cauchy distribution. The given distributions are characterized by thicker tails of the distributions and greater spikiness in the centers compared to the normal distribution. All of these distributions have been proposed in the past by several authors as possible alternatives to the normal distribution, [1, 2, 3, 4, 5, 6]. In comparison with the mentioned articles, we deal with the investigation of the power of fit of selected probability distributions for data from entire selected time interval and sub periods selected by trends. In the first part of the paper, the authors describe the analyzed data and the results of data analysis. In the following section, they describe the testing of the goodness of fit of selected distributions and present the achieved results. The last part of the paper summarizes the results of the analysis.

2 Data

The data source for our analysis is portal finance.yahoo.com. The analyzed data represent information of the price levels of the Dow Jones Industrial Average components on a daily basis. In the analysis, we also use data of the price levels of the index itself. The daily values of the adjusted closing positions of individual shares and the index itself are for a period of 20 years, from 03.01.2000 to 31.12.2019. The evolution of the DJIA index values is shown in Figure 1. The daily returns of the individual stocks as well as of the index itself

[1] University of Economics in Bratislava, Department of Operations research and Econometrics, Dolnozemská 1/b, 852 35 Bratislava, Slovak Republic, juraj.pekar@euba.sk
[2] University of Economics in Bratislava, Department of Operations research and Econometrics, Dolnozemská 1/b, 852 35 Bratislava, Slovak Republic, mario.pcolar@euba.sk

are quantified in a continuous manner that is consistent with a continuous interest rate. Calculated as difference between the natural logarithms, the closing price level from day t and the closing price level in day $t-1$.

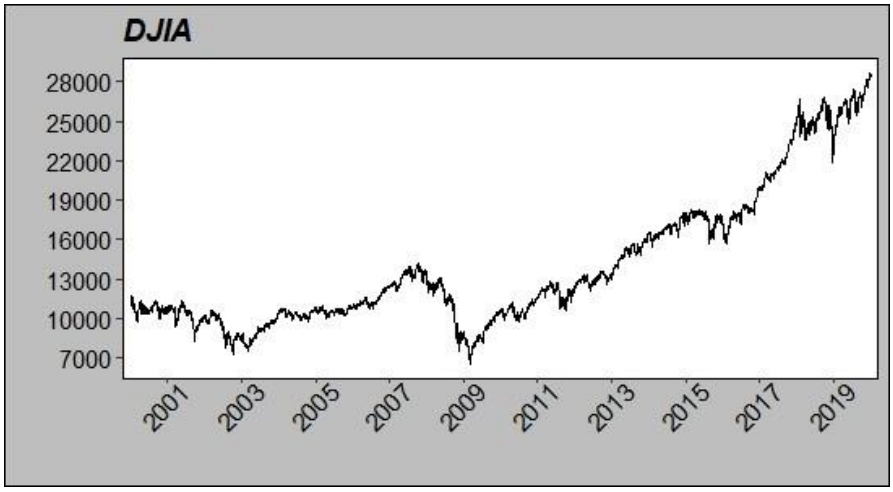


Figure 1 Development of DJIA levels in the analyzed period

The individual analyzed DJIA components included in our analysis are: *Dow, Inc.* (DOW), *Walgreens Boots Alliance, Inc.* (WBA), *Chevron Corporation* (CVX), *Walmart, Inc.* (WMT), *Verizon Communications, Inc*. (VZ), *International Business Machines Corporation* (IBM), *American Express Company* (AXP), *Apple, Inc*. (AAPL), *The Boeing Company* (BA), *Exxon Mobil Corporation* (XOM), *McDonald's Corporation* (MCD), *Johnson & Johnson* (JNJ), *Cisco Systems, Inc*. (CSCO), *The Coca-Cola Company* (KO), *NIKE, Inc.* (NKE), *The Travelers Companies, Inc*. (TRV), *JPMorgan Chase & Co.* (JPM), *Merck & Co., Inc*. (MRK), *The Home Depot, Inc.* (HD), *The Procter & Gamble Company* (PG), *The Walt Disney Company* (DIS), *3M Company* (MMM), *Pfizer, Inc.* (PFE), *United Technologies Corporation* (UTX), *Microsoft Corporation* (MSFT), *Caterpillar, Inc*. (CAT), *UnitedHealth Group Incorporated* (UNH), *Intel Corporation* (INTC), *The Goldman Sachs Group, Inc*. (GS), *Visa Inc.* (V).

2.1 Data analysis

From Figure 1, it is clear that the development of the values of the DJIA index in the period between 2000 and 2008 is characterized by a so-called "stable" behavior, stagnating around the level of 10000. Following the global financial and economic crisis, the trend is changing. In the period between 2009 and 2019, we observe a growth trend that persists throughout whole of the observed period. In accordance with such a finding, we extended the analysis of the empirical distribution of returns by the analysis of sub periods. The period 2000–2008 when the DJIA is characterized by stable behavior and the period 2009–2019 when the DJIA is characterized by more significant growth.

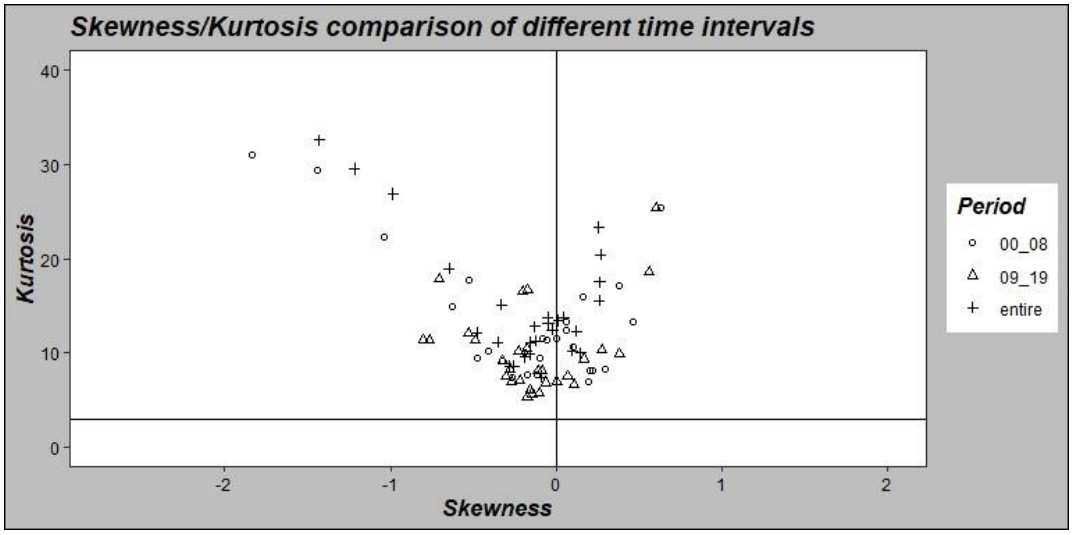


Figure 2 Comparison of Skewness and Kurtosis for each stocks and index itself with normal distribution in different periods

	WBA			CVX			WMT		
Mean	0.0002	−4.0E−05	0.0004	0.0004	0.0004	0.0003	0.0002	−3.6E−05	0.0004
Std	0.0173	0.0185	0.0163	0.0158	0.0180	0.0138	0.0148	0.0183	0.0112
Var	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002	0.0002	0.0003	0.0001
Skewness	−0.3518	−0.1928	−0.5280	0.0440	0.1610	−0.1738	0.0981	0.1952	−0.2007
Kurtosis	11.1324	10.0743	12.1639	13.7220	15.9960	5.4115	10.2785	6.9940	16.6086
Max	0.1539	0.1539	0.1114	0.1894	0.1894	0.0664	0.1050	0.1050	0.1034
Min	−0.1624	−0.1624	−0.1548	−0.1334	−0.1334	−0.0784	−0.1074	−0.0949	−0.1074
JarquBera test_p.value	0	0	0	0	0	0	0	0	0
Shapiro_test_p.value	—	0	0	—	0	0	—	0	0
	VZ			IBM			XOM		
Mean	0.0002	−0.0001	0.0004	0.0004	0.0004	0.0003	0.0002	0.0004	0.0001
Std	0.0153	0.0193	0.0112	0.0158	0.0180	0.0138	0.0150	0.0179	0.0122
Var	0.0002	0.0004	0.0001	0.0003	0.0003	0.0002	0.0002	0.0003	0.0001
Skewness	0.1485	0.2185	−0.1005	0.0440	0.1610	−0.1738	0.0140	0.0577	−0.1443
Kurtosis	10.0131	8.1800	5.7291	13.7220	15.9960	5.4115	13.4491	13.3701	5.6111
Max	0.1366	0.1366	0.0740	0.1894	0.1894	0.0664	0.1586	0.1586	0.0650
Min	−0.1261	−0.1261	−0.0644	−0.1334	−0.1334	−0.0784	−0.1503	−0.1503	−0.0639
JarquBera test_p.value	0	0	0	0	0	0	0	0	0
Shapiro_test_p.value	—	0	0	—	0	0	—	0	0
	MCD			JMJ			CSCO		
Mean	0.0004	0.0003	0.0005	0.0003	0.0002	0.0004	2.6E−05	−0.0005	0.0005
Std	0.0143	0.0181	0.0103	0.0120	0.0143	0.0097	0.0241	0.0310	0.0166
Var	0.0002	0.0003	0.0001	0.0001	0.0002	0.0001	0.0006	0.0010	0.0003
Skewness	−0.1594	−0.1754	0.0698	−0.6473	−0.5272	−0.8011	0.1228	0.2993	−0.7064
Kurtosis	9.8475	7.6496	7.6199	18.9493	17.7172	11.3698	12.3493	8.3583	17.9252
Max	0.0897	0.0897	0.0781	0.1154	0.1154	0.0524	0.2182	0.2182	0.1480
Min	−0.1372	−0.1372	−0.0517	−0.1725	−0.1725	−0.1058	−0.1769	−0.1405	−0.1769
JarquBera test_p.value	0	0	0	0	0	0	0	0	0

Shapiro_test_p.value	—	0	0	—	0	0	—	0	0
		KO			MRK			KD	
Mean	0.0002	−8.E−06	0.0004	0.0002	−0.0002	0.0005	0.0003	−0.0004	0.0009
Std	0.0128	0.0157	0.0099	0.0170	0.0207	0.0133	0.0193	0.0244	0.0137
Var	0.0002	0.0002	0.0001	0.0003	0.0004	0.0002	0.0004	0.0006	0.0002
Skewness	−0.0226	0.0993	−0.3231	−1.4295	−1.8350	0.2756	−0.9846	−1.0398	0.1079
Kurtosis	12.4683	10.5831	9.1994	32.5627	31.0319	10.4050	26.8288	22.2461	6.6657
Max	0.1300	0.1300	0.0729	0.1225	0.1225	0.1191	0.1316	0.1316	0.0996
Min	−0.1060	−0.1060	−0.0881	−0.3117	−0.3117	−0.0801	−0.3388	−0.3388	−0.0622
JarquBera test_p.value	0	0	0	0	0	0	0	0	0
Shapiro_test_p.value	—	0	0	—	0	0	—	0	0

Table 1 Descriptive statistics of selected daily stock returns and index itself

Table 1 includes descriptive statistics of 12 selected stocks returns as well as the index itself over the entire analysed period as well as within sub periods. The first column is always for the entire period, the second one is for 2000–2008, and the last one for 2009–2019. In addition to the usual statistics such as standard deviation, variance, min, max, or skewness and kurtosis, the table contains p.value of statistical normality tests. In the case of sub periods, in addition to the Jarque-Bera test, we also estimated the Shapiro-wilk test, which is a more powerful normality test, but on the other hand its limitation lies in the upper limit of the number of observations, which is satisfied only by sub periods. In all cases, the p.values are close to 0, which indicates the fact that we reject the hypothesis of a normal distribution of the daily returns of the individual components of the DJIA as well as the index itself. Descriptive statistics also show that DJIA's daily returns in a "stable" period have a greater kurtosis as well as greater variance. Such a finding also applies to the daily returns of individual components.

A comparison of skewness and kurtosis of the individual stocks and the index with the skewness and kurtosis characteristic for a normal distribution is shown in Figure 2. The black lines indicate the characteristic values of the skewness and kurtosis coefficients in the case of a normal distribution. From the Figure we omitted two observations in the case of a stable period as well as in case of the whole interval, as the values of these observations are significantly higher compared to the rest of the observations. The values of these observations are given in Table 1, namely values of the AAPL and PG. The comparison shows that in almost all cases, the observed values of skewness and kurtosis are higher than in the case of a normal distribution. Significant deviations from the normal distribution are especially in the values of the kurtosis, where the values of individual cases are significantly higher than in the case of the normal distribution kurtosis level equal to 3. Such a finding also supports the findings of normality tests. A comparison of the values for the individual sub periods again shows that the daily returns of the DJIA components are characterized in the stable period by higher values of the kurtosis compared to the growth period. The greatest influence on the high values of the kurtosis in the case of a "stable" period have significant outliers in the period between 2000 and 2002.

3 Comparison of goodness of fit of different models

In this part of the paper, we examine the goodness of fit of individual selected probability distributions in comparison with the normal distribution. In the analysis we will consider the following distributions: scalable Student's t-distribution, Laplace distribution, logistic distribution, power exponential distribution and Cauchy distribution. All of these distributions have been proposed in the past by several authors as possible alternatives to the normal distribution, as they are characterized by thicker tails of the distribution and greater spikiness in the center compared to the normal distribution. We use the maximum likelihood estimation method to estimate the parameters of individual distributions. We use RStudio software to quantify individual calculations.

One way to compare the goodness of fit of individual distributions compared to a normal distribution is to graphically compare the power of fit. A graphical comparison of the goodness of fit is illustrated by the data of Exxon Mobil Corporation (XOM) for the entire period under review. The Figure 3 shows a comparison of the normal distribution with selected alternative distributions. In the upper part there is a histogram and estimated density functions of individual distributions in comparison with the empirical distribution of

data, in the lower part from the left the QQ graph comparing data quantiles (y-axis) versus theoretical quantiles (x-axis), graph comparing empirical CDF and estimated CDF of individual distributions and PP graph comparing empirical and theoretical probabilities quantified for each point. The Q-Q graph provides additional information on the power of fit by focusing on the tails of the distribution. On the other hand, the P-P graph provides additional information on the power of the fit by focusing on the central part of the distribution.

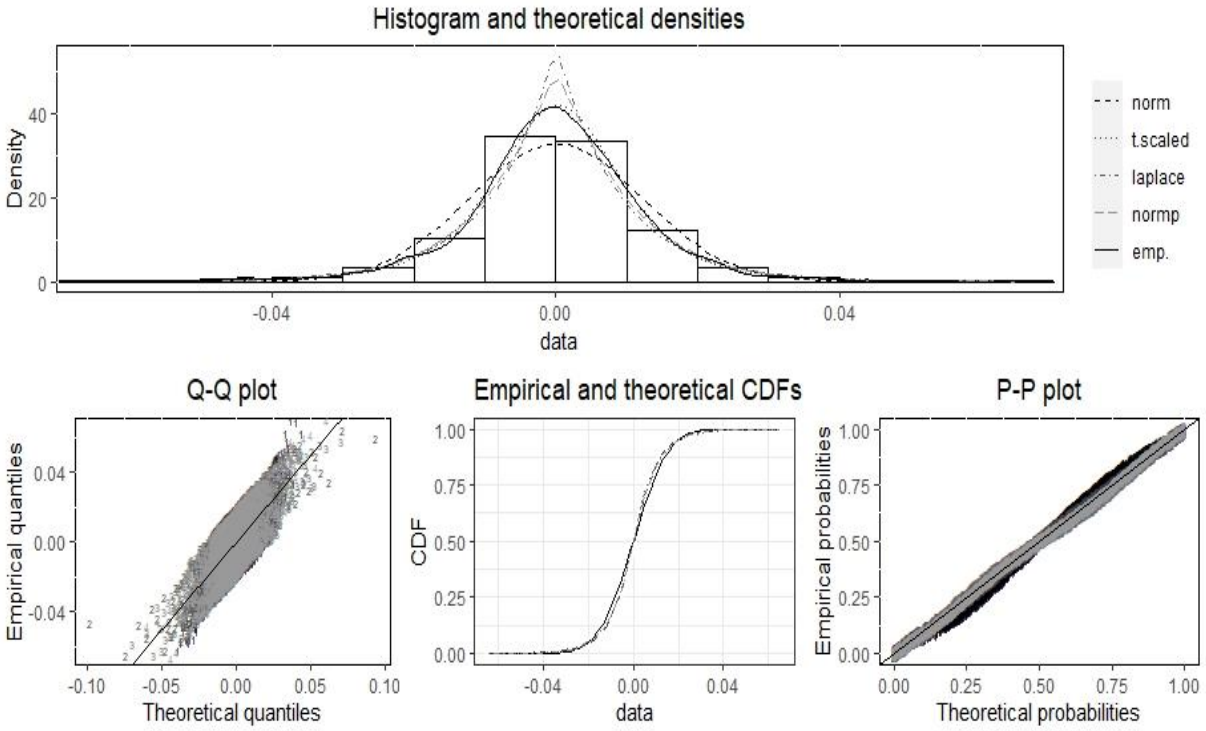


Figure 3 Graphical comparison of different distributions with normal distribution

A graphical comparison of individual models for individual stocks shows that the scalable student's t-distribution provides the greatest power of fit from selected distributions. The normal distribution underestimates the center and tails of the empirical distribution.

We will use Anderson-Darling statistics for a more exact comparison of the goodness of fit of individual selected distributions. In general, these statistics measure the deviations of the fit between the distribution function (CDF) $\hat{F}$ and the empirical distribution function F_n. Anderson-Darling statistic, which is calculated as a weighted average of square deviations $[F_n(x) - F(x)]^2$. The weights are chosen to place more emphasis on deviations at the tails of the distributions. Such a measure may be more appropriate for assessing the power of the fit if we pay particular attention to the power of fit at the tails of the distribution, which may be of particular interest for risk management needs. For these statistics, the smaller the value, the better the power of fit to the empirical distribution.

It is important to note that using these statistics to select the best distribution, from probability distributions with different numbers of parameters, misinterpretations can occur. The goodness of fit statistics alone describe only how well the estimated distribution describes the empirical distribution, the greater the number of parameters in the model increases the flexibility of the estimated distribution itself, on the other hand increases the complexity of the model which can lead to overfitting. Anderson-Darling statistics (AD) do not penalize the number of parameters in the model in any way. To overcome the above, we also present statistics based on information criteria that take into account the very complexity of the model. In the analysis, we present the Akaike Information Criterion (AIC).

	Anferson darling statistics						Akaike information criterion					
	Norm	**T.scaled**	**Logis**	**Lapl**	**GED**	**Cauch**	**Norm**	**T.scaled**	**Logis**	**Lapl**	**GED**	**Cauch**
WBA	Inf	**0.490**	7.935	1.764	1.516	34.897	−26517	**−27486**	−27298	−27439	−27438	−26603
CVX	Inf	**0.195**	5.179	5.357	3.541	42.453	−27445	**−28385**	−28224	−28259	−28270	−27364
WMT	82.347	**0.229**	14.580	3.311	3.367	32.402	−28104	**−29414**	−29077	−29311	−29309	−28647

	Anferson darling statistics						Akaike information criterion					
	Norm	**T.scaled**	**Logis**	**Lapl**	**GED**	**Cauch**	**Norm**	**T.scaled**	**Logis**	**Lapl**	**GED**	**Cauch**
VZ	Inf	**0.425**	10.515	2.137	2.064	33.877	−27742	**−28765**	−28541	−28708	−28706	−27931
IBM	88.264	**0.319**	18.096	3.269	3.268	29.922	−27249	**−28577**	−28216	−28484	−28482	−27858
AXP	Inf	**1.314**	35.255	6.436	6.448	19.712	−24147	**−26057**	−25429	−25934	−25932	−25557
AAPL	95.660	**1.548**	15.127	1.140	Inf	28.142	−22279	**−23984**	−23691	−23950	−23948	−23246
BA	44.618	**0.285**	6.135	3.629	2.216	39.393	−25670	**−26412**	−26282	−26340	−26353	−25448
XOM	Inf	**0.232**	5.732	4.444	3.137	41.258	−27949	**−28965**	−28781	−28836	−28842	−27975
MCD	68.449	**0.557**	12.620	3.076	3.087	33.376	−28433	**−29523**	−29265	−29466	−29464	−28723
JNJ	Inf	**0.691**	14.442	1.797	1.790	29.715	−30222	**−31529**	−31232	−31479	−31477	−30777
CSCO	Inf	**0.932**	32.223	5.684	5.618	21.669	−23181	**−24942**	−24367	−24840	−24838	−24406
KO	Inf	**0.318**	15.888	2.675	2.670	30.560	−29535	**−30898**	−30550	−30807	−30805	−30155
NKE	87.654	**0.431**	15.786	2.435	2.438	30.345	−25783	**−27200**	−26846	−27107	−27105	−26460
TRV	Inf	**0.384**	23.506	3.875	3.869	26.402	−26242	**−28138**	−27611	−27992	−27990	−27501
JPM	Inf	**1.160**	36.378	6.729	6.707	18.994	−23200	**−25392**	−24685	−25224	−25222	−24907
MRK	88.540	**0.182**	11.264	3.409	3.434	35.634	−26687	**−28265**	−27926	−28112	−28110	−27426
HD	89.217	**0.781**	18.411	1.763	1.766	26.947	−25454	**−26851**	−26513	−26809	−26807	−26162
PG	124.651	**0.332**	16.922	2.993	Inf	30.237	−29106	**−31247**	−30800	−31091	−31089	−30517
DIS	81.537	**0.445**	14.989	1.800	1.778	29.665	−25769	**−27080**	−26757	−27014	−27012	−26338
MMM	72.649	**0.792**	14.465	1.370	1.377	29.043	−28282	**−29403**	−29125	−29376	−29374	−28659
PFE	58.281	**0.516**	10.250	2.448	2.298	33.742	−27649	**−28569**	−28371	−28527	−28526	−27732
UTX	72.502	**0.895**	12.662	0.944	0.984	29.950	−26943	**−28166**	−27918	−28138	−28136	−27388
MSFT	Inf	**0.810**	19.159	1.951	1.952	26.667	−25604	**−27027**	−26648	−26974	−26972	−26353
CAT	45.654	**0.533**	7.212	1.873	1.168	35.363	−25040	**−25773**	−25640	−25749	−25753	−24869
UNH	Inf	**0.426**	11.147	1.949	1.943	32.887	−25325	**−26746**	−26447	−26644	−26642	−25929
INTC	80.319	**0.309**	15.326	2.447	2.458	30.880	−23570	**−24830**	−24513	−24756	−24754	−24079
GS	Inf	**0.274**	14.303	2.143	2.127	31.064	−23701	**−25214**	−24848	−25105	−25103	−24453
DJI	Inf	**1.927**	21.003	2.645	2.643	26.243	−30890	**−32168**	−31823	−32160	−32158	−31507

Table 2 Comparison of AIC and AD statistics for each model in entire period

From the comparison of AIC and AD statistics values given in Table 2, shows that all of the alternative distributions provide significantly better power of fit compared to the normal distribution. By far the best power of fit is provided by the scalable Student's t-distribution whose values are the lowest in all observations. Also, after taking into account the complexity of the models using AIC statistic, the Student's t-distribution is significantly the best distribution from tested one. The only exception is the case of daily returns of the DJIA index, where both in accordance with AD and AIC it is not possible to clearly identify the best model. Comparable good power of fit is provided by the Student's t-distribution, Laplace distribution and GED distribution.

Comparison in Figure 4 points out that the best power of fit from selected probability distributions, regardless of the long-term trend, is provided by the student's t-distribution. Cauchy distribution fits to empirical distribution better in the period 2000–2008 in comparison with other sub period. For the remaining distributions, the power of fit of individual distributions in individual periods varies depending on the stocks.

4 Conclusion

The aim of the analysis was to evaluate the goodness of fit of selected distributions for individual financial data represented by different time periods. The results of our analysis point to several findings. The empirical distributions of the daily returns of the individual stocks of the DJIA index and the index itself are spikier and have thicker tails compared to the normal distribution. The data of several stocks in the period 2000–2008 have a significantly higher kurtosis. The normal distribution of stocks daily returns was not found on any of the stocks examined, nor in the case of the index itself. From a comparison of selected probability distributions, it is clear that the scalable Student's t-distribution provides the best power of fit for individual stocks, regardless of the long-term trend of the DJIA index. In the case of the DJIA index alone, comparable power of fit is provided by the Student's t-distribution, GED and Laplace distribution. The goodness of fit of those distributions is comparable and the best distribution cannot be clearly identified. The above also applies to AIC. As it is clear from the data analysis that in several cases the data are characterized by a higher spikiness in the center of the distribution, thicker tails and the possible presence of asymmetry, it is possible to consider using more flexible distributions with more parameters in the future. The given analysis is the basis for future research into the adjustment of financial data for the needs of optimization models of portfolio selection.

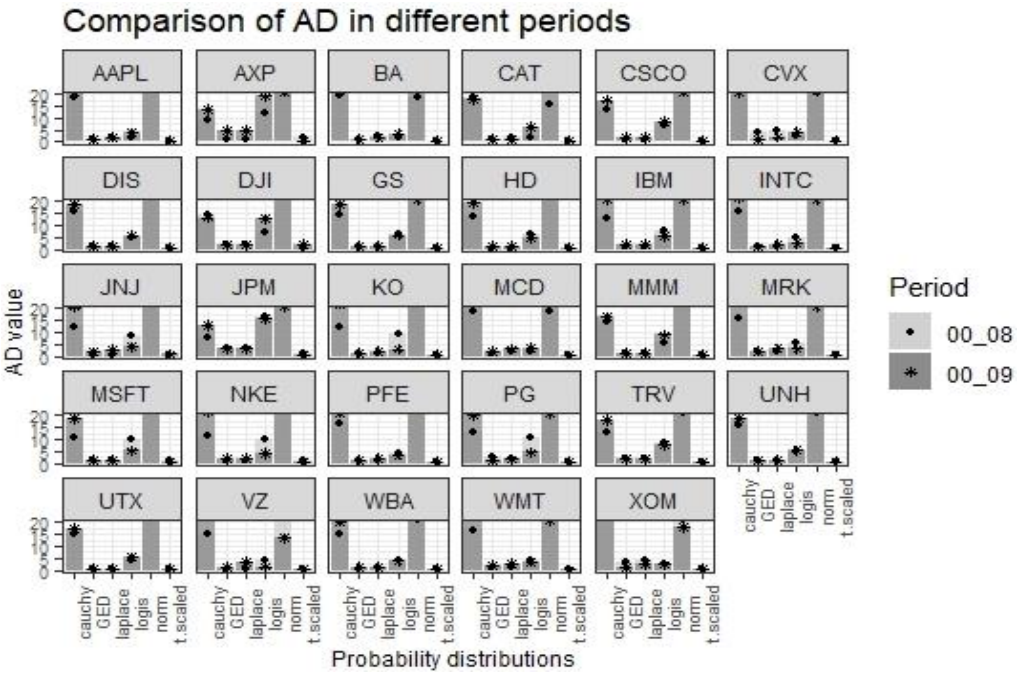


Figure 4 Graphical comparison of different distributions with normal distribution in different periods

Acknowledgements

This work was supported by the Grant Agency of Slovak Republic – VEGA grant no. 1/0339/20 "Hidden Markov Model Utilization in Financial Modeling".

References

[1] BLATTBERG, R. & GONEDES, N. (1974). A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices. *The Journal of Business*, 47(2), pp. 244–280.

[2] CORLU, C. G., METERELLIYOZ, M., & TINIÇ, M. (2016). Empirical distributions of daily equity index returns: A comparison. *Expert Systems with Applications*, 54, pp. 170–192.

[3] ELING, M. (2014). Fitting asset returns to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics*, 59, pp. 45–56.

[4] GRAY, J. B. & FRENCH, D. W. (2008). Empirical comparison of distributional models for stock index returns. *Journal of Business Finance & Accounting*, 17, pp. 451–459.

[5] HARRIS, R. D. F. & KÜÇÜKÖZMEN, C. (2001). The empirical distribution of stock returns: evidence from an emerging European market. *Applied Economics Letters,* 8(6), pp. 367–371.

[6] PRAETZ, P. D. (1972). The Distribution of Share Price Changes. *The Journal of Business*, 45(1), pp. 49–55.

Dynamic Vehicle Routing problem

Jan Pelikán[1]

Abstract. This article studies vehicle routing problem with vehicle routes carrying out in discrete time, e.g. days. The goal is to find vehicle routes for particular day for which nodes and their demand are given. There are two kinds of node every day: compulsory nodes and optional nodes. Compulsory (urgent) nodes have to be included in routes obligatory and optional nodes can be further postponed and included in transport on some next days together with new incoming transport demands. New transport demands arise continually and hold the information about node and its volume of demands as well as the time interval or deadline by which these transport demand needs to be included in transport. The aim is then to optimize the vehicle routes for each day which includes all compulsory nodes and those optional nodes which are the most effective fitting to be involved in the routes for the particular day. The objective function is total routes length per unit of the transported volume, this function is the linear-fractional function, Charles-Cooper transformation converts our nonlinear problem to a linear problem and the linear programming is used. Both mathematical model and heuristic method are proposed.

Keywords: vehicle routing problem, Charles-Cooper transformation, heuristic

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Vehicle routing problem (VRP hereafter) is a classical problem of operational research. There is a number of modifications to VRP coming from case studies assuming different transport conditions [1, 2, 3, 5]. This article proposes one of such modifications. The modification concerns in particular the continual vehicle route design which takes place in discrete time, e.g. vehicle routes for daily transport. It is necessary then to create the vehicle routes based on daily demands. Demands for transport gradually rise. Each demand is characterized by its location, the volume that is to be transported and the time interval (in days) for the demanded merchandise to be involved in the vehicle routes. The lower value of this interval corresponds to the merchandise being ready to be expedited whereas the upper value can have a connection with running out of supplies of the customer, node of the communication network. The nodes represent the final demand location, the customers, where the merchandise is transported from the depot. Both distances between the individual nodes and the nodes and the depot are also known.

Demands have to be assigned to particular day and designed vehicle routes for the actual transport. The transport date must correspond to the time interval for the particular demand. The transport demand involves the location – one of the nodes, the volume and the time interval for the demanded merchandise to reach the respective node. The demands arise gradually in time and the whole process proceeds continually. The aim is to minimize the transport cost.

Article [6] proposes mathematical solution and heuristic method for VRP in time-limited interval (week, month…). Given the fact that the demands for the particular time interval do not necessarily have to be released beforehand but instead they arise gradually the latter approach has to be unusable. The end of the time interval by which the demands need to be involved in the transport together with the possibility that the delivery time is exceeding the given time interval can lead to uneconomical vehicle use.

The solution is to optimize the routes for each day separately. Further we have to realize that the number of demands for a given day is comprised of new demands arising that day and older demands that have not been involved in the transport yet. We solve then the static VRP with a certain difference. The transport demands in the nodes not only involve the urgent demands, i.e. those that need to be involved in the transport immediately. They also involve the optional transport demands that can be further postponed unless their delivery time is not exceeded.

[1] University of Economics Prague, W. Churchill sq. 4, Prague 3, pelikan@vse.cz

The urgent demands need to be included in the transport as they are considered to be necessary, whereas in case of the rest of the demands one has to decide whether their transport can or cannot be further postponed. Each demand is defined by its location, i.e. the node where the shipment needs to be delivered to and the actual volume of the particular demand.

If we chose the objective function to be of standard type, i.e. the total sum of all route distances, then the optimal solution would not involve the optional nodes as that would increase the objective function itself. In case we chose the objective function to be represented by the total transported volume, which needs to be maximized, on the contrary the optimal solution with total distance object function would involve all optional nodes, which as noted before would lead to inefficiency in route distance planning as well as in use of vehicle capacity. Based on these reasons we propose a form of objective function that represents the average route distance per transported volume unit. This function, defined as ratio of total route distance and total transported volume, will be minimized. The proposed problem will be first modeled assuming non-linear objective function, that is represented by the linear-fractional function and set of constraints, which is essentially identical to that of classical VRP.

To find the solution for this non-linear problem we use Charles-Cooper transformation [4] that converts our problem to a linear programming one. Despite the fact that the transformation process treats the binary variables the binary condition does not have to hold. So additional conditions and variables have been included to the transformed model. Except for the mathematical model one can also modify the heuristic methods designed to solve the classical VRP. Finally, at the end of the article, we present a numerical solution to a problem demonstrating both approaches.

2 Mathematical model of VRP with obligatory and optional nodes

First let us propose the mathematical model which is result of modification of classical VRP. The difference is based on the objective function which is in this case represented by linear-fractional function.

Parameters of the model:

n number of nodes,

m number of compulsory nodes, nodes 2,3, …, *m* are compulsory nodes, nodes *m+1, m+2, …, n* are optional, node 1 is depot,

c_{ij} distance between node *i* and node *j*,

q_i demand of node *i*,

W capacity of vehicle.

Variables of the model are:

x_{ij} binary, equals 1 if a vehicle travels from node *i* to node *j*,

u_j variables in anti-cyclic constraints.

$$f(x) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} q_i x_{ij}} \rightarrow min \tag{1}$$

$$\sum_{i=1}^{n} x_{ij} = 1, \quad i = 1,2, \dots, m \tag{2}$$

$$\sum_{i=1}^{n} x_{ij} = \sum_{i=1}^{n} x_{ji}, \quad i = 1,2, \dots, n \tag{3}$$

$$u_i + q_j - W(1 - x_{ij}) \leq u_j \quad i = 1,2, \dots, n, \quad j = 2,3, \dots, n, \quad i \neq j \tag{4}$$

$$0 \leq u_j \leq W \quad j = 2,3, \dots, n \tag{5}$$

$$x_{ij} \text{ binary} \quad i, j = 1,2, \dots, n, \quad i \neq j. \tag{6}$$

The object function (1) is ratio with denominator total amount of loads of all routes and numerator total length of all routes. Equation (2) ensures that compulsory nodes will be entered and its demand q_j is covered. Equation (3) means condition: if vehicle enters a node it has to leave it. Anti-cyclic conditions are in (4). Inequality (5) assures that capacity of vehicles is not exceeded.

Model (1)–(6) is not linear in its objective function but can be transformed into a linear program rather easily using Charnes-Cooper transformation. The key is to substitute the variables x_{ij}, resp. u_i for the variables x'_{ij} and u'_i (respectively) in (1)–(5) using following relations $x'_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} q_i x_{ij}}, u'_j = \frac{u_j}{\sum_{i=1}^{n} q_i x_{ij}}, t = \frac{1}{\sum_{i=1}^{n} q_i x_{ij}}$ and new variable t. That way one can obtain a linear version of the mathematical model given as (7)–(12).

$$f'(x) = \sum_{i,j=1}^{n} c_{ij} x'_{ij} \to \min \tag{7}$$

$$\sum_{i=1}^{n} x'_{ij} = t, \quad i = 1,2,\dots,m \tag{8}$$

$$\sum_{i=1}^{n} x'_{ij} = \sum_{i=1}^{n} x'_{ji}, \quad i = 1,2,\dots,n \tag{9}$$

$$u'_i + q_j t - Wt + Wx'_{ij} \le u'_j, \quad i = 1,2,\dots,n, \quad j = 2,3,\dots,n, \quad i \ne j \tag{10}$$

$$0 \le u'_j \le Wt, \quad j = 2,3,\dots,n \tag{11}$$

$$x'_{ij} \ge 0, \quad i,j = 1,2,\dots,n, \quad i \ne j \tag{12}$$

$$t \ge 0,$$

where $x'_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} q_i x_{ij}}, u'_j = \frac{u_j}{\sum_{i=1}^{n} q_i x_{ij}}, t = \frac{1}{\sum_{i=1}^{n} q_i x_{ij}}$.

Original variables can be derived as $x_{ij} = x'_{ij}/t$ for all *i, j*. Binary conditions for variables x_{ij} can be ensured by additional conditions (13) and (14) where *M* is big number and additional binary variables y_{ij}. If $x'_{ij} = t$, then $x_{ij} = 1$, and if $x'_{ij} = 0$, then $x_{ij} = 0$.

$$-M(1 - y_{ij}) \le x'_{ij} - t \le M(1 - y_{ij}), \quad i \ne j \tag{13}$$

$$-My_{ij} \le x'_{ij} \le My_{ij}, \quad i \ne j \tag{14}$$

$$y_{ij} \text{ is binary for all } i \ne j. \tag{15}$$

The mathematical model (7)–(15) is binary linear program and can be solved using conventional LP packages like GUROBI, CPLEX, etc.

3 Numerical example

The proposed mathematical model was verified on an illustrative example. Consider 11 nodes where node 1 is a depot and $m = 6$. Capacity of each vehicle is $W = 100$. The requirements of the nodes are $q = (0\ 19\ 24\ 30\ 20\ 35\ 25\ 32\ 20\ 22\ 37)$. The distance matrix *C* is as below:

0	13	6	55	93	164	166	168	169	241	212
13	0	11	66	261	175	177	179	180	239	208
6	11	0	60	97	168	171	173	174	239	209
55	66	60	0	82	113	115	117	117	295	265
93	261	97	82	0	113	115	117	118	333	302
164	175	168	113	113	0	6	4	2	403	374
166	177	171	115	115	6	0	8	7	406	376
168	179	173	117	117	4	8	0	2	408	378
169	180	174	117	118	2	7	2	0	409	379

241	239	239	295	333	403	406	408	409	0	46
212	208	209	265	302	374	376	378	379	46	0

The optimal objective function of model (7)–(15) is $f'(x) = 3$, optimal value of variable $t = 0.0058$. Therefore, $x'_{ij} = 0.0058$, if $y_{ij} = 1$ otherwise $x'_{ij} = 0$. From optimal values of variables $x'_{ij}, y'_{ij} x_{ij}$ it is possible to derive that the optimal routes are:

1. route: 1-3-2-4-1 with transport volume 73 and length of the route 138,

2. route: 1-5-7-9-6-1 with transport volume 100 and length of the route 381,

The total length of all routes is 519, and the total load is 173. The length on one unit of load is 3 which is the optimal value of the objective function.

4 Heuristic method

As the studied problem is NP hard a modification of insertion heuristic is proposed. First, we create routes that contain only obligatory nodes. Further we focus on including the optional nodes in these routes whereas the vehicle capacity cannot be exceeded. If the initially designed routes do not provide us a possibility to add optional nodes due to lack of vehicle capacity new routes are designed. These new designed routes consist only of optional nodes. An optional node is involved in the new route design only if it causes the objective function value to decrease.

Step 1 (route design involving only obligatory nodes):

The *s*-th route involving obligatory nodes is denoted as $(v_1^s = 1, v_2^s, v_3^s, \dots v_{h(s)-1}^s, v_{h(s)}^s = 1)$, either a mathematical model or heuristic method are applied. The total demand of the obligatory nodes is $L = \sum_{i=1}^{m} q_i$ and the total route distance $D = \sum_{s=1}^{S} \sum_{i=1}^{h(s)-1} d_{v_i^s v_{i+1}^s}$. The objective function *r* value (involving no optional node) is then given as $r = L/D$.

Step 2 (adding of optional with the aim of decreasing the objective function value *r*):

Let us calculate the value $r' = \frac{L+q_k}{D+c_{ik}+c_{kj}-c_{ij}}$ for all of the optional nodes k (that for now are not involved in any designed vehicle route) and for each arc if route *(i,j)*. The value r' is result of putting the node k between the nodes *i* and *j*, assuming the vehicle capacity will not be exceeded as we add the demand q_k to the given route.

If $r' < r$ and is lowest possible, then the node k is added into the route and we set $r := r'$, iterate the step 2 until no more nodes can be added to the route due to vehicle capacity being exceeded.

Step 3 (a new route design):

If it is not possible to add optional nodes to the initially designed routes new routes will be designed so that the respective objective function value is minimal.

5 Numerical example (continued)

Let us use the data from the previous section.

Step 1: Nearest neighbor method gives us the following solution:

Route-1: 1-3-2-4-5-1 with the route distance 586 and the transported volume 128.

Route-2: 1-6-1 with the route distance 328 and the transported volume 35.

Total route distance is *D*=586, total transported volume is *L*=128, i.e. the objective function of this initial solution is *r=D/L*=586/128=4.5.

Step 2: Because the Route-1 does not allow for adding any nodes due to the vehicle capacity we will add the remaining optional nodes 7, 8, …, 11 only to the Route-2 between the node 1 and 6 (edge (1,6)). We reach the lowest value of r' by adding the node 7 resulting the route distance extension by 8 units and the increase of the transported volume by 25 units and the value of $r' = 586 + 8/128 + 25 = 594/153 = 3.8$. This particular value of r' is the lowest possible compared to all the values of r' being result of adding one of the

remaining optional nodes between the nodes 1 and 6. Therefore we accept this change to the Route-2 and the result is Route-2 in form 1-7-6-1 with $D = 594$, $L = 153$ and $r = 3.8$. Repeat step 2.

Step 2: The Route-2: 1-7-6-1 still has vehicle capacity for one of the nodes 8, 9, 10, 11 to be added. We will try to add one of them between the nodes 1 and 7 – edge (1,7), between the nodes 7 and 6 – edge (7,6), between the nodes 6 and 1 – edge (6,1). We found the lowest value of r' by inserting the node 8 between the nodes 7 and 6 (edge (7,6)). The resulting form of the Route-2 is then 1-7-8-6-1 extending the route distance by 9 units and the transported volume by 32 units. The value of r' is $r' = 594 + 9/153 + 32 = 603/185 = 3.25 < r$. Therefore we accept this change to the Route-2 and set $r := 3.25$ and repeat step 2.

Step 2: After adding the node 8 to the Route-2 the vehicle capacity has been reached. It is not possible to add more of the remaining optional nodes. Hence let us approach to the step 3.

Step 3: The remaining optional nodes 9, 10 and 11 could be eventually used to design a new vehicle route. However, based on the high increase of route distance and the low increase of the transported volume the objective function value *r*, given as the ratio of the latter, would grow. As our goal is to find a solution with the minimal value of *r* there is no reason to design such a route. We conclude the heuristics by stating that the remaining optional nodes 9, 10 and 11 will not be included in the solution and it is fully represented by the two following routes: 1-3-2-4-5-1 and 1-7-8-6-1. The transported volume is $L = 185$, the total route distance $D = 603$ and the objective function is $r = 603/185 = 3.25$.

6 Conclusion

VRP studied in the paper is based on everyday goods distribution problem at which goods can be divided into urgent delivery and non-urgent delivery, i.e. mandatory and optional nodes of vehicle routing problem. Object function represents mean costs per unit of transported goods. So, mathematical model is nonlinear with linear constraints and linear-fractional object function. The model is transformed to linear model using Charnes-Cooper transformation. Problem is NP hard therefore a modification of insert heuristic is proposed.

7 References

[1] BARROS, A. I. (1998). *Discrete and fractional programming techniques for location models*. Dordrecht: Kluwer Academic Publishers, DOI: 10.1007/978-1-4615-4072-4_1.

[2] BRAYSY, O. & GENDREAU, M. (2005) The vehicle routing problem with time windows, Part I. Route construction and local search algorithms. *Transportation Science* 39(1), 104–118, DOI: 10.1287/trsc.1030.0056.

[3] DESROCHERS, M., DESROSIERS, J., & SOLOMON, M. M. (1992). A new optimization algorithm for the vehicle routing problem with time windows. *Operations Research* 40, 342–354, DOI: 10.1287/opre.40.2.342.

[4] MARTOS, B. (1975). Nonlinear programming: Theory and methods. Amsterdam-Oxford: North-Holland Publishing, DOI: 10.1002/zamm.19760561116.

[5] LAPORTE, G. (1992). The vehicle routing problem: An overview of exact and approximate algorithms. *European Journal for Operations Research* 59, 345–358, DOI: 10.1016/0377-2217(92)90192-C.

[6] PELIKÁN, J. VRP with loading time window. In: *37th International Conference on Mathematical Methods in Economics 2019* [online]. České Budějovice, 11.09.2019 – 13.09.2019. České Budějovice: Ekonomická fakulta, Jihočeská univerzita v Českých Budějovicích, 2019, s. 20–24. ISBN 978-80-7394-760-6.

Evaluation of Financial Performance of the Textile Industry Companies in the Hradec Králové and Liberec Regions

Natalie Pelloneová[1]

Abstract. The aim of this paper is to evaluate the financial performance of business entities in the textile industry in the Hradec Králové and Liberec regions and to find out whether the business entities in these regions have different financial performance. The textile industry has been chosen because of its importance in the Hradec Králové and Liberec regions in the past. The analysis focuses on assessing the financial performance of textile enterprises doing business in the CZ-NACE 13100, 13200, 13300 and 13900 classes between years 2010 and 2017. For evaluating the financial performance, there were used number of employees and total assets as inputs. The output and the measure of financial performance was an EVA. The EVA was constructed according to the methodology of the Ministry of Industry and Trade. Using DEA model, there was determined a rate of technical efficiency for each business entity. For both samples, the Malmquist index (MI) values were calculated. With the help of MI, a total factor productivity change and its division into a technological change and technical efficiency change was calculated. In the last step the comparison of business entities in both regions was made.

Keywords: textile industry, financial performance, economic value added, DEA

JEL Classification: C61, L25, L67
AMS Classification: 90B90, 90C90

1 Introduction

Textile manufacturing and the textile industry are among the traditional industries of the Czech economy. The development of the textile industry in the Czech lands began as early as the late 18th century. The north and north-east of Bohemia held a dominant position among the Czech textile regions. The beginnings of the textile industry in the Hradec Králové and Liberec Regions date back to the 19th century, when several textile and engineering factories were established. The textile industry thus became the key driver of industrialisation and modernisation in both regions [9].

At present, textile manufacturing in the Hradec Králové and Liberec Regions is of national importance, and textile companies are among the most important local industrial companies. In the Hradec Králové Region, traditional textile manufacturing is present e.g. in Trutnov, Dvůr Králové nad Labem and in Náchod. In Trutnov, the beginning of the textile industry dates back to the 1860s, when the largest flax mill on the European continent was built in the town [12]. The first weaving mills that were built in Náchod at the turn of the 1970s and 1980s marked the start of the rapid growth of the town's textile industry [10]. Vamberk, another important town in this region, became famous internationally for the production of handmade lace. The largest textile companies are currently located in the Trutnov and Náchod areas. Trutnov-based JUTA a.s. is the largest textile company.

Historically, traditional textile manufacturing in the Liberec region has been present mainly in the Liberec, Rumburk and Varnsdorf areas. One of the best-known textile companies was Textilana, which was founded by Johann Liebieg in 1830. However, this company now longer exists. A similar fate befell a number of other companies such as Hedva, Tiba, Benar and Kolora. Unfortunately, the textile industry in the Liberec Region no longer carries such weight. Several smaller companies have been established in and around Liberec, which are mainly engaged in the sewing of textile products – e.g. Direct Alpine, s. r. o., Funstorm, s. r. o., Matějovský – bed linen, etc. In 2020, there were 56 textile companies registered in the Hradec Králové Region and 50 companies in the Liberec Region [2].

[1] Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, natalie.pelloneova@tul.cz.

2 Data Envelopment Analysis and Malmquist Index

Data Envelopment Analysis (hereinafter DEA) is a method that is used to evaluate the technical efficiency of production units. As a multi-criteria decision-making method, DEA is especially suitable for determining the technical efficiency of units that are mutually comparable [3]. Such units are compared with each other to determine which of them are efficient (lie on the efficiency frontier) and which are inefficient. In the case of inefficient units, it is possible to determine how the inefficient unit needs to reduce/increase its inputs/outputs in order to become efficient [8].

The shape of the efficiency frontier depends on the type of returns to scale. A distinction can be made between model that assume variable returns to scale (the efficiency frontier is formed by the convex envelope of the set of production possibilities) and those that assume constant returns to scale (the efficiency frontier is formed by a straight line) [4]. Depending on the type of returns to scale, two basic DEA models can be distinguished – CCR and BCC. Models that assume constant returns to scale are named after the authors Charnes, Cooper and Rhodes – CCR models. Models that assume variable returns to scale are named after the authors Banker, Charnes and Cooper – BCC models [14]. The technical efficiency of a company is measured relative to the other companies analysed, using an efficiency score [1]. The overall technical efficiency rate takes values ranging from 0 to 1. Technically efficient units achieve an efficiency rate of 1, for inefficient units the efficiency rate is less than 1 [5, 9].

It is also important for trends in technical efficiency to be evaluated over time. Since basic DEA models do not allow for efficiency evaluation over time, another tool needs to be used – the Malmquist Index (hereinafter MI). The MI is based on DEA models and it is an important indicator to measure changes in the relative efficiencies of DMUs in different time periods [6]. By implementing various rationalization measures, inefficient companies can become efficient over time. Conversely, some companies that are classified in the efficient category may find themselves among inefficient companies. However, since this fact is impossible to quantify using basic DEA models, it is appropriate to apply the MI. It is based on the assumption that a certain DMU consumes the vector m of inputs x^t and produces a vector s with outputs y^t. We suppose that (x^t, y^t) is an input-output pair of a certain DMU in period t and (x^z, y^z) is an input-output pair of the same unit in period z.

Different versions of the MI can be formulated [7]: it can be either input- or output-oriented, with constant, variable, non-increasing or non-decreasing returns to scale. In the input-oriented model, M_q measures the change in the efficiency of production unit q between two consecutive periods t and $t+1$ using the following formula (1).

$$M_q(x^{t+1}, y^{t+1}, x^t, y^t) = E_q T_q \tag{1}$$

Components E_q and T_q are given by equations (2), and (3). Where E_q represents the change in the relative efficiency of unit q relative to other units between periods t and $t+1$. The change in the production possibility frontier is expressed by T_q as a result of the development in technology between periods t and $t+1$.

$$E_q = \frac{D_q^{t+1}(x^{t+1}, y^{t+1})}{D_q^t(x^t, y^t)} \tag{2}$$

$$T_q = \sqrt{\frac{D_q^t(x^{t+1}, y^{t+1}) D_q^t(x^t, y^t)}{D_q^{t+1}(x^{t+1}, y^{t+1}) D_q^{t+1}(x^t, y^t)}} \tag{3}$$

3 Data and Methodology

The present research aims to evaluate the financial performance of companies in the textile industry in the Hradec Králové and Liberec Regions. This is achieved using the data envelopment analysis approach and the Malmquist index. Financial performance evaluation is based on the Economic Value Added (EVA) indicator. In both regions, the trends in financial performance are analysed in 2010 to 2017. The following research question was formulated: Is there a difference in the financial performance of selected textile companies in these two regions?

The research took place in the period from January 2020 to April 2020 and it was implemented within research samples that were obtained through the MagnusWeb database containing data for entities based in the Czech Republic [2]. The same source was used to obtain the number of employees. The research process can be divided into the following steps:

1. Compiling a list of enterprises to be assessed – the research samples were limited to companies operating in the Hradec Králové and Liberec Regions having the following characteristics: a legal entity, an economically active company, a textile company operating (according to statistical classification) in the sectors CZ-NACE 13100, 13200, 13300 and 13900. As at 4 April 2020, 56 companies in the Hradec Králové Region and 50 companies in the Liberec Region met the above criteria.

2. Gathering financial statements – it was necessary to obtain the required data from the above companies' balance sheets and profit and loss statements for 2010 to 2017. Not all companies comply with their legal obligation to publish selected data from their balance sheets and profit and loss statements in the collection of documents. Only 32 out of 56 companies in the Hradec Králové Region and 29 out of 50 companies in the Liberec Region published their complete financial data from 2010 to 2017.

3. Definition of inputs and outputs for data envelopment analysis – as the next step, it was necessary to define inputs and outputs for the purpose of data envelopment analysis. Number of employees and total assets were used as inputs; outputs were economic value added (EVA).

4. Determining the number of employees – data on the number of employees of companies for 2010 to 2017 were also obtained from the MagnusWeb database [2]. Where an interval was indicated, the centre of the interval was used for further calculation. If no value was indicated for the given year, the last available figure was used. Where a company indicate a zero number of employees, it counted as one employee (the owner as a person working on their own account). In terms of the number of employees and the amount of total assets, both research samples included 16 micro-enterprises, 22 small companies, 18 medium-sized companies and 5 large companies.

5. Calculating economic value added (EVA) – for companies with available financial statements, the values of the EVA indicator were calculated according to the methodology of the Ministry of Industry and Trade, see formula (4). Where *ROE* means return on equity, r_e means alternative cost of equity and *E* means equity. This calculation seems to be quite simple but the problem lies in the calculation of the alternative cost of equity r_e. To solve this problem, the INFA model is applied. The r_e can be calculated with the help of formula (5). Where r_f is risk free rate, $r_{company}$ is premium for business risk, r_{finstr} is premium for the risk arising from the capital structure, $r_{finstab}$ is premium for financial stability risk, r_{la} is premium for the insufficient liquidity of the share.

$$EVA = (ROE - r_e)E \tag{4}$$

$$r_e = r_f + r_{company} + r_{finstr} + r_{finstab} + r_{la} \tag{5}$$

The EVA indicator can only be meaningfully calculated for companies with a positive equity value. Therefore, companies with a negative equity value were excluded from the comparison. The research sample of companies for comparison was thus reduced to 30 companies in the Hradec Králové Region and 27 in the Liberec Region.

6. Determination of the values of the distance functions and Malmquist index – as the next step, the values of the distance functions and then the individual components of the Malmquist index were determined according to formulas (2) to (3). The value of the Malmquist index was calculated according to the formula (1). The MaxDEA 7 Ultra software was used to calculate the values of the distance functions.

7. Comparing the financial performance of textile companies in both regions – the financial performance of selected companies was compared using the non-parametric Kolmogorov-Smirnov test. The non-parametric test was chosen because it had been confirmed using the Shapiro-Wilk test that the values of the different variables did not have a normal distribution. Statistical testing was performed at a significance level of 5% using STATGRAPHICS Centurion XVIII.

4 Research Results and Discussion

The values of the Malmquist index for both research samples are indicated in Tables 1 and 2. Both tables show that the Malmquist index values rather fluctuated in the different years. At the same time, Tables 1 and 2 show that the overall change in the financial performance of textile companies over the period 2010–2017 is different in each of the two regions. The reasons for the change in financial performance are also different in each of the two regions.

Following an increase in 2010–2011, the sample of textile companies in the Hradec Králové Region experienced a decrease in 2011–2012. After that, financial performance increased in 2012–2014 and then, rather surprisingly, declined sharply in 2014–2015. In the period 2015–2016, performance slightly increased and

then, towards the end of the period under review, performance declined moderately in 2015–2016. Table 1 shows that companies in the Hradec Králové Region experienced their highest increase in performance in the period 2013–2014, when financial performance increased by almost 94% year-on-year. Performance growth was driven solely by improvement in the companies' internal efficiency. Internal efficiency increased by an average of 140% year-on-year.

In the second sample of textile companies in the Liberec Region, the value of the Malmquist Index declined in 2010–2012, which was followed by an increase in financial performance in 2012–2015 and yet another decline in 2015–2017. Table 2 shows that companies in the Liberec Region experienced their highest increase in performance in the period 2012–2013, when financial performance increased by almost 11% year-on-year. Performance growth was driven mainly by an improvement in the internal efficiency of member companies, namely an improvement of an average of 7% per year.

It can be assumed that the Czech Republic's overall economic situation may have – one way or another – affected the trends in the financial performance of textile companies in both regions. Between 2010 and 2011, the Czech economy experienced GDP growth. In the two years that followed, the Czech economy was in recession, and since 2014 it has been growing again. The sample of textile companies shows a decline in financial performance between 2011 and 2012. Following a slight improvement in performance in the period 2012–2014, companies stagnated or experienced a rather declining trend.

Period	M_q	E_q	T_q
2011/2010	1.0920	0.9827	1.1112
2012/2011	0.8752	0.7695	1.1373
2013/2012	1.3337	1.5177	0.8788
2014/2013	1.9389	2.3938	0.8100
2015/2014	0.5661	0.6003	0.9431
2016/2015	1.0015	0.7599	1.2291
2017/2016	0.9340	1.0383	1.0070
AVG	**1.0392**	**1.0383**	**1.0070**

Table 1 Malmquist index summary of annual means in the Hradec Králové region

Period	M_q	E_q	T_q
2011/2010	0.8665	0.8687	0.9974
2012/2011	0.9833	0.9982	0.9851
2013/2012	1.1056	1.0718	1.0315
2014/2013	1.0472	1.1693	0.8956
2015/2014	1.0520	0.9639	1.0914
2016/2015	0.8480	1.1437	0.7415
2017/2016	0.9780	0.9870	0.9909
AVG	**0.9788**	**1.0242**	**0.9556**

Table 2 Malmquist index summary of annual means in the Liberec region

It can be concluded that the financial performance of textile companies operating in the Hradec Králové Region increased year-on-year in the period 2010–2017. By contrast, the sample of textile companies in the Liberec Region experienced a decline in financial performance in 2010–2017. In 2010–2017, the financial performance of textile companies in the Hradec Králové Region increased by an average of 4% per year, while in the Liberec Region it decreased by an average of 2% per year. Tables 1 and 2 also show that the financial performance of textile companies was mainly driven by a change in internal technical efficiency. Textile companies in the Hradec Králové Region showed an increase in internal technical efficiency by an average of 4% per year, as well as a very modest increase in technological change. While textile companies in the Liberec Region also experienced an increase in internal technical efficiency, this only amounted to an

average of 2% per year. In contrast, the technological change of companies in the Liberec Region decreased by 5% year-on-year, which can be explained by the fact that textile companies in the Liberec Region have reduced their innovation.

Since the Shapiro-Wilk test confirmed that none of the different variables had a normal distribution, the differences were assessed using the non-parametric Kolmogorov-Smirnov test. Table 3 provides information on the significance of the differences in the magnitude of the MI and its E_q and T_q components between the two regions. For the period 2010–2017, it can be concluded that there are significant differences between the Hradec Králové and Liberec Regions in terms of both overall performance and the change in the E_q and T_q components.

Year	Mq	Eq	Tq
2011/2010	0.4259	0.7148	0.4222
	0.0115	**<0.0001**	**0.0126**
2012/2011	0.3666	0.4370	0.6000
	0.0438	**0.0087**	**<0.0001**
2013/2012	0.2888	0.3444	0.3370
	0.1866	0.0686	0.0792
2014/2013	0.3259	0.3037	0.3888
	0.0976	0.1454	**0.0271**
2015/2014	0.2888	0.4740	0.5000
	0.1866	**0.0033**	**0.0016**
2016/2015	0.4333	0.5370	0.7000
	0.0096	**0.0005**	**<0.0001**
2017/2016	0.3555	0.7074	0.6740
	0.0550	**<0.0001**	**<0.0001**
2010–2017	0.3222	0.3423	0.4529
	0.0004	**<0.0001**	**<0.0001**

Table 3 Kolmogorov-Smirnov test, DN statistic (P-Values) – Hradec Králové vs. Liberec region

5 Conclusion

The aim of the research was to evaluate the financial performance of companies operating in the textile industry in the Hradec Králové and Liberec Regions and compare them in order to determine whether there were differences in financial performance between companies in the two regions. The research was carried out on these two sample regions because their tradition in the textile industry dates back more than a century and – despite the decline in traditional textile manufacturing in the 1990s – it is still characterised by an above-average concentration of employment in this sector. For this purpose, 56 textile companies from the Hradec Králové Region and 50 companies from the Liberec Region were selected and their financial statements for the reference period 2010–2017 were analysed. For these years and for all companies selected, economic performance data were gathered for the accounting periods and the values of the EVA indicator were calculated according to the Ministry of Industry and Trade methodology. Data envelopment analysis and the Malmquist index were used in order to allow for a comparison of the various enterprises.

A total 49 companies for which it was not possible to obtain complete financial statements or which reported a negative value of equity in any year within the period under review had to be excluded from the comparison of the two regions, thus reducing the size of the research sample to 30 enterprises in the Hradec Králové Region and 27 enterprises in the Liberec Region. It is also important to note that there were six textile companies operating in both of the regions, which are members of the CLUTEX – cluster of technical textiles. Five of the companies evaluated were based in the Hradec Králové Region and one in the Liberec Region. The following companies from the Hradec Králové Region were included in the research sample: Z I T E X s.r.o., VEBA, textilní závody a.s., NYKLÍČEK a spol. s.r.o., MILETA a.s. and GRUND a.s. From the Liberec Region, the following company was included: Výroba stuh – ELAS. To some extent, this fact may have

contributed to the regions' overall financial performance. Indeed, as some experts [13] point out, clusters can make a significant contribution to improving the performance of the participating companies which, in turn, can better contribute to the performance of the regional economy.

As part of applying the DEA method, the average technical efficiency score was calculated for the period 2010–2017, and it was concluded that textile companies in the Hradec Králové Region achieved an average score of 0.3856, while the average score of companies in the Liberec Region was 0.3439. In order to achieve technical efficiency, textile companies would have to reduce their inputs by 61% in the Hradec Králové Region and by 66% in the Liberec Region. Based on the application of the Malmquist index, it can be concluded that textile companies in the Hradec Králové Region experienced a more robust increase in financial performance as compared to textile companies in the Liberec Region. The improvement in the financial performance of companies in the Hradec Králové Region was mainly due to an improvement in internal technical efficiency, namely an improvement of almost 4% per year, while technological change and innovations stagnated in this region. Also, enterprises in the Hradec Králové Region responded better to the aftermath of the 2012 and 2013 recession. In contrast, textile companies in the Liberec Region showed a 2% decline in financial performance per year in the period under review. In the case of textile companies in the Liberec Region, the decline in performance was caused solely by technological decline (a year-on-year decrease of 4%). On the other hand, internal technical efficiency grew by 2% year-on-year.

Based on the research, it can be concluded that the financial performance of textile companies in the Hradec Králové and Liberec Regions followed different trends. This fact was also confirmed by the Kolmogorov-Smirnov test that was performed and that confirmed statistically significant differences in the values of the MI and both of its components at a significance level of 5%. The different financial performance of textile companies in the two regions may have resulted from several factors. One of them is the fact that there are five textile companies that are based in the Hradec Králové Region and that are also members of the CLUTEX textile cluster. Many experts view cluster organisations as tools to improve the financial and innovation performance of companies. This might have had a positive impact on more significant technological changes and innovations in the Hradec Králové Region. The reason may be the fact that the very important company JUTA a.s. is based in the Hradec Králové Region and, in turn, may have significantly affected the performance of this region.

Acknowledgement

Supported by Czech Science Foundation grant no. GA18-01144S "An empirical study of the existence of clusters and their effect on the performance of member enterprises".

References

[1] BANKER, R. D., CHARNES, A., & COOPER, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9), 1078–1092.

[2] BISNODE. (2020). Magnusweb: Komplexní informace o firmách v ČR a SR [online]. Bisnode ČR, Prague, 2020 [cit. 2020-02-14]. Available at: https://magnusweb.bisnode.cz.

[3] CAVES, D. W., CHRISTENSEN, L. R., & DIEWERT, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 50(6), 1393–1414.

[4] CHARNES, A., COOPER, W. W., LEWIN, A. Y., & SEIFORD, L. M. (1994). *Data Envelopment Analysis: Theory, Methodology, and Application*. New York: Springer Science+Business Media.

[5] CHARNES, A., COOPER, W. W., LEWIN, A. Y., & SEIFORD, L. M. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.

[6] FÄRE, R., GRIFELL-TATJÉ, E., GROSSKOPF, S., & KNOX LOVELL, C. A. (1997). Biased Technical Change and the Malmquist Productivity Index. *The Scandinavian Journal of Economics*, 99, 119–127.

[7] JABLONSKÝ, J. & DLOUHÝ, M. (2004). *Modely hodnocení efektivnosti produkčních jednotek*. Prague: Professional Publishing.

[8] JI, Y. & LEE, C. (2010). Data Envelopment analysis. *The Stata Journal*, 10(2), 267–280.

[9] KRÁLOVÉHRADECKÝ KRAJ. (2018). Stručné dějiny Královéhradeckého kraje [online]. Královéhradecký kraj, 2018 [cit. 2020-03-10]. Available at: http://www.kr-kralovehradecky.cz/cz/kraj-volene-organy/kralovehradecky-kraj/historie/strucne-dejiny-kralovehradeckeho-kraje-35/.

[10] MĚSTO NÁCHOD. (2017). Náchod – město na česko-kladském pomezí [online]. Město Náchod, 2017 [cit. 2020-02-20]. Available at: https://www.mestonachod.cz/zivot-v-nachode/o-meste/historie/.

[11] RAY, S. C. (2004). *Data envelopment analysis: theory and techniques for economics and operations research*. New York: Cambridge University Press.
[12] ŠNAJDAR, H. (2014). Sláva a konec textilního průmyslu v Trutnově aneb Od prosperity k úpadku [online]. TN Média s.r.o., 2020 [cit. 2020-03-10]. Available at: https://trutnovinky.cz/zpravy/aktuality/2014/brezen/slava-a-konec-textilniho-prumyslu-v-trutnove-aneb-od-prosperity-k-upadku/.
[13] VITURKA, M. (2010). *Kvalita podnikatelského prostředí, regionální konkurenceschopnost a strategie regionálního rozvoje České republiky*. Praha: Grada Publishing.
[14] ZHU, J. (2014). *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets*. Cham: Springer International Publishing.

Use of Malmquist Index in Evaluating Financial Performance of Companies in Cluster of Czech Furniture Manufacturers

Natalie Pelloneová[1]

Abstract. This contribution deals with the influence of membership of a business entity in a cluster organization (CO) on its financial performance. The aim is to verify the hypothesis that the membership in a CO is connected with increasing financial performance of its members in time. In the research sample are included 17 members of the Cluster of Czech Furniture Manufacturers. The data are collected for years 2012–2017. For the assessment of financial performance, the DEA method is applied with two inputs and one output. For inputs were chosen the number of employees and the long-term capital, the economic value added was used as the output. For each company from the sample and periods, the Malmquist index values were calculated. With help of the Malmquist index it was possible to quantify the total factor productivity change and to decompose it to technological change and technical efficiency change. The development of the indicators was monitored in time. In conclusion, the results of the research are discussed. The research has shown an improvement in financial performance of companies in the Cluster of Czech Furniture Manufacturers.

Keywords: Malmquist index, cluster organization, economic value added, financial performance, cluster of Czech furniture manufacturers

JEL Classification: C61, L25, L68
AMS Classification: 90B90, 90C90

1 Introduction

Technical publications dealing with regional development or network economics use a variety of terms to describe agglomerations of economic entities. These may include e.g. clusters, industrial districts, or networks. The first mention of agglomerations is traditionally attributed to Alfred Marshall [12], who observed the concentration of specialized industries within a confined geographical area in Great Britain and termed these as industrial districts. However, the development of the term "cluster" did not begin until the early 1990s, when Michael Eugene Porter's ground-breaking book entitled "The Competitive Advantage of Nations" was published [3]. In this book, Porter formulated the first definition that drew clusters to the broader attention of not only the scientific community and is considered one of the most influential. Here, Porter [18] defines a cluster as a geographically proximate group of interconnected companies, suppliers, and associated institutions in a particular field as well as companies in related fields that compete and also cooperate with each other.

After more than twenty years, clusters are still an important topic for economists and economic policy makers. The concept of clusters is used to drive synergies and innovations in technology among major innovators – including entrepreneurs, universities, and research organizations – in order to support growth in regions and, in turn, entire countries. Due to their focus on activities related to research, development and innovation, clusters are considered to be the best and most effective solution to promote regional development, increase employment and support the implementation of new technologies. In the context of clusters, the concept of performance is very often used. As stated in the White Paper on cluster initiatives [1]: Clusters enable participating businesses to achieve a higher performance level and, in turn, improve their competitiveness. Many authors leave the question of cluster performance measurement and evaluation unanswered. During the formation of clusters, various experts and institutions defined various models that can be used to measure the performance of a cluster organization. These include e.g. the Cluster Initiative Performance Model, which describes performance using the social, political, and economic pillars. Another modern approach to measuring cluster performance is the "Cluster Benchmarking Project" implemented by the Nordic Innovation Centre.

[1] Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, natalie.pelloneova@tul.cz

The present paper examines the effect of business entities' membership in a selected cluster organization on their financial performance. It aims to evaluate the effect of the Cluster of Czech Furniture Manufacturers on its member business entities' financial performance, and to use the Malmquist Index to confirm the assumption that business entities' membership in the cluster organization translates into improved financial performance in a time series.

The DEA method is used to evaluate financial performance in research of several authors [11;17]. Ozcan and McCue [14] present an innovative method of measuring and assessing financial performance for hospitals using data envelopment analysis (DEA). Xu, Hui and Li [19] evaluate financial performance of coal companies using the DEA method. For example, Parupati and Chary [15] aims to analyze the time series performance of 40 IT companies based on DEA methodology. Operating expenses, total outside liabilities and shareholder funds are used as inputs for the DEA model. Net sales and cash flow from operating activities are used as outputs.

2 Malmquist index

The index was created by Sten Malmquist in 1953 to measure productivity. It was subsequently expanded and improved by several authors, including Forsund [9], Forsund and Hjalmarsson [10] and Fare el al. [8]. The use of the MI has several advantages. First, the index is used as part of the non-parametric DEA approach. Second, the index can be decomposed into two components, namely the change in internal technical efficiency and the change in technology. These properties make the MI a very popular tool to evaluate changes in efficiency. It is an important indicator to measure changes in the relative efficiencies of DMUs in different time periods [4].

Different versions of the MI can be formulated: it can be either input- or output-oriented, with constant, variable, non-increasing or non-decreasing returns to scale. In the input-oriented model, MI_q measures the change in the efficiency of production unit q between two consecutive periods t and $t+1$ using the following formula (1):

$$MI_q(x^{t+1}, y^{t+1}, x^t, y^t) = EFFCH_q TECH_q \tag{1}$$

Where x^t denotes inputs in t period, y^t denotes outputs in t period, x^{t+} denotes inputs in $t+1$ period, y^{t+} denotes outputs in $t+1$ period. Where D_q is production unit efficiency, $EFFCH_q$ represents the change in the relative efficiency of unit q relative to other units between periods t and $t+1$. The change in the production possibility frontier is expressed by $TECH_q$ as a result of the development in technology between periods t and $t+1$. Components $EFFCH_q$ and $TECH_q$ are given by equations (2), and (3).

$$EFFCH_q = \frac{D_q^{t+1}(x^{t+1}, y^{t+1})}{D_q^t(x^t, y^t)} \tag{2}$$

$$TECH_q = \sqrt{\frac{D_q^t(x^{t+1}, y^{t+1}) D_q^t(x^t, y^t)}{D_q^{t+1}(x^{t+1}, y^{t+1}) D_q^{t+1}(x^t, y^t)}} \tag{3}$$

If the input-oriented MI_q value is greater than 1, it means improvement (progress), if MI_q is equal to 1, there has been no change in productivity, and if the resulting MI_q value is less than 1, productivity has deteriorated. If $EFFCH_q$ is greater than 1, it means that the company has improved its relative technical efficiency. If $EFFCH_q$ equals 1, there has been no change in relative technical efficiency, and if $EFFCH_q$ is less than one, the company's relative technical efficiency has deteriorated. $TECH_q$ values can be interpreted in a similar manner: a value greater than 1 indicates technology progress or innovation; a $TECH_q$ value of 1 means there has been no change in technology, and if $TECH_q$ is less than 1, there has been regression in technology.

Since one of the variables used – economic value added – can also take negative values, the modified radial measure (VRM) model was used. An input-oriented VRM model operating under VRS is shown below in (4). Where λ_j are weights of all DMUs. The constraint $\sum \lambda_j = 1$ is the condition of convexity which is kept under VRS. The modification to the model consists in using the absolute values of inputs (outputs) instead of their actual values. It can be noticed that β measures how much an observed DMU should improve in order to reach the efficient frontier, in other words it represent the inefficiency measure [5].

$$\begin{gathered} \max \beta \\ X\lambda + \beta |x_q| \leq x_q \end{gathered} \tag{4}$$

$$Y\lambda \geq y_q$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda \geq 0$$

3 Data and methodology

Established in 2006 in the legal form of a cooperative, the Cluster of Czech Furniture Manufacturers was selected for the purpose of financial performance evaluation. The cluster is based in Brno and operates mainly in the South Moravian Region. This cluster organization is the largest cluster organization focusing on the furniture and woodworking industries and interior design. Its members are mainly Czech furniture companies that are internationally competitive and well-established in foreign markets [6]. The basic data source was information from the financial statements of the member business entities of the Cluster of Czech Furniture Manufacturers for the period 2012–2017, which were obtained from the MagnusWeb database [2]. The period under review was selected taking into account the development of the cluster organization and the beginning of its operation, while allowing for the fact that the effects of cluster membership can be expected to manifest with a certain delay. Most companies have not yet published their financial results in the Commercial Register, which is why the time series ends in 2017. The research was carried out in the following steps:

1. Compiling a list of the companies to be evaluated – as the first step, a database of member business entities of the Cluster of Czech Furniture Manufacturers was created. In the period analyzed, the cluster had 27 members in total. Since the research focused on evaluating financial performance, only business entities were included. In the period analyzed, the Cluster of Czech Furniture Manufacturers had a total of 23-member business entities. In the research conducted, it is only possible to compare companies that have been members of the cluster organization for the same period of time – these are the only companies can be considered the homogeneous core of the cluster. The core of the cluster organization consists of 17 business entities in the sectors CZ-NACE 161000, 310900 and 433200.

2. Gathering financial statements – it was necessary to obtain the required data from the above business entities' financial statements for 2012–2017. The success rate of obtaining the financial statements was 100%. Financial statements were obtained for all 17 business entities for each year.

3. Calculating economic value added – for each of the business entities, the economic value added indicator (hereinafter EVA) was subsequently calculated according to the formula (5), where EVA is defined as the product of equity E and 'spread' (return on equity ROE minus alternative cost of equity r_e). EVA can take both positive and negative values. A positive EVA means that the company generates value for its owners. If EVA is negative, the value of the company decreases [13].

$$EVA = spread \cdot E$$
$$EVA = (ROE - r_e) \cdot E \tag{5}$$

The CAPM method was used to estimate the cost of equity (see formula 6). Where r_f is the risk-free rate of return, often taken as the rate of return on treasury bills published by the Ministry of Industry and Trade of the Czech Republic; β_n is the quantity used to measure the systematic risk of the asset, the values will be obtained from the Damodaran website [7]; r_m is the expected rate of return in the market. National stock indices are most often used to determine the expected rate of return in the market r_m [13].

$$r_e = r_f + \beta_n(r_m - r_f) \tag{6}$$

4. Input and output specification – number of employees and long-term capital have been selected as inputs for the DEA model. Long-term capital is the sum of the following balance sheet items: equity, long-term bonds issued and long-term bank loans. The output is EVA.

5. Determination of technical efficiency values and calculation of the Malmquist index – for each enterprise within the set, the MaxDEA 7 Ultra software was used to calculate an efficiency score D_q under VRS, and the values for distance functions and each component of the Malmquist index were determined using formulas (2) and (3). Finally, the value of the Malmquist index was calculated using formula (1).

4 Research results

The cluster's average efficiency score (D) for each years is shown in Table 1 (the arithmetic mean was used for the calculation). For the entire period under investigation, the average technical efficiency score for companies within the Cluster of Czech Furniture Manufacturers was not high (0.3306). Out of a total of 17 business entities, 2 to 5 companies were at the efficiency frontier each year. Two business entities remained at the efficiency frontier throughout the entire period under review, namely the companies brychta.org and drevotvar.com. The third relatively successful company was Kronospan CR, which was classified as efficient in all years except 2013. The largest number of efficient business entities was identified in 2014 (5 in total). Within the period under review, the highest average technical efficiency score was also achieved in 2014. The trends in technical efficiency may have been influenced by economic cycles. In 2012 and 2013, the Czech economy was in recession, and it was in these years that – according to Table 1 – the lowest average technical efficiency scores within the entire period under review were achieved. However, this is difficult to prove scientifically as a fact. Interesting changes in trends were observed for the following companies: BLANÁŘ NÁBYTEK, a.s, FMP-lignum, manufacturing cooperative, GERBRICH s.r.o., TAUBENHANSL s.r.o. and Trachea, a.s.

Company	2012	2013	2014	2015	2016	2017	AVG
BLANÁŘ NÁBYTEK, a.s	0.070	0.007	1.000	0.299	0.004	0.004	**0.231**
brychta.org cooperative	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**
DŘEVODÍLO Rousínov, manufacturing cooperative	0.238	0.723	0.561	0.592	0.913	0.339	**0.561**
Dřevojas, manufacturing cooperative	0.034	0.048	0.033	0.167	0.029	0.025	**0.056**
Dřevotvar cooperative	0.016	0.036	0.019	0.018	0.015	0.016	**0.020**
drevotvar.com cooperative	1.000	1.000	1.000	1.000	1.000	1.000	**1.000**
Dřevozpracující cooperative	0.003	0.003	0.002	0.003	0.003	0.003	**0.003**
FMP-lignum, manufacturing cooperative	0.093	0.335	0.282	0.426	0.664	1.000	**0.467**
GERBRICH s.r.o.	0.328	0.076	0.217	0.828	0.337	0.204	**0.332**
JITONA a.s.	0.003	0.005	0.004	0.004	0.004	0.005	**0.004**
Kili, s.r.o.	0.015	0.036	0.021	0.021	0.016	0.016	**0.021**
KRONOSPAN CR, spol. s r. o.	1.000	0.003	1.000	1.000	1.000	1.000	**0.834**
NADOP – VÝROBA NÁBYTKU, a.s.	0.037	0.046	0.036	0.040	0.035	0.263	**0.076**
Resonanční pila, a.s.	0.043	0.088	0.054	0.060	0.056	0.069	**0.062**
TAUBENHANSL s.r.o.	0.117	0.124	1.000	0.269	0.105	0.397	**0.335**
Trachea, a.s.	1.000	0.040	0.087	0.490	0.186	0.270	**0.345**
V L A B O, s.r.o.	0.158	0.417	0.321	0.280	0.222	0.239	**0.273**
AVG	**0.303**	**0.235**	**0.390**	**0.382**	**0.329**	**0.344**	**0.331**

Table 1 Pure technical efficiency scores

Table 2 also provides an overview of trends in the average values for MI and its components for the cluster organisation in the period from 2012 to 2017. Given that the MI is constructed as a multiplicative model, the geometric mean was used to calculate these average values. Table 2 shows that the MI values rather fluctuated in the different years. The sample of the cluster's member companies shows a decline in financial performance in 2012–2013 (may be attributable to the recession). The financial performance of member companies decreased by 43%. The decrease in performance resulted from a decrease in both components of the MI. This was followed by a relatively robust growth in financial performance in 2013–2015. Surprisingly, financial performance then declined again in 2015–2016 and picked up again towards the end of the period under review. Within the entire period under review, the biggest year-on-year increase in the MI value (an increase of 81%) took place in 2013–2014. The index decomposition then shows that, in the same period, the technical efficiency of cluster members increased by 94% on average, while there also was a 7% technical regress.

Period	MI	EFFCH	TECH	GDP	IPI
2013/2012	0.5661	0.7085	0.7991	0.9952	1.0407
2014/2013	1.8145	1.9367	0.9369	1.0272	1.0443
2015/2014	1.6225	1.1862	1.3678	1.0531	1.0516

Period	MI	EFFCH	TECH	GDP	IPI
2016/2015	0.6975	0.5904	1.1815	1.0245	1.0583
2017/2016	1.0696	1.2085	0.8850	1.0435	1.0417
Geomean	1.0445	1.0303	1.0138	1.0285	1.0473

Table 2 Malmquist index summary of annual means in the Cluster of Czech Furniture Manufacturers

The reasons underlying the fluctuation in the MI and its components are not entirely clear. The Czech Republic's economic situation (see GDP column) may have contributed to changes in financial performance. Until 2014, the Czech Republic was in recession, which may have also affected – to some extent – the cluster's individual member companies. Since 2014, the Czech Republic has experienced economic growth and, since that year, the trends in the MI also point to an improvement in the financial performance of the member companies. The fact that financial performance declined in the penultimate reporting period 2015–2016 was a relatively surprising finding. Even though the Czech Republic experienced economic growth in that period, the GDP growth rate slowed down from 5.3% to 2.5%. Another possible reason for fluctuations in MI values may be individual changes in the member companies' performance.

The last row of Table 2 below indicates the total MI value, which shows the overall change in the member companies' financial performance. Overall performance has increased by an average of approximately 4% per year. This increase also corresponds to the average industrial production index (see IPI column) in the field of furniture manufacturing. In 2012–2017, the IPI showed a very similar growth rate. Table 2 also shows that the performance of companies was driven both by technological progress (innovation) and by improvement in the internal efficiency of member companies. The improvement in the cluster members' technical efficiency contributed more (3% per year) while the technological change only accounted for about 1% per year.

5 Conclusion

The paper dealt with the evaluation of the trends in the financial performance of 17 members of the Cluster of Czech Furniture Manufacturers in the 2012–2017 reference period. The paper used DEA analysis, namely the input-oriented BCC model expanded to include the Malmquist index, as a tool to compare the trends in efficiency over time. The number of employees and the amount of long-term capital were selected as inputs. First, an individual analysis of the trends in the efficiency of selected cluster members was performed and possible causes of significant changes in the trends were discussed.

It can be concluded that the financial performance of the companies within the Cluster of Czech Furniture Manufacturers increased over the 2012–2017 period under review. Technical efficiency – i.e. the ability to efficiently transform inputs into outputs – increased by 3% year-on-year. In addition, there was technological change and innovation as well, although their growth of 1% was rather modest. This supports the claim that business entities' membership in a cluster organization translates into their improved financial performance in a time series. Furthermore, the research confirmed the assumption that cluster organizations have a positive effect on innovation, as formulated in previous research [16]. The fact that financial performance declined in the penultimate reporting period 2015–2016, despite the Czech Republic's continued economic growth, was a relatively surprising finding. In both samples of companies, this decline was caused by a deterioration in technical efficiency, while technological progress stagnated.

In comparison with previous research [17], it can be concluded that the impact of a cluster organization varies across industries and the impact of the different components of the MI also varies. In the aerospace industry [17], research did not confirm a positive impact of business entities' membership in a cluster organization on their financial performance.

While the reference period for this study ends in 2017, the author of the paper will continue the research in subsequent years and will expand it to include additional sectors.

Acknowledgement

Supported by Czech Science Foundation grant no. GA18-01144S "An empirical study of the existence of clusters and their effect on the performance of member enterprises".

References

[1] ANDERSSON, T., SERGER, S. S., SORVIK, J., & HANSON, W. E. (2004). *The Cluster Policies White Book*. Malmo, Sweden: International Organization for Knowledge Economy and Enterprise Development.

[2] BISNODE. (2020). Magnusweb: Komplexní informace o firmách v ČR a SR [online]. Bisnode ČR, Prague, 2020 [cit. 2020-02-14]. Available at: https://magnusweb.bisnode.cz.

[3] BOJA, C. (2011). Clusters Models, Factors and Characteristics. *International Journal of Economic Practices and Theories*, 1(1), 34–43.

[4] CAVES, D. W., CHRISTENSEN, L. R., & DIEWERT, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 50(6), 1393–1414.

[5] CHENG, G., ZERVOPOULOS, P., & QIAN, Z. (2013). A variant of radial measure capable of dealing with negative inputs and outputs in data envelopment analysis. *European Journal of Operational Research*, 225, 100–105.

[6] CLUSTER OF CZECH FURNITURE MANUFACTURERS. (2014). About us [online]. Cluster of Czech Furniture Manufacturers, 2014 [cit. 2020-03-18]. Available at: http://www.furniturecluster.cz/about-us-EN122.

[7] DAMODARAN, A. (2019). Archived data – Levered and Unlevered Betas by Industry [online]. Stern School of Business, New York [cit. 2019-08-07]. Available at: http://pages.stern.nyu.edu/~adamodar/New_Home_Page/dataarchived.html.

[8] FARE ET AL. (1992). Productivity changes in Swedish pharmacies 1980–1989: A non-parametric Malmquist approach. *Journal of Productivity Analysis*, 3(12), 85–101.

[9] FORSUND, F. R. (1990). *The Malmquist productivity index*. Memorandum from Department of Economics, University of Oslo.

[10] FORSUND, F. R. & HJALMARSSON, L. (1974). On the measurement of productive efficiency. *Swedish Journal of Economics*, 76, 141–145.

[11] JAVADIAN, R., RAMEZANZADEH, S., FARD, P. F., ROZEII, M., & HAGHU, M. (2019). Assessing Financial Performance of NAJA Hospitals Using Data Envelopment Analysis Technique. *Journal of Police Medicine*, 8(4), 155–160.

[12] MARSHALL, A. (1920). *Principles of Economics*. London: Macmillan.

[13] MPO. (2017). Finanční analýza podnikové sféry za rok 2017 [online]. Ministerstvo průmyslu a obchodu, Prague, 2017 [cit. 2018-11-28]. Available at: https://www.mpo.cz/cz/rozcestnik/analyticke-materialy-a-statistiky/analyticke-materialy/financni-analyza-podnikove-sfery-za-rok-2017--237570/.

[14] OZCAN, Y. A. & MCCUE, M. J. (1996). Development of a Financial Performance Index for Hospitals: DEA Approach. *JORS*, 47(1), 18–26.

[15] PARUPATI, S. K. & CHARY, T. S. (2017). Financial Performance Evaluation of IT Industry through DEA Window Analysis Approach. *Journal of Management*, 4(2), 40–49.

[16] PAVELKOVÁ, D. (2009). *Klastry a jejich vliv na výkonnost firem*. Praha: Grada Publishing.

[17] PELLONEOVÁ, N. & ŠTICHHAUEROVÁ, E. (2019). Influence of Membership in the Moravian Aerospace Cluster on the Financial Performance of its Members: Malmquist Index Approach. In *Proceeding of the 37th International Conference on Mathematical Methods in Management 2019* (pp. 275–280). České Budějovice: University of South Bohemia in České Budějovice.

[18] PORTER, M. E. (1998). Clusters and the New Economics of Competition. *Harvard Business Review*, 76(6), 77–90.

[19] XU, Q., HUI, L., & LI, O. (2013). The Financial Performance Assessment on the Listed Coal Companies by Using DEA Model. *Advanced Materials Research*, 869, 603–611.

Two Objective Problems of Time and Capacity Overloaded Centers

Štefan Peško[1], Zuzana Borčinová[2]

Abstract. We study the problem motivated by a task of placing a limited number of ambulances providing health services, with the effort to exceed critical ambulance arrival time only in exceptional cases. The basic model of overloaded centers detects in a transport network of q-service centers with p-vehicles (with given capacity of customers/vehicles) with goal to minimize the number of inaccessible customer locations (those with exceeded critical arrival time from an assigned center). In the other two models we consider the fairest/fair approach to inaccessible customers. First we formulate a lexicographic problem, where the first criterion is the number of inaccessible customers, and the second is the total amount of time of customers that are unavailable, with capacity constraints. Next we propose its heuristic version, where we reduce the space of feasible solutions. We present the possibility to solve these problems exactly and heuristically (MILP) using the real data from Slovak regions for the task of placing ambulance of rescue service.

Keywords: overloaded centers, fair approach, MILP, Slovak regions

JEL Classification: 90C10, 90C15
AMS Classification: C02, C61, C65, C68

1 Introduction

We deal with solving a problem motivated by the task of placing a limited number of ambulances providing health services, with the effort to exceed critical ambulance arrival time only in exceptional cases. This issue is being intensively addressed by my departmental colleagues. I will mention at least some of them Janáček and Kvet in [2],[3],[6], Jánošíková, Gábrišová and Ježek [4], [5], Borčinová, Majer [7],[8].

A similar problem can occur in different services, not just in rescue ones.

Our basic model of overloaded centers detects in a transport network of q-service centers with p-vehicles (with given capacity of customers / vehicles) with goal to minimize the number of inaccessible customer locations (those with exceeded critical arrival time from an assigned center, see figure 1). In the other two

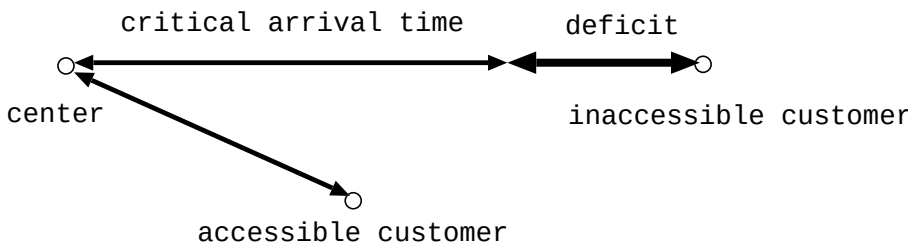


Figure 1 Accessible and inaccessible customers

models we consider the fairest/fair approach to inaccessible customers. First we formulate a lexicographic MILP, where the first criterion is the number of inaccessible customers, and the second is the total amount of time customers are unavailable, with capacity constraints. Next we propose its heuristic MILP version, where the capacity of conditions of the vehicles are relaxed.

2 Notation

We will use following notation in presented models. The transport network is represented by a sparse matrix $T = (t_{ij}), i \in I, j \in J$ where

[1] University of Žilina, Univerzitná 8215/1, 010 026 Žilina, Slovakia, stefan.pesko@fri.uniza.sk
[2] University of Žilina, Univerzitná 8215/1, 010 026 Žilina, Slovakia, zuzana.borcinova@fri.uniza.sk

- J is a set of customers where $n = |J|$ is a number of places (nodes) in the network,
- I is a set of feasible centers, $I \subset J$,
- t_{ij} is driving time from center i to customer j in minutes,
- w_j is weight of a customer $j \in J$,
- p is a number of vehicles,
- κ is capacity of vehicles,
- τ is critical arrival time in minutes,
- $A_i = \{j \in J : t_{ij} \leq \tau\}$ is a set of accessible customers from center $i \in I$.

We define for $i \in I, j \in A_i$ the binary variable

$$x_{ij} = \begin{cases} 1 & \text{if accessible customer } j \text{ is assigned to center } i, \\ 0 & \text{otherwise} \end{cases}$$

In addition, the value of the variable $x_{ii} = 1$ means that i is the center.

The variable $y_i \in \{0, 1, \ldots, q\}, i \in I$ indicates the number of vehicles in the center i. The selection variable $z_j \in \{0, 1\}, j \in J$ decides whether is a customer accessible $z_j = 0$ or inaccessible $z_j = 1$.

We can now approach the formulation of a basic model of customer time availability

3 Model of customer time availability

In this model, we will consider three parameters only, the number of required centers q, the critical arrival time τ and weights of the customers w . The aim is to find at most q centers covering a set of all customers with a minimum weighted number of inaccessible customers as the first criterion and a minimum weighted travel time of accessible customers as the second criterion. This problem can be formulated as a following problem of binary programming (CTA):

$$LEX\Big(\sum_{j \in J} w_j z_j, \sum_{i \in I} \sum_{j \in A_i} w_j t_{ij} x_{ij}\Big) \to \min \tag{1}$$

s.t.

$$\sum_{i \in I} x_{ii} \leq q \tag{2}$$

$$\sum_{i \in I : j \in A_i} x_{ij} + z_j = 1 \qquad \forall j \in J, \tag{3}$$

$$x_{ij} \leq x_{ii} \qquad \forall i \in I, \forall j \in A_i - \{i\}, \tag{4}$$

$$x_{ij} \in \{0, 1\} \qquad \forall i \in I, \forall j \in A_i, \tag{5}$$

$$z_j \in \{0, 1\} \qquad \forall j \in J. \tag{6}$$

The object function (1) is a vector of total weighted number of inaccessible customers and total driven time to this customers which is minimized lexicographically. Constraint (2) ensures that it will be created maximally q centers. If j is an inaccessible customer then constraint (3) does not assign a center to the customer. If j is an accessible customer then constraint (3) assign one center. Constraints (4) allows to assign a customer to a selected center. Constraints (5) and (6) are obligatory.

Note that summation in constraints (3) is performed for pairs (i, j) defined in sparse matrix T.

The CTA model does not assign vehicles with a given capacity to centers yet. The following capacitated model solves this problem.

4 Capacitated model of customer time availability

In this model we add two parameters, the number of available vehicles p with the capacity of κ customers. We have the same aim as in the CTA but in addition we want to distribute vehicles between chosen centers. This problem can be also formulated as following problem of integer programming (CCTA):

$$LEX\Big(\sum_{j\in J} w_j z_j, \sum_{i\in I}\sum_{j\in A_i} w_j t_{ij} x_{ij}\Big) \to \min \tag{7}$$

s.t.

$$\sum_{i\in I} y_i = p, \tag{8}$$

$$\sum_{j\in A_i} w_j x_{ij} \le \kappa y_i \qquad \forall i \in I, \tag{9}$$

$$q x_{ii} \ge y_i \ge x_{ii} \qquad \forall i \in I, \tag{10}$$

$$y_i \in \{0, 1, \dots, p\} \qquad \forall i \in I, \tag{11}$$

$$(2) - (6).$$

Constraint (8) distributes given number of vehicles between chosen centers. Constraints (9) ensure sufficient capacity in the centers. Constraint (10) allows to assign at least one vehicle to each selected center and no vehicle otherwise. Last new constraint is obligatory. The disadvantage of the CCTA model is that it does not answer the question to which center should be assigned unavailable customers. This disadvantage we propose to solve via following extended capacitated model.

5 Extended capacitated model of customer time availability

In this model we add index i in binary variable z_j to get the assignment variable $z_{ij}, i \in I, j \in J$

$$z_{ij} = \begin{cases} 1 & \text{if inaccessible customer } j \text{ is assigned to center } i, \\ 0 & \text{otherwise} \end{cases}$$

Note that the cause of inaccessibility of a customer in this model might be not only the time but also the total capacities in the centers.

The goal we have is the same as in the CCTA except that we will add a third criterion for change for the total deficit time of inaccessible customers (see fig.1). This problem can be also formulated as a following problem of integer programming (ECCTA):

$$LEX\Big(\sum_{i\in I}\sum_{j\in J} w_j z_{ij}, \sum_{i\in I}\sum_{j\in A_i} w_j t_{ij} x_{ij}, \sum_{i\in I}\sum_{j\in J} w_j (t_{ij} - \tau) z_{ij}\Big) \to \min \tag{12}$$

s.t.

$$\sum_{i\in I} x_{ii} \le q \tag{13}$$

$$\sum_{i\in I: j\in A_i} x_{ij} + \sum_{i\in I} z_{ij} = 1 \qquad \forall j \in J, \tag{14}$$

$$\sum_{i\in I} y_i = p, \tag{15}$$

$$\sum_{j\in A_i} w_j x_{ij} + \sum_{j\in J} w_j z_{ij} \le \kappa y_i \qquad \forall i \in I, \tag{16}$$

$$q x_{ii} \ge y_i \ge x_{ii} \qquad \forall i \in I, \tag{17}$$

$$x_{ij} \le x_{ii} \qquad \forall i \in I, \forall j \in A_i - \{i\}, \tag{18}$$

$$z_{ij} \le x_{ii} \qquad \forall i \in I, \forall j \in J - \{i\}, \tag{19}$$

$$x_{ij} \in \{0, 1\} \qquad \forall i \in I, \forall j \in A_i, \tag{20}$$

$$y_i \in \{0, 1, \dots, p\} \qquad \forall i \in I, \tag{21}$$

$$z_{ij} \in \{0, 1\}, \qquad \forall i \in I, \forall j \in J. \tag{22}$$

The object function (12) is a three dimensional vector consisting of total weight of inaccessible customers, weight of total driven time to accessible customers, and weight of total deficit time to inaccessible customers, which is minimized lexicographically. Constraints (13), (14) and (18) ensures the same conditions as (2), (3)

and (4) in the model CTA. Constraints (15) – (17) ensures the same condition as in model CCTA. Conditions (19) assign center for inaccessible customers. Last conditions (20), (21) and (22) are obligatory.

The ECCTA model contains too many binary variables, which could be of concern. I might raise an issue to solve the problem in an acceptable time.

6 Reduced heuristics for ECCTA

One of the ways to find a heuristic solution via an exact methods is to interrupt its calculation and declare the best achieved solution as heuristic solution. Another way is to reduce the set of feasible solutions or decompose the problem into more simple ones.

We suggest solving our problem ECCTA heuristically as follows:

- **Step 1** Reduction of set of feasible solutions: Here we assume that centers have a lot of weight and so we will choose them from set of nodes whose weight is bigger than their median. Let I^* be a set of hot candidates for centers (reduced set of candidates)

$$I^* = \{i \in I : w_i \geq w_{med}\}$$

  where w_{med} is median of a set $\{w_i : i \in I\}$.
- **Step 2** Reduction of the model CCTA via the candidate I^* as following problem (RCCTA):

$$LEX\Big(\sum_{j\in J} w_j z_j, \sum_{i\in I^*}\sum_{j\in A_i} w_j t_{ij} x_{ij}\Big) \to \min \tag{23}$$

s.t.

$$\sum_{i\in I^*} x_{ii} \leq q \tag{24}$$

$$\sum_{i\in I^*: j\in A_i} x_{ij} + z_j = 1 \qquad \forall j \in J, \tag{25}$$

$$x_{ij} \leq x_{ii} \qquad \forall i \in I^*, \forall j \in A_i - \{i\}, \tag{26}$$

$$\sum_{i\in I^*} y_i = p, \tag{27}$$

$$\sum_{j\in A_i} w_j x_{ij} \leq \kappa y_i \qquad \forall i \in I^*, \tag{28}$$

$$q x_{ii} \geq y_i \geq x_{ii} \qquad \forall i \in I^*, \tag{29}$$

$$y_i \in \{0, 1, \ldots, q\} \qquad \forall i \in I^*, \tag{30}$$

$$x_{ij} \in \{0, 1\} \qquad \forall i \in I^*, \forall j \in A_i, \tag{31}$$

$$z_j \in \{0, 1\} \qquad \forall j \in J. \tag{32}$$

  The change in the RCCAT model compared to the CCTA model is that we have $I = I^*$ from which we promise a significant acceleration of the computation time of the RCCAT.
- **Step 3** Assigning inaccessible customers to selected centers: We use two object functions 1 and 3 from the CCTA with with relevant variables z_{ij}, y_i.

$$LEX\Big(\sum_{(i,j)\in K_x} w_j z_{ij}, \sum_{(i,j)\in K_x} w_j (t_{ij} - \tau) z_{ij}\Big) \to \min \tag{33}$$

s.t.

$$\sum_{i\in I_x} y_i = p, \tag{34}$$

$$\sum_{i\in U_j} z_{ij} = 1 \qquad \forall j \in J_x, \tag{35}$$

$$\sum_{j\in V_i} w_j z_{ij} + w_i^* \leq \kappa y_i \qquad \forall i \in I_x, \tag{36}$$

$$z_{ij} \in \{0, 1\} \qquad \forall (i,j) \in K_x, \tag{37}$$

$$y_i \in \{0, 1, \ldots, p\} \qquad \forall i \in I_x. \tag{38}$$

The object function (33) is two dimensional vector of total weight of inaccessible customers and weight of total deficit time to inaccessible customers, which is minimized lexicographically. Constraints (34) distribute vehicles for fixed centers. Constraints (35) ensures that all inaccessible customers are assigned. Conditions (36) ensure sufficient capacity in the centers. Last conditions (37) and (38) are obligatory.

In this step, we fix the selection of centers in the set I_x by definition $I_x = \{i \in I^* : x_{ii} = 1\}$ from step 2. We will also need the following sets:

- $J_x = J - \{j : i \in I_x, j \in A_i : x_{ij} = 1\}$ – a set of inaccessible customers,
- $K_x = \{(i,j) \in I_x \times J_x : t_{ij} \leq \rho\tau\}$ – a set of feasible pairs (i,j), where ρ is maximum extension of a critical driving time (for example $\rho = 3$),
- $U_j = \{i \in I_x : (i,j) \in K_x\}, j \in J_x, V_i = \{j \in J_x : (i,j) \in K_x\}, i \in I_x$, – auxiliary index sets,
- $w_i^* = \sum_{j \in A_i} w_j, i \in I_x$ – cumulative weights of accessible customers assigned to the center.

7 Computational experiments

Our experiments were conducted on PC Workstation (processor 8-core i7-5960X 3GHz, RAM 32GB) with OS Linux (Debian/stretch). We used Python-based tools and the Python interface to commercial mathematical programming solver Gurobi [1].

				CAT				CCAT ($\kappa = 2500$)			
Inst.	n	p	q	δ	obj_1	obj_2	$Time[min.]$	δ	obj_1	obj_2	$Time[min.]$
BA	87	25	9	5	44	31561	< 0.009	5	44	31561	< 0.009
BB	515	46	46	8	10	27376	0.04	22	34	28611	0.06
KE	460	48	46	0	0	24117	0.02	5	17	34540	0.09
NR	350	36	35	1	3	34072	0.05	1	3	34072	0.05
PO	664	44	44	0	0	24467	0.03	37	72	38910	0.06
TN	276	26	26	0	0	23476	0.01	1	1	29298	0.01
TT	249	22	22	1	4	27142	0.02	8	30	27635	0.03
ZA	315	36	32	6	17	33757	0.01	6	17	33757	0.01
SR	2916	283	260	36	58	277231	6.18	36	58	277231	6.30

n – number of nodes in network, p – number of vehicles q – number of centers, δ – number of inaccessible nodes, κ – capacity of vehicles, obj_1 – weighted number of inaccessible customers, obj_2 – weighted travel time of accessible customers, *Time* – computational time.

Table 1 Computational results for models CAT and CCAT

	ECCAT ($\kappa = 2500$)						HCCAT ($\kappa = 2500$)					
Inst.	δ	$\overline{\Delta}$	obj_1	obj_2	obj_3	$Time[min.]$	δ	$\overline{\Delta}$	obj_1	obj_2	obj_3	$Time[min.]$
BA	5	4.16	44	31790	183	0.02	5	3.51	51	40198	179	< 0.009
BB	22	3.00	34	28611	102	0.72	26	3.09	45	59828	139	0.05
KE	5	3.18	17	34560	54	0.56	–	–	–	–	–	–
NR	1	2.00	3	34072	6	0.37	4	4.00	13	46143	52	0.02
PO	37	2.60	72	39291	187	0.97	41	2.84	90	97295	256	0.09
TN	1	2.00	1	29373	2	0.08	6	4.88	17	35102	83	0.02
TT	8	8.87	38	27635	266	0.23	11	5.68	41	47581	233	0.02
ZA	6	4.12	17	33757	70	0.12	16	3.19	57	53098	182	0.02
SR	35	4.12	58	277547	156	59.14	91	3.88	215	1152501	835	2.03

δ – number of inaccessible nodes, $\overline{\Delta} = obj_3/obj_1$ – mean of deficit time of inaccessible costumers, κ – capacity of vehicles, obj_1 – weighted number of inaccessible customers, obj_2 – weighted travel time of accessible customers, *Time* – computational time.

Table 2 Computational results for models ECCAT and HCCAT

To demonstrate the performance of the proposed algorithm, we tested it on seven self-governing regions of Slovak republic i.e Bratislava (BA), Banská Bystrica (BB), Košice (KE), Nitra (NR), Trenčín (TN), Trnava (TT), Žilina (ZA) and in throughout the region (SR). We implemented our models in Python 3.4 on a PC equipped by processor Intel i7-5960X,3.00 GHz, 32 GB RAM with 8 cores.

First results are summarized in table 1 for the models CAT and CCAT. As we can see in table 1, the components of vectors (δ, obj_1, obj_2) in CAT are at most equal corresponding in the CCAT. We expected such results, we were surprised only by small computational time for SR.

The main computing results are summarized in table 2 for the ECCAT. We can again see in table 2, that the components of vectors $(\delta, obj_1, obj_2, obj_3)$ in the regions SR we got under one minute of the calculation. Only the Instance of the Slovak Republic was calculated in an hour, which we consider an acceptable time.

A great disappointment is the heuristics RCCAP&HCCAP proposed for us. In the case of the KE region, they did not even get an admissible solution. The reason was the fact that one potencial center did not appear in the hot centers, which led to an inadmissible solution of the HCCAP. On the other hand, we were pleasantly surprised by the instances of TT, ZA, SR where the mean of deficit time of inaccessible costumers – $\overline{\Delta}$ is lower than in the exact model.

8 Conclusion and future research

In this paper we studied the models of overloaded centers detects in a transport network of q-service centers with p-vehicles (with given capacity of customers / vehicles) with goal to minimize the number of inaccessible customer locations (those with exceeded critical arrival time from an assigned center). We are gradually developing the basic lexicographic model of CPA through capacitated CCPA to the target model of ECCPA, for which we propose a reduced heuristic version of HCCPA. Computation experiments via solver Gurobi show that even the role of the largest instance of the Slovak Republic could be calculated in real time.

In further research we want to focus on an uncertainty capacity constraints in our models. We would like to chat the future results of over solution in simulation models for large instances.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-19-0441 and research grants VEGA 1/0776/20 Vehicle routing and scheduling in uncertain conditions and VEGA 1/0342/18 Optimal dimensioning of service systems.

References

[1] Gurobi Optimization Inc. *Gurobi optimizer reference manual*, 2020, http://www.gurobi.com.

[2] Janáček, J., Gábrišová, L. (2019). User-fair designing emergency service systems. *Transport*, 34(4), 499–507.

[3] Kvet, M., Janáček, J. (2018), Kernel-like research for robust emergency system designing, *Croatian Operation Research Review*, CRORR, 1–7.

[4] Jánošíková, Ľ., Gábrišová, L., Ježek, B. (2015). Load balancing location of emergency medical service stations In *Ekonomie a management, vědecký ekonomický časopis*, Roč.18, č.3, 30–40.

[5] Jánošíková, Ľ., Haviar, M. (2015). Hybrid genetic algorithms for the capacitated P-median problem. In *PSOR ´15 :proceedings of the 13th international Symposium on Operational Research*, Ljubljana, Slovenian Society Informatika, 176–181.

[6] Kvet, M., Janáček, J. (2020). Robust emergency system design using reengineering approach In *ICORES 2020 proceedings of the 9th international conference on Operations research and enterprise systems*, SciTePress, 172–178.

[7] Peško, Š., Borčinová, Z. (2019). A modified micro-genetic algorithm for robust emergency system designing, In *ICCC 2019, 20th International Carpathian Control Conference*. Danvers: IEEE Industry Applications Society, 2019., 1–4.

[8] Peško, Š., Majer, T. (2015). Robust p-median problem in changing networks, *Transport Problems, international scientific journal*, Vol. 10, iss. 3, 125–130.

Approximation of Oriented Behavioural Present Value by Trapezoidal Oriented Fuzzy Number

Krzysztof Piasecki[1], Anna Łyczkowska-Hanćkowiak[2]

Abstract. In this paper the model of imprecise quantity information is an oriented fuzzy number (OFN). By imprecision we mean the composition of indistinctness and ambiguity. In our paper, the OFN ambiguity is evaluated by of ambiguity index determined as an extension of energy measure introduced by de Luca and Termini. The OFN indistinctness is evaluated by indistinctness index determined as an extension of Czogała-Gottwald-Pedrycz entropy measure. Oriented Behavioural Present Value (O-BPV) is described such OFN which is not trapezoidal one. On the other hand, arithmetic operations on OFNs are much more complicated than arithmetic operations on trapezoidal OFNs. Therefore, the purpose of our study is approximation O-BPV by the nearest trapezoidal OFN. This way, we can simplify arithmetical operations on any linear transformation of O-BPV. The set of feasible trapezoidal OFNs is limited by the combination of following conditions: invariance of ambiguity index and invariance of indistinctness index.

Keywords: oriented fuzzy number, behavioural present value, ambiguity, indistinctness

JEL Classification: C02, C44, G11, G40
AMS Classification: 03E72, 91G30, 91B86

1 Introduction

Ordered fuzzy number (FN) is such model of imprecise number that subtraction of ordered FNs is the inverse operator to their addition [3]. Oriented FN (OFN) [8] is defined as some kind of ordered FN. The evolution process from ordered FN to OFN is briefly described in [10].

In [7, 14] the behavioural present value (BPV) as such approximation of fair price which is made under impact of behavioural factors. Behavioural premises cause that this approximation is an imprecise number. FNs are commonly accepted model of an imprecise number. Therefore, BPV is an FN. In [5] the information described by BPV is supplemented with a forecast of the price trend. This forecast was implemented in the model BPV as an orientation of FN. In this way the BPV was replaced by oriented BPV (O-BPV) described by an ordered FN. In [12], the formal approach to O-BPV is modified in this way that it is described by OFN.

Any financial analysis requires many linear transformations. For any OFN, implementation of these transformations is very complicated [9]. On the other hand, we can simplify the OFN linear transformations by limiting the calculations only to the case of trapezoidal OFN (TrOFN). O-BPV is not TrOFN. Therefore, the purpose of our article is to propose a method for approximating O-BPV by TrOFN. This approximation task uses approximation methods described in [9, 11]. In our the best knowledge, the problem of aproximation any OFN by TrORF is disscused only in these papers.

2 Oriented fuzzy numbers – selected facts

Dubois and Prade [1] define FN as such a fuzzy subset in the real line which fulfils some axioms. Any FN may be interpreted as imprecise approximation of a real number. The space of all FNs we denote by the symbol $\mathbb{F}$. OFN is an extension of the FN concept. OFN usefulness follows from the fact that it is interpreted as FN with additional information about the location of the approximated number. This information is given OFN orientation.

[1] Poznań University of Economics and Business, Department of Investment and Real Estate, al. Niepodległości 10, 61-875 Poznań, Poland, krzysztof.piasecki@ue.poznan.pl
[2] WSB University in Poznań, Institute of Finance, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, Poland, anna.lyczkowska-hanckowiak@wsb.poznan.pl

Definition 1 [8]**.** For any monotonic sequence $(a,b,c,d) \subset \mathbb{R}$, the oriented fuzzy number (OFN) $\overleftrightarrow{\mathcal{L}}(a,b,c,d,S_L,E_L) = \overleftrightarrow{\mathcal{L}}$ is the pair of orientation $\overrightarrow{a,d} = (a,d)$ and FN $\mathcal{L} \in \mathbb{F}$ described by membership function $\mu_L(\cdot\,|a,b,c,d,S_L,E_L) \in [0,1]^{\mathbb{R}}$ given by the identity

$$\mu_L(x|a,b,c,d,S_L,E_L) = \begin{cases} 0, x \notin [a,d] \equiv [d,a], \\ S_L(x), x \in [a,b[\equiv]b,a], \\ 1, x \in [b,c] \equiv [c,b], \\ E_L(x), x \in]c,d] \equiv [d,c[, \end{cases} \tag{1}$$

where the starting function $S_L \in [0,1[^{[a,b[}$ and the ending function $E_L \in [0,1[^{]c,d]}$ are upper semi-continuous monotonic ones meeting the condition

$$\left[\overleftrightarrow{\mathcal{L}}\right]_{0^+} = \lim_{\alpha \to 0^+} \{x \in \mathbb{R} : \mu_L(x|a,b,c,d,S_L,E_L) \geq \alpha\} = [a,d]. \tag{2}$$

The space of all OFN is denoted by the symbol $\mathbb{K}$. If $a < d$ then OFN $\overleftrightarrow{\mathcal{L}}(a,b,c,d,S_L,E_L)$ has the positive orientation $\overrightarrow{a,d}$. For any $z \in [b,c]$, the positively oriented OFN $\overleftrightarrow{\mathcal{L}}(a,b,c,d,S_L,E_L)$ is a formal model of linguistic variable "about or slightly above z". If $a > d$, then OFN $\overleftrightarrow{\mathcal{L}}(a,b,c,d,S_L,E_L)$ has the negative orientation $\overrightarrow{a,d}$. For any $z \in [c,b]$, the negatively oriented OFN $\overleftrightarrow{\mathcal{L}}(a,b,c,d,S_L,E_L)$ is a formal model of linguistic variable "about or slightly below z". If $a = d$, then OFN $\overleftrightarrow{\mathcal{L}}(a,b,c,d,S_L,E_L) = [\![a]\!]$ describes un-oriented real number $a \in \mathbb{R}$.

To estimate the distance between any pair of OFNs we introduce a pseudo-metrics $\delta: \mathbb{K}^2 \longrightarrow \mathbb{R}_0^+$ determined by a following identity

$$\delta\left(\overleftrightarrow{\mathcal{L}}(a,b,c,d,S_L,E_L), \overleftrightarrow{\mathcal{L}}(e,f,g,h,F_L,L_L)\right) = \sqrt{(a-e)^2 + (b-f)^2 + (c-g)^2 + (d-h)^2}. \tag{3}$$

Any linear operations on OFNs have a high level of formal complexity [10]. Due to that, in many practical applications researchers limit the use of OFN only to a form presented below.

Definition 2. [8] For any monotonic sequence $(a,b,c,d) \subset \mathbb{R}$ the trapezoidal OFN (TrOFN) $\overleftrightarrow{Tr}(a,b,c,d)$ is defined as the OFN determined by its membership function $\mu_{\overleftrightarrow{Tr}}(\cdot\,|a,b,c,d) \in [0;1]^{\mathbb{R}}$ given by the identity

$$\mu_{\overleftrightarrow{Tr}}(x|a,b,c,d) = \begin{cases} 0, x \notin [a,d] = [d,a], \\ \dfrac{x-a}{b-a}, x \in [a,b[=]b,a], \\ 1, x \in [b,c] = [c,b], \\ \dfrac{x-d}{c-d}, x \in]c,d] = [d,c[. \end{cases} \tag{4}$$

The symbol $\mathbb{K}_{Tr}$ denotes the space of all TrOFNs.

After Klir [2] we understand imprecision as a superposition of ambiguity and indistinctness of information. Ambiguity can be interpreted as a lack of a clear recommendation between one alternative among various others. Indistinctness is understood as a lack of explicit distinction between recommended and not recommended alternatives. Any OFN is a particular kind of imprecision information. An increase in information imprecision reduces suitability of this information. Therefore, it is logical to consider the problem of imprecision assessment.

In [13] it is justified that a proper tool for measuring the OFN ambiguity is the ambiguity index $a \in \mathbb{R}^{\mathbb{K}}$ assessing the ambiguity of any OFN $\overleftrightarrow{\mathcal{L}} = \overleftrightarrow{L}(a,b,c,d,L_L,R_L)$ in following way

$$a\left(\overleftrightarrow{L}(a,b,c,d,L_L,R_L)\right) = \int_a^d \mu_L(x|a,b,c,d,L_L,R_L)dx, \tag{5}$$

where $\mu_L \in [0,1]^{\mathbb{R}}$ is the membership function determining OFN $\overleftrightarrow{\mathcal{L}}$. Proposed ambiguity index is an extension of energy measure proposed by de Luca and Termini [4].

Moreover, in [13] it is justified that a proper tool for measuring the OFN indistinctness is the indistinctness index $g \in \mathbb{R}^{\mathbb{K}}$ assessing the indistinctness of any OFN $\overleftrightarrow{\mathcal{L}} = \overleftrightarrow{L}(a,b,c,d,L_L,R_L)$ in following way

$$\mathcal{g}\left(\overleftrightarrow{\mathcal{L}}(a,b,c,d,L_L,R_L)\right) = \int_a^d \min\{\mu_{\mathcal{L}}(x|a,b,c,d,L_L,R_L), 1-\mu_{\mathcal{L}}(x|a,b,c,d,L_L,R_L)\}\,\mathrm{d}x, \tag{6}$$

where $\mu_L \in [0,1]^{\mathbb{R}}$ is the membership function determining OFN $\overleftrightarrow{\mathcal{L}}$. Proposed indistinctness index is an extension of Czogała–Gottwald–Pedrycz entropy measure introduced in [6].

3 Oriented Behavioural Present Value

We take into account given financial asset. The price $\check{P}$ of this asset may fluctuate over time. Therefore, we can consider a price trend. In technical analysis, we always assume that in the nearest short time period this trend has a fixed point P_0 called the balanced price. Current value of balanced price P_0 is substantively justified by fundamental analysis. Then we conclude that the price $\check{P}$ should converge to the balanced price P_0. Considered asset is characterized by the vector $\mathcal{v} = \left(\check{P}, P_0, V_{\min}, V_{\max}\right)$, where

- $V_{\min}$ – the greatest lower bound of the PV evaluation,
- $V_{\max}$ – the least upper bound of the PV evaluation.

