

MME2021

Mathematical Methods in Economics

39th International Conference on Mathematical Methods in Economics

**Faculty of Economics and Management
Czech University of Life Sciences Prague**

8th - 10th September 2021

Conference Proceedings



Czech University Of Life Sciences Prague

**Faculty of Economics
and Management**

Czech University of Life Sciences Prague
Faculty of Economics and Management

Proceedings of the 39th International Conference on
Mathematical Methods in Economics

MME2021
Mathematical Methods in Economics

8th - 10th September 2021
Prague, Czech Republic, EU

Editor: Robert Hlavatý
Cover: Jiří Fejfar
Technical editors: Jiří Fejfar, Michal Hruška
Publisher: Czech University of Life Sciences Prague
Kamýcká 129, Prague 6, Czech Republic

Publication is not a subject of language check.

Papers are sorted by authors' names in alphabetical order.

All papers passed a peer review process.

© Czech University of Life Sciences Prague

© Authors of papers

ISBN 978-80-213-3126-6

Programme Committee

doc. RNDr. Ing. Miloš Kopa, Ph.D.

*President of the Czech Society for Operations Research
Charles University, Faculty of Mathematics and Physics*

prof. Dr. Ing. Miroslav Plevný

*Vice-president of the Czech Society for Operations Research
University of Economics in Prague, Faculty of Informatics and Statistics*

prof. RNDr. Helena Brožová, CSc.

Czech 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

prof. Ing. Jana Hančlová, CSc.

Technical University of Ostrava, Faculty of Economics

prof. Ing. Josef Jablonský, CSc.

University of Economics in Prague, Faculty of Informatics and Statistics

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

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

doc. Ing. Tomáš Šubrt, Ph.D.

Czech University of Life Science in Prague, Faculty of Economics and Management

Ing. Miroslav Vavroušek, Ph.D.

Technical University of Liberec, Faculty of Mechanical Engineering

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. Robert Hlavatý, Ph.D. (chair)

Czech University of Life Science in Prague, Faculty of Economics and Management

Ing. Jiří Fejfar, Ph.D.

Czech University of Life Science in Prague, Faculty of Economics and Management

Ing. Martina Houšková Beránková, Ph.D.

Czech University of Life Science in Prague, Faculty of Economics and Management

Ing. Igor Krejčí, Ph.D.

Czech University of Life Science in Prague, Faculty of Economics and Management

Ing. Tereza Jedlanová

Czech University of Life Science in Prague, Faculty of Economics and Management

Ing. Tereza Sedlářová Nehézová

Czech University of Life Science in Prague, Faculty of Economics and Management

Technical Editors

Ing. Jiří Fejfar, Ph.D.

Czech University of Life Science in Prague, Faculty of Economics and Management

Ing. Michal Hruška, Ph.D.

Czech University of Life Science in Prague, Faculty of Engineering

Foreword

It is a pleasure to present you the Proceedings of the 39th International Conference on Mathematical Methods in Economics - MME2021. The conference was held by the Czech University of Life Sciences Prague (CZU), Faculty of Economics and under the auspices of the Czech Society for Operational Research on September 8-10th 2021. The conference hosted nearly 120 participants, both onsite and online. The proceedings contain 89 reviewed contributions.

The conference MME is a traditional event that brings together researchers and practitioners in the field of Operations research and Econometrics, and it returned to CZU after 12 years has passed since being organised here. It is not hard to notice that many things have changed since then – the number of contributions has nearly doubled, the sessions have become online, and the old research questions have turned into new research questions. Fortunately, the most important aspect – the social role of the conference - has remained. It is still an event that represents an opportunity for the best experts in the field to meet and spend some time together.

The year 2021 was marked by issues and challenges where the mathematical methods will play an important role: the pandemic, the new green deal and the turbulent weather no less. On this 39th birthday of the conference, let us wish this event more successful years and ideas that will contribute to dealing with those challenges. Finally, let me express my sincere thanks to the members of the programme committee for securing the smooth review process and the members of the organising committee whose effort and hard work have made this event possible.

September 2021

Robert Hlavatý

Table of Contents

The competitiveness of V4 countries in the context of EU member states Markéta Adamová, Jana Klicnarová, Nikola Soukupová	6
Economic Policy Uncertainty and Stock Markets Co-movements Peter Albrecht, Svatopluk Kapounek, Zuzana Kučerová	12
Optimal consumption with irreversible investment in the context of the Ramsey model Nastaran Ansari, Adriaan Van Zon, Olaf Sleijpen	18
Impact of incorporating and tailoring PRINCE2 into the project-oriented environment Jan Bartoška, Jan Rydval, Tereza Jedlanová	24
Assessment of personal ambiguity attitude in a series of online experiments Simona Bažantová, Vladislav Bína, Václav Kratochvíl, Klára Šimůnková	30
An original two-index model of the multi-depot vehicle routing problem Zuzana Borčinová, Štefan Peško	36
Portfolio selection via a dynamic moving mean-variance model Adam Borovička	42
A shadow utility of portfolios efficient with respect to the second order stochastic dominance Martin Branda	48
Efficient Values of Selected Factors of DMUs Production Structure Using Particular DEA Model Helena Brožová, Milan Vlach	54
On the measurement of risk of some cointegrated trading strategies Michal Černý, Vladimír Holý, Petra Tomanová, Lucie Beranová	60
Statistical Analysis of ICT Utilization in Marketing Andrea Čížků	66
On the crossing numbers of join of one graph on six vertices with path using cyclic permutation Emília Draženská	72
Efficiency of Credit Risk Management and Their Determinants in Central European Banking Industries Xiaoshan Feng	83
Models of Technology Coordination Petr Fiala, Renata Majovská	89
The Impact of Covid-19 on Mutual Relations of Czech Macro-aggregates: Effect of Structural Changes Jakub Fischer, Kristýna Vltavská	95
Productivity analysis in the Mexican food industry Martin Flégl, Carlos Alberto Jiménez Bandala, Isaac Sánchez-Juárez, Edgar Matus	100
Evaluation and testing of non-nested specifications of spatial econometric models Tomáš Formánek	106
The link between DEA efficiency and return to assets Lukáš Frýd, Ondřej Sokol	112
The geography of most cited scientific publications: Mixed Geographically Weighted Regression approach Andrea Furková	117
Bilevel Linear Programming under Interval Uncertainty Elif Garajová, Miroslav Rada, Milan Hladík	123

An Efficiency Comparison of the Life Insurance Industry in the Selected OECD Countries with Three-Stage DEA Model Biwei Guan	129
Determination of Wages in Forestry Depending on the Occurrence of Natural Disasters David Hampel, Lenka Viskotová, Antonín Martiník	135
Efficiency evaluation of the health care system during COVID-19 pandemic in districts of the Czech Republic Jana Hančlová, Lucie Chytilová	141
Analysis of uneven distribution of diseases COVID – 19 in the Czech Republic Jakub Hanousek	147
Determinants of company indebtedness in the construction industry Jana Heckenbergerová, Irena Honková, Alena Kladivová	155
Robust Slater’s Condition in an Uncertain Environment Milan Hladík	161
Sensitivity of small-scale beef cattle farm’s profit under conditions of natural turnover Robert Hlavatý, Igor Krejčí	167
The Use of Genetic Algorithm in Clustering of ARMA Time Series Vladimír Holý, Ondřej Sokol	173
Enumerative Core of Pólya’s Theorems on Random Walk and the Role of Generating Functions in their Proofs Richard Horský	179
Numerical Valuation of the Investment Project with Expansion Options Based on the PDE Approach Jiří Hozman, Tomáš Tichý	185
Housing Submarkets: The case of the Prague Housing Markets Petr Hrobař, Vladimír Holý	191
Forecasting Czech unemployment rate using dimensional reduction approach Filip Hron, Lukáš Frýd	197
Health index for the Czech districts calculated via methods of multicriteria evaluation of alternatives Dana Hübelová, Beatrice-Elena Chromková Manea, Alice Kozumplíková, Martina Kuncová, Hana Vojáčková	202
Modelling of PX Stock Returns during Calm and Crisis Periods: A Markov Switching Approach Michaela Chocholatá	208
The Performance Assessment of Different Types of Investment Funds Using Markowitz Portfolio Theory Zuzana Chvátalová, Oldřich Trenz, Jitka Sládková	214
SBM models in data envelopment analysis: A comparative study Josef Jablonský	220
Swap Heuristics for Emergency System Design with Multiple Facility Location Jaroslav Janáček, Marek Kvet	226
The minimal network of hospitals in terms of transportation accessibility Ludmila Jánošíková, Peter Jankovič	232
Multifractal approaches in econometrics and fractal-inspired robust regression Jan Kalina	238
LTPD variables inspection plans and effect of wrong process average estimates Nikola Kaspříková	244

Flexible Job Shop Schedule generation in Evolution Algorithm with Differential Evolution hybridisation František Koblasa, Miroslav Vavroušek	249
Distortion risk measures in portfolio optimization Miloš Kopa, Juraj Zelman	255
The goal programming approach to investment portfolio selection during the COVID-19 pandemic Donata Kopańska-Bródka, Renata Dudzińska-Baryła, Ewa Michalska	261
Robust First Order Stochastic Dominance in Portfolio Optimization Karel Kozmík	269
System Dynamic Model of Beehive Trophic Activity Kratochvílová Hana, Rydval Jan, Bartoška Jan, Chamrada Daniel	275
An Analysis of Dependence between German and V4 Countries Stock Market Radmila Krkošková	281
Evaluation of the Construction Sector: A Data Envelopment Analysis Approach Markéta Křetínská, Michaela Staňková	287
The path-relinking based search in unit lattice of m -dimensional simplex Marek Kvet, Jaroslav Janáček	293
Portfolio discount factor evaluated by oriented fuzzy numbers Anna Łyczkowska-Hanćkowiak, Krzysztof Piasecki	299
Comparing TV advertisement in the year 2019 using DEA models Jan Malý, Petra Zýková	305
Efficiency of tertiary education in EU countries Klára Mašková, Veronika Blašková	312
Application of robust efficiency evaluation method on the Czech life and non-life insurance markets Markéta Matulová, Lucia Kubincová	318
Stochastic reference point in the evaluation of risky decision alternatives Ewa Michalska, Renata Dudzińska-Baryła	325
Structure of the threshold digraphs of convex and concave Monge matrices in fuzzy algebra Monika Molnárová	331
Weak Solvability of Max-plus Matrix Equations Helena Myšková	337
The Impact of Technical Analysis and Stochastic Dominance Rules in Portfolio Process David Neděla	343
Information Retrieval System for IT Service Desk for Production Line Workers Dana Nejedlová, Michal Dostál	349
Permanent Income Hypothesis with the Aspect of Crises. Case of V4 Economies Václava Pánková	355
Multi Vehicle Routing Problem Depending on Vehicle Load Juraj Pekár, Zuzana Čičková, Ivan Brezina	359
Using Parametric Resampling in Process of Portfolio Optimization Juraj Pekár, Mário Pčolár	364
Optimal routing order-pickers in a warehouse Jan Pelikán	370
Performance of the CLUTEX Cluster Applying the DEA Window Analysis Natalie Pelloneová	375

New models for a return bus scheduling problem Štefan Peško, Stanislav Palúch, Tomáš Majer	381
Different ways of extending order scales dedicated to credit risk assessment Krzysztof Piasecki, Aleksandra Wójcicka-Wójtowicz	387
Interval two-sided (max, min)-linear equations Ján Plavka	393
Forecasting of agrarian commodity prices by time series methods Alena Pozdílková, Jaromír Zahrádka, Jaroslav Marek	399
Discrete Time Optimal Control Problems with Infinite Horizon Pavel Pražák, Kateřina Frončková	405
Bankruptcy Problem Under Uncertainty of Claims and Estate Jaroslav Ramík, Milan Vlach	411
Searching for a Unique Good Using Imperfect Comparisons David M. Ramsey	417
Dealing with uncertainty by Fuzzy evaluation and robust optimization Tereza Sedlářová Nehézová, Michal Škoda, Helena Brožová	423
The System Dynamics approach to creation of a recovery model of an urban object Anna Selivanova	429
Fuzzy discount factor parametrized by logarithmic return rate Joanna Siwek, Krzysztof Piaseck	435
Consumption Expenditures and Demands of Ageing Population Jaroslav Sixta, Jakub Fischer	440
Central Moments and Risk-Sensitive Optimality in Markov Reward Processes Karel Sladký	446
Construction and optimization of HFT systems with the use of binary-temporal state model Michał Dominik Stasiak	452
Possibilistic median of a fuzzy number Jan Stoklasa, Pasi Luukka	458
Asymmetric Transmission of Crude Oil Prices to Retail Gasoline and Diesel Prices in U.S. Market Karol Szomolányi, Martin Lukáčik, Adriana Lukáčiková	464
DEA Window Analysis of Engineering Industry Performance in the Czech Republic Eva Štichhauerová, Miroslav Žižka	469
Work Contour Models in Projects Tomáš Šubrt, Jan Bartoška, Daniel Chamrada	475
Visualising HR data concerning the performance of university staff - career development assessment and assistance perspective Tomáš Talášek, Jana Stoklasová, Jan Stoklasa, Jana Talašová	481
Combined Time Coordination of Connections in Public Transport Dušan Teichmann, Michal Dorda, Denisa Mocková, Pavel Edvard Vančura, Vojtěch Graf, Ivana Olivková	487
Measuring Efficiency of Football Clubs: DEA approach Michal Tomíček, Natalie Pelloneová	493
Analysis of profitability of cryptocurrencies trading strategy based on technical analysis Quang Van Tran, Jaromír Kukal	499

GLMM Based Segmentation of Czech Households Using the EU-SILC Database Jan Vávra	505
State-space modeling of claims reserves in non-life insurance Petr Vejmělka	511
Forecasting third party insurance: A comparison between Random forest and Generalized Additive Model Lukáš Veverka	517
Efficiency Verifications of Classical Portfolio Optimization Models Anlan Wang	523
Scoring Applications in Early Collection Jiří Witzany, Anastasiia Kozina	529
On Modelling Dependencies between Criteria in PROMETHEE František Zapletal	535
Consumer Dividend Aristocrats: Dynamic DEA Approach Petra Zýková	541

The competitiveness of V4 countries in the context of EU member states

Markéta Adamová¹, Jana Klicnarová², Nikola Soukupová³

Abstract. Even though there is no generally accepted definition and understanding of the concept of competitiveness, this issue is currently of interest to many economic analyses as a basic measurement for countries' macroeconomic performance. The measurement of efficiency is subject to different methods and data operationalization, but in the European Union, the competitiveness concept has not been uniquely defined yet. In our previous research, we provided an analysis of the efficiency of the European Union (EU) Member states using the Data Envelopment Analysis (DEA) and Malmquist index of the efficiency of all EU-27 Member States during the period 2013-2019 given the total unemployment rate, general gross government debt, gross capital formation and GDP per capita as macroeconomic indicators. This paper provides an analysis of the position of V4 countries within the evaluation of the EU Member states with aims to analyze the efficiency progress of Visegrad countries (V4) in comparison with EU member states.

Keywords: competitiveness, EU member states, economic efficiency, DEA, Malmquist index

JEL Classification: C61, R11, R15

AMS Classification: 91B82

1 Introduction

The concept of competitiveness is currently of interest to economic theorists' attention as a basic measurement for countries' macroeconomic performance, even though there is no generally accepted definition, understanding and measurement system of this concept due to its complexity and different perceptions [8, 21]. According to [22], as a mirror of competitiveness could be understood the concept of economic efficiency as a commonly applied instrument to help identify the strengths and weaknesses of the evaluated states. The measurement of economic efficiency, which closely related to the use of resources in the economy, is subject to different methods and data operationalization, but in the European Union, the concept has not been uniquely defined yet, although the EU Member States efficiency is the source of national competitiveness [1, 4].

The interest in measuring the efficiency of states has led to the development of different methods and data operationalization applied to the evaluation of efficiency. A frequently used research method is Data Envelopment Analysis (DEA) method or Malmquist index productivity because DEA is suitable for determining efficiency of units that are comparable to each other, e. g. selected macroeconomic indicators of countries [1, 4, 22]. Table 1 shows researches where the efficiency of states is evaluated by the data envelopment analysis (DEA) method.

Authors	Year	Main findings
Fare et al. [7]	1994	Innovation contributes to competitiveness growth more than improvements.
Martic and Savic [13]	2000	Only 17 regions in the EU can be classified as effective by DEA.
Staničková and Melecký [22]	2016	The best efficiency changes in competitiveness were achieved by NUTS 2 regions belonging to the group of the Visegrad countries.
Moutinho, Madaleno and Robaina [18]	2017	Resources productivity shows a positive and significant influence independently of the country technical and eco-efficiency level.

¹ University of South Bohemia in České Budějovice, Department of Management, Studentská 13, 370 05 České Budějovice, adamovam@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, nsoukupova@ef.jcu.cz.

Authors	Year	Main findings
Łozowicka [11]	2020	In inefficient countries can be identified weak areas and indicate the action that should be taken to improve their efficiency.
Stanković, Marjanović and Stojković [23]	2021	26 out of 28 EU members state do not achieve satisfactory levels of socio-economic efficiency.

Table 1 DEA method used for evaluation of efficiency of states
Source: own processing

In the EU there are still significant economic, social, and territorial disparities that lead to cohesion concerns in the expanding and further integrating EU [24]. After the second world war, western European countries had to undergo turbulent times. In countries of central Europe, there was communism which affects its national economies [20]. Many researchers define the term new EU countries (post-2004) and old/traditional (pre-2004) EU countries, e. g. [3, 20, 17, 15]. Between old EU countries Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden belong. As new EU countries are labeled Bulgaria, Cyprus, Estonia, Latvia, Lithuania, Malta, Slovenia, Romania, Croatia, and all of the V4 countries [26]. According to [3], the present competitive advantage of new EU member states is a low-cost-based economy which is attractive for foreign investors.

V4 countries belong to the group of transition countries [12]. The V4 group consists of four Central European countries – Slovakia, Czechia, Hungary, and Poland. V4 countries have a common history, culture, and geographical position [2]. This may lead to the idea that they are comparable. According to [17], the V4 countries have undergone rapid growth accompanied by restructuring and modernization.

Based on the research of [9] within the V4 countries the Czechia is the most successful country in terms of indicators (reference years 2010 -2014). The differences in the economic performance of V4 countries are being narrowed. According to [14] *"development in V4 countries has a trend towards advanced countries, such as Austria and Germany. There was a growth in their performance, increasing trend in effective use of their advantages and improve in competitive position."*

This paper provides an analysis of the position of V4 countries within the evaluation of the EU Member states with aims to analyze the efficiency progress of Visegrad countries (V4) in comparison with EU member states.

2 Materials and methods

In our paper [10] we have analyzed the EU countries according to values of GDP per capita, Gross Capital Formation, General Government Debt, and Unemployment rate between 2013 and 2019. We applied Data Envelopment Analysis to identify countries with optimal values under all these criteria and used the Malmquist index (MI) to measure progress in this evaluation. If we focus on V4 countries, we can see from results DEA that only Czechia belongs to states with an optimal combination of outputs and inputs, in DEA language we called these states efficient. The position of Czechia might be influenced by the low level of the unemployment rate (it was oscillating around 2%) and decreasing level of government general gross debt (Czechia dept is decreasing from 2013) [6]. Results of a DEA are based only on data from the chosen year, therefore, it could be supplemented by the Malmquist index to add an overview of development over time (2013-2019).

However, if we study a MI based on data from 2013 and 2019, we can see that all V4 countries achieved quite good results, their position is in the first half within the order of EU member states. The Czech Republic took 2nd place (5.07), Hungary took 3rd place (4.2), Poland ranked 9th place similar to Slovakia (11th place, score 2.42). The results follow the idea that there is a different dynamic between old and new EU member states

DMUs	MI	DMUs	MI	DMUs	MI	DMUs	MI
Ireland	5.29	Netherlands	2.69	Belgium	1.86	Luxembourg	1.34
Czechia	5.07	Poland	2.58	Denmark	1.83	Sweden	1.32
Hungary	4.2	Malta	2.49	Germany	1.77	Spain	1.26
Croatia	3.44	Slovakia	2.42	Latvia	1.75	Bulgaria	1.22

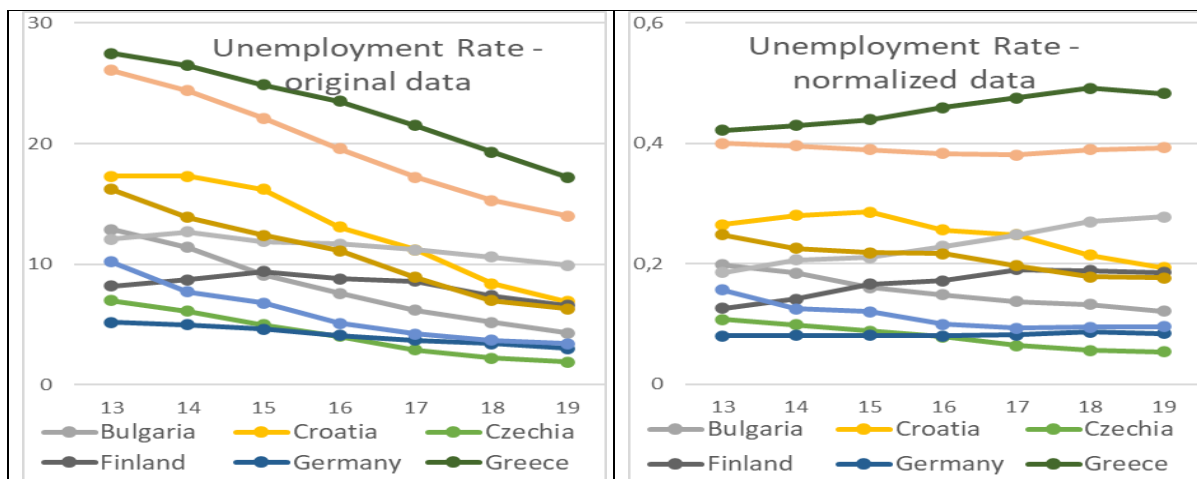
DMUs	MI	DMUs	MI	DMUs	MI	DMUs	MI
Portugal	3.34	Romania	2.28	Estonia	1.75	Italy	1.15
Slovenia	3.27	Lithuania	1.96	Finland	1.52	France	1.04
Cyprus	2.95	Greece	1.87	Austria	1.38		

Table 2 Malmquist Index Score of DMUs

Source: own processing

Now, our aim is to focus in more detail on the progress of V4 countries in chosen parameters in comparison to other EU countries. For our analyses we have used Unemployment Rate, General Government Debt, Gross Domestic Product per capita (hereinafter GDP), and Gross Capital Formation (hereinafter GCF), all of the indicators were ILO modeled estimates to ensure comparability across countries and over time. The data were obtained from the World Bank [25] and the International Monetary Fund and the reference period was chosen 2013-2019 due to the elimination of the effect of the financial crisis and COVID pandemic situation. For DEA we applied Unemployment Rate, General Government Debt as proxies for inputs and Gross Domestic Product per capita, and Gross Capital Formation as proxies for outputs.

To compare the progress of countries under these parameters. First, it was necessary to standardize data to get comparable values under different variables. Since we use cluster analysis with Euclidian metrics, we normalize data into vectors with unit L_2 -norms (under variables, i.e. the vector of GDP per capita in 2013 has a unit norm, for example). Our aim is to compare all values and progresses in all factors, therefore we applied cluster on all normalized data – values of all variables in all studied years - i.e. we used values of GDP per capita, GCF, Unemployment rate, and Debt from 2013 to 2019. In the following graphs, we can compare original and normalized data. We can see that the order of countries' values for each year stays the same, however, the proposition of the positions can change, it is given by the normalization.



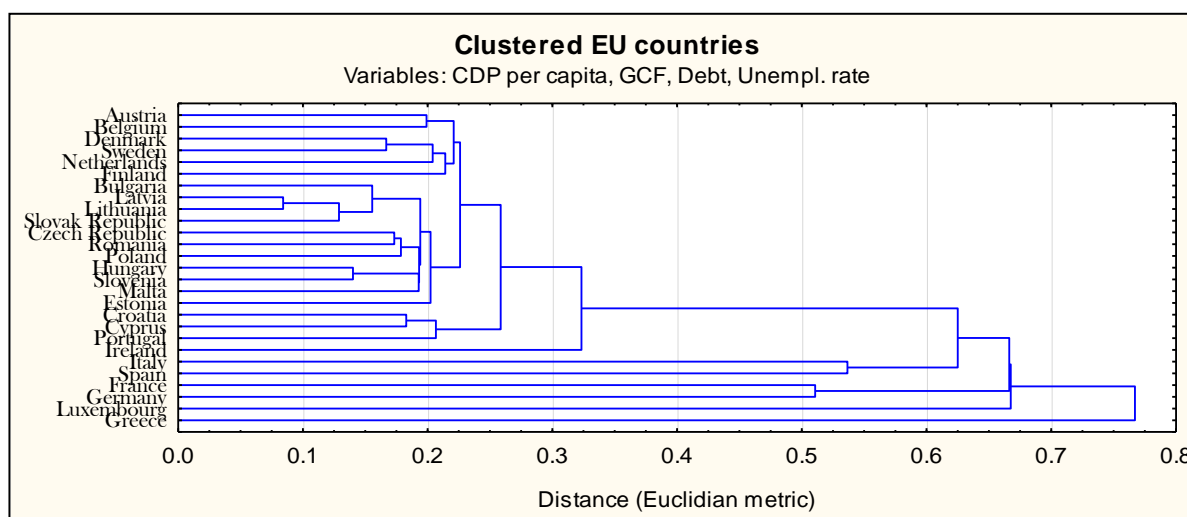
Graph 1 Comparison of original and standardized data (on the example of Unemployment Rate)

Source: own processing

First, we applied Hierarchical clustering to get an idea about the position of V4 countries and then we ran a k-means analysis, to discuss the properties of clusters in more detail.

3 Results

The hierarchical clustering was used to illustrate the clustering of EU countries, see graph 2. First, the old EU countries come together, except Germany and France, which form a separate cluster as Italy and Spain. This fact is due to the high level of gross capital formation (GCF) compared to other states. Luxembourg forms a separate cluster due to its high GDP per capita compared to others EU states. Greece also forms a separate cluster, but due to the high level of debt. The new countries of the European Union tend to come together. The hierarchical clustering shows that the positions and developments of V4 countries are quite similar.



Graph 2 Clustered EU countries
Source: own processing

To analyze the similarities and differences among countries in detail, we applied k-means cluster analysis, according to results in hierarchical methods, we decided for the choice of 8 clusters, the division into clusters, we can see in the following table 3.

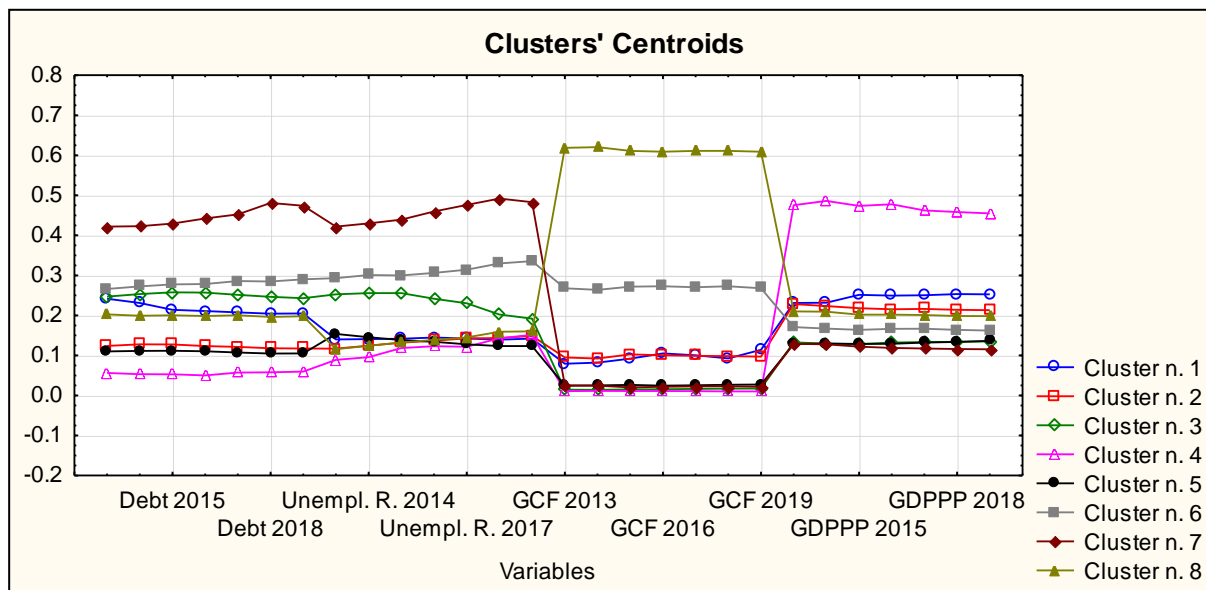
Cluster n. 1	Austria, Belgium, Ireland
Cluster n. 2	Denmark, Finland, Netherlands, Sweden
Cluster n. 3	Croatia, Cyprus, Portugal
Cluster n. 4	Luxembourg
Cluster n. 5	Bulgaria, Czechia, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovak Republic, Slovenia
Cluster n. 6	Italy, Spain
Cluster n. 7	Greece
Cluster n. 8	France, Germany

Table 3 Clusters of EU states
Source: own processing

As the results show, the division into old and new states of the European Union is still evident – e.g. Austria, Belgium, Denmark, Finland, Netherlands, Sweden form cluster n. 1 and n. 2, Italy and Spain together form cluster n. 6 or France and Germany together form cluster n. 8. Cluster number 5 consists of 11 cases including V4 countries. All these cases belong to new EU member states. Traditional member states create clusters together (instead of Portugal, which is with Malta and Cyprus). It supports the idea that differences between the traditional and new EU member states may be influenced by their position in a global economy. The new member states are attractive for multinational companies due to lower labor and production expenditures [15]. Even new members attract companies by tax competition [19]. According to [16], the new EU member states are approaching to results of the old member states (reference years 1996-2017, indicator GDPP), different in the speed of moving.

In all EU member states, the unemployment rate between 2013 – 2019 was decreasing. The position of V4 countries (the reference year 2016) was under the average of EU (6,7%) and Euro area (6,9). The Czechia had the lowest rate of the whole EU (2%). Poland and Hungary unemployment rates were oscillating around 3%. Slovakia was placed in the order of EU members in the second half with 5,8% (Eurostat, 2019). GDPP had an increasing trend in all EU countries between 2013 and 2019. The highest growth was recorded in Czechia from V4 countries. A similar improvement had Poland and Hungary and Slovakia placed last position [5, 6, 25]. The less indebted country from V4 is Czechia. The highest debt from V4 countries has Hungary, even Hungary exceeds the limit from the Maastricht agreement (60% of GDP). All V4 countries have general gross government debt under the EU average [6].

We can see that all V4 countries are in one cluster. To study the specification of individual clusters, see graph 3, where the values of clusters' centroids are displayed.



Graph 3 Clusters' centroids
Source: own processing

In more detail, graph 3 shows centroids of each cluster. Therefore, we can see that countries of cluster n. 5 have one of the lowest debts. Moreover, their debt is relative compared to the debt of other countries, decreasing. The same for the unemployment rate. On the other hand, also GDP per capita and GCF are the lowest ones (in GDP, only Greece (cluster n. 7) is worse than the centroid of cluster n. 5). If we run k-means analysis for 9 clusters, it divides cluster number 5 into two separate clusters, the first cluster would consist of the Czech Republic, Hungary, Romania, Malta, Poland, and Slovenia and the second cluster from Baltic countries together with Slovakia and Bulgaria. The first cluster has a better level in macroeconomic indicators like the unemployment rate, GDP per capita and gross capital formation than the second cluster. In contrast, the second cluster has a better level of general government gross debt.

4 Conclusions

The V4 countries reach worse values in GDP and GCF than old EU member states. But both indicators are relatively (in comparison with other EU countries) increasing within reference years in V4 countries. Their good results are influenced by a low indebtedness and a low unemployment rate under the EU average. In the next years, these good results could be affected by the economic situation accompanying the pandemic situation of COVID-19.

The position of V4 countries is in a cluster with the lowest debt, even relatively decreasing, the same unemployment rate is low and it is decreasing. V4 countries have many similarities due to the common history, the best position from them takes place Czechia. To compare with EU member states, V4 countries have the lowest employment rate and low level of indebted (except Hungary).

The V4 countries are approaching the results of the old member states what complies with the findings of [16]. V4 countries are leaders of new EU member states and have become an important source of improving the international competitiveness of the largest economy in the EU.

Acknowledgements

This contribution was supported by research grant GAJU No. 121/2020/S “Principles of circular economics in regional management leading to increased efficiency of systems”.

References

- [1] Anderson, H. J., Stejskal, J. (2019). Diffusion efficiency of innovation among EU member states: a data envelopment analysis. *Economies*, 7(2), 34.

- [2] Bacik, R., Kloudova, J., Gonos, J., & Ivankova, V. (2019). Management of Competitiveness and Economic Performance Based in the V4 countries.
- [3] Civelek, M., Ključnikov, A., Krajčík, V., Žufan, J. (2019) The Importance of Discount Rate and Trustfulness of A Local Currency for the Development of Local Tourism. *Journal of Tourism and Services*, 10(19), 77-92, <https://doi.org/10.29036/jots.v10i19.117>.
- [4] Cook, W. D., Seiford, L. M. (2009). Data envelopment analysis (DEA)–Thirty years on. *European journal of operational research*, 192(1), 1-17.
- [5] Eurostat. (2019). Statistics explained: Unemployment by sex and age – annual data. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Unemployment_statistics_and_beyond
- [6] Eurostat. (2020). Main page. Available at: <https://ec.europa.eu/eurostat/web/main/home>
- [7] Fare, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *The American economic review*, 66-83.
- [8] Hermoso-Orzáez, M. J., García-Alguacil, M., Terrados-Cepeda, J., Brito, P. (2020). Measurement of environmental efficiency in the countries of the European Union with the enhanced data envelopment analysis method (DEA) during the period 2005–2012. *Environmental Science and Pollution Research*, 1-25.
- [9] Ivanová, E., & Masárová, J. (2018). Performance evaluation of the Visegrad Group countries. *Economic research-Ekonomska istraživanja*, 31(1), 270-289.
- [10] Klicnarová, J., Adamová, M., & Soukupová N. (2021). Efficiency analysis of EU Member States in context of population aging. Submitted to *Central European Journal of Operations Research*.
- [11] Łozowicka, A. (2020). Evaluation of the Efficiency of Sustainable Development Policy Implementation in Selected EU Member States Using DEA. The Ecological Dimension. *Sustainability*, 12(1), 435.
- [12] Marková, J., & Švihlíková, I. (2019). Comparison of External Indebtedness and Debt Sustainability Development in V4 Countries. *European Journal of Transformation Studies*, 7(2), 113-127.
- [13] Martić, M., Savić, G. (2001). An application of DEA for comparative analysis and ranking of regions in Serbia with regards to social-economic development. *European Journal of Operational Research*, 132(2), 343-356.
- [14] Melecký, L., & Staničková, M. (2012). National efficiency evaluation of Visegrad countries in comparison with Austria and Germany by selected DEA models. In *Proceedings of 30th International Conference Mathematical Methods in Economics* (pp. 575-580).
- [15] Melecký, L., Staničková, M., & Hančlová, J. (2019). Nonparametric Approach to Evaluation of Economic and Social Development in the EU28 Member States by DEA Efficiency. *Journal of Risk and Financial Management*, 12(2), 72.
- [16] Młynarzewska-Borowiec, I. (2020). Income Gap between the New and Old EU Member States and Its Determinants in the Period 1996–2017. *Ekonomista*, (3), 401-430.
- [17] Molendowski, E., & Folfas, P. (2019). Effects of the Pillars of Competitiveness on the Competitive Positions of Poland and the Visegrad Group Countries in the Post Accession Period. *Comparative Economic Research. Central and Eastern Europe*, 22(2), 55-67.
- [18] Moutinho, V., Madaleno, M., & Robaina, M. (2017). The economic and environmental efficiency assessment in EU cross-country: Evidence from DEA and quantile regression approach. *Ecological Indicators*, 78, 85-97.
- [19] Podvieszko, A., Parfenova, L., & Pugachev, A. (2019). Tax competitiveness of the new EU member states. *Journal of Risk and Financial Management*, 12(1), 34.
- [20] Schwarcz, P., Kováčik, M., & Valach, M. (2021). The Development of Economic and Social Indicators in V4 Countries. *Acta Polytechnica Hungarica*, 18(2).
- [21] Segota, A., Tomljanović, M., Hudek, I. (2017). Contemporary approaches to measuring competitiveness—the case of EU member states. *Journal of Economics and Business*, 35(1), 123-150.
- [22] Staničková, M., Melecký, L. (2016). Malmquist index approach to efficiency analysis in selected old and new EU Member States. *Ekonomická revue – Central European Review of Economic*, 19 (87-104). 1805-9481. DOI: 10.7327/cerei.2016.09.02.
- [23] Stanković, J. J., Marjanović, I., & Stojković, N. (2021). DEA Assessment of Socio-economic Development of European Countries. *Management: Journal of Sustainable Business & Management Solutions in Emerging Economies*, 26(1).
- [24] Stephan, A., Happich, M., & Geppert, K. (2005). Regional disparities in the European Union: convergence and agglomeration (No. 2005, 4). Working Paper Series.
- [25] Worldbank. (2020). World Development Indicators. Available at: <https://databank.worldbank.org/>
- [26] Yamen, A., Allam, A., Bani-Mustafa, A., & Uyar, A. (2018). Impact of institutional environment quality on tax evasion: a comparative investigation of old versus new EU members. *Journal of International Accounting, Auditing and Taxation*, 32, 17-29.

Economic Policy Uncertainty and Stock Markets Co-movements

Peter Albrecht¹, Svatopluk Kapounek², Zuzana Kučerová³

Abstract. We empirically examine co-movements between the Economic Policy Uncertainty (EPU) index and selected stock market indices (S&P500, UK100, Nikkei225, and DAX30) at different investment horizons. We show significant but time-variant co-movements between EPU and stock markets employing wavelet analysis. Moreover, we identify EPU as a leading indicator of stock market drops, especially in the US, Japan, and Germany by using the time-varying domain based on the wavelet coherence. The lag between the changes of EPU and selected stock markets is from 4 months up to 32 months for longer investment horizons. We identify the co-movements between the EPU and stock markets also in times of decreased uncertainty but only to a small extent.

Keywords: Economic policy uncertainty, wavelet analysis, stock markets, investment horizons

JEL Classification: G01, G41

AMS Classification: 91G15

1 Introduction

There is a growing body of literature on the relationship between uncertainty and capital market returns. The uncertainty is generally measured using the Economic Policy Uncertainty (EPU) index published by Baker et al. (2016). As Pástor and Veronesi (2013) state, political uncertainty can be interpreted as uncertainty perceived by investors and connected with future policy actions of the government. Authors demonstrate that stock returns are affected by both fundamental economic and political shocks and show that political uncertainty increases the volatility of stock returns.

Increasing uncertainty in markets is related to higher exchange rate risk (Abid, 2020; Albuлесcu et al., 2019), but also FDI's (Canh et al., 2019). Uncertainty affects the exchange rates of emerging countries (Abid, 2020), i.e. that in time of higher uncertainty exchange rate risk appears which could impact fund, stock, and bond investments in emerging markets. Firms are not sure what impact it may have on demand, so they are more risk-averse (Tran, 2019). They decrease innovations also because of their risk aversion (He et al., 2020). Anyway, uncertainty is also significantly related to firms asset structures – companies prefer to own fewer assets in foreign currencies during times of higher uncertainty (Huang et al., 2019).

Increasing uncertainty is often related to decreasing stock prices (e.g. Antonakakis et al., 2013; Tiwari et al., 2019; Luo and Zhang, 2020), however, there is still some gap in literature and we focus on comovements of uncertainty and stock market returns at different investment horizons⁴. We follow Karp and Vuuren (2019) who show that investors react to uncertainty in financial markets in different ways and with different lags.

We follow this stream of literature and provide contribution in several ways. First, we employ a continuous and discrete wavelet transformation to identify the persistence (i.e. the cyclical behavior) of the EPU index and stock markets indices at different investment horizons. Second, we empirically examine time-varying comovements between the EPU index and stock market drops, especially during following events: Black Monday in 1987, the Gulf wars in 1990 and 2003, the Japan crisis in 1989, the DotCom bubble, the terrorist attacks in the US in 2001, the Global financial crisis in 2007, the Greek debt crisis in 2009, the Brexit referendum in 2016, and the

¹ Mendel University in Brno, Faculty of Business and Economics, Zemědělská 1, 613 00 Brno, Czech Republic, peter.albrecht@mendelu.cz.

² Mendel University in Brno, Faculty of Business and Economics, Zemědělská 1, 613 00 Brno, Czech Republic, kapounek@mendelu.cz.

³ Mendel University in Brno, Faculty of Business and Economics, Zemědělská 1, 613 00 Brno, Czech Republic, zuzana.kucerova@mendelu.cz.

⁴ Investment horizons are given by fractal dynamics. Effects of market fractions are significant when buying and selling orders are not efficiently cleared very often (Peters, 1994).

president Trump election in 2016. Third, we employ phase shift and wavelet cross-correlation sequences to provide detailed analysis of lags. We confirm prevailing co-movements at frequencies below 32 months and lags between 2 and 6 months. Thus, our results identify the EPU index as a leading indicator of stock market drops for long investment horizons.

2 Data and Methods

The co-movements between EPU and stock market returns is observed for the stock markets of major developed countries, the US, Great Britain, Japan, and Germany in the period 1985–2019 (monthly data)⁵. Stock markets are represented by the following indices: S&P500, UK100, Nikkei225, and DAX30.

We employ logarithmic differences to ensure stationarity for all the selected time series. We examine co-movements between EPU and stock markets employing wavelet analysis. This approach allows us to differentiate between frequencies representing different investment horizons (Fidrmuc et al., 2019).

3 Results

Most of the previous papers describe only periods when the uncertainty is rising (Luo and Zhang, 2020; Stolbov and Shchepeleva, 2020; Chen and Chiang, 2020) but this paper analyzes not only periods with rising uncertainty but also periods when there are peak seasons or the periods with decreasing uncertainty. As we observe, the relationship between stock indices returns and the uncertainty is very volatile, and that the uncertainty is related to stock markets returns differently for each market fraction. Even in times of higher uncertainty, it is not so possible to strictly conclude that the rising uncertainty causes the drop in stock market returns. As such, there is not a single causality between uncertainty and stock market returns. In this context, the contribution of this paper is very clear as this paper defines mutual coherence of stock indices returns and uncertainty for every period in time and every market fraction separately.

A more detailed analysis of stock market returns, taking the existence of uncertainty (using the EPU index) into account, is performed in the time-frequency domain applying the Wavelet coherence enabling the detection of co-movement; results are presented in Figure 1.

An important finding of the analysis of the returns of S&P500 and the uncertainty (Figure 1, upper-left sector) applying wavelet coherence and co-movement is that there is no exact co-movement in statistically significant periods. There are mostly two situations prevailing in the figure and that is an arrow pointing either to the bottom-left direction (an increase of the uncertainty indicates a decrease of the stock market index, i.e. the uncertainty is a leading indicator of the stock market index) or an arrow pointing to the top-left direction (the stock market index is a leading indicator of the uncertainty, therefore, when the stock market grows, the uncertainty decreases). The second type of behavior is interpreted in two ways: (1) the investors can see the positive potential in financial markets and, as such, the uncertainty about future growth decreases (Bonsall, et al. 2020); (2) the long-term investors look at different fundamentals and before these fundamentals are announced in newspaper articles, the markets have already incorporated this information into prices. That originates in the fact that long-term investors look at the information more difficult to process and release. Both explanations offer reasonable arguments so they could be valid both at the same time.

In Figure 1 we can see that the EPU index is a leading indicator of future returns of the S&P500 index for short-investment horizons (about 16-months and shorter), except for the period of the Global financial crisis in 2007. There is also no significant relationship between the EPU index and stock market returns for market fractions lower than 4 months.

Periods of significant historical coherence of the uncertainty and stock market returns occur in four periods while three of these periods are related with crises. The first period, characterized by the biggest one-day drop of indices, is identified the Black Monday in 1987 and our results show that the EPU index influences future returns of the S&P500 index in short investment horizons (frequency scales 4-8 months).

⁵ The US data begin in January 1985; Japan and the EU data begin in January 1987 and the data for the GB begin in January 1998. We use European EPU index as a proxy measurement of economic policy uncertainty in Germany.

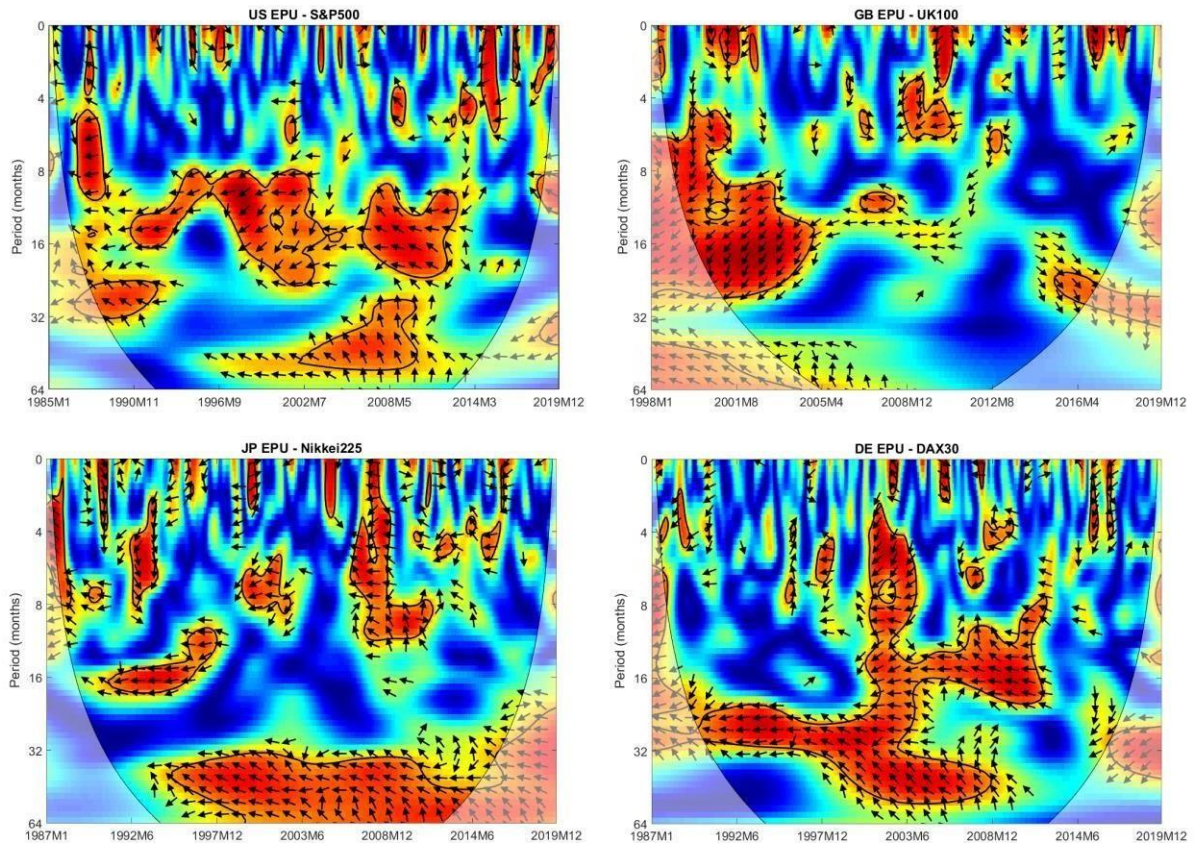


Figure 1: Wavelet Coherence for S&P500, UK100, Nikkei225, and DAX30.

Note: The color scales represent wavelet coherencies; the black contours denote insignificance at five percent against red noise, and the light shading shows the regions probably influenced by the edge effects. The direction of the relationship (the leading indicator) is represented by arrows (a left arrow denotes an antiphase (180°) while a right arrow denotes in-phase (0° or 360°). A downward-pointing arrow indicates as a leading indicator of stock market returns.

Source: own estimation.

In the case of longer horizons, the S&P500 index is a leading indicator for the frequency scales from 16 to 32 months. This relationship is valid for the whole tested period which could mean that the EPU index increases the uncertainty temporarily up to 16 months but in the case of long-term horizon, the stocks are more capable to determine future market direction and to determine the perception of the uncertainty itself on the markets.

The second significant period is the period from 1997 to 2002; this period includes several economic policy shocks having an impact on the rise of uncertainty (the Russian crisis, the President Bush's election, the DotCom bubble, and the terrorist attacks in the US). The uncertainty on the markets is high because of these events and, as Figure 1 shows, the EPU index is a leading indicator of the S&P500 index. Thus, the increasing uncertainty weakened the decision-making of companies and as a result, the stocks dropped.

The third significant period dates from 2002 to 2007 and can be characterized as the period of booming markets when the performance of stocks affected future uncertainty changes. The more detailed explanation is that the uncertainty rises very quickly during economic shocks and then decreases rapidly. Based on the fact, that there are not any negative events, investors perceive that the economy is in a good condition without any signs of possible present or future uncertainty.

The last period started in 2007. The S&P500 index affected the uncertainty during the Global financial crisis in 2007 and 2008 but the explanation seems to be different in this case. The uncertainty was high, but this crisis was affected by the whitewashing of information by banks and companies. Therefore, the crisis appeared when it was clear that the companies and banks had held very poisonous assets and it forced the stock prices to drop severely when markets noticed the problem (Ramskogler, 2015).

When we look at the results for wavelet coherence of the UK100 index and the GB uncertainty (Figure 1, upper-right sector), it is apparent that there are surprisingly very small significant areas of coherence; the values

of coherence for frequency scales shorter than 4 months are very fragmented. When compared with the S&P500 index, our results show similar characteristics for frequency scales from 4 to 16 months when the EPU index can be considered to be a leading indicator of the UK100 returns for most of the analyzed period.

The most significant area is partially shortened by a cone of influence; it is particularly the period during the DotCom bubble and the terrorist attacks in 2001. When we look closer at this coherence area, however, we can see that the coherence of the UK100 index and the EPU index is significant until 2005. It is probably due to the Gulf war 2 that started in 2003 and the GB army participated in this conflict.

The UK100 index was a leading indicator of the EPU index changes for 64 and higher frequency scales after the Global financial crisis in 2007 and the EPU index was a leading indicator of the UK100 returns for fractions up to 6 months. But there is still a notable area with no significant coherence between the EPU index and stock market returns.

We can also see the periods of low uncertainty with no significant coherence in the post-crisis period after 2011. The significant area of coherence is related to the Brexit referendum in 2016 when the only significant period is detected with a positive impact of the higher uncertainty on the UK100 returns. The explanation of this surprising phenomenon is quite simple; after the GB had decided to leave the EU, the GBP/EUR exchange rate dropped immediately by 10% of its value but no trade agreements with the EU were canceled. This depreciation of the British pound started an uptrend on stocks of the GB companies. However, after the first recalculations of the impact of Brexit on the GB economy had been released the prices were corrected back on the previous values.

Therefore, the impact of the EPU index on the UK100 index was not direct as it affected the GBP value first and the depreciation helped the stock values to rise.

The coherence between the uncertainty and stock returns is the most significant in Japan (Figure 1, bottom-left sector). A remarkable fact is that the significant relationship is observable particularly for the frequency scales above 32 months, and it is valid for almost the entire analyzed period. The Nikkei225 index can be considered to be a leading indicator of the EPU index for the whole period that includes the DotCom bubble, the Global financial crisis in 2007, and Brexit that was also important for Japan.

Besides that, there are significant areas for frequency scales lower than 16 months, and these areas are significant in the case of the increased uncertainty; these areas are related to the 1990 real estate bubble, then Japan came through several difficult moments to revive the Japanese economy from 1991 to 1999. The significant coherence is observable from 2000 to 2002 after the DotCom bubble, then also during the Global financial crisis in 2007 and to a lesser extent during the Brexit referendum.

The fact that the mutual coherence of the EPU index and the Nikkei225 index is significant only in times of the increased uncertainty indicates that it is of high importance to focus on this relationship during times when economic policy shocks occur. The EPU index serves as a leading indicator of the Nikkei225 index in frequency scales up to 16 months except for the Global financial crisis in 2007 and these findings are very similar to that of the US and the GB markets. The uncertainty did not take into account information affecting companies because this information was published when firms had financial problems or bankrupted (e.g. Lehman Brothers). As such, the uncertainty itself appeared very late when problems of big firms, banks, and funds were serious, and as a result stock returns became a leading indicator for the uncertainty during this period (Ramskogler, 2015).

In the case of the EU results (Figure 1, bottom-right sector), we can detect results similar for all four analyzed regions, and it is the existence of significant coherence particularly during economic shocks. We can also state that EPU index is a leading indicator of the DAX30 index in most cases for frequency scales from 4 to 16 months. When compared with the other markets for long-term investors – the DAX30 index is a leading indicator of the EPU index for frequency scales higher than 16 months. Information for long-term investors is more difficult to process as it is released in newspapers later when the market has already reacted to other economic fundamentals (Peters, 1994).

For frequency scales from 4 to 16 months, the EPU index affects DAX30 returns negatively. Again, this finding concerning the European financial markets is similar to the other three markets; these results are significant particularly in crisis periods (in 2000-2002 or in 2007). We identify the returns of the DAX30 indices to be a leading indicator of the EPU index (and not the uncertainty to be a leading indicator for returns) for this particular period. This specific information is more related to the DAX30 index because this information whitewash was operated by German banks and companies on a large-scale. These companies were buying securitized securities issued by the US banks, and these banks held subprime loans and mortgages in their portfolios. Therefore, the DAX30 index dropped, and as a result, the crisis began, and the uncertainty boomed (Ramskogler, 2015).

4 Conclusion

Our results confirm significant but time-variant co-movements between the EPU index and stock markets, especially during times of turbulence (Black Monday in 1987, the DotCom bubble, terrorist attacks in 2001, the Global financial crisis in 2007 etc.). We interpret frequency scales as investment horizons and confirm time-varying co-movements of the EPU index and stock market. Our results detect the co-movement for investment horizons between 4 and 32 months. Thus, we confirm that the EPU index serves as a leading indicator of stock market drops at short-term investment horizons.

Therefore, we offer some investment recommendations for periods of uncertainty increases as there is an opportunity to speculate on stock indices drops through hedge funds, options, and other instruments. However, it is necessary to make this investment for at least 16 months ahead; for more risk-averse investors, these findings might at least indicate the time to close buy positions or to sell the assets. Investors look at different trading fundamental indicators in case of long-term investments (Peters, 1994), and the prices of stocks are moved sooner after this information is released.

Third, we employ the phase shift and wavelet cross-correlation sequences to examine the lags (at different frequency scales) between EPU and stock market indices. We confirm prevailing lags between 2 and 6 months as such, these results signal that the EPU index serves as a leading indicator of stock market drops.

Our results are in line with recent literature (Baker et al., 2016; Tiwari et al., 2019; Luo and Zhang, 2020; Stolbov and Shchepeleva, 2020) and provide robust evidence of coherence between the uncertainty and stock market returns. However, we contribute with distinguishing investment horizons separately.

Acknowledgements

This research was funded by the internal grant agency of the Mendel University, grant no. PEF_TP_2020008. Svatopluk Kapounek was supported by the Czech Science Foundation, grant No. 20-17044S.

References

- [1] Abid, A. (2020). Economic policy uncertainty and exchange rates in emerging markets: Short and long runs evidence. *Finance Research Letters*, vol. 37, article no.101378.
- [2] Albuлесcu, C. T., Demirer, R., Raheem I. D., and Tiwari, A. K. (2019). Does the U.S. economic policy uncertainty connect financial markets? Evidence from oil and commodity currencies. *Energy Economics*, vol. 83(C), pp. 375-388.
- [3] Antonakakis, N., Chatziantoniou I., and Filis G. (2013). Dynamic co-movements of stock market returns, implied volatility and policy uncertainty. *Economics Letters* vol. 120, no. 1, pp. 87-92.
- [4] Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics*, vol. 131, no 4, pp. 1593-1636.
- [5] Bonsall, S. B., Green J., and Muller K. A. III. (2020). Market uncertainty and the importance of media coverage at earnings announcements. *Journal of Accounting and Economics*, vol. 69, no. 1, article no. 101264.
- [6] Canh, N. P., Binh, N. T., Thanh, S. and Schinckus, Ch. (2019). Determinants of foreign direct investment inflows: The role of economic policy uncertainty. *International Economics*, vol. 161, pp. 159-172.
- [7] Fidrmuc, J., S. Kapounek, and Junge, F. (2019). Cryptocurrency Market Efficiency: Evidence from Time-Frequency Analysis. *Czech Journal of Economics and Finance*, vol. 70, no. 2, pp. 121-144.
- [8] Chen, X., and Chiang, T. C. (2020). Empirical investigation of changes in policy uncertainty on stock returns—Evidence from China’s market. *Research in International Business and Finance*, vol. 53, article no. 101183.
- [9] He, F., Y. Ma and Zhang, X. (2020). How does economic policy uncertainty affect corporate Innovation?—Evidence from China listed companies. *International Review of Economics & Finance*, vol. 67, pp. 225- 23.
- [10] Huang, J., Luo, Y. and Peng, Y. (2019). Corporate financial asset holdings under economic policy uncertainty: Precautionary saving or speculating? *International Review of Economics & Finance*, In Press.
- [11] Karp, A., and van Vuuren, G. (2019). Investment implications of the fractal market hypothesis. *Annals of Financial Economics*, vol. 14, no. 1, article no. 1950001.
- [12] Luo, Y. and Zhang, C. (2020). Economic policy uncertainty and stock price crash risk. *Research in International Business and Finance*, vol. 51, article no. 101112.
- [13] Peters, E.E. (1994). *Fractal market analysis: Applying Chaos Theory to Investment and Economics*. John Wiley & Sons.

- [14] Ramskogler, P. 2015. Tracing the origins of the financial crisis. *OECD Journal: Financial Market Trends*, vol. 2014/2, pp. 47-61.
- [15] Stolbov, M. and Shchepeleva, M. (2020). Systemic risk, economic policy uncertainty and firm bankruptcies: Evidence from multivariate causal inference. *Research in International Business and Finance*, vol. 52, no. 101172.
- [16] Tiwari, A. K., Jana, R. K., Roubaud, D. (2019). The policy uncertainty and market volatility puzzle: Evidence from wavelet analysis. *Finance Research Letters*, vol. 31, pp. 278-284.
- [17] Tran, Q. T. (2019). Economic policy uncertainty and corporate risk-taking: International evidence. *Journal of Multinational Financial Management*, vol 52–53, no. 100605, pp. 1-9.

Optimal consumption with irreversible investment in the context of the Ramsey model

Nastaran Ansari¹, Adriaan Van Zon², Olaf Sleijpen^{3,4}

Abstract. The Ramsey model is widely used in macroeconomic studies, but only few studies consider the irreversibility of investment which is an important aspect of real-life investment. In this paper, we consider a Ramsey model with irreversible investment. While the main focus of the current literature is on a qualitative discussion of such problems, our paper provides a framework for quantitative analysis of the transition path. Finding the optimal transition path in a Ramsey model with irreversible investment requires solving a multistage optimal control problem with two kinds of stages. These stages are called ‘free’ and ‘blocked’ intervals in the literature, with zero gross investment in the blocked interval and positive gross investment in the free interval. We show that the optimality conditions for such a problem imply the continuity of the control variable along the transition path, which is an important feature in finding the switching moments between free and blocked intervals. We use this feature and the backward integration method for a quantitative analysis of the transition path.

Keywords: Ramsey model, Optimal control theory, multistage optimal control problem, Irreversibility of investment

JEL Classification: P28, C610

AMS Classification: 49K04

1. Introduction

The Ramsey model is one of the most popular and widely used models in macroeconomic studies. It has been constructed by Ramsey [1] to overcome one of the shortcomings of the Solow- Swan model by removing the constant saving rate assumption and allowing households to optimize their saving and consumption behaviour.

In the Ramsey model, output is a function of the capital stock and can be used for consumption and investment purposes. In this model, due to discounting, a unit of consumption which occurs at a later time leads to less utility. However, investment through postponing consumption, increases the capital stock which leads to higher consumption in the future. Intertemporal optimization is used to determine the optimal amount of consumption over time under a macro-economic budget constraint.

Most of the studies using a Ramsey-type model are based on the assumption of reversible investment. It means that consumption can temporarily exceed output, because it is possible to decumulate capital for consumption purposes. This assumption is unrealistic in many cases. In reality, once investment in fixed capital has taken place, it can't be used for consumption purposes anymore, hence, gross investment in capital must be non-negative.

Arrow and Kurtz [2] implement the irreversibility of investment in the Ramsey model by means of a non-negativity constraint on gross investment. In order to find the optimal transition path toward the steady state, they define two kinds of intervals: free intervals, where the non-negativity constraint on gross investment is not binding and blocked intervals, where the non-negativity constraint is binding. They consider the order of the free and blocked intervals but they effectively disregard the duration of these intervals as well as the timing of the switching moments between the intervals.

Rozenberg et al. [3] discuss a Ramsey model with irreversible investment and a constraint on accumulated pollution. They provide a qualitative discussion based on optimal control theory but in order to find a quantitative solution they use the GAMS solver as a black box.

With respect to the literature, there is still a lack of a systematic framework that facilitates quantitative analysis of the transitional dynamics in the Ramsey model with irreversible investment. Our paper aims to shed more light on handling this problem by formulating it as a multistage optimal control problem.

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.

To this end, we will provide a relatively simple picture of the problem by representing the Ramsey model with irreversible investment as a three-stage optimal control problem. First, we show that this perspective allows us to specify the switching moments between free and blocked intervals. Secondly, we use the backward integration method, introduced by Brunner and Strulik in [4], to analyze the transition path toward the steady-state in a quantitative manner.

2. Method

2.1. The Ramsey model with irreversible investment

In this section we discuss a Ramsey model with irreversible investment. The per capita Cobb-Douglas production function is given by⁵:

$$y = f(k) = A \cdot k^\alpha, \quad (1)$$

where y is output, k is capital, A represents total factor productivity, while α is the partial output elasticity of capital. Utility, $u(c)$, is a function of consumption, c , is given by:

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma}, \quad (2)$$

where σ is the elasticity of substitution⁶. The central planner's goal is to maximize the discounted value of total utility:

$$\max \int_0^\infty \frac{c^{1-\sigma}}{1-\sigma} e^{-\rho t} dt, \quad (3)$$

Subject to:

$$\dot{k} = i - \delta \cdot k, \quad (4)$$

$$y = i + c, \quad (5)$$

$$i \geq 0. \quad (6)$$

equations (4), (5) and (6) represent the equation of the motion for capital, the budget constraint and irreversibility constraint, respectively, where δ is the depreciation rate, and i represents the rate of gross investment in fixed capital. The Hamiltonian of this problem is:

$$H = \frac{c^{1-\sigma}}{1-\sigma} \cdot e^{-\rho t} + \lambda \cdot \dot{k}, \quad (7)$$

The lagrangean of the problem is:

$$L = H + \mu \cdot (A \cdot k^\alpha - c), \quad (8)$$

From the first order conditions (FOC) of this Hamiltonian problem we have:

$$\frac{\partial L}{\partial c} = 0 \Rightarrow c = (e^{-\rho t} (\lambda + \mu))^{\frac{1}{\sigma}}, \quad (9)$$

$$\dot{\lambda} = -\frac{\partial L}{\partial k} = (\delta - A \cdot k^{-1+\alpha} \cdot \alpha) \cdot \lambda + \mu (\alpha \cdot A \cdot k^{\alpha-1}), \quad (10)$$

$$\mu \cdot (A \cdot k^\alpha - c) = 0, \quad (11)$$

where λ is the co-state variable associated with capital. μ represents the lagrange multiplier of the non-negative gross investment constraint and equation (11) implies that it is zero for $c < y$, i.e., for strictly positive gross investment we have $\mu = 0$. Equation (10) represents the equation of motion for λ .

⁵ For the sake of simplicity, we set the labour force equal to L_0 . Hence equation (1) can be regarded as a per capita production function.

⁶ We assume $\sigma > 1$ and $\sigma \neq 1$.

2.2. A three-stage optimal control version of the Ramsey model with irreversible investment

In this section we consider the Ramsey model with irreversible investment as a multistage optimal control problem. According the Arrow and Kurtz paper [2] in such a problem the transition path must adhere to proposition 1:

Proposition 1: “there exists \underline{k} such that for all $k \leq \underline{k}$ the optimal policy coincides with the reversible case while investment is zero for all $\underline{k} \leq k$ “[2].

The situation in which gross investment is strictly positive while $k \leq \underline{k}$, coincides with the irreversible case, is called a free interval. The situation in which the non-negativity constraint on gross investment is binding and $\underline{k} \leq k$ is called a blocked interval.

Arrow and Kurtz [2] do not actually specify the value of \underline{k} . However, we show that proposition 1 can be transformed to a three-stage optimal control problem that includes transversality conditions which can be used to obtain the value of \underline{k} .

By keeping the terminology used in this paper close to the one used by Arrow & Kurtz [2], we define a three-stage model as follows: the first stage is associated with a blocked interval; the second stage is associated with a free interval and the third stage starts when we obtain the steady state. Stage one and two together define the transition path. We assume the $T1$ represents the switching moment between the first stage and the second stage, and the $T2$ represents the switching moment between the second stage and the third stage. So, $T2$ also represents the moment that the steady state arrives. In the three-stage model, the social planners’ goal is to maximise total utility W subject to (4)-(6), where W is given by:

$$W = W_B + W_F + \phi(k_{SS}, T2) \quad (12)$$

In equation (12) W_B and W_F are the total welfare accumulated during the first stage which is a blocked interval and the second stage, which is a free interval, respectively. $\phi(k_{SS}, T2)$ represents the scrap value function which shows the total utility gained by remaining in the steady state from time $T2$ onward. $\phi(k_{SS}, T2)$ shows the maximum value of the welfare integral of future utility starting from time $T2$ with an initial capital stock k_{SS} that remains unchanged during the steady state. Equation (12) can now be rewritten as follows:

$$W = \int_0^{T1} \frac{c_B^{1-\sigma}}{1-\sigma} \cdot e^{-\rho t} dt + \int_{T1}^{T2} \frac{c_F^{1-\sigma}}{1-\sigma} \cdot e^{-\rho t} dt + \int_{T2}^{\infty} \frac{c_{SS}^{1-\sigma}}{1-\sigma} \cdot e^{-\rho t} dt, \quad (13)$$

where c_B , c_F and c_{SS} represent the time paths of consumption during the three-stage defined above. obviously, c_{SS} is constant during the steady state, while c_B and c_F vary over time.

Now we consider the optimal consumption path and the Hamiltonians associated with each stage. In the first stage, investment i is zero. So, we have⁷:

$$H_B = \frac{c_B^{1-\sigma}}{1-\sigma} \cdot e^{-\rho t} + \lambda_B \cdot (-\delta \cdot k_B), \quad (14)$$

$$y = c_B, \quad (15)$$

$$\dot{k}_B = -\delta \cdot k_B, \quad (16)$$

Equation (16) is the equation of the motion for the capital stock during the blocked interval. It follows that $k_{B,t}$ is given by:

$$k_{B,t} = e^{-\delta \cdot t} \cdot k_0, \quad (17)$$

where k_0 is the initial value of the capital stock. In order to find the optimal consumption path during the second stage, we use the FOC given by equations (9)-(11). In the second stage investment is positive and $\mu = 0$. Hence we have:

$$H_F = \frac{c_F^{1-\sigma}}{1-\sigma} \cdot e^{-\rho t} + \lambda_F \cdot (A \cdot k_F^\alpha - c - \delta \cdot k_F), \quad (18)$$

$$\frac{\partial L_F}{\partial c_F} = 0 \Rightarrow c_F = (e^{-\rho t} \cdot \lambda_F)^{1/\sigma}, \quad (19)$$

⁷ From now on we will be using the subscripts, B , F , and SS to denote a blocked interval, a free interval and the steady state, respectively.

$$\dot{\lambda}_f = -\frac{\partial L_F}{\partial k_F} = (\delta - A \cdot k_F^{-1+\alpha}) \cdot \lambda_f, \quad (20)$$

$$\hat{c}_F = -\frac{\hat{\lambda}_F + \rho}{\sigma}, \quad (21)^8$$

$$\hat{\lambda}_F = \delta - A \cdot k_F^{-1+\alpha} \alpha \quad (22)$$

Using equations (18)-(22), the equations of motion for the capital stock and the level of consumption during the free interval are given by:

$$\dot{c}_F = \frac{c_F \cdot (A k_F^{\alpha-1} \alpha - \delta + \rho)}{\sigma} \quad (23)$$

$$\dot{k}_F = A k_F^\alpha - c_F - \delta \cdot k_F \quad (24)$$

In the third stage, we are at the steady state and $\dot{k}_{SS} = 0$ and $\dot{c}_{SS} = 0$.

2.3. Finding the steady state

The steady state would be obtained at the end of the second stage. In the steady state \dot{c}_{SS} and \dot{k}_{SS} both should be zero. So, from equation (23) and (24) and assuming $k_{SS} > 0$ and $c_{SS} > 0$, the steady state value of c and k , (c_{SS}, k_{SS}) , are given by:

$$(c_{SS}, k_{SS}) = \left(\frac{\delta \cdot (1 - \alpha) + \rho}{\alpha} \cdot \left(\frac{A \cdot \alpha}{\delta + \rho} \right)^{\frac{1}{1-\alpha}}, \left(\frac{A \cdot \alpha}{\delta + \rho} \right)^{\frac{1}{1-\alpha}} \right) \quad (25)$$

2.4. Transversality conditions and switching moments between intervals.

In this section we show that the transversality conditions of the three-stage model that pertain to the optimal switching moment between the stages imply continuity of consumption. This means that \underline{k} in proposition 1 is on the point of intersection of the stable arm and the $y = c$ line.

In order to find the optimum switching moment between the stages, the transversality conditions require the equality of the Hamiltonians of two adjoining stages at the switching moment between them (see [5] and [7]). The economic interpretation of this transversality condition is that staying one more unit of time in each stage adds the (utility) value of the Hamiltonian at that moment to the total utility. So, if the Hamiltonians are not equal at the switching moment, then it means that staying longer or shorter in one of the stages may actually improve total utility. The gains and losses from lengthening each stage by one unit of time are as follows⁹:

$$\frac{\partial W_B}{\partial T1} = H_{B,T1} \quad (26)$$

$$\frac{\partial W_F}{\partial T1} = -H_{F,T1} \quad (27)$$

$$\frac{\partial W_F}{\partial T2} = H_{F,T2} \quad (28)$$

$$\frac{\partial \phi(k_{SS}, T2)}{\partial T2} = -H_{SS,T2} \quad (29)$$

At the switching moment between the blocked and free interval the Hamiltonians are given by:

$$H_{B,T1} = \frac{c_{B,T1}^{1-\sigma}}{1-\sigma} e^{-\rho t} + \lambda_{B,T1} \cdot (-\delta \cdot k_{B,T1}) \quad (30)$$

$$H_{F,T1} = \frac{c_{F,T1}^{1-\sigma}}{1-\sigma} e^{-\rho t} + \lambda_{F,T1} \cdot (y_{F,T1} - c_{F,T1} - \delta \cdot k_{F,T1}) \quad (31)$$

In equations (30) and (31) $\lambda_{B,T1}$ and $\lambda_{F,T1}$ shows the shadow price of capital immediately before and after the switching moment. In the current problem there is no pure state constraint and in such a situation it can be shown

⁸ $\hat{x} = \frac{d \ln(x)}{dt}$, i.e., \hat{x} is the instantaneous proportional growth rate of x .

⁹ We add time subscripts $T1$ and $T2$.

that there is no jump in the co-state variable at the switching moment between the stages implying that $\lambda_{B,T1} = \lambda_{F,T1}$. So, since $y_{B,T1} = c_{B,T1}$ (and $y_{B,T1} = y_{F,T1}$) the equality of the Hamiltonians at the switching moment implies that $c_{B,T1} = c_{F,T1}$. So, there is no jump in the consumption.

Based on proposition 1, the free interval coincides with the reversible case [2]. In the phase diagram of the Ramsey model with reversible investment the only path that leads to the steady state is the stable arm [1]. Hence, the continuity of the consumption implies that the switching moment between the blocked and the free interval is at the moment that the stable arm leading toward the steady state from the North-East intersects the $y = c$ line in the c, k phase-plane. Hence \underline{k} in proposition 1 is the value of k in the point of intersection of the stable arm and the $y = c$ line in the phase diagram of the Ramsey model.

In the Ramsey model with irreversible investment, it takes an infinite amount of time to reach the steady state, so in practice we have two stages only. The reason to introduce the scrap value function associated the third stage is that it can very effectively cover situations in which a steady state arrives at some finite date in the future, for example if there is a fixed ceiling on CO_2 emissions, which we have included in more extensive version of the current model (for more details see [6]).

In order to find the \underline{k} in proposition 1 we need to travel along the stable arm coming from the North-East in the Ramsey phase diagram. In order to do this, we use the backward integration method described by Brunner and Strulik in [4]. This method traces the solution from a value arbitrarily close to the steady state back to the initial condition by making time run backwards. In the steady state \dot{c} and \dot{k} are exactly equal to zero so a backward solution starting exactly from the steady state would never get away from it. In order to find the initial value for moving backward, first the slope of the stable arm should be determined then the initial value for the backward integration can be found as the point of intersection of a straight line through the steady state, with the slope of stable arm, and a circle with radius ε to the North-East of the steady state (for more details see [4]).

In order to make the discussion clearer, the next section presents a numerical example. The system of differential equations shown in equations (23)-(24) describes the optimal transition path toward the steady state. However, it is nonlinear and cannot be solved analytically, but it is possible to obtain a numerical solution as will be shown in the next section.

3. A numerical example

Figures 1 and 2, which are generated by using Mathematica (© Wolfram), show the phase diagram and the optimal consumption path, respectively, for the parameter set given below:

$$\{\delta = 0.05, \rho = 0.02, \alpha = 0.6, A = 4, \sigma = 0.3, k_0 = 20000\}$$

where k_0 represents the initial value of the capital stock. In Figure 1 point E shows the equilibrium point and the red line shows the optimal transition path from $(c=y, k = k_0)$ to the steady state. The transition path consists of a blocked interval at the beginning and then it switches to the free interval at the point of intersection of the stable arm and the $y = c$ line. In Figure 1 the $\dot{c} = 0$ and $\dot{k} = 0$ loci, obtained from equations (23) and (24), divides the phase plane into four regions. Black arrows show the direction of the motions of c and k in each region. Blue arrows show the resultant vectors of these motions. Because of the irreversibility of investment, the levels of consumption that are above the $y = c$ line are infeasible.

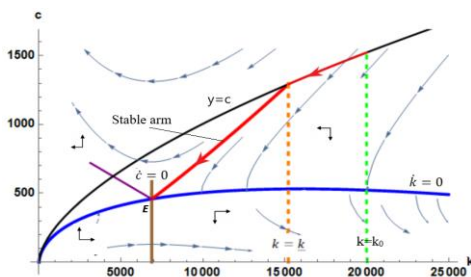


Figure 1 Phase diagram

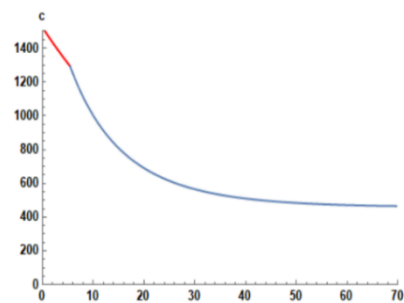


Figure 2 Consumption over time

4. A short discussion of the optimal order of intervals

In this section we try to clarify the reasons for the order of intervals proposed in proposition 1 by using the characterisations of the Ramsey phase diagram. If $k_0 > \underline{k}$ then the optimal transition path consists of two intervals, a blocked interval followed by a free interval. Now the question is that why this is the order of the intervals in proposition 1, why not to start with a free interval? If $k_0 > \underline{k}$, there are three possibilities to initiate the move to the steady state from the right-hand side of \underline{k} :

1. $c_0 > y$, however, this is ruled out by the irreversibility of investment assumption;
2. $c_0 = y$, which means staying in the blocked interval at the beginning and then switching to the free interval later.
3. $c_0 < y$, which means starting with a free interval. The direction of the motions, as implied by equation (23) and (24) implies that for $k_0 > \underline{k}$ choosing such a level of consumption would result in a move of the arrows far away from the steady state (as shown by the black and blue arrows in Figure 1). So, in order to hit the steady state, there must be a jump in consumption from the free interval to either the $y = c$ line or on to the stable arm directly.

As shown before, the optimality conditions imply continuity of consumption along the optimal transition path. Hence, it is not optimal to start with a free interval (like the third possibility) and the only remaining feasible option for choosing c_0 is the second option where $c_0 = y$.

5. Conclusion

Our paper provides a framework for the quantitative analysis of the transitional dynamics in the Ramsey model with irreversible investment. We look at the problem as a multistage optimal control problem with two kinds of stages namely free and blocked intervals. This perspective allows us to use a number of the transversality conditions to specify the switching moment between different types of intervals.

We show that transversality conditions in the proposed multistage problem imply that at the switching moment between the intervals there is no jump in consumption. This feature determines the optimal order of free and blocked intervals and also it implies that the optimal switching moment between the intervals is defined by the duration of the movement along the $y = c$ line up to the point of intersection with the stable arm.

By adding more constraints to the Ramsey model with irreversible investment, such as an emission ceiling or a constraint on available resources, the number and the order of the intervals could be different from the current problem. However, our paper provides a perspective on the handling of the irreversibility of investment constraint in the context of the Ramsey model in the simplest possible setting. This setting can easily be modified to handle more sophisticated versions of the problem where the optimal timing of the stages is a crucial aspect of the solution. In addition, if the steady state arrives at some finite moment in time, its optimal arrival time and its optimal initial stocks can be derived using the scrap value function that captures the utility value of the steady state in combination with transversality conditions regarding the optimality of this arrival time as well as the optimality of initial stocks, as we will show in a follow-up version of this paper.

References

- [1] Barro, R. J. and Sala-i-Martin, X. (2004) *Economic Growth*(2nd ed.), The MIT Press.
- [2] Arrow, K. J., & Kurtz, K. (1970). Optimal Growth with Irreversible Investment in a Ramsey Model. *Econometrica*, 38(2), 331–344.
- [3] 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>.
- [4] Brunner, M., Strulik, H. (2002), Solution of perfect foresight saddle point problems: a simple method and applications, *Journal of economic dynamics & control*, 26 (2002) 737-753
- [5] Leonard, D. Van Long, N. (1995) ‘Optimal control theory and static optimization in economics’, Cambridge university press.
- [6] Ansari, N. Van Zon, A. (2021) ‘Ramsey model with irreversible investment and emission ceiling’, Working paper.
- [7] Van Zon, A. and David, P. (2012) ‘Optimal multi-phase transition paths toward a global green economy’, *Unu-MERIT Working paper*, (31), pp. 1–51.