In [7, 14], we can find some method determining the bounds of PV evaluation. If $\check{P} > P_0$, then considered asset is overvalued. For the case $\check{P} < P_0$, considered asset is undervalued. Both of these cases we call the state of financial disequilibrium. Financial equilibrium is state in which asset prices satisfy condition $\check{P} = P_0$. Then considered asset is fully valued. The state of financial equilibrium/disequilibrium is characterized by relative deviation of the price $\check{P}$ from the balanced price P_0 determined as follows

$$\delta P = \frac{\check{P} - P_0}{\check{P}} \tag{7}$$

A behavioural present value (BPV) is defined as such evaluation of present value which is dependent on some behavioural factors [7, 14]. Then BPV is FN $\widetilde{BPV}(\boldsymbol{\mathcal{v}}) \in \mathbb{F}$ given by its membership function $\mu_{BPV}(\cdot\,|\boldsymbol{\mathcal{v}}) \in [0;1]^{\mathbb{R}^+}$ in following way

$$\mu_{BPV}(x|\mathcal{v}) = \begin{cases} 0, x \notin [V_{\min}, V_{\max}], \\ h(x|\mathcal{v}), x \in \left[V_{\min}, \check{P}\right[, \\ 1, x \in \left[\check{P}, \check{P}\right], \\ k(x|\mathcal{v}), x \in \left]\check{P}, V_{\max}\right], \end{cases} \tag{8}$$

where the functions $h(\cdot\,|\mathcal{v}) \in [0,1[^{[V_{\min},\check{P}[}$ and $k(\cdot\,|\mathcal{v}) \in [0,1[^{]\check{P},V_{\max}]}$ are determined by the identities

$$h(x|\mathcal{v}) = \begin{cases} \dfrac{(x - V_{\min})(1+\delta P)}{\check{P} - V_{\min} + (x - V_{\min})\cdot \delta P}, \delta P > 0, \\ \dfrac{x - V_{\min}}{\check{P} - V_{\min} - (x - V_{\min})\cdot \delta P}, \delta P \le 0, \end{cases}. \tag{9}$$

$$k(x|\mathcal{v}) = \begin{cases} \dfrac{V_{\max} - x}{V_{\max} - \check{P} + (V_{\max} - x)\cdot \delta P}, \delta P > 0, \\ \dfrac{(V_{\max} - x)(1-\delta P)}{V_{\max} - \check{P} - (V_{\max} - x)\cdot \delta P}, \delta P \le 0. \end{cases}. \tag{10}$$

A price forecast may be implemented in the BPV model as an orientation of FN. In this way the BPV is replaced by oriented BPV (O-BPV) described by an OFN $\overleftrightarrow{BPV} = \overleftrightarrow{\mathcal{L}}\left(V_s, \check{P}, \check{P}, V_e, L_{BPV}, R_{BPV}\right)$ determined by its membership function $\mu_{BPV}(\cdot\,|\mathcal{v})$ given by the identity (8) [13], where

- $[V_s, V_e] = [V_{\min}, V_{\max}] \equiv [V_{\max}, V_{\min}] \subset \mathbb{R}^+$ is interval of all possible BPV' values,
- (L_{BPV}, R_{BPV}) is any ordered pair of starting-function and ending-function.

Positive O-BPV orientation predicts rise in assets prise. Then O-BPV is given by the formula

$$\overleftrightarrow{BPV} = \overleftrightarrow{\mathcal{L}}\left(V_{\min}, \check{P}, \check{P}, V_{\max}, h, k\right). \tag{11}$$

Negative O-BPV orientation predicts fall in asset price. Then O-BPV is given by the formula

$$\overleftrightarrow{BPV} = \overleftrightarrow{\mathcal{L}}\left(V_{\max}, \check{P}, \check{P}, V_{\min}, k, h\right). \tag{12}$$

4 Approximation

To simplify linear transformations of any O-BPV $\overleftrightarrow{BPV} = \overleftrightarrow{\mathcal{L}}(V_s, \check{P}, \check{P}, V_e, L_{BPV}, R_{BPV})$, we can replace it with the nearest TrOFN. Hence, the main criterion of approximation is to determine such TrOFN $\overleftrightarrow{Tr}(p_0, q_0, r_0, s_0)$, that will satisfy the following condition

$$\begin{gathered}\delta\left(\overleftrightarrow{Tr}(p_0, q_0, r_0, s_0), \overleftrightarrow{\mathcal{L}}(V_s, \check{P}, \check{P}, V_e, L_{BPV}, R_{BPV})\right) = \\ = \min\left\{\delta\left(\overleftrightarrow{Tr}(p, q, r, s), \overleftrightarrow{\mathcal{L}}(V_s, \check{P}, \check{P}, V_e, L_{BPV}, R_{BPV})\right) : \overleftrightarrow{Tr}(p, q, r, s) \in \mathbb{K}_{Tr}\right\}.\end{gathered} \tag{13}$$

When processing imprecise values, we use OFN only to follow the influence of imprecision on the quality of obtained information. Due to in our approximation problem we can impose the requirements of estimating by such TrOFN which retains the imprecision estimates of an approximated value. Therefore, ambiguity and indistinctness of approximating TrOFN meet the conditions

$$a\left(\overleftrightarrow{Tr}(p_0, q_0, r_0, s_0)\right) = A = a\left(\overleftrightarrow{\mathcal{L}}(V_s, \check{P}, \check{P}, V_e, L_{BPV}, R_{BPV})\right). \tag{14}$$

$$g\left(\overleftrightarrow{Tr}(p_0, q_0, r_0, s_0)\right) = G = g\left(\overleftrightarrow{\mathcal{L}}(V_s, \check{P}, \check{P}, V_e, L_{BPV}, R_{BPV})\right). \tag{15}$$

Using (5) and (6), for any positively O-BPV (11) we obtain

$$A = \begin{cases} \frac{1}{\delta P} \cdot \left(1 - \frac{\ln|1 + \delta P|}{\delta P}\right) \cdot \left((\check{P} - V_{\min}) \cdot (1 + \delta P) + (V_{max} - \check{P})\right), \delta P > 0, \\ \frac{V_{\max} - V_{\min}}{2}, \delta P = 0, \\ \frac{-1}{\delta P} \cdot \left(1 + \frac{\ln|1 - \delta P|}{\delta P}\right) \cdot \left((\check{P} - V_{\min}) + (V_{\max} - \check{P}) \cdot (1 - \delta P)\right), \delta P < 0, \end{cases} \tag{16}$$

$$G = \begin{cases} \frac{1}{(\delta P)^2} \cdot \ln\left|\frac{(2 - \delta P)^2 \cdot (\delta P + 1)}{4}\right| \cdot \left((\check{P} - V_{\min}) \cdot (1 + \delta P) - (V_{\max} - \check{P})\right), \delta P > 0, \\ \frac{V_{\max} - V_{\min}}{4}, \delta P = 0, \\ (V_{\max} - V_{\min}) + \frac{1}{(\delta P)^2} \cdot \ln\left|\frac{(2 - \delta P)^2 \cdot (1 - \delta P)}{4}\right| \cdot (2 \cdot \check{P} - V_{\max} - V_{\min}), \delta P < 0. \end{cases} \tag{17}$$

In analogous way, for any negatively oriented O-BPV (12), we obtain

$$A = \begin{cases} \frac{-1}{\delta p} \cdot \left(1 - \frac{\ln|1 + \delta P|}{\delta p}\right) \cdot \left((\check{P} - V_{\min}) \cdot (1 + \delta P) + (V_{\max} - \check{P})\right), \delta P > 0, \\ \frac{V_{\min} - V_{\max}}{2}, \delta P = 0, \\ \frac{1}{\delta P} \cdot \left(1 + \frac{\ln|1 - \delta P|}{\delta P}\right) \cdot \left((\check{P} - V_{\min}) + (V_{\max} - \check{P}) \cdot (1 - \delta P)\right), \delta P < 0, \end{cases} \tag{18}$$

$$G = \begin{cases} \frac{1}{(\delta P)^2} \cdot \ln\left|\frac{(2 - \delta P)^2 \cdot (\delta P + 1)}{4}\right| \cdot \left((\check{P} - V_{\min}) \cdot (1 + \delta P) - (V_{\max} - \check{P})\right), \delta P > 0, \\ \frac{V_{\min} - V_{\max}}{4}, \delta P = 0, \\ (V_{\min} - V_{\max}) + \frac{-1}{(\delta P)^2} \cdot \ln\left|\frac{(2 - \delta P)^2 \cdot (1 - \delta P)}{4}\right| \cdot (2 \cdot \check{P} - V_{\min} - V_{\max}), \delta P < 0. \end{cases} \tag{19}$$

At the moment, we are unable to implement the condition "sequence (p_0, q_0, r_0, s_0) is monotonic" in the approximation task (13). Therefore, we are only trying to solve the task (13) restricted by the conditions (14) and (15). In [12], it is shown that if the sequence

$$(p_0, q_0, r_0, s_0) = \varpi_1 = \left(\frac{V_s + V_e - A - G}{2}, \frac{2 \cdot \check{P} - A + G}{2}, \frac{2 \cdot \check{P} + A - G}{2}, \frac{V_s + V_e + A + G}{2}\right). \tag{20}$$

is monotonic then $\overleftrightarrow{Tr}(\varpi_1)$ is the unique solution of the approximation task (13) with restricting conditions (14) and (15).

Example 1. For considered security we observe price $\check{P} = 60$. Substantially justified its balanced price is given as follows $P_0 = 40$. The greatest lower and the least upper bounds of PV evaluation are respectively $V_{min} = 30$ and $V_{max} = 80$. We have

$$\delta P = \frac{60 - 40}{60} = \frac{1}{3}.$$

Additionally, we predict fall in price of considered security. Then O-BPV is described as negatively OFN

$$\overleftrightarrow{BPV} = \overleftrightarrow{\mathcal{L}}(80,60,60,30, k, h)$$

where

$$h(x) = \frac{4x - 120}{x + 60} \quad \text{for} \quad x \in]60; 30],$$

$$k(x) = \frac{3x - 240}{x - 140} \quad \text{for} \quad x \in [80; 60[$$

Using (18) and (19), we obtain

$$a\left(\overleftrightarrow{BPV}\right) = A = -24.6517,$$

$$g\left(\overleftrightarrow{BPV}\right) = G = -13.8530.$$

Dependence (20) shows that TrOFN $\overleftrightarrow{Tr}(74.2524, 65.3994, 54.6007, 35.7477)$ is the unique solution of approximation task (13) under restricting conditions (14) and (15). It ist the best approximation of considered O-BPV by TrOFN.

If the sequence ϖ_1 is not monotonic then we can take into account less restrictive approximation task (13) with only condition (14). In [10], it is shown that if sequence

$$(p_0, q_0, r_0, s_0) = \varpi_2 = \left(\frac{3 \cdot V_s + V_e - 2 \cdot A}{4}, \frac{-V_s + 4 \cdot \check{P} + V_e - 2 \cdot A}{4}, \frac{V_s + 4 \cdot \check{P} - V_e + 2 \cdot A}{4}, \frac{V_s + 3 \cdot V_e + 2 \cdot A}{4}\right). \tag{21}$$

is monotonic then $\overleftrightarrow{\boldsymbol{Tr}}(\boldsymbol{\varpi_2})$ is the unique solution of approximation task (13) with restricting condition (14). Then the error of skipping the condition (15) can be estimated using a relative deviation

$$\delta G = \frac{g\left(\overleftrightarrow{Tr}(\varpi_2)\right) - G}{G} \tag{22}$$

Example 2. We try to apply approximation task (13) with condition (14) for O-BPV described in Example 1. Using (21), we obtain the sequence

$$\varpi_2 = (79.8259, 59.8259, 60.1742, 30.1742)$$

which is not monotonic. So therefore, we cannot determine the solution of task (13) with condition (14).

When both sequences ϖ_1 and ϖ_2 are not monotonic then we can only apply the unconditional approximation task (13). This task always has a solution. The TrOFN $\overleftrightarrow{Tr}(V_s, \check{P}, \check{P}, V_e)$ is the unique solution of unconditional task (13). Then the error of skipping the conditions (14) and (15) can be estimated using relative deviations (22) and

$$\delta A = \frac{a\left(\overleftrightarrow{Tr}(\varpi_2)\right) - A}{A}. \tag{23}$$

Example 3. We apply unconditional approximation task (13) with for O-BPV described in Example 1. Then O-BPV is approximate by TrOFN $\overleftrightarrow{Tr}(80,60,60,30)$. The error of skipping the conditions (14) and (15) is estimated by relative deviations (22) and 923) as follows

$$\delta G = \frac{g\left(\overleftrightarrow{Tr}(80,60,60,30)\right) - (-13.8530)}{-13.8530} = \frac{\frac{30-80}{4} - (-13.8530)}{-13.8530} = -0.098,$$

$$\delta A = \frac{a\left(\overleftrightarrow{Tr}(80,60,60,30)\right) - (-24.6517)}{-24.6517} = \frac{\frac{30-80}{2} - (-24.6517)}{-24.6517} = 0.014.$$

We suppose that $\overleftrightarrow{Tr}(80,60,60,30)$ may be accepted as approximation of considered O-BPV.

5 Final remarks

Presented approximation methods might lack the solution of incidental O-BPV. If, during implementation of a chosen approximation method of a given O-BPV we will obtain a non-monotonic sequence of parameters then such an approximation problem has no solution. Attempts should be made to approximate a given O-BPV by another approximation method.

In the context of experience gathered in [9, 11], it should be stated that the most faithful O-BPV approximation method is described by the task (13) with conditions (14) and (15). In this situation, this approximation should be the first choice method. And only in the absence of a solution designated by this method should be attempted use of approximation task (13) only with condition (14). If this approximation task has no solution, we must apply unconditional task (1). The drawbacks of UC-approximation method are the discrepancies between the characteristics of imprecision of approximated O-BPV and approximating TrOFN. It distorts the usefulness evaluation of processed information. The main advantage of unconditional approximation taks (13) is the fact that it is the only discussed methods which delivers a solution of any O-BPV.

The problems shown in examples are justification for an attempt to create such approximation methods of OFN by TrOFN that such a constraint that a "vector of solution parameters is a monotonic sequence" will be included.

The results obtained in this article will facilitate the use of O-BPV in portfolio analysis.

References

[1] DUBOIS, D. & PRADE, H. (1978). Operations on fuzzy numbers. *International Journal System Sciences*, 9, 613–629. https://doi.org/10.1080/00207727808941724.

[2] KLIR, G. J. (1993). Developments in uncertainty-based information, *Advances in Computers* 36, 255–332, doi:10.1016/s0065-2458(08)60273-9.

[3] KOSIŃSKI, W. PROKOPOWICZ, P. & ŚLĘZAK, D. (2002). Fuzzy numbers with algebraic operations: algorithmic approach. In M. KLOPOTEK, S. T. WIERZCHOŃ, & M. MICHALEWICZ, (Eds.) *Proc. IIS'2002* (pp. 311–320). Sopot, Poland, Physica Verlag, Heidelberg.

[4] DE LUCA, A. & TERMINI, S. (1979). Entropy and energy measures of fuzzy sets, In M. M. GUPTA, R. K. RAGADE & R. R. YAGER [Eds.], *Advances in fuzzy set theory and applications* (pp. 321–338). New York.

[5] ŁYCZKOWSKA-HANĆKOWIAK, A. (2017). Behavioural present value determined by ordered fuzzy number, *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.2988243.

[6] PEDRYCZ, W., GOTTWALD, S. & CZOGAŁA, E. (1982). Measures of fuzziness and operations with fuzzy sets, *Stochastica* 3, 187–205.

[7] PIASECKI, K. (2011). Behavioural Present Value, *SSRN Electronic Journal*, https://dx.doi.org/10.2139/ssrn.1729351.

[8] PIASECKI, K. (2018). Revision of the Kosiński's Theory of Ordered Fuzzy Numbers, *Axioms* Vol. 7(1), https://doi.org/10.3390/axioms7010016.

[9] PIASECKI, K. & ŁYCZKOWSKA-HANĆKOWIAK, A. (2018). On approximation of any ordered fuzzy number by a trapezoidal ordered fuzzy number, *Symmetry* 10(10), 26, https://doi.org/10.3390/sym10100526.

[10] PIASECKI, K. & ŁYCZKOWSKA-HANĆKOWIAK, A. (2020). Representation of Japanese Candlesticks by Oriented Fuzzy Numbers, Econometrics 8(1), 1, https://doi.org/10.3390/econometrics8010001.

[11] PIASECKI, K. & ŁYCZKOWSKA-HANĆKOWIAK, A., (2020). New method of approximation any oriented fuzzy number by trapezoidal one. In W. SZKUTNIK, A. SĄCZEWSKA-PIOTROWSKA, M. HADAŚ-DYDUCH & J. ACEDAŃSKI [Eds.] *13th International Scientific Conference "Analysis of International Relations 2020. Methods and Models of Regional Development. Winter Edition". Conference Proceedings* (pp. 14–21), Publishing House of the University of Economics in Katowice.

[12] PIASECKI, K. & ŁYCZKOWSKA-HANĆKOWIAK, A. (2020). Behavioural Present Value determined by oriented fuzzy number. In W. SZKUTNIK, A. SĄCZEWSKA-PIOTROWSKA, M. HADAŚ-DYDUCH, J. ACEDAŃSKI [Eds.] *14th International Scientific Conference "Analysis of International Relations 2020. Methods and Models of Regional Development. Summer Edition" Conference Proceedings*, (pp. 134–144) Publishing House of the University of Economics in Katowice.

[13] PIASECKI, K. & ŁYCZKOWSKA-HANĆKOWIAK, A. (2020). Imprecision indexes of oriented fuzzy numbers, In J. Kacprzyk [Ed.] *Uncertainty and Imprecision in Decision Making and Decision Support: New Challenges, Solutions and Perspectives*, Lecture Notes: Soft Computing, Springer (accepted for publication).

[14] PIASECKI, K. & SIWEK, J. (2015). Behavioural Present Value Defined as Fuzzy Number a new Approach, *Folia Oeconomica Stetinensia*, 15(2), 27–41, https://doi.o rg/10.1515/foli-2015-0033.

An Imprecise Image of Principal's Preferences for INSPIRE Negotiation Support System

Krzysztof Piasecki[1], Anna Łyczkowska-Hanćkowiak[2],
Ewa Roszkowska[3], Tomasz Wachowicz[4]

Abstract. Preferential information may be visualized in many different ways, which is an important issue in a principal-agent decision-making context, e.g., in representative negotiations. In the INSPIRE negotiation support system, the principal's preferences are visualized by circles with different radii. Negotiators digitize these preferences using numbers directly proportional to the size of the circles drawn by the principal to determine their negotiation offer scoring systems. The way negotiators understand the concept of the circle size is unknown, yet it may significantly affect the scoring systems they build and the quality of negotiation support that is offered to them by the INSPIRE system. Individual negotiators may differ in their understanding of this concept. It means that the notion "circle size" is a linguistic variable that may be described by a fuzzy set. Cited empirical studies show that the size of the circle is the value between the radius and the area of this circle. In this paper, we define the principal's preference as a fuzzy preorder between fuzzy "circle sizes". All considerations are illustrated by means of a short case study based on INSPIRE data.

Keywords: preference visualization**,** fuzzy ranking, negotiation problem, negotiation offer scoring systems

JEL Classification: C62, C65

AMS Classification: 03E72, 91B43, 91B86

1 Introduction

An agent's understanding of the principal's preferences is one of the topics of the principal-agent theory [7]. Among other things, the principal's preferences may be visualized in different ways as graphs, bar charts, pie charts, cartograms, bars, circles [2, 4, 5, 6, 9, 11]. However, there is no strict consensus regarding the best way of such visualization.

In our paper, we consider the principal's preferences, as described in the INSPIRE negotiation support system [3]. In this system, the principal's preferences are visualized by circles with different radii. For INSPIRE, agents' understanding of the principal's preferences has already been studied in [10, 12]. There, the rankings determined by agents were compared with visualization of the principal's preference. The main objective of this research is to determine the representation of the principal's preferences by fuzzy order relation.

2 Negotiation problem

In the formal model of INSPIRE, the negotiation template can be described by means of the ordered pair $(\mathcal{F}, \mathcal{X})$ where $\mathcal{F} = (f_i)_{i=1}^n$ is a sequence of negotiation issues f_i and $\mathcal{X} = (X_i)_{i=1}^n$ is a sequence of options lists X_i related for issue f_i. Then any negotiation package is given as the vector

$$\bar{P} = \left(x_{1,j(p)}, x_{2,j(p)}, \dots, x_{n,j(p)}\right) \in X_1 \times X_2 \times \dots \times X_n = \mathbb{P} \tag{1}$$

[1] Poznań University of Economics and Business, Department of Investment and Real Estate, al. Niepodległości 10, 61-875 Poznań, Poland, krzysztof.piasecki@ue.poznan.pl

[2] WSB University in Poznań, Institute of Finance, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, Poland, anna.lyczkowska-hanckowiak@wsb.poznan.pl

[3] University of Białystok, Institute of Quantitative Methods, ul. Warszawska 63, 15-062 Białystok, Poland, e.roszkowska@uwb.edu.pl

[4] University of Economics in Katowice, Department of Operations Research, ul. 1 Maja 50, 68, 40-287 Katowice, Poland tomasz.wachowicz@uekat.pl

Example 1: [10] We observe a negotiation, in which a musician and a broadcasting company "KAET-music" talk over the terms of a potential contract. The negotiation template is defined using four issues, each having a predefined list of options that allow to build 240 various offers (see Table 1).

Negotiations issues	Lists of predefined options
Number of new songs (introduced and performed each year)	11; 12; 13; 14 or 15 songs
Royalties for CDs (in percent)	1.5; 2; 2.5 or 3%
Contract signing bonus (in dollars)	$125 000; $150 000 or $200 000
Number of promotional concerts (per year)	5; 6; 7 or 8 concerts

Table 1 Example of negotiation template

We assume that at least one of the negotiating parties is an agent representing their Principal. The Principal visualizes its preferences using circles of various sizes. This is done separately for issues where the sequence $\left(\mathcal{C}_{i,0}\right)_{i=1}^{n}$ of circles visualizes the importance of individual issues. The guiding principle here is, the more important the issue f_i, the larger a size of the circle $\mathcal{C}_{i,0}$. Then, for each list X_i of predefined options, Principal separately visualizes the preferences between options by the sequence $\left(\mathcal{C}_{i,j}\right)_{j=1}^{n_i}$ of circles. The rule is that the better the option $x_{i,j}$, the larger a size of the circle $\mathcal{C}_{i,j}$.

Example 2: [10] In the negotiation described in Example 1, the management board of "KAET-music" is the Principal. Its preferences are visualized using circles, as shown in Table 2.

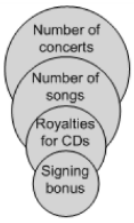
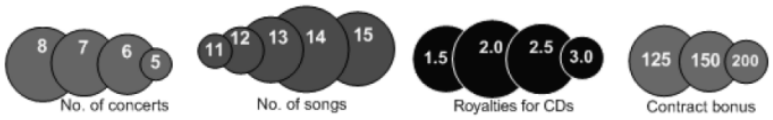
Issue importance	Preferences between predefined options
	

Table 2 Visualization of Principal's preferences.

The symbol $\mathbb{O}$ denotes the family of all circles. An agent subjectively interprets the notion "circle size" by nonnegative numbers. The rule is that the larger circle size, the higher number. In this way, each agent subjectively defines the "circle size" function $V: \mathbb{O} \to \mathbb{R}_0^+$.

In the INSPIRE system, any negotiation package is evaluated by scoring function $S: \mathbb{P} \longrightarrow [0,1]$ given as follows

$$S(\bar{P}) = S\left(x_{1,j(p)}, x_{2,j(p)}, \dots, x_{n,j(p)}\right) = \sum_{k=1}^{n} V\left(\mathcal{C}_{k,j(p)}\right) \tag{2}$$

Example 2 shows that issues' importance and preferences may be visualized using different scales. We can only notice that each circle $\mathcal{C}_{i,j}$ $(i = 1,2,\dots,n; j = 0,1,2,\dots,n_i)$ is uniquely represented by its radius $R_{i,j}$. For the needs of the INSPIRE system, these visualizations are standardized as follows:

- for visualization of issue importance, we calculate weights

$$\forall_{i-1,2,\dots,n}: \qquad w_i = \frac{R_{i,0}}{\sum_{k=1}^{n} R_{k,0}} \tag{3}$$

- for visualizations preferences between predefined options, we calculate standardized radii

$$\forall_{i-1,2,\dots,n}\ \forall_{j=1,2,\dots,n_i}: \qquad r_{i,j} = \frac{R_{i,j}}{\max\left\{R_{i,k}: k = 1,2,\dots,n_k\right\}} \tag{4}$$

Such standardization results from the assumption regarding the required range of any scoring function.

Example 3: In Table 3, we find the results of measuring the radius of circles drawn by the Principal in Table 2. The standardized radii are presented in Table 4.

Issue	Issue importance	Preferences between options				
Songs	4,74	2.00	2.70	3.70	4.90	4.20
Royalties	3.54	3.80	4.50	4.00	2.90	
Bonus	2.89	4.00	3.40	2.50		
Concerts	5.59	4.30	3.85	3.45	1.85	

Table 3 Original radii in preference visualization

Issue	Issue weights	Standardized radii for preference visualization				
Songs	0.282816	0.408163	0.551020	0.755102	1	0.857143
Royalties	0.211217	0.844444	1	0.888889	0.644444	
Bonus	0.172434	1	0.850000	0.625000		
Concerts	0.333532	1	0.895349	0.802326	0.430233	

Table 4 Standardized preference visualization

The agent interprets the "circle size" separately for issue importance and for each list of predefined options. In the INSPIRE system, all agent's interpretations are standardized in the same way as the principal's visualization of the principal's preferences. For this reason, the final evaluation of any negotiation package depends on the agent's understanding of the Principal's preferences.

3 Understanding the Principal's preferences

Brinton [1] recognized some problems with using circles as a tool for information presentation. We describe his investigation in modern language. The guiding principle of the method considered by him was that the higher utility of a characterized object, the larger the size of the representing circle. He showed that circle sizes evaluated by circle radius or by circle area make the reader misperceive the relative utility of the objects described by these circles. Brinton noticed that:

- comparison between radii causes overestimation of the relative utility of the worse object;
- comparison between areas causes underestimation of the relative utility of the worse object.

Many authors confirm these observations. They conclude accordingly that the relative sizes of a circle are misperceived, and these mistakes are systematic (see in [5]). Therefore, they propose such a formula of the "circle relative size" function, which allows "psychologically correct" circle sizes to be calculated. Their proposition implies that the "circle size" function $V(\cdot\,|\gamma): \mathbb{O} \to \mathbb{R}_0^+$ is given by the identity

$$V(\mathcal{C}|\gamma) = \alpha \cdot r^{\gamma} \tag{5}$$

where $\alpha \in \mathbb{R}^+$ is a size of benchmark circle with unit radius and $r \in \mathbb{R}_0^+$ is the radius of the circle $\mathcal{C} \in \mathbb{O}$. The exponent γ characterizes an agent's understanding of the Principal's preferences. Brinton's [1] observations prove that $\gamma \in [1, 2]$. The exponent γ derived in the empirical studies varies from 1.6 to 1.82 [5].

Due to (2) and (5), we can pre-specify the form of the scoring function $S(\cdot\,|\gamma): \mathbb{P} \to [0,1]$ as follows

$$S(\bar{P}|\gamma) = S\big(x_{1,j(p)}, x_{2,j(p)}, \dots, x_{n,j(p)}\big|\gamma\big) = \sum_{k=1}^{n} w_i \cdot r_{i,j(p)}^{\gamma} \tag{6}$$

It is evident that the rating of negotiation packages $\{\bar{P}_1, \bar{P}_2, \dots, \bar{P}_m\} \subset \mathbb{P}$ obtained with the use of the values $S(\bar{P}_k|\gamma)$ is the same as rating obtained with the use of the values $\sqrt[\gamma]{S(\bar{P}_k|\gamma)}$. Moreover, the values $\sqrt[\gamma]{S(\bar{P}_k|\gamma)}$ and $\sqrt[\delta]{S(\bar{P}_k|\delta)}$ can be compared for $\gamma \neq \delta$, because they are expressed in the same measurement unit. For these reasons, we propose the following final form of the scoring function $\hat{S}(\cdot\,|\gamma): \mathbb{P} \to [0,1]$ given by the identity

$$\hat{S}(\bar{P}|\gamma) = \hat{S}\left(x_{1,j(p)}, x_{2,j(p)}, \dots, x_{n,j(p)}\middle|\gamma\right) = \sqrt[\gamma]{\sum_{k=1}^{n} w_i \cdot r_{i,j(p)}^{\gamma}} \tag{7}$$

Let us note that the Jensen inequality implies

$$\forall_{\bar{P}\in\mathbb{P}}\ \forall_{\gamma\in[1,2]}: \qquad A(\bar{P}) \ = \ \hat{S}(\bar{P}|1) \leq \ \hat{S}(\bar{P}|\gamma) \leq \ \hat{S}(\bar{P}|2) = B(\bar{P}) \tag{8}$$

Moreover, for any negotiation package $\bar{P}$ the function $\hat{S}(\bar{P}|\ \cdot): [1,2] \to [1,2]$ is nondecreasing and continuous one. We conclude that for any negotiation package $\bar{P}$, all its possible scoring ratings form the interval

$$I(\bar{P}) \ = [A(\bar{P}), B(\bar{P})] \tag{9}$$

which is called scoring interval. These intervals will be the basis for comparisons between negotiation packages.

Example 4: We will compare the negotiation packages listed in Table 5**.** These packages will be compared using the scoring intervals, also described in Table 5.

Symbol	Package content	Negotiation package	$A(\bar{P})$	$B(\bar{P})$
$\bar{P}_1$	$(x_{1,1}, x_{2,1}, x_{3,1}, x_{4,1})$	$(5, 11, 1.5, 125\ 000)$	0.799762	0.838867
$\bar{P}_2$	$(x_{1,1}, x_{2,1}, x_{3,1}, x_{4,2})$	$(5, 11, 1.5, 150\ 000)$	0.773897	0.809844
$\bar{P}_3$	$(x_{1,1}, x_{2,1}, x_{3,1}, x_{4,3})$	$(5, 11, 1.5, 200\ 000)$	0.735100	0.773707
$\bar{P}_4$	$(x_{1,1}, x_{2,1}, x_{3,3}, x_{4,3})$	$(5, 11, 2.5, 200\ 000)$	0.744488	0.784151
$\bar{P}_5$	$(x_{1,1}, x_{2,2}, x_{3,3}, x_{4,3})$	$(5, 12, 2.5, 200\ 000)$	0.78489	0.808484
$\bar{P}_6$	$(x_{1,3}, x_{2,3}, x_{3,4}, x_{4,1})$	$(7, 13, 3.0, 125\ 000)$	0.789709	0.797568

Table 5 Considered negotiation packages and their scoring intervals

We take into account a pair $(\bar{P}, \bar{Q}) \in \mathbb{P}^2$ of negotiation packages evaluated by scoring function (7). On the space $\mathbb{P}$, we can consider the relation $\bar{P}.NW.\bar{Q}$, which reads:

$$\text{Package } \bar{P} \text{ is not worse than package } \bar{Q} \tag{10}$$

For fixed $\gamma \in [1,2]$, the relation is equivalent to the inequality

$$\hat{S}(\bar{P}|\gamma) \geq \ \hat{S}(\bar{Q}|\gamma) \tag{11}$$

It is evident that the fulfilment of condition (11) depends on the value γ. Therefore, we define the relation (10), as fuzzy one determined by its membership function $\mu_{NW}: \mathbb{P} \times \mathbb{P} \to [0,1]$ given as follows

$$\mu_{NW}(\bar{P}, \bar{Q}) = \int\limits_{\{\gamma\in[1,2]:\hat{S}(\bar{P}|\gamma)\geq\hat{S}(\bar{Q}|\gamma)\}} \mathrm{d}x \tag{12}$$

For the cases $B(\bar{Q}) \leq A(\bar{P})$ or $B(\bar{P}) < A(\bar{Q})$, the value $\mu_{NW}(\bar{P}, \bar{Q})$ is obvious. Considering the other cases, we will apply the following linear approximation of the scoring function

$$\hat{S}(\bar{P}|\gamma) \approx (2-\gamma) \cdot A(\bar{P}) + (\gamma - 1) \cdot B(\bar{P}) \tag{13}$$

Then the inequality (11) is replaced by inequality

$$(2-\gamma) \cdot A(\bar{P}) + (\gamma - 1) \cdot B(\bar{P}) \geq (2-\gamma) \cdot A(\bar{Q}) + (\gamma - 1) \cdot B(\bar{Q}) \tag{14}$$

The solution of inequality (14) shows that the approximation of the membership function μ_{NW} is given by means of identity

$$\mu_{NW}(\bar{P},\bar{Q}) = \begin{cases} 1, & A(\bar{P}) \geq A(\bar{Q}) \,\&\, B(\bar{P}) \geq B(\bar{Q}), \\ \left(\dfrac{A(\bar{P}) - A(\bar{Q})}{B(\bar{Q}) - B(\bar{P})} + 1\right)^{-1}, & A(\bar{P}) \leq A(\bar{Q}) < B(\bar{Q}) < B(\bar{P}), \\ \left(\dfrac{B(\bar{Q}) - B(\bar{P})}{A(\bar{P}) - A(\bar{Q})} + 1\right)^{-1}, & A(Q) \leq A(\bar{P}) < B(\bar{P}) < B(\bar{Q}), \\ 0, & A(\bar{P}) < A(\bar{Q}) \,\&\, B(\bar{P}) \leq B(\bar{Q}). \end{cases} \tag{15}$$

From the viewpoint of multivalued logic, the value $\mu_{NW}(\bar{P},\bar{Q})$ is interpreted as the truth-value of the sentence (10). The relation $.NW.$ is a fuzzy preorder in a sense given by Orlovsky [7]. This preorder is linear because we have

$$\forall_{(\bar{P},\bar{Q})\in\mathbb{P}^2}: \qquad \max\{\mu_{NW}(\bar{P},\bar{Q}), \mu_{NW}(\bar{Q},\bar{P})\} \geq \frac{1}{2} \tag{16}$$

Therefore, for any subsets $\mathbb{P}^* \subset \mathbb{P}$, the subset $\max \mathbb{P}^*$ of its maximal elements is distinguished in the following way

$$\max \mathbb{P}^* = \left\{\bar{P}_i \in \mathbb{P}^* : \forall_{\bar{P}_j\in\mathbb{P}^*}: \ \bar{P}_i . NW . \bar{P}_j\right\} \tag{17}$$

The set $max\ \mathbb{P}^*$ is a fuzzy one. Due to the Zadeh's Extension Principle, this fuzzy subset is determined by its membership function $\mu_{max\,\mathbb{P}^*}: \mathbb{P} \to [0,1]$ given as follows

$$\mu_{\max \mathbb{P}^*}(\bar{P}_i) = \min_{\bar{P}_j\in\mathbb{P}^*}\{\mu_{NW}(\bar{P}_i, \bar{P}_j)\} \tag{18}$$

From the viewpoint of multivalued logic, the value $\mu_{max\,\mathbb{P}^*}(\bar{P})$ is interpreted as truth-value of the sentence:

$$\text{Package } \bar{P} \text{ is the best.} \tag{19}$$

A set of maximal elements is a very useful tool for evaluating any sets of non-dominated negotiation packages. We often meet such situations in the subsequent negotiation phase when the parties submit alternative offers.

Example 5: We consider the space $\mathbb{P}_e$ of negotiation packages listed in Example 4. Using scoring intervals presented in Table 5, we determine the fuzzy preorder. The membership function of this relation is presented in Table 6.

	$\bar{P}_1$	$\bar{P}_2$	$\bar{P}_3$	$\bar{P}_4$	$\bar{P}_5$	$\bar{P}_6$
$\bar{P}_1$	1	1	1	1	1	1
$\bar{P}_2$	0	1	1	1	0.110145	0.437097
$\bar{P}_3$	0	0	1	0	0	0
$\bar{P}_4$	0	0	1	1	0	0
$\bar{P}_5$	0	0.889855	1	1	1	0.693759
$\bar{P}_6$	0	0.562903	1	1	0.306241	1

Table 6 The membership function of fuzzy preorder $.NW.$ on $\mathbb{P}_e$.

In the next step, we restrict our consideration to the set $\mathbb{P}_e^* = \{\bar{P}_2, \bar{P}_5, \bar{P}_6\}$ of non-dominated negotiation packages. The relation $.NW.$ on the space $\mathbb{P}_e^*$ and the set $max\ \mathbb{P}_e^*$ are presented in Table 7.

	Relation $.NW.$ on the space $\mathbb{P}_e^*$			**max $\mathbb{P}_e^*$**
	$\bar{P}_2$	$\bar{P}_5$	$\bar{P}_6$	
$\bar{P}_2$	1	0.110145	0.437097	0.110145
$\bar{P}_5$	0.889855	1	0.693759	0.693759
$\bar{P}_6$	0.562903	0.306241	1	0.306241

Table 7 The membership function of fuzzy preorder. $NW.$ on $\mathbb{P}_e^*$ and generated by it set of maximal elements.

4 Final remarks

In this paper, we reconsidered the problem of understanding the Principal's preferences as visualized in the INSPIRE negotiation support system. In the approach proposed by us, the premise "an agent's understanding" is replaced by a more general premise "any agent's understanding". In this way, we received a more reliable rating method with the use of scoring intervals. The price for raising the rating credibility was the reduction of order precision. A fuzzy relation discloses an imprecision of this order.

On the other hand, the results obtained in Example 5 show that disclosed imprecision may have limited coverage. It means that an agent's understanding of the Principal preferences may not impact the comparisons of most packages. In many papers, the negotiation case considered in our examples is a reference point for discussions on INSPIRE. Therefore, the conclusions presented in this paragraph are very important for future scientific discussion.

Econometric verification of the model (5) is an important direction for future research on INSPIRE or other negotiation support systems that use the principal's preference visualizations. Moreover, we suggest using the relation (10) to determine the multiple criteria comparison describing the preferences of both negotiating parties.

The presented model is a normative one. In the future, a discussion should be started about the use of the membership function (15) in negotiation practice.

Acknowledgements

Ewa Roszkowska and Tomasz Wachowicz research was supported by the grant from the Polish National Science Centre (2016/21/B/HS4/01583). We thank professor Gregory Kersten for access to the historical data of the Inspire negotiations.

References

[1] BRINTON, W. C. (1914). *Graphic methods for presenting facts*. New York: The Engineering Magazine Company.

[2] CROXTON, F. E. & STEIN, H. (1932). Graphic comparisons by bars, squares, circles, and cubes. *Journal of the American Statistical Association 27*, 54–60.

[3] KERSTEN, G. E. & NORONHA, S. J. (1999). WWW-based negotiation support: design, implementation, and use *Decision Support Systems* 25, 135–154.

[4] KORHONEN, P. & WALLENIUS, J. (2008). Visualization in the multiple objective decision-making framework. In J. BRANKE, K. DEB, K. MIETTINEN & R. SŁOWIŃSKI (Eds.) *Multiobjective optimization*. (pp. 195–212). Springer, Berlin.

[5] MACDONALD-ROSS, M. (1977). How numbers are shown. *Audio-Visual Communication Review* 25, 359–409.

[6] MIETTINEN, K. (2014). Survey of methods to visualize alternatives in multiple criteria decision making problems. *OR Spectrum* 36, 3–37.

[7] ORLOVSKY, S. A. (1978). Decision making with a fuzzy preference relation. *Fuzzy Sets Systems 1*, 155–167.

[8] PRATT, J. W. & ZECKHAUSER, R. J. (1985) Principals and Agents: An Overview. In: J. W. PRATT & R. J. ZECKHAUSER, (Eds.) *Principals and Agents: The Structure of Business*. (pp. 1–35), Cambridge MA, Harvard Business School Press.

[9] ROSELLI, L. R. P., FREJ, E. A. & DE ALMEIDA, A. T. (2018). Neuroscience experiment for graphical visualization in the FITradeoff decision support system. In: Y. CHEN, G. E. KERSTEN, R. VETSCHERA & H. XU, (Eds.) *Group decision and negotiation in an Uncertain World*. *GDN 2018*. (pp. 56–69) Lecture Notes in Business Information Processing, vol. 315, Cham, Springer.

[10] ROSZKOWSKA, E. & WACHOWICZ, T. (2015). Inaccuracy in Defining Preferences by the Electronic Negotiation System Users, In: B. KAMINSKI, G. E. KERSTEN & T. SZAPIRO, *Outlooks and Insights on Group Decision and Negotiation GND 2015*, (pp. 131–143), Lecture Notes in Business Information Processing, vol. 218, Heidelberg, Springer.

[11] SPENCE, I. & LEWANDOWSKY, S. (1991). Displaying proportions and percentages. *Application Cognitive Psychology 5*, 61–77.

[12] WACHOWICZ, T., KERSTEN, G. E., & ROSZKOWSKA, E. (2019). How do I tell you what I want? Agent's interpretation of principal's preferences and its impact on understanding the negotiation process and outcomes. *Operational Research*, *19*(4), 993–1032.

Discrete Model of Optimal Growth on a Finite Time Horizon as a Boundary Value Problem

Pavel Pražák[1], Kateřina Frončková[2]

Abstract. The aim of this paper is to introduce a way to transform a discrete growth model with a finite time horizon to a system of nonlinear equations that can be solved by a numerical method. Neoclassical growth model is usually presented in continuous time. If an objective utility function is given the growth model can be formulated as optimal control problem. This paper considers a discrete-time growth model on a finite time horizon. First necessary conditions for optimal solution to the problem are introduced. Then Euler equation is developed. The final model can be expressed by a system of two nonlinear difference equations with two boundary values based on these relations. Unfortunately, such a problem cannot be solved analytically. Therefore, the given system of difference equations is rewritten using the system of nonlinear equations which is subsequently solved by a suitable numerical method.

Keywords: difference equations, growth model, numerical methods, system of nonlinear equations

JEL Classification: C61, C63, E21
AMS Classification: 37N40

1 Introduction

Discrete neoclassical model forms the basis for the analysis of economic growth in current macroeconomics textbooks, [3] or [8]. The consumer savings rate s, $s \in [0, 1]$, is considered as an exogenous variable in this model. This fact does not allow consumers to optimize their consumption and can lead to inefficient savings. The more realistic model can assume that the savings rate s is not constant and is determined by optimizing a consumer behavior so that the optimal value of utility from consumption is achieved. Such ideas were proposed in 1928 by Frank Ramsey in [12]. His paper introduced the optimization of consumer behavior in economics. However, this ideas were received much later, only in the 1960s. In this period the optimal control theory was also integrated into economic models. Tjalling C. Koopmans published paper [7] in 1963 and David Cass published paper [4] in 1965. They developed the first endogenous neoclassical model. Using the Ramsey analysis of consumer optimization, they obtained an endogenous determination of the savings rate. We deal with a discrete variant of this growth model on the finite time horizon, cf. also [10], [11], in this paper. The final model is described using a system of two nonlinear difference equations with boundary values that cannot be solved analytically. Therefore the system is rewritten into the system of nonlinear equations and then its numerical solution is introduced.

2 Model

The construction of the model can be described in two steps. First the dynamics of the problem is derived and then the optimization problem is introduced. It is also assumed that the economy is closed.

2.1 Dynamics of Capital and Output

We suppose the dynamic as in the neoclassical growth model, [3]. Consequently we assume that the labour force L_t, $L_t \geq 0$, at the beginning of period t, $t \in \mathbb{N}_0$, grows at the rate n, $n \in [0, 1]$, and that the capital K_t at period t is depreciated by the rate of δ, $\delta \in [0, 1]$. It is then possible in a given closed economy to write the equation for resource constraints

$$C_t + K_{t+1} - (1 - \delta)K_t = F(K_t, L_t),$$

[1] University of Hradec Králové, Faculty of Informatics and Management, Department of Informatics and Quantitative Methods, Rokitanského 62, Hradec Králové, Czech Republic, pavel.prazak@uhk.cz
[2] University of Hradec Králové, Faculty of Informatics and Management, Department of Informatics and Quantitative Methods, katerina.fronckova@uhk.cz

which means that the consumption C_t, $C_t \geq 0$, at period t plus the investment $K_{t+1} - (1-\delta)K_t$ at period t equals to the output $F(K_t, L_t)$ of the given closed economy at period t. The model assumes that the production function F is neoclassical, [8]. Capital stocks at the end of the period t are denoted as K_{t+1}. They are used for production in the following period $t+1$. If we divide the equation by the current labour force L_t, we gain

$$\frac{C_t}{L_t} + \frac{K_{t+1}}{L_{t+1}} \cdot \frac{L_{t+1}}{L_t} - (1-\delta)\frac{K_t}{L_t} = \frac{F(K_t, L_t)}{L_t} = F\left(\frac{K_t}{L_t}, 1\right) = f\left(\frac{K_t}{L_t}\right),$$

where we use the fact that the production function F is a homogeneous function of degree one, which means that it complies with the assumption of constant returns to scale, [8]. If we denote $k_t = K_t/L_t$ capital equipment of labour or capital-labour ratio, $c_t = C_t/L_t$ consumption per unit of labor and realize that $1+n = L_{t+1}/L_t$, the previous equation can be shortly written as

$$c_t + (1+n)k_{t+1} - (1-\delta)k_t = f(k_t).$$

This equation can be further rewritten to the form of the difference equation

$$k_{t+1} = \frac{1}{1+n}(f(k_t) + (1-\delta)k_t - c_t). \tag{1}$$

It expresses the value of capital equipment of labour at the end of each period t, $t \in \mathbb{N}_0$, as a function of current capital equipment and current consumption.

2.2 Objective function

The key problem is to distribute consumption optimally over the finite time horizon of the length T period so that consumers get the most utility from their consumption. In other words, the model enables a central planner who is interested in consumer satisfaction and can influence their behavior to plan and select the optimal consumption c_t, $t \in [0 : T-1]$ to maximize the utility of the economy. Suppose, that the aggregate utility from consumption in each period t, $t \in [0 : T-1]$, is to be maximized

$$U(c_0, c_1, \ldots, c_{T-1}) = u(c_0) + \beta u(c_1) + \cdots + \beta^{T-1}u(c_{T-1}) = \sum_{t-0}^{T-1} \beta^t u(c_t), \tag{2}$$

where β, $\beta \in [0, 1]$, is a discount factor expressing consumer time preferences and u is a utility function. It is assumed that the utility function u is an increasing concave function defined for all $c \in [0, \infty)$ for which the Inada conditions $\lim_{c\to 0+} u(c) = \infty$, $\lim_{c\to\infty} u(c) = 0$ hold, for more details see [1].

2.3 Optimization Problem

All the previous consideration can be summarized to gain the following discrete-time optimal control problem

$$\max \sum_{t=0}^{T-1} \beta^t u(c_t), \tag{3}$$

subject to

$$k_{t+1} = \frac{1}{1+n}(f(k_t) + (1-\delta)k_t - c_t), \tag{4}$$

with initial condition

$$k_0 = \underline{k}, \tag{5}$$

and with conditions of nonnegativity

$$k_t \geq 0, \qquad t \in [1 : T], \tag{6}$$

$$c_t \geq 0, \qquad t \in [0 : T-1]. \tag{7}$$

The capital-labour ratio k_t is the **state variable** and the consumption c_t is the **control variable** in this problem. It is necessary to find the optimal process $(\widehat{k}_t, \widehat{c}_t)$, where $t \in [0 : T-1]$.

3 Results and Discussions

Our possible solution to the problem (3) – (7) is mainly based on the Fritz John and Karush-Kuhn-Tucker conditions, [2],[13]. Particularly a set of necessary conditions is formulated and analyzed step by step. Finally a system of two difference equations is derived and solved.

3.1 Necessary Conditions for Optimal Solution

A discrete variant of Maximum principle, [9], can be used to find necessary conditions for the optimal process of the problem (3) – (7). The Hamiltonian function of the problem is

$$H(t, k_t, c_t, \lambda_0, \lambda_{t+1}) = \lambda_0 \beta^t u(c_t) + \lambda_{t+1} \cdot \frac{1}{1+n}(f(k_t) + (1-\delta)k_t - c_t).$$

Nonnegative conditions (6) and (7) can be considered in the functional form $m(t, k_t) = -k_t \leq 0$ and $n(t, k_t, c_t) = -c_t \leq 0$ respectively. If there is the optimal solution, there also exists a nonnegative constant $\widehat{\lambda}_0$ and finite sequences of real numbers

$$\begin{aligned} \widehat{\lambda} &= (\widehat{\lambda}_1, \widehat{\lambda}_2, \ldots, \widehat{\lambda}_T), \\ \widehat{\mu} &= (\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_T), \\ \widehat{\nu} &= (\widehat{\nu}_0, \widehat{\nu}_1, \ldots, \widehat{\nu}_{T-1}), \end{aligned}$$

such that the following relations hold

- nonzero vector of multipliers

$$(\widehat{\lambda}_0, \widehat{\lambda}, \widehat{\mu}, \widehat{\nu}) \neq 0 \in \mathbb{R}^{3T+1}, \tag{8}$$

- the fist-order conditions for all $t \in [0 : T-1]$

$$\begin{aligned} &\frac{\partial H}{\partial c_t}(t, \widehat{k}_t, \widehat{c}_t, \widehat{\lambda}_0, \widehat{\lambda}_{t+1}) - \widehat{\nu}_t \frac{\partial n}{\partial c_t}(t, \widehat{k}_t, \widehat{c}_t) \\ &\quad = \widehat{\lambda}_0 \beta^t u'(\widehat{c}_t) - \widehat{\lambda}_{t+1} \cdot \frac{1}{1+n} + \widehat{\nu}_t \quad = \quad 0, \end{aligned} \tag{9}$$

- adjoint difference equation for $t \in [1 : T-1]$

$$\begin{aligned} \widehat{\lambda}_t &= \frac{\partial H}{\partial k_t}(t, \widehat{k}_t, \widehat{c}_t, \widehat{\lambda}_0, \widehat{\lambda}_{t+1}) - \widehat{\mu}_t \frac{\partial m}{\partial k_t}(t, \widehat{k}_t) - \widehat{\nu}_t \frac{\partial n}{\partial k_t}(t, \widehat{k}_t, \widehat{c}_t) \\ &= \widehat{\lambda}_{t+1} \cdot \frac{1}{1+n}(f'(\widehat{k}_t) + (1-\delta)) + \widehat{\mu}_t, \end{aligned} \tag{10}$$

- transversality condition

$$\widehat{\lambda}_T = -\widehat{\mu}_T \frac{\partial m}{\partial k_T}(T, \widehat{k}_T) = \widehat{\mu}_T, \tag{11}$$

- conditions of complementarity

$$\widehat{\mu}_t \geq 0 \text{ and } \widehat{\mu}_t \cdot \widehat{k}_t = 0, \quad t \in [1 : T], \tag{12}$$

$$\widehat{\nu}_t \geq 0 \text{ and } \widehat{\nu}_t \cdot \widehat{c}_t = 0, \quad t \in [0 : T-1]. \tag{13}$$

To start the analysis of introduced necessary conditions we consider that there is a period t, $t \in [0 : T-1]$, such that $c_t = 0$. Then $u'(c_t) = \infty$ and from the condition (9) it can be found that $\widehat{\lambda}_0 = 0$. It means that the problem (3) – (7) is not regular. We do not deal with this unregular problem and therefore we can assume that $c_t > 0$ for all t, $t \in [0 : T-1]$. This condition can be interpreted in such a way that the consumer optimizes his consumption when there is a nonzero consumption in each period. It also means that the consumer avoids starving in each period. If we consider the condition of complementarity (13) it means that $\widehat{\nu}_t = 0$ for every $t \in [0 : T-1]$. Now it is possible to show that the problem is regular. At the beginning let us consider that $\widehat{\lambda}_0 = 0$. It means that $\widehat{\lambda}_{t+1} = 0, \quad t \in [0 : T-1]$ according to the stationarity conditions (9). If transversality condition (11) is used then $\widehat{\mu}_T = \widehat{\lambda}_T = 0$. Moreover, it is clear from adjoint equation (10) that $\widehat{\mu}_t = 0$ for all $t \in [1 : T-1]$. To summarize the previous findings, all multipliers are zero, which is a contradiction to the condition (8). It means that the opposite is true and $\widehat{\lambda}_0 = 1$. The problem (3) – (7) with the condition $c_t > 0$ for all t, $t \in [0 : T-1]$, is therefore regular.

One of the attribute of utility function u is $u'(\cdot) > 0$. If the first-order conditions (9) are used it is possible to observe that $\widehat{\lambda}_{t+1} = (1+n)\beta^t u'(\widehat{c}_t) > 0$ for all periods t, $t \in [0 : T-1]$. In particular for $t = T-1$ the condition $\widehat{\lambda}_T > 0$ is true. The direct consequence of this observation and the transversality condition (11) is that $\widehat{\mu}_T > 0$. It also means according to the condition of complementarity (12) that

$$\widehat{k}_T = 0. \tag{14}$$

This terminal value can be interpreted in the following way: at the end of the time horizon T all stocks of capital k are consumed. As it holds $\widehat{c}_t > 0$, the state equation (4) implies that

$$\widehat{c}_t = f(\widehat{k}_t) + (1-\delta)\widehat{k}_t - (1+n)\widehat{k}_{t+1} > 0, \quad t \in [0 : T-1,].$$

It means that

$$f(\widehat{k}_t) + (1-\delta)\widehat{k}_t > (1+n)\widehat{k}_{t+1} \geq 0.$$

As the condition $f(0) = 0$ holds for the intensive production function f, it is necessary that $\widehat{k}_t > 0$ for all $t \in [1 : T-1]$. It means that $\mu_t = 0$, for all t, $t \in [1 : T-1]$.

On the basis of the condition of complementarity (12). The given considerations enable to simplify necessary conditions (9) – (13) for optimal proces of the problem (3) – (7). It is possible to write the first order conditions as

$$\widehat{\lambda}_{t+1} = (1+n)\beta^t u'(\widehat{c}_t) > 0, \quad t \in [0 : T-1], \tag{15}$$

instead of these conditions and adjoint equation as

$$\widehat{\lambda}_t = \widehat{\lambda}_{t+1} \cdot \frac{1}{1+n}(f'(\widehat{k}_t) + (1-\delta)), \quad t \in [1 : T-1]. \tag{16}$$

If the shift $t \mapsto t+1$ is assumed in adjoint equation (16) and the relation (15) is substituted there the following equation can be derived

$$u'(\widehat{c}_t) = \beta \cdot \frac{1 + f'(\widehat{k}_{t+1}) - \delta}{1+n} u'(\widehat{c}_{t+1}), \quad t \in [0 : T-2]. \tag{17}$$

That is the Euler equation whose economical interpretation can be found e.g. in [10].

3.2 Optimal Process as Solution to Discrete Boundary Value Problem

The optimal process can be found as a solution to the system of equations (4) and (17) with the initial value of the capital-labour ratio (5) and the terminal value of capital-labour ratio (14). To continue the analysis it is useful to specify the utility function u from consumption c as $u(c) = \ln c$ and the production function f of per-capita k as Coob-Douglas profuction function $f(k) = Ak^\alpha$, where A, $A > 0$, is a given level of technology and α, $\alpha \in (0, 1)$. Then the system (4) and (17) has the form

$$\widehat{k}_{t+1} = \frac{1}{1+n}(A\widehat{k}_t^\alpha + (1-\delta)\widehat{k}_t - \widehat{c}_t), \tag{18}$$

$$\widehat{c}_{t+1} = \frac{\beta}{1+n}(A\alpha\widehat{k}_t^{\alpha-1} + 1 - \delta)\widehat{c}_t \tag{19}$$

with the given boundary values

$$\widehat{k}_0 = \underline{k}, \ \widehat{k}_T = 0. \tag{20}$$

To solve the system it is necessary to find a value $\widehat{c}_0$ such that the process that starts at $\widehat{k}_0 = \underline{k}$ terminates at $\widehat{k}_T = 0$.

3.3 Steady-State

It is possible to find a steady-state (c°, k°) of the system (18) and (19) regardless of boundary values (20). The constant solution to this system is

$$k^\circ = \left[\frac{1}{A\alpha}\left(\frac{1+n}{\beta} - (1-\delta)\right)\right]^{\frac{1}{\alpha-1}}, \ c^\circ = A(k^\circ)^\alpha - (n+\delta)k^\circ. \tag{21}$$

3.4 Numerical Solution to System of Nonlinear Equations

The system (18), (19) with boundary conditions (20) can be recursively solved setting $n = 0$ and $\delta = 1$, which means that the labour force does not increase during the time and also that the capital is totally depreciated at each time period. Details of this special solution can be found in [10]. Full depreciation of capital at each period is rather a rare situation and such an assumption is not very realistic. If $\delta \in [0, 1)$ and $n \in (0, 1]$ then it is not possible to simplify difference equations (18), (19) and to find a solution directly. It would be desirable to find a solution that is based on a numerical technique instead. If the shift $t \to t+1$

in (18) is done and (19) is substituted in this new equation the following nonlinear system of equations can be found

$$F(\widehat{k}_t, \widehat{k}_{t+1}, \widehat{k}_{t+2}) = 0, \quad t \in [0 : T-2] \tag{22}$$

where

$$F(k_t, k_{t+1}, k_{t+2}) = \frac{Ak_{t+1}^{\alpha} + (1-\delta)k_{t+1} - (1+n)k_{t+2}}{Ak_t^{\alpha} + (1-\delta)k_t - (1+n)k_{t+1}} - \frac{\beta}{1+n}\left(A\alpha k_{t+1}^{\alpha-1} + (1-\delta)\right) \tag{23}$$

and with the boundary values (20). The system (22) contains $T-1$ nonlinear equations with $T-1$ unknowns $\widehat{k}_1, \widehat{k}_2, \ldots, \widehat{k}_{T-1}$. It can be solved by a suitable numerical method, [2]. The function `fsolve` that is implemented in Matlab was successfully applied. In this function the Trust-Region-Dogleg algorithm is employed, [5]. The particular Matlab implementation of the system of nonlinear equations (22) is introduced in the Appendix. As soon as the optimal path of capital-labour ratio $\widehat{k}$ is known the relation (18) can be used to find optimal development of consumption $\widehat{c}$. In particular

$$\widehat{c}_t = A\widehat{k}_t^{\alpha} + (1-\delta)\widehat{k}_t - (1+n)\widehat{k}_{t+1}, \; t \in [0 : T-1].$$

Having values of parameters determined the optimal proces can be calculated, cf. Figure 1.

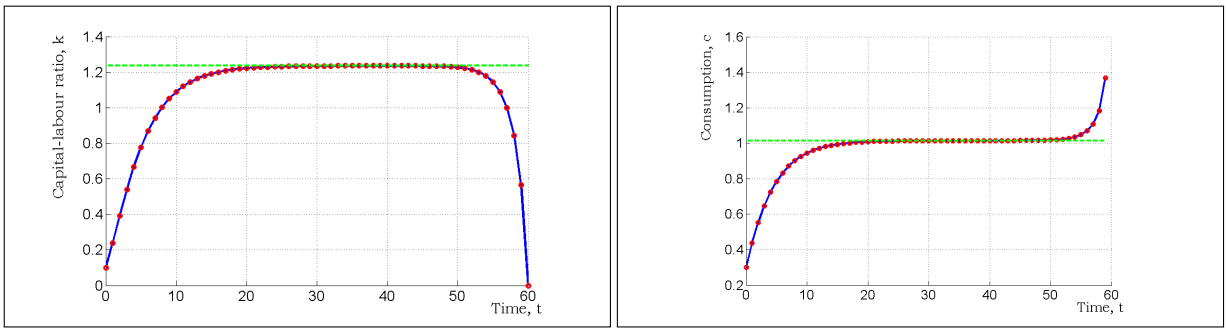


Figure 1 The left panel depicts the time graph of optimal capital-labour ratio $\widehat{k}$. Calculations for the initial value of capital-labour ratio $k_0 = 0.1$, the discount factor $\beta = 0.8$, the coefficients $\alpha = 0.35$ and $n = 0.02$, the coefficient of capital depreciation $\delta = 0.03$ and the time horizon $T = 60$ were performed and are presented here. The dash line represents the steady-state $k^{\circ} \doteq 1.24$. The right panel depicts the time graph of optimal consumption $\widehat{c}$ with the steady-state $c^{\circ} \doteq 1.02$.

4 Conclusion

The paper introduced one way to transform the discrete optimal control problem on a finite time horizon to the system of nonlinear equations that can be subsequently solved by an adequate numerical method. In particular, a simulation model was designed for a special type of growth model. Then it was shown that the optimal process is very close to the steady-state for most periods, but this state is not achieved, see Figure 1. The consumption coordinator cannot immediately select the steady-state values in each period for several reasons: the initial value of capital-labour ratio $k_0 = \underline{k}$ is likely different from the stationary value k°, further, the terminal value of the capital-labour ratio $\widehat{k}_T = 0$ is also different from k° and finally the process is on the finite time horizon T, in which case it is necessary to adapt the course to boundary conditions. This means that the optimal process will only be the best approximation to the steady state. Figure 1 confirms these observations. The capital-labour ratio starts to rise rapidly at the beginning, then it remains constant and falls rapidly to zero at the end of the time horizon to fulfill the end condition. Such an optimum solution is called the **turnpike** property in the literature, see [6]. Results for different values of parameters of the model will be studied in future work.

5 Appendix

Matlab implementation of the system (22) and the nonlinear function (23) has three parts. First the function with all parameters is written. Then the inline function with specific values parameters and one independent variable k, $k \in \mathbb{R}^{T-2}$, is introduced. Finally the function `fsolve` is used to solve the given system of nonlinear equations. The function with the above-introduced parameters $k_0, k_T, T, A, \alpha, \beta, \delta$ and n can be written as follows

```
function y=capital(k, k0, kT, T, A, alpha, beta, delta, n)
equation=zeros(T-1, 1);
k(T)=kT;
for i=[1:T-1]
  if i==1
     equation(1)=((A*k(1)^alpha+(1-delta)*k(1)-(1+n)*k(2))...
        /(A*k0^alpha+(1-delta)*k0-(1+n)*k(1)))...
         -beta/(1+n)*(A*alpha*k(1)^(alpha-1)+(1-delta));
  else
     equation(i)=((A*k(i)^alpha+(1-delta)*k(i)-(1+n)*k(i+1))...
        /(A*k(i-1)^alpha+(1-delta)*k(i-1)-(1+n)*k(i)))...
         -beta/(1+n)*(A*alpha*k(i)^(alpha-1)+(1-delta));
  end
end
y=equation;
end
```

Having set the particular values of parameters $k_0, k_T, T, A, \alpha, \beta, \delta$ and n it is possible to introduce a special inline function with these parameters

```
F=@(k) capital(k, k0, kT, T, A, alpha, beta, delta, n);
```

Now it is sufficient to call the following sequence of Matlab commands

```
init=0.4*ones(T-1, 1);
kstred=fsolve(F, init);
```

where `init` is the initial estimation of sequence $\widehat{k}_1, \widehat{k}_2, \ldots, \widehat{k}_{T-1}$ that is necessary to use the function `fsolve` correctly. The optimal trajectory of capital-labour rate $\widehat{k}$ is found with the boundary values (20) based on the results of numerical solution to system (22) and it is saved in the variable `kstred`.

Acknowledgements

Support of the Specific research project of the Faculty of Informatics and Management of University of Hradec Králové is kindly acknowledged. Special thanks belong to prof. Antonín Slabý for his valuable recommendations.

References

[1] Barro, R. & Sala-i-Martin, X. (2004). *Economic Growth.* Cambridge MA: MIT Press.

[2] Bazaraa, M. S., Sherali, H.D. & Shetty, C.M. (2006). *Nonlinear Programming, Theory and Algorithms*, New Jersey: John Wiley & Sons.

[3] Blanchard, O. (2000). *Macroeconomics*, 2nd ed. New Jersey: Prentice Hall.

[4] Cass, D. (1965). Optimum Growth in an Aggregative Model of Capital accumulation. *The Review of Economic Studies*, 32(3), 233–240.

[5] Forsythe, G. E., Malcolm, M.A. & Moler, C. B. (1976). *Computer Methods for Mathematical Computations*, Prentice-Hall.

[6] Heer, B. & Maußner, A. (2005). *Dynamic General Equilibrium Modelling, Computational Methods and Application*. Berlin-Heidelberg: Springer.

[7] Koopmans, T. C. (1963). On the Concept of Optimal Economic Growth, *Cowles Foundation Discussion Papers* 163, Cowles Foundation for Research in Economics, Yale University.

[8] Lovell, M. C. (2006). *Economics with calculus.* Hackensack, NJ: World Scientific Pub.

[9] Miao, J. (2014). *Economic Dynamics in Discrete Time.* Cambridge MA: MIT Press.

[10] Pražák, P. (2013). Discrete Neoclassical Model of Optimal Growth on a Finite Time Horizon. In Jedlička, P. (Ed.), *International Scientific Conference on Hradec Economic Days 2013 – Economic Development and Management Region* (pp. 117–124). Hradec Kralove: Gaudeamus.

[11] Pražák, P. (2014). Discrete Dynamic Economic Models in Excel. In Houska, M., Krejci, I., Flegl, M. (Eds.), *11th International Conference on Efficiency and Responsibility in Education* (pp. 595–601). Prague: Czech University of Life Sciences.

[12] Ramsey, F. P. (1928). A Mathematical Theory of Saving, *The Economic Journal*, 38(152), 543–559.
[13] Sydsaeter, K., Hammond, P., Seierstad, A. & Strøm, A. (2005). *Further Mathematics for Economic Analysis.* Harlow: Prentice Hall.

A Note on Geometry of Jaeckel's Dispersion Function

Miroslav Rada[1], Michal Černý[2]

Abstract. Rank estimators of parameters of linear regression models have been designed as estimators robust to outliers. The R-estimator is defined as a minimizer of Jaeckel's dispersion function (with respect to a particular score function). We investigate the geometry of Jaeckel's dispersion function and discuss its geometric properties essential for optimization algorithms, admitting the score function to belong to a very general class of functions. We also point out some drawbacks of the package `Rfit` implemented in `R` for robust linear regression. We introduce a Python+LaTeX tool `RankStat` for visualization of Jaeckel's function allowing a user to study how Jaeckel's function behaves in various situations, such as when the addition of a new observation to the regression model (e.g. in the case of streamed data) or a change of the score function.

Keywords: robust regression, rank estimator, Jaeckel dispersion function

JEL Classification: C13; C61
AMS Classification: 62F35; 90-04

1 Introduction

In linear regression, rank-estimators (R-estimators for short) play an important role due to their insensitivity to outliers. The R-estimator is defined as a minimizer of Jaeckel's dispersion function $\mathfrak{D}$, a special measure of variability of residuals with interesting geometric and combinatorial properties. First we define $\mathfrak{D}$ formally; then we interpret its geometry in terms of arrangements of hyperplanes, an important concept in combinatorial geometry. We also summarize some of its algorithmic properties necessary for design of efficient algorithms for its minimization (i.e., computation of the R-estimator) and show drawbacks of the implementation of the `R`-package `Rfit` for R-estimated linear regression models. Finally, we present a new Python+LaTeX tool `RankStat` for visualization of the geometry of $\mathfrak{D}$ for regression models with two explanatory variables (and a constant) and demonstrate some of its capabilities.

2 The score function and Jaeckel's dispersion function $\mathfrak{D}$

Linear regression. We consider the linear regression model

$$y = X\theta^* + \varepsilon,$$

where X stands for the $(n \times p)$-matrix of regressors, possibly including the all-one column corresponding to the constant term, $y \in \mathbb{R}^n$ stands for the (observations of) the dependent variable, $\theta^* \in \mathbb{R}^p$ stands for the vector of regression parameters to be estimated and $\varepsilon \in \mathbb{R}^n$ stands for the vector of error terms. In the entire text, the number of observations will be denoted by n and the number of regression parameters will be denoted by p. As usual, tuple (X, y) is input data and the goal is to estimate θ^*. For the sake of brevity, we skip the discussion on the assumptions on the error terms ε under which the R-estimator "works well" and we refer the reader to Hettmansperger's and McKean's book [8].

Score function. An R-estimator $\widehat{\theta}$ of regression parameters θ^* is defined relatively with respect to an in-advance fixed *score function* $\varphi : (0, 1) \to \mathbb{R}$, which is assumed to be continuous. In statistical literature, some additional properties of φ are often required (for details see [8]). In this text it will be sufficient to distinguish two cases, whether or not φ is non-decreasing. Some examples of score functions used in practice will be discussed below in this Section. Later, in Section 5, we will admit an arbitrary score function.

[1] University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, miroslav.rada@vse.cz
[2] University of Economics, Faculty of Informatics and Statistics, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, cernym@vse.cz

Consistent permutations of residuals. Let the ith row of X be denoted by x_i^{T}. Let $\mathcal{S}_n$ stand for the collection of permutations of the set $\{1, \ldots, n\}$. If $\theta \in \mathbb{R}^p$ and $\pi \in \mathcal{S}_n$ satisfies

$$y_{\pi(1)} - x_{\pi(1)}^{\mathrm{T}}\theta \leq y_{\pi(2)} - x_{\pi(2)}^{\mathrm{T}}\theta \leq \cdots \leq y_{\pi(n)} - x_{\pi(n)}^{\mathrm{T}}\theta, \tag{1}$$

we say that π is a permutation *consistent* with θ. If θ is interpreted as a candidate estimate of θ^*, then $y_{\pi(i)} - x_{\pi(i)}^{\mathrm{T}}\theta$ can be understood as the ith residual after sorting the residuals in a nondecreasing sequence, or *residual with rank i* (with respect to θ). This notion justifies why $\widehat{\theta}$ is called "rank-estimator".

If all of the inequalities in (1) are satisfied strictly, then there exists a single permutation consistent with θ. This permutation is denoted by π^θ. If some residuals are equal, then there exist multiple consistent permutations. In this case, π^θ will stand for *any* of the consistent permutations (say, for example, the lexicographically smallest one). It will be apparent that the forthcoming results are independent of the choice of the consistents permutation if multiple exist.

Definition of Jaeckel's $\mathfrak{D}$. Given a score function φ and data $X \in \mathbb{R}^{n\times p}, y \in \mathbb{R}^n$, Jaeckel's dispersion function is defined as

$$\mathfrak{D}(\theta) \equiv \mathfrak{D}_{\varphi,X,y}(\theta) = \sum_{i=1}^{n} \varphi\left(\frac{i}{n+1}\right)\left(y_{\pi^\theta(i)} - x_{\pi^\theta(i)}^{\mathrm{T}}\theta\right).$$

Indeed, it is easy to see that $D(\theta)$ does not change if a permutation consistent with θ is replaced by another permutation consistent with θ (if multiple consistent permutations exist).

The value $\mathfrak{D}(\theta)$ can be interpreted as a weighted sum of residuals, where the score function determines the weights, and the consistent permutation assigns the weights to the residuals according to their ranks. For example, if $\varphi(\xi) = \xi - 1/2$ for $0 < \xi < 1$, $n = 5$ and θ is such that $r_3 < r_2 < r_5 < r_1 < r_4$, where $r_i = y_i - x_i^{\mathrm{T}}\theta$, then $\mathfrak{D}(\theta) = -\frac{1}{3}r_3 - \frac{1}{6}r_2 + \frac{1}{6}r_1 + \frac{1}{3}r_4$ (the weight corresponding to r_5 is zero).

R-estimator. The R-estimator $\widehat{\theta}$ is a minimizer of $\mathfrak{D}(\theta)$. In general, if a minimizer exists, it need not be unique; then, any representative $\widehat{\theta}$ such that $\mathfrak{D}(\widehat{\theta}) = \min_{\theta\in\mathbb{R}^p} \mathfrak{D}(\theta)$ is taken as the estimate of θ^*.

The R-estimator it is defined relatively to a particular score function φ. Some of well-known representatives of score functions include the *sign score function* $\varphi(\xi) = \mathrm{sgn}(\xi - 1/2)$, *Wilcoxon score function* $\varphi(\xi) = \sqrt{12}(\xi - 1/2)$ or *van der Waerden score function* $\varphi(\xi) = \Phi^{-1}(\xi - 1/2)$, where Φ stands for the cumulative distribution function of $N(0, 1)$. Observe that all of these functions are nondecreasing and satisfy

$$\sum_{i=1}^{n} \varphi\left(\tfrac{i}{n+1}\right) = 0. \tag{2}$$

Property (2) implies that $\mathfrak{D}$ is bounded from below; i.e., a minimizer exists (although it need not be unique).

3 Geometry of $\mathfrak{D}$

An arrangement of hyperplanes and cells. Consider the family of hyperplanes, often referred to as *hyperplane arrangement* in combinatiorial geometry,

$$\mathcal{H}_{ij} := \{\theta \in \mathbb{R}^p \mid (x_j - x_i)^{\mathrm{T}}\theta \leq y_j - y_i\}, \quad 1 \leq i < j \leq n \tag{3}$$

in the space $\mathbb{R}^p$ of regression parameters. The hyperplanes cut the space into a finite number of regions called *faces (of an arrangement)*. The fulldimensional regions are called *cells*. More formally, $\mathbb{R}^p \setminus \bigcup_{1\leq i<j\leq n} \mathcal{H}_{ij}$ is a union of a finite number of connected regions. If C' is such a region, the closure of C' is called *a cell*. The family of all cells is denoted by $\mathfrak{C}$.

Property 1. If $C_1, C_2 \in \mathfrak{C}$ and $C_1 \neq C_2$, then $C_1 \cap C_2$ is a zero-measure set (in Lebesgue sense) and $\bigcup_{C\in\mathfrak{C}} C = \mathbb{R}^p$.

Property 2. A cell is a convex polyhedron.

Property 3. The function $\mathfrak{D}$ is cell-wise linear. In particular, if the domain of $\mathfrak{D}$ is restricted onto a cell $C \in \mathfrak{C}$, then $\mathfrak{D}$ is a linear function.

Property 4. A cell is characterised by a unique permutation $\pi \in \mathcal{S}_n$: if $C \in \mathfrak{C}$ and θ is an interior point of C, then there exists a unique θ-consistent permutation π_0 and $\mathfrak{D}(\theta) = \sum_{i=1}^{n} \varphi\left(\frac{i}{n+1}\right)(y_{\pi_0(i)} - x_{\pi_0(i)}^{\mathrm{T}}\theta)$ for all $\theta \in C$.

Property 5. The function $\mathfrak{D}$ is continuous on the entire space $\mathbb{R}^p$. In addition, if the score function φ is nondecreasing, then $\mathfrak{D}$ is convex.

4 Minimization of $\mathfrak{D}$

The convex case. The crucial property for minimization of $\mathfrak{D}$ (i.e., computation of the R-estimator) is convexity (recall Property 5). Indeed, the score functions used in robust statistics are always non-decreasing (such as the sign function, Wilcoxon function or van der Waerden function). In the convex case there exists a polynomial-time algorithm for minimization of $\mathfrak{D}$, see [3].

Remark. Even if an algorithm is known, this research area still remains open: the method from [3] is based on the Ellipsoid Method for linear programming with oracle-given separation procedure (along the construction in [6]). It is a polynomial-time procedure in theory, which is extremely hard to implement for numerical reasons (due to the significant role of "Big-L" arguments) and suffers from well-known drawbacks of the Ellipsoid Method, such as computation times close to the worst-case bound. It is a challenging problem whether this version of the Ellipsoid Method can be converted into an Interior Point Method which would yield both a formal proof of polynomiality *and* practical usefulness.

An implementation in `R`. In `R`, there is a package `Rfit` for minimization of $\mathfrak{D}$ with Wilcoxon scores a default. What the package does is minimization of $\mathfrak{D}$ using an approximate numerical method based on the BFGS algorithm (details can be found in [9]). The output is an *approximate* minimizer of $\mathfrak{D}$, with no guarantee of the error

$$\|\theta_{\text{TrueMinimizer}} - \theta_{\text{FoundBy}\texttt{Rfit}}\|. \tag{4}$$

Particularly interesting is the fact that *the `R`-implementation cannot recognize whether or not it has found a minimizer*. For example, if a true minimizer is supplied as an initial guess to the package, it starts iterating without recognizing that there is no work to be done. This is a serious drawback and has a deeper reason. In [2], Property 6 has been proven:

Property 6. Checking whether a given θ is a minimizer of $\mathfrak{D}$ is a **P**-complete problem. (See [5] as a reference book on **P**-completeness theory.)

Property 6 can be loosely interpreted that "checking whether a minimizer has been found cannot be done without linear programming" (and `Rfit` does *not* use LP). To conclude, these two undesirable properties – the inability to recognise a minimizer and no bound on the error (4) – lead to a natural challenge to implement a new package.

The general case. On the contrary to the convex case, if φ is not non-decreasing, then minimization of $\mathfrak{D}$ becomes a difficult task. (This is an informal statement; we currently do not have a proof if the problem is, say, **PSPACE**-hard.) There can exist multiple local minima and it is not easy to enumerate them. As far as we are aware, the only known algorithmic method is based on the enumeration of all cells $C \in \mathfrak{C}$.

Observe that given $C \in \mathfrak{C}$, the optimization problem $\min_{\theta \in C} \mathfrak{D}(\theta)$ can be written as a linear programming problem. Thus, the minimization method requires (i) a method for an enumeration of cells $C \in \mathfrak{C}$ plus (ii) solving a linear programming problem per cell C. A direct method for (i) would be an enumeration of all $\pi \in \mathcal{S}_n$ (due to Property 4); however, much better approaches (based on the theory of arrangements of hyperplanes) are known [1, 4, 10, 3], which can solve the problem in polynomial time if $p = O(1)$ (and $n \to \infty$).

5 `RankStat`: a Python+LaTeX tool for visualization of Jaeckel's $\mathfrak{D}$

This section is devoted to a new Python+LaTeX tool `RankStat` allowing to solve various tasks regarding Jaeckel's $\mathfrak{D}$. Both cases – the convex one and the general one – are covered.

The tool efficiently implements RSIncEnu algorithm from [3], allowing for visualizing $\mathfrak{D}$ corresponding to the linear regression model

$$y_i = \theta_0 + \theta_1 u_i + \theta_2 v_i + \varepsilon_i, \quad i = 1, \dots, n$$

with $p = 3$. Here, $x_i^{\mathrm{T}} = (1, u_i, v_i)$. Observe that in (3), the inequality $(x_j - x_i)^{\mathrm{T}}\theta \leq y_j - y_i$ does not depend on θ_0. Thus, without loss of generality, it is possible to fix $\theta_0 = 0$ (say) and plot the arrangement (3) in the 2-dimensional space of parameters θ_1, θ_2. (To avoid confusion: if condition (2) is satisfied, then $\mathfrak{D}$ does not depend on θ_0 at all. If condition (2) does not hold, then $\mathfrak{D}$ is unbounded. Hence, it makes sense to plot an arbitrary instance in $(p-1)$-dimensional space.) In the following examples, θ_0 is omitted and X is of dimension $n \times 2$.

Here, we demonstrate the tool on three tasks: consecutive adding of observations (i.e. what happens with the arrangement and $\mathfrak{D}$ when observations are added one-by-one, e.g. in a stream), changing $\phi(t)$ (i.e. how $\mathfrak{D}$ is sensitive on changing the weights of residuals).

Technical remarks. The computational logic of the tool is implemented in Python. To create high quality graphics, our tool prepares LaTeX code based on drawing package Ti*k*Z/PGF [11]. The LaTeX code can be either compiled to `pdf`, or output as an `svg` file; the conversion is performed by PGF package.

Another noteworthy dependency is Gurobi solver [7] with its Python module `gurobipy`. It is used for solving linear programs, that occur in the cell enumeration procedure (needed for visualization and for optimizing the noncovex case) and in the procedure for finding the minimum in the convex case.

Since $\mathfrak{D}$ is linear in on every $C \in \mathfrak{C}$, a precise heatmap of $\mathfrak{D}$ with smooth color transitions can be visualized. The tool prepares description of the color transition to be drawn in a concrete cell. Then, Ti*k*Z/PGF translates the description to instructions for a `pdf` or `svg` renderer.

5.1 Increasing number of observations

Here, we consider instance with increasing number of observations. Each entry of $y \in \mathbb{Z}^n$ and $X \in \mathbb{Z}^{n \times 2}$ is sampled from the uniform distribution over $[-50, 50]$. As a score function to determine weights of residuals, the van der Waerden score function is used.

In Figure 1, instances with $n = 4$ to $n = 9$ are visualized, showing how the heatmap evolves when new observations are added. Note that the arrangement is a bit more complex, some cells are cut of the displayed region. Our implementation heuristically removes outlier intersections of hyperplanes and only visualizes the "interesting" part of the heatmap of $\mathfrak{D}$.

Gray lines are hyperplanes defined by (3). Every hyperplane is labeled by indices of the pair of observations determing the hyperplane. The black circle in the middle of every image stands for the global minimum. Also, there is a thin dotted line segment in every cell. These line segments head from the midpoints of cells to the minima over cells.

Note that for 9 observations, the shape of contours of $\mathfrak{D}$ starts to resemble a smooth function. This might be an explanation of the often satisfactory behavior of heuristic algorithms for minimizing $\mathfrak{D}$, c.f. the implementation in `Rfit` package.

5.2 Changing the weights of residuals

Comparison of the common score functions. Here, we take the instance with $n = 6$ from Subsection 5.1 and plot the heatmap of $\mathfrak{D}$ for the score functions mentioned in Section 2: for van der Waerden's score function, for Wilcoxon's score function and for the sign score function.

The sign score function is the "sharpest" one, as one would expect. Note that the contours of van der Warden's and Wilcoxon's score function cannot be easily distinguished. This does not mean that the absolute values of $\mathfrak{D}$ are the same, it rather means that the relative value is almost the same.

The nonconvex case. We take the sign score function and X and y for $n = 5$ from Subsection 5.1.

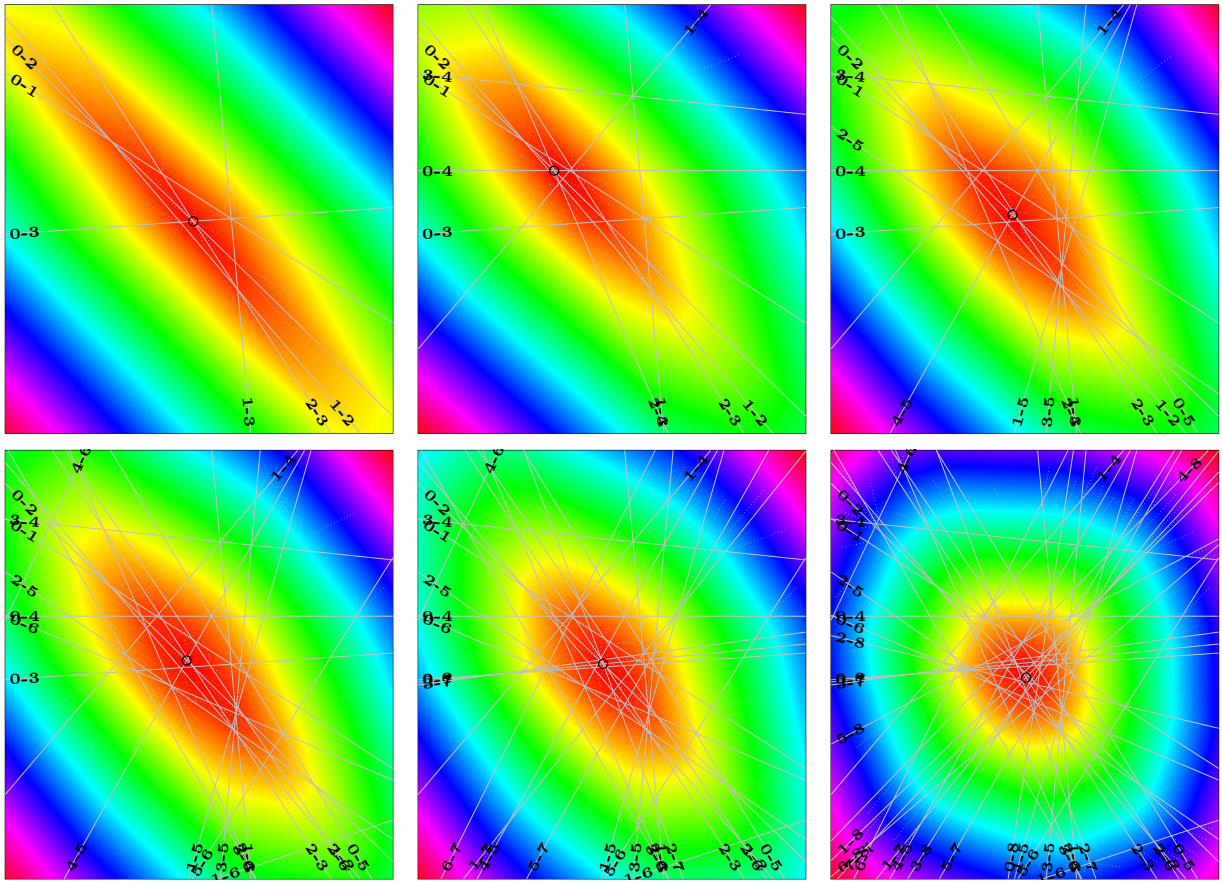


Figure 1 Evolution of the hyperplane arrangement and heatmap of $\mathfrak{D}$ corresponding to 4 to 9 observations.

For $n = 5$, the original sign score function assigns to residuals weights $(-1, -1, 0, 1, 1)^{\mathrm{T}}$. To demonstrate sensitivity of $\mathfrak{D}$ on monotonicity of weights, we examine what happens when we make just one transposition in the weight vector. We interchange the second and the fourth weight (resulting in weights $(-1, 1, 0, -1, 1)^{\mathrm{T}}$; the middle plot in Figure 3) and the third and the fifth weight (to obtain $(-1, -1, 1, 1, 0)^{\mathrm{T}}$; the last plot in Figure 3). The resulting heatmaps are plotted in Figure 3. The first plot is $\mathfrak{D}$ with the original sign score function.

Note that for the weight vector $(-1, 1, 0, -1, 1)^{\mathrm{T}}$, there are multiple global minima, marked with black circles. For the weight vector $(-1, -1, 1, 1, 0)^{\mathrm{T}}$, there is one global minimum, however, the multiple local minima can be found by examination of the dotted line segments.

6 Conclusion

In this paper, we point out some geometric properties of Jaeckel's dispersion function. We illustrate the nice geometry of the function by a new Python+LaTeX visualization tool, called `RankStat`, which is able to plot the function in case of two explanatory variables. The tool also implements efficient algorithms for minimization of Jaeckel's function, both for the convex case (usual in data analysis) and the general case. It is worth noting that the minimization algorithm for the convex case is able to find the precise minimum, unlike the `R` implementation `Rfit`, which has no guaranteed precision of the returned solution.

We demonstrate the tool on several artificial instances, showing how the heatmap of Jaeckel's function changes when adding new observations. We also show how transpositions in the weight vector influence the convexity of the function.

Acknowledgements

The work was supported by the Czech Science Foundation under project 19-02773S.

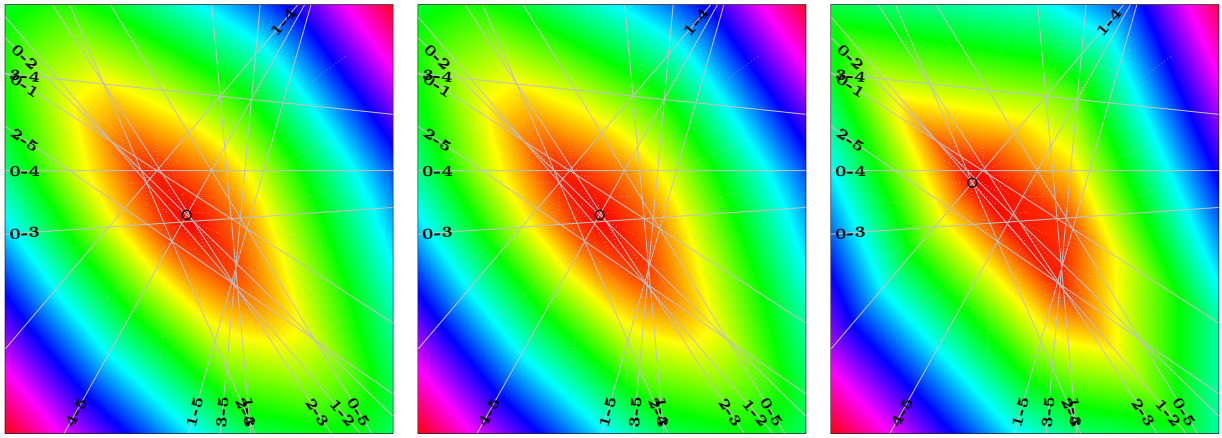


Figure 2 Comparison of heatmaps for van der Waerden's, Wilcoxon's and the sign score function. Here, $n = 6$. Van der Waerden's and Wilcoxon's function cannot be distinguished.

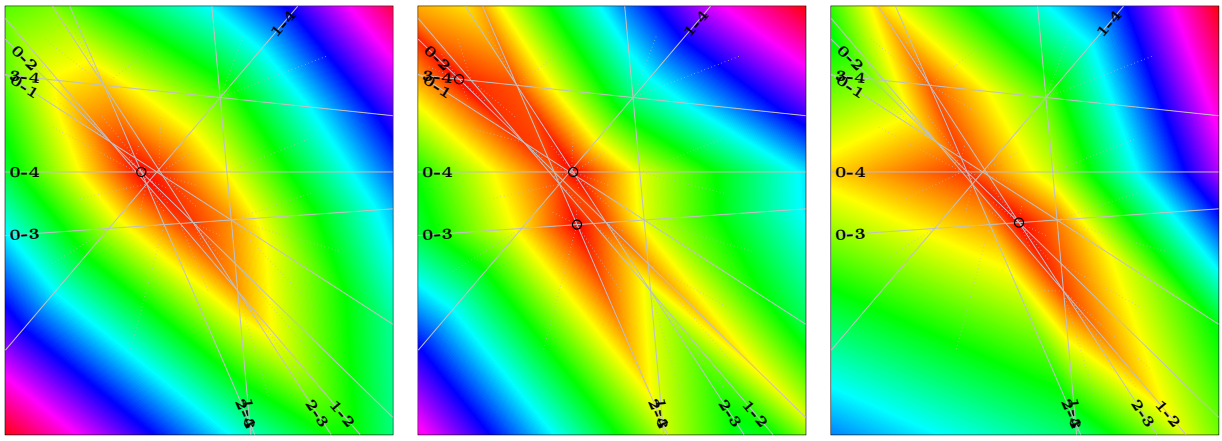


Figure 3 Comparison of heatmaps for sign score function with some weights swapped, details in text. $\mathfrak{D}$ on the first plot is convex. Note the multiple global minima in the second plot. On the third plot, there are also multiple minima, however, only one global.

References

[1] Avis, D., and Fukuda, K.: Reverse search for enumeration. *Discrete Applied Mathematics* **65** (1996), 21–46. URL `http://dx.doi.org/10.1016/0166-218X(95)00026-N`.

[2] Černý, M., Hladík, M., and Rada, M.: Walks on hyperplane arrangements and optimization of piecewise linear functions. *arXiv:1912.12750 [math]* (2019). URL `http://arxiv.org/abs/1912.12750`.

[3] Černý, M., Rada, M., Antoch, J., and Hladík, M.: A class of optimization problems motivated by rank estimators in robust regression. *arXiv:1910.05826 [math]* (2019). URL `http://arxiv.org/abs/1910.05826`.

[4] Ferrez, J. A., Fukuda, K., and Liebling, T. M.: Solving the fixed rank convex quadratic maximization in binary variables by a parallel zonotope construction algorithm. *European Journal of Operational Research* **166** (2005), 35–50. URL `http://www.sciencedirect.com/science/article/pii/S0377221704003352`.

[5] Greenlaw, R., Hoover, J. H., and Ruzzo, W. L.: *Limits to Parallel Computation: P-Completeness Theory*. Oxford University Press, New York, 1995.

[6] Grötschel, M., Lovász, L., and Schrijver, A.: *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, 1993.

[7] Gurobi Optimization, L.: *Gurobi Optimizer Reference Manual*, 2020. URL `http://www.gurobi.com`.

[8] Hettmansperger, T., and McKean, J.: *Robust Nonparametric Statistical Methods, Second Edition*. CRC Press, 2010. URL `https://www.taylorfrancis.com/books/9781439809099`.

[9] Kloke, J., D., and McKean, J., W.: Rfit: Rank-based Estimation for Linear Models. *The R Journal* **4** (2012),

57. URL https://journal.r-project.org/archive/2012/RJ-2012-014/index.html.

[10] Rada, M., and Černý, M.: A New Algorithm for Enumeration of Cells of Hyperplane Arrangements and a Comparison with Avis and Fukuda's Reverse Search. *SIAM Journal on Discrete Mathematics* **32** (2018), 455–473. URL https://epubs.siam.org/doi/10.1137/15M1027930.

[11] Tantau, T.: *The TikZ and PGF Packages - Manual for Version 3.0.0*, 2013. URL http://sourceforge.net/projects/pgf/.

Desirable Properties of Weighting Vector in Pairwise Comparisons Matrix With Fuzzy Elements

Jaroslav Ramík[1]

Abstract. We deal with pairwise comparisons matrix with fuzzy elements (FPCM). Fuzzy elements are appropriate whenever the decision maker (DM) is uncertain about the value of his/her evaluation of the relative importance of elements in question, or, when aggregating crisp pairwise comparisons of a group of decision makers in the group DM problem. The problem is formulated in a general setting investigating pairwise comparisons matrices with elements from abelian linearly ordered group (alo-group). Such an approach enables extensions of traditional multiplicative, additive or fuzzy approaches. Continuing our research in [8], here we propose a new order preservation concept based on alpha-cuts. Then we define an innovative concept of (weak) consistency of FPCMs, propose some desirable properties of priority vectors, and derive necessary and sufficient conditions for the existence of coherent vector (CV) and intensity vector (IV) of a FPCM. Finally, we formulate the optimization problem and derive the priority vector with the desirable properties. Illustrating examples are presented and discussed.

Keywords: multi-criteria optimization, pair-wise comparisons matrix, fuzzy elements, alo-group, consistency, priority vector.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The problem we consider here is as follows: The goal of the decision maker (DM) is to rank the alternatives $\mathcal{C} = \{c_1, c_2, ..., c_n\}$ from the best to the worst (or, vice versa, which is equivalent), using the information given by the decision maker in the form of an $n \times n$ pairwise comparisons matrix (PCM). The ranking of the alternatives is determined by the priority vector of positive numbers $w = (w_1, w_2, ..., w_n)$ which is calculated from the corresponding PCM.

Fuzzy sets as the elements of the pairwise comparisons matrix can be applied in the DM problem whenever the decision maker is not sure about the preference degree of his/her evaluations of the pairs in question. Fuzzy elements are useful in order to capture uncertainty stemming from subjectivity of human thinking and from incompleteness of information that is an integral part of multi-criteria decision-making problems. Fuzzy elements may be also useful as aggregations of crisp pairwise comparisons of a group of decision makers in the group DM problem.

In this paper we analyze the concept of COP from [1], and [6] to FPC matrices defining mean coherent vector, and mean intensity vector of a FPCM based on α-cuts. Then we reconsider Generalized Geometric Mean Method (GGMM) and show that the criteria under a generalized GMM depend on the locally defined inconsistency. Furthermore, we solve the problem of finding optimal priority vector satisfying the desirable properties by minimizing the corresponding global error index. Finally, we solve a numerical example in order to illustrate the proposed concepts and properties.

2 Preliminaries

This paper is a continuation of the paper in [8]. The reader can find the corresponding basic definitions, concepts and results there. Here, we summarize some necessary concepts, however, for a more detailed information we refer to [8], or, [9].

A *fuzzy subset* S of a nonempty set X (or a *fuzzy set* on X) is a family $\{S_\alpha\}_{\alpha\in[0;1]}$ of subsets of X such that $S_0 = X, S_\beta \subset S_\alpha$ whenever $0 \leq \alpha \leq \beta \leq 1$, and $S_\beta = \cap_{0\leq\alpha<\beta} S_\alpha$ whenever $0 < \beta \leq 1$. The *membership*

[1] Silesian University Opava, School of Business Administration in Karvina, University Sq. 1934/3, 733 40 Karvina, Czechia, ramik@opf.slu.cz

function of S is the function μ_S from X into the unit interval $[0;1]$ defined by $\mu_S(x) = \sup\{\alpha \mid x \in S_\alpha\}$. Given $\alpha \in]0;1]$, the set $[S]_\alpha = \{x \in X \mid \mu_S(x) \geq \alpha\}$ is called the α-*cut of fuzzy set* S. A fuzzy subset S of $\mathbf{R}^* = \mathbf{R} \cup \{-\infty\} \cup \{+\infty\}$ is a *fuzzy interval* whenever S is normal and its membership function μ_S satisfies the following condition: S is closed, compact and convex, i.e. the α-cut $[S]_\alpha$ are closed, compact and convex subset of X for every $\alpha \in]0;1]$, respectively. A bounded fuzzy interval S called the *triangular fuzzy number* is denoted by $S = (a, b, c)$. Notice that each crisp number is also a bounded fuzzy interval $S = (a, b, c)$ with $a = b = c$.

In order to unify various approaches and prepare a more flexible presentation, we apply alo-groups, see [7]. Recall that an abelian group, [4], is a set, G, together with an operation $\odot$ and corresponding "group axioms" that combine any two elements $a, b \in G$ to form another element in G denoted by $a \odot b$, see [4]. The well known examples of alo-groups can be found bellow, more examples in [4], or, [5].

Example 1. *Additive alo-group* $\mathcal{R} = (\mathbf{R}, +, \leq)$ is a continuous alo-group with: $e = 0,\ a^{(-1)} = -a$.

Example 2. *Multiplicative alo-group* $\mathcal{R}_+ = (\mathbf{R}_+, \bullet, \leq)$ is a continuous alo-group with: $e = 1,\ a^{(-1)} = a^{-1} = 1/a$. Here, by $\bullet$ we denote the usual operation of multiplication.

Example 3. *Fuzzy additive alo-group* $\mathcal{R}_a$=$(\mathbf{R}, +_f, \leq)$, see [4], is a continuous alo-group with: $a +_f b = a + b - 0.5,\ e = 0.5,\ a^{(-1)} = 1 - a$.

Example 4. *Fuzzy multiplicative alo-group* $\mathcal{R}_m$=$(]0;1[, \bullet_f, \leq)$, see [7], is a continuous alo-group with: $a \bullet_f b = \frac{ab}{ab+(1-a)(1-b)}, e = 0.5, a^{(-1)} = 1 - a$.

3 FPC matrices, reciprocity and consistency

A general approach based on alo-groups is useful, as it unifies various important approaches known from the literature. This fact has been already demonstrated on 4 examples presented above, where the well known alo-groups are shown. Particularly, all concepts and properties which will be presented in the sequel can be easily applied to any alo-group. Before we shall investigate PC matrices with fuzzy elements we remember some concepts and properties of PC matrices on alo-group with crisp elements.

A crisp PC matrix $A = \{a_{ij}\}$ is said to be $\odot$-*reciprocal*, if the following condition holds: For every $i, j \in \{1, ..., n\}$

$$a_{ij} \odot a_{ji} = e, \text{ or, equivalently, } a_{ji} = a_{ij}^{(-1)}. \tag{1}$$

A crisp FPC matrix $A = \{a_{ij}\}$ is $\odot$-*consistent* if for all $i, j, k \in \{1, ..., n\}$

$$a_{ik} = a_{ij} \odot a_{jk}, \text{ or, equivalently, } a_{ij} \odot a_{jk} \odot a_{ki} = e. \tag{2}$$

Remember that an $\odot$-consistent PC matrix $A = \{a_{ij}\}$ is $\odot$-reciprocal, but not vice-versa. The following equivalent condition for consistency of PC matrices is well known, see e.g. [4], [10].

A PCM $A = \{a_{ij}\}$ is $\odot$-consistent if and only if (shortly: iff) there is a vector $w = (w_1, ..., w_n)$, $w_i \in G$, such that

$$a_{ij} = w_i \div w_j \text{ for all } i, j \in \{1, 2, ..., n\}. \tag{3}$$

Here, $w_i \div w_j = w_i \odot w_j^{(-1)}$.

In [7], we extended the above stated definition of $\odot$-reciprocity and $\odot$-consistency to non-crisp matrices with fuzzy elements (shortly, FPCMs). In particular, we introduced a new concept of reciprocity and consistency based on α-cuts: α-$\odot$-reciprocity and α-$\odot$-consistency. Let us start with the α-$\odot$-reciprocity in the fuzzy case.

Let $\mathcal{G} = (G, \odot, \leq)$ be a divisible and continuous alo-group over an open interval G of $\mathbf{R}$, see [4]. Let $\alpha \in [0;1]$, $\tilde{A} = \{\tilde{a}_{ij}\}$ be an $n \times n$ matrix, where each element is a bounded fuzzy interval of the alo-group $\mathcal{G}$, let $[\tilde{a}_{ij}]_\alpha = [a_{ij}^L(\alpha), a_{ij}^R(\alpha)]$ be an α-cut of $\tilde{a}_{ij}$.
Matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is said to be α-$\odot$-*reciprocal*, if the following two conditions hold for each $i, j \in \{1, ..., n\}$:

$$a_{ii}^L(\alpha) = a_{ii}^R(\alpha) = e, \tag{4}$$

$$a_{ij}^L(\alpha) \odot a_{ji}^R(\alpha) = e. \tag{5}$$

If $\tilde{A} = \{\tilde{a}_{ij}\}$ is α-$\odot$-reciprocal for all $\alpha \in [0;1]$, then it is called $\odot$-*reciprocal*.
From now on we assume that $\tilde{A} = \{\tilde{a}_{ij}\}$ is $\odot$-reciprocal. Then $\tilde{A} = \{\tilde{a}_{ij}\}$ is called the *fuzzy pairwise comparisons matrix*, *fuzzy PC matrix, FPC matrix*, or, shortly, *FPCM*.

Now, we turn to the concept of consistency of FPC matrices. We start with the definition of *weak* α-$\odot$-consistent FPC matrix. In [7] this concept was named "α-$\odot$-consistency", here, we use, however, the name *"weak* α-$\odot$-consistency", as later on, we shall define a new and stronger concept, particularly, an α-$\odot$-consistency of FPC matrix.

Definition 1. Let $\alpha \in [0;1]$. A FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is said to be *weak* α-$\odot$-*consistent*, if the following condition holds:
There exists a crisp matrix $A' = \{a'_{ij}\}$ with $a'_{ik} \in [\tilde{a}_{ik}]_\alpha$, $a'_{ij} \in [\tilde{a}_{ij}]_\alpha$, $a'_{jk} \in [\tilde{a}_{jk}]_\alpha$, such that $A' = \{a'_{ij}\}$ is consistent, i.e. for each $i, j, k \in \{1, ..., n\}$ it holds

$$a'_{ik} = a'_{ij} \odot a'_{jk}. \tag{6}$$

The FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is said to be *weak* $\odot$-*consistent*, if $\tilde{A}$ is weak α-$\odot$-consistent for all $\alpha \in [0;1]$.
If for some $\alpha \in [0;1]$ the FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is not weak α-$\odot$-consistent, then $\tilde{A}$ is called α-$\odot$-*inconsistent.*
If for all $\alpha \in [0;1]$ the FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is α-$\odot$-inconsistent, then $\tilde{A}$ is called $\odot$-*inconsistent.*

Ramarks. Let $\alpha, \beta \in [0;1], \alpha \geq \beta$.
- For a crisp PCM, definitions of $\odot$-reciprocity and $\odot$-consistency coincide with the classical definitions.
- If $\tilde{A} = \{\tilde{a}_{ij}\}$ is weak α-$\odot$-consistent, then it is weak β-$\odot$-consistent.
- If $\tilde{A} = \{\tilde{a}_{ij}\}$ is β-$\odot$-inconsistent, then it is α-$\odot$-inconsistent.
- (5) holds for all $i, j \in \{1, ..., n\}$ if and only if (5) holds for all $i, j \in \{1, ..., n\}, 1 \leq i < j \leq n$.
- (6) holds for all $i, j, k \in \{1, ..., n\}$ iff (6) holds for all $i, j, k \in \{1, ..., n\}, 1 \leq i < j < k \leq n$.

The next proposition gives an equivalent condition for a FPC matrix to be weak α-$\odot$-consistent, see e.g. [7].

Proposition 1. *Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix, $[\tilde{a}_{ij}]_\alpha = [a^L_{ij}(\alpha), a^R_{ij}(\alpha)]$ be an α-cut of $\tilde{a}_{ij}$, $\alpha \in [0;1]$.*
FPCM $\tilde{A} = \{\tilde{a}_{ij}\}$ is weak α-$\odot$-consistent iff there exists a vector $w = (w_1, ..., w_n)$ with $w_i \in G, i \in \{1, ..., n\}$, such that for each $i, k \in \{1, ..., n\}$, it holds:

$$a^L_{ik}(\alpha) \leq w_i \div w_k \leq a^R_{ik}(\alpha). \tag{7}$$

A stronger concept of α-$\odot$-consistency (without adjective "weak") similar to (2) has been defined in [8], [9].

Definition 2. Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix, $\alpha \in [0;1]$. For each $i, j \in \{1, ..., n\}$, $[\tilde{a}_{ij}]_\alpha = [a^L_{ij}(\alpha), a^R_{ij}(\alpha)]$ be an α-cut. Denote

$$a^m_{ij}(\alpha) = (a^L_{ij}(\alpha) \odot a^R_{ij}(\alpha))^{(\frac{1}{2})}. \tag{8}$$

A crisp $n \times n$ - matrix $A^m(\alpha) = \{a^m_{ij}(\alpha)\}$ is called α-$\odot$-*mean matrix* associated to FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$.

Definition 3. A FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is said to be α-$\odot$-*consistent*, $\alpha \in [0;1]$, if the following condition holds:

$$a^m_{ik}(\alpha) = a^m_{ij}(\alpha) \odot a^m_{jk}(\alpha) \text{ for all } i, j, k \in \{1, ..., n\}. \tag{9}$$

Moreover, if $\tilde{A} = \{\tilde{a}_{ij}\}$ is α-$\odot$-consistent for all $\alpha \in [0;1]$, then $\tilde{A}$ is said to be $\odot$-*consistent*.

The following proposition gives a characterization of α-$\odot$-consistent FPC matrices, see [8].

Proposition 2. *Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix, $[\tilde{a}_{ij}]_\alpha = [a^L_{ij}(\alpha), a^R_{ij}(\alpha)]$ be an α-cut, $\alpha \in [0, 1]$.*
$\tilde{A} = \{\tilde{a}_{ij}\}$ is α-$\odot$-consistent iff there exists a vector $w(\alpha) = (w_1(\alpha), ..., w_n(\alpha))$ with $w_j(\alpha) \in G, j \in \{1, ..., n\}$, such that for each $i, k \in \{1, ..., n\}$, it holds:

$$a^m_{ik}(\alpha) = w_i(\alpha) \div w_k(\alpha). \tag{10}$$

As it holds $a^L_{ij}(\alpha) \leq (a^L_{ij}(\alpha) \odot a^R_{ij}(\alpha))^{(\frac{1}{2})} \leq a^R_{ij}(\alpha))$, it is clear that the following proposition holds.

Proposition 3. *Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix, $\alpha \in [0;1]$.*
If $\tilde{A}$ is α-$\odot$-consistent then $\tilde{A}$ is weak α-$\odot$-consistent. Moreover, if $\tilde{A}$ is $\odot$-consistent then $\tilde{A}$ is weak $\odot$-consistent.

Remark 1. Notice that α-$\odot$-consistency of FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is equivalent to $\odot$-consistency of the associated crisp matrix, particularly, α-$\odot$-mean matrix associated to $\tilde{A}$, with elements being given by (8). This property will be advantageous in deriving a corresponding priority vector of the FPC matrix in Section 5.

Example 5. Consider the additive alo-group $\mathcal{R} = (\mathbf{R}, \odot, \leq)$ with $\odot = +$, see Example 1. Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be given by triangular fuzzy number elements as follows:

$$\tilde{A} = \begin{bmatrix} (0,0,0) & (1,3,4) & (4,6,8) \\ (-4,-3,-1) & (0,0,0) & (2,4,5) \\ (-8,-6,-4) & (-5,-4,-2) & (0,0,0) \end{bmatrix} \simeq \begin{bmatrix} [0;0] & [1+2\alpha; 4-\alpha] & [4+2\alpha; 8-2\alpha] \\ [4+\alpha; -1-2\alpha] & [0;0] & [2+2\alpha; 5-\alpha] \\ [-8+2\alpha; -4-2\alpha] & [-5+\alpha; -2-2\alpha] & [0;0] \end{bmatrix}.$$

Moreover, α-$+$-mean matrix $A^m(\alpha)$ associated to $\tilde{A}$ for $\alpha \in [0,1]$ is calculated by (8) as

$$A^m(\alpha) = \begin{bmatrix} 0 & \frac{5+\alpha}{2} & 6 \\ -\frac{5+\alpha}{2} & 0 & \frac{7+\alpha}{2} \\ -6 & -\frac{7+\alpha}{2} & 0 \end{bmatrix}.$$

Checking equality (9), we obtain that $\tilde{A}$ is α-$+$-consistent for $\alpha = 0$.

4 Desirable properties of the priority vector

Pairwise comparisons matrices may violate some desirable properties of multiple criteria decision making: e.g. the best alternative with respect to DM's preferences is selected from the set of non-dominated alternatives, on condition this set is non-empty. The other PCMs may violate the conditions of order of preferences (the so called COP conditions), see [1], [4].

Definition 4. Let $A = \{a_{ij}\}$ be a PC matrix with crisp elements. A priority vector $w = (w_1, \cdots, w_n)$ is said to be the *coherent vector (CV) of PC matrix* $A = \{a_{ij}\}$ if the following condition holds

$$a_{ij} > e \text{ iff } w_i > w_j. \tag{11}$$

A priority vector $w = (w_1, \cdots, w_n)$ is said to be the *intensity vector (IV) of PC matrix* $A = \{a_{ij}\}$ if the following condition holds

$$a_{ij} > a_{kl} \text{ iff } w_i \div w_j > w_k \div w_l. \tag{12}$$

From (11) in the above definition it is evident that any CV w of a crisp PC matrix $A = \{a_{ij}\}$ satisfies the POP condition with respect to priority vector w, see [8], and [1]. The same holds for any IV of $A = \{a_{ij}\}$, resp. POIP condition, of $A = \{a_{ij}\}$. The opposite is, evidently, not true.

Let $A = \{a_{ij}\}$ be a crisp consistent PC matrix, and let $w = (w_1, \cdots, w_n)$ be a priority vector associated with A satisfying (3). Then it is obvious that conditions (11) and (12) are satisfied. Now, we are going to define the CV and IV of a FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ based on $A^m(\alpha) = \{a^m_{ij}(\alpha)\}$ called α-$\odot$-*mean matrix* associated to FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$.

Definition 5. Let $c_i, c_j \in \mathcal{C}, \tilde{A} = \{\tilde{a}_{ij}\}$ be an FPC matrix on the alo-group $\mathcal{G} = (G, \odot, \leq), \alpha \in [0;1]$.
We say that c_i *α-mean dominates* c_j, if $a^m_{ij}(\alpha) > e$, where $a^m_{ij}(\alpha) = (a^L_{ij}(\alpha) \odot a^R_{ij}(\alpha))^{(\frac{1}{2})}$.
Moreover, we say that c_i *mean dominates* c_j, if $a^m_{ij}(\alpha) > e$, for all $\alpha \in [0;1]$.

Now, we define the mean CV and IV depending on the mean dominance of alternatives, see [4].

Definition 6. Let $c_i, c_j \in \mathcal{C}, \tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix on the alo-group $\mathcal{G} = (G, \odot, \leq), w = (w_1, w_2, \cdots, w_n)$, $w_i \in G$, be a priority vector, $\alpha \in [0;1]$.
We say that the vector w is *mean α-coherent vector of FPC matrix $\tilde{A}$ (mean-α-CV)* if:
c_i α-mean dominates c_j, iff $w_i > w_j$.
Moreover, the vector w is *mean coherent vector of FPC matrix $\tilde{A}$ (mean-CV)* if:
c_i mean dominates c_j, iff $w_i > w_j$.

Definition 7. Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a FPC matrix on the alo-group $\mathcal{G} = (G, \odot, \leq), w = (w_1, w_2, \cdots, w_n), w_i \in G$, be a priority vector, $\alpha \in [0;1]$.
We say that the vector w is *mean α-intensity vector of FPC matrix $\tilde{A}$ (mean-α-IV)* if it holds:
$a^m_{ij}(\alpha) > a^m_{kl}(\alpha)$ iff $w_i \div w_j > w_k \div w_l$.
Moreover, the vector w is *mean intensity vector of FPC matrix $\tilde{A}$ (mean-IV)* if it holds:
$a^m_{ij}(\alpha) > a^m_{kl}(\alpha)$ iff $w_i \div w_j > w_k \div w_l$, for all $\alpha \in [0;1]$.

Remark 2. Notice, that the concepts of the α-mean CV/IV is introduced for a given FPC matrix $\tilde{A}$ and given priority vector w. If $\tilde{A} = \{\tilde{a}_{ij}\}$ is a crisp FPC matrix, then the priority vectors CV/IV introduced in Definitions 6 and 7 coincide with the usual CV/IV vectors defined by Definition 4.

Remark 3. By setting $k = l$ in Definition 7 we can see that each α-mean IV is α-mean CV of a given FPC matrix $\tilde{A}$. The opposite is, however, not true.

Definition 8. Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be an FPC matrix on alo-group $\mathcal{G} = (G, \odot, \leq)$. For each pair $i, j \in \{1, \ldots, n\}$, and a priority vector $w = (w_1, w_2, \cdots, w_n)$, $w_i \in G$, $\alpha \in [0; 1]$ and for $a_{ij}^m(\alpha) = (a_{ij}^L(\alpha) \odot a_{ij}^R(\alpha))^{(\frac{1}{2})}$ let us denote

$$\varepsilon^m(i, j, w, \alpha) = \max\{a_{ij}^m(\alpha) \odot w_j \div w_i, (a_{ij}^m(\alpha) \odot w_j \div w_i)^{(-1)}\}. \tag{13}$$

Definition 9. Let $\alpha \in [0; 1]$. The *global error index* $\mathcal{E}(\tilde{A}, w, \alpha)$, for a FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ and a priority vector $w = (w_1, \ldots, w_n)$ is defined as the maximal element of matrix of local errors $\varepsilon^m(w, \alpha)$, i.e.

$$\mathcal{E}(\tilde{A}, w, \alpha) = \max_{i,j \in \{1,\ldots,n\}} \varepsilon^m(i, j, w, \alpha). \tag{14}$$

Proposition 4. *Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be a reciprocal FPC matrix and $w = (w_1, \ldots, w_n)$ be a priority vector, $\alpha \in [0; 1]$. Then the global error index satisfies*

$$\mathcal{E}(\tilde{A}, w, \alpha) \geq e. \tag{15}$$

Moreover, it holds

$$\mathcal{E}(\tilde{A}, w, \alpha) = e \tag{16}$$

iff $\tilde{A} = \{\tilde{a}_{ij}\}$ is α-$\odot$-consistent.

5 Deriving priority vectors and measuring inconsistency of FPC matrices

In this section we propose a method for calculating the priority vector of $n \times n$ FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ for the purpose of rating the alternatives $c_1, \ldots, c_n \in \mathcal{C}$. Here, we do not follow the way of calculating the *fuzzy* priority vector proposed e.g. in [3], and is different to the previous method presented in [8]. Here, we generate a crisp priority vector, therefore, no defuzzification is necessary for final ranking of the alternatives. The proposed method for calculating the priority vector can be divided into two steps as follows. In the first step, we check by (7) whether the given FPC matrix is weak α-$\odot$-mean-consistent for some α. Then we calculate the maximal such α denoted by α^*. Otherwise, we change the PCM $\tilde{A}$. In the second step, we use the following algorithm.

Algorithm.
Consider an $n \times n$ FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$, given $\alpha^* \in]0; 1]$.
Calculate α^*-$\odot$-mean matrix $A^m(\alpha^*) = \{a_{ij}^m(\alpha^*)\}$ associated to FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$, by (8) for all $i, j \in \{1, \ldots, n\}$, i.e.

$$a_{ij}^m(\alpha^*) = (a_{ij}^L(\alpha^*) \odot a_{ij}^R(\alpha^*))^{(\frac{1}{2})}.$$

Based on this PCM, we define the following index set:

$$I(A^m(\alpha^*)) = \{(i, j, k, l) | i, j, k, l \in \{1, \ldots, n\}, a_{ij}^m(\alpha^*) > a_{kl}^m(\alpha^*)\}, \tag{17}$$

The global error index, $\mathcal{E}(\tilde{A}, w, \alpha^*)$, of $A^m(\alpha^*) = \{a_{ij}^m(\alpha^*)\}$ and $w = (w_1, \ldots, w_n)$ have been already defined by (14) as

$$\mathcal{E}(\tilde{A}, w, \alpha^*) = \max_{\{i,j \in \{1,\ldots,n\}\}} \{a_{ij}^m(\alpha^*) \odot w_j \div w_i, (a_{ij}^m(\alpha^*) \odot w_j \div w_i)^{(-1)}\}. \tag{18}$$

The following optimization problem (the variables are w_i) is solved
(P)

$$\mathcal{E}(\tilde{A}, w, \alpha^*) \longrightarrow \min; \tag{19}$$

subject to

$$a_{ij}^L(\alpha^*) \leq w_i \div w_j \leq a_{ij}^R(\alpha^*) \text{ for all } i, j \in \{1, \ldots, n\}, \tag{20}$$

$$\bigodot_{k=1}^{n} w_k = e, \tag{21}$$

$$w_k \in G, \text{ for all } k \in \{1, ..., n\}. \tag{22}$$

$$w_r \geq w_s \odot \eta \ \forall (r,s) \in I(A^m(\alpha^*)), \tag{23}$$

where $\eta > e$ is a preselected constant.
If optimization problem (P) has no feasible solution, then it means that FPC matrix $\tilde{A} = \{\tilde{a}_{ij}\}$ is weak α^*-$\odot$-consistent for no $\alpha \in [0; 1]$.