Impact of incorporating and tailoring PRINCE2 into the project-oriented environment

Bartoška Jan¹, Rydval Jan², Jedlanová Tereza³

Abstract. The paper proposes the use of the analytical network process (ANP) for quantification of the impact of incorporating the main topics of the international project management standard PRINCE2 into a semantic model, which displays the project management environment in a commercial project-oriented organisation. The semantic model is derived based on the organizational structure and the life cycle of a projects. The ANP network creates the basis for analysis of the preferences of the project roles, project documents, and other elements of the project management environment of the organization and analysis of their individual relationship to organization units. Using the ANP analysis the impact of incorporating the main topics of PRINCE2 into the project management environment of the organization is estimated and quantified. Furthermore, the ANP is used to perform a sensitivity analysis of the preferences of individual components of the semantic model when changing the importance of individual incorporated topics of PRINCE2 in the organization. The authors of the article build on their previous work and research activities in the field of project management.

Keywords:

Analytic Network Process; Corporate Organization; Multi-criteria Decision Making; PRINCE2; Project Management; Project Team Roles; Semantic Model; Sensitivity Analysis.

JEL Classification: C44

AMS Classification: 90C35

1 Introduction

International standard of project management PRINCE2 (PProject IN Controlled Environment) is the leading structured project management method in the United Kingdom, and it is used across the whole world in the private and public sector [14]. According to [6], PRINCE2 is a process based on project management methodology. PRINCE2 was developed to gain control at the start, during the progress and at the completion of projects. It is project management based on three constraints (time, quality and cost connecting in the Project Management Triangle). PRINCE2 divides projects into manageable and controllable stages. It is necessary to understand how to make all three project constraints adjust to each other to deliver a project within the scope.

The basis of the PRINCE2 method is to describe project management as planning, delegating, monitoring and control of all aspects of the project. It is necessary to achieve the project objectives within the expected performance targets for time, cost, quality, scope, benefits and risk. PRINCE2 clearly defines the roles and responsibilities of the project team members [6] and focuses on the product that has to be delivered. PRINCE2 has improved the product value especially in the field of IT projects [7], but also in other areas, for instance in the field of e-learning [2].

Nowadays, according to [7] business companies are shifting within their project-oriented environment to use the international project management methodology PRINCE2, not necessarily just in the field of IT projects. However, the implementation of PRINCE2 can also be a very difficult process, especially in terms of the dividing project into manageable and controllable stages [12]. And because the introduction of new ways and methods of project management can be a difficult process, it is necessary to define and describe the project environment of the organisation. Various methods and tools can be used for this, such as the semantic model, which has been used by

¹Czech 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, rydval@pef.czu.cz.

³ Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, jedlanova@pef.czu.cz.

[10], [16] or [5] and it is also possible to quantify individual parts of the project environment. The introduction and tailoring of the PRINCE2 into an organization always has a certain effect and impacts on the organization.

The aim of this paper is to quantify and evaluate the impact of the international project management standard PRINCE2's incorporation into an organization's project-oriented environment. The paper proposes the use of the Analytical Network Process (ANP) for quantification of the PRINCE2 incorporation's impact on to a commercial project-oriented organization. After PRINCE2 incorporation, the project team members must still decide on their own project management framework. Therefore, preferences of the project team roles are quantified and the stability of these preferences values are tested using sensitivity analysis. The results of the sensitivity analysis show the stability of the individual roles weights due to the changes in particular PRINCE2 principles' importance.

2 Materials and Methods

2.1 International standard of project management PRINCE2

PRINCE2 (PROjects IN CONTROLLED ENVIRONMENTs) is a structured project management method and practitioner certification programme [6]. PRINCE2 divide projects into manageable and controllable stages. PRINCE2 is a project management methodology of 7s. The principles, themes and processes all follow this model. PRINCE2 derives its methods from 7 core principles. Collectively, these principles provide a framework for good practice. The Principles are [6]: Continued Business Justification, Learn from Experience, Define Roles and Responsibilities, Manage by Stages, Manage by Exception, Focus on Products, Tailor to the Environment. Themes provide insight into how the project should be managed [6]. They can be thought of as knowledge areas, or how principles are put into practice. They are set up at the beginning of the project and then monitored throughout. Projects are kept on track by constantly addressing these themes – Business Case, Organisation, Quality, Plans, Risks, Changes and. Progress The PRINCE2 method also separates the running of a project into 7 processes. Each one is overseen by the project manager and approved by the project board. Here is a breakdown of each stage [6]: Starting Up a Project, Initiating a Project, Directing a Project, Controlling a Stage, Managing Product Delivery, Managing Stage Boundaries, Closing a Project.

2.2 Semantic model

A semantic model consists of a semantic (associative) network that [11] define as "natural graph representation". In the semantic network, each node represents individual objects of described world and edges connecting these nodes and represent relationships between these objects [21]. The term "semantic network" was for the first time used by [13] in his dissertation on the representation of English words and according to [15] are semantic networks suitable for displaying and expressing big information resources, management structures and processes or other areas.

2.3 Analytic Network Process

Many decision problems cannot be decomposed and structured hierarchically, i.e. they cannot be structured into an Analytic Hierarchy Process (AHP) model (see [18] for introduction to AHP theory), because they involve many interactions and dependencies of higher-level elements in a hierarchy on lower-level elements, i.e. these problems can be structured into a network. ANP is represented by a network, then as a hierarchy ([17], [19], [20]). Therefore, the Analytic Network Process (ANP) is a generalization of the Analytic Hierarchy Process. ANP can include the dependencies between the elements of different levels of the hierarchy as well as of the elements of the same level of the hierarchy (higher-level and lower-level elements in a hierarchy). The ANP model can reflect the increasing complexity of a network structure, where the network can be created from different groups of elements. Each group of elements creates a network cluster, which includes a homogeneous set of elements. Connections can exist between clusters as well as between the elements i.e. between the elements inside the cluster and between the elements from different clusters. In AHP for hierarchical trees, synthesized global priorities are calculated by multiplying the local priorities, which are determined via pairwise comparisons of the priority of the parent element. In ANP this process is replaced by the Limit Matrix calculation ([1], [17], [19], [20]).

The basic steps of the ANP method according to [19]:

- The first step is to create a network, which describes the decision problem. The ANP network shows dependency among decision elements.
- The second step is to conduct the pairwise comparisons of the elements within the clusters and among the clusters. ANP prioritizes not only decision elements but also their groups or clusters as it is often the case in

the real world. The consistency of these comparisons should be controlled. The consistency is measured by a consistency index defined by Saaty:

$$I_s = \frac{l_{\max} - n}{n - 1}, \quad (1)$$

where l_{\max} is the largest eigenvalue of Saaty's matrix and n is the number of criteria. Saaty's matrix is considered to be sufficiently consistent if $I_s < 0.1$.

- The third step is to construct the Supermatrix. The priorities derived from the pairwise comparisons are entered into the appropriate position in the Unweighted Supermatrix. This Supermatrix has to be normalized using clusters weights, the Weighted Supermatrix is calculated:

$$W = \begin{matrix} & C_1 & C_2 & \dots & C_N \\ \begin{matrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{matrix} & \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \dots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \dots & W_{nn} \end{bmatrix} \end{matrix} \quad (2)$$

where each block W_{ij} of the supermatrix consists of:

$$W_{ij} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix} \quad (3)$$

where:

$$\sum_1^n w_{ij} = 1, j \in \langle 1, n \rangle \quad (4)$$

- The fourth step is to compute the Limit Matrix and global preferences of decision elements are obtained. The Limit Matrix is used to obtain stable weights from Weighted Supermatrix. Raising the Weighted Supermatrix to powers generates the Limit Matrix, the powers will converge to a given matrix (the Limit Matrix), or the powers will converge to a cycle of matrices (the Limit Matrix is the average of these matrices). From Limit Matrix the final global values of priority (preferences) are gained. These preferences present the best decision selection. More information on standard steps of the Limit Matrix calculation are in [19] and see [1] for more information about algorithms for computing the Limit Matrix.
- Sensitivity analysis in the ANP. Sensitivity analysis is used to check the results obtained through the ANP model, i.e. to check the stability of preferences [8]. To start with the sensitivity analysis, elements having the highest preferences of the observed cluster are identified first. The impact of the increased value of the preference should be observed on all other elements of the cluster. The interpretation of the data obtained from the sensitivity analysis is, that as the input value, the priority of selected nod in the Unweighted Supermatrix, changes from 0 to 1 and the corresponding priorities of the alternatives, are computed from the Limit Supermatrix ([17], [19]). ANP in this paper processed by the SuperDecisions software ([22]). The software that implements the Analytic Network Process based on Thomas L. Saaty's work, was developed by William J. Adams working with Rozann W. Saaty.

3 Results and Discussion

3.1 Case study: Incorporating and tailoring of PRINCE2 in a commercial unit

The research in a commercial unit (the bank organization) took place from 2016 to 2020. The chosen organization is the international banking company with an extensive portfolio of banking services for corporate or personal clients - represents a typical corporate environment with developed project management. Within the research, a basic semantic model of project management was created as stated in [3] or [4] and further described and interpreted in [5], [10] or [16] – the model includes a complete network of project roles, departments, project documentation, project restrictions, etc.

The ANP for analysing the impact of incorporating the PRINCE2 into a project environment of a commercial unit is used as follows: The description of the project environment of a commercial unit is created using the semantic model of the this environment. Based on this description, a network for ANP is created consisting of clusters (sets

of indicators and elements) describing the project environment. It consists of five clusters: strategies, organisation, project documentation, limits, and team roles as stated in the previous research [3], [4], [10], [16] or [5], which we follow up with this article. Then pairwise comparison of individual elements within the cluster and between clusters is performed according to the ANP network settings. This yields the weights of the individual elements. If the comparison values are left equal to 1, i.e. the weights of the individual elements depend only on the network structure (the project environment structure). If the project environment is then changed, e.g. by implementing a new project management standard (PRINCE2), which can change the relationships between the cluster elements of the project environment, the pairwise comparison is performed again and the impact of the introduction of the project management standard on the organisation is determined.

So that after the incorporation of the basic principles of PRINCE2 into the project environment of the organization, the changes are reflected in the ANP model showing the project-oriented environment. These changes are mainly the setting of new relationships between the elements of the clusters: project documents, project team roles, and project constraints, especially in relation to the strategy of the organization, which now take into account the basic principles of PRINCE2. The preferences of individual project team roles (shown in Table 1) before the incorporation of the PRINCE2 principles are presented by the Neutral model and after the introduction of the PRINCE2 principles are presented by the PRINCE2 model. In the Total column are the values of preferences from the Limit Matrix of ANP, the column Normal shows the values of preferences normalized for Project Team Roles cluster, and the column Ideal shows the preferences obtained by dividing the values in the Total column by the largest value in the column.

Project Team Roles	Total		Normal		Ideal		Ranking		Diff.
	Neutral Model	PRINCE2 Model	Neutral Model	PRINCE2 Model	Neutral Model	PRINCE2 Model	Neutral Model	PRINCE2 Model	
Business Analyst (BAN)	0.00003	0.00471	0.00018	0.02646	0.00023	0.03790	8	7	1
Business Architect (BAR)	0.00015	0.00248	0.00101	0.01392	0.00124	0.01994	7	8	-1
IT Delivery Manager (ITDM)	0.01439	0.01687	0.09772	0.09473	0.12038	0.13572	2	2	0
Project Manager (PM)	0.11951	0.12430	0.81183	0.69800	1.00000	1.00000	1	1	0
Senior Supplier (SeS)	0.00408	0.00526	0.02773	0.02956	0.03416	0.04235	5	5	0
Senior User (SU)	0.00408	0.00526	0.02773	0.02956	0.03416	0.04235	4	4	0
Solution Architect (SAR)	0.00001	0.00515	0.00004	0.02891	0.00005	0.04142	9	6	3
Sponsor (Sp)	0.00089	0.00178	0.00602	0.01001	0.00741	0.01434	6	9	-3
Team Manager (TM)	0.00408	0.01226	0.02774	0.06886	0.03417	0.09865	3	3	0

Table 1 Quantification of the cluster Project Team Roles

In the project management environment in the commercial unit, the Project Manager (PM) is evaluated in the ANP model as the most important role, no other role is so important. However, after incorporating the PRINCE2 principles, the importance of the PM decreased, especially in favor of the project role Team Manager (TM), although the overall ranking of these roles remained the same. Furthermore, incorporating PRINCE2 had a significant effect on reducing the importance of the Business Architect role (BAR) in favor of Business Analyst (BAN) and in reducing the importance of the role Sponsor (Sp) in favor of the Solution Architect (SAR) (shown in Table 1). This occurred when the changes arose in roles and in roles' responsibility in case of partial transformation of the organization. Especially, there is important change (exchange of position/location) at SP and SAR in which case is the consequence of the upper involvement of role at the agile method of management project (SAR is becoming in agile teams Product Owner usually), meanwhile the role SP is suppressed.

Although the project role PM remained the most important role of the project team, its importance decreased the most compared to the other roles (see differences of weights in Figure 1). This is because the principles of PRINCE2 puts great emphasis on various roles in individual project management processes and not just on PM. The highest increase in the importance of the project roles is thus evident by the TM (Figure 2). Again, this change came in the case of a change of responsibility at partial agile transformation of the organization. At the role of TM came to increased responsibility in the consequence of increasing meaning of teams at their agile management.

One of PRINCE2's main principles is focused on products, therefore, preference values' stability of project team roles PM, TM, and IT DM has to be tested using sensitivity analysis. Sensitivity analysis was performed both in the model before the incorporating of PRINCE2 (Neutral model) and in the model after the incorporating of PRINCE2 into the project environment of the company (PRINCE2 model). The stability of the preferences of these roles was examined in terms of increasing the significance of PRINCE2 key elements Configuration (Figure 2 left) and Issue (Figure 2 right). As the importance of the PRINCE2's element Configuration (x-axis) increased, in the Neutral model the importance of PM increased and the importance of the IT DM decreased (y-axis). In the PRINCE2 model, the opposite is true and the importance of TM increases. Furthermore, as the importance of the PRINCE2's element Issue increases, in the Neutral model, the importance of the PM also increased at the expense of other roles, but in the PRINCE2 model, the importance of the TM increases.

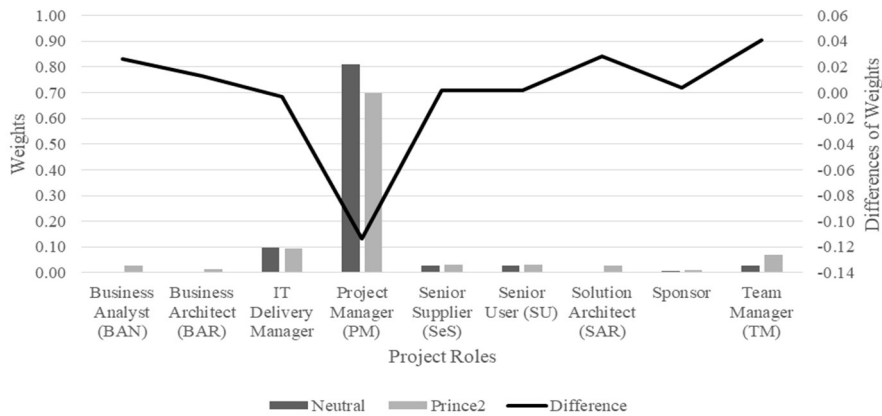


Figure 1 Change of preferences after the introduction of the PRINCE2

Sensitivity analysis is a suitable tool for monitoring the possible change of preferences of selected project team roles in the event of a change in the particular PRINCE2 principal importance in the company's project environment [16]. On the other hand, results of team roles importance as well as results from the sensitivity analysis obtained by ANP methods, refer only to the project-oriented environment in a specific commercial unit. As Ziemia [23] states in his work, without further research, the results cannot be generalized. In this research, the sensitivity analysis shows that the importance of particular project team roles is more sensitive in the PRINCE2 model than in the Neutral model. However, this sensitivity analysis was conducted only due to the change in the importance of the individual PRINCE2 principles. Furthermore, when the decision-makers can decide the importance of ANP structure elements, they can be sometimes inconsistent while filling in the pairwise comparison matrix. Then the consistency of the matrix can reach over a feasible consistency limit and the information value of the data can be ruined ([9]).

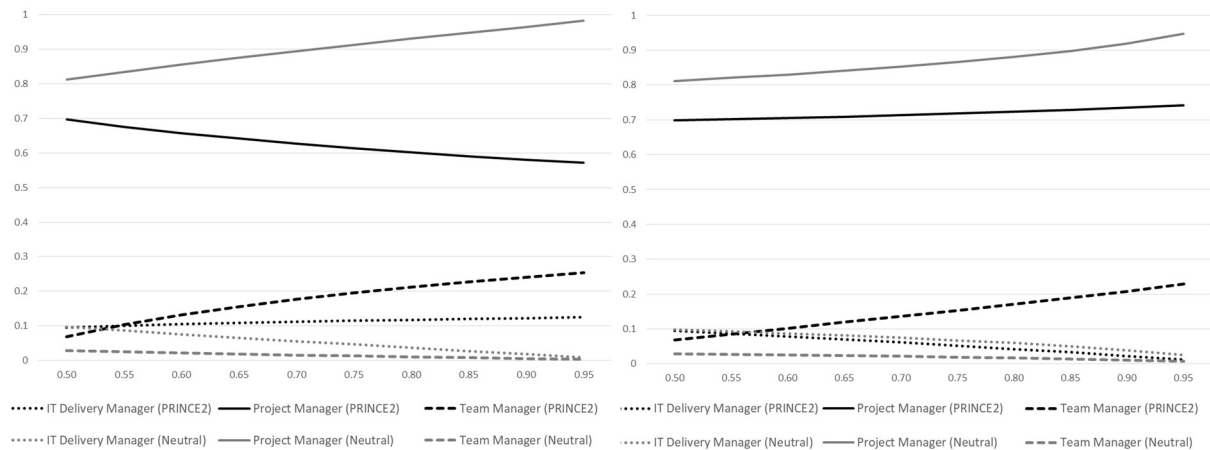


Figure 2 ANP sensitivity analysis of project team roles

4 Conclusion

The paper presents the use of semantic models of project management in the commercial sector (the banking sector), specifically in the context of international methodology PRINCE2 and in progress agile transformation in organization. The presented results are derived from partial results of the authors' own research from 2016 to 2020 and follow previous author's research. [3], [4], [5], [10] or [16]. Results of the case give evidence of changes in roles and in responsibility at the agile transformation of organization in connection with well-established methodology PRINCE2. Monitoring changes in the case of project roles is made possible using semantic model of project management. Especially, using ANP seems to be as significantly useful for quantification of variables just as it is for practice. Importance of international standards and methodology of project management is crucial for corporate practice, especially at agile approach in project management. Implementation and development of international standards and methodology of project management in the project environment of organization depends and is tailored to a specific environment of organization. For using semantic models of project management, the ANP is a very useful instrument. The ANP enables quantification of variables in project environment and monitoring their changes.

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] Adams, B. (2011). *SuperDecisions Limit Matrix Calculations*. Decision Lens Inc.
- [2] Axinte, S.-D., Petrica, G., Barbu, I.-D. (2017). Managing a software development project complying with PRINCE2 standard. *Proceedings of the 9th International Conference on Electronics, Computers and Artificial Intelligence*, ECAI 2017 ISBN: 978-150906457-1.
- [3] Bartoška, J. (2016). *A semantic model for the management of projects*. Habilitation work (thesis). DSE FEM CULS Prague.
- [4] 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.
- [5] Bartoška, J., Jedlanova, T., Rydval, J. (2019) Semantic Model of Project Management in Corporate practice In: *Proceedings of the 37th International Conference Mathematical Methods in Economics*. ISBN: 978-80-7394-760-6
- [6] Bennett, N. (2019). *Managing Successful Projects with PRINCE2*. Axelos ISBN 9780113315567.
- [7] De La Cámara Delgado, M., Marcilla, Fco.J.S., Calvo-Manzano, J.A., Vicente, E.F. (2012) Project management and IT governance Integrating PRINCE2 and ISO 38500. *7th Iberian Conference on Information Systems and Technologies, CISTI 2012*, ISBN: 978-989962477-1
- [8] Farman H., Javed H., Jan B., Ahmad J., Ali S., Khalil FN, et al. (2017) Analytical network process based optimum cluster head selection in wireless sensor network. *PLoS ONE* 12(7): e0180848. <https://doi.org/10.1371/journal.pone.0180848>.
- [9] Hlavatý, R., (2014). Saaty's matrix revisited: Securing the consistency of pairwise comparisons In *Proceedings of the 32th International Conference Mathematical Methods in Economics*. 287-292 , 2014.
- [10] 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.
- [11] Mařík, V. (1993). *Umělá inteligence: Díl I*. Praha: Academia, 264 s. ISBN 80-200-0496-3.
- [12] McGrath, S., Whitty, S.J. (2020). The suitability of PRINCE2 for engineering infrastructure. In: *Journal of Modern Project Management* (Vol 7, no 4) pp 312-347.
- [13] Quillian, R. (1968). *A recognition procedure for transformational grammars*. Doctoral dissertation. Cambridge (MA): Massachusetts Institute of Technology.
- [14] Rupali, P.P., Kirti, N.M. (2017). Benefits and Issues in Managing Project by PRINCE2 Methodology. In: *International Journal of Advanced Research in Computer Science and Software Engineering*. DOI: 10.23956/ijarcsse/V7I3/0134.
- [15] 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.
- [16] Rydval, J., Bartoška, J., Jedlanova, T. (2019) Sensitivity Analysis of Priorities of Project Team Roles Using the ANP Model In: *Proceedings of the 37th International Conference Mathematical Methods in Economics*. ISBN: 978-80-7394-760-6.
- [17] Saaty, T. L. (1996). *Decision Making with Dependence and Feedback: The Analytic Network Process*, ISBN 0-9620317-9-8, RWS.
- [18] Saaty, T. L. (2000). *Fundamentals of the Analytic Hierarchy Process*. RWS Publications, Pittsburgh, 2000.
- [19] Saaty, T. L. (2001). *Decision Making with Dependence and Feedback: The Analytic Network Process*, The Analytic Hierarchy Process Series. Pittsburgh: IX, RWS Publications.
- [20] Saaty, T. L. (2003). *The Analytic Hierarchy Process (AHP) for Decision Making and the Analytic Network Process (ANP) for Decision Making with Dependence and Feedback*, Creative Decisions Foundation.
- [21] Sowa, J. F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing Co., Pacific Grove, CA.
- [22] *SuperDecisions Software for Decision-Making*. (2018). URL <<http://www.superdecisions.com/>>.
- [23] Ziemba, P. (2019). Inter-Criteria Dependencies-Based Decision Support in the Sustainable wind Energy Management. *Energies* 2019, 12, 749; doi:10.3390/en12040749.

Assessment of personal ambiguity attitude in a series of online experiments

Simona Bažantová¹, Vladislav Bína², Václav Kratochvíl³, Klára Šimůnková⁴

Abstract. The paper deals with the issue of personal ambiguity attitude as a vital characteristic of decision-making. Similarly, as in the case of risk attitude, a person can seek ambiguity, show a neutral attitude, or have an ambiguity aversion. At the end of 2020 and in January 2021, an experiment consisting of a series of four lotteries was conducted. Each lottery consisted of 14 questions concerning bets on the results of drawing certain types of balls from an urn. Financial incentives for repeated attendance stimulated the participants, and the five most successful won the (gradated) prizes. The experiment was designed to assess personal attitude to ambiguity together with the measuring of this attitude and contained two variants of questions inspired by Ellsberg's experiments. Observing the pandemic limitations in place at the time, the series of lotteries were held online through MS Teams software and a questionnaire website, which showed the particular game settings and saved the participants' answers. The paper analyses the behaviour in each of the four lotteries and shows the bet changes in time as the participants' learning effect and characteristics.

Keywords: ambiguity, attitude, online experiment, decision-making

JEL Classification: D01

AMS Classification: 91B03

1 Introduction

Research on ambiguity has a long tradition. This phenomenon can be seen in everyday situations, and it is well-known that ambiguity (as a situation with no well-defined or vague probabilities) affects decision-making on a large scale. In some cases, people prefer risky bets to ambiguous ones, and this tendency is called *ambiguity aversion*. On the other hand, there is also the opposite situation, where people are *seeking ambiguity* as another way to express their ambiguity attitude [15]. Publications concerning the ambiguity aversion and identifying its consequences and the causes dealing with, for example, demographic factors (e.g. age effect, see Spröten et al. [14], gender differences, see Schubert [13]), or individuals' characteristics and experiences (see, e.g., Buhr & Dugas [2]), sometimes also with regard to the context of the decision or situation factors. It is also evident that people's ambiguity attitudes can be affected by many factors (for review, see Furnham & Marks [7]).

It appears that ambiguity attitude plays an important role in everyday decision-making and a particularly important role in decisions concerning economic problems (see, e.g., Bianchi & Tallon [1] or Dimmock et al. [5]). During the 20th century, the role of ambiguity was rather marginalized and risk attitude stayed in the forefront as the crucial personal characteristic of a decision-maker. This dates back to Savage's axiomatic formalization leading to the expected utility theory (see Savage [12] and Mongin [11]). The importance of ambiguity attitude was stressed in later studies showing it as the second most important characteristic of human decision making (see Cohen [4] and Lauriola [10]). More recent papers analyse the thought experiments using non-incentivized variants of the Ellsberg approach [6] although the incentivized version of assessing ambiguity attitude appears to be important (for an example supporting this assertion, see Cavatorta & Schroder [3]). The series of experiments summarized in this paper can be added to this stream of literature.

¹ Prague University of Economics and Business - Faculty of Management, Jarošovská 1117/II, Jindřichův Hradec, simona.bazantova@vse.cz.

² Prague University of Economics and Business - Faculty of Management, Jarošovská 1117/II, Jindřichův Hradec, vladislav.bina@vse.cz.

³ Prague University of Economics and Business - Faculty of Management, Jarošovská 1117/II, 377 01, J. Hradec & Institute of Information Theory and Automation, Prague, Czech Republic, velorex@utia.cas.cz

⁴ Prague University of Economics and Business - Faculty of Management, Jarošovská 1117/II, Jindřichův Hradec, klara.simunkova@vse.cz.

2 Methods and Data

This paper focuses on behaviour under ambiguity and deals with personal ambiguity attitude as a vital decision-making characteristic. Furthermore, we analysed the changes in the total amount of bets in lotteries in time as the participants' learning effect and personal characteristics.

Experimental design and questions in each online lottery follow experiments conducted under research project GAČR 19-06569S. Previous experiments (within the mentioned research project) were in offline form. However, due to COVID-19 restrictions, at the very end of 2020 and in January 2021, the experiments (consisting of a series of four lotteries) were conducted online through MS Teams software and a questionnaire website, showing the particular game settings and saving the participants' answers.

Each online lottery consisted of 14 questions concerning bets on the results of drawing certain types of balls from an urn (the urns were presented in a random order to participants). These questions can be divided into three categories: One Red Ball Example, 6-Colour Example and Ellsberg's Example. For the specific and detailed text of all questions, see Jiroušek & Kratochvíl [8]. For the present study, four games (questions) were most critical (mentioned in [8, pp. 54] and marked with abbreviations **F1**, **F2**, **I1** and **I2**) due to the calculation of the personal coefficients of ambiguity aversion. Questions F1 and F2 were the simulation of decision under risk (probability may be calculated). On the other hand, questions I1 and I2 were the simulation of decision under ambiguity (probability may not be calculated).

Ambiguity aversion was measured using the following formula (1) from Jiroušek & Kratochvíl [9]:

$$\alpha_F = \frac{a_1 - b_1}{a_1} \text{ and } \alpha_G = \frac{a_2 - b_2}{a_2} \quad (1)$$

Note: a_1 (a_2) refers to willing to bet a_1 (a_2) points for taking part at lottery F1 (F2) and b_1 (b_2) refers to willing to bet b_1 (b_2) points for taking part at lottery I1 (I2)

The semantics of coefficient α is “*the higher the aversion, the higher the coefficient*” [9, pp. 81]. In this paper, the personal coefficient of ambiguity aversion for each of the participants was calculated by the average values of α_F and α_G .

It is essential to mention the critical difference between the present online experiments and previous offline experiments. In the online experiments, the respondents did not bet their own money (the respondents play only for points in the game). In each lottery, won (or lost) game points were added to (or subtracted from) the personal account, and once all four lotteries were completed, we calculated the total number of game points to determine the winner.

The measured variables in the present experiments are illustrated in Table 1.

Nickname	Sex	Game	Result	Total Bet	Ambiguity coefficient	Previous Result	Change of ambiguity coefficient
Anettve	F	1	-60	107	0,35	NA	NA
Anettve	F	2	64	181	0	-60	-0,35
Anettve	F	3	0	104	0,29	64	0,29
Anettve	F	4	0	126	0,48	0	0,19
BlackJack	M	1	110	120	-1,50	NA	NA
BlackJack	M	2	0	100	0	110	1,5
BlackJack	M	3	65	140	0	0	0

Table 1 Dataset with measured variables

- *Ambiguity coefficients* were calculated by using the formula mentioned above. This indicated the ambiguity seeking or ambiguity aversion of the participants (numerical variable)
- *Total Bets* were defined as the total amount of the bet (game points) for each participant in each lottery (numerical variable)
- *Change of ambiguity coefficients* was calculated as the difference between α in the present lottery and α in the previous lottery (numerical variable)
- *Sex of respondent* (categorical variable)

- *Game* contains numerical marking of the lottery and shows the number of previous experiences with experimental situations (games or lotteries)
- *Result* is the variable that shows the number of winning or losing game points in the present lottery
- *Previous results* are the variable that shows the number of winning or losing game points in the previous lottery
- *Nickname* as the identified respondents

Data were analysed in statistical software R, and multiway ANOVA was used to examine the relationship between the measured variables. *Ambiguity coefficients*, *Total Bets* and *Change of ambiguity coefficients* were independent variables and the others were factors in ANOVA models. Due to the principals of ANOVA, we categorized the *Ambiguity coefficients* (values -2, -0.5, 0, 0.5, 2 into the categories "negative", "rather negative", "rather positive", "positive"), *Result* (values -800, -100, 0, 100, 800 into the categories "negative", "rather negative", "rather positive", "positive") while *Game* was also defined as factor.

The research sample consists mainly of students of the Faculty of Management, Prague University of Economics and Business. Financial incentives for repeated attendance stimulated the participants, and the five most successful won the (gradated) prizes. In total, 43 respondents participated in a series of four lotteries. However, only students and respondents who participated in two or more lotteries were included in further analyses. Three respondents were excluded due to unfulfilled conditions, and thirteen respondents were excluded due to participation in only one lottery. This reduction was used mainly due to the increased consistency of the sample and calculation with previous experiences. Therefore, 27 participants (10 male, 17 female) were included in further analyses.

3 Results

This paper analyses behaviour under ambiguity and risk in each of four lotteries, and shows changes in bets and ambiguity coefficients in time as a learning effect in the participants' behaviour. We specifically investigated the effect of previous experience on the willingness to bet in the next lottery following a successful or unsuccessful one, in connection to the sex of the respondents, the number of previous experiences and the ambiguity coefficient. Furthermore, we examined the effects of measured variables on the personal coefficient of ambiguity aversion.

3.1 Personal coefficient of ambiguity aversion

This section illustrates some experimental results of the personal coefficient of ambiguity aversion (α). The measuring of this coefficient was proposed above (see the section Methods and data). The average values of the personal coefficients of ambiguity aversion stay within the range of 0.22 to 0.29 in four online experimental lotteries ($\bar{\alpha}_{\text{lottery1}}=0.2516$, $\bar{\alpha}_{\text{lottery2}}=0.2239$, $\bar{\alpha}_{\text{lottery3}}=0.2153$, $\bar{\alpha}_{\text{lottery4}}=0.2946$). These values are similar to the previous experiment (0.2133) [9], even so, these experiments were in an online environment and the respondents did not bet their own money (the respondents played only for points in the game, see the Methods and Data section). At the same time, the positive values of the coefficient indicate the respondents' **ambiguity aversion**. However, the negative coefficient α for some of the participants was identified. The participants' coefficients of ambiguity aversion were in the range of -1 to 1 in all four lotteries. Moreover, negative values are interpreted as **seeking ambiguity**. In Figure 1, we compare each of the parts of coefficient α computed from bets in the games F1, I1 and F2, I2 (i.e. α_F and α_G).

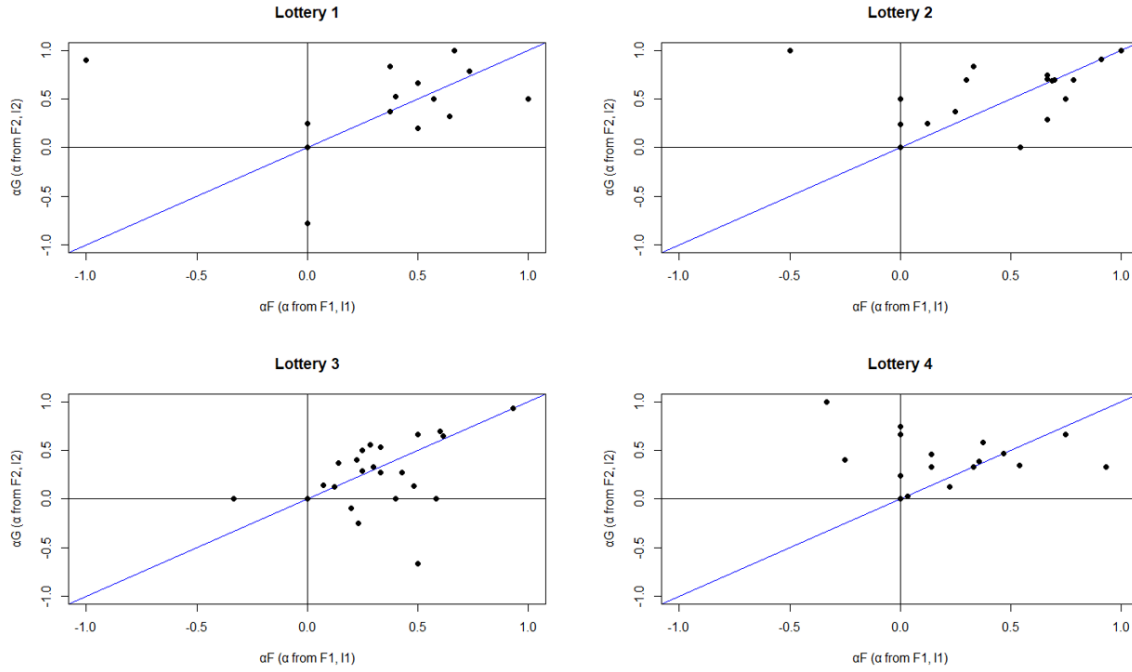


Figure 1 Comparison of participants' ambiguity aversion coefficient (α_F and α_G) in each lottery

Figure 1 indicates several interesting facts. Points on the diagonal represent consistent results of the participants' coefficient of ambiguity aversion. On the other hand, very skewed points from diagonal and the points on the axes indicate the participants' ambiguity aversion only in one pair of situations. These dots imply a certain illogicality in the decision making or the respondents' inclination to more intuitive behaviour.

3.2 Factors affecting the total amount of bets

Based on multiway ANOVA, we examined the effect of the previous results on the total amount of bets in the following lottery (for the learning effect determination). Moreover, the other measured (independent) variables were included in the model (such as the sex of the respondent, personal ambiguity coefficient and the number of experiences). Furthermore, the model was set up to include iterations between the factors (the mentioned independent variables) and with regard to blocks created by the respondents' nicknames.

The results indicated a significant effect of the sex of the respondent (p -value = 0.0003), personal ambiguity coefficient (p -value < 0.0001), previous result (p -value = p -value < 0.0001) and the number of experiences (p -value = 0.0005) on the total amount of the bet in each lottery. However, this model did not indicate a significant effect of iterations. Moreover, the results show a significant effect of the participant's nickname (p -value < 0.0001), so a high level of individuality can be considered for each one. The following table (Table 2) shows the size of the effects of the variables with a significant effect on the total amount of the bet. Partial eta squares indicate the large effects of the included factors on the bets in the model.

Parameter	Eta2 (partial)	90% CI
Sex	0.48	[0.21, 0.66]
Ambiguity coefficient	0.74	[0.53, 0.83]
Previous result	0.93	[0.88, 0.96]
Game	0.53	[0.24, 0.69]
nickname	0.90	[0.77, 0.93]

Table 2 Size effects of variables with significant effect on amount of bet

We generated boxplots for all variables included in the model. However, the following must be mentioned.

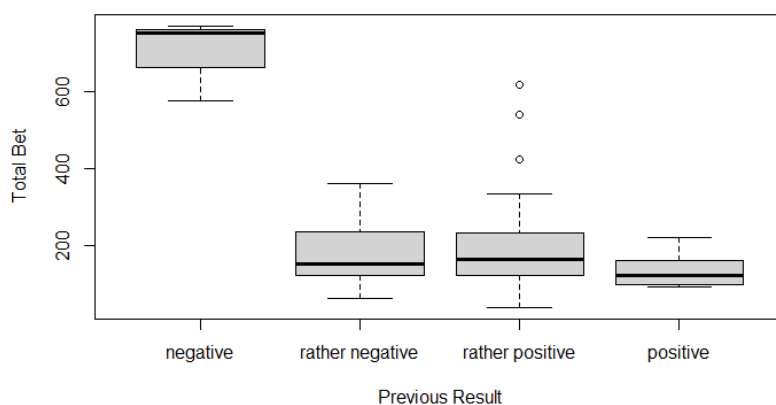


Figure 2 Dependence of the total amount of the bet and the previous result

Figure 2 shows that the negative result in the previous lottery (losing more than 100 game points) affects the total bet in the following lottery, where the bet is considerably higher. It can be caused by a change of participant’s game strategy (i.e. an ‘all-or-nothing’ basis) because participants did not endanger their own money in these online experiments.

3.3 Factors affecting the participants’ ambiguity aversion coefficient

As in the case of the previous analysis, we examined the effects of the respondents’ sex, the previous results and the number of experiences on the personal ambiguity aversion coefficient (the model including iterations between the factors and blocks created by the respondents’ nicknames). Contrary to our expectations, the analysis did not indicate any significant effect of the examined variables (all p-values was > 0.05 during the systematic iterative reduction of the model).

A similar analysis of the effects of the measured variables on the change of ambiguity coefficients across the lotteries ended similarly. In this context, we examined whether the change of personal ambiguity aversion coefficient was affected by the previous results (experiences). The other factors were the respondents’ sex and previous experiences (including iterations between the variables and the blocks based on the respondents’ nicknames). Contrary to expectations, the analysis did not detect any significant factor affecting ambiguity coefficient changes across the lotteries.

4 Conclusion

The key findings of this study can be summarized as follows: First, the present paper suggests a significant effect of a negative result in the previous lottery on the total bet in the following one, which could be caused by a change in the participant’s game strategy (i.e. an ‘all-or-nothing’ basis). Future research may examine this effect in a game with endangered own money. Second, in our online experimental lotteries, the mean value of the personal coefficients of ambiguity aversion stays within the range of 0.22-0.29, indicating that participants are inclined to ambiguity aversion. However, personal coefficients α stay within -1 to 1, which is interesting because, in our experiments, the participants were relatively homogeneous (regarding education and age). Despite the similarities of our sample, the alfa coefficient range was extensive, and we did not indicate the effects of the sex of the respondent, the previous result (a successful or unsuccessful game), and the number of experiences on the level of personal ambiguity aversion coefficient. Therefore, it appears that individuals’ ambiguity aversion may be more influenced by various internal factors such as attitude and motivation to play the game rather than any variables included in the present paper. Therefore, in further research, we shall turn our attention to a more profound examination of the internal factors of the participants. The primary limitation to the generalization of the results in this study is a relatively small sample size. In addition, the sample of respondents is not representative of the population.

Acknowledgements

This study was supported by the National Science Foundation of the Czech Republic (GAČR) under project no. 19-06569S.

References

- [1] Bianchi, M., & Tallon, J. M. (2019). Ambiguity preferences and portfolio choices: Evidence from the field. *Management Science*, 65(4), 1486-1501.
- [2] Buhr, K., & Dugas, M. J. (2006). Investigating the construct validity of intolerance of uncertainty and its unique relationship with worry. *Journal of anxiety disorders*, 20(2), 222-236.
- [3] Cavatorta, E., & Schröder, D. (2019). Measuring ambiguity preferences: A new ambiguity preference survey module. *Journal of Risk and Uncertainty*, 58(1), 71-100.
- [4] Cohen, M., Jaffray, J. Y., & Said, T. (1987). Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses. *Organizational behavior and human decision processes*, 39(1), 1-22.
- [5] Dimmock, S. G., Kouwenberg, R., Mitchell, O. S., & Peijnenburg, K. (2016). Ambiguity aversion and household portfolio choice puzzles: Empirical evidence. *Journal of Financial Economics*, 119(3), 559-577.
- [6] Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The quarterly journal of economics*, 643-669.
- [7] Furnham, A. & Marks, J. (2013). Tolerance of Ambiguity: A Review of the Recent Literature. *Psychology*, 4, 717-728.
- [8] Jiroušek, R. & Kratochvíl, V. (2019). Preliminary Results from Experiments on the Behavior under Ambiguity. In: M. Inuiguchi, R. Jiroušek & V Kratochvíl (Eds.), *Proceedings of the 22nd Czech-Japan Seminar on Data Analysis and Decision Making (CJS'19)* (pp. 53-64). Nový Svetlov, Czech Republic
- [9] Jiroušek, R. & Kratochvíl, V. (2020). On subjective expected value under ambiguity. *International Journal of Approximate Reasoning*, 127, 70–82.
- [10] Lauriola, M., & Levin, I. P. (2001). Relating individual differences in attitude toward ambiguity to risky choices. *Journal of Behavioral Decision Making*, 14(2), 107-122.
- [11] Mongin, P. (1997). *Expected utility theory*. Handbook of economic methodology, 342350, 234-350.
- [12] Savage, L. J. (1954). *The Foundations of Statistics*. New York: John Wiley & Sons Inc
- [13] Schubert, R., Gysler, M., Brown, M., & Brachinger, H. W. (2000). *Gender specific attitudes towards risk and ambiguity: An experimental investigation* (No. 00/17). Economics Working Paper Series.
- [14] Sproten, A., Diener, C., Fiebach, C., & Schwierien, C. (2010). *Aging and decision making: How aging affects decisions under uncertainty* (No. 508). Discussion Paper Series.
- [15] Trautmann, S. T. & van de Kuilen, G. (2015). *Ambiguity attitudes*. *The Wiley Blackwell handbook of judgment and decision making*. 1: 89–116. Chichester: John Wiley & Sons Ltd.

An original two-index model of the multi-depot vehicle routing problem.

Zuzana Borčinová¹, Štefan Peško²

Abstract. Vehicle routing problem (VRP) is a family of combinatorial optimization problems which aim to determine the lowest cost vehicle routes to serve a set of customers. In a solution of VRP, the vehicle routes originate from one or several depots and should return to the same depot they started with while ensuring that the total demand on each route does not exceed the vehicle capacity. In this paper, we present a two-index vehicle-flow formulation for the multi-depot vehicle routing problem (MDVRP) including a new constraint used to forbid routes to have the starting and ending points at two different depots. Computational experiments on several instances of varying depots and customer sizes showed that the optimal solutions were obtained by proposed formulation in a lower CPU time compared to the three-index formulations.

Keywords: vehicle routing problem, multiple depots, mixed-integer linear programming model, vehicle-flow formulation

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

Vehicle routing problem (VRP) is a family of combinatorial optimization problems which have many practical applications in the fields of transportation, distribution, and logistics. VRP is defined as designing routes for a fleet of vehicles to serve a set of customers with known demands. Each route is assumed to start and end at a depot and each customer is to be fully serviced exactly once. The primary objective is to minimize the total distance traveled by all vehicles. In a large number of practical situations, additional constraints are usually defined for variants of the VRP. For example, Capacitated VRP, or CVRP (every vehicle has a limited capacity), VRP with time windows, or VRPTW (the service at each customer must start within a given time interval), Multi-depot VRP, or MDVRP (vehicles are located in several different depots), Pickup and Delivery VRP, or PDVRP (customers may require a pickup or a delivery service), Periodic VRP, or PVRP (customers require repeated visits during the planning horizon), etc. This paper is focused on MDVRP. As mentioned above, MDVRP is a variant of the VRP where more than one depot is considered. Given the locations of depots and customers, the MDVRP requires the assignment of customers to depots and the vehicle routing for visiting these customers such that:

- (1) each vehicle route starts and ends at the same depot,
- (2) each customer is serviced exactly once by a vehicle,
- (3) the total demand of each route does not exceed the vehicle capacity,
- (4) the total cost of the distribution is minimized.

Even for relatively small size instances, the MDVRP is NP-hard and difficult to solve to optimality. Therefore, most solution methods proposed for the MDVRP are heuristics and metaheuristics. For example, Tabu Search has been used in [5]. An Adaptive Large Neighborhood Search approach was introduced in [12]. Genetic Algorithm is used in [14] and a hybrid algorithm based on Iterated Local Search is applied in [13]. Other hybrid metaheuristic algorithms combining Greedy Randomized Adaptive Search Procedure, Iterated Local Search, and Simulated Annealing are proposed in [1]. In [11], the authors present a parallel coevolutionary algorithm based on evolution strategy, and in [3] an algorithm based on the General Variable Neighborhood Search is proposed.

Only a few exact algorithms exist for the MDVRP, and these are only practical for relatively small problem sizes. Laporte et al. [8] were the first to report optimal solutions for problem sizes up to 50 customers and 8 depots by use of a branch-and-bound technique. Another exact approach for asymmetric MDVRPs by Laporte et al. [9] first transformed the problem into an equivalent constraint assignment problem, and then applied a branch-and-bound method to problem instances containing up to 80 customers and 3 depots. Contardo and Martinelli [4] proposed

¹ University of Žilina, Faculty of management science and informatics, Department of Mathematical Methods and Operations Research, Univerzitná 8215/1, Žilina, Slovakia, zuzana.borcinova@fri.uniza.sk

² University of Žilina, Faculty of management science and informatics, Department of Mathematical Methods and Operations Research, Univerzitná 8215/1, Žilina, Slovakia, stefan.pesko@fri.uniza.sk

an exact method based on ad-hoc vehicle-flow and set-partitioning formulations and solved the first by the cutting planes methods and the second by column-and-cut generation. A recent survey of exact and heuristic methods for solving MDVRP can be found in [10] and in [6].

Our contributions lie in introducing a new formulation for MDVRP under capacity constraint. Specifically, we propose a model based on a two-index vehicle-flow formulation, to which we include a new constraint used to forbid routes to have the starting and ending points at two different depots. To validate our approach, we also consider VRP with single-depot (CVRP) as a particular case of MDVRP. The remainder of this paper is organized as follows. *Section 2* describes the MDVRP. A new two-index mixed-integer linear programming formulation for MDVRP is proposed in *Section 3*. In *Section 4*, a computational comparison of different formulations for MDVRP on several families of instances is reported. Finally, some conclusions are drawn in *Section 5*.

2 Multi-Depot Vehicle Routing Problem

The MDVRP can be formalized as follows. Let $G = (V, H)$ be a complete directed graph with the node-set V and the arc-set $H = \{(i, j) : i, j \in V, i \neq j\}$. The nodes are partitioned into two subsets: the set of customers to be served V_C and the set of depots V_D , with $V_C \cup V_D = V$ and $V_C \cap V_D = \emptyset$. The positive traveling cost c_{ij} is associated with each arc $(i, j) \in H$. Each customer $i \in V_C$ has a certain positive demand d_i . In every depot, there is a fleet of p vehicles with the same capacity, denoted as Q , whereas K represents the set of all vehicles.

Figure 1 shows a MDVRP with twelve customers, in the form $V_C = \{1, 2, \dots, 12\}$ and two depots, named $V_D = \{13, 14\}$. In this example, depot 13 has three routes, that serves customers 1, 2, 3, 4, 5, 11, and 12, while depot 14 has two routes, serving customers 6, 7, 8, 9, and 10.

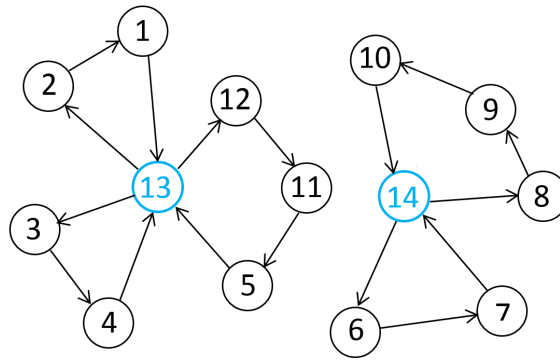


Figure 1 Example of MDVRP

In the following, we introduce a mathematical formulation for MDVRP that extends CVRP [7] to consider multiple depots. Let three-index binary decision variables x_{ijk} are equal to 1 when vehicle k visits node j immediately after node i , and 0 otherwise. Auxiliary variables $y_{ik} \in \langle d_i, Q \rangle$ indicate upper bound on the load already distributed by vehicle k upon leaving customer i .

The three-index mathematical model is defined as follows:

MDVRP-3i

$$\text{Minimize } \sum_{(i,j) \in H} \sum_{k \in K} c_{ij} x_{ijk}, \quad (1)$$

Subject to

$$\sum_{i \in V, i \neq j} \sum_{k \in K} x_{ijk} = 1, \quad \forall j \in V_C, \quad (2)$$

$$\sum_{i \in V, i \neq j} x_{ijk} - \sum_{i \in V, i \neq j} x_{jik} = 0, \quad \forall k \in K; j \in V_C, \quad (3)$$

$$\sum_{i \in V_D} \sum_{j \in V, i \neq j} x_{ijk} \leq 1, \quad \forall k \in K, \quad (4)$$

$$\sum_{i \in V_D} \sum_{j \in V, i \neq j} x_{jik} \leq 1, \quad \forall k \in K, \quad (5)$$

$$y_{ik} + d_j - y_{jk} \leq (1 - x_{ijk}) Q, \quad \forall i \in V_C; j \in V_C, i \neq j; k \in K, \quad (6)$$

$$x_{ijk} \in \{0, 1\}, \quad \forall (i, j) \in H; k \in K, \quad (7)$$

$$d_i \leq y_{ik} \leq Q, \quad \forall i \in V_C; k \in K. \quad (8)$$

Objective (1) minimizes the total traveling cost. Constraints (2) guarantee that each customer is visited exactly once. Constraints (3) impose the degree balance of each node, including both customers and depots. Constraints (4) and (5) establish that each vehicle starts and finishes at most in one depot. Constraints (6) eliminate the sub-tours in the solutions and ensure that the vehicle capacity is not exceeded. Finally, (7) and (8) are obligatory constraints.

Three-index formulations for VRP have limited practical use due to their large number of variables. In the next section, we propose a two-index vehicle-flow formulation for MDVRP that naturally eliminates the number of variables.

3 Two-index mathematical model of MDVRP

In this section, we first describe a new constraint used to assign the customers to the depots and then present the formulation itself.

3.1 Assignment constraint

Using the same notation as for three-index formulation, for every arc $(i, j) \in H$, we define a binary variable x_{ij} equal to 1 if arc (i, j) is used. For every customer $i \in V_C$, let $z_i \in V_D$ be an integer decision variable that determines which depot customer i is assigned to, according to the following constraints:

$$M(x_{ij} - 1) \leq z_i - z_j \leq M(1 - x_{ij}), \quad \forall (i, j) \in H, \quad (9)$$

where M is a sufficiently large positive constant. For each depot $i \in V_D$ is explicitly $z_i = i$.

Proposition 1. *Constraints (9) impose that if the arc (i, j) is used by a vehicle route, then customers i and j are assigned to the same depot.*

Proof. Let the arc (i, j) is used by a vehicle route, i.e. $x_{ij} = 1$. Then according to (9) is

$$0 \leq z_i - z_j \leq 0,$$

what implies that $z_i = z_j$.

Otherwise, if the arc (i, j) is not used, i.e. $x_{ij} = 0$, then

$$-M \leq z_i - z_j \leq M,$$

it means that z_i and z_j can but may not be the same. □

For example, in Figure 2, $x_{3,4} = 1$ and $z_3 = z_4 = 13$. In other case, $x_{4,5} = 0$ and both z_4, z_5 are equal to 13, while also $x_{4,6} = 0$, but z_4 and z_6 are different ($z_4 = 13$ and $z_6 = 14$).

Note that (9) also forbid routes to have the starting and ending points at two different depots.

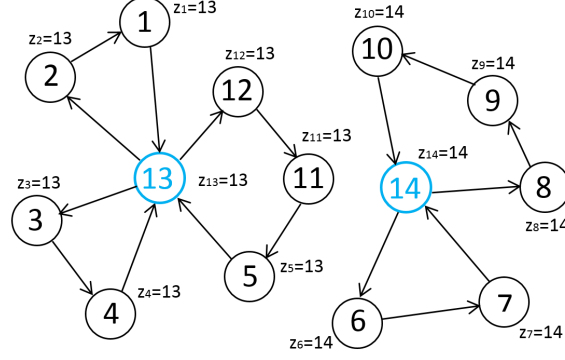


Figure 2 Assignment variables

3.2 A two-index vehicle-flow formulation

Analogously as in three-index formulation, auxiliary variable $y_i \in \langle d_i, Q \rangle$ is also associated with every customer and used in the sub-tour elimination constraints that ensure the continuity of the vehicle route in terms of the demand.

Now, we can formulate the MDVRP by the following mixed-integer linear programming model:

MDVRP-2i

$$\text{Minimize } \sum_{(i,j) \in H} c_{ij} x_{ij}, \quad (10)$$

Subject to

$$\sum_{j \in V, i \neq j} x_{ij} = 1, \quad \forall i \in V_C, \quad (11)$$

$$\sum_{j \in V, i \neq j} x_{ji} = 1, \quad \forall i \in V_C, \quad (12)$$

$$\sum_{i \in V_C} x_{ij} \leq p, \quad \forall j \in V_D, \quad (13)$$

$$M(x_{ij} - 1) \leq z_i - z_j \leq M(1 - x_{ij}), \quad \forall (i, j) \in H \quad (14)$$

$$y_i + d_j - y_j \leq (1 - x_{ij})Q, \quad \forall i \in V_C, j \in V_C, i \neq j \quad (15)$$

$$x_{ij} \in \{0, 1\}, \quad \forall (i, j) \in H, \quad (16)$$

$$d_i \leq y_i \leq Q, \quad \forall i \in V_C, \quad (17)$$

$$z_i = i, \quad \forall i \in V_D, \quad (18)$$

$$\min(V_D) \leq z_i \leq \max(V_D), \quad \forall i \in V_C. \quad (19)$$

Constraints (11) – (13) are degree constraints for customers and depots, respectively. The assignment constraints of customers to depots are given by (14). Constraints (15) are the capacity and sub-tour elimination constraints. At last, the obligatory constraints (16) – (19) define different kinds of decision variables.

4 Computational experiments

In this section we describe the implementation of proposed mathematical models and the results obtained on a set of problem instances. The mathematical models were coded in Python 3.7 and solved by the solver Gurobi 8.1. A computer equipped with an Intel Core i7-5960X, 3 GHz processor, and 32 GB of RAM was used to perform the computational experiments. The problem instances used in our experiments are generated based on Cordeau's instances [5]. Namely, for generating an instance of n customers and m depots, we randomly extract from the larger instance p01 n customer locations and their demands as well as m depots locations. The number of available vehicles per depot p is varied from 1 to 3. The rationale behind this is to provide instances where the solver can find an optimal solution within the short time limit to have the possibility of comparing the performance of both formulations proposed in this work, i.e. MDVRP-3i and MDVRP-2i, respectively.

<i>id</i>	<i>Instance</i>				<i>MDVRP – 3i</i>			<i>MDVRP – 2i</i>		
	<i>n</i>	<i>m</i>	<i>p</i>	<i>Q</i>	<i>Cost</i>	<i>Gap(%)</i>	<i>Time(s)</i>	<i>Cost</i>	<i>Gap(%)</i>	<i>Time(s)</i>
1.	20	2	1	160	314.973	0.000	6.73	314.973	0.000	1.62
2.	20	2	2	80	365.902	0.000	970.15	365.902	0.000	97.74
3.	20	2	3	80	360.997	0.000	1 547.47	360.997	0.000	25.47
4.	20	3	1	160	308.103	0.000	1.56	308.103	0.000	1.03
5.	20	3	2	80	326.138	0.000	101.68	326.138	0.000	6.54
6.	20	3	3	80	326.138	0.000	460.61	326.138	0.000	3.37
7.	20	4	1	80	340.714	0.000	1 508.49	340.714	0.000	28.56
8.	20	4	2	80	323.487	0.000	360.81	323.487	0.000	9.34
9.	20	4	3	80	323.487	0.008	325.54	323.487	0.000	3.95
10.	25	2	1	240	345.251	0.000	2 325.33	345.251	0.000	5.35
11.	25	2	2	160	367.284	0.000	1 399.47	367.284	0.000	199.33
12.	25	2	3	160	367.284	0.000	5 272.81	367.284	0.000	395.75
13.	25	3	1	160	358.960	0.003	128.02	358.960	0.000	11.53
14.	25	3	2	160	358.960	0.000	224.18	358.960	0.000	15.11
15.	25	3	3	160	358.960	0.000	288.44	358.960	0.000	17.67
16.	25	4	1	160	375.638	0.000	249.26	375.638	0.000	26.42
17.	25	4	2	160	357.411	0.000	134.68	357.411	0.000	28.26
18.	25	4	3	80	388.732	0.000	3 680.19	388.732	0.000	84.87

Table 1 Comparison of computational times

<i>Instance</i>	<i>n</i>	<i>p</i>	<i>Q</i>	<i>BKS</i>	<i>Cost</i>	<i>Gap(%)</i>	<i>Time(s)</i>
P-n16-k8	15	8	35	450	450.000	0.000	1.49
P-n19-k2	18	2	160	212	212.000	0.000	320.79
P-n20-k2	19	2	160	216	216.000	0.000	2825.22
P-n21-k2	20	2	160	211	211.000	0.000	519.59
P-n22-k2	21	2	160	216	216.000	0.000	1225.74
P-n22-k8	21	8	3000	603	603.000	0.000	165.35
P-n23-k8	22	8	40	529	529.000	0.000	3065.47

Table 2 Results for CVRP

Table 1 reports the computational results obtained by both optimization models. In this case, the first column, divided into five sub-columns, shows the main characteristics of the test problems: *id* – the identifier of instance to solve, *n* – the number of customers, *m* – the number of depots, *p* – the number of vehicles per depot, and *Q* – capacity of the vehicle. The next columns show the results obtained by the three-index model *MDVRP-3i* and two-index model *MDVRP-2i*. In each case, columns *Cost*, *Gap(%)*, and *Time(s)* present the objective value, gap, and computational time in seconds reported by solver, respectively. As shown in this table, both models solved all instances to optimality. However, model *MDVRP-2i*, in general, is much faster when compared to model *MDVRP-3i*.

To validate our two-index model of MDVRP, we also consider the CVRP as a particular case of the MDVRP where the number of depots $m = 1$. We conduct computational experiments on several instances from the literature, namely class P proposed by Augerat [2]. These instances can be found in <http://vrp.atd-lab.inf.puc-rio.br/index.php/en/>. We present the computational results obtained with *MDVRP-2i* in Table 2. In this table, the column labeled *Instance* represents the name of the instance, and columns *n*, *p*, and *Q* show the number of customers, the number of vehicles, and the vehicle capacity, respectively. Column *BKS* represents the best-known solution to the problem, as reported in the literature. Under columns labeled *Cost*, *Gap(%)*, and *Time(s)* we report the results obtained by *MDVRP-2i*. These results confirm that the proposed two-index model with a new assignment constraint can solve exactly two classes of problems, the MDVRP and the CVRP.

5 Conclusion

In this work, we have presented a new two-index formulation for a multi-depot vehicle routing problem under capacity constraint. The problem is modeled using vehicle-flow formulation including a new assignment constraint of the customers to the depots, that is shown to forbid routes to have the starting and ending points at two different depots. Presented results allow us to conclude that our model considerably improves the computational speed. Although our model has been tested only for the MDVRP and the CVRP, it can be extended to other variants of the VRP. This would be a possible topic of our future research.

Acknowledgements

This work was supported by the research grants VEGA 1/0776/20 „Vehicle routing and scheduling in uncertain conditions“ and the Slovak Research and Development Agency under the Contract no. APVV-19-0441 “Allocation of limited resources to public service systems with conflicting quality criteria”.

References

- [1] Allahyari, S., Salari, M., Vigo, D. (2015). A hybrid metaheuristic algorithm for the multi-depot covering tour vehicle routing problem. *Eur. J. Oper. Res.* 242 (3). (pp. 756–768).
- [2] Augerat, P. (1995). *Approche polyédrale du problème de tournées de véhicules*. Ph.D. Thesis, France: Institut National Polytechnique de Grenoble.
- [3] Bezerra, S.N., de Souza, S.R., Souza, M.J.F. (2018). A GVNS algorithm for solving the multi-depot vehicle routing problem. *Electron. Notes Discrete Math.* 66, 5th International Conference on Variable Neighborhood Search, (pp. 167–174).
- [4] Contardo, C., Martinelli, R. (2014). A new exact algorithm for the multi-depot vehicle routing problem under capacity and route length constraints. *Discret. Optim.* 12 (pp. 129–146).
- [5] Cordeau, J.F., Gendreau, M., Laporte, G. (1997). A tabu search heuristic for periodic and multi-depot vehicle routing problems. *Networks* 30(2) (pp. 105–119).
- [6] Jayarathna, D.G.N.D., Lanel, G.H.J., Juman, Z.A.M.S. (2021). Survey on Ten Years of Multi-Depot Vehicle Routing Problems: Mathematical Models, Solution Methods and Real-Life Applications. *Sustainable Development Research; Vol. 3, No. 1* DOI: 10.30560/sdr.v3n1p36.
- [7] Kulkarni, R.V., Bhawe, P.R. (1985). Integer programming formulations of vehicle routing problems. *European Journal of Operational Research* 20 (pp. 58–67).
- [8] Laporte, G., Nobert, Y., Arpin, D. (1984). Optimal solutions to capacitated multidepot vehicle routing problems. *Congressus Numerantiwn* 44 (pp. 283–292).
- [9] Laporte, G., Nobert, Y., Taillefer, S. (1988). Solving a family of multi-depot vehicle routing and location-routing problems. *Transp. Sci.* 22 (pp. 161–172).
- [10] Montoya-Torres, J.R., Franco, J.L., Isaza, S.N., Jiménez, H.F., Herazo-Padilla, N. (2015). A literature review on the vehicle routing problem with multiple depots. *Comput. Ind. Eng.* 79 (pp. 115–129).
- [11] de Oliveira, F.B., Enayatifar, R., Sadaei, H.J., Guimarães, F.G., Potvin, J.Y. (2016). A cooperative coevolutionary algorithm for the multi-depot vehicle routing problem. *Expert Syst. Appl.* 43 (pp. 117–130).
- [12] Pisinger, D., Ropke, S. (2007). A general heuristic for vehicle routing problems. *Comput. Oper. Res.* 34 (8) (pp. 2403–2435).
- [13] Subramanian, A., Uchoa, E., Ochi, L.S. (2013). A hybrid algorithm for a class of vehicle routing problems. *Comput. Oper. Res.* 40 (10) (pp. 2519–2531).
- [14] Vidal, T., Crainic, T.G., Gendreau, M., Lahrichi, N., Rei, W. (2012). A hybrid genetic algorithm for multidepot and periodic vehicle routing problems. *Oper. Res.* 60 (3) (pp. 611–624).

Portfolio selection via a dynamic moving mean-variance model

Adam Borovička¹

Abstract. Investment decision making, or portfolio selection, is not usually an easy process. Development in the capital market can be influenced by many factors. Prediction in this field becomes difficult. This uncertainty is reflected in the performance of the investment. Many approaches and methods supporting investment decision making reflect the uncertainty through the risk of not meeting the expected investment profit. Notoriously known Markowitz model measures the risk by a variance. However, one fact is neglected by (not only) this model. Risk, or return, is unstable over time. This dynamics should be taken into account to make a representative, robust decision. Dynamizing the process is performed via a ‘moving’ form of both characteristics. The dynamic version of Markowitz model called a moving mean-variance model is proposed. Both characteristics are enumerated through a developed approach in all overlapping subperiods from which moving means and (co)variances are calculated. The designed model is applied to select a portfolio from popular open unit trusts. To demonstrate the benefit of a dynamized model, the result is confronted with the output of a ‘static’ mean-variance model.

Keywords: dynamic, moving mean-variance, portfolio, unit trust

JEL Classification: C44, C61, G11

AMS Classification: 90B50, 90C30, 90C70

1 Introduction

The portfolio selection problem is still a ‘hot’ topic. Why? At first, more and more people are considering investing to appreciate their free funds. This effort is accelerated by an ever-expanding range of investment instruments covering wide investor audience. Secondly, a development in the capital market is not usually predictable, which makes a decision making complicated. These phenomena of the world of investment encourage the development of user-friendly methods and approaches that would be a significant support for making representative, robust investment decisions.

One of the most essential questions within the investment decision making process is a portfolio selection. For this purpose, the Markowitz mean-variance concept, based on a diversification idea, is widely applied [5,6]. This approach can take into account two most important characteristics of the investment – return and risk. Another benefit is its easy applicability for a wider range of investment instruments. On the other side, the already massive use of this concept is sometimes limited by a few shortcomings. Besides a detailed discussed assumption of normally distributed returns or a penalization of positive deviation from average return, one aspect (especially in the context of the capital market environment) is, in my opinion, neglected.

This is an assumption of stability of risk, or return of the portfolio over time. However, a level of the uncertainty in the capital market may not be stable over time. It means that the monitored characteristics, measured by mean and variance, would not have to be considered as static elements. How to take this aspect into account in the model? Answering this crucial question can move a model applicability closer to the portfolio making reality. One way to do this can be through the fuzzy set theory. In current literature, a mean-variance model is fuzzified through the returns represented as (triangular) fuzzy numbers. Then, the mean and (co)variance of fuzzy numbers are calculated, e.g. Huang [2]. This process quantifies (‘defuzzifies’) both characteristics on the vague (fuzzy) data in the crisp form. This concept, however, does not fully follow the instability of investment characteristics over time. A better way, how to represent a variable uncertainty, or instable (return) volatility, is an expression of mean and variance of the portfolio return as triangular fuzzy numbers on the crisp data [1]. However, this sophisticated approach may weaken its applicability due to a slightly more complex algorithm. So, I dare to propose an even friendlier approach based on ‘moving’ concept. Then the instability of mean and variance over time is concluded by dynamizing the process using a moving average. The observed historical period is divided into a few smaller

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

parts shifting by one-time subperiod. In these time-overlapping periods, both characteristics are calculated. Finally, the general mean and variance of the portfolio return are computed from partial characteristics of the subperiods via simple or weighted averaging. The main benefit of such a concept is less data complexity and user-friendliness.

Thus, the main aim of this article is to improve the original mean-variance concept to take into account a present instability of uncertainty (volatility) through by dynamizing the process using moving average concept. Novel concept is called a moving mean-variance model. The second aim is to demonstrate an application power of the proposed concept in a portfolio selection process in the capital market with increasingly popular open unit trusts. Composition of the portfolios made from the unit trusts offered Česká spořitelna is analyzed. The result is compared by the output of original mean-variance model to demonstrate the algorithm-application differences.

The rest of the article is structured as follows. Section 2 describes a portfolio selection procedure based on the proposed moving mean-variance model. Section 3 deals with a practical selecting a portfolio from the open unit trusts via a developed approach. Section 4 summarizes the main contributions of the article and outlines some ideas for future research.

2 Portfolio selection procedure using moving mean-variance concept

This section proposes a novel version of mean-variance model based on the ‘moving’ principle. This model is embedded in the whole portfolio selection process that is described in the following several steps.

2.1 Step 1: Determination of the investment policy

At the beginning of the investment decision making process, the investment policy must be declared [4,7]. The first very important aspect is a purpose of the investment – financing of the study, financial protection of the pension age, loan repayment, creation a fund for unexpected events, etc. The investment horizon is closely related to the investment intention. The amount of available financial funds also affects the investment. The frequency of the investment is also important (on-time, continuous). The investment is significantly influenced by the attitude to the risk which is also related to the expectation of investment performance. The form of a portfolio management (passive, active) also shapes the investment policy, experiences or financial literacy as well.

2.2 Step 2: Data collection and characteristic calculation

After a preselecting suitable investment instruments (assets) based on the investment policy, all necessary data (price, fee, volume of trades, etc.) must be collected. All required characteristics of the assets (return, risk, etc.) are then calculated. Financial data are mostly publicly available, which is a positive aspect of a decision making in the capital market. Assuming known historical prices, return and risk of the assets can be calculated.

Let p defines the number of equally long time periods with n observations of returns for m assets. Then return as a mean of the i -th asset in the t -th period is calculated as follows

$$\bar{r}_{it} = \frac{\sum_{k=1}^n r_{itk}}{n} \quad i = 1, 2, \dots, m, t = 1, 2, \dots, p, \quad (1)$$

where $r_{itk}, i = 1, 2, \dots, m, t = 1, 2, \dots, p, k = 1, 2, \dots, n$, represents the k -th observation of a return of the i -th asset in the t -th period. The difference between the beginning, or end, of two consecutive periods is constant throughout the considered history. Neighboring periods overlap between the beginning of one period and the end of the previous one. Then the return of the i -th asset as moving mean can be proposed in a simple, or weighted form as follows

$$r_i = \frac{\sum_{t=1}^p \bar{r}_{it}}{p}, \text{ or } r_i = \sum_{t=1}^p w_t \bar{r}_{it} \quad i = 1, 2, \dots, m, \quad (2)$$

where $w_t, t = 1, 2, \dots, p$, is the weight of the t -th period. The weights are standardized, so the following holds $\sum_{t=1}^p w_t = 1$. The weights can reflect an importance of the particular subperiods. For instance, it is possible to consider a stronger effect of last subperiods to the development at the beginning of the investment horizon. Over time, the influence declines, and development occurs over a longer time horizon, which, however, is always more

or less influenced by the recent past. Therefore, the values of weights of the subperiods have a declining trend towards the past. Weights can be subjectively determined using (e.g.) scoring method.

The covariance of return of the i -th and j -th asset in the t -th period is computed through the following formula

$$\sigma_{ijt} = \frac{\sum_{k=1}^n (r_{itk} - \bar{r}_{it})(r_{jtk} - \bar{r}_{jt})}{n} \quad i, j = 1, 2, \dots, m, t = 1, 2, \dots, p, \quad (3)$$

where r_{itk} , or r_{jtk} , $i, j = 1, 2, \dots, m, t = 1, 2, \dots, p, k = 1, 2, \dots, n$, denotes the k -th observation of return of the i -th, or j -th asset in the t -th period. Then the moving (co)variance of return of the i -th and j -th asset can be developed in a simple, or weighted form as follows

$$\sigma_{ij} = \frac{1}{p} \sum_{t=1}^p \sigma_{ijt}, \text{ or } \sigma_{ij} = \sum_{t=1}^p w_t \sigma_{ijt} \quad i, j = 1, 2, \dots, m. \quad (4)$$

Now, the return and risk of the portfolio can be designed, the model with all investment conditions as well.

2.3 Step 3: Portfolio making

After processing data over all periods, the following moving mean-variance model is proposed. This is Markowitz model (without short sales) with specially prepared data (made in the previous step)

$$\begin{aligned} \min \quad & \mathbf{x}^T \Sigma \mathbf{x} \\ \mathbf{r}^T \mathbf{x} \geq & r' \\ \mathbf{e}^T \mathbf{x} = & 1 \\ \mathbf{x} \geq & \mathbf{0} \end{aligned}, \quad (5)$$

where $\mathbf{x}^T = (x_1, x_2, \dots, x_m)$ is the vector of variables denoted as $x_i, i = 1, 2, \dots, m$, representing a share of the i -th asset in the portfolio. The vector $\mathbf{r}^T = (r_1, r_2, \dots, r_m)$ contains returns measured by moving mean marked as $r_i, i = 1, 2, \dots, m$, for the i -th asset. $\Sigma = (\sigma_{ij})$ is the matrix with the generic elements $\sigma_{ij}, i, j = 1, 2, \dots, m$, reflecting the mutual influence of i -th and j -th asset return measured by moving (co)variance. \mathbf{e}^T is a vector of ones only serving for making a portfolio as a whole. r' denotes a minimum required level of portfolio return (reference return). The designed model can be supplemented through another necessary investment conditions, e.g. minimum/maximum share of one asset in the portfolio, that can be mathematically formulated within the set X .

This version of the model minimizes investment risk, denoted as $\mathbf{x}^T \Sigma \mathbf{x}$ measured by moving variance of the portfolio return, under the conditions of minimum required return due to its user friendliness. Setting a return is certainly easier and more practical than determining the risk. The reference value of portfolio return can be inspired by its minimum and maximum possible level. Denote $\mathbf{r}^T \mathbf{x}_r^{\max}$, $\mathbf{r}^T \mathbf{x}_v^{\min}$ as a maximum, or minimum attainable portfolio return (formalized as $\mathbf{r}^T \mathbf{x}$ measured by moving mean) representing its ideal and basal value on the set of all necessary investment conditions (formulated in model (5) and eventually in the set X). Then the following holds

$$\begin{aligned} \mathbf{x}_r^{\max} = \arg \max \quad & \mathbf{r}^T \mathbf{x} & \mathbf{x}_v^{\min} = \arg \min \quad & \mathbf{x}^T \Sigma \mathbf{x} \\ \mathbf{e}^T \mathbf{x} = 1 & & \mathbf{e}^T \mathbf{x} = 1 & \\ \mathbf{x} \geq \mathbf{0} & & \mathbf{x} \geq \mathbf{0} & \\ \mathbf{x} \in X & & \mathbf{x} \in X & \end{aligned}, \text{ or } \quad (6)$$

The basal value is reasonably determined in the context of (the best) risk value. Finally, the minimum required level of portfolio return can be determined through the following proposed formula

$$r' = \mathbf{r}^T \mathbf{x}_v^{\min} + h(\mathbf{r}^T \mathbf{x}_r^{\max} - \mathbf{r}^T \mathbf{x}_v^{\min}), \quad (7)$$

where $h \in (0, 1)$ actually measures a rate of risk aversion in the spirit of shabby “higher return-higher risk” (closer to 0 means higher risk aversion). Of course, the reference level can also be set in another way. In any case, the interval $\langle \mathbf{r}^T \mathbf{x}_v^{\min}, \mathbf{r}^T \mathbf{x}_r^{\max} \rangle$ should guide preferences in the portfolio selection process. The effective frontier can then be drawn on this return interval. Thus, after a reference level determination, the model (5) with any additional

conditions (included in the set X) can be solved (in any optimization software supporting quadratic programming) to make a portfolio.

2.4 Step 4: Portfolio evaluation and revision

The passive investor lets the portfolio “to live its own life”. The active investor regularly monitors a portfolio performance. The composition of the portfolio is revised as needed. To reoptimize the portfolio, model (5) with the actual data and investor preferences can be applied very effectively. The preferences may change with regard to the current life situation of the investor and his surroundings.

3 Selecting the portfolio of unit trusts via moving mean-variance model

Let us introduce a real-life portfolio selection problem in the Czech capital market with open unit trusts. The analysis focuses on the most often investment situation reflecting longer-time investment made by investors having a smaller amount of free funds. To select the most suitable investment portfolio, a designed portfolio selection procedure using proposed moving mean-variance model is applied.

3.1 Step 1: Investment policy specification

The investment is conceived as longer-term. Its purpose can be a financial protection in pension age. Or more generally, this is the intention of appreciation of available funds, which will not be needed in the foreseeable future. Such an intention determines a rather conservative investment approach. This investor is not able to take too much risk. He would rather be satisfied with a smaller, but ‘surer’ return. Many investors of this category are not very experienced in the capital market. His role will be rather passive. Under all mentioned circumstances, the open unit trust is a suitable investment instrument. Although a composition of the portfolio will not be subject to excessive changes, it should be clear to the investor. Therefore, the portfolio should not consist of too many funds. Based on a personal investment experiences of the author and discussion of the investment consultant, an adequate number of assets is three to six.

3.2 Step 2: Price collection, return and risk calculation

To invest in the open unit trusts, the investor usually turns to his home bank. Today, most banks in the Czech market already offer open unit trusts, or at least mediate trading with them. As a long-term client of Česká spořitelna with a lot of practical experiences with its investment instruments, the open unit trusts offered and managed by Česká spořitelna are chosen. In addition, the value of property in the Česká spořitelna funds is the second largest in the Czech market. The offer of open unit trusts is very wide. Based on the investment policy, thirteen open unit trusts are preselected. There are five bond funds (Sporinvest, Sporobond, Trendbond, Corporate Bond, High Yield Bond), five mixed funds (Fund of Controlled Yields, Equity Mix, Dynamic Mix, Balanced Mix, Conservative Mix) and three equity funds (Sporotrend, Global Stocks, Top Stocks). These unit trusts have a sufficiently long history. In order to represent a long-term price development, the period from 2011 to 2019, representing price falls, ups and also calmer times, is selected. Prices from the last trading day of the month are downloaded from the Česká spořitelna Investment Center [3]. To reflect a dynamic instability, a nine-year period is divided into five overlapping five-year subperiods gradually shifted by one year, thus from 2011-2015 to 2015-2019 period. In each subperiod, the mean (return) and (co)variances of open unit trusts are calculated from 60 observations through formulas (1) and (3). Then the moving means and (co)variances are computed by (2) and (4). The weights are determined based on the idea of more significant influence of recent development (about this idea see more in Section 2.2). Then the weights of five subperiods are chronologically determined as 0.1 (for period 2011-2015), 0.1 (2012-2016), 0.2 (2013-2017), 0.25 (2014-2018) and 0.35 (2015-2019) based on subjective discretion reflecting personal (analytic) investment experiences supported by scoring method. Both essential characteristics (in %), gradually from bond, through mixed to equity funds, are shown below in the matrix and vector (9).