Remark 4. In general, problem (P) is a nonlinear optimization problem that may be solved by a numerical method, e.g. by one of available optimization methods, see e.g. [2]. By solving optimization problem (P) we obtain a corresponding priority vector with our desirable properties, that is mean α^*-CV, and IV. By the value of the optimal solution $w^* = (w_1^*, ..., w_n^*)$ of (P), the global consistency index (error) of $\tilde{A}$, is measured. By Proposition 4 $\mathcal{E}(\tilde{A}, w^*, \alpha^*) = e$ iff $\tilde{A} = \{\tilde{a}_{ij}\}$ is α^*-$\odot$-consistent. Otherwise, $\mathcal{E}(\tilde{A}, w^*, \alpha^*) > e$.

Remark 5. In general, the existence and uniqueness of optimal solution of (P) is not saved. Depending on the particular operation $\odot$, problem (P) may have multiple optimal solutions which is an unfavorable fact from the point of view of the DM. In this case, the DM should reconsider particular (fuzzy) evaluations in the original fuzzy pairwise comparison matrix.

Example 6. Consider the multiplicative alo-group $\mathcal{R}_+ = (\mathbf{R}_+, ., \leq)$ with $\odot = \cdot$, see Example 2. Let $\tilde{A} = \{\tilde{a}_{ij}\}$ be given by triangular fuzzy number elements as follows:

$$\tilde{A} = \begin{bmatrix} (1,1,1) & (1,2,3) & (7,8,9) \\ (\frac{1}{3}, \frac{1}{2}, 1) & (1,1,1) & (2,3,4) \\ (\frac{1}{9}, \frac{1}{8}, \frac{1}{7}) & (\frac{1}{4}, \frac{1}{3}, \frac{1}{2}) & (1,1,1) \end{bmatrix} \simeq \begin{bmatrix} [1;1] & [1+\alpha; 3-\alpha] & [7+\alpha; 9-\alpha] \\ [\frac{1}{3-\alpha}; \frac{1}{1+\alpha}] & [1;1] & [2+\alpha; 4-\alpha] \\ \frac{1}{9-\alpha}; \frac{1}{7+\alpha}] & [\frac{1}{4-\alpha}; \frac{1}{2+\alpha}] & [1;1] \end{bmatrix}.$$

Here, $\tilde{A}$ is a 3×3 PC matrix with triangular fuzzy number elements and the corresponding membership functions.
Given $\alpha^* = 0.683$. It can be verified that this is the maximal level of α such that $\tilde{A}$ is weak-α-consistent.
By solving problem (P), we obtain the optimal solution as $w^* = (w_1^*, w_2^*, w_3^*) = (2.509, 1.147, 0.347)$, moreover, $\mathcal{E}(\tilde{A}, w^*, 0.683) = 1.126 > 1$, i.e. $\tilde{A}$ is inconsistent. The corresponding ranking of alternatives is $c_1 > c_2 > c_3$.
It can be easily verified that w^* is both mean 0.683-CV and mean 0.683-IV.

6 Conclusion

This paper deals with PC matrices with fuzzy elements. We defined the concept of priority vector satisfying some desirable properties – the coherence and intensity – an extension of the well known concepts in the crisp case. They were used for ranking the alternatives. Such an approach allows for extending various approaches known from the literature.

Acknowledgements

This research has been supported by GACR project No. 18-01246S.

References

[1] Bana e Costa, A.A., Vasnick, J.A. (2008). A critical analysis of the eigenvalue method used to derive priorities in the AHP. *European Journal of Operational Research* 187 3, 1422–1428.

[2] Boyd, S., Vandenberghe, L. (2004). *Convex optimization.* Cambridge University Press, Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, Sao Paolo, Delhi.

[3] Csutora, R., Buckley, J.J. (2001). Fuzzy hierarchical analysis: the Lambda-Max method. *Fuzzy Sets and Systems* 120, 181–195.

[4] Cavallo, B. (2019). Coherent weights for pairwise comparison matrices and a mixed-integer linear programming problem. *Journal of Global Optimization* 75, 143–161.

[5] Cavallo, B., D'Apuzzo, L. (2020). Preservation of preferences intensity of an inconsistent Pairwise Comparison Matrix. *Internat. J. of Approximate Reasoning* 116, 33–42.

[6] Kulakowski, K., Mazurek, J., Ramik, J., Soltys, M. (2019). When is the condition of order preservation met? *European Journal of Operational Research* 277, 248–254.

[7] Ramik, J. (2015). Pairwise comparison matrix with fuzzy elements on alo-group. *Information Sciences* 297, 236–253.

[8] Ramik, J. (2018). Condition of Order Preservation in Pairwise Comparisons Matrix With Fuzzy Elements. In: 36th International Conference Mathematical Methods in Economics Conference Proceedings. ČSOV, VŠE FM J. Hradec, 457–462.

[9] Ramik, J. (2020). *Pairwise comparisons method: Theory and Applications in Decision Making.* Switzerland, Cham-Heidelberg-New York-Dordrecht-London: Springer Internat. Publ. 253 p.

[10] Saaty, T. L. (1980). *Analytic Hierarchy Process*. McGraw-Hill, New York.

On the Effect of Errors in Pairwise Comparisons during Search Based on a Short List

David Mark Ramsey[1]

Abstract. Consumers may have a huge amount of information available when making a decision. However, due to constraints on their powers of cognition, they cannot efficiently process this information. Heuristic rules are often used in order to apply near-optimal strategies. One such heuristic rule inolves using a short list. Such rules are effective when there are a large number of offers and some information can be gained on each offer at very little cost (e.g. via the Internet). However, to obtain a clear assessment of an offer, it must be inspected more closely. By using a short list, a searcher chooses a relatively small number of promising offers based on initial information. The searcher then inspects the offers on the short list, before accepting the one on this list that is assessed to be the best. This article investigates the effect of errors in perceiving the relative values of offers on the optimal length of a short list and effectiveness of such rules when comparisons are made sequentially.

Keywords: search, heuristics, short list, pairwise comparisons, imperfect perception

JEL Classification: C44
AMS Classification: 90B40

1 Introduction

This paper considers a two-stage problem of searching for a valuable resource in which a decision maker (DM) gathers information about available offers. For example, suppose a DM is looking for a new flat using an Internet database of offers. In the first stage, the DM can observe partial information about a large number of offers. On the basis of this information, a short list of flats are chosen to be viewed in real life. Hence, it is assumed that the search costs in the first round are low, but high in the second round. After closely inspecting the offers on the short list, the offer considered to be the best according to all the information gathered is accepted.

A short list is a useful heuristic tool to DMs when some information about offers can be obtained at low cost, but an accurate appraisal of the suitability of an offer is relatively costly. A successful heuristic must be adapted to the abilities of a decision maker and the structure of the information gained during search (see Simon [11], Todd and Gigerenzer [12], Bobadilla-Suarez and Love [2]). Much work has been published recently on the concept of short lists, which are useful when the costs of exhaustive search are high or the amount of information available exceeds the cognitive abilities of DMs (see Masatlioglu et al. [6] and Lleras et al. [5]). Bora and Kops [3] model search processes where DMs make short lists based on information from their peers. However, not much work has been published on the question of how long short lists should be according to the parameters of a search problem and the structure of the information. One other question regards the effect of imperfect perception on the effectiveness of such rules. The model presented here is intended to be a step towards answering such questions.

Analytis et al. [1] presented a similar model where searchers can ascribe (expected) values to offers at both stages of the search process. First, offers are ranked on the basis of initial information. Secondly, the DM closely observes offers in descending order of expected value and stops when the value of an offer given the new information exceeds the reward expected from future search. The model presented here extends Ramsey's model [9]. A fixed number of offers are ranked according to initial information. The k best ranked offers are then inspected closely, after which the best ranked offer based on all the information gained is accepted. The signals observed in these two rounds are described by two random variables (X_1, X_2) from a joint continuous distribution. The DM does not observe these values, but can rank offers based on these signals. The DM's payoff is the sum of the signals minus the search costs. This paper investigates how robust

[1] Wrocław University of Science and Technology, Faculty of Computer Science and Management, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland, david.ramsey@pwr.edu.pl

the optimal strategy of this form is to the rate of errors in pairwise comparisons when comparisons are made sequentially.

Section 2 presents the model. Some previous results related to this model and ranking via pairwise comparisons are presented in Section 3. The simulations are described in Section 4. Section 5 presents the results from these simulations. Conclusions and directions for future research are given in Section 6.

2 Model

A decision maker (DM) must choose one of n offers. Offer i is characterized by two quantitative signals, the random variables $X_{i,1}$ and $X_{i,2}$. The pairs of signals $(X_{1,1}, X_{1,2}), (X_{2,1}, X_{2,2}), \ldots, (X_{n,1}, X_{n,2})$ are assumed to be independent realizations from a joint normal distribution such that $E(X_{i,j}) = 0, j = 1, 2$, $Var(X_{i,j}) = \sigma_j^2$ and the coefficient of correlation between $X_{i,1}$ and $X_{i,2}$ is ρ, where $\rho \geq 0$. The value of offer i is given by V_i, where $V_i = X_{i,1} + X_{i,2}$. Hence, $X_{i,1}$ and $X_{i,2}$ are signals of the value of an offer (called Signal 1 and Signal 2, respectively) and the relative importance of a signal is defined by the variance in that signal relative to the variance in the other signal. As in the search problem descibed in the Introduction, the costs of observing the first (initial) signal are much lower than the costs of observing the second signal. It should be noted that the assumption of normality is made for convenience. Other joint distributions may be considered, as well as various functional forms describing the value of an offer on the basis of the signals.

The DM does not observe the value of these signals, but ranks offers according to imperfect pairwise comparisons on the basis of the information available. Let $p(x, y)$, where $x \geq y$, denote the probability that the DM assesses x to be greater than y, when x and y come from a distribution with variance σ^2. It is assumed that

$$p(x, y) = 1 - \frac{\exp(-r[x - y]/\sigma)}{2}, \tag{1}$$

where $r > 0$. Hence, when x and y take the same value, then comparison is random. When $x - y \to \infty$, then the probability of correct comparison tends to 1. The parameter r measures the accuracy of perceiving which of two signals from a given distribution is larger. For example, such errors may arise from flat hunters not being able to see flaws in the construction of a building. One could also give the probability, p, of correct comparison when the difference between the values x and y is equal to the standard deviation of the distribution they come from. It follows that $p = 1 - 0.5e^{-r}$. Equivalently, $r = -\ln(2 - 2p)$.

Here, we consider strategies that form a short list of k offers, where $k \geq 2$ on the basis of the first signal. Thus, the DM only observes the second signal of the k offers assessed to be the best on the basis of Signal 1. After observing the second signal for the offers on the short list, the DM accepts the offer from the short list assessed to be best overall. Hence, a two-stage search procedure is used. In the first stage, the DM sequentially observes Signal 1 for all the offers, which appear in random order. The first k offers are ranked on the basis of Signal 1 according to the optimal procedure for ranking (see Section 3), but with imperfect perception, as described by Equation 1 to form an initial short list. Later offers are first compared on the basis of Signal 1 with the offer currently in k-th place in the ranking, which is denoted by D_k. If the current offer is not preferred to D_k, then the current short list remains unchanged and the DM proceeds to the next offer (if one remains). If the current offer is preferred to D_k, then it replaces D_k on the current short list and is ranked with respect to the remaining $k - 1$ members of the short list, before the DM proceeds to the next offer (if one remains). Once Signal 1 has been observed for all of the offers, then the current short list becomes the official short list.

In the second stage of the search procedure, the DM observes Signal 2 for the offers on the official short list, i.e. closely inspects the offers on the short list. Here, we consider two approaches. According to the first approach, offers on the short list are observed from the best to worst ranked according to the ranking in stage one (the "in order" protocol). According to the second approach, the offers are observed from lowest to highest ranked according to the ranking in stage one (the "in reverse order" protocol). The first offer observed in stage two is the initial candidate. Each successive offer is compared with the current candidate on the basis of the value of these two offers (i.e. the sum of the two signals). If the current offer is assessed to be better than the current candidate, then it replaces the current candidate. After all the offers on the short list have been closely inspected, the DM selects the current candidate. The algorithm for the search procedure in stage one is illustrated by Fig. 1. The procedure for the search procedure in stage two is illustrated in Fig. 2.

The reward obtained by the DM is assumed to be the value of the offer accepted minus the search costs, which are defined to be $C = m_1c_1 + m_2c_2$, where m_i is the number of pairwise comparisons carried out in

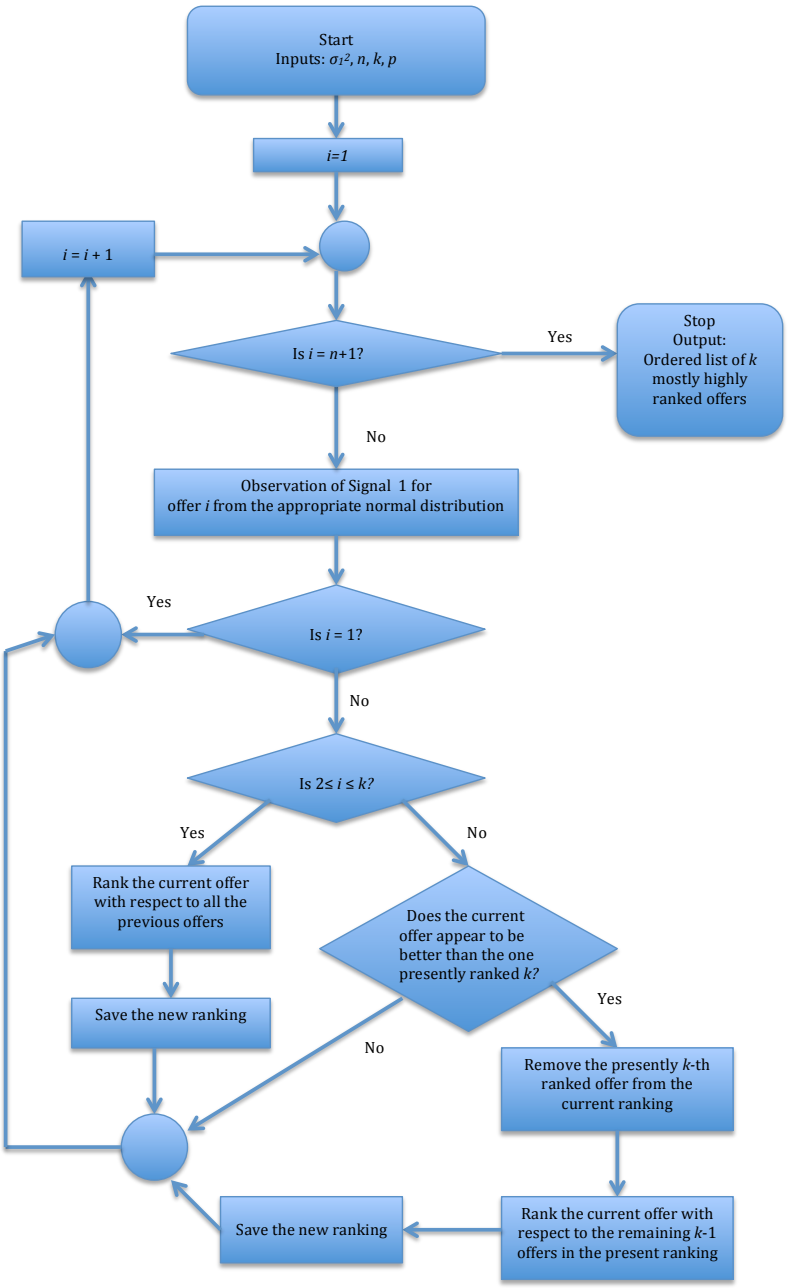


Figure 1 Algorithm for the search procedure in the first stage.

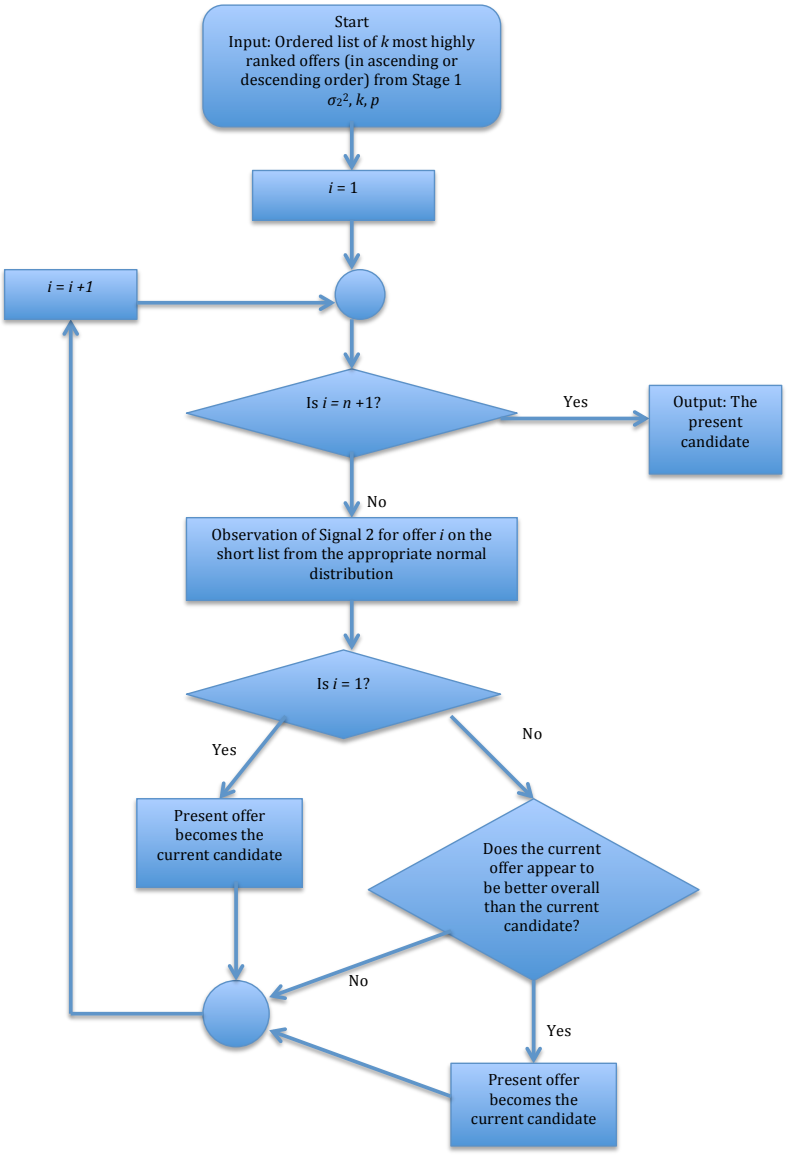


Figure 2 Algorithm for the search procedure in the second stage.

stage i and c_i is the cost of a pairwise comparison in stage i. Since the number of pairwise comparisons in stage two is $k-1$, one could assume that the costs of inspecting an item in stage two are c_2 and the optimal strategy of this form would be unchanged (the search costs would be simply be increased by c_2 in any given search procedure). It is assumed that c_1 is very small in comparison to c_2. The goal of the DM is to maximize the expected reward from search. Hence the strategy used by the DM is determined by the choice of both k (the length of the short list) and the order in which the offers on the short list are observed in the second stage.

The following section recalls some previous results regarding the optimal length of a short list for a similar problem of optimal choice and the optimal procedure for ranking a set of numbers via pairwise comparison.

3 Results on Short Lists and Ranking via Pairwise Comparisons

Ramsey [9] described a model of choice based on a short list with error-free assessment in which the marginal cost associated with increasing the length of a short list from k to $k+1$, M_k, is non-decreasing in k. Under fairly general conditions, it can be shown that the marginal increase in the expected value of the offer ultimately accepted from increasing the length of the short list from k to $k+1$, G_k, is decreasing in k. Under these conditions, the optimal length of the short list is given by the smallest value of k such that $M_k \geq G_k$. Hence, under the given assumptions the optimal length of the short list is given by a threshold rule.

Determination of the short list relies on the optimal procedure for ranking a set of k numbers, described e.g. in Knuth [4]. Such a ranking is constructed iteratively by ranking the i-th object to appear relative to the first $i-1$ offers, for $i = 2, 3, \ldots, k$. Let T_k be the expected number of pairwise comparisons needed to produce a full ranking of k offers and E_i the expected number of such comparisons required to rank the i-th offer with respect to the first $i-1$ offers. Hence, $T_k = \sum_{i=2}^{k} E_i$. Note that $T_2 = E_2 = 1$, since one pairwise comparison is required to compare the first two offers. In general, the i-th item is initially compared with a median ranked item of the first $i-1$ offers.

When i is even, regardless of whether the i-th offer is better or worse than this median offer, further comparison reduces to comparison with $\frac{i}{2}-1$ previous offers. It follows that for even i, $E_i = 1 + E_{i/2}$.

When i is odd, the i-th (current) offer can be compared to the $\frac{i-1}{2}$-th ranked offer. The current offer is higher ranked than this $\frac{i-1}{2}$-th ranked offer with probability $\frac{i-1}{2i}$ and further comparison reduces to comparison with $\frac{i-3}{2}$ other offers. Otherwise, further comparison reduces to comparison with $\frac{i-1}{2}$ other offers. It follows that for odd i

$$E_i = 1 + \frac{i-1}{2i} E_{(i-1)/2} + \frac{i+1}{2i} E_{(i+1)/2}.$$

From the $k+1$-th offer onwards, the short list is controlled by comparing each new offer with the offer presently ranked k on the short list. If a new offer is preferred to this last offer on the short list, then the new offer replaces it and is ranked with respect to the other $k-1$ offers presently on the short list. Hence, for $i = k+1, k+2, \ldots, n$ at least one comparison is always made and with probability k/i, on average E_k additional comparisons are made. Hence, the expected number of comparisons made from offer $k+1$ onwards is given by $U_{k,n}$, where

$$U_{k,n} = n - k + \sum_{i=k+1}^{n} \frac{kE_k}{i}.$$

Hence, when there are no errors in comparisons, the expected total number of comparisons in Stage 1 of the search procedure is $W_{k,n} = T_k + U_{k,n}$. When there are errors in comparisons, this could be used as an approximation for the number of pairwise comparisons made (note that. for example, when $i > k$, the probability of adding the i-th offer to the current short list is not k/i). This approximation become less accurate as the error rate increases.

Assuming n is fixed, the marginal cost of increasing the size of the short list from k to $k+1$, M_k, satisfies $M_k = W_{k+1,n} - W_{k,n}$. These marginal costs are not necessarily decreasing in k, e.g. for sufficiently large n, $M_2 < M_3 > M_4$, since the ranking algorithm works most efficiently when k is a power of two. However, calculations indicate that these marginal costs tend to increase for larger values of k (for n much larger than k). Thus the length of the short list determined according to Ramsey [9] does not necessarily define the optimal strategy based on pairwise comparisons. However, such a strategy will be at least near-optimal from this class of strategies (both in the sense that the length of the short list given by the threshold rule described above, k^t is very similar to the optimal length of the short list, k^* and the expected reward obtained using a short list of length k^t is very similar to the expected reward obtained using a short list of length k^*).

The following section describes the simulations used to study the robustness of such search rules to errors in comparison. One important aspect is the relation between such errors and the optimal strategy from the set of strategies considered. Another question is how the position in which an offer is observed affects the probability that it is placed on the short list and that it is ultimately accepted. When there are no errors in perception and offers appear in random order in stage one, then the probability of the i-th offer being ultimately accepted is equal to $1/n$. The probability that it is placed on the short list is k/n. Also, the probability of an offer being ultimately accepted depends on its rank according to the initial inspection, but is otherwise independent of the order in which the offers on the short list are observed in stage two.

4 Simulations

The search procedure is simulated for problems in which 20 offers were available. The two signals are independent observations from the normal distribution with $\sigma_1^2 = 1$ and $\sigma_2^2 \in \{\frac{1}{25}, \frac{1}{9}, 1, 9, 25\}$. Let $\sigma^2 = \sigma_1^2 + \sigma_2^2$ denote the overall variance in the value of the offers. Each pairwise comparison in stage one costs $c_1 = 0.001$. The initial ranking of the first k offers is carried out as described in Section 3 under the assumption that when a new offer is compared to an even number of offers, the next comparison is made with the highest ranked of the two medians. For example the following procedure is used to rank the seventh offer to appear.

a) The current offer is first compared with the third ranked of the first six offers. If it assessed as better than the third ranked item go to b), else go to c).

b) It remains to compare the current offer with the two presently highest ranked offers. Compare the current offer with the highest rank so far. If the current offer is assessed to be better, it is placed in first position in the new ranking. If the current offer is assessed to be worse than the best ranked so far, it is then compared with the second ranked offer. If the current offer is assessed to be better than the second ranked offer, then it is given rank two in the new ranking. If the current offer is assessed to be worse than the second ranked offer, then it is given rank three in the new ranking. Go to d).

c) It remains to compare the current offer with the offers presently ranked from four to six. The current offer is first compared to the offer currently ranked fifth. If the current offer is assessed to be better than the fifth ranked offer, then it is compared with the fourth ranked. In this case, the current offer is ranked fourth when it is assessed to be better than the currently fourth ranked, otherwise it is ranked fifth. If the current offer is assessed to be worse than the fifth ranked offer, then it is compared to the offer currently ranked sixth. In this case, the current offer is ranked sixth when it is assessed to be better than the offer currently ranked sixth, otherwise it is ranked seventh.

d) Given that the rank of the current offer is j. The ranks of the offers previously ascribed a rank of $\geq j$ is increased by one.

Each pairwise comparison in stage two costs $c_2\sigma_2$, where $c_2 \in \{0.02, 0.05, 0.1\}$. The costs in stage two are assumed to be proportional to the standard deviation of the second signal, since under sequential search with perfect information based on one signal, the expected number of offers seen when the search costs are proportional to σ is independent of σ (see Ramsey [8]). Hence, changes in the length of the optimal short list reflect the relative importance of the second signal rather than the overall variance in the signal. The accuracy of comparisons is described by the parameter p, the probability of correctly ranking two observations whose difference is equal to the standard deviation of the distribution they come from (σ_1 in stage one and σ in stage two). The values of p considered are in the set $\{0.9, 0.99, 0.999, 0.9999, 1\}$.

For each combination of the parameters and a search protocol ("in order" or "in reverse order"), 100 000 sets of 20 offers described by a pair of signals from the appropriate binormal distribution were generated. For each of these sets of offers, the search procedure was simulated for each length of short list k, where $2 \leq k \leq 19$, according to the procedures illustrated in Fig. 1 and Fig. 2. The expected value of the offer accepted under such a procedure, $\overline{V}_k$ was estimated by the average value of the offers accepted in these realizations. The expected number of pairwise comparisons in Stage 1 of the search, $W_{k,n}$, was estimated by counting the number of comparisons carried out in each realization of the search procedure and then calculating the mean number of comparisons made, $\overline{W}$. The number of pairwise comparisons in Stage 2 is always $k-1$. Based on this, the estimate of the expected search costs is $C_k = c_1\overline{W} + (k-1)c_2$. Hence, the estimate of the expected reward from search using a short list of length k can be estimated as $r_k = \overline{V}_k - C_k$. The estimate of the optimal expected reward for a given set of parameters is given by r^*, where $r^* = \max_{2\leq k\leq 19} r_k$ and the empirically derived optimal length of the short list, k^*, is the value of k for which this maximum is attained. It should be noted that r^* and k^* depend on the parameters of the search problem, σ_2^2, c_2 and p, as well as the protocol used ("in order" or "in reverse order"). However, to keep the notation simple, the notation does not reflect this.

Search under the empirically derived optimal strategy was then simulated for 100 000 sets of 20 offers as above. These simulations recorded the probabilities of being placed on the official short list and being ultimately accepted given the position of an offer in the sequence in stage one, as well as the probability of being ultimately accepted given an offer's ranking after stage 1 (which defines its position in the sequence in round two). The following section describes the results from these simulations.

5 Results from the Simulations

Tables 1 and 2, respectively, present the empirically found optimal lengths of the short list, k^*, and the expected reward, r^*, under strategies using the "in order" protocol. Tables 3 and 4 present k^* and r^*, respectively, under strategies using the "reverse order" protocol. Each cell in these tables corresponds to a given relative importance of the second signal (σ_2^2) and relative cost of search in the second stage, c. The five results given in each cell correspond to increasing levels of the accuracy of comparison, namely $p = 0.9, 0.99, 0.999, 0.9999, 1$, respectively.

	$c = 0.02$	$c = 0.05$	$c = 0.10$
$\sigma_2^2 = 1/25$	(2, 2, 2, 4, 2)	(2, 2, 2, 2, 2)	(2, 2, 2, 2, 2)
$\sigma_2^2 = 1/9$	(2, 4, 4, 4, 3)	(2, 2, 2, 2, 2)	(2, 2, 2, 2, 2)
$\sigma_2^2 = 1$	(4, 7, 7, 7, 6)	(4, 4, 4, 4, 4)	(3, 3, 3, 3, 3)
$\sigma_2^2 = 9$	(9, 12, 12, 12, 11)	(6, 7, 7, 7, 7)	(4, 4, 4, 4, 4)
$\sigma_2^2 = 25$	(11, 15, 15, 15, 14)	(7, 8, 8, 8, 8)	(4, 5, 5, 5, 5)

Table 1 Empirically derived optimal lengths of the short list for the "in order" protocol. The five results given in each cell correspond to increasing levels of accuracy of perception $p = 0.9, 0.99, 0.999, 0.9999, 1$, respectively.

	$c = 0.02$	$c = 0.05$	$c = 0.10$
$\sigma_2^2 = 1/25$	(1.696,1.826,1.850,1.860,1.872)	(1.690,1.815,1.849,1.854,1.860)	(1.681,1.813,1.831,1.844,1.855)
$\sigma_2^2 = 1/9$	(1.718,1.866,1.906,1.914,1.917)	(1.710,1.852,1.877,1.890,1.900)	(1.695,1.835,1.863,1.874,1.883)
$\sigma_2^2 = 1$	(2.085,2.370,2.422,2.439,2.422)	(2.000,2.247,2.288,2.307,2.304)	(1.883,2.107,2.147,2.155,2.172)
$\sigma_2^2 = 9$	(4.039,4.772,4.883,4.933,4.892)	(3.504,4.076,4.163,4.191,4.187)	(2.953,3.346,3.451,3.525,3.533)
$\sigma_2^2 = 25$	(6.249,7.476,7.667,7.726,7.697)	(5.146,6.100,6.252,6.283,6.293)	(4.118,4.835,4.950,4.995,5.031)

Table 2 Empirically estimated optimal rewards for the "in order" protocol. The five results given in each cell correspond to increasing levels of accuracy of perception $p = 0.9, 0.99, 0.999, 0.9999, 1$, respectively.

	$c = 0.02$	$c = 0.05$	$c = 0.10$
$\sigma_2^2 = 1/25$	(5, 4, 4, 4, 2)	(4, 2, 2, 2, 2)	(2, 2, 2, 2, 2)
$\sigma_2^2 = 1/9$	(5, 4, 4, 4, 3)	(4, 4, 2, 2, 2)	(2, 2, 2, 2, 2)
$\sigma_2^2 = 1$	(8, 7, 7, 7, 6)	(4, 4, 4, 5, 4)	(3, 3, 3, 3, 3)
$\sigma_2^2 = 9$	(10, 12, 13, 12, 11)	(6, 7, 7, 7, 7)	(4, 4, 4, 4, 4)
$\sigma_2^2 = 25$	(12, 14, 14, 14, 14)	(7, 8, 8, 8, 8)	(4, 5, 5, 5, 5)

Table 3 Empirically derived optimal lengths of the short list for the "in reverse order" protocol. The five results given in each cell correspond to increasing levels of accuracy of perception $p = 0.9, 0.99, 0.999, 0.9999, 1$, respectively.

Firstly, we consider the similarities between the results obtained for the two protocols. Both k^* and r^* are increasing in the variance of the second signal and decreasing in the relative costs of searching. The expected reward is generally increasing in the accuracy of the comparisons. However, when the search costs are relatively low ($c = 0.02$ or $c = 0.05$) and the second signal is of similar or greater importance than the first ($\sigma_2^2 \geq 1$), then a small error rate may even be favourable to the searcher. This will be considered in the Conclusion.

On the other hand, k^* is very weakly dependent on the accuracy of comparisons, particularly when the search costs in stage two are relatively high ($c \geq 0.05$). When these costs are low ($c = 0.02$), in comparison to

	$c = 0.02$	$c = 0.05$	$c = 0.10$
$\sigma_2^2 = 1/25$	(1.726,1.838,1.857,1.868,1.872)	(1.714,1.819,1.841,1.854,1.860)	(1.678,1.809,1.834,1.844,1.855)
$\sigma_2^2 = 1/9$	(1.769,1.881,1.906,1.916,1.917)	(1.734,1.856,1.876,1.881,1.900)	(1.692,1.835,1.859,1.875,1.883)
$\sigma_2^2 = 1$	(2.180,2.382,2.424,2.440,2.422)	(2.054,2.247,2.291,2.303,2.304)	(1.902,2.107,2.149,2.162,2.172)
$\sigma_2^2 = 9$	(4.172,4.793,4.889,4.939,4.892)	(3.569,4.077,4.169,4.198,4.187)	(2.998,3.411,3.500,3.523,3.533)
$\sigma_2^2 = 25$	(6.416,7.493,7.627,7.710,7.697)	(5.235,6.079,6.241,6.272,6.293)	(4.162,4.844,4.953,5.007,5.031)

Table 4 Empirically estimated optimal rewards for the "in reverse order" protocol. The five results given in each cell correspond to increasing levels of accuracy of perception $p = 0.9, 0.99, 0.999, 0.9999, 1$, respectively.

the optimal strategy when there are no errors, the introduction of a small error rate is associated with an increase in k^*. Under the "in order" protocol, when the error rate becomes relatively large, then k^* falls. Under the "reverse order" protocol, when the second signal is at most as important as the first ($\sigma_2^2 \leq 1$), then k^* increases as the error rate increases. For $\sigma_2^2 > 1$, when the error rate becomes large, then k^* falls, but by less than under the "in order" protocol. These subtle differences seem to be related to the effectiveness of these protocols (see the following paragraph).

When the error rate is low, there is no significant difference between the effectiveness of these protocols. However, when the rate of errors increases, the "reverse order" protocol is more effective, unless both the search costs are high ($c = 0.10$) and the second signal is unimportant (in this case, the optimal length of the short list is always 2). The explanation of this phenomenon depends on the importance of the second signal. When the second signal is relatively important (the optimal size of the short list is large), it is likely that the best offer on the short list must be compared with several other offers before being accepted. Hence, the probability of rejecting the best offer on the short list is relatively high. By inspecting the offers in ascending order of attractiveness (the "reverse order" protocol), this probability is minimized. Secondly, when the second signal is relatively unimportant and the length of the short list is low, the probability of not placing the best offer overall on the short list is relatively high. However, when a short list of moderate length is used in tandem with the "in order" protocol, there is a relatively large probability of rejecting the best offer on the short list. When the error rate is relatively large, then using a short list of moderate size in tandem with the "reverse order" protocol provides a good compromise, ensuring that the best offer is on the official short list with a large probability, but is rejected in stage two with a small probability.

The effect of an offer's position in the sequence in stage one depends on the length of the short list, k^*. When k^* is relatively small, $k^* \leq 4$, later offers are more likely to be placed on the short list. This effect is visible for small error rates and becomes stronger as the error rate increases. For low or moderate error rates, there is no visible relation between the position of an offer in stage one and the probability of ultimate acceptance. However, for high error rates, offers at the end of the list in stage one are more likely to be ultimately picked. When k^* is intermediate, $4 < k^* \leq 10$, the positional effect on the probability of being placed on the short list or being ultimately picked is very small. For high error rates, the offers appearing just after position k^* are the least likely to be placed on the short list. This is probably due to the facts that a) attractive offers appearing in such a position may be immediately rejected and b) unattractive offers that are placed on the current short list are likely to be removed before the official short list is completed. Additionally, the third offer observed in stage one has a slightly increased probability of being placed on the short list. This is due to the ranking algorithm used. The third offer is first compared with the best of the first two offers. This means that even relatively weak offers in this position are relatively likely to initially occupy the first position in the current short list. The probability of rejection in stage one in this case is very low. When k^* is large, $k^* > 10$, again the positional effects are very weak. When the error rate is high, offers appearing in the middle of stage one are the most likely to appear in the short list and have a marginally increased probability of being accepted. Later offers are the least likely to be placed on the short list or ultimately accepted. Due to the algorithm used to control the short list, a good offer appearing at the end of the list is quite likely to be immediately rejected by mistake.

The protocol used in round two is clearly associated with the probability of an offer of given rank after stage one being ultimately accepted, especially when the error rate is relatively large. Suppose that the rank of an offer after round one is s and the probability that such an offer is ultimately accepted is P_s. When no errors are made, there is always a strong correlation between s and P_s (those of high rank after stage one are most likely to ultimately chosen). When the "reverse order" protocol is used, this correlation becomes weaker

as the error rate increases, but always remains clear. However, when the "in order" protocol is used, this correlation disappears and when the length of the short list used is moderate or large can even reverse. For example, when $p = 0.9, c = 0.1$ and $\sigma_2^2 = 1$, the empirically found optimal length of the short list is three and it is more likely that the third offer on the short list is ultimately chosen rather than the second offer. Moreover, when $p = 0.9, c = 0.02$ and $\sigma_2^2 = 25$, the empirically found optimal length of the short list is 11 and the lowest ranked offer after stage one is more than twice as likely to be ultimately picked than the highest ranked offer after stage one.

6 Conclusion

This article has looked at the effect of errors in perception on search rules based on the short list heuristic. Two protocols for observing the short list are considered: a) "in order" – offers on the short list are inspected in descending order of attractiveness and b) "reverse order" – offers on the short list are inspected in increasing order of attractiveness. Errors in perception have little effect on the lengths of short lists that should be used, but subtle effects on the optimal rule of this form can be observed when the search costs in stage two are low. In psychological terms, the "reverse order" protocol may also be appropriate due to the recency effect, which means that DMs may show a bias towards observations that appeared recently (see Murphy [7]). However, it should be noted that the errors in perception described here do not distinguish between which offer was observed first and which second. Future research should look more carefully at both types of error.

The fact that in certain scenarios a small error rate does not lower the expected payoff of a searcher (compared to search without errors in comparisons) and may even lead to a slight increase in the expected payoff is of interest. When there are no errors in comparison, fixing the length of a short list may well be a very inflexible strategy. More advanced strategies based on short lists should allow the length of the short list to vary when, e.g. a) a small number of offers are clearly better than all the others on the basis of the first signal, or b) there is very little difference between two offers according to the first signal when only one of the offers would be placed on a short list of fixed length. Future research such investigate flexible short list strategies when the results of pairwise comparison are described by a richer range of outcomes (see Saaty [10]).

When errors are made in the comparison of two offers, making a larger number of comparisons would increase the search costs, but may lead to accepting a better offer. Future research such compare the effectiveness of the search rules considered here to rules where a larger number of comparisons are used in order to compare objects, e.g. each object on the short list might be compared to all of the other offers on that list. Future research should also consider the effect of the form of the joint distribution of the signals and the correlation between them on the optimal length of short lists in the presence of errors of perception.

Acknowledgements

The author is grateful for the financial support of the research project "Logistics, Trade and Consumer Decisions in the Age of the Internet" by the Polish National Science Centre via grant no. 2018/29/B/HS4/02857.

References

[1] Analytis, P. P., Kothiyal, A. and Katsikopoulos, K. (2014). Multi-attribute utility models as cognitive search engines. *Judgment and Decision Making, 95*, 403–419.

[2] Bobadilla-Suarez, S. and Love, B. C. (2018). Fast or frugal, but not both: Decision heuristics under time pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44(1)*, 24.

[3] Borah, A. and Kops, C. (2019). Rational choices: an ecological approach. *Theory and Decision, 86(3-4)*, 401–420.

[4] Knuth, D. E. (1973). *The Art of Computer Programming, Volume 3, Sorting and Searching*. Reading, Mass.: Adison-Wesley.

[5] Lleras, J. S., Masatlioglu, Y., Nakajima, D. and Ozbay, E. Y. (2017). When more is less: Limited consideration. *Journal of Economic Theory, 170*, 70–85.

[6] Masatlioglu, Y., Nakajima, D. and Ozbay, E. Y. (2012). Revealed attention. *American Economic Review, 102(5)*, 2183–2205.

[7] Murphy, J., Hofacker, C. and Mizerski, R. (2006). Primacy and recency effects on clicking behavior. *Journal of Computer-Mediated Communication, 11(2)*, 522–535.

[8] Ramsey, D. M. (2015). On a sequential decision process where offers are described by two traits, *Multiple Criteria Decision Making, 10*, 141–154.

[9] Ramsey, D, M. (2019). Optimal selection from a set of offers using a short list. *Multiple Criteria Decision Making, 14*, 75–92.

[10] Saaty, T. L. (1980) *The Analytic Hierarchy Process*, New York: McGraw-Hill, New York.

[11] Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63(2)*, 129.

[12] Todd, P. M. and Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences, 23(5)*, 727–741.

Measuring Performance of European Airports

Jana Rejentová[1], Markéta Matulová[2]

Abstract. This paper presents a performance evaluation of European airports, based on the application of Data Envelopment Analysis (DEA). We have evaluated 115 busiest airports in Europe according to the number of checked-in passengers in 2018. We used four inputs, including the number of Terminals, Runways, Boarding gates, and Aircraft stands. Three variables were used on the output side, namely Passengers, Movements, and Cargo. First, we estimated the airport efficiencies using the slack based measure super-efficiency model for the pooled dataset. Then we compared the performance of airports in public, private, and mixed ownership by applying the program evaluation procedure. It is based on separate efficiency evaluation within the groups followed by the comparison of the frontier values of the separate models. Kruskal-Walis test revealed statistically significant differences in the performance of fully public, partially public, or privately owned airports.

Keywords: airport benchmarking, data envelopment analysis, ownership, slack based DEA models, super-efficiency

JEL Classification: C61; C44; C24; R40
AMS Classification: 90C15

1 Introduction

Airport benchmarking serves the stakeholders as a useful tool to support their decision-making. Managers can use it to identify best practices, and develop new concepts for improvement and governments can assess the impact of policy decisions. The aviation industry developed rapidly over the last years, and different ways of managing airports evolved, which have brought the opportunity for the application of various benchmarking techniques. In our study, we evaluate the relative efficiency of 115 European airports using the data from the year 2018. Two airports of former Czechoslovakia are included in the sample, namely Prague and Bratislava airports, so we pay special attention to the performance of these two units comparing to the rest of the sample.

One of the essential factors determining airport performance is the form of ownership. In the past, the majority of airports were owned publicly, which is still true in some cases. However, during the last two decades of the 20th century, a process of privatization took place in many countries. With this change, competition between airports has also begun to grow. However, the majority of privatized airports in Europe remain subject to an economic regulation to some extent [8]. Current measures taken by European countries in connection with the coronavirus pandemic will have a very severe impact on transportation companies, including airlines and airports. Some governments are planning to nationalize major airlines to save them from bankruptcy, and similar measures may affect the airports. The change in the airports' ownership can probably have a significant impact on their productivity and efficiency. In the second part of our analysis, we provide a comparison of the performance of airports with different forms of ownership.

The paper is organized as follows: State of the art is presented in the second section. The data sample, as well as the model for the evaluation of efficiency and statistical analysis, are described in the third section. The results are presented and discussed in the fourth section and the last section is concluding.

2 State of the Art

Since the late 1990s, a lot of academic research emerged, applying quantitative approaches to assess the productivity and efficiency of airports. Liebert mentions in her survey 38 studies applying parametric methodology (mostly stochastic frontier analysis) and 29 nonparametric studies (mostly data envelopment analysis). Some authors however, debate the relevance of airport performance evaluation due to inconsistencies in the data selection and model specifications consequencing in its limited value for managers. Different

[1] Masaryk University, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, Czech Republic , 436972@mail.muni.cz
[2] Masaryk University, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, Czech Republic, Marketa.Matulova@econ.muni.cz

techniques and the airports' heterogeneous character may lead to controversies across studies [10]. The most of the research has been limited in the number of units compared, and the regional specification; only 10 papers out of 67 covered by the survey of Liebert [9] are based on the sample including European airports, and their number ranges only from 25 to 48, so our article aims to fill this gap.

The effects of different ownership forms on efficiency were analyzed by many empirical studies, but the results have not reached clear conclusions. Parker [15] utilizes DEA to estimate the technical efficiency of the British airports covering the period pre and post-privatization. He finds no evidence that complete privatization leads to improved technical efficiency. Barros and Dieke [3] analyze data on 31 Italian airports from 2001 to 2003 to reveal that private airports operate more efficiently than their partially private counterparts. However, Lin and Hong [11] find no connection between ownership form and efficiency after analyzing a dataset of worldwide airports for the years 2001 and 2002. Oum et al. [13] distinguish between public airports owned by public corporations and those owned by more than one public shareholder and conclude that different ownership and governance structures affect the quality of managerial performance. Conversely, Oum et al. [14] assess a sample of 100 airports worldwide covering the years 2001 to 2003 and they reach the conclusion that the productivity of a public corporation is not significantly different from that of a major private airport. So the effect of ownership on the efficiency of airports remains so far an open question.

3 Material and Methods

3.1 Data for the analysis

There are 1093 airports with assigned IATA (International Air Transport Association) code in Europe, see Airport Database 2019 [2]. For our analysis, we intended to cover 150 busiest European airports according to the number of passengers transported in 2018. The airport traffic report ACI EUROPE 2018 shows that most passengers came through London Heathrow Airport, followed by Charles Airport de Gaulle in Paris and Schiphol in Amsterdam. Prague airport ranked 37th, and Bratislava airport ranked 129th [1]. Data collection proved to be quite problematic. No summary was available for many airports, therefore data were searched individually, but relevant information was missing in some cases. That is why we had to exclude 35 airports, so finally our data sample covers 115 of them. The map in Figure 1 shows all the airports originally considered for analysis. It is color-coded for airports with complete data and airports excluded from the analysis. We can see that almost every European country has at least one representative in the sample.

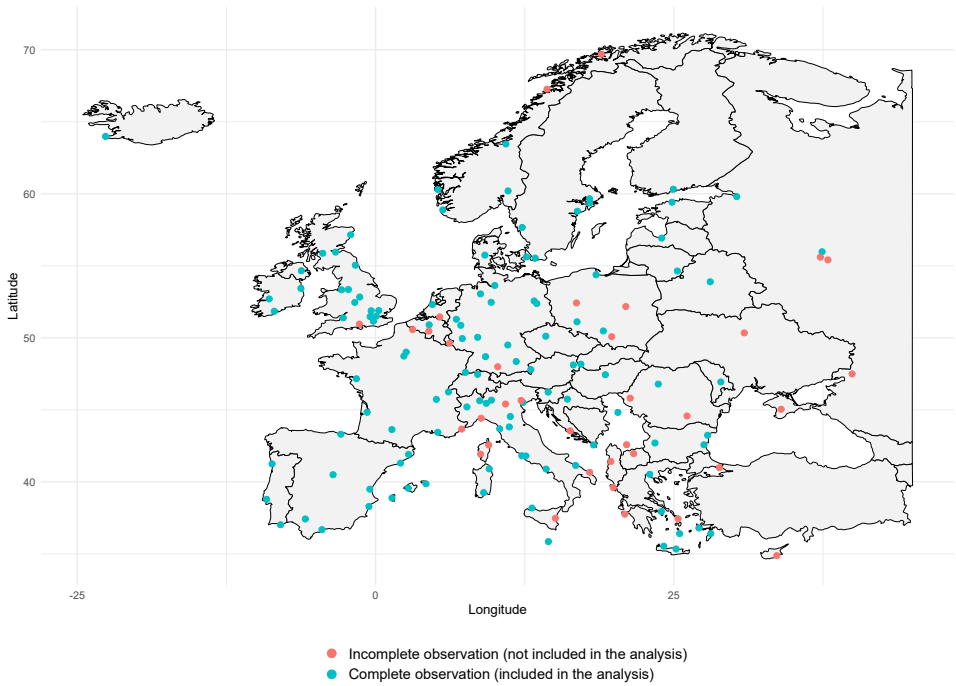


Figure 1 Map of airports locations

We have decided to express the output side of the analysis by the variables, including the number of pas-

sengers, aircraft movements and cargo. The number of runways, terminals, gates, and aircraft stands were selected for the inputs. When selecting relevant variables for the analysis, we built on a literature review. We surveyed 27 major studies on airport efficiency and found out that the inputs and outputs were represented by our selected variables or their equivalents very often. The number of terminals was used twice (plus 9 times total area of terminals and twice the capacity of terminals), the number of runways was used 7 times (plus 3 times their area and 4 times their total length), boarding gates 5 times, and aircraft stands 6 times (plus 3 times their area). On the output side, 20 studies used the number of passengers, 18 number of aircraft movements, 17 the total tuns of cargo. In addition to the characteristics we use, there are sometimes used also other variables, like total costs, staff, or check-in counters on the input side and revenues on the output side [10]. Descriptive statistics of inputs and outputs of units included in our study are in Table 1.

	Terminals	Runways	BoardingGates	AircraftStands	Passengers	Movements	Cargo [t]
Minimum	1.000	1.00	4.00	6.00	1677661	13195	51
1st quartile	1.000	1.00	14.50	24.00	4310005	42853	3800
Median	1.000	1.00	23.00	40.00	6962040	76995	18543
Mean	1.687	1.67	43.52	60.85	14309037	117317	141920
3rd quartile	2.000	2.00	53.00	75.00	17900050	149867	93922
Maximum	5.000	6.00	269.00	340.00	80124537	512115	2176389

Table 1 Descriptive statistics of inputs and outputs

Our sample comprises of 21 fully private airports, 44 fully public, and 50 airports in mixed ownership.

3.2 Super-efficiency slack based DEA model

Basic models of data envelopment analysis are used to evaluate the relative performance of decision making units (DMUs) in a homogeneous sample of n units. The efficiency of DMUs is expressed on the basis of m inputs and r outputs; their values for i-th DMU can be denoted by x_i and y_i, $i = 1, \ldots, n$. Tone [18] has proposed a DEA model that measures the efficiency of the units under evaluation using slack variables (Tone's Slacks Based Measure model; SBMT model). As the goal of many benchmarking studies is to obtain a unique ranking of the DMUs, it is not desirable to use a method that identifies a large portion of the units as fully efficient. This situation can often occur when we use basic radial DEA models, especially when we use many inputs and outputs and assume variable returns to scale. So the SBMT model has been further modified [19] to super SBMT model having more discriminatory power for the units identified as efficient by the basic model. The super-efficiency SBMT model removes the evaluated unit DMU_q from the set of DMUs and looks for its projection on the efficient frontier (DMU^*) determined by remaining $n-1$ units. The super-efficiency measure is defined as the distance of the units DMU_q and DMU^* in their input and output space. The virtual inputs x_i^*, $i = 1, \ldots, m$, and outputs y_k^*, $k = 1, \ldots, r$, are determined by the fractional linear program

$$
\begin{aligned}
&\min_{\forall \lambda_i, x_j^*, y_k^*} \frac{\frac{1}{m}\sum_{j=1}^{m} \frac{x_j^*}{x_{qj}}}{\frac{1}{r}\sum_{k=1}^{r} \frac{y_k^*}{y_{qk}}}, \\
&\text{subject to} \quad \sum_{i=1}^{n} x_{ij}\lambda_i \le x_j^*, \quad j = 1, \ldots, m, \\
&\qquad \sum_{i=1}^{n} y_{ik}\lambda_i \ge y_k^*, \quad k = 1, \ldots, r, \\
&\qquad x_j^* \ge x_{qj}, \quad j = 1, \ldots, m, \\
&\qquad y_k^* \le y_{qk}, \quad k = 1, \ldots, r, \\
&\qquad \lambda_i \ge 0, \quad i = 1, \ldots, n, i \ne q, \\
&\qquad \lambda_q = 0, \\
&\qquad \sum_{i=1}^{n} \lambda_i = 1 \quad .
\end{aligned}
\tag{1}
$$

The last constraint corresponds to the variable returns to scale assumption. When applying the method to

our data, we followed the two-phase procedure recommended by [7]:

1. Use the basic SBMT model (without excluding the unit under evaluation from the set of DMUs in the model) to divide units into efficient and inefficient ones, for which we compute the SBMT efficiency values.
2. Compute the super-efficiency according to the super SBMT model 1 for efficient units from the first phase.

By this procedure referred to as SBMT/superSBMT model we get values greater than 1 for efficient units and less than 1 for inefficient DMUs.

3.3 Program evaluation procedure

The purpose of this procedure is to distinguish between the degree of inefficiency caused by the uneconomical operation of the units and that caused by belonging to a particular group. Program evaluation procedure outlined by Brockett and Golany [5] and Sueyoshi and Aoki [17] includes four steps:

1. Split the group of all DMUs ($j = 1, \ldots, n$) into p programs consisting of $n_1, \ldots, n_p$ DMUs ($\sum_{i=1}^{p} n_i = n$). Run DEA separately for the individual groups.
2. In each of the p groups separately, adjust inefficient DMUs to their "level if efficient" values by projecting each DMU onto the efficiency frontier of its group.
3. Run a pooled DEA with all n DMUs at their adjusted efficient levels.
4. Apply a statistical test (Mann-Whitney for two groups or Kruskal-Wallis for $p > 2$) to the results of the previous step to determine if the groups have the same distribution of efficiency values within the pooled DEA set.

4 Results and discussion

We started the analysis by benchmarking of the selected airports. The computations were performed using the `R` package `deaR` and its procedures `model_sbmeff` and `model_sbmsupereff` [16]. After the application of the SBMT/superSBMT model, Hahn airport (Europe's 136th busiest airport) appears to be the most efficient; London Heathrow (Europe's 1st busiest airport) is the second most efficient and Amsterdam (Europe's 4th busiest airport) is the third most efficient. They are followed by Reykjavík and Kós airports on the 4th and the 5th position. The worst five performers included Bristol, Girona, Faro, Cork, and Olbia airports. Prague airport ranked 61st, and Bratislava ranked 66th, so they are both in the middle of the ranking. For these two airports of particular interest, we tried to find their peer units. The virtual unit for Prague airport is a combination of the following airports: London Luton (with a coefficient of 0.4651), Stavanger (with a coefficient of 0.2107), Palma de Mallorca (with a coefficient of 0.2063), Berlin Tegel (with a coefficient of 0.0898), and Frankfurt am Main (with a coefficient of 0.0280). For Bratislava, we obtain a virtual unit combining the airports of Rhodes (with a coefficient of 0.9863) and Amsterdam (with a coefficient of 0.0137). The scores of the ten busiest airports, together with Prague and Bratislava, are given in Table 2.

Next, we analyzed the effect of ownership on airport efficiency measures using the procedure presented in subsection 3.3. In Figure 2 we can see two graphs showing the boxplots of efficiency for individual types of ownership. The black line in the box diagram corresponds to the median, the black dot then to the mean efficiency. The graph on the left shows the situation when the program evaluation procedure of subsection 3.3 was not used. It is therefore a comparison of efficiencies between different ownership groups using an overall model. If we perform the Kruskal-Wallis test for these values, we obtain a p-value of 0.762, and thus we do not reject the null hypothesis that there is no effect of ownership on efficiency. However, this evaluation includes possible aspects of the uneconomical operation of individual airports. The boxplot on the right side of Figure 2 is obtained by the proposed procedure for group comparison using SBMT/superSBMT models. The graph show differences between groups. In this case, the Kruskal-Wallis test returns a p-value of 0.064, and therefore we reject the null hypothesis that there is no effect of airport ownership on the efficiency level at the significance level of 10%. On average, privately owned airports seem to be the most efficient, followed by a group of mixed-owned airports, and public airports appear to be the least efficient.

5 Conclusions

In this paper, we have evaluated the performance of 115 European airports. We have used the slack based measure super-efficiency DEA model based on inputs, including the number of terminals, runways, boarding gates, and aircraft stands, and outputs represented by the number of passengers, aircraft movements, and

ID	Airport	Ownership	SBMT/superSBMT efficiency	Ranking
1	London Heathrow	Fully private	1.266	2
2	Paris Charles de Gaulle	Mixed	1.011	36
3	Amsterdam Schiphol	Mixed	1.258	3
4	Frankfurt am Main	Mixed	1.152	11
6	Madrid	Mixed	0.310	64
7	Barcelona	Mixed	0.222	71
8	Munich	Fully public	1.071	18
9	London Gatwick	Fully private	1.089	14
10	Moscow Sheremetyevo	Mixed	1.054	22
...	...	...	...	...
37	Prague	Fully public	0.353	61
...	...	...	...	...
129	Bratislava	Fully public	0.267	66
...	...	...	...	...

Table 2 Results of the SBMT/superSBMT model for ten busiest European airports + Prague and Bratislava airports

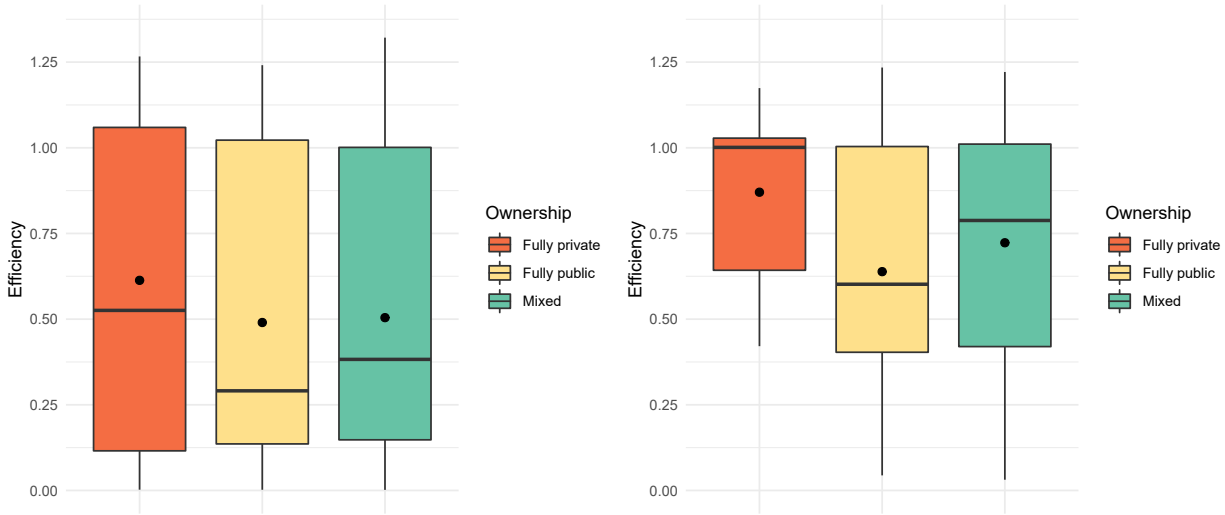


Figure 2 Efficiency comparison; on the left model for the original data, on the right model for the corrected values

tons of cargo. We identified Hahn, London Heathrow, Amsterdam Schiphol, Reykjavík, and Kós airports as the five top performers. Prague and Bratislava airports ranked in the middle of the sample.

Using the naive approach comparing the efficiency scores of individual airports, we found no statistically significant difference in the performance of airports with different forms of ownership. After applying advanced approach controlling for the influence of inefficiencies caused by individual managerial failures we obtained quite contrasting results identifying private airports as the most efficient and public airports as the least efficient.

Limitation of our study is given by the sensitivity of results to methodology employed. The values of efficiency scores are influenced by the model specification including the selection of a particular model from the portfolio of parametric or nonparametric approaches, and determination of inputs and outputs, economies of scale, etc. The results are not suitable for benchmarking purposes concerning possible reductions of inputs unless we include in the model also the constraints requiring integer values of some inputs (e.g. terminals or runways). In addition, some annual reports of airport performance show evidence of inconsistencies over time and there is also high possibility of occurrence of other biases caused e.g. by longer-term effects of lumpy capital investments. [12].

References

[1] ACI EUROPE, 2018. Airport traffic report (December, Q4, H2 & Full Year 2018). [online]. [accessed 2019-07-15]. Available from: https://www.aci- europe.org/policy/positionpapers.html

[2] AIRPORTDATABASE, 2019. [online]. [accessed 2019-07-15]. Available from: https://airportdatabase.net/europe-airports.html.

[3] Barros, C. P. & Dieke, P. U. (2007). Performance evaluation of Italian airports: a data envelopment analysis, *Journal of Air Transport Management*, 13(4), 184–191.

[4] Bogetoft P. & Otto L. (2011). *Benchmarking with DEA, SFA, and R*. New York: Springer.

[5] Brockett, P. L. & Golany B. (1996) Using rank statistics for determining programmatic efficiency differences in data envelopment analysis, *Management Science*, 42(3), 466–472.

[6] Coelli, T. J., Rao, D. S. P., O'Donnell, C. J. & Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. New York: Springer.

[7] Dlouhý, M., Jablonský, J. & Zýková, P. (2019). *Analýza obalu dat*. Praha: Professional publishing.

[8] Gillen, D. (2010). The evolution of airport ownership and governance, *Journal of Air Transport Management*, 17, 3–13.

[9] Liebert, V. (2011). *Airport Benchmarking: An Efficiency Analysis of European Airports from an Economic and Managerial Perspective*, Bremen: Jacobs University.

[10] Liebert, V. & Niemeier H. M. (2013). A Survey of Empirical Research on the Productivity and Efficiency Measurement of Airports, *Journal of Transport Economics and Policy*, 47, 157–189.

[11] Lin, L. C. & Hong, C. H. (2006). Operational performance evaluation of international major airports: an application of data envelopment analysis, *Journal of Air Transport Management*, 12(6), 342–351.

[12] Morrison, W. G. (2009). Understanding the complexities and challenges of airport performance benchmarking, *Journal of Airport Management*, 3, 145–158.

[13] Oum, T. H., Adler, N. & Yu, C. (2006). Privatization, corporatization, ownership forms and their effects on the performance of the world's major airports, *Journal of Air Transport Management*, 12(3), 109–121.

[14] Oum, T. H., Yan, J. & Yu, C.(2008). Ownership forms matter for airport efficiency: A stochastic frontier investigation of worldwide airports, *Journal of Urban Economics*, 64(2), 422–435.

[15] Parker, D. (1999). The performance of BAA before and after privatisation: A DEA study, *Journal of Transport Economics and Policy*, 33(2), 133–145.

[16] R package deaR [online]. [accessed 2019-07-15]. Available from: https://cran.r-project.org/web/packages/deaR.

[17] Sueyoshi, T., & S. Aoki (2001) A use of a nonparametric statistic for DEA frontier shift: the Kruskal and Wallis rank test, *Omega*, 29(1), 1–18.

[18] Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis *European Journal of Operational Research*, 130(3), 498–509.

[19] Tone, K. (2002). A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, 143(1), 32–41.

The Time Augmented Cobb-Douglas Production Function as a Tool of the Potential Output Estimation

Lenka Roubalová[1],

Abstract. The main aim of this paper is to propose and verify the most appropriate specification of the time augmented Cobb-Douglas production function and use it as an alternative way for the potential output and production gap estimation in the Visegrád Four countries, Austria and Germany during 1995–2015. We also compare the results based on the time augmented Cobb-Douglas production function including all parameters estimated and the approach based on the commonly used assumption that the labour factor intensity parameter is defined as a share of labour compensation in GDP. All the models are estimated via Nonlinear Least Squares technique and the time series data are sourced from the EU KLEMS and OECD database. Based on the results we evaluate the individual specifications with respect to the economic output development in each of the investigated countries. The results also show this function can be used as an alternative way for the purpose of the potential output analysis.

Keywords: Cobb-Douglas Production Function, Nonlinear Least Squares, Potential Output

JEL Classification: C51, E23, E24
AMS Classification: 62J02, 91B38

1 Introduction

The production function (PF) is linked to Turgot and the Classical economists [7] mainly, but also the Keynesian or rather Neokeynesian school is worth mentioning in association with it. The Cobb-Douglas production function (CDPF) is used as a tool of the economic analysis in various areas, e.g. for technological progress or for potential output and production gap analysis.

Since the potential output can't be observed directly, several ways how to estimate it has been proposed. The methodologies based on PF are: *Structural techniques* based on model of the supply-side economy using the economic theory explaining the relationship among output (Y), level of technology and input factors, labour (L) and capital (K), see [3]. As Cotis et al. [8] present, the OECD approach combines the CDPF with Harrod neutral technological progress (employing the hours worked and a NAIRU concept) and statistical filtering to calculate the participation rates trend, hours worked trend and total factor productivity (TFP) trend. The EU Commission approach is based on using information on wages that is preferred to the prices so a NAWRU is calculated instead of a NAIRU, for more details regarding NAWRU and NAIRU see [8].

The CDPF approach is widely used, also due to the link to the theoretical economic framework, see [8] or [23]. However, [2] mentions a lack of the reliable data what is problematic in case of developing countries. Due to the uncertainty, such measurement is usually accompanied by another ways of investigations like *Statistical univariate methods* using just one variable (output) based on smoothing out the trend usually by Hodrick-Prescott filter [11] that extract the trend component from the output time series or Baxter-King [5] filter based on long-term fluctuations (trend) and short-term fluctuations (cycle) separation. *Statistical multivariate methods,* usually Vector Autoregressive Models or the Kalman Filter [16] that derive information also from related series, so its development and the relationship to the trend is monitored.

The main aim of this paper is to find the most appropriate specification of the time augmented CDPF (see also [19] and [20]) and use it as an alternative way for the potential output estimation in the Visegrád Four countries, Austria and Germany during 1995–2015. We also compare the results based on the time augmented CDPF including all parameters estimated and the approach based on the commonly used assumption that the labour factor intensity parameter is defined as a share of labour compensation in GDP. All the models are evaluated, compared and discussed with respect to each other as well as with respect to the real GDP development.

[1] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operational Analysis, Zemědělská 1, 61300 Brno, xkravack@node.mendelu.cz

2 Material and Methods

The CDPF in its basic form is given by an equation $Y = \gamma \cdot K^{\alpha} \cdot L^{\beta}$; $(0 < \alpha < 1, 0 < \beta < 1)$ where Y is used for the economic output, K and L are input factors, capital and labour, and γ represents TFP. The factor intensity is represented by the parameters α (capital factor intensity) and β (labour factor intensity). Usually, the assumption of the constant returns to scale $\alpha + \beta = 1$ for the potential output estimation is adopted, so $\beta = 1 - \alpha$. Only two parameters, γ and α, are estimated in this case. Another way how the labour intensity parameter can be determined is using a labour share in the economic output. In this case β is calculated as an average share of the employee compensation in GDP. As the constant returns to scale are assumed, then $\alpha = 1 - \beta$.

The time augmented function assumes the parameters are not constant, but they are expressed as a linear function of time. As a first step, we estimate the parameters of the following time augmented versions of the CDPF: firstly, the version where only γ is time augmented, factor intensities are constant and secondly, parameters of the time augmented version of the CDPF where either γ or α are time augmented and obtained as an estimation. Such equations are given as:

$$Y_t = (c_0 + c_1 t) K^{\alpha} L^{(1-\alpha)}, \tag{1a}$$

$$Y_t = (c_0 + c_1 t) K^{(1-\beta)} L^{\beta}, \tag{1b}$$

$$Y_t = (c_0 + c_1 t) K^{(a_0 + a_1 t)} L^{(1 - a_0 - a_1 t))}, \tag{2}$$

where α is estimated constant parameter and β is calculated constant parameter given as a share defined in the previous paragraph[2], meanwhile $\gamma_t = (c_0 + c_1 t)$, $\alpha_t = (a_0 + a_1 t)$ represent a time augmented TFP and capital intensity parameters estimated via Nonlinear Least Squares (NLS) models. All models are evaluated from the point of view of the quality of the prediction using coefficient of determination R^2 and Theil´s U statistics, see [21]. Although, we don't verify the classical assumptions for NLS models as usual, we check the statistical significance using t-test p-values for the time augmented parameters c_1 and a_1 to monitor the importance of these 'extra' parameters added in CDPF.

In this paper we use Nonlinear Least Square techniques to estimate the CDPF parameters. While using NLS the adequate starting parameters are needed, for details see [22]. Then we obtain unique parameters – unique CDPF for each individual country. Next, we use the estimated parameters for each of the countries together with K and L_t^* in the equations (1a), (1b), (2) to obtain the equation expressing the potential output, where L_t^* is potential employment calculated as $L_t^* = p \cdot P \cdot \left(1 - U_t^{\mathrm{NAWRU}}\right)$, where p represents the participation rate of the productive population in the working process, P is working age population and $NAWRU$ is an unemployment rate with non-accelerating wage rate given as $U_t^{\mathrm{NAWRU}} = U_t - (\frac{\Delta U_t}{\Delta^2 wg_t}) \Delta wg_t$ and U_t represents unemployment rate and wg_t growth rate of wages, see [12]. After these steps, the potential output estimation is obtained.

As the data we use yearly time series for period 1995–2015 of selected Central European countries – Visegrád Four countries, Germany and Austria. For Poland and for Hungary the final period is shorter due to the unavailability of relevant data for full range. The data is sourced from EU KLEMS and OECD databases. For Y we use GDP in current basic prices in millions expressed in national currencies, for K we use gross fixed capital formation in millions expressed in national currencies, L is labour defined as number of employees. Such CDPF parameters based on time series data are suitable for an average stage of economy – the labour share dynamics or the TFP shifts that happen over time are not captured. Due to these shortcomings we modify the standard approach and we propose the time augmented CDPF, where both or one of the parameters are supposed to be a linear function of time. For details regarding the time augmented CDPF see [20]. Calculations are performed via MATLAB R2018b computational system, a significance level of 0.05 is held.

3 Results

First, we focused on the estimation of both versions of the CDPF augmented only in γ parameter, equations (1a) and (1b). Results are summarized in Tab. 1 and 2 showing that all the estimated models are in accordance with the economic theory. Theil's U indicates the models of the function (1a) are of better quality than

[2] Despite the fact that for both, the equations (1a) and (1b) we assume constant returns to scale we keep using two names of the parameters, either α or β to distinguish the way how the value for each parameter is obtained (estimation versus calculation).

(1b). **R^2** coefficient is quite high that indicates the relationship among *K, L, Y* is described sufficiently in both cases – models do not differ from the point of view of the prediction quality but do differ from viewpoint of the estimated parameters values. It is necessary to mention that we deal with nonstationary time series. According to [4] the spurious regression in case of nonlinear models has not been extensively examined. Nevertheless, in previous research we found the series were linearly cointegrated, see [19], and in this paper we check the stationarity for residuals of each individual nonlinear model. Our results prove the residuals are stationary, so we avoid the issue of the spurious regression even in case of NLS. The parameter of TFP (γ composed of c_0 and c_1) is for all cases either (1a) or (1b) positive that indicates the technological progress and isoquants shift inwards during the observed period.

	PL	CZ	SK	HU	AT	DE
$\widehat{c_0}$	116.11	11.22	11.87	65.34	73.68	66.99
$\widehat{c_1}$	6.81	$2.93 \cdot 10^{-1}$	1.13	2.54	3.15	1.77
$\hat{\alpha}$	$1.11 \cdot 10^{-1}$	$8.86 \cdot 10^{-1}$	$4.71 \cdot 10^{-1}$	$6.70 \cdot 10^{-1}$	$7.20 \cdot 10^{-2}$	$1.19 \cdot 10^{-1}$
p-val. c_1	0.586	<0.001	<0.001	<0.001	0.792	0.013
R^2 [%]	98.27	99.11	9874	98.37	98.49	98.05
Theil's U	0.38	0.43	0.42	0.39	0.59	0.58

Table 1 Parameters of the CDPF given by equation (1a), p-values, R2 and Theil's U.

	PL	CZ	SK	HU	AT	DE
$\widehat{c_0}$	3.63	49.12	10.58	217.66	26.07	30.64
$\widehat{c_1}$	$8.97 \cdot 10^{-1}$	2.23	$8.45 \cdot 10^{-1}$	13.61	$7.89 \cdot 10^{-1}$	6.57
β	$4.42 \cdot 10^{-1}$	$4.38 \cdot 10^{-1}$	$4.26 \cdot 10^{-1}$	$5.21 \cdot 10^{-1}$	$5.33 \cdot 10^{-1}$	$5.68 \cdot 10^{-1}$
p-val. c_1	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
R^2 [%]	97.12	98.65	98.68	98.21	98.36	97.10
Theil's U	0.55	0.53	0.52	0.52	0.63	0.71

Table 2 Parameters of the CDPF given by equation (1b), p-values, R2 and Theil's U.

In Tab. 3 there is an overview presenting capital and labour intensity parameters estimated and calculated. In case of equation (1b) a value of the capital intensity exceeds labour factor intensity for CZ, PL, SK and the labour factor intensity value exceeds the capital intensity for HU, AT and DE. However, the values of both parameters do not differ significantly across the whole group of countries. However, for results of (1a) the values differ significantly either among capital and labour intensities or among countries.

equation	parameter	PL	CZ	SK	HU	AT	DE
(1a)	$\hat{\alpha}$	0.111	0.886	0.471	0.670	0.072	0.119
	$1-\hat{\alpha}$	0.889	0.114	0.529	0.330	0.928	0.881
(1b)	$1-\beta$	0.558	0.562	0.574	0.479	0.467	0.432
	β	0.442	0.438	0.426	0.521	0.533	0.568

Table 3 Summary of the capital intensity parameters estimated α (and labour intensity parameters derived $1-\alpha$) obtained via (1a) and labour intensity calculated β (and capital intensity derived as $1-\beta$) obtained via (1b).

In comparison with intensities derived form (1a) in case of (1b) also the major impact of one factor intensity is different, especially for HU and PL, that is reflected in the isoquant slope – for PL, DE, HU in case of (1a) isoquants are steeper, for CZ and AT (1a) isoquants are much more flat than for (1b) and only for SK both versions of the CDPF produce very similar results. As this part of the analysis shows the models with all the parameters estimated (1a) are of better quality than (1b) we estimate also the CDPF time augmented in all parameters in Tab. 4.

Tab. 4 shows the specification (2) is better (evaluated by Theil's U from the viewpoint of the prediction quality) than the previous specifications for almost all countries with exception of SK, where (1a) version is better. For SK we can see very similar models across all these three specifications. If we compare the capital and labour intensity values, the values obtained via specification (2) are close to (1a) for SK and AT. For CZ and HU are values closer rather to (1b) version.

	PL	CZ	SK	HU	AT	DE
$\widehat{c_0}$	395.32	64.18	10.28	395.90	61.38	22.02
$\widehat{c_1}$	−21.30	−2.05	1.65	−14.44	−1.97	1.43
$\widehat{a_0}$	$-3.31 \cdot 10^{-1}$	$5.11 \cdot 10^{-1}$	$4.56 \cdot 10^{-1}$	$3.77 \cdot 10^{-1}$	$1.31 \cdot 10^{-1}$	$5.24 \cdot 10^{-1}$
$\widehat{a_1}$	$4.50 \cdot 10^{-2}$	$1.50 \cdot 10^{-2}$	$-4.79 \cdot 10^{-3}$	$1.15 \cdot 10^{-2}$	$2.70 \cdot 10^{-2}$	$-7.00 \cdot 10^{-3}$
p-val. c_1	0.119	<0.001	0.003	0.071	0.032	0.002
p-val. a_1	<0.001	<0.001	0.036	<0.001	<0.001	<0.001
R^2 [%]	98.97	99.39	99.01	99.39	99.66	98.83
Theil's U	0.31	0.35	0.43	0.28	0.31	0.46

Table 4 Parameters of the CDPF given by equation (2), p-values, R2 and Theil's U.

For PL we obtained negative value for a_0 estimation that means that capital intensity[3] is negative for first 6 years of the observed period. Based on values a_0 and a_1 it is worth mentioning that for model based on equation (2) capital intensity value exceeds the labour intensity value during the first three years of the observed period. With use of relevant data, a NAWRU time series is calculated that is essential for L^* calculation. Using L^* and K the potential output time series based on these models are obtained for each country, see Fig. 1.

4 Discussion

In Fig. 1 we can see for **CZ** that the real product is very close to the estimated Y^* the whole period. In [13] there is an evidence of Y above Y^* between 1994–2007 and 2000–2003 that is visible also in our results obtained via (1a) and (1b). According to [13] the Y^* development was driven mainly by capital that is in accordance mainly with our model (1a) where the capital intensity parameter is dominant. Also, looking further at the model (2) and the parameters a_0 and a_1 we reveal the capital (labour) intensity parameter trend is increasing (decreasing) during the period, however the impact of capital intensity parameter prevails even in the first year. In 2008 we can see a sharp decline of Y that reflects impact of actual economic issue – beginning of the crisis. Our results indicate this jump is not described by the models sufficiently. In case of model (2) we do not observe such a sharp decline of the Y^* in 2009 that indicates the model is better than (1a) and (1b). Model (2) shows the Y development under Y^* for the rest of the period. However, according to (1a) and (1b) it is above Y^*. According to [14] we can see the Y^* estimation depends on the method[4]. In [1] the conventional methods are used, and authors highlight that the different methods can lead to conflicting conclusions regarding the actual stage of the economic cycle.

For **SK** the development of the Y^* is very similar for all three CDPF versions that indicate Y is almost all the time under Y^* with exception of the first stage of the period. Results of [23] conclude that structural reforms in 1998 led to an increase of unemployment till 2001. As a result, the labour contribution to potential output growth increased. Using estimated parameters a_0 and a_1 we derive the labour intensity parameter in each year of the analyzed period increases**.** However, in [23] a short period when Y was above Y^* that cannot be confirmed by our results. On the other hand, in [23] several approaches were tested: the frequently used conventional ones and a new one based on MV Kalman filtering and the results obtained for SK differ. It is pointed out here that the output gap narrowed slightly in 2006 and the size of the positive output gap in 2007 is much smaller than that obtained by the conventional methods. They conclude in [15] non-conventional methods might be advantageous because the conventional approaches are interrelated with slack on labour market.

[3] Capital intensity here α composed by $a_0 + a_1 t$.
[4] Using conventional CDPF estimates the real product is below the potential level between 2009–2014 but statistical methods show there is a short period when the output gap is positive.

The development of the Y^* in case of **DE** shows, that according to models (1a) and (2) Y is under Y^* during the whole period. Looking at the development given by (1b) we can see Y is slightly above Y^* in 2003–2008. As the results show, the production gap, based on (2), also has become wider since 2008. In [3] authors show a recovery in Y^* growth, mainly due to the contributions of labour factor it became closer to the estimated pre-crisis growth rates. For 2008–2013 the crisis primarily affected capital and labour components of potential output. Paper [3] also shows the share of labour and capital contribution to the Y^* are very similar in 2002–2012 that is in accordance rather with our results resulting from (2) and (1b). According to (1a) the impact of labour is dominant.

For **HU** we can see the estimated Y^* differs across the models. The results obtained by (1a) and (1b) are almost identical exceeding Y till 2005, then the production gap is minimal and then, for 2009–2015 all three models produce very similar output and the production gap is positive in spite of the fact the model (2) estimation shows positive production gap during the whole period that differs from the development described by the previous two specifications. In [6] there is an evidence referring to Y slightly above Y^* in 2001. It is in accordance with our results (1a) and (1b) that show labour intensity exceeding the capital intensity parameter during the whole time period, meanwhile model (2) shows the shares are reversal since 2007. In [10] very low Y^* growth is affected by the labour market weaknesses.

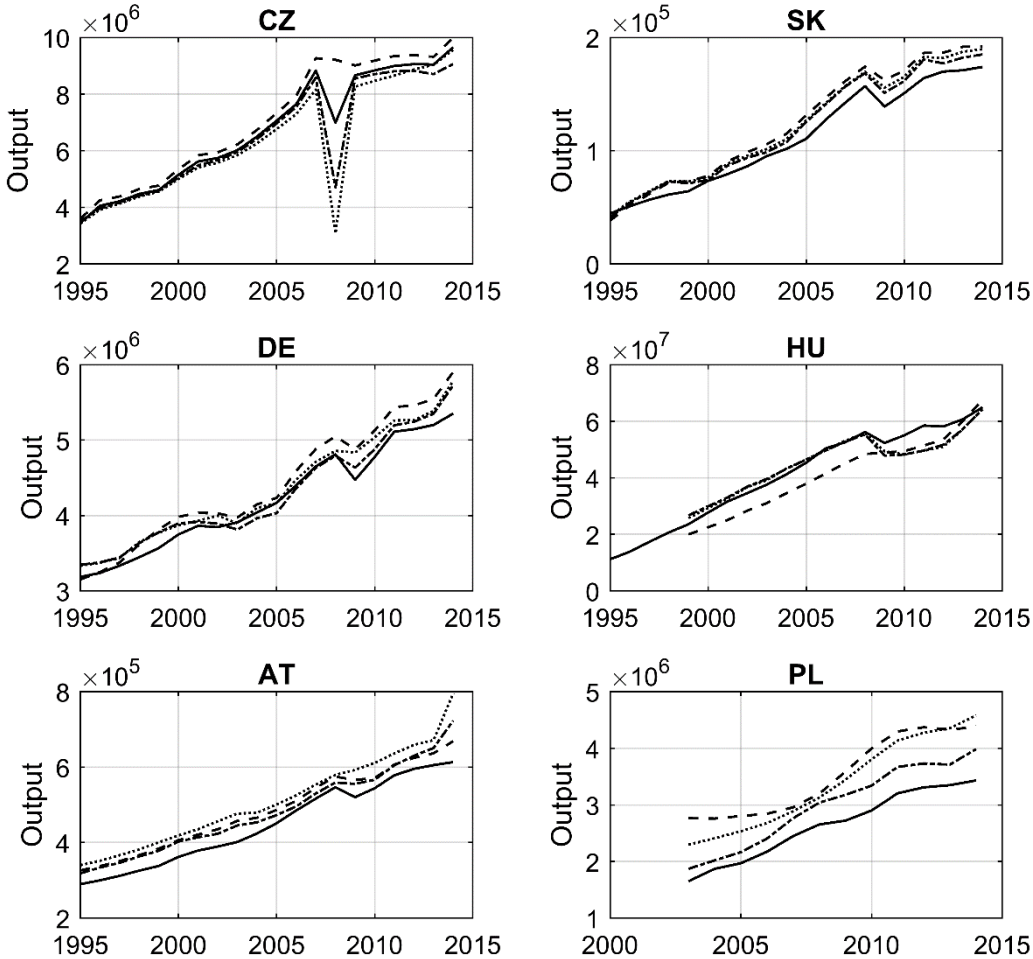


Figure 1 Real output in thousands of national currencies Y (solid line) versus Y^* estimated via various specifications of the CDPF – (1a) dotted line, (1b) dash-dotted line, (2) dashed line – for individual countries.

For **AT** we can see Y under Y^* during the whole period and the estimation of the Y^* do not differ significantly. By [17] the TFP has slowed down that is reflected in model (2) by the negative c_1, that may be result of wider dispersion of productivity across economic subjects due to slower diffusion of innovation. Employing the time variable (2) we find out the capital intensity parameter increases over time while the labour intensity parameter trend is decreasing. Such result is supported in [17] arguing that labour productivity also has slowed down.

For **PL** all three models show *Y* under *Y**. Estimation obtained via (2) and (1a) shows bigger negative production gap than we can see in case of (1b). As mentioned above, capital intensity is negative for first 6 years of the observed period that contradicts the basic economic theory. Such failing of the model indicates considerable impact of labour factor in 2003–2009. Results of [9] present labour contribution *Y** increasing in this period reflecting a decrease in the NAIRU from 2004. On the other hand, [18] shows for 2007–2008 the unemployment rate fell below its equilibrium level as *Y* exceeded *Y** level that is not observable in our results.

5 Conclusion

In this paper we propose an approach of *Y** estimation based on a time augmented CDPF. We revealed the results differ for each of the verified versions. We prove that the time variable can improve the basic model. This approach is able to identify quickly the contribution of all the components and *Y** development. Compared with the results of the conventional methods we found some minor discrepancies (but these papers often use value added instead of GDP or NAIRU instead of NAWRU that may cause deviations too). In conclusion, we recommend this way just for a quick insight into the macroeconomic summary or use it as a supplement to standard methods of *Y** estimation. For further research we recommend also further investigation of an output gap and looking at the topic from the point of view of growth of *Y* and *Y** and focus on the future predictions as well.

Acknowledgements

This article was supported by the grant No. PEF/TP/2020002 of the Grant Agency IGA PEF MENDELU.

References

[1] ADAMEC, V. & STŘELEC, L. (2012). A Study of Potential Output and Output Gap in the Czech Republic. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 60(2), 9–16.

[2] ALKHAREIF, R. M., BARNETTT, W. A., & ALSADOUN, N. A. (2017). Estimating the Output Gap for Saudi Arabia. *International Journal of Economics and Finance,* 9, 81–90.

[3] ANDERTON, R. (2014). Potential Output from Euro Area Perspective. *Occasional Paper Series,* 156, 1–64. [Online]. Available at: https://www.ecb.europa.eu/pub/pdf/scpops/ecbop156.en.pdf. [Accessed: 2020, April 11].

[4] BASS, F. M. & SRINIVASAN, S. (2002). A Study of Spurious Regression and Model Discrimination in the Generalized Bass Model. *Econometric Models in Marketing,* 16, 295–315.

[5] BAXTER, M. & KING, R. G. (1999). Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. *Review of Economics and Statistics,* 81, pp. 575–593.

[6] BENK, S., JAKAB, Z. M. &VADAS, G. Potential Output Estimations for Hungary: A Survey of Different Approaches. *MNB Occasional Papers,* 43, 1–48. [Online]. Available at: https://www.mnb.hu/letoltes/op-43.pdf. [Accessed: 2020, April 11].

[7] BLAUG, M. (1997). *Economic Theory in Retrospect*. 5. Cambridge: Cambridge University Press.

[8] COTIS, J. P., ELMESKOV, J. & MOUROUGANE, A. (2005). Estimates of Potential Output: Benefits and Pitfalls from a Policy Perspective. In Rechling, L. *(Ed.). Euro Area Business Cycles: Stylized Facts and Measurement Issues.* London: CEPR.

[9] EPSTEIN, N. & MACCHIARELLI, C. (2010). Estimating Poland's Potential Output: A Production Function Approach. *IMF Working Papers,* 10, 1–20.

[10] EUROPEAN COMMISSION. (2012). Macroeconomic Imbalances – Hungary. European Economy Occasional Papers, 106, 1–29. [Online]. Available at: https://ec.europa.eu/economy_finance/publications/occasional_paper/2012/pdf/ocp106_en.pdf. [Accessed: 2020, April 11].

[11] HODRICK, R. J. & PRESCOTT, E. C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 29, 1–16.

[12] HOLDEN, S. & NYMOEN, R. (2002). Measuring Structural Unemployment: NAWRU Estimates in the Nordic Countries*. The Scandinavian Journal of Economics*, 104(1), 87–104.

[13] HURNÍK, J. & NAVRÁTIL, D. (2005). Potential Output in the Czech Republic: A Production Function Approach. *Prague Economic Papers,* 2005(3), 253–266.

[14] IMF. (2016). Czech Republic Article IV Consultation – Press Release; Staff Report and Statement by the Executive Director for Czech Republic. *IMF Country Report,* 213(16), 1–73. [Online]. Available at: https://www.imf.org/external/pubs/ft/scr/2016/cr16213.pdf. [Accessed: 2020, April 11].

[15] KONUKI, T. (2008). Estimating Potential Output and the Output Gap in Slovakia. *IMF Working Papers,* 275(8*)*, 122.
[16] MAYBECK, P. S. (1979). *Stochastic Models, Estimation, and Control*. Academic Press: New York.
[17] OECD. (2017) *OECD Economic Surveys: Austria.* [Online]. Available at: https://www.oecd-ilibrary.org/economics/oecd-economic-surveys-austria-2017_eco_surveys-aut-2017-en. [Accessed: 2020, April 11].
[18] OECD. (2014). *OECD Economic Surveys: Poland*. [Online]. Available at: http://www.oecd.org/economy/surveys/Overview_Poland_2014.pdf [Accessed: 2020, April 11].
[19] ROUBALOVÁ, L. & VISKOTOVÁ, L. (2018). Productivity Development in Selected Central European Countries Measured by the Sato Production Function. *Review of economic perspectives*, 18(4), 353–370.
[20] ROUBALOVÁ, L. & VISKOTOVÁ, L. (2019). The Time Augmented Cobb-Douglas Production Function. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 67(5), 1347–1356.
[21] THEIL, H. (1966). *Applied Economic Forecasting*. Amsterdam: North-Holland.
[22] VISKOTOVÁ, L., ROUBALOVÁ, L., & HAMPEL, D. (2019). Starting Parameters for The Estimation of the Production Functions, *Proceedings of the 16th International Conference of Numerical Analysis and Applied Mathematics* (pp. 13–18). Melville: American Institute of Physics (AIP).
[23] ZIMKOVÁ, E. & BAROCHOVSKÝ, J. (2007). Odhad potenciálneho produktu a produkčnej medzery v slovenských podmienkach. *Politická ekonomie*, 4, 473–489.

Amortization Schedule via Linear Difference Equations

Dana Říhová[1]

Abstract. The main aim of this paper is to show an application of difference equations in the field of finance. We deal with the loan repayment of constant annuities and derive formulas which are used to create an amortization schedule. We focus especially on calculation the amount of interest and the amount reducing the outstanding principle as well as the loan balance in each payment period.
All necessary formulas are obtained by solving difference equations, contrary to the practice of financial mathematics where sequence properties are used. It is shown that recursion between two consecutive elements of considered sequences constitutes actually the first order linear difference equation with constant coefficients. As the mentioned formulas used in amortization schedule represent the rules for calculating an arbitrary element of such sequences to find them means to solve the appropriate difference equations which is demonstrated in this contribution.

Keywords: amortization schedule, loan repayment, linear difference equation

JEL Classification: G51
AMS Classification: Primary 39A06, Secondary 97M30

1 Introduction

The loan repayment is the process by which a loan is repaid by a sequence of periodic payments called annuities consisting of part payment of interest and part payment to reduce the outstanding principal. In the following text we will deal with the loan repayment of constant annuities.

An amortization schedule is a complete table with periodic loan repayments which includes the amount reducing the outstanding principal and the amount of interest in each payment and also the outstanding debt during the entire loan term until the debt is repaid at the end. Early in the schedule the majority of each payment is interest, later the majority of each payment covers the loan principal. More information you can find in [8], [12] or [14].

Let r denote the annual interest rate and k stand for the number of interest periods in a year. In this case the number of periodic payments per one interest period equals to one. The fraction $\frac{r}{k}$ equals to the interest rate per one compounding period and will be denoted i. For simplicity, we will not consider tax on interest income.

2 Loan repayment

Let us assume that the debt to be repaid is D and the periodic payments are constant over time and equal to a. The payments are made at the end of each period. Suppose that an annual interest rate is r and the loan repayments are made k times a year. Further, we assume that n is the total number of periods during which the loan is repaid. Let D_j represent the outstanding debt after the j-th payment. The model of loan repayment (see also [3] and [9]) can be expressed by the following recursion

$$D_{j+1} = (1+i)\,D_j - a, \qquad j = 0, 1, 2, \ldots, n-1 \tag{1}$$

where the initial element is given by the initial debt

$$D_0 = D. \tag{2}$$

We can see according to [7] that the above recurrence represents the first order nonhomogeneous linear difference equation with constant coefficients

$$D_{j+1} - (1+i)\,D_j = -a, \qquad j = 0, 1, 2, \ldots, n-1. \tag{3}$$

[1] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, dana.rihova@mendelu.cz

To find the functional notation for the j-th element of the above sequence means to solve this difference equation.

In the same way as in [9], [10] and [13] we obtain the general solution of (3) according to the superposition principle (see [6]) as the sum of the general solution to the appropriate homogeneous equation and an arbitrary particular solution to the nonhomogeneous equation. As $1+i$ is the root of the corresponding characteristic equation of (3) the general solution of the appropriate homogeneous difference equation is given by

$$C(1+i)^j, \quad C \in \mathbb{R}.$$

The root $1+i$ is not part of the right-hand side of (3) , which is the constant $-a$, therefore a particular solution can be estimated only by constant $b \in \mathbb{R}$, for more details see [5]. Substituting b into (3) and solving for b we obtain

$$b = \frac{a}{i}.$$

Hence the general solution of (3) can be written as the sum

$$D_j = C(1+i)^j + \frac{a}{i}, \quad C \in \mathbb{R}.$$

Constant C can be specified from the initial condition (2) and choosing $j = 0$ into above equation

$$C = D - \frac{a}{i}.$$

Thus the general solution of (3) takes form

$$D_j = \left(D - \frac{a}{i}\right)(1+i)^j + \frac{a}{i},$$

which implies the loan repayment formula (compare with [2], [4] and [1])

$$D_j = D(1+i)^j - \frac{a}{i}\left((1+i)^j - 1\right), \quad j = 1, 2, \ldots, n. \tag{4}$$

We assume that the loan is to be repaid in exactly n payments therefore we take into account the condition

$$D_n = 0.$$

Similarly to [4], [2], [1] and [9] the total amount of the debt D can be determined from (4) setting $j = n$

$$D(1+i)^n = \frac{a}{i}\left((1+i)^n - 1\right).$$

Thus we obtain the following formula for the debt

$$D = \frac{a}{i}\left(1 - \frac{1}{(1+i)^n}\right), \tag{5}$$

from which we can calculate the annuity a

$$a = D\frac{i}{1 - \frac{1}{(1+i)^n}}. \tag{6}$$

The formulas (5) and (6) are fundamental in calculating the loan repayments.

3 Amortization Schedule

3.1 Loan Balance

To calculate the outstanding debt after the j-th payment we substitute (5) into (4). In the same way as [11] we get

$$\begin{aligned} D_j &= \frac{a}{i}\left(1 - \frac{1}{(1+i)^n}\right)(1+i)^j + \frac{a}{i}\left(1 - (1+i)^j\right) = \\ &= \frac{a}{i}\left((1+i)^j - \frac{1}{(1+i)^n}(1+i)^j + 1 - (1+i)^j\right) = \\ &= \frac{a}{i}\left(1 - \frac{(1+i)^j}{(1+i)^n}\right). \end{aligned}$$

Hence we derived the formula

$$D_j = \frac{a}{i}\left(1 - \frac{1}{(1+i)^{n-j}}\right), \quad j = 1, 2, \ldots, n \tag{7}$$

which is applied in the amortization schedule to specify the loan balance.

The equality (7) is used in finance to determine the outstanding debt after the j-th payment if you know the loan amount D, the total number of all payments n and the annual interest rate r under the assumption that payments are paid k times a year, where $i = \frac{r}{k}$.

3.2 Interest

Now we derive difference equation for calculation of interest. Let U_j denote the amount of interest in the j-th period. Let us multiply the equation (1) by interest i

$$iD_{j+1} = i\,(1+i)\,D_j - ia.$$

Because interest is paid only from the remaining part of the debt each period, we have relation

$$U_{j+1} = iD_j, \qquad j = 0, 1, 2, \ldots n-1.$$

Combining both above equations and by renumbering we get the following recursion

$$U_{j+1} = (1+i)U_j - ia, \qquad j = 1, 2, \ldots, n-1$$

where the initial element is given by

$$U_1 = iD. \tag{8}$$

The reccurrence represents the first order nonhomogeneous linear difference equation with constant coefficients

$$U_{j+1} - (1+i)U_j = -ia, \qquad j = 1, 2, \ldots, n-1. \tag{9}$$

Its general solution is according to the superposition principle the sum of the general solution to the appropriate homogeneous equation

$$C(1+i)^j, \quad C \in \mathbb{R},$$

where $1+i$ is the root of the characteristic equation of (9), and an arbitrary particular solution to the nonhomogeneous equation which can be estimated by constant $b \in \mathbb{R}$. It is because the root $1+i$ does not appear on the right-hand side $-ia$ of the difference equation (9) which is a constant. More detailed information on solving similar difference equations can be found in [5]. We substitute b into (9) and obtain

$$b - (1+i)b = -ia.$$

Thus

$$b = a$$

and using superposition principle we can write general solution as follows

$$U_j = C(1+i)^j + a, \quad C \in \mathbb{R}. \tag{10}$$

From initial condition (8) and (5) we have

$$U_1 = a\left(1 - \frac{1}{(1+i)^n}\right). \tag{11}$$

The constant C can be now specified considering the equalities (11) and (10) with index $j = 1$

$$a\left(1 - \frac{1}{(1+i)^n}\right) = C(1+i) + a.$$

Hence

$$C = -a\frac{1}{(1+i)^{n+1}}.$$

Substituting the previous relation into (10) we get

$$U_j = -a\frac{1}{(1+i)^{n+1}}(1+i)^j + a,$$

which gives

$$U_j = a\left(1 - \frac{1}{(1+i)^{n-j+1}}\right), \qquad j = 1, 2, \ldots, n. \tag{12}$$

This formula is suitable for calculating the amount of interest in the j-th payment. Note that the amount of interest in the last installment is expressed by the following relation

$$U_n = a\left(1 - \frac{1}{1+i}\right). \tag{13}$$

3.3 Principal

The recursive rule for determination of principal used in finance (see [8])

$$M_{j+1} = (1+i)M_j, \qquad j = 1, 2, \ldots, n-1 \tag{14}$$

with initial element

$$M_1 = a - iD \tag{15}$$

can be derived from the fact that each payment consists of interest and principle

$$a = U_j + M_j, \qquad j = 1, 2, \ldots, n.$$

Substituting $U_j = a - M_j$ into (9) we have

$$a - M_{j+1} - (1+i)\,(a - M_j) = -ia$$

and by multiplying we obtain

$$a - M_{j+1} - a + M_j - ia + iM_j = -ia.$$

This equality implies (14) which is the first order homogeneous linear difference equation with constant coefficients

$$M_{j+1} - (1+i)M_j = 0, \qquad j = 1, 2, \ldots, n-1. \tag{16}$$

Its general solution has the form

$$M_j = C(1+i)^j, \quad \text{where} \quad C \in \mathbb{R} \tag{17}$$

because $1 + i$ is the root of the characteristic equation of (16). Further details concerning the solution of such linear difference equations you can find in [5]. From initial condition (15) and (5) we get

$$M_1 = a - a\left(1 - \frac{1}{(1+i)^n}\right),$$

hence

$$M_1 = a\,\frac{1}{(1+i)^n}. \tag{18}$$

From the above relation and (17) with $j = 1$ we can specify the constant C

$$a\,\frac{1}{(1+i)^n} = C(1+i),$$

thus

$$C = a\,\frac{1}{(1+i)^{n+1}}.$$

Therefore substituting this in (17) we have

$$M_j = a\,\frac{1}{(1+i)^{n-j+1}}, \qquad j = 1, 2, \ldots, n \tag{19}$$

which is the formula for calculation of principle amount in the j-th payment. Let us note that the principle amount in the final payment with index $j = n$ is equal to

$$M_n = a\,\frac{1}{1+i}. \tag{20}$$

4 Conclusions

All previously derived formulas (5), (7), (12), (19) and (11), (13), (18), (20) can be included in the following table called amortization schedule which is commonly used in finance. Note that the sequence of the amount of principal M_j is increasing, on the contrary the sequence of the amount of interest U_j is decreasing for $j = 1, 2, \ldots, n$. The outstanding debt is equal to zero at the end of loan repayment.

It should be emphasized that the above formulas were obtained by solving difference equations and not by the properties of sequence as is common in financial mathematics.

No. of installment j	Annuity	Interest U_j	Principal M_j	Loan Balance D_j
0	–	–	–	$\frac{a}{i}\left(1-\frac{1}{(1+i)^n}\right)$ (5)
1	a	$a\left(1-\frac{1}{(1+i)^n}\right)$ (11)	$a\frac{1}{(1+i)^n}$ (18)	$\frac{a}{i}\left(1-\frac{1}{(1+i)^{n-1}}\right)$
2	a	$a\left(1-\frac{1}{(1+i)^{n-1}}\right)$	$a\frac{1}{(1+i)^{n-1}}$	$\frac{a}{i}\left(1-\frac{1}{(1+i)^{n-2}}\right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	a	$a\left(1-\frac{1}{(1+i)^{n-j+1}}\right)$ (12)	$a\frac{1}{(1+i)^{n-j+1}}$ (19)	$\frac{a}{i}\left(1-\frac{1}{(1+i)^{n-j}}\right)$ (7)
$j+1$	a	$a\left(1-\frac{1}{(1+i)^{n-j}}\right)$	$a\frac{1}{(1+i)^{n-j}}$	$\frac{a}{i}\left(1-\frac{1}{(1+i)^{n-j-1}}\right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n-1$	a	$a\left(1-\frac{1}{(1+i)^2}\right)$	$a\frac{1}{(1+i)^2}$	$\frac{a}{i}\left(1-\frac{1}{1+i}\right)$
n	a	$a\left(1-\frac{1}{1+i}\right)$ (13)	$a\frac{1}{1+i}$ (20)	0

Table 1 Amortization schedule

References

[1] Banasiak, J. (2013). *Mathematical Modelling in One Dimension: An Introduction via Difference and Differential Equations*. Cambridge University Press.

[2] Elaydi, S. (2004). *An Introduction to Difference Equations*. Springer.

[3] Fulford, G., Forrester, P., Jones, A. (1997). *Modeling with Differential and Difference Equations*. Cambridge University Press.

[4] Goldberg, S. (2010). *Introduction to Difference Equations: With Illustrative Examples from Economics, Psychology, and Sociology*. Dover Publications Inc.

[5] Moučka, J., Rádl, P. (2015). *Matematika pro studenty ekonomie*. Praha: Grada Publishing Inc.

[6] Neusser, K. (2012). *Difference Equations for Economists*. Bern: University of Bern. Available at http://www.neusser.ch/downloads/DifferenceEquations.pdf.

[7] Pražák, P. (2013). *Diferenční rovnice s aplikacemi v ekonomii*. Hradec Králové: Gaudeamus.

[8] Radová, J., Dvořák, P., Málek, J. (2005). *Finanční matematika pro každého*. Praha: Grada Publishing Inc.

[9] Říhová, D., Viskotová, L. (2015). Loan Amortization and Linear Difference Equations. In *Proceedings of 18th Conference of Enterprise and Competitive Environment*, Brno: Mendel University, 783–790.

[10] Říhová, D., Viskotová, L. (2014). Some Applications of Linear Difference Equations in Finance with Wolfram|alpha and Maple. *Ratio Mathematica*, 27, 81–90.

[11] Říhová, D. (2019). Several Products of Financial Mathematics as Linear Difference Equations. In *Proceedings of Conference Aplimat 2019*, Bratislava: Slovenská technická univerzita, 992–1000.

[12] Šoba, O., Širůček, M., Ptáček, R. (2013). *Finanční matematika v praxi*. Praha: Grada Publishing Inc.

[13] Viskotová, L., Říhová, D. (2016). Annuity Due and Ordinary Annuity Using Linear Difference Equations and CAS. In *Proceedings of Conference Aplimat 2016*, Bratislava: Slovenská technická univerzita, 1093–1104.

[14] Investopedia. Available at http://www.investopedia.com.

Robust Optimization Approach in Chinese Postman Problem

Tereza Sedlářová Nehézová[1]

Abstract: Chinese postman problem is one of the most known arc routing problems with multiple uses in real application. It is similar to well-known Traveling salesman problem, but unlike it, the goal is to visit all edges, not nodes. This paper deals with using Chinese postman problem and it shows how to deal with a situation when travel times on given edges are not exactly known. In real life problems, it is very common that not every part of the model is precisely known. This can be caused by traffic complications or other uncertain circumstances and mathematical programming allows us to handle uncertainty in multiple ways. One of the frequently used methods is stochastic optimization, or there is robust optimization approach, that is used in this paper. The robust approach allows one to identify deviations of deterministic values, find robust-optimal solution and keeps the optimization model relatively simple. This paper describes robust optimization approach towards Chinese postman problem in detail with example in the end.

Keywords: Chinese postman problem, graph theory, Eulerian path, robust optimization, uncertainty

JEL Classification: C61
AMS Classification: 90C05, 90C08,90C11

1 Introduction

Chinese postman problem (CPP) [14, 10], generally defined as a route of minimal cost that visits each edge of given connected undirected graph at least once, and was introduced in paper of Kwan [14] with following problem: A postman has to deliver letters to a given neighborhood. He needs to walk through all the streets in the neighborhood and back to the post-office. How can he deign his route so that he walks the shortest distance? If Eulerian cycle exists in the given graph, it is the optimal solution. Otherwise the optimization problem is to find possible minimum number of edges to duplicate (with smallest total weights) in order to get a multigraph that includes Eulerian cycle [17]. Finding such an Eulerian graph is described by van Bevern *et al.* [7] and finding a tour if one exists in linear time was described by Berge [4]. Edmonds and Johnson [11] proved that CPP in its basic form is solvable in polynomial time. However, even small changes in constraints that seem innocuous can make CPP NP-Hard [7]. Nowadays, there is numerous modifications of CPP designed for specific purposes, for example with time varying variations [9] or time restrictions [1] or even special case of CPP called Rural Postman Problem, where visiting all edges is not required [15]. Bibliography of arc routing problems, including CPP and related studies since 2010 till 2017 was published by Mourão and Pinto [16]. From recent work can be mentioned paper of Corberán et al. [8], that introduces variant of CPP where the cost of traversing an edge depends not only on its length, but even on the weight of the vehicle used for transportation.

If generic optimization model, it is assumed, in most cases, that all parameters are exactly known and are equal to some nominal values. However, this approach does not consider the possible impact of uncertainty on the admissibility of the model. Therefore, because of the differences in data values and denominations, several model constraints may be disrupted and solutions obtained using nominal data may no longer be optimal or even feasible. This implies the need to construct models that are immune to data-side uncertainty. Soyster [18] was the first author who addressed the issue of robust optimization and also designed a linear programming optimization model that provides a solution that is resistant to deviations caused by the data uncertainty. Brief description of robust optimization, especially in linear programming was made by Ben-Tal and Nemirovski [3]. In the last 10 years, there has been a great growth in the application of robust optimization in various sectors [12].

[1] Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, nehezova@pef.czu.cz

So-called Γ-scenario set used in this paper and developed by Bertsimas and Sim [6] is one of the most adopted robust optimization approaches. The uncertainty is described by symmetric ranges of deviation from nominal values and control the number of deviating coefficients by parameter Γ. This approach keeps the model relatively simple, which is an important computational advantage, especially when solving large scale problems. Furthermore, the model still keeps the nature of linear programming.

This paper shows possibility of using CPP for finding shortest route in given graph while considering travel times to be uncertain. A way how to reformulate CPP into its robust counterpart assuming deviations in cost coefficients will be shown and this will be used for solving the given practical problem.

2 Materials and methods

2.1 Formulation of Chinese postman problem (CPP)

CPP in graph theory can be described as connected graph $G(V, E)$, where $V = \{V1, V2, \dots, Vn\}$ is a set of vertices (intersections), E is the set of edges between the vertices (streets), $E = \{e = (V_i, V_j)$; edge e between vertices V_i and $V_j\}$, each edge $e \in E$ with non-negative weight $C(e) = c(V_i, V_j)$ meaning the length of the edge (street). The problem is to determine a circle within G, which passes over each edge at least once and makes the minimum of total weight of the circle [14, 10, 13]. A strategy to find effective path of Chinese postman is based on observation that any desired tour induces an Eulerian supergraph of G by multiplying every edge of G according to the number of times it is traversed by the tour. Finding such an Eulerian graph is described, for example, by [7]. If the graph contains a vertex with odd degree, to find Eulerian path it must be edited in such way to get every vertex to even degree.

This problem can be formulated as linear integer program as shown by [13]:

$$\min \sum_{(i,j)\in E} c_{ij} x_{ij} \tag{2.1}$$

$$s.t.$$

$$\sum_{\substack{j \\ (i,j)\in E}} x_{ij} - \sum_{\substack{j \\ (j,i)\in}} x_{ji} = 0, i = 1,2, \dots, n \tag{2.2}$$

$$x_{ij} + x_{ji} \geq 1, \forall (i,j) \in E \tag{2.3}$$

$$x_{ij} \epsilon \{0; 1\}, \forall (i,j) \in E \tag{2.4}$$

where c_{ij} is nonzero weight of edge e_{ij} and $c_{ij} = c_{ji}$, x_{ij} is integer variable ($x_{ij} = 1$ meaning that path from V_i along the edge e_{ij} to V_j is realized), constraint (2.3) ensures that every edge is visited at least once, and equivalent number of "in" and "out" (even degree of vertices) arcs of every vertex in the model is secured by constraint (2.2).

2.2 Robust optimization approach in linear programming (LP)

In this paragraph it is shown how the generic linear optimization model can be transformed into its robust counterpart, which means to include uncertainty into it. Let's assume following model:

$$\max \sum_{j=1}^{n} c_j x_j \tag{2.5}$$

$$\text{s.t.}$$

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i, i = 1, \dots, m \tag{2.6}$$

$$x_j \epsilon \{0; 1\}, i = 1,2, \dots, m \tag{2.7}$$

The presence of uncertainty is usually considered as that some of the coefficients a_{ij}, b_j, c_j are not precisely defined (later we assume uncertainty in c_j). This leads to a new problem:

$$\max \sum_{j=1}^{n} (c_j + \delta_j^c) x_j \tag{2.8}$$

s.t.

$$\sum_{j=1}^{n} (a_{ij} + \delta_{ij}^a) x_j \leq b_i + \delta_i^b \,, i = 1, \dots, m \tag{2.9}$$

$$x_j \epsilon \{0; 1\}, j = 1,2, \dots, n \tag{2.10}$$

The (2.8) (2.9) and (2.10) is reformulation of (2.5) (2.6) and (2.7)where the uncertainty is expressed using deviations for any coefficient if need be. Deviations represented as δ are assumed to be any nonzero number. Any slight change in the original coefficient value may affect the optimal solution in adverse way and even can cause the infeasibility as illustrated by Ben-Tal, EL Ghaoui and Nemirovski [2]. Before a robust counterpart is constructed, there are 4 assumptions proposed by Bertsimas and Sim [5], that must be held:

For each coefficient c_j one is able to define expected (deterministic) value and maximum deviation δ_{ij}^c from its deterministic value.

1. Deterministic value c_j then belongs to the symmetric interval $\left[c_j - \delta_j^c, c_j + \delta_j^c\right]$.
2. The uncertain coefficients are stochastically independent random coefficients, each of its own deviation range.
3. For each constraint i it is possible to define maximum number of coeficients Γ that will simultaneously deviate from its deterministic value the constraint i.

We consider the following objective function with the deviations in c_j as in (2.8), which can be represented also as:

$$\min K \tag{2.11}$$

s.t.

$$\sum_{i=1}^{n} (c_j + \delta_j^c) x_j \leq K \tag{2.12}$$

Then the robust counterpart of (2.8) (2.9) (2.10) is constructed:

$$\min K \tag{2.13}$$

s.t.

$$\sum_{j=1}^{n} c_j \, x_j + \Gamma_i z_i + \sum_{j \in U_i} p_{ij} \leq K \,, i = 1, \dots, m \tag{2.14}$$

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i \,, i = 1, \dots, m \tag{2.15}$$

$$z_i + p_{ij} \geq \delta_j^c x_j, \forall j \in U_i \tag{2.16}$$

$$z_i \geq 0, i = 1, \dots, m \tag{2.17}$$

$$p_{ij} \geq 0, i = 1, \dots, m, \forall j \in U_i \tag{2.18}$$

$$x_j \epsilon \{0; 1\}, j = 1, \dots, n \tag{2.19}$$

where the parameter $\Gamma_i, 0 \leq \Gamma_i \leq |U_i|, i = 1, \dots, m$ controls the protection against uncertainty in the constraint i, value of Γ sets the maximum number of coefficients c_j that are assumed to deviate by a maximum of δ_j^c, p_{ij} is an auxiliary variable for each c_j that is considered uncertain, z_i is another auxiliary variable merely preserving a relationship between the first and second constraint and U_i indicates a set of indices j of those c_j which are actually considered to be uncertain.

3 Results and discussion

When both the integer model of CPP and robust counterpart of LP are defined, composing robust counterpart of CPP can be defined.

3.1 Robust optimization approach to Chinese postman problem

The uncertainty in CPP used in this paper will be laid on objective function coefficients c_{ij}, which in CPP represents nonzero weight of edge e_{ij}, as defined in (2.1) and if we assume it as time, it can perform uncertain behavior. The deviation of the expected value of c_{ij} has to be included to the objective function as δ_{ij}^{c}. We assume objective function including uncertainty in c_{ij}.

It is now possible to define the robust counterpart of CPP with uncertainty laid in cost coefficients:

$$\min K \quad \text{s.t.} \tag{3.1}$$

$$\sum_{(i,j)\in E} c_{ij}x_{ij} + \Gamma z + \sum_{(i,j)\in U} p_{ij} \leq K \tag{3.2}$$

$$\sum_{\substack{j \\ (i,j)\in E}} x_{ij} - \sum_{\substack{j \\ (j,i)\in E}} x_{ji} = 0, i = 1,2,\dots,n \tag{3.3}$$

$$p_{ij} + z_i \geq \delta_{ij}^{c} x_{ij}, \forall (i,j) \in U \tag{3.4}$$

$$x_{ij} + x_{ji} \geq 1, \forall (i,j) \in E \tag{3.5}$$

$$z \geq 0 \tag{3.6}$$

$$p_{ij} \geq 0, \forall (i,j) \in U \tag{3.7}$$

$$x_{ij} \epsilon \{0;1\}, \forall (i,j) \in E \tag{3.8}$$

where δ_{ij}^{c} is maximal deviation of c_{ij}, parameter $\Gamma, 0 \leq \Gamma \leq |U|$ allows one to control the uncertainty for of those $c_{ij}, (i,j) \in U$ of the CPP, parameter value Γ indicates the maximum number of c_{ij} which are expected to deviate by no more than δ_{ij}^{c}, p_{ij} and z has the same meaning as in (2.14) and (2.16).

3.2 Practical example

Given problem is to find the shortest route in given graph while using every edge at least once. For calculation purposes is given connected undirected graph $G(V,E), |V| = 44, |E| = 61$ (122 if all used both ways), which is inspired by paths in Prague zoo with aim to see every exhibit – use every edge. The resulting graph looks like the following:

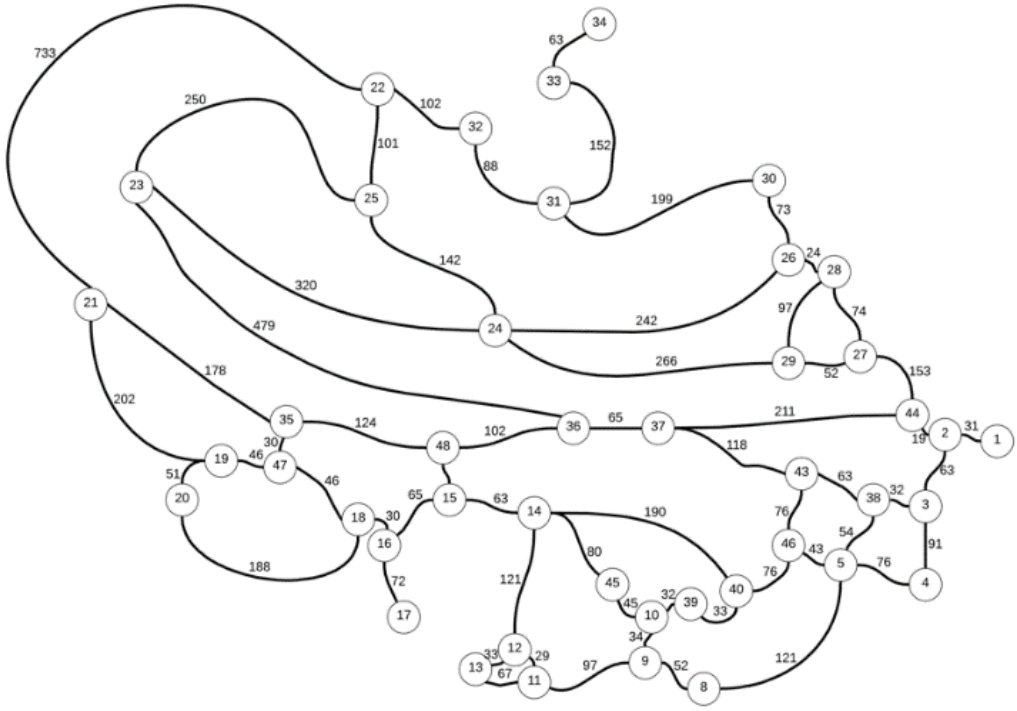


Figure 1 Given graph with deterministic durations of travel on each edge

When assuming deterministic case (chapter 2.1) of given problem and application of OpenSolver for excel with integrated Gurobi optimizer, optimal solution is 8619 m long, or 103,428 minutes at walking speed 5 km/h (In further calculations, only time spent on the edge will be considered), used edges can be seen in table 3.

If any change of any cost coefficient appear – Γ of unspecified coefficients change from its deterministic value (chapter 3), the shortest path is very likely going to change. The uncertainty of edge value is considered as maximum amount of time that one will spend on certain edge on top of its deterministic value. If assume scenario with $\Gamma = 25$, that means at most 25 coefficients c_{ij} will deviate from its deterministic value at max by δ_{ij}^{c}, which is, in this case, considered 10–30 %. Comparison of tested scenarios with different values of Γ can be seen in Table 1.

Γ value	**Objective value**	**Minutes in total**	**Difference from $\Gamma = 0$ (min)**
0	1,717	103,032	0
25	1,948	116,866	13,834
50	2,020	121,223	18,191
75	2,052	123,119	20,087
100	2,052	123,119	20,087
122	2,121	127,26	24,228

Table 1 Different robust solutions

Deviation of time travel values on edges is going to affect not only objective function, but structure of optimal solution too. Values of x_{ij} for different Γ are shown in table 3. As can be seen, cases with included uncertainty are different from the deterministic one ($\Gamma = 0$).

	0	**25**	**50**	**75**	**100**	**122**		**0**	**25**	**50**	**75**	**100**	**122**		**0**	**25**	**50**	**75**	**100**	**122**		**0**	**25**	**50**	**75**	**100**	**122**
x1_2	1	1	1	1	1	1	**x14_40**	1						**x25_23**		1	1	1	1	1	**x37_43**	1					1
x2_1	1	1	1	1	1	1	**x14_45**		1	1	1	1	1	**x25_24**	1						**x37_44**		1	1	1	1	1
x2_3	1	1	1	1	1	1	**x15_14**		1	1	1	1		**x26_24**	1					1	**x38_3**	1			1	1	1

	0	25	50	75	100	122		0	25	50	75	100	122		0	25	50	75	100	122		0	25	50	75	100	122
x2_44	1	1	1	1	1		**x15_16**	1	1	1	1	1	1	**x26_28**	1	1	1	1	1	1	**x38_5**		1	1			1
x3_2	1	1	1	1	1		**x15_48**	1					1	**x26_30**		1	1	1	1		**x38_43**	1	1	1	1	1	
x3_4	1			1	1	1	**x16_15**	1	1	1	1	1	1	**x27_28**		1	1	1	1		**x39_10**	1					
x3_38		1	1			1	**x16_17**	1	1	1	1	1	1	**x27_29**	1	1	1	1	1	1	**x39_40**		1	1	1	1	1
x4_3		1	1				**x16_18**	1	1	1	1	1	1	**x27_44**	1					1	**x40_14**		1	1	1	1	1
x4_5	1			1	1	1	**x17_16**	1	1	1	1	1	1	**x28_26**	1	1	1	1	1		**x40_39**	1					
x5_4		1	1				**x18_16**	1	1	1	1	1	1	**x28_27**	1					1	**x40_46**	1	1	1	1	1	1
x5_8	1					1	**x18_20**	1						**x28_29**		1	1	1	1	1	**x43_37**		1	1	1	1	1
x5_46		1	1	1	1	1	**x18_47**		1	1	1	1	1	**x29_24**		1	1	1	1	1	**x43_38**	1	1	1	1	1	1
x5_38	1			1	1		**x19_20**		1	1	1	1	1	**x29_27**	1	1	1	1	1		**x43_46**	1					
x8_5		1	1	1	1		**x19_21**	1					1	**x29_28**	1					1	**x44_2**	1	1	1	1	1	1
x8_9	1					1	**x19_47**	1	1	1	1	1		**x30_26**	1					1	**x44_27**		1	1	1	1	1
x9_10	1	1	1	1	1	1	**x20_18**		1	1	1	1	1	**x30_31**		1	1	1	1		**x44_37**	1					
x9_8		1	1	1	1		**x20_19**	1						**x31_30**	1					1	**x45_10**		1	1	1	1	1
x10_9	1	1	1	1	1	1	**x21_19**		1	1	1	1	1	**x31_32**		1	1	1	1		**x45_14**	1					
x9_11	1					1	**x21_35**	1	1	1	1	1		**x31_33**	1	1	1	1	1	1	**x46_5**	1					
x10_39		1	1	1	1	1	**x21_22**	1					1	**x32_22**		1	1	1	1		**x46_40**	1	1	1	1	1	1
x10_45	1						**x22_25**	1	1	1	1	1	1	**x32_31**	1					1	**x46_43**		1	1	1	1	1
x11_9		1	1	1	1		**x22_32**	1					1	**x33_31**	1	1	1	1	1	1	**x47_18**	1					
x11_12	1	1	1	1	1	1	**x22_21**		1	1	1	1		**x33_34**	1	1	1	1	1	1	**x47_19**	1	1	1	1	1	1
x11_13	1					1	**x23_24**		1	1	1	1	1	**x34_33**	1	1	1	1	1	1	**x47_35**		1	1	1	1	1
x12_11	1	1	1	1	1	1	**x23_25**	1						**x35_21**	1	1	1	1	1	1	**x48_15**		1	1	1	1	
x12_13		1	1	1	1		**x23_36**	1	1	1	1	1	1	**x35_47**	1					1	**x48_35**	1					1
x12_14	1					1	**x24_25**		1	1	1	1	1	**x35_48**		1	1	1	1		**x48_36**	1	1	1	1	1	1
x13_11		1	1	1	1		**x24_23**	1					1	**x36_23**	1	1	1	1	1								
x13_12	1					1	**x24_26**		1	1	1	1	1	**x36_37**	1	1	1	1	1	1							
x14_12		1	1	1	1		**x24_29**	1						**x36_48**	1	1	1	1	1	1							
x14_15	1					1	**x25_22**	1	1	1	1	1	1	**x37_36**	1	1	1	1	1								

Table 2 Solutions for different tested values of Γ

4 Conclusion

First, the formulation of CPP as linear integer problem under terms of uncertain travel time on edges using Γ-robustness approach was proposed and it was shown, that uncertainty in tested data have an impact to optimal solution, especially on order of edges. Both deterministic and robust case of CPP were found and compared and different scenarios showed deterioration of objective function up to 24%. Also, this approach is less demanding than comparison and evaluation of all possible cases with different values and presented robust optimization model still retains the nature of linear programming and thus is easy computationally traceable, which is considered as great advantage in case of already complex models like arc routing problems. Furthermore, there can be focus in improving given model to reflect conditions and requirements more realistically, there is a possibility to include constraints limiting total travel time, creating ways in graphs without having to visit every edge and include prioritization of certain edges over others.

Acknowledgements

This paper was supported by Internal Grant Agency (IGA) of FEM CZU, project No. 2020A0006.

References

[1] AMINU, U. F. & EGLESE, R. W. (2006). A constraint programming approach to the Chinese postman problem with time windows. *Computers & Operations Research*, 33(12), pp. 3423–3431. doi: 10.1016/j.cor.2005.02.012.

[2] BEN-TAL, A., EL GHAOUI, L., & NEMIROVSKI, A. (2009). *Robust Optimization*. Princeton University Press.
[3] BEN-TAL, A. & NEMIROVSKI, A. (1999). Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1), pp. 1–13. doi: 10.1016/S0167-6377(99)00016-4.
[4] BERGE, C. (1989). *Graphs*. North Holland.
[5] BERTSIMAS, D. & SIM, M. (2003). Robust discrete optimization and network flows. *Mathematical Programming*, 98(1–3), pp. 49–71. doi: 10.1007/s10107-003-0396-4.
[6] BERTSIMAS, D. & SIM, M. (2004). The price of robustness. *Operations Research*, 52(1), pp. 35–53. doi: 10.1287/opre.1030.0065.
[7] VAN BEVERN, R. ET AL. (2015). Chapter 2: The Complexity of Arc Routing Problems. In *Arc Routing*. Philadelphia, PA: Society for Industrial and Applied Mathematics, pp. 19–52. doi: 10.1137/1.9781611973679.ch2.
[8] CORBERÁN, A. ET AL. (2018). The Chinese postman problem with load-dependent costs. *Transportation Science*, 52(2), pp. 370–385. doi: 10.1287/trsc.2017.0774.
[9] DROR, M. (2000). *Arc Routing*. Edited by M. DROR. Boston, MA: Springer US. doi: 10.1007/978-1-4615-4495-1.
[10] EDMONDS, J. (1965). The Chinese Postman Problem. *Operations Research*, 1, pp. 373–b77.
[11] EDMONDS, J. & JOHNSON, E. L. (1973). Matching, Euler tours and the Chinese postman. *Mathematical Programming*, 5(1), pp. 88–124. doi: 10.1007/BF01580113.
[12] GABREL, V., MURAT, C., & THIELE, A. (2014). Recent advances in robust optimization: An overview. *European Journal of Operational Research*, 235(3), pp. 471–483. doi: 10.1016/j.ejor.2013.09.036.
[13] JUNWEN, F. (2014). Integer programming modeling for the Chinese Postman Problem. *Prestige International Journal of Management & IT-Sanchayan*, 3(2), p. 42.
[14] KWAN, M. (1960). Programming method using odd or even pints. *Acta Mathematica Sinica*, 10, pp. 263–266.
[15] LENSTRA, J. K. & KAN, A. H. G. R. (1981). Complexity of vehicle routing and scheduling problems. *Networks*, 11(2), pp. 221–227. doi: 10.1002/net.3230110211.
[16] MOURÃO, M. C. & PINTO, L. S. (2017). An updated annotated bibliography on arc routing problems. *Networks*, 70(3), pp. 144–194. doi: 10.1002/net.21762.
[17] ROBERTS, F. & TESMAN, B. (2009). *Applied Combinatorics (Second Edition)*. Chapman and Hall.
[18] SOYSTER, A. L. (1973). Convex Programming with Set-Inclusive Constraints and Applications to Inexac Linear Programming. *Operations Research*, 21(5), pp. 1154–1157. doi: 10.1287/opre.21.5.1154.

Sensitivity Analysis of a System Dynamics Decontamination Model

Anna Selivanova[1]

Abstract. This work is dedicated to tests of a simulation model of selected decontamination scenarios. The mathematical model was created using the System Dynamics methods. The first version of the model was designed within the Project of the Ministry of the Interior of the Czech Republic, VH172020015 – Recovery Management Strategy for Affected Areas after Radiation Emergency. The model allowed to prepare a cost-benefit analysis of chosen decontamination methods for a grassed recreation meadow.

In order to broaden the possibilities of the model, several improvements were added (e.g. corrections in dose conversion coefficients or an option to change personal protective equipment). Thereafter, selected scenarios were simulated with varied surface activities, decontamination factors and waste production parameters. Results of the carried out analysis were applicable to estimate e.g. lower thresholds of a surface contamination to implement grass removal, soil stripping and turf harvesting scenarios. Moreover, decontamination costs related to one square meter of the meadow were strongly affected by a volume of the produced waste.

Keywords: cost-benefit analysis, system dynamics, computer simulations, sensitivity analysis, nuclear accident, decontamination

JEL Classification: C63, Q51
AMS Classification: 90B99

1 Introduction

After the Chernobyl disaster and the Fukushima Daiichi accident, remediation and decontamination strategies were broadly developed within many international collaborations. Practical clean-up experience and created decontamination procedures were summarized e.g. in [9] and [16]. Obtained knowledge about remediation strategies have been implemented in the model ERMIN (European model for inhabited areas) in the JRodos software [10]. However, input decontamination parameters and costs could vary depending on site conditions [15].

Owing to many parameters and variables engaged in remediation strategies, their interrelations and a time-dependent behavior of some variables, decontamination problems could be also solved using the System dynamics methodology. For instance, considering contamination with radionuclides, nonlinearity of their radioactive decays is a typical characteristic of dynamic complex models [12]. Therefore, the System dynamics is a suitable tool to understand and solve these problems using computer simulations. Moreover, decontamination scenarios tailored exactly for specific cases could be whenever changed and applied for other problems considering actual requirements (e.g. costs of labor or equipment, team sizes or working speeds). Therefore, the System dynamics approach would simplify a decision-making process and help to choose the final site-specific remediation strategy.

This paper is focused on the creation of three decontamination scenarios of a large-scale recreation meadow located in Prague, Czech Republic. The meadow did not require any remediation actions and was selected for demonstration purposes only. Hence, the proposed scenarios could be modified by changing of chosen parameters for another object (e.g. playground), potentially or indeed contaminated. Afterwards, these scenarios could be used as resource materials for decision makers in radiation/nuclear post-emergency phases.

The mathematical model was designed on the basis of knowledge from Chernobyl and Fukushima [9], [16]. Using the system dynamics methods, the model connected both time-dependent physical processes and

[1]Faculty of Economics and Management, Czech University of Life Sciences Prague, Kamýcká 129, Prague, 165 00, Czech Republic, selivanova@pef.czu.cz;
National Radiation Protection Institute (SÚRO), Bartoškova 1450/28, Prague, 140 00, Czech Republic.

economical calculations. Relevant parameters were adopted from published/available on-line sources and corrected to 2019.

2 Material and methods

2.1 System Dynamics approach

The system dynamics approach allows to simulate complex dynamic systems and to integrate many variables or parameters in one assembly. Variables interconnected among themselves create specific patterns of non-linear behavior, "archetypes" described using "feedback loops" [7]. For instance, a radioactive decay is a dynamic exponential decay which corresponds to a negative (self-correcting) feedback loop. An opposite process, e.g. a population exponential growth, could be represented as a positive (self-reinforcing) feedback loop [12].

Due to work in the Vensim software [17], results presented in this paper are depicted as stock and flow diagrams. An example of a stock variable (time dependent) is a surface activity of radionuclides deposited on the soil surface. Inflows correspond to the activity increase (e.g. as a result of the radionuclides release), outflows describe its decrease (e.g. due to physical processes). Therefore, stock variables could be expressed using definite integrals:

$$s = \int_{T_0}^{T} (i - o)\mathrm{d}t + s_{T_0}, \tag{1}$$

where s is a stock variable, i represents all inflows, o corresponds to all outflows, T_0 is an initial time, T is a chosen time interval. Graphically stock variables could be depicted as boxes. Both inflows and outflows could be shown as pipes connected with boxes, adding to stocks, resp. subtracting from them [12].

2.2 Model description

The first version of the model was created within the Project of the Ministry of the Interior of the Czech Republic, VH172020015 – Recovery Management Strategy for Affected Areas after Radiation Emergency. The detailed description could be found in [8], [11]. Contrary to the previous model, corrections in the dosimetry calculations were implemented, as well as improvements in existing scenarios (including new countermeasures).

In order to simplify the model, only two radionuclides were considered, ^{134}Cs and ^{137}Cs, with longer half-lives. Initial surface activities of ^{134}Cs and ^{137}Cs were equal to 1 MBq m^{-2} (altogether 2 MBq m^{-2}). Within the modelling, the meadow area was roughly 60 000 m^2 [4]. A group of the most irradiated inhabitants consisted of 631 persons was assessed on the basis of data by the Czech Statistical Office [2, 3].

2.3 Physical processes

In order to estimate benefits, a calculation of doses obtained by inhabitants was required. Activities of both cesium isotopes were converted to annual effective doses using the modified formula [1]:

$$E = CC \cdot \frac{1}{\lambda_r + \lambda_w + \lambda_d} \cdot A_0 \cdot \left(1 - e^{-(\lambda_r + \lambda_w + \lambda_d) \cdot t_1}\right) \cdot (\Delta_{outdoor} + \Delta_{indoor} \cdot SF), \tag{2}$$

where CC is a dose rate conversion coefficient for ground contamination, $\Delta_{outdoor}$ is a correction for time spent outdoor, Δ_{indoor} is a correction for time spent indoor, SF is a shielding factor, λ_r is a decay constant, λ_w is a natural dispersion rate, λ_d is an estimated decontamination rate.

Effective external gamma dose rate conversion coefficients CC were adopted from [9]. Values of coefficients were set equal to $1.4 \cdot 10^{-12}$ (Sv h^{-1})/(Bq m^{-2}), or $3.6 \cdot 10^{-12}$ (Sv h^{-1})/(Bq m^{-2}) for ^{137}Cs, resp. ^{134}Cs. Owing to a major type of buildings around the meadow, the shielding factor SF was set 0.2. The natural dispersion rate λ_w was set to 0.05 year^{-1} [6]. The decontamination rate λ_d was estimated for both cesium isotopes using values of decontamination factors, the meadow area and working speeds adopted mainly from [9]. According to a recommendation [13], a value of 7 000 hours per year was expected for the time spent indoor and used for the correction Δ_{indoor} equal to 0.8. A total effective dose in each scenario was a sum of annual effective doses from ^{137}Cs and ^{134}Cs.

2.4 Remediation scenarios

Three decontamination scenarios were implemented, as well as one reference scenario (without any remediation). All created scenarios considered the meadow demarcation using fences, warning tapes and boards. In addition, decontamination of all workers and vehicles using water was included in all scenarios. Countermeasure scenarios anticipated grass removal carried out by two workers. Both soil stripping and turf harvesting itself assumed two agricultural machinery operators. However, the soil stripping countermeasure considered six assistants providing manual collection of waste residues (grass and soil), while the turf harvesting scenario anticipated only one assistant [9]. The soil stripping and the turf harvesting scenarios also included a water tie-down to limit a resuspension during the surface removal [9]. Moreover, all decontamination scenarios considered a waste handling [9], [16]. All scenarios were simulated using additional variables – scenario switches.

2.5 Total costs estimation

Total costs were a sum of labor costs, costs of equipment, costs of consumer materials and consumption of fixed capital. Labor costs included two wage rates: 500 CZK h^{-1} for vehicles operators and 400 CZK h^{-1} for assistants. Wage rates were set on the basis of data by ČEZ, a. s. [5] and corrected with regard to inflation and workers' qualifications. Moreover, according to [14], wages of workers carrying out the Fukushima clean-up were 12 000 Yen per day. Considering the inflation correction and an 8-hour working day, in 2019 workers would have the wage rate equal to approximately 380 CZK h^{-1}, which roughly corresponds with the model. Detailed costs of personal protective equipment (tyveks, gloves etc.), demarcation equipment (fences, warning tapes and boards) and auxiliary equipment (shovels, brooms, garden carts and waste bags) could be found in [11], as well as waste transportation costs and consumption of fixed capital (tractors, sod harvesters etc.).

2.6 Radiation impact and benefits assessment

Radiation impact of each scenario was assessed as a cost of a health detriment of all inhabitants. The health detriment cost was a product of a collective effective dose (the annual dose multiplied by the number of inhabitants) and a financial coefficient equal to 2.5 mln. CZK/Sv [13]. Afterwards, benefits of all scenarios were estimated as a difference between the health detriment costs of the reference scenario and the relevant countermeasure scenario.

2.7 Sensitivity analysis

Within the sensitivity analysis, three parameters were tested: decontamination factors (DF), surface activities and waste production. DF could be defined as a ratio of an initial surface activity and a surface activity after decontamination. According to [9] or [16], DF could vary depending e.g. on implementation time after deposition. Hence, required higher DF could be observed soon after deposition.

In order to estimate threshold activities for implementation of selected countermeasures, a total initial activity range of 200 kBq m^{-2} – 2 MBq m^{-2} was tested. In case of the grass removal, three values of the DF were assumed: 1.05, 5.5 and 10. For the soil stripping countermeasure, DF was equal to 5, 17.5 and 30, while for the turf stripping scenario these values were 1.8, 5.9 and 10. In case of the sensitivity analysis for the soil stripping and the turf harvesting countermeasures, the DF of the grass removal was conservatively equal to 1.05. Default waste production parameters were set conservatively: $2 \cdot 10^{-2}$ m^3 m^{-2} for grass removal, $5 \cdot 10^{-12}$ m^3 m^{-2} for soil stripping and turf harvesting. Then, the sensitivity analysis for varied waste production was carried out. Three values of waste production parameters were tested. Within the test, neither the default total surface activity nor DF were varied. The total surface activity was constant (2 MBq m^{-2}), as well as DF: 1.05 for the grass removal, 1.8 for the turf harvesting and 5 for the soil stripping scenario. For the grass removal, the waste production was $1 \cdot 10^{-4}$ – $2 \cdot 10^{-2}$ m^3 m^{-2}. In case of the turf harvesting countermeasure, values of the waste production were $2 \cdot 10^{-5}$ – $5 \cdot 10^{-2}$ m^3 m^{-2}. The waste production parameter for the soil stripping scenario was equal to $2 \cdot 10^{-2}$ – $7 \cdot 10^{-2}$ m^3 m^{-2}.

3 Results and discussion

3.1 Mathematical model

The final model consisted of 18 working layers dedicated to dosimetry calculations, economical assessments and a summary page with output results (e.g. annual effective doses or remediation costs). Figure 1

represents the dose assessment layer using stock and flow diagrams. Activities of ^{134}Cs and ^{137}Cs are presented as stock variables (boxes) and decrease over time as a result of three processes: radioactive decay, natural dispersion (weathering) and countermeasures. The radioactive decay process and weathering are depicted as one outflow (to the right) due to their natural essence, whilst decontamination actions are shown separately, as the second outflow (to the left). Scenario switches are added to the second outflow description. These variables are depicted as simple titles. Variables located on other layers of the model are shown as titles in angle brackets. Owing to many variables and parameters, polarities/feedback loops have not been sketched. Changes in surface activities of ^{137}Cs and ^{134}Cs for all scenarios with default set-ups are shown in Figure 2 (limited to 50 days due to better readability).

3.2 Demarcation

The demarcation duration was 1 day of work. Depending on the total surface activity (200 kBq m^{-2} – 2 MBq m^{-2}), annual effective doses were in a range of 1–11 mSv. Demarcation costs were 0.4 mln. CZK (7 CZK m^{-2}). Considering this scenario as the reference scenario, benefits were mathematically equal to zero.

3.3 Grass removal

The grass removal scenario took 4 days. Costs of the grass removal countermeasure were roughly 2 mln. CZK, or 35 CZK m^{-2}. For the DF of 1.05 (e.g. for late implementation time after deposition), annual effective doses after the grass removal were almost the same, as for the reference scenario, with the dose reduction of 5%. In case of the lowest DF, benefits of the scenario were significantly lower than its costs for all considered surface activities.

Assuming higher DF of 5.5 and 10, all annual effective doses were below 1 mSv (dose reductions, DR > 80%) corresponding to the dose limit for the public. Then, benefits of grass removal were higher than costs for the surface activity ≥ 400 kBq m^{-2}. Hence, the grass removal effectiveness significantly depended on the implementation time after surface contamination and corresponded with data in [9].

Varying the waste production parameter, while the total surface activity was 2 MBq m^{-2} and the DF was 1.05, benefits of the grass removal countermeasure were higher than its costs in case of the lowest waste production ($1 \cdot 10^{-4}$ m^3 m^{-2}), although the DR was again ≤ 5%. Considering these parameters, the total cost was roughly 0.5 mln. CZK, or 8 CZK m^{-2}, which was very similar to costs in the reference scenario (paragraph 3.2). For the waste production parameter equal to $1 \cdot 10^{-2}$ m^3 m^{-2}, implementation costs were 1.3 mln. CZK (22 CZK m^{-2}). In case of the value of $2 \cdot 10^{-2}$ m^3 m^{-2}, the results corresponded to the model default set-up and costs of 2 mln. CZK (35 CZK m^{-2}). Both costs were higher than benefits. Therefore, costs in the grass removal scenario were very sensitive to waste handling.

3.4 Soil stripping

The soil stripping countermeasure duration was 41 days. However, the duration could be shortened using more units of the agricultural machinery. Costs of the soil stripping scenario were 6.7 mln. CZK, or 116 CZK m^{-2}. Soil stripping costs in the Fukushima were in an interval of 290–710 Yen m^{-2} for areas > 1 000 m^2 (depending on techniques and equipment) [16]. This interval could be equal to costs in a range of 69–170 CZK m^{-2} being in a good agreement with the simulation results of soil stripping presented in the current paper.

Considering all DF values for the soil stripping scenario and benefits higher than implementation costs, the lower threshold of the total surface activity was 1 MBq m^{-2}. For the DF of 5, annual effective doses were ≤ 2 mSv (DR ≈ 80%). In case of higher DF (17.5 and 30), effective doses were < 1 mSv per year (DR ≥ 94%). Therefore, owing to high values of DF, the soil stripping scenario seemed to be applicable any time after the surface contamination. However, the scenario would be required for higher surface activities.

Assuming the lowest DF equal to 5 and the surface activity of 2 MBq m^{-2}, changing of the waste production ($2 \cdot 10^{-2}$ – $7 \cdot 10^{-2}$m^3 m^{-2}) did not affect an applicability of the soil stripping countermeasure, since the benefits were higher than the implementation costs. Nevertheless, total costs were sensitive to the waste production and were in a range of 4.6–8.2 mln. CZK, or from 79–140 CZK m^{-2}. According to [16], the produced waste volume considerably affected remediation costs. Hence, the revealed sensitivity to the waste production (resulting in a wide range of waste volumes and final costs) agreed with the published data.

3.5 Turf harvesting

The turf harvesting scenario took 17 days. As well as for the soil stripping, the working speed could be increased with a use of more operators and machinery. Costs of implementation were 6.1 mln. CZK, or 105 CZK m^{-2}. In accordance with the report [16], the Fukushima implementation costs of turf removal in areas > 1 000 m^2 were equal to 470 Yen m^{-2} (≈ 113 CZK m^{-2}) for parks or playgrounds and 1 500 Yen m^{-2} for gardens (≈ 360 CZK m^{-2}). The simulated scenario results and data from the Fukushima clean-up were of the same order of magnitude, although costs from simulations were commonly cheaper than the costs of the clean-up in Japan.

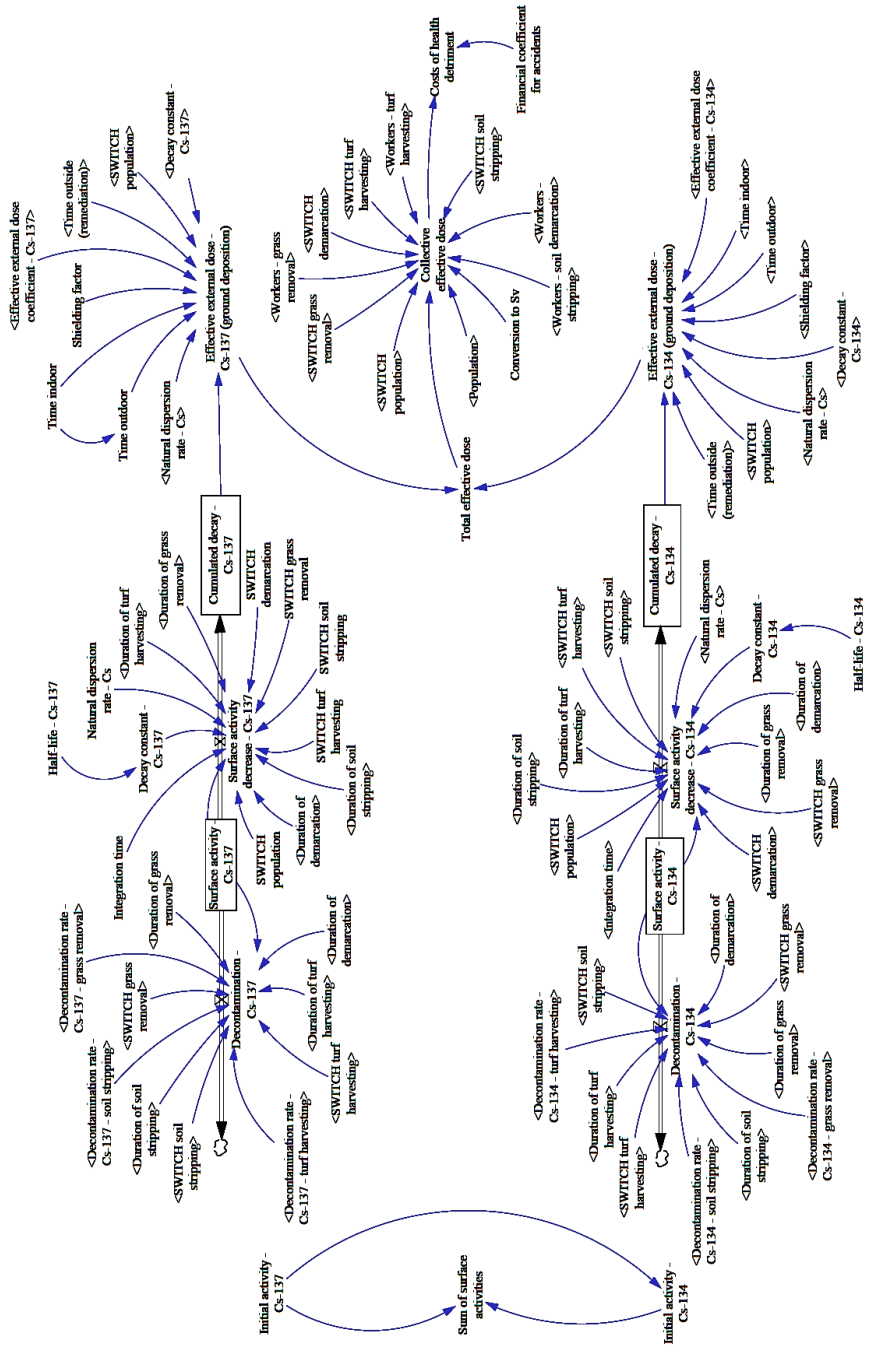


Figure 1 Stock and Flow diagram of Dose assessment layer

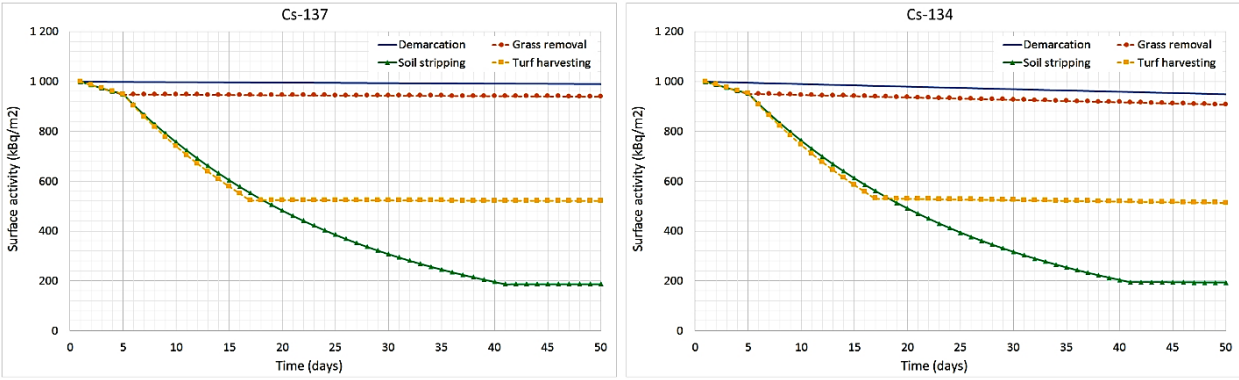


Figure 2 Surface activity decrease

For the DF of 1.8 (DR ≈ 55%), annual effective doses were < 6 mSv and the surface activity implementation threshold was 1.6 MBq m^{-2}. In case of the DF equal to 5.9 (DR ≈ 83%), effective doses did not exceed 2 mSv per year. For this DF value, the implementation threshold of surface activities was 1 MBq m^{-2}. Anticipating the DF of 10 (DR ≈ 90%), the most irradiated inhabitant would obtain less than 1 mSv per year and the threshold surface activities were at least 800 kBq m^{-2}. Therefore, different DF values affected possibilities of the turf harvesting implementation (activity thresholds) and this countermeasure would probably require more detailed consideration in relation to site conditions.

As well as in the soil stripping scenario and the conservative set-up of the model, varied values of the waste production parameter ($2 \cdot 10^{-2}$ – $5 \cdot 10^{-2}$ m^3 m^{-2}) did not have an impact on the scenario applicability (benefits were greater than costs), but substantially affected remediation costs: 4.0–6.1 mln. CZK, or 69–105 CZK m^{-2}.

4 Conclusion

Using the system dynamics methods, the mathematical model of the meadow decontamination was created. The model could be also applied for other contaminated areas with various values of parameters. Four countermeasure scenarios were simulated in the Vensim software. According to the results, implementation costs were in a good agreement with the Fukushima clean-up experience. Within the model tests for the chosen object, surface activity thresholds of scenarios implementation were estimated. Moreover, total costs of remediation were significantly affected by volumes of the produced waste.

References

[1] AHN, J., CARSON, C., JENSEN, M., & JURAKU, K. (2014). Reflections on the Fukushima Daiichi Nuclear Accident: toward social-scientific literacy and engineering resilience. Springer Berlin Heidelberg.

[2] CZSO. (2001). Úroveň bydlení [Characteristics of housing]. Czech Statistical Office. https://www.czso.cz/csu/czso/13-2130-03--6_5_uroven_bydleni [1 February 2020].

[3] CZSO. (2012). Základní výsledky sčítání lidu, domů a bytů – Česká republika – 2011 Basic results of Census – Czech Republic – 2011: Tab. 15 Houses by inhabitants, type and persons in dwellings]. Czech Statistical Office. https://www.czso.cz/documents/10180/20534544/zvcr015.pdf/a591687b-7a78-4233-ba12-e37739f0d7dc?version=1.0 [1 February 2020].

[4] EAGRI. (2019). Veřejný registr půdy – LPIS [Public land register – LPIS (Land Parcel Identification System)]. Portal eAGRI – resortní portál Ministerstva zemědělství. http://eagri.cz/public/app/lpisext/lpis/verejny2/plpis/ [5 May 2020].

[5] FOLTÝN, P. (2008). Optimalizace radiační ochrany s použitím metody "cost-benefit". Výpočet pro pracoviště ETE ČEZ a.s. University of South Bohemia České Budějovice.

[6] IAEA. (2001). Generic models for use in assessing the impact of discharges of radioactive substances to the environment (Safety Reports Standards No. 19). Vienna, International Atomic Energy Agency.

[7] MILDEOVÁ, S. & VOJTKO, V. (2008). Systémová dynamika [System Dynamics] (Vyd. 2.). Prague, Oeconomica.

[8] NERIS. (2019). 5th NERIS Workshop, 3–5 April 2019, Roskilde, Denmark. https://www.eu-neris.net/activities/workshops/roskilde-2019.html [5 May 2020].

[9] NISBET, A., BROWN, J., CABIANCA, T., JONES, A. L., ANDERSSON, K. G., HÄNNINEN, R., IKÄHEIMONEN, T., KIRCHNER, G., BERTSCH, V., & HIETE, M. (2010). Generic handbook for assisting in the management of contaminated inhabited areas in Europe following a radiological emergency (Version 2).
[10] RASKOB, W., TRYBUSHNYI, D., IEVDIN, I., & ZHELEZNYAK, M. (2011). JRODOS: Platform for improved long term countermeasures modelling and management. Radioprotection, 46(6), 731–736.
[11] SELIVANOVA, A. (2019). Cost-benefit analysis of decontamination in late-phase recovery from a nuclear accident. Czech University of Life Sciences Prague.
[12] STERMAN, J. D. (2000). Business dynamics: systems thinking and modeling for a complex world. Irwin/McGraw-Hill.
[13] SÚJB. (2016). Implementing decree 422/2016 Coll., on Radiation Protection and Security of a Radioactive Source. Prague, State Office for Nuclear Safety.
[14] THE GUARDIAN. (2011). Fukushima cleanup recruits "nuclear gypsies" from across Japan. 13 July 2011. https://www.theguardian.com/environment/2011/jul/13/fukushima-nuclear-gypsies-engineers-labourers [3 May 2020].
[15] TOTOKI, Y. (2014). Use of simulation capabilities of the ERMIN by Japanese after Fukushima. NERIS-TP Dissemination Workshop – 22–24 January 2014, Oslo (Norway).
[16] U.S. EPA. (2016). Current and Emerging Post-Fukushima Technologies, and Techniques, and Practices for Wide Area Radiological Survey, Remediation, and Waste Management. Office of Research and Development, Homeland Security Research Center.
[17] VENTANA SYSTEMS. (2015). Vensim. Vensim software.

Management of Imprecision Risk for Two-Assets Portfolio

Joanna Siwek[1], Krzysztof Piasecki[2]

Abstract: We consider a portfolio of financial assets that are evaluated by a fuzzy expected discount factor (EDF), given as a trapezoidal fuzzy number. Applied fuzziness implies imprecision risk that burdens the decision-making. By imprecision we understand the composition of indistinctness and ambiguity. In the following paper, the ambiguity of fuzzy EDF is evaluated by energy measure introduced by de Luca and Termini. On the other hand, the indistinctness of fuzzy EDF is evaluated by Czogała-Gottwald-Pedrycz entropy measure. From this, the risk characteristic is defined as a value vector of the energy and entropy measures. We show relationships between risk characteristic of portfolio and risk characteristic of its component. For greater clarity of the discussion, we restrict the considerations to the case of a two-assets portfolio.

Keywords: fuzzy numbers, portfolio, ambiguity, indistinctness

JEL Classification: C02, C44, G11, G40
AMS Classification: 03E72, 91G30, 91B86

1 Introduction

When considering the rapidly changing and growing financial markets, both theorists and practitioners noticed the problem of imprecise return rates. Imprecision results from a lack of a clear recommendation for one alternative over various others, as well as from the lack of an explicit distinction between recommended and not recommended alternatives. Fuzzy numbers (FNs) are commonly accepted model of imprecise numbers. The answer to this issue came naturally, since in the same time, rapid development of fuzzy mathematics lead to the creation of a fuzzy portfolio analysis and various fuzzy portfolio models. The idea behind the most popular approach was to apply the existing portfolio theory and fuzzify some of its parameters, such as the return rate or present value (PV) [14].

In this paper we restrict our considerations to the case when the imprecise return rate is determined by combining the PV given by a trapezoidal FN and the random future value [13]. Moreover, the paper proves that the fuzzy expected discount factor stands as a better tool for appraising considered securities than the fuzzy expected return rate.

When processing imprecise values, we use FNs only to track the influence of the input data imprecision on the general imprecision of the obtained result. In [13], the relationships between imprecision risk burdening portfolio components and imprecision risk burdening the whole portfolio is considered, but excluding the possibility of short positions. This kind of a condition is very limiting in a portfolio analysis. In the case under consideration, to the best of our knowledge, the problem of short selling has not yet been considered. Therefore, the main goal of this paper is to present relationships between imprecision risk burdening portfolio components and imprecision risk burdening the whole portfolio, while allowing for short positions. For a greater clarity of presentation, we restrict our considerations to the case of a two-assets portfolio.

2 Fuzzy numbers – some facts

Dubois and Prade [3] define the fuzzy (FN) as such a fuzzy subset in the real line that fulfils some axioms. The space of all FNs we be denoted by $\mathbb{F}$. From the results of research obtained in [4], we have that any FN can be equivalently defined as follows:

[1]Adam Mickiewicz University, Department of Artificial Intelligence, ul. Uniwersytetu Poznańskiego 4,61-614 Poznań, Poland, jsiwek@amu.edu.pl
[2] Poznań University of Economics and Business, Department of Investment and Real Estate, al. Niepodległości 10, 61-875 Poznań, Poland, krzysztof.piasecki@ue.poznan.pl

Theorem 1: [2] For any FN $\mathcal{L}$ there exists such a non-decreasing sequence $(a,b,c,d) \subset \mathbb{R}$ that $\mathcal{L}(a,b,c,d,L_L,R_L) = \mathcal{L} \in \mathcal{F}(\mathbb{R})$ is determined by its membership function $\mu_L(\cdot\,|a,b,c,d,L_L,R_L) \in [0,1]^{\mathbb{R}}$ described by the identity

$$\mu_L(x|a,b,c,d,L_L,R_L) = \begin{cases} 0, x \notin [a,d], \\ L_L(x), x \in [a,b[, \\ 1, x \in [b,c], \\ R_L(x), x \in]c,d], \end{cases}. \tag{1}$$

where the left reference function $L_L \in [0,1]^{[a,b]}$ and the right reference function $R_L \in [0,1]^{[c,d]}$ are upper semi-continuous monotonic ones meeting the condition

$$\left[\overleftrightarrow{\mathcal{L}}\right]_{0^+} = \lim_{\alpha \to 0^+} \{x \in \mathbb{R} : \mu_L(x|a,b,c,d,S_L,E_L) \geq \alpha\} = [a,d]. \tag{2}$$

For any $z \in [b,c]$, a FN $\mathcal{L}(a,b,c,d,L_L,R_L)$ is a formal model of linguistic variable "about z". Understanding the phrase "about z" depends on the applied pragmatics of the natural language. We restrict our considerations to the following kind of FN.

Definition 1. For any non-decreasing sequence $(a,b,c,d) \subset \mathbb{R}$, a trapezoidal FN (TrFN) is a FN $\mathcal{T} = Tr(a,b,c,d) \in \mathbb{F}$ defined by its membership function $\mu_T \in [0,1]^{\mathbb{R}}$ in the following way

$$\mu_T(x) = \mu_{Tr}(x|a,b,c,d) = \begin{cases} 0, x \notin [a,d], \\ \dfrac{x-a}{b-a}, x \in [a,b[, \\ 1, x \in [b,c], \\ \dfrac{x-d}{c-d}, x \in]c,d]. \end{cases} \tag{3}$$

The space of all TrFNs is denoted by $\mathbb{F}_{Tr}$. In line with Zadeh's Extension Principle, the scalar multiplication $\odot$ on $\mathbb{F}_{Tr}$ is given by identity

$$\alpha \odot Tr(a,b,c,d) = \begin{cases} Tr(\alpha \cdot a, \alpha \cdot b, \alpha \cdot c, \alpha \cdot d)\ \alpha \geq 0 \\ Tr(\alpha \cdot d, \alpha \cdot c, \alpha \cdot b, \alpha \cdot a)\ \alpha < 0 \end{cases}. \tag{4}$$

Then, the addition $\oplus$ on $\mathbb{F}_{Tr}$ is determined by

$$Tr(a,b,c,d) \oplus Tr(e,f,g,h) = Tr(a+e, b+f, c+g, d+h). \tag{5}$$

After Klir [6] we understand imprecision as superposition of ambiguity and indistinctness. Ambiguity can be interpreted as a lack of a clear recommendation between one alternative among various others. Indistinctness, on the other hand, as a lack of explicit distinction between recommended and not recommended alternatives.

A proper tool for measuring ambiguity is the energy measure [10], which for any TrFN is determined by

$$d\big(Tr(a,b,c,d)\big) = \int_a^d \mu_T(x|a,b,c,d)\mathrm{d}x = \frac{1}{2} \cdot (d+c-b-a). \tag{6}$$

In a portfolio analysis, the most convenient tool for measuring indistinctness is Czogała-Gottwald-Pedrycz entropy measure introduced in [11]. For any TrFN, this measure can be calculated as

$$e\big(Tr(a,b,c,d)\big) = \int_a^d \min\{\mu_T(x|a,b,c,d), 1-\mu_T(x|a,b,c,d)\}\,\mathrm{d}x = \frac{1}{4} \cdot (d-c+b-a). \tag{7}$$

Energy and entropy measures together are called imprecision measures.

3 Expected discount factor

Any asset can be evaluated by the present value (PV), which is defined as a current equivalent of a present or future cash flow [12]. Therefore, PV can be imprecise. It implies that PV should be described by a FN. Buckley [1], Gutierrez [5], Kuchta [8] and Lesage [9] prove the sensibility of using TrFNs as fuzzy financial arithmetic tools. For these reasons, we assume that PV is estimated by TrFN

$$PV = Tr(V_s, V_f, V_l, V_e), \tag{8}$$

where the non-decreasing sequence $(V_s, V_f, \check{P}, V_l, V_e)$ is defined as follows:

- $\check{P}$ is the asset price,
- $[V_s, V_e]$ is an interval of all possible PVs,
- $[V_f, V_l] \subset [V_s, V_e]$ is an interval of all prices which do not perceptibly differ from the price $\check{P}$.

The expected return rate is a widely accepted tool for analysing financial assets. In practice of financial markets analysis, any asset is characterized by probability distribution of the return rate determined for a PV equal to the price $\check{P}$. Today, we have extensive knowledge about this subject. We assume that the expected value of a return rate probability distribution exists. This way we can determine the expected return rate (ERR) $\bar{r}$ as expected value of this distribution. In [13] it is shown that if PV is given by (8), then ERR is such a FN which is not a TrFN. On the other side, in [13] it is proven that expected discount factor (EDF) $\mathcal{V}$ is, contrarily, represented by a TrFN

$$\mathcal{V} = Tr\left(V_s \cdot \frac{\bar{v}}{\check{P}}, V_f \cdot \frac{\bar{v}}{\check{P}}, V_l \cdot \frac{\bar{v}}{\check{P}}, V_e \cdot \frac{\bar{v}}{\check{P}}\right). \tag{9}$$

The EDF $\bar{v}$ calculated using the return rate $\bar{r}$ is given by identity

$$\bar{v} = \frac{1}{1 + \bar{r}}. \tag{10}$$

Thus, the criterion of ERR maximization is replaced by the criterion of minimizing the EDF, with same theoretical conclusions.

An increase in ambiguity of EDF $\mathcal{V} \in \mathbb{F}_{Tr}$ suggests a higher number of alternative recommendations to choose from. This may result in making a decision, which will be *ex post* associated with a profit lower than maximal, that is with a loss of chance. This kind of risk is called an ambiguity risk. The ambiguity risk burdening EDF $\mathcal{V}$ is measured by energy measure $d(\mathcal{V})$.

An increase in the indistinctness of EDF $\mathcal{V} \in \mathbb{F}_{Tr}$, on the other hand, suggests that the differences between recommended and not recommended decision alternatives are harder to differentiate. This leads to an increase in the indistinctness risk, which is a risk of choosing a not recommended option. The indistinctness risk burdening EDF $\mathcal{V}$ is measured by entropy measure $e(\mathcal{V})$. Imprecision risk consists of both ambiguity and indistinctness risk, combined.

4 Imprecision risk burdening the portfolio

Let us consider the case of a two-asset portfolio π, consisting of financial assets Y_1 and Y_2. The asset Y_i $(i = 1; 2)$ is evaluated by

$$PV(Y_i) = Tr(V_s^{(i)}, V_f^{(i)}, V_l^{(i)}, V_e^{(i)}) \tag{11}$$

and by ERR $\bar{r}_i$ determined using asset price $\check{P}_i \in \mathbb{R}^+$. For each asset Y_i, we appoint its EDF

$$\mathcal{V}_i = Tr\left(V_s^{(i)} \cdot \frac{\bar{v}_i}{\check{P}_i}, V_f^{(i)} \cdot \frac{\bar{v}_i}{\check{P}_i}, V_l^{(i)} \cdot \frac{\bar{v}_i}{\check{P}_i}, V_e^{(i)} \cdot \frac{\bar{v}_i}{\check{P}_i}\right), \tag{12}$$

where $\bar{v}_i$ is an EDF appointed by (10), with use of ERR $\bar{r}_i$. In this section, we will consider relationships between imprecision risk burdening portfolio components and imprecision risk burdening the whole portfolio. Using (6) and (7), we obtain

$$d(\mathcal{V}_i) = \frac{\bar{v}_i}{2 \cdot \check{P}_i} \cdot \left(V_e^{(i)} + V_l^{(i)} - V_f^{(i)} - V_s^{(i)}\right), \tag{13}$$

$$e(\mathcal{V}_i) = \frac{\bar{v}_i}{4 \cdot \check{P}_i} \cdot \left(V_e^{(i)} - V_l^{(i)} + V_f^{(i)} - V_s^{(i)}\right), \tag{14}$$

In [13] authors considered a portfolio with excluded short sales. Then, portfolio PV is calculated as follows

$$PV(\pi) = Tr\left(V_s^{(1)}, V_f^{(1)}, V_l^{(1)}, V_e^{(1)}\right) \oplus Tr\left(V_s^{(2)}, V_f^{(2)}, V_l^{(2)}, V_e^{(2)}\right) = \\ = Tr\left(V_s^{(1)} + V_s^{(2)}, V_f^{(1)} + V_f^{(2)}, V_l^{(1)} + V_l^{(2)}, V_e^{(1)} + V_e^{(2)}\right). \tag{15}$$

The share p_i of the asset Y_i in the portfolio π is given by

$$p_i = \frac{\check{P}_i}{\check{P}_1 + \check{P}_2}. \tag{16}$$

Portfolio ERR $\bar{r}_\pi$ and portfolio ERR $\bar{v}_\pi$ are represented by equations

$$\bar{r}_\pi = p_1 \cdot \bar{r}_1 + p_2 \cdot \bar{r}_2, \tag{17}$$

$$\bar{v}_\pi = \left(\frac{p_1}{\bar{v}_1} + \frac{p_2}{\bar{v}_2}\right)^{-1}. \tag{18}$$

The portfolio EDF $\mathcal{V}_\pi$ equals

$$\mathcal{V}_\pi = \left(\frac{p_1}{\bar{v}_1} + \frac{p_2}{\bar{v}_2}\right)^{-1} \odot \left(\left(\frac{p_1}{\bar{v}_1} \odot \mathcal{V}_1\right) \oplus \left(\frac{p_2}{\bar{v}_2} \odot \mathcal{V}_2\right)\right). \tag{19}$$

Let us consider the case when portfolio π contains short position. Without loss of generality, we can assume that the asset Y_2 is a short position. According to the practice of the financial markets, we preclude arbitrage. Therefore, we have $\check{P}_2 < \check{P}_1$. In agree with (4), PV of such portfolio π is calculated as follows

$$PV(\pi) = Tr\left(V_s^{(1)}, V_f^{(1)}, V_l^{(1)}, V_e^{(1)}\right) \oplus \left((-1) \odot Tr\left(V_s^{(2)}, V_f^{(2)}, V_l^{(2)}, V_e^{(2)}\right)\right) = \\ = Tr\left(V_s^{(1)}, V_f^{(1)}, V_l^{(1)}, V_e^{(1)}\right) \oplus Tr\left(-V_e^{(2)}, -V_l^{(2)}, -V_f^{(2)}, -V_s^{(2)}\right) = \\ = Tr\left(V_s^{(1)} - V_e^{(2)}, V_f^{(1)} - V_l^{(2)}, V_l^{(1)} - V_f^{(2)}, V_e^{(1)} - V_s^{(2)}\right). \tag{20}$$

Then share p_i of the asset Y_i in the portfolio π is given by

$$p_1 = \frac{\check{P}_1}{\check{P}_1 - \check{P}_2}, \quad p_2 = \frac{-\check{P}_2}{\check{P}_1 - \check{P}_2} \tag{21}$$

Portfolio ERR $\bar{r}_\pi$ and portfolio EDF $\bar{v}_\pi$ are given respectively by equations (17) and (18). When comparing conditions (15) and (16) with conditions (20) and (21), we can see significant differences between the portfolio without the short sale and the portfolio including a short position. We have:

Theorem 1. If portfolio π contains short position, then its EDF $\mathcal{V}_\pi$ fulfils the condition (19).

Proof: By (9), we calculate the EDF of the portfolio π:

$$\mathcal{V}_\pi = Tr\left(\left(V_s^{(1)} - V_e^{(2)}\right) \cdot \frac{\bar{v}_\pi}{\check{P}_1 - \check{P}_2}, \left(V_f^{(1)} - V_l^{(2)}\right) \cdot \frac{\bar{v}_\pi}{\check{P}_1 - \check{P}_2}, \left(V_l^{(1)} - V_f^{(2)}\right) \cdot \frac{\bar{v}_\pi}{\check{P}_1 - \check{P}_2}, \left(V_e^{(1)} - V_s^{(2)}\right) \cdot \frac{\bar{v}_\pi}{\check{P}_1 - \check{P}_2}\right) =$$

$$= \bar{v}_\pi \odot Tr\left(p_1 \cdot \frac{V_s^{(1)}}{\check{P}_1} - p_2 \cdot \frac{V_e^{(2)}}{-\check{P}_2}, p_1 \cdot \frac{V_f^{(1)}}{\check{P}_1} - p_2 \cdot \frac{V_l^{(2)}}{-\check{P}_2}, p_1 \cdot \frac{V_l^{(1)}}{\check{P}_1} - p_2 \cdot \frac{V_f^{(2)}}{-\check{P}_2}, p_1 \cdot \frac{V_e^{(1)}}{\check{P}_1} - p_2 \cdot \frac{V_s^{(2)}}{-\check{P}_2}\right) =$$

$$= \left(\frac{p_1}{\bar{v}_1} + \frac{p_2}{\bar{v}_2}\right)^{-1} \odot Tr\begin{pmatrix} \frac{p_1}{\bar{v}_1} \cdot \left(\bar{v}_1 \frac{V_s^{(1)}}{\check{P}_1}\right) + \frac{-p_2}{\bar{v}_2} \cdot \left(\bar{v}_2 \cdot \frac{V_e^{(2)}}{-\check{P}_2}\right), \frac{p_1}{\bar{v}_1} \cdot \left(\bar{v}_1 \frac{V_f^{(1)}}{\check{P}_1}\right) + \frac{-p_2}{\bar{v}_2} \cdot \left(\bar{v}_2 \cdot \frac{V_l^{(2)}}{-\check{P}_2}\right), \\ \frac{p_1}{\bar{v}_1} \cdot \left(\bar{v}_1 \frac{V_l^{(1)}}{\check{P}_1}\right) + \frac{-p_2}{\bar{v}_2} \cdot \left(\bar{v}_2 \cdot \frac{V_f^{(2)}}{-\check{P}_2}\right), \frac{p_1}{\bar{v}_1} \cdot \left(\bar{v}_1 \frac{V_e^{(1)}}{\check{P}_1}\right) + \frac{-p_2}{\bar{v}_2} \cdot \left(\bar{v}_2 \cdot \frac{V_s^{(2)}}{-\check{P}_2}\right) \end{pmatrix} =$$

$$= \left(\frac{p_1}{\bar{v}_1} + \frac{p_2}{\bar{v}_2}\right)^{-1} \cdot \begin{pmatrix} \frac{p_1}{\bar{v}_1} \odot Tr\left(\bar{v}_1 \frac{V_s^{(1)}}{\check{P}_1}, \bar{v}_1 \frac{V_f^{(1)}}{\check{P}_1}, \bar{v}_1 \frac{V_l^{(1)}}{\check{P}_1}, \bar{v}_1 \frac{V_e^{(1)}}{\check{P}_1}\right) \oplus \\ \frac{-p_2}{\bar{v}_2} \odot Tr\left(\bar{v}_2 \cdot \frac{V_e^{(2)}}{-\check{P}_2}, \bar{v}_2 \cdot \frac{V_l^{(2)}}{-\check{P}_2}, \bar{v}_2 \cdot \frac{V_f^{(2)}}{-\check{P}_2}, \bar{v}_2 \cdot \frac{V_s^{(2)}}{-\check{P}_2}\right) \end{pmatrix} =$$

$$= \left(\frac{p_1}{\bar{v}_1} + \frac{p_2}{\bar{v}_2}\right)^{-1} \cdot \left(\frac{p_1}{\bar{v}_1} \odot \mathcal{V}_1 \oplus \frac{p_2}{\bar{v}_2} \odot Tr\left(\bar{v}_2 \cdot \frac{V_s^{(2)}}{\check{P}_2}, \bar{v}_2 \cdot \frac{V_f^{(2)}}{\check{P}_2}, \bar{v}_2 \cdot \frac{V_l^{(2)}}{\check{P}_2}, \bar{v}_2 \cdot \frac{V_e^{(2)}}{\check{P}_2}\right)\right) =$$

$$= \left(\frac{p_1}{\bar{v}_1} + \frac{p_2}{\bar{v}_2}\right)^{-1} \odot \left(\left(\frac{p_1}{\bar{v}_1} \odot \mathcal{V}_1\right) \oplus \left(\frac{p_2}{\bar{v}_2} \odot \mathcal{V}_2\right)\right). \textbf{QED}$$

By (19), (6) and (7), we obtain that the imprecision measure of the portfolio EDF is, in fact, a linear combination of analogous imprecision measures calculated for each of component assets:

$$d(\mathcal{V}_\pi) = \left(\frac{p_1}{\bar{v}_1} + \frac{p_2}{\bar{v}_2}\right)^{-1} \cdot \left(\frac{p_1}{\bar{v}_1} \cdot d(\mathcal{V}_1) + \frac{p_2}{\bar{v}_2} \cdot d(\mathcal{V}_2)\right), \tag{22}$$

$$e(\mathcal{V}_\pi) = \left(\frac{p_1}{\bar{v}_1} + \frac{p_2}{\bar{v}_2}\right)^{-1} \cdot \left(\frac{p_1}{\bar{v}_1} \cdot e(\mathcal{V}_1) + \frac{p_2}{\bar{v}_2} \cdot e(\mathcal{V}_2)\right). \tag{23}$$

The linear combination mentioned above includes weights for imprecision measures of the component EDFs. The weight calculated for the asset Y_i increases with its share $p_i$in the portfolio and, respectively, decreases with the value of its discount factor $\bar{v}_i$. This fact leads to a conclusion that when trying to minimize the imprecision risk of a portfolio, one should focus on:

- minimizing the ambiguity and indistinctness of such portfolio's long positions, which are characterized by the highest expected return rates;
- maximizing the ambiguity and indistinctness of such portfolio's short positions, which are characterized by the highest expected return rates.

On the other hand, the shares of an asset in the whole portfolio are, according to the theory, appointed post factum after gathering available information on said assets. It means that ambiguity and indistinctness of portfolio components are determined prior to the portfolio analysis.

Conditions (22) and (23) show that if short sales are excluded, then the portfolio diversification only "averages" the risk of imprecision. Moreover, we can see that the use of short sales in the portfolio diversification may increase the imprecision risk. This is in line with the practice of financial markets, since use of short selling increases the risk for the investor.

Example 1. Let us consider financial assets Y_1 and Y_2. The asset Y_1 is evaluated by

$$PV(Y_1) = Tr(150, 250, 400, 500)$$

and by ERR $\bar{r}_1 = 0.09$ determined by using asset price $\check{P}_1 = 300$. Applying ERR $\bar{r}_1$, we appoint the EDF

$$\bar{v}_1 = \frac{1}{1 + 0.09} = 0.9174.$$

If we take into account $PV(Y_1)$, then EDF can be calculated in the following way

$$\mathcal{V}_1 = Tr\left(150 \cdot \frac{0.9174}{300}, 250 \cdot \frac{0.9174}{300}, 400 \cdot \frac{0.9174}{300}, 500 \cdot \frac{0.9174}{300}\right) = Tr(0.4587, 0.7645, 1.223, 1.5291).$$

The imprecision risk burdening the asset Y_1 is evaluated by energy measure $d(\mathcal{V}_1) = 0.7645$ and entropy measure $e(\mathcal{V}_1) = 0.1529$. Asset Y_2 is evaluated by

$$PV(Y_2) = Tr(100, 150, 225, 250)$$

and by ERR $\bar{r}_2 = 0.05$ determined by asset price $\check{P}_2 = 200$. Using ERR $\bar{r}_2$ we appoint EDF

$$\bar{v}_2 = \frac{1}{1 + 0.05} = 0.9524.$$

If we consider $PV(Y_2)$, then we calculate the EDF as follows

$$\mathcal{V}_2 = Tr\left(100 \cdot \frac{0.9524}{200}, 150 \cdot \frac{0.9524}{200}, 225 \cdot \frac{0.9524}{200}, 250 \cdot \frac{0.9524}{200}\right) = Tr(0.4762, 0.7143, 1.071, 1.1905).$$

The imprecision risk burdening the asset Y_2 is evaluated by energy measure $d(\mathcal{V}_2) = 0.5357$ and entropy measure $e(\mathcal{V}_2) = 0.0892$.

Let us consider portfolio π_1 containing the long positions Y_1 and Y_2. This portfolio is evaluated by

$$PV(\pi_1) = Tr(150 + 100, 250 + 150, 400 + 225, 500 + 250) = Tr(250, 400, 625, 750)$$

The π_1 portfolio's structure is characterized by shares

$$p_1^{(1)} = \frac{300}{300 + 200} = 0.6, p_2^{(1)} = \frac{200}{300 + 200} = 0.4.$$

Portfolio's ERR $\bar{r}_\pi^{(1)}$ and ERR $\bar{v}_\pi^{(1)}$ are equal to

$$\bar{r}_\pi^{(1)} = 0.6 \cdot 0.09 + 0.4 \cdot 0.05 = 0.074, \bar{v}_\pi^{(1)} = \left(\frac{0.6}{0.9174} + \frac{0.4}{0.9524}\right)^{-1} = 0.9311.$$

If we consider $PV(\pi_1)$ then EDF is

$$\mathcal{V}_\pi^{(1)} = 0.9311 \odot \left(\left(\frac{0.6}{0.9174} \odot \mathcal{V}_1\right) \oplus \left(\frac{0.4}{0.9524} \odot \mathcal{V}_2\right)\right) = Tr(0.4656, 0.7449, 1.1639, 1.3967).$$

The imprecision risk burdening the asset π_1 is evaluated by energy measure $d(\pi_1) = 0.6750$ and entropy measure $e(\pi_1) = 0.1280$. These results illustrate the imprecision risk averaging.

Let us consider a portfolio π_2 containing a long position Y_1 and a short Y_2. This portfolio is evaluated by

$$PV(\pi_2) = Tr(150 - 250, 250 - 225, 400 - 150, 500 - 100) = Tr(-150, 25, 250, 750)$$

The portfolio's π_2 structure is characterized by shares

$$p_1^{(2)} = \frac{300}{300 - 200} = 3, p_2^{(2)} = \frac{-200}{300 - 200} = -2.$$

Portfolio's ERR $\bar{r}_\pi^{(2)}$ and ERR $\bar{v}_\pi^{(2)}$ are equal to

$$\bar{r}_\pi^{(2)} = 3 \cdot 0.09 + (-2) \cdot 0.05 = 0.17, \bar{v}_\pi^{(2)} = \left(\frac{3}{0.9174} + \frac{(-2)}{0.9524}\right)^{-1} = 0.8547.$$

If we consider $PV(\pi_2)$ then we can appoint EDF in following way

$$\mathcal{V}_\pi^{(2)} = 0.8547 \odot \left(\left(\frac{3}{0.9174} \odot \mathcal{V}_1\right) \oplus \left(\frac{-2}{0.9524} \odot \mathcal{V}_2\right)\right) = Tr(-1.2821, 0.2127, 2.1368, 6.4103).$$

The imprecision risk burdening the asset π_2 is evaluated by energy measure $d(\pi_2) = 3.0983$ and entropy measure $e(\pi_2) = 0.5876$. These results illustrate the possibility to increase the imprecision risk due to the inclusion of a short position in the portfolio.

5 Final remarks

In this paper, we explicitly described the relationships between imprecision risk burdening portfolio components and the same risk burdening the two –assets portfolio, with short position allowed. By using mathematical induction, we can easily generalize the obtained results to the general portfolio case.

From a formal point of view, these kind of relationships described for the case with short sale acceptance are identical with those obtained for the case of short sale rejection. On the other hand, introduction of a short sale radically changes the assessment of the portfolio imprecision risk. To a large extent, these changes are caused by the specificity of the definition (4) of scalar multiplication for TrFN. For e.g. ordered FN [7] the definition of scalar multiplication can be different. Because of this, ordered FN's suitability for managing imprecision risk of a portfolio should be investigated, which is a promising direction for future research.

Acknowledgements

The research was financed by the Polish National Science Centre (grant no. 2015/17/N/HS4/00206).

References

[1] BUCKLEY, I. J. (1987). The fuzzy mathematics of finance, *Fuzzy Sets and Systems* 21, 257–273.

[2] DELGADO, M., VILA, M. A., & VOXMAN, W. (1998). On a canonical representation of fuzzy numbers, *Fuzzy Sets and Systems* 93, 125–135.

[3] DUBOIS, D. & PRADE, H. (1978). Operations on fuzzy numbers. International Journal System Sciences, 9, 613–629. https://doi.org/10.1080/00207727808941724.
[4] GOETSCHEL, R. & VOXMAN, W. (1986). Elementary fuzzy calculus. *Fuzzy Sets and Systems* 18, 31–43. https://doi.org/10.1016/0165-0114(86)90026-6.
[5] GUTIERREZ, I. (1989). Fuzzy numbers and Net Present Value, *Scandinavian Journal of Management* 5 (2), 149–159.
[6] KLIR, G. J. (1993). Developments in uncertainty-based information, *Advances in Computers* 36, 255–332, doi:10.1016/s0065-2458(08)60273-9.
[7] KOSIŃSKI, W., PROKOPOWICZ, P., & ŚLĘZAK, D. (2002). Fuzzy numbers with algebraic operations: algorithmic approach. In M. KLOPOTEK, S. T. WIERZCHOŃ, & M. MICHALEWICZ, (Eds.) *Proc. IIS'2002*, (pp. 311–320) Sopot, Poland, Physica Verlag, Heidelberg.
[8] KUCHTA, D. (2000). Fuzzy capital budgeting, *Fuzzy Sets and Systems*, 111, 367–385.
[9] LESAGE, C. (2001). Discounted cash-flows analysis. An interactive fuzzy arithmetic approach. *European Journal of Economic and Social System*s 15(2), 49–68.
[10] DE LUCA, A. & TERMINI, S. (1979), Entropy and energy measures of fuzzy sets, In M. M. GUPTA ET AL (Eds.) *Advances in fuzzy set theory and applications* 20, (pp. 321–338), Amsterdam, North Holland.
[11] PEDRYCZ, W., GOTTWALD, S., & CZOGAŁA, E. (1982). Measures of fuzziness and operations with fuzzy sets, *Stochastica* 3, 187–205.
[12] PIASECKI, K. (2012) Basis of financial arithmetic from the viewpoint of the utility theory, *Operations Research and Decisions* 22(3), 37–53.
[13] PIASECKI, K. & SIWEK, J. (2018). Multi-asset portfolio with trapezoidal fuzzy present values, *Przegląd Statystyczny (Statistical Review)* LXV(2), 183–199, DOI: 10.5604/01.3001.0014.0535.
[14] TSAUR, R. (2013). Fuzzy portfolio model with different investor risk attitudes, *European Journal of Operational Research*, vol. 227., DOI: 10.1016/j.ejor.2012.10.036.

Risk-Sensitivity and Average Optimality in Markov and Semi-Markov Reward Processes

Karel Sladký[1]

Abstract. This contribution is devoted to risk-sensitivity in long-run average optimality of Markov and semi-Markov reward processes. Since the traditional average optimality criteria cannot reflect the variability-risk features of the problem, we are interested in more sophisticated approaches where the stream of rewards generated by the Markov chain that is evaluated by an exponential utility function with a given risk sensitivity coefficient. Recall that for the risk sensitivity coefficient equal to zero (i.e. the so called risk-neutral case) we arrive at traditional optimality criteria, if the risk sensitivity coefficient is close to zero the Taylor expansion enables to evaluate variability of the generated total reward. Observe that the first moment of the total reward corresponds to expectation of total reward and the second central moment to the reward variance. In this note we present necessary and sufficient risk-sensitivity and risk-neutral optimality conditions for long run risk-sensitive average optimality criterion of unichain Markov and semi-Markov reward processes.

Keywords: Markov and semi-Markov reward processes, exponential utility function, risk sensitivity, long run optimality.

JEL Classification: C44, C61, C63
AMS Classification: 90C40, 60J10, 93E20

1 Formulation and Notation

Consider a controlled semi-Markov reward process $Y = \{Y(t), t \geq 0\}$ with finite state space $\mathcal{I} = \{1, 2, \ldots, N\}$ along with the embedded Markov chain $X = \{X_n, n = 0, 1, \ldots\}$. The development of the process $Y(t)$ over time is the following: At time $t = 0$ if $Y(0) = i$ the decision maker selects decision from an infinite (compact) set $\mathcal{A}_i \equiv [0, K_i] \subset \mathbb{R}$ of possible decisions (actions) in state $i \in \mathcal{I}$. Then state j is reached in the next transition with a given probability $p_{ij}(a)$ after random time $\eta_{ij}(a)$. Let $F_{ij}(a, \tau)$ be a non-lattice distribution function representing the conditional probability $\mathrm{P}(\eta_{ij} \leq \tau)$. We assume that for $\ell = 1, 2$, $0 < d_{ij}^{(\ell)} = \int_0^\infty \tau^\ell \, \mathrm{d}F_{ij}(a, \tau) < \infty$ hence also $0 < d_i^{(\ell)} = \sum_{j=1}^N p_{ij}(a) d_{ij}^{(\ell)}(a) < \infty$. Finally, one-stage transition reward $r_{ij} > 0$ will be accrued to transition from state i to state j, and reward rate $r_i(a)$ per unit of time incurred in state i is earned. We assume that each $p_{ij}(a)$ and $r_i(a)$ is a continuous function of $a \in \mathcal{A}_i$.

A (Markovian) policy controlling the semi-Markov process Y, say $\pi = (f^0, f^1, \ldots)$, is identified by a sequence of decision vectors $\{f^n, n = 0, 1, \ldots\}$ where $f^n \in \mathcal{F} \equiv \mathcal{A}_1 \times \ldots \times \mathcal{A}_N$ for every $n = 0, 1, 2, \ldots$, and $f_i^n \in \mathcal{A}_i$ is the decision (or action) taken at the nth transition if the embedded Markov chain X is in state i. Let π^k be a sequence of decision vectors starting at the k-th transition, hence $\pi = (f^0, f^1, \ldots f^{k-1}, \pi^k)$. Policy which selects at all times the same decision rule, i.e. $\pi \sim (f)$, is called stationary; $P(f)$ is transition probability matrix with elements $p_{ij}(f_i)$. Stationary policy $\tilde{\pi}$ is randomized if there exist decision vectors $f^{(1)}, f^{(2)}, \ldots, f^{(m)} \in \mathcal{F}$ and on following policy $\tilde{\pi}$ we select in state i action $f_i^{(j)}$ with a given probability $\kappa_i^{(j)}$ (of course, $\kappa_i^{(j)} \geq 0$ with $\sum_{j=1}^N \kappa_i^{(j)} = 1$ for all $i \in \mathcal{I}$). For details see e.g. [1, 9, 10].

Let ξ_n be the cumulative reward obtained in the n first transitions of the considered embedded Markov chain X. Since the process starts in state X_0, $\xi_n = \sum_{k=0}^{n-1} [r_{X_k} \cdot \eta_{X_k, X_{k+1}} + r_{X_k, X_{k+1}}]$. Similarly let $\xi_{(m,n)}$ be reserved for the cumulative (random) reward, obtained from the mth up to the nth transition (obviously, $\xi_n = r_{X_0} \cdot \eta_{X_0, X_1} + r_{X_0, X_1} + \xi_{(1,n)}$, we tacitly assume that $\xi_{(1,n)}$ starts in state X_1).

For the (random) reward earned up to time t, say $\xi(t)$ we have $\xi(t) := \left[\int_0^t r_{Y(s)} \mathrm{d}s + \sum_{k=0}^{N(t)} r_{Y(\tau_k^-), Y(\tau_k^+)}\right]$, with $Y(s)$, denoting the state of the system at time s, $Y(\tau_k^-)$ and $Y(\tau_k^+)$ the state just prior and after the kth

[1] Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, sladky@utia.cas.cz

jump, $N(t)$ the number of jumps up to time t, and $\mathcal{R}_i(f,t) := \mathsf{E}_i^{\pi}\xi(t)$ denote the expected total reward of the semi-Markov process $Y(t)$ up to time t given its initial state at time $t=0$ if policy $\pi \sim (f)$ is followed.

In this note, we assume that the stream of rewards generated by the Markov processes is evaluated by an exponential utility function, say $\bar{u}^{\gamma}(\cdot)$, i.e. a separable utility function with constant risk sensitivity $\gamma \in \mathbb{R}$. For more details see e.g. [2, 3, 4, 7, 14]. Then the utility assigned to the (random) outcome ξ is given by

$$\bar{u}^{\gamma}(\xi) := \begin{cases} (\text{sign}\,\gamma)\ \exp(\gamma\xi), & \text{if } \gamma \neq 0, \quad \text{risk-sensitive case,} \\ \xi \qquad \text{for } \gamma = 0 & \qquad\qquad\quad \text{risk-neutral case.} \end{cases} \tag{1}$$

Observe that exponential utility function $\bar{u}^{\gamma}(\cdot)$ is separable, and multiplicative if the risk sensitivity $\gamma \neq 0$ or additive for $\gamma = 0$. In particular, for $u^{\gamma}(\xi) := \exp(\gamma\xi)$ we have $u^{\gamma}(\xi_1+\xi_2) = u^{\gamma}(\xi_1)\cdot u^{\gamma}(\xi_2)$ if $\gamma \neq 0$ and $u^{\gamma}(\xi_1+\xi_2) \equiv \xi_1+\xi_2$ for $\gamma = 0$.

The certainty equivalent corresponding to ξ, say $Z^{\gamma}(\xi)$, is given by

$$\bar{u}^{\gamma}(Z^{\gamma}(\xi)) = \mathsf{E}\,[\bar{u}^{\gamma}(\xi)] \qquad \text{(the symbol } \mathsf{E} \text{ is reserved for expectation).} \tag{2}$$

From (1), (2) we can immediately conclude that

$$Z^{\gamma}(\xi) = \begin{cases} \gamma^{-1}\ \ln\{\mathsf{E}\ u^{\gamma}(\xi)\}, & \text{if } \gamma \neq 0 \\ \mathsf{E}\,[\xi] & \text{for } \gamma = 0. \end{cases} \tag{3}$$

Considering Markov decision process X, then if the process starts in state i, i.e. $X_0 = i$ and policy $\pi = (f^n)$ is followed, for the expectation of utility assigned to (cumulative) random reward ξ_n obtained in the n first transitions we get by (1)

$$\mathsf{E}_i^{\pi}\ \bar{u}^{\gamma}(\xi_n) := \begin{cases} (\text{sign}\,\gamma)\,\mathsf{E}_i^{\pi}\ \exp(\gamma\xi_n), & \text{if } \gamma \neq 0, \quad \text{risk-sensitive case} \\ \mathsf{E}_i^{\pi}\ \xi_n \quad \text{for } \gamma = 0 & \qquad\qquad\quad \text{risk-neutral case.} \end{cases} \tag{4}$$

In what follows let

$$\bar{U}_i^{\pi}(\gamma,n) := \mathsf{E}_i^{\pi}\bar{u}^{\gamma}(\xi_n), \quad U_i^{\pi}(\gamma,n) := \mathsf{E}_i^{\pi}\exp(\gamma\xi_n), \quad v_i^{\pi}(n) := \mathsf{E}_i^{\pi}\ (\xi_n). \tag{5}$$

In this note we focus attention on risk-neutral and risk-sensitive optimality of so called unichain models, i.e. when the underlying Markov chain contains a single class of recurrent state and present characterization of control policies by discrepancy functions. Discrepancy functions were originally introduced in [8] for risk-neutral unichain models, possible extension to multichain case can be found in [11, 12]. To this end we make

Assumption GA. There exists state $i_0 \in \mathcal{I}$ that is accessible from any state $i \in \mathcal{I}$ for every $f \in \mathcal{F}$.

Obviously, if Assumption GA holds, then the resulting transition probability matrix $P(f)$ is *unichain* for every $f \in \mathcal{F}$ (i.e. $P(f)$ has no two disjoint closed sets).

2 Risk-Neutral Optimality in Unichain Semi-Markov Processes

At first we focus attention on the embedded Markov chains and slightly extend some results reported in [13]. To this end, on introducing for arbitrary $g, w_j \in \mathbb{R}$ $(i,j \in \mathcal{I})$ and decision $f \in \mathcal{F}$, the discrepancy functions

$$\tilde{\varphi}_{i,j}(w^{\mathrm{c}},g^{\mathrm{c}},f) := d_i(f_i)\cdot r(i) + r_{ij} - w_i^{\mathrm{c}} + w_j^{\mathrm{c}} - g^{\mathrm{c}}, \qquad \bar{\varphi}_{i,j}(w^{\mathrm{t}},g^{\mathrm{t}},f) := d_i(f_i) - w_i^{\mathrm{t}} + w_j^{\mathrm{t}} - g^{\mathrm{t}} \tag{6}$$

for the random reward obtained, resp. time elapsed, up to the nth transition we have

$$\xi_n = ng^{\mathrm{c}} + w_{X_0}^{\mathrm{c}} - w_{X_n}^{\mathrm{c}} + \sum_{k=0}^{n-1}\tilde{\varphi}_{X_k,X_{k+1}}(w^{\mathrm{c}},g^{\mathrm{c}},f) \ \text{ resp. } \ \eta_n = ng^{\mathrm{t}} + w_{X_0}^{\mathrm{t}} - w_{X_n}^{\mathrm{t}} + \sum_{k=0}^{n-1}\bar{\varphi}_{X_k,X_{k+1}}(w^{\mathrm{t}},g^{\mathrm{t}},f). \tag{7}$$

Hence by (7) for the expectation of ξ_n, $\mathsf{E}_i^{\pi}\xi_n =: v_i^{\pi}(n)$, resp. of η_n, with $\mathsf{E}_i^{\pi}\eta_n =: t_i^{\pi}(n)$, we get

$$v_i^{\pi}(n) = ng^{\mathrm{c}} + w_i^{\mathrm{c}} + \mathsf{E}_i^{\pi}\{\sum_{k=0}^{n-1}\tilde{\varphi}_{X_k,X_{k+1}}(w^{\mathrm{c}},g^{\mathrm{c}},f) - w_{X_n}^{\mathrm{c}}\}, \tag{8}$$

$$t_i^{\pi}(n) = ng^{\mathrm{t}} + w_i^{\mathrm{t}} + \mathsf{E}_i^{\pi}\{\sum_{k=0}^{n-1}\bar{\varphi}_{X_k,X_{k+1}}(w^{\mathrm{t}},g^{\mathrm{t}},f) - w_{X_n}^{\mathrm{t}}\}. \tag{9}$$

Now we show how to express average reward generated by the semi-Markov process $Y(t), t \geq 0$ in terms of the embedded Markov chain X_n. Considering policy $\pi \sim (f)$, let

$$\tilde{\varphi}_i(w^{\mathrm{c}},g^{\mathrm{c}},f) \quad := \quad \sum_{j\in\mathcal{I}} p_{ij}(f_i)\tilde{\varphi}_{i,j}(w^{\mathrm{c}},g^{\mathrm{c}},f) = \sum_{j\in\mathcal{I}} p_{ij}(f_i)[d_i(f_i)\cdot r(i) + r_{ij} - w_i^{\mathrm{c}} + w_j^{\mathrm{c}} - g^{\mathrm{c}}], \tag{10}$$

$$\bar{\varphi}_i(w^{\mathrm{t}},g^{\mathrm{t}},f) \quad := \quad \sum_{j\in\mathcal{I}} p_{ij}(f_i)\bar{\varphi}_{i,j}(w^{\mathrm{t}},g^{\mathrm{t}},f) = \sum_{j\in\mathcal{I}} p_{ij}(f_i)[d_i(f_i) - w_i^{\mathrm{t}} + w_j^{\mathrm{t}} - g^{\mathrm{t}}] \tag{11}$$

It is well-known from the dynamic programming literature (cf. e.g. [1, 6, 9, 10]) that for every $f \in \mathcal{F}$ and arbitrary transition costs $s_{ij} = d_i(f_i)r(i) + r_{ij}, i,j, \in \mathcal{I}$, there exist numbers $g(f)$ and $w_i(f), i \in \mathcal{I}$ (unique up to additive constant) such that

$$w_i(f) + g(f) = d_i(f_i)r(i) + \sum_{j\in\mathcal{I}} p_{ij}(f_i)[r_{ij} + w_j(f)], \ \ (i \in \mathcal{I}), \quad \text{i.e.} \tag{12}$$

$$\sum_{j\in\mathcal{I}} p_{ij}(f_i)\,\varphi_{i,j}(w,g) = 0 \ \text{ where } \quad \varphi_{i,j}(w,g) := d_i(f_i)r(i) + r_{ij} - w_i(f) + w_j(f) - g(f).$$

In particular, for suitable selected $w_j^{\mathrm{c}}(f)$, resp. $w_j^{\mathrm{t}}(f)$, we have

$$v_i^{\pi}(n) = ng^{\mathrm{c}}(f) + w_i^{\mathrm{c}}(f) - \sum_{j\in\mathcal{I}} p_{ij}(f)\cdot w_j^{\mathrm{c}}(f), \qquad t_i^{\pi}(n) = ng^{\mathrm{t}} + w_i^{\mathrm{t}} - \sum_{j\in\mathcal{I}} p_{ij}(f)\cdot w_j^{\mathrm{t}}(f), \quad \text{where} \tag{13}$$

$$w_i^{\mathrm{c}}(f)+g^{\mathrm{c}}(f) = d_i(f_i)\cdot r(i)+\sum_{j\in\mathcal{I}} p_{ij}(f_i)[r_{ij}+w_j^{\mathrm{c}}(f)], \text{ resp. } w_i^{\mathrm{t}}(f)+g^{\mathrm{t}}(f) = d_i(f_i)+\sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot w_j^{\mathrm{t}}(f), \quad (i \in \mathcal{I}). \tag{14}$$

After some manipulation we obtain from (13)

$$w_i^{\mathrm{t}}(f)\cdot\frac{g^{\mathrm{c}}(f)}{g^{\mathrm{t}}(f)} + g^{\mathrm{c}}(f) = d_i(f_i)\cdot\frac{g^{\mathrm{c}}(f)}{g^{\mathrm{t}}(f)} + \sum_{j\in\mathcal{I}} p_{ij}(f_i)\cdot w_j^{\mathrm{t}}(f)\frac{g^{\mathrm{c}}(f)}{g^{\mathrm{t}}(f)} \tag{15}$$

and by subtracting (15) from (14) we get

$$w_i(f) = \bar{r}_i(f) + \sum_{j\in\mathcal{I}} p_{ij}(f_i)w_j(f) - d_i(f_i)g(f) \qquad \text{where} \tag{16}$$

$$w_i(f) := w_i^{\mathrm{c}}(f) - w_i^{\mathrm{t}}(f)\cdot\frac{g^{\mathrm{c}}(f)}{g^{\mathrm{t}}(f)}, \quad g(f) := \frac{g^{\mathrm{c}}(f)}{g^{\mathrm{t}}(f)}, \quad \bar{r}_i(f) = d_i(f_i)\cdot r(i) + \sum_{j\in\mathcal{I}, j\neq i} p_{ij}(f_i)r_{ij}.$$

On introducing matrix notations $P(f) = [p_{ij}(f_i)]$, $\quad D(f) = \text{diag}\,[d_i(f_i)]$, (square matrices)
$\bar{r}(f) = [\bar{r}_i(f)]$, $\quad w(f) = [w_i(f)]$, $\quad \bar{g}(f) = [g(f)]$ (column vectors) $\qquad$ equation (16) can be written as

$$w(f) = \bar{r}(f) + P(f)w(f) - D(f)g(f) \Rightarrow g(f) = D^{-1}(f)\bar{r}(f) + [D^{-1}(f)P(f) - I]\cdot w(f)]. \tag{17}$$

Let

$\tilde{r}(f) := D^{-1}(f)\bar{r}(f), \qquad \tilde{w}(f) := D^{-1}(f)w(f), \qquad \tilde{P}(f) := D^{-1}(f)\cdot P(f)\cdot D(f)$

Then for the elements of $\tilde{r}(f)$, $\tilde{w}(f)$, $\tilde{P}(f)$ we have

$$\tilde{r}_i(f) = \bar{r}(i) + [d_i(f_i)]^{-1}r_{ij}, \quad \tilde{p}_{ij}(f_i) := p_{ij}(f_i)\frac{[d_j(f_j)]}{[d_i(f_i)]}, \quad \tilde{w}_i(f) := [d_i(f_i)]^{-1}w_i(f).$$

In particular, let us consider continuous-time Markov decision chain with transition intensities $\mu_{ij}(f_i)$, where $\sum_{j\in\mathcal{I},j\neq i}\mu_{ij}(f_i) = -\mu_{ii}(f_i)$ and $\mu_i(f_i) = -\mu_{ii}(f_i)$ is the intensity of jumps from state i. Obviously, this is a very special case of semi-Markov processes with transition probabilities $p_{ij}(f) = \frac{\mu_{ij}(f_i)}{\mu_i(f_i)}$, and expected holding time $d_i(f_i) = \frac{1}{\mu_i(f_i)}$ in state i. Then on replacing in (17) transition probabilities and expected holding times by transition intensities for the average reward per unit of time of the considered continuous-time Markov process we conclude that

$$g(f) = r(i) + \sum_{j\neq i}\mu_{ij}(f_i)r_{ij} + \sum_{j}\mu_{ij}(f_i)w_{ij}(f) \tag{18}$$

the standard equation for average reward of a continuous time Markov reward chain (cf. e.g. [6]).

3 Risk-Sensitive Optimality in Unichain Semi-Markov Processes

In this section we assume that the risk sensitivity coefficient $\gamma \neq 0$ and the transition probability matrix $P(f)$ is unichain for every $f \in \mathcal{F}$, i.e. Assumption GA is fulfilled. We show how the discrepancy functions can be employed for finding optimality conditions for risk-sensitive Markov and semi-Markov processes. These results slightly extend some previous results reported in [12, 14, 15, 16, 17].

Similarly to the risk-neutral models, let for real g, w_i's $(i \in \mathcal{I})$

$$\varphi_{ij}(w,g,f) := r_{ij} + d_i(f)\cdot[r(i) - g] + w_j - w_i, \text{ where } w' = \min_{i\in\mathcal{I}} w_i,\ w'' = \max_{i\in\mathcal{I}} w_i. \tag{19}$$

Then if policy $\pi = (f^n)$ is followed we get by (5),(19) for the risk-sensitive case

$$U_i^\pi(\gamma,n) = \mathsf{E}_i^\pi \mathrm{e}^{\gamma\sum_{k=0}^{n-1}[d_{X_k}(f_{X_k})\cdot r(X_k)+r_{X_k,X_{k+1}}]} = \mathrm{e}^{\gamma w_i}\times \mathsf{E}_i^\pi \mathrm{e}^{\gamma[\sum_{k=0}^{n-1}\varphi_{X_k,X_{k+1}}(w,g,f)-w_{X_n}]}. \tag{20}$$

Hence for a given $\gamma \neq 0$ there exist numbers $\bar{w}, \tilde{w}$ such that for any policy $\pi = (f^n)$

$$\mathsf{E}_i^\pi \mathrm{e}^{\gamma[\sum_{k=0}^{n-1}\varphi_{X_k,X_{k+1}}(w,g,f)-\bar{w}]} \leq \frac{U_i^\pi(\gamma,n)}{\mathrm{e}^{\gamma w_i}} \leq \mathsf{E}_i^\pi \mathrm{e}^{\gamma[\sum_{k=0}^{n-1}\varphi_{X_k,X_{k+1}}(w,g,f)-\bar{w}]}. \tag{21}$$

In what follows we show that under certain conditions it is possible to choose w_i's and g such that for stationary policy $\pi \sim (f)$ and any $i \in \mathcal{I}$

$$\sum_{j\in\mathcal{I}} p_{ij}(f_i)\,\mathrm{e}^{\gamma[r_{ij}+d_i(f_i)r(i)+w_j(f)]} = \mathrm{e}^{\gamma[d_i(f_i)g(f)+w_i(f)]} \quad\text{or}\quad \mathsf{E}_i^\pi \mathrm{e}^{\gamma\,\varphi_{X_k,X_{k+1}}(w,g,f)} = 1. \tag{22}$$

Moreover, on introducing new variables

$$v_i(f) := \mathrm{e}^{\gamma w_i(f)}, \quad \rho(f) := \mathrm{e}^{\gamma g(f)}, \quad q_{ij}(f) := p_{ij}(f)\mathrm{e}^{\gamma[r_{ij}+d_i(f_i)r(i)]} \tag{23}$$

from (22) we arrive at the following set of equations

$$\sum_{j\in\mathcal{I}} q_{ij}(f_i)\,v_j(f) = \rho(f)^{[d_i(f_i)]}\cdot v_i(f). \qquad (i\in\mathcal{I}) \tag{24}$$

Observe that if all $d_i(f_i)$'s are equal to some constant, say d, then (24) is a well-known formula for finding spectral radius (or so called Perron eigenvalue) of a nonnegative matrix, $v_i(f)$'s are elements of the corresponding Perron eigenvector (cf. [5]). In particular, if for the all $i \in \mathcal{I}$ and $f \in \mathcal{F}$ the values $d_i(f_i)$'s are equal to one the considered semi-Markov reward process is reduced to a Markov reward chain and Eq.(24) to formulas for calculating γ-risk average reward/cost optimality equation of the Markov reward chain. Unfortunately, comparing with the risk-neutral model, unichain property itself (cf. Assumption GA), cannot guarantee positivity of the Perron eigenvector; however, Perron eigenvector is strictly positive if the respective matrix is irreducible.

In what follows we focus attention on finding stationary policies $\pi^* \sim (f^*)$, resp. $\hat{\pi} \sim (\hat{f})$, such that for any $f \in \mathcal{F}$ it holds $\rho(f^*) \geq \rho(f)$, resp. $\rho(\hat{f}) \leq \rho(f)$. We show that the policies $\pi^* \sim (f^*)$, resp. $\hat{\pi} \sim (\hat{f})$, can be found by policy iterations. To specify the policy iteration algorithm, it will be convenient to use the following matrix notation.[1]

[1] In vector inequalities $a \geq b$ denotes that $a_i \geq b_i$ for all elements of the vectors a, b, and $a_i > b_i$ at least for one i, but not for all i's, and $a > b$ if and only if and $a_i > b_i$ for all i's. Using matrix notations the symbol I is reserved for identity matrix, e denotes unit (column) vector.

On introducing the $N \times N$ matrices $Q(f) = [q_{ij}(f_i)]$ and (column) N-vector $v(f) = [v_i(f)]$ along with diagonal $N \times N$ matrix $B(f) = \mathrm{diag}[\rho(f)^{[d_i(f_i)]}]$, from (24) we get

$$B(f) \cdot v(f) = Q(f) \cdot v(f) \quad \Longleftrightarrow \quad v(f) = [B(f)]^{-1} \cdot Q(f) \cdot v(f). \tag{25}$$

Obviously, from (25) for the i-element of $v(f)$ it holds $v_i(f) = \rho(f)^{-d_i(f_i)} \cdot \sum_{j=1}^{N} q_{ij}(f_i) \cdot v_j(f)$.

If stationary policy $\pi \sim (f)$ is followed, policy improvement routine can be used for finding an improved decision in any state (such approach slightly extends policy iteration method reported in [7] for finding maximal possible spectral radius of a family of controlled ergodic Markov reward chains).

a) $\sum_{j=1}^{N} q_{ij}(h_i) \cdot v_j(f) \geq v_i(f)$ if maximal $\rho(f)$ is seeking, resp.,

b) $\sum_{j=1}^{N} q_{ij}(h_i) \cdot v_j(f) \leq v_i(f)$ if minimal $\rho(f)$ is seeking.

Repeating the above procedure we can generate a sequence of stationary policies with increasing, resp. decreasing, sequence of the values $\rho(f)$'s converging to maximal, resp. minimal, value of $\rho(f)$'s.

To this end, since $Q^{(B)}(f) := [B(f)]^{-1} \cdot Q(f)$ is an irreducible nonnegative matrix, let for some $h \in \mathcal{F}$ $z^{(B)}(h)$ be the left Perron eigenvector of $Q^{(B)}(h)$, i.e. $z^{(B)}(h) \cdot Q^{(B)}(h) = \rho^{(B)}(h) \cdot z^{(B)}(h)$. Since the matrix $Q^{(B)}(\cdot)$ is irreducible, the vectors $z^{(B)}(h)$, $v(f)$ are strictly positive, hence also their product $z^{(B)}(h) \times v(f)$ is positive.

In particular, let

$$\psi(h,f) := [Q^{(B)}(h) - Q^{(B)}(f)] \cdot v(f), \qquad \bar{\psi}(h,f) := B(f) \cdot [Q^{(B)}(h) - Q^{(B)}(f)] \cdot v(f)$$

be column N-vectors with elements $\psi_i(h,f)$, resp. $\bar{\psi}_i(h,f)$. Then by (25) we can conclude that

$$Q^{(B)}(h) \cdot v(h) - Q^{(B)}(f) \cdot v(f) = Q^{(B)}(h) \cdot v(h) - Q^{(B)}(h) \cdot v(f) + \bar{\psi}(h,f)$$

and on premultiplying by $z^{(B)}(h)$ we can conclude that

$$[\rho^{(B)}(h) - \rho^{(B)}(f)] \cdot z^{(B)}(h) \cdot v(f) = z^{(B)}(h) \cdot \bar{\psi}(h,f).$$

Then $\bar{\psi}(h,f) > 0$ implies that $\rho^{(B)}(h) > \rho^{(B)}(f)$, if $\bar{\psi}(h,f) < 0$ then $\rho^{(B)}(h) < \rho^{(B)}(f)$.
Observe that if all $d_i(f)$'s are equal to one (or at least to some constant number) the value $\rho(f)$ in (25) is the Perron eigenvalue (equal to the spectral radius) of the nonnegative matrix $Q(f) = [q_{ij}(f)]$. To this end we can expect that for not too much different elements of the diagonal matrix $B(f)$ if $\bar{\psi}(h,f) > 0$, resp. $\bar{\psi}(h,f) < 0$, then also $\psi(h,f) > 0$, resp. $\psi(h,f) < 0$.

Repeating the above improvements, we arrive at stationary policies $\pi^* \sim (f^*)$, resp. $\hat{\pi} \sim (\hat{f})$, such that

$$v_i(f^*) = \max_{f \in \mathcal{F}} \sum_{j \in \mathcal{I}} q_{ij}(f_i)\, [\rho(f)]^{[-d_i(f_i)]} v_j(f_i) = \sum_{j \in \mathcal{I}} q_{ij}(f_i^*)\, [\rho(f^*)]^{[-d_i(f_i^*)]} v_j(f_i^*) \tag{26}$$

$$v_i(\hat{f}) = \min_{f \in \mathcal{F}} \sum_{j \in \mathcal{I}} q_{ij}(f_i)\, [\rho(f)]^{[-d_i(f_i)]} v_j(f_i) = \sum_{j \in \mathcal{I}} q_{ij}(\hat{f}_i)\, [\rho(\hat{f})]^{[-d_i(\hat{f}_i)]}\, v_j(\hat{f}_i) \tag{27}$$

Moreover, from (26), (27), we get the following set of nonlinear equations

$$\begin{aligned} \mathrm{e}^{\gamma w_i^*} &= \max_{f \in \mathcal{F}} \sum_{j \in \mathcal{I}} p_{ij}(f_i)\, \mathrm{e}^{\gamma [r_{ij} + d_i(f_i)[r(i) - g(f)] + w_j(f)]} \\ &= \mathrm{e}^{\gamma d_i(f_i^*)[r(i) - g(f^*)]} \sum_{j \in \mathcal{I}} p_{ij}(f_i^*)\, \mathrm{e}^{\gamma [r_{ij} + w_j(f^*)]} \quad (i \in \mathcal{I})) \end{aligned} \tag{28}$$

$$\begin{aligned} \mathrm{e}^{\gamma \hat{w}_i} &= \min_{f \in \mathcal{F}} \sum_{j \in \mathcal{I}} p_{ij}(f_i)\, \mathrm{e}^{\gamma [r_{ij} + d_i(f_i)[r(i) - g(f)] + w_j(f)]} \\ &= \mathrm{e}^{\gamma [d_i(\hat{f}_i)[r(i) - g(\hat{f})]} \sum_{j \in \mathcal{I}} p_{ij}(\hat{f}_i)\, \mathrm{e}^{\gamma [r_{ij} + w_j(\hat{f})]} \quad (i \in \mathcal{I}) \end{aligned} \tag{29}$$

Eqs. (28),(29) can be called the γ-*risk average reward/cost optimality equation* for semi-Markov processes.

Similar results can be also formulated for the corresponding certainty equivalents, see (2), (3). To this end, let $Z_i^{\pi}(\gamma, n) := \ln U_i^{\pi}(\gamma, n)$ hence in virtue of (20), (21) we can conclude from Eqs. (28),(29) that for $\hat{\pi} \sim (\hat{f})$,

$\pi^* \sim (f^*)$ and any $\pi = (f^n)$

$$Z_i^{\pi}(\gamma, n) = \mathsf{E}_i^{\pi} \sum_{k=0}^{n-1} [d_{X_k}(f_{X_k}) \cdot r(X_k) + r_{X_k, X_{k+1}}] = \sum_{j \in \mathcal{I}} q_{ij}(f_i) \, [\rho(f)]^{[-d_i(f_i)]} \, v_j(f_i) \tag{30}$$

$$Z_i^{\hat{\pi}}(\gamma, n) \leq Z_i^{\pi}(\gamma, n) \leq Z_i^{\pi^*}(\gamma, n) \tag{31}$$

Special case: continuous-time Markov chain.

Let us consider continuous-time Markov decision chain with transition intensities $\mu_{ij}(f_i)$, for $i, j \in \mathcal{I}, j \neq i$, $\sum_{j \in \mathcal{I}, j \neq i} \mu_{ij}(f_i) = -\mu_{ii}(f_i)$ and $\mu_i(f_i) = -\mu_{ii}(f_i)$ is the intensity of jumps from state i. Obviously, this is a very special case of semi-Markov processes with transition probabilities $p_{ij}(f_i) = \frac{\mu_{ij}(f_i)}{\mu_i(f_i)}$, for $j \neg i$, $p_{ii}(f_i) = 0$ and expected holding time $d_i(f_i) = \frac{1}{\mu_i(f_i)}$ in state i. Then by (22) for the considered continuous-time Markov process after some manipulation we conclude that

$$\sum_{j \in \mathcal{I}, j \neq i} \mu_{ij}(f_i) \, \mathrm{e}^{\gamma [r_{ij} + \frac{1}{\mu_i(f_i)} r(i) + w_j(f)]} = \mu_i(f_i) \cdot \mathrm{e}^{\gamma [\frac{1}{\mu_i(f_i)} g(f) + w_i(f)]}. \tag{32}$$

Moreover, on introducing new variables (recall that $d_i(f_i) = \frac{1}{\mu_i(f_i)}$)

$$\bar{v}_i(f) := \mathrm{e}^{\gamma w_i(f)}, \quad \bar{\rho}(f) := \mathrm{e}^{\gamma g(f)}, \quad \bar{q}_{ij}(f) := \mu_{ij}(f) \mathrm{e}^{\gamma [r_{ij} + d_i(f_i) r(i)]} \tag{33}$$

from (32) we arrive at the following set of equations

$$\sum_{j \in \mathcal{I}, j \neq i} \bar{q}_{ij}(f_i) \, \bar{v}_j(f) = \bar{\rho}(f)^{[d_i(f_i)]} \cdot \bar{v}_i(f) \qquad (i \in \mathcal{I}). \tag{34}$$

Acknowledgements

This research was supported by the Czech Science Foundation under Grant 18-02739S.

References

[1] Bertsekas, D. P. (2007). *Dynamic Programming and Optimal Control, Volume 2, Third Edition.* Belmont, Mass.: Athena Scientific.

[2] Cavazos-Cadena, R. & Montes-de-Oca, R. (2003). The value iteration algorithm in risk-sensitive average Markov decision chains with finite state space, *Math. Oper. Res.*, 28, 752–756.

[3] Cavazos-Cadena, R. (2003). Solution to the risk-sensitive average cost optimality equation in a class of Markov decision processes with finite state space, *Math. Methods Oper. Res.,* 57, 253–285.

[4] Cavazos-Cadena, R. & Hernández-Hernández, D. (2005). A characterization of the optimal risk-sensitive average cost infinite controlled Markov chains, *Ann. Appl. Probab.,* 15, 175–212.

[5] Gantmakher, F. R. (1959). *The Theory of Matrices.* London: Chelsea.

[6] Howard, R. A. (1960). *Dynamic Programming and Markov Processes.* Cambridge, Mass.: MIT Press.

[7] Howard, R. A. & Matheson, J. (1972). Risk-sensitive Markov decision processes, *Manag. Sci.*, 23 , 356–369.

[8] Mandl, P. (1971). On the variance in controlled Markov chains, *Kybernetika*, 7, 1–12.

[9] Puterman, M. L. (1994). *Markov Decision Processes – Discrete Stochastic Dynamic Programming*. New York: Wiley.

[10] Ross, S. M. (1983). *Introduction to Stochastic Dynamic Programming.* New York: Academic Press.

[11] Sladký, K. (1974). On the set of optimal controls for Markov chains with rewards, *K*ybernetika, 10, 526–547.

[12] Sladký, K. (1980). Bounds on discrete dynamic programming recursions I, *Kybernetika*, 16, 526–547.

[13] Sladký, K. (2005). On mean reward variance in semi-Markov processes, *Math. Methods Oper. Res.*, 62, 387–397.

[14] Sladký, K. (2008). Growth rates and average optimality in risk-sensitive Markov decision chains, *Kybernetika*, 44, 205–226.

[15] Sladký, K. (2012). Risk-sensitive and average optimality in Markov decision processes. In J. Ramík & D. Stavárek (Eds.), *Proc. 30th Internat. Conference Mathematical Methods in Economics 2012, Part II* (pp. 799–804). Karviná: Silesian University, School of Business Administration.

[16] Sladký, K. (2018). Risk-sensitive average optimality in Markov decision processes, *Kybernetika*, 54, 1218–1230.

[17] van Dijk, N. M. & Sladký, K. (2006). On the total reward variance for continuous-time Markov reward chains, *J. Appl. Probab.*, 43, 1044–1052.

Implications of Population Aging for Economic Growth in V4 Countries

Nikola Soukupová[1], Jana Klicnarová[2], Markéta Adamová[3]

Abstract. Population aging is a typical feature of population development in Europe. The workforce is decreasing, while the 65+ population is increasing. This fact presents challenges for economies, public finance, health, and pension systems. The post-socialist countries (including V4 countries) are relatively younger than the European average; their age transformation is running – a delay of 30 years is expected, but with faster and intensive impact. The paper examines the impact of V4 countries' population aging on their economic growth. The population aging is represented by the total dependency ratio and gross domestic product per capita as proxies for economic growth. Correlation and regression analysis is applied to model the dependence of the growth of gross domestic product on population factors; data from 1993–2018 are used.

Keywords: aging, total-age dependency ratio, economic growth, V4 countries

JEL Classification: C32, J11
AMS Classification: 90C15

1 Introduction

In the 20th century, there is a significant rise in life expectancy. Increasing life expectancy combined with declining birth rates have led to a structural change of the age, called population aging [21], [17]. The aging is the result of demographic processes, where the life expectancy is increasing. The level of fertility is low; therefore, the share of population 65+ exceeds the share young age group [26]. Population aging in industrialized countries has been identified as the main topic regarding future economic development [25].

Population aging will lead to many negative impacts (decreasing social security and problems with public healthcare service, utilization of existing infrastructure, and human capital) [6]. Population aging will lead to lower workforce participation and savings rates, thus raising anxieties about a future slowing of economic growth [17]. On the other hand, it is expecting an improvement in the employment of the older population and the women [7].

"Population aging has been a major demographic trend in Europe" [13]. Especially EU11 countries are more endangered to the negative consequence of an aging population than other countries in Europe [17]. However, special attention must be paid to the countries of Central and Eastern Europe, where significant changes in the political, economic and social life after the democratic changes in 1989 meant that the aging of the population in this area takes place with an interval of approximately 20-30 years, but faster and more intensely [21]. The post-socialist countries (including e.g., V4 countries) are relatively younger than the European average, but their age transformation is running [23].

Current population development in V4 countries is undergoing many demographic changes – especially population aging [21]. In the development of all V4 countries, we can recognize many similarities (history, economic level, location, culture, political changes, and the development after II World War). However, there are some economic and political differences, but their demographic behaviour was influenced by a similar population policy [21], [23]. V4 countries have a combined population of 64.3 million inhabitants [21]. For the next years, 2020–2021, the are many prognoses. For the Czech Republic, economic growth is projected, and there is a problem of labour shortage. In Slovakia, economic growth is expected to slow down; hence it is necessary to solve the challenge of population aging. A robust economic recovery is forecasted for Hungary; they have to deal with labour shortage, too. Poland expects strong economic growth

[1] University of South Bohemia in České Budějovice, Department of Management, Studentská 13, 370 05 České Budějovice, nsoukupova@ef.jcu.cz.
[2] University of South Bohemia in České Budějovice, Department of Applied Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, janaklic@ef.jcu.cz.
[3] University of South Bohemia in České Budějovice, Department of Management, Studentská 13, 370 05 České Budějovice, adamovam@ef.jcu.cz.

and a decreasing unemployment rate [19]. Based on these trends, we can see that V4 countries face some challenges within labour market in a situation when their economic growth will be increasing. According to the global competitiveness index 2018, the Czech Republic has the highest score (70.85), next is Poland (68.89), then Slovakia (66.77); the worst result has Hungary (66.08) [19]. So, we can see that V4 economies are similar and comparable.

Population aging in V4 countries is not homogenous; it means there is a different level of population aging [19]. But for the future, they should have even similar state about population aging.

The World Health Organization (WHO) defines an aging society as society where more than 7% of the population is aged 65 years or above, an aged society as one in which this age group accounts for more than 14% of the total population, and a hyper-aged society as society wherein this rate is greater than 20% [9].

Country	7%	14%	20%
Czech Republic	1945[4]	2005	2020
Hungary	1941	1993	2020
Poland	1966	2012	2030
Slovakia	1962	2015	2027

Table 1 Time trend for population aging in V4 countries

"Aging populations are rapidly transforming economies by shifting demand for both goods and services to other sectors of the economy" [5]. Many researches proved the importance of age structure on economic growth [1, 4, 11, 12, 19, 15, 16, 20, 22, 24].

In many studies [9, 3, 20] the old-age dependency ratio is used as the standard indicator of population ageing – *"it takes the number of people who have reached the state pension age and divides it by the number of working age (16–64 years) adults in order to estimate the proportion of older people relative to those who pay for them"* [27].

The aim of this paper is to determine how population aging is related to economic growth as measured by real GDP per capita in V4 countries. This study is to address the following questions: how is the old-age dependency ratio related to economic growth? This study answers this question in relation to V4 countries.

2 Materials and methods

This paper examines the impact of V4 countries' population aging on their economic growth. The population aging is represented by the total dependency ratio – the sum of the youth-dependency ratio (hereinafter YADR) and old-dependency ratio (hereinafter OADR) and gross domestic product per capita (hereinafter GDPPP) as proxies for economic growth.

$$\text{Total dependency} = \left(\frac{\text{Population under 15 years} + \text{Population 65 years and over}}{\text{Population 15 to 64}}\right) \times 100 \tag{1}$$

$$\text{Old-age dependency} = \left(\frac{\text{Population 65 years and over}}{\text{Population 15 to 64}}\right) \times 100 \tag{2}$$

$$\text{Youth-age dependency} = \left(\frac{\text{Population under 15 years}}{\text{Population 15 to 64}}\right) \times 100 \tag{3}$$

$$\text{GDPPP} = \left(\frac{\text{Real GDP}}{\text{Population}}\right) \tag{4}$$

$$\text{GEAP} = \left(\frac{\text{Population 15 to 64}}{\text{Population}}\right) \times 100 \tag{5}$$

Variable GDPPP is often used for measuring of economic growth in many studies [12], [15], [16], [20] because "*GDP per capita is especially useful for making cross country comparisons as well as for making comparisons over time*" [14].

[4] Estimated.

The old-age dependency ratio (OADR) is the ratio of the number of elderly people at an age when they are generally economically inactive (i.e. aged 65 and over), compared to the number of people of working age (i.e. 15–64 years old). The OADR represented basic variable for population aging. The age limit of 65+ is the most used in aging research within Central East Europe [26]. Subnational proportional age structures were then used to calculate combined dependency ratios for young-age and old-age for each of V4 countries.

Due to [3], we add the growth rate of economically active population (GEAP) for methodical correctness. The variable definitions show Table 2.

Variable	Definition	Period
ODAR	Old-age dependency ratio: population aged 65 years and above divided by population aged 15–64 years (%)	1993–2018
YADR	Youth-age dependency ratio: population aged 0–14 years divided by population aged 15–64 years (%)	1993–2018
GEAP	Growth rate of economically active population (%)	1993–2018
GDPPP	Gross Domestic Product (GDP) per capita (NT$/person)	1993–2018

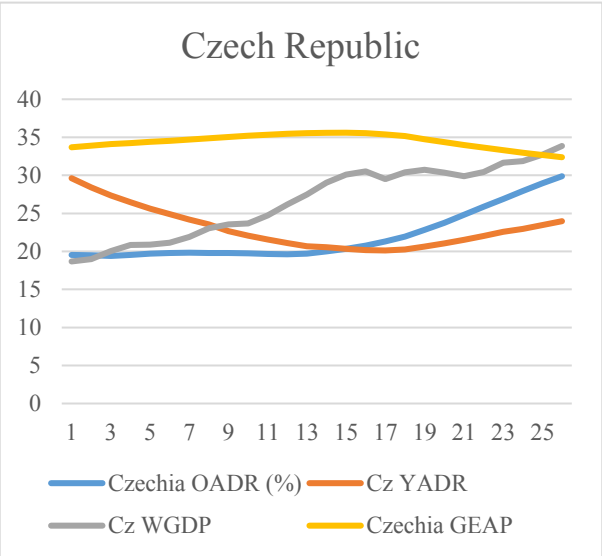
Table 2 Variable Definitions

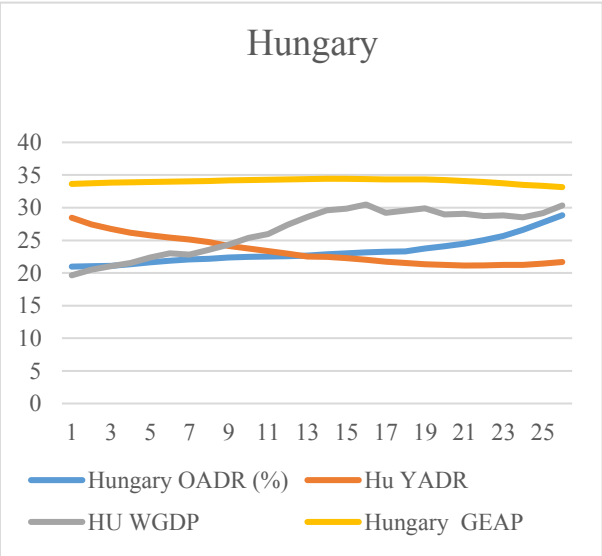
This paper examines data from 1993–2018. Data sources include EUROSTAT, OECD, WorldBank, and National Statistical Offices. The aim of this paper is to determine how population aging is related to economic growth as measured by real GDP per capita in V4 countries. This study is to address the following questions: how is the old-age dependency ratio related to economic growth? This study answers this question in relation to V4 countries.

Before we use correlation and regression analysis to describe the relevance between population aging and economic growth, let us display examined series into graphs. Each graph shows trajectories of examined variables in one of the V4 countries for years from 1993 to 2018 (the units on an axe *y* differ for variables but are the same for different countries).

From the graphs, we can see that there is a change in behaviour of series after the year 2000, since that time in all countries is OADR growing up and GEAP is decreasing in all countries. The starting pressure will reach a peak between 2015 and 2035 when the baby boom generation enters retirement. This fact presents challenges for sustainable public finances, especially the financing of health care and pensions [8], [18]. The baby boom was followed by another population wave, so-called "Husákovy děti" born 1970 in the Czechia and Slovakia. They will be retired too in 2035; thus, it will be intense pressure on these economics [10].

According to the European Commission (2017), the population aging in V4 countries will be increasing. From V4 countries in 2070, Poland (86.7%) and Slovakia (82.7%) will have the highest total age dependency ratio. Czechia will have 75.6% and Hungary will have 78.6% [7], [29].





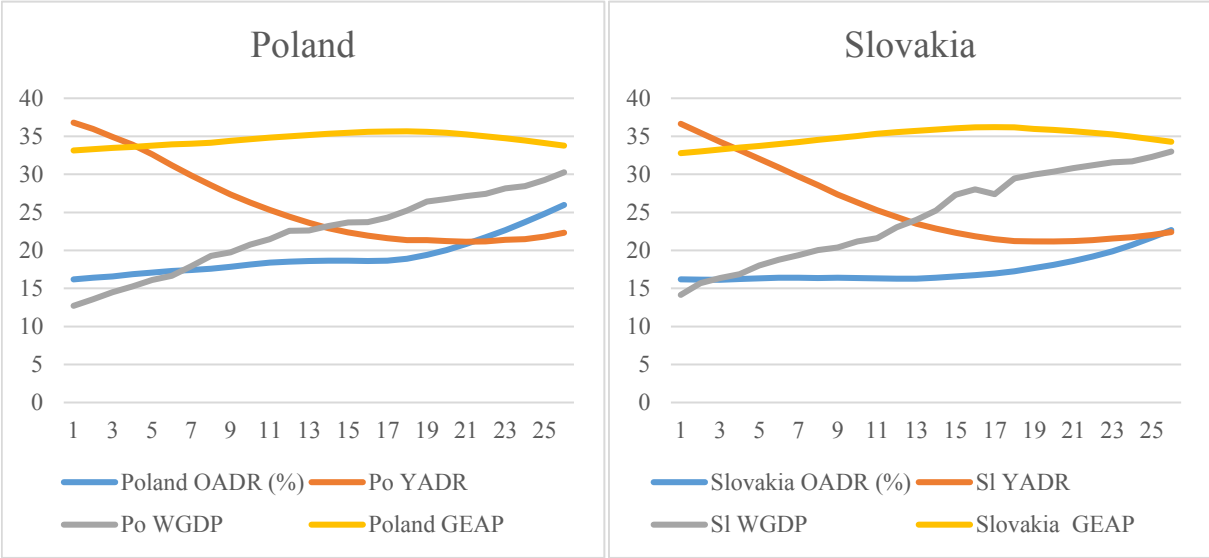


Figure 1 Variables of V4 countries for years from 1993 to 2018

Now, let us apply correlation and regression analysis to model the dependence of the growth of gross domestic product on population factors.

3 Results

First, we supposed to apply methods introduced by Bloom and Williamson in the year 1998 to construct models to describe economic growth by population aging factors. We decided to use their models to the data of V4 countries data in the years 1993–2018. More precisely, we aimed to explain the growth rate of real GDP per capita by GEAP, the growth rate of OADR, and the growth rate of YADR. However, these methods failed. Due to the graphs given above, we tried to change the time of the examination. We constructed the models based only on data between 1993 and 2008 (before the economic crises and the change in the OADR and GEAP behaviour). However, no significant results were achieved, too.

If we applied correlation analysis on variables used by Bloom and Williamson, there was no significant correlation between the logarithm of the rate of GDPPP and the logarithm of the rate of OADR, YADR; nor logarithm of GEAP (for both examined periods). The same results we got if we applied correlation coefficients on original rates (without logarithm).

Hence, since the multiplicative models are supposed in demographics, see e.g. research Huang et al. (2018); Oliver (2015), we decided to check the relationship between our variables in their log-forms. First, we applied correlation analysis. From graphs (Fig. 1), we can see significant linear trend of GDPPP, resp. ln(GDPPP) and trajectories of population aging factors, resp. their logarithms. Therefore, there is probably a significant correlation of examined variables with the time factor. Hence, we decided not to use correlation coefficients (which could be false significant) but to apply a partial correlation coefficient with controlling variable time, resp. ln(time). We supposed the correlation coefficients to be more significant for lagged variables, i.e. when we compare GDPPP (t+1) with population characteristics in time t. However, the results do not confirm such assumption – not depending on the examined period, the correlation coefficients were higher for non-lagged variables. As was supposed from the graphs, correlation coefficients constructed for the period 1993–2008 were the most significant (very often we received coefficients with absolute value in an area of 0.9). In the case of other chosen period, the correlation coefficients stayed significant, but their values were lower.