3.3 Step 3: Portfolio making

Through the models (6) with the additional conditions $0.15\mathbf{y} \leq \mathbf{x} \leq 0.4\mathbf{y}, \mathbf{y} \in \{0,1\}$ included in the set X ensuring the requirement for a limited number of funds in the portfolio, the basal and ideal value of a portfolio return is determined. To make a portfolio, the model (5) with additional conditions formulated as follows

$$\begin{aligned}
& \min \mathbf{x}^T \Sigma \mathbf{x} \\
& \mathbf{r}^T \mathbf{x} \geq 0.347 \\
& \mathbf{e}^T \mathbf{x} = 1 \\
& 0.15\mathbf{y} \leq \mathbf{x} \leq 0.4\mathbf{y} \\
& \mathbf{x} \geq \mathbf{0} \\
& \mathbf{y} \in \{0,1\}
\end{aligned} \tag{8}$$

where $\mathbf{x} = (x_1, x_2, \dots, x_{13})^T$ represents shares of the funds in the order indicated in Step 2, i.e. from $i = 1 \approx$ Sporinvest to $i = 13 \approx$ Top Stocks. With the same order of elements, the matrix of (co)variances and vector of returns are constructed in the following form

$$\Sigma = \begin{bmatrix} 0.016 & 0.042 & 0.069 & 0.049 & 0.077 & 0.020 & 0.140 & 0.110 & 0.084 & 0.047 & 0.133 & 0.130 & 0.178 \\ 0.042 & 0.374 & 0.612 & 0.138 & 0.292 & 0.063 & 0.379 & 0.369 & 0.341 & 0.212 & 0.586 & 0.500 & 0.254 \\ 0.069 & 0.612 & 2.755 & 0.774 & 0.764 & 0.137 & 1.078 & 0.953 & 0.842 & 0.500 & 3.476 & 1.502 & 0.373 \\ 0.049 & 0.138 & 0.774 & 1.727 & 0.934 & 0.120 & 1.254 & 0.948 & 0.658 & 0.344 & 3.468 & 0.851 & 1.575 \\ 0.077 & 0.292 & 0.764 & 0.934 & 1.447 & 0.166 & 2.195 & 1.638 & 1.175 & 0.589 & 3.638 & 1.988 & 2.833 \\ 0.020 & 0.063 & 0.137 & 0.120 & 0.166 & 0.037 & 0.309 & 0.235 & 0.175 & 0.093 & 0.392 & 0.343 & 0.413 \\ 0.140 & 0.379 & 1.078 & 1.254 & 2.195 & 0.309 & 6.624 & 4.719 & 3.154 & 1.464 & 7.024 & 7.034 & 9.954 \\ 0.110 & 0.369 & 0.953 & 0.948 & 1.638 & 0.235 & 4.719 & 3.408 & 2.309 & 1.089 & 5.158 & 5.002 & 6.952 \\ 0.084 & 0.341 & 0.842 & 0.658 & 1.175 & 0.175 & 3.154 & 2.309 & 1.605 & 0.778 & 3.633 & 3.374 & 4.461 \\ 0.047 & 0.212 & 0.500 & 0.344 & 0.589 & 0.093 & 1.464 & 1.089 & 0.778 & 0.400 & 1.767 & 1.583 & 1.972 \\ 0.133 & 0.586 & 3.476 & 3.468 & 3.638 & 0.392 & 7.024 & 5.158 & 3.633 & 1.767 & 17.762 & 6.534 & 8.209 \\ 0.130 & 0.500 & 1.502 & 0.851 & 1.988 & 0.343 & 7.034 & 5.002 & 3.374 & 1.583 & 6.534 & 10.583 & 10.440 \\ 0.178 & 0.254 & 0.373 & 1.575 & 2.833 & 0.413 & 9.954 & 6.952 & 4.461 & 1.972 & 8.209 & 10.440 & 22.447 \end{bmatrix}, \mathbf{r} = \begin{bmatrix} -0.006 \\ 0.130 \\ -0.077 \\ 0.150 \\ 0.246 \\ -0.059 \\ 0.378 \\ 0.258 \\ 0.219 \\ 0.110 \\ 0.143 \\ 0.692 \\ 0.854 \end{bmatrix}. \tag{9}$$

The reference level of portfolio return is computed in the spirit of a longer-term rather conservative strategy. Thus, $h = 0.5$ from (7) reflects a more risk-averse attitude. Then the reference return level is computed as $r' = 0.0001 + 0.5(0.695 - 0.0001) = 0.347\%$.

The solution of model (8) represents the portfolio with a following composition: 38.21% Sporobond, 29.22% High Yield Bond and 32.57% Global Stocks. The significant share of Sporobond is not surprising. This fund has solid positive return and mainly has a strong diversification ability through low covariances of returns with other funds. Equity fund Global Stocks significantly helps to achieve a reference return of the investment by the second greatest monthly expected return. Its greater risk is compensated by the third fund High Yield Bond with solid return and diversification power with the other two funds. In case of greater risk aversion, mixed fund Sporinvest will inevitably become (despite virtually zero return) part of the portfolio due to sovereignly lowest covariances. On the other side, with a declining risk aversion, the equity fund Top Stocks will start participating in the portfolio thanks to the greatest return, especially at the expense of Sporobond.

Finally, how did the application of the dynamized version of the mean-variance model (proposed ‘moving’ form) affect the investment decision making? Let us compare the efficient frontiers made by the original mean-variance model, moving mean-variance model with simple and weighted characteristics (named as simple and weighted moving mean-variance model) as shown in Figure 1.

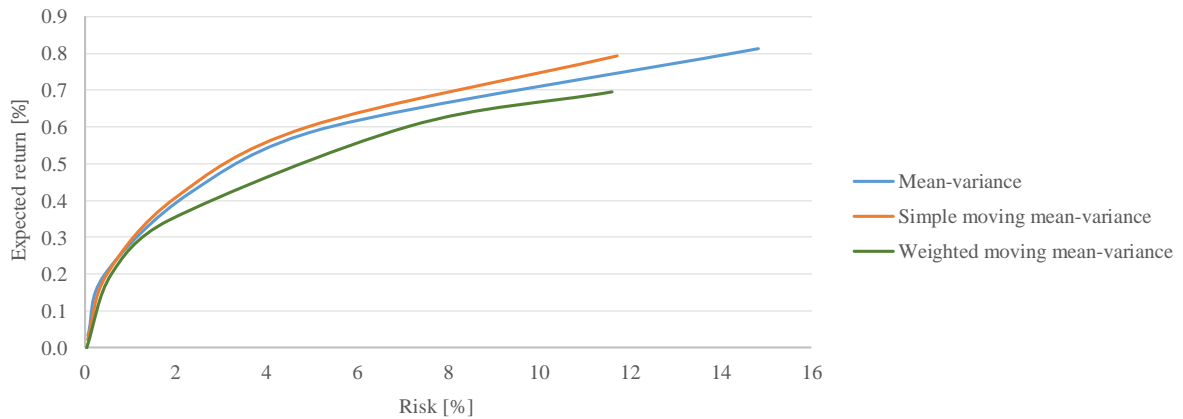


Figure 1 Effective frontiers made by three various form of mean-variance model

It is obvious that a dynamic version of a mean-variance model provides a different set of portfolios than the original Markowitz form. The set of funds participating in the portfolios is quite stable (approximately 6 funds), but their shares sometimes differ significantly. The efficient frontier from the simple moving mean-variance model quite copies the efficient frontier generated by mean-variance model. The main reason is an applied simple form of moving average which does not significantly change the input data of the original model. The weighted moving form provides a much different result. As can be seen that the portfolios with the same risk mostly provide a lower level of return than the portfolios made by mean-variance model. The main reason is the fact that the last five-year subperiod 2015-2019 showing lower returns has significantly highest weight. Emphasizing the impact of the recent subperiod proves to be appropriate, which ultimately confirms the beginning of ‘life’ of the investment in 2020. Although the time is too short for a longer-term investment, the performance from January 2020 to March 2021 is positive (4.78%). So far, the investment decision based on the moving mean-variance model seems to be the right one. However, only time will tell about the overall success of the investment.

4 Conclusion

The article deals with a modification of the mean-variance model to take into account an instable uncertainty (volatility) in the capital market. The proposed dynamic version of the Markowitz model is named moving mean-variance model. The observed historical period is divided into shorter overlapping subperiods within which a mean and (co)variance are calculated. Gradual shifting by one time period from the start to the end of the historical period can capture the changing return instability. Inclusion of this non-negligible feature of the capital market is reflected in the composition of the investment portfolio(s) which is more or less different from the portfolio made by the original mean-variance concept. The proposed concept has proven itself algorithmically and practically. Adequate inclusion of the variable instability has proven to be beneficial, in particular when selecting a portfolio of open unit trusts. The proposed moving mean-variance model is clearly trying to get closer to the reality of capital market processes which makes an investment decision more representative and robust.

Further research could be focused on the weights of particular subperiods. As they influence the result, i.e. the final decision, determining their values would deserve even more detailed analysis. Another interesting research area would be to compare a ‘moving’ version of the model with its fuzzy form also expressing the changing uncertainty, both from an algorithmic and application point of view.

Acknowledgements

The research project was supported by Grant No. F4/42/2021 of the Internal Grant Agency, Faculty of Informatics and Statistics, Prague University of Economics and Business.

References

- [1] Borovička, A. (2021). Stock portfolio selection under unstable uncertainty via fuzzy mean-semivariance model. Submitted to *Central European Journal of Operations Research*.
- [2] Huang, X. (2007). Portfolio selection with fuzzy returns. *Journal of Intelligent & Fuzzy Systems*, 18, 383–390.
- [3] Investment Center. (2021). *Archiv prodejných cen fondů*. [online] Available at: https://cz.products.erstegroup.com/Retail/cs/Ke_stauC5uBEenuC3uAD/Dokumenty_ke_stauC5uBEenuC3uAD/Archiv_prodejnuC3uADch_cen_fonduC5uAF/index.phtml, [Accessed 10 May 2021]
- [4] Levy, H. (1999). *Introduction to investments*. Cincinnati: International Thomson Publishing.
- [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] Steigauf, S. (2003). *Fondy – jak vydělávat pomocí fondů*. Praha: Grada Publishing.

A shadow utility of portfolios efficient with respect to the second order stochastic dominance

Martin Branda¹

Abstract. We consider diversification-consistent DEA models which are consistent with the second order stochastic dominance (SSD). These models can identify the portfolios which are SSD efficient and suggest the revision of portfolio weights for the inefficient ones. There is also a way how to reconstruct the utility of particular investors based on efficient portfolio which they hold. We apply the above mentioned approaches to industry representative portfolios and discuss the risk aversion of the investors. We focus on the sensitivity with respect to various levels of the risk aversion.

Keywords: Data envelopment analysis, diversification, second order stochastic dominance, risk aversion, shadow utility, sensitivity

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

Data envelopment analysis (DEA), introduced in [13], is nowadays an important class of models which serve to access efficiency of decision making units which consume given set of inputs to produce several outputs. The applications ranges from bank branches up to country regions efficiency, cf. [23]. A special attention has been paid to applications in finance, especially to efficiency of mutual funds and investment opportunities in general. Since the seminar work [26], many papers has been published on applications as well as methodology, see, e.g., [5, 11, 14, 25]. Recently, new class of DEA models with diversification, known also as diversification-consistent DEA (DC DEA), was introduced in [19]. These new models overcame the drawback of the traditional DEA models which does not take into account diversification effect between considered investment opportunities. In other words, if risk measures were considered as the inputs, the traditional DEA model underestimated the risk of the combination of investment opportunities and classified some of them as efficient, even though some improvement in the risk criterion is possible. Since the work [19], several DC DEA classes of models were investigated. Note that previously several attempts can be found in the literature, in particular in [11, 12, 17] which were focused on mean–variance, and mean–variance–skewness efficiency. They also introduced shadow utility functions based on the moment criteria. Paper [6] dealt with DC DEA models based on general deviation measures and investigated the strength of the proposed models as well as inclusion of condition on sparsity of portfolios. In [7], the models were generalized and the analysis was extended to coherent risk measures using the directional distance measures. Bootstrap technique was employed to investigate the empirical properties and stability of the models and resulting scores. The dynamic extension was introduced by [21]. They decomposed the overall efficiency of mutual funds over the whole investment period into efficiencies at individual investment periods taking into account dependence among the periods. Paper [8] studied models with Value at Risk inputs and proposed tractable reformulations. Traditional DEA models were used to approximate the efficient frontier and to assess performance of portfolios by [24]. In [22], two directional distance based diversification super-efficiency models for discriminating efficient funds were proposed. Paper [29] was focused on robustness and integrated parameter uncertainty into diversification-consistent DEA models leading to bi-level problems which were then transformed into equivalent single-level DEA problems. Note that in many cases, for discretely distributed returns and proper choices of risk measures, the authors showed that the proposed models can be formulated as linear programming problems which enables to solve even large instances of the obtained problems to optimality.

An important research topic is relation of DEA efficiency and stochastic dominance efficiency. The efficiency with respect to stochastic dominance is a well established concept in financial mathematics since [15, 16], see also [20]. In DEA literature, we can find several papers which were investigating relations to stochastic dominance efficiency, in particular [25] introduced several models which are consistent with second order stochastic dominance (SSD), whereas [18] proposed equivalent models. In [9], an equivalence between new class of diversification-consistent

¹ Charles University in Prague
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Sokolovská 83, Prague 186 75, Czech Republic
tel.: +420 221 913 404, fax: +420 222 323 316, branda@karlin.mff.cuni.cz

DEA models and stochastic dominance efficiency tests with respect to SSD was shown. This relation was further elaborated by [7]. The equivalences were then generalized to N -order stochastic dominance efficiency tests in [10].

Recently, [2] proposed new approach how to incorporate the risk aversion of a particular investor into the DC DEA framework which is equivalent with SSD efficiency. They derived a shadow utility which renders the DC-DEA/SSD efficient portfolios as optimal for the investor. We will focus on this approach and propose an additional sensitivity analysis with respect to the investor risk aversion. The approach relies on spectral risk measures which were proposed by [1] as a special class of coherent risk measures [4]. Using the proper choice of the risk spectra, we can identify the optimal investment opportunity for any risk-averse investor, see [28].

The paper is organized as follows. Section 2 reviews the DC DEA models and the basic notation of efficiency with respect to the second order stochastic dominance. In Section 3, an approach to risk aversion based on spectral risk measures is summarized. Section 4 provides a numerical study with a special attention to sensitivity with respect to the risk aversion.

Below we will assume that n assets with random rates of return R_i are available and we can use any (nonnegative¹) combination to compose a portfolio. This leads to the following sets of available investment opportunities, or simply portfolios:

$$\mathcal{X} = \left\{ \sum_{i=1}^n x_i R_i : \sum_{i=1}^n x_i = 1, x_i \geq 0 \right\}. \quad (1)$$

2 Diversification consistent DEA models

First, we review the general formulation of a diversification-consistent DEA model as it was proposed in [7]. It employs J return measures \mathcal{E}_j as the outputs and K coherent risk measures \mathcal{R}_k as the inputs. Coherent risk measures were proposed in [4] as real functionals on $\mathcal{L}_p(\Omega)$ space with finite p -th moment (usually $p \in \{1, 2\}$), which fulfill the following axioms:

- (R1) translation equivariance: $\mathcal{R}(X + c) = \mathcal{R}(X) - c$ for all $X \in \mathcal{L}_p(\Omega)$ and constants $c \in \mathbb{R}$,
- (R2) positive homogeneity: $\mathcal{R}(0) = 0$, and $\mathcal{R}(\lambda X) = \lambda \mathcal{R}(X)$ for all $X \in \mathcal{L}_p(\Omega)$ and all $\lambda \geq 0$,
- (R3) subadditivity: $\mathcal{R}(X_1 + X_2) \leq \mathcal{R}(X_1) + \mathcal{R}(X_2)$ for all $X_1, X_2 \in \mathcal{L}_p(\Omega)$,
- (R4) monotonicity: $\mathcal{R}(X_1) \leq \mathcal{R}(X_2)$ when $X_1 \geq X_2$, $X_1, X_2 \in \mathcal{L}_p(\Omega)$.

Note that the axioms (R2) and (R3) imply convexity. We say that \mathcal{E} is a return measure if there exists a coherent risk measure \mathcal{R} such that $\mathcal{E} = -\mathcal{R}$. Since both coherent risk as well as return measures can take positive as well as negative values, the DC DEA models proposed by paper [7] were based on the directional distance measures where, for a benchmark portfolio $X_0 \in \mathcal{X}$, the directions are defined as

$$e_j(X_0) = \max_{X \in \mathcal{X}} \mathcal{E}_j(X) - \mathcal{E}_j(X_0), d_k(X_0) = \mathcal{R}_k(X_0) - \min_{X \in \mathcal{X}} \mathcal{R}_k(X). \quad (2)$$

These directions quantify the maximal possible improvements over the risk and return measures for the benchmark portfolio X_0 to reach the efficient frontier. The frontier corresponds to the strong Pareto–Koopmans efficiency, i.e. we say that X_0 is efficient, if there is not other portfolio $X_1 \in \mathcal{X}$ such that

$$\mathcal{R}_k(X_1) \leq \mathcal{R}_k(X_0), \forall k, \mathcal{E}_j(X_1) \geq \mathcal{E}_j(X_0), \forall j,$$

with at least one inequality strict. This efficiency can be then accessed by the following diversification-consistent DEA model based on directional distance measure:

$$\begin{aligned} \min_{\theta_k, \varphi_j, x_i} & \frac{1 - \frac{1}{K} \sum_{k=1}^K \theta_k}{1 + \frac{1}{J} \sum_{j=1}^J \varphi_j} \\ \text{s.t. } & \mathcal{E}_j \left(\sum_{i=1}^n R_i x_i \right) \geq \mathcal{E}_j(X_0) + \varphi_j \cdot e_j(X_0), \quad j = 1, \dots, J, \\ & \mathcal{R}_k \left(\sum_{i=1}^n R_i x_i \right) \leq \mathcal{R}_k(X_0) - \theta_k \cdot d_k(X_0), \quad k = 1, \dots, K, \\ & \sum_{i=1}^n x_i = 1, \quad x_i \geq 0, \quad \varphi_j \geq 0, \quad \theta_k \geq 0, \end{aligned} \quad (3)$$

¹ Short-sales are not allowed.

where φ_j denotes the fraction of the improvement of the optimal portfolio $\sum_{i=1}^j R_i x_i$ over the maximal possible improvement in return \mathcal{E}_j . Similarly, θ_k denotes the fraction of the improvement of the optimal portfolio $\sum_{i=1}^j R_i x_i$ over the maximal possible improvement in risk \mathcal{R}_j . The objective function quantifies the mean improvement in risks in the numerator and the mean improvements in returns in the denominator. If the optimal is equal to one, then X_0 is identified as efficient, otherwise it is inefficient and the optimal solution (weights x_i) corresponds to an efficient portfolio which can be seen as a projection to an efficient frontier and used to revise the inefficient portfolio. However, the projection need not be in relation with the investor's risk aversion.

Formal definition of the second-order stochastic dominance (SSD) efficiency over $\mathcal{L}_1(\Omega)$ space is based on the twice cumulative probability distribution function of $X \in \mathcal{L}_1(\Omega)$ defined by

$$F_X^{(2)}(t) = \int_{-\infty}^t F_X(\eta) d\eta,$$

where $F_X(t) = P(X \leq t)$ is cdf. We say that X dominates \tilde{X} with respect to the second-order stochastic dominance (SSD), $\tilde{X} \leq_{SSD} X$, if and only if

$$F_X^{(2)}(t) \leq F_{\tilde{X}}^{(2)}(t), \quad \forall t \in \mathbb{R}, \quad (4)$$

The relation is strict, i.e. $\tilde{X} <_{SSD} X$, if the inequality is strict for at least one $t \in \mathbb{R}$. We say that $X \in \mathcal{X}$ is SSD efficient if there is no other $\tilde{X} \in \mathcal{X}$ for which it holds $X <_{SSD} \tilde{X}$. Paper [7] showed that if the distribution of random return is discrete with S equiprobable realizations, the inputs in DC DEA model (3) correspond to Conditional Value at Risks (CVaRs, [27]) on levels $1/S, 2/S, \dots, 1$ and the output is the expected return, then the resulting DC DEA model is equivalent to SSD efficiency tests. In other words, the portfolio is SSD efficient if and only if it is DC DEA efficient.

3 Risk aversion and spectral risk measures

Spectral risk measures (SRM), cf. [1], is a special subclass of the coherent risk measures. They are defined as the weighted quantiles of the random returns

$$M_\phi(X) = - \int_0^1 F_X^{-1}(p) \phi(p) dp \quad (5)$$

where we consider the quantile function

$$F_X^{-1}(p) = \min\{x : F_X(x) \geq p\}, \quad p \in [0, 1], \quad (6)$$

and an admissible risk spectrum which must be:

(A1)positive: for all $I \subseteq [0, 1]$ holds

$$\int_I \phi(p) dp \geq 0,$$

(A2)non-increasing: for all $q \in (0, 1)$ and $\varepsilon > 0$ such that $[q - \varepsilon, q + \varepsilon] \subset [0, 1]$, holds

$$\int_{q-\varepsilon}^q \phi(p) dp \geq \int_q^{q+\varepsilon} \phi(p) dp,$$

(A3)normalized:

$$\|\phi\| = \int_0^1 \phi(p) dp = 1.$$

Note that Conditional Value at Risk on level α can be obtained as a special case for risk spectrum for the risk spectrum

$$\phi(p) = \frac{1}{1-\alpha} \mathbb{I}\{0 \leq p \leq 1-\alpha\}.$$

In general, investors can identify their risk aversion by choosing the risk spectrum ϕ and derive the admissible empirical risk spectrum using the formula

$$\phi_s = \frac{\phi(s/S)}{\sum_{s=1}^S \phi(s/S)}. \quad (7)$$

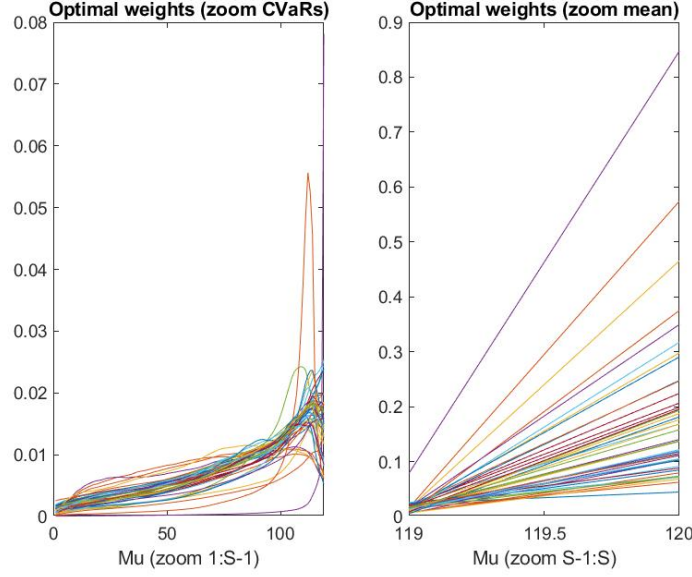


Figure 1 Optimal weights μ_s for efficient projected portfolios

Each empirical spectral risk measure (5) can be then expressed as

$$M_\phi^S(X) = \sum_{s=1}^S \mu_s \text{CVaR}_{1-s/S}^S(X) = \mu_S \mathbb{E}(-X) + \sum_{s=1}^{S-1} \mu_s \text{CVaR}_{1-s/S}^S(X), \quad (8)$$

where the weights can be derived from the empirical risk spectrum using the relation

$$\mu_s = s(\phi_s - \phi_{s+1}), \quad s = 1, \dots, S, \quad (9)$$

together with $\phi_{S+1} \equiv 0$. If the investors identify the risk spectrum corresponding to their risk aversion, they can obtain an ideal portfolio by solving the problem (8).

From Proposition 4.1 in [7] we know that the optimal solution of DC DEA model (3) under mean–CVaRs choice of the output–inputs corresponds to the SSD efficient portfolio and using Corollary 3.3 in [2] we can construct the shadow empirical risk spectrum which renders the projection as optimal in SRM minimization problem (8). If \bar{x} are the optimal portfolio weights for the benchmark portfolio X_0 and $\bar{X} = \sum_{i=1}^n R_i \bar{x}_i$ the corresponding random return of the optimal (efficient) portfolio, then weights can be obtained as

$$\begin{aligned} \mu_S &= \frac{1}{e(X_0)} \sum_{s=1}^{S-1} \frac{\text{CVaR}_{s/S}^S(\bar{X}) - \text{CVaR}_{s/S}^S(X_0) + d_s(X_0)}{d_s(X_0)}, \\ \mu_s &= \frac{1}{d_{S-s}(X_0)} \frac{\mathbb{E}(\bar{X}) - \mathbb{E}(X_0) + e(X_0)}{e(X_0)}, \quad s = 1, \dots, S-1, \end{aligned} \quad (10)$$

and the shadow empirical risk spectrum as

$$\phi_s = \sum_{t=s}^S \frac{\mu_t}{t}, \quad s = 1, \dots, S. \quad (11)$$

Figure 1 shows the weights obtained for the projected industry representative portfolios which are analyzed in the numerical study. We can compare the obtained empirical risk spectrum with various theoretical risk spectra which model various levels of investor’s risk aversion.

4 Empirical study and sensitivity analysis

In this section, we access efficiency of the industry representative portfolios of US stock market which are listed in the Kenneth French online library. We consider monthly returns between 2012 and 2020. We apply the above

	$k \in \{1, 2, 3, 4\}$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$	$k = 11$	$k = 12$	$k \in \{13, 14, 15\}$
Clths	6	6	6	5	3	17	43	42	42	42
Hlth	21	21	21	21	21	18	15	25	26	27
MedEq	5	5	5	4	2	14	41	43	43	43
Chems	16	16	16	16	15	8	27	30	32	32
Txtls	3	3	2	1	5	25	45	45	45	45
BldMt	17	17	17	17	16	9	26	32	31	31
Cnstr	36	36	36	36	36	35	12	15	13	12
FabPr	20	20	20	20	19	16	23	29	28	28
Mach	18	18	18	18	17	10	29	31	30	30
Ships	1	2	3	7	20	40	47	47	47	47
BusSv	7	7	7	6	4	13	40	41	41	41
Comps	25	25	25	25	25	22	18	26	25	23
LabEq	2	1	1	2	9	32	46	46	46	46
Banks	19	19	19	19	18	12	16	28	29	29
Insur	10	10	10	10	8	1	33	37	38	38
REst	9	9	9	9	7	5	37	38	39	39
Oil	47	47	47	47	47	47	42	5	5	1

Table 1 Representative portfolios with most changes in the ranking with respect to the risk aversion

introduced approaches, in particular we will investigate the sensitivity of the proposed DC DEA model and its solutions with respect to various risk aversions expressed by exponential risk spectrum:

$$\phi_k(p) = \frac{k \cdot e^{-k \cdot p}}{1 - e^{-k}}, \quad k > 0, \quad (12)$$

We will compare the derived shadow risk spectrum and the investor's one. We consider parameters $k \in \{1, 2, \dots, 15\}$ which cover most of the realistic risk aversions of real investors.

Table 1 contains the industry representative portfolios with most changes in ranking according to the risk aversion and using the distance between the ideal and projected portfolios. We can observe that portfolio Ships is the best according to the risk aversion represented by parameters $k \in \{1, 2, 3, 4\}$, whereas is one the worst for $k \in \{10, \dots, 15\}$. On the other hand, portfolio Oil, which is the best for $k \in \{13, 14, 15\}$ is very far from the ideal portfolio for $k \in \{1, \dots, 10\}$. To summarize, we can observe that the ranking is highly dependent of the risk aversion level.

5 Conclusions

In this paper, we have reviewed the diversification-consistent DEA models which are equivalent to the stochastic dominance tests with respect to the second order stochastic dominance. We have proposed a sensitivity analysis of the ranking of considered industry representative portfolios with respect to the various levels of the investor's risk aversion. In particular, we have compared distances between the shadow and the theoretical (empirical) risk spectra showing high dependence of the ranking on the risk aversion parameters. More demanding models are postponed as a topic for future research where they can be solved using the numerical technique proposed in [3].

Acknowledgements

This work was supported by the Grant Agency of the Czech Republic under the grant project 19-28231X.

References

- [1] Acerbi, C.: Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance* **26** (2002), 1505–1518.
- [2] Adam, L., Branda, M.: Risk-aversion in data envelopment analysis models with diversification. *Omega* **102** (2021), 102338.

- [3] Adam, L., Branda, M., Heitsch, H., Henrion, R.: Solving joint chance constrained problems using regularization and Benders' decomposition. *Annals of Operations Research* **292** (2020), 683–709.
- [4] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. *Mathematical Finance* **9** (1999), 203–228.
- [5] Basso, A., Funari, S.: A data envelopment analysis approach to measure the mutual fund performance. *European Journal of Operational Research* **135** (3) (2001), 477–492.
- [6] Branda, M.: Diversification-consistent data envelopment analysis with general deviation measures. *European Journal of Operational Research* **226** (3) (2013), 626–635.
- [7] Branda, M.: Diversification-consistent data envelopment analysis based on directional-distance measures. *Omega* **52** (2015), 65–76.
- [8] Branda, M.: Mean-value at risk portfolio efficiency: approaches based on data envelopment analysis models with negative data and their empirical behaviour. *4OR* **14** (1) (2016), 77–99.
- [9] Branda, M., Kopa, M.: On relations between DEA-risk models and stochastic dominance efficiency tests. *Central European Journal of Operations Research* **22** (1) (2014), 13–35.
- [10] Branda, M., Kopa, M.: DEA models equivalent to general N-th order stochastic dominance efficiency tests. *Operations Research Letters* **44** (2) (2016), 285–289.
- [11] Briec, W., Kerstens, K., Lesourd, J.-B.: Single period Markowitz portfolio selection, performance gauging and duality: a variation on the Luenberger shortage function. *Journal of Optimization Theory and Applications* **120** (1) (2004), 1–27.
- [12] Briec, W., Kerstens, K., Jokung, O.: Mean-variance-skewness portfolio performance gauging: A general shortage function and dual approach. *Management Science* **53** (2007), 135–149.
- [13] Charnes, A., Cooper, W., Rhodes, E.: Measuring the efficiency of decision-making units. *European Journal of Operational Research* **2** (1978), 429–444.
- [14] Chen, Z., Lin, R.: Mutual fund performance evaluation using data envelopment analysis with new risk measures. *OR Spectrum* **28** (2006), 375–398.
- [15] Hadar, J., Russell, W.R.: Rules for ordering uncertain prospects. *American Economic Review* **9** (1969), 25–34.
- [16] Hanoch, G., Levy H.: The efficient analysis of choices involving risk. *Review of Economic Studies* **36** (3) (1969), 335–346.
- [17] Joro, T., Na, P.: Portfolio performance evaluation in a mean–variance–skewness framework. *European Journal of Operational Research* **175** (2006), 446–461.
- [18] Kuosmannen, T.: Performance measurement and best-practice benchmarking of mutual funds: combining stochastic dominance criteria with data envelopment analysis. *Journal of Productivity Analysis* **28** (2007), 71–86.
- [19] Lamb, J.D., Tee, K-H.: Data envelopment analysis models of investment funds. *European Journal of Operational Research* **216** (3) (2012), 687–696.
- [20] Levy, H.: *Stochastic dominance: Investment decision making under uncertainty*. Second edition, Springer, New York, 2006.
- [21] Lin, R., Chen, Z., Hu, Q., Li, Z.: Dynamic network DEA approach with diversification to multiperiod performance evaluation of funds. *OR Spectrum* **39** (2017), 821–860.
- [22] Lin, R., Li, Z.: Directional distance based diversification super-efficiency DEA models for mutual funds. *Omega* **97** (2020), 102096.
- [23] Liu, J.S., Lu, L.Y.Y., Lu, W.-M., Lin, B.J.Y.: Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega* **41** (1) (2013), 3–15.
- [24] Liu, W., Zhou, Z., Liu, D., Xiao, H.: Estimation of portfolio efficiency via DEA. *Omega* **52** (2015), 107–118.
- [25] Lozano, S., Gutiérrez, E.: Data envelopment analysis of mutual funds based on second-order stochastic dominance. *European Journal of Operational Research* **189** (2008), 230–244.
- [26] Murthi, B.P.S., Choi, Y.K., Desai, P.: Efficiency of mutual funds and portfolio performance measurement: a non-parametric approach. *European Journal of Operational Research* **98** (2) (1997), 408–418.
- [27] Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* **26** (2002), 1443–1471.
- [28] Wächter, H.P., Mazzoni, T.: Consistent modeling of risk averse behavior with spectral risk measures. *European Journal of Operational Research* **229** (2) (2013), 487–495.
- [29] Xiao, H., Ren, T., Zhou, Z., Liu, W.: Parameter uncertainty in estimation of portfolio efficiency: Evidence from an interval diversification-consistent DEA approach. *Omega* **103** (2021), 102357.

Efficient Values of Selected Factors of DMUs Production Structure Using Particular DEA Model

Helena Brožová¹, Milan Vlach²

Abstract. We suggest a particular application of the data envelopment analysis method for estimation of selected inputs and outputs values simultaneously to create all decision-making units as efficient as possible. It is an iterative approach allowing the decision-maker to find an efficient pattern for all decision making units. The proposed method is based on the Cook, Green and Zhu [6] approach and it is illustrated by simple example.

Keywords: Efficiency, efficient value of input, efficient value of output, dual role factor, CCR model

JEL Classification: C44, C61, D24

AMS Classification: 90B50, 90C05, 90C90

1 Introduction

A number of methods have been developed for the analysis of decision making units' (DMU) efficiency and evaluation of their inputs and outputs. In general, the more efficient the unit is, the less inputs it consumes and the more outputs it produces. It is a multiple attribute analysis problem, so the problem can be approach by the multiple criteria decision methods. It can be also approach by parametric methods based on statistical tools like regression analysis. There are also nonparametric approaches that use mathematical programming, like the data envelopment analysis (DEA) proposed and developed by Charnes, Cooper and Rhodes [4] and Banker, Charnes and Cooper [1].

The DEA measures the productive efficiency of the production process of DMUs on the basis of their own inputs and outputs. Typically, these models give an efficiency index for monitored DMU, which means to achieve an efficient pattern for the DMU by all inputs decreasing or all outputs increasing. However, changing all the parameters of a monitored DMU leading to its efficiency may not always be possible and we may assume that it would often be necessary to determine changes in only some selected factors. Therefore, different problems occur; namely, how to find the selected inputs, outputs or both values maintaining the other parameters of the DMU at which the unit will be efficient or to find this value for all monitored DMUs to be efficient. Cook, Green and Zhu [6] dealt with the similar 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 value 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].

In this paper, we suggest a specific DEA method to create efficient patterns for all DMUs by estimation of selected inputs and outputs values simultaneously. The method is based on the Cook, Green and Zhu [6] approach and its application is demonstrated by the small example consists of 5 DMUs with 2 inputs and 1 output. One input and one output can be always changed. This example is used, because it can be represented also graphically.

The paper is organized as follows. Section 2 introduces briefly the Charnes, Cooper and Rhodes (CCR) DEA model and the dual-role factors DEA model. Section 3 describes our suggested method for evaluation of selected inputs and outputs to achieve efficiency of all DMUs. Section 4 contains example illustrating how the method can be applied. Section 5 concludes the paper by summarizing the paper findings.

¹ ICzech University of Life Sciences, Faculty of Economics and Management, Department of Systems Engineering, Kamýcká 129, Prague 6, Czech Republic, brozova@pef.czu.cz

² Charles University, Faculty of Mathematics and Physics, Department of Theoretical Computer Science and Mathematical Logic, Malostranské náměstí 25, Prague 1, Czech Republic, milan.vlach@mff.uni.cz

2 DEA models and efficient inputs and outputs values

The DEA models are generally used to evaluate efficiency of units and to set the necessary changes in inputs (or outputs) for inefficient units to increase their efficiency and to reach the efficiency frontier. Charnes, Cooper and Rhodes [4] introduced a model, now called the CCR DEA model, for evaluation of the efficiency of a set of DMUs that transforms multiple inputs into multiple outputs under the assumption of the constant returns to scale. The outputs and inputs can be of various characteristics and of variety of forms that can be sometimes difficult to measure.

The input oriented CCR DEA model [4] measuring efficiency of DMU is constructed as the maximization of the ratio of weighted outputs to weighted inputs under the constraints that, for all DMUs, this ratio is less than or equal to 1 and its value for the efficient DMU is equal to 1. Let H be the DMU which efficiency is measured. Linearization of CCR DEA model leads to the following linear programming problem (envelopment model):

$$\begin{aligned} \text{Maximize } & \sum_{j=1}^n u_{jH} y_{jH} & \text{subject to } & \sum_{i=1}^m v_{iH} x_{iH} = 1 \\ & & & - \sum_{i=1}^m v_{iH} x_{ik} + \sum_{j=1}^n u_{jH} y_{jk} \leq 0, k \in K = \{1, 2, \dots, p\} \\ & & & u_{jH} \geq 0, j = 1, 2, \dots, n \\ & & & v_{iH} \geq 0, i = 1, 2, \dots, m \end{aligned} \quad (1)$$

and the corresponding dual problem (multiplier model):

$$\begin{aligned} \text{Minimize } & \Phi_H & \text{subject to } & \Phi_H x_{iH} - \sum_{k \in K} \lambda_k x_{ik} - s_i^- = 0, i = 1, 2, \dots, m \\ & & & y_{jk} - \sum_{k \in K} \lambda_k y_{jk} + s_j^+ = 0, j = 1, 2, \dots, n \\ & & & \lambda_k \geq 0, k \in K \end{aligned} \quad (2)$$

where variables u_{jk} and v_{ik} are weights of outputs and inputs, λ_k are multipliers, s_i^-, s_j^+ are slack variables, y_{jk} is the value of j^{th} output from unit k , and x_{ik} is the value of i^{th} input to k^{th} DMU, H is index of the evaluated DMU. This notation is used through the whole paper.

Let $(\mathbf{u}^*, \mathbf{v}^*, \boldsymbol{\lambda}^*, \mathbf{s}^{*-}, \mathbf{s}^{*+})$, $i = 1, 2, \dots, m, j = 1, 2, \dots, n$, be an optimal solution of the problems (1) and (2), and let Φ_H^* be the optimal value of both objective functions.

There are two basic results of the DEA models. First, the **efficiency score** Φ_H^* for the DMU H being evaluated. Second, the **input and output levels that the individual DMU should theoretically use to be as efficient** as the other efficient DMUs, which will be calculated by

$$\begin{aligned} x'_{iH} &= \Phi_H^* x_{iH} - s_i^{*-}, i = 1, 2, \dots, m \\ y'_{jH} &= y_{jH} + s_j^{*+}, j = 1, 2, \dots, n \end{aligned} \quad (3)$$

However, changing all the parameters of a monitored DMU leading to its efficiency may not always be possible and we may assume that it would often be necessary to determine changes in only some particular factors. The problem of possible reallocation of one factor was investigated by Cook, Green and Zhu [6] who studied the cases where some factors of the DMUs can play input or output role and if some reallocation of this factor among the DMUs improves their efficiency. They improve the idea of Beasley [2] who incorporated the dual-role factor twice into the DEA model, once as input and once as output. Cook, Green and Zhu [6] also suggested a model for optimal allocation of dual-role factor under the condition that the whole amount of such factor has to be exploited.

Brozova and Bohackova (2020) dealt with the simple problem of determining the values of selected inputs while maintaining the number of other inputs and outputs so that the DMUs are efficient. Suggested model was inspired by the model of Cook, Green and Zhu [6].

Let $f_{lk}, l = 1, \dots, p, k \in K$ be the inputs whose values we wish to determine so that DMUs were as efficient as possible. The proposed model can be linearized using new variables $\delta_{lk} = \alpha_l f_{lk}, l = 1, \dots, p, k \in K$. Linear form of model of Cook, Green and Zhu [6] has the following form:

$$\begin{aligned}
\text{Maximize } & \sum_{k \in K} \sum_{j=1}^n u_j y_{jk} \quad \text{subject to} \\
& \sum_{k \in K} \left(\sum_{i=1}^m v_i x_{ik} + \sum_{l=1}^p \delta_{lk} \right) = 1 \\
& - \sum_{i=1}^m v_i x_{ik} - \sum_{l=1}^p \delta_{lk} + \sum_{j=1}^n u_j y_{jk} \leq 0, k \in K \\
& \alpha_l w_{lk}^l \leq \delta_{lk} \leq \alpha_l w_{lk}^u, l = 1, \dots, p, k \in K \\
& u_j \geq 0, j = 1, 2, \dots, n \\
& v_i \geq 0, i = 1, 2, \dots, m \\
& \alpha_l \geq 0, \delta_{lk} \geq 0, l = 1, \dots, p, k \in K
\end{aligned} \tag{4}$$

3 Estimation of the efficient values of selected inputs and outputs

Suppose now we want to change only selected parameters of a monitored DMU leading to efficiency of all DMUs. To achieve it, we propose a method how to find the optimal values of selected inputs and outputs with the aim the all DMUs are efficient which is based on the model (4). Let $f_{lk}, l = 1, \dots, p, k \in K$ are the inputs and $g_{qk}, q = 1, \dots, r, k \in K$ are the outputs whose values we wish to determine so that DMUs become as efficient as possible. The proposed model (an extension of linear form of model (4)) is as follows:

$$\begin{aligned}
\text{Maximize } & \sum_{k \in K} \left(\sum_{j=1}^n u_j y_{jk} + \sum_{q=1}^r \mu_q g_{qk} \right) \quad \text{subject to} \\
& \sum_{k \in K} \left(\sum_{i=1}^m v_i x_{ik} + \sum_{l=1}^p \alpha_l f_{lk} \right) = 1 \\
& - \sum_{i=1}^m v_i x_{ik} - \sum_{l=1}^p \alpha_l f_{lk} + \sum_{j=1}^n u_j y_{jk} + \sum_{q=1}^r \mu_q g_{qk} \leq 0, k \in K \\
& w_{lk}^l \leq f_{lk} \leq w_{lk}^u, l = 1, \dots, p, k \in K \\
& z_{qk}^l \leq g_{qk} \leq z_{qk}^u, q = 1, \dots, r, k \in K \\
& u_j \geq 0, j = 1, 2, \dots, n \\
& v_i \geq 0, i = 1, 2, \dots, m \\
& \alpha_l \geq 0, f_{lk} \geq 0, l = 1, \dots, p, k \in K \\
& \mu_q \geq 0, g_{qk} \geq 0, q = 1, \dots, r, k \in K
\end{aligned} \tag{5}$$

Using new variables $\delta_{lk} = \alpha_l f_{lk}, \sigma_{qk} = \mu_q g_{qk} \geq 0, q = 1, \dots, r, l = 1, \dots, p, k \in K$, we obtain the following linearization of (5):

$$\begin{aligned}
\text{Maximize } & \sum_{k \in K} \left(\sum_{j=1}^n u_j y_{jk} + \sum_{l=1}^r \sigma_{qk} \right) \quad \text{subject to} \\
& \sum_{k \in K} \left(\sum_{i=1}^m v_i x_{ik} + \sum_{l=1}^p \delta_{lk} \right) = 1 \\
& - \sum_{i=1}^m v_i x_{ik} - \sum_{l=1}^p \delta_{lk} + \sum_{j=1}^n u_j y_{jk} + \sum_{l=1}^r \sigma_{qk} \leq 0, k \in K \\
& \alpha_l w_{lk}^l \leq \delta_{lk} \leq \alpha_l w_{lk}^u, l = 1, \dots, p, k \in K \\
& \mu_q z_{qk}^l \leq \sigma_{qk} \leq \mu_q z_{qk}^u, q = 1, \dots, r, k \in K \\
& u_j \geq 0, j = 1, 2, \dots, n \\
& v_i \geq 0, i = 1, 2, \dots, m \\
& \alpha_l \geq 0, \delta_{lk} \geq 0, l = 1, \dots, p, k \in K \\
& \mu_q \geq 0, \sigma_{qk} \geq 0, q = 1, \dots, r, k \in K
\end{aligned} \tag{6}$$

Lemma 1 Both model (5) and model (6) have always optimal solutions.

Proof. The feasible set of model (6), that is, the set of points satisfying all constraints of model (6) is bounded and closed, and the objective function is linear. Therefore the maximum of the objective function attains its maximum at some point of feasible set. The feasible set of model (5) is also bounded and closed, it is enough to put $f_{kl} = \delta_{lk}/\alpha_l$ and $g_{qk} = \sigma_{qk}/\mu_q, l = 1, \dots, p, q = 1, \dots, r, k \in K$ and again its linear objective function has its maximum. ■

Note: In order to be able to calculate values f_{kl}, g_{qk} in the optimal solution of model (6), it is necessary that α_l, μ_q be larger than some very small positive value $\varepsilon \geq 0$. So, we have to use the constraints $\alpha_l \geq \varepsilon, l = 1, \dots, p$ and $\mu_q \geq \varepsilon, q = 1, \dots, r$.

Lemma 2 All DMUs are efficient if and only if, for the optimal solution of model (5) or (6), we have $\Phi^* = 1$.

Proof. A) Let all DMUs be efficient. It means $\sum_{i=1}^m v_i x_{ik}^t + \sum_{l=1}^p \alpha_l f_{lk} = \sum_{j=1}^n u_j y_{jk} + \sum_{q=1}^r \mu_q g_{qk}$.

From the constraint of aggregated inputs $\sum_{k \in K} (\sum_{i=1}^m v_i x_{ik}^t + \sum_{l=1}^p \alpha_l f_{lk}) = 1$ it follows that also aggregate outputs $\sum_{k \in K} (\sum_{j=1}^n u_j y_{jk} + \sum_{q=1}^r \mu_q g_{qk}) = 1$ and consequently $\Phi^* = 1$.

B) Suppose now the optimal aggregate efficiency $\Phi^* = \sum_{k \in K} (\sum_{j=1}^n u_j y_{jk} + \sum_{q=1}^r \mu_q g_{qk}) = 1$ and of course for feasible solution the constraint $\sum_{k \in K} (\sum_{i=1}^m v_i x_{ik}^t + \sum_{l=1}^p \alpha_l f_{lk}) = 1$ is fulfilled.

Suppose the inefficient DMU h exists, which means that $\sum_{i=1}^m v_i x_{ih}^t + \sum_{l=1}^p \alpha_l f_{lh} > \sum_{j=1}^n u_j y_{jh} + \sum_{q=1}^r \mu_q g_{qh}$. In this case some DMU d exists, for which has to be $\sum_{i=1}^m v_i x_{id}^t + \sum_{l=1}^p \alpha_l f_{ld} < \sum_{j=1}^n u_j y_{jd} + \sum_{q=1}^r \mu_q g_{qd}$. In this case $-\sum_{i=1}^m v_i x_{id}^t - \sum_{l=1}^p \alpha_l f_{ld} + \sum_{j=1}^n u_j y_{jd} + \sum_{q=1}^r \mu_q g_{qd} > 0$. It means, that at least one constraint is not valid. Such solution is infeasible and input oriented efficiency of such DMU would be greater than 1. This implies that all DMUs have to be efficient. ■

Lemma 3 There are such lower and upper bounds $w_{lk}^L, w_{lk}^U, z_{qk}^L, z_{qk}^U, l = 1, \dots, p, q = 1, \dots, r, k \in K$ that the optimal solution of model (5) and also (6) have value $\Phi^* = 1$.

Proof. If $\Phi^* = 1$, then for all DMUs is $\sum_{i=1}^m v_i x_{ik}^t + \sum_{l=1}^p \alpha_l f_{lk} = \sum_{j=1}^n u_j y_{jk} + \sum_{q=1}^r \mu_q g_{qk}$ (lemma 2) and also $\sum_{k \in K} (\sum_{i=1}^m v_i x_{ik}^t + \sum_{l=1}^p \delta_{lk}) = 1$.

Because $w_{lk}^L \leq f_{lk} \leq w_{lk}^U$, it has to be $\alpha_l w_{lk}^L \leq \delta_{lk} \leq \alpha_l w_{lk}^U$ and also because $z_{qk}^L \leq g_{qk} \leq z_{qk}^U$ it is also $\mu_q z_{qk}^L \leq \sigma_{qk} \leq \mu_q z_{qk}^U$.

It is possible to find the values of the lower and upper limits to which following inequalities hold for $k \in K$

$$\sum_{l=1}^p \alpha_l w_{lk}^L - \sum_{q=1}^r \mu_q z_{qk}^U \leq \sum_{l=1}^p \alpha_l f_{lk} - \sum_{q=1}^r \mu_q g_{qk} = \sum_{j=1}^n u_j y_{jk} - \sum_{i=1}^m v_i x_{ik}^t \leq \sum_{l=1}^p \alpha_l w_{lk}^U - \sum_{q=1}^r \mu_q z_{qk}^L.$$

Because of nonnegativity of all parameters, for instance, it is possible to set for all DMUs

- $w_{lk}^L = \text{small}W, l = 1, \dots, p, k \in K$ where $\text{small}W \leq \min_{l=1, \dots, p, k \in K} (f_{lk})$ and $w_{lk}^U = f_{lk}, l = 1, \dots, p, k \in K$,
- $z_{qk}^L = g_{qk}, q = 1, \dots, r, k \in K$ and $z_{qk}^U = \text{big}Z, q = 1, \dots, r, k \in K$, where $\text{big}Z \geq \max_{q=1, \dots, r, k \in K} (g_{qk})$. ■

Although it is theoretically possible to find the required values of inputs and outputs so that all DMUs are effective, the results obtained in this way may not always be feasible in practice. Therefore, the following iterative procedure is often terminated even if not all DMUs are effective. In this case, it is necessary to relax the limits for the searched values of inputs and outputs; that is, to reduce the lower limits of inputs and increase the upper limits of outputs

3.1 Iterative procedure

The procedure for obtaining the values of the selected inputs and outputs so that all units are as efficient as possible has the following steps:

1. For efficient DMUs and the selected inputs set the values $w_{lk}^L = w_{lk}^U = f_{lk}, l = 1, \dots, p, k \in K$ *efficient*, and the selected outputs set $z_{qk}^L = z_{qk}^U = g_{qk}, q = 1, \dots, r, k \in K$ *efficient*.
2. For nonefficient DMUs and the selected inputs choose arbitrary the values $w_{lk}^L = \text{small}W = \min_{l=1, \dots, p, k \in K} (f_{lk}), l = 1, \dots, p, k \in K$ *nonefficient* and $w_{lk}^U = f_{lk}, l = 1, \dots, p, k \in K$ *nonefficient*.
3. For nonefficient DMUs and the selected outputs choose arbitrary the values $z_{qk}^U = \text{big}Z, q = 1, \dots, r, k \in K$ *nonefficient* and $z_{qk}^L = g_{qk}, q = 1, \dots, r, k \in K$ *nonefficient*.
4. Solve the model (6) and find the optimal solution
5. If the optimal value $\Phi^* < 1$ or if you are not satisfied with the results, then
 - set the new higher upper bounds z_{qk}^U and/or smaller lower bounds z_{qk}^L for outputs and higher upper bounds w_{lk}^U for inputs and/or smaller lower bounds w_{lk}^L and go to the step 4,
 - else continue to step 6.
6. If the $\Phi^* = 1$ or is near enough to 1, end the iterative procedure.
7. Calculate the optimal values of the selected inputs as $f_{lk} = \delta_{lk} / \alpha_l, l = 1, \dots, p, k \in K$ and outputs as $g_{qk} = \sigma_{qk} / \mu_q, q = 1, \dots, r, k \in K$ and calculate the efficiency of each DMUs according to the CCR DEA model.

4 Application of suggested procedure

Suggested iterative procedure is illustrated on the following small example. The efficiency of five DMUs with two inputs and one output is evaluated. Although this example contains few DMUs regarding to the number of inputs and outputs, it is suitable for its clarity and the possibility of graphical representation of DMUs transformed them to unit of output.

The initial state of DMUS and their CCR efficiency are in the Table 1 and Figure 1 - Initial data. It is seen that only 2 DMUs are efficient (DMU 2, DMU 3) and that efficiency of other DMUs are low (DMU5, 0.71) or very low (DMU1, 0.57, DMU4, 0.43). Graphically, the nonefficient DMUs lay far from the efficiency frontier (Figure 1 – Initial data).

Results of input oriented, output oriented resp. CCR model give also the necessary changes of inputs, output resp. of the nonefficient DMUs. The fact that only two DMUs are efficient means also that virtual efficient units are derived from DMU2 and DMU3. Graphical representation shows, that except DMU3 all other units have use the same production pattern as DMU 2 to be efficient, their virtual DMU is DMU 2 (Table 1 and Figure 1 – CCR results).

	Initial data				Input oriented CCR			Output oriented CCR		
	In 1	In 2	Out 1	Efficiency	In 1	In 2	Out 1	In 1	In 2	Out 1
DMU1	1	1	1	0.57	0.57	0.29	1	1	0.5	1.75
DMU2	4	2	7	1.00	4	2	7	4	2	7
DMU3	4	1	6	1.00	4	1	6	4	1	6
DMU4	8	5	6	0.43	3.43	1.71	6	8	4	14
DMU5	4	2	5	0.71	2.86	1.43	5	4	2	7
<i>SUM</i>	<i>21</i>	<i>11</i>	<i>25</i>		<i>14.86</i>	<i>6.43</i>	<i>25</i>	<i>21</i>	<i>9.5</i>	<i>35.75</i>

Table 1 DMU with initial and efficient values of inputs and outputs using CCR models

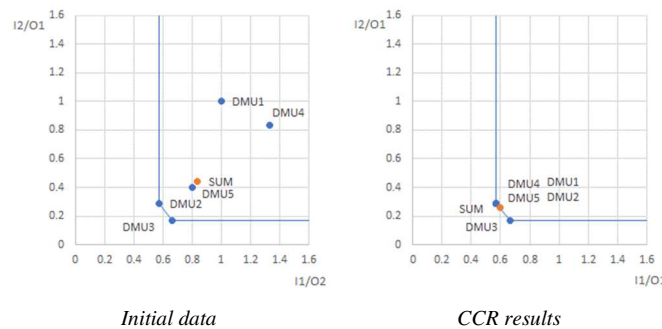


Figure 1 DMU with initial and efficient values of inputs and outputs using CCR models

Note: The orange point represents unit as sum of inputs and outputs ($\sum_{i=1}^m x_{ik}$, $\sum_{l=1}^p f_{lk}$, $\sum_{j=1}^n y_{jk}$, $\sum_{q=1}^r g_{qk}$).

Now, the efficient values of input I2 and output O1 will be calculated while the value of input I1 remains the same. Suggested approach allows to change selected inputs or outputs regardless of orientation of CCR model.

For the first iteration cycle, the lower bounds of O1 and upper bounds of I2 are set the same as the initial values of this input and output. For efficient units the lower bounds are equal to upper bounds. For non-efficient units the lower bounds for I2 was set as minimum of all I2 values, in this example it is 1, and upper bound for O1 is higher than maximum of all O1 values, in this example it is set to 10 (higher than maximum of all O1). The overall efficiency Φ^* is equal to 0.97 and efficiency of DMU 2 is 0.88.

According to the Lemma 3 it is possible to find such lower and upper bounds, that all DMUs are efficient. If the upper bound of outputs is increased to 11, the overall efficiency Φ^* is equal to 1 and all DMUs are efficient. The new values of selected factors that all units are as efficient as possible are in the column In 2* and Out 2* (Table 2, Figure 2, column First iteration).

For the second iteration cycle, the lower bounds of O1 and upper bounds of I2 are set the same and the lower bounds for I2 is equal to 1 (minimum of all I2). If the upper bound for O1 is equal to 13.9 (higher than previous upper bound but necessary for efficient DMUs) for all DMUs, then the overall efficiency Φ^* is equal to 1. All DMUs are again efficient, one initially efficient unit DMU 2 increases its O2 (Table 2, Figure 2, column Second iteration).

From graphical representation seems, that the first iteration cycle provides more realistic results as DMUs production structure is sufficiently individual, specific to each of them. Comparing results of the first and second iteration cycles, the dominated DMUs are received in the second iteration cycle. Only nondominated units is DMU 4 as shown Figure 2, Second iteration.

	Initial data				First iteration (nonefficient)				Second iteration (all DMU)			
	In 1	In 2	Out 1	Efficiency	In 1	In 2*	Out 1*	Efficiency	In 1	In 2*	Out 1*	Efficiency
DMU1	1	1	1	0.57	1	1	2.25	1.00	1	1	1.77	1.00
DMU2	4	2	7	1.00	4	2	7	0.98	4	2	7	1.00
DMU3	4	1	6	1.00	4	1	6	1.00	4	1	6.97	1.00
DMU4	8	5	6	0.43	8	1	11	1.00	8	1	13.9	1.00
DMU5	4	2	5	0.71	4	2	7	1.00	4	1	6.97	1.00
SUM	21	11	25		21	7	33.25		21	6	36.6	

Table 2 DMU data with initial and changed values of selected input I2 and output O1

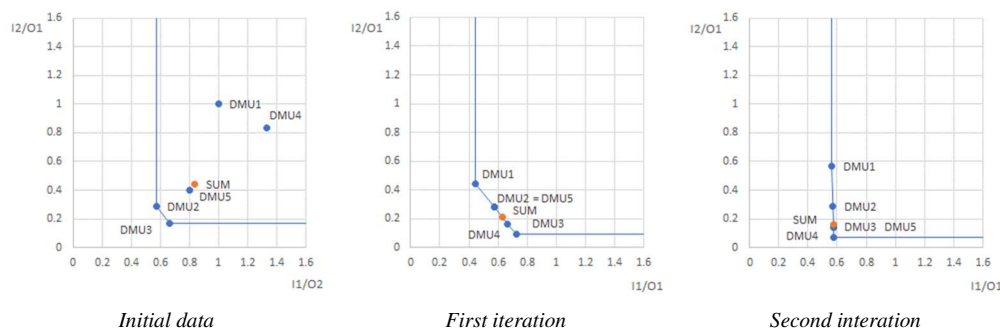


Figure 2 DMU with initial and changed values of selected input I2 and output O1

Note: The orange point represents unit as sum of inputs and outputs ($\sum_{i=1}^m x_{ik}, \sum_{l=1}^p f_{lk}, \sum_{j=1}^n y_{jk}, \sum_{q=1}^r g_{qk}$).

In this example, it is appropriate to recommend DMUs to change their production structure according to the results of the first iteration cycle, all units are efficient and, in addition, the individuality of their production process is largely preserved (given the permitted and non-permitted changes of inputs and output values).

5 Conclusion

The DEA method is widely used to analyze the efficiency of production units, and also allows the determination of the necessary changes in inputs (or outputs) to make inefficient DMUs still efficient. However, such changes lead to the erasure of differences (individualities) between the production processes of individual DMUs, because virtual units are formed as linear combinations of effective DMUs. In addition, it is often not possible to change either all inputs or all outputs of monitored DMUs. The proposed method of searching for suitable values only for selected inputs and outputs (simultaneously inputs and outputs) therefore allows to preserve the values of unchangeable parameters of DMUs and at the same time to set changeable parameters so that DMUs are as efficient as possible. In addition, by setting the suitable lower and upper limits for the values of individual changeable parameters, it is possible to find such values of inputs and outputs of DMUs that they are all effective. Since the suggested method is iterative, it allows the decision-maker to experiment and adjust the lower and upper bounds for changeable inputs and outputs so that the obtained efficient DMUs are best suited to his/her ideas and requirements.

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. (2020). Efficient Amount of Selected Input Using DEA Approach. In Kapounek, S. & Vránová, H. (Eds.), *Mathematical Methods in Economics 2020* (pp. 66 – 73). Mendelu, Brno.
- [4] Charnes, A., Cooper, W. & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research* 2, 429–444.
- [5] Chen, W-Ch.: Revisiting dual-role factors in data envelopment analysis: derivation and implications, *IIE Transactions*, 46 (2014), 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] Farrell, M. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society, Series A* 120, 253–281.

On the measurement of risk of some cointegrated trading strategies

Michal Černý¹, Vladimír Holý², Petra Tomanová³, Lucie Beranová⁴

Abstract. We consider optimal trading strategies with cointegrated portfolios, i.e. when the price processes of a family of assets have a stationary linear combination. We further assume that the portfolio price process can be modeled as the Ornstein-Uhlenbeck process and consider trading strategies of Bertram's type, meaning that positions are open on a below-average price level and closed again on an above-average price level and the adjustment of positions is associated with transaction costs. This is a kind of arbitrage trading strategy where the crucial random variable is the length of a trade cycle, i.e., the time to collection of profit. Here, profitability of a strategy is measured as profit-per-time unit in the limit over the infinite time horizon. The optimization problem reduces to finding optimal entry and exit thresholds maximizing the per-time-unit profit. We augment this optimization problem by various kinds of constraints measuring riskiness of a trading strategy, such as expected length of a trade cycle (i.e., expected time to profit collection), volatility of the length of a trade cycle or per-time-unit volatility of per-time-unit profit.

Keywords: asset trading; nonconvex optimization; cointegrated portfolio; stationary process; risk measure

JEL Classification: C13; C61

AMS Classification: 91B60

1 Introduction

Trading with cointegrated portfolios allows an investor to take advantage of mean reversion. In literature, the case with two assets is often considered and is referred to as *pairs trading*; however, in this text the trading strategy can be easily formulated for portfolios with a general number of assets (some possibly held short). Examples of cointegrated portfolios include related commodities, such as oil held long plus kerosene held short, or Brazilian coffee held long with Colombian coffee held short, or portfolios containing an asset with derivatives on the same or similar underlying, such as an asset held long plus a sale forward contract thereof.

There is a rich literature on cointegrated trading or pairs trading; recall e.g. the distance approach [5], the cointegration approach [11], the stochastic spread approach [4], the stochastic control approach [9], the machine learning approach [7], the copula approach [8] or the PCA approach [1]. For a comprehensive literature review see [10].

We follow a kind of the stochastic spread approach and extend the previous work of Bertram [2], [3] and Holý and Černý [6]. Here, the focus is on optimizing the price thresholds when positions are open and closed. This work builds also on [4]. The main building block is Bertram's optimization problem [2], where maximization of profit-per-time unit as a function of the thresholds is formulated as an unconstrained optimization problem. In this text we extend some ideas from [6] and augment the problem with constraints bounding the riskiness (measured in various ways) of the strategy.

¹ Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, cernym@vse.cz

² Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, vladimir.holy@vse.cz

³ Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, petra.tomanova@vse.cz

⁴ Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, lucie.beranova@vse.cz

2 The trading strategy

2.1 Cointegrated portfolios

Assume that there exists a family of assets A_1, \dots, A_n with continuous-time price processes $P_1(t), \dots, P_n(t)$ and that it is possible to construct a portfolio $\eta = (\eta_1, \dots, \eta_n)^T \in \mathbb{R}^n$ such that its price process

$$P(t) := \sum_{i=1}^n \eta_i P_i(t)$$

is stationary. Let us define $\mu := \mathbb{E}P(t)$ (by stationarity, μ does not depend on t), choose two levels $a < \mu < b$, called *entry threshold* and *exit threshold* and consider the following trading strategy.

- {1} Wait until time t_1 when $P(t_1) = a$. Then open positions, i.e. buy portfolio η .
- {2} Wait until time $t'_1 > t_1$ when $P(t'_1) = b$. Then close the positions (i.e. sell potfolio η , or, equivalently, buy portfolio $-\eta$). Then, repeat {1} and {2} forever.

This strategy generates a sequence of times $t_1 < t'_1 < t_2 < t'_2 < \dots$, where the time interval $[t_i, t'_i]$ refers to *open positions* and the interval $[t'_i, t_{i+1}]$ refers to *closed positions*.

Remark. This is just a convention; the strategy could be alternatively reformulated as follows.

- {1*} Wait until time t_1 when $P(t_1) = a$. Then buy portfolio η .
- {2*} Wait until time $t'_1 > t_1$ when $P(t'_1) = b$. Then buy portfolio -2η .
- {3*} Wait until time $t_2 > t'_1$ when $P(t_2) = a$. Then buy portfolio 2η and iterate forever.

This is a *symmetric version* of the trading with no idle times in time intervals $[t'_i, t_{i+1}]$. Its analysis is analogous; namely, in the limit it is a version with doubled profit compared to strategy {1}–{2}. This is why we can restrict our attention only to {1}–{2}.

Disregarding some pathological cases, the stationarity phenomenon allows us to take advantage of *mean reversion*, meaning that once the price process $P(t)$ deviates from its mean, it returns to μ in finite time. If the thresholds a, b are chosen “reasonably”, then in time t'_i we collect deterministic profit

$$\pi_0 := P(t'_i) - P(t_i) = b - a.$$

In this setup, *choice* of a trading strategy reduces to the choice of the thresholds a, b . The pair (a, b) is simply referred to as *strategy*. What is random here is the the length of a trade cycle

$$T_i := t_{i+1} - t_i.$$

The tradeoff is between the deterministic profit π_0 and the time to its collection, which is measured by the mean value of the trade cycle as

$$\mathcal{T} := \mathbb{E}T_i.$$

Thus it makes sense to standardize the profit *per unit of time*, i.e., to measure profit as

$$\Pi_0 = \lim_{t \rightarrow \infty} t^{-1} \pi_0 N(t),$$

where $N(t)$ is the counting process for the number of trade cycles in timer window $[0, t]$. By stationarity of $P(t)$, the process $N(t)$ grows at a linear rate and it follows that Π_0 is well-defined.

2.2 Transaction costs

We will assume that a transaction is associated with *transaction cost* c_i *per unit of asset* A_i , meaning that adjustment of a position in asset A_i costs c_i dollars. Then, the profit per trade cycle is

$$\pi := b - a - c,$$

including transaction costs

$$c := 2 \sum_{i=1}^n |\eta_i| c_i$$

(the factor 2 corresponds to the fact that positions are open and then closed which requires a pair of transactions). Then, profit-per-time-unit reduces to

$$\Pi \equiv \Pi(a, b) = \lim_{t \rightarrow \infty} t^{-1} \pi N_t.$$

2.3 Ornstein-Uhlenbeck process and its standardization

To find expressions for Π_0 , the crucial quantity is the random process $N(t)$. To derive concrete results, it is necessary to make further assumptions about the particular form of the price process $P(t)$. We follow the work [2], [6] and others and assume that it follows the Ornstein-Uhlenbeck equation

$$dP(t) = \tau(\mu - P(t)) + \sigma dW(t),$$

with parameters μ (mean value), $\tau > 0$ (speed of mean reversion) and $\sigma > 0$ (volatility), where $W(t)$ stands for the standard Wiener process. In the sequel, if not stated otherwise, all unreferenced statements follow [2] and [6].

Here we make an essential assumption: *all of the parameters η, μ, τ, σ are assumed to be known exactly*, meaning that they are *not* econometric estimates. This is an idealistic assumption, which is however frequent in portfolio theory. (In practice, the parameters are estimated from finite-sample observations of the price processes $P_1(t), \dots, P_n(t)$ which means that they suffer from statistical errors; this fact is disregarded in this text but is surely worth investigating in depth).

Then we can use Itô's Lemma and consider a *standardized version* \tilde{P} of the Ornstein-Uhlenbeck process given by substitution

$$\tilde{t} = \tau t, \quad \tilde{P}(\tilde{t}) = (2\tau)^{1/2} \sigma^{-1} (P(t) - \mu).$$

Now we can assume, without loss of generality, that $\tau = 1$, $\mu = 0$ and $\sigma = \sqrt{2}$; this is equivalent to saying that the price process $P(t)$ is standardized. In this setup, the profit $\Pi(a, b)$ is indeed a function of a, b only (and not μ, σ, τ). And the task reduces to the optimization problem "how to choose a, b to maximize $\Pi(a, b)$ "?

3 Bertram's optimization problem

3.1 The case with no risk constraints

The optimization problem from the last paragraph,

$$\max_{a,b} \Pi(a, b) \text{ subject to } b - a - c \geq 0, \quad (1)$$

is referred to as *Bertram's (unconstrained) problem*. It has the following properties:

- (a) Optimization problem (1) is convex.
- (b) Its optimum is *symmetric*, meaning that its optimal solution (a^*, b^*) satisfies $b^* = -a^*$.
- (c) The function $\Pi(a, b)$ is not elementary; the optimum (a^*, b^*) cannot be expected to have a closed algebraic form. In particular, the "best" known expression is (using $\tilde{k} := 2k - 1$ as a shorthand)

$$\Pi(a, b) = \frac{\pi}{\mathcal{T}}, \quad \mathcal{T} = \sum_{k=1}^{\infty} 2^{\tilde{k}/2} \Gamma(\tilde{k}) \frac{a^{\tilde{k}} - b^{\tilde{k}}}{\tilde{k}!}. \quad (2)$$

3.2 The case with bounded risk

It is natural to augment (1) with risk constraints. If $R(a, b)$ is a risk measure, it is assumed that an exogenous admissible risk level R^{\max} is given and the risk-constrained problem takes the form

$$\max_{a,b} \Pi(a, b) \text{ subject to } b - a - c \geq 0, \quad R(a, b) \leq R^{\max}. \quad (3)$$

First of all, it is natural to consider strategies with long expected trading cycles to be "more adverse" than strategies with shorter cycles. Thus, it is natural to consider a bound on \mathcal{T} given by expression (2):

$$R_1 := \mathcal{T}.$$

As a second alternative, a natural choice is the volatility of profit $\text{var}(\pi N_t)$ in the long run $t \rightarrow \infty$. Unfortunately,

$$\lim_{t \rightarrow \infty} \text{var}(\pi N_t) = 0,$$

and thus it is a trivial measure. Thus we need standardization in the form of *per-time-unit volatility of profit* defined as

$$R_2 \equiv R_2(a, b) := \lim_{t \rightarrow \infty} t^{-1/2} \text{var}(\pi N_t)^{1/2}. \quad (4)$$

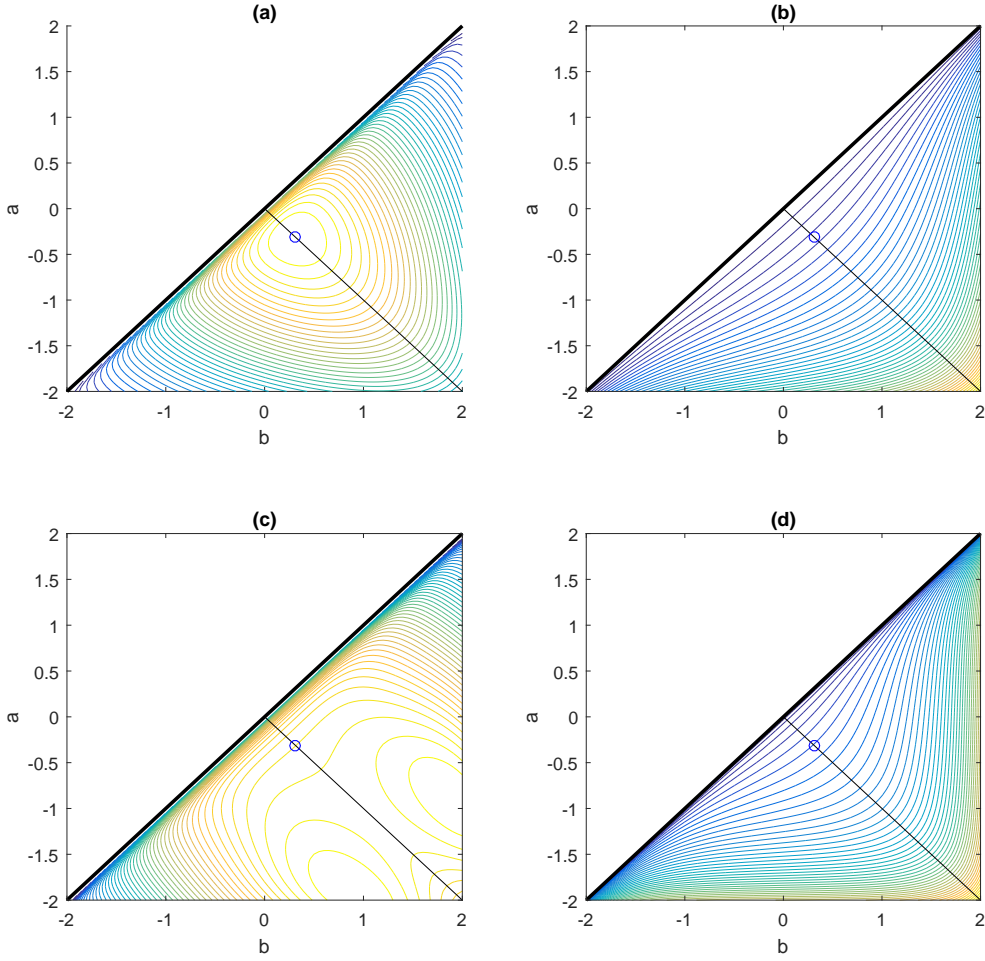


Figure 1 The space (a, b) of feasible strategies satisfying $b - a - c \geq 0$ with fixed costs $c = 0.02$, the symmetry line $a = -b$ and the optimal strategy for the unconstrained problem (1) depicted by a blue circle. Figure (a): Contour plot of the profit function $\Pi(a, b)$. Figure (b): Contour plot of risk measure $R_1(a, b)$, corresponding to the expected length of a trade cycle. Figure (c): Contour plot of risk measure $R_2(a, b)$, corresponding to the per-time-unit volatility of profit. Figure (d): Contour plot of risk measure $R_3(a, b)$, corresponding to the volatility of the length of a trade cycle.

It is easy to show that the normalization of the variance by $t^{-\omega}$ in (4) results in a nontrivial risk measure only for $\omega = -1/2$. The available expressions for R_2 are non-elementary again: we have

$$R_2 = \pi \mathcal{V} \mathcal{T}^{-3/2},$$

where, denoting the digamma function by $\varphi(\cdot)$,

$$\begin{aligned} \mathcal{V} &:= \text{var}(T_i)^{1/2} = (w_1(a) - w_1(b) - w_2(a) + w_2(b))^{1/2}, \\ w_1(x) &= \frac{1}{4} \left(\sum_{k=1}^{\infty} \Gamma(k/2) \frac{2^{k/2} x^k}{k!} \right)^2 - \frac{1}{4} \left(\sum_{k=1}^{\infty} (-1)^k \Gamma(k/2) \frac{2^{k/2} x^k}{k!} \right)^2, \\ w_2(x) &= \sum_{k=1}^{\infty} \Gamma(k - 1/2) \varphi(k - 1/2) \frac{2^{k-1/2} x^{2k-1}}{(2k-1)!}. \end{aligned}$$

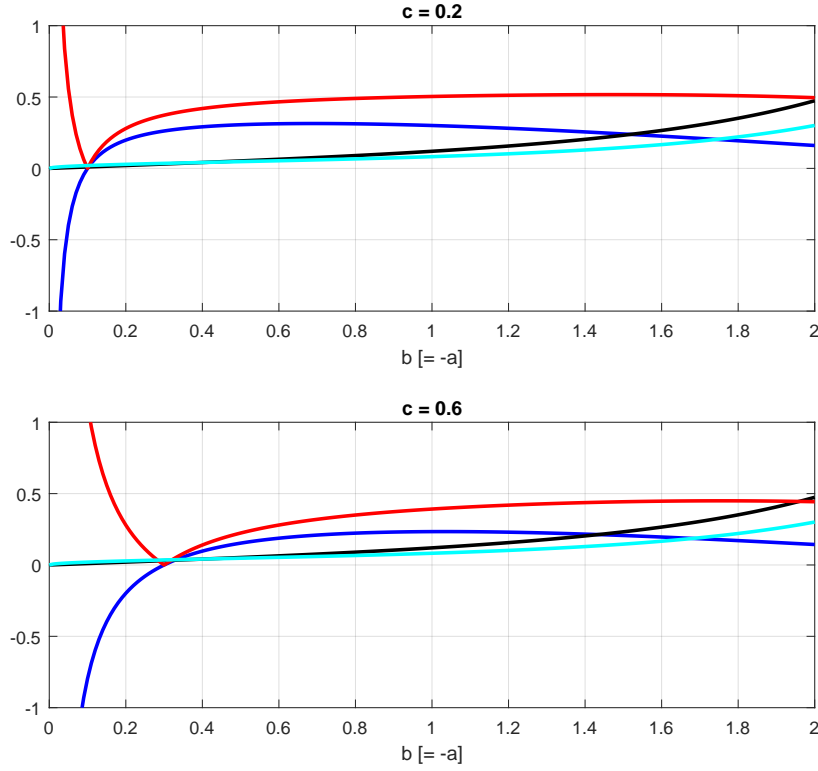


Figure 2 Symmetric strategies satisfying $b = -a$ for fixed costs $c = 0.2$ (upper plot) and $c = 0.6$ (lower plot). The profit function $\Pi(-b, b)$ in blue and the three risk measures: $\frac{1}{50}R_1(-b, b)$ in black, $R_2(-b, b)$ in red, $\frac{1}{50}R_3(-b, b)$ in cyan. The scaling factor $\frac{1}{50}$ is introduced just for better readability of the plot.

With the available expression for \mathcal{V} , the volatility of the length of a trade cycle, it is also natural to put a bound on it:

$$R_3 := \mathcal{V}.$$

Also further risk measures can be constructed analogously, such as \mathcal{T}/\mathcal{V} , the variation coefficient of the length of a trade cycle.

On the other hand, currently it seems implausible to derive expressions for quantiles (value-at-risk) and expected shortfall (conditional value-at-risk) of profit as representatives of other frequently considered risk measures from literature. This is an interesting problem for further analysis.

3.3 Geometry of the risk constraints

It is instructive to visualize the objective function $\Pi(a, b)$ and the risk measures R_1, R_2, R_3 as in Figure 1 by contour plots. It is visible that in all three cases of risk measures, the constraint $R_i \leq R^{\max}$ is non-convex. It is also apparent that the risk measures are symmetric along the line $a = -b$.

Figure 2 depicts the same functions for symmetric strategies, where $b \in [0, 2]$ and $b = -a$. The figure illustrates the effect of transaction costs (with $c = 0.2$ in the upper plot and $c = 0.6$ in the lower plot). Observe that R_1 and R_3 do not depend on c , while R_2 does.

Solving the risk-constrained problem (3) for symmetric strategies can be essentially based on the Binary Search algorithm from [6]. Since it is a problem in a single variable, there are no serious numerical issues. The only point why the problem has to be solved numerically is that the optimal solution seems to have no explicit algebraic expression (most probably it is not an elementary function of the input data).

4 Conclusions

Cointegrated trading strategies take the advantage of the fact that it is possible to combine a portfolio of assets the price process of which is stationary. Here, the mean reversion feature can be utilized for an arbitrage strategy (“wait for a below-mean price, open positions, then wait for an above-mean price and close positions”). The risk is introduced here through the random length of a trade cycle. If combined with per-unit-of-asset transaction costs and a portfolio consisting of a pair of assets, the strategy reduces to Bertram’s pairs trading strategy. We formulated this strategy in more general way, with a general portfolio of assets (affecting the transaction costs), and with multiple kinds of risk constraints.

However, still, the analysis is performed under the unrealistic assumption that all parameters of the model — the cointegration coefficients and the parameters of the Ornstein-Uhlenbeck process — are known exactly. This is violated in practice where we have a (large) family of tradable assets and we need (i) to test whether a cointegration combination exists, (ii) to estimate the cointegration coefficients, (iii) to test whether the price path can be modeled as Ornstein-Uhlenbeck process, (iv) to estimate its parameters. All of the tests and estimates are made from a finite sample of observations of points on the trajectory of the price processes P_1, \dots, P_n and suffer from statistical errors. If a portfolio manager works with the “wrong” (statistically estimated) values instead of the true parameters, this leads to a suboptimal strategy in the sense of the optimization problem (3). It can be intuitively expected that small errors in parameter estimates will lead to small errors in the trading strategy, measured e.g. by the decrease in per-time-unit profit compared to the profit of a theoretical investor who would know the true parameters exactly. Quantification of this loss is one of interesting problems for further research.

Acknowledgements

The work was supported by the Czech Science Foundation under project 19-02773S.

References

- [1] Avellaneda, M. and Lee, J. H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance* **10**(7), 761–782, doi:10.1080/14697680903124632.
- [2] Bertram W. K. (2009). Optimal trading strategies for Itô diffusion processes. *Physica A: Statistical Mechanics and Its Applications* **388**(14), 2865–2873, doi:10.1016/j.physa.2009.04.004.
- [3] Bertram W. K. (2010). Analytic solutions for optimal statistical arbitrage trading. *Physica A: Statistical Mechanics and its Applications* **389**(11), 2234–2243, doi:10.1016/j.physa.2010.01.045.
- [4] Elliott, R. J., Van Der Hoek J. and Malcolm, W. P. (2005). Pairs trading. *Quantitative Finance* **5**(3), 271–276, doi:10.1080/14697680500149370.
- [5] Gatev E., Goetzmann W. N. and Rouwenhorst K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies* **19**(3), 797–827, doi:10.1093/rfs/hhj020.
- [6] Holý V. and Černý M. (2021). Bertram’s pairs trading strategy with bounded risk. Submitted in: *Central European Journal of Operational Research*. Preprint (arXiv): <https://arxiv.org/abs/2102.04160>.
- [7] Huck, N. (2009). Pairs selection and outranking: An application to the S&P 100 index. *European Journal of Operational Research* **196**(2), 819–825, doi:10.1016/j.ejor.2008.03.025.
- [8] Liew, R. Q. and Wu, Y. (2013). Pairs trading: A copula approach. *Journal of Derivatives & Hedge Funds* **19**(1), 12–30, doi:10.1057/jdhf.2013.1.
- [9] Jurek, J. W. and Yang, H. (2007). Dynamic portfolio selection in arbitrage. *EFA 2006 Meeting Paper*, doi:10.2139/ssrn.882536.
- [10] Krauss, C. (2017). Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys* **31**(2), 513–545, doi:10.1111/joes.12153.
- [11] Vidyamurthy, G. (2004) Pairs Trading: Quantitative Methods and Analysis. 1st edition, Wiley, Hoboken, <https://www.wiley.com/en-us/Pairs+Trading+%3A+Quantitative+Methods+and+Analysis-p-9780471460671>

Statistical Analysis of ICT Utilization in Marketing

Andrea Čížku¹

Abstract. The paper investigates and quantifies the utilization of online marketing tools by Czech companies during the current Covid-19 crisis, which stressed the importance of exploitation of online tools in business. Companies are constantly seeking out ways of enhancing their business and online marketing tools offer a way of achieving these goals. A wide range of online marketing tools are taken into account in the statistical analysis performed in this paper. A survey approach based on questionnaires was applied to collect the data. Statistical analysis is performed by applying chi-square tests within the contingency table framework. An important finding is that there are important differences in the utilization of online marketing tools by small craft businesses and other companies. Dynamical analysis of changes in trends in the utilization of these tools are also investigated and analyzed in a disaggregated way to describe the dynamic trends for different online marketing tools.

Keywords: online marketing, small craft businesses, covid-19 crisis, social networks, ICT, e-business

JEL Classification: M15, M31, M37

AMS Classification: 91B82

1 Introduction

According to Kingsnorth (2016), online marketing tools represent an environment in which marketing strategies are combined with the application of new technologies. Eid, El-Gohary (2013) states that companies have already integrated these new technologies during the first two decades of the 21st century and online marketing tools are already a part of communication strategies with customers. Online marketing tools and their utilization is not only a relevant practical issue, but it is also a topic discussed thoroughly in the marketing literature nowadays. Cetlová, Velimov (2019) performed a pilot project to investigate the use of online marketing activities in Czech companies. Cetlová, Marciník, Velimov (2020) performed a statistical analysis of the relationship between the use of online marketing tools and the size of companies and provided empirical evidence that ICT tools are an essential part of online marketing in large companies while the usage of these tools in small self-employed businesses is rather weak. Civelek et al. (2020) performed a detailed statistical analysis of 1156 Czech, Slovakian and Hungarian small and medium-sized enterprises (SME). Their findings are that the Hungarian SMEs apply social media platforms in their operations more than Czech and Slovakian SMEs. Breckova, Karas (2020) found out that there is a dependence between the use of online marketing tools and the export focus of companies. Companies operating on foreign markets use a wider range of online tools for their business than companies focused on the domestic market.

The goal of this paper is to contribute to this literature by (1) analyzing a wide range of marketing tools in a disaggregated way - utilization of mobile applications, e-shop applications, built-in contact forms, chat-boxes, Google Analytics, SEO (Search Engine Optimization), SEM (Search Engine Marketing), CRM (Customer Relationship Management), social networks, PPC (Pay Per Click), (2) investigating the differences in utilizing this wide range of online marketing tools between small craft businesses and other companies, (3) describing dynamic trends in utilization of these tools during the current Covid-19 crisis.

The structure of the paper is as follows. Firstly, a range of analyzed online marketing tools is summarized in chapter 2. Comparison of small businesses with other companies is then presented in chapter 3. Section 4 presents an analysis of dynamic trends in the utilization of online marketing tools during the current Covid-19 crisis. The final chapter 5 concludes.

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.

2 Online Marketing Tools

Online marketing comprises many tools and methods. For this reason, the first questionnaire from January 2020 contained the following questions:

- 1) Do you have an idea of what online marketing is?
- 2) Do you have a web page for your business?
- 3) Is there a mobile application on your web page?
- 4) Does your web page contain a mobile e-shop application?
- 5) Does your web page have a built-in contact form for communication with customers?
- 6) Is there a chat box on your web page for communication with visitors?
- 7) Do you monitor the visit rate of your business web page?
- 8) Do you exploit Google Analytics or a similar tool?
- 9) Do you utilize SEO for web optimization?
- 10) Do you exploit SEM (Search Engine Marketing)?
- 11) Do you have a special e-mailing tool for communication with (potential) customers?
- 12) Do you utilize CRM (Customer Relationship Management)?
- 13) Are you presenting your business on social networks?
- 14) Do you measure the effectiveness of marketing activities on social networks?
- 15) Have you advertised your products and services on the internet by using PPC advertising (Pay Per Click)?

The same questions were asked once more later in September 2020 to obtain information on how the utilization of online marketing tools changes in time during the current Covid-19 crisis. Moreover, the following additional questions were added to the questionnaire in September 2020 to obtain information specific to the Covid-19 crisis:

- 16) Do you plan to increase the budget for online marketing activities in the next three months?
- 17) Do you consider online marketing tools to have a positive influence on the reputation of your business?
- 18) Does the utilization of online marketing tools bring savings compared to the traditional marketing communication?
- 19) Have you begun to use other online technologies for communication with customers, suppliers or employees during the current Covid-19 crisis?
- 20) Do you plan to utilize these online technologies for communication in the near future?

Data were collected from both these questionnaires by Cetlová et al. (2020) within the project implemented by the Central Bohemian Association of Managers and Entrepreneurs (STAMP). Cetlová et al. (2020) also analyzed these data. Nonetheless, their analysis was based only on simple descriptive statistics. The presented paper extends their analysis by performing a rigorous statistical analysis based on chi-square tests within the framework of contingency tables. Results obtained from this analysis are described in the following chapters 3 and 4.

3 Comparison of small businesses with other companies

Utilization of online marketing tools is compared between small self-employed businesses and other companies by testing the dependency between two categorical variables in a contingency table. Specifically, the following hypotheses were statistically tested by chi-square test within contingency tables:

H_0 : In the population, the two categorical variables are independent.

H_A : In the population, the two categorical variables are dependent.

Contingency tables are tools for analyzing the relationship between two categorical variables. It is a special type of frequency distribution table, where two variables are shown simultaneously. The following contingency table with absolute frequencies was obtained from the raw data from January 2020 to answer the first above-mentioned question: “Do you have an idea of what online marketing is?”

Observed frequencies	Do you have an idea of what online marketing is?		Total sum
	No	Yes	
Small craft businesses	11	22	33
Other enterprises	14	117	131
Total sum	25	139	164

Table 1 Contingency table with absolute frequencies for categorical variables *type of business* and *awareness of online marketing* for data from January 2020.

Table 1 shows that there are 33 small craft businesses and 11 of them do not have an idea of what online marketing is. There are also 131 other enterprises and only 14 of them do not have an idea of what online marketing is. Therefore, there are 11+25=25 enterprises that do not have an idea of what online marketing is. The remaining numbers in the contingency table are interpreted in a similar way.

The following hypotheses are tested by chi-square test in the contingency table:

H_0 : There is no relationship between the type of business and awareness of online marketing.

H_A : There is a relationship between these two categorical variables.

Chi-square statistic is calculated as follows:

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^n \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right], \quad (1)$$

where O_{ij} is the observed frequency in the i -th row and j -th column of a contingency table,

E_{ij} is the expected frequency in the i -th row and j -th column of a contingency table,

m is the number of rows and n represents the number of columns in a contingency table.

Expected frequencies are in contingency tables calculated according to the formula:

$$E_{ij} = \frac{O_{i\cdot} \cdot O_{\cdot j}}{N}, \quad (2)$$

where $O_{i\cdot} = \sum_{j=1}^n O_{ij}$ is the sum of observed frequencies in the i -th row,

$O_{\cdot j} = \sum_{i=1}^m O_{ij}$ is the sum of observed frequencies in the j -th column,

$N = \sum_{i=1}^m \sum_{j=1}^n O_{ij}$ is the total number of observations.

If the null hypothesis H_0 is true, then the chi-square statistic χ^2 has a chi-square distribution with $k = (m-1) \cdot (n-1)$ degrees of freedom, i.e. $\chi^2 \sim \chi^2(k)$. P-value of the chi-square test statistic calculated from table 1 was found to be 0.001, meaning that the hypothesis H_0 is refused on 1% level of significance ($0.001 < 0.01$). Therefore, there is a relationship between the type of business and the awareness of online marketing. Obviously, small craft businesses are much less aware of online marketing than other enterprises. The ratio of craftsmen not aware of online marketing is $11/33 = 0.3$ while the ratio of other enterprises not aware of online marketing is much smaller and equal to $14/131 = 0.11$.

Described statistical procedure was performed for all remaining questions and for the questions analyzed later in September 2020. The results are summarized in the following table:²

Question number	1	2	3	4	5	6	7	8	9	10
P-value (January 2020)	0.001 ***	0.031 **	0.109	0.078 *	0.003 ***	0.195	0.047 **	0.066 *	0.025 **	0.012 **
P-value (September 2020)	0.000 ***	0.000 ***	0.004 **	0.283	0.000 ***	0.030 **	0.001 ***	0.002 ***	0.152	0.206

Question number	11	12	13	14	15	16	17	18	19	20
P-value (January 2020)	0.013 **	0.006 ***	0.029 **	0.013 **	0.167					
P-value (September 2020)	0.000 ***	0.001 ***	0.000 ***	0.003 ***	0.071 *	0.003 ***	0.000 ***	0.000 ***	0.083 *	0.000 ***

Table 2 Summary of the results of the chi-square test of independence between categorical variable *type of business* and other categorical variables defined by the questions 1-20 for the data from January 2020 as well as September 2020.

² Symbols ***, ** and * indicate that the result of the statistical test is significant on 1%, 5% and 10% level of significance, respectively.

Legend: (short description of questions 1-20)

- | | |
|----------------------------------|--|
| 1. awareness of online marketing | 11. emailing tool |
| 2. web pages | 12. CRM |
| 3. mobile application | 13. social networks |
| 4. e-shop | 14. effectiveness on social networks |
| 5. contact form | 15. PPC |
| 6. chat box | 16. budget for online marketing |
| 7. visit rate | 17. reputation |
| 8. Google Analytics | 18. savings |
| 9. SEO | 19. new online technologies |
| 10. SEM | 20. new online technologies in the near future |

The results indicate a statistically important relationship between the type of business and the utilization of a given online marketing tool (specified in questions 1-20) in most cases. This provides empirical evidence for the validity of the hypothesis that the utilization of online marketing tools is different in small craft self-employed businesses than in other enterprises. There are few exceptions in which statistical dependency was not confirmed at none of the standard statistical levels of significance. These are questions numbers 3, 6 and 15 for data from January 2020 and questions 4, 9 and 10 for data from January 2021. The utilization of a given marketing tool by small self-employed craft businesses and by other enterprises can be considered the same in these cases.

Let's now investigate the ratio of small businesses and other enterprises with a positive answer "yes" to a given question, which is summarized by the following figure:

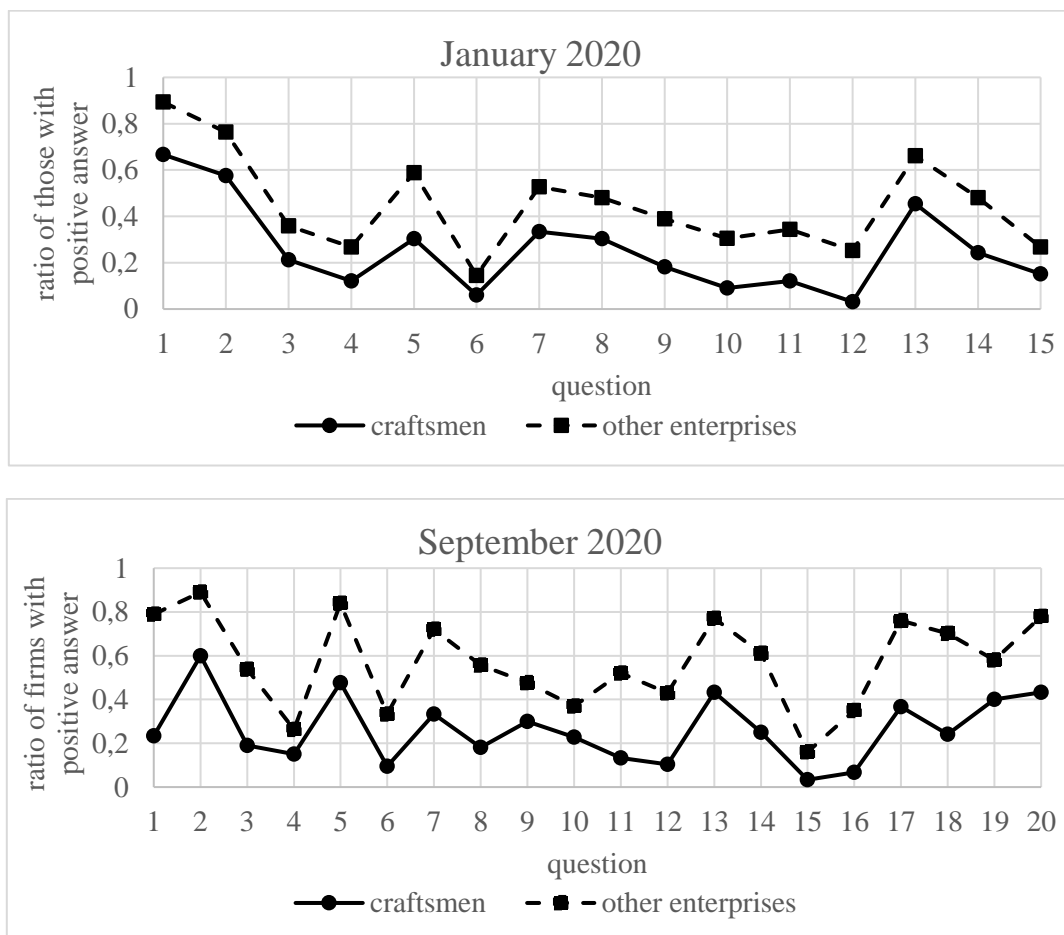


Figure 1 Ratio of small craft businesses and other enterprises with a positive answer to questions 1-20 for the data from January 2020 and September 2020.

The figure documents that the ratio of small businesses with a positive answer to the given questions is lower than the corresponding ratio of other enterprises. This result is quite robust as it holds for all questions and for both

datasets, which represents another empirical evidence for the hypothesis that the degree of utilization of online marketing tools in small self-employed craft businesses is significantly smaller than in other enterprises. It can also be seen from Figure 1 that there is a relatively high ratio of positive answers in small craft businesses as well as other enterprises in the following questions:

- 1. awareness of online marketing
- 2. web pages
- 5. contact form
- 7. visit rate
- 13. social networks
- 17. reputation
- 19. new online technologies
- 20. new online technologies in the near future

Thus, the most popular online marketing tools are web pages, contact form and social networks. This observation suggests that there is a potential for promoting the utilization of other online marketing tools between Czech companies (mobile applications, e-shop, chat box, Google Analytics, SEO, SEM, special emailing tools, CRM and PPC). Quite a relatively high percentage of companies monitor the visit rate of their web pages, consider online marketing tools as a way to promote their reputation and begun to use new online technologies and plan to do so in the near future as well. Nonetheless, a rather low percentage of companies are willing to increase their budget for online marketing activities during the current Covid-19 crisis (question number 16).

4 Dynamics trends in utilization of online marketing tools

Let’s now investigate the dynamics in time. The ratio of firms with a positive answer to the above mentioned questions is now in figure 2 displayed in such a way that enables us to visually analyze changes in time.

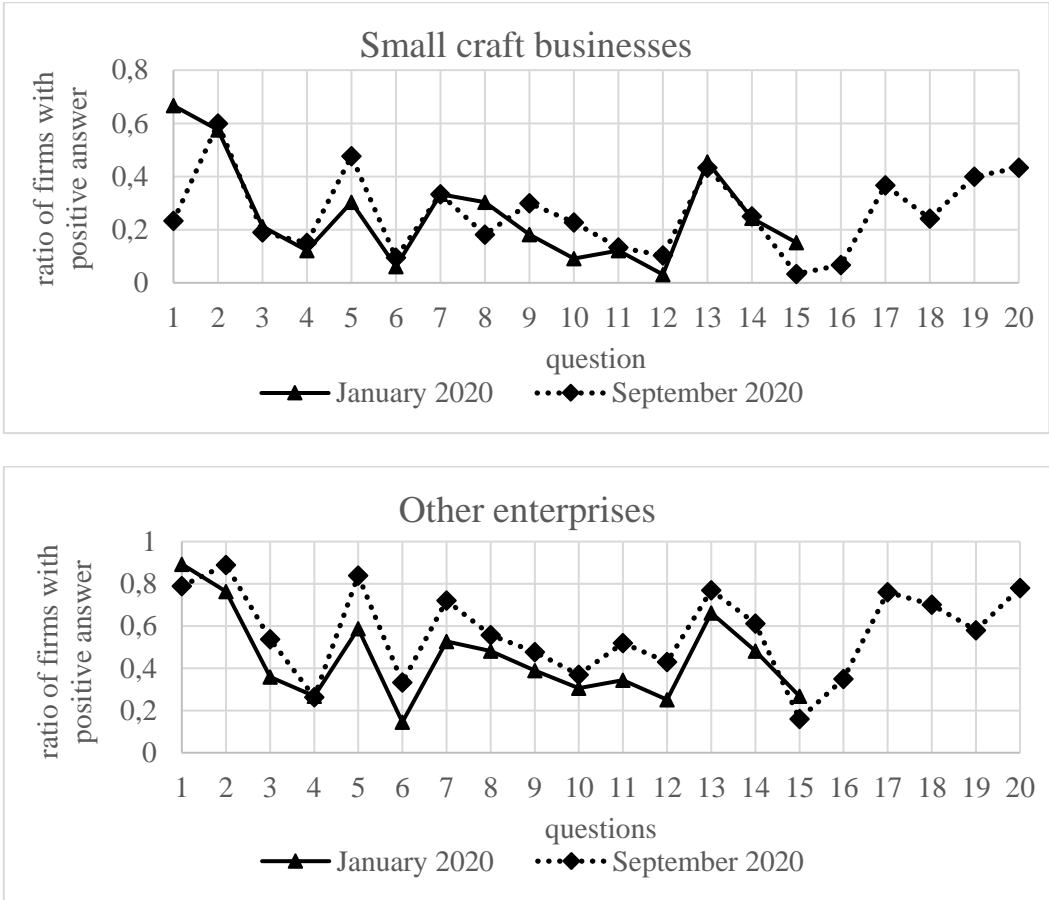


Figure 2 Dynamics of the ratio of small craft self-employed businesses and other enterprises with a positive answer to questions 1-20.

The first graph of figure 2 shows that the percentage of small craft businesses with a positive answer to above-mentioned questions is roughly the same in January as in September 2020. From this point of view, the utilization of online marketing tools among small craft businesses did not change much during the current Covid-19 crisis. Nonetheless, the exploitation of online marketing tools did change among other enterprises, and the change had a positive sign for most of the questions. Therefore, the differences in the utilization of online marketing tools between small craft self-employed businesses and other enterprises increased during the current Covid-19 crisis.

5 Conclusion

McKinsey (2020) states that the decline in the global economy due to Covid-19 measures has already overcome the Great Recession of 2009. McKinsey conducted a research showing that only businesses using new technologies addressing the changing environment will stay competitive and will be able to adapt. Their results suggest that the only way out of the global economic crisis is to accelerate the use of new technologies including online marketing tools. Results from this paper show that this is not true for the segment of the Czech small craft businesses. Utilization of online marketing tools by small self-employed companies in the Czech Republic is very low and the current Covid-19 crisis did not change this trend. Utilization of online marketing tools by other Czech companies was considerably higher at the beginning of this crisis and there is also a growing trend in the exploitation of these online tools by larger companies. Thus, the current Covid-19 crisis has even increased the difference between small craft businesses and other enterprises regarding their utilization of online marketing tools. The application of e-business within firms not only improves business processes, administration, sales, financial management, human resources, and service quality, but also an exchange in communication between companies, customers, suppliers, banks, and public administration (Cetlová, Velimov, 2019). Nonetheless, small self-employed craft businesses in the Czech Republic have not yet overcome the technological barriers associated with the application of these tools despite the current Covid-19 crisis which emphasized the importance of these online tools.

References

- [1] Breckova, P. Karas, M. (2020). Online Technology and Promotion Tools in SMEs. *Innovative Marketing*, 16(3), 85-97. [http://dx.doi.org/10.21511/im.16\(3\).2020.08](http://dx.doi.org/10.21511/im.16(3).2020.08)
- [2] Cetlová, H., Velimov, E. (2019). Online Marketing Activities and Marketing Communication Tools in Czech Small and Medium-Sized Enterprises. *Marketing Identity*, 7(1), 803-815.
- [3] Cetlová, H., Marciník, R., Velimov, E. (2020). Research Into Online Marketing Activities and Tools for Measuring their Efficiency by Entrepreneurs in the Czech Republic. *Socioekonomické a humanitní studie*, 11(1), 41-54.
- [4] Civelek, M., Gajdka, K., Světlík, J., Vavrečka, V. (2020). Differences in the Usage of Online marketing and Social Media Tools: Evidence from Czech, Slovakian and Hungarian SMEs. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 15(3), 537-563. <https://doi.org/10.24136/eq.2020.024>
- [5] Eid, R., El-Gohary, H. (2013). The Impact of E-Marketing Used on Small Business Enterprises' marketing success. *The Service Industries Journal*, 33(1), 31-50. <https://doi.org/10.1080/02642069.2011.594878>
- [6] Kingsnorth, S. (2016). *Digital Marketing Strategy: An Integrated Approach to Online Marketing*. Kogan Page Publishers. ISBN 978-0-7494-7470-6
- [7] McKinsey COVID-19: Implications for business. May 7 2021. Executive Briefing. [online]. [cit. 2021-07-05]. Available at: <https://www.mckinsey.com/business-functions/risk/our-insights/covid-19-implications-for-business>

On the crossing numbers of join of one graph on six vertices with path using cyclic permutation

Emília Draženská¹

Abstract.

The crossing number, $cr(H)$, of a simple graph H is the minimal number of edge crossings over all good drawings of H in the plane. In general, compute the crossing number for a given graph is a very difficult problem. The crossing numbers of a few families of graphs are known. One of them are join products of special graphs. In the paper, we extend known results concerning crossing number of the join product $G + P_n$, where the graph G consists of 5-cycle whose one vertex is adjacent to another vertex of this cycle and one other isolated vertex, and P_n is the path on n vertices. The methods used in the paper are based on combinatorial properties of cyclic permutations and the proof is done with the help of software.

Keywords: graphs, drawings, crossing numbers, cyclic permutation, join product.

JEL classification: C02

AMS classification: 05C10; 05C38

1 Introduction

Let the graph H be a simple, undirected and connected graph with vertex set V and edge set E . The crossing number, $cr(H)$, of a graph H is the minimum number of edge crossings in any drawing of H in the plane. (For the definition of a *drawing*, see [7].) The drawing with a minimum number of crossings (an *optimal* drawing) must be a *good drawing*, meaning that each two edges have at most one point in common, which is either a common end-vertex or a crossing.

The problem of reducing the number of crossings on the edges in the drawings of graphs was studied in many areas and the most important area is VLSI technology. The crossing numbers have been studied to improve the readability of hierarchical structures and automated graph drawings. The visualized graph should be easy to read and understand. For the understandability of graph drawings, the reducing of crossings is likely the most important. The investigation on the crossing number of a given graph is very difficult problem. Garey and Johnson proved [4] that computing $cr(H)$ is an NP-complete problem.

The exact values of crossing numbers are known for several special classes of graphs. One of them is a join products of two graphs. The *join product* $H_1 + H_2$ of two graphs $H_1 = (V_1, E_1)$ and $H_2 = (V_2, E_2)$ is obtained from the vertex-disjoint copies of H_1 and H_2 by adding all edges between $V(H_1)$ and $V(H_2)$. For $|V(G_1)| = m$ and $|V(G_2)| = n$, the edge set of $H_1 + H_2$ is the union of disjoint edge sets of the graphs H_1 , H_2 , and the complete bipartite graph $K_{m,n}$. Let D_n denote the *discrete graph* on n vertices, let P_n and C_n be the path and the cycle on n vertices. In the proofs of the paper, we will often use the term “region” also in nonplanar drawings. In this case, crossings are considered to be vertices of the “map”.

The exact values for crossing numbers of $H + P_n$ and $H + C_n$ for all graphs H of order at most four are given in [10], and the crossing numbers of the graphs $H + D_n$, $H + P_n$, and $H + C_n$ are also known for some graphs H of order five and six, see [2], [3], [7], [8], [9], [11], [12], [13], [14], and [15].

In this paper we extend these results by giving the exact values of the crossing numbers for join product for a special graph G on six vertices with the path P_n .

¹Technical University in Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Némcovcej 32, 042 00 Košice, Slovak Republic, e-mail: emilia.drazenska@tuke.sk

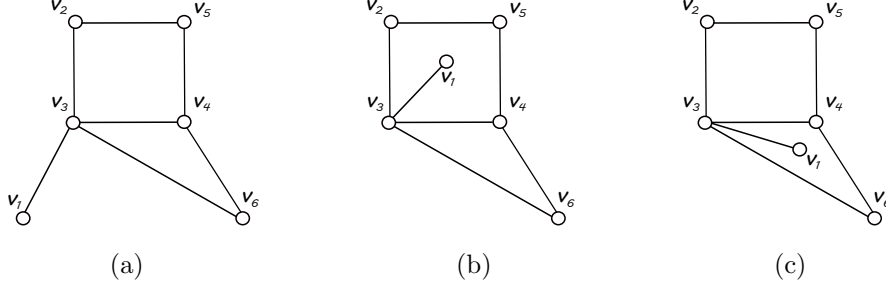


Figure 1: Three planar drawings of the graph G and the vertex notation of G

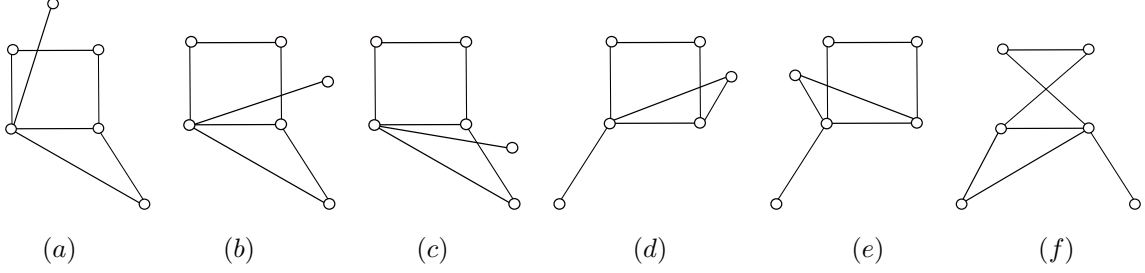


Figure 2: Six drawings of the graph G with one crossing

D ($D(H)$) be a good drawing of the graph H . We denote by $\text{cr}_D(H)$ the number of crossings among edges of H in the drawing D . Let H_i and H_j be two edge-disjoint subgraphs of H . We denote by $\text{cr}_D(H_i, H_j)$ the number of crossings between the edges of H_i and edges of H_j , and the number of crossings among edges of H_i in D by $\text{cr}_D(H_i)$.

In the paper, some proofs will be also based on the Kleitman's result on crossing numbers of the complete bipartite graphs [6]. More precisely, he proved that

$$\text{cr}(K_{m,n}) = \left\lfloor \frac{m}{2} \right\rfloor \left\lfloor \frac{m-1}{2} \right\rfloor \left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor, \quad \text{if } \min\{m, n\} \leq 6. \quad (1)$$

2 The crossing number of $G + P_n$

Let G be the connected graph of order six consisting of the four-cycle and three-cycle with a common edge and one more vertex adjacent with a vertex of degree three (see Figure 1).

We consider the join product of the graph G with the discrete graph D_n . The graph $G + D_n$ consists of one copy of G and of n vertices t_1, \dots, t_n . Each vertex t_i , $i = 1, \dots, n$, is adjacent to every vertex of G . Let T^i , $i = 1, \dots, n$, denote the subgraph that is uniquely induced by the six edges incident with the fixed vertex t_i . This means that the graph $T^1 \cup \dots \cup T^n$ is isomorphic to the complete bipartite graph $K_{6,n}$ and

$$G + D_n = G \cup K_{6,n} = G \cup \left(\bigcup_{i=1}^n T^i \right). \quad (2)$$

We also use the same definition and notation for the good drawing D of the graph $G + P_n$. The graph $G + P_n$ contains $G + D_n$ as a subgraph, and therefore let P_n^* denote the path induced on vertices of $G + P_n$ not belonging to the subgraph G . The path P_n^* consists of the vertices t_1, \dots, t_n and the edges $\{t_i, t_{i+1}\}$, for $i = 1, \dots, n-1$. Thus,

$$G + P_n = G \cup K_{6,n} \cup P_n^* = G \cup \left(\bigcup_{i=1}^n T^i \right) \cup P_n^*. \quad (3)$$

Let D be a good drawing of the graph $G + P_n$. The *rotation* $\text{rot}_D(t_i)$ of a vertex t_i in the drawing D is the cyclic permutation that records the (cyclic) counter-clockwise order in which the edges leave t_i , see [5]. We use the notation (123456) if the counter-clockwise order the edges incident with the vertex t_i is $t_iv_1, t_iv_2, t_iv_3, t_iv_4, t_iv_5$, and t_iv_6 . For $i, j \in \{1, \dots, n\}$, $i \neq j$, every subgraph $T^i \cup T^j$ of the graph $G + P_n$ is isomorphic with the graph $K_{6,2}$. We will study the minimum number of crossings between the edges of T^i and the edges of T^j in a subgraph $T^i \cup T^j$ induced in the drawing D of $G + P_n$ depending on the rotations $\text{rot}_D(t_i)$ and $\text{rot}_D(t_j)$.

D.R.Woodall [16] published that in the subdrawing of $T^i \cup T^j$ induced by D holds $\text{cr}_D(T^i, T^j) \geq 6$ if $\text{rot}_D(t_i) = \text{rot}_D(t_j)$. If $Q(\text{rot}_D(t_i), \text{rot}_D(t_j))$ denotes the minimum number of interchanges of adjacent elements of $\text{rot}_D(t_i)$ required to produce the inverse cyclic permutation of $\text{rot}_D(t_j)$, then $\text{cr}_D(T^i, T^j) \geq Q(\text{rot}_D(t_i), \text{rot}_D(t_j))$. By P we will understand the inverse cyclic permutation to the cyclic permutation P . In this paper, some parts of proofs can be done with the help of software that generates all cyclic permutations of six-element set in [1].

We will separate the subgraphs T^i for $i \in \{1, \dots, n\}$ of the graph $G + D_n$ into three mutually disjoint subsets. For $i \in \{1, \dots, n\}$, let $R_D = \{T^i : \text{cr}_D(G, T^i) = 0\}$ and $S_D = \{T^i : \text{cr}_D(G, T^i) = 1\}$. Every other subgraph T^i crosses the edges of G at least twice in D . For $T^i \in R_D$, let F^i denotes the subgraph $G \cup T^i$ of $G + P_n$, and let $D(F^i)$ be the subdrawing induced by D .

According to the arguments in the proof of Theorem 1, in the optimal drawing D of $G + P_n$, there is a subgraph T^i whose edges cross the edges of the graph G at most once. Thus, we will deal with only drawings of G with a possibility of existence of a subgraph $T^i \in R_D \cup S_D$.

Assume first a good subdrawing of the graph G in which there is no crossing on the edges of G . In this case, we obtain three nonisomorphic planar drawings of G shown in Figure 1. The vertex notation of the graph G in Figure 1 will be justified later.

Let us first assume the drawing of G with the corresponding vertex notation in such a way as shown in Figure 1(a). We need to list all possible rotations $\text{rot}_D(t_i)$ which can appear in D if the edges of T^i do not cross the edges of G . Since there is only one subdrawing of $F^i \setminus \{v_2, v_3\}$ represented by the rotation (1546) , there are two ways of obtaining the subdrawing of F^i depending on the region in which the edges t_iv_2 and t_iv_3 are placed. We denote these two possibilities under our consideration by $\mathcal{A}_1^*, \mathcal{A}_2^*$, summarized in Table 1. As for our considerations it does not play a role in which of the regions is unbounded, assume the drawings shown in Figure 3. Let us discuss all possible rotations $\text{rot}_D(t_i)$ which can appear in D if the edges of T^i cross the edges of G exactly once. The vertex t_i must be placed in the region with at least five vertices of G on its boundary, which means that the vertex t_i be placed in outer region of G . The edge v_4v_6 can be crossed only by the edge t_iv_3 , and the edge v_1v_3 can be crossed either by the edge t_iv_2 or the edge t_iv_6 . So, there are three configurations represented by the cyclic permutations $\mathcal{A}_1 = (125436), \mathcal{A}_2 = (154632)$ and $\mathcal{A}_3 = (163254)$. The edge v_4v_5 can be crossed either by the edge t_iv_2 or the edge t_iv_3 . And, there are three configurations represented by the cyclic permutations $\mathcal{A}_4 = (135246), \mathcal{A}_5 = (125346)$ and $\mathcal{A}_6 = (152463)$. Two edges, either t_iv_3 or t_iv_4 , can cross the edge v_2v_5 . And we have three configurations represented by the cyclic permutations $\mathcal{A}_7 = (123546), \mathcal{A}_8 = (132456)$ and $\mathcal{A}_9 = (124563)$. The edge v_3v_6 can be crossed by the edge t_iv_4 , so there exist two configurations represented by the cyclic permutations $\mathcal{A}_{10} = (132564), \mathcal{A}_{11} = (125643)$. The last possibility is that the edge v_2v_3 can be crossed either by the edge t_iv_4 or the edge t_iv_5 . Thus, we have four configurations represented by the cyclic permutations $\mathcal{A}_{12} = (142563), \mathcal{A}_{13} = (134256), \mathcal{A}_{14} = (135246)$ and $\mathcal{A}_{15} = (152463)$ (see Figure 4).

Now assume the drawing of G with the corresponding vertex notation in such a way as shown in Figure 1(b) or 1(c). Since in both drawings there are at most 5 vertices of G in the boundary of any region, the edges of any T^i cross the edges of G at least once. Consider all possible drawings of $G \cup T^i$ in which the edges of T^i cross the edges of G exactly once. First, assume the drawing of G as shown in Figure 1(b). Let the vertex t_i be placed in the same region as the vertex v_1 . Since there is only one subdrawing of $G \cup T^i \setminus \{v_3, v_6\}$ represented by the rotation (1452) , there are two and four possibilities how to obtain the subdrawing $G \cup T^i$ depending on which region the edge t_iv_3 is placed and which edge of G is crossed by the edge t_iv_6 . These $2 \times 4 = 8$ possibilities under our consideration are denoted by $\mathcal{B}_1 = (134652), \mathcal{B}_2 = (146523), \mathcal{B}_3 = (136452), \mathcal{B}_4 = (164523), \mathcal{B}_5 = (145263), \mathcal{B}_6 = (134526), \mathcal{B}_7 = (134562)$ and $\mathcal{B}_8 = (145623)$. If the vertex t_i is placed in the outer region, there is the only subdrawing of $G \cup T^i \setminus \{v_1\}$ represented by the rotation (25463) . And, there are three possibilities how to obtain the subdrawing $G \cup T^i$ depending on which edge of G is crossed by the edge t_iv_1 . These three possibilities

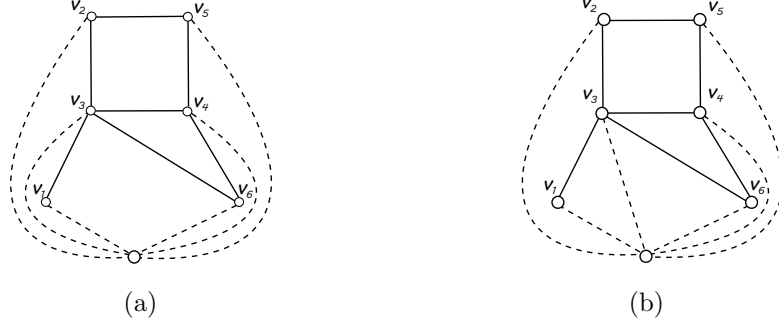


Figure 3: Drawings of all possible configurations of the graph F^i , if $cr(G) = 0$

$\mathcal{A}_1^* : (132546)$	$\mathcal{A}_2^* : (125463)$
------------------------------	------------------------------

Table 1: Configurations of graph $G \cup T^i$, $T^i \in R_D$

are $\mathcal{B}_9 = (154632)$, $\mathcal{B}_{10} = (146325)$ and $\mathcal{B}_{11} = (125463)$ (see Figure 5). Second, assume the drawing of G as shown in Figure 1(c). The vertex t_i must be placed in the outer region. There is the only subdrawing of $G \cup T^i \setminus \{v_1\}$ represented by the rotation (32546) and there are two possibilities how to obtain the subdrawing $G \cup T^i$ depending on which edge of G is crossed by the edge $t_i v_1$. These two possibilities are denoted by $\mathcal{C}_1 = (163254)$ and $\mathcal{C}_2 = (132546)$ (see Figure 6).

Assume the drawing of the graph G with one crossing among its edges. We will consider only such drawings of G for which there is a possibility of the existence of a subgraph $T^i \in R_D$, because of arguments in the proof of Theorem 1. Since there is $T^i \in R_D$, all vertices of G are placed in the same region. There are six possibilities how to obtain a crossing between two edges of G . The edge incident with a vertex of degree one crosses either one of the two edges of 4-cycle which are not adjacent to it, or with an edge of 3-cycle which is not adjacent to it. Another crossing between the edges of the graph G occurs when the edge of 4-cycle crosses the edge of 3-cycle. And the last possibility how to create a crossing is an crossing of two edges of 4-cycle (see Figure 2). If in the drawing of the graph G the edge incident with a vertex of degree one is crossed, there is exactly one possible drawing of the subgraph $G \cup T^i$ for $T^i \in S_D$. It is not difficult to see that for each of the remaining three diagrams of the graph G there are two possible drawings of the subgraph $G \cup T^i$ for $T^i \in S_D$ (see Figure 7).

The set \mathcal{M} consists of all configurations related to the concrete drawing of the graph G which is subdrawing $G \cup T^i$ induced in any good drawing of the graph $G + P_n$. And in the set \mathcal{M}_D there are all configurations from \mathcal{M} that exist in the drawing D . Let X, Y be the configurations from \mathcal{M}_D . We shortly denote by $cr_D(X, Y)$ the number of crossings in D between T^i and T^j for different $T^i, T^j \in R_D$ such that $G \cup T^i, G \cup T^j$ have configurations X, Y , respectively. Finally, let $cr(X, Y) = \min\{cr_D(X, Y)\}$ over all good drawings of the graph $G + P_n$. The previous can apply to $T^i, T^j \in S_D$. Our aim is to established $cr(X, Y)$ for all pairs $X, Y \in \mathcal{M}$.

We are ready to find the necessary numbers of crossings between T^i and T^j for the configurations $G \cup T^i$ and $G \cup T^j$ from \mathcal{M} . Consider the set of configurations $\mathcal{M} = \{\mathcal{A}_1^*, \mathcal{A}_2^*\}$. Assume the subgraph $G \cup T^i$ with the configuration \mathcal{A}_1^* . The configurations $\mathcal{A}_1^*, \mathcal{A}_2^*$ are represented by the cyclic permutations (125463) , (132546) , respectively. Since the minimum number of interchanges of adjacent elements of (125463) required to produce cyclic permutation $(132546) = (164523)$ is five, any subgraph T^j with the configuration \mathcal{A}_2^* crosses the edges of T^i at least five times, i.e., $cr(\mathcal{A}_1^*, \mathcal{A}_2^*) \geq Q(\text{rot}(t_i), \text{rot}(t_j)) = 5$. Clearly, also $cr(\mathcal{A}_i^*, \mathcal{A}_i^*) \geq 6$ for $i = 1, 2$. The lower bounds of number of crossings of two configurations from set $\{\mathcal{A}_1^*, \mathcal{A}_2^*\}$ are in the Table 2.

–	\mathcal{A}_1^*	\mathcal{A}_2^*
\mathcal{A}_1^*	6	5
\mathcal{A}_2^*	5	6

Table 2: Lower bounds of numbers of crossings for two configurations from $\{\mathcal{A}_1^*, \mathcal{A}_2^*\}$

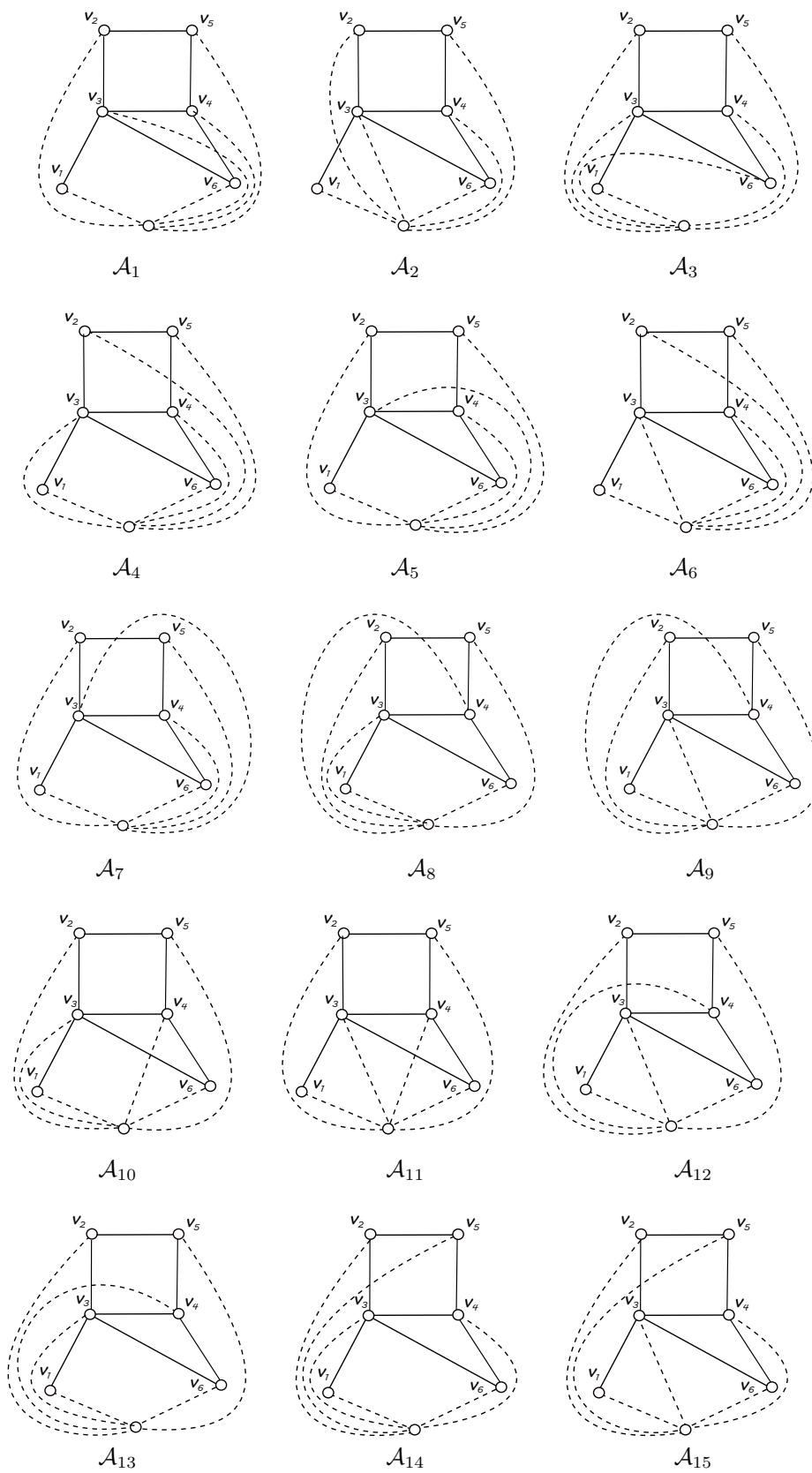


Figure 4: Fifteen drawings of the graph $G UT^i$ for $T^i \in S_D$, if G has a drawing as in Figure 1(a)

-	\mathcal{A}_1	\mathcal{A}_2	\mathcal{A}_3	\mathcal{A}_4	\mathcal{A}_5	\mathcal{A}_6	\mathcal{A}_7	\mathcal{A}_8	\mathcal{A}_9	\mathcal{A}_{10}	\mathcal{A}_{11}	\mathcal{A}_{12}	\mathcal{A}_{13}	\mathcal{A}_{14}	\mathcal{A}_{15}
\mathcal{A}_1	6	4	3	3	5	4	4	3	4	4	4	3	2	3	4
\mathcal{A}_2	4	6	4	4	3	4	3	3	4	3	4	4	4	4	4
\mathcal{A}_3	3	4	6	4	3	3	4	4	3	4	4	2	3	4	3
\mathcal{A}_4	3	4	4	6	3	5	4	4	3	4	4	2	3	6	5
\mathcal{A}_5	5	3	3	3	6	3	5	4	4	3	3	4	3	3	3
\mathcal{A}_6	4	4	3	5	3	6	4	3	4	4	4	3		5	6
\mathcal{A}_7	4	3	4	4	5	4	6	4	3	4	3	4	4	4	4
\mathcal{A}_8	3	3	4	4	4	3	4	6	5	4	3	4	5	4	3
\mathcal{A}_9	4	4	3	4	4	4	3	5	6	3	4	5	4	4	4
\mathcal{A}_{10}	4	3	4	4	3	4	4	4	3	6	5	3	4	4	4
\mathcal{A}_{11}	4	4	4	4	3	4	3	3	4	5	6	4	3	4	4
\mathcal{A}_{12}	3	4	2	2	4	3	4	4	5	3	4	6	5	2	3
\mathcal{A}_{13}	2	4	3	3	3	2	4	5	4	4	3	5	6	3	2
\mathcal{A}_{14}	3	4	4	6	3	5	4	4	4	4	4	2	3	5	6
\mathcal{A}_{15}	4	4	3	5	3	6	4	3	4	4	4	3	2	6	5

Table 3: Lower bounds of numbers of crossings between two different T^i and T^j with configurations \mathcal{A}_k and \mathcal{A}_l of $G \cup T^i$ and $G \cup T^j$, respectively

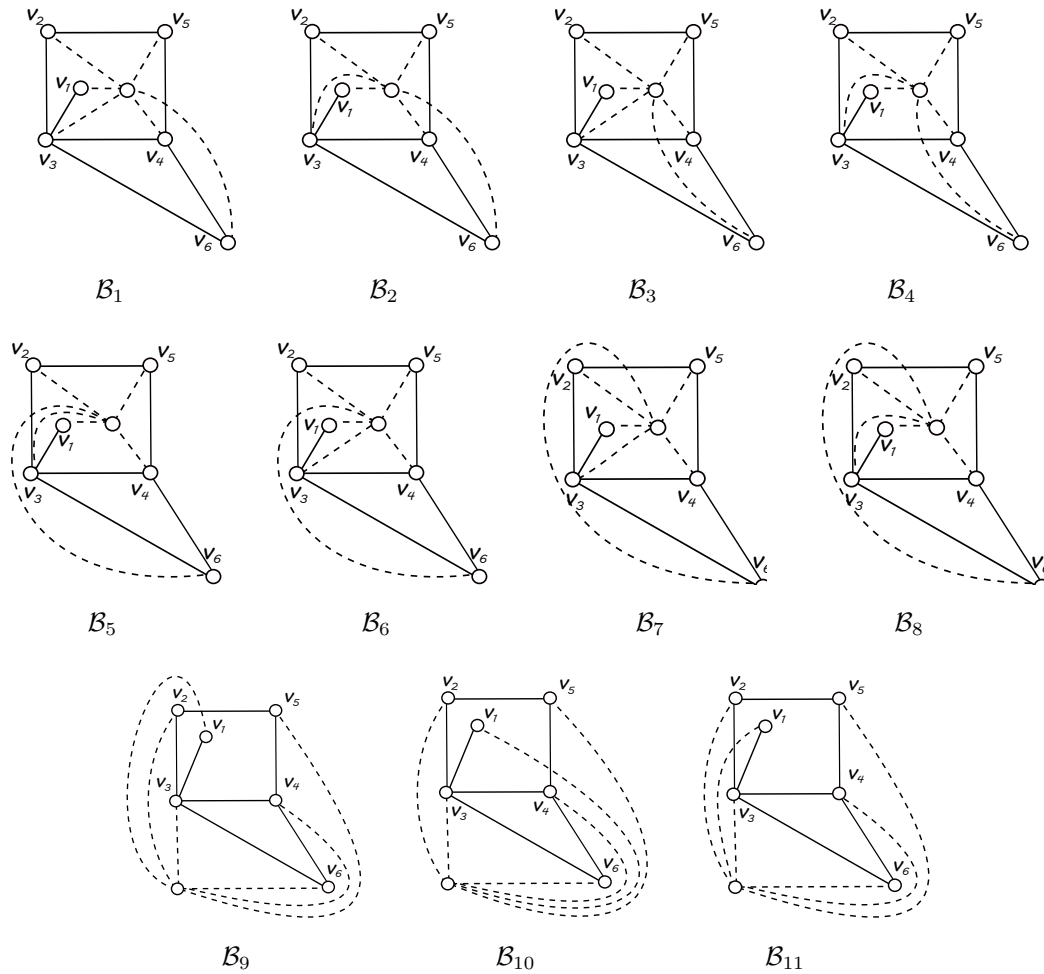


Figure 5: Eleven drawings of the graph $G \cup T^i$ for $T^i \in S_D$, if G has a drawing as in Figure 1(b)

–	\mathcal{B}_1	\mathcal{B}_2	\mathcal{B}_3	\mathcal{B}_4	\mathcal{B}_5	\mathcal{B}_6	\mathcal{B}_7	\mathcal{B}_8
\mathcal{B}_1	6	5	5	4	3	4	5	4
\mathcal{B}_2	5	6	4	5	4	3	4	5
\mathcal{B}_3	5	4	6	5	3	4	4	3
\mathcal{B}_4	4	5	5	6	4	3	3	4
\mathcal{B}_5	3	4	3	4	6	5	4	5
\mathcal{B}_6	4	3	4	3	5	6	5	4
\mathcal{B}_7	5	4	4	3	4	5	6	5
\mathcal{B}_8	4	5	3	4	5	4	5	6

Table 4: Lower bounds of numbers of crossings between two different T^i and T^j with configurations \mathcal{B}_k and \mathcal{B}_l of $G \cup T^i$ and $G \cup T^j$, respectively

–	B_9	B_{10}	B_{11}
B_9	6	5	5
B_{10}	5	6	4
B_{11}	5	4	6

Table 5: Lower bounds of numbers of crossings between two different T^i and T^j with configurations \mathcal{B}_k and \mathcal{B}_l of $G \cup T^i$ and $G \cup T^j$, respectively

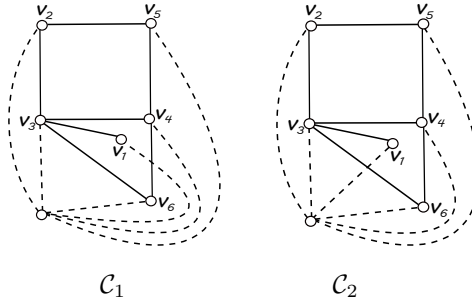


Figure 6: Two drawings of the graph $G \cup T^i$ for $T^i \in S_D$, if G has a drawing as in Figure 1(c)

–	\mathcal{C}_1	\mathcal{C}_2
\mathcal{C}_1	6	5
\mathcal{C}_2	5	6

Table 6: Lower bounds of numbers of crossings between two different T^i and T^j with configurations \mathcal{C}_k and \mathcal{C}_l of $G \cup T^i$ and $G \cup T^j$, respectively

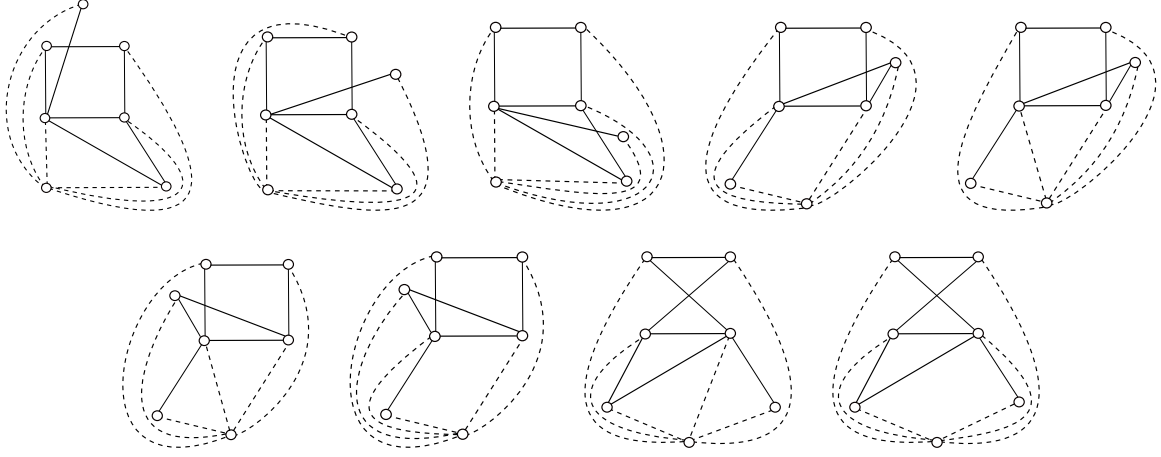


Figure 7: Nine drawings of the graph F^i , if $cr(G) = 1$

Now, consider the set of configurations $\mathcal{N} = \{\mathcal{A}_i, i = 1, \dots, 15\}$ and subset of this set which contains configurations occurring in the considered drawing D of $G + P_n$. The verification of the lower bounds for the number of crossings of two configurations from \mathcal{N} proceeds in the same way as before (the drawings of subgraph $G \cup T^i$ are in Figure 4). The resulting lower bounds for the number of crossings of configurations from \mathcal{N} are summarized in the Table 3. Take into consideration the set of configurations $\{\mathcal{B}_i, i = 1, \dots, 8\}$ and subset of this set which contains configurations occurred in the studied drawing D of $G + P_n$. In the Table 4 are summarized lower bounds of numbers of crossings for any two configurations of the subgraphs $G \cup T^i$ and $G \cup T^j$ with drawings in Figure 5. Assume the subset of the set $\{\mathcal{B}_9, \mathcal{B}_{10}, \mathcal{B}_{11}\}$ contains configurations occurring in examined drawing D of $G + P_n$. In the Table 5, there are lower bounds of numbers of crossings for any two configurations of the subgraphs $G \cup T^i, G \cup T^j$ with drawings in Figure 5. As the edges of the path P_n are not crossed in D , $\{\mathcal{B}_1, \dots, \mathcal{B}_8\}$ is disjoint with $\{\mathcal{B}_9, \mathcal{B}_{10}, \mathcal{B}_{11}\}$. Consider set of configurations $\{\mathcal{C}_1, \mathcal{C}_2\}$ and subset of this set contains configurations occurred in concrete drawing D of $G + P_n$. The lower bounds of numbers of crossings for any two configurations of the subgraphs $G \cup T^i, G \cup T^j$ with drawings in Figure 6 are given in the Table 6.

Lemma 1. Let D be a good drawing of $G + P_n$, $n \geq 2$ with a vertex notation of the graph G as in Figure 1(a). If $T^n \in R_D$ such that F^n has configuration $A_i^* \in \mathcal{M}_D$ for $i = 1, 2$, then $cr_D(T^n, T^k) \geq 3$ for any $T^k \in S_D$.

Proof. Let in D the graph F^n have configuration A_1^* . Since $T^k \in S_D$, the vertex t_k cannot be placed in the region bounded either by 4-cycle or 3-cycle of the graph G . If the vertex t_k is placed in the region with vertices v_1, v_3, v_6, t_n on its boundary (Figure 3(a)), it is easy to see, that $cr_D(T^n, T^k) \geq 3$. Moreover, if the vertex t_k is placed in another region, there are exactly two vertices of the graph G on its boundary, then $cr_D(T^n, T^k) \geq 3$. The same idea can be used for configuration A_2^* . \square

Theorem 1. $cr(G + P_n) = 6 \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor + 2 \lfloor \frac{n}{2} \rfloor + 1$ for $n \geq 2$.

Proof. There is a drawing of $G + P_n$ (see Figure 8) with $6 \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor + 2 \lfloor \frac{n}{2} \rfloor + 1$ crossings. Thus, we have $cr(G + P_n) \leq 6 \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor + 2 \lfloor \frac{n}{2} \rfloor + 1$. We prove the reverse inequality by induction on n . Using algorithm on the website <http://crossings.uos.de/>, we can prove that the result is true for $n = 2$ and $n = 3$. Suppose now that, for $n \geq 4$, there is a drawing D with

$$cr_D(G + P_n) < 6 \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor + 2 \lfloor \frac{n}{2} \rfloor + 1, \quad (4)$$

and let

$$cr_D(G + P_m) \geq 6 \lfloor \frac{m}{2} \rfloor \lfloor \frac{m-1}{2} \rfloor + 2 \lfloor \frac{m}{2} \rfloor + 1 \quad \text{for any positive integer } m < n. \quad (5)$$

As the graph $G + D_n$ is a subgraph of the graph $G + P_n$ and $cr_D(G + D_n) = 6 \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor + 2 \lfloor \frac{n}{2} \rfloor$ (see [13]), it implies $cr_D(G + P_n) = 6 \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor + 2 \lfloor \frac{n}{2} \rfloor$. Thus, no edge of the path P_n is crossed in D and all vertices t_1, \dots, t_n are placed in the same region of the subdrawing $D(G)$.

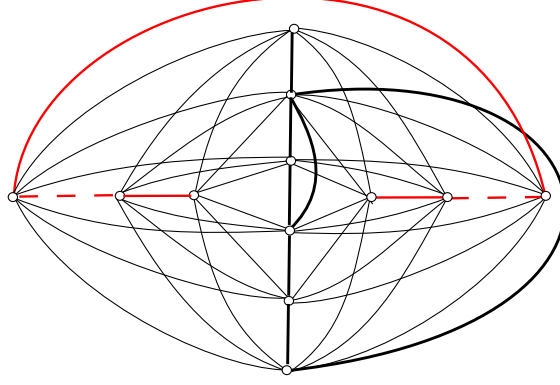


Figure 8: The graph of $G + P_n$ with $6 \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor + 2 \lfloor \frac{n}{2} \rfloor + 1$ crossings

First, we prove that the considered drawing D must be antipodal-free, that is $\text{cr}_D(T^i, T^j) \neq 0$ for all $i, j, i \neq j$. As a contradiction suppose that, without loss of generality, $\text{cr}_D(T^{n-1}, T^n) = 0$. If $\text{cr}(G) = 0$, using Table 2, both subgraphs T^n and T^{n-1} are not from the set R_D . Assume there is exactly one of T^n or T^{n-1} in the set R_D . Without loss of generality, $T^n \in R_D$. As every region in the drawing of $G \cup T^n$ contains at most four vertices of G on its boundary and $\text{cr}_D(T^{n-1}, T^n) = 0$, we have $\text{cr}_D(T^{n-1}, G) \geq 2$. If both T^n and T^{n-1} are from the set S_D , then $\text{cr}_D(G, T^{n-1} \cup T^n) \geq 2$. If $\text{cr}(G) \geq 1$, one can easily see that T^n and T^{n-1} are from the set S_D , thus $\text{cr}_D(G, T^{n-1} \cup T^n) \geq 2$. The fact that $\text{cr}(K_{6,3}) = 6$ implies that any $T^k, k = 1, 2, \dots, n-2$, crosses $T^{n-1} \cup T^n$ at least six times. So, for the number of crossings, in D , we have

$$\begin{aligned} \text{cr}_D(G + P_n) &= \text{cr}_D(G + P_{n-2}) + \text{cr}_D(T^{n-1} \cup T^n) + \text{cr}_D(K_{6,n-2}, T^{n-1} \cup T^n) + \\ &+ \text{cr}_D(G, T^{n-1} \cup T^n) \geq 6 \lfloor \frac{n-2}{2} \rfloor \lfloor \frac{n-3}{2} \rfloor + 2 \lfloor \frac{n-2}{2} \rfloor + 1 + 6(n-2) + 2 = \\ &= 6 \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor + 2 \lfloor \frac{n}{2} \rfloor + 1. \end{aligned}$$

It contradicts with (4). So, D must be an antipodal-free.

Moreover, our assumption on D together with $\text{cr}(K_{6,n}) = 6 \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor$ implies that

$$\text{cr}_D(G) + \text{cr}_D(G, K_{6,n}) \leq 2 \lfloor \frac{n}{2} \rfloor.$$

Let us denote $r = |R_D|$ and $s = |S_D|$. Then,

$$\text{cr}_D(G) + 0r + 1s + 2(n - r - s) \leq 2 \lfloor \frac{n}{2} \rfloor. \quad (6)$$

If $\text{cr}_D(G) = 0$, then $2r + s \geq 2n - 2 \lfloor \frac{n}{2} \rfloor$. Moreover, if $r = 0$, then $s = n$.

Case 1: $\text{cr}_D(G) = 0$.

First, we can choose the vertex notation of the graph G as shown in Figure 1(a). Let us consider that $n = s$. We have two possibilities.

(i) For every $i, j, i \neq j$: $\text{cr}_D(T^i, T^j) \geq 3$.

Without loss of generality let $T^n \in S_D$ and let us fix $G \cup T^n$. Then we have

$$\begin{aligned} \text{cr}(G + P_n) &\geq \text{cr}_D(K_{6,n-1}) + \text{cr}_D(K_{6,n-1}, G \cup T^n) + \text{cr}_D(G \cup T^n) \geq \\ &\geq 6 \lfloor \frac{n-1}{2} \rfloor \lfloor \frac{n-2}{2} \rfloor + 4(n-1) + 1 > 6 \lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor + 2 \lfloor \frac{n}{2} \rfloor. \end{aligned}$$

(ii) Using Table 3, there is not $T^i, T^j \in S_D, i \neq j$ such, that $\text{cr}_D(T^i, T^j) = 1$, we assume that there are $T^i, T^j \in S_D, i \neq j$: $\text{cr}_D(T^i, T^j) = 2$.

If there exists also such $T^k \in S_D$ that for every $T^l \in S_D$, $l \neq k$: $\text{cr}_D(T^k, T^l) \geq 3$, we fixed $G \cup T^k$. And the same inequalities as in previous case (i) hold.

If there is not such $T^k \in S_D$, without loss of generality, let for $T^n, T^{n-1} \in S_D$ is $\text{cr}_D(T^{n-1}, T^n) = 2$ and let us fix $G \cup T^{n-1} \cup T^n$. In this step we are interested in all possible drawings of the subgraph $G \cup T^i$ for some $T^i \in S_D$. We have $\text{cr}_D(T^{n-1} \cup T^n, T^i) \geq 6$ for every T^i with $i \neq n, n-1$, by summing the values in all columns in the considered two rows of Table 3. Thus,

$$\begin{aligned} \text{cr}(G + P_n) &\geq \text{cr}_D(K_{6,n-2}) + \text{cr}_D(K_{6,n-2}, G \cup T^n \cup T^{n-1}) + \text{cr}_D(G \cup T^n \cup T^{n-1}) \geq \\ &\geq 6 \left\lfloor \frac{n-2}{2} \right\rfloor \left\lfloor \frac{n-3}{2} \right\rfloor + 7(n-2) + 4 > 6 \left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2 \left\lfloor \frac{n}{2} \right\rfloor. \end{aligned}$$

Let us consider that $n \neq s$, that is, $n \geq s+1$. Using (6), we have $r \geq 1$. Let us assume that $T^n \in R_D$ with F^n having configuration either \mathcal{A}_1^* or \mathcal{A}_2^* . We will discuss two possibilities over congruence n modulo 2.

- Let n be even. By fixing the graph $G \cup T^n$ and using Table 2, Lemma 1, we have

$$\begin{aligned} \text{cr}(G + P_n) &\geq \text{cr}_D(K_{6,n-1}) + \text{cr}_D(K_{6,n-1}, G \cup T^n) + \text{cr}_D(G \cup T^n) \geq \\ &\geq 6 \left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n-2}{2} \right\rfloor + 5(r-1) + 4s + 3(n-r-s) + 0 = \\ &= 6 \left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n-2}{2} \right\rfloor + 3n + (2r+s) - 5 \geq \\ &\geq 6 \left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n-2}{2} \right\rfloor + 3n + \left(2n - 2 \left\lfloor \frac{n}{2} \right\rfloor\right) - 5 > 6 \left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2 \left\lfloor \frac{n}{2} \right\rfloor. \end{aligned}$$

- Let n be odd. By fixing the subgraph T^n ,

$$\begin{aligned} \text{cr}(G + P_n) &\geq \text{cr}_D(G + P_{n-1}) + \text{cr}_D(G + P_{n-1}, T^n) \geq \\ &\geq 6 \left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n-2}{2} \right\rfloor + 2 \left\lfloor \frac{n-1}{2} \right\rfloor + 1 + 5(r-1) + 3s + 1(n-r-s) + 0 = \\ &= 6 \left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n-2}{2} \right\rfloor + 2 \left\lfloor \frac{n-1}{2} \right\rfloor + n + 2(2r+s) - 4 \geq \\ &\geq 6 \left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n-2}{2} \right\rfloor + 2 \left\lfloor \frac{n-1}{2} \right\rfloor + n + 2 \left(2n - 2 \left\lfloor \frac{n}{2} \right\rfloor\right) - 4 > 6 \left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2 \left\lfloor \frac{n}{2} \right\rfloor. \end{aligned}$$

Consider the drawing of G as in the Figure 1(b) or 1(c). As we mentioned above, the edges of every T^i cross the edges of G . As there does not exist $T^i \in R_D$, i.e., $r = 0$, using (6), we have $s = n$. We use the results from the Tables 4, 5, 6, and we have $\text{cr}_D(T^i, T^j) \geq 3$ for every i, j . Thus, by fixing the graph $G \cup T^n$ we have

$$\begin{aligned} \text{cr}(G + P_n) &\geq \text{cr}_D(K_{6,n-1}) + \text{cr}_D(K_{6,n-1}, G \cup T^n) + \text{cr}_D(G \cup T^n) \geq \\ &\geq 6 \left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n-2}{2} \right\rfloor + 4(n-1) + 1 > 6 \left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2 \left\lfloor \frac{n}{2} \right\rfloor. \end{aligned}$$

Case 2: $\text{cr}_D(G) = 1$.

Based on equation (5), $r \geq 1$. Without loss of generality, we assume that $T^n \in R_D$. With respect to drawings of $G \cup T^n$ (see Figure 7), it is possible to verify, that the edges of T^i cross the edges of $G \cup T^n$ at least four times for every $i = 1, \dots, n-1$. So, by fixing the graph F^n we have

$$\begin{aligned} \text{cr}(G + P_n) &\geq \text{cr}_D(K_{6,n-1}) + \text{cr}_D(K_{6,n-1}, G \cup T^n) + \text{cr}_D(G \cup T^n) \geq \\ &\geq 6 \left\lfloor \frac{n-1}{2} \right\rfloor \left\lfloor \frac{n-2}{2} \right\rfloor + 4(n-1) + 1 > 6 \left\lfloor \frac{n}{2} \right\rfloor \left\lfloor \frac{n-1}{2} \right\rfloor + 2 \left\lfloor \frac{n}{2} \right\rfloor. \end{aligned}$$

Case 3: $\text{cr}_D(G) \geq 2$.

We use the same idea as in previous case for all possible drawings of the graph G with a possibility of an existence of a subgraph $T^i \in R_D$ in the considering drawing D . It completes the proof. \square

References

- [1] Berežný, Š., Buša, J. Jr., Staš, M. (2018). *Software solution of the algorithm of the cyclic-order graph*, Acta Electrotechnica et Informatica, 18, No. 1, 3–10.
- [2] Berežný, Š., Staš, M. (2017). *On the crossing number of the join of five vertex graph G with the discrete graph D_n* , Acta Electrotechnica et Informatica, 17, No. 3, 27–32.
- [3] Berežný, Š., Staš, M. (2018). *Cyclic permutations and crossing numbers of join products of symmetric graph of order six*, Carpathian J. Math., 34, No. 2, 143–155.
- [4] Garey, M. R., Johnson, D. S. (1983). *Crossing number is NP-complete*. SIAM J. Algebraic. Discrete Methods, 4, 312–316.
- [5] Hernández-Vélez, C., Medina, C., Salazar, G. (2014). *The optimal drawing of $K_{5,n}$* . Electronic Journal of Combinatorics, 21(4), 29.
- [6] Kleitman, D. J. (1970). *The crossing number of $K_{5,n}$* . J. Combinatorial Theory, 9, 315–323.
- [7] Klešč, M. (2010). *The crossing number of join of the special graph on six vertices with path and cycle*, Discrete Math., 310, 1475–1481.
- [8] Klešč, M. (2007). *The join of graphs and crossing numbers*, Electron. Notes in Discrete Math., 28, 349–355.
- [9] Klešč, M., Schrötter, Š. (2012). *The crossing numbers of join of paths and cycles with two graphs of order five*, Combinatorial Algorithms, Sprinder, LNCS, 7125, 160–167.
- [10] Klešč, M., Schrötter, Š. (2011). *The crossing numbers of join products of paths with graphs of order four*, Discuss. Math. Graph Theory, 31, 321–331.
- [11] Klešč, M., Valo, M. (2012). *Minimum crossings in join of graphs with paths and cycles*, Acta Electrotechnica et Informatica, 12, No. 3, 32–37.
- [12] Staš, M. (2017). *On the crossing number of the join of the discrete graph with one graph of order five*, J. Math. Model. and Geometry, 5, No. 2, 12–19.
- [13] Staš, M. (2018). *Cyclic permutations: Crossing numbers of the join products of graphs*, Proc. Aplimat 2018: 17th Conference on Applied Mathematics, 979–987.
- [14] Staš, M. (2019). *Determining crossing number of one graph of order five using cyclic permutations*, Proc. Aplimat 2019: 18th Conference on Applied Mathematics, 1126–1134.
- [15] Staš, M. (2019). *Determining crossing number of join of the discrete graph with two symmetric graphs of order five*, Symmetry, 11, No. 2.
- [16] Woodall, D. R. (1993). *Cyclic-order graphs and Zarankiewicz's crossing number conjecture*, J. Graph Theory, 17, 657–671.

Efficiency of Credit Risk Management and Their Determinants in Central European Banking Industries

Xiaoshan Feng¹

Abstract. Credit risk is one of the major risks in commercial banks. Therefore, whether commercial banks conduct effective credit risk management and employ technology changes with the times are essential. The main aim of this study is to evaluate the performance of credit risk management and productivity change and identify the determinants. We employ the Data Envelopment Analysis (DEA) on selected commercial banks in the Czech Republic, Germany, Republic of Austria, Poland, and Hungary. Based on valid data from 2012 to 2019, selected banks received lower efficiency scores using variable returns to scale in line with expectations. Additionally, strong evidence from the Malmquist Index demonstrated that the selected banking industries have various improvements during the past 8 years due to innovation in credit risk management. Furthermore, logistic regression results emphasized the significant differences among banking industries and suggested the credit risk measurement (SA/IRB), size, capital adequacy, and ownership have significant influences on the likelihood of banks being efficient on credit risk management.

Keywords: Credit risk management, non-performing loans ratio, macroeconomic variables, bank performance indicator, central Europe, data envelopment analysis, Malmquist index, logistic regression.

JEL Classification: G21, C31, C67, C80, C61, C58

AMS Classification: 62M10, 91G40, 91G70, 90C05

1 Introduction

For the past few years, accompanied by the recovery of the whole economy, robust growth in the volume of banking lending activities achieved. Thus, commercial banks need to be more cautious about the quality of their assets. While the outbreak of the global pandemic (COVID-19) lightened an incoming challenge for the global economy. Not only bank industry should prepare for the possible forthcoming deterioration, but policymakers and regulation institutions also need to implement a more effective framework to reduce the relevant risks.

As one of the most important risks, credit risk can lead to a huge failure in banks. Subsequently, investigating the efficiency of credit risk management of commercial banks becomes one of the most important steps to measure the overall soundness of the banking industry. Simultaneously, it is necessary to get to the bottom of the possible determinants in credit risk management efficiency.

The objective of this paper is to evaluate how the efficiency of credit risk management is influenced by the macroeconomic and bank itself determinants for selected banking industries in Central Europe. In this paper, we selected 10 commercial banks from each of 5 countries in Central Europe, which are, Czech Republic, Germany, Republic of Austria, Poland, and Hungary, respectively.

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.

2 Literature Review

In general, prior research typically investigated the operational efficiency of the banking industry, while the credit risk management efficiency has received limited attention in this area. Additionally, studies which employed the logistic regression model are scarce especially from the standpoint of efficiency. In this section, we will summarize and compare the relevant research.

The most widely used technique to measure efficiency is Data Envelopment Analysis (DEA). Much of the current literature conducts DEA to evaluate banking efficiency. Řepková [11] suggested that the efficiency scores from

¹ VSB - Technical University of Ostrava, Department of Finance, Sokolská třída 33 702 00 Ostrava, Czech Republic, xiaoshan.feng@vsb.cz.

the BCC model have higher values than from the CCR model due to the elimination of deposit management inefficiency. The study applied the Malmquist index to estimate the efficiency change in the Czech banks over time from 2001 to 2010. The negative growth in efficiency indicates the industry has lacked innovation or technological progress during the time.

Several studies incorporate non-performing loans ratio (NPLs) as a proxy of credit risk when measuring the efficiency of credit risk management in the banking sector. Undesirable outputs like NPLs may present in the banking sector which prefers 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 the NPLs ratio as an output but use the inverse value. The second method is to treat this undesirable output as input, which applied in other studies (Puri and Yadav [10]; Toloo and Hančlova [14]). 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.

Gaganis et al. [5] included loan loss provisions (LLPs) as an input variable to 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. More recent attention has focused on the determinants of credit risk management, when the NPLs ratio is widely incorporated as a proxy of credit risk (Messai and Jouini [13]; Škarica [12]; Louzis, et al. [8]).

One of the most important steps before credit risk management is to measure the credit risk. Since the enforcement of Basel II, banks calculate their minimum capital requirements under Pillar I use risk weights provided by the standardized approach (SA) or the internal ratings-based approach (IRB).

Cucinelli et al. [4] suggested that banks using IRB were able to curb the increase in credit risk driven by the macroeconomic slowdown better than banks under the standardized approach, which provide the study of using IRB shows better performance than using SA. Hakenes and Schnabel [7] published an analysis of bank size and risk-taking under Basel II, they found out although bank can choose between SA and IRB, while makes larger banks a competitive advantage and pushes smaller banks to take higher risks. This may even lead to higher aggregate risk-taking. Thus far, few studies have incorporated credit risk measurement as one of the impact factors for credit risk management efficiency.

There exhibits abundant literature focus on the bank's operational efficiency and lack of timeliness. It is rare to investigate efficiency from the view of credit risk management of banks. On top of that, when many studies examine the determinants of credit risk management, dynamic panel data analysis is widely used in previous literature. Scarce research specified the determinants of credit risk efficiency from external and internal impacts. Hence, the knowledge gap can be addressed.

This study will first measure the efficiency scores from DEA, subsequently, examine the determinants of banking credit risk management efficiency using a logistic regression model. To diversify the choices of determinants, in addition to macroeconomic perspective, we will incorporate the bank relevant factors which are, respectively, profitability factor, size, ownership (domestic or foreign-owned), capital adequacy, and the credit risk measurement (IRB or SA).

3 Methodology and Data collection

Following the previous literature, we apply Data Envelopment Analysis and the Malmquist index to measure the efficiency of credit risk management and productivity change of selected 10 banks from each of the five countries, respectively, Austria, the Czech Republic, Germany, Hungary, and Poland during the period from 2012-2019. Moreover, we employ the logistic regression model to investigate the possible internal and external determinants of credit risk management efficiency.