Based on this analysis, we constructed regression models based on data from 1993 to 2008 as follows:

$$\ln\left(\mathrm{GDPPP}_{\{t\}}\right) = \beta_0 + \beta_1 \ln(\mathrm{OADR}_t) + \beta_2 \ln(\mathrm{YADR}_t) + \beta_3 \ln(\mathrm{GEAP}_t). \tag{6}$$

For all four countries, we got a regression model with a non-significant parameter of the variable GEAP (in fact, it is strongly negatively correlated with OADR and YADR), and we obtained models with R2 (and adjusted R2) in area of 0.98. On the other hand, the partial correlation coefficients were the highest ones between variable ln(GDPPP) and ln(GEAP) – the most significant correlation was achieved for Hungary – 0.96 and the smallest one for Poland 0.79. These results show the strong dependence of gross domestic product

per capita on characteristics of population aging. On the other hand, in these demographics models, the possible non-stationary of examined series is not taken into account, so in fact, the dependence could not be so strong. For more solid modelling, it is necessary to check the stationarity of the series and also to study which other factors and which lags have a significant impact on gross domestic product per capita.

Due to the changes in the last years (which can we see from the graph), we decided not to construct models for the whole period (however, the models are significant). However, we can state that also in V4 countries, there is a dependence GDPPP on population aging factors.

To model the changes in the last years, we tried to apply the model for years 2009 – 2018 (it is not too long period, but it does not include the changing periods). Similarly, to the previous period, the variable ln(GEAP) is not significant except the model for Poland, where all variables are significant, andR^2 (and adjusted R^2) is 0.99. Significant models we also get for Slovakia (R^2=0.97) and the Czech Republic (R^2=0.79), but for Hungary, the model has only R^2=0.45. From the research results, it is clear that the model has not yet managed to stabilize, and long-term changes are still running. The model will be possible to apply in a long-term perspective.

4 Conclusions

These results prove the strong dependence of gross domestic product per capita on characteristics of population aging. These findings are corresponding with previously worldwide studies (Tab. 3).

Authors	Year	Findings
Lindh and Malmberg [15]	1999	The 65+ age group was negatively related to the GDP in 21 OECD countries for the period 1950–1990.
Tang and MacLeod [28]	2006	Older workers are less productive than younger workers; aging has negative impact on productivity growth.
Bloom et al. [2]	2008	The effect of old age on growth is negative in the short run.
Maestas et al. [16]	2016	Population aging decreases the growth rate of GDP per capita.
Pham and Vo [22]	2019	Overall economic performance in development countries could be affected by growth in the share of the elderly.

Table 3 Similar studies to our results

The results showed the similarity of the relationship population aging and GDPPP in V4 countries with comparison with the previous worldwide studies. It could be summarized that also in the V4 countries there is a strong dependence on population aging factors and GDPPP.

The share of workforce (GEAP) and total dependency ration (OADR+YADR) have a negative relationship. According to population projections in V4 countries, the total dependency ration will be increasing, even in 2070 it will reach the following values: Poland (86.7%), Slovakia (82.7%), Czechia, (75.6%) and Hungary (78.6%). This fact will be accompanied by a decline of GEAP, which will put considerable pressure on the economies, health and pension systems of the V4 countries.

Acknowledgements

This contribution was supported by research grant GAJU No. 121/2020/S “Principy cirkulární ekonomiky v regionálním managementu vedoucí ke zvýšení efektivnosti systémů”.

References

[1] BLOOM, D. E., CANNING, D., & FINK, G. (2010). Implications of population ageing for economic growth. *Oxford review of economic policy*, 26(4), 583–612.

[2] BLOOM, D. E. & CANNING, D. (2008). Global demographic change: dimensions and economic significance. *Population and Development Review*, 34, 17–51.

[3] BLOOM, D. E. & WILLIAMSON, J. G. (1998). Demographic transitions and economic miracles in emerging Asia. *World Bank Economic Review*, 12(3), 419–455.

[4] BRUNOW, S. & HIRTE, G. (2006) Age structure and regional growth, *Jahrbuch fur Regionalwissenschaft/Review of Regional Research*, 26(1), 3–23.
[5] CRISTEA, M., NOJA, G. G., STEFEA, P., & SALA, A. L. (2020). The impact of population aging and public health support on EU labor markets. *International Journal of Environmental Research and Public Health*, 17(4), 1439.
[6] EUROPEAN COMMISSION. (2016). Ready for the Demographic Revolution? Measuring Active Ageing. Available at: https://ec.eropa.eu/scial/main.jsp?lagId=en&catId=752&furtherNews=yes&newsId=2430.
[7] EUROPEAN COMMISSION. (2017). The 2018 Ageing Report: Underlying Assumptions and Projection Methodologies. Publications Office of the European Union: Luxembourg.
[8] GYÔRI, L. (2013). *Challenges of Ageing Societies in the Visegrad Countries*. Budapest: Hungarian Charity Service of the Order of Malta.
[9] HUANG, W. H., LIN, Y. J., & LEE, H. F. (2019). Impact of Population and Workforce Aging on Economic Growth: Case Study of Taiwan. *Sustainability*, 11(22), 6301.
[10] JEŘÁBKOVÁ, V. (2009). Zdravotně-sociální péče o seniory v České republice. Reprodukce lidského kapitálu/vzájemné vazby a souvislosti/II. ročník. Vysoká škola ekonomická v Praze. Praha: Oeconomica.
[11] KÁČEROVÁ, M. & ONDAČKOVÁ, J. (2015). The process of population ageing in countries of the Visegrad Group (V4). *Erdkunde*, 69(1), 49–68.
[12] KELLEY, A. C. & SCHMIDT, R. M. (2005), Evolution of recent economic-demographic modeling: a synthesis. *Journal of Population Economics*, 18(2), 275–300.
[13] KULCSÁR, L. J. & BROWN, D. L. (2017). Population Ageing in Eastern Europe: Toward a Coupled Micro-Macro Framework. *Regional Statistics*, 7(1), 115–134.
[14] LEQUILLER, F. (2004) Is GDP a satisfactory measure of growth? *OECD Observer*, No 246–247.
[15] LINDH, T. & MALMBERG, B. (1999). Age Distributions and the Current Account – A Changing Relation? (No. 1999: 21). Working Paper.
[16] MAESTAS, N., MULLEN, K. J., & POWELL, D. (2016). The effect of population aging on economic growth, the labor force and productivity (No. w22452). National Bureau of Economic Research.
[17] MAREŠOVÁ, P., MOHELSKÁ, H., & KUČA, K. (2015). Economics aspects of ageing population. In *Procedia Economics and Finance*, 23 (pp. 534–538).
[18] MUNNELL, A. H. (2004). *Population aging: It's not just the baby boom* (Vol. 16). Boston, MA: Center for Retirement Research at Boston College.
[19] OECD. (2019). Economic Forecast Summary November 2019, Available at: http://www.oecd.org/economy/czech-republic-economic-snapshot/; http://www.oecd.org/economy/poland-economic-snapshot/; https://www.oecd.org/economy/hungary-economic-snapshot/; http://www.oecd.org/economy/czech-republic-economic-snapshot/
[20] OLIVER, M. (2015). Population ageing and economic growth in Japan. *International Journal of Sociology and Social Policy*. 35(11), 841 – 863
[21] ONDAČKOVÁ, J., KÁČEROVÁ, M., MLÁDEK, J., POPJAKOVÁ, D., & VANČURA, M. (2018). Population age structure transformation in the capitals of the Visegrad Group countries. *Geographia Polonica*, 91(3), 281–299.
[22] PHAM, T. N. & VO, D. H. (2019). Aging Population and Economic Growth in Developing Countries: A Quantile Regression Approach. *Emerging Markets Finance and Trade*, 1–15.
[23] PIOTROWICZ, W. (2015). Information technology and systems in the Visegrád group of countries (Czech Republic, Hungary, Poland, and Slovakia): a literature review. *Journal of Global Information Technology Management*, 18(2), 7.
[24] POOT, J. (2008). Demographic change and regional competitiveness: the effects of immigration and ageing. *International Journal of foresight and innovation Policy*. 4(1–2), 129–145.
[25] PRETTNER, K. (2013). Population aging and endogenous economic growth. *Journal of population economics*. 26(2), 811–834.
[26] SPÉDER, Z. & BÁLINT, L. (2013). Aging in the Visegrad countries: selected demographic and sociological aspects. Budapest: Hungarian Charity Service of the Order of Malta.
[27] SPIJKER, J. & MACINNES, J. (2013). Population ageing: the timebomb that isn't? bmj, 347, f6598.
[28] TANG, J. & MACLEOD, C. (2006) Labour force ageing and productivity performance in Canada. *Canadian Journal of Economics*. 39(2), 582–603.
[29] WORLD HEALTH ORGANIZATION (WHO). (2017). Global Strategy and Action Plan. on Ageing and Health. World Health Organization: Geneva, Switzerland.

Efficiency Assessment of the UK Travel Agency Companies – Data Envelopment Analysis Approach

Michaela Staňková[1], David Hampel[2]

Abstract. Thomas Cook travel agency, one of the oldest travel agencies in the world, surprised the whole Europe when it went bankrupt in September 2019. This article focuses on efficiency evaluation of British companies operating in the same industry as Thomas Cook. Data envelopment analysis method was selected to calculate the efficiency. We use annual accounting data of the companies that are available in the Orbis database. Material costs, costs of employees, capital and assets are selected as input variables. Net income, turnover, working capital and solvency ratio represents output variables. Efficiency rating models are compiled for constant returns to scale as well as for variable returns to scale. This makes it possible to evaluate the area of the scale efficiency. All models are radial input oriented and constructed separately for each of the three monitored periods. It was found that the evaluation of efficiency through the data envelopment analysis method revealed some problems of Thomas Cook travel agency in the years preceding bankruptcy. Particularly, this travel agency ranked among companies with the lowest scale efficiency.

Keywords: company accounting data, data envelopment analysis, efficiency, linear programming, UK travel agency sector

JEL Classification: C67, D24
AMS Classification: 90B50, 90C08

1 Introduction

Last year's collapse of one of the most famous and also the oldest travel agency, stirred up the debate whether it was possible to predict this situation. Bankruptcy of Thomas Cook's travel agency surprised many economic analysts, inter alia, because the company did not enter insolvency proceedings, but ended up in liquidation directly. Many studies focus on areas of bankruptcy prediction in which different methods can be found, like in [8, 9, 11 and 13]. Models based on different methods differ in their predictive capabilities (see [14]), but all models have one property in common – the use of financial indicators. Binary classification models typically require information about whether the subject (company) is active (healthy) or can be considered as bankrupted (at a particular point in time). Based on this information, the learning and also construction of a model for the prediction of bankruptcy is made. However, in exploring the travel agency sector in the UK, it raises a problem with the availability of required data. Focusing on commonly used variables such as liquidity ratio, ROA, EBIT margin, profit margin it was found that only active (healthy) companies would remain in the data set. Therefore, it was necessary to look at this issue from a different perspective. It can be assumed that bankrupt companies cannot compete with other companies and are therefore less efficient in their activities. It is therefore possible to focus on efficiency evaluation instead of typical bankruptcy prediction. This is because the evaluation of efficiency is done through other methods and typically it is based on different variables, like in [12]. For this purpose, the data envelopment analysis (DEA) method seems appropriate. Since the DEA method is not primarily used to classify bankruptcy companies, it also does not require the supply of information about which companies can now be considered bankrupted. The use of the DEA method in evaluating efficiency is not even conditioned by the inclusion of bankruptcy companies in the data file, unlike e.g. logistic regression.

DEA method is a technique originally developed in operational research. It is a special case of linear programming, which can be used for efficiency evaluation, like in [12]. The basis of the DEA method is stated in Farrell's work [3]. Later, his ideas were developed and bring more models. The three most known DEA models are the CCR model by Charnes, Cooper, and Rhodes [6]; the BCC model by Banker, Charnes, and

[1] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, michaela.stankova@mendelu.cz.
[2] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, david.hampel.uso@mendelu.cz.

Cooper [1] and the Additive model (and later derived SBM models) by Charnes et al. [5]. DEA is non-parametric method that compares the observed inputs and outputs of each unit with that of the most performing unit in the information dataset. In all DEA models, the efficiency frontier is constructed, which is made up of fully efficient units (companies). In the case of efficiency evaluation, input variables are typical production inputs (such as labour and capital). Output variables include company outputs (i.e. production). Financial variables are often used in studies focused directly on the efficiency of travel agencies. For example, in [4], the number of employees, annual expenses and having service potential were used as inputs and the output was represented by the number of customers served. These variables were also used in the [10], with the difference that average spends per customer were added.

2 Material and Methods

The financial (annual accounting) data was collected from the Orbis database. We deal with companies within the travel agency, tour operator and other reservation service and related activities sector (NACE code 79) from the United Kingdom with data available in years 2014, 2015, 2016 and 2017. With regard to the theory, we chose seven characteristics representing both production and cost characteristics often used for efficiency assessment and one variable that is typically used for bankruptcy prediction with percentage expression, see Table 1. Solvency ratio is a ratio indicator that takes into account the financing strategy of the company and provides information regarding the credit burden of the company (from the inverse point of view it indicates the company's indebtedness). The use of ratio variables is not typical within the DEA method, as it could lead to an ideal (efficiency) state outside the variable's real values. But in assessing the financial health of the company rations are important and widely used. Together with the DEA models even this types of variables can provide important information, see [7] or [13]. Therefore, we decided to keep this indicator in these DEA models.

In Table 1, there are basic characteristics of financial variables used in our analysis. To keep maximum companies in the data set, we chose to use unbalanced dataset. A total of 336 companies in 2014, 387 companies in 2015, 370 companies in 2016 and 118 companies in 2017 were included in the efficiency evaluation. Unfortunately, for 2017 it was possible to obtain data for a much smaller group of companies, which was due to the availability of data in the Orbis database. The possible impact of this problem will be discussed in this article.

		Capital	**Costs of employees**	**Costs of goods sold**	**Total assets**	**Net income**	**Solvency ratio**	**Turnover**	**Working capital**
		Mil EUR	Mil EUR	Mil EUR	Mil EUR	Mil EUR	%	Mil EUR	Mil EUR
	Char.	**(Input)**	**(Input)**	**(Input)**	**(Input)**	**(Output)**	**(Output)**	**(Output)**	**(Output)**
2014	Min	0.00	0.00	0.00	0.01	–151.73	–85.94	0.03	–1262.72
	Median	0.08	2.01	15.76	10.13	0.27	25.64	21.46	–0.14
	Mean	2.03	10.88	90.14	78.26	1.62	28.17	114.05	–6.87
	Max	88.72	1207.42	8643.56	7450.28	208.26	91.86	11042.98	46.06
	T. Cook	88.72	1207.42	8643.56	7450.28	–151.73	4.92	11042.98	–1262.72
2015	Min	0.00	0.00	0.01	0.01	–218.43	–88.50	0.00	–1584.61
	Median	0.09	2.46	16.93	11.80	0.30	27.72	24.73	–0.18
	Mean	3.99	11.87	93.04	85.93	1.61	28.95	118.11	–7.04
	Max	371.73	1202.03	8210.69	8083.16	231.40	96.83	10628.31	56.03
	T. Cook	93.61	1202.03	8210.69	8083.16	31.20	6.18	10628.31	–1584.61
2016	Min	0.00	0.00	0.00	0.01	–23.97	–75.18	0.11	–1529.40
	Median	0.07	2.38	15.23	11.13	0.35	27.07	20.94	–0.23
	Mean	5.52	10.52	78.51	89.50	3.33	27.57	102.62	–6.36
	Max	1336.63	1069.54	6949.08	8062.74	294.96	99.21	9069.57	65.33
	T. Cook	80.13	1069.54	6949.08	8062.74	4.65	4.70	9069.57	–1529.40

	Char.	Capital Mil EUR (Input)	Costs of employees Mil EUR (Input)	Costs of goods sold Mil EUR (Input)	Total assets Mil EUR (Input)	Net income Mil EUR (Output)	Solvency ratio % (Output)	Turnover Mil EUR (Output)	Working capital Mil EUR (Output)
2017	Min	0.00	0.01	0.00	0.01	–408.12	–88.39	0.01	–1613.20
	Median	0.11	2.60	14.28	9.97	0.24	29.01	19.01	–0.11
	Mean	13.38	21.74	154.76	174.46	–0.34	29.97	200.10	–16.88
	Max	964.75	1137.06	7942.43	7499.17	235.46	97.02	10210.88	38.71
	T. Cook	78.22	1137.06	7942.43	7499.17	14.74	4.23	10210.88	–1613.20

Table 1 Minimum, maximum, mean and median of each selected variable in individual years

The DEA models were constructed separately for particular years. To calculate the technical efficiency of companies, input oriented models were selected. The main reason for this choice is that a possible correction for the inefficient units seems to be more appropriate through a reduction in inputs than an increase in outputs in this sector. To calculate the scale efficiency CCR model[3]

$$\begin{aligned}
\max \quad & E_H = \sum_{j=1}^{n} v_{jH} y_{jH}, \\
\text{subject to} \quad & \sum_{i=1}^{m} u_{iH} x_{iH} = 1, \\
& -\sum_{i=1}^{m} u_{iH} x_{ik} + \sum_{j=1}^{n} v_{jH} y_{jk} \leq 0, \forall k = 1,2,\dots,p, \\
& u_{iH} \geq \varepsilon, \forall i = 1,2,\dots,m, \\
& v_{jH} \geq \varepsilon, \forall j = 1,2,\dots,n,
\end{aligned}$$

with the constant returns to scale (CRS) assumption and also input BCC model[4]

$$\begin{aligned}
\max \quad & E_H = \sum_{j=1}^{n} v_{jH} y_{jH} + \mu_H, \\
\text{subject to} \quad & \sum_{i=1}^{m} u_{iH} x_{iH} = 1, \\
& -\sum_{i=1}^{m} u_{iH} x_{ik} + \sum_{j=1}^{n} v_{jH} y_{jk} + \mu_H \leq 0, \forall k = 1,2,\dots,p, \\
& u_{iH} \geq \varepsilon, \forall i = 1,2,\dots,m, \\
& v_{jH} \geq \varepsilon, \forall j = 1,2,\dots,n, \\
& \mu_H \text{ free},
\end{aligned}$$

assuming variable returns to scale (VRS) was selected. Scale efficiency can be expressed as the ratio of the CRS efficiency to the VRS efficiency. Technical details about all above stated models can be found in [2]. All calculations were performed in computational system MATLAB R2019a and DEA SolverPro version 15.

[3] The model is constructed for unit H, which in one of the p units. Input resp. output variable is arranged in matrix $X = \{x_{ik}, i = 1, 2, \dots, m, j = 1, 2, \dots, p\}$ resp. $Y = \{y_{ik}, i = 1, 2, \dots, n, j = 1, 2, \dots, p\}$. The ε is the so-called infinitesimal constant.

[4] The model is constructed for unit H, which in one of the p units. Input resp. output variable is arranged in matrix $X = \{x_{ik}, i = 1, 2, \dots, m, j = 1, 2, \dots, p\}$ resp. $Y = \{y_{ik}, i = 1, 2, \dots, n, j = 1, 2, \dots, p\}$. The ε is the so-called infinitesimal constant and μ_H represents the magnitude of the deviation from the constant returns to scale.

3 Results

Technical efficiency results under CRS and VRS during the whole time period are summarized in Table 2. It has been found that the most companies are at least 50% efficient. When comparing the results within the VRS and CRS models, it can be stated that under the VRS assumption, the number of fully efficient companies is more than doubled. In 2017, the growth of fully efficient companies is apparent. However, this growth may be due to a reduced number of DMUs in 2017.

	2014		2015		2016		2017	
Efficiency (%)	**CRS (%)**	**VRS (%)**	**CRS (%)**	**VRS (%)**	**CRS (%)**	**VRS (%)**	**CRS (%)**	**VRS (%)**
⟨0; 20)	8.33	5.95	7.49	3.17	5.95	1.89	7.63	3.39
⟨20; 40)	5.36	3.87	5.48	4.61	4.86	4.86	8.47	6.78
⟨40; 60)	8.33	6.25	9.80	7.20	7.30	5.68	3.39	4.24
⟨60; 80)	25.00	15.18	27.67	17.29	31.62	20.00	13.56	6.78
⟨80; 100)	40.18	39.58	38.04	36.02	37.30	33.51	46.61	35.59
100	12.80	29.17	11.53	31.70	12.97	34.05	20.34	43.22

Table 2 Distribution of companies' efficiency under CRS and VRS in all periods

In Table 3, the basic characteristics for the calculated sectoral efficiency score are summarized together with the Thomas Cook's travel agency's results. In models where the assumption of CRS has been done, Thomas Cook's travel agency has always been identified as inefficient. Until 2017, this travel agency obtained an efficiency score lower than the industry average. Furthermore, it was found that throughout the reporting period, this travel agency has a lower efficiency than the median but higher efficiency than the lower quartile. It can therefore be stated that more than 50% of companies in this sector are in a better position than Thomas Cook's travel agency. Conversely, there are more than a quarter of companies in this sector that operate even less effectively than Thomas Cook's travel agency.

When evaluating the situation of Thomas Cook's travel agency based on VRS models, the situation is different. Unlike VRS models, this travel agency has been identified as efficient throughout the period. This score is higher than the average and median in this sector. However, as already stated based on Table 2 values, many other companies in this type of model have achieved the same efficiency score.

	2014		2015		2016		2017	
Char.	**CRS (%)**	**VRS (%)**	**CRS (%)**	**VRS (%)**	**CRS (%)**	**VRS (%)**	**CRS (%)**	**VRS (%)**
Mean	73.06	80.34	72.75	81.79	73.84	82.90	77.10	85.10
St. deviation	26.83	24.54	25.33	22.52	23.61	20.32	28.20	23.77
Lower quartile	66.38	77.44	64.40	75.75	67.71	76.32	72.25	82.20
Median	81.02	87.82	79.67	88.56	80.03	87.03	87.82	96.30
Upper quartile	90.77	100.00	89.40	100.00	88.02	100.00	96.34	100.00
T. Cook	71.53	100.00	68.37	100.00	68.74	100.00	80.44	100.00

Table 3 Mean, standard deviation and quartiles of sector efficiency under CRS and VRS compared to Thomas Cook company results in all periods

Based on efficiency results with different assumptions about scale returns, scale efficiency values can be calculated as the ratio of the CRS efficiency to the VRS efficiency. The results of this type of efficiency are summarized in Table 4. Most companies were found to be scale inefficient. Between 2014 and 2016, this represents only about 14% of companies whose size of operations can be considered optimal. Most companies in this sector are scale inefficient, but over half of them have an efficiency score over 80%.

Given the fact that the Thomas Cook's travel agency was considered VRS-efficient in models with VRS but not CRS-efficient in models with CRS, it can be stated that this travel agency is scale inefficient, see Table 5. The scale inefficiency is due to the company dimension. Thomas Cook's travel agency should adjust the

volume of its activities to make it more productive. It has been found that the Thomas Cook's travel agency is performing poorly in terms of scale efficiency in comparison to other companies in this sector. In all reporting periods, this travel agency ranked among 25% of the companies with the lowest scale efficiency.

Scale efficiency %	2014 (%)	2015 (%)	2016 (%)	2017 (%)
⟨0; 20)	2.08	2.88	2.70	4.24
⟨20; 40)	4.17	0.29	2.43	3.39
⟨40; 60)	2.08	4.32	2.70	2.54
⟨60; 80)	7.74	8.65	7.84	0.85
⟨80; 100)	69.94	70.61	70.81	65.25
100	13.99	13.26	13.51	23.73

Table 4 Distribution of companies' scale efficiency in all periods

Char.	2014 (%)	2015 (%)	2016 (%)	2017 (%)
Mean	89.25	89.23	88.04	88.95
St. deviation	20.07	19.42	19.87	23.29
Lower quartile	89.38	86.36	86.99	93.61
Median	97.54	97.93	95.74	96.70
Upper quartile	99.92	99.96	98.86	99.92
T. Cook	71.53	68.37	68.74	80.44

Table 5 Mean, standard deviation and quartiles of sector scale efficiency compared to Thomas Cook company results in all periods

The radial measurement procedure in selected models focuses on proportional change (reduction) in all inputs for inefficient units (companies). However, it is possible to look at the problem from the perspective of so-called slacks (i.e. surpluses on the input side or shortages on the output side). These slacks are needed to push the inefficient companies to the frontier. Companies on the frontier have slacks equal to zeros, but inefficient companies have some nonnegative slacks. These slacks are individual for each variable and each unit. Number of identified slacks in individual models in all periods is presented in Table 6. In this sector, working capital slack was most often identified. However, the values of this slack ranged from thousands of Euros to several million Euros. On the other hand, turnover slack was the least frequent, but there was always a shortage of tens or hundreds of millions of Euros.

	Model	Capital (Input)	Costs of employees (Input)	Costs of goods sold (Input)	Total assets (Input)	Net income (Output)	Solvency ratio (Output)	Turnover (Output)	Working capital (Output)
2014	CRS	0.35	0.06	0.03	0.03	0.59	0.58	0.01	0.59
	VRS	0.45	0.54	0.09	0.17	0.73	0.49	0.06	0.62
2015	CRS	0.31	0.06	0.01	0.00	0.65	0.43	0.01	0.59
	VRS	0.43	0.23	0.10	0.15	0.66	0.42	0.07	0.64
2016	CRS	0.21	0.06	0.01	0.03	0.67	0.59	0.01	0.58
	VRS	0.53	0.24	0.10	0.17	0.81	0.48	0.08	0.60
2017	CRS	0.26	0.15	0.07	0.04	0.66	0.67	0.04	0.56
	VRS	0.26	0.36	0.17	0.19	0.76	0.69	0.11	0.53

Table 6 Relative frequency of slacks in individual models in all periods

Within the VRS assumed models, the Thomas Cook's travel agency had only zero slacks as it was on the efficiency frontier. Based on slacks from CRS models it can be stated that Thomas Cook's travel agency should change more things. The surpluses on the capital side were identified in all periods. This is a reduction in capital of tens of millions of Euros. Deficiencies in turnover, net income and solvency ratio variables also occurred in all periods. However, the size of these slacks varies from period to period.

4 Discussion

In the case of Thomas Cook's travel agency there were problems with the level of capital as well as with the solvency ratio. The capital structure is usually planned with regard to the company's strategy. In view of these findings, it can be concluded that this travel agency has chosen a financial (capital) strategy that has not led to efficient operation compared to its competitors. The influence of the choice of financing strategy was also demonstrated in the construction industry in article [12]. Examination of slacks also revealed that although there were no shortcomings in turnover at any time, so in the case of net income. This finding points to the problem of paying generous bonuses, which economists today report as one of the big problems of company.

Unfortunately, in 2017 it was possible to obtain data for a smaller group of companies than in previous years for analysis. And in 2017 there are some differences – for example, a larger number of effective units have been identified here, or in that year, the scale efficiency Thomas Cook's travel agency was very close to the industry average. Reducing the data file could have reshaped the efficiency frontier to such an extent that it had an impact on the increased number of efficient companies. However, the scale efficiency values indicate that the model reflects the economic situation of the time at least for the Thomas Cook's travel agency. In 2017 Thomas Cook announced the sale of its Belgian airline operations to Lufthansa. As a result, Thomas Cook Airlines Belgium was shut down by November 2017. It is this sale that has moved this company closer to the most productive scale size.

5 Conclusion

In this article, the DEA method was used to calculate the efficiency of the companies in the travel agency sector in the UK to assess the situation of the Thomas Cook's travel agency. It was found that although standard approaches to bankruptcy prediction could not be used due to the unavailability of data, the efficiency evaluation could bring substantial information in this area. The results of this paper revealed problems in particular with the volume of Thomas Cook's travel agency activities. Its size of operations is not optimal; modifications on its size can render this scale inefficient company more efficient. In addition, it was found that the travel agency did not apply the appropriate capital and financial strategy.

It was found that, based on an efficiency evaluation of the UK travel sector; it was possible to detect signals that the Thomas Cook's travel agency was not working properly. However, there is still room for further research. It would be possible to look at the changes in efficiency in each year through the Malmquist index and examine whether the Thomas Cook's travel agency is moving away from the efficiency frontier. For a more insight into the inefficiency of this company, it would be helpful to extend the period under review. Unfortunately, there are no studies to compare the results for this sector so specified with DEA models. However, the fact that the Thomas Cook's travel agency achieves below-average technical efficiency and lags behind in particular in terms of scale efficiency these models can be considered as a real picture of the situation.

Acknowledgements

This article was supported by the grant No. PEF/TP/2020002 of the Grant Agency IGA PEF MENDELU.

References

[1] BANKER, R., CHARNES, A., & COOPER, W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, *30(9),* 1078–1092.

[2] COOPER, W. W., SEIFORD, L. M., & TONE, K. (2007). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software*. 2nd edition. New York: Springer Science & Business Media.

[3] FARRELL, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, 120(3),* 253–290.
[4] FUENTES, R. (2011). Efficiency of travel agencies: A case study of Alicante, Spain. *Tourism Management, 31(1),* 75–87.
[5] CHARNES, A., COOPER, W. W., GOLANY, B., SEIFORD, L., & STUTZ, J. (1985). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, *30(1–2),* 91–107.
[6] CHARNES, A., COOPER, W., & RHODES, E. (1987). Measuring the efficiency of decision making units. *Management science*, 2(6), 429–444.
[7] JANOVÁ, J., VAVŘINA J., & HAMPEL, D. (2012). DEA as a tool for bankruptcy assessment: the agribusiness case study. *Proceedings of the 30th International Conference Mathematical Methods in Economics 2012.* Karviná: Silesian University in Opava, School of Business Administration in Karviná, 379–383.
[8] KLEPÁČ, V. & HAMPEL, D. (2016). Prediction of Bankruptcy with SVM Classifiers Among Retail Business Companies in EU. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis,* 64, 627–634.
[9] KLEPÁČ, V. & HAMPEL, D. (2018). Predicting Bankruptcy of Manufacturing Companies in EU. *Economics and management*, *21(1),* 159–174.
[10] KÖKSALA, C. D. & AKSU, A. A. (2007). Efficiency evaluation of A-group travel agencies with data envelopment analysis (DEA): A case study in the Antalya region, Turkey. *Tourism Management, 28(3),* 830–834.
[11] STAŇKOVÁ, M. & HAMPEL, D. (2017). Identification of bankruptcy factors for engineering companies in the EU. *Mathematical Methods in Economics 2017: Conference Proceedings.* Hradec Králové: Univerzita Hradec Králové, 714–719.
[12] STAŇKOVÁ, M. & HAMPEL, D. (2018). Efficiency Comparison in the Development of Building Projects Sector. *Mathematical Methods in Economics 2018: Conference Proceedings.* MatfyzPress: Praha, 503–508.
[13] STAŇKOVÁ, M. & HAMPEL, D. (2019). Bankruptcy Prediction Based on Data Envelopment Analysis. *Mathematical Methods in Economics 2019: Conference Proceedings.* České Budějovice: Jihočeská univerzita v Českých Budějovicích, 31–36.
[14] STAŇKOVÁ, M. & HAMPEL, D. (2019). Bankruptcy prediction of engineering companies in the EU using classification methods. *Acta Universitatis agriculturae et silviculturae Mendelianae Brunensis, 66(5),* 1347–1356.

Cobb-Douglas or Translog Production Function in Efficiency Analysis?

Marie Šimpachová Pechrová[1], Ondřej Šimpach[1]

Abstract. The aim of this paper is to assess two types of specification of production function when estimating the technical efficiency with Stochastic Frontier Analysis – Cobb-Douglas (CD) and translog (TL) production functions and their influence on the results. CD production function is easy to be estimated. It is a special case in a more general class of production functions with constant elasticity of substitution, where is the mostly used a TL function.

We compared both specifications on the unbalanced panel data of agricultural holdings in the Czech Republic. There were 517 farms and 1708 observations for years 2013 to 2016. Accounting data were taken from Albertina database purchased from Bisnode, s. r. o., company and data about acreage were taken from Land Parcel Identification System. Production was approximated by sales, production factors were: consumption of material and energy, fixed assets, number of employees and acreage of agricultural land. Technical efficiencies were calculated by JLMS (Jondrow, Lovell, Materov, Schmidt, 1982 [11]) method and their differences tested by Wilcoxon signed-rank test and correlated by Spearman's rank coefficient.

The technical efficiency in CD specification was slightly higher (85.69%) than in TL specification (85.12%), but there were found statistically significant differences. Spearman's $\rho = 0.9597$ was close to 1 pointing on high positive dependence. Using likelihood ratio test, the CD specification was rejected in favor of more general specification of TL function.

We can conclude that TL function is more appropriate to be used as a production function in SFA, but the effect on the technical efficiency is mild and almost negligible (despite that the test revealed statistically significant differences, the dependence was very high.)

Keywords: agricultural holdings, Cobb-Douglas, production function, stochastic frontier analysis, translog production function

JEL Classification: C23, C52
AMS Classification: 90C15

1 Introduction

The concept of technical efficiency was firstly introduced by Farrell. The idea is to compare so-called decision- making units according to their ability to transform the inputs to outputs. Those units (typically firms) that are able to produce the given output with the lowest possible combination of inputs (input-oriented models) or are able to attain the highest possible output with given inputs (output-oriented models) are laying on the frontier and are 100% efficient. Other firms are less technically efficient and shall improve their performance. There are two approaches to efficiency measurement: non-parametric and parametric. The well-known are two representants of each group – Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA).

Data envelopment analysis was first proposed by Charnes et al. [6] "It uses linear programming techniques to build a non-parametric efficiency frontier of the data sample." [22]. It can include more inputs and outputs, but its usage on panel data is limited. As we have unbalance panel of agricultural holdings, we rather use parametric method, particularly SFA.

The SFA was simultaneously developed by Meeusen and Van den Broek [15] and Aigner et al. [1]. It estimates a parametric frontier of the best possible practices given a standard cost or profit function. [22]. Its advantage is that it can be used on panel data, can include determinants of technical inefficiency in the

[1] Institute of Agricultural Economics and Information, Mánesova 75, Prague 2, simpachova.marie@uzei.cz, simpach.ondrej@uzei.cz

model and assess them in one step, and takes into account statistical noise, so the differences among firms are given not only due to inefficiency, but also because of other stochastic factors. On the other hand, it is necessary to correctly specify the production technology and the distribution of the inefficiency term which is not always easy. Inefficiency term can have exponential, gamma, half-normal or truncated normal distribution. "Notice that the half-normal is a special case of the truncated normal." [13]. Since the SFA measures the efficiency of a firm's production relative to the estimated production frontier, it is important that the frontier is appropriately specified. [12].

Three functional types prevail within the SFA – Cobb-Douglas (CD), CES and translogaritmic (TL) functions. The most general is translog specification of frontier technology as it provides a good first-order approximation to a broad class of functions, including the CES, and includes the Cobb-Douglas as a special case. [12]. CD production function is easy to be estimated (in the linearized form it can be done by ordinary least squares method), and interpreted, because the coefficients are elasticities. The limitations of CD production function are "the inherent assumption of constant elasticity of substitution between the inputs which also implies a constant percentage of income distribution across them." [16]. Unfortunately, the Cobb-Douglas still fits the data well in cases where some fundamental assumptions are violated. [16].

Hence, there are other types of functions used. CD is a special case in a more general class of production functions with constant elasticity of substitution. The translog function relaxes the assumption of constant elasticity of substitution and reduces to the Cobb Douglas function in case there is constant elasticity of substitution. [20].

Several authors compared the usage of both, CD and TL, specifications. For example, Constantine, Martine and Rivera [7] tested the hypothesis that all the second order coefficients and the cross products in TL function are equal to zero. They rejected the null hypothesis, so the specification in the form of a TL function is preferred to the CD. Duffy and Papageorgiou [8] who examined panel data for 82 countries in a period of 28 years also rejected the Cobb-Douglas specification. Similarly, Kneller and Stevens [12] preferred TL specification to CD.

2 Methods

The CES is a natural extension of the Cobb-Douglas in that it permits the elasticity of substitution to be something other than unity – the elasticity of scale can change with output and/or factor proportions. [9] We compare the technical efficiency calculated with Cobb-Douglas production function estimated as (1) and general translog production function estimated as (2). Both functions have the advantage that they can be linearized by natural logarithms. We do not include time t as an explanatory variable, so we cannot assess technical change (change of technology in time), but in this way, the CD and TL models are better comparable.

$$\ln y_{it} = \sum_{k=1}^{K} \beta_k \ln x_{k,it} + \varepsilon_{it} \tag{1}$$

$$\ln y_{it} = \sum_{k=1}^{K} \beta_k \ln x_{k,it} + \frac{1}{2} \sum_{k=1}^{K} \sum_{l=1}^{K} \beta_{kl} \ln x_{k,it} \ln x_{l,it} + \varepsilon_{it} \tag{2}$$

where y_{it} denotes the production of farm i (i = 1, 2, ..., N, where N is total number of farms) in time t (t = 1, 2, ..., T, where T is total number of observed years). $x_{k,it}$ stands for the input k (k = 1, 2, ..., K, where K is total number of production factors, 4 in our case) of firm i in time t; β_k are the estimated parameters of inputs; ε_{it} is the synthetic error term that consists of inefficiency term u_{it} and stochastic term v_{it}. The division on $\varepsilon_{it} = v_{it} - u_{it}$ was introduced by Aigner et al. [1] who proposed a method that distinguishes the inefficiency from other sources of disturbance that cannot be influenced by the firm. The distribution of the inefficiency was assumed to be truncated normal $u_{it} \sim N^{+}(\mu, \sigma_u^2)$ – with mean μ and variance σ_u^2.

We used True fixed-effects model. It is time-varying model which means that firms' technical inefficiency can develop over time. The parameters of stochastic frontier function are estimated by the maximum likelihood method. Then the elasticities were calculated. "Input elasticities measure the sensitivity of output to an increase in inputs." [2] Coefficients in Cobb-Douglas power function can be already interpreted as elasticities. In translog function are derived by partial derivation of ln y over ln x. The elasticities are calculated at the mean value of the inputs. An example of partial derivation for $x1$ is written in [3].

$$\frac{\delta \ln y}{\delta \ln x_1} = \beta_1 + \beta_{11} \ln x_1 + \frac{1}{2}\beta_{12} \ln x_2 + \frac{1}{2}\beta_{13} \ln x_3 + \frac{1}{2}\beta_{14} \ln x_4 \quad (3)$$

After the estimation of the model, the efficiency is calculated using Jondrow et al. [11] method which measures the contribution of uit to εit $E[u_{it}|\varepsilon_{it}]$. Technical efficiency is then calculated as $\exp[-E(u_{it}|\varepsilon_{it})]$.

Both models were compared by Likelihood ratio test. The differences in technical efficiency were tested by Wilcoxon signed-rank test. In order to understand the effect of specification on efficiency scores, it is also appropriate to examine the correlations between the series by Spearman's rank coefficient ρ. Value of the coefficients takes value between −1 to 1, where values below 0 stays for indirect dependency and above 0 for direct. The calculations were done in Stata IC 15.

Data for the unbalanced panel accounting data of agricultural holdings in the Czech Republic were taken from Albertina database (business register) purchased from Bisnode, s. r. o., company and data about acreage were taken from Land Parcel Identification System. Production of *i*-th farm in time t was approximated by sales of production and services which was adjusted by index of agricultural producers' prices to eliminate the influence of inflation. Explanatory variables – production factors – were: $x_{1,it}$ – consumption of material and energy, $x_{2,it}$ – fixed assets – both variables were adjusted by the index of industry producers' prices in order to eliminate the influence of inflation, $x_{3,it}$ – number of employees and $x_{4,it}$ – acreage of agricultural land. There were 517 farms and 1708 observations for years from 2013 to 2016. Number of observations ranged from 2 to 4 with average 3.3 observations per farm. The description of the variables is given in Tab. 1. Annual sales (58 mil. CZK) were on average higher than the median (40 mil. CZK). Similarly, material and energy consumption was on average 29 mil. CZK with median 21 mil. CZK. Fixed assets were 69 mil. CZK on average, but the median was lower – almost 57 mil. CZK. Average number of employees was 42 with median 38, maximum employees 225.

Variable	Mean	Std. dev.	Median	Min	Max
y_{it} – sales of own products and services (adjusted)	58,453	64,245	40,303	27	498,372
x_{1it} – material and energy consumption (adjusted)	29,057	30,932	21,404	1	315,756
x_{2it} – fixed assets (adjusted)	69,011	90,522	56,698	1	1,200,000
x_{3it} – number of employees [–]	42	42	38	3	225
x_{4it} – acreage [ha]	49,447	1,156	739	0	10,381

Source: own elaboration

Table 1 Description of the variables (in thous. CZK if not stated otherwise)

3 Results

The results of estimated true fixed effects model – with translog and Cobb-Douglas production functions are displayed in table 2. Both models were statistically significant according to Wald χ^2 test. Log likelihood in CD was 1181.8458 and in TL as 1028.4140.

The coefficients in Cobb-Douglas function are interpreted as elasticities. Increase of material and energy consumption by 1% cause increase of production by 0.2803%. Increase of the fixed assets use by 1% cause increase of production by 0.2614%. The highest elasticity is in case of the employees. Increase of their number by 1% cause increase of production by 0.5629%.

The elasticities in translog production function are evaluated at the mean value of the inputs. An increase of material and energy consumption by 1% cause increase of production by 0.3539%. Increase of the fixed assets use by 1% cause increase of production by 0.2627%. Again, the highest elasticity is in case of the employees almost 49.80%. Increase of the acreage of the farm by 1% cause increase of the production by 24.21%. The elasticity is almost similar in case of fixed assets. In case of material and acreage it is higher in translog function. Elasticity of employees is higher in CD function. Sum of the coefficients indicates the returns to scale. In this case it is equal to 1.2085 which shows that firms exhibit increasing returns to scale.

Models were compared using LR test. Model CD is special case of TL model. Likelihood ratio test assumed under H_0 that $\beta_{11} = \beta_{12} = ... = \beta_{44} = 0$, i.e. that production function is better described by the CD function. The likelihood ratio test has test criterion $\chi^2 = 141.95$ that is significant at the 0.05% level. That means that the

probability that H_0 holds is almost zero, hence, the H_0 is rejected. At least one parameter is statistically significantly different from zero. TL is preferred over the CD. Kneller and Stevens [12], Duffy and Papageorgiou [8] and Constantine, Martine and Rivera [7] also rejected the CD specification of production in favor of more general TL.

Then the technical efficiency was estimated using JLMS method [11]. The results are displayed in table 3.

	Cobb-Douglas production function		Translog production function		
ln y	**Coef.** sign. level	**(Std. error)**	**Coef.** sign. level	**(Std. error)**	**Elasticities**
Frontier					
β_1 [ln x_1]	0.2803***	$(8.26 \cdot 10^{-6})$	0.1824***	$(3.46 \cdot 10^{-4})$	0.3537
β_2 [ln x_2]	0.2614***	$(5.44 \cdot 10^{-6})$	0.1744***	$(1.45 \cdot 10^{-4})$	0.2627
β_3 [ln x_3]	0.5629***	$(2.55 \cdot 10^{-5})$	0.1371***	(.)	0.4980
β_4 [ln x_4]	0.1040***	$(3.69 \cdot 10^{-6})$	0.0297***	$(3.07 \cdot 10^{-4})$	0.2421
β_{11} [ln x_1 ln x_1]			0.0088***	$(2.47 \cdot 10^{-5})$	
β_{12} [ln x_1 ln x_2]			0.0030***	$(1.36 \cdot 10^{-5})$	
β_{13} [ln x_1 ln x_3]			0.0205***	$(3.09 \cdot 10^{-4})$	
β_{14} [ln x_1 ln x_4]			0.0048***	$(1.00 \cdot 10^{-5})$	
β_{22} [ln x_2 ln x_2]			0.0060***	$(1.94 \cdot 10^{-5})$	
β_{23} [ln x_2 ln x_3]			0.0220***	$(1.61 \cdot 10^{-4})$	
β_{24} [ln x_2 ln x_4]			0.0039***	$(2.08 \cdot 10^{-5})$	
β_{33} [ln x_3 ln x_3]			0.0249***	$(3.52 \cdot 10^{-4})$	
β_{34} [ln x_3 ln x_4]			0.0074***	$(3.35 \cdot 10^{-5})$	
β_{44} [ln x_4 ln x_4]			0.0141***	$(1.61 \cdot 10^{-5})$	
μ_u					
constant			−14.6404***	(4.5526)	
Variance of inefficiency term					
constant	−3.2042***	(0.0484)	1.0078***	(0.3022)	
Variance of stochastic term					
constant	−32.8654***	(92.6917)	−33.1247	(44.8065)	
σ_u	0.2015***	(0.0049)	1.6552	(0.2501)	
σ_v	$7.30 \cdot 10^{-8}$	$(3.38 \cdot 10^{-6})$	0.0000	$(1.44 \cdot 10^{-6})$	
λ	2759543***	(0.0049)	25800000	(0.2501)	

Source: own elaboration; Note: *** denotes significance at 1%, ** at 5% and * at 10%

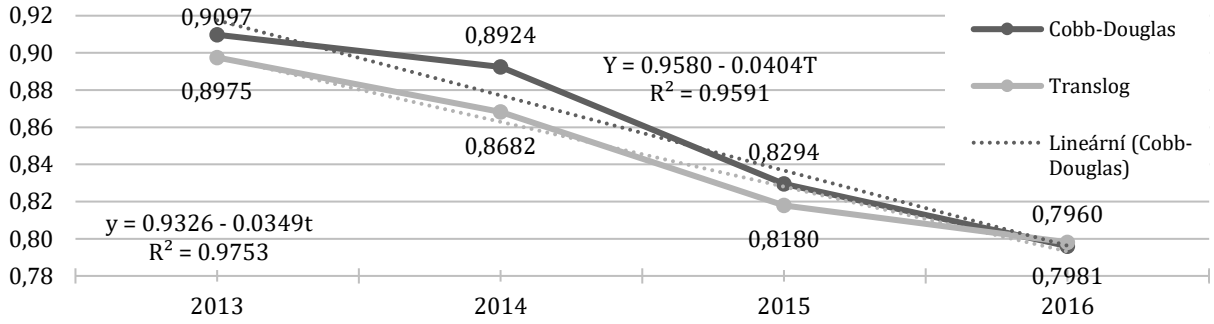
Table 2 True fixed-effects model estimates, distribution of inefficiency term: CD – truncated normal, TL – exponential, distribution of random error: CD, TL – normal

Average technical efficiency in case of CD function was slightly higher (85.69%) than in TL (84.57%). Median was similar in both specifications – 90%. There were 545 100% efficient firms (with efficiency higher than 99%) in CD and 549 in TL. Because the technical efficiency is not normally distributed among farms, the correlation had to be assessed by Spearman correlation coefficient. Spearman's rho was equal to 0.94. H_0 that the efficiencies are independent was rejected. There is positive and strong dependence between both variables.

Production function	Mean	Median	Std. dev.	Min	Max	100% efficient farms
Cobb-Douglas	0.8569	0.8999	0.1668	0.0810	1.0000	545
Translog	0.8457	0.9009	0.1813	0.0693	1.0000	549

Table 3 Technical efficiency of farms based on two specifications of production function (Source: own elaboration)

We can also look on the development of both technical efficiencies in time. From Fig 1, it can be seen that CD efficiency was higher in all years with exception of 2016. It decreased in all years and was the lowest in the last year. This decreasing trend is rather surprising. If we estimate the trend of CD efficiency by trend function, $y = 0{,}9580 - 0{,}0404t$ with rather high fit – adjusted coefficient of determination was $\hat{R}^2 = 93.76\%$. Also, TL trend function $y = 0.9326 - 0.0349t$ had high fit of data – $\hat{R}^2 = 96.29\%$. Average decrease during the whole period 2013 to 2016 was 4.35% in CD function and 3.84% in TL function.



Source: own elaboration

Figure 1 Development of technical efficiency estimated by CD and TL function in period 2013–2016

4 Discussion

Kneller and Stevens [12] concluded "that the effect of the specification of the frontier on measures of productive efficiency is less important." As same as in our case, when TL is preferred over the CD, Kneller and Stevens [12], Duffy and Papageorgiou [8] and Constantine, Martine and Rivera [7] also rejected the Cobb-Douglas specification in favor of translog. The effect of functional form on the estimated efficiency terms was relatively minor. Higher impact had different specification of production function labor and its measurement. Kumbhakar, Heshmati and Hjalmarsson [14] considered the issues related to the specification and estimation of various models incorporating time-varying technical efficiency. They compared Cornwell, Schmidt and Sickles model, Lee and Schmidt model and Battese and Coelli model and found that the efficiency scores vary substantially among the models. "Therefore, an important issue is to what extent the different models generate reasonable, or rather unreasonable, results." [14]. In our case, all models bring reasonable and very similar results. Holmgren [10] compared the efficiency estimated by stochastic cost frontier models with various specification of the explained variable. Particularly he used different expressions of "production of public transport" and concluded that "models using only vehicle-kilometers or only passenger trips tend to underestimate efficiency compared to a model using both at the same time" [10], so also the specification of the variables is important to obtain correct results.

Regarding the height of the technical efficiency in the Czech Republic, it is usually estimated similarly as in our research (around 85%). For example, Pechrová [19] found out that the average efficiency of Czech farms in years 2007–2013 was 86.74%, median was higher (91.03%) as same as in our research. Rudinskaya et al. [21] estimated the technical efficiency of the Czech farms in years 2011–2015 in height of 71.5 %. Lower efficiency was estimated by Nowak, Kijek and Domanska [18] because they used DEA. Under constant returns to scale they calculated average efficiency of Czech agriculture between 2007–2011 in height 69.2% and under variable returns to scale 69.6%. Similarly, in research of Bojnec et al. [3] DEA technical efficiency score of Czech farms for years 2001–2006 varied between 54% in the first year to 65% in the last year of this period.

Regarding the development in time, Čechura et al. [5] found for the CR a trend function y = 0.140 – 0.038t, but with very low fit $\hat{R}^2 = 40\%$. The trend was negative and also average change of technical efficiency during observed period 2004–2011 was negative (−0.029). According to Čechura [4] efficiency in Czech agriculture was considerably volatile in years 2004–2007. Technical efficiency increased in 2005 and decreased in the next year. In 2007, the level of technical efficiency returned to roughly the level it had in 2004. Those two results correspond to our results for the period 2013–2016, so we can see that the decreasing trend of the technical efficiency continued. Naglová and Šimpachová Pechrová [17] found out that food processing companies in the Czech Republic were the least technically efficient in 2020 (average efficiency was 59.21%) and the increase after this year was relatively slow – up to 69.56% in 2014, but then the decrease followed on 65.46% in 2017.

Conclusion

The paper evaluated two types of production function – Cobb-Douglas (CD) and translog (TL) – when estimating the technical efficiency with Stochastic Frontier Analysis. Both specifications were compared on the unbalanced panel data of agricultural holdings in the Czech Republic. Technical efficiencies were calculated by JLMS method and their differences tested by Wilcoxon signed-rank test and correlated by Spearman's rank coefficient.

There were found statistically significant differences in the technical efficiency in CD specification (85.69%) and in TL specification (85.12%), but Spearman's $\rho = 0.9597$ was close to 1 pointing on high positive dependence. Therefore, likelihood ratio test was used to choose the best specification. The CD specification of production function was rejected in favor of more general specification of TL function. We can conclude that TL function is more appropriate to be used as a production function in SFA, but the effect on the technical efficiency is mild and almost negligible despite the statistically significant differences. This is a desirable feature as the specification of the production function does not influence the results of technical efficiency. The challenge for future research is to examine the other determinants of the technical efficiency of the agricultural holdings in the Czech Republic.

Acknowledgements

The paper was supported by the Ministry of Agriculture of the Czech Republic, institutional support MZE-RO0918, Internal Research Project no. 1113/2020.

References

[1] AIGNER, D., LOVELL, C. & SCHMIDT, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*, 21–37.

[2] ARCHAMBAULT S. ET AL. (2014). Nannochloropsis sp. algae for use as biofuel: Analyzing a translog production function using data from multiple sites in the southwestern United States. *Algal Research*, *6*, 124–131.

[3] BOJNEC, Š., FERTŐ, I., JAMBOR A. & TÓTH, J. (2014). Determinants of technical efficiency in agriculture in new EU member states from Central and Eastern Europe. *Acta Oeconomica*, *64(2),*197–217.

[4] ČECHURA, L. (2012). Technical efficiency and total factor productivity in Czech agriculture. *Agricultural Economics – Czech*, *58(4)*, 147–156.

[5] ČECHURA, L. ET AL. (2015). Comparison of Technology and Technical Efficiency in Cereal Production among EU Countries. *Agris on-line Papers in Economics and Informatics*, *VII(2)*, 27–37.

[6] CHARNES, A., COOPER, W. & RHODES, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, *3*, 429–444.

[7] CONSTANTIN, P. D., MARTIN, L. M. & RIVERA, E. B. B. R. (2009). Cobb-Douglas, Translog Stochastic Production Function and Data Envelopment Analysis in Total Factor Productivity in Brazilian Agribusiness. Journal of Operations and Supply Chain Management, 2(2), 20–34.

[8] DUFFY, J.; PAPAGEORGIOU, C. (2000). A cross-country empirical investigation of the aggregate production function specification. *Journal of Economic Growth*, *5*, 87–120.

[9] HEATHFIELD, D. F. & WIBE, S. (1987). T*he Translog Function. In: An Introduction to Cost and Production Functions*. Palgrave, London.

[10] HOLMGREN, J. (2018). The effects of using different output measures in efficiency analysis of public transport operations. *Research in Transportation Business and Management*, 28, 12-22. DOI: 10.1016/j.rtbm.2018.02.006

[11] JONDROW, J., LOVELL, C. A. K., MATEROV, I. S., SCHMIDT, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics, 19(1982), 233–238.

[12] KNELLER, R. & STEVENS, P. A. (2003). The specification of the aggregate production function in the presence of inefficiency. *Economics Letters*, *81*, 223–226.

[13] KUMBHAKAR, S. C. & TSIONAS, E. G. (2006). Estimation of stochastic frontier production functions with input-oriented technical efficiency. *Journal of Econometrics*, *133*, 71–96.

[14] KUMBHAKAR, S. C., HESHMATI, A. & HJALMARSSON, L. (1997). Temporal patterns of technical efficiency: Results from competing models. *International Journal of Industrial Organization*, *15(5)*, 597–616.

[15] MEEUSEN, W. & VAN DEN BROEK, K. (1977). Efficiency estimation from Cobb–Douglas production functions with composed error. *International Economic Review*, *18*, 435–444.

[16] MILLER, E. (2008). An assessment of CES and Cobb-Douglas Production Functions. [on-line] Congres. Budget Office. Available from https://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/94xx/doc-9497/2008-05.pdf.
[17] NAGLOVÁ, Z. & ŠIMPACHOVÁ PECHROVÁ, M. (2019) Subsidies and technical efficiency of Czech food processing industry. *Agricultural Economics*, *65(4)*, 151–159.
[18] NOWAK, A., KIJEK, T. & DOMAŇSKA, K. (2015). Technical efficiency and its determinants in the European Union agriculture. *Agricultural Economics – Czech, 6(16),* 275–283.
[19] PECHROVÁ, M. (2015). Impact of the Rural Development Programme Subsidies on the farms' inefficiency and efficiency. *Agricultural Economics – Czech*, *61(5)*, 197–204.
[20] PRASAD, R. (2015). The production function methodology for estimating the value of spectrum. *Telecommunications Policy*, *39(1)*, 77–88.
[21] RUDINSKAYA, T., HLAVSA, T. & HRUŠKA, M. (2019). Estimation of technical efficiency of Czech farms operating in less favoured areas. *Agricultural Economics – Czech*, *65(10)*, 445–453.
[22] SILVA, T. C. ET AL. (2017). A comparison of DEA and SFA using micro- and macro-level perspectives: Efficiency of Chinese local banks. Physica A: Statistical Mechanics and its Applications, 469, 216–223.

Fuzzy Evaluation of Passability of Routes During Crisis Situation

Michal Škoda[1], Helena Brožová[2]

Abstract. The importance of securing a timely help during natural disasters is very easy to understand, but not so easy to achieve. Especially in cases when the area affected by the disaster is hard to reach and there are not many details known about the situation. The aim of this paper is to help to identify passable routes during natural disasters in the first moments when the situation occurs. The entire approach is based on fuzzy set theory, which is widely used in representing uncertain knowledge.

First, fuzzy linguistic scales and criteria are chosen. Second, these routes and natural disaster are evaluated based on the chosen criteria. Based on innovative approach of using α-level cut the effect of natural disaster on routes is considered. At the end, using Hamming distance, the passability of routes are calculated and interpreted.

Keywords: evaluation, fuzzy number, scale, criteria, crisis, hamming distance

JEL Classification: C44, D70, D80
AMS Classification: 90B50, 62C86

1 Introduction

Unfortunately, crisis situations caused by accidents or natural disasters are nothing rare in today's world. Among natural disasters we could include for instance earthquakes, floods, whirlwinds, severe wind gusts, hailstorms, droughts or fires [23]. Natural disasters may cause significant damage to environment, infrastructure and human life [2].

In field of crisis management most papers usually focus on crisis preparedness and evaluation of crisis management as in Boersma et al. [4] or Vichova and Hromada [20], but in this paper will be the focus on logistic in crisis situations caused by natural disasters as for example in Nikoo et al. [15], Francini et al. [10], Zahedi et al. [22], Shahparvari et al. [17].

On the research of logistics in crisis situations, many papers have discussed it from the perspectives of characteristics, concept, system construction and safeguard mechanism [15]. Almost all crisis situations are closely connected with uncertainty, which can take many forms [1, 3, 8]. Also, Sheu [19] pointed out that logistics in crisis situations are different from commercial logistics in that it has such significant characteristics as unpredictability, non-conventionality and uncertainty.

Fuzzy theory is very helpful to deal with the uncertainty and vagueness of human thoughts in making decisions [14, 21]. The theory assumes that people do not think in exact values as “yes” or “no”, but rather distinguish a range of blurry values as for example “absolutely”, “nearly”, “rather yes” or “maybe no” [6, 11]. Therefore, fuzzy logic provides a simple method to reach a definite conclusion based on ambigous, imprecise or vague information. This is especially usefull way to assess situation in cases where experts or participants do not have enough reliable data [12].

The aim is to present an evaluation approach based on fuzzy theory focusing on passability of routes in crisis situations that is user-friendly, flexible, precise and working with vague expressions. More specifically the evaluation approach will focus on passability of routes after earthquakes.

The structure of the paper is as follows: the second section describes the used tools and methods with the main focus on tools from area of fuzzy set theories. The linguistic scales, fuzzification and interpretation

[1] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague 6, mskoda@pef.czu.cz.

[2] Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague 6, brozova@pef.czu.cz.

methods are picked. In the third section, the whole evaluation approach is described in detail. The fourth section provides a case study and, in the conclusion, the proposed approach and results are evaluated.

2 Fuzzy number and fuzzy linguistic scale

Fuzzy numbers are fuzzy sets in a set of real numbers $\mathbb{R} = (-\infty, +\infty)$. For the purposes of this article triangular and trapezoidal fuzzy numbers will be used.

A fuzzy number A is a triangular fuzzy number, if its membership function $\mu_A: \mathbb{R} \to [0,1]$ has the following form

$$\mu_A(x) = \begin{cases} 0 & \text{if} \quad x < a_1 \\ \dfrac{x - a_1}{a_2 - a_1} & \text{if} \quad a_1 \leq x < a_2 \\ \dfrac{a_3 - x}{a_3 - a_2} & \text{if} \quad a_2 < x \leq a_3 \\ 0 & \text{if} \quad x > a_3, \end{cases} \tag{1}$$

(1)

where a_1, a_2, a_3 are real numbers and $a_1 \leq a_2 \leq a_3$. We denote a triangular fuzzy number by $A = (a_1, a_2, a_3)$.

A fuzzy number A is a trapezoidal fuzzy number, if its membership function $\mu_A: \mathbb{R} \to [0,1]$ has the following form:

$$\mu_A(x) = \begin{cases} 0 & \text{if} \quad x < a_1 \\ \dfrac{x - a_1}{a_2 - a_1} & \text{if} \quad a_1 \leq x < a_2 \\ 1 & \text{if} \quad a_2 \leq x \leq a_3 \\ \dfrac{a_4 - x}{a_4 - a_3} & \text{if} \quad a_3 < x \leq a_4 \\ 0 & \text{if} \quad x > a_4, \end{cases} \tag{2}$$

where a_1, a_2, a_3, a_3 are real numbers and $a_1 \leq a_2 \leq a_3 \leq a_4$. We denote it by $A = (a_1, a_2, a_3, a_4)$ [13].

Let $A = (a_1, a_2, a_3, a_4)$, $B = (b_1, b_2, b_3, b_4)$, $C = (c_1, c_2, c_3, c_4)$ and $D = (d_1, d_2, d_3, d_4)$ be four trapezoidal fuzzy numbers. Then, the average of these fuzzy numbers is the following:

$$A_{avg} = \left(\frac{1}{n}\sum_{i=1}^{n} a_i, \frac{1}{n}\sum_{i=1}^{n} b_i, \frac{1}{n}\sum_{i=1}^{n} c_i, \frac{1}{n}\sum_{i=1}^{n} d_i \right). \tag{3}$$

This operation yields trapezoidal fuzzy numbers as a result [5].

***α**-transformation of Triangular Fuzzy Number to Trapezoidal Fuzzy Number*

Suppose a triangular fuzzy number and a real number $\alpha \in [0,1]$. The new trapezoidal fuzzy number, which kernel is equal to the α-cut of original triangular fuzzy number, will be called the α-transformed fuzzy number.

The α-transformation of triangular fuzzy number $A = (a_1, a_2, a_3)$ is a trapezoidal fuzzy number $A_\alpha = (a_1, a_{2\alpha}, a_{3\alpha}, a_4)$, where the element a_1 remain the same and element a_3 from the original triangular fuzzy number will become element a_4 of the transformed trapezoidal fuzzy number. Using following formula, the new values of element $a_{2\alpha}$ and element $a_{3\alpha}$ will be calculated:

$$a_{2\alpha} = a_1 + \alpha \times (a_2 - a_1) \tag{4}$$

$$a_{3\alpha} = a_3 - \alpha \times (a_3 - a_2) \tag{5}$$

Graphical representation of this transformation is shown in the Figure 1.

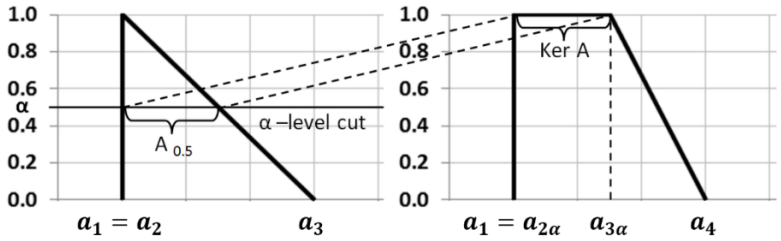


Figure 1 α-transformation of triangular to trapezoidal fuzzy number

Fuzzy Linguistic Scale

For the purposes of expression of vague terms, a linguistic fuzzy scale will be used. There are many different linguistic scales described for example in [7] or [9]. In this article, the Likert scale will be used.

Interpretation of fuzzy numbers

In order to reach a linguistic term, the resulting fuzzy numbers will be interpreted through the Hamming distance which is method for determination of distance between fuzzy numbers. The Hamming distance for trapezoidal fuzzy numbers is defined as follows.

$$d_H(A,B) = \sum_{j=1}^{n} |a_j - b_j|, \tag{6}$$

where $A = (a_1, a_2, a_3, a_4)$ and $B = (b_1, b_2, b_3, b_4)$ are two trapezoidal fuzzy numbers. The Hamming distance value equal to 0 refers to the ideal solution [18].

3 Steps of process of the passability evaluation

Step 1: Determination of Criteria

In this step the determination of criteria, by which the passability of routes will be determined, should be done. For purposes of the Case study the following three criteria will be used:

- First criterion: Distance from the epicentre – It is assumed that the further the route is from epicentre, the more passable it will be.
- Second criterion: Intensity of the earthquake in the epicentre – It is assumed that the greater the intensity is, the lower the passability of routes will be.
- Third criterion: Route resistance to earthquakes – It is assumed that the better the materials and the better the subsoil, the higher the passability of the routes will be.

Step 2: Selection of the Main Fuzzy Linguistic Scales

In the second step of the evaluation approach the linguistic scales and related fuzzy numbers are defined. For the simplicity the same scale will be used for all three criteria in the Case Study (Table 1 and **Chyba! Nenalezen zdroj odkazů.**)

Criteria			**Fuzzy number**
Distance from the epicentre	**Intensity of the earthquake**	**Route resistance**	
Short	Big	Nondurable	(0.00, 0.00, 0.40)
Medium	Medium	Normal	(0.10, 0.50, 0.90)
Long	Small	Durable	(0.60, 1.00, 1.00)

Table 1 Scales and coresponding fuzzy numbers

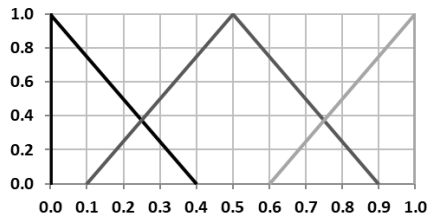


Figure 2 Graphical representation of the main scale

Step 3: The Interpretative Fuzzy Number

During the third step the fuzzy number that will be used for transformation and final interpretation should be determined. The shape of the interpretative fuzzy number is crucial to the final outcome, since it is used to reflect the effect of earthquake to routes. The bigger the slope of the interpretative fuzzy number is the greater the distance between individual resulting variants. For the purposes of Case study, the fuzzy number (0.00, 0.90, 1.00) with significant slope will be used (**Figure 3**).

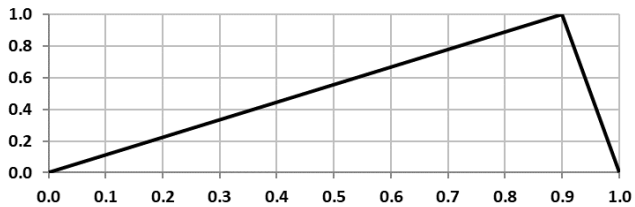


Figure 3 Graphical representation of the interpretative fuzzy number

Step 4: The Calculation of Derived Scales

In order to interpret the final results, the derived scales must be calculated. For the purposes of Case Study two derived scales will be used. First derived scale will be used for interpretation of α-level and the second derived scale will be used for interpretation of final results. In order to construct derived scales, two extreme values of possible results in previous step are calculated and the rest of values is evenly distributed between them. The derived fuzzy linguistic scale for interpretation of α-level is shown in Table 2 and Figure **4**).

α	**Fuzzy Number**
0.1	(0.00, 0.00, 0.40)
0.2	(0.07, 0.11, 0.47)
0.3	(0.13, 0.22, 0.53)
0.4	(0.20, 0.33, 0.60)
0.5	(0.27, 0.44, 0.67)
0.6	(0.33, 0.56, 0.73)
0.7	(0.40, 0.67, 0.80)
0.8	(0.47, 0.78, 0.87)
0.9	(0.53, 0.89, 0.93)
1	(0.60, 1.00, 1.00)

Table 2 Scale expresing α-level

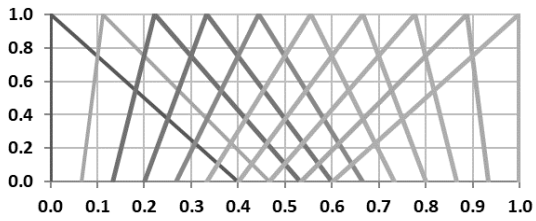


Figure 4 Scale expresing α-level

The derived fuzzy linguistic scale for interpretation of final results is shown in Table 3 and Figure 5.

Passability	Fuzzy Number
Impassable	(0.00, 0.09, 0.99. 1.00)
Almost impassable	(0.00, 0.33, 0.97. 1.00)
Partly passable	(0.00, 0.57, 0.95. 1.00)
Passable	(0.00, 0.82, 0.92. 1.00)
Seamlessly passable	(0.00, 0.90, 0.90. 1.00)

Table 3 Scale expresing passability

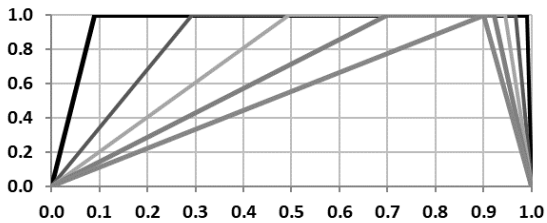


Figure 5 Scale expresing passability

Step 5: The Calculation and Determination of Passability

First the situation for each route is evaluated based on selected criteria using predetermined fuzzy linguistic scale. Then the average of corresponding fuzzy numbers is calculated for each route. Then the minimal Hamming distance of these average fuzzy numbers and fuzzy numbers from selected fuzzy linguistic scale for α-level is calculated. α-level with the minimum distance is picked and used to transform predetermined interpretative fuzzy number for each route. The lower the α is, the greater effect of transformation will be. Finally, the Hamming distance of α −transformed fuzzy number and fuzzy numbers in derived fuzzy linguistic scale interpreting the passability is calculated. The fuzzy number with minimum distance from transformed fuzzy number is picked like final evaluation of passability of the route.

4 Case study

The designed approach for the evaluation of passability of routes is applied on a following crisis situation:

The earthquake strikes a densely populated area. Initial estimates speak of the intensity of 8 on the Richter magnitude. The goal is to quickly evaluate the passability of the routes in area so that the necessary assistance can be provided as soon as possible. The plan of the area with epicentre is shown in the Figure 6 below.

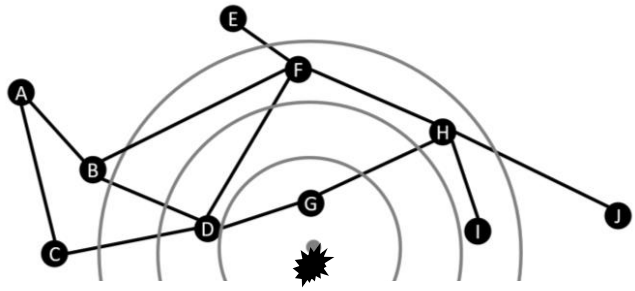


Figure 6 The plan of the area (with final indentification of passable routes)

In **Table 4** the evaluation of the situation based on predetermined criteria on predetermined fuzzy linguistics scales is shown. Based on the linguistic scale, the earthquake intensity is evaluated as Big.

Route	Distance from the Epicentre		Intensity of the Earthquake	Route Resistance to Earthquakes	
AB	Long	(0.60, 1.00, 1.00)	Big (0.00, 0.00, 0.40)	Durable	(0.60, 1.00, 1.00)
AC	Long	(0.60, 1.00, 1.00)		Nondurable	(0.00, 0.00, 0.40)
BD	Medium	(0.10, 0.50, 0.90)		Nondurable	(0.00, 0.00, 0.40)
BF	Medium	(0.10, 0.50, 0.90)		Durable	(0.60, 1.00, 1.00)
CD	Medium	(0.10, 0.50, 0.90)		Nondurable	(0.00, 0.00, 0.40)
DF	Medium	(0.10, 0.50, 0.90)		Durable	(0.60, 1.00, 1.00)
DG	Short	(0.00, 0.00, 0.40)		Nondurable	(0.00, 0.00, 0.40)
EF	Long	(0.60, 1.00, 1.00)		Normal	(0.10, 0.50, 0.90)
FH	Medium	(0.10, 0.50, 0.90)		Normal	(0.10, 0.50, 0.90)
GH	Short	(0.00, 0.00, 0.40)		Normal	(0.10, 0.50, 0.90)
HI	Medium	(0.10, 0.50, 0.90)		Durable	(0.60, 1.00, 1.00)
HJ	Long	(0.60, 1.00, 1.00)		Normal	(0.10, 0.50, 0.90)

Table 4 Evaluation of the situation

For each route the average fuzzy number is calculated. The α-levels were selected based on the minimal Hamming distance of average fuzzy number and fuzzy numbers in derived fuzzy linguistic scale interpreting α-level. The selected α-levels were used to transform predetermined interpretative fuzzy number. The Hamming distance of transformed fuzzy numbers and fuzzy numbers in the fuzzy linguistic scale interpreting the passability is calculated. The results of each route can be seen in **Table 5** below.

Route	Average Fuzzy Number	α	Transformed Fuzzy Number	Passability
AB	(0.40, 0.67, 0.80)	0.70	(0.00, 0.63, 0.93, 1.00)	**Passable**
AC	(0.20, 0.33, 0.60)	0.40	(0.00, 0.36, 0.96, 1.00)	**Almost impassable**
BD	(0.03, 0.17, 0.57)	0.25	(0.00, 0.23, 0.98, 1.00)	**Almost impassable**
BF	(0.23, 0.50, 0.77)	0.55	(0.00, 0.50, 0.95, 1.00)	**Partly passable**
CD	(0.03, 0.17, 0.57)	0.25	(0.00, 0.23, 0.98, 1.00)	**Almost impassable**
DF	(0.23, 0.50, 0.77)	0.55	(0.00, 0.50, 0.95, 1.00)	**Partly passable**
DG	(0.00, 0.00, 0.40)	0.10	(0.00, 0.09, 0.99, 1.00)	**Impassable**
EF	(0.23, 0.50, 0.77)	0.55	(0.00, 0.50, 0.95, 1.00)	**Partly passable**
FH	(0.07, 0.33, 0.73)	0.40	(0.00, 0.36, 0.96, 1.00)	**Almost impassable**
GH	(0.03, 0.17, 0.57)	0.25	(0.00, 0.23, 0.98, 1.00)	**Almost impassable**

Route	Average Fuzzy Number	α	Transformed Fuzzy Number	Passability
HI	(0.23, 0.50, 0.77)	0.55	(0.00, 0.50, 0.95, 1.00)	**Partly passable**
HJ	(0.23, 0.50, 0.77)	0.55	(0.00, 0.50, 0.95, 1.00)	**Partly passable**

Table 5 Calculation steps and final evaluation of the routes

From the results it is obvious that there are no problems with route AB. Four routes are partly passable and the rest is almost impassable or impassable (**Figure 6**).

5 Conclusions

It can be assumed that the suggested approach is generally usable for many other types of crisis situations caused by disasters such as floods, fires or volcano eruption. The approach can be also considered as simple and accurate. However, further research is needed to support these claims.

The possible weakness regarding accuracy of results can be found in the setting of the Fuzzy Linguistic Scales and Interpretative Fuzzy Numbers. The slope and position of the fuzzy numbers is essential and should be set by an expert or group of experts who should determine what values a specific fuzzy number can reach.

It is important to note that linear fuzzy numbers do not need to be sufficient. In some cases, it would be necessary to use curved or non-linear fuzzy numbers for more accurate results.

In this case study it was confirmed that this approach can be applicable on a crisis situation caused by disaster. But it was only a brief example of a specific situation. Practical use of the suggested approach would require greater robustness.

Acknowledgements

This research is supported by the grant No. 2020A0006 "Využití robustních a fuzzy přístupů v krizovém managementu" of the Internal Grant Agency of the University of Life Sciences Prague.

References

[1] AFIFI, WALID A., ET AL. (2014) Uncertainty and Control in the Context of a Category-Five Tornado. *Research in Nursing & Health*, 37(5), 358–366, DOI: 10.1002/nur.21613.

[2] AHMAD, K., RIEGLER, M., POGORELOV, K., CONCI, N., HALVORSEN, P., & DE NATALE, P. (2017). Jord: a system for collecting information and monitoring natural disasters by linking social media with satellite imagery. *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing,* page 12. ACM.

[3] BENALI, AKLI, ET AL. (2017) Fire Spread Predictions: Sweeping Uncertainty under the Rug. *Science of The Total Environment*, 592, 187–196., DOI: 10.1016/j.scitotenv.2017.03.106.

[4] BOERSMA, F. K., PASSENIER, D. F., MOLLEE, J. S., & VAN DER WAL, C. N., (2012) Crisis Management Evaluation: Formalisation & Analysis of Communication During Fire Incident In Amsterdam Airport Train Tunnel. *Proceedings of the 26th European Conference on Modelling and Simulation*, ECMS, 325-331.

[5] BOJADZIEV, G. & BOJADZIEV, M. (2007). Fuzzy Logic for Business, Finance, and Management. *Advances in Fuzzy Systems — Applications and Theory*. DOI: 10.1142/6451.

[6] BUCKLEY, J. J., ESLAMI, E., & FEURING, T. (2006). *Fuzzy mathematics in economics and engineering*. New York: Physica-Verlag Publishing.

[7] CASILLAS, J. (2011). *Accuracy improvements in linguistic fuzzy modeling*. Berlin: Springer.

[8] CHENG, Y. & ZHENG, X. (2019). Effect of uncertainty on cooperative behaviors during an emergency evacuation. *Communications in Nonlinear Science and Numerical Simulation,* 66, 216–225. DOI: 10.1016/j.cnsns.2018.06.022.

[9] DONG, Y. & XU, J. (2019). *Linguistic decision making: numerical scale model and consistency-driven methodology*. Singapore: Springer.

[10] FRANCINI, M., ARTESE, S., GAUDIO, S., PALERMO, A., & VIAPIANA, M. F. (2018). To support urban emergency planning: A GIS instrument for the choice of optimal routes based on seismic hazards. *International Journal of Disaster Risk Reduction*, 31, 121–134. DOI: 10.1016/j.ijdrr.2018.04.020.

[11] HERRERA, F., ALONSO, S., CHICLANA, F., & HERRERA-VIEDMA, E. (2009). Computing with words in decision making: foundations, trends and prospects. *Fuzzy Optimization and Decision Making*, *8*(4), 337–364. DOI: 10.1007/s10700-009-9065-2.
[12] JIANG, W., DENG, L., CHEN, L., WU, J., & LI, J. (2009). Risk assessment and validation of flood disaster based on fuzzy mathematics. *Progress in Natural Science*, *19*(10), 1419–1425. DOI: 10.1016/j.pnsc.2008.12.010.
[13] KLIR, G. J. & YUAN, B. (2015). *Fuzzy sets and fuzzy logic: theory and applications*. India: Pear-son.
[14] LIN, C.-J. & WU, W.-W. (2008). A causal analytical method for group decision-making under fuzzy environment. *Expert Systems with Applications*, *34*(1), 205–213. DOI: 10.1016/j.eswa.2006.08.012.
[15] NIKOO, N., BABAEI, M., & MOHAYMANY, A. S. (2018). Emergency transportation network design problem: Identification and evaluation of disaster response routes. *International Journal of Disaster Risk Reduction,* 27, 7–20. DOI: 10.1016/j.ijdrr.2017.07.003.
[16] QI, C. & HU, L. (2020). Optimization of vehicle routing problem for emergency cold chain logistics based on minimum loss. *Physical Communication*, 40, 101085. DOI: 10.1016/j.phycom.2020.101085.
[17] SHAHPARVARI, S., ABBASI, B., CHHETRI, P., & ABARESHI, A. (2017). Fleet routing and scheduling in bushfire emergency evacuation: A regional case study of the Black Saturday bushfires in Australia. *Transportation Research Part D: Transport and Environment.* DOI: 10.1016/j.trd.2016.11.015.
[18] SHALLIT, J. (2009). Hamming distance for conjugates. *Discrete Mathematics*, *309*(12), 4197–4199. DOI: 10.1016/j.disc.2008.11.001.
[19] SHEU, J.-B. (2007). Challenges of emergency logistics management. *Transportation Research Part E: Logistics and Transportation Review*, *43*(6), 655–659. DOI: 10.1016/j.tre.2007.01.001.
[20] VICHOVA, K. & HROMADA, M. (2018) The evaluation module of the crisis preparedness for the hospitals. *International Journal of Biology and Biomedical Engineering*, 12, 178–185.
[21] ZADEH, L. A. (1964). *Fuzzy Sets*. DOI: 10.21236/ad0608981.
[22] ZAHEDI, A., KARGARI, M., & KASHAN, A. H. (2020). Multi-objective decision-macking model for distribution planning of goods and routing of vehicles in emergency. *International Journal of Disaster Risk Reduction,* 101587. DOI: 10.1016/j.ijdrr.2020.101587.
[23] ZUZÁK, R. & FAJFAROVÁ, M. (2009). *Krizové řízení podniku*. Prague, Grada, ISBN 978-80-247-3156-8.

Impact of Cluster Organizations on Financial Performance in Selected Industries: Malmquist Index Approach

Eva Štichhauerová[1], Miroslav Žižka[2]

Abstract. The article deals with the evaluation of differences in financial performance of companies in relation to membership in a cluster organization. The aim is to determine whether member companies in a cluster organization reach higher financial performance than other companies. The research included companies from two sectors: the automotive and engineering industries. The companies were divided into three groups: member companies of a cluster organization, non-member companies operating in the same region as a cluster organization, and companies operating in other regions. Financial performance development was evaluated using the Malmquist index based on DEA window scores. Equity, liabilities, revenues from own products and services and economic value added were used as inputs and outputs of the model. The overall change in financial performance was monitored in terms of changes in technical efficiency and technological change. The change in technical efficiency was further examined in terms of pure technical efficiency and scale efficiency. Differences between individual groups of companies were evaluated using the Games-Howell nonparametric post-hoc test, applied to the logarithm data. In the automotive industry, natural clustering has proved to be more favorable in terms of the financial performance of companies. In the engineering industry, the best performing companies were members of cluster organizations.

Keywords: Adjacent Malmquist Index, cluster organization, economic value added, financial performance, automotive industry, engineering industry.

JEL Classification: C61, L25, L62, L64.
AMS Classification: 90B90, 90C90

1 Introduction

Clusters have been an important tool of industrial policy since the 1990s. Cluster policy guru is considered to be Michael Porter [11] who sees the cluster as: *"Clusters are geographic concentrations of interconnected companies and institutions in a particular field. Clusters encompass an array of linked industries and other entities important to competition. They include, for example, suppliers of specialized inputs such as components, machinery, and services, and providers of specialized infrastructure. Clusters also often extend downstream to channels and customers and laterally to manufacturers of complementary products and to companies in industries related by skills, technologies, or common inputs. Finally, many clusters include governmental and other institutions—such as universities, standards-setting agencies, think tanks, vocational training providers, and trade associations—that provide specialized training, education, information, research, and technical support".* A large number of studies (for example [5, 10]) can be found in the literature showing the positive impact of clusters on innovation and competitiveness of companies and regions. Nevertheless, critical articles (for example [4, 12]) can be found to point out the limited or even negative impact of clusters on performance. It is also necessary to draw attention to the definition of cluster. Clusters can naturally arise from historical ties between companies and other actors in the region. In this case, natural or Porterian clusters are mentioned [10]. However, clusters can also arise as a result of the organized efforts of a cluster initiative [7]. The result is then a cluster organization as a formalized entity bringing together member organizations, managing joint activities and providing members with various services. A cluster organization is often a subset of the natural cluster on which its "mycelium" originated.

[1] Technical University of Liberec, Faculty of Economics, Department of Business Administration and Management, Studentska 2, 461 17 Liberec, Czech Republic, eva.stichhauerova@tul.cz.
[2] Technical University of Liberec, Faculty of Economics, Department of Business Administration and Management, Studentska 2, 461 17 Liberec, Czech Republic, miroslav.zizka@tul.cz.

The research presented in this article aims to determine whether the member companies of cluster organizations in two selected sectors are performing better than other companies. The other companies are divided into two groups. They are, on the one hand, companies which operate in the same region and the same sector as the cluster organization but are not its members. They can be understood as natural cluster entities. Another group consists of companies from the same two sectors, which are dispersed in other regions of the Czech Republic. We assume that this group of companies, that does not take advantage of the specificities of clusters, should have the lowest performance of the three groups.

The research builds on the previous results published in the article [14], which compared the performance of companies in the textile and nanotechnology sectors, but only for member companies of cluster organizations. Further research, in addition to expanding the scope to include two other sectors, adds two more groups of companies, differentiated according to their location.

2 Theoretical background

The Malmquist Index (furthermore MI), designed in 1953 by Malmquist, is a measure for assessing the change in relative productivity between different time periods. It is based on the assumption that a certain decision-making unit (DMU), which may be, for example, a company, consumes the vector m of inputs $\boldsymbol{x}^t$ and produces a vector s with outputs $\boldsymbol{y}^t$. We suppose that (x^t, y^t) is an input-output pair of a certain DMU in period t and (x^z, y^z) is an input-output pair of the same unit in period z.

Then the MI of this DMU between the periods z and t can be expressed by (1). $D^t(x^t, x^t)$ and $D^z(x^t, x^t)$ are distances between DMUs with inputs and outputs in periods t and efficient frontiers in periods t and z. $D^z(x^z, x^z)$ and $D^t(x^z, x^z)$ are distances between DMUs with inputs and outputs in the period z and efficient frontiers in periods z and t. The distance of the DMU to the frontier can be defined as the maximum possible expansion of the outputs [13].

Data Envelopment Analysis (DEA) can be used to determine the distance functions D. We measure the radial distance of input and output vectors in two periods, given the technology. The period spacing may not be unitary, it may be wider, then it is a window-analysis.

To determine the causes of changes in efficiency over time, MI can be decomposed into the product of two components – a change in the technical efficiency $EFFCH^{t,z}$ and a technological change $TECH^{t,z}$, see relation (2). The component $EFFCH^{t,z}$ is defined as the ratio of DMU efficiency in period z to its efficiency in period t, considering the frontier in the same period. It expresses the internal efforts of the DMU to improve performance [6]. These are various managerial measures to improve work organization and production efficiency. The second component $TECH^{t,z}$ shows the frontier shift effect. Value of $TECH^{t,z}$ lower than one means that the frontier in period z shifted undesirable towards the frontier in period t. Value of $TECH^{t,z}$ greater than one means the opposite. Basically, *EFFCH* points to a change in efficiency caused by a single DMU. *TECH* points to a group efficiency change induced by all DMUs [13]. Efficiency frontier shift is mainly due to industrial sector innovations. If index values are higher than one, performance, efficiency and innovation will increase. The change in technical efficiency of *EFFCH* can be further broken down into the product of change in pure technical efficiency *PECH* (under conditions of variable returns to scale) and change in scale efficiency *SECH*.

$$MI^{t,z}(x^z, y^z, x^t, y^t) = \sqrt{\frac{D^t(x^z, y^z)D^z(x^z, y^z)}{D^t(x^t, y^t)D^z(x^t, y^t)}} \tag{1}$$

$$MI^{t,z}(x^z, y^z, x^t, y^t) = \underbrace{\frac{D^z(x^z, y^z)}{D^t(x^t, y^t)}}_{EFFCH^{t,z}} x \underbrace{\sqrt{\frac{D^t(x^z, y^z)D^t(x^t, y^t)}{D^z(x^z, y^z)D^z(x^t, y^t)}}}_{TECH^{t,z}} \tag{2}$$

Since one of the outputs – economic value added – can take both positive and negative values, an input-oriented model with a variant radial measure (VRM) was used to calculate window scores [3]. Equations (3) of the model under CRS conditions are shown below. This model uses absolute values of inputs and outputs instead of their actual values. The difference $(1 - \beta)$ expresses the efficiency of a given DMU. The β indicates the degree of improvement required to achieve frontier reduction in proportional input reduction. Therefore, β is a measure of inefficiency. $\boldsymbol{X}$ is the matrix of inputs, $\boldsymbol{Y}$ is the matrix of outputs, $\boldsymbol{x_q}$ is the vector of inputs of q, $\boldsymbol{y_q}$ is the vector of outputs of q and λ is the vector of weights assigned to each DMUs.

$$\max \beta \tag{3}$$

$$\boldsymbol{X}\lambda + \beta|\boldsymbol{x}_q| \leq \boldsymbol{x}_q$$
$$\boldsymbol{Y}\lambda \geq \boldsymbol{y}_q$$
$$\lambda \geq 0$$

3 Data and methodology

Two sectors were chosen for the current research – automotive and engineering (machinery), in which cluster organizations exist at the maturity stage. The Czech Machinery Cluster was established as early as in 2003 as one of the first entities of this kind in the Czech Republic. The Moravian-Silesian Automotive Cluster was established in 2006. Both clusters have been active for quite a long time and therefore it can be expected that the positive impact on performance of member companies should occur. The research was divided into the following steps.

Step 1: Establishing a list of companies in both sectors – for each cluster, core activities were identified according to the NACE statistical classification. These are the objects of business of the main members of the cluster organization. In the case of engineering cluster, these are the activities of NACE 251 and 28x. For the automotive cluster, these are the activities of NACE 293. Companies in these sectors were divided into three groups:

1. Member companies of both cluster organizations – in the case of an engineering cluster there are 10 companies, in the automotive cluster there are 11 companies in the above defined activities;
2. Companies operating in the same region where the cluster organization operates but are not members of the cluster – engineering cluster operates in the Moravian-Silesian, Olomouc and South Moravian regions (1,011 companies operate in this sector outside the cluster), automotive cluster is active in the Moravian-Silesian Region (another 25 companies operate outside the cluster organization);
3. Companies operating in defined branches in other regions of the Czech Republic – in the engineering industry there are 1,836 companies and in the automotive industry there are 135 companies.

Step 2: Definition of inputs and outputs for Data Envelopment Analysis – the choice of inputs was influenced by the intention to expand the further research and include comparisons with other industries, such as the service sector. While it can be accepted that the indicator of the size of tangible fixed assets could be considered relevant for the application of comparisons within the manufacturing and industrial sector, its use would not be justified for the assessment of companies from the service sector (e.g. IT sector). Regardless of the economic sector, however, all the company's assets are covered by a source of financing – capital, in a certain structure. Therefore, equity and liabilities were used as inputs. Outputs were revenues from own products and services and economic value added. Economic value added (furthermore EVA) was determined in accordance with the methodology of the Ministry of Industry and Trade [8] as the product of equity and spread between return on equity and alternative cost of equity.

Step 3: Collection of data for individual groups of companies – for the sets of companies defined according to Step 1, accounting data were obtained from the balance sheets and profit and loss accounts for the time series 2009 to 2016. The data sources used were MagnusWeb [2] and collections of documents from public register [9]. Unfortunately, many companies fail to comply with the legal obligation to publish financial statements and annual reports. Some, especially smaller companies did not publish any statements or reports for the period, or there were missing data in the time series. Complete time series were obtained in the engineering industry for 6 companies from the cluster core (group labelled as EC), 167 companies in the cluster region (EN) and 198 companies from other regions (EO). The automotive industry consisted of 9 companies from the cluster core (AC), 13 companies from the region (AN) and 52 companies from other regions (AO).

Step 4: Calculation of the Malmquist index and its components – for companies (DMUs), divided into three groups (AC, AN, AO and EC, EN, EO) for each industry, the Adjacent Malmquist Index was calculated based on DEA scores and its four components (technical efficiency change *EFFCH*, technological change *TECH*, pure efficiency change *PECH* and scale efficiency change *SECH*). An input-oriented model, with radial distances and in variants with constant (CRS) and variable returns to scale (VRS), was used to determine the scale efficiency. The calculation of adjacent MI based on the DEA window-analysis is based on the principle that each DMU in a given period is independent. Thus, the performance of one and the same unit is compared between different period t and z. The width of the window determines the length of the period analyzed, in which case the window width was considered one year. Subsequently, geometric means and

medians of the Malmquist index and its components were calculated for each period and for all samples. MaxDEA 7 Ultra SW was used to calculate efficiency scores.

Step 5: Comparison of the Malmquist index and its components between samples – the Games-Howell test applied to the logarithm data was used to identify significant differences between all 6 company groups. The reason for this procedure was that the data did not come from a normal distribution and Levene's variance check test revealed that the homoscedasticity condition was violated in some companies groups. The Games-Howell post doc test is a nonparametric approach to compare multiple observation groups. It works with the order of the original values and examines the means differences of individual groups. This test can be applied with significantly different numbers of observations in groups. In order to test the significance of the differences between the geometric means of the Malmquist index and its components, the original values of the indicators were logarithmized. As [1] states, the geometric mean is a monotonic function of the mean of the logarithms, and therefore the identification of a significant difference between logarithm means indicates the existence of a significant difference between the geometric means of the original variables. All tests were performed at significance level alpha 5%.

4 Research Results

Table 1 contains, for each period and for the Malmquist index and its components, the results of the Levene's heteroscedasticity p-value test, followed by a summary of all statistically significant differences identified by the Games-Howell test across company groups. Table 1 shows the cases outlined in bold which confirm the hypothesis that the increase in the performance of companies from the core cluster is higher than the growth in the performance of non-member companies in the same region and both indicators outweigh the increase in the performance of other companies in the same sector. Groups that have shown a significant difference in the opposite direction than expected were highlighted in italics.

Year	Test	MI	EFFCH	TECH	PECH	SECH
2010/2009	Levene´s P	0.25	0.19	**<0.01**	0.11	0.67
	Games-Howell	AN>EN, AN>EO	*AN>AC, AO>AC*, EC>AC	**AN>AO**, *EO>EC*, AN>EN, AN>EO, AN>EC, EN>AO, EO>AO	*AO>AC*	AN>EN
2011/2010	Levene´s P	0.44	0.15	**<0.01**	0.11	**0.03**
	Games-Howell	×	**EN>EO**, AN>EO, AO>EO	EO>AC, EO>AO, EO>EN	**EN>EO**, AN>EO, AO>EO	×
2012/2011	Levene´s P	0.32	0.13	**<0.01**	**0.03**	0.10
	Games-Howell	×	×	AN>EO, AO>EO	×	×
2013/2012	Levene P	0.38	0.18	**<0.01**	**0.04**	0.09
	Games-Howell	**AC>AO**	*EO>EN*, EO>AN	**AN>AO**, **EN>EO**, *EN>EC*, AN>EN, AN>EO, AN>EC, EN>AO,	×	×
2014/2013	Levene´s P	0.17	0.26	**<0.01**	0.09	0.14
	Games-Howell	AC>EN, AO>EN	**AC>AN**, *AO>AN*, EN>AN, EO>AN, EC>AN	*AN>AC*, **AN>AO**, AC>EN, AC>EO, AN>EN, AN>EO, AO>EN, AO>EO	EN>AN, EC>AN	*EO>EN*
2015/2014	Levene P	0.03	**0.01**	**<0.01**	**<0.01**	0.30
	Games-Howell	*EN>EC, EO>EC*, AO>EC	**AN>AO**, AN>EN, AN>EO, AN>EC	*EN>EC*, *EO>EC*, AO>EO, AO>EC	*EN>EC*, *EO>EC*, AO>EO, AO>EC	×

Year	Test	MI	EFFCH	TECH	PECH	SECH
2016/2015	Levene´s P	0.65	0.27	**<0.01**	**0.03**	**0.03**
	Games-Howell	**AN>AO**	**EN>EO**	*AN>AC*, AN>EN, AN>EO	**AN>AO**	AC>EO, AO>EO

Table 1 Significant differences between groups of companies in individual years

It can be stated that it was only partially possible to confirm the assumption stated in the introduction of the article on higher performance of clusters. This assumption was confirmed only in some years and some performance indicators, more often in the automotive sector (for 8 pairs of compared groups). The AC group achieved significantly higher MI values in 2013/2012 than the AO group, and in 2014/13 higher *EFFCH* values compared to the AN group. The case where the AN group showed on average higher efficiency than the AO group was relatively more frequent: the MI value was higher for the AN group in 2016/2015; higher *EFFCH* value in 2015/2014. Significantly higher value of the *TECH* indicator in the AN group compared to AO was proved in 2010/2009, 2013/2012 and 2014/2013.

Fewer statistically different pairs were identified among the groups in the engineering industry to confirm the assumption (only 4 in total). Significantly higher average value for the EN group compared to the EO was identified in the following cases: for the *EFFCH* indicator for the period of 2011/2010 and 2016/2015, for the *TECH* indicator for the period of 2013/2012 and finally for the *PECH* indicator for the period of 2011/2010.

Table 2 shows the geometric means and median values of MI and its components in individual company groups. Table 3 then adds information on which groups of companies show significant differences. It is apparent from Table 2 that performance was growing faster in the automotive sector over the whole reporting period. On average, by around 4% per year, while in the engineering sector just by around half a percent per year. Internal technical efficiency was the main source of performance growth in the automotive industry. In the engineering sector, the low increase in performance was only due to an improvement in internal efficiency; the technological component decreased by more than half a percent per year. In terms of scale of efficiency, there was a slight decrease in both sectors.

From the perspective of individual groups, it is clear that the performance of naturally clustered companies in the automotive industry grew the fastest (by almost 11% per year). The drivers of performance in this group were mainly innovations. It can be seen from Table 3 that in this respect, the group of naturally clustered firms in the automotive industry significantly outperformed naturally clustered firms from the engineering industry (and also non-clustered engineering companies). The weakest performance growth was then reported by other automotive companies (the difference between AN and AO is significant, see Table 3), where the growth was driven only by an improvement in internal efficiency.

Group	MI		EFFCH		TECH		PECH		SECH	
	AVG	Med.	AVG	Med.	AVG	Med.	AVG	Med.	AVG	Med.
AC	1.082	1.063	1.018	1.000	1.063	1.030	1.005	1.000	1.013	1.000
AN	1.108	1.095	1.020	1.000	1.086	1.171	1.045	1.000	0.977	1.000
AO	1.011	1.000	1.030	1.000	0.981	1.001	1.050	1.000	0.981	1.000
Automotive total	**1.036**	**1.020**	**1.027**	**1.000**	**1.009**	**1.032**	**1.043**	**1.000**	**0.984**	**1.000**
EC	1.073	0.993	1.104	1.000	0.972	0.966	1.020	1.000	1.082	1.000
EN	0.984	1.000	1.006	1.000	0.978	0.999	1.103	1.000	0.913	0.991
EO	1.021	1.013	1.015	1.000	1.006	1.000	1.061	1.000	0.957	1.000
Engineering total	**1.005**	**1.002**	**1.013**	**1.000**	**0.993**	**1.000**	**1.079**	**1.000**	**0.939**	**0.999**
Total	**1.010**	**1.007**	**1.015**	**1.000**	**0.995**	**1.000**	**1.073**	**1.000**	**0.946**	**1.000**

Table 2 Malmquist index summary of group geometric means and medians for period 2016/2009

In the engineering sector, companies in the cluster organization showed the strongest performance growth, averaging over 7% per year. The source of performance growth in this group was the improvement of in-

ternal efficiency and especially scale efficiency. In the group of non-member companies and in other companies, the performance growth was very weak, or in the EN group, performance even declined. However, the differences in performance development between groups of engineering companies are not significant.

Tests	MI	EFFCH	TECH	PECH	SECH
Levene´s P	<0.01	<0.01	<0.01	<0.01	<0.01
Games-Howell differences	**AN>AO**	x	**AN>AO** AN>EN AN>EO	x	x

Table 3 Significant differences between company groups for the period of 2016/2009

5 Conclusions

In terms of the impact of clustering on the development of corporate performance, it can be stated that a certain positive trend was demonstrated. In the case of the automotive industry, natural clustering seems more favorable. This means that companies are using existing industrial links and facilities in the region, including the presence of research and education institutions, to boost their performance growth. The establishment of a cluster organization has not yet contributed to further growth improvement. On the other hand, companies operating "dispersed" outside the region showed only very poor performance improvements.

In the engineering industry, the establishment of a cluster organization for enterprises might seem to be beneficial (see faster performance growth in Table 2 compared to other company groups). The effect of a cluster organization was particularly evident in the area of scale efficiency. However, for a small number of companies in a cluster organization sample, the difference compared to other groups is not significant.

Compared to previous research [14], it can be stated that the different impact of clustering across industries and its different impact on individual performance components are confirmed. In the textile sector, the cluster organization had a positive impact, in particular in promoting technological growth. In the nanotechnology sector, scale efficiency has improved and, as a result, internal technical efficiency has improved. Recent research shows the positive impact of clustering on technological progress in the automotive industry. Moreover, in engineering, economies of scale with a positive impact on internal efficiency prevail. At the same time, it appears that in terms of public support for clusters, it is necessary to consider when it makes sense to support the establishment of cluster organizations from the outside. Indeed, in some sectors, natural clustering can yield better results than deliberately encouraged establishment of cluster organizations.

Acknowledgements

Supported by the grant No. GA18-01144S "An empirical study of the existence of clusters and their effect on the performance of member enterprises" of the Czech Science Foundation.

References

[1] ALF, E. F. & GROSSBERG, J. M. (1979). The geometric mean: Confidence limits and significance tests. *Perception &Psychophysics*, *26,* 419–421.

[2] BISNODE: *Magnusweb: Komplexní informace o firmách v ČR a SR* [online]. Bisnode ČR, Prague, 2020 [cit. 2020-03-13]. Available at: https://magnusweb.bisnode.cz.

[3] CHENG, G., ZERVOPOULOS, P., & QIAN, Z. (2013). A variant of radial measure capable of dealing with negative inputs and outputs in data envelopment analysis. *European Journal of Operational Research*, *225*, 100–105.

[4] KUKALIS, S. (2010). Agglomeration Economies and Firm Performance: The Case of Industry Clusters. *Journal of Management*, *36*, 453–481.

[5] LEI, H. & HUANG, CH. (2014). Geographic clustering, network relationships and competitive advantage: Two industrial clusters in Taiwan. *Management Decision*, *52*, 852–871.

[6] LI, Z., CROOK, J., & ANDREEVA, G. (2017). Dynamic prediction of financial distress using Malmquist DEA. *Expert Systems with Applications, 80*, 94–106.

[7] LINDQVIST, G., KETELS, CH., & SÖLVELL, Ö. (2013). *The Cluster Initiative Greenbook 2.0.* Stockholm: Ivory Tower Publishing.
[8] MPO: *Finanční analýza podnikové sféry za rok 2017.* Ministry of Industry and Trade, Prague, 2018.
[9] MSP: *Veřejný restřík a Sbírka listin* [online]. Ministry of Justice CR, Prague, 2015 [cit. 2020-03-13]. Available at: https://or.justice.cz/ias/ui/rejstrik.
[10] PAVELKOVÁ, D., ET AL. (2019). *Towards Smart Cluster Policies in V4 Countries.* Zlín: Tomas Bata University.
[11] PORTER, M. E. (1998). Clusters and the New Economics of Competition. *Harvard Business Review*, *76*, 77–90.
[12] RULAND, W. (2013). Does Cluster Membership Enhance Financial Performance? *iBusiness*, *5*, 1–11.
[13] WANG, D. D. (2019). Performance assessment of major global cities by DEA and Malmquist index analysis. *Computers, Environment and Urban Systems, 77*, 101365.
[14] ŽIŽKA, M. (2019). Inter-Branch Comparison of Cluster Company Performance Using Malmquist Index. In M. HOUDA & R. REMEŠ (Eds.), *37th International Conference on Mathematical Methods in Economics 2019. Conference Proceedings* (pp. 37–42) České Budějovice: University of South Bohemia in České Budějovice.

On the Issue of Comparisons with Reference Fuzzy Numbers – a Graphical Method to Guide the Choice of a Suitable Method in Managerial Decision-making

Tomáš Talášek[1], Jan Stoklasa[2]

Abstract. This paper proposes a mathematical method for the analysis of the behavior of similarity measures of fuzzy numbers in the task of comparing a fuzzy output of a mathematical model with a reference fuzzy number. It is based on the visualisation of the set of all fuzzy numbers similar with the reference fuzzy number. This approach provides the user with the information concerning the whole set of objects that are considered equally similar to the reference object on the given level of similarity. This novel approach allows the used to see how compact or diverse the set of "equally similar" objects is. It thus allows the user to see how reasonable it might be to rely on the given similarity measure of fuzzy numbers in the given application problem. The proposed method is intended for laymen users of fuzzy model, while providing valuable insights for the designers of these models as well. The selection of appropriate similarity measure can be beneficial in a wide set of practical problems using models that provide fuzzy outputs, ranging from evaluation tasks through the ordering of alternatives (managerial decision-making) to fuzzy optimization.

Keywords: similarity measure, reference fuzzy number, visualisation, comparison

JEL Classification: D81, C44
AMS Classification: 90B50, 91B06

1 Introduction

The ability to compare two objects is crucial in economics and management. When these objects are the quantities, values of variables or even complex formal representations of alternatives provided by mathematical models, the comparison can get complicated. One of the reasons might be the abstraction required from mathematical models, which translates into the necessity of comparing mathematical entities instead of actual well known objects (see e.g. [3]). Another issue potentially complicating the comparisons might lie in the possibility to choose different measures of "similarity" or "closeness" of the compared entities. The possibility to choose an appropriate measure of e.g. distance or similarity of the objects can easily translate into a hard choice problem, as roadmaps for the selection of appropriate measures or suggestions of "measures of choice" for different situations are not widely available. For example Talášek points out this problem and suggests some solutions thereof in [8]. The problem of the selection of appropriate methods for the comparison of fuzzy objects and quantities has recently been receiving some attention, but in general this topics remains understudied.

This paper aims to contribute to the literature on the selection of appropriate methods for the comparisons of fuzzy objects by suggesting an analytical framework stemming from [8, 10, 7] the results of which can be represented graphically. This makes the method applicable also by laymen and beginners in the field of fuzzy modeling. As such it aims to ease one of the well known problems connected with fuzzy modelling, i.e. the freedom to choose appropriate measures (but also the necessity to do so correctly, or at least not to choose completely inappropriate methods). The proposed method approaches the whole problem from a novel perspective by asking the question *"What (other) fuzzy numbers from the given family have the same similarity with (or distance from) the reference fuzzy number?"*. This allows for the assessment of the possible diversity within the set of fuzzy numbers "equally similar to the reference one". When e.g. the diversity is too

[1] Palacký University Olomouc, Faculty of Education, Department of Mathematics, Žižkovo nám. 5, 77140 Olomouc, Czech Republic and Lappeenranta-Lahti University of Technology LUT/School of Business and Management, Yliopistonkatu 34, 53850 Lappeenranta, Finland, tomas.talasek@upol.cz.
[2] Lappeenranta-Lahti University of Technology LUT/School of Business and Management, Yliopistonkatu 34, 53850 Lappeenranta, Finland and Palacký University Olomouc, Faculty of Arts, Department of Applied Economics, Křížkovského 12, 77900 Olomouc, Czech Republic, jan.stoklasa@lut.fi.

high or when the fuzzy numbers in this set are of very different cardinalities and this fact is counter-intuitive in the given application task, the particular method should be discarded.

2 Preliminaries

Let U be a nonempty set (*the universe of discourse*). A *fuzzy set* A on the universe U is defined by the mapping $A : U \to [0,1]$. A family of all fuzzy sets on U is denoted by $\mathcal{F}(U)$. For each $x \in U$ the value $A(x)$ is called the *membership degree* of the element x in the fuzzy set A and $A(.)$ is called a *membership function* of the fuzzy set A. Let A be fuzzy set on the same universe U. The set $\mathrm{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes the *kernel* of A, $A_\alpha = \{x \in U | A(x) \geq \alpha\}$ denotes an α-*cut* of A for any $\alpha \in [0,1]$, $\mathrm{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A. $\mathrm{Hgt}(A) = \sup\{A(x) | x \in U\}$ denotes a height of fuzzy set.

A *fuzzy number* is a fuzzy set A defined on the set of real numbers which satisfies the following conditions: $\mathrm{Ker}(A) \neq \emptyset$ (A is *normal*); A_α are closed intervals for all $\alpha \in (0,1]$; and $\mathrm{Supp}(A)$ is bounded. A fuzzy number A is said to be defined on $[a,b] \subset \mathbb{R}$, if $\mathrm{Supp}(A)$ is a subset of the interval $[a,b]$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1, a_4] = \mathrm{Cl}(\mathrm{Supp}(A))$ and $[a_2, a_3] = \mathrm{Ker}(A)$, where $\mathrm{Cl}(\mathrm{Supp}(A))$ denotes a closure of $\mathrm{Supp}(A)$.

The fuzzy number A is called *trapezoidal* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$ and $a_1 \neq a_4$; for such fuzzy numbers we will use a simplified notation $A \sim (a_1, a_2, a_3, a_4)$. Trapezoidal fuzzy number A for which $a_2 = a_3$ is called *triangular* fuzzy number and can be denoted by shortened notation $A \sim (a_1, a_2, a_4)$. Fuzzy set A_G on U is called *generalized trapezoidal fuzzy number* if there exists a trapezoidal fuzzy number A and $w_A \in [0,1]$ for which $A_G(x) = w_A \cdot A(x), x \in U$. Let A be a fuzzy number on $[a,b]$ for which $a_1 \neq a_4$. Then the *center of gravity* of A on $[a,b]$ is defined by the formula $\mathrm{COG}(A) = \int_a^b xA(x) / \int_a^b A(x)dx$. For more information about fuzzy sets see e.g. [5].

3 Selected similarity measures used for comparison

In this paper two similarity measures for generalized fuzzy numbers were selected. These measures were already studied in several previous study (e.g. [7, 12, 11, 10]) and this paper provides extension of previous findings. Because the numerical experiment will be performed on non-generalized fuzzy numbers, presented formulas are adjusted for the sake of clarity.

Let $A \sim (a_1, a_2, a_3, a_4)$ and $B \sim (b_1, b_2, b_3, b_4)$ be trapezoidal fuzzy numbers on $[0,1]$, then the studied similarity measures are:

- *similarity measure* s_1 proposed by Wei and Chen in 2009 [13]:

$$s_1(A,B) = \left(1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}\right) \cdot \frac{\min\{Pe(A), Pe(B)\} + 1}{\max\{Pe(A), Pe(B)\} + 1}, \tag{1}$$

  where $Pe(A) = \sqrt{(a_1 - a_2)^2 + 1} + \sqrt{(a_3 - a_4)^2 + 1} + (a_3 - a_2) + (a_4 - a_1)$ is the perimeter of A, $Pe(B)$ is defined analogously.
- *similarity measure* s_2 proposed by Hejazi, Doostparast and Hosseini in 2011 [2]:

$$s_3(A,B) = \left(1 - \frac{\sum_{i=1}^{4} |a_i - b_i|}{4}\right) \cdot \frac{\min\{Pe(A), Pe(B)\}}{\max\{Pe(A), Pe(B)\}} \cdot \frac{\min\{Ar(A), Ar(B)\} + 1}{\max\{Ar(A), Ar(B)\} + 1}, \tag{2}$$

  where $Ar(A) = \frac{1}{2} \cdot (a_3 - a_2 + a_4 - a_1)$ is area of A, $Ar(B)$ area of B and is computed analogously, $Pe(A)$ and $Pe(B)$ are computed identically as in the s_1.

Although the numerical experiment is performed on these two similarity measures, any similarity measure for (symmetrical triangular) fuzzy numbers can be used instead (e.g. [1, 14, 6, 4]).

4 The proposed graphical method and an example of its use via numerical experiment

In this section the new graphical method for the analysis of behaviour of similarity measures of fuzzy numbers is proposed. The method was designed to provide easy-to-understand graphical comparison of similarity measures with respect to selected reference fuzzy number, which represents a general application context.

For the purposes of the numerical experiment three triangular fuzzy numbers $R_1 \sim (0.4, 0.5, 0.6), R_2 \sim (0.4, 0.5, 0.5)$ and $R_3 \sim (0, 0.5, 1)$ were selected as a *reference fuzzy numbers*. The goal is to observe the similarity that these prototype fuzzy numbers have with other fuzzy numbers on $[0, 1]$ interval. To be more specific the goal is to visualize different sets of fuzzy numbers that have the same similarity with the reference one. To enable an easy-to-read visualization of the results, only symmetrical triangular fuzzy number are compared with these prototype fuzzy number. This selection is based on a property that symmetrical triangular fuzzy number $A \sim (a_1, a_2, a_3)$ can be unambiguously represented by two-tuple $(a_2, a_3 - a_1)$, where a_2 represents the center of gravity and $a_3 - a_1$ represents the length of the support of fuzzy number A (see [10, 8]). Therefore, each fuzzy can be represented as a point in a two-dimensional space.

Although it is possible to randomly generate two-tuples that represents symmetrical triangular fuzzy numbers from $[0, 1]$ and use them for the comparison (see e.g. [9]), this paper utilizes the *grid approach* (see [10, 8]) that enables us to generate fuzzy numbers in a way that the representation of these fuzzy numbers uniformly cover the $[0, 1] \times [0, 1]$ space. Both $[0, 1]$ intervals are uniformly divided into 1 001 points - each point represents center of gravity/length of the support of fuzzy number. Then using the Cartesian product we obtain 1 002 001 two-tuples such that each represents a different symmetrical triangular fuzzy number. However, this set contains also fuzzy numbers that are not defined on $[0, 1]$ interval. After these fuzzy numbers are excluded, we obtain set $Out = \{O_1, \ldots, O_{500\,000}\}$ that contains half a million of a "uniformly distributed" symmetrical triangular fuzzy numbers on $[0, 1]$.

Both similarity measures s_1 and s_2 between each reference fuzzy number R_1, R_2, R_3 and each symmetrical triangular fuzzy number from the set Out are calculated. For each reference fuzzy numbers we iteratively select symmetrical triangular fuzzy numbers from the set Out for which $s_m(R_j, O_i) = k$, where $m = 1, 2$; $j = 1, 2, 3$; $i = 1, \ldots, 500\,000$; $k = 0,\ 0.05,\ 0.1, \ldots, 1$ are stored in the set $Out_k^{s_m, R_j}$. Also, a unique color is assigned to each value of k. Finally, the last step is the final visualization of the results. For similarity measure s_m and reference fuzzy number R_j the fuzzy numbers from the set $Out_i^{s_m, R_j}$ are depicted as points in two dimensional space using the color that was assigned to value k. This way we obtain the "contour line" representing the fuzzy numbers that have the same similarity with the respective reference fuzzy number[1]. Results for both similarity measures and all the reference fuzzy numbers considered in this paper are depicted in Figure 1. Please note that the results form a triangle in the two dimensional space (depicted by the black line) that represents all the triangular symmetrical fuzzy numbers on the [0, 1] interval. Area outside of this triangle represents symmetrical triangular fuzzy numbers, that are defined outside the $[0, 1]$ interval.

Figure 1 allows users of the graphical model (including laymen/non-experienced users) to easily compare the behaviour of selected similarity measures with respect to the selected reference fuzzy number. The main advantage over the "traditional approach" that is typically based on listing several properties of similarity measures is that the user receives the "whole picture". Several observations from the Figure 1 can be concluded:

- From the managerial point of view, both similarity measures s_1 and s_2 are not convenient for comparisons, when the reference fuzzy number is R_1 or R_2. In each of these cases, there exist two fuzzy numbers that have the same similarity measures with $R_1(R_2)$, have the same center of gravity but the length of the support of these two fuzzy numbers is different. Because the length of the support is proportional to the cardinality of fuzzy numbers[2] which in managerial models represent the uncertainty, there is a danger, that the model evaluates two alternatives as equally good/bad even though these alternatives have the same "mean value" (represented by the center of gravity), but different uncertainty. It can also be observed that the case where two different fuzzy numbers with the same center of gravity but different uncertainty have the same similarity to the reference fuzzy number is more frequent for s_2 than for s_1.
- On the other hand, if the reference fuzzy number is R_3, the issue from the previous item does not occur. If two symmetrical triangular fuzzy numbers have identical similarity with fuzzy number R_3, they do not have the same center of gravity.
- The case with the reference fuzzy number R_2 clearly shows, how the choice of an asymmetrical triangular fuzzy number as a reference fuzzy number affects the resulting similarity measures. Therefore the proposed method provides a powerful tool that can be used to examine how the shape of the reference fuzzy number can affect the behaviour of the mathematical model, and thus also the appropriateness of the similarity measures considered therein.
- Similarity measure s_2 differentiates "faster" between fuzzy numbers. This can be clearly seen from the

[1] To provide an easy-to-read visualization of the results, it is recommended to round the resulting similarity measures to two decimal places.

[2] The length of the support of triangular fuzzy numbers is equal to half the cardinality of the fuzzy number.

number of "contour lines" that is higher in the case of similarity measure s_2 for each reference fuzzy number. Therefore the similarity s_2 is more suitable in the models, where their users demand higher differentiation between alternatives.

The proposed graphical method enables us to easily compare the behaviour of two similarity measures and identify cases when these similarity measure can be applied but also the cases when their use should be avoided.

5 Conclusion

In this paper we have presented and analysed an analytical method for the visualisation of subsets of the set of fuzzy numbers of the given type that have the same similarity with the chosen reference fuzzy number. As can be seen in Section 4 the graphical outputs provided by the method are self explanatory and can be used and understood by non-professionals in the field of fuzzy modeling after only a short explanation/introduction of the outputs. Within the framework of our numerical experiment, we have identified a potential problem with the two analyzed similarity measures for the purpose of managerial decision-making. The experiment has shown that two different outputs with the same COG but different uncertainty can have the same similarity with the reference fuzzy numbers in two of the three cases of reference fuzzy numbers. From this one can clearly see that the suitability of a similarity measure is problem and context dependent. It is not correct to conclude that the analyzed measures are problematic overall, as there are settings (e.g. represented by the choice of the reference fuzzy number as R_3) where the identified problems with the similarities are no longer an issue. The main added value of the proposed method is that such a conclusion can be reached by a non-professional fuzzy analyst. The method therefore provides means for a specific type of sensitivity analysis of the suitability of the similarity measures for the given problems/tasks.

Acknowledgements

This research would like to acknowledge the funding received from the Finnish Strategic Research Council, grant number 313396 / MFG40 – Manufacturing 4.0 and by the Specific University Research Grant IGA_FF_2020_001, as provided by the Ministry of Education, Youth and Sports of the Czech Republic in the year 2020.

References

[1] Chen, S.-J., and Chen, S.-M.: Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *IEEE Transactions on Fuzzy Systems* **11** (2003), 45–56.
Retrieved from `http://ieeexplore.ieee.org/document/1178065/`.

[2] Hejazi, S., Doostparast, A., and Hosseini, S.: An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers. *Expert Systems with Applications* **38** (2011), 9179–9185.
Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S0957417411001217`.

[3] Jandová, V., Krejčí, J., Stoklasa, J., and Fedrizzi, M.: Computing Interval Weights for Incomplete Pairwise-Comparison Matrices of Large Dimension—A Weak-Consistency-Based Approach. *IEEE Transactions on Fuzzy Systems* **25** (2017), 1714–1728.
Retrieved from `http://ieeexplore.ieee.org/document/7762089/`.

[4] Khorshidi, H. A., and Nikfalazar, S.: An improved similarity measure for generalized fuzzy numbers and its application to fuzzy risk analysis. *Applied Soft Computing* **52** (2017), 478–486. Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S1568494616305336`.

[5] Klir, G. J., and Yuan, B.: *Fuzzy Sets and Fuzzy Logic: Theory and Aplications*. Prentice Hall, New Jersey, 1995.

[6] Patra, K., and Mondal, S. K.: Fuzzy risk analysis using area and height based similarity measure on generalized trapezoidal fuzzy numbers and its application. *Applied Soft Computing* **28** (2015), 276–284. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S1568494614006061`.

[7] Stoklasa, J., Talášek, T., and Stoklasová, J.: Executive summaries of uncertain values close to the gain/loss threshold – linguistic modelling perspective. *Expert Systems with Applications* **145** (2020), 113108.
Retrieved from `https://www.sciencedirect.com/science/article/pii/S0957417419308255`.

[8] Talášek, T.: *The linguistic approximation of fuzzy models outputs*. Ph.D. thesis, Lappeenranta-Lahti University of Technology LUT, Lappeenranta, 2019.

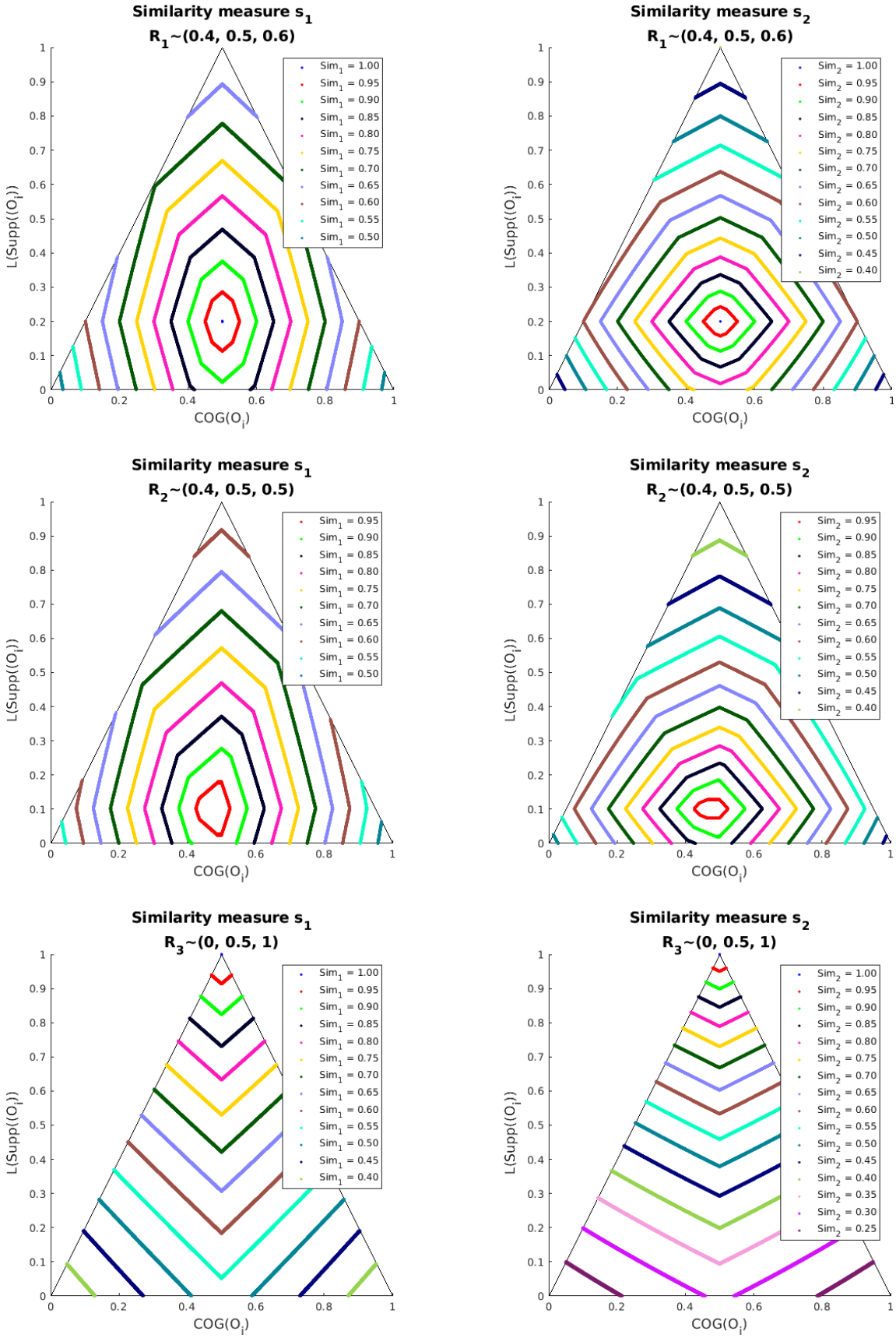


Figure 1 The result of the numerical experiment representing the different behaviour of similarity measures s_1 (left) and s_2 (right) with respect to different reference fuzzy numbers $R_1 \sim (0.4, 0.5, 0.6)$ (top), $R_2 \sim (0.4, 0.5, 0.5)$ (middle) and $R_3 \sim (0, 0.5, 1)$ (bottom). Each color represents fuzzy numbers (represented as a two-tuple in two dimensional space) that have the same similarity measure with selected reference fuzzy number.

[9] Talášek, T., and Stoklasa, J.: The role of distance/similarity measures in the linguistic approximation of triangular fuzzy numbers. In: *Proceedings of the international scientific conference Knowledge for Market Use 2016* (Slavíčková, P., ed.). Societas Scientiarum Olomucensis II, Olomouc, 539–546.

[10] Talášek, T., and Stoklasa, J.: A Numerical Investigation of the Performance of Distance and Similarity Measures in Linguistic Approximation under Different Linguistic Scales. *Journal of Multiple-Valued Logic and Soft Computing* **29** (2017), 485–503.

[11] Talášek, T., and Stoklasa, J.: Ordering of fuzzy quantities with respect to a fuzzy benchmark – how the shape of the fuzzy benchmark and the choice of distance / similarity affect the ordering. In: *Proceedings of the 36th International Conference on Mathematical Methods in Economics* (Váchová, L., and Kratochvíl, O., eds.). MatfyzPress, Jindřichův Hradec, 573–578.

[12] Talášek, T., and Stoklasa, J.: Comparison of similarity measures for generalized trapezoidal fuzzy numbers. In: *Proceedings of the 37th International Conference on Mathematical Methods in Economics* (Houda, M., and Remeš, R., eds.). University of South Bohemia in České Budějovice, České Budějovice, 482–486.

[13] Wei, S.-H., and Chen, S.-M.: A new approach for fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. *Expert Systems with Applications* **36** (2009), 589–598.
Retrieved from `http://linkinghub.elsevier.com/retrieve/pii/S0957417407004630`.

[14] Wen, J., Fan, X., Duanmu, D., and Yong, D.: A modified similarity measure of generalized fuzzy numbers. *Procedia Engineering* **15** (2011), 2773–2777.
Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S1877705811020236`.

Minimising Number of Shunting Tracks at Tram Turning Loops

Dušan Teichmann[1], Michal Dorda[2], Denisa Mocková[1], Pavel Edvard Vančura[1], Vojtěch Graf[2], Ivana Olivková[2]

Abstract. Tram transport infrastructure and its subsequent operation is a very costly matter for each transport operator. It is therefore purposeful for the infrastructure to be built only to the extent which is necessary. The article deals with the optimisation of the necessary number of shunting tracks under predefined volume of tram transport which is given by the number of tram lines and a frequency of their connections. This is a problem that is specific to rail transport, when changing the order of individual vehicles is organizationally complicated and is feasible only with appropriate construction of the rail yard (so called turning loops), in some cases it is not possible at all. One of the ways how to reduce the number of necessary shunting tracks at tram turning loops is to adjust the arrival and departure times of tram vehicles serving the individual lines – time coordination. Mathematical programming will be used to deal with the problem.

Keywords: Optimisation, tram transport, turning loop, tram turnarounds, shunting tracks

JEL Classification: C44
AMS Classification: 90C15

1 Introduction – motivation to solve the problem

In public transport, tram transport is one of the most demanding in terms of the infrastructure costs. Therefore, it is necessary to build the tram infrastructure on the operated network only to the extent which is necessary. For the current volume of traffic (especially in large cities), the situation cannot be solved by building single-track tram lines with passing loops, because such tram tracks have problems with their safety, capacity and in case of an infrastructure problem, traffic may be delayed and it may take a long time to eliminate the delay.

One of the ways to save the costs for the construction and operation of the tram infrastructure is to minimise the railway infrastructure which is located outside the tram tracks between the final stops. Such unnecessary infrastructure can be found either in tram depots or at turning loops. One of the possibilities to achieve the minimum number of the necessary shunting tracks at the turning loop is time coordination of arrivals and departures of connections of individual tram lines using the turning loop. In our article, we continue in the article [7], which dealt with the minimisation of the number of the used shunting tracks at the tram turning loop equipped with two tracks under the assumption that two tram lines use the turning loop.

In the presented article, we deal with the minimisation of the number of the shunting tracks at the tram turning loops which are equipped with higher numbers of the shunting tracks (more than 2) and with higher volumes of tram traffic.

2 Analysis of the state of the art

Many articles have been published on the topic of time coordination in the past. We can state that prof. Jan Černý had systematically dealt with many issues of time coordination. He published his findings concerning

[1]CTU in Prague, Faculty of Transportation Studies, Department of Logistics and Management of Transport, Konviktská 20, 110 00 Prague, teichdus@fd.cvut.cz, mockova@fd.cvut.cz, vancurap@dpp.cz

[2]VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, 708 33 Ostrava Poruba, michal.dorda@vsb.cz, vojtech.graf@vsb.cz, ivana.olivkova@vsb.cz

node and section (line) coordination mainly in publications [1], [2] and [3]. In these publications it is possible to find many methods especially for solving time coordination in isolated nodes and isolated sections of the transport networks.

Some authors followed in the further development of the methods presented in the above-mentioned publications. We can mention for example publications [5] and [6], which dealt with more complicated assumptions on the sections of the transport networks, such as simultaneous coordination in both directions. Another publication which is related to the problem is publication [4]. However, none of the referenced publications deals with the application of the principles of time coordination for the tram infrastructure at the turning loops. Our article was thematically preceded by publication [7].

3 Technological assumptions applied in the proposed model

When creating the model, we considered the following technological assumptions:

1. We assume that the presented mathematical model should be applied mainly in the phase when a new tram transport timetable is being planned.
2. We consider that all the tram lines are operated according to a regular-interval timetable with the same interval for all the lines. Therefore, it holds that the optimisation problem regularly repeats after elapsing the time that corresponds to the tram interval.
3. The time of the arrival at the final stop and the time when parking of the tram on the track at the turning loop begins are different in real conditions. The exact time, which we consider to be the beginning of the parking of each tram, can be calculated by adding the time required for passengers to get out the tram plus the driving time of the tram to the parking place at the turning loop to the actual arrival time of the tram at the final stop according to the timetable. An analogous assumption is used for the difference between the time the tram leaves its parking place at the turning loop and the time when the tram departs from the turning loop to serve a new connection. For simplicity, let us consider that the actual arrival time of the tram at the final stop is the same as the time of parking of the tram on any track of the turning loop. However, we will increase the minimum turnaround time at the final stop by the time the passengers need to get out or board the tram and the driving time of the tram to the parking place at the turning loop. The minimum turnaround time may be further increased by a time reserve for the elimination of the delays.
4. When a tram fleet is homogeneous from the point of view of the tram type, then the capacity of each shunting track at the turning loop can be expressed by the total number of the trams that can be parked on the track at the same time. When the tram fleet is heterogeneous, then the capacity of the shunting track can be expressed by the number of so-called unit trams that can be parked on the track at the same time. The number of the unit trams can be calculated as follows. One tram type is chosen (usually the one with the shortest or the most common length) and its length determines the unit tram. The trams of other types are converted to the unit tram by dividing their length by the length of the unit tram. If more than one tram type is used on the same tramline, then it is necessary for the purposes of capacity calculations to consider the tram type that is determined by the maximum number of the unit trams. In the following text, let us assume a more general variant – the heterogeneous tram fleet serving the connections of the tram lines using the turning loop.
5. We assume that only single tram serving the tram line use the turning loop at the same time – that means it is not possible to park more than one tram for the same tram line at the turning loop.
6. We consider that it may happen that several trams serving the different tram lines arrive at the turning loop at the same time. An analogous situation may also happen for the departures of the trams from the turning loop.

4 Problem formulation

Let us consider a tram turning loop consisting of a set K of shunting tracks. A set I of tram lines which use the turning loop is defined. For each tram line, the earliest possible arrival time of the tram serving the connection of the tram line – t_i^{ear}, the earliest possible departure time of the subsequent connection of the same tram line in the opposite direction served by the same vehicle – T_i^{ear}, the minimum turnaround time of the tram at the turning loop – ω_i, the maximum allowed postponement of the arrival time of the line $i \in I$ – a_i and the maximum allowed postponement of the departure time of the subsequent connection of the line $i \in I$ in the opposite direction served by the same tram – b_i. For each track $k \in K$ of the turning loop its

capacity c_k is known, the capacity is given by the maximum number of the unit trams that can be found on the track at the same time.

The task is to decide on the time shifts of the arrivals of the individual connections at the final stop (before entering the turning loop) and on the time shifts of the departures of the individual connections from the final stop (after leaving the turning loop) and on the assignment of the shunting tracks to the trams of the individual lines so that the proposed time shifts take their values within pre-defined limits, the minimum turnaround times of the trams at the turning loops are satisfied and the total number of the shunting tracks being used by the trams for their turnarounds is as minimal as possible.

5 Mathematical model

Let us first summarise all the sets, constants and variables used in the proposed mathematical model.

The sets:

K The set of the shunting tracks of the turning loop.

I The set of the tram lines which use the turning loop for their turnarounds.

The constants:

t_i^{ear} The earliest possible arrival time of the connection of the line $i \in I$ at the final stop.

T_i^{ear} The earliest possible departure time of the consequent connection of the line $i \in I$ in the opposite direction served by the same tram.

ω_i The minimal turnaround time of the line $i \in I$ at the turning loop.

a_i The maximum allowed time shift of the arrival of the connection of the line $i \in I$ at the final stop.

b_i The maximum allowed time shift of the departure from the final stop of the consequent connection of the line $i \in I$ in the opposite direction served by the same tram.

c_k The capacity of the track $k \in K$.

κ_i The length of the tram serving the connection of the line $i \in I$ expressed in unit trams.

M A large enough constant.

The variables:

z_{ki} A binary variable which models the occupation of the track $k \in K$ by the tram of the line $i \in I$. If it holds that $z_{ki} = 1$, then the tram serving the connection of the line $i \in I$ will occupy the track $k \in K$ after its arrival at the turning loop, in the opposite case $z_{ki} = 0$.

w_k A binary variable modelling the occupation of the track $k \in K$. If it holds that $w_k = 1$, then the track $k \in K$ will be used for the turnarounds of some tram lines, in the opposite case $w_k = 0$.

u_{kij} An auxiliary binary variable – when the connection of the line $i \in I$ arrived at the turning loop and began to occupy the track $k \in K$ before the arrival of the connection of the line $j \in I$, then $u_{kij} = 1$, in the opposite case $u_{kij} = 0$.

v_{kij} An auxiliary binary variable – when the connection of the line $i \in I$ departed from the track $k \in K$ before the departure of the connection of the line $j \in I$, then $v_{kij} = 1$, in the opposite case $v_{kij} = 0$.

x_{ki} The time shift of the arrival of the connection of the line $i \in I$ on the track $k \in K$ (taken from the earliest possible arrival time).

y_{ki} The time shift of the departure of the connection of the line $i \in I$ from the track $k \in K$ (taken from the earliest possible departure time).

The mathematical model that minimises the number of the tracks at the turning loop by means of time coordination of the arrival and departure times can be written in the following form:

$$\min f(u, v, w, x, y, z) = \sum_{k \in K} w_k \tag{1}$$

subject to:

$$\sum_{k \in K} z_{ki} = 1 \qquad i \in I \qquad (2)$$

$$t_i^{ear} z_{ki} + x_{ki} \leq M w_k \qquad k \in K, i \in I \qquad (3)$$

$$T_i^{ear} z_{ki} + y_{ki} \leq M w_k \qquad k \in K, i \in I \qquad (4)$$

$$t_j^{ear} z_{ki} + x_{kj} - (t_i^{ear} z_{ki} + x_{ki}) \leq M \left(u_{kij} + \sum_{\substack{l \in K \\ l \neq k}} z_{li} \right) \qquad k \in K, i \in I, j \in I, j \neq i \qquad (5)$$

$$T_j^{ear} z_{ki} + y_{kj} - (T_i^{ear} z_{ki} + y_{ki}) \leq M \left(v_{kij} + \sum_{\substack{l \in K \\ l \neq k}} z_{li} \right) \qquad k \in K, i \in I, j \in I, j \neq i \qquad (6)$$

$$u_{kij} \leq z_{ki} \qquad k \in K, i \in I, j \in I, j \neq i \qquad (7)$$

$$u_{kij} \leq z_{kj} \qquad k \in K, i \in I, j \in I, j \neq i \qquad (8)$$

$$v_{kij} \leq z_{ki} \qquad k \in K, i \in I, j \in I, j \neq i \qquad (9)$$

$$v_{kij} \leq z_{kj} \qquad k \in K, i \in I, j \in I, j \neq i \qquad (10)$$

$$u_{kij} + u_{kji} \leq 1 \qquad k \in K, i \in I, j \in I, j \neq i \qquad (11)$$

$$v_{kij} + v_{kji} \leq 1 \qquad k \in K, i \in I, j \in I, j \neq i \qquad (12)$$

$$u_{kij} + v_{kji} \leq 1 \qquad k \in K, i \in I, j \in I, j \neq i \qquad (13)$$

$$t_i^{ear} z_{ki} + x_{ki} + \omega_i z_{ki} \leq T_i^{ear} z_{ki} + y_{ki} \qquad k \in K, i \in I \qquad (14)$$

$$\sum_{i \in I} \kappa_i z_{ki} \leq C_k \qquad k \in K \qquad (15)$$

$$x_{ki} \leq a_i \qquad k \in K, i \in I \qquad (16)$$

$$y_{ki} \leq b_i \qquad k \in K, i \in I \qquad (17)$$

$$u_{kij} \in \{0,1\} \qquad k \in K, i \in I, j \in I \qquad (18)$$

$$v_{kij} \in \{0,1\} \qquad k \in K, i \in I, j \in I \qquad (19)$$

$$w_k \in \{0,1\} \qquad k \in K \qquad (20)$$

$$x_{ki} \in R_0^+ \qquad k \in K, i \in I \qquad (21)$$

$$y_{ki} \in R_0^+ \qquad k \in K, i \in I \qquad (22)$$

$$z_{ki} \in \{0,1\} \qquad k \in K, i \in I \qquad (23)$$

Function (1) represents the optimisation criterion used in the model – the total number of the shunting tracks used for the turnarounds of the trams. Constraints (2) ensure that only single track is assigned to each tram. Constraints (3) and (4) model necessary links between the constraints and the optimisation criterion. If the tram of the line $i \in I$ is assigned to the track $k \in K$, then the track is included in the value of the optimisation criterion.

Constraints (5) ensure that if the tram serving the connection of the line $i \in I$ began to occupy the same track $k \in K$ before the tram serving the connection of the line $j \in I$, then it holds for the auxiliary variable $u_{kij} = 1$. Constraints (6) ensure that if the tram serving the connection of the line $i \in I$ left the same track $k \in K$ before the tram serving the connection of the line $j \in I$, then it holds for the auxiliary variable $v_{kij} = 1$.

Constraints (7) and (8) are related to the arrivals of the trams and assure that the variable u_{kij} is equal to 1 only if the track $k \in K$ is used for the turnarounds of the trams serving the connections of the lines $i \in I$ and $j \in I$ together. Constraints (9) and (10) are related to the departures of the trams and model that the variable v_{kij} is equal to 1 only if the track $k \in K$ is used for the turnarounds of the trams serving the connections

of the lines $i \in I$ and $j \in I$ together. Without constraints (7) up to (10) and adding the term $\sum_{\substack{l \in K \\ l \neq k}} z_{li}$ into constraints (5) and (6) the model may provide incorrect solution from the point of view of the resulting values of the variables u_{kij} and v_{kij}.

Constraints (11) ensure that for each pair of the trams serving the connections of the different lines with the different arrival times and sharing the same turning loop track it will not happen that the tram serving the connection of the line $i \in I$ arrived before the tram serving the connection of the line $j \in I$ and concurrently the tram serving the connection of the line $i \in I$ will arrive after the tram serving the connection of the line $j \in I$. Constraints (12) have an analogous meaning as constraints (11), but they correspond to the departures of the trams.

Constraints (13) ensure that it will not happen for the track $k \in K$ that the tram serving the connection of the line $i \in I$ arrived on the track $k \in K$ before the arrival of the tram serving the connection of the line $j \in I$ and concurrently the tram serving the connection of the line $i \in I$ will depart from the track $k \in K$ after the tram serving the connection of the line $j \in I$ and vice versa.

Constraints (14) model that for each tram it will not happen that the tram should depart from the track before the minimum turnaround time elapses. Constraints (15) ensure that the capacity of any track is not exceeded from the point of view of the number of the trams found on the track at the same time. Constraints (16) and (17) guarantee that the time shifts of the arrivals and the departures of the individual connections (taken from their earliest possible time positions) will not exceed their maximum allowable limits.

And finally, constraints (18) up to (23) define domains of definition of the variables used in the model.

6 Optimisation experiments with the proposed model

Optimisation experiments presented in the article were carried out by observing the following rules:

1. The numbers of the shunting tracks forming the turning loop $|K|$ were chosen to correspond to real conditions, the minimum number $|K|$ is equal to 2 and the maximum number $|K|$ equals to 5. We can say that the maximum value is enough also for tram transport operators which serve large tram transport networks.
2. The numbers of the tram lines using the turning loop correspond to real conditions as well. The minimum number of $|I|$ equals to $|K|$ because for $|I| < |K|$ it is not purposeful to solve the problem. Testing the model for the individual numbers of the tracks $|K|$ was initiated for cases when $|I| = |K|$. This is because the earliest possible arrival and departure times may be set so that each line must have a separate turning loop track (this may occur when it is not possible to move with the arrival and departure times, that means $a_i = b_i = 0$ for $i = 1, \dots, |I|$).
3. The maximum number of the lines $|I|$ is chosen to be $|K| + 2$, because in practice we do not meet with the turning loops that are used by more than $|K| + 2$ lines in the Czech Republic.
4. For different combinations of $|K|$x$|I|$ different input data sets were generated – the earliest possible arrival and departure times, the input data sets remain the same for all the values of the maximum time shifts 0, …, 5.
5. When generating input data, we observed the rule that the earliest possible departure time of any connection of any line cannot be lower than the sum of the earliest possible arrival time plus the minimum turnaround time (that means the minimum turnaround time must be satisfied).

Results of the optimisation experiments are summarised in Table 1.

K	I	t ear [min]				T ear [min]				ai, bi	ΣWk	Time [s]
		1	2	3	4	1	2	3	4			
2	2	13	17			21	25			0	1	0,1
		23	30			29	37			1	1	0
		21	23			29	30			2	1	0
		5	12			10	19			3	1	0
		10	22			20	27			4	1	0
		36	37			41	42			5	1	0
	3	15	20	19		22	28	25		0	2	0
		5	11	21		20	18	26		1	2	0
		21	30	10		34	43	22		2	2	0
		5	5	15		11	17	22		3	2	0
		10	27	27		37	37	32		4	2	0
		8	20	29		25	40	40		5	2	0
	4	x	x	x	x	x	x	x	x	0	x	0
		25	30	24	32	40	36	40	38	1	2	0
		32	20	28	9	42	25	38	42	2	2	0
		26	18	33	13	35	35	40	20	3	2	0
		19	6	26	19	26	16	42	34	4	2	0
		13	22	19	19	30	30	25	25	5	2	0

K=2

K	I	t ear [min]					T ear [min]					ai, bi	ΣWk	Time [s]
		1	2	3	4	5	1	2	3	4	5			
3	3	17	28	19			25	36	25			0	2	0,1
		16	6	10			25	16	19			1	2	0
		9	31	37			17	40	50			2	2	0
		23	7	19			31	22	24			3	2	0,1
		7	15	19			22	20	32			4	2	0,1
		14	26	3			24	42	8			5	2	0
	4	6	10	30	27		40	38	35	45		0	3	0
		27	35	38	21		35	47	45	40		1	2	0
		40	22	31	22		50	35	40	45		2	2	0,1
		45	40	7	3		55	59	15	35		3	2	0
		10	51	20	54		35	60	50	60		4	2	0
		0	18	1	3		15	25	8	20		5	2	0
	5	9	23	16	9	13	25	28	28	20	20	0	2	0
		x	x	x	x	x	x	x	x	x	x	1	x	x
		15	9	5	5	17	25	17	19	25	23	2	2	0,2
		12	6	12	15	18	20	25	18	22	25	3	2	0
		9	9	19	4	12	19	22	25	19	20	4	2	0
		15	9	3	12	17	24	24	25	24	25	5	2	0

K=3

K	I	t ear [min]						T ear [min]						ai, bi	ΣWk	Time [s]
		1	2	3	4	5	6	1	2	3	4	5	6			
4	4	12	25	6	25			20	30	11	35			0	2	0,1
		7	7	11	14			22	15	19	19			1	2	0,1
		16	22	17	25			27	27	22	30			2	2	0,1
		2	18	10	2			20	25	22	25			3	2	0
		0	17	12	9			8	25	25	15			4	2	0,1
		25	1	22	4			30	10	35	24			5	2	0
	5	23	15	10	20	19		29	33	25	28	29		0	3	0,1
		27	27	10	34	23		32	33	40	39	35		1	3	0,1
		31	11	13	31	22		39	25	40	36	28		2	2	0,1
		27	34	18	27	13		40	42	43	32	36		3	2	0
		26	15	26	14	20		40	40	35	42	42		4	3	0,1
		34	26	14	27	32		42	39	30	30	40		5	2	0,1
	6	15	20	30	37	40	17	37	35	39	43	45	26	0	3	0
		20	32	34	20	36	35	42	37	40	35	41	45	1	3	0,1
		27	36	20	19	32	36	45	41	32	42	45	41	2	3	0,1
		34	17	24	26	24	34	40	45	32	45	45	39	3	3	0,2
		15	28	37	24	21	23	44	35	45	44	26	29	4	3	0,1
		29	29	35	38	37	15	45	35	42	45	42	45	5	3	0,1

K=4

K	I	t ear [min]							T ear [min]							ai, bi	ΣWk	Time [min]
		1	2	3	4	5	6	7	1	2	3	4	5	6	7			
5	5	13	18	24	28	32			35	33	30	33	38			0	3	0,2
		30	15	10	22	22			35	40	33	30	28			1	2	0
		21	22	31	17	23			37	32	37	40	32			2	3	0,1
		31	17	23	33	26			36	40	36	38	31			3	3	0,1
		29	25	34	15	20			35	37	39	43	43			4	3	0,1
		23	13	29	29	14			40	34	38	40	40			5	2	0,1
	6	1	19	4	13	8	22		27	25	23	21	16	30		0	3	0,1
		22	6	6	3	5	9		27	28	28	30	26	26		1	3	0,1
		5	11	11	7	25	22		21	28	17	14	30	28		2	3	0,1
		22	14	8	17	12	21		29	32	35	29	32	29		3	3	0,1
		21	2	13	5	14	21		29	7	22	29	29	30		4	3	0,1
		3	4	23	22	14	8		14	26	28	30	25	25		5	3	0,1
	7	26	22	34	33	22	44	24	32	42	46	40	28	49	31	0	3	0,1
		39	32	29	33	32	26	22	46	47	50	47	47	52	45	1	5	0,2
		29	34	23	43	43	22	45	51	39	50	50	49	51	51	2	3	0,1
		40	23	34	37	38	40	40	45	44	40	44	50	45	45	3	3	0,2
		22	44	24	39	33	29	33	51	49	51	49	45	38	38	4	3	0,1
		22	39	33	24	44	24	39	52	49	53	53	52	43	45	5	3	0,2

K=5

K	I	t ear [min]								T ear [min]								ai, bi	ΣWk	Time [min]
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8			
6	6	30	13	6	7	25	10			35	37	40	38	35	24			0	4	0,2
		12	19	4	24	22	12			25	24	30	30	27	28			1	3	0,7
		18	20	3	15	2	10			24	29	20	26	26	16			2	3	0,3
		10	1	6	15	3	11			15	19	13	25	21	19			3	3	0,1
		15	3	15	18	3	15			21	16	29	23	22	29			4	3	0,2
		14	0	3	8	19	10			19	18	28	16	24	24			5	3	0,2
	7	25	18	22	24	30	16	25		32	36	27	32	35	29	36		0	3	0,2
		30	16	25	14	21	24	21		35	30	30	40	36	32	34		1	4	0,2
		28	17	12	29	11	15	21		33	37	37	34	17	30	31		2	3	0,2
		29	16	11	25	28	18	30		34	39	43	34	37	39	35		3	4	0,1
		26	14	27	22	14	11	12		31	39	33	39	45	29	36		4	3	0,1
		19	10	15	26	14	18	30		30	30	20	40	32	30	35		5	3	0,1
	8	20	16	22	21	15	10	19	18	39	30	33	36	35	42	25	24	0	5	0,1
		10	19	18	5	23	23	16	20	25	24	25	33	28	31	34	25	1	4	0,1
		17	21	20	16	12	8	25	24	31	27	26	33	31	32	33	39	2	4	0,1
		19	10	7	17	5	7	13	17	27	17	25	23	38	31	18	27	3	4	0,1
		21	22	24	6	24	21	19	21	28	29	29	23	31	28	35	28	4	4	0,2
		22	19	12	19	13	21	10	6	27	26	38	27	28	32	32	37	5	4	0,1

K=6

K	I	t ear [min]									T ear [min]									ai, bi	ΣWk	Time [min]
		1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9			
7	7	8	9	22	7	12	10	23			25	26	27	29	28	26	28			0	4	0,1
		8	17	10	19	25	6	21			26	22	24	25	30	33	26			1	4	0,2
		22	7	7	24	24	21	15			27	28	31	29	29	27	20			2	3	0,2
		13	17	6	12	22	6	10			25	22	21	25	27	23	19			3	3	0,2
		14	15	14	9	8	9	16			19	23	27	31	27	34	25			4	3	0,2
		13	17	10	13	7	23	21			23	27	27	21	29	28	29			5	3	0,2
	8	25	26	21	25	17	24	11	15		31	31	26	30	24	37	32	33		0	4	0,1
		17	11	28	26	15	28	10	16		24	29	37	32	29	35	35	35		1	4	0,2
		30	21	20	16	15	17	16	28		35	30	25	33	35	36	35	33		2	4	0,2
		29	11	22	20	11	12	22	24		35	35	33	25	37	36	31	36		3	4	0,5
		27	24	17	16	21	30	23	22		38	31	30	38	30	35	38	30		4	4	0,3
		23	25	20	23	12	21	23	15		28	30	28	28	35	26	33	35		5	4	0,3
	9																			0		
																				1		
																				2		
																				3		
																				4		
																				5		

K=7

Table 1 Summary of the results

7 Conclusions

The presented article proposes the mathematical model for minimising the number of shunting tracks at the tram turning loops. Reducing the number of the tracks can be achieved by using time coordination of the arrival and departure times of the individual connections which use the turning loop.

The presented article demonstrates that linear programming can be used to solve the given problem. To verify the functionality, computational experiments were performed with the proposed mathematical model in the Xpress-IVE optimisation software. Model data were used in the computational experiments and the performed computations proved the full functionality of the proposed model.

To make the model applicable for practice, the problem needs to be solved in tasks related to larger tram networks. The tasks with larger networks involve multiple tram turning loops which are interconnected by the operated tram lines. This is the main problem we would like to address in the future.

Acknowledgements

This work was supported by project CK01000043 System for supporting network time coordination of connections at interchange nodes.

References

[1] ČERNÝ, J. (1991). *Základy matematickej teórie dopravy*. Bratislava: Veda.

[2] ČERNÁ, A. & ČERNÝ, J. (2004). *Teorie řízení a rozhodování v dopravních systémech*. Pardubice: Institut Jana Pernera.

[3] ČERNÁ, A. & ČERNÝ, J. (2014). *Manažerské rozhodování o dopravních systémech*. Pardubice: Univerzita Pardubice.

[4] KHANI, A. & SHAFAHI, Y. (2011). Transfer Optimization in Transit Networks: Headway and Departure Time Coordination. In *Proceedings of 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)* (pp. 1531–1536). Campus George Washington Univ (GWU), Washington, DC.

[5] KOZEL, P. (2011). The time coordination of bus links in both transport directions. In *Proceedings of 29th International Conference on Mathematical Methods in Economics 2011, parts I & II* (pp. 357–362). Janska Dolina, Slovakia.

[6] KOZEL, P. & MICHALCOVA, S. (2017). Time coordination of bus arrivals on a set of bus stops. In *Proceedings of 12th International Conference on Strategic Management and its Support by Information Systems (SMSIS)* (pp. 267–273), Ostrava, Czech Republic.

[7] TEICHMANN, D., DORDA, M., MOCKOVÁ, D., OLIVKOVÁ, I., GRAF, V., & DVOŘÁČKOVÁ, A. (2020). Of Approach to Minimize the Number of Shunting Tracks at Turning Loops. Article accepted for publication.

Robustness of Dynamic Score-Driven Models Utilizing the Generalized Gamma Distribution

Petra Tomanová[1]

Abstract. Time series processes might exhibit a complex dynamic structure that cannot be captured by static linear models. However, once model parameters are allowed to be time-varying with possibly complex non-linear updating mechanisms, a huge number of model specification choices arise. A modern framework of generalized autoregressive score (GAS), which lies on the idea that at each time the local fit of the model is improved by the scaled score of the log conditional observation density, helps to overcome this issue. Despite reduced ambiguity of model specification, some arbitrary aspects remain. We investigate the impact of the choice of parametrization, link and scaling functions on the model performance and thus we assess the robustness of the GAS methodology. In this paper, we restrict ourselves to the non-negative time series processes for which we utilize a flexible generalized gamma distribution. Using two data sets of retail and financial durations, we demonstrate that these somewhat arbitrary aspects do not dramatically impact the performance of the GAS models unless numerical issues are encountered.

Keywords: generalized autoregressive score model, generalized gamma distribution, link function, parametrization, scaling function

JEL Classification: C22, C46
AMS Classification: 91G70

1 Introduction

In order to capture the dynamic behavior of time series processes, it might be inevitable to allow model parameters to vary over time. Generally, one of two possible directions might be taken: (i) model parameters are viewed as stochastic processes with their own source of error and thus, they are not perfectly predictable based on past observations; (ii) model parameters are functions of lagged dependent and exogenous variables. We focus on the latter which brings benefits such as perfect predictability based on past information, straightforward evaluation of likelihood, and possible natural extensions to other more complicated dynamics. To avoid the ambiguous choice of updating mechanism for time-varying parameters, we utilize a modern framework of the *generalized autoregressive score (GAS)* proposed by Creal et al. [5].

The GAS models assume that the time-varying parameter of any underlying probability distribution follows a recursion consisting of the autoregressive term and the scaled score of the logarithmic observation density. In spite of the fact that the updating mechanism follows a defined rule, four decisions have to be made – to choose (i) a proper underlying distribution, (ii) its parametrization, (iii) link function, and (iv) scaling function of the score. The underlying distribution should be sufficiently flexible to correctly model the time series process. Its selection is rather natural, however, the "right" choice of its parametrization might be ambiguous. It is commonly driven by criteria such as favourability of the parameter interpretation and ability to ensure that the parameter values satisfy possible constraints such as positiveness. Parametrization alongside with link and scaling function is rather arbitrary with an unclear impact on model performance. In this paper, we investigate this issue for non-negative time series processes. We utilize a flexible generalized gamma distribution which contains several common distributions such as gamma, Weibull, and exponential as special cases.

Our robustness analysis lies in a simulation study for which the time series processes reflect real data behavior. The first dataset is taken from Tomanová and Holý [9] where authors analyze arrivals in queueing systems. Second, we use transactions of a major company traded in the Nordic stock market publicly available in R package `ACDm` [1]. Both datasets are related to the duration analysis, however, the first one concerns the retail data and the second one the financial market data. For a survey of financial duration analysis, see Pacurar [6] and Bhogal and Thekke Variyam [2].

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 1938/4, 130 67 Prague 3, Czechia, petra.tomanova@vse.cz

The aim of this paper is to analyze whether the arbitrary choices in GAS methodology substantially affect the model performance. In Section 2, the generalized autoregressive score framework is recalled. Then, two specifications of generalized gamma distribution, their scores and link functions are discussed in Section 3. Two simulation studies based on two datasets are performed in Section 4 and 5. Section 6 concludes.

2 Generalized Autoregressive Score Framework

Let $\{y_t\}_{t\in\mathbb{Z}}$ be a real-valued stochastic sequence of observations with conditional probability density

$$p_y\left(y_t|h(f_t(\theta)); g\right),$$

for all $t \in \mathbb{Z}$, where $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ represents a scalar time-varying parameter that depends on a vector of time-invariant parameters $\theta \in \Theta$, $h : \mathbb{R} \to \mathbb{R}$ is a link function, and $g \in \mathcal{G}$ is a vector of time-invariant parameters that indexes the conditional density p_y. Under the *generalized autoregressive score (GAS)* framework of [5], the dynamic process for the time-varying parameter $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ is given by the autoregressive updating equation

$$\begin{aligned} f_{t+1}(\theta) &= c + bf_t(\theta) + as_t(f_t(\theta); g), \\ s_t(f_t(\theta); g) &= S(f_t(\theta); g)\nabla_t(f_t(\theta); g), \\ \nabla_t(f_t(\theta); g) &= \frac{\partial \log p_y(y_t|f; g)}{\partial f}\Big|_{f=f_t(\theta)}, \end{aligned}$$

where c, b, and a are time-invariant parameters that together with g form the parameter vector θ belonging to the parameter space Θ, and $S(f_t(\theta); g)$ is a univariate scaling factor for the score $\nabla_t(f_t(\theta); g)$ of the conditional observation density p_y. Scaling function is typically set to the k-th power of the inverse of the Fisher information matrix

$$\begin{aligned} S(f_t(\theta); g) &= (\mathcal{I}_t(f_t(\theta); g))^{-k}, \\ \mathcal{I}_t(f_t(\theta); g) &= \mathrm{E}_{t-1}[\nabla_t(f_t(\theta); g)\nabla_t(f_t(\theta); g)'], \end{aligned}$$

where typically $k \in \{0, 0.5, 1\}$, see [4] and [5] for further details and the explanation.

For the purpose of this paper, let the probability density function be defined by the observation equation

$$y_t = h(f_t(\theta))u_t \quad \forall t \in \mathbb{Z},$$

where for all $g \in \mathcal{G}$, $\{u_t\}_{t\in\mathbb{Z}}$ is an independently identically distributed sequence with u_t independent of f_t for every t and $u_t \sim p_{u,g}(u_t)$. Then a probability measure for $\{y_t\}$ is defined when a point (h, p_u, S, θ) is chosen,

$$(h, p_u, S, \theta) \in \mathcal{H} \times \mathcal{P}_u \times \mathcal{S} \times \Theta,$$

where $\mathcal{H}$ denotes the space of link functions, $\mathcal{P}_u$ the space of families of densities p_u for u_t, and $\mathcal{S}$ the space of scaling functions [4]. In this paper, we restrict ourselves to non-negative $\{y_t\}_{t\in\mathbb{Z}}$. As a density p_u, we utilize a flexible generalized gamma distribution and investigate the impact of a link function h and a scaling function S to the model performance while estimating the parameter vector θ by the *maximum likelihood method*.

3 Generalized Gamma Distribution

The *generalized gamma distribution* is a continuous probability distribution for non-negative variables proposed by Stacy [8]. It is a three-parameter generalization of the two-parameter gamma distribution and contains several common distributions such as the exponential and the Weibull distribution as special cases. The distribution has the scale parameter $\beta > 0$ and the shape parameters $\psi > 0$ and $\varphi > 0$. The conditional probability density function is

$$p_y(y|\beta, \psi, \varphi) = \frac{1}{\Gamma(\psi)}\frac{\varphi}{\beta}\left(\frac{y}{\beta}\right)^{\psi\varphi-1} e^{-\left(\frac{y}{\beta}\right)^{\varphi}}$$

for $y \in \mathbb{R}^+$, where $\Gamma(\cdot)$ is the gamma function. Special cases include the gamma distribution for $\varphi = 1$, the Weibull distribution for $\psi = 1$ and the exponential distribution for $\psi = 1$ and $\varphi = 1$. The expected value

and variance is

$$\mathrm{E}[Y] = \beta \frac{\Gamma\left(\psi + \varphi^{-1}\right)}{\Gamma\left(\psi\right)},$$
$$\mathrm{var}[Y] = \beta^2 \frac{\Gamma\left(\psi + 2\varphi^{-1}\right)}{\Gamma\left(\psi\right)} - \left(\beta \frac{\Gamma\left(\psi + \varphi^{-1}\right)}{\Gamma\left(\psi\right)}\right)^2.$$

For $y \in \mathbb{R}^+$, the score and the Fisher information for the parameter β is

$$\nabla_\beta(y, \beta, \psi, \varphi) = \frac{\partial \log p_y(y|\beta, \psi, \varphi)}{\partial \beta} = \frac{\varphi}{\beta}\left(y^\varphi \beta^{-\varphi} - \psi\right),$$
$$\mathcal{I}_\beta(\beta, \psi, \varphi) = \mathrm{E}\left[\nabla_\beta(y, \beta, \psi, \varphi)^2 \middle| \beta, \psi, \varphi\right] = \frac{\psi\varphi^2}{\beta^2}.$$

As a scaling function, we utilize the inverse of the Fisher information ($k = 1$), the square root of the inverse of the Fisher information ($k = 0.5$), and the unit scaling ($k = 0$). Each choice for the scaling function gives rise to a new GAS model.

We analyze the impact of the most common link function $h(\cdot)$ – logarithmic transformation – which ensures positiveness of the time-varying parameter. The link function $h(\beta)$ affects the score and the Fisher information

$$\nabla_{h(\beta)}(y, h(\beta), \psi, \varphi) = \dot{h}^{-1}(\beta)\nabla_\beta(y, \beta, \psi, \varphi),$$
$$\mathcal{I}_{h(\beta)}(h(\beta), \psi, \varphi) = \dot{h}'^{-1}(\beta)\mathcal{I}_\beta(\beta, \psi, \varphi)\dot{h}^{-1}(\beta),$$

where $\dot{h} = \partial h(\beta)/\partial \beta'$. When $h(\beta) = \log \beta$ then $\dot{h} = 1/\beta$. Thus, the score and the Fisher information for the parameter $\beta^* = \log \beta$ is

$$\nabla_{\beta^*}(y, \beta^*, \psi, \varphi) = \frac{\partial \log p_y(y|\beta^*, \psi, \varphi)}{\partial \beta^*} = \varphi\left(y^\varphi e^{-\varphi\beta^*} - \psi\right),$$
$$\mathcal{I}_{\beta^*}(\beta^*, \psi, \varphi) = \mathrm{E}\left[\nabla_{\beta^*}(y, \beta^*, \psi, \varphi)^2 \middle| \beta^*, \psi, \varphi\right] = \psi\varphi^2,$$

for $y \in \mathbb{R}^+$. Note that the Fisher information for β^* is not dependent on β^* itself. Thus, the scaling function choice for this model has no impact on model performance.

An alternative parametrization of the generalized gamma distribution can be utilized. Prentice [7] reparameterized and extended the distribution of the logarithm of a generalized gamma variate. In this paper, we consider the parametrization that is related to the common one as

$$\beta = \exp(\mu)\lambda^{\frac{2\sigma}{\lambda}}, \qquad \psi = \frac{1}{\lambda^2}, \qquad \varphi = \frac{\lambda}{\sigma},$$
$$\mu = \log \beta + \frac{1}{\varphi}\log \psi, \qquad \sigma = \frac{1}{\varphi\sqrt{\psi}} \qquad \lambda = \frac{1}{\sqrt{\psi}},$$

where $\mu \in \mathbb{R}$, $\sigma > 0$ and $\lambda > 0$. The score for the parameter μ is

$$\nabla_\mu(y, \mu, \sigma, \lambda) = \frac{\partial \log p_y(y|\mu, \sigma, \lambda)}{\partial \mu} = \frac{y^{\frac{\lambda}{\sigma}} \exp\{-\mu\}^{\frac{\lambda}{\sigma}} - 1}{\lambda\sigma}.$$

To summarize, we consider three common scaling functions (the inverse of the Fisher information, the square root of the inverse of the Fisher information, and the unit scaling), two different link functions (level and logarithmic transformation), and two different parametrizations of the distribution function (common and the alternative one). The analyses are performed on two different data set – retail data and financial market data.

4 Empirical Example of Retail Data

The first data set is taken from Tomanová and Holý [9]. The data are originally obtained from the database of an online bookshop with one brick-and-mortar location in Prague, Czechia. It contains information about order arrivals (number of ordered books) such as time and volume. The aim is to model times between

arrivals (durations) for the purpose of process simulation and optimization. The durations are diurnally and seasonally adjusted, see [9] for further details.

The data set covers the period of June 8 – December 20, 2018, resulting in 28 full weeks and 5,753 observations. For adjusted data the unconditional mean is 1.00, the unconditional variance is 1.07 and the distribution is illustrated in Figure 1. The generalized gamma distribution is clearly favored over the exponential distribution for the raw data, however, the difference is not so striking for the adjusted data. However, the conditional generalized gamma distribution provides a considerably better fit even for the adjusted data, see [9] for further details. The estimated GAS model with various specifications is presented in Table 1. The maximum likelihood is obtained for the unit scaling with common parametrization and no transformation (level). However, the other specifications have almost the same log-likelihood.

Furthermore, we conduct a simulation study in order to investigate the effects of model misspecification. We simulate 1,000 observations using the models given by Table 1 and estimate models with various parametrizations, link and scaling functions. We repeat the simulation 20,000 times and report the mean differences in log-likelihoods between the estimated models and the true model in Table 2. All differences between the model performances are negligible. Naturally, the largest differences in log-likelihoods are in the case of the true inverse scaling when assuming unit scaling and vice versa. The square-root scaling offers a middle way. The logarithmic model and the model with the alternative parametrization (which also contains the logarithm of β) behave almost identically.

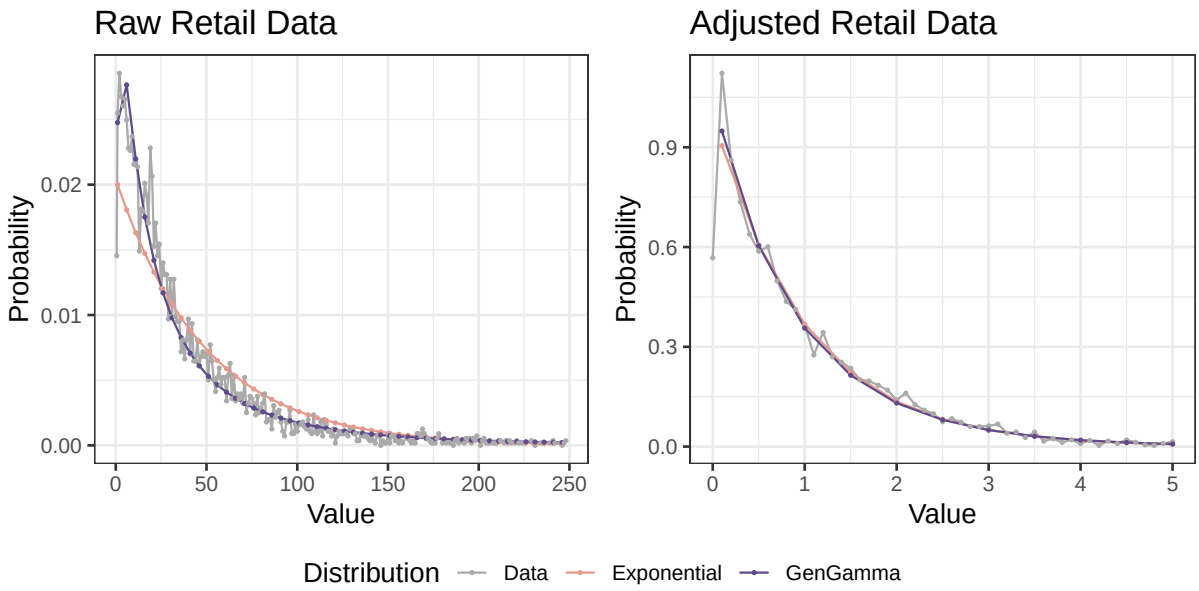


Figure 1 The unconditional probability density function for the retail data.

Model	c	b	a	ψ	φ	Log-Lik	AIC
Level/Unit	0.28	0.65	0.05	1.17	0.89	−5721.49	11452.97
Level/Sqrt	0.26	0.68	0.07	1.16	0.89	−5721.73	11453.46
Level/Inverse	0.23	0.72	0.07	1.15	0.90	−5722.17	11454.33
Log/Unit	−0.06	0.72	0.07	1.15	0.90	−5723.31	11456.62
Alt/Unit	−0.01	0.72	0.07	0.93	1.04	−5723.31	11456.62

Table 1 The estimated coefficients with the log-likelihoods and the AIC for the retail data.

	Estimated Model				
True Model	level/Unit	level/Sqrt	level/Inv	log/Unit	Alt/Unit
level/Unit	0.00	−0.04	−0.19	−0.13	−0.13
level/Sqrt	−0.04	0.00	−0.06	−0.04	−0.04
level/Inv	−0.15	−0.02	0.00	−0.02	−0.02
log/Unit	−0.09	−0.01	−0.02	0.00	0.00
Alt/Unitt	−0.10	−0.02	−0.03	−0.00	0.00

Table 2 The differences in the log-likelihoods for various models based on the retail data.

5 Empirical Example of Financial Market Data

The second data set includes transactions of a major company traded in the Nordic stock market which is publicly available in R package `ACDm` [1]. The data set covers two weeks of intraday transactions recorded at 1-second precision and includes the transaction date, timestamp, price and volume. The total number of transactions is 99,330 from which the price durations are calculated. Price duration is measured as a time until the price process changes by a given level. Our data set can be replicated by using function `ACDm::computeDurations(transData, type = 'price', priceDiff = .01)`. Then the price durations are adjusted by cubic splines with a vector of nodes: `c(seq(600, 1105, 120), 1105)`.

The number of price durations is 2,054, the unconditional mean is 0.99, the unconditional variance is 2.16 and the distribution is illustrated in Figure 2. Again the generalized gamma distribution provides a better fit than the exponential distribution. Table 3 contains the estimated coefficients with log-likelihoods and AIC values for various model specifications. The best models are based on (i) logarithmic transformation, and (ii) alternative parametrization, both with unit scaling. The models based on common parametrization without logarithmic link function perform distinctively worse. The model with unit scaling produces the worst results due to problems with numerical optimization. The source of the problem might be the constrain $\beta > 0$ which is not naturally satisfied as in the case of logarithmic transformation of β.

We conduct a simulation study with 1,000 replications and report the differences in log-likelihoods in Table 4. The results show that the model with logarithmic transformation and the model with alternative parametrization exhibit very similar results in line with the previous study. The model utilizing the common parametrization with inverse scaling performs comparably with those two. The performance of models with unit and square root scaling is very poor due to numerical issues. Thus, when the time-varying parameter has to be positive, the logarithmic transformation should be preferred even if the true model is different since the numerical problems might produce much more severe drops in performance than the model misspecification.

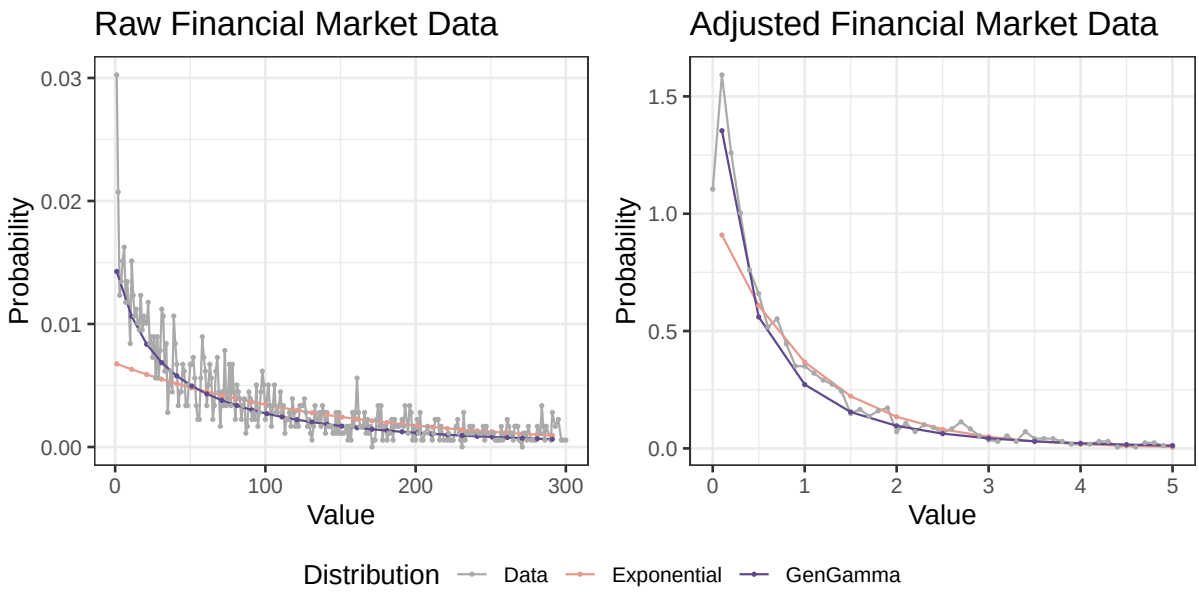


Figure 2 The unconditional probability density function for the financial market data.

Model	c	b	a	ψ	φ	Log-Lik	AIC
Level/Unit	0.01	0.78	0.00	3.14	0.43	−1801.49	3612.98
Level/Sqrt	0.01	0.86	0.01	2.93	0.47	−1724.94	3459.88
Level/Inverse	0.00	0.95	0.21	3.22	0.44	−1722.06	3454.11
Log/Unit	−0.33	0.91	0.32	3.75	0.41	−1718.38	3446.75
Alt/Unit	−0.05	0.91	0.32	0.52	1.26	−1718.38	3446.75

Table 3 The estimated coefficients with the log-likelihoods and the AIC for the financial market data.

	Estimated Model				
True Model	level/Unit	level/Sqrt	level/Inv	log/Unit	Alt/Unit
level/Unit	0.00	1.49	1.36	1.42	1.42
level/Sqrt	−23.48	0.00	−3.19	−2.24	−2.24
level/Inv	−115.06	−30.27	0.00	−1.86	−1.45
log/Unit	−93.31	−16.72	−2.32	0.00	0.00
Alt/Unitt	−92.75	−17.88	−1.88	−0.00	0.00

Table 4 The differences in the log-likelihoods for various models based on the financial market data.

6 Discussion and Conclusion

We investigate the impact of parametrization, link function, and scaling function choices on the GAS model performance, and thus, we address the robustness of the GAS methodology. Our results show that the differences in performance are negligible unless the numerical issues are encountered. This might be easily avoided by utilizing the logarithmic link function or suitable parametrization without causing substantial drops in model performance since our results show that the impact of parametrization or link function is negligible. Thus, the practitioners can select these aspects of GAS models purely based on convenience without a worries that the choice might lead to an unwanted drop in model performance.

Our findings are in line with the paper [3] where authors analyzed financial durations based on GAS models and found that there are no significant differences between the three considered scaling functions. However,

they only considered the logarithmic link function and one parametrization. Moreover, the authors used a discrete approach – specifically they utilized the zero-inflated negative binomial distribution. A more extensive study is still missing in the literature, thus in the future work we will analyze other model specifications as well and explore this topic more thoroughly.

Acknowledgements

The work on this paper was supported by the grant No. F4/53/2019 of the Internal Grant Agency of University of Economics, Prague.

References

[1] Belfrage, M. (2016). Package 'ACDm': Tools for Autoregressive Conditional Duration Models. `https://CRAN.R-project.org/package=ACDm`.

[2] Bhogal, S. K., & Thekke Variyam, R. (2019). Conditional Duration Models for High-Frequency Data: A Review on Recent Developments. *Journal of Economic Surveys*, 33(1), 252–273.

[3] Blasques, F., Holý, V., & Tomanová, P. (2020). Zero-Inflated Autoregressive Conditional Duration Model for Discrete Trade Durations with Excessive Zeros. `https://arxiv.org/pdf/1812.07318.pdf`.

[4] Blasques, F., Koopman, S. J., & Lucas, A. (2014). Stationarity and ergodicity of univariate generalized autoregressive score processes. *Electronic Journal of Statistics*, 8(1), 1088–1112.

[5] Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics*, 28(5), 777–795.

[6] Pacurar, M. (2008). Conditional Duration Models in Finance: A Survey of the Theoretical and Empirical Literature. *Journal of Economic Surveys*, 22(4), 711–751.

[7] Prentice, R. L. (1974). A log gamma model and its maximum likelihood estimation. *Biometrika*, 61(3), 539-544.

[8] Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of mathematical statistics*, 33(3), 1187–1192.

[9] Tomanová, P., & Holý, V. (2020). Clustering of Arrivals in Queueing Systems: Autoregressive Conditional Duration Approach. `https://arxiv.org/pdf/2003.14382.pdf`.

Modelling Sustainable Investment

Oldřich Trenz[1], Zuzana Chvátalová[2], Jitka Sládková[3],
Oldřich Faldík[4], František Ostřížek[5]

Abstract. SRI is a long-term investment approach integrating ESG factors in the research, analysis and selection process of securities in investment portfolio. It evaluate of ESG factors in order to better capture long-term return for investors, and benefit the society by influencing the behavior of companies. This article aims to describe the sustainable assessment model, its restrictive aspects and different possibilities of its deployment. This model combines different sets of indicators starting with economic ones and trying to engage all available social, environment and governance indicators. This engagement is performed by applying DEA model then combining its result with SVA.The model reflects the specifics requirements of the country and industry in which the company operates. SVA can be implemented by calculating weights and benchmark values for each sector of companies. The suggested model will be a backbone of web portal (WEBRIS 2) which is designed for quick and efficient data aggregation and assessment. Finally, the results visualization will be presented in the case study for selected sector.

Keywords: SRI, DEA, efficiency, sustainability assessment, SVA.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Sustainability performance can be said to be an ability of an organization to remain productive over time and hold on to its potential for maintaining long-term profitability. It shouldn't be acted only on the basis of economic results, whereas it should take into consideration non-financial factors [3], [17], [24], [25]. Because of that, all the current trends and frameworks depend on more comprehensive sustainability pillars; environmental, social, economic and governance performance. Therefore, by following these frameworks and integrating associated activities, companies will be able to achieve long-term benefits [4], [7]. This will be done by engaging different companies in disclosure of their overall economic, environmental, social, and governance (addition of economic indicator to ESG) impacts and helping them in understanding, measuring and communicating their sustainability performance (SP) [19].

One of the most famous and common methods for measuring corporate sustainability which incorporates three ESG dimensions is called the triple bottom line approach [10], [24]. It calculates the value by using not only financial but also non-financial resources. This value is called sustainable value added (SVA). This approach simplifies the measurements and enables SP to be measured in monetary terms depending on the data availability on the enterprise level as well as on the benchmark [20]. It shows how much value or damage is created as a result of using economic, environmental and social resources, compared to a benchmark [11]. In order to use this method, the benchmark company should be chosen. In this research, Data Envelopment Analysis (DEA) is used for this purpose.

Data Envelopment Analysis (DEA) which is "data oriented" approach for evaluating the efficiency of a number of producers. It combines the measurements of multiple inputs into any satisfactory overall measure of

[1] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, oldrich.trenz@mendelu.cz.
[2] Zuzana Chvátalová, Brno University of Technology/ Faculty of Business and Management/ Institute of Informatics, Kolejní 2906/4, 612 00 Brno, The Czech Republic, chvatalova@fbm.vutbr.cz.
[3] Jitka Sládková, Sting Academy/ Department of Economics and Management, M. Horákové 2031/1, 602 00, Brno, Czech Republic, sladkova.jit@gmail.com.
[4] Oldřich Faldík, Mendel University in Brno/ Faculty of Business and Economics/ Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, oldrich.faldik@mendelu.cz.
[5] Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, frantisek.ostrizek@mendelu.cz.

efficiency. In DEA the companies are usually referred to as the Decision Making Units (DMUs) which convert multiple inputs into multiple outputs.

Sustainable and Responsible Investment (SRI) is a long-term investment approach integrating Environmental, Social and Governance (ESG) factors in the research, analysis and selection process of securities within a given investment portfolio. It combines fundamental analysis and engagement with evaluation of ESG factors in order to better capture long term return for investors, and benefit the society by influencing the behavior of companies [9].

SRI, however, must incorporate multiple elements: an investment strategy for major environmental problems necessarily presupposes the exclusion of securities with poor management both for environmental and social risks. There are seven categories of SRI strategies [9]: 1) Sustainability themed investments; 2) Best-in-Class investment selection; 3) Exclusion of holdings from the investment universe; 4) Norms-based screening; 5) Integration of ESG factors in financial analysis; 6) Engagement and voting on sustainability matters; 7) Impact investing.

EUROSIF [9] analyzed 13 European Union (EU) markets: Austria, Belgium, Denmark, Finland, France, Germany, Italy, Netherlands, Poland, Spain, Sweden, Switzerland and the United Kingdom with the conclusion that the concept of SRI and the means by which investors can integrate ESG considerations in the investment process have greatly evolved over the last two decades, making it increasingly challenging to decide on the best method of achieving this. The important role will be played by information and communication technologies (ICT), including open linked data and public access to company reporting of ESG data.

Our previous research [12], [15], [19] resulted in assembling a model which integrating environmental, social, economic and corporate governance indicators. It combines several indicators from different areas and enables enterprises to effectively compare their performance. We verified a pilot version of this model in Czech brewery segment [18], [14] and Czech agricultural biogas plants segment [13].

Currently, it's the applicability of the model is being verified under practical conditions. The synthesis was then performed and a web application designed as required. The implementation and testing of the application followed.

The paper aims to introduce the sustainable assessment, model the application of the newly developed prototype of the portal WEBRIS 2 (Web Information System for Corporate Performance Evaluation and Environmental Reporting) for sustainability evaluation, modelling and benchmarking in small and medium-sized companies. The design and implementation of the portal WEBRIS 2 is discussed in this paper. This portal is based on cloud technologies and promotes sustainable development of enterprises and benchmarking.

2 Material and Methods

2.1 Data envelopment analysis (DEA)

As the field of Data Envelopment Analysis has grown, varieties of models and analyses are implemented. Since DEA was introduced, researchers in a number of fields have quickly recognized that it is an excellent and easily used methodology for modelling operational processes and performance evaluation. Various forms of DEA models have been used for entities, such as hospitals, universities, and companies.

One of the famous models is called CCR model [5] used for relative efficiency calculation. It produces multiple optimal solutions. This non-uniqueness of solution hampers the use of cross-efficiency evaluation. Therefore, [8], [23] suggested introducing a secondary goal to optimize the input and output weights while keeping unchanged the CCR model-efficiency of the target DMU. The formulated model is known as the aggressive formulation for cross-efficiency evaluation. Other different models as: minimizing the total deviation from ideal point, minimizing the maximum deviation, and minimizing the mean absolute deviation [21] were also implemented. These models are all established on the basis of an unrealistic ideal point, which defines the best relative efficiency of one as the target efficiency of each DMU. This target efficiency, however, is only realizable to DEA efficient DMUs, but not realizable to non-DEA efficient units. In order to solve this problem, set of models suggested to define the CCR model-efficiencies of the n DMUs as their target efficiencies, which are all realizable to DEA efficient units and non-DEA efficient units. These models are mentioned as minimizing or maximizing the total deviation from the ideal point, minimizing or maximizing the squared sum of deviations from the ideal point, and minimizing the mean absolute deviation from the ideal point [26].

The newest models to measure the sustainability management and SP based on advanced DEA model are approaches combine dual-role factors and a cross-efficiency technique [6]. It assesses the conflicts and trade-off among environmental, economic and social interests by using three continuous multi-criteria approaches and a set of different weights.

2.2 Sustainable Value Added (SVA)

The Sustainability Value Added (SVA) method is an effective method for sustainability assessment (SA). It plays a strategic role in decision making. It encourages the companies to deal with resources more effectively and efficiently. SVA value represents the extra value created as a result of using economic, environmental and social resources, compared to a benchmark. It is expressed in absolute monetary terms. According to the method published by [7]. The SVA value calculation can be expressed as follows: The gross value added (GVA) of the company should be calculated (in unit €). After that, the amount of each environment or social resources should be determined (e.g. t, m^3, etc.). Then efficiency computed by dividing the GVA on the amount of resources (unit €/t, €/m^3). The same steps should be done for the benchmark. Finally, the last two values are subtracted from each other and the result multiplied by the amount of considered indicator. This process is depicted in Fig. 1.

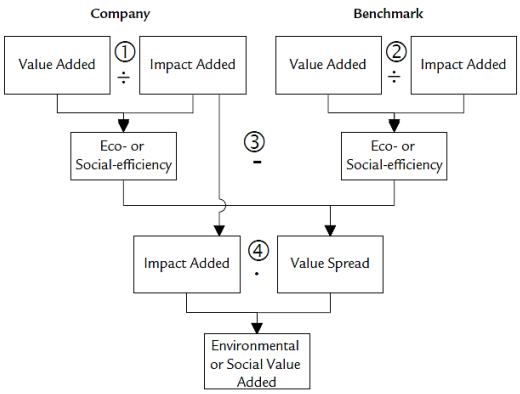


Figure 1 Evaluation steps of the Environmental or Social Value Added [7]

The calculation of SVA value for the company in two different times t_1 and t_0 is presented in Eq. (1) where n is the number of relevant environmental indicators. $EIA_{i,t0}$, $EIA_{i,t1}$ describe the eco-effectiveness of environmental impacts in t_0 and t_1. $EE_{i,b}$ is eco-efficiency of the benchmark for i environmental resource. This value calculated using Eq. (2) where b refers to benchmark. EG = (GVA_{t1} – GVA_{t0}) representing economic growth between t_0 and t_1.

$$\mathrm{SVA} = \mathrm{EG} - \frac{1}{N}\sum_{i=1}^{N} \mathrm{EE}_{i,b} \cdot (\mathrm{EIA}_{i,t1} - \mathrm{EIA}_{i,t0}) \tag{1}$$

$$\mathrm{EE}_b = \frac{\mathrm{VA}_b}{\mathrm{EIA}_b} \tag{2}$$

2.3 Web Information System for Corporate Performance Evaluation and Sustainable and Responsible Investment

The purpose of the Web Information System for Corporate Performance Evaluation and Sustainable and Responsible Investment (WEBRIS) is to enable the authentication and authorization. There are two types of WEBRIS 2 users: (1) Company. WEBRIS allows login or registration of a new user – a company. The registered user is offered to complete a new identification form or is presented with the original identification downloaded from open data in the Business register of managers in the Czech Republic [2]. (2) Web administrator. It sets up user rights and broadens with an access to the forms of any company.

The portal WEBRIS includes modules with a clearly defined contract, which is a list of provided and required services. In addition, this modular architecture is combined with Model-View-Controller (MVC) architecture [22]. A questionnaire template designed for the detection of specific information necessary for indicators calculation is included in each module. Furthermore, this module can communicate with another module with the aim of obtaining information from linked open data of public sources in the Czech Republic [2].

3 Results

During WEBRIS system upgrading (for sustainable investment), the process of refactoring was strictly separated from adding new functionality. In the first step, refactoring of the system was performed. Within WEBRIS 2 system refactoring, the architecture was completely modified. One of the main objectives of refactoring was to improve the design of the architecture system to facilitate the extension of the system by installing additional modules. The designed architecture allows easy replacement, removal and the insertion of modules. Façade Layer unites more complex logic, which means that one method of Facade Layer calls multiple methods of Service Layer.

The resulting module was implemented in PHP language and Nette framework. This module interface, through which it provides services, was clearly defined as well as the other modules. Original implementation of WEBRIS was extended by the chart JavaScript framework Fig. 2 a) and Fig. 2 b) show these charts in benchmarking.

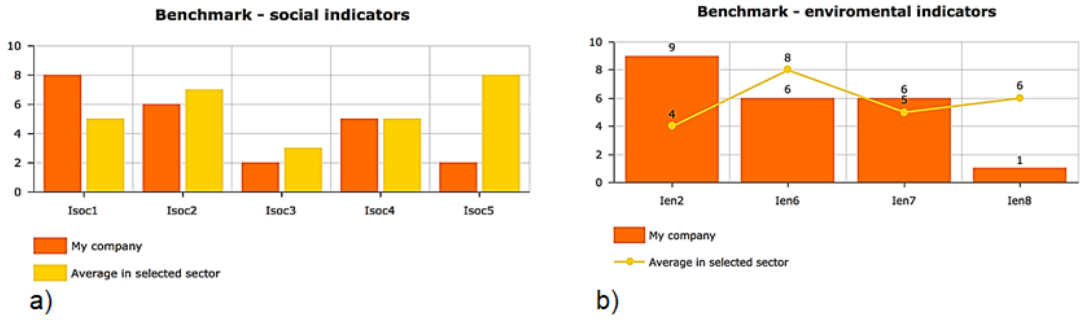


Figure 2 a) Benchmark – clustered graph; b) Benchmark – column and line graph

Proposed method of sustainability assessment providing an active participation in decision-making. Whereas the improvements include several modifications, in order to achieve the following factors:

- comprehensive sustainability assessment: we focused our efforts on developing a comprehensive sustainability assessment. Therefore, environmental, social, economic and corporate governance indicators should be integrated. In this case, the proposed model won't only deal with financial indicators but should also include non-financial ones;
- suitability: The assessment should be done for different companies in the Czech Republic. However, the model can't be universal, because the indicators should reflect the specifics of the industry in which the company operates. Therefore, different available sustainability frameworks are used and specific set of indicators are chosen for each sector (e.g. agriculture, manufacture, breweries company, ...);
- simplicity and applicability: The modified model should be easy, simple, suitable and accurate. It reflects not only three dimensions (economic, environment, and social), but also the corporate governance pillar is added. As mentioned above, EVA is the most important and measured indicator which combines all the basic components required to describe the economic situation of the company. For this reason, the gross value added (VA) is replaced by Economic Value Added to describe the financial situation of the companies more efficiently.

After applying all mentioned modifications on the Eq. (1), it is presented in Eq. (3).

$$\mathrm{SVA} = 3\mathrm{EVA}_c - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathrm{EI}_{i,c}}{\mathrm{EI}_{i,b}}\mathrm{EVA}_b\right) - \frac{1}{M}\sum_{j=1}^{M}\left(\frac{\mathrm{SI}_{j,c}}{\mathrm{SI}_{j,b}}\mathrm{EVA}_b\right) - \frac{1}{K}\sum_{l=1}^{K}\left(\frac{\mathrm{GI}_{l,c}}{\mathrm{GI}_{l,b}}\mathrm{EVA}_b\right) \tag{3}$$

where EI, SI, GI, w_{Ei}, w_{Ej}, w_{Gl} are values and the weights of environment, social and governance indicators, respectively. Symbol b refers to benchmark, while symbol c refers to the studied company. According to Eq. (3), increasing the value of environment indicator "for example the amount of hazardous waste is increased", will effect negatively on SVA. Increasing the economic value added of the company, in turn increases SVA value. In order to improve our proposed model, we supposed that different indicators don't effect equally on enterprises score. That means, each indicator should have a different weight on sustaina-

bility calculation. This weight differs according to the country, size and sector of the company. The implementation of this improvement can be done using Eq. (4), where w_l is the average weight for i's indicator after applying DEA model.

$$\mathrm{SVA} = 3\mathrm{EVA}_c - \frac{1}{N}\sum_{i=1}^{n}\left(w_{i,av}\frac{\mathrm{EI}_{i,c}}{\mathrm{EI}_{i,b}}\mathrm{EVA}_b\right) - \frac{1}{M}\sum_{j=1}^{m}\left(w_{j,avg}\frac{\mathrm{SI}_{j,c}}{\mathrm{SI}_{j,b}}\mathrm{EVA}_b\right) - \frac{1}{K}\sum_{l=1}^{k}\left(w_{l,av}\frac{\mathrm{GI}_{l,c}}{\mathrm{GI}_{l,b}}\mathrm{EVA}_b\right) \quad (4)$$

In our model, the benchmark value and the weight of each indicator are calculated using DEA model. By applying this model, the most efficient company can be determined which is taken as a benchmark company. More information about these calculations can be found in [18].

In order to apply this method, an example of five Czech biogas plant A, B, C, D, E (South-Moravian region) is used. The important data is extracted from Amadeus database [1]. After that, without loss of generality, few KPIs are chosen. We considered KPIs, water consumption (ENV3) in m^3, social cost (SO1) in CZK, sickness (SO3) in percentage and remuneration of board members (GOV1) in CZK as organizations' inputs. The Organizations' output is Economic Value Added (EC1) in CZK, produced heat (ENV1) in mega joule and generated electrical power (ENV2) in megawatt-hours. In additional, a dual-role factor is considered as employee training (SO2) in CZK.

As Mentioned before, the weight of each indicator and the benchmark values will be calculated using DEA model. This model computes efficiency score of selected organizations (DMUs) using the linear programing in Maple [16]. Tab. I summarize the output of DEA model which consists of several parameters. These parameters are the efficiency scores of biogas stations and the weight of each indicator used in their sustainability assessment.

Company	**Efficiency Score**	w_{ENV3}	w_{SOC1}	w_{SOC3}	w_{GOV1}	w_{SOC2}	**SVA**
A	1.00	0	0.0047	0	0.0124	0	1629
B	0.95	0	3.3333	0	0	0.0427	70
C	1.00	0	0.0084	0	0.0219	0	52073
D	0.99	0.0614	0	0	0.0027	0.0413	78
E	0.85	0.0444	0	0.0117	0.0121	0.324	79

Table 1 Efficiency score, input (DEA model) and Sustainability value added (SVA)

The above-mentioned DEA model is solved five times, ones for each target biogas stations. As a result, there are five sets of input and output weights. The five efficiency values are then averaged as the overall performance of the biogas stations sector. The highest efficiency score, company C, will be used as benchmark values in Eq. (3) for sustainability calculation. By applying Eq. (3) the sustainability value added of each enterprise is calculated.

Table 1 (section SVA) presents the final assessment of studied companies. These values make us understand which company has better contributed to sustainability development. The results present the sustainability value assessments with applying the averaged weights Eq. (4) respectively. According to presented results company C is the most sustainable company, while B or D is the less sustainable one using SVA model. Fig. 3 presents the proved version of sustainability value added which takes into consideration the weight of each indicator.

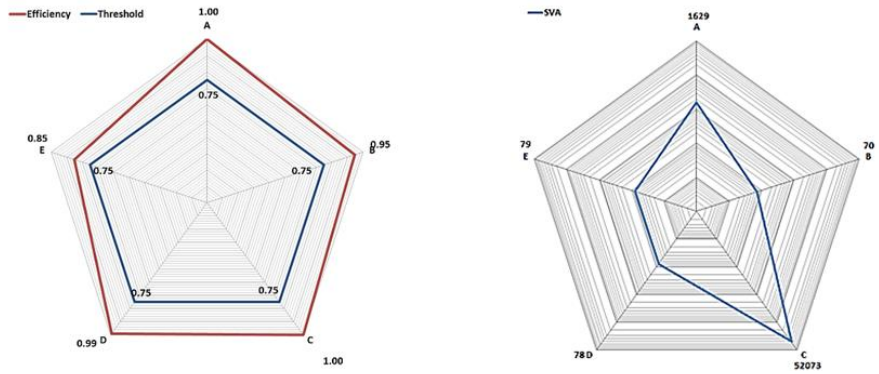


Figure 2 The level of efficiency of the five Czech biogas stations

4 Conclusions

Sustainability assessment and sustainable investment is a comprehensive process to achieve the best performance and determine the weak points of the studied organization. An expensive data collecting and managing, difficulty of determining the appropriate sustainability indicators and capturing reliable data-information are the main barriers that face different organizations. In order to overcome them, WEBRIS 2 system is suggested. WEBRIS 2 is a combination of different information and communication technologies which can be used for quick and efficient data aggregation and assessment. The backbone of this system is a model used for sustainability assessment.

This paper aims to propose an improved method of sustainability assessment, sustainable investments. It employs important and widely used financial value (e.g. SVA, EVA) and new data oriented analysis (e.g. DEA) for evaluating the efficiency of number of producers. This work is supported by make it reflects the specifics requirements of the country and industry (biogas stations). The weights and benchmark values are calculated and used in SVA model. The results visualization is presented in the case study for biogas stations sector. According to our proposed method each company from different sectors can assess their sustainability in easy fast way. Then compare their results with other companies in the same sector.

Acknowledgements

This paper is supported by the Czech Science Foundation. Project name: Modelling and simulation of sustainable investment decision-making [GA17-23448S] and the project Research Infrastructure for Young Scientists, this is co-financed from Operational Programme Research, Development and Education [CZ.02.1.01/0.0/0.0/16_017/0002334].

References

[1] Amcharts – Chart Maker. (2020). [cit. 2020-4-21] Available at: https://www.amcharts.com/javascript-charts.

[2] Business register. (2020). [cit. 2020-4-21] Available at: https://or.justice.cz/ias/ui/rejstrik.

[3] CARROLL, A. B. (2000). Ethical challenges for business in the new millennium: Corporate social responsibility and models of management morality. *Business Ethics Quarterly*. 10, 33–42.

[4] CHABOWSKI, B. R., MENA, J. A., & GONZALES-PARDON, T. L. (2011). The structure of sustainability research in marketing. *Journal of the Academy of Marketing Science.* 39, 55–70.

[5] CHARNES, A., COOPER, W., & RHODES, E. (1978). Measuring the efficiency of decision making units. European *Journal of Operational Research*. 2, 429–444.

[6] COOK, W. D., GREEN, R. H., & ZHU, J. (2006). Dual-role factors in data envelopment analysis. *IIE Transactions*. 38, 105–115.

[7] CRUZ, L. B, PEDROZO, E. Á., & ESTIVALETE, B. F. (2006). Towards sustainable development strategies: a complex view following the contribution of Edgar Morin. *Management Decision*. 44, 871–891.

[8] DOYLE, J. & GREEN, R. (1994). Efficiency and cross-efficiency in DEA: derivations, meanings and uses. *Journal of the Operations Research Society*. 45, 567–578.

[9] EUROSIF. European SRI Study 2016. (2016). [cit. 2020-4-21] Available at: http://www.eurosif.org/wp-content/uploads/2016/11/SRI-study-2016-HR.pdf.

[10] FIGGE, F. & HAHN, T. (2004). Sustainable value added-measuring corporate contributions to sustainability beyond eco-efficiency. *Ecological economics*. 48, 173–187.
[11] HART, S. L. & MILSTEIN, M. B. (2003). Creating sustainable value. *The Academy of Management Executive*. 17, 56–67.
[12] HODINKA, M., ŠTENCL, M., HŘEBÍČEK, J., & TRENZ, O. (2012). Current Trends of Corporate Performance Reporting Tools and Methodology Design Of Multifactor Measurement Of Company Overall Performance. *Acta univ. Agric. et silvic. Mendel. Brun*. 2, 85–89.
[13] HŘEBÍČEK, J., FALDÍK, O., CHVÁTALOVÁ, Z., KASEM, E., & TRENZ, O. (2016). Sustainability assessment of biogas plants. *Proceedings Venice 2016*, Sixth International Symposium on Energy from Biomass and Waste.
[14] HŘEBÍČEK, J., FALDÍK, O., KASEM, E., & TRENZ, O. (2015). Determinants of Sustainability Reporting in Food and Agriculture Sectors. *Acta univ. Agric. et silvic. Mendel. Brun*. 2, 539–552.
[15] HŘEBÍČEK, J., TRENZ, O., & VERNEROVÁ, E. (2013). Optimal set of agrienvironmental indicators for the agricultural sector of Czech Republic. *Acta univ. Agric. et silvic. Mendel. Brun*. 7, 2171–218.
[16] HŘEBÍČEK, J., TRENZ, O., CHVÁTALOVÁ, Z., & SOUKOPOVÁ, J. (2014). Optimization of corporate performance using data envelopment analysis with Maple. *In Aurelio Araujo et al. Engineering Optimization*. 763–767.
[17] KAPLAN, R. S. & NORTON, D. P. (2001). Transforming the balanced scorecard from performance measurement to strategic management: Part I. *Accounting horizons*. 15, 87–104.
[18] KASEM, E., TRENZ, O., HŘEBÍČEK, J., & FALDÍK, O. (2015). Key Sustainability Performance Indicator Analysis for Czech Breweries. *Acta univ. Agric. et silvic. Mendel. Brun*. 6, 1937–1944.
[19] KOCMANOVÁ, A., HŘEBÍČEK, J., ET AL. (2013). Měření podnikové výkonnosti. Brno: Littera.
[20] KUOSMANEN, T. & KUOSMANEN, N. (2009). How not to measure sustainable value (and how one might). *Ecological Economics*. 69, 235–243.
[21] LIANG, L., WU, J., COOK, W. D., & ZHU, J. (2008). Alternative secondary goals in DEA cross efficiency evaluation. *International Journal of Production Economics*. 113, 1025–1030.
[22] MVC Framework Tutorial. (2016). [cit. 2020-4-21] Available at: http://www.tutorialspoint.com/mvc_framework.
[23] SEXTON, T. R., SILKMAN, R. H., & HOGAN, A. J. (1986). Data envelopment analysis: critique and extensions. In: Measuring Efficiency: An Assessment of Data Envelopment Analysis. 73–105.
[24] SIEW, R. (2015). A review of corporate sustainability reporting tools (SRTs). *Journal of Environmental Management*. 164, 180–195.
[25] WADDOCK, S. & SMITH, N. (2000). Corporate responsibility audits: Doing well by doing good. *Sloan Management Review*. 41, 75–83.
[26] WANG, Y. M. & CHIN, K. S. (2010). Some alternative models for DEA cross-efficiency evaluation. International *Journal of Production Economics*. 128, 332–338.

Consumption Development of V4 Countries over the Business Cycle

Terézia Vančová[1], Luboš Střelec[2]

Abstract. This paper estimates marginal propensities to consume out of current and lagged disposable income for individual Visegrad Four countries using quarterly data covering the period from 1995(1) until 2018(4). To analyse the effect of business cycle on the consumption and disposable income development, these time series are divided into two subsamples according to the growth or decline of the real GDP. There are calculated growth rates of consumption expenditures and disposable income for the whole time series and separately for the different phases of business cycle. The average quarterly growth rates of consumption series are higher than the average quarterly growth rates of disposable income series. During expansion the average grow rates of consumption are even higher than of disposable income and during recessions the consumption decreased further then disposable income. Consumption is excessively sensitive to current income and consumption is not protected against business-cycle fluctuations.

Keywords: business cycle, consumption, disposable income, growth rates, time series, Visegrad Group (V4) countries

JEL Classification: C01, C22, E10, E21, E32
AMS Classification: 62M10, 62P20, 91B62, 91B55, 91B84

1 Introduction

Final consumption expenditure of households comprises more than 50% of GDP in the V4 countries and therefore the study of consumption dynamics is of great importance for the macroeconomics policy adjustment and predictions (Eurostat [6]).

The recent empirical analyses have shown that traditional models of Permanent Income Hypothesis and Life-cycle Hypothesis do not reflect consumption behaviour expressed through the consumption function (Jappelli and Pistaferri [12]; Broda and Parker [1]; Parker [21] and Kueng [16]). Empirical evidence finds consumption to be excessively sensitive to measured income and excessively smooth to unpredictable current income change (Flavin [7] and Deaton [2]).

Dossche et al. [4] confirm a significant role of Euro area consumption in the economic expansion that started in 2013. The consumption growth has followed the growth in household income. Household income and wealth explain the largest part of consumption growth and only during the period of the Great Recession was private consumption lower than this simple relationship with income and wealth would suggest (Dossche et al. [4]).

This paper is structured as follows: section 2 offers a brief overview of the theoretical approach to the consumption function and the recent literature discussion in the context of the business cycle. Data and methods are introduced in the section 3. Section 4 presents the results of the empirical analysis and the discussion and section 5 is dedicated to the conclusions of the analysis.

2 Consumption function

The importance of studying consumption expenditures originates from the policy implications on the trend in output, economic fluctuations and smoothing the business cycles. In the original Keynesian literature, consumption has been considered a linear function of disposable income (Keynes [15]). Changes in taxes or transfer payments can thus play a substantial role in the macroeconomic stabilization policy. Then with an oncoming of Friedman's [8] Permanent Income Hypothesis (PIH) and Franco Modigliani and Richard Brumberg's [18] Life Cycle Hypothesis (LCH), an assumption of high marginal propensity to consume out of

[1]Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, Czech Republic, e-mail: terezia.vancova@mendelu.cz.
[2]Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, Brno, Czech Republic, e-mail: lubos.strelec@mendelu.cz.

current income has been questioned. The Permanent Income-Life Cycle model suggests that consumers maintain a steady level of consumption expenditures determined by consumers' lifetime resources.

Hall [11] has shown that consumption has under rational expectations the characteristic of a random walk process. Flavin [7] has discovered a discrepancy in this statement because of the non-zero correlation between the change in consumption and the lagged changes in income. The presence of borrowing constraints is a standard answer to this excess sensitivity phenomenon (Gourinchas and Parker [9]; Kaplan and Violante [13], [14]; Jappelli and Pistaferri [12]; Einian and Nili [5]).

2.1 Consumption in the business cycle context

The business cycle can be defined as short-run fluctuations in output and employment caused by shocks to aggregate demand and aggregate supply (Mankiw [17]). Empirical analyses of business cycles have led to the finding that the alternation of expansions and recessions is highly irregular.

Various approaches to the detection of a recession are based on a quite different assumption. A recession is widely defined in terms of negative growth of real GDP in at least two successive quarters. National Bureau of Economic Research [19] offer an alternative definition "A recession is a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales."

According to Permanent Income Hypothesis, permanent income is smoother than current income so that consumption is smoother than current income. Permanent Income Hypothesis implies that both consumption and labour income trend upward over time, but income fluctuates more about its trend than consumption. The saving ratio is pro-cyclical, rising in the boom, and falling in the slump, so that consumption is to some extent protected against business-cycle fluctuations in income (Deaton [3]).

Paap and Dijk [20] show that the average growth rates of US per capita consumption and income differ in recession and expansion periods. Recessions are more pronounced in the income series than in the consumption series and expansions are also more manifested in the income series than in the consumption series (The average quarterly growth rate of per capita income is higher than the average quarterly growth rate of per capita consumption.). This observation that consumption is smoother than income is consistent with the Permanent Income-Life Cycle Hypothesis.

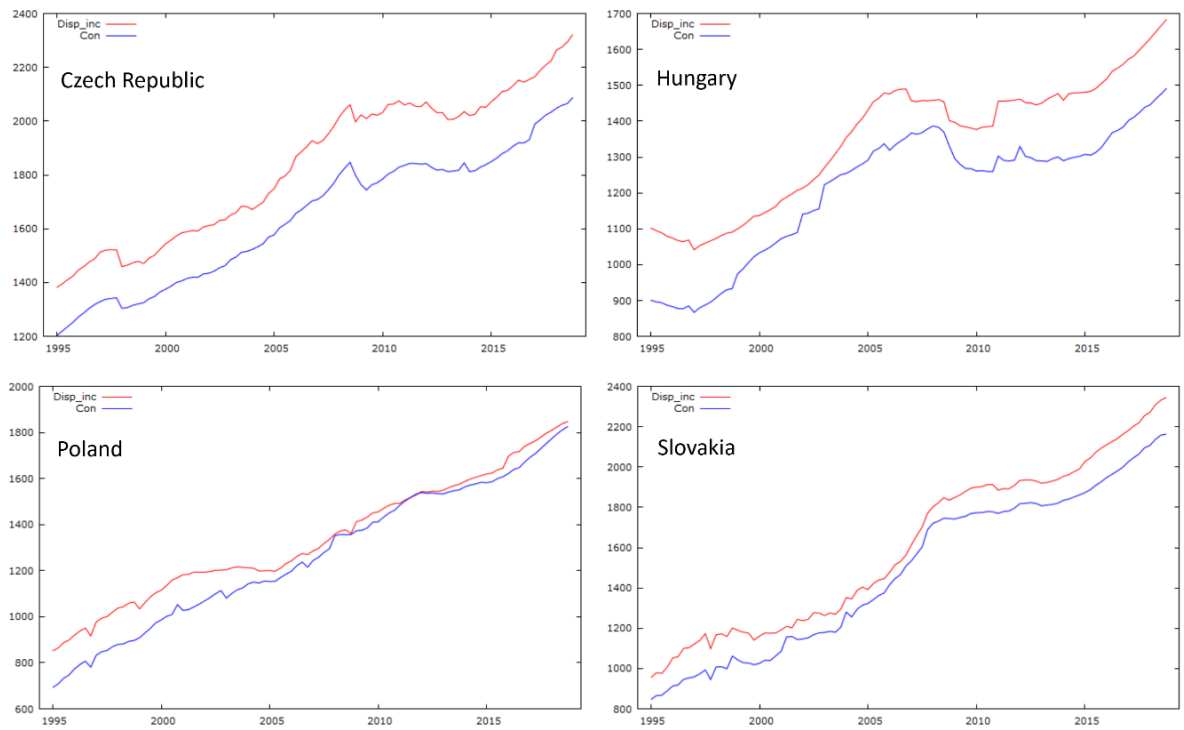


Figure 1 Real per capita consumption expenditure and disposable income: Visegrad Group, 1995: 1–2018: 4

Figure 1 illustrates the development of per capita consumption expenditure and per capita disposable income of Visegrad Group countries covering the period from the first quarter of 1995 till the final quarter of 2018. Consumption series as well as income series are increasing over the sample time with some shorter periods of decline. The value of aggregate disposable income exceeds the value of consumption expenditure that creates a space for savings generating in all V4 countries almost through the entire period. Poland is the only exception with some short periods of not generating aggregate savings. It seems like consumption in the V4 countries is not as smooth as Permanent Income Hypothesis assumes.

3 Data and Methods

The empirical study is based on the time series analysis using individual Visegrad Four countries data covering the period from the first quarter of 1995 until the final quarter of 2018. Eurostat is the data source for time series of gross domestic product, GDP implicit deflator, gross disposable income and final consumption expenditure of households. The time series are measured as per capita aggregates in euro currency and a GDP implicit deflator is used to acquire the real terms of employed variables, the base year being 2010. There is not available the quarterly frequency of disposable income variable. The segregation of annual figures into quarterly figures can be calculated by external and economically related variable. Real GDP per capita is in this context a relevant time series because there is a strong positive relationship between these selected time series. The correlation coefficient between annual time series of real GDP and disposable income reaches value over 0.98 and is statistically significant at the 0.01 level in case of every Visegrad Four country. The annual time series of disposable income is disaggregated into a quarterly time series by computing of the proportional shares of each real GDP quarter to its annual value. Chosen economic variables are seasonally adjusted by the TRAMO/SEATS procedure.

The empirical analysis is divided into two part. The first part is devoted to the V4 countries models of consumption function. A dynamic regression model is individually applied for every Visegrad Group country. These Autoregressive distributed lag models are estimated in the most suitable way of lagged consumption and various lags of disposable income (1). The Augmented Dickey Fuller (ADF) test with no intercept is to test the null hypothesis that a unit root is present in residuals.

$$C_t = \beta_0 + \beta_1 C_{(t-1)} + \beta_2 Yd_{it} + \beta_3 Yd_{(t-1)} + \ldots + \beta_k Yd_{(t-k)} + \varepsilon_t \tag{1}$$

The second part is then devoted to the development of consumption expenditures and disposable income during the different phases of business cycle. A recession is identified as the negative growth of GDP in at least two successive quarters. A peak is defined as the last period of an expansion and a trough as the last period of a recession. Average growth rates of consumption expenditures as well as of disposable income are calculated particularly for the phase of expansion and for the phase of recession.

4 Results

In the first part of the empirical analysis, the consumption functions of every Visegrad Group country in the form of dynamic regression model (1) are estimated. ARDL(1,1) model is the most suitable form of dynamic model for each of V4 countries, more lags are not significant. Beside the lagged consumption, current and one lagged disposable income help to explain current consumption. These results show that a lagged disposable income has a predictable value on current consumption in every Visegrad Four country. This reveals an evidence against the Permanent Income Hypothesis/Random Walk Hypothesis and points out to the excess sensitivity of consumption to current income. This conclusion is consistent with the traditional work of Flavin [7] as well as more recent studies like Grochová and Střelec [10], Parker [21] or Kueng [16]. Jappelli and Pistaferri [12], Kaplan and Violante [14] or Einian and Nili [5] find the explanation of these circumstances in the presence of liquidity constraints or precautionary savings. There are parameters estimations of the dynamic models for individual V4 countries in the Table 1.

Country	β_0	β_1	β_2	β_3
Czechia	1.05	0.89	0.63	−0.53
Hungary	8.86	0.95	0.72	−0.68
Poland	−9.48	0.94	0.60	−0.53
Slovakia	2.51	0.96	0.77	−0.73

Table 1 The coefficients estimations of dynamic regression models for individual V4 countries

Because of possible concerns over spurious regression, residuals of dynamic regression models for individual V4 countries are tested using ADF test. Residuals of created models are stationary and not autocorrelated and the threat of spurious regression is excluded in every case (Table 2).

Country	ADF test	Durbin–Watson test	Ljung–Box test
Czechia	<0.001	0.051	0.182
Hungary	<0.001	0.342	0.289
Poland	<0.001	0.410	0.706
Slovakia	<0.001	0.116	0.124

Table 2 P-values of ADF test, Durbin–Watson test and Ljung–Box test for residuals of dynamic regression models for individual V4 countries

The estimated parameters of each model are quite similar in terms of interpretation. Coefficient β_1 reaches nearly an unity value, only in Czech Republic is moderately lower. A negative coefficient of the first lag of disposable income is slightly outweighed by a positive coefficient of current disposable income. The highest positive effect of disposable income is captured in the Czech Republic, followed by Poland.

There has been an evidence on the excess smoothness of consumption of individual V4 countries in the previous research (Vančová [23]). At once the consumption expenditures of the single V4 counties are both excessively smooth to a sudden shock to the lifetime income and excessively sensitive to the lagged income growth. Consumption responds less to unanticipated changes (innovations, shocks) than Permanent Income Hypothesis would expect. Vančová and Střelec [24] show the degree of persistence of consumption in the V4 countries is high that means consumption tends to stay above or below its trend after a shock for a longer time.

Romer [22] confirms the coherence of the excess smoothness and sensitivity because excess sensitivity deals with anticipated changes in income and excess smoothness with unanticipated changes in income. Changes in consumption are related to the average of previous changes and just moderately to the innovations in income.

In the second part of the empirical analysis, the development of consumption expenditures and disposable income during the different stages of business cycle is examined. The period of the first quarter of 1995 until the final quarter of 2018 is divided into two variant phases: the upward phase, expansion, and the downward phase, recession. The identification of the recession is based on the negative growth of real GDP in at least two successive quarters. The beginning and the end of the several recessions in the Czech Republic, Hungary, Poland and Slovakia are presented in the Table 3. There are two identified recession periods in Hungary, Poland and Slovakia and three classified recession periods in the Czech Republic. The timings of downward phases of V4 countries are quite similar because their economics are interconnected and are the part of open global economic network.

Country	Phase of recession	
	The beginning	The end
Czechia	1997:2	1998:1
	2008:3	2009:3
	2011:4	2012:4
Hungary	2008:2	2009:2
	2011:4	2012:4
Poland	2000:4	2001:2
	2012:2	2012:4
Slovakia	1999:1	1999:4
	2008:4	2009:1

Table 3 The phases of identified recessions in the V4 countries

The average growth rates of consumption expenditures and of disposable income are calculated for the entire time series and separately for the selected phases of recessions and of expansions (Table 4).

Country	Whole time series		Expansion phase		Recession phase	
	Disposable income	Consumption	Disposable income	Consumption	Disposable income	Consumption
Czechia	1.0055	1.0058	1.0071	1.0076	0.9961	0.9953
Hungary	1.0045	1.0053	1.0064	1.0077	0.9940	0.9930
Poland	1.0082	1.0103	1.0085	1.0111	1.0036	0.9980
Slovakia	1.0095	1.0099	1.0107	1.0111	0.9917	0.9916

Table 4 Average growth rates of consumption expenditures and of disposable income in the V4 countries

The series of consumption expenditures and disposable income are mainly increasing over the sample time with moderately short periods of decline. For the whole sample, the average quarterly growth rate of consumption is quite higher than the average quarterly growth rate of disposable income. To evaluate the effect of the business cycle phases on the consumption and disposable income development, the sample is divided into two subsamples. The first subsample contains quarters corresponding to expansion observations: the average growth rates of consumption expenditures and disposable income in expansion are as expected higher. The differences in the average growth rates of consumption expenditures and the average growth rates of disposable income also reach higher values in expansion phase. The second subsample corresponds to quarters which are labelled as a recession: this downward phase is more pronounced in the consumption series than in the disposable income series.

Both consumption and disposable income trend upward over the period of 1995–2018 but consumption expenditures grow faster than disposable income. The differences in growth rates of consumption and disposable income in each stage of the business cycle display that consumption is less protected against business cycle.

5 Conclusions

This paper indicates the excess sensitivity of consumption to disposable income in all Visegrad Four countries. Dynamic regression models for individual Visegrad Four countries over the period from 1995(1) until 2018(4) shows that current disposable income as well as a lagged disposable income has predictive value on current consumption expenditures. Consumption seems to be not that smooth as Permanent Income-Life Cycle suggests. The countercyclical fiscal policy on a regular basis aims to boost consumption expenditures of household through reductions in taxes or increases in social benefits. There is a suitable opportunity for the fiscal policy in the contribution to the growth of household consumption expenditures because changes in current income that are predictable can have a significant effect on consumption.

The average growth rates of consumption and disposable income in both stages of the business cycle are various. During the expansion consumption expenditures grow faster than disposable income and during the recession consumption expenditures again decrease faster than disposable income. This suggests a higher exposure of consumption expenditures to economic fluctuations. Fiscal policy measures affecting household current income can possibly change consumption more effectively during the recovery phase.

Acknowledgements

Supported by the grant No. PEF_DP_2020030 of the Grant Agency IGA PEF MENDELU.

References

[1] BRODA, CH. & PARKER, J. (2014). The Economic Stimulus Payments of 2008 and the Aggregate Demand for Consumption. *Journal of Monetary Economics*, **68**, 20–36.

[2] DEATON, A. S. (1987). Life-Cycle Models of Consumption: Is the Evidence Consistent with the Theory? Advances in Econometrics. Fifth World Congress. Vol. II. Cambridge and New York: *Cambridge University Press*, **2**, 121–148.

[3] DEATON, A. S. (1992). *Understanding Consumption*. 1st edition. New York: Oxford University Press.

[4] DOSSCHE, M., FORSELLS, M., ROSSI L., & STOEVSKY, G. (2018). Private consumption and its drivers in the current economic expansion. *ECB Economic Bulletin*, Issue 5/2018, 86–110.

[5] EINIAN, M. & NILI, M. (2020). Excess sensitivity and borrowing constraints: Evidence from Iranian households. *Econ Transit Institut Change*, **28**, 137–160.
[6] EUROSTAT. (2019). Final consumption expenditure of households and non-profit institutions serving households. National accounts and GDP.
[7] FLAVIN, M. (1981). The Adjustment of Consumption to Changing Expectations about Future Income. *Journal of Political Economy*, **89**(5), 974–1009.
[8] FRIEDMAN, M. (1957). *A Theory of the Consumption Function*. 1st edition. Princeton: Princeton University Press.
[9] GOURINCHAS, P.-O., & PARKER, J. A. (2002). Consumption over the Life Cycle. *Econometrica*, **70**(1), 47–89.
[10] GROCHOVÁ, L. & STŘELEC, L. (2013). Quantitative modelling of consumption patterns in the CEEC. *Proceedings of the 31st International Conference Mathematical Methods in Economics 2013*. Jihlava: College of Polytechnics Jihlava, 207–212.
[11] HALL, R. E. (1978). Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence. *Journal of Political Economy*, **86**(6), 971–987.
[12] JAPPELLI, T. & PISTAFERRI, L. (2011). The Consumption Response to Income Changes. *Annual Review of Economics*, **2**, 479–506.
[13] KAPLAN, G. & VIOLANTE, G. L. (2010). How much Consumption Insurance beyond Self-insurance? *American Economic Journal: Macroeconomics*, **2**(4), 53–87.
[14] KAPLAN, G. & VIOLANTE, G. L. (2014). A Model of the Consumption Response to Fiscal Stimulus Payments. *Econometrica*, **82**(4), 1199–1239.
[15] KEYNES, J. M. (1936). *The general theory of employment, interest, and money*. United Kingdom: Palgrave Macmillan.
[16] KUENG, L. (2018). Excess Sensitivity of High-Income Consumers. *The Quarterly Journal of Economics*, **133**(4), 1693–1751.
[17] MANKIW, N. G. (2002). *Macroeconomics*, 5th Edition. New York: Worth Publishers.
[18] MODIGLIANI, F. & BRUMBERG, R. (1954). Utility analysis and the consumption function: an interpretation of cross-section data. *Post-Keynesians Economics, K Kurihara edition. New Brunswick: Rutgers University Press*, 388–436.
[19] NATIONAL BUREAU OF ECONOMIC RESEARCH. (2008). The NBER's Recession Dating Procedure. Business Cycle Dating Committee, *National Bureau of Economic Research*.
[20] PAAP, R. & DIRK, H. (2003). Bayes Estimates of Markov Trends in Possibly Cointegrated Series, *Journal of Business and Economic Statistics*, **21**(4), 547–563.
[21] PARKER, J. (2017). Why Don't Households Smooth Consumption? Evidence from a 25 Million Dollar Experiment. *NBER Working Paper*, No. 21369.
[22] ROMER, D. (1996). *Advanced Macroeconomics*. 1st edition. New York: The McGraw-Hill.
[23] VANČOVÁ, T. (2019). The excess smoothness and sensitivity of consumption in the V4 countries. *Acta Universitatis agriculturae et silviculturae Mendelianae Brunensis = Acta of Mendel University of agriculture and forestry Brno = Acta Mendelovy zemědělské a lesnické univerzity v Brně*. **67**(6), 1653–1663.
[24] VANČOVÁ, T. & STŘELEC, L. (2019). The Persistence of Consumption and Disposable Income in the V4 Countries. In *Mathematical Methods in Economics 2019: Conference Proceedings*. České Budějovice: Jihočeská univerzita v Českých Budějovicích. 109–114.

Identification of Temporal Patterns in Income and Living Conditions of Czech Households: Clustering Based on Mixed Type Panel Data from the EU-SILC Database

Jan Vávra[1], Arnošt Komárek[2]

Abstract. The EU-SILC database contains annually gathered rotating-panel data on a household level covering indicators of monetary poverty, severe material deprivation or low work household intensity. Data are obtained via questionnaires leading to outcome variables of diverse nature: *numeric*, *binary*, *ordinal* being gathered at each occasion in each household. Only limited number of approaches exist in the literature to analyze such *mixed-type* panel data. We present a statistical model for such type of data which is built on a thresholding approach to link binary or ordinal variables to their latent numeric counterparts. All, possibly latent, numeric outcomes are then jointly modelled using a multivariate version of the linear mixed-effects model. A mixture of such models is then used to model heterogeneity in temporal evolution of considered outcomes across households. A Bayesian variant of the Model Based Clustering (MBC) methodology is finally exploited to classify households into groups with similar evolution of indicators of monetary poverty, material deprivation or low work household intensity. The method is applied to socially-economic focused dataset of Czech households gathered in a time span 2005–2016.

Keywords: multivariate panel data, mixed type outcome, model based clustering, classification

JEL Classification: C33, C38
AMS Classification: 62H30

1 Data and research problem

Throughout the EU states the poverty and social exclusion is measured using indicators of monetary poverty, severe material deprivation and very low work household intensity. Relevant data are gathered within *The European Union Statistics on Income and Living Conditions* project (EU-SILC, https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions. This is an instrument with the goal to collect timely and comparable cross-sectional and *longitudinal multidimensional* microdata on income, poverty, social exclusion and living conditions. Data are obtained via questionnaires leading to outcome variables of diverse nature: *numeric* (e.g., income), *binary* (e.g., affordability of paying unexpected expenses) and *ordinal* (e.g., level of ability to make ends meet). It is our primary aim to use such longitudinally gathered outcomes towards segmentation of households according to typical patterns of their temporal evolution.

To this end, we propose a statistical model capable of joint modelling of longitudinal outcomes of diverse nature (*numeric, binary, ordinal*) while taking potential dependencies as well longitudinal as among different outcomes obtained at each occasion into account. This is a topic of Section 2. Consequently, we use the model within a Bayesian model based clustering (MBC) procedure to perform unsupervised classification of study units (households) into groups whose characteristics are not known in advance. This part of methodology is described in Section 3. The final Section 4 describes in detail the use of this methodology on the Czech subset of the EU-SILC dataset. The paper is finalized by conclusions in Section 5

2 Joint model for mixed type panel data

In general, we have data on n units/panel members (e.g., households) at our disposal containing $R \geq 1$ longitudinally gathered outcomes (e.g., income, affordability of week holiday and level of a financial burden

[1] Charles University, Faculty of Mathematics and Physics, Dept. of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic, vavraj@karlin.mff.cuni.cz
[2] Charles University, Faculty of Mathematics and Physics, Dept. of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic, komarek@karlin.mff.cuni.cz

of housing). Let $\mathbf{Y}_i = \left(\mathbf{Y}_{i,1}^\top, \ldots, \mathbf{Y}_{i,R}^\top\right)^\top$ stand for a vector consisting of all the values $\mathbf{Y}_{i,r} = (Y_{i,r,1}, \ldots, Y_{i,r,n_i})^\top$ of the rth outcome ($r = 1, \ldots, R$) of the ith unit ($i = 1, \ldots, n$) obtained at n_i occasions. Let $\mathcal{C}_i$ stand for available covariates (the times of measurements, possibly other explanatory variables) of i-th unit. Finally, let $g(\mathbf{y}_i; \mathcal{C}_i, \boldsymbol{\theta})$ represent the assumed distribution of the outcome vector $\mathbf{Y}_i$ which possibly depends on the covariates $\mathcal{C}_i$ and also on a vector $\boldsymbol{\theta}$ of unknown parameters. It is assumed that this distribution is built from the following hierarchical model.

First, if the rth, $r = 1, \ldots, R$, longitudinal outcome vector $\mathbf{Y}_{i,r}$ is composed of *ordinal* or *binary* variables, we will take a natural thresholding approach (see, e.g., Dunson [1]) and will assume that each element of $\mathbf{Y}_{i,r}$, $Y_{i,r,j}$, $j = 1, \ldots, n_i$, is determined by corresponding latent continuous variable $Y^\star_{i,r,j}$, which is covered by one of the intervals given by the set of thresholds $\boldsymbol{\gamma}_r$. That is,

$$Y_{i,r,j} = l, \qquad \text{iff} \quad \gamma_{r,l} < Y^\star_{i,r,j} \leq \gamma_{r,l+1}, \qquad l = 0, \ldots, L_r, \tag{1}$$

where $\boldsymbol{\gamma}_r = \left(\gamma_{r,1}, \ldots, \gamma_{r,L_r}\right)^\top$ are unknown thresholds such that $-\infty = \gamma_{r,0} < \gamma_{r,1} < \cdots < \gamma_{r,L_r} < \gamma_{r,L_r+1} = \infty$. In the following, denote these latent continuous counterparts by $\mathbf{Y}^\star_{i,r}$. In case the rth longitudinal outcome is directly observed as *continuous*, we set $\mathbf{Y}^\star_{i,r} = \mathbf{Y}_{i,r}$.

Further, for each $\mathbf{Y}^\star_{i,r}$, $r = 1, \ldots, R$, a classical linear mixed model (LMM) is assumed. That is,

$$\mathbf{Y}^\star_{i,r} = \mathbb{X}_{i,r}\boldsymbol{\beta}_r + \mathbb{Z}_{i,r}\mathbf{B}_{i,r} + \boldsymbol{\varepsilon}_{i,r}, \tag{2}$$

where $\mathbb{X}_{i,r}$ and $\mathbb{Z}_{i,r}$ are model matrices derived from the covariate information $\mathcal{C}_i$, $\boldsymbol{\beta}_r$ is a vector of unknown parameters. Further, $\mathbf{B}_{i,r}$ is a vector of random effects related to the rth longitudinal outcome and $\boldsymbol{\varepsilon}_{i,r}$ is an error term vector for which a classical normality assumption is exploited, i.e., $\boldsymbol{\varepsilon}_{i,r} \sim \mathcal{N}_{n_i}\left(0, (\tau_r)^{-1}\mathbf{I}_{n_i}\right)$. The residual variance $(\tau_r)^{-1}$ is unknown.

Dependencies among the R longitudinal outcomes $\mathbf{Y}_{i,1}, \ldots, \mathbf{Y}_{i,R}$ are taken into account by considering a joint distribution for the random effects vector $\mathbf{B}_i = \left(\mathbf{B}_{i,1}^\top, \ldots, \mathbf{B}_{i,R}^\top\right)^\top$ which joins the random effect vectors from the mixed models for all R longitudinal measurements. Namely, a multivariate normal distribution is assumed here, i.e., $\mathbf{B}_i \sim \mathcal{N}_q\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, where both the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ are unknown parameters.

Finally, let $\boldsymbol{\zeta}$ be the set of unknown parameters of interest, i.e. $\boldsymbol{\zeta} = \{\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$, where $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ stand for sets of parameters $\boldsymbol{\beta}_r$ and τ_r across all outcomes $r = 1, \ldots, R$. Then, the density of (latent) continuous outcomes of the i-th individual is given by integration of product of a multivariate normal density related to the LMM and a density of $\mathcal{N}_q\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, which is known to lead to the density $g^\star\left(\mathbf{y}_i^\star; \mathcal{C}_i, \boldsymbol{\zeta}\right)$ of multivariate normal distribution. To obtain the density of actually observed outcomes $g(\mathbf{y}_i; \mathcal{C}_i, \boldsymbol{\theta})$ we need to separate $\mathbf{y}_i^\star$ into numeric (N) and ordinal (O) parts (including binary):

$$g\left(\mathbf{y}_i; \mathcal{C}_i, \boldsymbol{\zeta}, \boldsymbol{\gamma}\right) = \int t\left(\mathbf{y}_i^{\mathsf{O}} \,\middle|\, \mathbf{y}_i^{\mathsf{O},\star}; \boldsymbol{\gamma}\right) g^\star\left(\mathbf{y}_i^\star; \mathcal{C}_i, \boldsymbol{\zeta}\right) \mathrm{d}\mathbf{y}_i^{\mathsf{O},\star}, \tag{3}$$

where $t(\cdot|\cdot)$ represents the thresholding procedure.

3 Model based clustering

We first assume that K (the number of groups into which we intend to classify the units) is known in advance and $K \geq 2$. The classification proceeds by using the model outlined in Section 2 within the Bayesian model based clustering procedure (MBC, Fraley and Raftery [2]). Hence, it is assumed that the overall model, f, is given as a finite mixture of certain group-specific models f_k, $k = 1, \ldots, K$. That is, $f(\mathbf{y}_i; \mathcal{C}_i, \boldsymbol{\theta}) = \sum_{k=1}^K w_k f_k(\mathbf{y}_i; \mathcal{C}_i, \boldsymbol{\psi}, \boldsymbol{\psi}^k)$, where $\mathbf{w} = \left(w_1, \ldots, w_K\right)^\top$ are the mixture weights (proportions of the groups in the population), $\boldsymbol{\psi}$ is a vector of unknown parameters common to all groups and $\boldsymbol{\psi}^k$, $k = 1, \ldots, K$, are vectors of group-specific unknown parameters. Hence the vector $\boldsymbol{\theta}$ of all unknown parameters is $\boldsymbol{\theta} \equiv \left\{\mathbf{w}, \boldsymbol{\psi}, \boldsymbol{\psi}^1, \ldots, \boldsymbol{\psi}^K\right\}$.

Using the notation from previous section we set the group-specific density f_k to be the density g, however, depending on parameter $\boldsymbol{\zeta}^k$ elements of which ($\boldsymbol{\beta}_r^k$, τ_r^k, $\boldsymbol{\mu}^k$, $\boldsymbol{\Sigma}^k$) may (or may not) be group-specific, i.e. different value of the parameter is considered to be in different groups. For example, if we suppose that the groups differ only in the covariate effects, then

$$f(\mathbf{y}_i; \mathcal{C}_i, \boldsymbol{\theta}) = \sum_{k=1}^K w_k g\Big(\mathbf{y}_i; \mathcal{C}_i, \underbrace{\boldsymbol{\tau}, \boldsymbol{\Sigma}, \boldsymbol{\gamma}}_{\boldsymbol{\psi}}, \underbrace{\boldsymbol{\beta}^k, \boldsymbol{\mu}^k}_{\boldsymbol{\psi}^k}\Big).$$

Further, let $U_i \in \{1, \ldots, K\}$ denote the unobserved allocation of the ith unit into one of the K groups. As it is usual with the mixture models, the group-specific distribution $f_k(\mathbf{y}_i; \mathcal{C}_i, \psi, \psi^k)$, $k = 1, \ldots, K$, can be viewed as a conditional distribution of the outcome $\mathbf{Y}_i$ given $U_i = k$ while the mixture weights $\mathbf{w}$ determine the distribution of the allocations, i.e., $\mathsf{P}(U_i = k) = w_k$, $k = 1, \ldots, K$. Classification of the ith unit can then be based on suitable estimates of the conditional individual allocation probabilities $p_{i,k}(\boldsymbol{\theta})$, $k = 1, \ldots, K$, calculated by the Bayes rule:

$$p_{i,k}(\boldsymbol{\theta}) \;=\; \mathsf{P}\big(U_i = k \,\big|\, \mathbf{Y}_i = \mathbf{y}_i; \mathcal{C}_i, \boldsymbol{\theta}\big) \;=\; \frac{w_k\, g\left(\mathbf{y}_i; \mathcal{C}_i, \psi, \psi^k\right)}{f(\mathbf{y}_i; \mathcal{C}_i, \boldsymbol{\theta})}. \tag{4}$$

Calculation of such probabilities requires performing the integration (3), which is in fact the integration of multivariate normal density over an ($n_i \times$# ordinal outcomes)-dimensional interval, bounds of which are determined by the measured levels of ordinal outcomes $\mathbf{y}_i^{\mathsf{O}}$ and threshold parameter $\boldsymbol{\gamma}$. A method for computing such possibly highly dimensional integrals needs to be chosen carefully with respect to not only the precision but the computation time as well, since for one set of parameters $\boldsymbol{\theta}$ we need to use it at least $(n \times K)$-times. Moreover, we can limit ourselves to first j observations only, $j = 1, \ldots, n_i$, to capture the evolution of classification probability as the amount of available information increases.