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. 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_{rj} are the matrixes of inputs and outputs respectively. The efficiency rate of such a unit can be expressed as:

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

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:

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

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

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

$$v_1, v_2, \dots, v_i \geq 0, (i = 1, 2, \dots, m) \quad (5)$$

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.

$$MI = (Catch - up) \times (Frontier - shift) \quad (6)$$

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.

a) 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 [6], then developed by Bowlin [3], 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 (Toloo and Tichý [15]).

In this study, we will collect data of 10 representative commercial banks from each of the five industries, which includes the large size bank, medium and small size. All data are from the annual report of each bank on a consolidated basis.

This study aims to investigate the efficiency of credit risk management in the banking industries. Therefore, we will apply the intermediation approach to measure the efficiency of credit risk management based on the DEA model. To assess credit risk modeling in the banking industry, Berg et al. [2] suggested using non-performing loans as a proxy of credit risk in a nonparametric study of the bank production, Altunbas et al. [1] incorporated loan loss provisions to analyze the efficiency of Japanese banks.

Therefore, based on previous literature, we proxy credit risk by the ratio of non-performing loans to total gross loans, so-called NPLs ratio. Then incorporate loan loss provision ratio to represents the ratio of provision and non-performing loans, 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 10 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, NPLs ratio is transformed as *Input* x_2 .

b) 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:

$$Y_{i,t} = \beta_0 + X'_{i,t}\beta + \epsilon_{i,t}, i = 1, \dots, N, t = 1, \dots, T. \quad (7)$$

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 $x_{i,t}$:

$$\pi(x_{i,t}) = p(Y_{i,t} = 1|X'_{i,t} = x_{i,t}) \quad (8)$$

The logistic regression model can be constructed as follows:

$$\log \frac{\pi(x_{i,t})}{1 - \pi(x_{i,t})} = \beta_0 + X'_{i,t}\beta, i = 1, \dots, N, t = 1, \dots, T. \quad (9)$$

c) Data Selection for Logistic Regression Model

To investigate the determinants of the efficiency of banks' credit risk management, the dependent variable in the regression model is the efficiency score obtained from the previously 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. Risk-weighted assets calculated by the Standardized Approach (SA), which is calculated as the ratio of RWAs under SA to total RWAs. Ownership of bank is the binary variable, in which 1 represents foreign-owned, 0 represents domestic-owned. The binary outcomes are measured by 1 and 0, which represent DMU is efficient and inefficient, respectively.

4 Empirical results

4.1 Efficiency of selected banking industries

In this session, we will compare the five banking industries from efficiency results and the Malmquist index. The efficiency obtained from DEA SolverPro™. Generally, selected countries exhibited different levels of asymmetry within each banking industry, in which the Czech Banking industry has the largest one, TE ranges from 0.11-1. Fig. 1 shows TE calculated under the CCR model. The efficiency score ranges from 0.17-0.62. German banking industry fluctuated sharply among 5 selected industries, the efficiency score dramatically fell in 2015 due to the record-setting loss of the largest bank – Deutsche Bank, after experienced the Brexit and following events, the German banking system has a relatively low performance of credit risk management and started making a recovery based on the potential of retail banking service and the investing of IT upgrades. Hungary has relatively low efficiency scores, due to the importance of the banking industry in Hungary is surprisingly low, and relatively low amount of loans and advances to the household. Similar to Řepková [1], we obtained higher scores from BCC due to the variable returns to scale. Fig. 2 provides strong evidence that the average PTE large range from 0.32-0.92. During the year 2015, Hungary, Poland, and Germany experienced a sharp decline in credit risk management.

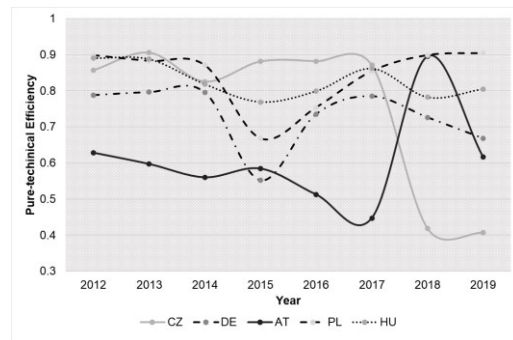
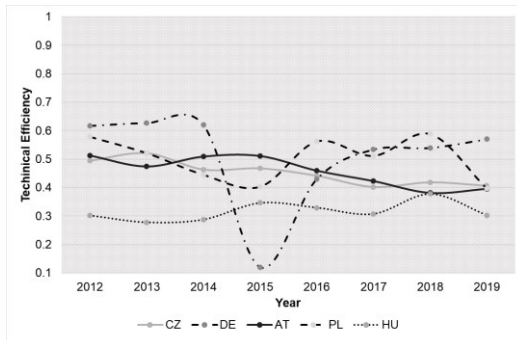


Figure 1 Technical efficiency among selected industries **Figure 2** Pure-technical efficiency among selected industries

In addition, the Austrian banking industry suffered a continuous collapse during 2012-2017, but a jump in credit risk management efficiency through technology progress in 2018. In contrast, the Czech banking industry has

relatively high efficiency of credit risk management till 2017, while a sharp collapse exhibited which might be caused by the announcement, that is, to exit from the exchange rate commitment. This is the first time in eleven years of the Czech intervention currency market. After the CNB announced the decision, the Czech koruna fell 3.2% to the euro, it was the biggest decline after 2010. At the same time, the CNB also announced to keep interest rates unchanged. This decline can correspond to the technology progress decline from 2017.

The Malmquist index measures the productivity change of a DMU. If MI is bigger than 1, it will indicate progress in the productivity change of a DMU from period 1 to 2. We can see from Fig. 3, only the German and the Austrian banking industries have sharp fluctuations and MI larger than 2 and 8 during the period 2014 to 2015, and 2015 to 2016 respectively, which indicate these two industries have huge productivity changes during specific periods. Except for Poland, the rest of the selected banking industries have not improved their development in credit risk management but have steady improvements for recent years.

While under the assumption of variable returns to scale, only the German banking industry has exceptional fluctuation during the period 2014 to 2015. But during the period 4 and 5, except the Czech Republic and Hungary, other banking industries did not have progress.

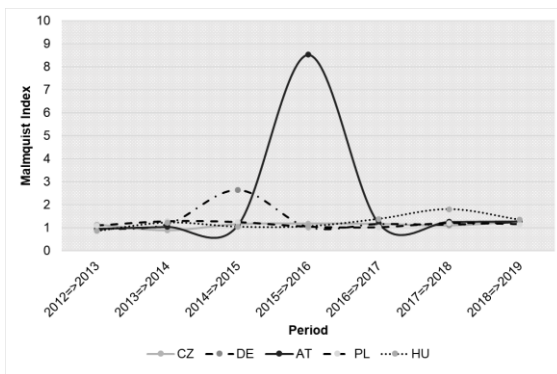


Figure 3 Malmquist index (CRS) among selected industries

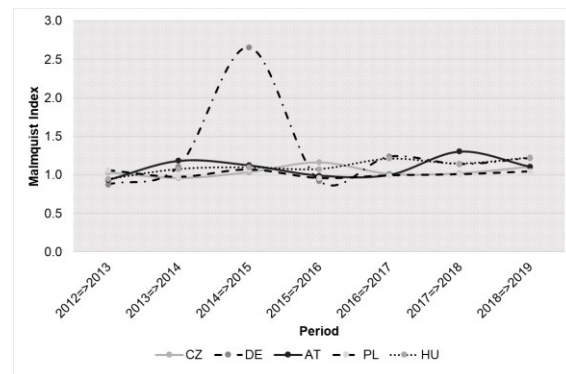


Figure 4 Malmquist index (VRS) among selected industries

4.2 Determinants of Efficient Credit Risk Management

Under logistic regression, we estimated the likelihood of a DMU is efficient given the determinants SIZE, CAR, ROAA, SA, OWN and GDPG. Strong evidence proves that size, capital adequacy, the use of standardized approach to calculate RWAs, and profitability were statistically significant for the efficiency of credit risk management, this finding valid for the Czech Republic, Germany, Austria, and Poland. Put differently, the probability of a bank being efficient increases with larger size and less exposure which calculated under standardized approach to meet the capital requirement, higher capital adequacy ratio, and lower profitability.

The results can be strongly supported by the fact that lower asset quality under a flourishing economy. Moreover, larger size banks proved to have a relatively higher ability to conduct credit risk management just as the efficient DMU- ČSOB in the Czech Republic and Erste Group Bank in Austria. Based on our efficiency scores, the efficient DMU is one of the largest banks in the industry. In Austria and Poland, regression results indicate the flourishing economy will increase the probability of banks acting efficient in credit risk management, while in Hungary, Germany, and the Czech Republic, it has surprisingly different opposite results. Simultaneously, ownership of banks only has a significant impact on the Polish and Austrian banking industry, which corresponding to both governments that encourage to reduce the percentage of foreign-owned domestic banks. Additionally, less use of standardized approach corresponds to Cucinelli [4] that using IRB shows better performance than using SA, and the statistical significance indicates that credit risk measurement has proved as one of the important factors of efficient credit risk management in selected banking industries.

5 Conclusion

The objective of this study is to evaluate the efficiency of bank credit risk management among five Central European countries, then get insight into the possible determinants of the likelihood of a bank managing credit risk efficiently.

The DEA model provides evidence that the large asymmetry of bank credit risk management within each banking industry especially in the Czech banking sector. The use of the DEA model reflects the characteristics of unit invariant. This paper selects the relevant indicators that can obtain the efficiency comparison of each DMU within the industry. Moreover, the weight in the DEA model is generated by means of an optimizing calculation. Therefore, the efficiency score is reasonable because it is not affected by human subjective factors. The essence behind

the DEA model form reference bank, thus, to provide a way for other banks in the industry to improve efficiency for inefficient banks. Although the DEA model has certain limitations, whether it is from the perspective of the number of DMU or the undesirable output in this paper, the DEA model still provides theoretical and empirical support for the advanced credit risk management of the selected banking industry.

Meanwhile, the Malmquist index demonstrates selected countries that have achieved productivity change especially the German banking industry. MI provides an intuitive and reliable trace that the development achieved through advanced technology has also had a positive impact on the efficiency of bank credit risk management.

Furthermore, taking advantage of the logistic regression model that the dependent variable is binary and the reasonable acceptance of multicollinearity, the result highlights the significant impact of size, approaches of credit risk measurement, capital adequacy, and profitability. Meanwhile, the significant differences among banking industries can be emphasized.

Generally, the probability of banks acting efficiently on credit risk management increasing with the larger size, a lower percentage of using standardized approach, adequate capital, and less profitability. Based on our study, we aim to provide the banking industry and policymakers a clear picture of the current circumstance of credit risk management in selected countries, simultaneously, we suggest the banks strengthening internal regulation and ensure adequate capital reserves through a comprehensive risk management framework, prepare for the incoming weakening economy and the possible deterioration of asset quality.

Acknowledgements

The author was supported through the Czech Science Foundation (GACR) under project 20-25660Y and moreover by SP2021/15, an SGS research project of VSB-TU Ostrava. The support is greatly acknowledged.

References

- [1] Altunbas, Y., Liu, M., Molyneux, P. and 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. and 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] Bowlin, W.F. (1998) 'Measuring Performance: An Introduction to Data Envelopment Analysis (DEA) '. *Journal of Cost Analysis* 7, pp. 3-27.
- [4] Cucinelli, D., Battista, M.L.D., Marchese, M. and Nieri, L. (2018). Credit risk in European banks: The bright side of the internal ratings-based approach. *Journal of Banking & Finance* 93(C): 213–229. <https://doi.org/10.1016/j.jbankfin.2018.06.014>.
- [5] Gaganis, C., Liadaki, A., Doumpos, M. and 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>.
- [6] Golany, B. and Roll, Y. (1989) 'An Application Procedure for DEA', *Omega* 17, pp. 237-250.
- [7] Hakenes, H. and Schnabel, I. (2011). Bank size and risk-taking under Basel II. *Journal of Banking & Finance*, 35, issue 6: 1436–1449. <https://doi.org/10.1016/j.jbankfin.2010.10.031>.
- [8] Louzis, D. P., Vouldis, A. T. and Metaxas, V. L. (2012) 'Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios', *Journal of Banking & Finance*. North-Holland, 36(4), pp. 1012–1027.
- [9] Paradi, J. C. and 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] Puri, J. and 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.
- [11] Řepková, I. (2012) 'Measuring the efficiency in the Czech banking industry : Data Envelopment Analysis and Malmquist index', pp. 781–786.
- [12] Škarica, B. (2014) 'Determinants of Non-Performing Loans in Central and Eastern European Countries', *Financial Theory and Practice, Institute of Public Finance*, vol. 38(1), pp. 37-59.
- [13] Messai, A.S. and Jouini, F. (2013) 'Micro and Macro Determinants of Non-performing Loans', *International Journal of Economics and Financial Issues*, 3(4), pp. 852–860.
- [14] Toloo, M. and 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>
- [15] Toloo, M. and Tichý, T. (2015) 'Two alternative approaches for selecting performance measures in data envelopment analysis', *Measurement*, Vol. 65, pp. 29-40.

Models of Technology Coordination

Petr Fiala¹, Renata Majovská²

Abstract. Changes in the current economy are caused by changes in the technological field, especially in information and communication technologies. These changes cause a strong network of economic subjects. The network becomes the basic structure of relationships and brings some specifics to which it is necessary to respond if economic entities want to stay in the current business. Specific network effects can be observed for both physical networks and virtual networks of users of the same products. Individually rational technology decisions can lead to collectively inefficient results when network externalities exist. Game theory provides tools for coordination analysis. The paper presents models of technology coordination based on a coordination game. Arthur's basic model is presented and generalized to model with converters and models with prices. Technology compatibility affects the selection of standards and social welfare.

Keywords: networks, technology, coordination, games

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

The network becomes the basic structure of relationships and brings some specifics to which it is necessary to respond if economic entities want to stay in the current business (see [4], [7]). Changes in the current economy are caused by changes in the technological field, especially in information and communication technologies. The adoption of new technologies in networks suggests that individually rational decisions regarding technologies can lead to collectively inefficient results in the presence of network externalities. However, the use of converter technologies may change these results. Converters allow previously unattainable network externalities to arise and, in some cases, prevent locking. These issues play an important role, especially in managing the development of computer networks and Internet standards.

Competing technologies such as Apple and Microsoft compete for market share based on several characteristics, including network size. Computers have their own value, allowing you to perform a number of activities independently. In addition, they have a network value that increases with the number of computers connected to the network. Network products exhibit network externalities that can be defined as increments of the benefits that a user derives from using the product as the number of users of the same type of product grows. Although positive externalities receive the most attention in the network literature, negative network externalities can also arise. The problem of locking when using a technology arises when this technology is used by more users than another technology, although this other technology may have a higher own value. Switching to another technology can be caused by its increasing network value, even though its own value is lower. However, switching to another technology also requires certain costs, depending on their amount and the increase in value that this change will bring. Due to the complementarity of the individual components of information and communication systems, their compatibility is required. This means that complementary components must work with the same standards. This creates a problem of coordination as firms agree on standards. Game theory provides tools for coordination analysis (see [6]). We will use simple tools such as a coordination game, Arthur's model and their modifications and generalizations to analyze the problem of technology coordination in networks.

¹ Prague University of Economics and Business, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, pfiala@vse.cz

² University of Finance and Administration, Prague, Department of Computer Science and Mathematics, Estonská 500, 101 00 Praha 10, Czech Republic, renata.majovska@mail.vsfs.cz

2 Coordination game

We will use a coordination game to analyze the coordination of technology selection. Suppose that two firms 1 and 2 faces the selection of two technologies A and B . The technology firms are shown in the table1. Both technologies exhibit network externalities, where it is assumed that the values apply

$$a > c, b > d. \quad (1)$$

The table 1 defines the so-called coordination game, where the goal is to coordinate both firms so that they use the same technology, where both firms achieve higher values than when each uses a different technology. It is actually a non-cooperative game of two players (firms 1 and 2) with two strategies (selection of technology A and B). The basic concept of solving non-cooperative games is to find the so-called Nash equilibrium, when changing the strategy of any of the players, provided that the strategies of other players remain unchanged, this player can only be damaged. In the coordination game there are two Nash equilibrium solutions; equilibrium (A, A) - both firms choose technology A and equilibrium (B, B) - both firms choose technology B .

		Firm 2	
		Technology A	Technology B
Firm 1	Technology A	$a; a$	$d; c$
	Technology B	$c; d$	$b; b$

Table 1 Values of technologies for firms

The existence of a number of equilibrium solutions raises the question of how firms will coordinate their activities. If technology A is new technology and technology B is old, the following two types of market failure can be observed. If it holds that the value $a > b$ and is chosen equilibrium solution (B, B) , ie. firms will stay with the worse rated old technology B , then there is talk of excess inertia. If it holds that the value $b > a$ and is chosen equilibrium solution (A, A) , ie. firms move to the worse-rated new technology A , then there is talk of excess momentum.

Both symmetric equilibrium solutions can be suitable candidates for selection. Another refinement of the notion of Nash equilibrium solution is the Pareto-dominant and risk-dominant equilibrium solution. An equilibrium solution is Pareto dominant when there are no other strategies for which the value of the solution is better for at least one player and not worse for the other players. An equilibrium solution is risk-dominant if the best response of both players remains unchanged until the opponent selects an equilibrium strategy with a probability of at least 0.5 (see [5]).

If applicable

$$a > b > c > d > 0 \text{ and } (b - d) > (a - c), \quad (2)$$

then the equilibrium solution (A, A) is Pareto-dominant and the equilibrium solution (B, B) is risk-dominant.

3 Arthur's basic model

Consider the Arthur's basic model [1] with two firms 1 and 2 and two technologies A and B . Each firm makes the decision to purchase the technology according to the initial preferred own value of the technology and the network externalities associated with each technology. These values are summarized in Table 2, where a_1 is the original preferred value of technology A for firm 1, n_A is the size of the network using technology A , s is the parameter of the network value. Analogous designations apply to firm 2 and technology B .

	Technology A	Technology B
Firm 1	$a_1 + s n_A$	$b_1 + s n_B$
Firm 2	$a_2 + s n_A$	$b_2 + s n_B$

Table 2 Values for technology selection

Increasing returns from scale are considered, so the parameter s is always positive. Firm 1 initially prefers technology A and firm 2 initially prefers technology B , so this is true

$$a_1 > b_1 \text{ and } a_2 < b_2. \quad (3)$$

According to Arthur's model, one firm buys one technology in each time period. The firm comes to market in the period t_i . The type of incoming firm is a random component in the model, both firms have the same probability of arrival. The selection of technology by firm i is determined by a combination of three factors:

- type of firm (random component);
- the initial preferred value of the technology;
- the number of previous selections of each technology.

The model assumes a two-component value of the technology. The first component is the own value of the technology, the second component is the network value. Arthur's shows that under these conditions, users will be locked in one of the technologies. This result can be easily derived from the matrix in Table 2. Firm 1 will initially prefer technology A, given the initial preferred own value and no network value. Firm 1 will switch to technology B as soon as it takes effect

$$b_1 + s n_B > a_1 + s n_A. \quad (4)$$

The inequality can be rewritten in the form of a so-called switching inequality

$$n_B - n_A > \frac{(a_1 - b_1)}{s}. \quad (5)$$

This inequality, together with a similar inequality for firm 2, determines the absorption barriers. Once the difference in network size with technology B exceeds the size of network with technology A by a certain value, determined by the initial preferred values and the parameter of the network value s , users will be locked by technology B, which becomes the standard. Firm 1 will give up technology A if the size of the network with technology B is such that the benefit from technology B exceeds the original preferred value for firm 1. At this point, both firms 1 and 2 will only buy technology B and the size of the network with technology A will not magnify. This analysis is expressed graphically in Figure 1.

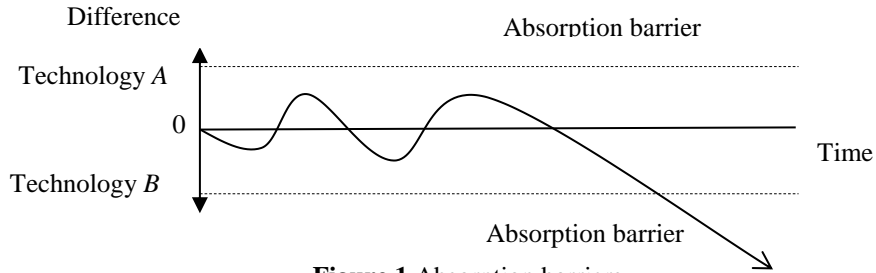


Figure 1 Absorption barriers

4 Generalization of Arthur's model

Arthur's basic model allows for certain generalizations (see [2]). We present a model with the possibilities of using converters so that the technologies become compatible. Next, we present models where prices for the purchase of technologies are introduced and an impact on social welfare is analyzed.

4.1 Model with converters

The introduction of converters, which allow compatibility between technologies, leads to interesting changes in Arthur's model. Let us introduce the compatibility parameters k_{AB} and k_{BA} , from the interval between zero and one, which measure the compatibility of technology A with technology B, respectively the compatibility of technology B with technology A. The values are summarized in Table 3, from which we derive some relations for absorption barriers.

	Technology A	Technology B
Firm 1	$a_1 + s n_A + k_{AB} s n_B$	$b_1 + s n_B + k_{BA} s n_A$
Firm 2	$a_2 + s n_A + k_{AB} s n_B$	$b_2 + s n_B + k_{BA} s n_A$

Table 3 Values for technology selection with converters

Assume a reciprocal converter that allows compatibility in both directions and $k_{AB} = k_{BA}$. Then the switching inequality has a shape

$$n_B - n_A > \frac{(a_1 - b_1)}{s(1 - k_{AB})}. \quad (6)$$

Compared to the basic Arthur's model, the absorption barriers are multiplied by a coefficient

$$\frac{1}{(1 - k_{AB})}. \quad (7)$$

Depending on the value of the compatibility parameter k_{AB} , the following situations may occur:

- For completely incompatible technologies, compatibility parameter $k_{AB} = 0$, we get the basic Arthur's model.
- For a partially compatible reciprocal converter, compatibility parameter $0 < k_{AB} < 1$, the absorption barriers are wider than in the situation without converters. As the value of the k_{AB} compatibility parameter increases, the absorption barriers expand.
- For a fully compatible reciprocal converter, compatibility parameter $k_{AB} = 1$, absorption barriers will be removed and users will not be locked by any technology. The situation can be captured graphically in Figure 2.

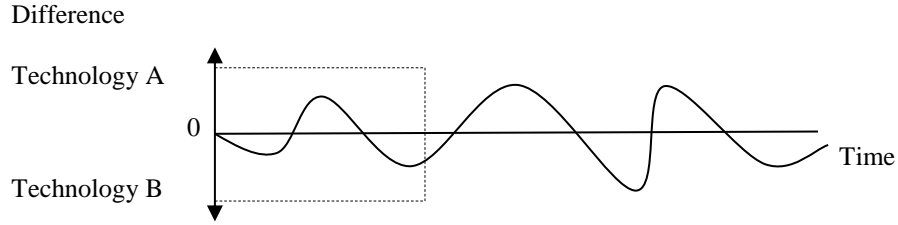


Figure 2 Fully compatible reciprocal converter

We can perform a similar analysis for a bidirectional converter where the compatibility parameters are not the same ($k_{AB} \neq k_{BA}$).

Let's further examine the situation when introducing a one-way converter. Assume that technology A has access to technology B ($0 < k_{AB} \leq 1$) using a one-way converter and technology B does not have access to technology A ($k_{BA} = 0$).

Depending on the value of the k_{AB} compatibility parameter, the following situations may occur:

- For a partially compatible one-way converter, compatibility parameter $0 < k_{AB} < 1$:
Firm 1 will switch from the originally preferred technology A to technology B as soon as it takes effect

$$(1 - k_{AB}) n_B - n_A > \frac{(a_1 - b_1)}{s}. \quad (8)$$

Firm 2 will switch from the originally preferred technology B to technology A as soon as it takes effect

$$n_A - (1 - k_{AB}) n_B > \frac{(b_2 - a_2)}{s}. \quad (9)$$

The absorption barrier for technology A is approaching and the absorption barrier for technology B is moving away.

- For a fully compatible one-way converter, compatibility parameter $k_{AB} = 1$:
Firm 1 would switch from the originally preferred technology A to technology B as soon as it takes effect

$$-n_A > \frac{(a_1 - b_1)}{s}, \quad (10)$$

however, this inequality never occurs because the left side is always negative and the right side is positive.

Firm 2 will switch from the originally preferred technology B to technology A as soon as it takes effect

$$n_A > \frac{(b_2 - a_2)}{s}. \quad (11)$$

As the number of users grows, Firm 2 will switch to Technology A, which is becoming the standard. The absorption barrier for technology A gets even closer and the absorption barrier for technology B ceases to exist.

4.2 Model with prices

So far, we have examined the situation without considering the prices for the acquired technology. Let us give a simple generalization of the Arthur's model, which explains how the standardization process works if we include in the model technology suppliers who strategically set prices. The modification has the same assumptions as the basic Arthur's model, in addition to introducing some other conditions. It assumes the existence of two suppliers A, B , each of which sponsors its own technology. These firms have market power to set prices, different from the marginal costs of manufacturing technology. Each technology is associated with fixed costs, which are always incurred and marginal costs are incurred only if the supplier sells the unit in a given period. Each supplier starts production with an initial subsidy to establish a fund. The supplier strategically sets the prices for the technology units sold, which flow to the suppliers in the fund when the unit is sold. Pricing strategically influences consumer behavior, subsidizing low prices attracts more people. However, once the fund is exhausted, the supplier goes bankrupt and production ends.

Let denote as p_A the price per unit of technology A , as p_B the price per unit of technology B . The values, including the own value of the technology and the network value, less the price paid for the technology, are summarized in Table 4. If the prices $p_A = p_B = 0$, we get the basic Arthur's model.

	Technology A	Technology B
Firm 1	$a_1 + s n_A - p_A$	$b_1 + s n_B - p_B$
Firm 2	$a_2 + s n_A - p_A$	$b_2 + s n_B - p_B$

Table 4 Values for technology selection with prices

Firm 1, for which we assume the original preference for technology A , will switch to technology B as soon as it takes effect

$$n_B - n_A > \frac{[(a_1 - b_1) - (p_A - p_B)]}{s}. \quad (12)$$

A similar switching inequality applies to firm 2. The consumer's decision to switch to another technology is given by comparing the difference between the initial preferred values of the technologies, the difference in prices and the size of the networks. If the supplier can sell at a sufficiently lower price than the competition, the effects of the difference between the initial preferred technology values and the network effects may be outweighed in the consumer's decision. An important feature of pricing decisions is the lack of an accurate forecast of a competitor's prices in each period.

4.3 Model with an impact on social welfare

The selection of standards and issues of technology compatibility have an impact on social welfare. It is possible to formulate a coordination game where firms are in a conflict situation when distributing revenues from standardization. A situation where there are more type 1 firms and more type 2 firms in the economy can be modeled as a game with a single firm 1 and a single firm 2, which compete in several rounds. In each round, it is a matter of determining the equilibrium for two firms and their contribution to social welfare, given by the sum of the achieved values for both firms. Both firms 1 and 2 have a selection of two strategies, selection of technology A or technology B . Consider in the payoff matrix the actual values of technologies and network values, including the possibility of using the converter for compatibility with other technologies, ie the same values as the model with converters. The values for the coordination game are given in Table 5.

		Firm 2	
		Technology A	Technology B
Firm 1	Technology A	$a_1 + s n_A + k_{AB} s n_B;$ $a_2 + s n_A + k_{AB} s n_B$	$a_1 + s n_A + k_{AB} s n_B;$ $b_2 + s n_B + k_{BA} s n_A$
	Technology B	$b_1 + s n_B + k_{BA} s n_A;$ $a_2 + s n_A + k_{AB} s n_B$	$b_1 + s n_B + k_{BA} s n_A;$ $b_2 + s n_B + k_{BA} s n_A$

Table 5 Impact on social welfare

The influence of converter efficiency on social welfare can be demonstrated on specific values. For lower values of the compatibility parameter, the game has coordinating Nash equilibrium solutions, equilibrium (A, A) and/or equilibrium (B, B) . As the compatibility parameter increases, the values for mixed networks containing both A and B technologies increase. As the compatibility parameter is further increased, the values for mixed networks increase and the social value also increases.

5 Conclusions

Coordination game, Arthur's model, their modifications and generalizations are used to analyze the problems of technology coordination in networks. Important conclusions can be drawn from the analyzed simple models. These conclusions concern changes in the selection and adoption of technology in the use of converters and strategic pricing. The influence of technology compatibility on the selection of standards and on social welfare is also analyzed.

The introduction of converters will change the locking of technologies and the adoption of standards. The specific changes depend on the type of converter, whether it is unidirectional or bidirectional and whether it is partially or fully compatible. In the case of a bidirectional converter, the introduction of a bidirectional partially compatible converter will delay the adoption of the technology standard, while the introduction of a bidirectionally fully compatible converter will completely prevent the adoption of the technology standard. In the case of a one-way converter, the introduction of a one-way partially compatible converter means a tendency towards locking by the preferred technology, and the introduction of a one-way fully compatible converter will cause the preferred technology to be adopted as standard.

The introduction of strategic technology pricing will significantly affect the selection and adoption of technology as a standard. If the financial resources of technology suppliers are highly asymmetric, then strong companies, by setting low prices, outweigh the disadvantages of the own value of the technology and can gain an increasing number of users and thus achieve the adoption of their own technology as a standard. For weak companies, this strategy is very risky and can lead to financial problems even if the technology's own value benefits.

Technology compatibility affects the selection of standards and social welfare. As the compatibility parameter increases, the values for mixed networks increase, and the social value increases and the deviation from coordinating solutions increases.

Applied coordination principles for the use of technologies in networks can be modified and adapted to other problems on networks, such as the coordination of resources in supply chains or project portfolios (see [3]).

Acknowledgements

This work was supported by the grant No. IGA F4/42/2021, Faculty of Informatics and Statistics, Prague University of Economics and Business. 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] Arthur, W. (1989). Competing Technologies, Increasing Returns, and Lock-in by Historical Events. *Economic Journal* 99, 116-131.
- [2] Fiala, P. (2016). *Dynamické vytváření cen a alokace zdrojů v sítích*. Praha: Professional Publishing.
- [3] Fiala, P. (2018). Project portfolio designing using data envelopment analysis and De Novo optimisation. *Central European Journal of Operations Research* 26, 847-859.
- [4] Gottinger, H.-W. (2006). *Economies of Network Industries*. London: Routledge.
- [5] Harsanyi, J.C. & Selten, R. (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge: MIT Press.
- [6] Kreps, D. (1991). *Game Theory and Economic Modelling*. Oxford: Oxford University Press.
- [7] Shy, O. (2001). *The Economics of Network Industries*. Cambridge: Cambridge University Press.

The Impact of Covid-19 on Mutual Relations of Czech Macro-aggregates: Effect of Structural Changes

Jakub Fischer¹, Kristýna Vltavská²

Abstract. One of the assumptions of economic analyses and prognoses consists of the stability of mutual relations of macro-aggregates. For example, there is a close connection between gross domestic product and gross national income. As another example, the share of the value-added on the (gross) production does not change quarter-to-quarter in standard time. Nevertheless, covid-19 pandemia and governmental measures led to the change in these relations. While ICT is going stronger, manufacturing remains stable, and services like accommodation and food service activities were almost entirely closed. Furthermore, the sectorial change (from non-financial institutions to government institutions, from market to non-market production) influences GDP and current taxes. The paper aims to analyse the impact of structural changes in 2020 on the mutual relations of the Czech macro-aggregates and ratio indicators. As the primary method, we use index decomposition, particularly the decomposition of total indicators into the levels effect and substitution effect. In 2020, based on the preliminary data, the substitution effect positively influences year-on-year changes in labour productivity and hourly wages and salaries (both by 0.5 p.p.). It has no impact on monthly wages and salaries' development.

Keywords: macro-aggregates, gross domestic product, labour productivity, covid-19, index decomposition

JEL Classification: C43, E24, O47

AMS Classification: 62P20

1 Introduction

The stability of mutual relations between key macro-aggregates is one of the assumptions of economic analyses and prognoses. As examples, we can mention a close connection between gross domestic product and gross national income. Similarly, the share of the value-added on the (gross) production does not change quarter-to-quarter in standard time. Nevertheless, covid-19 pandemia and governmental measures have brought substantial structural changes, and these relations have been broken. ICT activities and health services are going stronger, manufacturing remains stable, and services like accommodation and food service activities were almost entirely closed due to the lockdown. Furthermore, the sectorial change (from non-financial institutions to government institutions, from market to non-market production) influences GDP and current taxes.

Several recent papers focus on the post-covid-19 economic development and the perspective of individual industries. Kotz et al. [4] aim at the productivity and growth: they mention the potential of pandemic-related productivity acceleration potential in several industries like healthcare, construction, ICT and retail of 1.2–3.0% in years 2019–2024, driven by telemedicine, operational efficiency, industrialisation, digital construction, warehouse automation and e-commerce. A similar forecast is presented by Mischke et al. [6], who predict a potential to accelerate annual productivity growth by about one p.p. to 2024. They mention that this acceleration is twice comparing to the pre-pandemic rate of productivity growth. Bloom et al. [1] bring a detailed bottom-up analysis and prognosis of the impact of covid-19 on productivity in the United Kingdom. The authors consider micro-drivers of macro productivity, use business survey results, and review many papers within the area of COVID-19 impact on productivity changes. They conclude the total factor productivity will be reduced by up to 5% in 2020 Q4 and by around 1% in 2022 and beyond.

Our paper aims to analyse the impact of structural changes in 2020 on some ratio indicators like labour productivity and average wages (wage per FTE, wage per hour). While the cited authors analyse and forecast the development of individual industries or the economy as a whole, we try to calculate the impact of changes in the economy's

¹ Prague University of Economics and Business, Department of Economic Statistics, nám. W. Churchilla 4, 130 67 Prague, fischerj@vse.cz

² Prague University of Economics and Business, Department of Economic Statistics, nám. W. Churchilla 4, 130 67 Prague, kristyna.vltavska@vse.cz

structure. As the primary method, we use index decomposition, particularly the decomposition of total indicators into the levels effect and substitution effect.

The paper is divided as follows. Chapter 2 explains preliminary data available four months after the end of the reference year 2020 and presents the methodology based on the index decomposition. Chapter 3 brings the results and their brief discussion.

2 Data and Methodology

2.1 Data

In spring 2021, about five months after the reference period, limited data sources from the national accounts are available for the last reference year. We do not have data based on annual surveys but just the preliminary data from quarterly national accounts. Using quarterly data implies limited interpretation strength of the results.

For our analysis, we use the data from the quarterly time series: supply side of GDP (structure of GDP sources by NACE activities), nominal wages and salaries and data on employment (FTE and hours worked).

As an example of original data, we present gross value added (GVA) in table 1. This indicator is close to the gross domestic product (GDP), but we consider it better for an industrial analysis as indirect taxes like VAT do not influence it. The structure of GVA and its change between 2019 and 2020 is described in table 2. Employment can be measured as a number of persons or as a number of hours worked. The number of persons is more accurate while the number of hours worked is more reliable.

Furthermore, we can use total employment (employees + self-employed) or just the employees. For the analysis of labour productivity, we use total employment (self-employed persons also contribute to the economic output like GVA). However, we consider just employees to analyse wages and salaries because self-employed persons receive mixed-income and not wages.

	A	B+C+D+E	F	G+H+I	J	K
GVA 2019 (current prices)	111,331	1,516,825	291,555	966,055	305,100	217,294
GVA 2020 (current prices)	109,212	1,473,134	306,728	894,923	320,935	210,271
GVA 2020 (previous years prices)	116,655	1,409,086	281,356	854,460	309,820	211,310
	L	M+N	O+P+Q	R+S+T+U	Total	
GVA 2019 (current prices)	483,701	387,387	799,881	110,537	5,189,666	
GVA 2020 (current prices)	495,071	361,556	864,901	102,394	5,139,125	
GVA 2020 (previous years prices)	468,581	351,600	810,673	97,188	4,910,729	

Table 1 Gross value added, mil. CZK, source: Czech Statistical Office³.

Note: A - Agriculture, forestry and fishing, B+C+D+E - Manufacturing, mining and quarrying and other industry, F - Construction, G+H+I - Trade, transportation, accommodation and food service, J - Information and communication, K - Financial and insurance activities, L - Real estate activities, M+N - Professional, scientific, technical and administrative activities, O+P+Q - Public administration, education, health and social work, R+S+T+U - Other service activities

year/industry	A	B+C+D+E	F	G+H+I	J
2019	2.1	29.2	5.6	18.6	5.9
2020	2.1	28.7	6.0	17.4	6.2
year/industry	K	L	M+N	O+P+Q	R+S+T+U
2019	4.2	9.3	7.5	15.4	2.1
2020	4.1	9.6	7.0	16.8	2.0

Table 2 Gross value added by NACE activities (%), source: Czech Statistical Office, own computation.

³ https://www.czso.cz/csu/czso/hdp_ts

2.2 Methodology

Firstly, we use the simple analysis of the contribution to growth (CTG): we can compute the direct impact of individual NACE activity i to the change in the gross value added according to the following equation:

$$CTG_i = \left(\frac{GVA_{i,t} - GVA_{i,t-1}}{\sum GVA_{i,t-1}} - 1 \right) \cdot 100 \quad (1)$$

Secondly, we can compute fundamental ratio indicators:

- labour productivity as the GVA in 2019 prices to the total number of hours worked,
- average hourly wages as a ratio of nominal wages and salaries to the hours worked of employees,
- average monthly wages as a ratio of nominal wages and salaries to the number of employees.

Finally, we will decompose the total change in ratio indicators (labour productivity, average hourly wages and average monthly wages) to the effect of changes in individual industries (levels effect) and changes in industrial structure (substitution effect). The methodology of the decomposition is described in detail by Fischer et al. [2]. The authors also bring some terminology remarks using by Lippe [5], Shorrocks [7] and others. Moreover, the decomposition of labour productivity in a particular industry is fully described in [3].

For labour productivity, the essential formulas are as follows.

Levels effect:

$$I_{LE}^{(L)} = \frac{\frac{\sum_{i=1}^n lp_{2020,i} HW_{2019,i}}{\sum_{i=1}^n HW_{2019,i}}}{\frac{\sum_{i=1}^n lp_{2019,i} HW_{2019,i}}{\sum_{i=1}^n HW_{2019,i}}} = \frac{\sum_{i=1}^n lp_{2020,i} HW_{2019,i}}{\sum_{i=1}^n lp_{2019,i} HW_{2019,i}} \quad (2)$$

Substitution effect:

$$I_{SE}^{(P)} = \frac{\frac{\sum_{i=1}^n lp_{2020,i} HW_{2020,i}}{\sum_{i=1}^n HW_{2020,i}}}{\frac{\sum_{i=1}^n lp_{2020,i} HW_{2019,i}}{\sum_{i=1}^n HW_{2019,i}}} \quad (3)$$

where

$lp_{t,i}$... the specific labour productivity for activity i in time t ,

$HW_{t,i}$... the total number of hours worked in activity i and time t .

The total change in labour productivity can be decomposed as follows:

$$\bar{I}_{lp} = I_{LE}^{(L)} \cdot I_{SE}^{(P)} \quad (4)$$

The decomposition formulas for average wages and salaries are based on the same principle.

3 Results and Discussion

Table 3 shows the contribution of industries to the decline in GVA. One can see that two industries (mining & manufacturing, trade & transportation & accommodation & food services) contribute by two thirds to the GVA decline. On the other hand, there is a positive impact of public services (0.19 p.p.), agriculture (0.09 p.p.) and ICT (0.08 p.p.). These results are critical for estimates of tax revenues. While manufacturing and trade belong to commercial industries, public services do not. Hence, we can expect that the relation between GDP and tax revenues will get worse.

Table 4 describes year-on-year changes in productivity (GVA per hours worked). The total change is +0.5 with differentiation between industries (from -2.10% in construction through +3.72% in finance and insurance to

+6.39% in agriculture). Average hourly wages (Table 5) increased by 5.70% (much higher than the labour productivity!). The hourly wages increased in all industries, from 2.51% in agriculture to 13.11% in real estate activities. Average monthly wages (Table 6) increased by 1.58%, from -3.87% in trade, transportation, accommodation and food services to +12.90% in real estate. The results comply with the economic situation.

Table 7 brings the results of the decomposition. We can explain the total increase in labour productivity (+0.5%) by the structural changes (+0.6%). The effect of changes in specific rates of productivity was marginal. The opposite effects occurred in average monthly wages: levels effect contributes by 1.3% to the total increase of 1.6%. Finally, the total change in hourly wages and salaries (+5.7%) is explained by the change in levels by +5.2%. Structural change contributes by 0.5%.

	A	B+C+D+E	F	G+H+I	J	K	L
contribution	0.09	-1.87	-0.18	-1.94	0.08	-0.10	-0.26
	M+N	O+P+Q	R+S+T+U	GVA	Taxes on products	Subsidies on products	GDP
contribution	-0.62	0.19	-0.23	-4.85	-0.70	0.05	-5.60

Table 3 Contribution of industries to variation in Gross value added

year/industry	A	B+C+D+E	F	G+H+I	J	K
2019	358	570	285	389	934	1,489
2020	381	570	279	383	917	1,545
year/industry	L	M+N	O+P+Q	R+S+T+U	Total	
2019	1,939	415	370	296	490	
2020	1,922	408	377	299	492	

Table 4 Hourly labour productivity (GVA per total employment of hours worked), CZK

year/industry	A	B+C+D+E	F	G+H+I	J	K
2019	190	241	199	206	399	426
2020	195	249	204	213	424	470
year/industry	L	M+N	O+P+Q	R+S+T+U	Total	
2019	204	244	270	191	241	
2020	231	251	295	204	255	

Table 5 Average hourly wage of employees, CZK

year/industry	A	B+C+D+E	F	G+H+I	J	K
2019	29,080	33,890	30,435	30,931	59,892	61,738
2020	29,572	33,616	30,547	29,735	63,772	65,537
year/industry	L	M+N	O+P+Q	R+S+T+U	Total	
2019	30,485	35,291	38,671	28,193	35,046	
2020	34,417	35,732	41,360	27,479	35,599	

Table 6 Average monthly wages, CZK

	Labour productivity	Monthly wages and salaries	Hour wages and salaries
Changes in specific rates (levels effect)	-0.1	1.3	5.2
Structural change (substitution effect)	0.6	0.2	0.4
Total change	0.5	1.5	5.7

Table 7 Index decomposition (%)

We present the very first flash estimate of structural changes on key economic indicators. No data from annual sources are available for this early analysis, so the interpretation strength is limited. Interpretation obstacles due to the measurement constraints occur: the short-term statistics and the quarterly national accounts are based on some crucial assumptions. One of the basic assumptions is that the stable economy's structure in a short period. The validity of this assumption is a question.

4 Conclusion

Pandemia of covid-19 brought challenges for economic analysts and the data measurement as well. This paper tries to use preliminary data for the preliminary estimates of structural changes on some ratio indicators like labour productivity and average wage.

In 2020, which was the first pandemic year, labour productivity increased by 0.5 % year-on-year, fully explained by the structural changes. Average monthly wages increased by 1.6 %, almost entirely explained by the levels changes. Average hourly wages, despite the pandemia, increase by 5.7%. Structural changes contributed to the growth by just a tenth (+0.5%) while wages within industries increased by 5.2%.

We plan to improve and refine the analysis using additional data for 2020, which we expect to become available in the following months and years. In addition, we plan to analyse some national accounts data like gross output, GDP, GNI and taxes and examine changes in their mutual relations.

References

- [1] Bloom, N. et al. (2020). The impact of COVID-19 on productivity. *NBER Working Paper Series*. NBER, Cambridge, December 2020.
- [2] Fischer, J., Flusková, H. & Vltavská, K. (2020). At-Risk-of-Poverty Rate or Social Exclusion in Visegrad Countries 2005-2017: Impact of Changes in Households' Structure. *Statistika*, 4, 351–364.
- [3] Fischer, J., Vltavská, K., Doucek, P. & Hančlová, J. (2012). The Influence of Information and Communication Technologies on Labour Productivity and Total Factor Productivity in the Czech Republic. *Politická ekonomie*, 5, 653-674.
- [4] Kotz, H. H., Mischke, J., Smit, S. (2021). Pathways for productivity and growth after the COVID-19 crisis. *VOX, CEPR Policy Portal*. May 2021. <https://voxeu.org/article/pathways-productivity-and-growth-after-covid-19-crisis>
- [5] Lippe, P. (2007). *Index Theory and Price Statistics*. Essen: Peter Lang.
- [6] Mischke, J. et al. (2021). Will productivity and growth return after the COVID-19 crisis? Special Report. McKinsey, March 2021. <https://www.mckinsey.com/industries/public-and-social-sector/our-insights/will-productivity-and-growth-return-after-the-covid-19-crisis>.
- [7] Shorrocks, A.F. (2013). Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. *J Econ Inequal*, 11, 99–126.

Productivity analysis in the Mexican food industry

Martin Flegl¹, Carlos Alberto Jiménez Bandala², Isaac Sánchez-Juárez³, Edgar Matus⁴

Abstract. Food industry represents an important part in the Mexican economy, including more than 400,000 companies and a Gross Domestic Product of around 16 billion of Mexican pesos in 2019. For that reason, this paper has the objective to analyze its productivity using data from the 2014 and 2019 Economic Censuses related to 2013 and 2018 economic indicators. The paper presents results of a productivity analysis of 1,672 municipalities from 32 Mexican states grouped in eight regions using Data Envelopment Analysis. The results indicate significant differences between regions and, also, an important growth of the productivity in 2018.

Keywords: Data Envelopment Analysis, Food Industry, Regional Development, Economic Asymmetries, Regional Polarization.

JEL Classification: C44, E23, L66

AMS Classification: 90-08, 90C05

1 Introduction

The food industry in Mexico represents 4.6% of the national economy [9]. In the last trimester of 2020, the food industry generated 4.35 billion of Mexican pesos (1 MXN is approximately .05 USD, thereafter pesos) in Gross Domestic Product (GDP), representing a growth of 5.88% compared to the same period of the previous year. As Figure 1 shows, the GDP of the Food industry has been constantly increasing during the last almost 20 years, reaching its highest value in 2019 with 16.9 billion of pesos. In 2019, the whole industry included 433,370 economic units (companies), where the highest number of the economic units were registered in Estado de México (27,070), Oaxaca (21,493) and Puebla (17,958). The biggest gross production is reported in Jalisco (2.25 billion of pesos), followed by Estado de México (1.92 billion of pesos) and Guanajuato (1,43 billion of pesos). Finally, the industry employs 1.9 million of workers (47.4% of males and 52.6% of females), with an average monthly salary of 4,370 pesos [4].

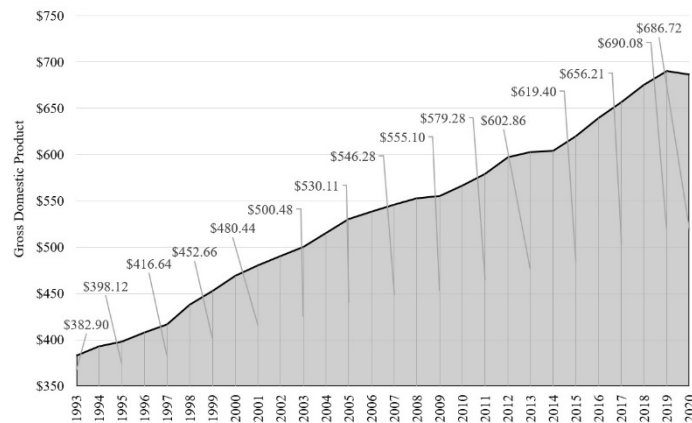


Figure 1 Evolution of the GDP in billions of pesos in the food industry in Mexico. Constant 2013 prices (own elaboration based on data from [9]).

The Mexican economy is characterized by the fact that many of the companies are micro, small and medium companies (MSMEs). The size of the companies is defined by a combination of the number of workers and total sales; MSMEs have less than 250 workers and have annual sales of less than 250 million Mexican pesos [8]. In

¹ Tecnológico de Monterrey, School of Engineering and Science, Calle Puente 222, Coapa, Arboledas del Sur, Tlalpan, 14380, Mexico City, Mexico, martin.flegl@tec.mx. ORCID: 0000-0002-9944-8475

² Universidad La Salle México, Facultad de Negocios, Benjamín Franklin 47, Col. Condesa, 06140, Mexico City, Mexico, carlos_jime-nez@lasalle.mx. ORCID: 0000-0003-4431-0054

³ Universidad Autónoma de Ciudad Juárez, Department of Social Sciences, Avenida Universidad S/N, Zona Chamizal, Ciudad Juárez, 32310, Chihuahua, Mexico, isaac.sanchez@uacj.mx. ORCID: 0000-0002-1975-5185

⁴ Universidad La Salle México, Facultad de Negocios, Benjamín Franklin 47, Col. Condesa, 06140, Mexico City, Mexico, ma-caed00@gmail.com.

the food industry, there were 420,862 (97.11%) companies with 0-10 employees, 9,312 (2.15%) companies with 11-50 employees, 1,134 (.26%) companies with 51-100 employees and 2,062 (.48%) companies with 101+ employees [4]. It should be noted that the Mexican economy is one of the most unequal in the American continent [5][6]. The North of the country represents a more developed industry, while the development in the South lags behind. This has been explained historically by a lack of investment resources in the South, but also by the existence of internal colonialist structures, in which the South transfers value to the North [10]. In this case, there are two hypotheses of such difference. The first hypothesis is linked to a duality that is defined by socio-cultural elements that suppose a low labor performance in the South and a lack of industrial productive vocations. This duality has justified the non-intervention of the State to encourage productive chains in the South of the country, which is why it has been assigned an eminent primary economic task with the intention of sending agricultural products to the North for processing [12]. The second hypothesis assumes that Southern industries being more labor intensive, transferring value to Northern industries. However, companies from the South would be just as productive as those from the North and, therefore, require support and investment to grow just like those from the North [14].

In this sense, the objective of the paper is to analyze the productivity in the Mexican food industry with an application of the Data Envelopment Analysis model. The secondary objective is to verify whether we can observe difference between Northern and Southern regions of the country. The analysis includes two periods 2014 and 2019 to observe how the productivity in the industry changed within a time. We selected the food industry because it is the branch of manufacturing that is closest to the primary sector, which is the most developed sector in less advanced regions and, therefore, is a branch with a lower demand for capital than other branches such as metallurgy, chemicals, or electronics.

2 Materials and methods

2.1 Data Envelopment Analysis

The Data envelopment Analysis (DEA) allows to evaluate several decision-making units (DMU) regarding their capabilities to convert multiple inputs into multiple outputs [3]. Each DMU can have m different input quantities to produce different outputs. If the model assumes variable returns to scale, you can use the so-called BCC model [2]. The BCC output-oriented model for DMU_o is formulated as follows:

Maximize

$$q = \sum_{r=1}^s u_r y_{ro} - u_o \quad (1)$$

subject to

$$\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} - u_o \leq 0, j = 1, 2, \dots, n,$$

$$\sum_{i=1}^m v_i x_{io} = 1, \quad (2)$$

$$\mu_r, v_i \geq \varepsilon, \varepsilon > 0, u_o \text{ free in sign.}$$

where x_{ij} is the quantity of the input i of the DMU_j , y_{rj} is the amount of the output r of the DMU_j , and μ_r and v_i are the weights of the inputs and outputs $i = 1, 2, \dots, m, j = 1, 2, \dots, n, r = 1, 2, \dots, s$ and ε is the so-called non-Archimedean element necessary to eliminate zero weights of the inputs and outputs. DMU is 100% efficient if $q = 1$, i.e., there is no other DMU that produces more outputs with the same combination of inputs, whereas DMU is inefficient if $q > 1$.

2.2 Data

For the analysis we used the economic indicators related to the Mexican food industry from the 2014 and 2019 Economic Censuses carried out by Instituto Nacional de Estadística y Geografía [7][8]. Each censuses includes information linked to manufacturing, commercial and service activities from companies operating in Mexico. The 2014 Economic Census refers to the data for 2013 and the 2019 Economic Census refers to data for 2018. In the food industry, we included the information related to the following subsectors in the food industry: agriculture-related services; preparation of animal feed; grinding grains and seeds and obtaining oils and fat; manufacture of sugars, chocolates, sweets and the like; preservation of fruits, vegetables and prepared foods; manufacture of dairy

products; slaughter, packing and processing of meat from cattle, poultry and other edible animals; preparation and packaging of fish and shellfish; preparation of bakery products and tortillas; other food industries; and branches grouped by the principle of confidentiality.

This information is linked to the Mexican municipalities as it is not possible to identify companies due to the confidentiality of the Economic Censuses. Moreover, to be able to compare the productivity between 2013 and 2018, we only included municipalities that appear in both Economic Censuses. In the end, the analysis includes 1,672 municipalities that represents 67.91% of the whole Mexico. Table 1 displays the division of the municipalities among the 32 Mexican states⁵. These 1,672 municipalities include information from 164,558 economic units in 2013 and from 189,590 economic units in 2018. Moreover, the results from both Censuses are comparable as we used constant prices of 2018 for the 2014 Economic Census.

State	# of municipalities	State	# of municipalities
Aguascalientes	9 (11) – 81.82%	Morelos	28 (33) – 84.85%
Baja California	5 (5) – 100.00%	Nayarit	19 (20) – 95.00%
Baja California Sur	5 (5) – 100.00%	Nuevo León	31 (51) – 60.78%
Campeche	11 (11) – 100.00%	Oaxaca	218 (570) – 38.25%
Chiapas	74 (122) – 60.66%	Puebla	156 (217) – 71.89%
Chihuahua	34 (67) – 50.75%	Querétaro	16 (18) – 88.89%
Ciudad de México	16 (16) – 100.00%	Quintana Roo	7 (11) – 63.64%
Coahuila	25 (38) – 65.79%	San Luis Potosí	42 (58) – 72.41%
Colima	10 (10) – 100.00%	Sinaloa	16 (18) – 88.89%
Durango	30 (39) – 76.92%	Sonora	33 (72) – 45.83%
Estado de México	114 (125) – 91.20%	Tabasco	17 (17) – 100.00%
Guanajuato	41 (46) – 89.13%	Tamaulipas	26 (43) – 60.47%
Guerrero	67 (81) – 82.72%	Tlaxcala	49 (60) – 81.67%
Hidalgo	67 (84) – 79.76%	Veracruz	159 (212) – 75.00%
Jalisco	108 (125) – 86.40%	Yucatán	75 (106) – 70.75%
Michoacán	104 (113) – 92.04%	Zacatecas	42 (58) – 72.41%

Table 1 Division of the municipalities among the Mexican states.

2.3 Structure of the model

The input part of the DEA model summarizes the resources of each municipality in the food industry:

- *Personnel*: Hours worked by the personnel in thousands of hours (HWP). This variable represents the labor factor of the production, therefore, a greater number of hours worked by the personnel for the same level of production would indicate lower productivity.
- *Material*: Raw materials and materials in millions of pesos (RMM), Number of economic units (NEU). These variables indicate the material inputs of the production necessary for the transformation. Higher productivity is associated with less use of materials and economic units.
- *Finance*: Total expenditures in millions of pesos (TE), Total personnel remunerations in millions of pesos (TPR). These variables represent the salary expenses of the industry and all expenses used in the production. Therefore, it is implicit that the higher the expenditure with the same level of production, the lower the productivity is.

The output part of the DEA model includes:

- Total gross production in millions of pesos (TGP). This variable measures the economic results of each municipality in terms of volume.

The selection of the inputs and outputs follows the common structure of DEA models in the agricultural productivity analysis [1][11][15]. We used the BCC output-oriented model as the intention is to analyze the productivity level of each municipality related to their economic results. The BCC model is used as we consider a direct competition in the food industry. Finally, we used MaxDEA 7 Ultra software for all the calculations. The importance of the inputs and outputs with $\varepsilon = .5$, which best balanced the model, is as follows: HWP 2.59%, RNM 6.67%, NEU 3.96%, TE 83.60%, TPR 3.18%, and 100% in case of the 2014 model, and HWP 4.46%, RNM 19.11%, NEU 5.74%, TE 63.69%, TPR 6.99%, and TGP 100% in case of the 2019 model.

⁵ Mexico is divided into 2446 municipalities and Mexico City (Ciudad de México) is divided into 16 parts.

3 Results

The average productivity of the Mexican municipalities in 2013 is .3498 with standard deviation of .156 and 30 municipalities reached the productivity of 1.0 (representing 1.79% of the analyzed sample). This result indicates very low productivity in the food industry in Mexico. As Figure 2 in the Appendix shows, we cannot identify a region (state) with very high productivity. The municipalities with the 1.0 productivity are placed across Mexico. To understand a little bit more the obtained results, we divide the municipalities according to their geographical dependence⁶. Table 2 shows that the highest average productivity in the food industry in 2013 is reported in the Southeast region (.4003), one of the biggest considering the number of municipalities, but it is also a region with the highest variability measured by the standard deviation (.1895). What is more, the Southeast region is the only one evaluated above the country average. On the other hand, the West region reported the lowest average productivity in Mexico (.3179) with the lowest variability (.1090). Applying the Games-Howell test⁷, the differences in productivity between the regions are statistically significant ($p < .001$). More specifically, the average productivity of the municipalities in the Southeast region is statistically higher compared to the rest of the regions (considering the confidence level of 99%). The rest of the differences are not statistically significant, except the difference between the East region and the Center North ($p = .083$) and West ($p = .012$) regions.

Regions	Mean	Std. Deviation	N
Center North	.3238	.1464	150
Center South	.3324	.1305	158
East	.3490	.1528	431
Northeast	.3248	.1390	82
Northwest	.3360	.1540	123
Southeast	.4003	.1895	377
Southwest	.3440	.1475	110
West	.3179	.1090	241
Average	.3498	.1557	1,672

Table 2 Average productivity by geographical regions in 2013

In 2018, the average productivity of the Mexican municipalities increased by +.2343 up to .5841 with a standard deviation of .1163. In this case, 36 municipalities reached the 1.0 productivity (representing 2.15% of the analyzed sample). As Figure 3 in the Appendix illustrates the improvement in the productivity of the industry can be seen all around the country, which resulted that the difference between 2013 and 2018 is statistically significant ($p < .001$). The best evaluated region is now Northeast (.6342) whose average productivity increased by +.3094, followed by the Northwest region (.6223, +.2863). The Southeast region that was evaluated as the best region in 2013 is evaluated as the 3rd worst region in 2018, with the average productivity of .5729, because its productivity improved in the smallest proportion (+.1726). The worst evaluated region is the East region with the average productivity of .5586, with an improvement of +.2096 compared to 2013. Five out of the eight regions are evaluated above the Mexican average. The Games-Howell test indicates statistically significant differences between the regions, for example the Northeast and Northwest regions compared to the rest of the regions ($p < .001$), and West and Center South regions compared to the East, Southeast and Southwest regions (confidence level of 95%) in almost all cases (only few exceptions can be observed).

The changes in the productivity of the periods can be explained by the economic structure of Mexico itself. The Southern regions, eminently agricultural, send their largest production to the Northern regions for processing. In 2013-2014, the international oil prices increased the gasoline prices and, as the major transportation of goods in Mexico is done by roads, it is a reason why the Northern regions, further away from agricultural production, were less productive than those in the South closer to the agricultural centers. With the above we can point out that developing industrial centers, particularly food centers, in agricultural areas would have positive results in productivity. This, together with an active industrial development policy, would have a positive impact on regional economic development and would combat regional asymmetries [13].

⁶ Mexico is divided into eight geographical regions: Northwest (Baja California, Baja California Sur, Chihuahua, Durango, Sinaloa and Sonora), Northeast (Coahuila, Nuevo León and Tamaulipas), West (Colima, Jalisco, Michoacán and Nayarit), East (Hidalgo, Puebla, Tlaxcala and Veracruz), Center North (Aguascalientes, Guanajuato, Querétaro, San Luis Potosí and Zacatecas), Center South (Ciudad de México, Estado de México and Morelos), Southeast (Chiapas, Guerrero and Oaxaca) and Southwest (Campeche, Quintana Roo, Tabasco and Yucatán).

⁷ Games-Howell is a nonparametric test that does not assume equal variances and the same sample size. In our case, there are significant differences regarding the number of municipalities between the regions and the variances are different (Levene's test $p < .001$). All the statistical tests presented in this article are based on this test.

Regions	Mean	Std. Deviation	N	Difference 2018-2013
Center North	.5880	.1022	150	+.2642
Center South	.6003	.0877	158	+.2679
East	.5586	.1038	431	+.2096
Northeast	.6342	.1318	82	+.3094
Northwest	.6223	.1231	123	+.2863
Southeast	.5729	.1410	377	+.1726
Southwest	.5701	.1105	110	+.2261
West	.6038	.0974	241	+.2859
Average	.5841	.1163	1,672	+.2343

Table 3 Average productivity by geographical regions in 2018

4 Conclusions

The objective of the paper was to analyze the productivity in the Mexican food industry using data from the 2014 and 2019 Economic Censuses. The results revealed that in 2013, the highest productivity was reported in the municipalities of the South regions of Mexico, which did not confirm the historical observation of a lower development of these regions in Mexico. However, this could have been caused by the rise of the 2013-2014 international oil prices that negatively affected the transportation of goods in Mexico. Regarding the obtained results in 2018, we can observe significant improvements in the productivity in the food industry across the whole country. But the improvements were higher in the municipalities of the Northern regions of Mexico, which correspond to the historical development of Mexican economy, where the Northern regions concentrate more developed industry. The further research could extend the analysis including data from level of investments in the industry. This piece of information would explain whether higher productivity leads to higher investments in the Mexican food industry.

5 Acknowledgements

This research was carried out within the framework of the project “Patterns of success and failure in the economic evolution of businesses identified from data mining and artificial neural networks”, A3-S-129311, of the CONACYT-INEGI Sector Fund.

6 References

- [1] Arita, S., and Leung, P.S.: A Technical Efficiency Analysis of Hawaii's Aquaculture Industry. *Journal of the World Aquaculture Society* **45(3)** (2014), 312-321. <https://dx.doi.org/10.1111/jwas.12124>
- [2] Banker, R. D., Charnes, A., and Cooper, W. W.: Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science* **30(9)** (1984), 1078-1092. <http://dx.doi.org/10.1287/mnsc.30.9.1078>
- [3] Cooper, W., Seiford, L., and Zhu, J.: *Handbook on data envelopment analysis*. Nueva York: Springer, 2011.
- [4] DataMÉXICO: *Industria alimentaria [Food industry]*. Secretaría de Economía [Ministry of Economy], 2021, [Online], available: <https://datamexico.org/es/profile/industry/industria-alimentaria?yearSelectorGdp=timeOption0> [1 June 2021].
- [5] Dávila, E., Kessel, G., and Levy, S.: El sur también existe: un ensayo sobre el desarrollo regional de México. *Economía Mexicana. Nueva Época* **11(2)** (2002), 205-260.
- [6] García-Almada, R.: *Liberalización comercial, descentralización territorial y polarización económica en México*. Ciudad Juárez: Universidad Autónoma de Ciudad Juárez, 2012.
- [7] INEGI: *México - Censos Económicos 2014*. Instituto Nacional de Estadística y Geografía [National Institute of Statistics and Geography], 2014, [online], available: <https://www.inegi.org.mx/programas/ce/2014/> [1 June 2021].
- [8] INEGI: *México - Censos Económicos 2019*. Instituto Nacional de Estadística y Geografía [National Institute of Statistics and Geography], 2019, [online], available: <https://www.inegi.org.mx/rnm/index.php/catalog/547> [1 June 2021].
- [9] INEGI: *México – Sistema de Cuentas Nacionales*. Instituto Nacional de Estadística y Geografía [National Institute of Statistics and Geography], 2020, [online], available: <https://www.inegi.org.mx/sistemas/bic/> [1 June 2021].

- [10] Jiménez-Bandala, C. A.: Development in Southern Mexico: Empirical Verification of the “Seven Erroneous Theses about Latin America”. *Latin American Perspectives* **45(2)** (2018), 129-141. <https://dx.doi.org/10.1177/0094582X17736036>
- [11] Marcikić Horvat, A., Matkovski, B., Zekić, S., and Radovanov, B.: Technical efficiency of agriculture in Western Balkan countries undergoing the process of EU integration. *Agricultural Economics* **66(2)** (2020), 65-73. <https://doi.org/10.17221/224/2019-AGRICECON>
- [12] Myrdal, G.: *Economic Theory and Under-developed Regions*. London: Gerald Duckworth, 1957.
- [13] Revilla, D., García-Andrés, A., and Sánchez-Juárez, I.: Identification of key productive sectors in the Mexican Economy. *Expert Journal of Economics* **3(1)** (2015), 22-39.
- [14] Stavenhagen, R.: Seven erroneous theses about Latin America. *New University Thought* **4(4)** (1965), 25–37.
- [15] Toma, E., Dobre, C., Dona, I., and Cofas, E.: DEA Applicability in Assessment of Agriculture Efficiency on Areas with Similar Geographically Patterns. *Agriculture and Agricultural Science Procedia* **6** (2015), 704-711. <https://doi.org/10.1016/j.aaspro.2015.08.127>

7 Appendix

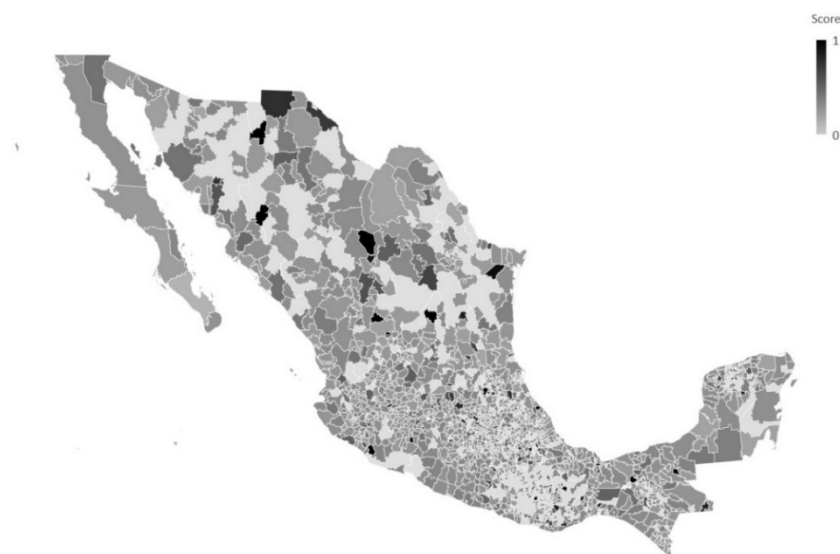


Figure 2 Productivity in the food industry by municipalities in 2013 (own elaboration using GeoNames, Microsoft tool).



Figure 3 Productivity in the food industry by municipalities in 2018 (own elaboration using GeoNames, Microsoft tool).

Evaluation and testing of non-nested specifications of spatial econometric models

Tomáš Formánek¹

Abstract. Spatial econometric models have generally non-nested specifications if they are based on different spatial setups (connectivity and weight matrices). For maximum likelihood estimators and non-nested models, the usual tests (likelihood ratio, Wald, et.c) generally cannot be used for model selection and/or testing. This article provides a structured discussion on estimation and evaluation (selection and/or testing) of non-nested spatial models. The distinction between model selection and model testing is important. While model selection algorithms approach all models symmetrically, the null and alternative models are treated differently for testing purposes.

The empirical part of this paper provides an illustrative application of the evaluation methods discussed. Emphasis is given to models estimated by maximum likelihood approaches and to Vuong's test, which is derived from the Kullback–Leibler information criterion.

Keywords: spatial model, model selection, non-nested models.

JEL Classification: C23, C31, C52, E66

AMS Classification: 91B72

1 Introduction

Spatial econometrics models address the existence of spatial dependency in observed data, as well as some general and theoretically defined interactions among variables. Should spatial effects be ignored or left without proper treatment, model estimation would lead to biased and inconsistent results. Typically, one would estimate parameters for a relevant and explicitly defined (prior) spatial dependency pattern – along with the “usual” model coefficients that describe macroeconomic dependencies [1].

In principle, spatial dependency is quite similar to autoregressive processes in time series. However, spatial dependency is neither “one-dimensional” (on a time axis) nor unidirectional (current observations being dependent on past values). In most economic applications, spatial units are interdependent and the strength of their interactions is defined in terms of distances, rather than oriented distances (although analyses with focus on core-periphery behavior and similar topics exist, where oriented distances are relevant).

Generally speaking, spatial dependency is often described as a non-continuous function, decreasing with distance between spatial units. The possibilities of measuring distances and distinguishing neighbors (close, interacting, spatially dependent units) from distant (non-interacting units) are numerous, with a multitude of possible approaches and their combinations. For example, one may evaluate distances given as shortest connections between two nearest point of given regions (i.e. polygons on a map) or use distances between geographical center-points. Similarly, either L_1 or L_2 norms may be used for measurements and neighbors can be cast through the application of a common border rule (contiguity-based neighborhood definition), etc. Nevertheless, the true pattern (functional form) of spatial dependency is not known in most empirical cases.

Hence, researches would typically have to consider several spatial structures (neighborhood definitions) in order to evaluate and select a “good” spatial setup for their analyses and/or predictions. Unfortunately, if two or more spatial econometric models are based on different spatial setups, they are generally non-nested. For the widely used likelihood-based estimators, the non-nested nature of spatial models based on alternative neighborhood definitions implies that most common statistics cannot be used for model selection and/or testing [8, 9]. This article provides a structured discussion on the specifics of estimation and testing of non-nested spatial models.

The remainder of this paper is structured as follows: Section two describes key features and aspects of spatial model estimation and testing, along with references to fundamental literature. Section three provides an illustrative application based on unemployment dynamics in selected EU countries for the time period 2014 – 2019. Section four and the list of references conclude this contribution.

¹ VŠE Praha, nám. W. Churchilla 4 Praha 3, Prague 130 67, Czech Republic, formanek@vse.cz

2 Estimation, evaluation and testing of spatial models

To formalize and estimate spatial models, researches start by defining the underlying spatial structure. To classify any two spatial units as either neighboring or distant, one would typically use a connectivity matrix \mathbf{C} . For example, individual units of the connectivity matrix may be cast as follows:

$$[c_{ij}] = \begin{cases} 0 & \text{if } i = j, \\ 0 & \text{if } d_{ij} > \tau, \\ 1 & \text{if } d_{ij} \leq \tau \text{ and } i \neq j, \end{cases} \quad (1)$$

where $d_{ij} = d_{ji}$ is the distance between two spatial units. For $c_{ij} = 1$, units i and j are neighbors, i.e. are sufficiently close to interact and influence each other mutually, and vice versa. Here, symmetry of the $(N \times N)$ matrix \mathbf{C} , (i.e. $c_{ij} = c_{ji}$) is implied from the use of distances d_{ij} . Zeros on the main diagonal of \mathbf{C} indicate that spatial units are not neighbors to themselves by definition. Parameter τ is a heuristically selected maximum distance (threshold) between two neighbors. The choice and changes in τ can have significant impacts on the results of spatial model estimators, which use spatial structure as prior information. As we deal with spatial units of non-zero surface area, we usually use representative points – centroids – to measure distances between regions. Alternatively, expression (1) can be evaluated using closest border-to-border aerial distances, based on transport infrastructure, etc.

Spatial econometric models are based on a transformation of the connectivity matrix \mathbf{C} from (1): a spatial weights matrix \mathbf{W} is calculated by row-wise standardization of \mathbf{C} . The transformation rule is relatively simple, each element of \mathbf{W} is calculated as $w_{ij} = c_{ij} / \sum_{j=1}^N c_{ij}$ so that all row sums in \mathbf{W} equal one: $\sum_{j=1}^N w_{ij} = 1$ for $\forall i$. Please note that \mathbf{W} matrices are no longer symmetric. Besides the maximum neighbor distances as outlined in expression (1), different plausible approaches exist for construction of both matrices – \mathbf{C} (common border rule, k -nearest neighbors approach, etc.) and \mathbf{W} (multiple standardization schemes can be applied). Detailed technical discussion of this topic is provided by [1].

As we gather available prior information for the construction of a spatial regression model, the actual neighborhood structure is largely unknown. While there may be quite a few useful leads that analysts can use, the choice of τ -parameter in (1) is arbitrary and different approaches and combinations of \mathbf{C} and \mathbf{W} construction are possible.

LeSage and Pace [5] show that theoretically well-defined models can be estimated and tested with adequate precision, even if the spatial structure used for estimation is somewhat inaccurate (i.e. if it slightly deviates from the actual yet unknown neighborhood definition). To summarize their findings, one can start from a cross sectional model with spatially dependent endogenous variable:

$$\mathbf{y} = \lambda \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{u}, \quad (2)$$

where \mathbf{y} is the $N \times 1$ dependent variable vector, \mathbf{X} is the usual $N \times K$ matrix of regressors (includes the intercept element), \mathbf{u} is the random element and \mathbf{W} is the spatial weights matrix. Model parameter λ is a scalar describing the strength of spatial dependency and $\boldsymbol{\beta}$ is a vector of parameters (say, describing economic dynamics). Besides the commonly used specification (2) with spatial dependency on \mathbf{y} , other types of spatial models exist – both in literature and in empirical applications [1, 5]. In most types of spatial models, the generalization from cross-sectional to panel data is relatively straight forward, very similar to the case of non-spatial models. Finally, it should be noted here that $\boldsymbol{\beta}$ -parameters are not the marginal effects. However, using \mathbf{W} and the estimates of λ and $\boldsymbol{\beta}$, direct and spillovers marginal effects can be calculated easily [3].

Under very general conditions [3], the maximum likelihood (ML) estimator may be used to produce estimates of all parameters of model (2): $\boldsymbol{\beta}$, λ and random element variance σ^2 . Assuming normal distribution of the error elements, ML function for equation (2) can be cast as

$$LL(\boldsymbol{\theta}) = -\frac{N}{2} \log(2\pi\sigma^2) + \log |\mathbf{I}_N - \lambda \mathbf{W}| - \frac{1}{2\sigma^2} \mathbf{u}' \mathbf{u}, \quad (3)$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda, \sigma^2)$, $\mathbf{u} = \mathbf{y} - \lambda \mathbf{W} \mathbf{y} - \mathbf{X} \boldsymbol{\beta}$ and $\det(\partial \mathbf{u} / \partial \mathbf{y}) = |\mathbf{I}_N - \lambda \mathbf{W}|$ is the Jacobian. Using eigenvalues κ of matrix \mathbf{W} , the condition $\lambda \in (\min(\kappa)^{-1}, \max(\kappa)^{-1})$ must be fulfilled to ensure model stability [3]. In theory, maximization of (3) is based on first order conditions

$$\partial LL / \partial \boldsymbol{\beta} = \mathbf{X}' (\mathbf{I}_N - \lambda \mathbf{W}) \mathbf{y} - \mathbf{X}' \mathbf{X} \boldsymbol{\beta} = \mathbf{0}, \quad (4)$$

$$\partial LL / \partial \lambda = -\text{tr} [(\mathbf{I}_N - \lambda \mathbf{W})^{-1}] \mathbf{W} - \frac{1}{2\sigma^2} \mathbf{u}' \mathbf{W} \mathbf{y} = 0, \quad (5)$$

$$\partial LL / \partial \sigma^2 = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \mathbf{u}' \mathbf{u} = 0. \quad (6)$$

The equation system (4) – (6) has no convenient analytical solution. However, an accessible estimation algorithm can be used to produce coefficient estimates – shown next in expressions (8) through (14). Regularity conditions holding, ML estimate is asymptotically efficient and reaches the Cramer–Rao lower bound for variance, given by the inverse of the information matrix:

$$[\mathbf{I}(\boldsymbol{\theta})]^{-1} = -E \left[\partial^2 LL(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' \right]^{-1} . \quad (7)$$

To estimate the parameters of equation (2), we start by OLS estimation of auxiliary regressions:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{u}_0 , \quad (8)$$

$$\mathbf{W}\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_d + \mathbf{u}_d , \quad (9)$$

so that

$$\hat{\boldsymbol{\beta}}_0 = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} ,$$

$$\hat{\boldsymbol{\beta}}_d = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\mathbf{y} ,$$

$$\mathbf{e}_0 = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0 ,$$

$$\mathbf{e}_d = \mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_d ,$$

where $\hat{\boldsymbol{\beta}}$ and \mathbf{e} denote estimated parameters and residuals respectively. By means of the auxiliary regressions (8) and (9), we can re-write the residuals of model (2) as

$$\mathbf{e}(\lambda) = \mathbf{e}_0 - \lambda \mathbf{e}_d . \quad (10)$$

Using $\mathbf{e}(\lambda)' \mathbf{e}(\lambda) = \mathbf{e}_0' \mathbf{e}_0 - 2\lambda \mathbf{e}_0' \mathbf{e}_d + \lambda^2 \mathbf{e}_d' \mathbf{e}_d$, we can formulate a concentrated version of the log-likelihood function (3) that only depends on the λ parameter:

$$LL(\lambda) = c + \log |\mathbf{I}_N - \lambda \mathbf{W}| - \frac{N}{2} \log [\mathbf{e}(\lambda)' \mathbf{e}(\lambda)] , \quad (11)$$

where c is a constant, independent of λ . To maximize the ML function (11), the interval $\lambda \in (\min(\kappa)^{-1}, \max(\kappa)^{-1})$ is split into multiple “tiny” intervals along $\lambda_1 < \lambda_2 < \dots < \lambda_q$ values and we evaluate (11) at each λ . Arbitrarily accurate λ estimates that maximize (11) can be produced by additional interval splitting and (11) evaluation. Subsequently, $\hat{\lambda}$ is used to produce the remaining parameters of (2) as follows:

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_0 - \hat{\lambda} \hat{\boldsymbol{\beta}}_d , \quad (12)$$

$$\hat{\sigma}^2 = N^{-1} \mathbf{e}(\hat{\lambda})' \mathbf{e}(\hat{\lambda}) , \quad (13)$$

$$\text{var}(\hat{\boldsymbol{\mu}}) = \hat{\boldsymbol{\Omega}} = \hat{\sigma}^2 [(\mathbf{I}_N - \hat{\lambda} \mathbf{W})' (\mathbf{I}_N - \hat{\lambda} \mathbf{W})]^{-1} . \quad (14)$$

Besides the approach described here, technical discussion and references to alternative estimators for different model generalizations are available from [1, 3].

2.1 Model evaluation and testing

Figure 1 illustrates a situation where we have two alternative spatial structures \mathbf{C} that are (generally) plausible, yet distinct. If such structures are used to cast spatial models in the general form of equation (2), they would be non-nested, which follows from the nature of the RHS element $\lambda \mathbf{W}\mathbf{y}$ – even if the $\mathbf{X}\boldsymbol{\beta}$ element of (2) does not change. After estimating such model specifications, the next logical step would consist of focusing on evaluation and comparison of such econometric models.