To infer on the model parameters and to perform related classification a Bayesian approach was adopted and implemented in the `R` software in combination with the `C` language and routines from the `R` package `mvtnorm` to calculate integrals (3). Monte Carlo Markov chain (MCMC) methods were used to obtain a sample from posterior distribution of $\boldsymbol{\theta}$ and consequently also from the posterior distribution of each of classification probabilities $p_{i,k}(\boldsymbol{\theta})$. Not only their posterior means but also their credible intervals were used for classification to quantify uncertainty in allocation of the study units into the groups.

4 Application to EU-SILC data

The methodology was applied to Czech households from the EU-SILC data while considering jointly

- logarithm of lowest income to make ends meet (to pay for its usual necessary expenses),
- affordability of paying unexpected expenses (required expense faced without help of anybody – only own resources used, e.g. surgery, a funeral, major repairs in the house, replacement of durables like washing machine, car) – binary variable:
  - Yes (household can afford unexpected expenses),
  - No (household cannot afford unexpected expenses),
- ability to make ends meet (respondent's feeling with respect to his household's income) – ordinal variable with six levels:
  1. With great difficulty,
  2. With difficulty,
  3. With some difficulty,
  4. Fairly easily,
  5. Easily,
  6. Very easily.

Each of the $n = 20\,299$ Czech households had been followed for $n_i = 4$ years induced by rotational design which replaces households that would exceed the 4-year limit with newly entering households. The set of households included in the following analysis consists of households entering the study after 2005 and leaving the study before 2016.

This year span covers also the economical crisis in 2009 that is suspected to influence the prosperity and social status of Czech households. In order to capture the possible change in the evolution of outcomes of interest we were discouraged from using simple linear trend and, therefore, were forced to use more flexible parametrization of time. Numeric (and the latent numeric counterparts of categorical outcomes) were modelled using B-spline parametrization of order 3 with knots at years 2008 and 2011. This piecewise-cubic parametrization leads to 6 coefficients (including the intercept term) and forms the crucial part of the fixed effect part of the model. Furthermore, it is extended by the weighted family size[1] that potentially could affect the outcomes of interest. Thus, the structure of fixed effects is in the form

$$\mathbf{X}_{i,j,r}^{\top}\boldsymbol{\beta}_r = \beta_{0,r} + \beta_{1,r}s_1(t_{i,j}) + \cdots + \beta_{5,r}s_5(t_{i,j}) + \beta_{6,r}w_{i,j},$$

[1] Each member of the household contributes to the family size by the following values: 1 for adult person in the role of the head of the family, 0.5 for other person older than 14 and 0.3 for person younger 14 years.

where t_{ij} is the time (in years) that has passed since 2005 at which the i-th household was interviewed for j-th time, $s_1(t), \ldots, s_5(t)$ then corresponds to above mentioned spline parametrization at time t and w_{ij} is the corresponding weighted family size at that time.

All three outcomes are linked through household-specific random intercept term $\mathbf{B}_{0,i}$ that follows trivariate zero-mean normal distribution with general covariance matrix $\boldsymbol{\Sigma}$. Combining fixed and random effects part we obtain the supposed mixed-effects model for each (latent) numeric outcome r:

$$Y^{\star}_{i,j,r} = B_{0,i,r} + \mathbf{X}^{\top}_{i,j,r}\boldsymbol{\beta}_r + \varepsilon_{i,j,r} = B_{0,i,r} + \beta_{0,r} + \beta_{1,r}s_1(t_{i,j}) + \cdots + \beta_{5,r}s_5(t_{i,j}) + \beta_{6,r}w_{i,j} + \varepsilon_{i,j,r}.$$

Moreover, each cluster $k = 1, \ldots, K$ is defined by its own set of fixed effects $\boldsymbol{\beta}^{(k)}$ for clustering purposes. Other parameters like matrix $\boldsymbol{\Sigma}$ or precisions of the error terms τ_r are considered to be the same among clusters and, therefore, do not help to differentiate them. Hence, the discovered clusters could be distinguished only by interpretation of differences between β coefficients and the shape of the spline curve they correspond to.

4.1 Results

Gibbs sampling procedure was applied to data on Czech households for several values of the total number of clusters K to determine which number will be the most suitable one. The choice of $K \in \{2, 3, 4, 5, 6\}$ can be supported by low values of deviance of the model defined as $D(\boldsymbol{\theta}; \mathbf{Y}_1, \ldots, \mathbf{Y}_n) = -2\sum_{i=1}^{n} \log g\left(\mathbf{Y}_i; \mathcal{C}_i, \boldsymbol{\theta}\right)$, which involves integration of all latent variables. Figure 1 presents estimates of the posterior distribution of deviance (viewed as a parametric function of $\boldsymbol{\theta}$). The deviance appears to decrease with higher value of K, with the exception of $K = 5$. Although, the choice of $K = 6$ seems to be the most beneficial, we should not blindly believe it. Let us explore the behaviour (and interpretation) of these clusters first.

Focusing on the numeric variable only we plot the estimated spline curves for each choice of K including $K = 1$ which corresponds to general evolution when no clustering was applied. Curves in Figure 2 are plotted for households of unit size (just one adult member) since the weighted family size should not be ignored as its effect may differ among clusters. In general it seems that the need for higher income has been increasing till 2009, after which this increasing trend has slowly stabilized or even begun to decrease.

As we begin to distinguish hidden clusters we always find a pair of clusters sharing the same shape of the evolution described above differing only in the level. It seems that sorting in more than two clusters in similar way is inefficient since curves of other clusters follow completely different shape. These clusters usually represent a very low percentage of households that behave extremely in some specific sense. For example violet cluster groups households with extreme growth of lowest income to make ends meet in years 2005–2008. However, the same cluster groups households with rapid decrease in the time span 2008–2010. Analogously, we could interpret even the blue and the brown cluster. This is not caused just by our chosen spline

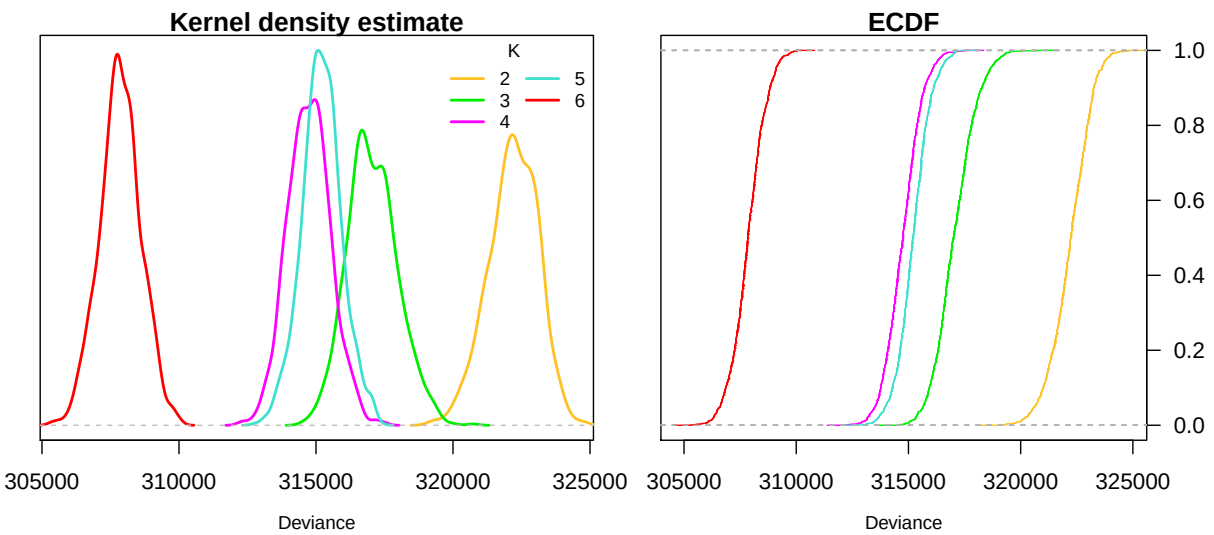


Figure 1 Estimated posterior distribution of deviances under several choices of number of clusters $K = 2, 3, 4, 5, 6$.

parametrization but mainly by the nature of the gathered data. We have to keep in mind, that households were questioned only four times in consecutive years. Therefore, for $K > 3$ we discover clusters of extreme behaviour which is usually limited to a certain time span. Moreover, we should not forget that discovered clusters differ in the evolution of categorical outcomes as well which may be even more extreme. Unless we are interested in these extremes, we should limit ourselves to lower count of clusters which still represent a considerable fractions of households and can characterize households in the whole time span. For this reason, we would recommend to use either $K = 2$ or $K = 3$.

Let us examine more the case of $K = 3$ which seems to be the most reasonable. Using Figures 2 (unit weighted family size) and 3 (weighted family size of 2) we can describe the found clusters in the following way:

2 This cluster (blue) represents about 7% of households that used to have very high living standard until the crisis in year 2009 came after which households in this cluster have the lowest lowest income to make ends meet. The need for higher income does not rise with enlarging the family size as fast as in other two clusters, since the estimated $\beta_{6,1}^{(2)} \doteq 0.36$ corresponding to weighted family size is the lowest among all $\beta_{6,1}^{(k)}$ parameters. This is supported by a gap between blue and other spline curves in Figure 3 compared to 2. On the other hand, about one third of households in this cluster could not afford to pay unexpected expenses in 2005–2010. After 2010 almost all households in this cluster were prepared to unfavourable circumstances. Similarly, after 2010 they were more able to make ends meet as the proportion of households easily making ends meet increased which relates to the low living standard.

1,3 Cluster 1 (red) shares the same evolution of the lowest income to make ends meet as the cluster 3 (green). In both clusters it seems to have increased until the crisis came and after which the actually needed amount of income stabilizes (maybe slightly decreases). Households in this cluster differ from the third one in the much higher increase of the needed income for an additional family member: $\beta_{6,1}^{(1)} \doteq 0.61 > 0.41 \doteq \beta_{6,1}^{(3)}$, which is supported by the switch of the red and the green spline curves in Figures 2 and 3. Clusters 1 and 3 can be further distinguished by the evolution of proportions of

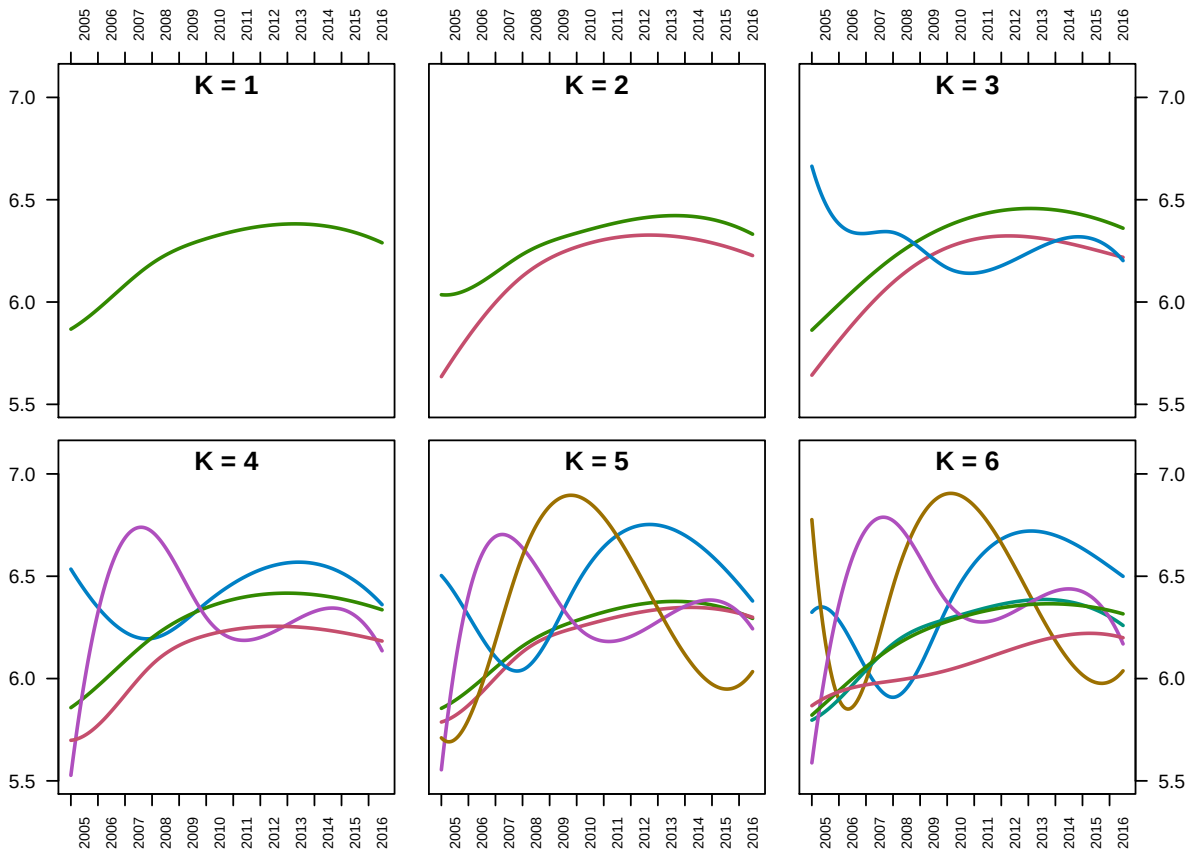


Figure 2 Spline curves for logarithm of the lowest income to make ends meet of households of unit weighted family size for different choice of the number of clusters K.

categorical variables. In 2005–2011, the cluster 1 represents households with increasing probability of being able to pay for unexpected expenses, whereas, this probability decreases in cluster 3. After 2011, clusters 1 and 3 switched the monotonicity in the evolution of this probability. Similarly, cluster 1 represents households with increasing difficulties to make ends meet until 2011, after which year these difficulties fade away. The cluster 3 reflects this behaviour in the completely opposite way as around 2011 it consists of households having no difficulties to make ends meet.

Households were classified by the rule based on highest posterior density intervals (HPD). If the lower bound for the maximal posterior probability of belonging to cluster is higher than upper bounds for all other probabilities, then the household is classified into the cluster that maximizes this probability. Otherwise, the household remains unclassified, which occurred in almost 23% of cases.

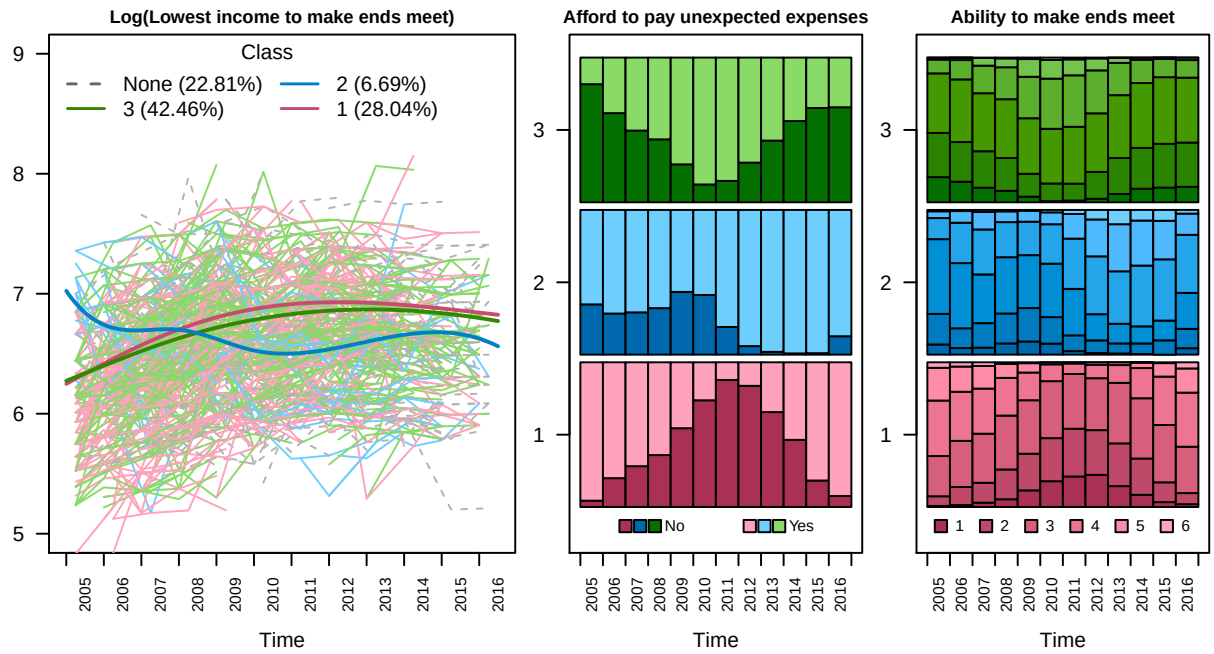


Figure 3 Longitudinal profiles of numeric, binary and ordinal outcomes of $n = 1\,000$ randomly selected Czech households. Bold curves on the left represent the estimated conditional expectation of response within $K = 3$ discovered groups for a household of weighted family size of 2. Categorical outcomes are presented by the proportions of categories in each year separately for the discovered groups. Some households remain unclassified.

5 Conclusions

We have developed a statistical model dealing with panel data of a mixed type. It was achieved by application of multivariate mixed effects model on numeric outcomes together with latent numeric outcomes which give rise to observed binary and ordinal outcomes. Mixture of such models was further used to discover different patterns in evolution of outcomes of interest. Using a fully Bayesian approach we were able to sort Czech households into three substantially different groups according to their ability to afford to pay for unexpected expenses, ability to make ends meet and the lowest needed income to do so.

Acknowledgements

This research was supported by the Czech Science Foundation (GAČR) grant 19-00015S.

References

[1] Dunson, D. B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B*, **62**, 355–366.

[2] Fraley, C. and Raftery, A. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, **97**, 611–631.

Recursive Estimation of IGARCH Model

Petr Vejmělka[1]

Abstract. For the financial time series modeling, the models with conditional heteroscedasticity GARCH are most commonly used. The estimation of GARCH models is already implemented in many software products. However, when working with high-frequency data such as stock market prices or index levels, classical estimation methods often fail, and it is necessary to use a recursive approach. In literature, recursive estimation formulas for several GARCH models have already been suggested. The aim of this contribution is to modify the GARCH model estimation algorithm for the model IGARCH (Integrated GARCH model), GARCH model with a unit root in the autoregressive polynomial in the volatility equation. This model proves to be useful in situations of high-frequency financial data, where it often happens that the sum of GARCH model parameters is close to one. This paper includes both mathematical derivation of a recursive algorithm for the IGARCH model and also an application of this algorithm.

Keywords: IGARCH, recursive estimation, financial time series, volatility

JEL Classification: C51, C58

1 Introduction

For the financial time series modeling, models with conditional heteroskedasticity GARCH are most commonly used. They are currently the strongest tool for the financial time series modeling and have not yet been overcome. A common method of estimating parameters in these models is a classical non-recursive estimate. However, this approach cannot be applied, for example, to high-frequency data such as stock market prices or index levels. The volume of such data is enormous and real-time parameter estimation would not be possible. For this reason, a recursive approach is more appropriate, not only in this case.

Several articles already dealt with a derivation of recursive algorithms for GARCH models, e.g. [1], [8] and [10]. These articles primarily focused on GARCH models. Later, other articles dealt with the modification of these algorithms for other models such as EWMA, see [9], their robustification or extension for multidimensional time series. The aim of this paper is to modify the algorithm for recursive estimation of the GARCH model parameters proposed in [8] for IGARCH, which is a special case of GARCH models. Its usefulness can be appreciated in situations in which we treat high-frequency financial data. In this case, the persistence in volatility occurs frequently and, therefore, the application of IGARCH model is reasonable.

The recursive estimates are obtained with the help of the algorithm based on minimizing the Taylor expansion of the loss function, in which both weights and the log-likelihood function for Gaussian IGARCH model are represented. Since it is necessary to calculate the first and second order derivatives, which are difficult to obtain, some approximations are made to simplify these calculations. The result is the recursive algorithm estimating the parameters of IGARCH model, which is able to provide the estimates online.

This paper contains the following sections. Section 2 presents IGARCH model, its essential features and the links to GARCH model. In Section 3, a brief mathematical derivation of the recursive algorithm used to estimate the parameters of IGARCH model is provided. Section 4 is devoted to simulations to verify the ability of the recursive algorithm to estimate the parameters numerically. Finally, Section 5 contains conclusions achieved in this paper.

2 IGARCH model

The integrated GARCH model of orders p and q, usually denoted as IGARCH(p,q), is a simple extension of GARCH(p,q) model. This modification was proposed in [6].

The IGARCH(p,q) process $\{y_t\}_{t\in\mathbb{Z}}$ is commonly defined as:

$$y_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{1}$$

[1] MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8, vejmelp@karlin.mff.cuni.cz

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of independent, identically distributed random variables with zero mean and unit variance.

Sufficient conditions for σ_t^2 being positive are $\alpha_0 > 0, \alpha_1, \ldots, \alpha_p \geq 0, \beta_1, \ldots, \beta_q \geq 0$. The only difference between GARCH and IGARCH model consits in condition $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$.

Unlike stationary GARCH model, that assumes $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ so that the volatility forecasts converge to the unconditional variance as the prediction horizon increases, the unconditional variance does not exist in IGARCH model. Due to the condition $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$, the impact of current information persists in the conditional volatility forecasts for all horizons.

For example, assuming the IGARCH(1,1) model, it is obvious that after applying the condition $\alpha_1 + \beta_1 = 1$, only one of these parameters remains in the model. Renaming the remaining parameter to β, we get

$$\sigma_t^2 = \alpha_0 + (1-\beta)y_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{2}$$

where $\alpha_0 > 0$ and $0 \leq \beta \leq 1$. The special case of IGARCH(1,1) model, when setting explicitly $\alpha_0 = 0$ in (2), is called EWMA *(Exponentially Weighted Moving Average)*. Nevertheless, in this case it is necessary to replace the restrictions for β by $\beta \in (0,1)$, to ensure the positiveness of the conditional variance.

The *l*-step-ahead volatility forecast in IGARCH(1,1) model is expressed as

$$\hat{\sigma}_{t+l|t}^2 = \hat{\sigma}_{t+1|t}^2 + (l-1)\alpha_0, \tag{3}$$

where

$$\hat{\sigma}_{t+1|t}^2 = \sigma_{t+1}^2 = \alpha_0 + (1-\beta)y_t^2 + \beta\sigma_t^2 \tag{4}$$

is the one-step-ahead volatility forecast.

3 Recursive estimation of IGARCH parameters

In this part, we deal with the recursive algorithm for estimating the parameters of IGARCH(*p*,*q*) model. First of all, the equation for the conditional variance in (1) must be rewritten in the form:

$$\sigma_t^2 = y_{t-p}^2 + \alpha_0 + \alpha_1(y_{t-1}^2 - y_{t-p}^2) + \cdots + \alpha_{p-1}(y_{t-p+1}^2 - y_{t-p}^2) + \beta_1(\sigma_{t-1}^2 - y_{t-p}^2) + \cdots + \beta_q(\sigma_{t-q}^2 - y_{t-p}^2), \tag{5}$$

to meet the condition $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$. In this way, we removed the parameter α_p from the equation.

For the purpose of the recursive algorithm it is suitable to rewrite this equation to the following form:

$$\sigma_t^2(\boldsymbol{\theta}) = y_{t-p}^2 + \boldsymbol{\varphi}_t^T(\boldsymbol{\theta})\boldsymbol{\theta}, \tag{6}$$

where

$$\boldsymbol{\theta} = (\alpha_0, \alpha_1, \ldots, \alpha_{p-1}, \beta_1, \ldots, \beta_q)^T, \tag{7}$$

$$\boldsymbol{\varphi}_t(\boldsymbol{\theta}) = (1, y_{t-1}^2 - y_{t-p}^2, \ldots, y_{t-p+1}^2 - y_{t-p}^2, \boldsymbol{\varphi}_{t-1}^T(\boldsymbol{\theta})\boldsymbol{\theta} + y_{t-1-p}^2 - y_{t-p}^2, \ldots, \boldsymbol{\varphi}_{t-q}^T(\boldsymbol{\theta})\boldsymbol{\theta} + y_{t-q-p}^2 - y_{t-p}^2)^T. \tag{8}$$

It can be verified that this construction really corresponds to (6).

The derivation of the recursive algorithm is based on the method suggested in [8] for the GARCH(*p*,*q*) model recursive estimation improving the previously proposed methods, see [1] and [10].

It uses a self-weighted approach based on minimalization of the loss function

$$V_t(\boldsymbol{\theta}) = \gamma_t \sum_{k=1}^{t} \left[\prod_{l=k+1}^{t} \lambda_l \right] F_k(\boldsymbol{\theta}), \qquad t \in \mathbb{N}, \tag{9}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{\Theta}$ is a parametric space, $\{\lambda_l\}$ is a sequence of weights that are positive and $\{\gamma_t\}$ is a sequence whose elements fulfill

$$\gamma_t = \frac{1}{\sum_{k=1}^{t} \left[\prod_{l=k+1}^{t} \lambda_l\right]} \tag{10}$$

and further $\sum_{t=1}^{\infty}\gamma_t=\infty$ and $\sum_{t=1}^{\infty}\gamma_t^2<\infty$. Due to the fact that in the equation (10) we have for $k=t$ an element $\prod_{l=t+1}^{t}\lambda_l$, we will accept the convention $\prod_{l=t+1}^{t}\lambda_l=1$.

The general function $F_k(\boldsymbol{\theta})$ in (9) will be considered as follows:

$$F_k(\boldsymbol{\theta})=\frac{y_k^2}{y_{k-p}^2+\boldsymbol{\varphi}_k^T(\boldsymbol{\theta})\boldsymbol{\theta}}+\log[y_{k-p}^2+\boldsymbol{\varphi}_k^T(\boldsymbol{\theta})\boldsymbol{\theta}]. \tag{11}$$

This structure is motivated by the log-likelihood function for Gaussian IGARCH(p,q) model.

To get an estimate $\widehat{\boldsymbol{\theta}}$ one must minimize the loss function $V_t(\boldsymbol{\theta})$. However, we will not minimize directly the function $V_t(\boldsymbol{\theta})$, but its Taylor expansion around $\widehat{\boldsymbol{\theta}}_{t-1}$. This approximation of $V_t(\boldsymbol{\theta})$ will have the following form:

$$V_t(\boldsymbol{\theta})\approx V_t(\widehat{\boldsymbol{\theta}}_{t-1})+[\frac{\partial}{\partial\boldsymbol{\theta}}V_t(\widehat{\boldsymbol{\theta}}_{t-1})](\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_{t-1})+\frac{1}{2}(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_{t-1})^T[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}V_t(\widehat{\boldsymbol{\theta}}_{t-1})](\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_{t-1}). \tag{12}$$

The right-hand side of the approximation (12) attains its minimum value for $\boldsymbol{\theta}$, which can be denoted as $\widehat{\boldsymbol{\theta}}_t$, since it is an estimate made at time t, given by the equation:

$$\widehat{\boldsymbol{\theta}}_t=\widehat{\boldsymbol{\theta}}_{t-1}-\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}V_t(\widehat{\boldsymbol{\theta}}_{t-1})\right]^{-1}\left[\frac{\partial}{\partial\boldsymbol{\theta}}V_t(\widehat{\boldsymbol{\theta}}_{t-1})\right]. \tag{13}$$

The application of (13) assumes that the gradient $\frac{\partial}{\partial\boldsymbol{\theta}}V_t(\widehat{\boldsymbol{\theta}}_{t-1})$ and the Hessian matrix $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}V_t(\widehat{\boldsymbol{\theta}}_{t-1})$ must be calculated. These values depend on $\frac{\partial}{\partial\boldsymbol{\theta}}F_t(\boldsymbol{\theta})$ and $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}F_t(\boldsymbol{\theta})$. Their expressions assume calculation of $\frac{\partial}{\partial\boldsymbol{\theta}}(\boldsymbol{\varphi}_t^T(\boldsymbol{\theta})\boldsymbol{\theta})$ and $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}(\boldsymbol{\varphi}_t^T(\boldsymbol{\theta})\boldsymbol{\theta})$. To simplify the notation, let us denote $\boldsymbol{\psi}_t(\boldsymbol{\theta})=\frac{\partial}{\partial\boldsymbol{\theta}}(\boldsymbol{\varphi}_t^T(\boldsymbol{\theta})\boldsymbol{\theta})$.

Because of the difficulty of evaluation $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}V_t(\boldsymbol{\theta})$, this element is approximated by the matrix $\mathbf{R}_t(\boldsymbol{\theta})$ fulfilling

$$\mathrm{E}\left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}V_t(\boldsymbol{\theta})-\mathbf{R}_t(\boldsymbol{\theta})\,\middle|\,\boldsymbol{\Omega}_{t-1}\right]=0,\qquad\forall t\in\mathbb{N}. \tag{14}$$

The matrix $\mathbf{R}_t(\boldsymbol{\theta})$ is chosen as follows:

$$\mathbf{R}_t(\boldsymbol{\theta})=(1-\gamma_t)\mathbf{R}_{t-1}(\boldsymbol{\theta})+\gamma_t\tilde{\mathbf{F}}_t(\boldsymbol{\theta}), \tag{15}$$

where

$$\tilde{\mathbf{F}}_t(\boldsymbol{\theta})=\frac{\boldsymbol{\psi}_t(\boldsymbol{\theta})\boldsymbol{\psi}_t^T(\boldsymbol{\theta})}{(y_{t-p}^2+\boldsymbol{\varphi}_t^T(\boldsymbol{\theta})\boldsymbol{\theta})^2}. \tag{16}$$

Due to (13), the approximation of $\frac{\partial}{\partial\boldsymbol{\theta}}V_t(\widehat{\boldsymbol{\theta}}_{t-1})$, (15), (16) and employing the approximations $\boldsymbol{\varphi}_t(\widehat{\boldsymbol{\theta}}_{t-1})\approx\widehat{\boldsymbol{\varphi}}_t$ and $\boldsymbol{\psi}_t(\widehat{\boldsymbol{\theta}}_{t-1})\approx\widehat{\boldsymbol{\psi}}_t$, we get the following recursive algorithm:

$$\widehat{\boldsymbol{\theta}}_t=\widehat{\boldsymbol{\theta}}_{t-1}+\gamma_t\mathbf{R}_t^{-1}\widehat{\boldsymbol{\psi}}_t\frac{y_t^2-y_{t-p}^2-\widehat{\boldsymbol{\varphi}}_t^T\widehat{\boldsymbol{\theta}}_{t-1}}{(y_{t-p}^2+\widehat{\boldsymbol{\varphi}}_t^T\widehat{\boldsymbol{\theta}}_{t-1})^2}, \tag{17}$$

$$\mathbf{R}_t=(1-\gamma_t)\mathbf{R}_{t-1}+\gamma_t\frac{\widehat{\boldsymbol{\psi}}_t\widehat{\boldsymbol{\psi}}_t^T}{(y_{t-p}^2+\widehat{\boldsymbol{\varphi}}_t^T\widehat{\boldsymbol{\theta}}_{t-1})^2}, \tag{18}$$

$$\begin{aligned}\widehat{\boldsymbol{\varphi}}_{t+1}=(&1,y_t^2-y_{t-p+1}^2,\ldots,y_{t+2-p}^2-y_{t-p+1}^2,\\&\widehat{\boldsymbol{\varphi}}_t^T\widehat{\boldsymbol{\theta}}_t+y_{t-p}^2-y_{t-p+1}^2,\ldots,\widehat{\boldsymbol{\varphi}}_{t+1-q}^T\widehat{\boldsymbol{\theta}}_{t+1-q}+y_{t+1-q-p}^2-y_{t-p+1}^2)^T,\end{aligned} \tag{19}$$

$$\widehat{\boldsymbol{\psi}}_{t+1}=\widehat{\boldsymbol{\varphi}}_{t+1}+\sum_{j=1}^{q}\widehat{\beta}_{j,t}\widehat{\boldsymbol{\psi}}_{t+1-j}, \tag{20}$$

for $t\in\mathbb{N}$.

This algorithm can be adjusted to a more computationally advantageous form by application the matrix inversion lemma, see [12]. The main purpose of this adjustment is to avoid the calculation of the inverse

matrix $\mathbf{R}_t^{-1}$ in each iteration. Therefore, denoting $\widehat{\mathbf{P}}_t = \gamma_t \mathbf{R}_t^{-1}$, we get the final form of the recursive algorithm

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \frac{\widehat{\mathbf{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t(y_t^2 - y_{t-p}^2 - \widehat{\boldsymbol{\varphi}}_t^T\widehat{\boldsymbol{\theta}}_{t-1})}{\widehat{\boldsymbol{\psi}}_t^T\widehat{\mathbf{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t + \lambda_t(y_{t-p}^2 + \widehat{\boldsymbol{\varphi}}_t^T\widehat{\boldsymbol{\theta}}_{t-1})^2}, \tag{21}$$

$$\widehat{\mathbf{P}}_t = \frac{1}{\lambda_t}\left[\widehat{\mathbf{P}}_{t-1} - \frac{\widehat{\mathbf{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t\widehat{\boldsymbol{\psi}}_t^T\widehat{\mathbf{P}}_{t-1}}{\widehat{\boldsymbol{\psi}}_t^T\widehat{\mathbf{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t + \lambda_t(y_{t-p}^2 + \widehat{\boldsymbol{\varphi}}_t^T\widehat{\boldsymbol{\theta}}_{t-1})^2}\right], \tag{22}$$

$$\begin{aligned}\widehat{\boldsymbol{\varphi}}_{t+1} = (&1, y_t^2 - y_{t-p+1}^2, \ldots, y_{t+2-p}^2 - y_{t-p+1}^2,\\ &\widehat{\boldsymbol{\varphi}}_t^T\widehat{\boldsymbol{\theta}}_t + y_{t-p}^2 - y_{t-p+1}^2, \ldots, \widehat{\boldsymbol{\varphi}}_{t+1-q}^T\widehat{\boldsymbol{\theta}}_{t+1-q} + y_{t+1-q-p}^2 - y_{t-p+1}^2)^T,\end{aligned} \tag{23}$$

$$\widehat{\boldsymbol{\psi}}_{t+1} = \widehat{\boldsymbol{\varphi}}_{t+1} + \sum_{j=1}^{q}\widehat{\beta}_{j,t}\widehat{\boldsymbol{\psi}}_{t+1-j} \tag{24}$$

for $t \in \mathbb{N}$.

Furthermore, this algorithm must be supplemented by the correct choice of weights $\{\lambda_t\}$. One of possible options, suggested in [12], is the recursive calculation of weights according to the formula

$$\lambda_t = \tilde{\lambda}\lambda_{t-1} + (1 - \tilde{\lambda}),\ t \in \mathbb{N}, \tag{25}$$

where recommended choices of constants $\tilde{\lambda} \in (0, 1)$ and λ_0 are $\tilde{\lambda} = 0.99$ and $\lambda_0 = 0.95$.

An important part of the recursive algorithm consits in setting the initial estimates. They can be set in various ways. One possibility is $\widehat{\boldsymbol{\theta}}_0 = (\frac{1}{n}\sum_{i=1}^{n} y_{1-i}^2[1 - (p - 1 + q)\varepsilon], \varepsilon, \ldots, \varepsilon, \varepsilon, \ldots, \varepsilon)^T$, where ε is a small positive number satisfying $(p - 1 + q)\varepsilon < 1$ for suitable n; $\widehat{\mathbf{P}}_0 = c\mathbf{I}$, where c is a large constant, e.g. $c = 10^5$ to minimize the influence of the initial estimate and to approach quickly the actual vector of parameters; $\widehat{\boldsymbol{\varphi}}_1 = (1, y_0^2 - y_{-p+1}^2, \ldots, y_{-p+2}^2 - y_{-p+1}^2, k, \ldots, k)^T$, where k is equal to a small positive constant; $\widehat{\boldsymbol{\psi}}_1 = \widehat{\boldsymbol{\varphi}}_1$ and $\widehat{\boldsymbol{\psi}}_i = 0, i = -q + 2, \ldots, 0$ for $i \leq 0$.

Finally, sufficient conditions for IGARCH model are established to ensure the positiveness of the conditional variance. They can be implemented by means of the following rule in the corresponding algorithm:

$$\widehat{\boldsymbol{\theta}}_t = \begin{cases}\widehat{\boldsymbol{\theta}}_t, & \text{if } \widehat{\boldsymbol{\theta}}_t \in \mathbf{D_S},\\ \widehat{\boldsymbol{\theta}}_{t-1}, & \text{if } \widehat{\boldsymbol{\theta}}_t \notin \mathbf{D_S},\end{cases} \tag{26}$$

where $\mathbf{D_S} = \{\boldsymbol{\theta} \in \mathbb{R}^{p+q} | 0 < \alpha_0, 0 \leq \alpha_1 \leq 1, \ldots, 0 \leq \alpha_{p-1} \leq 1, 0 \leq \beta_1 \leq 1, \ldots, 0 \leq \beta_q \leq 1\}$ is the set of vectors $\boldsymbol{\theta}$, whose components meet the above conditions and thus ensure the positiveness of the conditional variance $\sigma^2(\boldsymbol{\theta})$. In the case of getting outside of the set $\mathbf{D_S}$ in the next step of the algorithm, the new estimate is ignored and the previous one is used instead.

Another possible modification may be the robustification of the algorithm. The robustification of the recursive estimate of GARCH model was proposed in [4]. By its modifying, one can obtain a robustified recursive algorithm for IGARCH model.

4 Simulations

In this Section Monte Carlo simulations are realized to investigate whether the proposed algorithms are capable to estimate the parameters of IGARCH model focusing on the development of the estimates over time. Two IGARCH models are considered, namely IGARCH(1,1) and IGARCH(3,1). For both cases 1,000 time series of length 20,060 are simulated and then put into the algorithm. Time series of length 20,060 are simulated because of the fact that the first 60 observations are used for setting the initial estimates and the remaining 20,000 observations for the subsequent on-line estimation.

The obtained estimates are presented using boxplots, from which we removed the outliers for simplicity. The development of estimates over time is represented by displaying boxplots for estimates in times $T = 2{,}500$, $T = 5{,}000$, $T = 10{,}000$ and $T = 20{,}000$. Each box shows the range between the first and third quartile of obtained estimates, the white bar represents the median and the black line indicates the true value of the corresponding parameter.

Figure 1 shows boxplots of the IGARCH(1,1) parameters estimates with true values of parameters $\alpha_0 = 0.3$ and $\beta_1 = 0.8$. For IGARCH(1,1), one must estimate only two parameters, namely α_0 and β_1. If we use the

equation (1) for the computation of the conditional variance, the value of parameter α_1 can be calculated as $\alpha_1 = 1 - \beta_1$. The algorithm is accompanied by controlling, whether $\widehat{\boldsymbol{\theta}}_t \in \mathbf{D_S}$, where $\mathbf{D_S} = \{\boldsymbol{\theta} \in \mathbb{R}^2 | \alpha_0 > 0, 0 \leq \beta_1 < 1\}$. The initial estimates were set as follows: $\widehat{\boldsymbol{\theta}}_0 = (\frac{1}{60}\sum_{i=1}^{60} 0.9y_{1-i}^2, 0.1)^T$, $\widehat{\mathbf{P}}_0 = 10^5\mathbf{I}$, $\widehat{\boldsymbol{\varphi}}_1 = (1, 0.1)^T$ and $\widehat{\boldsymbol{\psi}}_1 = \widehat{\boldsymbol{\varphi}}_1$. For the weights, we put $\tilde{\lambda} = 0.99$ and $\lambda_0 = 0.95$.

Figure 2 shows boxplots of the IGARCH(3,1) parameters estimates with true values of parameters $\alpha_0 = 0.1$, $\alpha_1 = 0.2$, $\alpha_2 = 0.2$ and $\beta_1 = 0.5$. In this case a similar controlling process as for the previous model was used. The initial estimates were set alike: $\widehat{\boldsymbol{\theta}}_0 = (\frac{1}{60}\sum_{i=1}^{60} 0.7y_{1-i}^2, 0.1, 0.1, 0.1)^T$, $\widehat{\mathbf{P}}_0 = 10^5\mathbf{I}$, $\widehat{\boldsymbol{\varphi}}_1 = (1, y_0^2 - y_{-2}^2, y_{-1}^2 - y_{-2}^2, 0.1)^T$ and $\widehat{\boldsymbol{\psi}}_1 = \widehat{\boldsymbol{\varphi}}_1$. The values for weights are again $\tilde{\lambda} = 0.99$ and $\lambda_0 = 0.95$.

One can see from Figures 1 and 2 that as the number of available observations increases, the estimates gradually improve and the range of boxplots decreases. The differences between $T = 2{,}500$ and $T = 20{,}000$ are significant. In the considered cases, the length of the time series of 5,000 was already sufficient. Nevertheless, it should be noted that with the increasing orders of the model p and q the accuracy of the algorithm decreases and satisfactory results are achieved only for much longer time series.

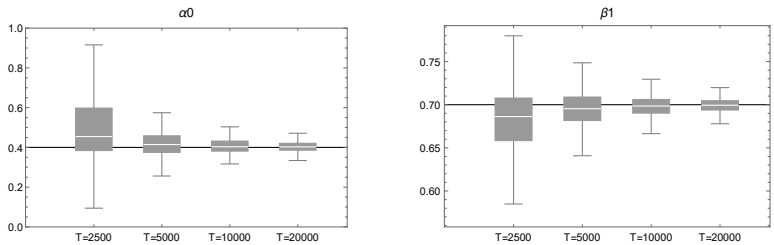


Figure 1 Boxplots of the IGARCH(1,1) estimates for $(\alpha_0, \beta_1) = (0.4, 0.7)$

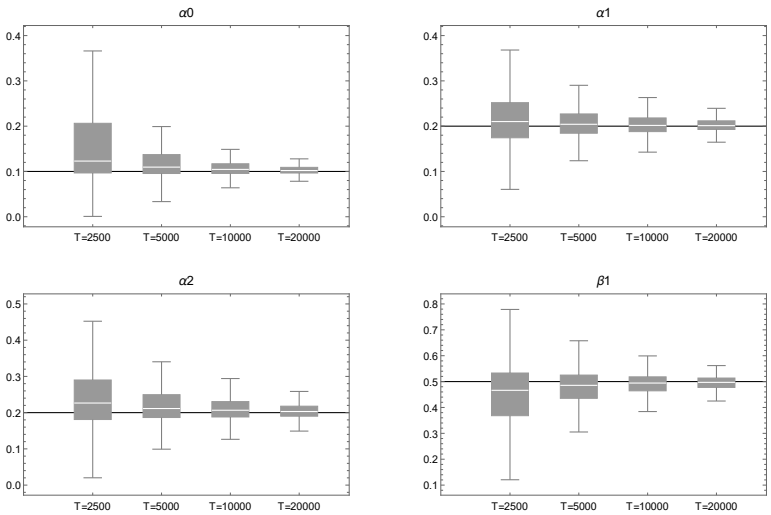


Figure 2 Boxplots of the IGARCH(3,1) estimates for $(\alpha_0, \alpha_1, \alpha_2, \beta_1) = (0.1, 0.2, 0.2, 0.5)$

5 Conclusion

This paper deals with the online modeling of time series, particularly the estimation of the IGARCH model parameters. We derived the form of the recursive algorithm with a possible choice of initial conditions that can be used, e.g. for modeling high-frequency data, where a classical non-recursive estimation would be insufficient. Subsequently, this algorithm was tested in the simulation study verifying whether this algorithm is capable to correctly estimate the unknown parameters of the model. The algorithm was demonstrated by the fundamental IGARCH(1,1) model and then for a more complex structure of IGARCH(3,1). In general, the recursive algorithms for estimating GARCH models are effective for low orders of p and q, and this fact has been confirmed for the IGARCH model during the simulations.

Acknowledgements

This paper was supported by the grant 19-28231X provided by the Czech Science Foundation.

References

[1] Aknouche, A. & Guerbyenne, H. (2006). Recursive estimation of GARCH models. *Communications in Statistics - Simulation and Computation* **35**, 925–938.

[2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

[3] Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods* (Second Edition). Springer-Verlag, New York, ISBN 978-0-387-97429-3.

[4] Cipra, T. &, Hendrych, R. (2018). Robust recursive estimation of GARCH models. *Kybernetika* **54**, 1138–1155.

[5] Cipra, T. &, Hendrych, R. (2019). Modelling of currency covolatilities. *Statistika* **99**, 259–271.

[6] Engle, R. F. & Bollerslev, T. (1986). Modelling the persistence of conditional variance. *Econometric Reviews* **5**, 1-50.

[7] Hamilton J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, ISBN 0-691-04289-6.

[8] Hendrych, R. & Cipra, T. (2015). Self-weighted recursive estimation of GARCH models. *Communications in Statistics - Simulation and Computation* **47**, 315–328.

[9] Hendrych, R. & Cipra, T. (2015). Robustified on-line estimation of the EWMA models: Simulations and applications. *Proceedings of the 33rd International Conference Mathematical Methods in Economics*, Plzeň, 237–242.

[10] Kierkegaard, J., Nielsen, J., Jensen, L. & Madsen, H. (2000). *Estimating GARCH models using recursive methods*. Working Paper, Technical University of Denmark, Lyndgby.

[11] Ljung L. (2002). *System Identification: Theory for the User* (Second Edition). Prentice Hall, Upper Saddle River, N.J., ISBN 978-0-136-56695-3.

[12] Ljung, L. & Söderström, T. (1983). *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, MA., ISBN 978-0-262-62058-1.

Micro Moments Marketing Model

Lukáš Veverka[1], Vladimír Holý[2]

Abstract. What if you would know exactly when and who needs your product. What if you would be able to stimulate these moments. In this analysis, we will examine whether TV can drive the micro-moments of interest. In the time of multi-screening, it is common to look for information immediately even during watching the TV. Therefore we focus on the organic webpage visits with a minute frequency obtained from Google Analytics and compare them with the time of advertisement. Firstly, it is necessary to remove the daily pattern of visits which differs in each day and then to separate the geometrically decreasing uplift of organic visits in the moment of the TV ad. After that, the effect of TV is obtained and it is possible to consider factors affecting its size. The results confirm the ability of TV to drive an immediate interest and show a clear dependency on the GRPs (TV impressions).

Keywords: data-driven marketing, micro-moments, kernel smoothing, dynamic models

JEL Classification: C510
AMS Classification: 62P20

1 Introduction

Communication of the right message to the right audience, at the right time, in context – that sounds like a perfect plan on how to sell anything. However, it is no longer just a fiction but an example of the moment marketing. Its spontaneity is not that spontaneous as it might look like. The trick is based on a prepared offline experience that drives online interaction.

People's intention to look for some information reflects their need. Google calls that moment a micro-moment. The whole theory and strategy around it expands on the moment marketing. TV can either raise the interest or remind a past need. In that micro-moment people turn to their phones to look up for more information. Since simultaneous television and internet consumption has rapidly increased [1] it is a good question of how these two media types cooperate. To answer these questions, two data sources were used – Google Analytics which provides data about organic webpage visits with a minute frequency and television advertising database. Then a procedure able to determine the relationship was developed and tested on an online retailer that communicates through TV.

The main motivation for this work is to separate the effect of each TV campaign for further research. The findings of this work can be utilized during the strategy creation for specific clients. For example, certain PPC keywords' highest cost might be increased for the moments just after the TV ad. Another example might be customization of the landing page based on which product was communicated or the whole TV buying process might be optimized.

2 Literature review

The effect of advertising has been studied already in a lot of research papers. However, the more data we have with the advent of the internet, the more precise and provable models can be created. This research is mostly inspired by the paper [2], where there are estimated the effects of TV ads on customers' online behaviour. Very similar work is [3] which uses Google search queries instead of webpage visits. Another example in this area is [5], which describes the dependency between TV ads and telephone calls. However, it works only with hourly frequency data which cannot cover and describe all the details of the people's behaviour.

The above-mentioned research [2] decomposes the time series with fixed effects. In detail, there is a vector containing 139 time fixed effects (actually it is a matrix since there are more product categories but it is not

[1] University of Economics, Prague, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, vevl00@vse.cz
[2] University of Economics, Prague, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, vladimir.holy@vse.cz

important for this case). There are FE for: weeks in a year (51 FE), day of the week (6), hour of the day (23), and minute of the hour (59). There is however no part for interaction among these variables. An example where this model reaches its limits might be this: constantly increasing webpage visits in the evening. The FE model will estimate the average effect of the hour but minutes are estimated through all hours which means they cannot describe the constantly increasing webpage visits in the certain hour. The best solution for this would be an interaction of these variables, though this would consume a lot of degrees of freedom. In addition, to cover all these interactions it would be necessary to have a record from the same day type (e.g. Monday), hour and minute. This is not possible to meet in most cases of advertisement. Moreover, the method would be possible to use only after a long time period. Therefore we propose this new attitude which allows us to estimate the effect with a smaller sample of data about advertisements and shorter periods.

3 Methodology

3.1 One-Dimensional Kernel Smoother

The course of webpage visits has a certain pattern through the day and it differs for each day type (most significant differences are between weekend and working days – see Figure 1). Therefore it is necessary to estimate the pattern and decompose the time series.

First of all, it is necessary to get all the levels of webpage visits on the same scale since their variance differ (some days before Christmas there was an extraordinary traffic). In case of not doing that the kernel smoother would be biased. It is assumed that each day can have a different standard deviation but each of the seven day types has the same diurnal pattern. Therefore, the values were standardized by each day with only a standard deviation (the mean 0 is not necessary) $y_i = \frac{z_i}{\sigma}$, where z_i are the original values of webpage visits and σ stands for the standard deviation of the certain day which is saved for future reverse standardization.

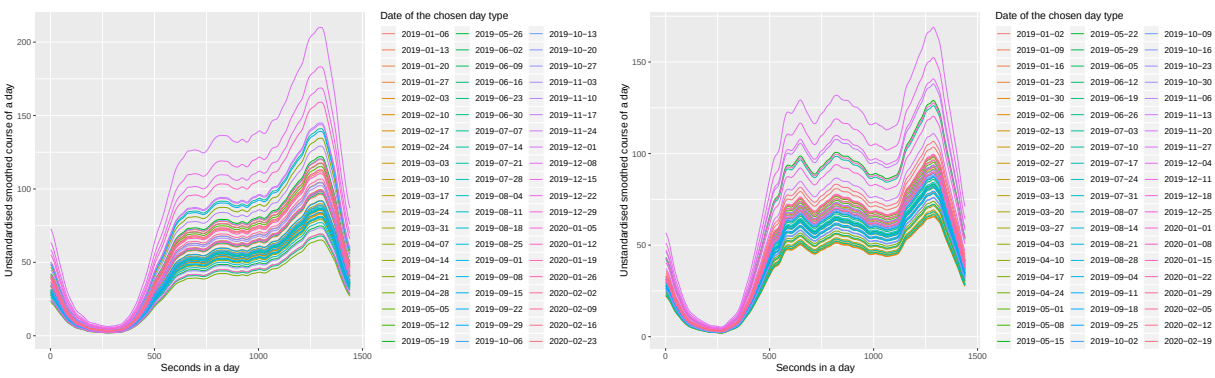
After this, the Nadaraya–Watson kernel-weighted estimation [6] is used to get the diurnal pattern for each day type. It is a locally weighted average, using a kernel as the weighting function. The equation for the approximation of the daily pattern is

$$\hat{f}(t_0) = \frac{\sum_{i=1}^{N} K_\lambda(t_0, t_i) y_i}{\sum_{i=1}^{N} K_\lambda(t_0, t_i)}, \tag{1}$$

where t stands for time (1 to 1440 corresponding to the minutes), y represents the standardized webpage visits. K_λ is a density function with parameter λ which represents the metric window [7]. The summation length N may vary with the different number of observation inside the metric window – which gives us the hint that index i represents the time within the metric window whereas the index 0 stands for the time for which the approximation is calculated. In this case, we utilize the Gaussian kernel given by

$$K_\lambda(t_0, t_i) = \exp\left(-\frac{(t_0 - t_i)^2}{2\lambda^2}\right). \tag{2}$$

After this procedure, the reverse standardization is applied for future easier interpretation. The resulting smoothed webpage visits might be seen in the already mentioned Figure 1. They are then subtracted from



(a) Sunday – weekend

(b) Wednesday – working day

Figure 1 Comparison of the smoothed courses by different day types

the original values which gives us only the deviation from the daily pattern. Explanation of these deviations with the TV advertisement is the main intention.

3.2 Dynamic models

Since the advertising has a lagged effect [4] it is necessary to calculate with that. In this research, it is assumed that the lagged effect is geometrically decreasing [8]. Since this research is intended to work with a lot of data it is necessary to create a window to lower the computational demands. Instead of computing the infinite distributed lag model for each campaign on the whole dataset (e.g. 500 thousand rows representing approximately one year), it is taken only a few minutes before and after the ad. It is assumed that most of the effect takes place within 10 minutes after the ad. This significantly lowers the computational demands since it is possible to calculate most of the effects separately. However, there is the exception of overlapping ads in the assumed time gap. The solution for this case which might happen is described in section 3.2.2. However, there is no methodological limitation of this model to extend the mentioned gap, it will only take more time to compute.

3.2.1 Estimation of one geometrically distributed lag

In case of just one ad within the 10 minutes gap, it is possible to perform the Koyck transformation. The advantage of this transformation is its ability to make a linear model from a non-linear one. An equation for an infinite distributed lag model is expressed as

$$y_t = \alpha + \sum_{i=0}^{\infty} (\beta_i x_{t-i}) + u_t. \tag{3}$$

Here y stands again for the webpage visits and x is a dummy variable on the location of the TV ad. Even though it is possible that there are two ads in the same time, we calculate it as if there was just one, since it is impossible to distinguish different effects of their qualities at this moment of calculation. We can estimate the separate effect of each ad (with at least 1 minute time difference) and then include them into the next analysis. The β coefficients in (3) represents the effect of x and they are assumed to be geometrically decreasing. Therefore they can be expressed as $\beta_i = \theta(\phi^i)$. Now it is possible to rewrite (3) to its lagged form:

$$y_{t-1} = \alpha + \theta \sum_{i=0}^{\infty} (\phi^i x_{t-i-1}) + u_{t-1}. \tag{4}$$

The lagged version is used in the transformation to the linearly estimable model. The process consists of subtracting (4) multiplied by ϕ from (3). The resulting linear model is specified as

$$y_t = \delta + \phi y_{t-1} + \theta x_t + v_t, \tag{5}$$

where ϕ represents the carry-over effect of an advertising and θ corresponds to the immediate effect [9]. The intercept δ stands for $\alpha(1-\phi)$ which is the result of the subtraction as well as v_t is in fact $u_t - \phi u_{t-1}$.

3.2.2 Estimation of more geometrically distributed lags

The Koyck transformation has a great limitation when estimating the effect of more variables with geometrically decreasing lags. The reason is that it is impossible to distinguish different carry-over effects for different communication (e.g. own and competitor's). The reason for that is seen in equation

$$y_t = \delta + \phi y_{t-1} + \theta_1 x_{1t} + \theta_2 x_{2t} + v_t, \tag{6}$$

where it is supposed to have two regressors x_1 and x_2. The dynamic marginal effect of x_1 on y would be then $\frac{\partial y_{t+s}}{\partial x_{1t}} = \theta_1 \phi^s$ where s stands for the infinite lag. The effect of x_2 on y would be $\frac{\partial y_{t+s}}{\partial x_{2t}} = \theta_2 \phi^s$. The Koyck transformation would only cause averaging of both carry-over coefficients as a result of estimating just one parameter (ϕ) [10]. Another attitude might be VAR models but in that case, we cannot guarantee the geometrically decreasing effect of advertisement. Therefore we decided to think of another method based on the non-linear optimization. It has the same assumptions as the method mentioned in 3.2.1, moreover it is estimated with OLS. The following equation describes the advertisement effect

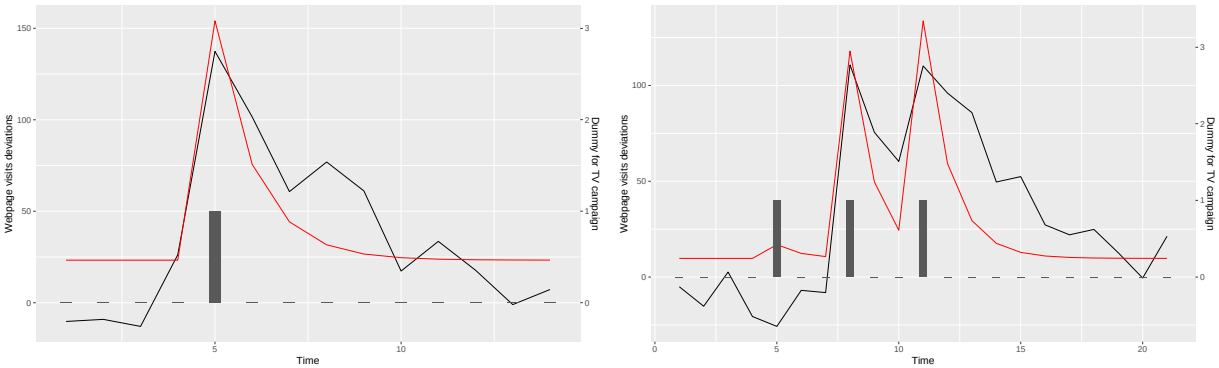
$$y_t = \alpha + \sum_{i=1}^{\infty} \phi_i^{t-s_i} \theta_i z_{it} + u_t. \tag{7}$$

It includes the geometrically decreasing lag for more variables and behaves non-linearly because of the carry-over effect coefficients. It must be calculated for each t within the selected time window covering the overlapping effects of ads and for each campaign (represented by i) inside the window (the total number of campaigns is N). For each campaign, it is estimated its carry-over coefficient ϕ_i and an immediate effect θ_i. The geometrically decreasing lag is written as $\phi_i^{t-s_i}$, where t represents the time for which it is calculated and s_i corresponds to the time when the campaign i starts. The result might be negative and therefore there is z_{it} in the equation which is a dummy telling whether the campaign already is on the air in the time t.

4 Results

4.1 Estimation of the effect

As it was described in section 3.2.1, the estimation of the effect of just one campaign is quite simple if it is only within the window of the assumed time when most of the effect takes place. An example of such an occurrence in the real data is visible in Figure 2a. The limitations of the Koyck transformation might be visible in Figure 2b where the size of the effect of each campaign would be estimated correctly however the speed of the diminishing effect significantly differs and averaging the carry-over coefficients would lead to bad results. Even though it is not clearly visible from the Figure 2, it is important to have the intercept in the equation (7) since it is possible that the average of the differences from the smoothed diurnal pattern is not 0. That would lead to a biased estimation of the ad effect (e.g. higher immediate effect with longer carry-over effect). However, due to the windows which allow for different values of the intercept even the variance of the error cannot be assumed to be the same in each window.



(a) One campaign in the window

(b) More overlapping campaigns in the window

Figure 2 Estimation of the geometrically decreasing effect of TV on the webpage visits deviation from the daily pattern. The bars stand for the dummy variable representing the location of the TV ad. The black line is the real deviations and the red line corresponds to the dynamic model estimate. The window represents a small part of the dataset necessary to estimate the effect.

Now, when the effect of each TV campaign is estimated, it is possible to focus on the reasons standing for their different size. First of all, in Table 1 there are some descriptive statistics about the results. The average webpage visits caused by the effect of TV ads was a bit more than 11 new visitors with quite a short average carry-over effect of 18%. It is convenient to calculate the total effect as Total effect $= \frac{\text{Immediate effect}}{1-\text{Carry over}}$. The size of the total effect might depend on the number of impressions caused by the ad which is represented in marketing as GRP (Gross Rating Point). This dependency confirms Figure 3.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Immediate effect	3,486	11.243	15.177	0.00000	0.413	16.346	136.754
Carry-over	3,486	0.186	0.166	0.000	0.00001	0.381	0.400

Table 1 Descriptive statistics – results of the estimation of each TV ad

However, the dependency of the size of the effect of the TV ad on other qualities of the ad (e.g. length, motive, buying day-part, premium position, ...) might be a great subject for future research. Just a preliminary

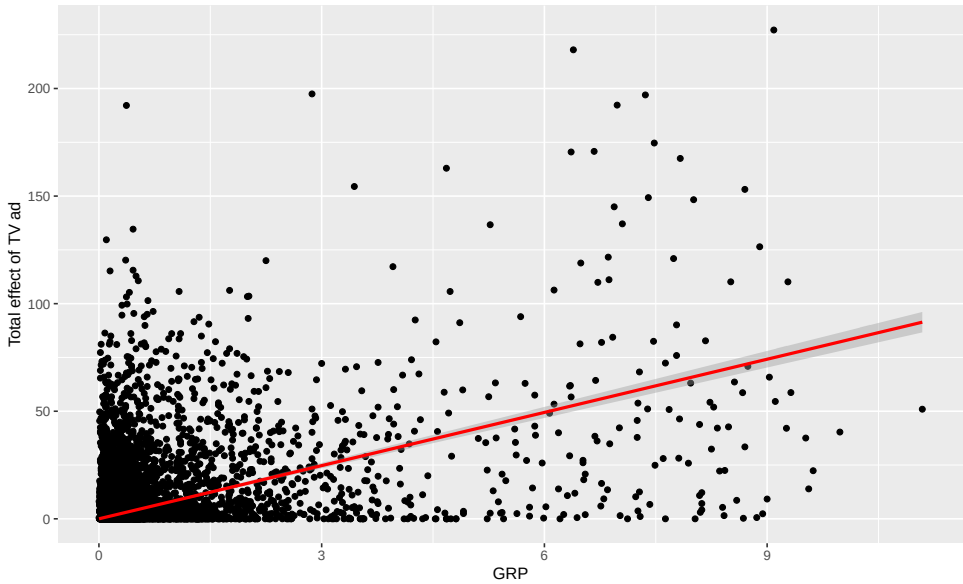


Figure 3 Dependency between the total effect of an ad and the GRP (amount of impressions)

regression showed key factors affecting the size (see Table 2). The dependency on the size of GRP (TV impressions) is clear and confirms that the results are meaningful. Sponsoring also brings more visitors to the website as its position is right before or after the program which makes it more probable to see. Last significant factor affecting the size is buying day-part. Prime time (19:00–23:00) is full of more attractive shows than the rest of the day. Super prime time (19:30–21:30) was not found significant mainly because there were only 2 observation.

	Dependent variable:
	Total effect
GRP	3.354***
	(0.186)
Advertisement type – Sponsoring	5.972***
	(1.891)
Buying day-part – Prime Time	4.118***
	(0.815)
Buying day-part – Super Prime Time	−22.934
	(15.684)
Observations	3,534
R^2	0.397
Adjusted R^2	0.396
Residual Std. Error	22.097 (df = 3529)
F Statistic	464.091*** (df = 5; 3529)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2 Factors affecting the size of the effect of the TV ad

5 Conclusion

The procedure of analysing the existence of the multi-screening phenomenon was described in this paper. It is based on the examination of an increase in organic webpage visits from the moment when a TV ad has been on the air. We focused on the time window of 10 minutes where it is assumed that most of the effect takes place. The results confirm that TV ads can stimulate people to look up the website. Furthermore it is possible to find patterns in the ad's qualities (e.g. amount of the impressions caused by an ad, advertisement type, ...) which generate higher effects. This allows us to optimize the TV and digital strategy.

Firstly, the daily pattern must have been removed from the webpage visits. For that reason, we used the Nadaraya–Watson smoother with Gaussian kernel weights which estimated the course through day. Secondly, the time windows were created with respect to the overlapping ads within the 10 minutes. Then dummy variables were put on the location of the TV ad and the effect (immediate as well as the carry-over) were estimated separately for each ad. This attitude provides us with the information about the effect of each and every TV ad which enables further research on which attributes of the ad cause the highest average increase.

6 Discussion

The main topic for discussion (and simulation) is whether and how the statistical significance of the effect of a TV ad should be measured within the time window since the variance in each window differs. Another topic for discussion is how to get as much knowledge as possible from the estimated effects and the ad qualities. In this paper a multivariate regression was applied, however it cannot find effects of combination of the qualities. Therefore other methods (such as random forest or neural networks) from the field of data mining might provide more interesting results.

Acknowledgements

This work was supported by the Internal Grant Agency of University of Economics, Prague under Grant IGA 19/2019.

References

[1] Lin, C., Venkataraman, S. & Jap, S. (2013). Media Multiplexing Behavior: Implications for Targeting and Media Planning. *Marketing Science*, 32(2), 310–324.

[2] Liaukonyte, J., Teixeira, T. & Wilbur, K. C. (2015). Television Advertising and Online Shopping. *Marketing Science*. 34(3), 311–330.

[3] Zigmond, D. & Stipp, H. (2010). Assessing a New Advertising Effect: Measurement of the Impact of Television Commercials on Internet Search Queries. *Journal of Advertising Research*. 50(2), 162–168.

[4] Tull, D. (1965). The Carry-Over Effect of Advertising. *Journal of Marketing*, 29(2), 46–53.

[5] Tellis, G., Chandy, R. & Thaivanich, P. (2006). Which Ad Works, When, Where, and How Often? Modeling the Effects of Direct Television Advertising. *Journal of Marketing Research*. 37(1), 32–46.

[6] Nadaraya, E. A. (1964). On Estimating Regression. *Theory of Probability & Its Applications*. 9(1), 141–142.

[7] Hastie, T., Tibshirani, R., Friedman, J. (2001). *The Elements of Statistical Learning*. New York, NY, USA: Springer New York Inc.

[8] Clarke, D. (1976). Econometric Measurement of the Duration of Advertising Effect on Sales. *Journal of Marketing Research*. 13(4), 345–357.

[9] Klein, L. R. (1955). Distributed Lags and Investment Analysis. *The Economic Journal*. 65(259), 523–525

[10] Veverka, L. & Frýd, L. (2019). Analysis of TV Advertisements. *37th International Conference on Mathematical Methods in Economics 2019*. 572–578.

Estimation of Selected Production Functions Using Starting Parameters Given by Stochastic Funnel Algorithm

Lenka Viskotová[1], David Hampel[2], Lenka Roubalová[3]

Abstract. In this paper we deal with the problem of finding starting parameters for numerical methods designed for estimation of nonlinear production functions. Appropriateness of these starting parameters has a major impact on success of finding stable final parameters estimation. Bad starting parameters can cause the estimate of production function unacceptable from an economic point of view, or divergence of the estimation iteration process. We are focused on two approaches of searching starting parameters: the first one is based on establishing 'self-starting' models and searching preliminary estimates systematically in grid or by random shooting. As the second approach we use so-called stochastic funnel algorithm. We verify deployment of this algorithm for production functions and extend its usage with the possibility of searching for negative parameters. To assess mentioned methods we provide simulation study based on selected production functions. Our results show that stochastic funnel algorithm is able to give starting parameters comparable to the tradition methods for the CES type production functions, and visibly better for the Sato production function.

Keywords: nonlinear least squares, production function, starting parameters, stochastic funnel algorithm

JEL Classification: C51, C63, D24
AMS Classification: 65L05, 90B30

1 Introduction

Production function (PF) is an important tool of economic research. The model of PF serves for analysing the relationships between inputs and outputs. The traditional inputs used in production theory are labour L and capital K. Then for the output Q we can write the production function in the long run as a function of two variables

$$Q = f(K, L).$$

The most PFs are nonlinear in parameters and the estimation of their parameters leads to nonlinear regression problems, where the 'good' starting values of parameters are required and influence the success of finding stable and economically reasonable parameters estimation. The estimates of parameters are obtained analogically to the linear models by minimizing the sum of squared errors

$$SSE(\hat{\theta}) = \sum_{i=1}^{n} \left[y_i - f(x_i, \hat{\theta}) \right]^2,$$

where $Y_i = f(x_i, \theta) + e_i$, $e_i \sim \text{iid}(0, \sigma^2)$, $i = 1, \ldots, n$ is the general nonlinear model with the vector x_i of independent variables related to the dependent variable y_i, and with the vector of parameters θ. The techniques of minimizing the SSE are based on searching local extremes and according to [2] has to challenge the problems of non-smooth objective functions with large flat areas or the problems of discontinuity of the functions. It is more than appropriate to try to find really good starting values not far away from minimizing parameters to guarantee that the SSE gets continuously downwards in a path to the minimum. Otherwise, the iterative process may converge to a wrong solution set (local minimum), or totally diverge, or collapse due to the overflow or underflow errors.

[1] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, 62100 Brno, lenka.viskotova@mendelu.cz
[2] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, 62100 Brno, david.hampel.uso@mendelu.cz
[3] Mendel University in Brno, Faculty of Business and Economics, Department of Statistics and Operation Analysis, Zemědělská 1, 62100 Brno, xkravack@node.mendelu.cz

The aim of this paper is to verify deployment of the so-called stochastic funnel algorithm (SFA) for estimation of production functions and confront the results with the traditional methods based on establishing 'self-starting' models and searching preliminary estimates systematically in grid or by random shooting. For this purpose, a simulation study is performed.

2 Material and Methods

2.1 Production Functions

In our paper we consider four special production functions differing in numbers of parameters. The Cobb-Douglas PF, firstly presented in [1] and exhibiting constant returns to scale, has the form

$$Q = cK^a L^{1-a}, \tag{1}$$

where c is a positive constant expressing the level of technology and $a \in (0, 1)$ is an output elasticity with respect to capital. The general form of the Cobb-Douglas PF

$$Q = cK^a L^b \tag{2}$$

possesses three parameters, newly $b \in (0, 1)$ representing an output elasticity with respect to labour. The sum $a + b > 1$ corresponds to increasing returns to scale, the sum $a + b < 1$ to decreasing returns to scale. The case $a + b = 1$ is equivalent to (1). The Cobb-Douglas PF is a representative of production functions called the constant elasticity of substitution (CES) PF. The most general form of the CES PF including cases with variable returns to scale was given by Kmenta in [3]:

$$Q = \gamma[\delta K^{-\rho} + (1 - \delta)L^{-\rho}]^{-v/\rho}, \tag{3}$$

where $\gamma > 0$ is a parameter of efficiency, $0 < \delta < 1$ is a parameter of distribution, $v > 0$ is a parameter of returns to scale and $\rho \in \langle -1, 0) \cup (0, \infty)$ is known as a parameter of substitution. The last investigated PF

$$Q = \frac{K^2 L^2}{aK^3 + bL^3}, \tag{4}$$

where a, b are positive constants, belongs to non-CES PFs and was introduced in [8] by Ryuzo Sato.

2.2 Simulation Study

Data

Within the outlined simulation study, we used the same scheme as we had established in [9] and which turned out to be suitable. 50 realizations of uniformly distributed random values of $K \sim U(10000, 50000000)$, resp. $L \sim U(1000, 5000000)$) were generated and the chosen values were assigned to the parameters of particular PFs as below:

- the basic Cobb-Douglas PF (1): $a = 0.1$ and $c = 5$;
- the general Cobb-Dougals PF (2): $a = 0.3$, $b = 0.4$ and $c = 5$;
- the CES PF (3): $\delta = 0.4$, $\rho = 0.6$, $v = 0.8$ and $\gamma = 11$;
- the Sato PF (4): $a = 0.5$ and $b = 0.5$.

Then the output Q was calculated following the PFs with given theoretical parameters and biased by the addition of $e \sim N(0, \sigma^2)$ for the selection $\sigma = 1000$ in the case of the CES PF and $\sigma = 10000$ otherwise.

'Self-Starting' Models

Establishing 'self-starting' models needs as a first step the linearization of the models (1), (2), (4). The equations of these linearized PF can be found in [9]. The CES PF (3) cannot be linearized and thus is approximated by so called Kmenta's approximation (see [3]). The second step is to use multivariate linear regression. This gives estimates of basic starting parameters (case BSP). The SSE value is calculated.

The first technique of improving starting parameters is regular splitting of the interval $I_i = (\beta_i - 0.1|\beta_i|, \beta_i + 0.1|\beta_i|)$ for each parameter β_i and searching such combinations of parameters on this grid, which give the least SSE (case GR). In our study the interval is split in 40 points.

Further technique is called 'random shooting' and the difference from the previous method is that we generate random uniformly distributed values from the intervals I_i. This gives $40^{\text{number of parameters}}$ combinations of parameters for the purpose of finding the least SSE again (case RS).

Stochastic Funnel Algorithm

The Stochastic Funnel Algorithm (SFA) introduced in [4] is the other approach for which we want to verify the ability to find good starting values (case SFA). It relies on a trust region within the landscape of the SSE. At the beginning it needs to be emphasized that the SFA in the cited paper assumes positive parameters because the search space is logarithmic. This is why we modify the SFA algorithm for positive and negative values as in the case of parameter ρ in (3). The SFA consists of several steps:

1. To start the SFA the population of 100 vectors of parameters of the PF is initialized. Each parameter falls in the interval $\left(10^{-6}, 10^{6}\right)$ and it is generated by uniformly distributed random values on the interval $(-6, 6)$. It means that each order of magnitude has approximately 8 values, i.e. the SFA does not work in any way with estimates obtained by linearization or linear approximation of the PF. In the case of parameters from $(0, 1)$ we replace the interval $(-6, 6)$ by the interval $(-6, 0)$.
2. The SSE values are associated with all generated vectors of the trust region and sorted on the basis of ascending SSE. The best 20 combinations of the parameters will help to define new trust region, the rest of vectors are omitted. O'Connor et al. [4] call this process elitism.
3. The updated trust region, at any iteration of the SFA, for any particular parameter β, is given as

$$\left\langle \max\left\{10^{-6}, \beta_{min} - R_D\left(\beta_{max} - \beta_{min}\right) - R_S\right\}, \min\left\{10^{6}, \beta_{max} + R_D\left(\beta_{max} - \beta_{min}\right) + R_S\right\}\right\rangle,$$

   resp. for parameters from $(0, 1)$

$$\left\langle \max\left\{10^{-6}, \beta_{min} - R_D\left(\beta_{max} - \beta_{min}\right) - R_S\right\}, \min\left\{10^{0}, \beta_{max} + R_D\left(\beta_{max} - \beta_{min}\right) + R_S\right\}\right\rangle,$$

   where the minimum β_{min} and maximum β_{max} of each parameter β are extreme values in the best 20 parameters combinations. Constants dynamic relaxation R_D and static relaxation R_S are set empirically to 0.1 and 10^{-12} excluding the CES PF where $R_D = 0.4$ improves the convergence behaviour in the nonlinear squares method (NLS). More about these constants can be found in [4].
4. The search space is converted to a log_{10} space, which guarantees that the new 100 vectors of parameters in the next generation have uniformly distributed exponents of parameters. The previous best 20 vectors are combined with the new 100 vectors and again sorted considering the SSE values to determine the new best 20 parameters combinations.
5. The entire process continues until a termination criterion is met. In this study it is given by the inequality

$$\frac{SSE_{max}}{SSE_{min}} \leq 1.001, \tag{5}$$

   where SSE_{min} and SSE_{max} are the extreme values from the best 20 parameters combinations of the current generation.

The CES PF (3) has the parameter $\rho \in \langle -1, 0) \cup (0, \infty)$, thus both positive and negative values are possible. The solution of this problem lies in setting two separate trust regions initialized as $\left(10^{-6}, 10^{0}\right)$ and $\left(10^{-6}, 10^{6}\right)$. When the termination criteria are met in both cases, the starting values are chosen according to the least SSE value. All the calculations for the values of the interval $(-1, 0)$ generate values from $(0, 1)$ and the SSE values have to be computed with the adequate changes of signs.

Estimation of Parameters of Production Functions

For the purpose of estimation of parameters the nonlinear least squares method is employed. The information about convergence, the number of iterations and the final SSE are stored and compared with the results obtained using different techniques of getting starting parameters. Comparisons are also made to the SSE values calculated before starting the NLS method.

Computations

All the calculations are made in computational system Matlab R2019b, the NLS method is performed using the function `fitnlm`. The entire simulation consists of 100 replications of the procedure described above for each type of the PF.

3 Results and Discussions

A convergence of iterative processes in estimation of nonlinear model parameters is an essential point and can be summarized as perfect in both cases of the Cobb-Douglas PF when all iterative processes terminated successfully. The iterative processes of the NLS method for the CES PF collapsed for about 5% of replications, no matter what kind of searching starting parameters had been selected. The most interesting situation can be viewed for the Sato PF, when the SFA provides all iterative processes to be convergent. In contrast, the other ways of getting starting values of the Sato PF lead in about 13% of divergent iterations.

The average numbers of iterations needed for convergence in particular replication in Table 1 bring an interesting insight into the iterative processes, where convergence was reached. The basic Cobb-Douglas PF with 2 parameters and the CES PF are estimated at approximately the same number of iterations across all cases. The results for the general Cobb-Douglas PFs are available much more slowly in the case of the SFA. In contrast, the SFA gives the estimation of the Sato PF really fast not only compared to the cases of 'self-starting' models, but also in general.

Production function	case BSP	case GR	case RS	case SFA
Basic Cobb-Douglas	3.80	3.97	3.94	3.40
General Cobb-Douglas	10.70	9.79	10.35	55.08
CES	12.00	13.01	13.16	13.15
Sato	13.40	13.63	13.77	2.97

Table 1 The average number of iterations

Table 2 and Table 3 bring an important assessment of the quality of the estimated production functions. Table 2 considers the average SSE values for models with starting parameters before employing the NLS method, Table 3 presents the average SSE values for models already estimated by the NLS method. As in Table 1, only replications where convergence of iterative process was reached are included. If we compare both approaches ('self-starting' models and the SFA) for the basic Cobb-Douglas PF, the general Cobb-Douglas PF and the CES PF, the SSE values in Table 3 are similar. As regards Table 2, differences in these three production functions can be observed there. When we focus on the Sato PF, the quality of the estimate with respect to the second approach (the case SFA) is significantly better then for the other cases. The tables indicate that the starting parameters for the Sato PF give satisfying results even before the NLS method.

Production function	case BSP	case GR	case RS	case SFA
Basic Cobb-Douglas	$1.07 \cdot 10^{11}$	$3.27 \cdot 10^{10}$	$3.24 \cdot 10^{10}$	$1.15 \cdot 10^{13}$
General Cobb-Douglas	$5.86 \cdot 10^{9}$	$8.16 \cdot 10^{11}$	$5.35 \cdot 10^{9}$	$1.26 \cdot 10^{11}$
CES	$1.14 \cdot 10^{8}$	$2.07 \cdot 10^{9}$	$5.59 \cdot 10^{7}$	$5.53 \cdot 10^{7}$
Sato	$1.50 \cdot 10^{20}$	$8.14 \cdot 10^{18}$	$8.18 \cdot 10^{18}$	$4.83 \cdot 10^{9}$

Table 2 The average SSE values for particular cases of starting parameters before employing the NLS method

Production function	case BSP	case GR	case RS	case SFA
Basic Cobb-Douglas	$4.72 \cdot 10^{9}$	$4.72 \cdot 10^{9}$	$4.72 \cdot 10^{9}$	$4.72 \cdot 10^{9}$
General Cobb-Douglas	$4.76 \cdot 10^{9}$	$4.76 \cdot 10^{9}$	$4.76 \cdot 10^{9}$	$4.76 \cdot 10^{9}$
CES	$4.70 \cdot 10^{7}$	$4.70 \cdot 10^{7}$	$4.70 \cdot 10^{7}$	$4.73 \cdot 10^{7}$
Sato	$3.63 \cdot 10^{13}$	$3.62 \cdot 10^{13}$	$3.55 \cdot 10^{13}$	$4.82 \cdot 10^{9}$

Table 3 The average SSE values for particular cases of starting parameters after employing the NLS method

As mentioned above, in the case SFA, in contrast to the other cases, all replications for the Sato PF are convergent. Moreover, each replication gives 'good' SSE values and its estimated parameters are close to those from data simulation scheme. In addition, the results for using the SFA differ from the results with the 'self-starting models' to such an extent that approximately 39% of all convergent replications of the cases BSP, GR, RS correspond to the situation when the given extreme is certainly not global, the SSE value is too large and the estimated parameters are far from those used to simulate the data.

An example of the values from these ineffective replications can be found in Table 4 and Table 5. Table 4 contains values before using the NLS method, Table 5 contains values after employing the NLS method and includes number of iterations. The advantage of using the SFA method in comparison with the methods using 'self-starting' models is more than obvious here. In the discussed example, 14 generations of the SFA were sufficient to find suitable starting parameters. Figure 1 demonstrates the speed of convergence of the maximal and the minimal values of the estimated parameters of the Sato PF in the given example.

Sato Production function	case BSP	case GR	case RS	case SFA
parameter a	-0.0121	-0.0133	-0.0133	0.4994
parameter b	-22.6602	-24.9262	-24.8521	0.5014
SSE	$1.17 \cdot 10^{15}$	$1.00 \cdot 10^{15}$	$1.00 \cdot 10^{15}$	$6.60 \cdot 10^{9}$

Table 4 An example of the starting parameters for estimation of the Sato PF

Sato Production function	case BSP	case GR	case RS	case SFA
parameter a	$1.06 \cdot 10^{12}$	$5.61 \cdot 10^{9}$	$5.36 \cdot 10^{9}$	0.4997
parameter b	$-1.49 \cdot 10^{18}$	$-7.90 \cdot 10^{15}$	$-7.53 \cdot 10^{15}$	0.5013
SSE	$1.03 \cdot 10^{14}$	$1.03 \cdot 10^{14}$	$1.03 \cdot 10^{14}$	$6.59 \cdot 10^{9}$
number of iterations	12	13	13	3

Table 5 An example of the estimated parameters of the Sato PF using the NLS method

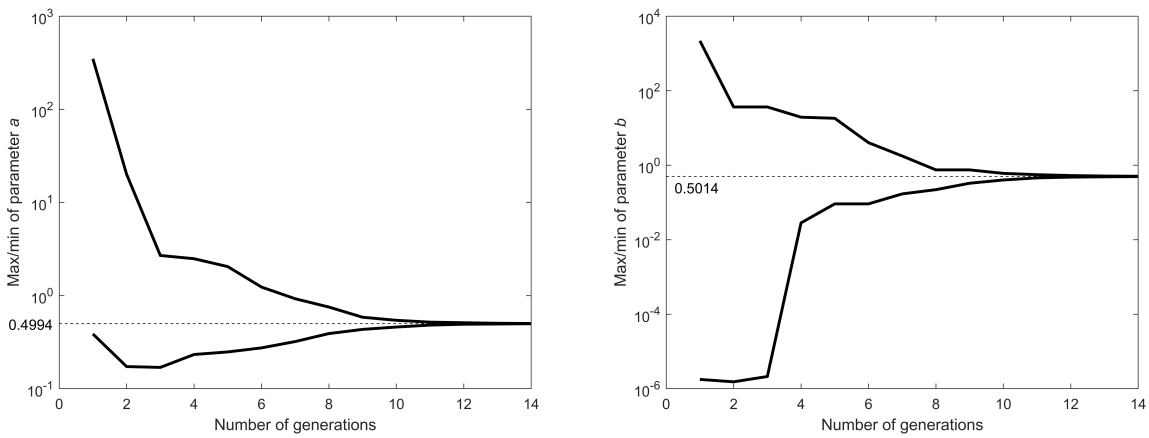


Figure 1 Convergence of max/min values of the parameters of the Sato PF in the SFA

The number of generations of the SFA in the previous example of the Sato PF has no general informative value, it always depends on the given functional form, as can be seen in Table 6. This table contains the average numbers of generations in the SFA for different PFs. These values must be also assessed with respect to unequal termination criteria specified for the particular PFs (see section 2.2). In the case of the Sato PF the limit 1.1 in (5) gives excellent results as well and within less generations of algorithm.

Production function	number of generations in the SFA
Basic Cobb-Douglas	16.42
General Cobb-Douglas	472.36
CES	1992.68
Sato	13.94

Table 6 The average numbers of generations for particular cases of searching starting parameters in the SFA

4 Conclusions

The simulation study aims to compare two different approaches of setting starting values of parameters for estimation production functions using the NLS method. We can conclude that the parameters of the Sato PF are estimated visibly better by the SFA than by the methods of the other cases. However, the other types of PFs give better or the same results in cases of an approach based on 'self-starting' models than with the SFA. The results point out that the appropriate way of searching starting values of parameters relies on the form of the production function and it plays a crucial role in the success of the NLS method.

The stochastic funnel algorithm makes an open space for the discussion on the assigning values to the constant R_D and R_S, as well as on the termination criteria (see [5]). The choice of these values influences searching of starting parameters and it cannot be unified across different types of production functions.

Further research could focus on estimates of selected production functions for various settings of initialization values of parameters, which are used for data simulation, and make comparisons of different approaches within each production function. The paper [6] and [7] extends the issue of the production functions by a time variable and the possible research in the field of starting values of parameters for these time-augmented PFs could bring beneficial conclusions.

Acknowledgements

This article was supported by the grant No. PEF/TP/2020002 of the Grant Agency IGA PEF MENDELU.

References

[1] Cobb, C. W. & Douglas, P. H. (1928). A Theory of Production. *The American Economic Review*, 18, 139–165.

[2] Henningsen, A. & Henningsen, G. (2012). On estimation of the CES production function—Revisited. *Economics Letters*, 115, 67–69.

[3] Kmenta, J. (1967). On Estimation of the CES Production Function. *International Economic Review*, 8(2), 180–189.

[4] O'Connor, A., Gao, J. & Louis, J. (2008). Using a Stochastic Funnel to find NLR Starting Values. In H. R. Arabnia & Y. Mun (Eds.), *WORLDCOMP '98: World Congress in Computer Science, Computer Engineering and Applied Computing* (pp. 96–102). USA: CSREA Press.

[5] O'Connor, A., Gao, J. & Louis, J. (2009). Termination Criteria for Evolutionary Algorithms. In *Proceedings of the 2009 International Conference on Genetic and Evolutionary Methods* (pp. 35–42). Las Vegas: CSREA Press.

[6] Roubalová L. & Viskotová, L. (2018). Productivity Development in Selected Central European Countries Measured by the Sato Production Function. *Review of economic perspectives*, 18(4), 353–370.

[7] Roubalová L. & Viskotová, L. (2019). The Time Augmented Cobb-Douglas Production Function. *Acta Universitatis agriculturae et silviculturae Mendelianae Brunensis*, 67(5), 1347–1356.

[8] Sato, R. (1964). Diminishing Returns and Linear Homogeneity: Comment. *The American Economic Review*, 54, 744–745.

[9] Viskotová, L., Roubalová, L. & Hampel, D. (2019). Starting Parameters for the Estimation of Production Functions. In *International Conference of Numerical Analysis and Applied Mathematics (ICNAAM 2018)*. Melville: American Institute of Physics (AIP).

Bivariate Geometric Distribution and Competing Risks: Statistical Analysis and Application

Petr Volf[1]

Abstract. The present contribution studies the statistical model for discrete time two-variate duration (time-to-event) data. The analysis is complicated by just partial data observation caused either by the right-side censoring or even by the presence of dependent competing events. The case is modeled and analyzed with the aid of a two-variate geometric distribution. The model identifiability is discussed and it is shown that the model is not identifiable without proper additional assumptions. The method of analysis is illustrated both on artificially generated example and on real unemployment data.

Keywords: bivariate geometric distribution, discrete time, event-history analysis, competing risks, unemployment data.

JEL Classification: C41, J64
AMS Classification: 62N02, 62P25

1 Introduction

In many time-to event data cases, more than one potential event can be associated with observed objects. For example, we can follow X and Y being the times to failure of a device caused by different reasons. To cope with such cases, various bivariate probability distributions have been introduced in the literature (see for example Marshall and Olkin [6] and many other sources). However, most of them were developed for continuous time cases. In the discrete-time lifetime data (often originated from continuous time processes with observations aggregated to intervals), it is assumed that the lifetimes X and Y are discrete random variables (attaining positive integer values, measured in corresponding time units). Then, the distribution of the time to event can be modeled by different variants of geometric distribution. Naturally, corresponding discrete bivariate distributions have been introduced in the literature as well, as the bivariate geometric distribution versions of Basu and Dhar [1]. However, it might be said that despite frequent discrete measuring of lifetime and other duration data, very common in applications, few papers related to discrete lifetime data appears in the literature (see for example, Grimshaw et al. [4], Davarzani et al. [3]). One of the reasons is that the continuous time models are often more "comfortable" both from the point of analysis and of theoretical knowledge (compare for instance the frequent use of the Cox regression model).

When each of followed events terminates the observation (like, for instance, a critical failure of a device) the events are competing, just the first occurring is observed. When corresponding random variables, times to these events, are independent, one variable censors (randomly, from the right side) the other. However, if they are dependent mutually, we deal with more complicated case of dependent competing risks. Discrete time case is further complicated (when compared to the continuous time setting) by potential occurrence of both events in the same time interval, though just one of them really happens. Consequently, in general, in such a case the model parameters are not identifiable. This could be easily shown by the analysis of log-likelihood, which is then evidently over-parametrized.

In the present contribution, first, a version of two-variate geometric distribution is recalled. Then, the setting of two competing risk is described, together with well know problems with model identification. The main part then studies the assumptions under which the model identifiability holds. The analysis method and the impact of additional assumptions will be illustrated on artificial data and the results discussed. Finally, a real data example will be presented.

2 Bivariate geometric distribution

Standard univariate geometric distribution corresponds to the order of the first "1" in the series of Bernoulli "0-1" attempts. Thus, when U is the Bernoulli r.v. with $P(U = 1) = p$, the geometric r.v. X "starting at 1" has $P(X = k) = (1 - p)^{k-1} \cdot p$, for $k = 1, 2, \ldots$, $\mathrm{E}X = 1/p$, $\mathrm{var}(X) = (1 - p)/p^2$.

[1] Department of Stochastic Informatics, ÚTIA AV ČR, Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic, volf@utia.cas.cz

Similarly, a basis for the bivariate geometric model is the bivariate Bernoulli random variable (U, V) with joint probability distribution p_{00}, p_{10}, p_{01}, p_{11}, where $p_{jk} = P(U = j, V = k)$, j, k are 0 or 1. The marginal probabilities are $p_{1.} = P(U = 1) = p_{10} + p_{11}$, $p_{.1} = P(V = 1) = p_{01} + p_{11}$ and the covariance $\mathrm{cov}(U, V) = p_{11} - p_{1.} \cdot p_{.1}$. Recall also that in the case $p_{11} = p_{1.} \cdot p_{.1}$, i.e. when U, V are non-correlated, they are already independent, too.

Let us denote corresponding geometric random variables X, Y. Then, as in Basu and Dhar [1], the following form of bivariate geometric distribution will be considered, for $s, t = 1, 2, \ldots$:

$$P(X = s, Y = t) = p_{00}^{s-1} \cdot p_{10} \cdot p_{.0}^{(t-s-1)} \cdot p_{.1} \quad \text{for } s < t,$$

$$P(X = s, Y = t) = p_{00}^{t-1} \cdot p_{01} \cdot p_{0.}^{(s-t-1)} \cdot p_{1.} \quad \text{for } s > t,$$

$$P(X = s, Y = t) = p_{00}^{s-1} \cdot p_{11} \quad \text{in the case } s = t. \tag{1}$$

Covariance can be also computed easily, its sign is the same as of $\mathrm{cov}(U, V)$. Namely

$$\mathrm{cov}(X, Y) = (p_{10} + p_{01}) * (1 + p_{00})/(1 - p_{00})^3 + (p_{10}/p_{.1} +$$

$$+ p_{01}/p_{1.})/(1 - p_{00})^2 + p_{11}(1 + p_{00})/(1 - p_{00})^3 - 1/(p_{.1}\, p_{1.}).$$

Correlation then equals $\mathrm{corr}(X, Y) = \mathrm{cov}(X, Y) \cdot p_{1.}\, p_{.1} / \sqrt{(1 - p_{1.})\,(1 - p_{.1})}$.

3 Competing risks problem

The interest in the problem of mutually dependent competing risks dates back to 70-ties of the last century. Formally, there are K random variables $T_j, j = 1, \ldots, K$, running simultaneously, each representing the time to certain event. The occurrence of events, therefore also corresponding random variables, can be dependent mutually. Observation is terminated at the minimum of them. Let $\overline{F}_K(t_1, \ldots, t_K) = P(T_1 > t_1, \ldots, T_K > t_K)$ be the joint survival function of $\{T_j\}$. However, instead the 'net' survivals T_j we observe just 'crude' data (sometimes called also 'the identified minimum') $Z = \min(T_1, \ldots, T_K)$ and the indicator $\delta = j$ if $Z = T_j$. Such data allow for direct estimation of the distribution of $Z^* = \min(T_1, \ldots, T_K)$, for instance its survival function $S(t) = P(Z^* > t) = \overline{F}_K(t, \ldots, t)$.

Generally, however, from data (Z_i, δ_i), $i = 1, \ldots, N$ it is not possible to identify the marginal or joint distributions of $\{T_j\}$. Tsiatis in [8] has shown that for arbitrary joint model we can find a model with independent components having the same incidences, i.e. we cannot distinguish among the models. It follows that it is necessary to make certain functional assumptions about the form of both marginal and joint distribution in order to identify them. Several such cases are specified for instance in Basu and Ghosh [2]. Later on the case of competing risks with covariates was studied by many other authors, in a more precise way, already with the aid of a copula describing the dependence. However, the results concern mostly the continuous time setting. On the contrary, our interest lies in the analysis of discrete models.

4 Bivariate geometric distribution under competing risks

As it has already been said, in the competing risks case only $T = \min(X, Y)$ is observed. However, in the discrete time setting a serious problem arises, that both X, Y may occur in the same time interval, i.e. $X = Y = t$, however just one of them is observed, the "first" one. The probability of such an instance equals $p_{11} \cdot \sum_{t=1}^{\infty} p_{00}^{t-1} = p_{11}/(1 - p_{00})$. In other words, if $X = t$ is observed, on Y we know only that $Y \geq t$ and the probability of such a result is between $p_{00}^{t-1} \cdot p_{10}$ and $p_{00}^{t-1} \cdot (p_{10} + p_{11})$. In fact, we are not able to evaluate it precisely, without an additional assumption. Let me recall here a similar problem from the analysis of discrete-time life tables with censoring. We know the number of item at the beginning of interval, and numbers of items failed or censored during. However, for accurate estimation of survival function, for instance by the Product Limit Estimate, also the order of them is necessary. There are two boundary instances, namely that censoring precedes failures, or vice versa, the reality is between (c.f. Prentice and Gloeckler [7]). In order to cope with this problem here, let us first formulate the following:

Assumption A1. Let us assume that the probabilities of observed data can be expressed with the aid of a (known) constant $C \in [0, 1]$ as

$$P(T = X = s, Y \geq T) = p_{00}^{s-1} \cdot (p_{10} + C\, p_{11}),$$

$$P(T = Y = t, X \geq T) = p_{00}^{t-1} \cdot (p_{01} + \bar{C}\, p_{11}).$$

Here $\bar{C} = 1 - C$.

Let us denote these two cases by an indicator δ: $\delta = 1$ in the first case, $\delta = 2$ in the second. Further, let $p_1 = p_{10} + Cp_{11}$, $p_2 = p_{01} + \bar{C}p_{11}$, $p_0 = 1 - p_1 - p_2 = p_{00}$. These probabilities are identifiable and characterize the **incidence** of really observed events.

The sense of Assumption 1 could be clarified, at least to certain extent, by following examples:

Example 1. Let us consider a case of employment data The events are "to leave voluntarily", X, or "to be fired", Y. One can imagine that one leaves the company just before being dismissed, i.e. $X = Y$, both are (Y just potentially) in the same time interval, but only X is observed. On the other hand, it is hard to imagine an opposite case that one is fired despite he already announced his decision to leave. In such a case, $C = 1$ could be a reasonable choice.

Example 2. X, Y are the times of the first (or next) goal in an ice-hockey match, of both teams. If one team scored first, still there was a potential possibility that the second team would score first, in the same interval (the same minute, say). Its chance can be estimated comparing "scoring strengths" of teams, for instance putting $C = p_{10}/(p_{10} + p_{01})$, $\bar{C} = p_{01}/(p_{10} + p_{01})$.

Nevertheless, Assumption 1 itself does not suffice to model identification, as it quantifies just relative proportions of p_{11}, not its relation to other probabilities. Let observed data consist in N independent replications of T, δ, $T = \min(X, Y)$, $\delta = 1, 2$. The likelihood of unknown probabilities p_{jk}, $j, k = 0, 1$, under these data is

$$\mathcal{L} = \prod_{i=1}^{N} \left(p_0^{(T_i - 1)} \cdot p_1^{I[\delta_i = 1]} \cdot p_2^{I[\delta_i = 2]} \right), \tag{2}$$

where $p_{00} = p_0$ and $p_1 = p_{10} + C\, p_{11}$, $p_2 = p_{01} + \bar{C}\, p_{11}$. Hence we have 3 unknown parameters related to 2 known (well estimated) incidence probabilities p_1, p_2. Therefore p_{jk} are not identifiable.

4.1 Special cases

The simplest case arises when $p_{11} = 0$, i.e. there is no chance of both events occurrence in one time interval. In such a case there is no identification problem. Notice also that then the $\mathrm{corr}(U, V)$ as well as $\mathrm{corr}(X, Y) < 0$.

Independent variables. Another particular case occurs when U, V are not correlated. They are then also independent and X, Y are independent as well. In this case Assumption 1 suffices to identification of their distribution, as there are in fact just two probabilities to be estimated, namely $p_{1.}$ and $p_{.1}$, the rest can be derived from them. This instance covers also the case of right-side random censoring: For example, when X is a variable of our interest and Y is censoring variable, we are in fact not interested in estimation of characteristics of Y. Nevertheless, we need to assume something about order of values X, Y when occurring potentially in the same time interval. The instance is often encountered when life tables are analyzed, and standardly it is assumed (as in Prentice, Gloeckler [7]) that censoring occurred at the intervals end. In our setting it means $C = 1$. Naturally, this is just an assumption, other border case can assume $C = 0$, the reality is between.

4.2 An assumption guarantying identifiability

Let us try to propose another kind of limitation to model parameters. In fact, the meaning of constant C is

$$C = P(\delta = 1 | X = Y), \quad \bar{C} = P(\delta = 2 | X = Y).$$

Using the Bayes formula, we are able to evaluate the opposite,

$$P(X = Y | \delta = 1) = \frac{P(\delta = 1 | X = Y) \cdot P(X = Y)}{P(\delta = 1)}.$$

Denote this by α. As $P(\delta = 1) = P(\delta = 1 | X < Y) \cdot P(X < Y) + P(\delta = 1 | X = Y) \cdot P(X = Y) + P(\delta = 1 | X > Y) \cdot P(X > Y) = 1 \cdot p_{10} + C \cdot P_{11} + 0$, then

$$\alpha = \frac{C\, p_{11}}{p_{10} + C\, p_{11}} \text{ and } C = \frac{\alpha\, p_{10}}{(1 - \alpha)\, p_{11}}.$$

Quite similarly, for $\beta = P(X = Y|\delta = 2)$ we have

$$\beta = \frac{\bar{C}\, p_{11}}{p_{01} + \bar{C}\, p_{11}} \quad \text{and} \quad \bar{C} = \frac{\beta\, p_{01}}{(1-\beta)\, p_{11}}. \tag{3}$$

Another consequence is that

$$p_{11} = \frac{\alpha}{1-\alpha} p_{10} + \frac{\beta}{1-\beta} p_{01} = \alpha\, p_1 + \beta\, p_2. \tag{4}$$

Constants $\alpha,\ \beta$ characterize the proportion of hidden events $X = Y$ in observed $\delta = 1$ or $\delta = 2$, resp. In fact, as seen also from (4), assumption on knowledge of $\alpha,\ \beta$ is much stronger that just Assumption 1 on C. Nevertheless, it can be considered to be realistic, obtained from a prior knowledge, experience, for instance. Let us formulate it as:

Assumption A2. Let us assume that both constants $\alpha,\ \beta$ defined above are known.

Notice that for instance $\alpha = 1$ means that $p_{10} = 0$, all cases with $\delta = 1$ are caused by events $X = Y$. Similarly, $\beta = 1$ is equivalent to $p_{01} = 0$, in fact we then deal with degenerate cases.

Now, it is easy to show that the model is identifiable. The log-likelihood is now

$$L = \sum_{i=1}^{N} \Big((T_i - 1) \ln p_{00} + I[\delta_i = 1] \ln(p_{10}/(1-\alpha)) + I[\delta_i = 2] \ln(p_{01}/(1-\beta)) \Big), \tag{5}$$

with $p_{00} = 1 - p_{10}/(1-\alpha) - p_{01}/(1-\beta)$. Hence, there are just 2 parameters to be estimated, the rest are then obtained from them. If both $\alpha, \beta < 1$, there is no problem to get consistent estimates of $p_{10}/(1-\alpha)$, $p_{01}/(1-\beta)$, and then of all original p_{jk}, also p_{11} from (4). Naturally, a wrong specification of α, β, leads to error in estimates of p_{jk}.

Let us also return to examples 1 and 2 from above: In Example 1, with $C = 1$, we obtain that $\alpha = p_{11}/(p_{10} + p_{11})$, $\beta = 0$, and $p_{11} = \alpha \cdot p_{10}/(1-\alpha)$. Similarly for $C = 0$, i.e. $\bar{C} = 1$.

In Example 2 we obtain that $\alpha = \beta = p_{11}/(1 - p_{00})$.

4.3 Artificial example

The data were generated from the bivariate geometric model (1) with parameters $p_{00} = 0.7,\ p_{10} = 0.1,\ p_{01} = 0.15,\ p_{11} = 0.05$. Further, we selected $C = 0.5$ which yields $\alpha = 0.2000$ and $\beta = 0.1429$, from (3). These values were taken as known constants, in accord with Assumption 2. The estimation followed the standard MLE scheme, with the aid of the Newton-Raphson iteration. As it uses the first and second derivatives of log-likelihood, then the Fisher information matrix, and, consequently, asymptotic variances of estimates, can be estimated, too.

Two results are displayed below, with $N = 100$ and 1000 generated values.

For $N = 100$:
Estimated $p_{00},\ p_{10},\ p_{01},\ p_{11} = 0.6732,\ 0.0967,\ 0.1765,\ 0.0536$.
Their asymptotic standard deviations were estimated as 0.0301, 0.0150, 0.0242, 0.0048.
Further, estimated covariance and correlation of X and Y was $\mathrm{cov}(X, Y) = 1.6816$, $\mathrm{corr}(X, Y) = 0.0719$, while the real 'true' values were 2.2222, 0.0808, resp.

For $N = 1000$ the following values were obtained:
Estimates $p_{00},\ p_{10},\ p_{01},\ p_{11} = 0.7032,\ 0.0969,\ 0.1506,\ 0.0493$,
with asymptotic standard deviations estimated as 0.0085, 0.0045, 0.0065, 0.0014. Finally, covariance and correlation of X and Y were estimated as $\mathrm{cov}(X, Y) = 2.3153$, $\mathrm{corr}(X, Y) = 0.0819$.

It is possible to say that the precision of estimates is rather good and increases with growing extent of data, N. In both cases the values of (2-dimensional here) score function, i.e. the first derivatives of the log-likelihood, which should be 0 at the log-likelihood maximum, were of order $1.0e - 004$.

Summary. In studied setting we are able to estimate consistently just incidence probabilities $p_1,\ p_2$ corresponding to observed events $\delta = 1$ or 2, resp., hence also $p_{00} = p_0 = 1 - p_1 - p_2$. Further, we know that $p_1 + p_2 = p_{10} + p_{01} + p_{11}$.

Under Assumption 1 $p_1 = p_{10} + C \cdot p_{11}$, $p_2 = p_{01} + \bar{C} \cdot p_{11}$, however even the knowledge of C does not suffice to reliable estimation of p_{jk}, except in the case of independent variables U, V. Just under much stronger Assumption 2 we have that $p_{10} = (1-\alpha) \cdot p_1$, $p_{01} = (1-\beta) \cdot p_2$, and $p_{11} = \alpha\, p_1 + \beta\, p_2$.

A wrong specification of α, β leads to biased estimation of p_{jk} (except p_{00}), while incidence probabilities p_1, p_2 are available directly from the maximization of likelihood (2). Simultaneously they suffice to replication of competing risk data, without knowledge of other parameters.

5 Application

Han and Hausman [5] have analyzed the data on unemployment duration, namely the records on 1051 people, collected there in Table III (with several insignificant misprints which were corrected). The time is discrete, as the information was gathered on a week basis, for 70 weeks. More about data origin can be found in the paper [5]. The data show certain non-regularities, visible also in Figures 1 and 2 below, for instance significantly larger numbers of events in 26-th week which is related to a change of support after the first half-year of unemployment. In general, the censoring (exit from the study) is caused mostly by the termination of unemployment insurance benefits (it concerns to weeks 39 and 52, too), and, naturally, by the end of study. Han and Hausman had also an information on several covariates which was not available to us. They used a discrete version of Cox regression model, with certain not fully justified approximations (e.g. substituting Gauss distribution instead the Gumbel one). The other problem are ties of events limiting the correct use of continuous time model. This problem, which we have discussed in previous sections, has been omitted.

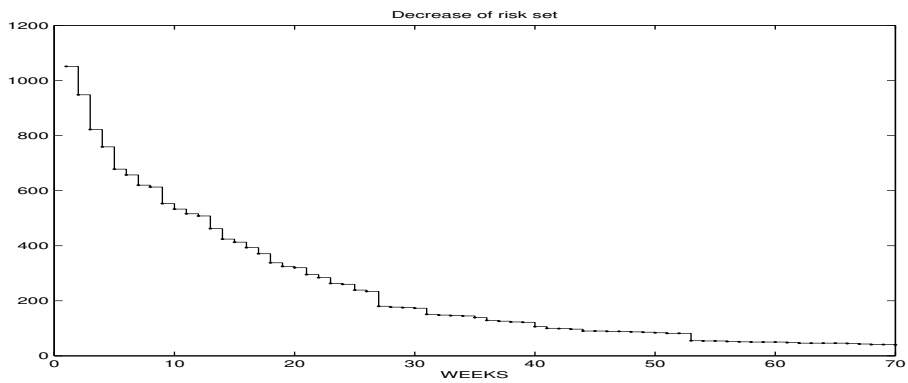


Figure 1 Decrease of number of persons at risk set

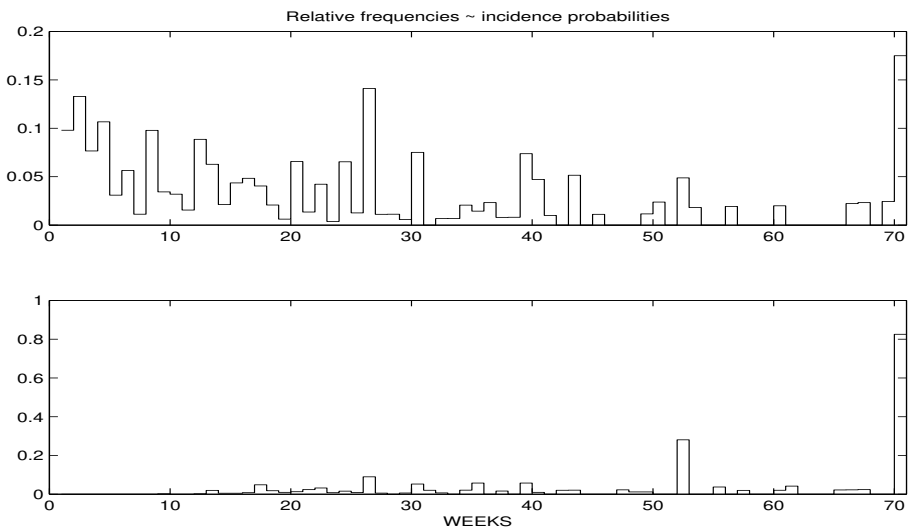


Figure 2 Incidence probabilities estimated by relative frequencies, of re-employment (above), censoring (below).

We have concentrated here to the analysis of unemployment termination and censoring as two independent competing events. Further, it was assumed that censoring occurred at the intervals end, hence constant $C = 1$. From this point of view, the situation was simplified and leading to a unique solution. On the other hand, distributions of Bernoulli random variables $U = U(t)$ for re-employment and $V = V(t)$ for censoring depended on time – weeks from 1 till $T = 70$. Let us denote N_t the number of people staying in the study at the t-th week beginning. Thus, $N_1 = 1051$, $N_{t+1} = N_t - n_t - m_t$,, where n_t, m_t are the numbers of persons re-employed and censored, respectively, in t-th week.

First, the incidence probabilities were estimated as $p_1(t) = n_t/N_t$, $p_2(t) = m_t/N_t$ for variables $U(t)$, $V(t)$, respectively. These estimates are displayed in Figure 2, while Figure 1 shows the decrease of N_t. In the next step, estimates of marginal probabilities $p_{1.}(t)$ of $U(t)$ and $p_{.1}(t)$ of $V(t)$ were computed. Due the assumptions of independence and of $C = 1$, it holds that $p_1(t) = p_{10}(t) + p_{11}(t) = p_{1.}(t)p_{.0}(t) + p_{1.}(t)p_{.1}(t) = p_{1.}(t)$ and $p_2(t) = p_{01}(t) = p_{0.}(t)p_{.1}(t) = (1 - p_{1.}(t))p_{.1}(t)$. Then

$$p_{1.}(t) = p_1(t), \;\; p_{.1}(t) = p_2(t)/(1 - p_1(t)).$$

On the basis of these results, distributions of two independent random variables, the time to re-employment, X, and the time to censoring, Y, were derived easily. Finally, we can construct a distribution of probabilities of competing events, i.e. of being re-employed at s, before censoring, or being censored at t, still unemployed. They are

$$P(X = s, Y \geq s) = \prod_{j=1}^{s-1} (p_{0.}(j)p_{.0}(j)) \cdot p_{1.}(s),$$

$$P(Y = t, X > t) = \prod_{j=1}^{t-1} (p_{0.}(j)p_{.0}(j)) \cdot p_{0.}(t) \cdot p_{.1}(t).$$

6 Concluding remarks

It has to be said that the presented results on identifiability under mutually dependent competing risks and discrete time are not satisfactory. Potential occurrence of unobserved events $X = Y$ is the cause of problems which were overcame just with rather strong assumptions. Naturally, the basic model with constant probabilities can be generalized. Thus, the final example considered time-dependent probabilities. In another setting, the probabilities can depend on explanation variables, covariates, for instance via the logistic regression model. In the continuous time cases, it was proved that the presence of covariates and assumption of a regression model type (e.g. the Cox one) can support the model identifiability (see e.g. Volf [9] for references). However, as the discrete time allows for more events at the same time interval, the situation is worse. In fact, we were not able to show any facilitation caused by the information provided by covariates.

Acknowledgements

The research was supported by the grant No. 18-02739S of the Grant Agency of the Czech Republic.

References

[1] Basu, A. P. & Dhar, S. (1995). Bivariate geometric distribution. *Journal of Applied Statistical Science*, 2, 33–44.

[2] Basu, A. P. & Ghosh, J. K. (1978). Identifiability of the multinormal and other distributions under competing risks model. *Journal of Multivariate Analysis*, 8, 413–429.

[3] Davarzani, N., Achcar, J. A., Smirnov, E. N. & Peeters, R. (2015). Bivariate lifetime geometric distribution in presence of cure fractions. *Journal of Data Science*, 13, 755–770.

[4] Grimshaw, S. D., McDonald, J., McQueen, G. R. & Thorley, S. (2005). Estimating hazard functions for discrete lifetimes. *Communications in Statistics, Part B-Simulation and Computation*, 34, 451–463.

[5] Han, A. & Hausman J. A. (1990). Flexible parametric estimation of duration and competing risk models. *J. of Applied Econometrics*, 5, 1–28.

[6] Marshall, A. W. & Olkin, I. (1985). A family of bivariate distributions generated by the bivariate bernoulli distribution. *Journal of the American Statistical Association*, 80, 332–338.

[7] Prentice, R. & Gloeckler, L. (1978). Regression analysis of grouped survival data with application to the breast cancer data. *Biometrics*, 34, 57–67.

[8] Tsiatis, A. (1975). A nonidentifiability aspects of the problem of competing risks. textitProc. Nat. Acad. Sci. USA, 72, 20–22.

[9] Volf, P. (2018). Problem of competing risks with covariates: Application to an unemployment study. In *Proceedings of the 36th International Conference Mathematical Methods in Economics, 2018.* FM VŠE Jindřichuv Hradec, 624–629.

Minimum Risk Portfolio Optimization with Twofold Efficiency Tests

Anlan Wang[1]

Abstract. Considering the risk aversion attitude to investing activities in the portfolio optimization problems, minimum risk portfolios are generated in our analysis under the framework of the Mean-Variance model. To test the efficiency of the minimum risk portfolios obtained from the in-sample data, twofold work is constructed. Firstly, the out-of-sample data which covered the 2007–2009 global financial crisis is applied specifically for testing the efficiency of risk-minimizing feature in reality; Secondly, we use the rolling-window approach to test the robustness of the strategy portfolios by applying a lager sample data. From our empirical analysis results, we find that the obtained strategies are efficient in the global financial crisis period, while the strategies obtained from the rolling window approach are not robust in the chosen period 2008–2019.

Keywords: portfolio optimization, minimum risk portfolio, Mean-Variance, Fuzzy Probability, Mean-MAD, Mean-CVaR, maximum drawdown, financial crisis, rolling window approach

JEL Classification: G11, G17
AMS Classification: 46N10, 62F03

1 Introduction

Financial portfolio optimization problems aim at finding the efficient investment strategies considering the optimal allocation of limited funds. Markowitz [7] proposed the Mean-Variance model which considered a portfolio's return and risk simultaneously. In financial markets, investors always balance the trade-offs between the risk and return of a portfolio according to their subjective preferences.

In this paper, the goal is to make the twofold efficiency tests of the minimum risk portfolio strategies which are obtained by the applied optimization models under the assumption of risk aversion. In section 2, we make the literature review based on the pioneers' studies. The review is constructed from different aspects, which includes the limitations and extensions of the classical Markowitz model as well as the review of the application of benchmarks in the previous portfolio optimization researches. In section 3, we describe the methodology applied in this paper with formulas. In section 4, we verify the efficiency of the obtained strategies by twofold tests based on the empirical analysis. The conclusion of this paper is made in section 5.

2 Literature Review

In the financial decision-making field, portfolio optimization involves the efficient investment strategies considering the optimal allocation of limited funds. On the premise of the liquidity and security of the investment funds, the common method of optimization is to balance the trade-offs between return and risk of a portfolio according to investors' subjective preferences. Markowitz [7] proposed the Mean-Variance model considering a portfolio's return and risk simultaneously, moreover, a higher degree of risk means a higher potential return.

In the last five decades, along with the considerations of real-life conditions and enhancements of algorithms, additional constraints have been developed to the early classical Markowitz model, for example, Fulga [3] presented an approach which incorporates the loss aversion preferences in the mean-risk framework, and the efficiency of the proposed approach is tested against the classical Markowitz model. Besides subjective preferences, more factors of portfolios are involved in later studies since 1950s, such as the liquidity of the portfolio, the transaction costs, the diversification degree of investments, the social responsibility and so on.

However, from later studies on portfolio optimization problems, the difficulties in the applications of the classical Markowitz model appeared. On one hand, there exists the computational complexity associated

[1] VŠB – Technical University of Ostrava, Department of Finance, Sokolská tř. 33, Ostrava 70200, Czech Republic, anlan.wang@vsb.cz.

with the covariance matrix, so, as an alternative to the risk measure in the Markowitz model, the variance is replaced with mean absolute deviation (henceforth MAD) in the model proposed by Konno and Yamazaki [5], which requires no computation or inversion of a covariance matrix, it solves a linear optimization rather than the quadratic optimization in Markowitz model. While in practical analysis, many cases studied by pioneers indicated that the Mean-Variance model and Mean-MAD model usually generated similar portfolio optimization strategies, and the common disadvantage for both these two models is that the investments to assets in a portfolio are not well diversified.

On the other hand, according to several existing studies of portfolio optimization problems, we find the future returns of stocks in a portfolio are usually obtained by estimating the historical data of the stocks, and the uncertainty associated with the returns of stocks or portfolio is described as randomness. However, recently, more studies highlight the importance of considering the uncertainty of returns as fuzziness. What's more, in Mansour et al., they pointed out that the investor is not able to define precisely his/her goals for stocks in his/her portfolio, and the fuzzy theory seems an alternative way to describe an imprecise or fuzzy environment, for example when the subjective behaviour of financial investors is taken into account. In Tanaka and Guo [10], as extensions to Markowitz model, they proposed two kinds of portfolio optimization models, the first model is based on fuzzy probabilities and aims to minimize the variance of the portfolio return while the other one utilizes possibility distributions and minimizes the spread of the portfolio return.