Evaluation and comparison of models can be approached using two distinct paradigms: model selection and models testing. Those are two relevant yet conceptually different tasks. Generally speaking, model selection treats all models symmetrically, while testing approaches the null and alternative models differently. When model selection is performed, a definitive output is generated – a model is selected. However, hypothesis testing does not seek nor does it provide such outcome: rejecting the null hypothesis (model specification) does not imply acceptance of the alternative specification. Importantly, the choice of null hypothesis may be arbitrary and it may also greatly affect the interpretation of test outcome [8].

Also, there is a firm empirical motivation for the distinction between model selection and hypothesis/model testing. The selection paradigm is more pertinent as a decision making tool. Hypothesis/model testing is mostly used for inferential applications (e.g. for assessing validity of theoretically determined predictions).

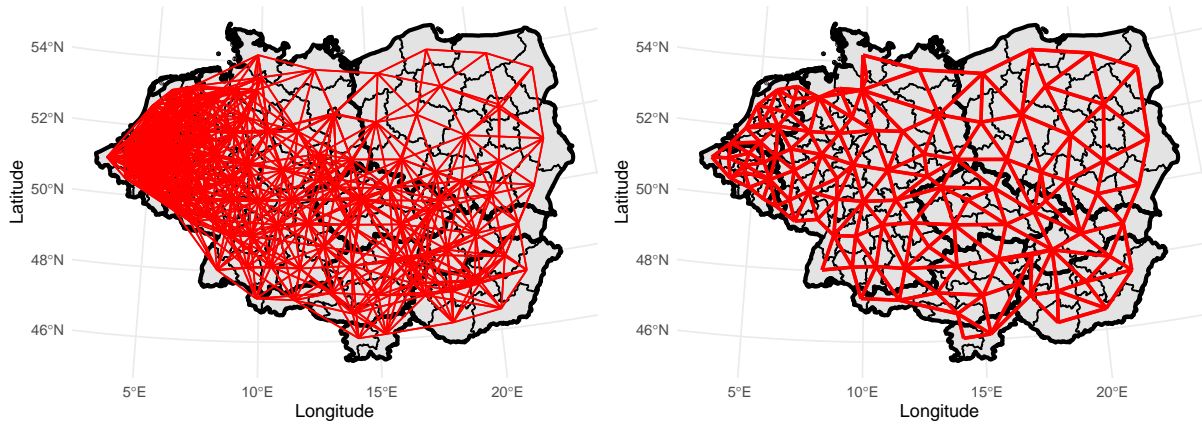


Figure 1 Alternative neighborhood structures used for model estimation: distance-based with maximum distance between neighbors set at 250 km (left) and contiguity-based (right)

Both the hypothesis testing and model selection approaches are valid and relevant. Pesaran [8] points out that model selection algorithms are mainly based on statistics of model fit to the data (maximized log-likelihoods, minimized information criteria or sum of squared residuals). The empirical (illustrative) section of this article features both model evaluation approaches: Vuong's test (16) is used for model testing and model selection is based on maximized log-likelihood values.

In general, various statistics can be used for comparing non-nested models. The J-test, JA-test, N-test and NT-test [8] can be listed among the most important and widely used statistics for non-nested models. However, those are suitable for OLS-estimated models. Their application extends to the instrumental variable regression but not towards ML-based estimators.

Vuong [9] provides a feasible method (statistic) for testing non-nested ML-estimated spatial models. His approach is based on the Kullback–Leibler information criterion (KLIC). The KLIC criterion reflects a maximized log-likelihood value difference between a ML estimator applied to a misspecified model, say $g(y|\mathbf{Z}, \boldsymbol{\beta})$ and to a true model $h(y|\mathbf{Z}, \boldsymbol{\alpha})$. Formally, KLIC can be cast as:

$$\text{KLIC} = E [\log h(y_i|z_i, \boldsymbol{\alpha})|h \text{ is true}] - E [\log g(y_i|z_i, \boldsymbol{\beta})|h \text{ is true}], \quad (15)$$

where z_i is a full set of regressors, $i = 1, \dots, N$ is used to identify individual observations, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are model parameters. Even if the true model specification $h(\cdot)$ is unknown, KLIC can be used for testing. Using Vuong's test (16), we can compare two alternative functions – say g_0 and g_1 – and determine whether they are equivalent or whether one of the functions is closer to the true (unobserved) specification. If expression (15) is produced for both g_0 and g_1 , then taking the difference of KLIC for the two g_1 and g_0 functions effectively eliminates the likelihood function of the true specification h . Hence, Vuong's test statistic V can be expressed as

$$V = \frac{\sqrt{N} \left(\frac{1}{N} \sum_{i=1}^N m_i \right)}{\sqrt{\frac{1}{N} \sum_{i=1}^N (m_i - \bar{m})^2}} = \sqrt{N} (\bar{m}/s_m), \quad m_i = \log L_{i,1} - \log L_{i,0}, \quad (16)$$

where $L_{i,1}$ and $L_{i,0}$ are likelihood functions of g_1 and g_0 evaluated at a given observation i . Elements \bar{m} and s_m refer to sample means and standard deviation of m_i . Under the null hypothesis of both tested models being equally good (i.e. equally distant from the true model h), V asymptotically follows standard Normal distribution. Interestingly, Vuong's test has a directional interpretation: If g_1 is substantially better than g_0 (i.e. closer to h), V diverges and $\text{plim } V = +\infty$ (and vice versa). Additional technical discussion covering (among other topics) nested and partially nested g_1 and g_0 specifications is provided by [8, 9].

3 Empirical illustration: unemployment dynamics

The above discussion is illustrated using a spatial econometric model focused on labor market dynamics. For estimation purposes, panel data were collected for the following economies: Austria, Belgium, Czechia, Denmark, Germany, Hungary, Luxembourg, the Netherlands, Poland, Slovakia and Slovenia (110 NUTS2 regions in total), with annual observations covering the period 2014 – 2019. All variables and geographic information was downloaded from Eurostat into R software for processing and estimation [2, 7]. Unemployment is modelled using the regional competitiveness paradigm [4]. Hence, unemployment is given as a function of GDP growth, industrial production prominence on the labor market, energy consumption and country-level effects (besides the NUTS2 unobserved effects that are intrinsic to panel data models).

The generalization from a cross-sectional model (2) to a panel data is discussed in detail by [3] and the regression equation used for modelling unemployment dynamics may be cast as follows:

$$\begin{aligned} \mathbf{y} &= \lambda (\mathbf{I}_T \otimes \mathbf{W}) \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} , \\ \mathbf{u} &= (\boldsymbol{\iota}_T \otimes \mathbf{I}_N) \boldsymbol{\mu} + \boldsymbol{\varepsilon} \end{aligned} \quad (17)$$

where \mathbf{y} is the vector of NT observations of the dependent variable and \mathbf{X} is a $NT \times K$ matrix of regressors. Elements \mathbf{I}_T and \mathbf{I}_N are identity matrices with dimensions given by their subscripts, $\boldsymbol{\iota}_T$ is a vector of ones and \otimes denotes Kronecker product. The elements \mathbf{W} , λ and $\boldsymbol{\beta}$ follow from equation (2), $\boldsymbol{\mu}$ are the unobserved individual effects and $\boldsymbol{\varepsilon}$ is a spatially random error term of the model.

Variables for our illustrative application are as follows: individual observations y_{it} of the dependent variable reflect unemployment rates in percentage points (drawn from Eurostat’s “lfst_r_lfu3rt” dataset) and matrix \mathbf{X} contains the following regressors: $\log(GDP_{it})$ is the log-transformed real GDP per capita (Eurostat’s “nama_10r_2gdp” is used for GPD and “ei_cphi_m” for inflation adjustment), $RE_B_E_{it}$ is the relative employment in industrial sectors (NACE rev.2 sectors B to E, drawn from “lfst_r_lfe2en2” dataset), variable $\log(Engy_{it})$ is the log-transformed total energy consumption (measurements available only at the NUTS0-level) and $\mathbf{NUTS0}_i$ is a vector of country-specific dummy variables (Austria is left out and serves as a reference/base unit).

Table 1 Comparison of model estimates for alternative spatial structures

Impact / λ	Estimate	Std. Error (simulated)	z -value (simulated)	Pr ($> z $) (simulated)
Distance-based spatial structure used, $\tau = 250$ km:				
$\log GDP_Direct_Imp$	-3.164	0.376	-9.597	0.000
$\log GDP_Indirect_Imp$	-10.381	2.609	-4.100	0.000
$RE_B_E_Direct_Imp$	-11.519	2.444	-4.697	0.000
$RE_B_E_Indirect_Imp$	-33.078	10.514	-3.229	0.001
λ	0.776	0.031	25.365	0.000
<i>Log likelihood (LL)</i> -863.908				
Contiguity-based spatial structure				
$\log GDP_Direct_Imp$	-4.360	0.401	-10.864	0.000
$\log GDP_Indirect_Imp$	-7.206	1.165	-6.210	0.000
$RE_B_E_Direct_Imp$	-9.555	2.693	-3.539	0.000
$RE_B_E_Indirect_Imp$	-15.791	5.011	-3.164	0.002
λ	0.670	0.030	22.626	0.000
<i>Log likelihood (LL)</i> -891.724				

The model was estimated in R by means of the `sp1m` package [6] and using two different spatial setups, as shown in Figure 1. The first spatial structure shown in Figure 1 (left) follows from expression (1) with maximum distances among neighbors (interacting units) set at $\tau = 250$ km. The alternative spatial structure is constructed along contiguity (common border) rules – two regions are considered as neighbours if they share a common border. As discussed in previous sections, spatial models with two different spatial structures – and \mathbf{W} matrices – are non-nested.

Table 1 shows estimation results for the two non-nested spatial panel models based on equation (17) and differing in their spatial prior information only. Table 1 includes the estimated λ coefficients, along with direct and indirect

(spillover) effects of the main regressors observed at NUTS2 level. Due to space constraints for this contribution and given their limited interpretation, state-level (NUTS0) regressors are omitted from Table 1. However, all the estimation output is available from the author upon request, along with corresponding R code and data.

From Table 1, we may clearly identify prominent differences among the estimated marginal effects. By simply comparing the maximized log-likelihood values, one would slightly prefer and select the specification that features distance-based neighbors. However, Vuong’s test based on expression (16) clearly favours the contiguity based structure, with $V = -30.435$ if contiguity based model is used as the base specification. Reversing the null and alternative “roles” for the two spatial setups just flips the sign of Vuong’s test with no consequence in terms of test interpretation. Based on the Vuong’s test, we conclude that contiguity-based setup is significantly closer to the true specification (this conclusion holds at any reasonable significance level).

4 Conclusions

In most empirical applications of spatial econometrics, spatial prior information is used to distinguish close units (neighbors) from distant units that are independent. However, spatial econometric models always face uncertainty with respect to the true yet unobservable spatial structure.

This article provides a structured and relatively simple approach towards estimation and testing of non-nested spatial models that are based on alternative spatial structures. The discussion provided covers both model estimation methodology and Vuong’s test for non-nested specifications, derived from the Kullback–Leibler information criterion.

Acknowledgements

Institutional research support provided by Faculty of Informatics and Statistics, University of Economics, Prague. Geo-data source: GISCO-Eurostat (European Commission), Administrative boundaries: © EuroGeographics.

References

- [1] Anselin, L. (1988). *Spatial econometrics: methods and models*. Dordrecht: Kluwer.
- [2] Bivand, R. & Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, **63**(18), 1–36.
- [3] Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-sectional Data to Spatial Panels*. New York: Springer.
- [4] Formánek, T., and Hušek, R. (2016). On the stability of spatial econometric models: Application to the Czech Republic and its neighbors. In A. Kocourek & M. Vavroušek (Eds.), *Mathematical Methods in Economics* (pp 213–218). Liberec: TU Liberec.
- [5] LeSage, J. P., & Pace, R. K. (2014). The biggest myth in spatial econometrics. *Econometrics*, **2**(4), 217–249.
- [6] Millo, G., & Piras, G. (2012). splm: Spatial panel data models in R. *Journal of Statistical Software* **47**(1), 1–38.
- [7] Pebesma, E. (2018). Simple features for R: standardized support for spatial vector data. *The R Journal*, **10**(1), 439–446.
- [8] Pesaran, M.H.: *Time Series and Panel Data Econometrics*. Oxford University Press, Oxford, 2015.
- [9] Vuong, Q.H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, **57**(2), 307–333.

The link between DEA efficiency and return to assets

Lukáš Frýd¹, Ondřej Sokol²

Abstract. The data envelopment analysis (DEA) is a standard tool in the analysis of determinants of the efficiency of economic agents. A two-stage efficiency analysis consisting of estimating DEA efficiency and then using it as dependent variable in regression with various determinants, is often used for this purpose. In this article, we focus on the relevance of this approach using empirical agricultural data. In particular, we show that the lagged DEA efficiency of agricultural companies is not correlated with return on assets indicator, and, similarly, lagged return on assets are not correlated with DEA efficiency. Low cross-correlation values indicate that a two-stage analysis using the DEA method can lead to misleading results.

Keywords: data envelopment analysis, agricultural, cross-correlation

JEL Classification: C50

AMS Classification: 90C90

1 Introduction

The data envelopment analysis (DEA) is one of the most popular efficiency estimator. Due to its non-parametric approach and the possibility of multiple outputs, the DEA is widely used in analysis of efficiency determinants, see overview of DEA models by Emrouznejad and Yang [3].

The resulting efficiency estimates are often used in various two-stage efficiency analysis. In the first stage, the efficiency is estimated. In the second stage, the statistical significance of the variables affecting the estimated efficiency in the first stage is studied. The ability to estimate efficiency is demonstrated in simulation studies where efficiency is simulated using non-linear production functions. However, differently specified production functions lead to different efficiency estimates. Furthermore, DEA efficiency estimates are sensitive to the presence of outliers and measurement errors, etc., [2, 7].

The possibilities of correcting these shortcomings are studied with respect to the distribution of the overall DEA efficiency. The methods do not take into account changes in the order of effective units and even small differences in the definition of inputs have a significant impact on the final ranking of DEA units. In this case, the two-stage method can provide considerably misleading results.

In this article, we focus on linking DEA efficiency and economic performance of agricultural companies – farms. We stem from the premise that efficient companies tend to follow better economics indicators trajectory in the long run than inefficient companies. If the DEA method is able to estimate the efficiency of a given farm, then there should be a positive relationship between efficiency at time t and economic results of the company at time $t + h$. At the same time, it can be assumed that economic results at time $t - h$ can positively affect efficiency at time t . Specifically, we are examining the relationship between the lagged production efficiency of farms at present and their return on assets (*ROA*) and conversely between lagged *ROA* and DEA efficiency. The analysis is performed on the data of Czech agricultural enterprises in the period between 2008 and 2015 and the results suggest that there is an unexpectedly weak correlation.

The structure of the paper is as follows. In Section 2, we describe the used dataset and the methodology used for obtaining efficiency estimates and correlation estimates. In Section 3, we discuss the estimated correlation between DEA efficiency and *ROA*.

¹ Department of Econometrics, Prague University of Economics and Business, Winston Churchill Square 4, 13067 Prague, Czech Republic, lukas.fryd@vse.cz

² Department of Econometrics, Prague University of Economics and Business, Winston Churchill Square 4, 13067 Prague, Czech Republic, ondrej.sokol@vse.cz

2 Data and methodology

2.1 DEA method

Let n_1 is the number of inputs, n_2 the number of outputs and $m + 1$ the number of decision making units (DMU). Consider

- $I_0 \in \mathbb{R}^{n_1}$ is the input nonnegative vector for DMU₀,
- $O_0 \in \mathbb{R}^{n_2}$ is the output nonnegative vector for DMU₀,
- $I \in \mathbb{R}^{m \times n_1}$ is the input nonnegative matrix for the other DMUs,
- $O \in \mathbb{R}^{m \times n_2}$ is the output nonnegative matrix for the other DMUs.

As we need to estimate the efficiency of each unit, we run DEA optimization model. In particular, we use the linear approximation of the model [5] to compute the efficiency of given DMU₀. Hladík's model uses the efficiency scale from 0 to 2. Similarly to common DEA approaches, the higher score of production unit means, that the unit remains efficient for larger variation of all data. Equally, the lower score means that unit would be inefficient for larger variation of all data. The borderline of efficient units is equal to 1. The score is based on the largest allowable variation of all input and output data such that unit remains efficient or the smallest variation of data to become efficient in case of inefficient unit [5]. While the model was published only recently, it was already used in several empirical studies, see [4, 6].

The linear model is as follows

$$\begin{aligned}
 & \delta^* = \max_{\tilde{u}, \tilde{v}} \delta \\
 \text{s.t.} \quad & O_0^T \tilde{u} \geq 1 + \delta, \\
 & I_0^T \tilde{v} \leq 1 - \delta, \\
 & O\tilde{u} - I\tilde{v} \leq 0, \\
 & \tilde{u}, \tilde{v} \geq 0,
 \end{aligned} \tag{1}$$

where $r = 1 + \delta^*$ is the resulting efficiency score of chosen DMU₀. As stated above, if $r \in (0, 1)$ then the DMU₀ is inefficient and if $r \in [1, 2)$ indicates that DMU₀ is efficient. In order to extract vectors of input and output weight, we can compute $\tilde{u} := u/(1 - \delta)$ and $\tilde{v} := v/(1 - \delta)$ with u and v represent the vectors of input and output weights, respectively.

2.2 Data

We use a data set of 220 Czech farms in the period 2008 to 2015. The data set is balanced panel.

The data come from *Farm Accountancy Data Network* (FADN). FADN is an agriculture database, maintained under auspices of the European Commission. FADN participation is voluntary for farms. Data from a sample of farms are sent to a national branch of FADN (so-called *liaison agent*), which transmit the data to the global FADN database and is responsible for the international comparability of the data, i.e., the methodology of the collecting the data shall be the same in every EU country.

The survey does not cover all the farms in the Union but only those which due to their size could be considered commercial. In total, the FADN sample consists of about 80 000 holdings and represent about 5 million farms using about 90 percent of the total utilized agricultural area and producing about 90 percent of agricultural marketable output. The database consists of ~ 5000 economic and other variables on an annual basis.

We use 4 inputs and 3 outputs for the DEA estimation of the technical efficiency based on similar recent studies (see for example [1]). We also used this approach in [4]. Our list of inputs consist of

1. Total labor input in annual work units (AWU),
2. Total utilized agricultural area in ha,
3. Depreciation and interest paid in Euro,
4. Total intermediate consumption in Euro,

and the outputs are following

1. Total output crops and crop production in Euro,
2. Total output livestock and livestock products in Euro,
3. Other output in Euro.

Other output is calculated as the sum of all other outputs.

2.3 Cross-correlation of DEA efficiency and ROA

Once we have the estimated efficiency for each farm for each year, we calculate the sample correlation between DEA farm efficiency and return on assets (*ROA*). Denote $r_{i,t}$ the DEA efficiency of farm i in time t and $ROA_{i,t}$ return on assets of farm i in time t .

$$C_{i,h} = \frac{\sum_t (r_{i,t} - \bar{r}_i)(ROA_{i,t+h} - \overline{ROA_{i,h}})}{(n-1)s_{r_i}s_{ROA_{i,h}}} \quad (2)$$

where \bar{r}_i and $\overline{ROA_{i,h}}$ are estimated means of $r_{i,t}$ and $ROA_{i,t+h}$ respectively, s_{r_i} and $s_{ROA_{i,h}}$ are sample standard deviations of $r_{i,t}$ and $ROA_{i,t+h}$ respectively, n is the number of valid observations with respect to h which is the given lag.

We consider all possible lags ranging from -6 to 6 , e.g. we compute $C_{i,h}$ for all i and $h = -6, \dots, 6$. Naturally, for h approaching zero, we have significantly more observations than for very low or very high values of h .

3 Results

Table 3 shows summary statistics for cross-correlation in individual delays. The quartile range (0.25 – 0.75) for lag = 1 is from -0.01 to 0.44 . A similar result is obtained for lag = -1 . The quartile ranges for the other lags are approximately in the range from -0.2 to 0.1 . The results show that efficiency and *ROA* are strongly contemporaneously correlated. Conversely, in the case of different lags, the correlation is very weak.

Table 1 Cross-correlation summary statistics

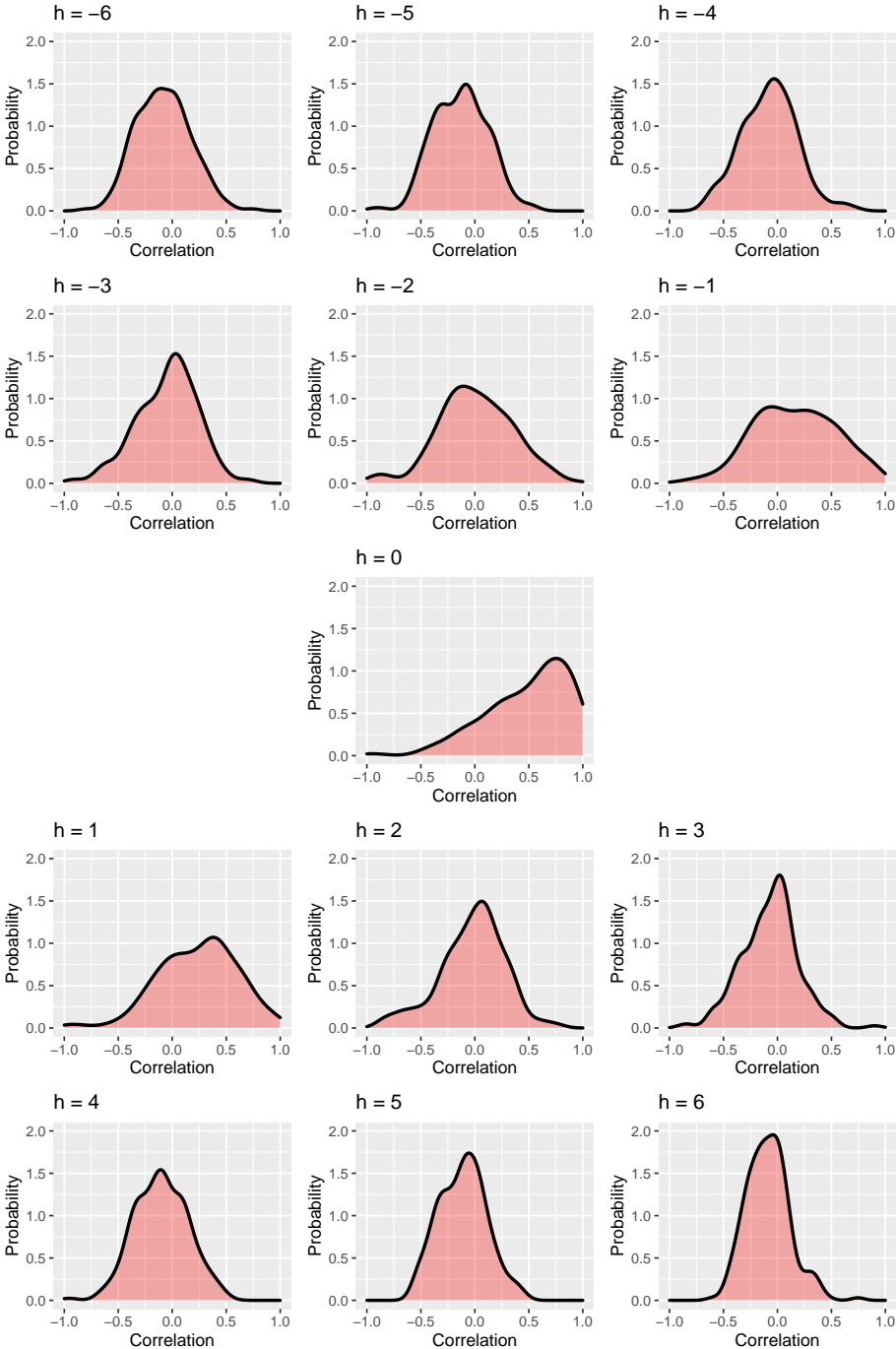
Lag (h)	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-6	-0.66	-0.26	-0.09	-0.08	0.07	0.63
-5	-0.74	-0.31	-0.12	-0.12	0.04	0.49
-4	-0.60	-0.24	-0.06	-0.07	0.10	0.62
-3	-0.73	-0.23	-0.01	-0.04	0.14	0.60
-2	-0.80	-0.21	-0.01	0.00	0.24	0.74
-1	-0.77	-0.13	0.18	0.16	0.45	0.88
0	-0.91	0.42	0.67	0.57	0.82	0.98
1	-0.79	-0.01	0.26	0.21	0.44	0.82
2	-0.70	-0.20	0.01	-0.02	0.16	0.64
3	-0.70	-0.22	-0.04	-0.07	0.07	0.72
4	-0.74	-0.28	-0.10	-0.10	0.08	0.46
5	-0.50	-0.27	-0.10	-0.11	0.03	0.43
6	-0.54	-0.23	-0.10	-0.08	0.03	0.63

We show the estimated density of correlation between and DEA efficiency estimates and return on assets for individual farms for $h = -6, \dots, 0, \dots, 6$ in Figure 3. Here, $h < 0$ represent the distribution of the DEA efficiency correlation at time t and *ROA* in the past – at time $t + h$. Conversely, $h > 0$ represent the distribution of the correlation for the DEA efficiency at time t and future *ROA* at time $t + h$.

With zero lag, $h = 0$, a clear positive correlation between the two variables can be seen. Most farms achieve a correlation in the range from 0.5 to 1. This result is expected, because the inputs are strongly correlated with the *ROA*. Hence, DEA efficiency should be strongly correlated with *ROA*.

However, in the case of any lag, $h \neq 0$ we can see a significant drop in estimated correlation. The positive, but still rather minor, correlation can be seen only for lag $(-1, 1)$. For higher absolute h , cross-correlation estimates are in the range from -0.5 to 0.5 and interquartile range from -0.3 to 0.2 . The mean of correlation is negative for $|h| > 2$, although we cannot reject the null hypothesis of 0. This is indeed a surprising result as we would expect a positive correlation.

Figure 1 Estimated density of correlation between and DEA efficiency estimates and *ROA* for various lag h .



4 Conclusion

The topic of the work is the suitability of using the DEA method as an estimator of efficiency with regard to its frequent use in the two-step method. We start from the premise that more efficient companies should achieve better economic results in the medium and long term than inefficient companies. We specifically analyze the correlation between DEA efficiency of Czech farms and the *ROA* indicator. The results show that the correlation between DEA efficiency and *ROA* reaches a correlation above 0.5 only contemporaneously. In contrast, the cross-correlation for the delay $h \neq 0$ is very low.

Acknowledgements

The work was supported by the Czech Science Foundation under grant 20-17529S and by the Internal Grant Agency of Prague University of Economics and Business under Grant F4/34/2020.

References

- [1] Davidova, S., and Latruffe, L.: Relationships between technical efficiency and financial management for czech republic farms. *Journal of Agricultural Economics* **58** (2007), 269–288.
- [2] Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., and Shale, E. A.: Pitfalls and protocols in dea. *European Journal of operational research* **132** (2001), 245–259.
- [3] Emrouznejad, A., and Yang, G.-l.: A survey and analysis of the first 40 years of scholarly literature in dea: 1978–2016. *Socio-Economic Planning Sciences* **61** (2018), 4–8.
- [4] Frýd, L., and Sokol, O.: Relationships between technical efficiency and subsidies for czech farms: A two-stage robust approach. *Socio-Economic Planning Sciences* (2021), 101059.
- [5] Hladík, M.: Universal Efficiency Scores in Data Envelopment Analysis Based on a Robust Approach. *Expert Systems with Applications* **122** (2019), 242–252.
- [6] Holý, V.: The impact of operating environment on efficiency of public libraries. *Central European Journal of Operations Research* (2020), 1–20.
- [7] Simar, L., and Wilson, P. W.: A general methodology for bootstrapping in non-parametric frontier models. *Journal of applied statistics* **27** (2000), 779–802.

The geography of most cited scientific publications: Mixed Geographically Weighted Regression approach

Andrea Furková¹

Abstract. The paper analyses spatial heterogeneity of top-level scientific publications of the European regions and try to answer the question which regions or groups of regions are the most innovative in this sense. We supposed that the response of innovation output (most cited scientific publications) to a change on innovation inputs (R&D expenditure and human capital) might be not homogeneous across all European regions. In addition, we hypothesize that there is still gap between post-socialist and “western” countries in terms of elite publications. Mixed geographically weighted regression (MGWR) model was used as a main tool for examining our research questions. MGWR model can produce parameter estimations that have global character and other parameters that have local character in accordance with observation location. We found out that the both innovation input parameters vary significantly across the European area and the gap between post-socialist and “western” countries in terms of elite publications was confirmed.

Keywords: Mixed geographically weighted regression, spatial heterogeneity, scientific publications, innovation

JEL Classification: O31, R12

AMS Classification: 91B72

1 Introduction

It is evident that the main objective of Research & Development (R&D) policy is to increase innovation outcomes. However, the problem arise when we want to measure the level of innovation activities and technological progress. Following the concept of the Regional Knowledge Production Function (RKPF) model (see e.g., [8]), two types of indicators are usually considered, i.e., technological innovation inputs and technological innovation outputs. Traditionally, R&D expenditure and human capital are recognized as significant innovation determinants. On the other hand, the number of patent applications, number of scientific publications and citations are accepted as innovation outputs. In this paper, we raise a different approach to evaluation of scientific activities. We turn our attention to the “elite publications”, i.e., scientific publications that are among the top 10% most cited publications worldwide and we will consider it as a proxy for an innovation output of the region. These top-level publications are considered as a measure for the efficiency of the research system as highly cited publications are assumed to be of higher quality. This indicator is also the part of a composite indicator, the Regional Innovation Index –RII (see [5]) which is one of the few options for the comparative assessment of the performance of European innovation systems at the regional level.

This paper will try to analyse which European regions or groups of regions have the largest share in the production of the top-level scientific publications and therefore are the most innovative regions in this sense. We suppose that the response of innovation output (most cited scientific publications) to a change on innovation inputs (R&D expenditure and human capital) might be not homogeneous across all European regions. In addition, we hypothesize that there is still gap between post-socialist and “western” countries in terms of elite publications. Mixed geographically weighted regression (MGWR) model seems to be a suitable tool for examining our hypotheses. This model is a combination of linear regression model and geographically weighted regression (GWR) model; therefore, MGWR model could produce parameter estimation that had global parameter estimation, and other parameter that had local parameter in accordance with its observation location.

The structure of the paper is as follows: section 2 provides data and study area descriptions and brief theoretical backgrounds of the study; empirical results are presented and interpreted in section 3. Main concluding remarks contain section 4 and the paper closes with references.

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 Methodology

The first part of this section provides an overview of a study area and description of the data. The second part of this section briefly introduces MGWR model relevant for the subsequent empirical analysis.

2.1 Data description and study area

The real distribution of most cited scientific publications of European regions in 2019 is presented in Figure 1. The map shows 238 regions of 23 the EU member states, Norway, Serbia² and Switzerland at different NUTS (Nomenclature of territorial units for statistics) levels, i.e., at NUTS 1 or NUTS 2 levels. Our innovation analysis will include 220 observations because of isolated observations (island regions); the data set reduction was done. According to RII, the regions presented in Figure 1, have been classified into four performance groups, i.e., groups of high, strong, moderate and low performers. Figure 1 indicates strong geographical performance differences. Scientific publications among the top-10 % most cited seems to be less spread within countries but more across countries. Many regions in Northern, Western and Central European countries such as Denmark, Norway, Sweden, Finland, Ireland, France, Germany, Belgium, and Austria are ranked as strong performers. Elite scientific publications are produced by the United Kingdom, Switzerland and the Netherlands, where the majority of regions consist of high performers. While, there might be a relatively small variety among regions in many European countries, for instance Greek regions show the highest level of variety with regard to the top 10% most cited publications. There can be found a high performer region and also low performer regions. It is interesting to mention a Portuguese autonomous region Madeira which is the only Southern European region represented in the top 10 group of European regions. Different trend can be seen as for the rest of Southern European regions and Eastern European regions. These regions are usually classified as low or moderate performers.

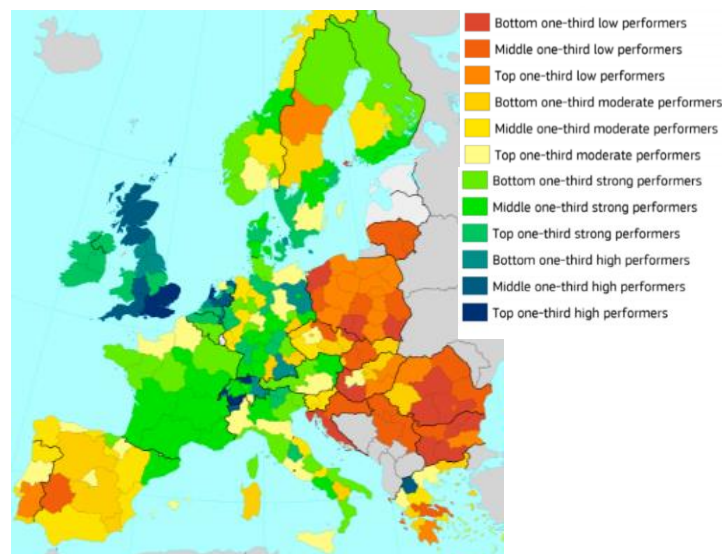


Figure 1 The distribution of scientific publications among the top 10% most cited publications worldwide
Source: author's elaboration based on the RIS 2019 [5]

It is obvious that the geographical position of the region and properties of the neighbourhood occupying a key role for creation of innovation in given areas. The geographical aspect also plays an important role in case of production of scientific publications, especially with regard to the most valued scientific publications. We hypothesize that the response of innovation output (most cited scientific publications) to a change on innovation inputs (R&D expenditure and human capital) might be not homogeneous across all European regions and we assume that there is still a gap between post-socialist and “western” countries in terms of elite publications. Thus, the problem of spatial heterogeneity as one of the spatial effects may be present in connection with the modelling of regional innovation activities. For this reason, we decided to apply MGWR model, which provides local parameters estimations while some parameters may be global. The next section briefly introduces this model.

² For Serbia, official NUTS codes are not yet available and therefore unofficial codes will be used (see [5]).

Table 1 gives a description of all variables (innovation output and innovation inputs) used in our empirical analysis. The selection of the data was influenced by the fact that at regional levels, the relevant data are limited. A brief reasoning for inclusion of the variables under consideration is provided in Table 1. Data related to variables *PUB* and *EXP* are obtained from RIS 2019 [5]. Variable *HRST* comes from regional Eurostat statistics database [9].

Category	Definition	Form	Abbr.
Dependent variable:			
Scientific publications among the top 10% most cited publications worldwide	The number of scientific publications among the top 10 % most cited publication worldwide per total number of scientific publications.	Normalized ³	<i>PUB</i>
Explanatory variables:			
R&D expenditure in the public sector Reasoning: R&D expenditure represents one of the major drivers of economic growth in a knowledge-based economy. R&D spending is essential for making the transition to a knowledge-based economy as well as for improving production technologies and stimulating growth.	All R&D expenditures in the government sector and the higher education sector as percentage of GDP.	Normalized	<i>EXP</i>
Human resources in science and technology (<i>HRST</i> ⁴) Reasoning: For scientific activities, human resources represent a knowledge base, which is a source of ideas.	<i>HRST</i> as percentage of active population.	Normalized ⁵	<i>HRST</i>
Regions of post-socialist countries (<i>PSOC</i>) Reasoning: Former political system of the country may influence innovative activities of the regions.	$DUM_{PSOC}=1$, if region belongs to <i>PSOC</i> , others=0.	Binary	DUM_{PSOC}

Table 1 Model variables description

2.2 Mixed Geographically Weighted Regression Model

First, let us pay a brief attention to the Geographically Weighted Regression (GWR) model. The goal of GWR methodology is to obtain local linear regression estimates for each point in the space, i.e., for each observation $i \in \{1, \dots, N\}$ we deal with different vector of local parameters $\beta(u_i, v_i)$. Coordinates (u_i, v_i) represents the longitude and latitude of observation i . GWR method requires the spatial kernel function and its bandwidth selection. Next, N dimensional diagonal weight matrix \mathbf{W}_i is constructed such that $\mathbf{W}_i = K(\mathbf{d}_i, h)$, where $K(\)$ is a spatial kernel function, \mathbf{d}_i is a distance vector between the central point and all neighbours, and h is a bandwidth or decay parameter (see e.g., [2]). The GWR model can be expressed as:

$$y_i = \beta_v(u_i, v_i; h) X_v + \varepsilon_i, \quad \forall i \in \{1, \dots, N\}, \quad (1)$$

where y is a vector of dependent variable, h is a bandwidth parameter that allows to define the local subsample around the coordinates of each point (u_i, v_i) using a given distance kernel $K(\)$, X_v represents k_v explanatory variables with spatially varying coefficients (β_v) and ε is an error term. The parameters of the GWR model are estimated by the weighted least squares approach and the estimation of the parameters in each location i is given by (see, e.g., [6]; [4]; [1]):

³ The data provided by RIS 2019 are already normalized. The minmax procedure was used and the maximum normalised score is equal to 1 and the minimum normalised score is equal to 0. For more details regarding normalising RII data see [5].

⁴ HRST are people who fulfil one or other of the following conditions: (1) have successfully completed a tertiary level education; (2) not formally qualified as above but employed in a scientific and technical occupation where the above qualifications are normally required (see [10]).

⁵ The minmax procedure was used.

$$\beta(u_i, v_i) = (X^T W_i X)^{-1} (X^T W_i y) \quad (2)$$

GWR model defined by formula (2) seems to be not sufficient for socioeconomic variables that have global effects and are independent from individual location. In addition, it appears inadequate for local categorical variables, since spatially varying parameters associated with such variables may have no meaning. For such situations, mixed GWR (MGWR) model was developed. This model can be formulated as follows [4]:

$$y_i = \beta_c X_c + \beta_v(u_i, v_i; h) X_v + u_i, \quad \forall i \in \{1, \dots, N\} \quad (3)$$

where X_c represents k_c explanatory variables with constant coefficients (β_c), and X_v represents k_v explanatory variables with spatially varying parameters (β_v). It should be noted that $k = k_c + k_v$. All remaining terms of model (3) were defined above. Already Fotheringham et al. [3] dealt with the issue of MGWR estimation defined in (3). They proposed seven-step estimation; however, this approach has appeared somewhat intensive in terms of computation. In [7], less demanding, a two steps methodology based on partial linear models can be found. Geniaux and Martinetti also used this methodology and we will follow this approach in our empirical analysis. For more details, see [4].

3 Empirical Results

The distribution of innovation processes (represented by most cited scientific publications worldwide) across the European regions suggests that, the strength of the influence of particular determinants of innovation may vary in given locations. Thus, the problem of spatial heterogeneity should be considered. Next, we will assume that the expected value of most cited scientific publications is a function of R&D expenditure in the public sector and Human resources in science and technology. In addition, we hypothesize that there is still a gap between post-socialist and “western” countries in terms of elite publications. For this reason, the model includes global dummy variable reflecting political history of the region and global interactive dummy variable with human resources variable.

The process of estimating of a MGWR model starts with weighting scheme selection. We decided for Gaussian weighting scheme with fixed⁶ bandwidth parameter h (see Table 2) calibrated by cross-validation optimization procedure. The selected results of MGWR estimation (minimum, lower quartile, median, mean, upper quartile, maximum) in comparison to OLS estimation are presented in Table 2. The evidence for spatial heterogeneity is already supported by basic statistics. For instance, parameter estimate of *HRST* (Human resources in science and technology) is varying even from negative values -0.0361 up to 0.6970 with median value 0.2460, while the global OLS parameter estimate is 0.3320. The minimum and maximum values of estimated MGWR parameters indicate how varied the influence of a given innovation input may be in a particular region. As for global OLS estimation, we can see that all parameters are statistically significant except the parameter associated with *EXP* variable. This was unexpected but at the same time, we can see based on the local regressions that this factor was significant in up to 54.55% of cases. Consequently, R&D spending in public sector seems to be essential for producing and improving regional scientific activities. The MGWR estimation results reveal that the most important determinant of most cited publications is *HRST* variable. Its significance was confirmed in 97.27% of cases.

Based on the MGWR model, we also examined the differences between post-socialist and “western” countries in terms of elite publications and the question whether the effect of human resources varies by political history of the region. The results verified our assumptions that the political history of the region still matter and that the effect of human resources varies by political history of the region. Both global dummy variables are statistically significant and they have expected negative signs.

The overview of the estimated local parameters is presented in the form of box and significance maps in Figure 2 and Figure 3.

⁶ According to preliminary estimates and analyses, we concluded to prefer the model with fixed weighting scheme to the model with adaptive one.

	MGWR (Gaussian kernel functions with fixed bandwidth, $h=3.793$)							Global (OLS)
	Min.	First Quartile	Median	Mean	Third Quartile	Max.	Percent of significant cases at 95 %	
β_0	0.0353	0.3218	0.3413	0.3310	0.3563	0.4704	100%	0.3556 (0.0000)
$\beta_1 (EXP)$	-0.1363	-0.0034	0.0855	0.0713	0.1220	0.4171	54.55%	0.0001 (0.9976)
$\beta_2 (HRST)$	-0.0361	0.1442	0.2460	0.2704	0.3762	0.6970	97.27%	0.3320 (0.0000)
$\beta_3 (DUM_{psoc})$	-0.1465 (0.0260)						-	-0.1504 (0.0000)
$\beta_4 (DUM_{psoc}HRST)$	-0.0848 (0.5332)						-	-0.2157 (0.0032)
AIC	-476.852							-395.826
R ²	-							0.6701

Table 2 Summary of MGWR estimation

Note: p -values in parenthesis.

Source: own calculations in RStudio

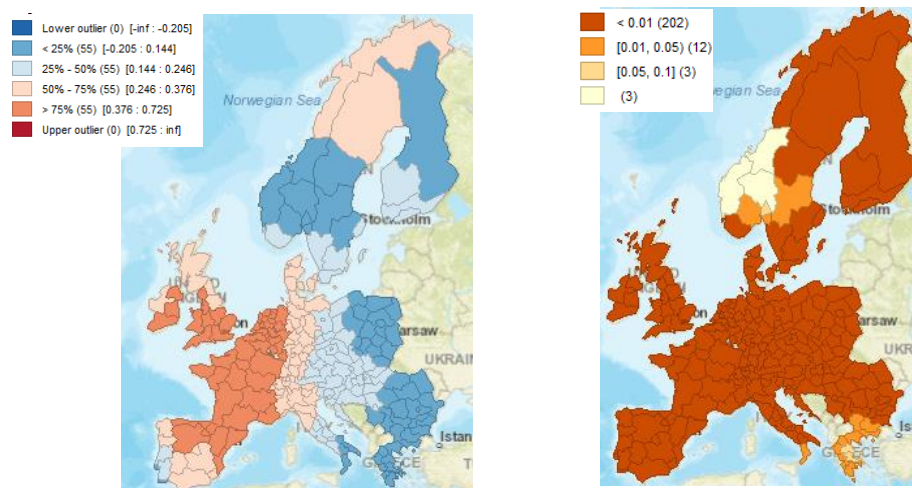


Figure 2 Human resources in science and technology: box map of local parameter estimates (left) and significance map (right)

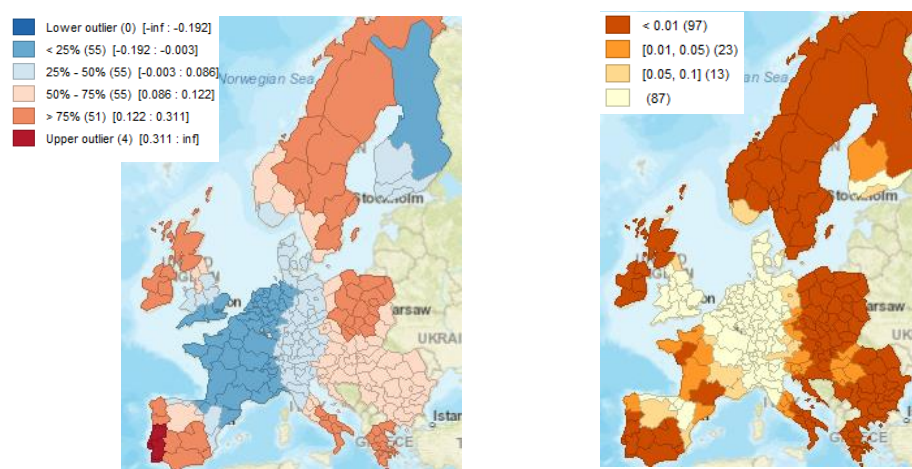


Figure 3 R&D expenditure in the public sector: box map of local parameter estimates (left) and significance map (right)

4 Conclusion

In this paper, MGWR approach has been exploited as a tool for analysing spatial heterogeneity of top-level scientific publications of the European regions. Based on the MGWR estimation, we found out that the both innovation input parameters vary significantly across the European area. The greatest effect of R&D expenditure (highest parameter values, see Figure 3 (left)) is evident for regions that are classified as low and moderate performers in terms of most cited publications (see Figure 1). These are mainly regions with post-socialistic history, regions of Scandinavian countries, and regions of Spain, Greece and south Italy. Almost opposite situation applies to the human resources variable as we recorded low parameter values (see Figure 2 (left)) for already mentioned regions and high parameter values are evident for high and strong performers in terms of most cited publications. In addition, we found out that the effect of human resources varies by political history of the region. Despite the fact that the countries of Central and Eastern European countries have undergone a difficult process of post-socialist transformation, these regions are still lagging. Our results invoke local R&D policy implication not region wide policy implication.

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 VEGA 1/0211/21 "Econometric Analysis of Macroeconomic Impacts of Pandemics in the World with Emphasis on the Development of EU Economies and Especially the Slovak Economy".

References

- [1] Anselin, L. & Rey, S.J. (2014). *Modern Spatial Econometrics in Practice*. Chicago: GeoDa Press LLC.
- [2] Chocholatá, M. (2020). Spatial Variations in the Educational Performance in Slovak Districts. *Statistics and Economy Journal*, 100(2), 193–203.
- [3] Fotheringham, A.S., Brunsdon, C., & Charlton, M. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, 39(3), 497–524.
- [4] Geniaux, G. & Martinetti, D. (2018). A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. *Regional Science and Urban Economics*, 72, 74–85.
- [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] LeSage, J. P. (1999). The theory and practice of spatial econometrics. Available at: <http://www.spatial-econometrics.com/html/sbook.pdf>. [Accessed 20 Feb. 2018].
- [7] Mei, C., He, S. & Fang, K. (2004). A note on the mixed geographically weighted regression model. *Journal of Regional Science*, 44(1), 143–157.
- [8] Moreno, R., Paci, R. & Usai, S. (2005). Spatial Spillovers and Innovation Activity in European Regions. *Environment and Planning A: Economy and Space*, 37(10), 1793–1812.
- [9] <http://ec.europa.eu/eurostat/>. [Accessed 1 Mar. 2020].
- [10] https://ec.europa.eu/eurostat/cache/metadata/en/hrst_esms.htm. [Accessed 1 Mar. 2020].

Bilevel Linear Programming under Interval Uncertainty

Elif Garajová¹, Miroslav Rada², Milan Hladík³

Abstract. Bilevel linear programming provides a suitable mathematical model for many practical optimization problems. Since the real-world data are often inaccurate or uncertain, we consider the model under interval uncertainty, in which only the lower and upper bounds on the input data are available and we assume that the uncertain coefficients can be perturbed independently within the given intervals.

Building on the theory of interval optimization and bilevel linear programming, we study the basic properties of bilevel interval linear programs from both a theoretical and a computational point of view. In our study, we focus on the main problems solved in interval optimization, such as computing the range of optimal values, checking the existence of feasible and optimal solutions and testing unboundedness of a scenario in the interval program.

Keywords: bilevel programming, interval uncertainty, optimality

JEL Classification: C44, C61

AMS Classification: 90C70

1 Introduction

Throughout the recent years, *bilevel programming* models [5, 2] have been successfully applied in solving a wide range of practical optimization problems. In such models, we consider a hierarchical structure of decision making consisting of two levels represented by two nested optimization problems—the leader (upper-level) problem and the follower (lower-level) problem. Mathematically, we solve a problem in the form

$$\begin{aligned} \min_{x,y} \quad & f(x, y) \\ \text{s. t.} \quad & (x, y) \in X, \\ & y \in \arg \min_y \{g(x, y) \text{ s. t. } (x, y) \in Y\}, \end{aligned}$$

for the given constraint sets X, Y . In this paper, we focus on the bilevel programming models with a linear objective function and linear constraints on both levels [1, 9]. Although this is perhaps the easiest special case, it is still difficult to tackle and several decision problems related to bilevel linear programming were, in fact, proved to be NP-hard [6].

Since uncertainty is an ever-present issue in real-world optimization problems, attention has also been devoted to exploring bilevel models with inexact, vague or imprecise data. Here, we adopt the approach of *interval programming*, assuming that only lower and upper bounds on the inexact data are known and that the values can be perturbed independently within these bounds. While the topic of single-level linear programming with interval data is quite well-studied (see e.g. [4, 15, 10] and references therein), only a handful of works are available for bilevel interval programming problems [11, 12, 14].

We derive several results on the theoretical and computational properties of bilevel interval linear programs. First, we build on and revise the former results [3, 13] on computing the best and the worst value, which is optimal for some scenario of the interval problem (this is also known as the problem of computing the *optimal value range*). Then, we prove that the decision problems of checking existence of feasible and optimal solutions are NP-hard for bilevel interval linear programming. Furthermore, we also show NP-hardness of checking unboundedness of at least one scenario of the interval program.

¹ Charles University, Faculty of Mathematics and Physics, Dept. of Applied Mathematics, Malostranské nám. 25, Prague, Czech Republic; Prague University of Economics and Business, Dept. of Econometrics, nám. W. Churchilla 4, Prague, Czech Republic, elif@kam.mff.cuni.cz

² Prague University of Economics and Business, Dept. of Econometrics & Dept. of Financial Accounting and Auditing, nám. W. Churchilla 4, Prague, Czech Republic, miroslav.rada@vse.cz

³ Charles University, Faculty of Mathematics and Physics, Dept. of Applied Mathematics, Malostranské nám. 25, Prague, Czech Republic; Prague University of Economics and Business, Dept. of Econometrics, nám. W. Churchilla 4, Prague, Czech Republic, hladik@kam.mff.cuni.cz

2 Bilevel Interval Linear Programming

Let us first introduce the essential notions and notation of interval programming and bilevel programming used throughout the paper. Note that all inequality relations on the set of matrices and vectors are understood element-wise.

Interval data. Let the symbol $\mathbb{I}\mathbb{R}$ denote the set of all closed real intervals. Given two real matrices $\underline{A}, \bar{A} \in \mathbb{R}^{m \times n}$ satisfying $\underline{A} \leq \bar{A}$, we define an *interval matrix* $\mathbf{A} \in \mathbb{I}\mathbb{R}^{m \times n}$ as the set

$$\mathbf{A} = [\underline{A}, \bar{A}] = \{A \in \mathbb{R}^{m \times n} : \underline{A} \leq A \leq \bar{A}\},$$

where \underline{A} is the *lower bound* and \bar{A} is the *upper bound* of the interval matrix. Alternatively, an interval matrix can be also determined by the *center* $A_c = \frac{1}{2}(\bar{A} + \underline{A})$ and *radius* $A_\Delta = \frac{1}{2}(\bar{A} - \underline{A})$. An *interval vector* $\mathbf{a} \in \mathbb{I}\mathbb{R}^n$ can be defined analogously as an $n \times 1$ interval matrix.

Bilevel interval programming. For given interval matrices $\mathbf{A}_1 \in \mathbb{I}\mathbb{R}^{m_1 \times n_1}$, $\mathbf{A}_2 \in \mathbb{I}\mathbb{R}^{m_2 \times n_1}$, $\mathbf{B}_1 \in \mathbb{I}\mathbb{R}^{m_1 \times n_2}$, $\mathbf{B}_2 \in \mathbb{I}\mathbb{R}^{m_2 \times n_2}$, interval objective vectors $\mathbf{c} \in \mathbb{I}\mathbb{R}^{n_1}$, $\mathbf{d} \in \mathbb{I}\mathbb{R}^{n_2}$, $\mathbf{a} \in \mathbb{I}\mathbb{R}^{n_2}$ and interval right-hand-side vectors $\mathbf{b}_1 \in \mathbb{I}\mathbb{R}^{m_1}$, $\mathbf{b}_2 \in \mathbb{I}\mathbb{R}^{m_2}$, we define a *bilevel interval linear program (BILP)* as the set of all bilevel linear programs in the form

$$\begin{aligned} \min_{x, y \geq 0} \quad & \mathbf{c}^T x + \mathbf{d}^T y \\ \text{s. t.} \quad & \mathbf{A}_1 x + \mathbf{B}_1 y \geq \mathbf{b}_1, \\ & y \in \arg \min_{y \geq 0} \{ \mathbf{a}^T y \text{ s. t. } \mathbf{B}_2 y \geq \mathbf{b}_2 - \mathbf{A}_2 x \}, \end{aligned} \quad (1)$$

where the coefficients of the bilevel program belong to the respective intervals ($A_1 \in \mathbf{A}_1$, $A_2 \in \mathbf{A}_2$ etc.). A specific bilevel linear program in the set is called a *scenario*. For simplicity of notation, we also write the former interval problem in the concise form with interval coefficients as

$$\begin{aligned} \min_{x, y \geq 0} \quad & \mathbf{c}^T x + \mathbf{d}^T y \\ \text{s. t.} \quad & \mathbf{A}_1 x + \mathbf{B}_1 y \geq \mathbf{b}_1, \\ & y \in \arg \min_{y \geq 0} \{ \mathbf{a}^T y \text{ s. t. } \mathbf{B}_2 y \geq \mathbf{b}_2 - \mathbf{A}_2 x \}. \end{aligned} \quad (2)$$

Dependency problem. Note that the formulation (2) of a bilevel interval linear programming problem is not the most general, since we only consider constraints expressed by inequalities with non-negative variables on both levels. When dealing with interval coefficients in the programs, it is not always possible to apply the standard transformations to convert a given problem into the desired form and programs in different forms may have to be treated separately (see [8] for details).

Feasibility and optimality. For a given solution $(x, y) \in \mathbb{R}^{n_1+n_2}$ to be feasible for the bilevel program (1), we need to ensure that it satisfies both the upper-level and the lower-level constraints, i.e. that it belongs to the *constraint region*

$$S = \{(x, y) \in \mathbb{R}^{n_1+n_2} : \mathbf{A}_1 x + \mathbf{B}_1 y \geq \mathbf{b}_1, \mathbf{A}_2 x + \mathbf{B}_2 y \geq \mathbf{b}_2, x, y \geq 0\}.$$

Furthermore, the vector y has to be a rational response of the follower to the leader's choice x (i.e., the vector y has to be an optimal solution of the lower-level problem for the fixed x):

$$M(x) = \{y \in \mathbb{R}^{n_2} : y \in \arg \min_{y \geq 0} \{ \mathbf{a}^T y \text{ s. t. } \mathbf{B}_2 y \geq \mathbf{b}_2 - \mathbf{A}_2 x, y \geq 0 \}\}.$$

Then, the set of all bilevel feasible solutions, also known as the *inducible region*, is the set of all pairs $(x, y) \in S$ such that $y \in M(x)$.

For the interval programming problem, we consider the feasible and optimal solutions in the weak sense: a given solution $(x, y) \in \mathbb{R}^{n_1+n_2}$ is *(weakly) feasible*, if it is a feasible solution for at least one scenario of the bilevel interval linear program. Analogously, a given (x, y) is *(weakly) optimal*, if it is optimal for some scenario of the BILP.

Example 1. Consider the bilevel interval linear program

$$\begin{aligned} \min_{x,y \geq 0} \quad & -x - y \\ \text{s. t.} \quad & y \in \arg \min_{y \geq 0} \{ [-1, 1]y \text{ s. t. } 3x - 2y \leq 12, 2x + y \leq 15, -3x + 5y \leq 10 \}. \end{aligned} \quad (3)$$

The polygon forming the constraint region of problem (3) is depicted in Figure 1. To find the weakly optimal solutions, we need to examine the possible scenarios of the interval program. In this case, there is only a single interval coefficient in the lower-level objective $\mathbf{a}y = [-1, 1]y$.

We can observe that for any choice of the objective coefficient $a \in [-1, 0)$, the optimal solution of the follower's problem will be on the upper boundary of the polygon. Similarly, for the values $a \in (0, 1]$, the optimal solutions (and thus also the points of the inducible region) lie on the lower boundary. For $a = 0$, the entire polygon is optimal. The corresponding optimal solutions from the inducible regions in the 3 cases are (5, 5), (6, 3) and (5, 5), respectively. Therefore, the weakly optimal solution set of BILP (3) is $\{(5, 5), (6, 3)\}$, with the best optimal value -10 and the worst optimal value -9 .

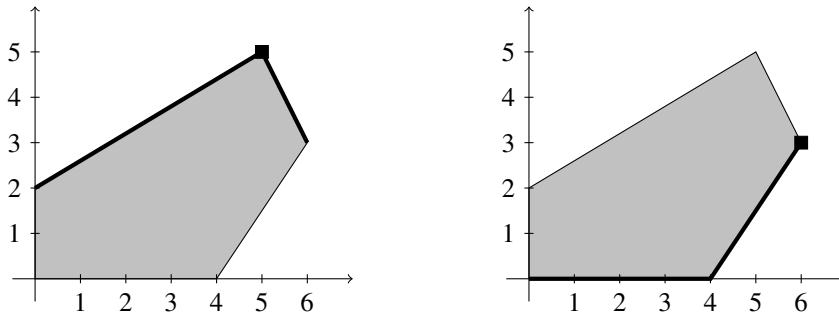


Figure 1 The constraint region of bilevel interval linear program (3). The bold lines depict the inducible region of the program for the lower-level objective $a < 0$ (left) and $a > 0$ (right), the square vertices represent the optimal solutions.

3 Properties of Bilevel Interval Linear Programs

Optimal value range. One of the main problems solved in all areas of interval programming is the so-called optimal value range problem, whose goal is to compute the best and the worst possible value that is optimal for some scenario of the given interval program. In the previous works [13], the authors proposed to compute the optimal values for a restriction of scenarios, for which some of the choices of interval coefficients are fixed:

$$\begin{aligned} \min_{x,y \geq 0} \quad & \underline{c}^T x + \underline{d}^T y \\ \text{s. t.} \quad & \overline{A}_1 x + \overline{B}_1 y \geq \underline{b}_1, \\ & y \in \arg \min_{y \geq 0} \{ \mathbf{a}^T y \text{ s. t. } \overline{B}_2 y \geq \underline{b}_2 - \overline{A}_2 x \}, \end{aligned} \quad (4)$$

as the best-value program, and,

$$\begin{aligned} \min_{x,y \geq 0} \quad & \overline{c}^T x + \overline{d}^T y \\ \text{s. t.} \quad & \underline{A}_1 x + \underline{B}_1 y \geq \overline{b}_1, \\ & y \in \arg \min_{y \geq 0} \{ \mathbf{a}^T y \text{ s. t. } \underline{B}_2 y \geq \overline{b}_2 - \underline{A}_2 x \}, \end{aligned} \quad (5)$$

as the worst-value program. The idea is similar to the computation used in single-level interval linear programming [4] and exploits non-negativity of the variables to find the extremal values. It should, however, be noted that these scenarios do not yield the best possible and the worst possible optimal value of the BILP, in general. This discrepancy is caused by the fact that expanding or reducing the lower-level feasible set may be beneficial or detrimental to the leader's objective value, depending on the specific objective function. We illustrate the issue through the following example.

Example 2. Consider two instances of the bilevel interval linear program (note that both x and y are onedimensional)

$$\begin{aligned} \min_{x,y \geq 0} \quad & dy \\ \text{s. t.} \quad & x \leq 2, \\ & y \in \arg \min_{y \geq 0} \{-y \text{ s. t. } x - y \geq -2, -2x - y \geq [-8, -5]\}, \end{aligned} \quad (6)$$

where the coefficient d in the upper-level objective function is either $d = 1$ or $d = -1$. The two extremal scenarios of the interval program (with right-hand-sides -8 and -5) are depicted in Figure 2.

Let us first examine the instance with $d = -1$, in which the upper-level objective is the same as the lower-level objective. In this case, the optimal solutions for the two extremal scenarios are $(2, 4)$ and $(1, 3)$, respectively, with the optimal values -4 and -3 . The best optimal value was achieved with the largest constraint set and the worst optimal value with the smallest constraint set, as proposed in (4) and (5). However, we will see that this is not always the case.

Consider now the instance with $d = 1$, where the upper-level and lower-level objectives are opposite. Here, the optimal solutions for the two extremal scenarios are $(0, 2)$ and $(2, 1)$, respectively. The best optimal value 1 was achieved in for the smallest constraint set, while the worst optimal value 2 was achieved for the largest constraint set.

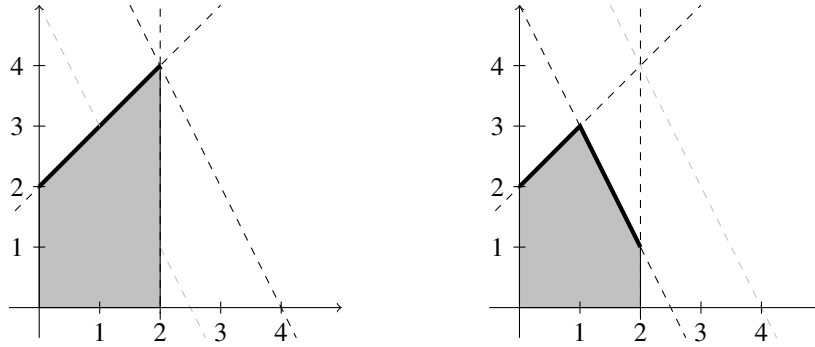


Figure 2 The two extremal scenarios of BILP (6) with the largest constraint set (left) and the smallest constraint set (right). The inducible regions are highlighted by the thick black line.

Example 2 shows that the choice of the lower-level coefficients in computing the best or the worst optimal value cannot be pre-determined, since it depends on the specific objective function considered in the problem. However, at the upper level, the choices can be made as proposed in the former work.

Proposition 1. *The best optimal value of BILP (2) can be computed as the best optimal value of the bilevel interval linear program*

$$\begin{aligned} \min_{x,y \geq 0} \quad & \underline{c}^T x + \underline{d}^T y \\ \text{s. t.} \quad & \overline{A}_1 x + \overline{B}_1 y \geq \underline{b}_1, \\ & y \in \arg \min_{y \geq 0} \{\underline{a}^T y \text{ s. t. } \mathbf{B}_2 y \geq \mathbf{b}_2 - \mathbf{A}_2 x\}. \end{aligned} \quad (7)$$

Proposition 2. *The worst optimal value of BILP (2) can be computed as the worst optimal value of the bilevel interval linear program*

$$\begin{aligned} \min_{x,y \geq 0} \quad & \overline{c}^T x + \overline{d}^T y \\ \text{s. t.} \quad & \underline{A}_1 x + \underline{B}_1 y \geq \overline{b}_1, \\ & y \in \arg \min_{y \geq 0} \{\overline{a}^T y \text{ s. t. } \mathbf{B}_2 y \geq \mathbf{b}_2 - \mathbf{A}_2 x\}, \end{aligned} \quad (8)$$

Both results can be proved in the same way as for single-level interval linear programming problems (see e.g. [4]), using non-negativity of the variables.

Computational complexity. Let us now examine the computational complexity of some decision problems related to bilevel interval linear programming. It is interesting to note that the form of a BILP considered in this paper (inequality constraints with non-negative variables) turned out to be the easiest in single-level linear programming in the sense that several of the generally NP-hard problems are easily solvable for programs in this special form.

However, since bilevel linear programming is difficult even with real coefficients and no uncertainty present in the model, there is little hope that the considered interval problems would be easy to solve. Indeed, we show that this is the case.

First of all, let us consider the feasible and optimal solutions. A natural question is to ask whether a given BILP even has a feasible (or optimal) solution for at least one scenario. It can be observed that this question leads to an NP-hard problem, in both cases, because checking the existence of an optimal solution is already NP-hard for the single-level interval linear programs [7]. Thus, we can prove NP-hardness of the considered decision problems simply by taking a BILP in the form

$$\begin{aligned} \min_{x, y \geq 0} \quad & 0^T x + 0^T y \\ \text{s. t.} \quad & y \in \arg \min_{y \geq 0} \{ \mathbf{a}^T y \text{ s. t. } \mathbf{B}_2 y \geq \mathbf{b}_2 - 0x \}. \end{aligned} \quad (9)$$

Now, BILP (9) has a feasible solution if and only if the lower-level interval linear programming problem has an optimal solution. This yields the following result:

Proposition 3. *Checking whether there exists a weakly feasible solution (or a weakly optimal solution) of a bilevel interval linear program is an NP-hard problem.*

Furthermore, we can also show that checking whether some feasible scenario of the BILP has an unbounded objective value is an NP-hard problem, as well. Again, we utilize a reduction from the problem of checking the existence of an optimal solution to a single-level interval linear program. In this case, we consider a BILP in the form

$$\begin{aligned} \min_{x, y \geq 0} \quad & -x \\ \text{s. t.} \quad & y \in \arg \min_{y \geq 0} \{ \mathbf{a}^T y \text{ s. t. } \mathbf{B}_2 y \geq \mathbf{b}_2 - 0x \}, \end{aligned} \quad (10)$$

which is unbounded if and only if it is feasible, i.e. if and only if the lower-level program has an optimal solution. Thus, we obtain the desired result.

Proposition 4. *Checking whether there exists a feasible scenario, in which the value of the upper-level objective function is unbounded, is an NP-hard problem.*

Although all three of the considered decision problems are NP-hard in the general case, there still may be special classes of BILPs (that are not fully interval), for which at least some of the problems can be efficiently solved.

4 Conclusion

We examined some of the basic properties of the bilevel linear programming problem with interval data. For the optimal value range problem, we have shown through a counterexample that the formerly proposed method does not always consider the scenarios yielding the best and the worst optimal values and we have revised the derived results accordingly. From a complexity-theoretical point of view, we have proved that three of the decision problems connected to the properties of bilevel interval linear programs are NP-hard, namely the problem of checking the existence of a weakly feasible or a weakly optimal solution and the problem of checking unboundedness of at least one scenario of the program.

Acknowledgements

The authors were supported by the Czech Science Foundation under Grant P403-20-17529S. E. Garajová and M. Hladík were also supported by the Charles University project GA UK No. 180420.

References

- [1] Bard, J. F.: An Efficient Point Algorithm for a Linear Two-Stage Optimization Problem. *Operations Research* **31** (1983), 670–684.
- [2] Bialas, W., and Karwan, M.: On two-level optimization. *IEEE Transactions on Automatic Control* **27** (1982), 211–214.
- [3] Calvete, H. I., and Galé, C.: Linear bilevel programming with interval coefficients. *Journal of Computational and Applied Mathematics* **236** (2012), 3751–3762.
- [4] Chinneck, J. W., and Ramadan, K.: Linear programming with interval coefficients. *J Oper Res Soc* **51** (2000), 209–220.
- [5] Dempe, S., Kalashnikov, V., Pérez-Valdés, G. A., and Kalashnykova, N.: Linear Bilevel Optimization Problem. In: *Bilevel Programming Problems: Theory, Algorithms and Applications to Energy Networks* (Dempe, S., Kalashnikov, V., Pérez-Valdés, G. A., and Kalashnykova, N., eds.), Energy Systems. Springer, Berlin, Heidelberg, 2015, 21–39.
- [6] Deng, X.: Complexity Issues in Bilevel Linear Programming. In: *Multilevel Optimization: Algorithms and Applications* (Migdalas, A., Pardalos, P. M., and Värbrand, P., eds.), Nonconvex Optimization and Its Applications. Springer US, Boston, MA, 1998, 149–164.
- [7] Garajová, E., and Hladík, M.: Checking weak optimality and strong boundedness in interval linear programming. *Soft Computing* **23** (2019), 2937–2945.
- [8] 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.
- [9] Hansen, P., Jaumard, B., and Savard, G.: New Branch-and-Bound Rules for Linear Bilevel Programming. *SIAM Journal on Scientific and Statistical Computing* **13** (1992), 1194–1217.
- [10] Hladík, M.: Optimal value range in interval linear programming. *Fuzzy Optim Decis Making* **8** (2009), 283–294.
- [11] Li, H., and Fang, L.: An Efficient Genetic Algorithm for Interval Linear Bilevel Programming Problems. In: *Ninth International Conference on Computational Intelligence and Security*. 41–44.
- [12] Li, H., and Fang, L.: An Evolutionary Algorithm Using Duality-Base-Enumerating Scheme for Interval Linear Bilevel Programming Problems. *Mathematical Problems in Engineering* **2014** (2014), e737515.
- [13] Mishmast Nehi, H., and Hamidi, F.: Upper and lower bounds for the optimal values of the interval bilevel linear programming problem. *Applied Mathematical Modelling* **39** (2015), 1650–1664.
- [14] Ren, A., and Wang, Y.: A cutting plane method for bilevel linear programming with interval coefficients. *Annals of Operations Research* **223** (2014), 355–378.
- [15] Rohn, J.: Interval linear programming. In: *Linear Optimization Problems with Inexact Data* (Fiedler, M., Nedoma, J., Ramík, J., Rohn, J., and Zimmermann, K., eds.). Springer US, Boston, MA, 2006, 79–100.

An Efficiency Comparison of the Life Insurance Industry in the Selected OECD Countries with Three-Stage DEA Model

Biwei Guan ¹

Abstract. In this paper, we use the three-stage data envelopment analysis model to evaluate the efficiency score of 12 life insurance markets from OECD and make a comparison of them. In the first stage, we used the basic DEA model, and in the second stage, we use stochastic frontier analysis slack regression to remove the impact of environmental effects and statistical noise on the efficiency score. After the adjustment according to the second stage, we recalculate the efficiency score of each market. We find the environmental factors have little effect on the German, Ireland and Italy life insurance market, they perform great. But the environmental factors have a heavy effect on Belgium, Greece and Hungary. After removing the influence of environmental factors, the technical efficiency of Belgium, Greece, and Hungary decreased significantly.

Keywords: Panel data, Three-stage DEA model, Life-insurance, OCED countries

JEL Classification: C67, G15, G22

AMS Classification: 60H99

1 Introduction and Literature Review

In recent years, efficiency measurement related to the insurance industry was popular and attracted many regulators' and investors' attention. Eling and Luhnen [7] mentioned from 2000 to 2010, more than 90 studies focused on the efficiency measurement of insurance industry. Kaffash et al. [13] pointed out that from 1993 to 2018, 132 studies on the application of DEA in insurance industry were published. Nowadays, the number of research on this topic has continued to grow. For the measurement of efficiency, there are two main methods: stochastic frontier analysis (SFA) and data envelopment analysis (DEA). In the beginning, Farrell [9] introduced the basic DEA model to evaluate the efficiency of modern companies; on this basis, Charnes et al. [2] and Banker et al. [1] introduced the CCR model and BCC model respectively. Later, some researchers pointed out that the traditional DEA model ignored the influence of environmental effects and statistical noise on decision-making units (DMUs). Fried et al. [10] proposed a three-stage DEA model to eliminate the influence of the above two factors.

The purpose of this paper is to compare the efficiency of 12 OECD life insurance industries from 2013 to 2019 through the three-stage DEA model. In the second stage, the stochastic frontier analysis (SFA) slack regression is helped to eliminate the influence of environmental effects and statistical noise on decision-making units, to obtain more accurate efficiency score. This paper is divided into four sections. Section 2 is data and methodology, here we will introduce the specific information of the three-stage DEA model, and explain the source of the data and why it was chosen. Section 3 is the empirical results, in this part, the efficiency score of each insurance markets will be shown, and the comparison of these results. The last part is the conclusion, which includes the main contribution and key findings of this paper.

In previous efficiency studies, the types and numbers of insurance markets selected are different. Diacon [4] applied the DEA model to analyze the technical efficiency (TE) of the 6 OECD general insurance markets; Diacon et al. [5] studied the pure technical efficiency (PTE) and scale efficiency (SE) of 15 OECD life insurance markets through the DEA model; Davutyan and Klumpes [3] studied the TE, PTE and SE of 7 OECD life insurance markets and non-life insurance markets under the DEA model. Huang and Eling [12] pointed out that the relationship between solvency and efficiency of insurance firms is positive; Hardwick et al. [11] mentioned that the cost efficiency (CE) of the life insurance companies is directly proportional to the existence of audit committees and inversely proportional to the existence of external directors; Jakob et al. [15] concluded that risk management is significantly related to investment management efficiency (IME). Therefore, efficiency measurement plays an important role in the analysis of the insurance industry and is the basis of all deeper analyses. In this paper, we will calculate the value of TE, PTE, and SE of 12 selected OECD life insurance industries. In the second stage of the DEA model, in addition to the inputs and outputs variables required in the traditional model, we also need to

¹ VŠB – Technical University of Ostrava, Department of Finance, Sokolská tř. 33, Ostrava 70200, Czech Republic, biwei.guan@vsb.cz

select environmental variables. Huang and Eling [12] also selected the ratio of shareholder equity to assets, liabilities to liquid assets ratio, and premiums to surplus ratio as indicators of the insurance market regulatory environment. In the existing research, the choice of input variables and output variables are also very different. Kaffash et al. [13] found that as output variables, "premiums" accounted for 50.82%, "losses and incurred losses" accounted for 22.13%, and "investment income" accounted for 21.31%, while as input variables, "the number of employees", "capital debt", "equity capital" and "materials and business services" accounted for 60.72%, 49.18%, 37.7%, and 32.79% respectively. The variables we selected are slightly different from the above variables, which we will explain in detail in the next chapter.

2 Data and Methodology

2.1 Data

This paper selects the relevant data of 12 OECD life insurance markets from 2013 to 2019, the data mainly from OECD (2014-2020). The *input variables* generally used in the previous studies are mainly divided into three categories, namely labor input, capital input, and other material input. (Eling and Jia [6]; Eling and Schaper [8]; and Eling and Luhn, [7]). However, we did not find the number of employees only in the life insurance market, thus, in this paper, the number of companies, debt capital and equity capital are chosen as the input indicators. When considering *output variables*, we can consider the social functions of insurance industry. One of the very important function undertaken by insurance is risk protection. In this paper, we choose the sum of net income plus gross technical provisions and the total investments as the output variables. In the selection of *environmental variables*, we consider the macroeconomic environment of the whole market and the industry environment of the life insurance market itself. We choose the growth of GDP, insurance density and market share as related variables.

In the Three-stage DEA model, we will use SFA slack regression in the second stage to remove the impact of environmental effects and try to adjust the selected DMUs to the same external environment. Table 1 presents the sample summary statistics. From Table 1, we can see that in the input variables, the degree of dispersion of the debt capital is very large; in the output variables, the degree of dispersion of both is very large; in the environment variables, the degree of dispersion of insurance density is largest. Through the observation of the results in Table 1, it is not difficult to find that the maximum values of most indicators are far higher than their average values. If we observe the original data, we will find that this is because the input and output values of the German life insurance market are much higher than those of other markets, which is also the main reason for the high dispersion of various indicators. At the same time, we also find that equity capital is much lower than debt capital, while the two output variables are close.

	Unit	Min	Mean	Max	Std. dev.
Panel A: Input variables					
Number of companies	1	2	35.10	93	24.61
Debt capital	Million US dollars	2761	215989	1264290	329010
Equity capital	Million US dollars	167	6927.70	25254	7321.54
Panel B: Output variables					
Total investments	Million US dollars	1407	197433	1427913	341112
Net income + Technical provisions	Million US dollars	2225	194739	1206903	304032
Panel C: Environmental variables					
Insurance density	US dollars	141	4953.74	48768	10536
Market share	%	0.10	1.41	8.40	1.74
Growth of GDP	%	-3.241	2.556	25.163	3.203
Number of observations			672		

Table 1 Summary of Sample Statistics of 12 OECD Life Insurance Industries

2.2 Three-Stage DEA Model

Data envelopment analysis is suitable for the evaluation of complex multi-output and multi-input problems. We can use DEA model to calculate many kinds of efficiency scores. In this paper, we mainly focus on TE, PTE and SE of the life insurance industry. Table 2 describes these three kinds of efficiency in detail.

Term	Description	Decomposition
Technical Efficiency	TE reflects the ability of a manufacturer to maximize output under a given input, the returns to scale are fixed.	TE=SE×PTE
Pure Technical Efficiency	PTE reflects the production efficiency of the inputs of the DMU at the optimal scale, the returns to scale can be changed.	PTE=TE/SE
Scale Efficiency	SE reflects the gap between the actual scale and the optimal production scale.	SE=TE/PTE

Table 2 DEA Efficiency Terms

The First Stage: Calculate Efficiency using Unadjusted Input or Output Variables

In this paper, we select the input-oriented BCC model (Banker et al. [1]) and input-oriented CCR model (Charnes et al. [2]) to calculate the required efficiency. The assumption of the CCR 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:

$$\min \theta \quad (1)$$

$$s.t. \sum_{j=1}^n \lambda_j x_{ij} \leq \theta x_{i0} \quad (2)$$

$$\sum_{j=1}^n \lambda_j y_{rj} \geq y_{r0} \quad (3)$$

$$\lambda_j \geq 0, i = 1, 2, \dots, n; j = 1, 2, 3, \dots, n; r = 1, 2, \dots, n \quad (4)$$

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 12 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.

BCC model has almost the same constraints as the CCR model. The only difference is that in the BCC 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:

$$\sum_{j=1}^n \lambda_j = 1 \quad (5)$$

The Second Stage: Adjusting Input or Output Variables with SFA Slack Regression

When using SFA slack regression to regress the slack variables in the first stage, we need to consider whether to adjust the input and output variables at the same time or to adjust only one of them. Fried et al. [10] proposed that this depends on the type of oriented we choose in the first stage. In this paper, we choose the input-oriented, so in the second stage, we only adjust the input variables. In addition, Fried et al. [10] mentioned that we should perform a separate regression for each different slack variable, which allows environmental variables to have different effects on different slack variables. We can construct the following SFA slack regression functions:

$$S_{ni} = f(Z_i; \beta_n) + v_{ni} + \mu_{ni} \quad (6)$$

$$i=1,2,\dots,I; n=1,2,\dots,N \quad (7)$$

where S_{ni} is the slack value of n -th inputs on i -th DMU; Z_i represents the environmental variables, β_n represents the coefficient of environmental variables; v_{ni} represents the statistical noise and μ_{ni} represents the managerial inefficiency. $v \sim N(0, \sigma_v^2)$ is the random error term, it can represent the influence of statistical noise on input slack variables; $\mu \sim N^+(0, \sigma_\mu^2)$ can represents the influence of managerial inefficiency on input slack variables.

As mentioned earlier, using SFA slack regression helps to eliminate the influence of statistical noise and environmental effects. Therefore, we need to adjust the input variables. The adjustment formula is as follows:

$$X_{ni}^A = X_{ni} + \left[\max(f(Z_i; \hat{\beta}_n)) - f(Z_i; \hat{\beta}_n) \right] + [\max(v_{ni}) - v_{ni}] \quad (8)$$

$$i=1,2,\dots,I; n=1,2,\dots,N \quad (9)$$

where X_{ni}^A represents the adjusted input variables; X_{ni} is the original input variables; $[\max(f(Z_i; \hat{\beta}_n)) - f(Z_i; \hat{\beta}_n)]$ represents the adjustment of input variables based on the environmental effects; $[\max(v_{ni}) - v_{ni}]$ represents the adjustment of input variables based on the statistical noise. To calculate the statistical noise, we have the following formulas:

$$E(\mu|\varepsilon) = \sigma_* \times \left[\frac{\phi(\frac{\lambda\varepsilon}{\sigma})}{\Phi(\frac{\lambda\varepsilon}{\sigma})} + \frac{\lambda\varepsilon}{\sigma} \right] \quad (10)$$

$$\sigma_* = \frac{\sigma_\mu \sigma_v}{\sigma} \quad (11)$$

$$\sigma = \sqrt{\sigma_\mu^2 + \sigma_v^2} \quad (12)$$

$$\lambda = \frac{\sigma_\mu}{\sigma_v} \quad (13)$$

$$E[v_{ni}|v_{ni} + \mu_{ni}] = S_{ni} - f(Z_i; \beta_n) - E[\mu_{ni}|v_{ni} + \mu_{ni}] \quad (14)$$

The Third Stage: Calculate Efficiency using Adjusted Input or Output Variables

In this stage, the adjusted input variables are re-applied to the DEA model of the first stage to obtain a new efficiency score. Compared with the results of the first stage, the adjusted results will be more accurate, because all DMUs are adjusted to the same external environment, and the impact of statistical noise is proposed.

3 Empirical Results

In the first stage, using DEAP 2.1 can help us get the initial efficiency score of TE, PTE, and SE; then in the second stage, using Frontier 4.1 can help us adjust the input variables, in the second stage, we use the input orientation and select the cost function; in the third stage, we use the adjusted input variables and use DEAP 2.1 recalculate the adjusted efficiency score. There are 12 life insurance industries as the DMUs, the period is seven years, from 2013 to 2019. When we use DEAP 2.1 to calculate the efficiency score, we first need to choose the specific model to use. We selected 12 decision-making units for seven years. Unlike the direct calculation of section data, when we use panel data, we have two options. We can split the panel data into cross-section data and calculate the efficiency score respectively and summarize them; or use the Malmquist model, directly use the panel data. The Malmquist model can be used to measure productivity change, the productivity change can be decomposed into technical change and technical efficiency change. Through the Malmquist model, we can also get TE and PTE of each market every year, and then we can calculate SE. But there is a problem: when we use the Three-stage DEA model, in the second stage, when we adjust the input variables, we need to use the input slacks. Using the Malmquist model can not get input slacks. Thus, we choose the first way to deal with the application of panel data in DEAP 2.1. Table 3 (a) presents the initial efficiency score of the DMUs.

	<i>(a) Initial efficiency score from stage 1</i>			<i>(b) Adjusted efficiency score from stage 3</i>		
	<i>TE</i>	<i>PTE</i>	<i>SE</i>	<i>TE</i>	<i>PTE</i>	<i>SE</i>
Belgium	0.961	0.992	0.969	0.527	0.999	0.528
Denmark	0.952	0.977	0.974	0.932	0.987	0.944
Finland	0.993	0.996	0.997	0.908	0.997	0.911
Germany	1	1	1	1	1	1
Greece	0.832	0.977	0.844	0.291	0.996	0.292
Hungary	0.978	1	0.978	0.232	1.000	0.232
Ireland	0.995	0.997	0.997	0.996	0.998	0.999
Italy	0.971	0.978	0.993	0.968	0.979	0.988
Luxembourg	0.982	0.985	0.997	0.946	0.995	0.951
Poland	0.983	0.984	0.998	0.753	0.989	0.762
Portugal	0.954	0.956	0.998	0.761	0.970	0.784
Spain	0.933	0.959	0.973	0.823	0.968	0.850

Table 3 Summary of Efficiency Score from Stage 1 and Stage 3

From Table 3 (a), we can see that all the initial efficiency scores of Germany are 1, which shows that in the German life insurance markets, input resources are not wasted, and all inputs are completely and effectively converted into output. Among them, Greece has the lowest TE and SE, Portugal has the lowest PTE; the PTE has the lowest degree of dispersion, indicating that the PTE of each industry is not very different. One of the reasons is when calculating PTE, returns to scale are variable. In the second stage, we use SFA slack regression to analyze the three input variables separately and calculate the new input variables after eliminating the influence of environmental effects and statistical noise. We compare the original input variables with the adjusted input variables, and the results are shown in the three graphs (a, b, c) in Figure 1.

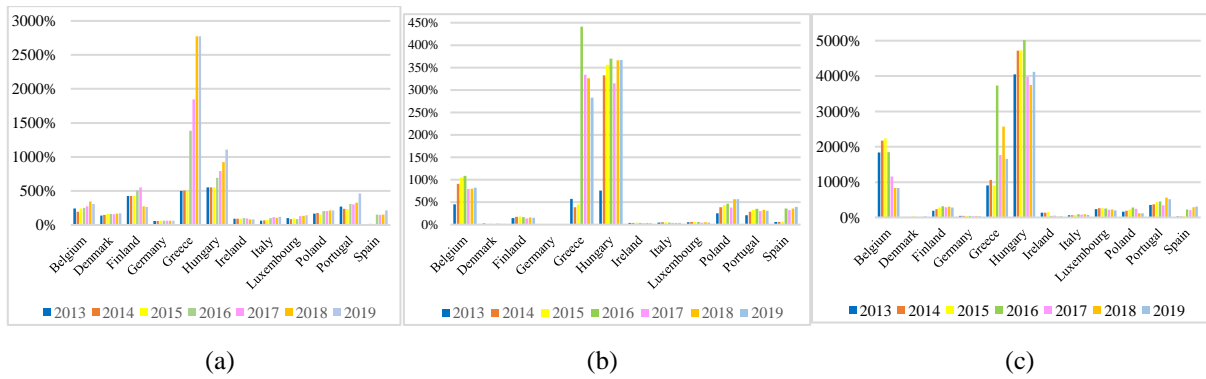


Figure 1 Comparison of the Changes of the Input Variables

For Figure 1(a), it is not difficult to find that after eliminating the influence of the external environment and statistical noise, the value of the number of companies has been greatly increased. In particular, in the Greece life insurance market, the number of companies increased by more than 250% in 2018 and 2019. Among them, the change in Germany is the most average and almost the smallest. Spain changed little in 2013 and 2014. From Figure 1(b), we can see that there is an obvious difference between the original debt capital and the adjusted ones, especially the Greece life insurance market and the Hungary life insurance market. Germany remains the least changed market, and Denmark, Ireland, Italy, and Luxembourg have barely changed. The biggest change is the value of the Greece life insurance market in 2016, which is about 450%. In all life insurance markets, the adjustment ratio of debt capital between 2013 and 2019 is similar every year, except Greece and Spain. From 2016 to 2019, the adjustment ratio of debt capital has increased significantly compared with the previous three years. From Figure 1(c), we can see that the adjustment of equity capital is roughly the same as that of debt capital. The difference is that the equity capital adjustment ratio of the Luxembourg life insurance market is much larger than that of debt capital. We can also see that for equity capital, the adjustment range of each market is significantly larger than that of debt capital, and the highest adjustment range is the Hungary life insurance market. In 2016, the adjustment ratio exceeded 5000%.

After obtaining the adjusted input variables, we calculate the relevant efficiency scores with the new input variables. The results are shown in Table 3(b). From Table 3(b), we can see that the efficiency score of the German life insurance market remains at 1. In order to feel the difference between the original efficiency score and the adjusted efficiency score more intuitively, we made Figure 2.

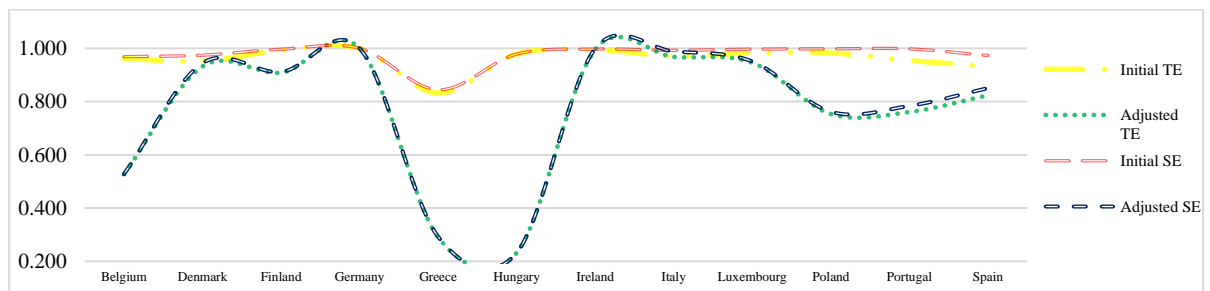


Figure 2 Comparison of Initial Results and Adjusted Results

From Figure 2, we can see that after adjusting the input variables, the efficiency score rankings of the Hungarian life insurance market and the Greek life insurance market are exchanged. Although other markets have changed in varying degrees, almost all of them remain in the original ranking. We also can find that the technical efficiency and scale efficiency of each market have decreased significantly, but the pure technical efficiency has increased slightly. The three most changing life insurance markets are Hungary, Greece, and Belgium. There is little change in the efficiency scores of Ireland and Italy, which shows that environmental variables have little influence on these two markets.

4 Conclusion

This paper analyzes the efficiency of 12 life insurance markets. As members of OCED and EU, these 12 markets have high comparability. We calculated and compared the technical efficiency, pure technical efficiency, and scale efficiency of these markets from 2013 to 2019.

In the first stage, we used the input-oriented CCR model and the input-oriented BCC model to get the original efficiency score; in the second stage, we used SFA slack regression to perform regression analysis on input variables only, after eliminating the impact of environmental effects and statistical noise, we maintained these 12 life insurance markets in a more similar environment, and got the adjusted new input variable values; in the third stage, we used the new variables to recalculate the relevant efficiency score, the original efficiency and the adjusted efficiency have more significant changes. By comparing the efficiency scores of life insurance markets in the first and third stages, we find that the Hungarian life insurance market and Greek life insurance market are most significantly affected by environmental variables, while the Irish life insurance market and Italian life insurance market are almost unaffected. In the first stage, the efficiency scores of Belgium, Greece, and Hungary are not bad, but after the adjustment in the second stage, their efficiency scores are greatly reduced. It can be seen that these three markets need to pay more attention to their own environment, so as to timely adjust input variables and reduce losses. In addition, the efficiency score of the German life insurance industry before and after the adjustment is both 1, which is very superior.

In future research, we may will add more environmental variables, and more in-depth study of the specific impact of each environmental variable on the efficiency value. We will also study the relationship between efficiency and profitability, as well as the relationship between efficiency and solvency.

Acknowledgements

The author was supported through the Czech Science Foundation (GACR) under project 20-25660Y and moreover by SP2021/15, 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] Charnes A., Cooper W. W. & Rhodes E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2 (6), 429-444.
- [3] Davutyan, N. & Klumpes, P.J.M. (2008). Consolidation and Efficiency in the Major European Insurance Markets: A Non Discretionary Inputs Approach. Working Paper, Imperial College London.
- [4] Diacon, S.R. (2001). The Efficiency of UK General Insurance Companies. Working Paper, University of Nottingham.
- [5] Diacon, S.R., Starkey, K. & O'Brien, C. (2002). Size and efficiency in European long-term insurance companies: an international comparison. *Geneva Papers on Risk and Insurance*, 27 (3), 444-466.
- [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] Fried, H. O., Lovell, C. A. K., Schmidt, S. S. & Yaisawarng, S. (2002). Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis. *Journal of Productivity Analysis*, 17, 157-174.
- [11] Hardwick, P., Adams, M. B. & Hong, Z. (2003). Corporate governance and cost efficiency in the United Kingdom life insurance industry. Swansea: European Business Management School, University of Wales.
- [12] Huang, W. & Eling, M. (2013). An efficiency comparison of the non-life insurance industry in the BRIC countries. *European Journal of Operational Research*, 226 (3), 577-591.
- [13] 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.
- [14] OECD. (2014-2020). OECD Insurance Statistics.
- [15] [17] Yakob, R., Yusop, Z., Radam, A. & Ismail, N. (2014). Two-stage DEA method in identifying the exogenous factors of insurers' risk and investment management efficiency. *Journal of Sains Malaysian*, 43 (9), 1439-1450.

Determination of Wages in Forestry Depending on the Occurrence of Natural Disasters

David Hampel¹, Lenka Viskotová², Antonín Martiník³

Abstract. This paper deals with determination of wages in the forestry sector in the Czech Republic for the last 15 years. The relation between wages and the total felling volume or salvage felling volume is explored with the aim to build a model usable in the reforestation simulation study currently being prepared. Results based on the Czech Republic data are validated by analogous analysis using relevant data from the Slovak Republic. Data about wages and the volume of logging are obtained from the Czech Statistical Office and the Statistical Office of the Slovak Republic. Multiple regression model and vector autoregression model are employed for assessing possible relationships. The results point to significant dependence of wages on total felling. Causality in the sense of Granger is verified for the Czech Republic. Differences between available data and results for the Czech Republic and the Slovak Republic as well as limitations of this study are discussed. Finally, it is possible to conclude that the realised model can be incorporated into the prepared reforestation simulation study to ensure realistic costs determination in such situations as an intensive bark beetle attack or occurrence of windstorm.

Keywords: forestry in the Czech Republic, forestry in the Slovak Republic, multivariate regression, natural disasters, salvage felling, wages in forestry

JEL Classification: C51

AMS Classification: 62J05

1 Introduction

The research team of our department deals with optimization in reforestation problem from 2012. In [7] and [8] we presented simulation approach for generating data about forestry economics and estimated cost and revenue functions entering optimal control problem solved in [10]. In [9] we reestimate underlying functions based on actual data and recognized that two important factors were omitted in [7]: drought effects and the bark beetle damage. To improve the simulation, it is necessary to include effects of huge bark beetle damage followed by high salvage felling volume on economic parameters of forestry.

Development of salvage felling and its composition is depicted in Figures 1 and 2 separately for the Czech Republic and the Slovak Republic. In detail, windstorm Kyril in 2007 and intensive bark-beetle attack are visible for the Czech Republic; windstorms in 2004 and 2014 followed by bark beetle attacks are present for the Slovak Republic. In the global view, we can see huge salvage felling volume increase in 2017–2019 in the Czech Republic and in 2004–2006 in the Slovak Republic. This dramatic change should at least affect labour demand and push to the wage growth in forestry. Development of forest workers wages is depicted in Figure 3, where we can see increased dynamics in the years associated with salvage felling increase.

Based on conducted literature research we can declare, that there is a lack of serious information and research on wages in forestry sector of the Czech Republic. This is in line with [5], where an employment in the Czech forestry is elaborated from a historical point of view. We can read there, that employment, quantity and structure of workers is one of the most important factors influencing the development and efficiency of forestry. Despite this fact the labor force in the forestry of the Czech Republic is the subject of research, in contrast to many developed countries of the world, only marginally, mostly as an economic cost factor. Systematic scientific research on human resources in forestry economy is on the outskirts of interest in the Czech Republic.

Paper [1] analyzes prices of works related to timber harvesting and skidding in the selected forest stands. Increasing wages of forest workers are mentioned there as the impulse for the investments to machinery. Further, study [3] provides a first-time series of cost factors to be used when modeling and evaluating the cost competitiveness of forest felling and processing operations on a global scale.

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

³ Department of Silviculture, Mendel University in Brno, Zemědělská 1, Brno, Czech Republic, antonin.martinik@mendelu.cz

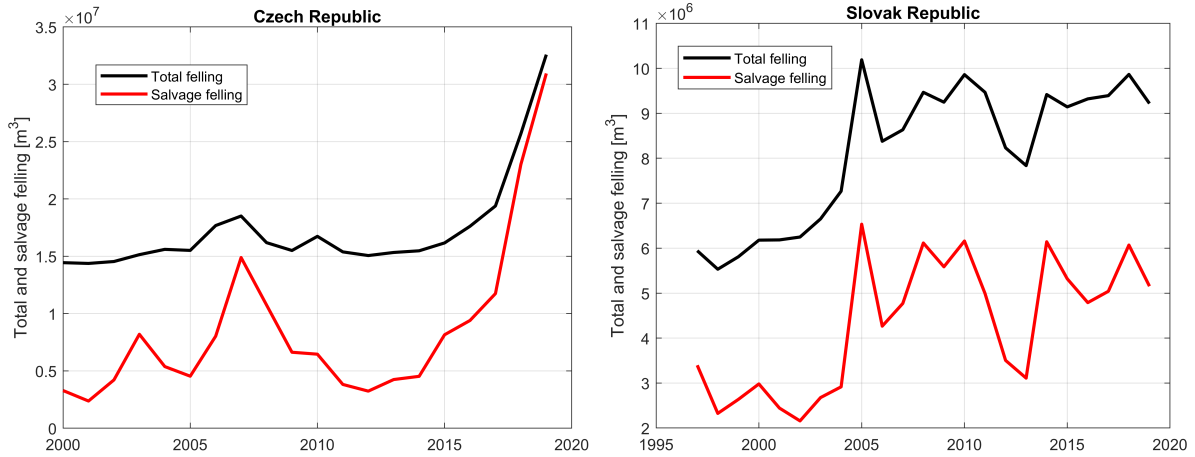


Figure 1 Total felling with the share of salvage felling in the Czech Republic (left graph) and in the Slovak Republic (right graph). Source of data: Czech Statistical Office and the Statistical Office of the Slovak Republic.

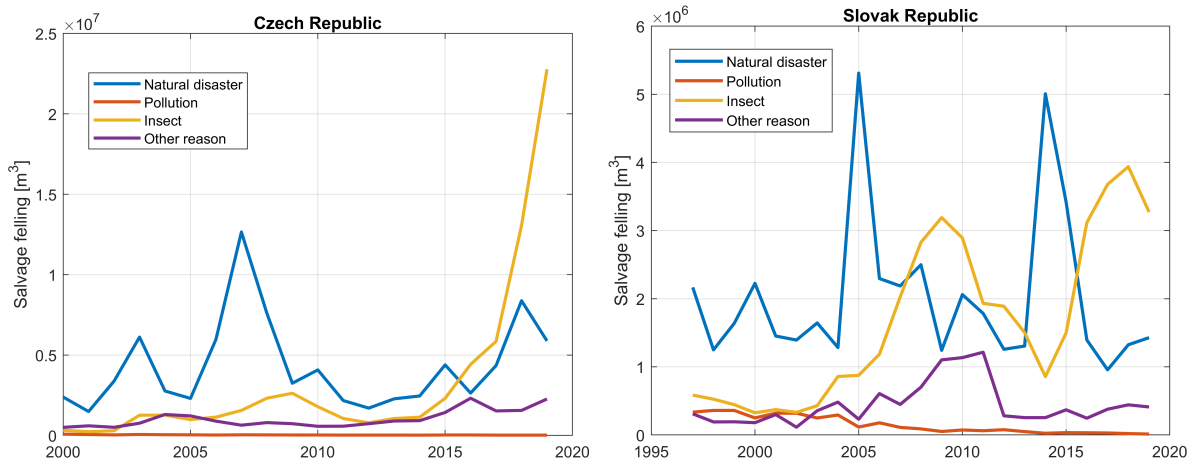


Figure 2 Share of salvage felling caused by specific reason in the Czech Republic (left graph) and in the Slovak Republic (right graph). Source of data: Czech Statistical Office and the Statistical Office of the Slovak Republic.

More specific information on wages in forestry can be found in professional journals. Aleš Erber, forest analyst and professional forest manager, stated in 2018 that forestry had been producing very good economic results in recent years, and without droughts, wind calamities and bark beetles, they would had been even higher. The below-average remuneration of forest workers, sole traders and foresters are therefore a paradox, the result of which is the transfer of human resources to other sectors (see [4]).

The aim of this paper is to estimate and verify the dependency of forest workers wages on the total felling volume in the Czech Republic and the Slovak Republic. The paper is organized as follows: section 2 describes data and methods used, section 3 comprises results and the last section concludes.

2 Material and methods

The data about development of forest workers wages and felling volume related to the Czech Republic were acquired from the Czech Statistical Office, publication *Forestry* (published annually), available years 2002–2019. Analogous data about the Slovak Republic were obtained from Statistical Office of the Slovak Republic, DATAcube database, section 4.1, available years 1997–2019. Because of unexpected behaviour of average wages of forest workers in total non-state forestry sector, we use wages from state forestry sector available only in the years 2000–2019. DATAcube database include interesting data on forestry economics including investment in machinery and economic result, which we include in our analysis for the Slovak Republic.

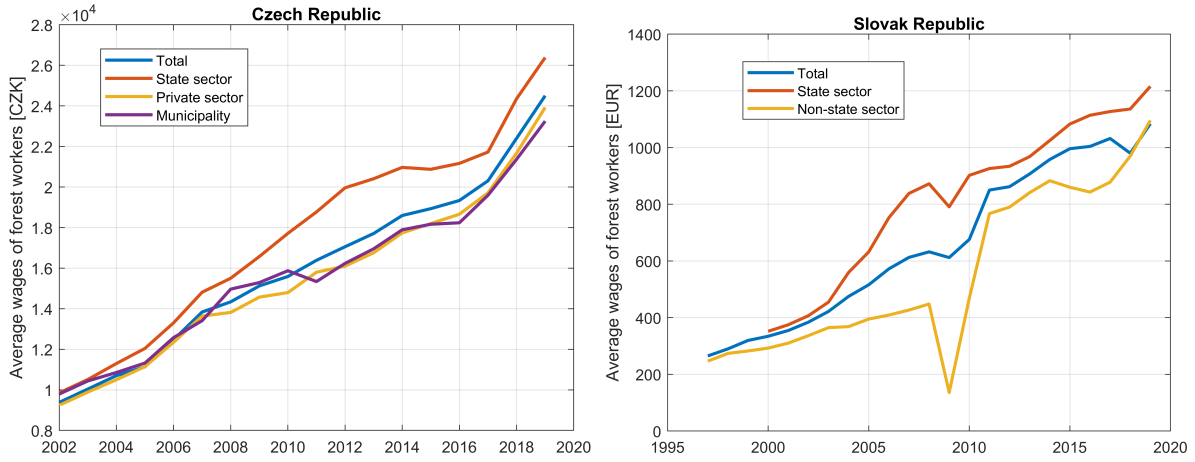


Figure 3 Development of average wages of forest workers for total forestry and for partial sectors in the Czech Republic (left graph) and in the Slovak Republic (right graph). Source of data: Czech Statistical Office and the Statistical Office of the Slovak Republic.

For description and verification of the dependency of forest workers wages on total felling volume we use regression model

$$Wages_t = \beta_0 + \beta_1 f(Felling_t) + \varepsilon_t, \quad (1)$$

where $Wages$ are average monthly wages of forest workers (in CZK for the Czech Republic, in EUR for the Slovak Republic), $Felling$ is total felling volume in m^3 , ε is a random term and t means time index, $t = 1, \dots, T$, T is length of time series. Function f is given in general, we explore characteristics turned out to be the best of common function forms as logarithm, inverse, quadratic and linear. For both countries, linear function form resulting as the best selection. We check stability of model parameters by Quandt likelihood ratio (QLR) test, see [2], with resulting presence of structural change. This might be incorporated into model as

$$Wages_t = \beta_0 + \beta_1 Felling_t + \beta_2 D_t + \beta_3 D Felling_t + \varepsilon_t, \quad (2)$$

where D is indicator of time range after structural change: $D_t = 0$ before structural change, $D_t = 1$ when structural change appears and later. $D Felling$ is element-wise product of $Felling$ and D . For the Slovak Republic we build also the model

$$Wages_t = \beta_0 + \beta_1 Felling_t + \beta_2 D + \beta_3 D Felling_t + \beta_4 Investment_t + \varepsilon_t, \quad (3)$$

where $Investment$ means total forestry investment in machinery. For the final estimated models so-called classical assumptions were tested. Fulfilling these assumptions ensure superior properties of ordinary least squares estimation method. We realize such testing based on following approaches:

- Correct specification of a model – RESET test;
- Homoskedasticity of an error term – White test, Breusch-Pagan test;
- Serial independency of an error term – Breusch-Godfrey test, Ljung-Box test;
- Normality of an error term – Shapiro-Wilk test, Lilliefors test, Jarque-Berra test.

When dealing with time series regression, it is necessary to avoid so-called spurious regression problem. For this purpose, we check cointegration of time series included into regression by testing stationarity of residuals. We use KPSS test for this purpose; stationarity of residues means cointegration of time series and in this case there is no risk of spurious regression. All mentioned tests are fully described for example in [6] and [11].

Another option how to deal with explored data is to estimate multivariate time series model. We employ vector autoregressive (VAR) model for this purpose. By this model we can describe and test dependencies between different lags of variables. Moreover, VAR model can be used for Granger causality testing. Detailed description of VAR model and related tools can be found in [12].

Level of significance was set to $\alpha = 0.05$. The analysis was performed in computational system MATLAB R2020b and Gretl 2020d.

3 Results and Discussion

The estimation of the relation of wages and volume of felling for the Czech Republic starts with model (1). Various forms of function f were examined, linear function was selected finally. The p-value <0.001 of the QLR test points to the existence of structural break in 2009. This is feasible break because of beginning of economics crisis impacts in the Czech Republic. This structural change was described by the model (2) and the results of the estimation are presented in Table 1. Classical assumptions of this model were tested, see Table 4, Model 1, and we cannot reject proper specification of the model, absence of heteroskedasticity and autocorrelation of the random term and normality of the random term. Moreover, the possibility of spurious regression was ruled out with stationary residuals, see KPSS test result. By this model, 93 % of wages variability was explained, what points out to reasonable quality of the model.

Variable	Coefficient	Std. Error	t -ratio	p-value
constant	-1307.86	3934.02	-0.332	0.745
<i>Felling</i>	$9.53 \cdot 10^{-4}$	$2.43 \cdot 10^{-4}$	3.931	0.002
<i>D</i>	14248.30	4043.11	3.524	0.003
<i>DFelling</i>	$-6.41 \cdot 10^{-4}$	$2.47 \cdot 10^{-4}$	-2.591	0.021

Table 1 Estimated model (2) for the Czech Republic

For the data from Slovak Republic, analogous procedure was employed: model (1) was estimated with several forms of function f resulting with linear function finally. The p-value <0.001 of the QLR test points to existence of structural break in 2006. This finding was surprising, but we can justify this by broader context of forest economics in Slovak Republic. It is visible in Figure 4, that windstorm at the end of 2004 followed by high volume of felling in 2005 and 2006 caused high investments in machinery for the state sector of forestry and high profit for the non-state sector of forestry (relatively also in state sector) in 2006. This situation was almost replicated in the years 2014–2016. It is possible to say, that since 2006 Slovak forestry has faced different problems to those before 2006, what reflects into forestry economics. We believe that this explanation confirms estimated structural change.

This structural change was described by the model (2) and estimated, see Table 2. Although this model explains almost 88 % of wages variability, we can not eliminate possibility of spurious regression, see KPSS test result supported by Breusch-Godfrey and Ljung-Box tests results pointing to the presence of autocorrelation in random term (see Table 4, Model 2).

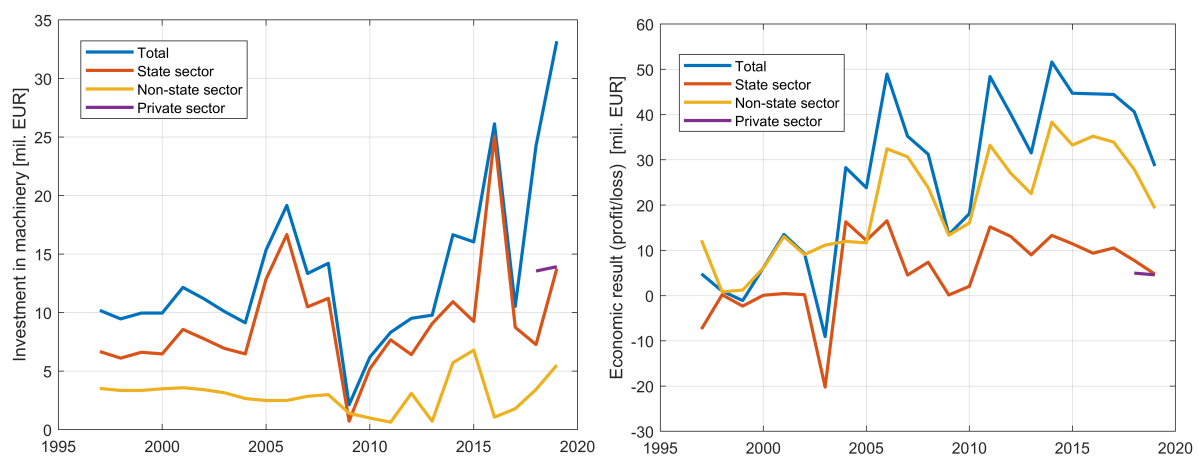


Figure 4 Investment in machinery (left graph) and economic result (right graph) for total forestry and for partial sectors in the Slovak Republic. Note that data for private sector were available for the years 2018 and 2019 only. Source of data: Statistical Office of the Slovak Republic.

Unsatisfactory verification of the estimated model (2) for the Slovak Republic motivates us to find a better model. When exploring further possibilities, we estimated models also with investment in machinery or economic results of forestry, including employing different lags of independent variables. As the best candidate model we selected and estimated model (3) enriched by the investment in machinery variable, see Table 3.

By this manner, we obtain the model with determination of almost 93 %, which does not suffer from spurious

Variable	Coefficient	Std. Error	<i>t</i> -ratio	p-value
const	260.12	128.66	2.022	0.059
<i>Felling</i>	$5.31 \cdot 10^{-5}$	$1.76 \cdot 10^{-5}$	3.020	0.008
<i>D</i>	265.37	49.69	5.340	<0.001

Table 2 Estimated model (2) for the Slovak Republic

Variable	Coefficient	Std. Error	<i>t</i> -ratio	p-value
const	246.72	101.24	2.437	0.027
<i>Felling</i>	$4.49 \cdot 10^{-5}$	$1.40 \cdot 10^{-5}$	3.200	0.006
<i>D</i>	258.50	39.12	6.607	<0.001
<i>Investment</i>	6.35	1.87	3.391	0.004

Table 3 Estimated model (3) for the Slovak Republic

regression or classical assumptions violation, see Table 4, Model 3. Usability of this model in subsequent reforestation simulation is not clear, because of uncertain development of investment in machinery. Moreover, such data are not available in the Czech Republic. On the other hand, the parameters of this model only slightly differs from those of the model in Table 2, what points to the potential usability of Model 2.

Test	Model 1	Model 2	Model 3
RESET test (2 nd and 3 rd powers)	0.243	0.784	0.602
RESET test (2 nd power)	0.134	0.834	0.584
RESET test (3 rd power)	0.166	0.819	0.556
White	0.601	0.263	0.577
Breusch-Pagan	0.879	0.208	0.672
Breusch-Godfrey	0.358	<0.001	0.219
Ljung-Box	0.272	0.004	0.264
Shapiro-Wilk	0.739	0.911	0.555
Lilliefors	0.390	0.570	0.310
Jarque-Berra	0.893	0.957	0.729
KPSS	>0.1	0.039	>0.1

Table 4 Verification of classical assumptions and cointegration relations of estimated models. Tabulated are p-values of the performed tests.

Finally, estimated models are presented graphically in Figure 5. For the Czech Republic, critical increase of wages in 2017–2019 is relatively well described. For the Slovak Republic, it is visible, that both models produce very similar estimate up to the year 2008. From this year, estimate based on investment to machinery shows visibly better results. Nevertheless, the critical wage increase happened in 2004–2006 is described in acceptable way by both models. Another possible model was suggested by reviewer of this paper. When we add *real GDP per capita* to the model (2), this variable appears as significant. For both countries, *Felling* remains significant with positive parameter. Spurious regression was discounted in these models.

Further, we made attempt to model the relationship between forest workers wages and volume of total felling by vector autoregression model. Because of nonstationarity of original time series we differentiate them and estimate VAR model on such differences. For the Czech Republic we estimate acceptable VAR(1) model with three significant parameters. With F-test p-value 0.028 we verify Granger causality from *Felling* to *Wages* with positive impact. For the Slovak Republic we did not find VAR model with significant parameters; we tried to employ *Investment* variable as the third time series with the same result. This can be caused by relatively short time series unsuitable for the VAR model. Vice versa, detected Granger causality in the case of the Czech Republic should not be overrated for the same reason.

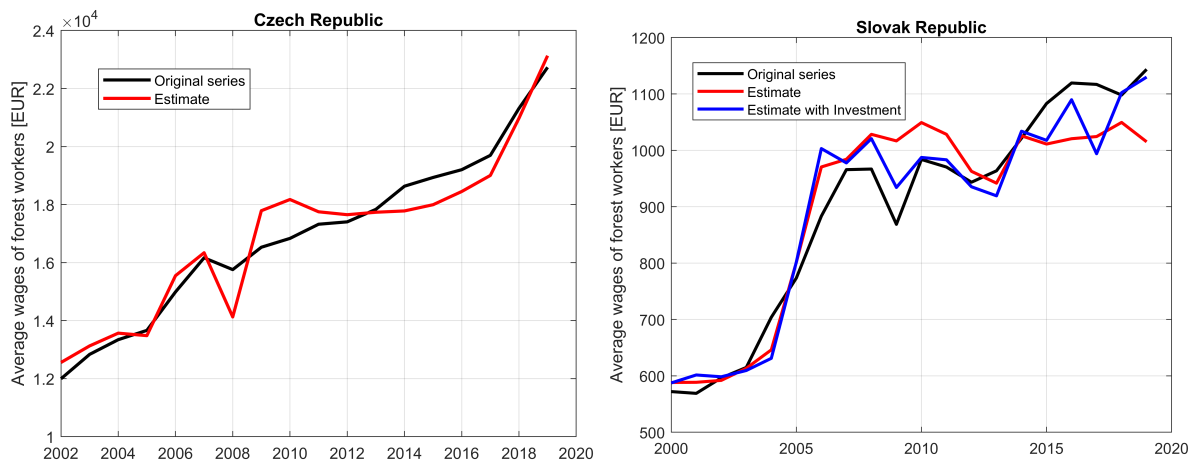


Figure 5 Original and estimated average wages expressed in 2015 prices for total forestry in the Czech Republic (left graph) and for state forestry sector in the Slovak Republic (right graph).

4 Conclusions

Based on the achieved results, it can be summarized that we have empirically demonstrated the relation between the volume of logging and the wages of forest workers in the Czech Republic. An analogous analysis for the Slovak Republic at least partially confirmed this relationship and, in addition, pointed out the importance of involving investment in machinery for a quality explanation of the development of forest workers' wages. This partial discrepancy is caused, among other things, by the different onset and consequences of calamities in the forestry of both countries. In conclusion, we can state that we have estimated a significant mechanism that will allow a more accurate simulation of the impact of major disasters on economic indicators of forestry.

Acknowledgements

This paper presents the results of the research supported by Czech Science Foundation grant GA18-08078S.

References

- [1] Bartoš, L., Máchal, P. & Skoupý, A. (2009). Possibilities of using price analysis in decision making on the use of harvester technology in forestry. *Acta univ. agric. et silvic. Mendel. Brun.*, LVII(4), 31–36.
- [2] Brooks, C. (2009). *Introductory econometrics for finance*. Cambridge: Cambridge University Press.
- [3] Di Fulvio, F., Abbas, D., Spinelli, R., Acuna, M., Ackerman, P. & Lindroos, O. (2017). Benchmarking technical and cost factors in forest felling and processing operations in different global regions during the period 2013–2014. *International Journal of Forest Engineering*, 28(2), 94–105.
- [4] Erber, A. (2018). Causes of low wages in forestry [in Czech]. *Zemědělec* 17/2018.
- [5] Fanta, A. & Šišák L. (2014). Analysis of structure development of employment in the Czech forestry sector from the 1950s to the present. *Zprávy lesnického výzkumu*, 59(3), 160–166.
- [6] Greene, W. H. (2020) *Econometric analysis*. Eighth edition, Global edition. Harlow: Pearson.
- [7] 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). UP v Olomouci.
- [8] Hampel, D., Janová, J. & Kadlec, J. (2015). Estimation of Cost and Revenue Functions for Reforestation System in Dražanská Highlands. In *Proceedings of the International Conference on Numerical Analysis and Applied Mathematics 2014*. Melville: American Institute of Physics (AIP).
- [9] Hampel, D. & Viskotová, L. (2020). Actual Revision of Cost and Revenue Functions for Reforestation System in Dražanská Highlands. In *MME 2020: Proceedings* (pp. 148–153). MENDELU v Brně.
- [10] 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.
- [11] Kmenta, J. (2011) *Elements of econometrics*. Second edition. Ann Arbor: University of Michigan Press.
- [12] Lutkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin: New York.

Efficiency evaluation of the health care system during COVID-19 pandemic in districts of the Czech Republic

Jana Hanclova¹, Lucie Chytilova²

Abstract. The article deals with the assessment and evaluation of the performance of the health care system in managing the COVID-19 pandemic, which reflects the situation in 77 districts of the Czech Republic at the time of the 3rd peak (March 6, 2021) with 9130 confirmed positive cases. Data envelopment analysis (DEA) is used for this research with 4 inputs (population, incidence in the previous 14 days, the incidence in the previous 14 days at age 65+, capacity of test facilities) and desired output (number of recovered patients) and undesirable output (number of deaths). The DEA model includes non-radial measures and non-proportional changes in output variables.

The results document that it is most appropriate to use the DEA model M3 with the desired reduction in deaths and an increase in the number of recoveries. For this model, 35% of districts in the Czech Republic has been effective. The main problem with failure to be effective was the high number of COVID-19-related deaths.

Keywords: COVID-19, Data envelopment analysis, districts, efficiency, health care system, undesirable output.

JEL Classification: C61, I00, C44

AMS Classification: 90B90

1 Introduction

The onset of the COVID-19 epidemic was December 31, 2019 in China. The degree of infection varies from country to country, from district to district. The first case appeared in the Czech Republic (CR) on March 1, 2020. COVID-19 has since resulted in a high number of deaths and the confirmed cases in the Czech Republic. Figure 1 presents the daily increments of confirmed infected cases from August 27, 2020 to May 25, 2021. Figure 1 also presents the development of the epidemiological situation as a seven-day moving average (red dashed line) with 3 peaks. The analysis of this article will focus on the latest peak around March 3, 2021.

Measuring the CR's COVID-19 response performance is an extremely important challenge for health care policymakers. Also, people and governments in the Czech Republic have been challenged by COVID-19 and its consequences. Social distancing and personal protective measures became the primary means of controlling the spread of COVID-19. There is a number of research questions that researchers are looking for answers to.

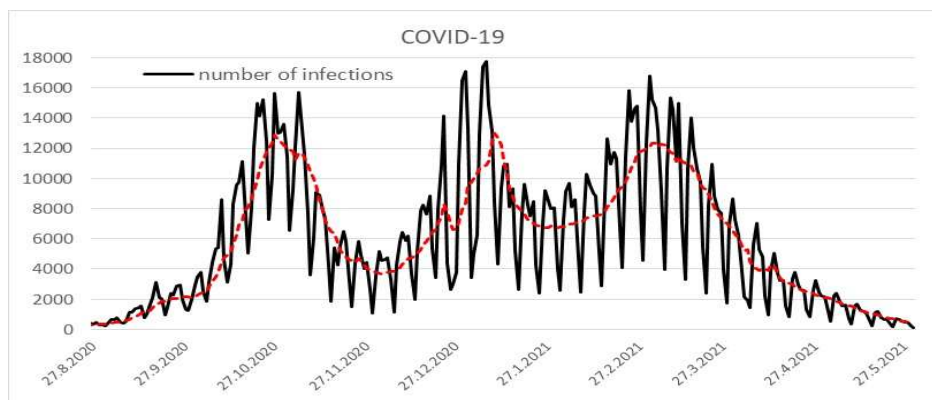


Figure 1 Development of the number of infected in the Czech Republic [<https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19>]

¹ VSB-Technical University of Ostrava, Faculty of Economics, Department of Systems Engineering, Sokolska tr. 2416/33, 702 000 Ostrava, Czech Republic, jana.hanclova@vsb.cz.

² VSB-Technical University of Ostrava, Faculty of Economics, Department of Systems Engineering, Sokolska tr. 2416/33, 702 000 Ostrava, Czech Republic, lucie.chytilova@vsb.cz.

In this study, we will focus on assessing and evaluating the efficiency of COVID-19 response performance in the districts in the Czech Republic during the 3rd peak period on March 6, 2021 with around 9130 infected confirmed cases.

What are the factors and the structure of the health management system of the COVID-19 pandemic with a focus on the use of data envelopment analysis (DEA)? Hamzan, Yu and See in [4] examined the relative efficiency level of managing COVID-19 in Malaysia using network data envelope analysis. A network process consists of 3 sub-processes - community surveillance, medical care I (health care associated with detected positive people) and medical care II (care associated with severe patients requiring intensive care). Dlouhy in [3] used a simple DEA to evaluate the health system efficiency in OECD countries. The researched system included 3 inputs (physicians, nurses, hospital beds) and 2 outputs (population, life expectancy). The authors documented that health resources have to be mobilized in a short time. In conclusion, most of the articles that examine the efficiency of systems during the COVID-19 pandemic are at the country level, respectively. States (USA), focus mainly on efficiency in hospitals and the structure corresponds to simple and network DEA systems.

This article aims to examine the efficiency of COVID-19 response performance in the 77 districts in the Czech Republic during March 2021. The results of this study fill gaps in the literature in terms of assessing the performance of the health system in managing COVID-19 in small districts using DEA with non-radial measures and non-proportional changes.

The rest of the paper is organized as follows. Section 2 introduces the DEA basic methodology, including non-radial measures with non-proportional changes. Section 3 describes the data and Section 4 presents the results. Section 5 includes conclusions and options for future research.

2 Data Envelopment Analysis

For modelling the N district system ($j = 1, 2, \dots, N$), we will consider a general conceptual DEA model with R desirable outputs y and S undesirable outputs b . We also assume the set of I multiple inputs x . The multiple-output production technology with emphasis on input-specific technology can be described as $T_1(x) = \{(y, b) : x \text{ can produce } (y, b)\}$.

To assess the efficiency of COVID-19 response performance, we will use the distance directional function (DDF), which Chung et al in [1] introduced as the joint production of desirable output y and undesirable output b :

$$\bar{D}_{T_1}(x, y, b, g^y, g^b) = \sup\{\beta | (y, b) + \beta(g^y, g^b) \in T_1(x)\}, \quad (1)$$

where the nonzero vector (g^y, g^b) is the *direction vector* and β expresses the intensity of the increase in the desired production while reducing unwanted production and is referred to as the *scaling factor*. At the same time, Toloo and Hanclova in [6] set the condition: $\bar{D}(x, y, b, g^y, g^b) \geq 0$ if and only if $(y, b) \in T_1(x)$. This DDF function moves the joint production (y, b) along the direction (g^y, g^b) to place it on the production frontier. Zhou et al showed in [7] that the radial method for measuring efficiency can be overestimated if we have non-zero slack variables, and therefore introduced non-radial measures. Toloo et al in [5] and Chytilova and Hanclova in [2] used the DEA model with non-radial measures, where the direction vector is $g = (g^x, g^y, g^b)' = (-x_o, y_o, -b_o)'$.

In our paper, we will use *output-oriented DEA* models with *non-proportional changes* in outputs using a general scaling vector $\beta' = (\beta^x, \beta^y, \beta^b)$, we can formulate a DEA model (2) to determine the efficiency of COVID-19 responsive performance in general:

$$(2)$$

$$\begin{aligned}
z &= \bar{D}_{T_2}(\mathbf{x}, \mathbf{y}, \mathbf{b}, \mathbf{g}^x, \mathbf{g}^y, \mathbf{g}^b) = \max \{ \mathbf{w}^x \cdot \boldsymbol{\beta}^x + \mathbf{w}^y \cdot \boldsymbol{\beta}^y + \mathbf{w}^b \cdot \boldsymbol{\beta}^b \} = \boldsymbol{\beta}^* \\
s.t. \quad & \sum_{j=1}^N \lambda_j x_{ij} \leq (1 - \beta_i^x) x_{io} \quad i = 1, \dots, I \\
& \sum_{j=1}^N \lambda_j y_{rj} \geq (1 + \beta_r^y) y_{ro} \quad r = 1, \dots, R \\
& \sum_{j=1}^N \lambda_j b_{sj} \leq (1 - \beta_s^b) b_{so} \quad s = 1, \dots, S \\
& VRS : \sum_{j=1}^N \lambda_j = 1 \quad \lambda_j \geq 0 \quad \forall i, r, s \\
& \beta_i^x, \beta_r^y, \beta_s^b \geq 0 \quad \forall i, r, s
\end{aligned}$$

where vector $\mathbf{w}' = (\mathbf{w}^x, \mathbf{w}^y, \mathbf{w}^b)$ is the normalized weight vector and we assume that the weight of all inputs, desirable and undesirable outputs is gradually 1/3. Model (3) is similar to the additive DEA model in the sense that both attempt to identify the potential slacks in inputs and outputs as much as possible. The non-radial directional distance function is based on T_2 :

$$\bar{D}_{T_2}(\mathbf{x}, \mathbf{y}, \mathbf{b}, \mathbf{g}^x, \mathbf{g}^y, \mathbf{g}^b) = \sup \{ \mathbf{w}' \boldsymbol{\beta} : |(\mathbf{x}, \mathbf{y}, \mathbf{b}) + \mathbf{g} \cdot \text{diag}(\boldsymbol{\beta}) \in T_2 \}, \quad (3)$$

In the empirical study, we will focus on the comparison of 3 variants of the DEA model (2):

- Model M1: $\mathbf{g} = (\mathbf{g}^x, \mathbf{g}^y, \mathbf{g}^b)' = (\mathbf{0}, \mathbf{0}, -\mathbf{b}_o)'$ \wedge $\mathbf{w} = (\mathbf{w}^x, \mathbf{w}^y, \mathbf{w}^b)' = (\mathbf{0}, \mathbf{0}, \mathbf{1})'$;
- Model M2: $\mathbf{g} = (\mathbf{g}^x, \mathbf{g}^y, \mathbf{g}^b)' = (\mathbf{0}, \mathbf{y}_o, \mathbf{0})'$ \wedge $\mathbf{w} = (\mathbf{w}^x, \mathbf{w}^y, \mathbf{w}^b)' = (\mathbf{0}, \mathbf{1}, \mathbf{0})'$;
- Model M3: $\mathbf{g} = (\mathbf{g}^x, \mathbf{g}^y, \mathbf{g}^b)' = (\mathbf{0}, \mathbf{y}_o, -\mathbf{b}_o)'$ \wedge $\mathbf{w} = (\mathbf{w}^x, \mathbf{w}^y, \mathbf{w}^b)' = (\mathbf{0}, \frac{1}{2}, \frac{1}{2})'$.

All M1 – M3 models use non-radial measures and non-proportional changes. Model M1 is a DEA model-oriented only to decrease undesirable outputs, model M2 is oriented only to increase desired outputs, and model M3 is oriented only to increase undesirable outputs and decrease desired outputs.

To evaluate the efficiency, we will have a total *beta index* and sub-indices $\beta^x, \beta^y, \beta^b$. The index $\boldsymbol{\beta} = \mathbf{0}$ indicates the effective unit, and the higher the beta, the less effective the unit (district). Evaluation using the *y-b performance index IBPI* can generally be determined on the basis of relation (4) according to the publication of authors Zhou et al in [7] in the case of the output-oriented model with undesirable output:

$$YBPI = \frac{(1 - \beta^{b*})}{(1 + \beta^{y*})} \quad (4)$$

The numerator in (4) represents the average proportion by which the undesirable output can be reduced, while the denominator expresses the degree to which the desired output can be increased. The *YBPI* index is standardized between 0 and 1, $YBPI = 1$ means that the district is located at the frontier of best practice.

3 Data

To determine the efficiency of COVID-19 response performance, we will use the DEA model (2) in the three mentioned variants of models M1 - M3.

The DEA model included 4 inputs (population, total incidence in the previous 14 days, the incidence of people over 65 in the previous 14 days, and testing capacity of the facility in the district) and 1 desired output (number of recovered patients) and 1 undesirable output (number of deaths associated with COVID-19). Table 1 summarizes the description of the individual variables in the DEA model. The source of the POP variable is the Czech Statistical Office, the source of data for indications IN14 and IN65 is closed data sets for predictive modelling, which were provided to us after sending an official request for access to the Ministry of Health of the Czech

Republic. The CAP data source is the open data set *Přehled odběrových míst*³ in the testing section. The source of data for both outputs Y and B are open data sets Anti-epidemic System of the Czech Republic (PES)⁴.

This study is devoted to the analysis of 3 variants of DEA models for 77 districts of the Czech Republic as of March 6th, 2021. For inputs IN14 and IN65, the sum of daily cases for the previous 14 days is cumulated, i.e. from February 20, 2021 to March 6, 2021. Conversely, for both outputs Y and B, the sum of the number of persons for 14 days later, i.e. from March 6, 2021 to March 20, 2021.

I/O	ID	Variable	Description	Unit per district
inputs	POP	population	number of persons	number of persons
	IN14	incidence in the last 14 days	number of confirmed positive cases in the last 14 days	number of persons
	IN65	incidence in the last 14 days for 65+	number of confirmed positive cases at the age of 65+ in the last 14 days	number of persons
desirable output	CAP	testing capacity	total maximum capacity of testing facilities	number of persons
desirable output	Y	recovered patients	number of patients recovered over the next 14 days	number of persons
undesirable output	B	deaths	number of deaths in the next 14 days	number of persons

Table 1 Description of variables in the DEA model [Source: ČSÚ³, MZ ČR⁴]

4 Results

All three variants of the model (3) were optimized using the GAMS software. The first part of the analysis is devoted to the comparison of the values of the average value BETA_M1, BETA_M2 and BETA_M3 for individual models M1, M2 and M3. A *zero-beta value* means that the district is located at the frontier of best practice, i.e. it is efficiency.

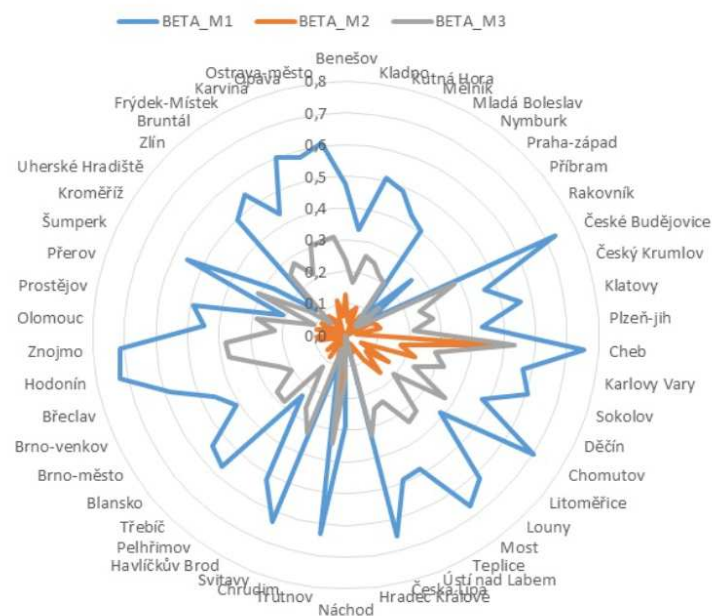


Figure 2 Comparison of levels BETA_M1 to BETA_M3 [own calculation]

³ <https://www.czso.cz/csu/czso/pocet-obyvatel-v-obcich-k-112021>

⁴ <https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19>

For the variant of the M1 model, where only a reduction in the number of deaths is considered, 27 efficiency units (35%) were indicated. For the variance of M2, where an increase in the number of recovered patients was allowed, there were 28 efficiency units (36%) and in comparison, with M1, the district of Hradec Králové was also effective. In the case of the M3 model, where it was desirable to increase the number of cured patients and reduce the number of deaths, 27 districts (35%) lie on the efficiency frontier. At the same time, 27 districts were efficient for all examined models M1 to M3. Figure 2 shows the level of inefficiency (beta) of districts according to the models M1 to M3.

For the M3 model, the districts of Cheb (with a beta value of 0.532), Znojmo (0.382), České Budějovice (0.379), Hodonín (0.374), Trutnov (0.344), Most (0.338), Svitavy (0.330), Louny (0.327), Česká Lípa (0.327), Sokolov (0.324) and Ostrava město (0.313) belonged to the quartile with the worst efficiency in managing the COVID-19 pandemic (at the time of the third peak in the number of infected persons in the Czech Republic around March 6, 2020). The main reason for not reaching the efficiency limit was the problem of the high number of deaths of patients compared to efficiency units. Here, in the future, it will be necessary to make further detailed analysis and look for the facts that caused this.

Further efficiency analysis was performed based on the *y-b performance of the IBPI index* from Equation (4). This index expresses the ratio of the average decrease in the number of deaths to the average increase in the number of recovered patients.

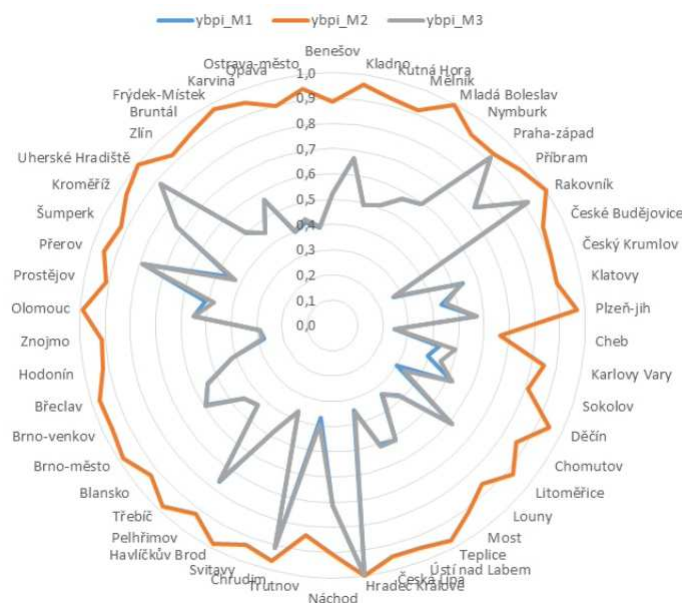


Figure 3 Comparison of YBPI_M1 to YBPI_M3 [own calculation]

Figure 3 shows the level of this index for inefficiency districts for the 3rd of the COVID-19 pandemic. The *IBPI* index is almost identical in terms of level M1 and M3, which is documented by a statistically significant correlation of 0.999. The number of effective districts corresponds to the previous analysis with beta coefficients. Furthermore, we will again focus on the least efficiency quartile according to the YBPI for the M3 model, which prefers a reduction in death and an increase in the number of recovered patients. The least efficiency districts managing the COVID-19 pandemic were confirmed by the districts of Cheb (0.252), České Budějovice (0.270), Hodonín (0.284), Znojmo (0.288), Chomutov (0.336), Most (0.229), Česká Lípa (0.356), Svitavy (0.365), Louny (0.387), Ostrava-město (0.393), Trutnov (0.398), Karviná (0.403) and Břeclav (0.419). In comparison with the evaluation of the least efficiency quartile according to beta, the districts of Karviná and Břeclav were included in the place of the district of Sokolov, which is also related to the calculation of quartile boundaries.

5 Conclusion

The presented paper is devoted to assessing and evaluating the management of the pandemic situation COVID-19 at the time of the peak (according to the number of infected persons) in the districts in the Czech Republic. An analysis of the data envelope with models of output-oriented, non-radial measures and non-proportional changes

of output variables were used for this research. The obtained results document that the proposed models can be used in practice with the intention of prevention in other pandemic situations.

The results showed that of the three proposed variants of the models, the most suitable model is the M3 model, which aims to reduce the number of patients who have died, but also to increase the number of recovered patients. In the evaluation of the health care system at the time of the peak of the COVID-19 pandemic (March 6, 2021), 35-36% of effective districts were demonstrated for all variants of the M1 to M3 models. The average beta inefficiency was rated 0.158 for the M3 model. The analysis of the results of the M3 model further showed that in order to improve the efficiency of the health care system, the main problem is to reduce the number of deaths, which is confirmed by the average inefficiency. The need to reduce undesirable output is significant compared to the average inefficiency of increasing the number of recovered patients. The empirical study also pointed to a group of districts in the "worst" quartile, i.e. with the worst level of efficiency, where policymakers, the Ministry of Health and other institutions need to pay attention to the prevention of possible similar pandemics. The obtained results also have their limits. This is mainly an analysis at the district level in the Czech Republic at the time of the peak of the COVID-19 pandemic and the possibility of obtaining relevant data files.

Further research in this area will focus on the possibility of extending the proposed DEA models to the DEA network group, spatially orienting to the state level due to more accessible data, examining the homogeneity of service units, reallocation of resources and development of health care infrastructure.

Acknowledgements

This research was supported by the Czech Science Foundation within the project GA 19-13946S and the Student Grant Competition (SGS) within the project SP2021/51.

References

- [1] Chung, Y. H., Färo, R. and Grosskopf, S. (1997). 'Productivity and undesirable outputs: A directional distance function approach'. *Journal of Environmental Management*, 51(3), 229–240.
- [2] Chytilova, L. and Hanclova J. (2019). 'Estimating the environmental efficiency of European cross-countries using a non-radial general directional distance function'. *Proceedings of 37th International Conference on Mathematical Methods in Economics (MME 2019)*, České Budějovice, Czech Republic, pp. 541-546.
- [3] Dlouhy, M. (2020). 'Health System Efficiency and the COVID-19 Pandemic'. *Proceedings of 38th International Conference on Mathematical Methods in Economics (MME 2020)*, Brno, Czech Republic, pp. 80-84.
- [4] Hamzan, N.M., Yu, M.M., See, K.F. (2021). 'Assessing the efficiency of Malaysia health system in COVID-19 prevention and treatment response'. *Health Care Management Science*, 24, pp. 273-285.
- [5] Toloo, M., Allahyar, M. and Hanclova, J. (2018). 'A non-radial directional distance method on classifying inputs and outputs in DEA: Application to banking industry'. *Experts Systems with Applications*, 92, pp. 495-506.
- [6] Toloo, M. and Hanclova, J. (2020). 'Multi-valued measures in DEA in the presence of undesirable outputs'. *Omega International Journal of Management Science*, 94, pp. 1-11.
- [7] Zhou, P., Ang, B.W., Wang, H. (2012). 'Energy and CO2 emission performance in electricity generation: A non-radial directional distance function approach'. *European Journal of Operational Research*, 221 (3), pp. 625-635.

Analysis of uneven distribution of diseases COVID – 19 in the Czech Republic

Jakub Hanousek¹

Abstract. The COVID – 19 pandemic affected more, or less each of us. The goal of this article is to measure regional uneven distribution in 77 districts in the Czech Republic. The data cover a period since March 2020 to March 2021 and comes from Institute of Health Information and Statistics of the Czech Republic. The regional variations are measured by dual DEA models with unwanted outputs. Data envelopment analysis is a method based on a linear programming that measure efficiency of production units. The advantage of this method is an optimization weights of each criteria (inputs, outputs) to maximize a score from each unit. The production unit in this paper is a one region in the Czech Republic. There are 77 units. In order to stop the spread of disease, the Government of the Czech Republic applied national prohibitions. The results of this study show us that the regional variation of spread of diseases are huge. The regional prohibition probably would be preferable and more effective in this situation.

Keywords: COVID – 19, DEA, unwanted outputs

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

Disease COVID – 19 is a new disease. This disease is caused by a new type of coronavirus SARS-CoV-2. Coronavirus SARS-CoV-2 was first detected at the end of the year 2019 in the chinese city Wuhan. SARS-CoV-2 have been widespread throughtout the word during the year 2020. The Czech Republic is the one of the most affected country in the world by SARS-CoV-2.

The goal of this article is to meassure inequalities in spread SARS-CoV-2 in the 77 regions of the Czech Republic. Inequalities are meassure by data evelopment analysis (DEA) [1,2]. DEA is based on the theory of linear programming and estimates the production frontier as the piecewise linear envelopment of the data. The production units which lies on a production frontier are effective. Production unit which are lies under the production frontier are inefficent. The production units which are effective have score 1. Inneficent production units in output oriented model have score > 1 . Score means how much proportionally increase outputs to became production unit effective [4].

The input is a population in the region. The oputputs are number of infected people with SARS-CoV-2 and number of death people with SARS-CoV-2. Effective production units in this article are the most affected regions in the Czech Republic. The inneficent production units are regions with better epidemic situation. The score at inneficent production units means how to proporcionally allow icrease outputs (infected and death). The production frontier in this article represent situation that all regions were same as the most affected regions.

2 Methods

DEA was developed by Charnes, Cooper and Rhodes in 1978 [3] and constructs the production frontier and evaluates the technical efficiency of production units. The production unit uses a number of inputs to produce outputs. The technical efficiency of the production unit is is defined as the ratio of its total weighted output to its total weighned input or, vice versa, as the ratio of its total weighned input to its total weighned output. DEA model permits each production unit to choose its input and output weights to maximize its technical efficiency score. A technically efficient production unit is able to find such weights that the production unit lies on the production frontier [5]. The production frontier represents the maximum amounts of output that is produced by given amounts

¹ University of Economist and Business, Prague, Faculty of Informatics and Statistics, Department of Econometrics, Winston Churchil Sq., Prague, Czech Republic, e-mail: xhanj52@vse.cz

of input (the output maximization DEA model) or, alternatively, the minimum amounts of inputs required to produce the given amount of output (the input minimization DEA model). This article deals with two DEA models. The first model is output maximization model with constant revenue from scale. The second model is output oriented model with variable revenue from scale [4].

The effective production units represent in this article the worst districts. The districts which are most effected with virus.

2.1 Output oriented model with constant return of scale

$$\begin{aligned}
 &\text{Maximize: } \varphi_q + \varepsilon (\mathbf{e}^T \mathbf{s}^+ + \mathbf{e}^T \mathbf{s}^-), \\
 &\quad \mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^- = \mathbf{x}_q, \\
 &\text{Subject to: } \mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^+ = \varphi_q \mathbf{y}_q, \\
 &\quad \boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^- \geq \mathbf{0}.
 \end{aligned} \tag{1}$$

φ_q is a variable which represents efficiency rate of a production unit. ε is an infinitesimal constant. The infinitesimal constant $\varepsilon = 10^{-8}$. $\mathbf{e}^T = (1, 1, 1, \dots, 1)$. \mathbf{s}^+ and \mathbf{s}^- are vectors of additional variables. \mathbf{X} is a matrix of inputs. \mathbf{Y} is a matrix of outputs. $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n)$ is a vector of weights which are assign to productions units. Weights are the variables in a model [4].

2.2 Output oriented model with variable constant of scale

$$\begin{aligned}
 &\text{Maximize: } \varphi_q + \varepsilon (\mathbf{e}^T \mathbf{s}^+ + \mathbf{e}^T \mathbf{s}^-), \\
 &\quad \mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^- = \mathbf{x}_q, \\
 &\text{Subject to: } \mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^+ = \varphi_q \mathbf{y}_q, \\
 &\quad \mathbf{e}^T \boldsymbol{\lambda} = \mathbf{1}, \\
 &\quad \boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^- \geq \mathbf{0}.
 \end{aligned} \tag{2}$$

φ_q is a variable which represents efficiency rate of a production unit. ε is an infinitesimal constant. The infinitesimal constant $\varepsilon = 10^{-8}$. $\mathbf{e}^T = (1, 1, 1, \dots, 1)$. \mathbf{s}^+ and \mathbf{s}^- are vectors of additional variables. \mathbf{X} is a matrix of inputs. \mathbf{Y} is a matrix of outputs. $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n)$ is a vector of weights which are assign to productions units. Weights are the variables in a model. $\mathbf{e}^T \boldsymbol{\lambda} = \mathbf{1}$ is a condition of convexity [4].

3 Application

The application of the described methods is illustrated on the data of the Czech Republic. The data comes from Institute of Health Information and Statistics of the Czech Republic.

The Czech Republic is divided into 77 regions and in 2021 it had 10.44 mil. Inhabitants [6]. The number of infected people with virus SARS-CoV-2 since March 2020 to March 2021 was 1,403,809 in the Czech Republic. The number of death people with virus SARS-CoV-2 since March 2020 to March 2021 was 24,331 in the Czech Republic [7]. The data are presented in Table 1.

District	Population	Number of infected	Number of death
Praha	1,268,796	159,523	2,220
Praha západ	131,231	18,920	159
Příbram	112,816	16,700	153
Rakovník	54,993	7,965	143
Benešov	95,459	16,923	327
Beroun	86,160	12,532	124
Kladno	158,799	23,540	347
Kolín	96,001	17,020	254
Kutná hora	73,404	10,911	186
Mělník	104,659	15,917	280
Mladá Boleslav	123,659	19,591	331
Nymburk	94,884	15,479	255
Praha východ	157,146	24,896	227
České Budějovice	186,462	22,556	359
Český Krumlov	60,516	6,434	137
Jindřichův Hradec	90,604	13,374	264
Písek	69,843	10,412	143
Prachatice	50,010	5,754	145
Strakonice	69,786	9,208	193
Tábor	101,115	13,434	323
Domažlice	59,926	8,588	170
Klatovy	85,726	12,580	249
Plzeň město	188,045	28,199	422
Plzeň jih	62,389	9,891	194
Plzeň sever	74,940	12,554	179
Rokycany	47,458	7,678	166
Tachov	51,917	8,449	190
Cheb	90,188	13,714	536
Karlovy Vary	115,446	15,002	416
Sokolov	89,961	14,234	382
Děčín	128,834	17,717	329
Chomutov	122,157	13,438	320
Litoměřice	117,278	16,158	275
Louny	85,191	10,903	202
Most	111,775	12,982	300
Teplice	125,498	14,693	233
Ústí nad Labem	118,228	15,532	222
Česká Lípa	100,756	14,093	277
Jablonec nad Nisou	88,200	15,490	251
Liberec	169,878	29,266	364

Semily	73,605	11,968	156
Hradec Králove	162,661	29,631	356
Jičín	79,702	12,969	189
Náchod	109,550	20,685	412
Rychnov nad Kněžnou	77,829	13,769	170
Trutnov	118,174	24,662	461
Chrudim	103,199	17,359	235
Pardubice	168,423	28,584	362
Svitavy	103,245	14,081	246
Ústí nad Orlicí	136,760	21,223	297
Havlíčkův Brod	94,217	14,413	277
Jihlava	110,522	13,833	230
Pelhřimov	71,914	10,087	182
Třebíč	111,693	12,525	209
Žďár nad Sázavou	117,219	15,501	202
Blansko	105,708	12,495	317
Brno mesto	385,913	39,935	692
Brno venkov	206,300	24,718	372
Břeclav	112,828	12,292	297
Hodonín	153,225	18,755	392
Vyškov	88,154	11,313	276
Znojmo	111,380	12,433	300
Jeseník	38,779	4,185	115
Olomouc	230,408	29,768	539
Prostějov	107,859	14,102	201
Přerov	130,082	18,050	257
Šumperk	121,299	13,035	263
Kroměříž	105,569	14,008	264
Uherské Hradiště	141,467	18,577	261
Vsetín	142,420	18,259	308
Zlín	190,488	26,310	457
Bruntál	92,693	10,404	263
Frýdek Místek	207,756	27,305	525
Karviná	256,394	30,945	598
Nový Jičín	148,074	18,281	270
Opava	174,899	27,279	398
Ostrava mesto	326,018	37,820	735

Table 1 Data

The results from model 1 and model 2 are in table 2. The most affected regions have score 1. The virtual outputs represent how much will increase outputs to become the regions to same level as the most affected regions. For example, region Praha has score 1.66. Region Prague have 159,523 infected people and 2,220 death people with COVID-19. If we work with constant return of scale region Praha will be increase to 264,788 infected and to 4,950 death people with COVID-19.

	Eff. Score constant return of scale	virtual output1	virtual output2	Eff. Score variable return of scale	virtual output1	virtual output2
Praha	1.66	264,788	4,950	1.00	159,523	2,220
Praha západ	1.45	27,387	512	1.38	26,192	481
Příbram	1.41	23,544	440	1.40	23,375	439
Rakovník	1.44	11,477	215	1.19	9,488	197
Benešov	1.16	19,695	381	1.13	19,131	370
Beroun	1.43	17,981	336	1.35	16,973	327
Kladno	1.41	33,140	619	1.25	29,424	523
Kolín	1.18	20,035	375	1.14	19,337	369
Kutná hora	1.40	15,319	286	1.27	13,910	274
Mělník	1.37	21,842	408	1.35	21,416	405
Mladá Boleslav	1.32	25,807	482	1.29	25,305	469
Nymburk	1.28	19,802	370	1.23	19,068	364
Praha východ	1.32	32,795	613	1.17	29,230	521
České Budějovice	1.73	38,913	727	1.45	32,666	565
Český Krumlov	1.87	12,057	257	1.66	10,667	227
Jindřichův Hradec	1.39	18,554	366	1.34	17,885	353
Písek	1.40	14,576	272	1.25	13,054	259
Prachatice	1.62	9,328	235	1.33	7,635	192
Strakonice	1.52	13,985	293	1.39	12,827	269
Tábor	1.43	19,223	462	1.42	19,058	458
Domažlice	1.43	12,259	243	1.24	10,672	218
Klatovy	1.39	17,538	347	1.33	16,742	331
Plzeň město	1.39	39,244	734	1.16	32,851	568
Plzeň jih	1.29	12,805	251	1.14	11,264	228
Plzeň sever	1.25	15,639	292	1.14	14,278	281
Rokycany	1.22	9,401	203	1.00	7,678	166
Tachov	1.20	10,132	228	1.01	8,570	193
Cheb	1.00	13,714	536	1.00	13,714	536
Karlovy Vary	1.38	20,680	573	1.29	19,279	535
Sokolov	1.15	16,343	439	1.12	15,993	429
Děčín	1.52	26,887	503	1.46	25,797	479
Chomutov	1.73	23,313	555	1.62	21,777	519
Litoměřice	1.51	24,475	458	1.51	24,447	457
Louny	1.63	17,779	332	1.54	16,740	323
Most	1.66	21,586	499	1.64	21,277	492
Teplice	1.78	26,190	490	1.74	25,520	472
Ústí nad Labem	1.59	24,673	461	1.59	24,668	461
Česká Lípa	1.47	20,664	406	1.44	20,261	398
Jablonec nad Nisou	1.19	18,407	344	1.13	17,463	336
Liberec	1.21	35,452	663	1.05	30,722	540
Semily	1.28	15,361	287	1.17	13,958	275
Hradec Králove	1.15	33,946	635	1.01	29,876	529
Jičín	1.28	16,633	311	1.19	15,422	301
Náchod	1.08	22,362	445	1.07	22,209	442

Rychnov nad						
Kněžnou	1.18	16,242	304	1.09	14,972	293
Trutnov	1.00	24,662	461	1.00	24,662	461
Chrudim	1.24	21,537	403	1.21	21,065	399
Pardubice	1.23	35,149	657	1.07	30,552	538
Svitavy	1.53	21,546	403	1.50	21,076	399
Ústí nad Orlicí	1.34	28,541	534	1.26	26,840	489
Havlíčkův Brod	1.35	19,475	374	1.31	18,872	363
Jihlava	1.67	23,065	431	1.65	22,824	429
Pelhřimov	1.49	15,008	281	1.34	13,552	268
Třebíč	1.86	23,309	436	1.84	23,105	434
Žďár nad						
Sázavou	1.58	24,463	457	1.58	24,433	457
Blansko	1.57	19,662	499	1.57	19,600	497
Brno město	2.02	80,537	1,505	1.33	52,959	918
Brno venkov	1.74	43,053	805	1.42	34,991	596
Břeclav	1.74	21,408	517	1.69	20,799	503
Hodonín	1.64	30,737	642	1.41	26,378	551
Vyškov	1.47	16,661	406	1.43	16,162	394
Znojmo	1.70	21,145	510	1.66	20,694	499
Jeseník	1.67	6,974	192	1.00	4,185	115
Olomouc	1.62	48,084	899	1.22	36,258	657
Prostějov	1.60	22,509	421	1.57	22,185	418
Přerov	1.50	27,147	507	1.44	26,058	479
Šumperk	1.89	24,646	497	1.84	23,925	483
Kroměříž	1.57	21,970	414	1.54	21,635	408
Uherské Hradiště	1.59	29,523	552	1.47	27,392	497
Vsetín	1.63	29,722	556	1.51	27,504	498
Zlín	1.51	39,753	743	1.25	33,014	573
Bruntál	1.66	17,267	436	1.63	16,928	428
Frýdek Místek	1.57	42,938	826	1.21	32,903	633
Karviná	1.71	52,897	1,022	1.21	37,479	724
Nový Jičín	1.69	30,902	578	1.54	28,166	507
Opava	1.34	36,500	682	1.15	31,311	548
Ostrava město	1.77	67,128	1,305	1.16	44,023	856

Table 2 Results of modul 1 and model 2

The total infected people in our period are 1,433,809 and number of total death people are 24,331. The variations between regions are large. The worst results in model with constant return of scale have regions Trutnov and Cheb. The three regions with the best results in model with constant return of scales are regions Brno – město, Šumperk and Třebíč. The regions Trutnov and Cheb lies on the production frontier. If we moved all regions to product frontier in model 1, the total infected people would increases to number 2,139,809 and the total death people increases to 42,089.

The worst results with model variable return of scales have regions Praha, Rokycany, Cheb, Trutnov, Jeseník. The three regions with the best results in model with variation return of scales are regions Šumperk, Třebíč and Břeclav. If we moved all regions to the product frontier in model 2, the total number of infected people would increases to 1,833,845 and the total death people increases to 35,102.

The results from model 1 are figure on figure 1.

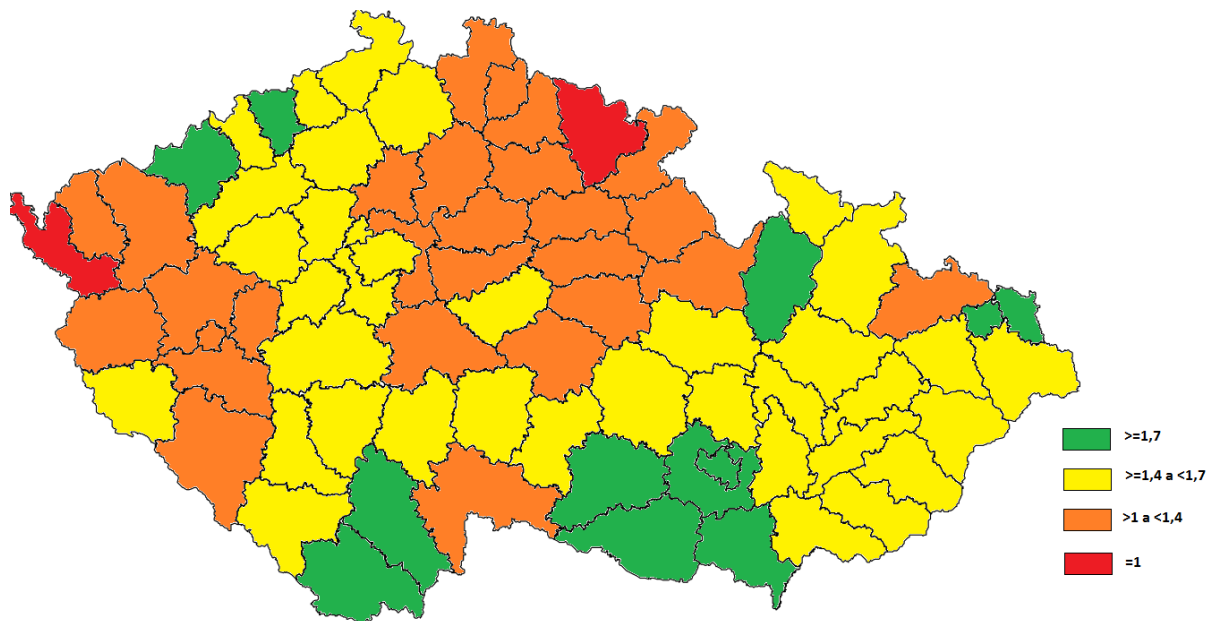


Figure 1 Map with results of model 1.

The most affected regions with COVID-19 are on the map of the Czech Republic have red color. The regions with the smallest incidence with COVID-19 have green color.

4 Conclusions

This paper focuses on measure of regional variation in spread of disease COVID-19 between 77 regions in the Czech Republic. The variations were measured by product frontier model based on data envelopment analysis. Models in this paper work with one input and two outputs. The input is population in the q region. The first output is number of infected people with COVID-19 in the q region and the second output is number of people death with COVID-19 in the q region. The regions which lie on the product frontier were the most affected regions with COVID-19. Variations were measured with two outputs oriented models. The first model works with constant return of scales and the second model works with variation return of scales. The second model show lower differences between regions and lower total infected and death people with COVID-19. If we moved all regions to product frontier calculated with model 1, the total infected people would increase from 1,433,809 to 2,139,809 and the total death people would increase from 24,331 to 42,089. If we moved all regions to product frontier calculated with model 2, The total infected people would increase from 1,433,809 to 1,833,845 and the total death people would increase from 24,331 to 35,102. The regional variations are large. The local prohibition in this situation could be more effective than national prohibition.

Acknowledgements

This work was supported by the project no. F4/42/2021 of the Internal Grant Agency, Faculty of Informatics and Statistics, University of Economics, Prague.