Although more and more enhanced portfolio optimization models are proposed in recent years, at the point of scientific precision, measuring the performance of the proposed optimization model is just one side of the work, for another side, it's significant to involve a benchmark in the analysis to verify the efficiency of the models, see Kresta and Wang [5]. Various benchmarks can be found in scientific literature, however, there are few groups of benchmarks, which are generally applied. The most commonly applied benchmark is the 1/N strategy, which is easy to implement because it does not rely on estimations of the asset returns, the components assets of the naive portfolio are invested at equal weights. DeMiguel et al. [2] evaluated and compared the performance of several optimization methods with respect to the performance of the 1/N strategy, they found that the effect of estimation error on return probability distribution is large in those optimization models, but this type of error can be avoided by using the 1/N weights. Another commonly used benchmark is the indices, for example, Solares, et al. [9] used the prices of Dow Jones Industrial Average index (DJI) as the benchmark in their research. In the portfolio optimization researches, the classical Mean-Variance model or alternatively Mean-VaR model are also usually applied as benchmarks to test the efficiency of the new proposed approaches, see. e.g. Fulga [3], Rankovic et al. [8], Lwin et al. [6] or Babazadeh and Esfahanipour [1]. In this paper, we apply the 1/N strategy and the DJI index as the benchmarks.

In our analysis, to verify the efficiency of the obtained strategy portfolios, the performances of the obtained portfolios are compared with those of the benchmarks, what's more, to make the verification more persuasive, 10,000 random-weights portfolios are also generated to make the hypothesis tests to compare the performances.

3 Portfolio Optimization Methodology under Mean-Risk Framework

3.1 Mean-Variance Model

In the Mean-Variance model, we analyse the inter-relationship between mean and variance of a portfolio's returns in a certain period. We denote x_i as the weight of asset i in a portfolio investment, and short sales are excluded, so the values of x_i satisfy $x_i \geq 0$ for all assets. We denote $E(R_i)$ as the expected return of asset i, and in our case we suppose that the expected stock return is identical to the average of the real returns within historical period, then the expected return of a portfolio $E(R_p)$ can be calculated as follow,

$$E(R_p) = \sum_{i=1}^{N} x_i \cdot E(R_i) = x^T \cdot E(R) \tag{1}$$

where $x = [x_1, x_2, \dots, x_N]^T$, $E(R) = [\,E(R_1), E(R_2), \dots, E(R_N)]^T$, the sum of x_i in a portfolio equals to 1, and the $E(R_p)$ is the weighted average of $E(R_i)$. The variance of a portfolio's returns in a certain period, which is denoted as ${\sigma_p}^2$, is regarded as the risk measure of a portfolio in the Mean-Variance model. ${\sigma_p}^2$ is calculated by the $N \times N$ covariance matrix $\boldsymbol{Q} = [\sigma_{i,j}, i = 1, 2, \dots, N, j = 1, 2, \dots, N]$ for all component asset pairs (i, j) in a portfolio, the calculation of ${\sigma_p}^2$ is shown in equation (2), where the σ_p in equation (3) is the standard deviation (henceforth STD) of a portfolio's returns.

$$\sigma_p{}^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i \cdot \sigma_{i,j} \cdot x_j = x^T \cdot \boldsymbol{Q} \cdot x \tag{2}$$

$$\sigma_p = \sqrt{\sigma_p{}^2} \tag{3}$$

Fuzzy Probability Strategy

Tanaka et al. [10] proposed a Fuzzy Probability model, which combines the probability distribution of stocks returns in Markowitz's model with fuzzy theory, in our analysis, we apply this model to handle uncertainty in the stocks returns probability distribution. In the Fuzzy Probability model, we consider not only the historical stocks returns $\{r_{i,m}, i = 1, \dots, N, m = i, \dots, M\}$, but also possibility grades $\{h_m, m = i, \dots, M\}$, which reflects a similarity degree between the future state of the stock market and the state of m_{th} sample offered by experts, and according to Tanaka et al. [10], we define h_m as,

$$h_m = 0.1 + 0.3 \cdot (m-1)/(M-1) \tag{4}$$

these grades are applied to determine the fuzzy expected returns of stocks and fuzzy covariance matrix for a given data. We denote $E(r_i^F)$ as the expected return of stock i under Fuzzy Probability strategy, if given the historical stocks returns $r_{i,m}$ and possibility grades h_m, the fuzzy weighted expected return of stock i can be calculated as equation (5), where m is the number of historical observations of returns of stock i. The fuzzy weighted covariance matrix $\boldsymbol{Q^F} = [\sigma_{i,j}^F, i = 1, \dots, N, j = 1, \dots, N]$ can be defined by equation (6), we can see when the given data have the same possibility grades, the Fuzzy Probability model is just equivalent to the Markowitz's model.

$$E(r_i^F) = \frac{\sum_{m=1}^{M} h_m \cdot r_{i,m}}{\sum_{m=1}^{M} h_m} \tag{5}$$

$$\sigma_{i,j}^F = \frac{\sum_{m=1}^{M}\left(r_{i,m} - E(r_i^F)\right) \cdot \left(r_{j,m} - E\left(r_j^F\right)\right) \cdot h_m}{\sum_{m=1}^{m} h_m} \tag{6}$$

For both Bayesian Strategy and Fuzzy Probability Strategy, the estimates such as $E(r_i^F)$ and $\boldsymbol{Q^F}$ can be directly applied in the portfolio optimization model under the Mean-Variance framework.

3.2 Mean-MAD Model

Comparing with the Mean-Variance model, the risk measure of the portfolio is replaced by mean absolute deviation of the portfolio's returns in the Mean-MAD model, the calculation of mean absolute deviation is shown in equation (7), where T is number of observations, $R_{i,t}$ is the return of asset i for each time t.

$$MAD = \frac{\sum_{t=1}^{T}\sum_{i=1}^{N} x_i \cdot \left|R_{i,t} - E(R_i)\right|}{T} \tag{7}$$

3.3 Mean-CVaR Model

Under the mean-risk framework, the Mean-CVaR model focuses on the measure of the expected shortfall for a portfolio. As we know, VaR is defined as the worst-case loss associated with a given probability and a time horizon. However, rather than the application of VaR, the CVaR which indicates the expected loss under the condition of exceeding VaR is applied as the risk measure in our analysis. The CVaR for a portfolio is defined as follow,

$$CVaR_\alpha(x) = \frac{1}{1-\alpha}\int_{f(x,y)\geq VaR_\alpha(x)} f(x,y)p(y)dy \tag{8}$$

where x is a portfolio satisfies $x \in X$ (X is the set of available portfolios), α is the probability level such as that $0 < \alpha < 1$, it implies that the probability of the portfolio's returns falling below the value $VaR_\alpha(x)$ is $1-\alpha$, and in our analysis we set $\alpha = 95\%$. $f(x,y)$ is a loss function for a portfolio x and asset return y, $p(y)$ is the probability density function for asset return y, VaR_α is the VaR of portfolio x at probability level α and it can be defined as follow.

$$VaR_\alpha(x) = \min\{\gamma: \Pr[f(x,Y) \leq \gamma] \geq \alpha\} \tag{9}$$

3.4 Minimum Risk Portfolio

The minimum risk portfolio is the portfolio which has the minimum risk when the risk is measured by ${\sigma_p}^2$, MAD and $CVaR_\alpha(x)$ in our case. The minimum risk portfolios are obtained by the following program.

$$\begin{gathered}\text{minimize } {\sigma_p}^2 \text{ (or } MAD \text{ or } CVaR_\alpha(x)) \\ \text{subject to} \\ \sum_{i=1}^{N} x_i = 1 \\ x_i \geq 0, i = 1, \dots, N\end{gathered} \tag{10}$$

3.5 Hypothesis Test

We verify the efficiency of strategies under the portfolio optimization models by generating random-weights portfolios and making hypothesis tests. The risk measures applied in our analysis are STD and MAD of the portfolio returns, the CVaR of the portfolio, and the maximum drawdown (henceforth MDD) which indicates the maximum loss from a peak to a trough of a portfolio investment's wealth evolutions is also applied.

Random-weights portfolio, as it literally means, the weights of assets in each random portfolio are generated randomly, in our case we set up 10,000 random portfolios, and in each portfolio the sum of the weights x_i equals to 1. For the generation of random weights, we choose $y \in [0,1]^{N-1}$ uniformly by means of $N-1$ uniform reals in the interval [0,1], and sort the coefficients so that $0 \leq y_1 \leq \cdots \leq y_{N-1}$, then x_i can be shown as in equation (11), because we can recover the sorted y_i by means of the partial sums of the x_i, the mapping $y \to x$ is $(N-1)!$ to 1.

$$x_i = (y_1, y_2 - y_1, y_3 - y_2, \dots, y_{N-1} - y_{N-2}, 1 - y_{N-1}) \tag{11}$$

As we know that a hypothesis test relies on the method of an indirect proof, that is, to prove the hypothesis that we would like to demonstrate as correct, we show that an opposing hypothesis is incorrect. In our case, the strategy portfolios under the optimization methods are more likely to be demonstrated as efficient, so according to the rule of hypothesis tests, we can make the null hypothesis and alternative hypothesis as follows:

$$\text{null hypothesis} — H_0: RMV_s = E(RMV_r),$$

$$\text{alternative hypothesis} — H_A: RMV_s < E(RMV_r)$$

where RMV_s is the risk measure value of the strategy portfolio, and $E(RMV_r)$ is the expected value of the risk measure of the random-weights portfolio. In our hypothesis test, the p-value is the proportion of the random-weights portfolios which meet the condition $E(RMV_r) < RMV_s$ in the total number of random-weights portfolios[2]. We set the significance level to 10%, when p-value is less than 10%, then we reject H_0, which means the performance of the strategy portfolio is better than that of the random-weights portfolio, so the strategy is efficient; when p-value is no less than 10%, then we fail to reject H_0, which means the performance of the strategy portfolio makes no difference from that of the random-weights portfolio, so the strategy is inefficient in this case.

4 Empirical Results

4.1 Global Financial Crisis

The chosen sample data is the daily closing prices of the components of Dow Jones Industrial Average index (DJI). In this case, there are 28 component stocks included, the two missing stocks is the stock of Visa Inc. and the stock of Dow Inc. due to the incomplete data in the chosen period. The chosen period in this case covers the 2007–2009 global financial crisis. We divide the whole sample data into two parts. The in-sample part is from January 3rd, 2006 to August 10th, 2007 and the out-of-sample part is from August 13th, 2007 to March 2nd, 2009.

[2] The lower value of the risk measure, the better performance of the portfolio.

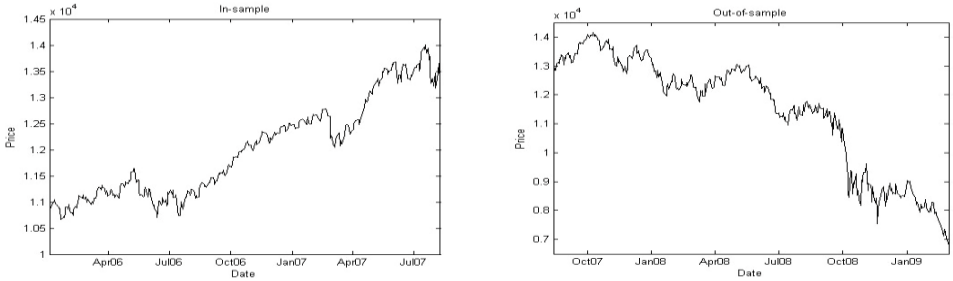


Figure 1 Historical price evolutions of DJI

In *Figure 1* we show the price evolutions of DJI in these two periods, we can see in the in-sample period the price of DJI shows an increasing trend, however, in the out-of-sample period the price kept decreasing resulting from the influence of the global financial crisis, so, in this sense, obtaining the minimum risk portfolios from the in-sample period and then testing the efficiency of them in the out-of-sample period has a practical significance.

The minimum risk portfolios are obtained based on the models introduced in section 3. For each minimum risk portfolio, it lies on the lowest point of the efficient frontier of the corresponding optimization model. We make the back-tests of the obtained minimum risk portfolios by applying the out-of-sample data, and we show the performances of them in *Table 1*. In *Table 1*, for the Fuzzy Probability extension, we find the risk performance measured by STD, MAD and CVaR are better than those of the classical Mean-Variance portfolio. What's more, comparing the performances of all the minimum risk portfolios we obtained, on one hand, we find the values of mean return of each portfolio are negative due to the decreasing prices during the out-of-sample period; on the other hand, concerning the risk in the out-of-sample period, we find that the risk is the lowest for minimum CVaR portfolio measured by all four risk measures, which is surprising in the case of STD and MAD measures because the other strategy portfolios minimize specifically these two measures in the in-sample period. As for the risk measures, we find that the strictest risk measure is the MDD, because when the risk is measured by MDD, except for the minimum CVaR strategy, the p-values of the hypothesis tests of all the other strategies indicate that these strategies are inefficient.

	Mean daily return	MDD (p-value)	STD (p-value)	MAD (p-value)	CVaR (p-value)
Minimum Variance	-0.07%	35.29% (**)	1.67% (****)	1.08% (****)	3.86% (****)
Minimum Variance - FP	-0.08%	36.40% (ns)	1.65% (****)	1.06% (****)	3.80% (****)
Minimum MAD	-0.07%	35.29% (**)	1.70% (****)	1.10% (****)	3.98% (****)
Minimum CVaR	-0.05%	32.30% (****)	1.59% (****)	1.05% (****)	3.63% (****)
DJI index	-0.15%	52.25%	2.11%	1.44%	5.14%
Naive strategy	-0.11%	44.94%	2.20%	1.48%	5.28%

Table 1 Minimum-risk portfolios' performances in the out-of-sample period
(ns p-value>0.05, * p-value<0.05, ** p-value<0.01, *** p-value<0.001, **** p-value<0.0001)

According to the results, we also know that the average value of the 10,000 random portfolios' mean returns is –0.11%, which underperforms the minimum risk portfolios' performances. What's more, the number of assets selected in each minimum risk portfolio is shown in Table 2, we can see the investment is not well diversified in each minimum risk portfolio considering there are 28 component stocks included in our analysis.

	Minimum Variance	Minimum MAD	Minimum CVaR	Minimum Variance - FP
Number of assets	13	15	8	14

Table 2 The number of assets selected in minimum risk portfolio

4.2 Rolling Window Approach

In order to test the robustness of the obtained strategy portfolios to the changes of the periods, we perform the tests on rolling one-year window basis. The chosen dataset in this case is the daily closing prices of 28 DJI components, and the missing components are Dow Inc. and Visa Inc. due to the incomplete data in the chosen period. We always take 3 years (750 days) as the in-sample period and 1 year (250 days) as the out-of-sample period, then we move the start of the out-of-sample one-year period day by day from December 24th, 2008 to January 2nd, 2019.

The models applied in this case are Mean-Variance model, Mean-MAD model and Mean-CVaR model. We show the results of the p-values under the rolling window approach in *Figure 2*. Firstly, we find the MDD and CVaR are the strictest risk measures, because for these two measures, there are long periods in which all the three strategies did not minimize the risk in the out-of-sample period efficiently. Secondly, the minimum MAD strategy performs more efficiently in the out-of-sample period with lower p-values' evolutions comparing to other two strategies, while the minimum CVaR strategy does not provide good results in the out-of-sample period. Last but not least, minimizing the chosen risk measure in the in-sample period does not guarantee its lowest value in the out-of-sample period (e.g. in 2016-2017 we get the best out-of-sample STD by minimizing in-sample MAD rather than minimizing in-sample STD).

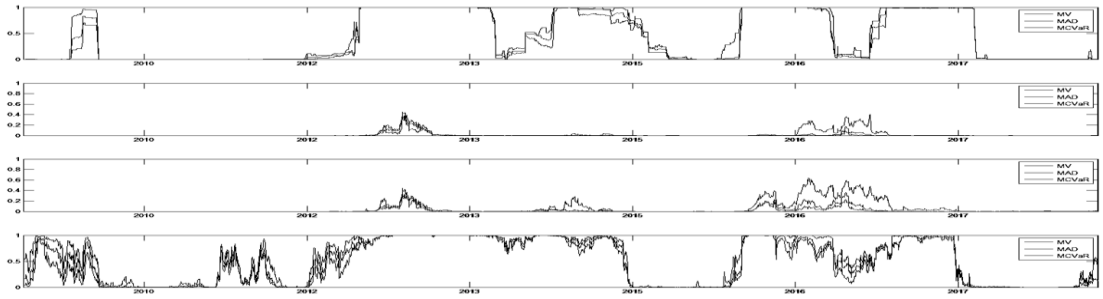


Figure 2 Rolling window p-values (statistics from top to down: MDD, STD, MAD, CVaR)

5 Conclusion

The goal of this paper is to make the twofold efficiency tests of the obtained minimum risk portfolio strategies under the assumption of risk aversion. From the analysis of the case which covers the global financial crisis, in the out-of-sample period, we find that the Fuzzy Probability extension partly improves the performance of the classical strategy, and we also find that the minimum CVaR portfolio has the best performance even when the portfolio risk is measured by all four applied risk measures. However, the results show that the investment is not well diversified in each minimum risk portfolio considering there are 28 component stocks included in our analysis. Based on the rolling window analysis, we conclude that the minimum MAD strategy has the best performance considering the rolling out-of-sample periods, however, overall, the results show that the strategies obtained from the rolling window approach are not robust in the chosen period 2008–2019.

Acknowledgements

The author was supported through the Czech Science Foundation (GACR) under paper 18-13951S and moreover by SP2020/11, an SGS research paper of VSB-TU Ostrava. The support is greatly acknowledged.

References

[1] BABAZADEH, H. & ESFAHANIPOUR, A. (2019). A novel multi period mean-VaR portfolio optimization model considering practical constraints and transaction cost. *Journal of Computational and Applied Mathematics, 361*, 313–342.

[2] DEMIGUEL, V., GARLAPPI, L., & UPPAL, R. (2009). Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy. *Review of Financial Studies*, 22, 1915–1953.

[3] FULGA, C. (2016). Portfolio optimization under loss aversion. *European Journal of Operational Research, 251*, 310–322.

[4] KONNO, H. & YAMAZAKI, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock-market. *Management Science, 37*, 519–531.

[5] KRESTA, A. & WANG, A. (2019). Efficiency test as the benchmark for minimum-risk portfolio optimization strategies. In Zadnik Stirn L. et al. (Eds.), *Proceedings of the 15th International Symposium on Operational Research* (pp. 417–422). Slovenian Society Informatika Press, Bled, Slovenia.

[6] LWIN, K. T., QU, R., & MACCARTHY, B. L. (2017). Mean-VaR portfolio optimization: A nonparametric approach. *European Journal of Operational Research, 260*, 751–766.

[7] MARKOWITZ, H. (1952). Portfolio selection. *The Journal of Finance, 7*, 77–91.

[8] RANKOVIC, V., DRENOVAK, M., UROSEVIC, B., & JELIC, R. (2016). Mean-univariate GARCH VaR portfolio optimization: actual portfolio approach. *Computers & Operations Research, 72*, 83–92.
[9] SOLARES, E., COELLO, C., FERNANDEZ, E., & NAVARRO, J. (2019). Handling uncertainty through confidence intervals in portfolio optimization. *Swarm and Evolutionary Computation, 44*, 774–787.
[10] TANAKA, H., GUO, P., & TURKSEN, B. (2000). Portfolio selection based on fuzzy probabilities and possibility distributions. *Fuzzy Sets and Systems, 111*, 387–397.

Can Experts' Knowledge in eNS Inspire Inspire Efficient Classification of Potential Debtors?

Aleksandra Wójcicka-Wójtowicz[1]

Abstract. The paper presents the results of a case-study on information perception in evaluation of potential debtors. It concerns the phase prior to the decision-making stage. The experts were asked to build a scoring system representing their preferences (e.g. qualitative features, soft factors etc.). The case-study is conducted in Inspire – electronic Negotiation System (eNS) which covers a pre-negotiation phase using a combination of Simple Additive Weighting method (SAW) and Ordered Fuzzy Numbers (OFN). The conducted analysis proves the utility of experts' knowledge in the pre-decision process of debtors' evaluation and classification.

Keywords: negotiation support, ordered fuzzy number; preference analysis, Simple Additive Weighting method

JEL Classification: B4, C9, C02, C38, G40
AMS Classification: 03B52, 03E72, 94D05, 91B08

1 Introduction

Credit risk is just one sub-group of banking risk but frequently it decides about default or survival. Banks cannot unreservedly grant funding to any potential borrower. The final decision must be preceded by the examination of a future debtor and the entity must be classified into an appropriate group. Various methods (internal ratings, standard credit risk models, neural networks, decision trees etc.) can be used. Although most of the implemented methods are quantitative, lately there has been a strong trend to include also qualitative factors in a form of experts' knowledge. However subjective those factors might seem, often they capture subtle nuances and complex processes in a way unattainable for quantitative models.

Lately, the interest in experts' knowledge has visibly increased. Its utilization is clear when one considers the steps of credit risk evaluation and process of granting or rejecting funding by the bank. Otherwise, how can a fact be explained that the final credit granting/rejecting decision does not based solely on the results of financial (quantitative) analysis but on the experience of Credit Committee members (experts) who, of course, can follow the recommendations of the analysts or reject them. It clearly proves that the final decision is a result of the knowledge, experience and preferences of experts.

There are various approaches to capture the above elements [11]. Among those we can mention linear and non-linear credit risk models, neural networks, decision trees and many hybrids. The modifications appear as a result of the shortcomings of existing models. The significance of experts' knowledge and experience, as well as other qualitative factors in credit risk assessment and debtors' classification, are recognized as increasingly influential and helpful in decision-making process. One of the recently increasingly utilized pair is a Simple Additive Weighting method (SAW) combined with the Ordered Fuzzy Numbers (OFNs) [10]. SAW is also used in eNS Inspire.

The paper consists of 5 sections. The first, being the Introduction, is followed by Section 2 which briefly presents the nature of negotiations in credit risk assessment process. In Section 3 the main assumptions of implemented methods and approaches are defined. It also covers the basic description of utilization of experts' knowledge in eNS Inspire. Section 4 presents the conducted experiment and Section 5 concludes the paper.

2 Negotiations in credit risk assessment process

Negotiations in credit risk assessment process are very specific as the bank is considered to be the dominant party as it is up to the bank whether the financing will (or not) be granted. However, the nature of negotiations, perceived as a kind of bargaining, comes from the main idea of a struggle between at least two parties of opposite objectives.

[1] Poznań University of Economics and Business/Institute of Informatics and Quantitative Economics, Department of Operations Research, Al. Niepodległości 10, 61-875 Poznań, aleksandra.wojcicka-wojtowicz@university.edu.

There are many negotiating systems amongst which we can list electronic Negotiation System (eNS) Inspire [4]. It is a system supporting multi-issue negotiations. It provides many options of the process, e.g. assessment of offers, management of communication and graphical display in a form of charts and graphs, but the most important issue is that it allows for including specific preferences concerning personal experience of the negotiating parties. In case of any negotiation and in this research, in case of debtors' assessment, the personal experience, preferences and knowledge play a significant role in the decision-making process. It is due to the fact that the final decision is taken at the meeting of a credit assessment committee, consisting of experts (higher level managers) who make their decisions basing on their personal, professional experience. Their main role is to use their ability to look at the future projections of the borrower's business and not just their past performance. Any underwriting agreements, financial projections and the health of the borrower's industry are all very important, as they will be leading indicators of potential volatility in loan payments, however, assessed features can frequently be conflicting or excluding one another. During negotiation process the participants must establish their objectives, preferences and potential (expected) outcome which may be modified in the process. Those assumptions are not revealed to the opposite party. The significant part is to define the factors which influence the negotiations proceedings and the whole negotiation context, e.g. previous experience (deadlocks or disputes).

3 Methodology

The conducted research is a combination of approaches and methods, namely Simple Additive Weighting (SAW) method and Linguistic Approach [11].

3.1 Linguistic approach

A linguistic approach is a way of expression of the imprecision occurring as a result of utilisation of natural language. The human perception and articulation of preferences regarding various issues tends to be inaccurate due to strengthening or weakening the main assessment by additional adjectives and adverbs such as more/less, better/worse etc. The process of linguistic approach implementation starts with the determination of granularity, i.e., the cardinality of the linguistic term set used for showing the information. The imprecision granularity indicates the capacity of distinction that may be expressed. The knowledge value is increasing with the increase in granularity. The typical values of cardinality used in the linguistic models are odd ones, usually between 3 and 13. It is worth to note that the idea of granular computing comes from Zadeh [12] who wrote fuzzy information granulation underlies the remarkable human ability to make rational decisions in an environment of imprecision, partial knowledge, partial certainty and partial truth. Also, Yao [11] pointed out that the consideration of granularity is motivated by the practical needs for simplification, clarity, low cost, approximation… For review variety of application linguistic models in decision-making see for example [2]. In general, [2], any linguistic value is characterized by means of a label with semantic value. The label is an expression belonging to a given linguistic term set. Finally, a mechanism of generating the linguistic descriptors is provided. In credit risk assessment, all linguistic assessments are linked with Tentative Order Scale (TOS) given as a sequence

$$TOS = \{\text{Bad, Average, Good}\} = \{C, B, A\} = \{V_1, V_2, V_3\}. \tag{1}$$

Any element of TOS is called a reference point and can be enlarged by intermediate values.

3.2 Simple Additive Weighting Method

The Simple Additive Weighting method (SAW) is used to facilitate a multi-criteria evaluation problem. The most significant part in this process is the determination of weights which include experts' personal and professional experience and preferences. In eNS Inspire it is achieved by a specific scoring approach as Inspire offers a SAW-based tool that helps negotiators (hereinafter also called experts) to analyze their preferences in a stage of pre-negotiations which establishes their priorities concerning the negotiation template.

SAW is a scoring method based on the concept of a weighted average of criterion ratings in which the individual criterion ratings can be expressed by Trapezoidal Ordered Fuzzy Numbers (TrOFNs). If such is the case then SAW method should be equipped with scoring function determined on the space $\mathbb{K}_{Tr}^{n} = \mathbb{K} \times \mathbb{K} \times \ldots \times \mathbb{K}$.

Roszkowska and Kacprzak [8] tentatively introduced the Oriented Fuzzy SAW (OF-SAW) method with criterion ratings expressed trapezoidal Kosiński's numbers. In [6] OF-SAW method is modified in a way that it

is compatible with the revised theory of ordered FNs [5]. In this case, criterion ratings are given as Trapezoidal Ordered Fuzzy Numbers (TrOFNs). To evaluate a borrower characterized by attributes record $\mathcal{A} \in \mathbb{A}$ where $\mathbb{A}$ is an anticipated set of potential borrowers, a 7-steps OF-SAW is implemented (see more in [8]).

The way the experts' knowledge can be utilized is via a scoring system implemented in Inspire. A certain number of points is available to the expert to attribute them to all distinguished issues. This way the importance of each issue is established and, in turn, also its individual weight. Each of the issues must obtain the score between 0 and the issue weight. Then Inspire presents the list of selected complete packages with total scores. It is, however, easy to implement changes in the results of an initial scoring system as with each change Inspire recalculates the rates of issues and options.

There can be also a graphical visualization of preferences which takes different forms. The most popular are circles. However, in case of circles, the size and radiuses of the circles are important as they indicate the significance of issues and options. Unfortunately, that being the case, may cause a few problems. It has been analyzed in Brinton [1] that circles as information presentation can cause the reader to misread the importance of the data by under- or overestimating the ratio between the area and the radius. There are also many other methods of visualization of the information in negotiating problems (e.g. [4]).

This paper focuses on analyzing the negotiation process prior to making the final decision of building a scoring system by means of SAW by the experts – utilizing their experience, knowledge and preferences.

4 Case study

A significant part of the case study is including the appropriate issues in the proposed template. In the chosen problem there are five chosen general issues, considered to be the most important, were implemented with predefined salient options originating from the above mentioned linguistic scale. The general issues are the result of a number of criteria indicated in advance by the members of the credit committee and gathered in specific groups. The issues (in relation to a potential borrower/debtor) with the available options are presented in Table 1.

Issues	Options
general risk level	high; moderate; low
diversification	extensive; moderate; minimal
prospects / projections	good; moderate; bad
management	exceptional; acceptable; unacceptable
range of operations	wide; average; minimal

Source: own elaboration

Table 1 Issues and options

The conducted experiment heavily relies on the preferences of experts-negotiators (2) who at this stage must attribute each individual issue with a number of scoring points (initially there are 100 points to be distributed between the issues). The experiment concerns two companies of: 1) construction, 2) pharmaceutical industry. The initial points for both experts and entities are presented in Table 2.

	company A		company B	
Issues	expert 1	expert 2	expert 1	expert 2
general risk level	35	40	25	30
diversification	15	15	20	15
prospects / projections	25	20	30	25
management	5	10	10	15
range of operations	20	15	15	15

Source: own elaboration – case study basing on experts' choices

Table 2 Distribution of rating points.

It is worth stressing that those groups include a number of factors and each expert can interpret a specific group differently, depending on their own personal experience. For instance, the issue group of general risk

level can include not only level of credit risk but also the level of the market, trade, suppliers or customers risk.

After rating the issues, the options in each issue must also be rated similarly. In the Inspire system, for each issue at least one option must be assigned the maximum rating for the issue and at least one option must be assigned a rating of zero (www1). The assessment of options for company A is presented in table 3 (expert 1) and table 4 (expert 2).

Issues	Options		
general risk level (max 35)	high (0)	moderate (20)	low (35)
diversification (max 15)	extensive (15)	moderate (10)	minimal (0)
prospects / projections (max 25)	good (25)	moderate (25)	bad (0)
management (max 5)	exceptional (5)	acceptable (5)	unacceptable (0)
range of operations (max 20)	wide (20)	average (15)	minimal (0)

Source: own elaboration basing on eNS Inspire template

Table 3 Rating options – expert 1

Issues	Options		
general risk level (max 40)	high (0)	moderate (30)	low (40)
diversification (max 15)	extensive (15)	moderate (10)	minimal (0)
prospects / projections (max 20)	good (20)	moderate (10)	bad (0)
management (max 10)	exceptional (10)	acceptable (10)	unacceptable (0)
range of operations (max 15)	wide (10)	average (10)	minimal (0)

Source: own elaboration basing on eNS Inspire template

Table 4 Rating options – expert 2

Similar calculations were carried out for the company 2.

Having experts' ratings for each issue and each option, eNS Inspire calculates ratings for complete packages that are the subject of consideration. A single unit of 'a package' consists of all five criteria (options), for example:

- low general risk level,
- extensive diversification,
- good prospects,
- exceptional management and
- wide range of operations'

is one complete package. Individual packages and their ratings describe the expert's particular preferences. It can be observed that in case of both experts there are criteria which individual options are indifferent, for instance good or moderate prospects (expert 1), exceptional or acceptable management (expert 1 and 2), wide or average range of operation (expert 2). That finding, in turn, can result in decreasing the number of options as a conclusion can be made that experts do not differentiate those two states.

In the next step the list of selected complete packages is presented with the global scores.

What is really important is the fact that in the course of negotiation, in the process of reaching agreement, experts can change and adjust their scores.

The confrontation of experts' opinions and expectation leads to adjustment movements. Each expert needs to reconsider their original choices and evaluate the issues and options. In a perfect world expert should have similar preferences (considering the same entity) and their expectations and scoring should be alike. However, the whole process really can have only two outcomes: a compromise or a deadlock (no solution). The length of the process depends only on the experts' willingness to achieve a success (a compromise) but on the other hand this compromise cannot be imposed on them.

Despite their differences, in the carried out experiment for company 1 (negotiations resembling a discussion during the credit committee meeting) after 3 rounds of negotiations the experts reached the following compromise – a package of:

- low general risk level,
- extensive diversification,
- moderate prospects,
- acceptable management and
- wide range of operations.

For company 2 (negotiations took 5 rounds) the experts reached the following compromise – a package of:

- moderate general risk level,
- extensive diversification,
- good prospects,
- acceptable management and
- average range of operations.

Due to the reached compromise, the debtor would be classified to the group of potential debtors (borrowers). In case of a deadlock, the application would be eventually rejected.

An important fact must be stressed – as the whole process is a negotiation, its main objective is to reach a specific compromise. The utilization of experts' knowledge and experience is invaluable. There is no *good* or *bad* solution in the sense of e.g. error types. The reached solution is unique.

5 Conclusions

Reaching a final decision is usually a complex and extensive process. It relies on quantified data and also on human preferences and experience. Those, in turn, are difficult to quantify – usually inaccurate and imprecise.

However, methods and tools of negotiations, such as Inspire, can be useful by aiding the experts to present and process their preferences and knowledge. It can be utilized not only to classify the potential debtors but also to rank them, not to mention ranking and rating the chosen options which are expressed by linguistic (imprecise) approach.

The fact that during the process of negotiations the number of possible options could be decreased to some major or main areas also facilitates the decision-makers, helping them to focus on the most significant issues. Using the experts' knowledge allows for ranking the options but consequently also the borrowers. This can lead to classifying them into a specific, individual credit rating group connected with ratings similar to those of rating agencies.

The further ongoing research aims at implementing (in Inspire) ordered fuzzy numbers (as the technique to cope with the imprecision and inaccurateness) and, furthermore OF-SAW.

References

[1] BRINTON, W. C. (1914). *Graphic methods for presenting facts*. The Engineering magazine company, 1914.

[2] HERRERA, F., & HERRERA-VIEDMA, E. (2000). Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems*, *115*(1), 67–82. DOI: 10.1016/s0165-0114(99)00024-x.

[3] KERSTEN, G. & NORONHA, S. (1999) WWW-based negotiation support: design, implementation and use. *Decis Support Sys 25(2)*, 135–144.
[4] MIETTINEN, K. (2014) Survey of methods to visualize alternatives in multiple criteria decision making problems. *OR Spectr* 36:3–37.
[5] PIASECKI, K. (2018) Revision of the Kosiński's Theory of Ordered Fuzzy Numbers. Axioms 7(1). DOI: 10.3390/axioms7010016.
[6] PIASECKI, K.; ROSZKOWSKA, E. (2018), On Application of Ordered Fuzzy Numbers in Ranking Linguistically Evaluated Negotiation Offers. *Adv. Fuzzy Syst. 1*, 1–12, DOI: 10.1155/2018/1569860.
[7] PIASECKI, K., ROSZKOWSKA, E., & ŁYCZKOWSKA-HANĆKOWIAK, A. (2019 a). Simple Additive Weighting Method Equipped with Fuzzy Ranking of Evaluated Alternatives, Symmetry 11(4). doi:10.3390/sym1104048.
[8] ROSZKOWSKA, E., KACPRZAK, D. (2016) The fuzzy SAW and fuzzy TOPSIS procedures based on ordered fuzzy numbers. Inf. Sci. 369:564–584. DOI: 10.1016/j.ins.2016.07.044.
[9] WÓJCICKA-WÓJTOWICZ, A. & PIASECKI, K. (2019). A Scale of Credit Risk Evaluations Assessed by Ordered Fuzzy Numbers, *SSRN Electronic Journal*, September, 1–7. DOI:10.2139/ssrn.3459822.
[10] WÓJCICKA-WÓJTOWICZ, A. (2020). How to Include Experts' Imprecision in Credit Risk Assessment? SZKUTNIK, W., SĄCZEWSKA-PIOTROWSKA, A., HADAŚ-DYDUCH, M., ACEDAŃSKI, J. (eds.), *13th International Scientific Conference "Analysis of International Relations 2020. Methods and Models of Regional Development. Winter Edition*". Conference Proceedings, Publishing House of the University of Economics in Katowice, 234–247.
[11] YAO, Y. Y. (2004) Granular computing, Computer Science (Ji Suan Ji Ke Xue), 31, 1–5.
[12] ZADEH, L. (1975). The concept of a linguistic variable and its application to approximate reasoning—I. *Information Sciences*, *8*(3), 199–249. DOI: 10.1016/0020-0255(75)90036-5s.

Selected Methods of Categorical Data Analysis and Their Application in Consumer Behaviour Research

Martina Zámková[1], Luboš Střelec[2], Stanislav Rojík[3],
Martin Prokop[4], Radek Stolín[5]

Abstract. This paper explores the methods of categorical data analysis, with special focus on their pros and cons and their possible application in the research of Czech consumers' shopping habits. To that end, a large survey mapping shopping preferences of Czech consumers (mainly the shopping behaviours of different genders in relation to their income, education, age, place of residence etc.) has been used as a source of data. The survey results were analyzed using the contingency tables analysis, including the Pearson's chi-squared test of independence in contingency tables. The dependence intensity was identified via Pearson's Contingency Coefficient and the impact of individual categories was assessed with the post-hoc residual test. The influence of several combined factors on the payment method was tested through logistic regression; the model parameters were estimated by the Maximum Likelihood Estimation and the ROC curve allowed for the assessment of the model's quality. Every method proved a statistically significant dependence of the most frequent payment method on the education and age of respondents, number of household members and the size of respondents' place of residence.

Keywords: logistic regression, Pearson's chi-squared test, ROC curve

JEL Classification: C30
AMS Classification: 62H99

1 Introduction

Consumer payment methods have been evolving in time. Up until the mid-20th century, there has basically been only one option to carry out payment transactions. The second half of the 20th century was marked by technological development and saw a gradual introduction of cashless – card-based – payment transactions. The turning point came in the 1990s with the rise of the Internet, when card payments globally started to become an equal alternative to cash payments [7]. The increasing popularity of card transactions (as opposed to cash payments) may be traced back to their technological advantages perceived by customers. Let us mention, for example, easier and better access to assets (liquidity) [6]. Then came another important milestone in the history of non-cash payments and the efficiency of credit and debit cards in transactions – the *e-commerce*, i.e. on-line shopping and on-line payments in general [8].

Considering the two major payment card types (debit and credit), the research shows a decreasing trend in the use of credit cards, while there has been an increase in debit card payments in recent years [3]. However, card-based transactions are not all positive. Compared to cash payments, credit and debit cards have been confirmed to pose an added potential danger – increased consumer spending [4].

Recently, due to the Covid-19 crisis and its economical and logistical implications, there has been a global increase in card-based transactions (as opposed to cash payments) and it is expected that the market share and international and domestic importance of card transactions would continue to go up. This paper therefore aims to draw a comparison between frequently used payment methods, with special focus on cash and debit and credit card payments. The research results allowed us to explore and statistically express the dependencies considering the most frequent shopping payment method and selected consumers' socio-demographic characteristics (e.g. education, age, household size etc.).

[1] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 58601 Jihlava, martina.zamkova@vspj.cz.
[2] Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 61300 Brno, lubos.strelec@mendelu.cz.
[3] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 58601 Jihlava, stanislav.rojik@vspj.cz.
[4] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 58601 Jihlava, martin.mrokop@vspj.cz.
[5] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 58601 Jihlava, radek.stolín@vspj.cz.

2 Methods and Materials

A large survey (757 respondents) mapping shopping preferences of Czech consumers has been used as a source of primary data. The quota sampling questionnaire has been sent out in early 2020. The survey data are categorical and thus processed exclusively with statistical methods suitable for the work with lexical variables.

Where the response variable proves to be categorical, logistic regression is used. Explanatory variables may be continuous as well as categorical. In a binary logistic regression, the response variable Y is dichotomous with the values of 1 and 0, indicating the presence or absence of an event A. The probability of the A's occurrence is $p = P(Y = 1)$. The probability of A occurring in certain conditions defined by X is $p(X) = P(Y = 1|X)$, and at the same time equals the mean value $E(Y|X)$. The probability is defined as the odds ratio of A occurrence and non-occurrence, i.e. odds $= p/(1 - p)$. The logarithm of the odds (logit) defined by logit $= \ln(p/(1 - p))$ is therefore linearly dependent on the conditions given by X, the sought regression model parameters shall be b_i. The probability of the A's occurrence, the odds and the logarithm of the odds are denoted by the following equations:

$$p(A) = \frac{1}{1 + \exp\left(-(b_0 + b_1x_1 + b_2x_2 + \cdots)\right)} \tag{1}$$

$$\frac{p(A)}{1 - p(A)} = \exp\left(b_0 + b_1x_1 + b_2x_2 + \cdots\right) \tag{2}$$

$$\ln\left(\frac{p(A)}{1 - p(A)}\right) = b_0 + b_1x_1 + b_2x_2 + \cdots \tag{3}$$

Regression model parameters are estimated by the Maximum Likelihood Estimation. The Wald statistics tests the statistical significance of regression coefficients. The model's quality could be assessed e.g. with the chi-square goodness of fit test. The model's predictive capacity is evaluated by the means of the confusion matrix and the determination of the sensitivity and specificity; the model's predictive ability may be graphically demonstrated by the ROC curve. Sensitivity is defined as the proportion of actual positives that are correctly identified, while specificity is the proportion of actual negatives that are correctly identified as such. The graph illustrates the dependence of sensitivity and 1-specificity for various selected thresholds. The quality of the model increases with the distance between the ROC curve and the diagonal line [5].

Additionally, contingency tables analysis proved to be an easy way to display the relationship between categorical data. The dependence intensity was identified via the Pearson's Contingency Coefficient and the impact of individual categories was assessed with the post-hoc residual test, for more details see [1], [2], [9], [10] and [11].

3 Research results

As for the binary response variable (How do you usually pay for goods in stores?) we focused on factors that significantly impact whether the respondent pays more frequently in cash (coded as 0) or by card (coded as 1). The explanatory variables (gender, age, education, income, members of the household, children, municipality size) were considered as categorical, the values of the variables were coded on an ordinal scale of 1, 2, 3 … corresponding to the increasing value of the variable. Only the nominal variable of gender was scaled with values 1 (female) and 0 (male). The testing of models containing different explanatory variables gradually filtered out the following variables – gender, income, number of children; their impact proved to be insignificant. Resulting regression model parameter estimates including the Weld statistics values, significance of individual coefficients and the reliability of 95% confidence intervals are listed in Table 1, indicating clear association between the payment method and age, education, household and municipality size. On average, the growing trend of card payments is occurring in the higher income group, higher age group, and larger municipality group; the card payment frequency decreases with the growing number of members in the household. The analysis of sub-dependencies in contingency tables enabled the identification of the exceptions to this average trend in some of the categories.

	Coefficient	Standard Error	Wald Statistics	Significance	Lower 95%	Upper 95%
Constant	−0.0099	0.3618	0.0008	0.9781	−0.7191	0.6992
Age	0.2181	0.0791	7.6074	0.0058	0.0631	0.3731
Education	0.3621	0.1079	11.2703	0.0008	0.1507	0.5736
Household Members	−0.2430	0.0709	11.7586	0.0006	−0.3819	−0.1041
Municipality Size	0.1499	0.0493	9.2464	0.0024	0.0533	0.2466

Table 1 Regression Model Parameters

The overall statistical significance of the model comes out of the nullity test of all regression coefficients (likelihood ratio test), as per Table 2: The model is statistically significant.

The predictive ability of the model is indicated in the confusion matrix (Tables 3 and 4) and the ROC curve (Fig. 1). The confusion matrix suggests that the regression model has high sensitivity and lower specificity, i.e. predicts well the actual positive values as opposed to the actual negative values. The total percentage of correctly classified cases equals the area under the ROC curve; it follows that the model's total predictive ability is slightly under 70% – the model classifies individual cases with rather satisfactory results.

−2 Log-likelihood:	
Initial model =	976.5312
Final model =	907.0126
Likelihood Ratio Statistics:	
Chi-square statistics =	69.5187
Degrees of freedom =	4
Right-tailed probability =	< 0.0001

Table 2 Model's Statistical Significance

The sensitivity/specificity correlation for different values of threshold probabilities is shown in Fig. 2.

Aside from the overall model and in order to provide a detailed description of individual sub-dependencies, there is a contingency table representing every one of them. The Pearson's chi-square test confirms the dependence. The dependence intensity was identified via the Pearson's Contingency Coefficient. The impact of individual categories on the response variable was assessed with the residual analysis. As an example, see an illustrated description of a selected dependence (the others are explained via comments).

Observed/Predict.	0	1
0	78	184
	29.77%	70.23%
1	54	441
	10.91%	89.09%

Table 3 Confusion Matrix

Threshold probability classification =	0.5
Sensitivity =	89.09%
Specificity =	29.77%
Correctly classified =	68.56%

Table 4 Sensitivity and specificity

The dependence between the payment method and the number of household members shall serve as the illustrative example. Table 5 clearly shows a decreasing trend of card payment frequency as the number of members in the household increases. Considering the 4 DoF, the dependence intensity is rather high; the Pearson's coefficient is 0.2. The coefficient is statistically significant. The card payment frequency significantly decreases with the increasing number of household members, from approx. over 70% (one- and two-member households) to 47% (5-member and 5+ households)

Different preferred payment methods in differently sized families are confirmed by the clustered bar charts, see Fig. 3. The more members the household has, the lower the preference for card payments, while the more members in the household, the higher is the frequency of cash payments.

Table 6 illustrates the post-hoc residual analysis. For the 5% level of significance and for the 10 cells in the contingency table the threshold residual value equals 2.807. The two categories (2-member and 5-member families) are the strongest contributors to the dependence; the category of 4-member families is on the edge.

According to the chi-square test as well as the logistic regression, the dependence of the payment method on gender is insignificant; approx. two thirds of men as well as women prefer to pay by card. The frequency of card transactions increases (from approx. 60 to 85%) with age up until the age of 45, and then it goes down under 70% – dependence on age is statistically significant. 45+ respondents are an exception to the rule of the generally increasing trend identified by logistic regression. Furthermore, payment method is significantly dependent on education with the card-based payments being more frequently carried out (60 to 80%) by people with higher education. On the other hand, there is no clear trend when it comes to the household income – the dependence is rather weak. The card payment frequency does not drop too much with the number of children, only in the households with 4 and more children, where the frequency is reported to drop under 40%. Cards are significantly more frequently used in larger municipalities – the values go up from approx. 50% (small municipalities of under 5,000 residents) to more than 85% (cities over 100,000 residents).

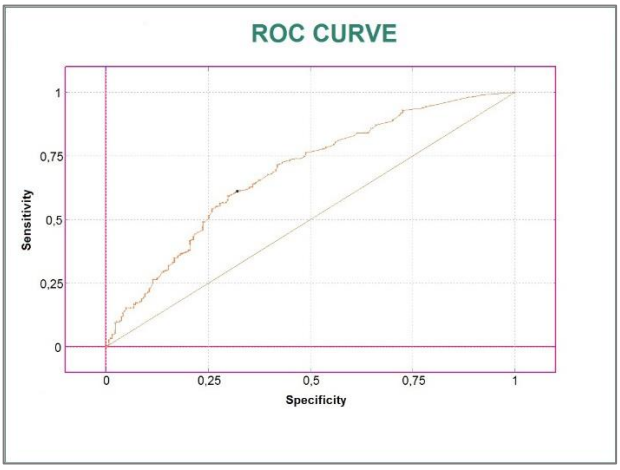


Figure 1 ROC curve

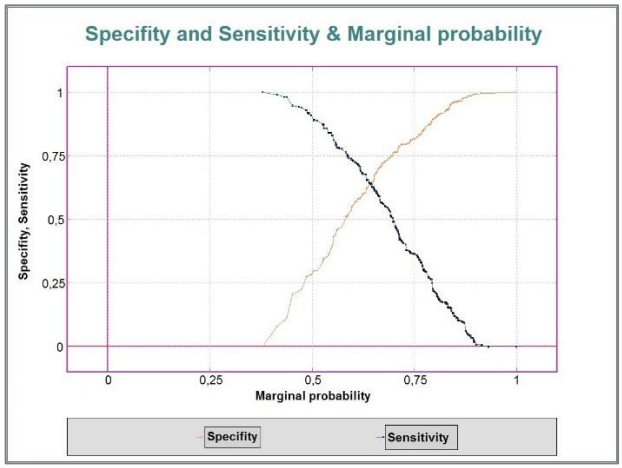


Figure 2 Sensitivity and specificity

Column %	1	2	3	4	5
Payment by Card	73.68%	75.86%	69.28%	58.37%	46.74%
Payment by Cash (or cash equivalent)	26.32%	24.14%	30.72%	41.63%	53.26%

Table 5 Column Relative Frequencies (variables *Payment Method* and *Household Members*)

		Household Members				
		1	**2**	**3**	**4**	**5**
Payment:	**Cash**	–1.6	–3.7	–1.1	2.7	4.0
	Card	1.6	3.7	1.1	–2.7	–4.0

Tab**le 6** Adjusted Standardized Residuals (variables *Payment Method* and *Household Members*), critical value 2.807

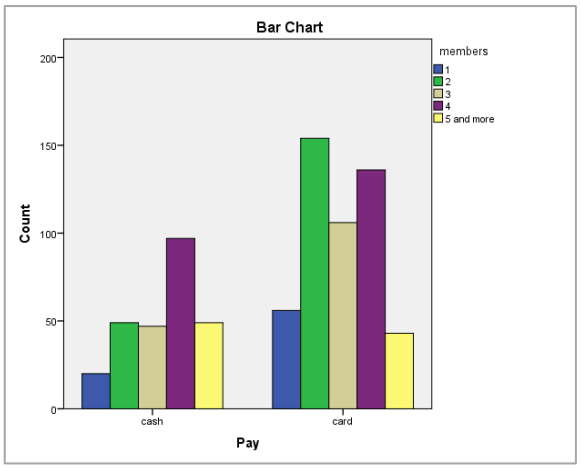


Figure 3 Clustered Bar Charts (variables *Payment Method* and *Household Members*)

4 Conclusion and Discussion

The intention behind this paper was to demonstrate various methods of categorical data processing on an example of the payment method most frequently used by customers in actual shops. First of all, the binary logistic regression model provided a general assessment of the impact of different factors on the payment method. The contingency table describes the sub-dependencies on separate factors; the chi-square test and the Pearson's Contingency Coefficient measure the existence and strength of the dependence; and the post-

hoc residual analysis shows the influence of individual categories. The paper features results for the dependence of the most frequent payment method on the number of household members, which show a decreasing card usage frequency as opposed to payments in cash in larger households. The research also shows that the selection of the payment method in stores does not depend on the customers' gender, and slightly depends on their income and number of children in the household. The card payment frequency increases with respondents' education, municipality size and age (up to 45 years, then it starts to decrease again).

Researches from all over the globe came to the conclusion that the payment method, or rather the frequency of card payments differs significantly according to the type of purchased goods. Purchases of goods such as beer of coffee are paid by card less often than grocery and other hauls [7]. A team of authors [7] earlier concluded that card holders are usually younger, which is in line with the results of this research indicating increasing frequency of card payments up to 45 years of age. Another team of authors [8] pointed out the growing number of card-based payment transactions. They identified the proportions of card payments compared to other possible payment methods recorded in e-commerce in 2011: 19% of transactions were carried out by card, while more than 45% transactions were paid in cash upon delivery. Looking forward, to the post-pandemic economy and the global implementation of cashless economies, it should be interesting to see the development of this trend which forms the topic of further research by this team of authors, among others.

Acknowledgements

This research was supported by the College of Polytechnics, Jihlava, under Grant no. 1170/4/2006 "Využití vícerozměrných statistických metod ve statistickém výzkumu".

References

[1] AGRESTI, A. (2002). *Categorical Data Analysis*. John Wiley a Sons, New York.
[2] ANDĚL, J. (2005). *Mathematical statistics*. Matfyzpress, Praha.
[3] ELLISON, A., WILLIAMSAND, S., & WHYLE, C. (2015). *The electronic payment needs of peopleon low incomes.* Payments Council, Toynbee Hall.
[4] GREENACRE, L. & AKBAR, S. (2019). The impact of payment method on shopping behaviour among low income consumers. *Journal of Retailing and Consumer Services*, *47*, 87–93. 2019. DOI: 10.1016/j.jretconser.2018.11.004.
[5] HOSMER, D. W. & LEMESHOW, S. (2000). *Applied Logistic Regression*. Wiley, New York.
[6] MELVIN, M. & NORRBIN, S. *(2017) Financial Management of the Multinational Firm. International Money and Finance. Elsevier.*
[7] RUNNEMARK, E., HEDMAN, J., & XIAO, X. (2015). Do consumers pay more using debit cards than cash? *Electronic Commerce Research and Applications*, *14(5)*, 285–291. DOI: https://doi.org/10.1016/j.elerap.2015.03.002.
[8] TAKAHASHI, M., AZUMA, H., & TSUDA, K. (2016). A Study on the Efficient Estimation of the Payment Intention in the Mail Order Industry. *Proceedings of the 20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems*. (pp. 1122–1128). Procedia Computer Science.
[9] Zámková, M. & Blašková, V. (2013). The differences in the marketability of organic products in Greece and the Czech Republic. *Agricultural Economics – Zemědělská ekomonika, 59*(5), 219–226.
[10] Zámková, M. & Prokop, M. (2014). Consumer behaviour of young people from Slovak Republic on the field of the bioproducts by using the correspondence and dependence analysis. *Proceedings of the Mathematical Methods in Economics.* (pp. 1113–1118). *Olomouc:* Palacký University.
[11] Zámková, M. & Prokop, M. (2013). The monitoring of Consumers Behavior on the Field of the Bioproducts by Using the Correspondence and Dependence Analysis. *Proceedings of the Mathematical Methods in Economics.* (pp. 1035–1040). *Jihlava:* College of Polytechnics Jihlava.

On Importance of Performance Values for Preference Degrees in PROMETHEE

František Zapletal[1]

Abstract. PROMETHEE methods are popular methods for multi-criteria decision-making. They achieved their popularity, especialy for their good traceability, mathematical simplicity and applicability for many different types of decision-making problems. These methods are based on pairwise comparison of alternatives in terms of each criterion, resulting in quantitative expression of user's preferences. In the original PROMETHEE algorithm, this preference degree is a function of the difference in performances of two alternatives, regardless the initial performance values. This issue simplifies the algorithm computationally, but does not necessarily reflect real preferences of the decision-maker. Based on the law of diminishing marginal utility, it is reasonable to consider not only a difference in performances, but also their values. The aim of this paper is to show that there are problems, for which it is reasonable to take the initial values into account, and to provide an extension of the original PROMETHEE algorithm dealing with this issue. The revised algorithm works with new input parameters, which must be set by the decision-maker. However, all these inputs have their economic interpretation.

Keywords: PROMETHEE, preference degree, multi-criteria decision-making, outranking method.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The family of PROMETHEE methods has been introduced in 1980's by Brans, Mareschal and Vincke [2]. The original PROMETHEE method was developed to get the ranking of alternatives. Since that time, many extensions were presented to solve also other types of decision-making problems, e.g., PROMETHEE V method (portfolio selection) [8], PROMETHEE sorting [5], Efficiency analysis [4], PROMETHEE clustering [6].

The PROMETHEE methods have already been applied in many fields of human doing, see the review of applications available in [1].

Despite the algorithms of each PROMETHEE method varies according to its aim, the logic and the first steps are always the same. Pairwise comparison of alternatives in terms of each criterion is used to derive the strength of preference between them. The problem, which is analysed in this paper, is that this strength of preference (so called preference degree, see Sec. 2) is based on the differences in performances only, i.e. not on performance values directly. The authors of [3] identified a couple of rules for the Even Swap decision-making method, which can be taken as recommendations for any decision making method. One of that is as follows: "Remember that the value of an incremental change depends on what you start with." When comparing PROMETHEE with other popular multi-criteria decision-making methods (MCDM) as, for instance, Analytical Hierarchy Process, TOPSIS, or Weighted Sum Approach [7], PROMETHEE is the only one method, where there is no possibility to take into account the values of performances.

This simplification goes against the very well-known law of diminishing marginal utility. The following two examples demonstrate the situations, in which the PROMETHEE algorithm fails to model the problem well:

- **Mobile phone selection; criterion: size of the screen** – let us suppose that this criterion is maximizing (greater values are preferred), when the screen is too small (e.g., 4"), it could be difficult for a user to read the text or type on the keyboard. In this case, one can expect that one additional inch of size brings substantially higher increase in utility in comparison with the increase from 6" to 7".
- **Ideal destination for the office; criterion: travelling distance from home** – let us suppose that this criterion is minimizing (less values are preferred) and that the decision-maker is willing to walk to the

[1] VŠB – Technical university of Ostrava, Sokolská 33, Ostrava, Czech Republic, frantisek.zapletal@vsb.cz

office if the distance from home is not greater than 3km. In this case, it is natural that the distance reduction of 2km will bring different utility if the initial value is 4 and 8km, respectively.

The algorithm of PROMETHEE assigns the same strength of preference for 5" over 4" screen as for 7" over 6" screen (the difference is 1); the same strength of preference for the distance of 2km over 4km as for 6km over 8km. It is apparent that this approach can be very counterintuitive.

The aim of this paper is to introduce a new extension of the PROMETHEE algorithm, which takes into account not only the differences in performance values, but also the performance values themselves.

2 Brief description of the original PROMETHEE algorithm

In this section, the basic two PROMETHEE algorithms for ranking alternatives are briefly recalled – the partial ranking in PROMETHEE I and the complete ranking in PROMETHEE II. For more details, see [2].

Let us have a decision-making problem with m alternatives and k criteria. The PROMETHEE ranking is found using the following steps:

Step 1: *Preference degrees $P_i(A_x, A_y) \in [0, 1]$ are calculated for all pairs of alternatives A in terms of each criterion $i = 1, , k$ using a preference function p_i (this function assigns a preference degree P_i to each difference in performance values). The preference degree says how strongly the decision maker prefers an alternative with better performance in terms of the given criterion to the one with worse performance. [2] proposed 6 different types of the preference functions, see Fig. 1.*

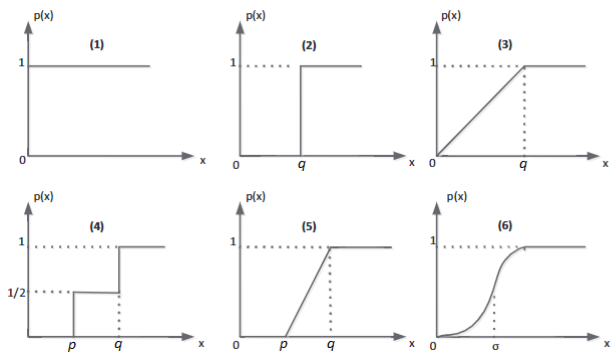


Figure 1 Types of preference functions [2]:(1) Usual, (2) U-shape, (3) V-shape, (4) Level, (5) Linear, (6) Gaussian

Step 2: *The preference degrees are aggregated to preference indices Π. This is done using the sum product of preference degrees and weights:*

$$\Pi(A_x, A_y) = \sum_{i=1}^{k} w_i \cdot P_i(A_x, A_y). \tag{1}$$

The preference index $\Pi(A_x, A_y)$ expresses how strongly A_x is preferred to A_y with respect to all the considered criteria. The weights w_i must be normalised, i.e. $\sum_{i=1}^{k} w_i = 1$. Note that $\Pi(A_x, A_y)$ is again in $[0, 1]$ for all variations of the alternatives.

Step 3: *The preference indices are aggregated to positive and negative flows ($\phi^+ \in [0, 1], \phi^- \in [0, 1]$) of each alternative, see (2) and (3). The positive flow of an alternative is a mean value of the preference indices comparing this alternative with the others (how much better is the alternative than the others), . The other way around, the negative flow of an alternative is a mean value of the preference indices comparing the remaining alternatives with the one under evaluation (how much worse is the alternative than the others):*

$$\phi^+(A_x) = \sum_{j=1, j\neq x}^{m} \Pi(A_x, A_j)/(m-1), \tag{2}$$

$$\phi^-(A_x) = \sum_{j=1, j\neq x}^{m} \Pi(A_j, A_x)/(m-1), \tag{3}$$

where m stands for the number of alternatives. Due to the fact that the positive and negative flows provide only a partial ranking, see Brans, Vincke and Mareschal [2], these partial flows can be further aggregated to the net flows $\phi \in [-1, 1]$:

$$\phi(A_x) = \phi^+(A_x) - \phi^-(A_x). \tag{4}$$

The partial ranking (PROMETHEE I) is based on the positive and negative flows and distinguishes three relations: **preference**

$$A_x \succ^I A_y \Leftrightarrow (\phi^+(A_x) \geq \phi^+(A_y)) \wedge (\phi^-(A_x) \leq \phi^-(A_y)),$$

where at least one of the inequalities is strict; **equivalence**

$$A_x =^I A_y \Leftrightarrow (\phi^+(A_x) = \phi^+(A_y)) \wedge (\phi^-(A_x) = \phi^-(A_y))$$

and **incomparability** $(A_x \mathcal{I}^I A_y)$ otherwise.

The complete ranking is based on the net flows ϕ and distinguishes only **preference**

$$A_x \succ^{II} A_y \Leftrightarrow \phi(A_x) > \phi(A_y),$$

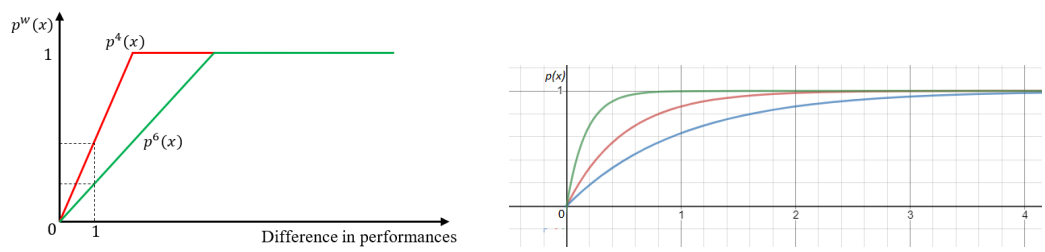
and **equivalence**

$$A_x =^{II} A_y \Leftrightarrow \phi(A_x) = \phi(A_y).$$

3 New proposed algorithm

There are two possible way-outs how to enrich the PROMETHEE algorithm with information about the performance values. Either the shape of the preference function can vary with the performance values, or the shape must be set with respect to some reference performance value.

The first option would be undoubtedly very demanding for the decision-maker. The shape (or at least the parameters of the preference function) should be reconsidered for each pairwise comparison separately. Let us assume the example of mobile phone selection from the introduction and V-shape function (see Fig. 1, (3)). As it was required, the strength of preference is considered stronger for 5" over 4" than for 7" over 6". Here, the decision maker would have to set the preference threshold value q^w (the only parameter of the V-shape function) for the given value of the worse performance in the pair p^w ($w = 4$ for the pairwise comparison between 5" and 4", and $w = 6$ for the pairwise comparison between 7" and 6"). Fig. 2a provides an example how this can be modelled (the difference of 1" brings twice higher value of the preference degree when comparing 5" with 4", than 7" with 6").



(a) Different preference function for each comparison

(b) Concave preference function referring to the worst value (for $\alpha = 1, 2, 6$)

Figure 2 Two options how to manage the preference functions

The second option works with only one preference function like the original algorithm, but reflects some reference performance value, which is fixed for all the comparisons. The reasonable choice is the worst performance value among all alternatives. For the example with 4 alternatives A, B, C, D (4", 5", 6" and 7"), the preference function would refer to the 4-inch option. Then, the comparison between 4" and 5" is managed in the same way as in the original algorithm because one of the pair corresponds to the reference (worst) value. However, when comparing 6" with 7", the distance of the alternatives under comparison from the reference (worst) value must be taken into account.

Because the latter option is easier to manage for the user (less additional inputs are needed), it will be considered further. Namely, the following revised algorithm is proposed:

Step 1: *Choose the preference function p_i^w for the given criterion, which assigns the preference degree between each alternative and the worst-performing alternative A_w (this preference degree is calculated in the same way as in the original PROMETHEE algorithm).*

Step 2: *Calculate the preference degrees $P_i^w(A_j, A_w)$, $\forall j$.*

Step 3: *Calculate the preference degrees for all pairs of the alternatives: $P_i(A_j, A_k), j \neq k$, where $P_i(A_j, A_k) = \max\{P_i^w(A_j, A_w) - P_i^w(A_k, A_w); 0\}$.*

Step 4: *Continue with the original PROMETHEE algorithm.*

If the algorithm proposed above is applied to the (piece-wise) linear type of the preference function (i.e. linear, V-shape), the only possible difference in results, which can occur in comparison with the original algorithm, is caused by the breakpoints of the functions (p and q values). In other words, if $P_i^w(A_j, A_w)$ and $P_i^w(A_k, A_w)$ are both in $]0; 1[$, then $P_i(A_j, A_k)$ will be identical with the original algorithm. On the other hand, even so, it can still help to make the model more realistic. For example, one wants to express that any improvement of the performance in terms of the screen size above the worst value 4" is important (this implies that the V-shape preference function is suitable here). But, if the size exceeds 6", any further enlargement does not bring added utility (preference), because it is "big enough" for the user. When considering the V-shape function with $q = 2$" and our alternatives A, B, C, D described above, the results of the original algorithm and the modified one (with the reference value $w = 4$") differ, see Tab. 1. It can be seen that the new algorithm respects, unlike the original one, the requirement that "if the size exceeds 6", any further enlargement does not bring added utility (preference)".

Original algorithm	A	B	C	D	New algorithm	A	B	C	D
A	–	0	0	0	A	–	0	0	0
B	0.5	–	0	0	B	0.5	–	0	0
C	1	0.5	–	0	C	1	0.5	–	0
D	1	1	0.5	–	D	1	1	0	–

Table 1 Preference degrees: comparison of the original and new algorithm (V-shape function)

Despite the modification proposed above can help to express the preferences in a better and easier way than the original algorithm, it was shown that its benefits are substantially decreased by linearity of the preference function. However, as written in Sec. 2, there is one standard preference function type, which is not linear – Gaussian type (see Fig. 1, no. (6)) given by:

$$f^G = 1 - \exp(-x^2/2\sigma^2), \tag{5}$$

where σ is the parameter of the function corresponding to the inflection point.

The results of the example considering the screen size and 4 alternatives for the Gaussian preference function with $\sigma = 1$ are displayed in Tab. 2.

Original algorithm	A	B	C	D	New algorithm	A	B	C	D
A	–	0	0	0	A	–	0	0	0
B	0.3935	–	0	0	B	0.3935	–	0	0
C	0.8647	0.3935	–	0	C	0.8647	0.4712	–	0
D	0.9889	0.8647	0.3935	–	D	0.9889	0.5945	0.1242	–

Table 2 Preference degrees: comparison of the original and new algorithm (Gaussian function)

It can be seen that, even if the difference in performance is the same between, B–C and C–D, the strength of preference of B over C is almost 4 times greater than between C and D. Due to the fact that the Gaussian preference function is increasing for any positive difference in performances, the requirement "if the size exceeds 6", any further enlargement does not bring added utility (preference)" cannot be fully met here. But, for a sufficiently low value of σ, the added utility is reduced almost to zero in this case.

There is the last issue, which has to be worked out – how to set the value of σ. The function (5) is convex for $x \leq \sigma$ and concave for $x \geq \sigma$. The satisfaction derived from improving the performance by another 1

unit starts to decrease if the decision-maker feels "satisfied enough". Therefore, σ can be understood as the value, by which the worst-performing alternative should be improved to become acceptable for the decision-maker. For example, if the 4" screen is to small for the user and 5" is the minimum value, which is acceptable, $\sigma = 5 - 4 = 1$. The problem is that σ in (5) must always be positive, thus if the worst-performing alternative performs better than the minimum acceptable value, Gaussian function fails. In such case, a function concave on $[0, \infty[$ has to be used instead. An example of the suitable function is (6), which contains the parameter α, instead of σ in the Gaussian function (see Fig. 2b).

$$f^L = 1 - \exp(-\alpha x), \tag{6}$$

As well as for σ in the Gaussian function, let us make some proposal how the α parameter in (6) can be determined by the user. α influences how fast the function approaches 1. When assuming that all the alternatives perform sufficiently well in terms of the given criterion (they exceed the acceptable value σ), one can ask, what performance value would bring the absolute satisfaction. Thus, the decision-maker has to set the value (d) by which the worst-performing alternative must be improved to bring the absolute satisfaction to the decision-maker. The absolute satisfaction should correspond to some value of the preference degree very close to 1 (f^L never reaches 1), e.g., p=0.99. Then, assuming $f^L(d) = 0.99$, it is easy to derive that

$$\alpha = -\frac{1}{d}\ln(1-p) = -\frac{1}{d}\ln 0.01, \tag{7}$$

where p is the value of the preference degree, which corresponds with the absolute preference.

To summarize, the modified PROMETHEE algorithm described in the beginning of this section can be even enriched with the non-linear preference functions. Then, the first step of the proposed algorithm should be replaced by the following (more specific) one: *Does the worst-performing alternative bring enough satisfaction? (i.e. whether this value is acceptable in terms of the given criterion) If it does so, choose the Gaussian function (5) and its parameter σ (how much the worst-performing alternative must be improved to become acceptable in terms of this criterion). If not, choose the concave function f^L (6) and its parameter α (how much the worst-performing alternative must be improved to bring the absolute satisfaction in terms of this criterion).*

4 Numerical example

Let us have a housing selection problem. A young couple wants to rent a flat based on only two criteria – C_1: size and C_2: Monthly rent. They decide among three alternatives (flats): A, B, C, see their performances in Tab. 3.

	A	B	C
Size [m^2]	50	70	90
Monthly rent [K CZK/month]	5	7	9

Table 3 Performances of the alternatives

The couple prefers to have a flat with at least 60m^2, which can be taken as the minimum acceptable value. Flat A does not reach this value (it is by 10m^2 smaller), thus, the Gaussian function with $\sigma = 10$ should be used. On the other hand, the most expensive alternative (Flat C with 9,000CZK/month) is still in accordance with the couple's budget, so even this value is acceptable. Therefore, the concave function (6) should be used here. (Almost) absolute satisfaction (0.99 preference degree) would be achieved when the rent is not greater than 4,000CZK/month per month, i.e. when the most expensive alternative would improve its performance by at least 5,000. According to (7), α is equal to 0.921 for $d = 5$ and $p = 0.99$.

Tab. 4 contains the preference degrees for both – the modified algorithm and the original algorithm. There are two changes. First, the increase by 20m^2 from 70 does not bring so great marginal preference as the same increase from 50. Second, the decrease by 2,000 from 7,000 does not bring so high increase in preference degree as the same decrease from 9,000. Both these changes are reasonable and fully respect the description of the problem.

Assuming that both criteria are equally important (their weights are equal to 0.5), the positive, negative and net flows of the alternatives can be derived using (1), see Tab. 5 and 6.

The partial ranking (PROMETHEE I) of the alternatives is substantially different. The results of the new modified algorithm are:

$$B \succ^I C \succ^I A,$$

New algorithm	$P(A,B)$	$P(A,C)$	$P(B,A)$	$P(B,C)$	$P(C,A)$	$P(C,B)$
C_1	0	0	0.8647	0	0.9997	**0.135**
C_2	**0.1334**	0.9749	0	0.8415	0	0
Original algorithm	$P(A,B)$	$P(A,C)$	$P(B,A)$	$P(B,C)$	$P(C,A)$	$P(C,B)$
C_1	0	0	0.8647	0	0.9997	**0.8647**
C_2	**0.8415**	0.9749	0	0.8415	0	0

Table 4 Preference degrees for the numerical example

	$\phi^+(A)$	$\phi^-(A)$	$\phi^+(B)$	$\phi^-(B)$	$\phi^+(C)$	$\phi^-(C)$
New algorithm	0.2771	0.4661	0.4266	0.0671	0.2837	0.4541
Original algorithm	0.4541	0.4661	0.4266	0.4266	0.4661	0.4541

Table 5 Positive and negative flows of the alternatives

whereas the original PROMETHEE I algorithm brings

$$C \succ^{I} A,$$

and B is incomparable with the others. Moreover, the differences between the partial flows are significantly greater for the new algorithm (st. dev = 0.1441) in comparison with the original algorithm (st. dev = 0.0165). For this example, the new algorithm has a greater distinguishing power than the original one.

There is also a rank-reversal in the complete ranking for both algorithm. Naturally, the complete ranking is the same as the partial ranking for the new algorithm (there were no incomparabilities between in the partial ranking):

$$B \succ^{II} C \succ^{II} A,$$

and the complete ranking for the original algorithm is

$$C \succ^{II} B \succ^{II} A.$$

Due to the greater distances between the flows, and better respecting the input requirements of the couple, the results obtained using the new modified algorithm are more reliable. Thus, the new algorithm is more suitable for this example than the original PROMETHEE algorithm.

	$\phi(A)$	$\phi(B)$	$\phi(C)$
New algorithm	-0.189	0.3595	-0.1704
Original algorithm	-0.012	0	0.012

Table 6 Net flows of the alternatives

5 Conclusions

This paper introduces an alternative algorithm of the PROMETHEE method, which takes into account not only the difference between performance values, but also these values themselves. An advantage of this modification is that a decision-maker has the possibility to express the diminishing strength of preference for each additional unit of performance improvement. The modified algorithm does not require the decision-maker to think about more input parameters – just one value is necessary for each of the two recommended preference functions (Gaussian and the concave function). Moreover, these input parameters have their economic interpretation, thus they are also relatively easy to assign. The modified algorithm has been demonstrated using the numerical example where the original PROMETHEE algorithm cannot capture the reality credibly enough. The future research should be devoted to the analysis of other suitable shapes of the preference functions. More general recommendations when to use the new and old algorithm would be also useful for decision-makers.

Acknowledgements

This paper was supported by SGS project No. SP2020/125. The support is greatly acknowledged.

References

[1] Behzadian, M., Kazemzadeh, R. B., Albadvi, A. and Aghdasi, M. (2010). PROMETHEE: A comprehensive literature review on methodologies and applications, *European journal of Operational research* **200(1)**, 198–215.

[2] Brans, J. P., Mareschal, B. and Vincke, P. (1984). A New Family of Outranking Methods in Multicriteria Analysis, *Operational Research*.

[3] Hammond, J. S., Keeney, R. L. and Raiffa, H. (1998). Even swaps: A rational method for making trade-offs. *Harvard business review* **76**, 137–150.

[4] Ishizaka, A., Resce, G. and Mareschal, B. (2018). Visual management of performance with PROMETHEE productivity analysis. *Soft Computing* **22(22)**, 7325–7338.

[5] Nemery de Bellevaux, P. and Lamboray, C. (2008). FlowSort: a flow-based sorting method with limiting or central profiles. *TOP* **16**, 90–113.

[6] Sarrazin, R., De Smet, Y. and Rosenfeld, J. (2018). An extension of promethee to interval clustering. *Omega* **80**, 12–21.

[7] Triantaphyllou, E. (2000). *Multi-criteria decision making methods: A comparative study*, Springer, Boston.

[8] Vetschera, R., and De Almeida, A. T. (2012). A PROMETHEE-based approach to portfolio selection problems. *Computers & Operations Research* **39(5)**, 1010–1020.

Interdependencies Among the Polish Capital Market and the Markets in Germany and the United States in the Years 1998–2019

Marek Zinecker[1], Michał Bernard Pietrzak[2], Tomáš Meluzín[3], Adam P. Balcerzak[4]

Abstract. After 1989, Poland began the transformation process of the country, which aimed to change the economy from centrally planned into a market one. Building an effective financial market was an important element of this process. In the 90s of the last century there was also a significant increase in links between capital markets in the world, which was mainly the result of the progressing globalization. The growing interdependence among financial markets is now an immanent feature of their dynamic development. In this context, it is important to analyse changes in the interrelationships between capital markets, as it allows taking into account the risks associated with their functioning. The current knowledge on the development of financial market mechanisms is also an important element of macroeconomic policy. The main purpose of the article is to conduct an empirical study that will allow identification and description of changes in the interdependencies among the capital market in Poland and the capital markets of Germany and the United States. The study was based on data from the period 1998–2019. DCC-GARCH model was applied here as the research tool.

Keywords: DCC-GARCH model, conditional variance, conditional correlation, Poland, capital market

JEL Classification: G12, G15, C58
AMS Classification: 91B84

1 Introduction

In the last 30 years, dynamic development of capital markets around the world has been observed. Undoubtedly, the main factor of this phenomenon has been the progressing globalization, the intensification of which has been observed since the 1990s [18, 19, 29, 30,45, 46]. Globalization processes influences the development of economies, affecting the generally understood socio-economic development of countries [24, 25, 37, 39], as well as the selected dimensions comprising it: the level of innovation [8, 21,22, 26, 47, 48], competitiveness of economies [27], institutional changes and changes in trade relations between the economies of individual countries [9,10, 23, 28, 32, 52] or the situation on selected markets (markets of goods and services or labour market [6; 15, 20]. However, globalization processes have the greatest impact on the development and functioning of capital markets [12, 17, 31, 40, 41]. This translates into increased links between financial markets, as demonstrated in many empirical studies [3, 4, 5, 11]. Changes in the interdependence between capital markets affect the effectiveness of stabilization measures related to macroeconomic policies [7, 33, 34, 38, 50]. Contagion processes on financial markets strengthen or limit the impact of macroeconomic policy effectiveness [16; 35, 36, 42, 43, 53]. The global financial crisis of 2008 has shown that in the realities of internationalization of capital markets due to unexpected disturbances there may be a significant reduction in their functioning or even the collapse of whole national financial systems [1, 49, 51, 54]. Therefore, the identification and continuous analysis of changes in the strength of these interdependencies over time is an important element of risk management in both micro and macroeconomic terms.

[1] Brno University of Technology, Faculty of Business and Management, Kolejní 2906/4, 612 00 Brno, Czech Republic, zinecker@fbm.vutbr.cz.
[2] Nicolaus Copernicus University in Toruń, Nicolaus Copernicus University in Toruń, Faculty of Economic Sciences and Management, Ul. Gagarina 13a, 87-100 Toruń, Poland, michal.pietrzak@umk.pl
[3] Brno University of Technology, Faculty of Business and Management, Kolejní 2906/4, 612 00 Brno, Czech Republic, meluzint@fbm.vutbr.cz
[4] Institute of Economic Research, Ul. Bilitewskiego 5, lok. 19; 10-693 Olsztyn, Poland, adam.p.balcerzak@gmail.com

At the beginning of the 1990s, Poland began a market transformation, where one of the main goals was to create an effective capital market. As a starting point of this process, the Warsaw Stock Exchange was established on April 14, 1991, which became the core of the financial market development in Poland. At the initial stage, the stock exchange was heavily dependent on information from global capital markets, including mainly information from the United States and relevant European markets. Over the years, Poland has significantly improved the quality of institutions [2], which has improved the valuation of assets on the Polish stock exchange. The strength of the relationship between the Warsaw Stock Exchange indices and the main global indices has decreased. However, it should be emphasized that there is still a significant impact of external information on the asset valuation process. Therefore, the purpose of the proposed article is to analyse the interdependence among the Polish capital market and the world's leading capital market of the United States and Germany that can be considered as the leading capital market for Central and Eastern Europe. The study was conducted in the period 1998–2019, where the DCC-GARCH model with the conditional t-student distribution was used to determine the relationship between the pointed capital markets.

2 The research on the relations between the capital market in Poland and selected capital markets

In accordance with the objective of the article, an analysis of the interdependencies among the capital market in Poland and the capital markets of the United States and Germany was made. The DCC-GARCH model was used to measure the strength of the interdependence [13, 14]. DCC-GARCH models allow to model conditional variance for individual assets or indexes, as well as conditional correlation [44]. In the DCC-GARCH model, the conditional variance depends on time-delayed conditional variances and on the square of return rates. However, in the case of the conditional correlation equation, the standardized residuals from the equation of variance and time-delayed conditional correlations make the independent variables.

In the empirical study time series of logarithmic rates of return for three stock indices (WIG, S&P 500 and DAX) in the period 1999–2019 were used[5]. In the first step of the study, the specification of the DCC-GARCH model was determined for three selected capital markets, where the AR model (1) was adopted in the equations for the conditional mean, and the GARCH (1,1) model was used in the equations for the conditional variance.

To estimate the parameters of the DCC-GARCH model, the maximum likelihood method with the conditional t-student distribution was used. Table 1 contains the results of the DCC-GARCH model parameter estimation, where a 5% significance level was used to assess the significance of the parameters. The parameters of the AR(1) model are statistically significant for each of the three equations of the conditional mean. Also the parameters of the GARCH model (1,1) for individual equations of conditional variance and parameters of the conditional correlation equation are statistically significant. The assessment of the parameters of the degrees of freedom of the t-student distribution takes values below 10, which indicates the correct selection of this conditional distribution in the estimation procedure.

Equations of conditional mean and conditional variation for index DAX				**Equations of conditional mean and conditional variation for index WIG**			
Model	Parameter	Estimate	p-value	Model	Parameter	Estimate	p-value
AR(1)	α_0	0.008	~0.00	AR(1)	α_0	0.005	~0.00
	α_1	–0.061	0.034		α_1	0.052	0.021
GARCH(1,1)	ω_0	0.002	0.63	GARCH(1,1)	ω_0	0.001	~0.00
	β_1	0.085	~0.00		β_1	0.054	~0.00
	γ_1	0.909	~0.00		γ_1	0.931	~0.00
	v	8.545	~0.00		v	8.472	~0.00
Equations of conditional mean and conditional variation for index S&P 500				**Equation of conditional correlation**			

5 The time series were obtained from: http://www.finance.yahoo.com.

Model	Parameter	Estimate	p-value	Model	Parameter	Estimate	p-value
AR(1)	α_0	0.006	~0.00		α	0.017	~0.00
	α_1	0.092	~0.01	AR(1)	β	0.981	~0.00
GARCH(1,1)	ω_0	0.001	~0.00		v	8.802	~0.00
	β_1	0.113	~0.00	Log–Likelihood		53572.114	
	γ_1	0.812	~0.00	Akaike criteria		–19.196	
	v	6.123	~0.00	Bayes criteria		–19.168	

Table 1 The results of parameters estimation of the DCC-GARCH model for the WIG, DAX and S&P 500 indexes

In the next step of the study, conditional correlation values between the examined stock indices were determined. The obtained correlations for subsequent pairs of indexes (DAX, WIG) and (S&P 500, WIG) are presented in Figure 1. The correlation values indicate strength of the relationship between the two indexes examined. The analysis of the conditional correlation values allows to determine development trends shaping the interdependence between capital markets.

The initial visual analysis of conditional correlations indicates that in the entire analysed period the relationships between the pair of markets (Poland-United States) and (Poland-Germany) are similar. It should also be noted that in selected periods in spite of the geographical distance and relations in the sphere of real economy the Warsaw Stock Exchange was more strongly correlated with the American market as compared to the German market.

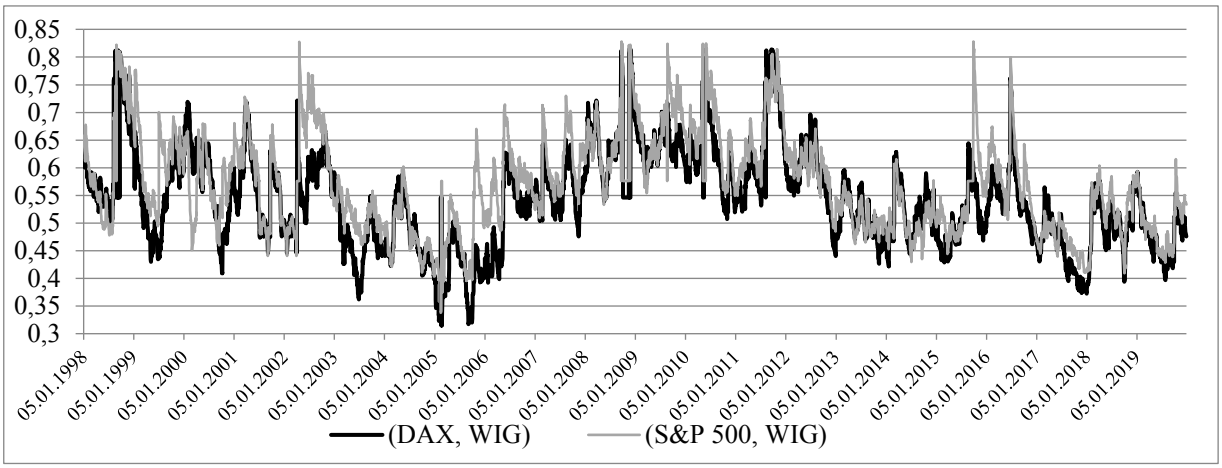


Figure 1 The conditional correlation between the WIG, DAX and S&P 500 indexes.

The visual assessment of the conditional values of the correlation between two pairs of indexes indicates that in the years 1998–2003 there were the strongest correlations between the Polish capital market and the capital markets of the United States and Germany. Then, there is a visible decrease in the strength of the interdependence among the studied indexes in 2004, which may be related to Poland's accession to EU structures. A significantly lower level of interdependence between selected capital markets occurs in 2004–2007. Membership in the European Union had a positive impact on the quality of institutions and the functioning of the capital market. As a result, the Warsaw Stock Exchange reduced dependence on external information in the asset valuation process. In 2008, due to the global financial crisis, the level of interdependence among the markets increased again. The impact of the financial crisis turned out to be long-term and it is only after 2012 that the capital markets returned to normal operations from before the crisis. After a period of high interdependence among the capital markets around the world, a period of significant decrease in interdependence should be noted in the years 2012–2019. It should be emphasized that in the period 2012–2019 there were significant shock events that contributed to a significant increase in the interdependence between capital markets (10.2015 and 06.2016). However, these events were only of a short-term nature and did not significantly affect the disruption of the functioning of capital markets and the occurrence of the contagion effect.

3 Conclusions

The subject of the article concerns the relationships between capital markets, whose strength and range are constantly changing. The phenomenon of interdependence among markets is becoming a significant exogenous factor affecting the efficiency of national economic policy and affects risk management processes at the microeconomic level. Growing links between markets become particularly important in crisis situations, when after a specific shock, successive shocks follow, slowing down the pace of markets returning to equilibrium. As a result of the growing interdependence and impact of shock transmission mechanisms between individual capital markets, the effectiveness of stabilization activities is often weakened. In such conditions, the accumulation of subsequent market shocks means that the crisis situation may develop in an unpredictable manner, often leading to the contagion effect. This situation took place in the years 2008–2012, where the global financial crisis led to the disruptions on the most of capital markets in the world. This means that the identification of interrelationships between markets is a long term significant research problem, and its undertaking may allow for establishing a strategy of action during the crisis and developing systemic tools, which can help to improve the functioning of financial markets. Therefore, the analysis of changes in the strength of these relationships over time is an important guide for domestic macroeconomic decision-makers and is becoming an important element of risk management related to the functioning of capital markets.

The aim of the current article was to examine the interrelations between the Polish capital markets and capital markets of the United States and Germany. The study made it possible to identify the interdependencies between selected markets, and then assess these interrelationships over time. The interdependencies established for Poland, both for the United States and Germany, developed similarly throughout the period considered. Despite the fact that the capital markets of individual European countries operate in a specific way, they react similarly to external information from leading capital markets. It can be concluded that the shock events on the neighbouring German capital market and the strongest US capital market in the world are similarly transferred to the Warsaw Stock Exchange.

The capital market in Poland can be described as developing, whose functioning is slowly approaching the state of mature capital markets. This fact is visible on the established trends for conditional correlation values. A period of high interdependence between selected capital markets in the years 1998–2003 was visible, in which the valuation of assets on the Warsaw Stock Exchange was largely dependent on external information. The situation, both macroeconomic and on the capital market in Poland significantly improved in 2004–2007. This was the result of many reforms that were implemented in order to prepare Poland for accession to the EU structures as well as further membership in the Community. Then in 2008–2012 there was a significant increase in the interdependence between the studied markets, which was the result of the contagion effect that occurred in the markets in connection with the global financial crisis. In the following years 2013–2019, the functioning of the Warsaw Stock Exchange was moving back to equilibrium.

Acknowledgements

This paper was supported by the Institute of Economic Research Toruń, Poland, and Brno University of Technology, Faculty of Business and Management, Czech Republic (grant number: 2019/1). The title of the project is “Public and Private Equity in CEE countries”.

References

[1] ABAKUMOVA, J. & PRIMIEROVA, O. (2018). Globalization, growth and inequality: testing causality and asymmetries. *Ekonomicko-manazerske spektrum*, 12(2), 83–95.

[2] BALCERZAK, A. P. (2020). Quality of institutions in the European Union countries. Application of TOPSIS based on entropy measure for objective weighting. *Acta Polytechnica Hungarica*, 17(1), 101–122.

[3] BAUR, D. (2003). Testing for Contagion – Mean and Volatility Contagion. *Journal of Multinational Financial Management*, 13, 405–422.

[4] BEKAERT, G. & WU, G. (2000). Asymmetric Volatility and Risk in Equity Markets. *Review of Financial Studies*, 13, 1–42.

[5] BEREZINETS, I. V., BULATOVA, L. A., ILINA, Y. B., & SMIRNOV, M. V. (2019). Reactions of emerging stock markets to dividend announcements during economic growth: evidence from India and Russia. *Euroasian Economic Review*, 9(1), 71–89.

[6] BIESZK-STOLORZ, B. & DMYTRÓW, K. (2019). Spatial diversity of effectiveness of forms of professional activisation in Poland in years 2008–2014 by poviats. *Oeconomia Copernicana*, 10(1), 113–130.

[7] BILLIO, M. & CAPORIN, M. (2010). Market Linkages, Variance Spillover and Correlation Stability: Empirical Evidences of Financial Contagion. *Computational Statistics & Data Analysis*, 54(11), 2443–2458.
[8] BRODOWSKA-SZEWCZUK, J. (2019). Determinants of the development of enterprise's innovativeness in the aspect of competitiveness of the economy. *Entrepreneurship and Sustainability Issues, 7(2),* 1279–1295
[9] CIEŚLIK, A. & HIEN TRAN, G. (2019). Determinants of outward FDI from emerging economies. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 14(2), 209–231.
[10] CIEŚLIK, A., MICHAŁEK, J. J., & SZCZYGIELSKI, K. (2019). What matters for firms' participation in Global Value Chains in Central and East European countries? *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 14(3), 481–502.
[11] CORSETTI, G., PERICOLI, M., & SBRACIA, M. (2005). Some Contagion, Some Interdependence: More Pitfalls in Testing for Contagion. *Journal of International Money and Finance*, 24, 1177–1199.
[12] DIMA, B., BARNA, F., & NACHESCU, M. L. (2018). Does rule of law support the capital market? *Economic Research-Ekonomska Istraživanja*, 31(1), 461–479.
[13] ENGLE, R. F. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics*, 20, 339–350.
[14] ENGLE, R. F. (2009). *Anticipating Correlations, A New Paradigm for Risk Management*. Princeton and Oxford: Princeton University Press.
[15] FEDULOVA, I., VORONKOVA, O., ZHURAVLEV, P., GERASIMOVA, E., GLYZINA, M., & ALEKHINA, N. (2019). Labour productivity and its role in sustainable development of economy: on the example of a region. *Entrepreneurship and Sustainability Issues*, 7(2), 1059–1073.
[16] FORBES, K. & RIGOBON, R. (2002). No Contagion, Only Interdependence: Measuring Stock Market Comovements. *Journal of Finance*, 57(5), 2223–2261.
[17] GAIGALIENE, A., JURAKOVAITE, O., & LEGENZOVA, R. (2018). Assessment of EU banking network regionalization during post-crisis period. *Oeconomia Copernicana*, 9(4), 655–675.
[18] HADAŚ-DYDUCH, M. (2016). Globalization in the context of cointegration selectes countries. In *16th International Scientific Conference Globalization and Its Socio-Economic Consequences, Conference Proceedings*. Zilina: University of Zilina, 587–594.
[19] HADAŚ-DYDUCH, M. (2017). Analysis of globalization by wavelet through the prism of the selected euro area countries. In *17th International Scientific Conference Globalization and Its Socio-Economic Consequences*, Zilina: University of Zilina, 605–612.
[20] HE, T.T., LI, W.X. B., & TANG, G.Y.N. (2019). Foreign institutional investors and stock price synchronicity of Chinese listed firms: further evidence. *Euroasian Economic Review*, 9(1), 107–120.
[21] JAKIMOWICZ, A. & RZECZKOWSKI, D. (2019). Do barriers to innovation impact changes in innovation activities of firms during business cycle? The effect of the Polish green island. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 14(4), 631–676.
[22] JAKIMOWICZ, A. & RZECZKOWSKI, D. (2019). Firm ownership and size versus innovation activities over the business cycle: near-zero inertia as a sign of the transition from the fifth to the sixth Kondratieff wave. *Oeconomia Copernicana*, 10(4), 689–741.
[23] JAKLIČ, A. & SVETLIČIČ, M. (2019). China and Central and Eastern European Countries within '16+1': Group or Bilateral Relations? *Entrepreneurial Business and Economics Review*, 7(2), 83–100.
[24] JANKIEWICZ, M. (2018a). The Influence of the Disposable Income on the Consumption Structure in European Countries – Spatio-Temporal Analysis. In M. STANÍČKOVÁ, L. MELECKÝ, E. KOVÁŘOVÁ, & K. DVOROKOVÁ (Eds.). *Proceedings of the 4th International Conference on European Integration 2018*, Ostrava: VŠB-Technical University of Ostrava, 537–544.
[25] JANKIEWICZ, M. (2018b). Spatio-temporal analysis of the convergence of consumption expenditures in the European Union countries. In C. BRĂTIANU, A. ZBUCHEA, & A. VIȚELAR (Eds.). *Challenging the status quo in management and economics: Strategica International Academic Conference*. Bucharest, 142–154.
[26] KIJEK, T. & MATRAS-BOLIBOK, A. (2019). The relationship between TFP and innovation performance: evidence from EU regions. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 14(4), 695–709.
[27] KISEĽÁKOVÁ, D., ŠOFRANKOVÁ, B., ONUFEROVÁ, E., & ČABINOVÁ, V. (2019). The evaluation of competitive position of EU-28 economies with using global multi-criteria indices. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *14*(3), 441–462.
[28] KOZUBIKOVA, L., KOTASKOVA, A., DVORSKY, J., & KLJUCNIKOV, A. (2019). The impact of political factors' perception on suitability of international business environment: The case of startups. *Economics and Sociology*, 12(1), 61–79.

[29] KUC, M. (2017). Is the regional divergence a price for the international convergence? The case of Visegrad group. *Journal of Competitiveness*, 9(4), 50–65.
[30] KUC-CZARNECKA, M., LO PIANO, S., & SALTELLI, A. (2020). Quantitative Storytelling in the Making of Composite Indicator. *Social Indicators Research*. Advance online publication. doi: https://doi.org/10.1007/s11205-020-02276-0.
[31] LEGENZOVA, R., GAIGALIENĖ, A., & JURAKOVAITĖ, O. (2019). Evaluation of the post-crisis EU banking network connectedness in the global context. *Oeconomia Copernicana*, 10(1), 37–53.
[32] LITAVCOVÁ, E., HEČKOVÁ, J., & CHAPČÁKOVÁ, A. (2017). Cross-border mergers and acquisitions as efficient managent tool of capital allocation within European area. *Polish Journal of Management Studies*, 16(1), 94–104.
[33] MOUDUD-UL-HUQ, S. (2019). Banks' capital buffers, risk, and efficiency in emerging economies: are they counter-cyclical? *Euroasian Economic Review*, 9(4), 467–492.
[34] ONOFREI, M., NICUŞOR CĂRĂUŞU, D., & LUPU, D. (2019). The role of the macroeconomic environment in shaping capital market co-movement in C.E.E. countries. *Economic Research-Ekonomska Istraživanja*, 32(1), 3813–3834.
[35] PRASAD, E. S., ROGOFF, K., WEI, S. J., & KOSE, M. A. (2007). Financial globalization, growth and volatility in developing countries. In *Globalization and poverty*. University of Chicago Press, 457–516.
[36] PRZEKOTA, G., REMBEZA, J., MENTEL, G., SZETELA, B. (2019). The relationship between the stock market and the economy: evidence form Central and Eastern European countries. *Transformations in Business & Economics*, 18(2A), 397–415.
[37] SÁNCHEZ-LÓPEZ, C., ACEYTUNO, M. T., & DE PAZ-BAÑEZ, M. A. (2019). Inequality and globalisation: Analysis of european countries. *Economics and Sociology*, 12(4), 84–100.
[38] SCHMUKLER, S. L. & ABRAHAM, F. (2017). *Financial Globalization*. Washington: World Bank.
[39] SHKOLNYK, I., KOZMENKO, S., KOZMENKO, O., & MERSHCHII, B. (2019). The impact of economy financialization on the level of economic development of the associate EU member states. *Economics and Sociology*, 12(4), 43–58.
[40] SKALICKÁ, M., ZINECKER, M., PIETRZAK, M. B., MELUZÍN, T., & DOHNAL, M., (2019b). Financial impact analysis of going public at the Warsaw Stock Exchange: Using Fuzzy Set Theory to understand behaviours of mature companies. *Management & Marketing. Challenges for the Knowledge Society*, 14 (1), 59–79.
[41] SKALICKÁ, M., ZINECKER, M., BALCERZAK, A. P., & MELUZIN, T. (2019a). Why Do Companies Go Public? Evidence from the Prague Stock Exchange. *Entrepreneurial Business and Economics Review*, 7(1), 181–199.
[42] SKARE, M. & PORADA-ROCHOŃ, M. (2019). Financial and economic development link in transitional economies: a spectral Granger causality analysis 1991–2017. *Oeconomia Copernicana*, 10(1), 7–35.
[43] SKARE, M. & PORADA-ROCHOŃ, M. (2019). Tracking financial cycles in ten transitional economies 2005–2018 using singular spectrum analysis (SSA) techniques. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 14(1), 7–29.
[44] SONG, Y., XU, R., LI, L., & WANG, R. (2019). Research on stock price linkages between the third-party payment companies: empirical analysis of companies in China. *Transformations in Business & Economics*, 18(3), 88–105.
[45] STAVYTSKYY, A., KHARLAMOVA, G., GIEDRAITIS, V., & SENGUL, E. C. (2019). Gravity model analysis of globalization process in transition economies. *Journal of International Studies*, 12(2), 322–341.
[46] STRÝČKOVÁ, L. (2017). The practice of capital structure choice in the Czech Republic: A comparative study based on the global data. *Journal of International Studies*, 10(2), 185–203.
[47] SZOPIK-DEPCZYŃSKA, K., CHEBA, K., BĄK, I., STAJNIAK, M., SIMBOLI, A., & IOPPOLO G. (2018). The study of relationship in a hierarchical structure of EU sustainable development indicators. *Ecological Indicators*, 90, 120–131.
[48] SZOPIK-DEPCZYŃSKA, K., KĘDZIERSKA-SZCZEPANIAK, A., SZCZEPANIAK, K., CHEBA, K., GAJDA, W., & IOPPOLO, G. (2018). Innovation in sustainable development: an investigation of the EU context using 2030 agenda indicators. *Land Use Policy'* 79, 251–262.
[49] THALASSINOS, I. E. & STAMATOPOULOS, V. T. (2015). The Trilemma and the Eurozone: A Pre-Announced Tragedy of the Hellenic Debt Crisis. *International Journal of Economics and Business Administration,* 3(3), 27–40.
[50] THALASSINOS, I. E. & THALASSINOS, E. P. (2006). Stock Markets' Integration Analysis. *European Research Studies Journal,* 9(3–4), 3–14.
[51] THALASSINOS, I. E. & THALASSINOS, E. Y. (2018). Financial Crises and e-Commerce: How Are They Related? *Available at SSRN:* https://ssrn.com/abstract=3330169.
[52] WAJDA-LICHY, M. & KAWA, P. (2018). Trade-Finance Nexus: Was it Distorted in the Aftermath of the Global Financial Crisis? *Entrepreneurial Business and Economics Review*, 6(3), 11–27.

[53] ZAREMBA, A. & NIKOROWSKI, J. (2019). Trading costs, short sale constraints, and the performance of stock market anomalies in Emerging Europe. *Economic Research-Ekonomska Istraživanja*, 32(1), 403–422.
[54] ZHANG, X., ZHAO, Y., QI, R., & USMAN, M. (2019). The impact of bank competition on stock price crash risk: evidence form the Chinese market. *Transformations in Business & Economics*, 18(2B), 900–920.

French Nuclear Power Plants in 2019 – DEA Approach

Petra Zýková[1]

Abstract. The paper deals with the efficiency analysis of 18 French nuclear power plants in 2019. The aim is to find efficient and inefficient nuclear power plants. The efficiency analysis is based on the use of data envelopment analysis (DEA) models. This paper applies the traditional DEA model with undesirable output. The analysis considers two inputs: nominal power in megawatts, number of employees, and two outputs. The first output is undesirable – fuel in tonnes and the second is desirable output – electric energy in terawatt per hours.

Keywords: data envelopment analysis, efficiency, nuclear power plant, France

JEL Classification: C44
AMS Classification: 90C05, 90C90

1 Introduction

Electric energy and its consumption in European states is often discussed. More often, than the consumption is the production of electric energy discussed. One possibility of the generation of electric energy is nuclear power plants. This article is focussing on nuclear power plants in France because France is very famous for its nuclear power plants and Frenchmen are relying on nuclear energy. Concretely this article study efficiency of Franch nuclear power plants in 2019. There are many ways how to solve this task. One of them is using DEA models. DEA models have been first developed by Charnes et al. [1] based on the concept introduced by Farrell [3]. Using DEA models, the efficiency scores of decision-making units are computed. This article uses for the analysis DEA model with undesirable output, which is shown in [2].

The paper is organised as follows. The next section presents the definition of DEA models generally and definition of DEA model with undesirable outputs, which is used in this study. Section 3 contains information about energy mix in France and about the results of used DEA model in this paper. The last section of the article concludes the results and discusses future research.

2 DEA models

DEA models are a general tool for efficiency and performance evaluation of the set of homogenous DMUs that spend multiple (w) inputs and transform them into multiple (t) outputs. The measure of efficiency (efficiency score) of this transformation is one of the main results of the application of DEA models. Let us denote $\mathbf{Y} = \left(y_{rj}, r = 1, \dots, t, j = 1, \dots n\right)$ a non-negative matrix of outputs and $\mathbf{X} = \left(x_{kj}, k = 1, \dots, w, j = 1, \dots n\right)$ a non-negative matrix of inputs. The efficiency score of the unit under evaluation DMU_{j_0} is derived as follows:

Maximise

$$U_{j_0} = \frac{\sum_{r=1}^{t} u_r y_{rj_0}}{\sum_{k=1}^{w} v_k x_{kj_0}}$$

subject to

$$\begin{aligned} \frac{\sum_{r=1}^{t} u_r y_{rj}}{\sum_{k=1}^{w} v_k x_{kj}} &\leq 1, \quad j = 1, \dots, n, \\ u_r &\geq \varepsilon, \quad r = 1, \dots, t, \\ v_k &\geq \varepsilon, \quad k = 1, \dots, w, \end{aligned} \tag{1}$$

where u_r is a positive weight of the r-th output, v_k is a positive weight of the k-th input, and ε is an infinitesimal constant. Model (1) is not linear in its objective function but may easily be transformed into a linear program. The linearised version of the input-oriented model (often called the CCR model) is as follows:

Maximise

[1] University of Economics Prague, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, petra.zykova@vse.cz.

$$U_{j_0} = \sum_{r=1}^{t} u_r y_{rj_0}$$

subject to

$$\begin{aligned} \sum_{k=1}^{w} v_k x_{kj_0} &= 1, \\ \sum_{r=1}^{t} u_r y_{rj} - \sum_{k=1}^{w} v_k x_{kj} &\le 0, \quad j = 1, \dots, n, \\ u_r &\ge \varepsilon, \quad r = 1, \dots, t, \\ v_k &\ge \varepsilon, \quad k = 1, \dots, w. \end{aligned} \tag{2}$$

Above mentioned DEA models analyse efficiency score of DMU for desirable maximising outputs. Sometimes there are undesirable outputs. These outputs are minimising character. All outputs O could be divided into two disjunctive subgroups: $O_D$desirable outputs and O_U undesirable outputs. Outputs together form set of outputs: $O = O_D \cup O_U$. Undesirable outputs are transformed to $\psi_{rj} = -y_{rj} + d_r,\ r \in O_U,\ j = 1, \dots, n$where $d_r = \max_j(y_{rj}) + 1$. These transformed ψ_{rj} outputs have maximising character. Including ψ_{rj} transforms model (2) to the following model:

Maximise

$$U_{j_0} = \sum_{r \in O_D} u_r y_{rj_0} + \sum_{r \in O_U} u_r \psi_{rj_0}$$

subject to

$$\begin{aligned} \sum_{k=1}^{w} v_k x_{kj_0} &= 1, \\ \sum_{r \in O_D} u_r y_{rj} + \sum_{r \in O_U} u_r \psi_{rj} - \sum_{k=1}^{w} v_k x_{kj} &\le 0,\ j = 1, \dots, n, \\ u_r &\ge \varepsilon,\ r = 1, \dots, t, \\ v_k &\ge \varepsilon,\ k = 1, \dots, w. \end{aligned} \tag{3}$$

3 Nuclear power plants in France

The consumption of electric energy in France in 2019 was 473 TWh [5]. There is France's energy mix in Figure 1. There are six different ways of electric energy generation or six types of power plants. It is evident from Figure 1 that nuclear power plants in France generate the most amount of electric energy; it is more than 70 %. That is why this article is investigated the efficiency evaluation of French nuclear power plants. The hydroelectric power plants generate the second biggest amount of electric energy; 11 %. After that thermal with 8 % and wind with 6 %. Solar and bioenergy generate only 2 % of annual electric energy consumption. The wind, solar, hydroelectric and bioenergy are renewable sources of energy. The French president Emmanuel Macron announced that France would have only 50 % of nuclear energy in energy mix 2050 [9]. The problem of renewable energy sources is not black and white, because we still cannot store the electrical energy and we need lots of are metals for renewable power plants. Still, the rare metals are not renewable; there is a shortage of rare metals in the world [8].

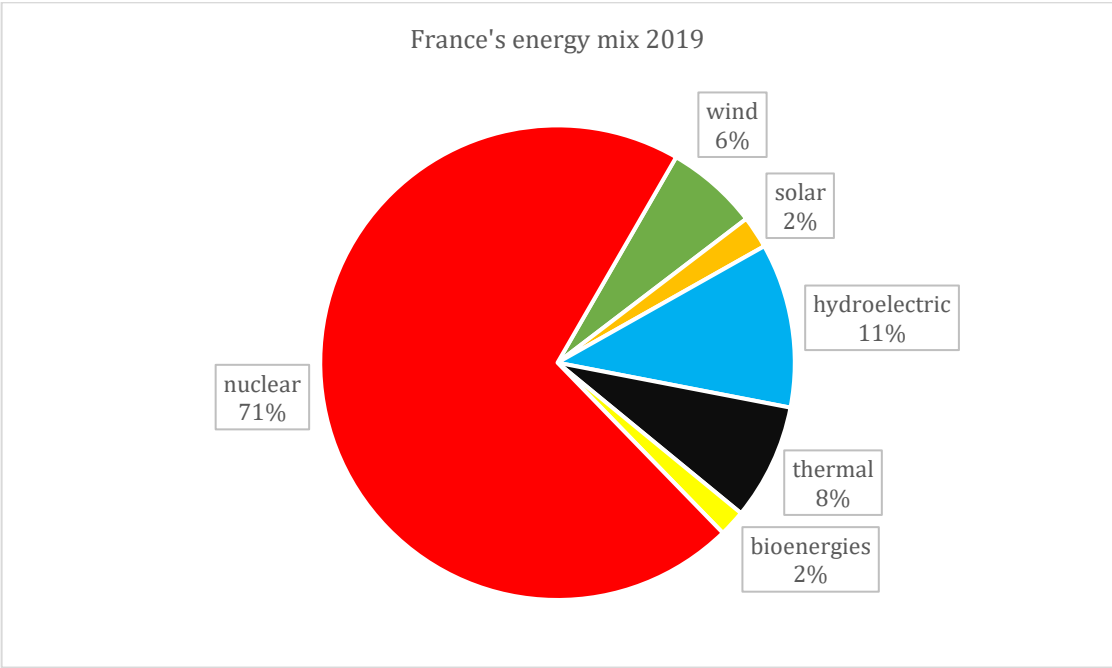


Figure 1 France's energy mix 2019, source [5]

There have been used data set about 18 active nuclear power plants in 2019. The locations of nuclear power plants in France are shown in Figure 2. There are also four not-active nuclear power plants which have been closed. There are now under the decomposition: Brennilis, Bugey, Creys – Malville, Marcoule.

There are two inputs: nominal power in megawatt, number of employees and two outputs. The spent fuel in tonnes, which is desirable output and electric energy in terawatt per hours.

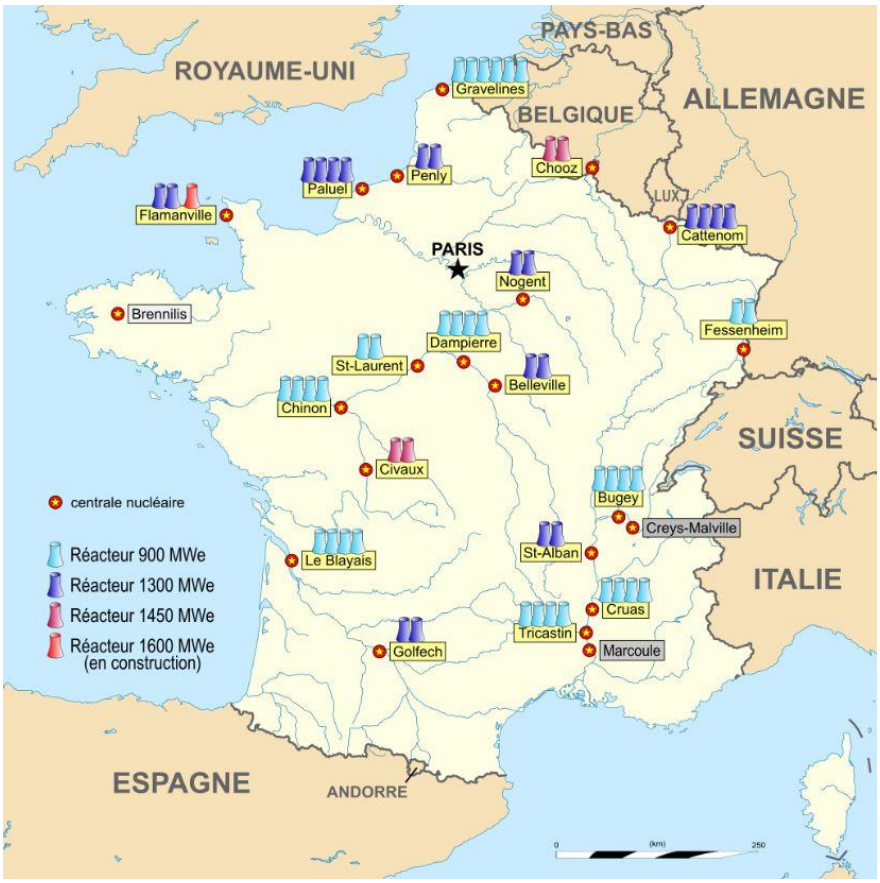


Figure 2 Locations of French nuclear power plants, source [7]

	Nuclear power plant	Nominal Power (MW)	Number of employees	Spent fuel (t)	Electric energy (TWh)
1	Belleville sur Loire	2600	1058	36.0	14.50
2	Blayais	3600	1998	60.0	26.50
3	Cattenom	5200	2044	96.5	30.95
4	Civaux	2900	1200	43.0	22.14
5	Cruas-Meysse	3600	1794	104.5	21.62
6	Dampierre en Burly	3600	2027	26.1	24.02
7	Fessenheim	1800	950	25.9	12.30
8	Flamanville	2600	1174	56.0	6.47
9	Golfech	2600	1020	45.0	17.00
10	Gravelines	5400	3000	126.0	32.10
11	Chinon	3600	2247	69.0	23.17
12	Chooz	2900	1040	108.0	17.90
13	Nogent sur Seine	2600	1181	47.0	16.26
14	Paluel	5200	2137	50.0	26.20
15	Penly	2600	1079	93.6	17.40
16	Saint-Alban	3600	1144	3.0	18.44
17	Saint-Laurent	1800	1065	36.0	10.60
18	Tricatin	3685	2008	47.0	20.21

Table 1 Data set – nominal power, number of employees, spent fuel and electrical energy in 2019 of 18 French nuclear power plants.[6]

There are results of the analysis of 18 French nuclear power plants in 2019 calculated by model (3) in table 2. There have been three power plants established as efficient. There are Civaux, Fessenheim and Saint-Alban. Interestingly, every efficient power plant is in other parts of France. There are in regions Nouvelle-Aquitaine, Grand Est, Auvergne-Rhône-Alpes respectively. Every efficient nuclear power plants in France in 2019 has two nuclear reactors. The power plants Fessenheim is going to be close in 2020 [9]. Other 15 nuclear power plants in France in 2019 were evaluated as inefficient. It is possible to divide them according to the relative efficiency score. There are four nuclear power plants with the efficiency score greater than 0.9. There are Blayais, Golfech, Chooz and Saint-Laurent. These nuclear power plants are located in Nouvelle-Aquitaine, Occitanie, Grand Est and Centre-Val de Loire respectively. These nuclear power plants have two nuclear reactors except power plant Blayais, which has four reactors. The less efficiency power plant is Flamanville in Normandie, which is on the coast of the English Channel. It was caused by the less production of electric energy, which was caused by the temporary shutdown of reactors in Flammanville. The results are also influenced by spent fuel in tonnes, which could be a little bit misleading. Because the exchange of fuel is a demanding process which goes on sections and takes time.

	Nuclear power plant	Model (3)	Ranking
4	Civaux	1.0000	1–3
7	Fessenheim	1.0000	1–3
16	Saint-Alban	1.0000	1–3
2	Blayais	0.9642	4
9	Golfech	0.9634	5
12	Chooz	0.9329	6
17	Saint-Laurent	0.9001	7
1	Belleville sur Loire	0.8842	8
6	Dampierre en Burly	0.8834	9

	Nuclear power plant	Model (3)	Ranking
15	Penly	0.8766	10
13	Nogent sur Seine	0.8436	11
11	Chinon	0.8430	12
3	Cattenom	0.8207	13
5	Cruas-Meysse	0.7866	14
10	Gravelines	0.7786	15
18	Tricatin	0.7215	16
14	Paluel	0.6645	17
8	Flamanville	0.5656	18

Table 2 Efficiency scores and ranking of 18 French nuclear power plants in 2019 computed by the model (3)

4 Conclusions

This paper dealt with an efficiency analysis of 18 French nuclear power plants in 2019. There have been used DEA models for this efficiency analysis, exactly the DEA model with undesirable outputs. There were used two inputs (nominal power and number of employees), and two outputs (spent fuel – undesirable output, electric energy – desirable output). This analysis divided French nuclear power plants into two subgroups – efficient and inefficient power plants. The efficient nuclear power plants in 2019 were Civaux, Fessenheim and Saint-Alban. There are also interesting French nuclear power plants with the efficiency score higher than 0.9. These are Blayais, Golfech, Chooz and Saint-Laurent. There were not any other studies of nuclear power plants in France in 2019 with which could be this study compared.

The numerical experiments were realised using original procedures written in the LINGO modelling language. Efficiency analysis of nuclear power plants is an interesting subject also for further research. In our future research, we plan to extend this efficiency analysis to an eco-efficiency analysis [4], which should be useful because of the planned closing of some French nuclear power plants.

Acknowledgements

This work was supported by the Internal Grant Agency of the Faculty of Informatics and Statistics, University of Economics, Prague, project F4/29/2020 (*Dynamic data envelopment analysis models in economic decision making*).

References

[1] CHARNES, A., COOPER, W., & RHODES, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6), pp. 429–444.
[2] DLOUHÝ, M., JABLONSKÝ, J., & ZÝKOVÁ P. (2018). *Analýza obalu dat*. Praha: Professional Publishing.
[3] FARRELL, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3): 253–290.
[4] MARTIN-GAMBOA, M., IRIBARREN, D., & DUFOUR, J. (2018). Environmental impact efficiency of natural gas combined cycle power plants: A combined life cycle assessment and dynamic data envelopment analysis approach. *Science of The Total Environment*. (615), 29–37.
[5] https://www.connaissancedesenergies.org/sites/default/files/pdf-actualites/Bilan%20%C3%A9lectrique%202019.pdf
[6] https://www.edf.fr/groupe-edf/edf-pres-de-chez-moi
[7] https://fr.wikipedia.org/wiki/Industrie_nucl%C3%A9aire_en_France#/media/Fichier:Nuclear_power_plants_map_France-fr_2.svg
[8] https://www.info.cz/svet/drabova-nemecka-energiewende-popira-zakony-fyziky-41632.html?utm_source=www.seznam.cz&utm_medium=sekce-z-internetu&fbclid=IwAR0kJucjD4gxwLoZGdfQDIDwa1xh984vDw4Vmt_onD9lO9ogrLwrltPwkjo
[9] https://www.lefigaro.fr/flash-eco/2018/11/27/97002-20181127FILWWW00096-emmanuel-macron-la-part-du-nucleaire-ramenee-a-50-en-2035.php

Index of Authors

38th International Conference
on Mathematical Methods in Economics (MME 2020)
Conference Proceedings

Responsible editors: Svatopluk Kapounek, Hana Vránová
Technical editors: Jan Přichystal, Jiří Rybička

Published by Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
www.mendelu.cz

Published by Mendel University Publishing Center,
Zemědělská 1, 613 00 Brno, Czech Republic

684 pages

First edition
Brno 2020

ISBN 978-80-7509-734-7

https://mme2020.mendelu.cz/

ISBN 978-80-7509-734-7

9 788075 097347