References

- [1] Dlouhý, M. and Hanousek, J. (2019). An assesment of Regional Variations: AnApplication to Polish Regions. In: 37 th International Conference on Mathematical Methods in Economics, České Budejovice: University of South Bohemia in ČeskéBudějovice, Faculty of Economics, pp. 281-286
- [2] Dlouhý, M. (2018). Measuring Geographic Inequalities: Dealing with multiple Health Resources by Data Envelopment Analysis. *Frontiers in Public health*, 6(53), pp. 1-6.
- [3] Charnes, A.,Cooper, W. W.&Rhodes, E. (1978). Measuring the Inefficiency of Decision Making Units. *European Journal of Operational Research*, 2, 429–444. doi: 10.1016/0377-2217(78)90138-8
- [4] Jablonský, J., Dlouhý, M.(2004): *Modely hodnocení efektivnosti produkčních jednotek*. Praha: Professional Publishing, pp. 183.
- [5] Kumbhakar, S. C., Lovell, C. A. K(2000). *Stochastic frontier analysis*. Cambridge: Cambridge University Press, pp. 355.
- [6] Český statistický úřad, [online] Available at: <https://www.czso.cz/csu/czso/pocet-obyvatel-v-obcich-k-112019> [Accessed 28.3.2021]
- [7] Ministerstvo zdravotnictví České republiky [online] Available at: <https://onemocneni-aktualne.mzcr.cz/covid-19/kraje> [Accessed 31.3.2021]

Determinants of company indebtedness in the construction industry

Jana Heckenbergerová¹, Irena Honková², Alena Kladivová³

Abstract.

The aim of this paper is to reveal the determinants of indebtedness in the construction industry companies. The construction industry is a specific sector where payment morale is generally poor. It gradually negatively affects other companies in the following sectors. Finding the essential determinants of corporate indebtedness can prevent liquidity problems.

Based on a literature review, the following determinants were selected for analyses: share of fixed assets, interest rate, return on assets, size of the company and its age. Correlation analysis and multiple linear regression analysis have been chosen to determine the influence of the determinants within years 2016-2019.

It was found that the generally recommended fixed asset share determinant was not an appropriate determinant and its possible effect on indebtedness was also proven to be insignificant. Surprisingly interest rates have also classified as insignificant. Significant determinants negatively affecting indebtedness for construction companies were determined as enterprise size and duration. The most important determinant was the return on assets with negative influencing outcome.

Keywords: indebtedness, capital structure, return on asset, construction industry

JEL Classification: M1

AMS Classification: 62, 91

1 Introduction

Construction industry is one of the key sectors of the economy. The share of the construction industry in the gross value added of the whole economy has been between 5% and 7% [11]. Therefore, it is considered as one of the important indicators of the development in the economy.

This industrial sector was deeply affected by the last economic crisis in 2009 and 2010, as evidenced by the proportion of failed loans of up to 28 %, the highest of all branches of industry [6]. Construction sales accelerated significantly year-on-year growth in 2018, but still did not reach the level of 2008. The return on equity (ROE) was 16,47% in 2018 and it is still less than 22,57% from the pre-crisis period of 2008 [11]. Consequences of this crisis are linked with indebtedness and liquidity in this sector. This is gradually negatively affecting other enterprises in following branches.

Identifying and analyzing factors affecting the indebtedness of construction companies could help with prediction of upcoming liquidity problems. Searching of mutual relations can confirm or deny the significance of analyzed determinants. Knowledge of significant factors affecting the indebtedness of companies can help creditors to evaluate the company rating. This eliminates further problems with the repayment of liabilities and secondary insolvency, and therefore it contributes to a healthy business environment.

¹ University of Pardubice/Faculty of Economics and Administration, Institute of Mathematics and Quantitative Methods, Studentská 95, Pardubice, jana.heckenbergerova@upce.cz

² University of Pardubice/Faculty of Economics and Administration, Institute of Business Economics and Management, Studentská 95, Pardubice, irena.honkova@upce.cz

³ University of Pardubice/Faculty of Economics and Administration, Studentská 95, Pardubice, alena.kladivova@student.upce.cz

2 Literature review and Problem statement

As the essential determinant of the capital structure it is usually mentioned the tax costs and the tax shield. Other factors are based on sector standards and various costs, for example the weight average cost of capital (WACC) and costs of financial distress. Another significant determinants are including, according to Křivská [9], profitability and stability of the company, the asset structure of the enterprise, the business sector, the management of the enterprise and its approach to risk, the structure of ownership and control over the enterprise, financial freedom, the amount of investment, the size of the enterprise, the goodwill and history of the enterprise, the requirements of the credit rating agencies.

Marks [10] deals with the factors of the capital structure in their publication as well. They consider that the approach of shareholders or owners, their requirements for the dividend payout ratio, their relationship to credit and risk, corporate philosophy and the sector, the business life phase, have a major influence on the capital structure. Růčková [14] argues that the capital structure is mainly influenced by the focus of the company's business. She summarizes other factors in four areas: business risk, corporate tax position, financial flexibility, and managerial conservatism and aggressiveness. Singh [15] and Chen & Chen [4] in their researches confirm the importance of the profitability, size and volatility of the enterprise. Oztekin [12] observes a context between indebtedness and company size, tangible assets, and profitability. He states that the capital structure reflects the institutional environment in which it operates.

Aulová and Hlavsa [2] focused on specific sector of Czech farms in their work. The size and asset collateral were identified as the most important determinants. Long-term indebtedness was most affected by size, asset collateral, tax shield and retained earnings. On the contrary, Viviani, J. [16] found out that there is no statistically significant dependence between indebtedness and the size of the enterprise, the structure of assets, the profitability of assets and the tax shield. Prášilová [13] found out in her research that the age of the company has a positive effect on the total indebtedness of Czech companies, and she observed a negative relationship with the profitability of assets. Only the share of fixed assets affected long-term indebtedness. In the ICT sector, it was found a negative relationship of total debt to the size of the enterprise and a positive relationship with the volume of retained earnings. Křivská [9] considers that larger enterprises generally show higher profits and that a higher level of liquid assets is less risky for investors. However, external influences, such as the level of the capital market, legislative processes, the economic policy itself and the mentioned above economic cycle or tax shield [7,8], affect total indebtedness as well.

In previously mentioned studies, the significant relationship to the capital structure was proven only for some determinants. Obviously, a few of described characteristics overlap and complement each other. Since the most of analyses were sector-specific, it is problematic to generalize the results as each sector has its own specificities. In our study, the main goal is to determine the direct effects on construction industry indebtedness.

Therefore based on above discussion, we selected the most common internal determinants: fixed asset share (SFA), interest rate (IR), return on assets (ROA), enterprise size (S) and duration (D). We neglected a lot of other determinants by the cause of one sector investigation only. External influences were excluded as well due to their general effect.

3 Source data and Methods

The source dataset, available in the public register [1], consist unconsolidated financial statements of fifty companies in the construction industry in the Czech Republic within years 2016-2019. Crucial fact for companies' selection is that they have not been liquidated till 31st August, 2020. Other entrance criterions as the legal form, size and duration, were not implemented.

Our research starts with a technical financial analysis, which evaluates main sample characteristics of the total indebtedness. The correlation analysis and multiple linear regression analysis are utilized to determine the effects and significance of individual determinants.

Powerful software tool Statistics 12 was helpful for our analyses, where the following abbreviations are used: Total Indebtedness (TI), Share of Fixed Assets (SFA), Interest Rate (IR), Return on Assets (ROA), Duration (D) and Size (S). In all presented results, normality is assumed and the significance level is pre-set to $\alpha = 0.05$.

4 Results

Although financial data from the construction industry have not yet reached the situation before the 2008 financial crisis [11], it is obvious that total indebtedness is already reaching recommended level. Sample distribution of total indebtedness, illustrated in Figures 1 and 2, is left skewed. This is confirmed by the average debt lower than the median as shown in Table 1. Moreover it is quite heavy tailed cause of a few companies with quadruple debt compared with average value. The average total indebtedness in the construction industry (51%) overreaches the recommended 40% level of total indebtedness. Nevertheless the median value of total indebtedness (41%) already corresponds to this recommended standard. As in any sector, there are companies with almost zero indebtedness and conversely over-indebted companies with negative equity.

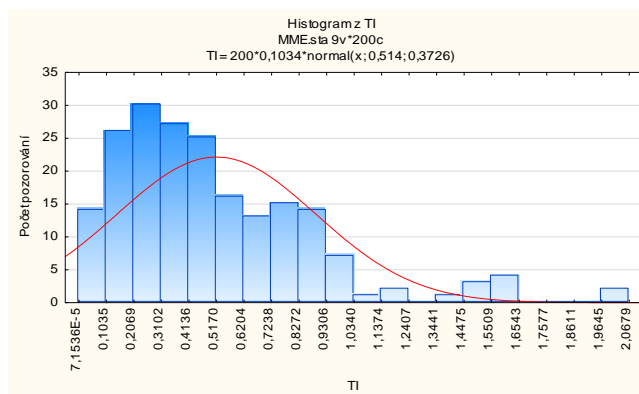


Figure 1 Histogram for the Total Indebtedness

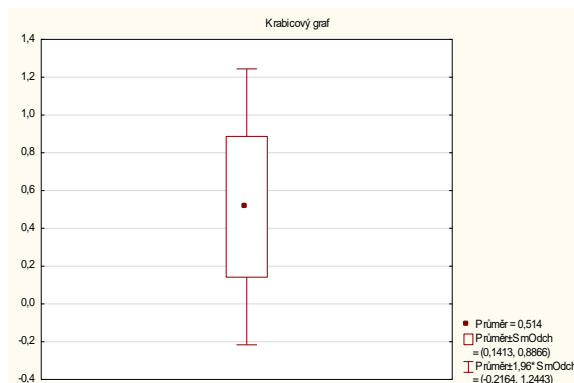


Figure 2 Boxplot of the Total Indebtedness

	N	Average	Median	Minimum	Maximum	Std	Var.coef.	Skew	Kurtosis
TI	200	0,514	0,418	0	2,068	0,373	72,504	1,484	2,886

Table 1 Sample characteristics of the Total Indebtedness

At the beginning of the analyses, correlation matrix within the individual determinants was evaluated to see whether the explaining variables for the total indebtedness were appropriate. Obviously from Table 2, where significant correlations are marked as red, the appropriate determinants are the interest rate (IR), return on assets (ROA), the duration (D), and the size (S). These determinants do not correlate with each other significantly. However, the share of fixed assets (SFA) is not very suitable as a determinant of indebtedness, as it points to a significant correlation with other determinants return on assets, duration and size.

	TI	SFA	IR	ROA	D	S
TI	1,000	-0,089	0,077	-0,187	-0,253	-0,225
SFA	-0,089	1,000	0,110	-0,231	0,434	0,481
IR	0,077	0,110	1,000	-0,117	0,036	0,059
ROA	-0,187	-0,231	-0,117	1,000	-0,089	0,067
D	-0,253	0,434	0,036	-0,089	1,000	0,423
S	-0,225	0,481	0,059	0,067	0,423	1,000

Table 2 Correlation matrix between the Total Indebtedness and selected determinants

Table 2 also shows correlations between total indebtedness (TI) and the influencing variables: fixed asset share (SFA), interest rate (IR), return on assets (ROA), duration (D) and enterprise size (S). There are significant correlations between total indebtedness (TI) and return on asset assets (ROA), duration (D) and size (S). All these determinants negatively affect total indebtedness. The effect of the fixed asset share (SFA) is insignificant and rather negative, the impact of interest rate (RI) is insignificantly positive.

It follows from the above that companies with higher return on assets have lower total indebtedness. It is also true that the older and larger the company the less indebted it is. However, behavior of the interest rate determinant (IR) is interesting as only an insignificantly positive correlation between this determinant and total indebtedness has been shown.

To reveal direct correlation between TI and its determinants without added effects, partial correlations are evaluated and summarized in Table 3. Significant partial correlations are marked as red and they are showing similar results. They confirm the significant negative impact of return on assets (ROA) and duration (D), and the medium and statistically insignificant impact of the interest rate (IR) and size (S). It is verified again that the share of fixed assets (SFA) has almost no effect on the total indebtedness in the construction industry.

Explaining variable	Dependent variable IT
	Partial correlation
SFA	0,019
IR	0,072
ROA	-0,184
D	-0,205
S	-0,115

Table 3 Partial correlations of the Total Indebtedness and selected determinants

The results of the multiple linear regression analysis, summarized in the Table 3, correspond to the previous conclusion as well. The return on assets (ROA) and duration (D) have the significant negative impact, while the share of fixed assets (SFA) and interest rate (IR) do not significantly affect total indebtedness (TI). Nevertheless, the value of the determination index $R^2=0,12463333$ is showing that regression model is unsuitable for further predictions. **It seems that some significant explaining determinant of indebtedness is missing. Uncovering this mystery will be goal of our upcoming research.**

	b	Std of b	t(194)	p-value
b0	0,839	0,077	10,825	0,000
SFA	0,031	0,116	0,267	0,790
IR	1,874	1,852	1,012	0,313
ROA	-0,436	0,168	-2,603	0,010
D	-0,011	0,004	-2,920	0,004
S	-0,001	0,000	-1,610	0,109

Table 4 Regression analysis coefficients and their significance (black–significant and red–nonsignificant)

5 Discussion and Conclusions

The following describing variables have been chosen for the construction industry in the Czech Republic: fixed asset share (SFA), interest rate (IR), return on assets (ROA), size of enterprise (S) and duration (D). Empirical research results have confirmed that the variable of fixed assets share (SFA) is not suitable as a determinant of total indebtedness (TI), as it is influenced by other determinants: return on assets (ROA), duration (D) and size (S). This determinant has been also classified as insignificant in the research.

Furthermore, another insignificant determinant interest rate (IR) has been identified. The finding that the interest rate is only increasing marginally with increasing indebtedness has been novel and surprising. The most important theories about the capital structure [3] claim that as indebtedness increases, the so-called costs of financial distress begin to infiltrate companies, when creditors (most often banks) demand a higher interest rate for higher risk.

The determinant of the size of the enterprise (S) has been classified as medium-significant with a negative effect on total indebtedness (TI). Correlation analysis identified this determinant as significant, while partial correlation and regression analysis showed medium significance. This is caused by significant correlation between duration (D) and size (S) itself.

Return on assets (ROA) and duration (D) have been determined as the most important determinants of the total indebtedness of construction companies. Both have had a significant negative effect on the indebtedness regardless to analyzing method.

The fact that the longer a company operates on the market, the less indebted it is, is not surprising. Long-term businesses are mostly capital stronger than the newly established companies and therefore, they already have enough capital to cover their assets. For a similar reason, the size of the enterprise (S) is the indebtedness determinant as well. A large enterprise has sufficient equity capital and does not necessarily need debt to finance its activities. These two determinants, duration (D) and size of the enterprise (S), are highly correlated with each other and it is not recommended to use them both in one regression analysis.

It has been confirmed that highly profitable companies have less total indebtedness. This fact is interesting in view of the effect of the tax shield. Economic theories generally recommend the involvement of debt for higher-profit enterprises. The interest on debt is a tax-efficient expenditure and it can reduce the tax base of profitable enterprises. It has been proven that if a company can borrow at an interest rate below the return on assets (ROA) the involvement of this debt increases the return on equity (ROE), i.e. leverage has a positive effect. To sum up, large, long-term highly profitable companies for construction industry do not adopt external capital even if they could benefit from a tax shield.

Acknowledgements

This contribution was supported by the Student Grant Competition No. SGS_2021_012 of University of Pardubice in 2020.

References

- [1] ARES (2020). *Ministerstvo financí ČR*. Available from: http://www.info.mfcr.cz/ares/ares_es.html.cz.
- [2] Aulová, R. & Hlavsa, T. (2013). Capital Structure of Agricultural Businesses and its Determinants. *Agris on-line Papers in Economics and Informatics*, 5, 23-36.
- [3] Brealey, R. & Myers, S. (2014). *Teorie a praxe firemních financí*. Brno: BizBooks.
- [4] Chen, S. & Chen L. (2011). Capital structure determinants: An empirical study in Taiwan. *African Journal of Business Management*, 5, 10974-10983.
- [5] CZ-NACE (2020). *Český statistický úřad*. Available from: <http://nace.cz>.
- [6] Honková, I. (2016). Use of External Sources of Financing in the Construction Industry. *Scientific papers of the University of Pardubice*, 36, 42-54.
- [7] Hrdý, M. (2011). Does the Debt Policy Theoretically and Practically Matter in Concrete Firm? *Český finanční a účetní časopis*, 1, 19-32.
- [8] Kislingerová, E. (2013). *Sedm smrtelných hříchů podniků: úpadek a etika managementu*. Praha: C.H. Beck.
- [9] Křivská, R. (2009). Determinants of capital structure and its optimization. *Dizertační práce*, 54-57.

- [10] Marks, K. (2009). *The handbook of financing growth: strategies, capital structure, and M&A transactions*. Hoboken: John Wiley.
- [11] Ministerstvo průmyslu a obchodu. *Stavebnictví 2019*. Available from: mpo.cz.
- [12] Oztekin, O. (2015). Capital Structure Decisions around the World: Which Factors Are Reliably Important? *Journal of Financial and Quantitative Analysis*, 50, 301-323.
- [13] Prášilová, P. (2012). Determinanty kapitálové struktury českých podniků. *Ekonomie a Management*, 1, 89-104.
- [14] Růčková, P. (2015). *Finanční analýza: metody, ukazatele, využití v praxi*. Praha: Grada Publishing.
- [15] Singh, D. (2016). A Panel Data Analysis of Capital Structure Determinants: An Empirical Study of Non-financial Firms in Oman. *International Journal of Economics and Financial Issues*, 6, 1615-1656.
- [16] Viviani, J. (2008). Capital structure determinants: An empirical study of French companies in the wine industry. *International Journal of Wine Business Research*, 20, 171-194.

Robust Slater's Condition in an Uncertain Environment

Milan Hladík¹

Abstract. Slater's condition is, no doubt, an important regularity condition used in nonlinear programming. It states that the feasible set must contain an interior point. We analyse this condition in an uncertain environment. We assume that uncertainty of the input data has the form of intervals covering the true values; we assume no other information about the uncertainty is known. Then Slater's condition holds robustly if it is satisfied for each possible realization of the interval values.

In particular, we investigate interval systems of linear equations and inequalities. Therein, Slater's condition has the form of strong solvability with strict inequalities. We present a finite characterization of this property and inspect its computational complexity – in some cases it is polynomial, but in some cases it is NP-hard. As an illustration, we apply our results in interval linear programming in the problem of testing boundedness of the optimal solution set.

Keywords: linear programming, interval analysis, interval system, robustness, NP-hardness

JEL Classification: C44, C61

AMS Classification: 90C05, 65G40, 15A39

1 Introduction

Slater's condition is a constraint qualification appearing in optimality conditions in convex optimization, among many other situations. Roughly speaking, Slater's condition requires an existence of an interior feasible point.

In this paper, we are concerned with Slater's condition in case the input data are uncertain. This is a common situation in many practical problems, including optimization problems. In particular, we deal with the problems, where uncertainty is represented by intervals. That is, the only information we have are upper and lower bounds on the true values. No other information (such as probability distribution or fuzzy shape) is known.

Interval data. Interval data are represented by interval vectors and matrices; we denote them by boldface. An interval matrix is by definition the set of matrices

$$\mathbf{A} = \{A \in \mathbb{R}^{m \times n}; \underline{A} \leq A \leq \overline{A}\},$$

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 corresponding terms are the midpoint matrix A_c and the radius matrix A_Δ of A defined respectively as

$$A_c = \frac{1}{2}(\underline{A} + \overline{A}), \quad A_\Delta = \frac{1}{2}(\overline{A} - \underline{A}).$$

For more on interval analysis, see, e.g., the books [3, 7, 11].

Given an interval system, it is called *weakly solvable* if it is solvable for at least one realization of interval coefficients, and it is called *strongly solvable* if it is solvable for every realization of interval data. For instance, an interval system of linear inequalities $Ax \leq b$ is weakly (strongly) solvable if $Ax \leq b$ is solvable for some (for every) $A \in \mathbf{A}$ and $b \in \mathbf{b}$.

Next, a *strong solution* to an interval system is a point that solves every realization of the system. Clearly, if an interval system possesses a strong solution, then it is strongly solvable. The converse implication does not hold in general.

The goal. We investigate Slater's condition of interval linear systems. Even though this constraint qualification is more used in nonlinear programming, we begin our investigation with the more simple case of linear constraints.

¹ Charles University, Department of Applied Mathematics, Malostranské nám. 25, 118 00, Prague & University of Economics, Department of Econometrics, nám. W. Churchilla 4, 13067, Prague, Czech Republic, hladik@kam.mff.cuni.cz

We say that *Slater's condition holds robustly* if it holds for every realization of interval data. This in turn leads to strong solvability of interval systems with strict inequalities. In particular, we focus on strong solvability of interval system $Ax < \mathbf{b}$ and interval system $Ax = \mathbf{b}, x > 0$. We present necessary and sufficient conditions for strong solvability and for existence of strong solutions, and we analyze the computational complexity of the problem in question, too.

Notice that systems of strict inequalities occur also in other situations than in Slater's condition. For more on this issue in the real case see, e.g., [2, 13].

Notation. For vectors $a, b \in \mathbb{R}^n$ we use $a \preceq b$ to denote $a \leq b, a \neq b$. We use $e = (1, \dots, 1)^T$ for the vector of ones (with convenient dimension) and $\text{diag}(s)$ for the diagonal matrix with entries s_1, \dots, s_n . The absolute value and the inequalities are understood entrywise.

2 Characterization and computational complexity

In this section we show that the interval Slater's condition is easy to verify for interval inequalities, but can be computationally hard in general for equality constrained problems. The following characterization is a modification of the result by Rohn [14] on strong solvability of a system without strict inequalities.

Theorem 1. *An interval system $Ax = \mathbf{b}, x > 0$ is strongly solvable if and only if the system*

$$(A_c + \text{diag}(s)A_\Delta)x = b_c - \text{diag}(s)b_\Delta, \quad x > 0 \quad (1)$$

is solvable for each $s \in \{\pm 1\}^m$.

Proof. The interval system $Ax = \mathbf{b}, x > 0$ is strongly solvable if and only if for each $A \in \mathbf{A}$ and $b \in \mathbf{b}$ the system $Ax = b, x > 0$ has a solution. The system $Ax = b, x > 0$ is equivalent (w.r.t. solvability) to system $Ax - by = 0, x \geq e, y \geq 1$. By Farkas' lemma, it is feasible if and only if the system

$$A^T u - v \geq 0, \quad -b^T u - w \geq 0, \quad -e^T v - w < 0, \quad v, w \geq 0$$

is infeasible for each $A \in \mathbf{A}$ and $b \in \mathbf{b}$. It equivalently reads $(A \mid -b)^T u \not\geq 0$. Thus $Ax = \mathbf{b}, x > 0$ is strongly solvable if and only if the system $(A \mid -b)^T u \not\geq 0$ is not weakly solvable. By [9], weak solvability of $(A \mid -b)^T u \not\geq 0$ is equivalent to solvability of

$$\begin{pmatrix} A_c^T + A_\Delta^T \text{diag}(s) \\ b_c^T - b_\Delta^T \text{diag}(s) \end{pmatrix} u \not\geq 0$$

for some $s \in \{\pm 1\}^m$. By Farkas' lemma again, we obtain the statement. \square

The exponential number in the characterization is not easy to avoid since the problem is intractable. To show it, we first present an auxiliary result, which is worth of stating it explicitly.

Proposition 2. *Checking weak solvability of $Ax \not\leq 0$ is an NP-hard problem even with interval entries in one row of A only.*

Proof. By [3], checking solvability of

$$|Ax| \leq e, \quad e^T |x| \geq 1 \quad (2)$$

is NP-hard. We claim that it is equivalent to checking solvability of the system

$$|Ax| \leq ey, \quad y \geq 0, \quad e^T |x| \geq y \quad (3)$$

with at least one inequality satisfied strictly. Obviously, if (2) has solution x^* , then x^* and $y^* := 1$ solve (3) as required. Conversely, let x^*, y^* be a solution to (3). If $y^* > 0$, then $\frac{1}{y^*}x^*$ solves (2). If $y^* = 0$, then $e^T |x^*| > y^* = 0$, and so we can put $y^* := e^T |x^*|$ and reduce this case to the previous one.

Now, by the Gerlach characterization of interval inequalities [3, 6], system (3) describes the solution set of the interval system

$$Ax - ey \leq 0, \quad -Ax - ey \leq 0, \quad -y \leq 0, \quad [-e, e]^T x + y \leq 0,$$

which has the desired form. \square

Now, we show that it is intractable to check if there is a Slater point in each realization of an interval system in the form $Ax = \mathbf{b}$, $x > 0$. That is, even when the intervals are situated in the right-hand side only, the problem is hard.

Theorem 3. *Checking strong solvability of an interval system $Ax = \mathbf{b}$, $x > 0$ is co-NP-hard.*

Proof. Similarly as in the proof of Theorem 1, interval system $Ax = \mathbf{b}$, $x > 0$ is strongly solvable if and only if the system $(A \mid -\mathbf{b})^T u \geq 0$ is not weakly solvable. However, checking weak solvability of this interval system is NP-hard by Proposition 2. \square

In contrast to strong solvability, deciding on existence of a strong solution is a simple problem. The fundamental drawback is that a strong solution for interval equations exists in rare situations. Indeed, as the following observation shows, it exists only if there are no interval coefficients, just real values!

Corollary 1. A vector x is a strong solution to an interval system $Ax = \mathbf{b}$, $x > 0$, if and only if

$$A_c x = b_c, \quad x > 0, \quad A_\Delta = 0, \quad b_\Delta = 0.$$

Proof. By [3, Thm. 2.16], a vector x is a strong solution to an interval system $Ax = \mathbf{b}$, $x \geq 0$, if and only if it satisfies

$$A_c x = b_c, \quad A_\Delta |x| = b_\Delta = 0.$$

Since $x > 0$, the condition $A_\Delta |x| = 0$ reads $A_\Delta x = 0$, which holds if and only if $A_\Delta = 0$. \square

For an interval system of linear inequalities, Slater's condition is characterized by an adaptation of the results of Rohn and Kreslová [15].

Theorem 4. *The interval system $Ax < \mathbf{b}$ is strongly solvable if and only if the system*

$$\overline{A}x^1 - \underline{A}x^2 < \underline{b}, \quad x^1 \geq 0, \quad x^2 \geq 0 \tag{4}$$

is solvable in variables x^1, x^2 .

Proof. The interval system $Ax < \mathbf{b}$ is not strongly solvable if and only if there are $A \in \mathbf{A}$ and $b \in \mathbf{b}$ such that $Ax < b$ is unsolvable. By Farkas' lemma, equivalently, the system

$$A^T u = 0, \quad b^T u \leq 0, \quad u \geq 0$$

is solvable. Thus we have that the interval system

$$A^T u = 0, \quad b^T u \leq 0, \quad u \geq 0$$

is weakly solvable. By the generalization of the Oettli–Prager and Gerlach theorems [9], the solution set is described by

$$\overline{A}^T u \geq 0, \quad -\underline{A}^T u \geq 0, \quad \underline{b}^T u \leq 0, \quad u \geq 0.$$

By Farkas' lemma again, the system (4) is unsolvable. \square

Theorem 5. *Suppose that the interval system $Ax < \mathbf{b}$ is strongly solvable, and define $x^* := x^1 - x^2$, where x^1, x^2 solves (4). Then x^* is a solution to $Ax < \mathbf{b}$ for every $A \in \mathbf{A}$ and $b \in \mathbf{b}$.*

Proof. Let $A \in \mathbf{A}$ and $b \in \mathbf{b}$ be arbitrary. Then

$$Ax^* = A(x^1 - x^2) = Ax^1 - Ax^2 \leq \overline{A}x^1 - \underline{A}x^2 < \underline{b} \leq b. \tag{4}$$

Corollary 2. An interval system $Ax < \mathbf{b}$ is strongly solvable if and only if it has a strong solution.

Adapting the results from [3], we can also state several equivalent characterizations of robust Slater's points.

Corollary 3. For a vector $x \in \mathbb{R}^n$, the following conditions are equivalent:

1. x is a strong solution to $Ax < \mathbf{b}$,
2. $A_c x + A_\Delta |x| < \underline{b}$,
3. $x = x^1 - x^2$, $\overline{A}x^1 - \underline{A}x^2 < \underline{b}$, $x^1, x^2 \geq 0$.

Proof. We already have the equivalence $1 \Leftrightarrow 3$ by Theorems 4 and 5. Equivalence $1 \Leftrightarrow 2$ follows from the fact that x is a strong solution to $Ax < \mathbf{b}$ if and only if $\max_{A \in \mathbf{A}} Ax < \underline{b}$. Notice that the value $A_c x + A_\Delta |x|$ is the entrywise maximum of Ax subject to $A \in \mathbf{A}$. \square

3 Consequences on boundedness of realizations of interval LP problems

Besides KKT optimality conditions, strict feasibility is important in many other issues as well. Herein, we show some consequences in checking boundedness of optimal solution set of an interval linear programming (LP) problem. By an *interval LP problem* we mean a family of LP problems

$$f(A, b, c) = \min c^T x \text{ subject to } x \in M(A, b), \quad (5)$$

where $M(A, b)$ is the feasible set, $A \in \mathbf{A}$, $b \in \mathbf{b}$, $c \in \mathbf{c}$, and $\mathbf{A}, \mathbf{b}, \mathbf{c}$ are given interval matrix and vectors. By $\mathcal{S}(A, b, c)$ we denote the optimal solution set corresponding to the particular realization $(A, b, c) \in (\mathbf{A}, \mathbf{b}, \mathbf{c})$. The set of all possible optimal solutions is then

$$\mathcal{S} = \bigcup_{(A, b, c) \in (\mathbf{A}, \mathbf{b}, \mathbf{c})} \mathcal{S}(A, b, c).$$

We usually write (5) shortly as

$$\min c^T x \text{ subject to } x \in M(\mathbf{A}, \mathbf{b}),$$

and we distinguish three canonical forms

$$\min c^T x \text{ subject to } Ax = \mathbf{b}, x \geq 0, \quad (\text{A})$$

$$\min c^T x \text{ subject to } Ax \leq \mathbf{b}, \quad (\text{B})$$

$$\min c^T x \text{ subject to } Ax \leq \mathbf{b}, x \geq 0. \quad (\text{C})$$

In the real case, one can consider any canonical form with no harm on generality because they can be equivalently transformed to each other. However, in the interval case, this is no more true [5]. That is why we have to consider the forms separately. Indeed, we will see later on that the computational complexity differs.

Interval linear programming was surveyed in [3, 8]. The optimal solution set in particular was addressed in [1, 4, 10, 12].

We say that an interval LP problem is *realization bounded* if $\mathcal{S}(A, b, c)$ is bounded for every realization $(A, b, c) \in (\mathbf{A}, \mathbf{b}, \mathbf{c})$. Notice that we consider an empty set as a bounded set. Obviously, if \mathcal{S} is bounded, then $\mathcal{S}(A, b, c)$ is bounded for every realization. The converse is not true in general, as the following example shows. It remains an open question, however, whether the converse implication is valid provided both primal and dual problems are strongly feasible.

Example 1. Consider the interval LP problem

$$\min x \text{ subject to } [0, 1]x = 1, x \geq 0.$$

Each realization taking a positive value $a \in [0, 1]$ has a unique optimal solution $x = 1/a$. Taking the value $a := 0 \in [0, 1]$ results in an infeasible LP problem. Thus in total we have $\mathcal{S} = [1, \infty)$, which is unbounded despite the fact that the problem is realization bounded.

Recall the characterization of bounded optimal solution set of a real-valued LP problem from [16].

Theorem 6. *Suppose that both primal and dual problems are feasible. The optimal solution set is bounded if and only if the dual problem contains a feasible solution satisfying the inequalities strictly.*

Infeasibility of the primal problem produces no optimal solution, so the boundedness is preserved. Hence we obtain a sufficient condition for realization boundedness of interval LP problems.

Corollary 4. An interval LP problem is realization bounded if for every realization the dual problem contains a feasible solution satisfying the inequalities strictly.

The assumption can be checked by the methods presented in Section 2, where strong feasibility of various interval systems was discussed. In the following, we show consequences for the particular forms of an interval LP problem.

Type (A). For this class of interval LP problems, the condition from Corollary 4 is easy to check.

Corollary 5. An interval LP problem of type (A) is realization bounded if the system

$$\overline{A}^T y^1 - \underline{A}^T y^2 < \underline{c}, y^1, y^2 \geq 0 \quad (6)$$

is feasible.

Example 2. Consider the interval LP problem from Example 1

$$\min x \text{ subject to } [0, 1]x = 1, x \geq 0.$$

We already observed that it is realization bounded. Indeed Corollary 5 confirms this because system (6), which reads

$$1y^1 - 0y^2 < 1, y^1, y^2 \geq 0,$$

is feasible.

On the other hand, consider a variation of the above problem (see [3])

$$\min -x \text{ subject to } [0, 1]x = 1, x \geq 0;$$

This problem is also realization bounded. Nevertheless, we cannot verify it by Corollary 5 because system (6), which reads

$$1y^1 - 0y^2 < -1, y^1, y^2 \geq 0,$$

is infeasible.

Type (B). Herein, the condition from Corollary 4 can be hard to check since strong solvability of $A^T y = b, y < 0$ is intractable; see Theorem 3.

Type (C). Similarly as for type (A), realization boundedness is polynomially decidable. The proof of the following statement is a simple adaptation of that for Corollary 5.

Corollary 6. An interval LP problem of type (C) is realization bounded if the linear system $\underline{A}^T y < \underline{c}, y < 0$ is feasible.

4 Conclusion

Robust Slater's condition in the context of interval linear systems basically means strict feasibility of every realization of the interval system. For an interval system of linear inequalities, the robust Slater's condition has a favourable characterization by means of linear inequalities, which makes the condition easy to check. Moreover, if the condition holds true, then there is a point which is the Slater's point for each realization of interval data.

In contrast, for an interval system of equations with nonnegative variables, the problem of checking robust Slater's condition is intractable. We presented a finite characterization by a reduction to 2^m linear systems, where m is the number of equations. Here, one could be interested in a computationally cheap sufficient condition.

Another research direction can be an extension of the presented results to a general interval system of mixed equations and inequalities. Let us also remind the open problem under which conditions the realization boundedness implies boundedness of the optimal solution set \mathcal{S} .

Acknowledgements

The author was supported by the Czech Science Foundation under project 19-02773S.

References

- [1] Allahdadi, M., and Nehi, H. M.: The optimal solution set of the interval linear programming problems. *Optim. Lett.* **7** (2013), 1893–1911.
- [2] Fajardo, M. D., Goberna, M. A., Rodríguez, M. M. L., and Vicente-Pérez, J.: *Even Convexity and Optimization. Handling Strict Inequalities*. EUROATOR. Springer, Cham, 2020.
- [3] Fiedler, M., Nedoma, J., Ramík, J., Rohn, J., and Zimmermann, K.: *Linear Optimization Problems with Inexact Data*. Springer, New York, 2006.
- [4] Garajová, E., and Hladík, M.: On the optimal solution set in interval linear programming. *Comput. Optim. Appl.* **72** (2019), 269–292.
- [5] 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.
- [6] Gerlach, W.: Zur Lösung linearer Ungleichungssysteme bei Störung der rechten Seite und der Koeffizientenmatrix. *Math. Operationsforsch. Stat., Ser. Optimization* **12** (1981), 41–43. In German.
- [7] Hansen, E. R., and Walster, G. W.: *Global Optimization Using Interval Analysis*. 2nd edition. Marcel Dekker, New York, 2004.
- [8] 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.
- [9] Hladík, M.: Weak and strong solvability of interval linear systems of equations and inequalities. *Linear Algebra Appl.* **438** (2013), 4156–4165.
- [10] Hladík, M.: Two approaches to inner estimations of the optimal solution set in interval linear programming. In: *Proceedings of the 2020 4th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence* (Deb, S., ed.), ISMSI 2020. Association for Computing Machinery, New York, USA, 99–104.
- [11] Moore, R. E., Kearfott, R. B., and Cloud, M. J.: *Introduction to Interval Analysis*. SIAM, Philadelphia, PA, 2009.
- [12] Rada, M., Garajová, E., Horáček, J., and Hladík, M.: A new pruning test for parametric interval linear systems. In: *Proceedings of the 15th International Symposium on Operational Research SOR'19, Bled, Slovenia, September 25-27, 2019* (Zadnik Stirn et al., L., ed.). BISTISK d.o.o., Ljubljana, Slovenia, 506–511.
- [13] Rodríguez, M. M. L., and José, V.-P.: On finite linear systems containing strict inequalities. *J. Optim. Theory Appl.* **173** (2017), 131–154.
- [14] Rohn, J.: Strong solvability of interval linear programming problems. *Comput.* **26** (1981), 79–82.
- [15] Rohn, J., and Kreslová, J.: Linear interval inequalities. *Linear Multilinear Algebra* **38** (1994), 79–82.
- [16] Roos, C., Terlaky, T., and Vial, J.-P.: *Interior Point Methods for Linear Optimization*. 2nd edition. Springer, New York, 2006.

Sensitivity of small-scale beef cattle farm's profit under conditions of natural turnover

Robert Hlavatý¹, Igor Krejčí²

Abstract. We continue our long-term research focused on beef herd management optimization from the perspective of a small-scale farmer in the Czech Republic. We have built a linear programming model of beef herd development spanning a ten-year period under the constraints of limited farm capacity with the main variables being the heifer acquisition and selection of heifers for either rearing or fattening process. In this paper, we use the aforementioned approach to determine the sensitivity of farmer's profit to the subsidies, input costs and selling prices. This time, we specifically focus on the natural turnover of the herd, meaning that no acquisition of cattle is possible. The sensitivity analysis shows that from the three factors influencing the profit, it is the change in input costs that plays a crucial role in the farmer's profit.

Keywords: beef cattle farm, linear programming, optimisation, sensitivity

JEL Classification: C61, Q12

AMS Classification: 90C05, 90C90

1 Introduction

Small farms, the focus of our research, represent the most common form of business in EU Agriculture [22]. The small farms model is the oldest kind of agriculture business model that retains the key role in the Common Agricultural Policy [14], [5]. The small size of the farms is the first aspect of our research. The second aspect shows the crucial difficulties of agriculture in general – the dependency on biological processes, long delays between action and reaction and seasonality [2]. From this perspective the cattle farming is considered one of the most complex due to the natural delays embodied in the relevant biological processes, especially breeding and fattening periods [20]. Moreover, the primary producers of meat commonly face the problems of low profitability and weak market position [19], [18], [7] and the Czech Republic is not an exception [28].

Authors who deal with cattle modelling typically examine the feeding and insemination strategies [21], [17], [3]. Another trend is based on genomic selection strategies [25], [11]. These strategies clearly aim at strong leverage points, however, these leverage points are commonly beyond the reach of the average farmer. Many of these politics are appropriate for the national level [1]. Other strategies for strengthening the position of the farmers and lowering the risk focus on diversification (commonly agritourism) [26], [23] or some kind of direct distribution of the products to final consumers [8], [15]. Despite the above-mentioned strategies prove to be efficient in many cases, they also require new skills and a different mindset, which is not typical for common farming [30]. Consequently, such requirement could result in need of changes in agricultural and rural education on individual and community level [12], [29].

Our current research is focused on different aspects. The goal is to examine the common practice of the Czech farmers, use the typical timing and common prices and identify the leverage points under conditions of the average farmer. We are not showing the benefits of a specific action and rather focus on common decision-making of the small-scale beef cattle farmers and show the benefits of optimal choices. The beginning of the whole research was based on interviews with small farmers and focus groups [24], [16]. We stayed in contact with the farmers through the whole modelling process and implemented their opinions and needs in the optimization model that describe the development of the beef herd throughout ten years period. The model outlines and brief analysis of optimal decision making were first presented by [9] and later fully described in extended form by [10]. One of the core leverage points was the optimal acquisition of heifers. In this paper we focus on the natural turnover, i.e. the model situation does not allow the purchasing of heifers and the growth of the herd is dependent only on the biological processes. We test how the changes of prices and subsidies influence the optimal distribution (fattening vs rearing) of the heifers in the growth phase of the business (where no limit of the herd size is considered).

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

2 Materials and methods

This section is divided into two chapters. The first describes the nature of the problem and introduces the individual variables and the second presents the linear programming model.

2.1 Problem description and variables

The beef herd consists of different categories that must be taken into consideration. The categories are represented by variables in 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
C_i^B	Set of all bull 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
F_BUL_i	Set of all fattening bulls in generation i of age $\in (7,24]$ months
B_HEI_i	Set of all breeding heifers in generation i of age $\in (7,26]$ 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

Table 1 Variables and beef herd categories

The *generation* is understood here as a set of all calves born in the same month and all older stages of its lifespan. The *generations* are indexed by i . The pregnancies of heifers (or cows, consequently) are indexed by j . The dimension of variable indices i and j is not specifically set as our modelling approach does not require enumerating it explicitly. Each generation starts at the moment a calf C_i is born (or set of calves in fact, however, for the sake of simplicity, each category will be referred to as a single animal in the description hereby). The calf C_i can be a female C_i^H or male C_i^B . The male calf C_i^B later grows into the mature fattening bull F_BUL_i which is later sold (slaughtered).

The female calf can either become a fattening heifer F_HEI_i which is slaughtered at the age between 18-24 months or it can become a breeding heifer B_HEI_i . The breeding heifer becomes pregnant and turns into $P_HEI_A_{ij}$, which denotes the first stage of the pregnancy (≤ 5 months) and then it enters the second stage of the pregnancy $P_HEI_B_{ij}$ (> 5 months) called a heavily pregnant heifer, resulting in calf $C_{(i+1)j}$ birth. The two stages of pregnancy must be distinguished in the model due to the different costs involved. After the first birth, a new pregnancy soon occurs after a service period and the heifer becomes a mature cow $P_COW_{i(j+1)}$ that is pregnant for $(j + 1)$ th time. The pregnancy results in another birth of the calf $C_{(i+1)(j+1)}$ and this process is repeated until n -th pregnancy. When any calf $C_{(i+1)j}, \forall i, j = 1, \dots, n$ is born, a new *generation* is started and the same cycle begins.

2.2 Linear optimization model

Concerning the variables and relationships between categories described in the previous subchapter, we construct the linear optimization model. The problem is generally described with the following optimization model:

$$\begin{aligned}
 &\text{maximise } P \\
 &\text{s.t. } H_k \in \Phi, \forall k \\
 &\quad H_k \subset \mathbb{R}^+
 \end{aligned} \tag{1}$$

The objective of the optimization problem is maximizing the profit P . H_k denotes the set of all variables that occur in month k and involves all variables described in Table 1. Φ is the polyhedral set of all constraint imposed on the problem. Note that all variables can attain real non-negative values and we do not assume integrality constraints in our model. This is because our approach is rather average-based and works with the entire categories of cattle instead of modelling single animals. This relaxation still allows observing the development of the beef herd without making the problem hard in terms of computational complexity. Before the detailed linear optimization model is described, it is necessary to introduce several parameters and cost coefficients that enter the model, as we describe them in Table 2.

Parameter	Description	Default ^d
ω	Calves' mortality	5%
ψ	Heifer and cow culling rate	15%
s^{FAT}	Monthly cost per cattle for fattening ^a	46 EUR
s^{COW}	Monthly cost per suckler cow, including the calve ^a	70 EUR
s^{HEI_A}	Monthly cost per heifer up to 5 months of gestation ^a	39 EUR
s^{HEI_B}	Monthly cost per heavily pregnant heifer ^a	42 EUR
r^{SUB}	Subsidies on beef calves ^b	138 EUR/calf
r^{COW}	Yearly subsidies on beef cows ^b	7 EUR/Live unit
r^{LAND}	Yearly average land subsidies ^b	344 EUR/ha
p^{FAT}	Fattening bull unit sale price ^c	999 EUR
p^{COW}	Fattening heifer/suckler cow unit sale price ^c	740 EUR
hsc	Hectares per suckler cow ^{b,c}	1.5 ha

Table 2 Parameters of the beef herd (Sources: ^a[13], ^b[4], [27], ^c[24] and ^d[6])

All prices are based on the data from 2016. The costs are already cleared of the labour costs as the small-scale farms depend mostly on their own family workforce. The objective function P maximises the profit over all beef herd categories using the coefficients from Table 2.

$$\begin{aligned}
P = & \sum_i \sum_{j=1}^n \left(r^{SUB} C_i^H + (p^{COW} - 14 * s^{FAT}) F_HEI_i \right. \\
& + \left(19 * \frac{1}{12} * hsc * r^{LAND} - 19 * s^{FAT} \right) B_HEI_i \\
& + \left(5 * \frac{1}{12} * hsc * r^{LAND} - 5 * s^{HEI_A} \right) P_HEI_A_{i1} \\
& + \left(7 * \frac{1}{12} * hsc * r^{LAND} - 7 * s^{HEI_B} + \psi p^{COW} \right) P_HEI_B_{i1} \\
& + \left(12 * \frac{1}{12} * hsc * (r^{LAND} + r^{COW}) - 12 * s^{COW} + \psi p^{COW} \right) P_COW_{j+1} \\
& + \sum_i \left(r^{SUB} C_i^B + (p^{FAT} - 14 * s^{FAT}) F_BUL_i \right)
\end{aligned} \tag{2}$$

The first double sum in the objective expresses the costs and subsidies for female calves, heifers and cows, the second sum is related to male calves and bulls. The parameters in Table 2 are expressed as monthly/yearly ratings and it is necessary to add various multipliers into the objective function to capture the real-time existence of each category in the herd. A detailed explanation is provided by [10].

The objective is maximised with respect to constraint set Φ which is formed of the following equations and inequalities:

$$(1 - \omega)C_i = C_i^H + C_i^B, \forall i \tag{3}$$

$$C_i^H = C_i^B, \forall i \tag{4}$$

$$C_i^H = F_HEI_i + B_HEI_i, \forall i \tag{5}$$

$$F_HEI_i \geq 0.5 C_i^H, \forall i \tag{6}$$

$$B_HEI_i \leq 0.5 C_i^H, \forall i \tag{7}$$

$$B_HEI_i = P_HEI_A_{ij} = P_HEI_B_{ij}, \forall i; j = 1 \tag{8}$$

$$P_COW_{i(j+1)} = (1 - \psi)P_HEI_B_{ij}, \forall i; j = 1 \tag{9}$$

$$P_COW_{i(j+1)} = (1 - \psi)P_COW_{i(j+2)}, \forall i; j = 1 \dots n - 2 \tag{10}$$

$$C_i^B = F_BUL_i, \forall i \tag{11}$$

Equations 3 and 4 express the even distribution of new-born calves to heifer calves C_i^H and bull calves C_i^B with the given mortality rate ω . Each heifer calf C_i^H later becomes either fattening heifer F_HEI_i or breeding heifer B_HEI_i as shown by Equation 5. Inequalities 6 and 7 show the distribution to fattening and breeding heifers which may be different for each generation, depending on the current profitability of that decision. Equation 8 only shows the progress in pregnancy of heifers. Equations 9 and 10 express the culling procedure which is done after a birth occurs in the case of heifer or mature cow. Equation 11 merely shows the ageing process of young bulls C_i^B , into mature ones F_BUL_i .

3 Results and discussion

We run the optimisation model described in the previous chapter for different scenarios. We used the *OpenSolver* extension for MS Excel in order to solve the problem and each run took approximately 120 seconds. The baseline scenario under conditions specified in chapter 2 results in the 10-years profit equal to 123,681 EUR. The strict constraint on the heifers' acquisition does not leave the farmer with big decision space. The herd develops similarly without the upward shifts typical for the situation when the farmer purchases heavily pregnant heifers. Despite the model is linear, the change of the profit does not necessarily change in a proportional way to the parameter changes. This happens when the change of the parameter results in the change of heifers' distribution.

Scenario	10-year Profit (EUR)	Heifers for fattening	Heifers for rearing	Relative profit to baseline
Baseline	123 681	64.92%	35.08%	X
Selling prices +5%	141 502	64.92%	35.08%	1.14
Selling prices +10%	159 323	64.92%	35.08%	1.29
Selling prices -5%	105 859	64.92%	35.08%	0.86
Selling prices -10%	88 038	64.92%	35.08%	0.71
Subsidies +5%	149 396	50.00%	50.00%	1.21
Subsidies +10%	170 441	50.00%	50.00%	1.38
Subsidies -5%	105 376	67.77%	32.23%	0.85
Subsidies -10%	90 118	72.33%	27.67%	0.73
Prices of inputs +5%	93 574	67.77%	32.23%	0.76
Prices of inputs +10%	66 098	72.33%	27.67%	0.53
Prices of inputs -5%	160 245	50.00%	50.00%	1.30
Prices of inputs -10%	192 139	50.00%	50.00%	1.55

Table 3 Scenarios comparison

The maximum share of the heifers for rearing is 50%. This distribution starts at the increase of the subsidies by 1.75% and decrease of the input prices by 1.9%. It is also worth mentioning that the change of the selling prices isn't necessary caused by the prices of the meat but it could be also caused by better nutrition, i.e. heavier cattle for slaughter. Table 3 compares the scenarios with the stress on a different distribution of heifers.

Figure 1 compares the influence of the changes on the profit. Because the growth and decrease of the selected parameters have different interpretation (growth of subsidies and selling prices is good but the growth of inputs' prices is undesirable from the producers' point of view), the change is called better and worse showing whether the change represents the improvement or deterioration of the farmers' situation. It is clear from the Figure 1 that the 10-years profit under conditions of the pure natural turnover sensitive on the prices of inputs at most. The possibility of heifers' acquisition allows the farmer to increase the size of the herd faster. Therefore, if the purchase of the heifers is possible, the sensitivity on subsidies and selling prices grows especially for the increase of the subsidies or prices of the sold product.

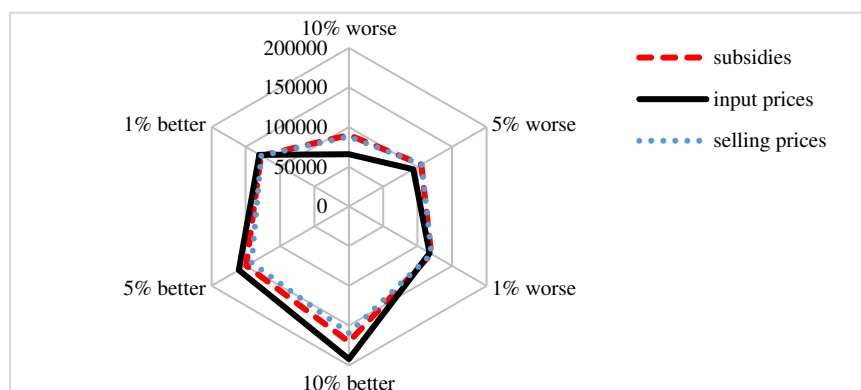


Figure 1 Impact of changes on profit

Two scenarios (subsidies -10% and input prices + 10%) show the behaviour that is typical for situations with low sustainability. It is also the weakness of this modelling approach that must be clearly expressed. Because the model has the 10-years perspective, the optimal distribution of heifers under conditions of such terminated process leads to the preference of heifers for fattening for the last two years when the ratio of the input costs and output prices is not favourable (this increased number of sold cattle and decrease the number of cattle, which are connected with the costs). The real farmer that wants to stay on the market would achieve even lower profit because this strategy isn't possible in the real life. It is also important to stress that the definition of the maximal herd size has also a significant impact on the optimal decisions.

The farm is the system, where everything is interconnected. In comparison when the farmer could purchase heavily pregnant heifers [10], the dependence only on the natural turnover in the growth phase is extremely risky. One must understand that such phase is typically connected with the investment into farms capacities (stables, storage, land) commonly accompanied by the loan repay. Even the baseline scenario shows that the heifers' distribution leads to slower capacity utilization, which in other words is postponing the maximal annual profit but the market situation (parameter values) leads to a preference of fattening to maximize the 10-years profit.

4 Conclusion

We presented the modelling approach based on the monthly cuts, which was beneficial when we compared the model behaviour with the real farms' development. In this paper, we have focused on the sensitivity of decision making regarding the heifers sorting. For this analysis we focused on three categories of parameters – subsidies, selling prices and prices of inputs. All categories consist of more model parameters but for paper purposes, we change all parameters from one category in a similar way. The modelling situation describes the farmer that does not want or cannot purchase heifers during the growth phase of the business. Under these conditions, the most crucial is the change of the input prices. The sensitivity testing showed that the subsidies and the prices of the inputs are close to the situation when the farmer prefers the fastest possible growth of the herd.

References

- [1] Amer, P.R. (2012). Turning science on robust cattle into improved genetic selection decisions. *Animal*, 6(4), 551–556.
- [2] Behzadi, G., O'Sullivan, M.J., Olsen, T.L., & Zhang, A. (2018). Agribusiness supply chain risk management: A review of quantitative decision models. *Omega*, 79, 21–42.
- [3] Canozzi, M.E.A., Marques, P.R., De Souza Teixeira, O., Pimentel, C.M.M.M., Dill, M.D., & Barcellos, J.O.J. (2019). Typology of beef production systems according to bioeconomic efficiency in the south of Brazil. *Ciencia Rural*, 49(10), 1–9.
- [4] Czech Beef Breeders Association (2019). Sazby podpor pro rok 2019 [online]. Available from: <http://www.cschms.cz/index.php?page=novinka&id=2745> [Accessed 24 Aug 2020].
- [5] 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].
- [6] 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].
- [7] Fousekis, P., Katrakilidis, C. & Trachanas, E. (2016). Vertical price transmission in the US beef sector : Evidence from the nonlinear ARDL model. *Economic Modelling*, 52, 499–506.
- [8] Govindan, K. (2018). Sustainable consumption and production in the food supply chain: A conceptual framework. *International Journal of Production Economics*, 195, 419–431.
- [9] Hlavatý, R. & Krejčí, I. (2020). Optimizing production in small-scale beef cattle farm. In S.Kapounek & H.Vránová (Eds.), *38th International Conference on Mathematical Methods in Economics conference proceedings* (pp. 166-172).
- [10] Hlavatý, R., Krejčí, I., Houška, M. Moulis, P., Rydval, J., Pitrova, J. Pilař, L. Horáková, T. & Tichá, I. (2021). Understanding the decision making in small-scale beef cattle herd management through a mathematical programming model. *International Transactions in Operational Research*, [in Press]
- [11] Hong, J., Mei, C., Raza, S.H.A., Khan, R., Cheng, G., & Zan, L. (2020). SIRT5 inhibits bovine preadipocyte differentiation and lipid deposition by activating AMPK and repressing MAPK signal pathways. *Genomics*, 112(2), 1065–1076.

- [12] Husák, J. & Hudečková, H. (2017). Conditions for Development of Rural Community Education in the Czech Republic. *Journal on Efficiency and Responsibility in Education and Science*, 10(3), 64–70.
- [13] 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].
- [14] Khalil, C.A., Conforti, P., Ergin, I. & Gennari, P. (2017). *Defining Smallholders To Monitor Target 2.3. of the 2030 Agenda for Sustainable Development*. Rome, No. ESS / 17-12.
- [15] Kim, S., Lee, S.K., Lee, D., Jeong, J. & Moon, J. (2019). The effect of agritourism experience on consumers' future food purchase patterns. *Tourism Management*, 70, 144–152.
- [16] Koláčková, G., Krejčí, I. & Tichá, I. (2017). Dynamics of the small farmers' behaviour – scenario simulations. *Agricultural Economics (Czech Republic)*, 63(3), 103–120.
- [17] Liu, J., Tian, K., Sun, Y., Wu, Y., Chen, J., Zhang, R., He, T. & Dong, G. (2020). Effects of the acid-base treatment of corn on rumen fermentation and microbiota, inflammatory response and growth performance in beef cattle fed high-concentrate diet. *Animal*, 14(9), 1876–1884.
- [18] Lopes, R.B., Canozzi, M.E.A., Canellas, L.C., Gonzalez, F.A.L., Corrêa, R.F., Pereira, P.R.R.X. & Barcellos, J.O.J. (2018). Bioeconomic simulation of compensatory growth in beef cattle production systems. *Livestock Science*, 216, 165–173.
- [19] Manevska-Tasevska, G., Hansson, H. & Rabinowicz, E. (2014). Input saving possibilities and practices contributing to more efficient beef production in Sweden. *Agricultural and Food Science*, 23(2), 118–134.
- [20] Mayberry, D., Ash, A., Prestwidge, D., & Herrero, M. (2018). Closing yield gaps in smallholder goat production systems in Ethiopia and India. *Livestock Science*, 214, 238–244.
- [21] Mourits, M.C.M., Huirne, R.B.M., Dijkhuizen, A.A., Kristensen, A.R. & Galligan, D.T. (1999). Economic optimization of dairy heifer management decisions. *Agricultural Systems*, 61(1), 17–31.
- [22] Neuenfeldt, S., Gocht, A., Heckeley, T. & Ciaian, P. (2019). Explaining farm structural change in the European agriculture: a novel analytical framework. *European Review of Agricultural Economics*, 46(5), 713–768.
- [23] Pitrova, J., Krejčí, I., Pilar, L., Moulis, P., Rydval, J., Hlavatý, R., Horáková, T. & Tichá, I. (2020). The economic impact of diversification into agritourism. *International Food and Agribusiness Management Review*, 23(5), 1–22.
- [24] 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.
- [25] Raza, S.H.A., Khan, S., Amjadi, M., Abdelnour, S.A., Ohran, H., Alanazi, K.M., Abd El-Hack, M.E., Taha, A.E., Khan, R., Gong, C., Schreurs, N.M., Zhao, C., Wei, D., & Zan, L. (2020). Genome-wide association studies reveal novel loci associated with carcass and body measures in beef cattle. *Archives of Biochemistry and Biophysics*, 694, 108543.
- [26] Schilling, B.J., Attavanich, W. & Jin, Y. (2014). Does Agritourism Enhance Farm Profitability? *Journal of Agricultural and Resource Economics*, 39(1), 69–87.
- [27] Syrůček, J., Krpálková, L., Kvpilík, J., and Vacek, M., 2017. *Kalkulace Ekonomických Ukazatelů v Chovu Skotu*. Prague: Institute of Animal Science.
- [28] Syrůček, J., Kvpilík, 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.
- [29] Tomšíková, K., Hudečková, H. & Tomšík, K. (2019). Enhancing Attractiveness of Secondary Agricultural Education in The Czech Republic. *Journal on Efficiency and Responsibility in Education and Science*, 12(4), 135–145.
- [30] Wesselink, R., Blok, V., van Leur, S., Lans, T. & Dentoni, D. (2015). Individual competencies for managers engaged in corporate sustainable management practices. *Journal of Cleaner Production*, 106, 497–506.

The Use of Genetic Algorithm in Clustering of ARMA Time Series

Vladimír Holý¹, Ondřej Sokol²

Abstract. Time series clustering is a well-covered topic in the data mining literature. In this paper, we assume that each of a large number of time series follows one of several autoregressive-moving-average (ARMA) models. We propose to jointly assign time series to clusters and estimate the ARMA coefficients in each cluster by a genetic algorithm. We also simultaneously determine the number of clusters by minimizing the Akaike information criterion (AIC). We illustrate our approach in an application to weekly product sales of a retail drugstore and focus on the specification of a genetic algorithm. First, we investigate the suitability of a k-means solution based on a distance between the ARMA coefficients as an initial solution. Second, we study the influence of the genetic algorithm parameters such as the number of generations, the size of the population, the probability of mutation, and the ratio of elite individuals.

Keywords: Time Series Clustering, ARMA Model, Genetic Algorithm, Retail Analytics

JEL Classification: C32, C38, C63
AMS Classification: 62M10, 62H30

1 Introduction

Xiong and Yeung [13, 14] proposed a method for clustering of time series using a mixture of autoregressive moving average (ARMA) models. This approach belongs to the class of model-based time series clustering. The task is to assign each observed time series to one of several clusters with common ARMA dynamics and estimate ARMA parameters in each cluster. For this purpose, [13, 14] utilized an expectation-maximization (EM) algorithm. A drawback of this iterative method is that it finds only a local maximum of the likelihood function. Although [14] addressed this issue by using a stochastic variant of the standard EM algorithm, they still operated within the EM framework.

We take another direction and adopt a genetic algorithm to tackle the optimization problem. A genetic algorithm is a nature-inspired metaheuristic that is able to escape local extrema. This approach can also simultaneously determine the number of clusters by maximizing the Akaike information criterion (AIC). In this paper, we study the impact of an initial solution and control parameters of a genetic algorithm.

For other applications of genetic algorithms in clustering problems, see [3, 5, 7, 9]. For a literature review of nature-inspired metaheuristics in clustering, see [4, 6, 8]. For a literature review of time series clustering, see [1, 12].

2 Model

Similarly to [13, 14], we assume that there are N time series of length T denoted as $Y_i = (y_{i,1}, \dots, y_{i,T})$, $i = 1, \dots, N$. Our setting can be simply extended to allow for different lengths of time series; however, we focus on the case with common T to ease the notation. Each time series $i = 1, \dots, N$ belong to one of K clusters; we denote this assignment as $\kappa_i \in \{1, \dots, K\}$. Each time series $i = 1, \dots, N$ also has the ARMA(P, Q) structure

$$y_{i,t} = \omega_{\kappa_i} + \sum_{j=1}^P \varphi_{\kappa_i,j} y_{i,t-j} + \sum_{j=1}^Q \theta_{\kappa_i,j} e_{i,t-j} + e_{i,t}, \quad e_{i,t} \stackrel{iid}{\sim} \mathbf{N}(0, \sigma_{\kappa_i}^2), \quad t = 1, \dots, T. \quad (1)$$

Each cluster k therefore has its own parameters $\Theta_k = \{\omega_k, \varphi_{k,1}, \dots, \varphi_{k,P}, \theta_{k,1}, \dots, \theta_{k,Q}, \sigma_k^2\}$. Again, we ease the notation by assuming that all time series has the common ARMA order P and Q but this can be straightforwardly relaxed.

¹ Prague University of Economics and Business, Department of Econometrics, Winston Churchill Square 1938/4, 130 67 Prague 3, Czechia, vladimir.holy@vse.cz

² Prague University of Economics and Business, Department of Econometrics, Winston Churchill Square 1938/4, 130 67 Prague 3, Czechia, ondrej.sokol@vse.cz

Our goal is to estimate the assignment $\kappa = \{\kappa_1, \dots, \kappa_N\}$ as well as the parameters $\Theta = \{\Theta_1, \dots, \Theta_K\}$. For this purpose, we use the maximum likelihood method. The estimates $\hat{\kappa}$ and $\hat{\Theta}$ are then obtained by maximizing the log-likelihood

$$\{\hat{\kappa}, \hat{\Theta}\} \in \arg \max_{\kappa, \Theta} \sum_{i=1}^N \ln f(Y_i; \kappa_i, \Theta_{\kappa_i}), \quad (2)$$

where $f(Y_i; \kappa_i, \Theta_{\kappa_i})$ is the joint density function for i -th time series. So far we have considered the number of clusters K to be given.

Alternatively, if K is unknown, we estimate it along with κ and Θ by minimizing the Akaike information criterion (AIC)

$$\{\hat{K}, \hat{\kappa}, \hat{\Theta}\} \in \arg \min_{K, \kappa, \Theta} 2K(2 + P + Q) - 2 \sum_{i=1}^N \ln f(Y_i; \kappa_i, \Theta_{\kappa_i}). \quad (3)$$

Both (2) and (3) are discontinuous optimization problems due to the presence of discrete variables κ (and K). To numerically find the optimal solution, [13, 14] used an expectation-maximization (EM) algorithm. In this paper, we use a genetic algorithm.

3 Genetic Algorithm

In this section, we highlight some specifics of our used genetic algorithm. For a textbook treatment of evolutionary computing, see e.g. [2].

We start by obtaining a suitable initial solution which can greatly improve computational performance. For this purpose, we use k-means clustering. First, we estimate ARMA models separately for each of N time series. Second, we partition N sets of $2 + P + Q$ estimated coefficients into K clusters by minimizing the within-cluster sum of squares. Third, we re-estimate ARMA models for each of K clusters, possibly containing multiple time series.

The genetic algorithm used in our study consists of the following steps. First, several possible candidate solutions are randomly generated. Together with the k-means solution, they form the initial population. Each candidate solution in the population is coded according to the renumbering procedure of [5] which labels clusters sequentially from 1 to K based on their first occurrence in $\kappa_1, \dots, \kappa_N$. This procedure is quite important as it prevents any ambiguity in the representation of candidate solutions. Unlike in [5], we use a fixed maximum number of clusters. However, as we are minimizing AIC which depends on the number of clusters, the resulting number of used clusters can be lower, i.e. there can be some empty clusters.

Next, new generations are iteratively generated. A fixed number of candidate solutions with the highest values of the fitness (objective) function, known as the elite individuals, are carried over from the previous generation to the next generation. The remaining new candidate solutions are generated using the linear rank selection, uniform crossover, and random mutation operators. For each new candidate solution, the linear rank selection operator randomly selects two parents from the previous generation (including the elite individuals) with probability proportional to their rank according to the fitness function. The uniform crossover operator then copies the cluster number from either parent with equal probability for each time series. The random mutation operator further changes some cluster numbers with a small given probability to completely random ones. Finally, the renumbering procedure is applied to each candidate solution. The computation is terminated after a fixed number of generations.

The algorithm requires four parameters to be determined – the number of generations, the size of the population, the probability of mutation, and the ratio of elite individuals.

4 Numerical Performance

To illustrate the applicability of the proposed genetic algorithm in clustering of ARMA times series, we cluster product categories in a Czech drugstore retail chain according to their weekly sales in 2018. We have $N = 65$ time series of length $T = 52$. A similar dataset was also analyzed and described in more detail in [3], [10], [11]. We assume that each cluster follows an ARMA(1,1) model.

This empirical study is inspired by the general problem of supplying the retail chain. Assume that the retail chain regularly supplies its stores from several supply centers. Each supply center focuses on only a few selected product categories. If the products in the supply center are similar in terms of sales dynamic, it is easier to optimize the inventory in the store and the frequency of deliveries. Hence, the aim is to determine which time series of product

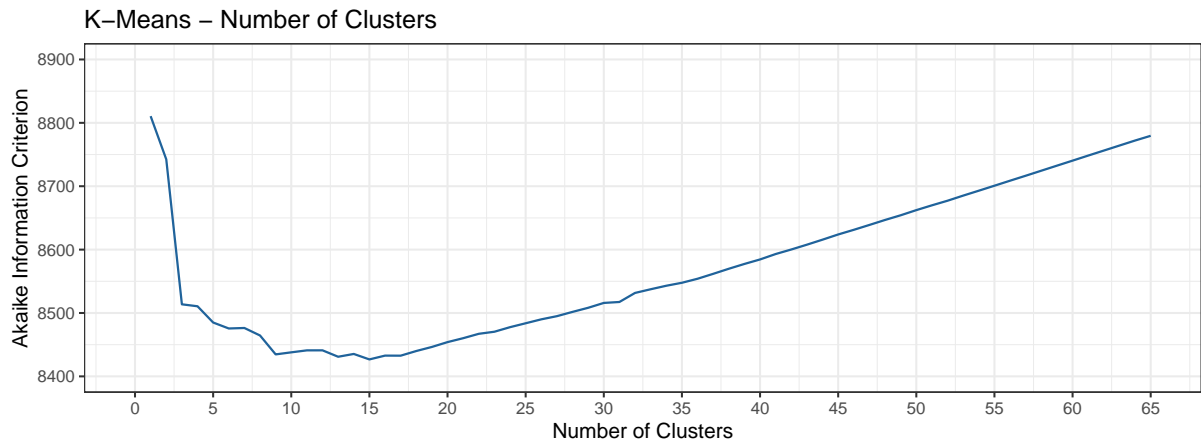


Figure 1 AIC of the solution obtained by the k-means algorithm with various number of clusters.

categories sales have similar dynamics – and should be stored in the same supply center. In other words, we aim to cluster product categories’ time series of sales by their dynamics.

We start with the k-means solution. Figure 1 shows the AIC of the solution for numbers of clusters K from 1 to 65. Note that the k-means method finds a solution according to the within-cluster sum of squares and not log-likelihood or AIC (which consequently causes a bumpy shape of the curve). The best solution is obtained for $K = 15$ clusters with AIC 8427.

Next, we investigate the performance of the proposed genetic algorithm. We set the number of generations to 1000, the size of the population to 1000, the probability of mutation to 0.01, and the ratio of elite individuals to 0.10. Figures 2 and 3 then show the achieved AIC for some changes in these parameter values. As genetic algorithms are random in nature, we repeat each computation 300 times and report mean performance. Note that to be able to compare computational performance for different population sizes in the first plots of figures 2 and 3, we show standardized generations – each containing exactly 1000 candidate solutions. For the population size of 1000, standardized generation is the same as regular generation. For the population size of e.g. 200, standardized generation refers to every fifth regular generation.

We assess the usability of the k-means solution as an initial solution for the proposed genetic algorithm. Figures 2 and 3 show that the algorithm with random initialization starts at much worse candidate solutions and requires a lot of generations to get closer to the algorithm starting with the k-means solution. A suitable initial solution can therefore significantly speed up computations.

Finally, we assess the influence of the control parameters. The first plots of figures 2 and 3 show development of AIC for several population sizes. We can see that higher population sizes require more standardized generations to achieve better results. As a rule of thumb, the number of generations should be higher than the population size. If this holds, however, the algorithm is not very sensitive to the ratio between the number of generations and the population size.

The second plots of figures 2 and 3 show development of AIC for several mutation probabilities. Clearly, when the mutation probability is zero, the algorithm gets quickly stuck in a local optimum. When the mutation probability is too high, on the other hand, the algorithm converges much slower. In our case, the value of 0.005 offers the best performance.

The third plots of figures 2 and 3 show development of AIC for several elites ratios. We omit the case with no elite individuals as it would not guarantee a non-increasing curve of AIC and would require a very different scale in Figure 3. Nevertheless, we can see that lower numbers of elite individuals lead to a slower convergence rate. The value of 10 percent is then sufficient for good performance.

The best average AIC over the 300 repetitions is 8387 obtained for the size of the population 1000, the probability of mutation 0.005, and the ratio of elite individuals 0.10. The best solution overall has AIC 8380 and $K = 10$ clusters. This solution was found in several control parameter settings, including even the random initialization. This result emphasizes the stochastic nature of the algorithm and encourages to repeat computations several times to determine the stability of the solution.

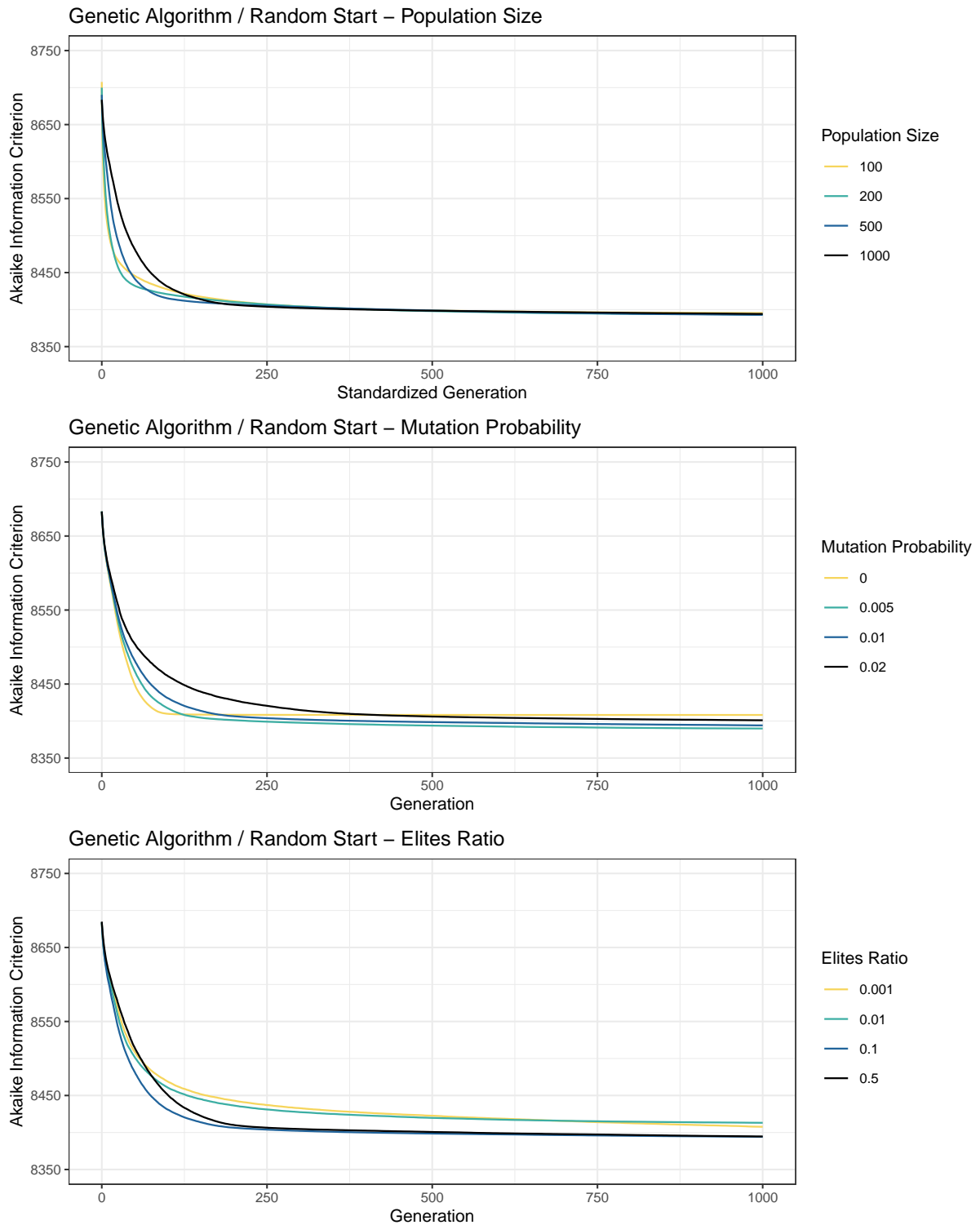


Figure 2 AIC of the solution obtained by the genetic algorithm with completely random initialization.

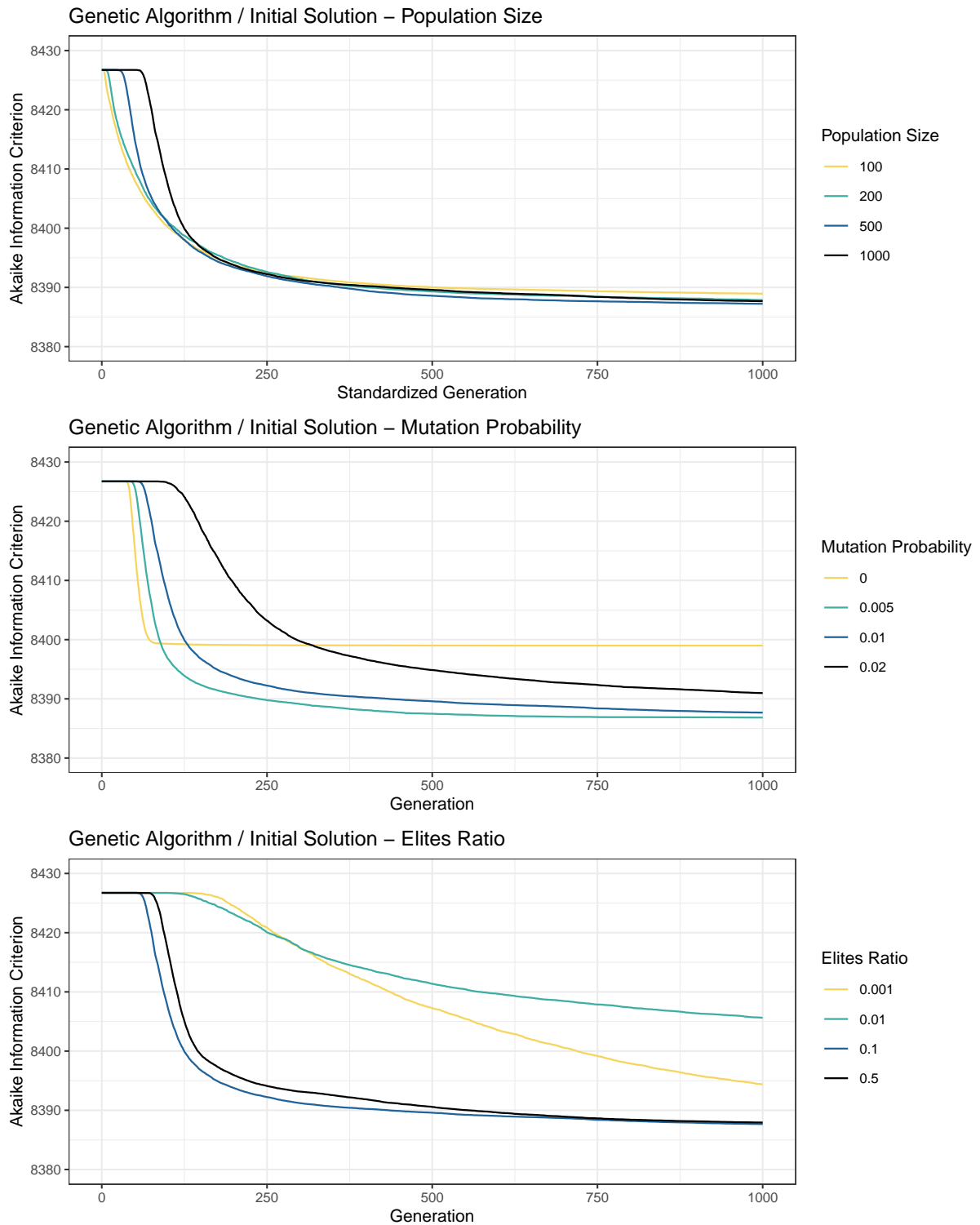


Figure 3 AIC of the solution obtained by the genetic algorithm with k-means solution as an initial solution.

5 Conclusion

We have demonstrated the use of a genetic algorithm in clustering of ARMA time series. We have shown that using the k-means method based on estimated coefficients of individual ARMA models yields a suitable initial solution which can significantly speed up computations. We have also suggested values for the size of the population, the probability of mutation, and the ratio of elite individuals parameters.

A limitation of our study is that we present results only for a moderate size of the problem. Specifically, in an application to a retail drugstore, we cluster 65 product categories using weekly sales time series with 52 observations. A more extensive empirical and simulation study is required to draw conclusions for larger problems. Furthermore, the comparison to the expectation-maximization (EM) algorithm is beyond the scope of this paper but remains a topic for our future research. Another direction we want to take in the future is extending the model specification to accommodate for various distributions and time-varying parameters; perhaps in the framework of generalized autoregressive score models.

Acknowledgements

This research was supported by the Internal Grant Agency of the Prague University of Economics and Business under project F4/27/2020. 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] Aghabozorgi, S., Seyed Shirshorshidi, A., & Ying Wah, T. (2015): Time-Series Clustering - A Decade Review. *Information Systems* **53**, 16–38.
- [2] Eiben, A. E., and Smith, J. E. (2015): *Introduction to Evolutionary Computing*. Second edition. Springer, Berlin, Heidelberg.
- [3] Holý, V., Sokol, O., & Černý, M. (2017): Clustering Retail Products Based on Customer Behaviour. *Applied Soft Computing* **60**, 752–762.
- [4] Hruschka, E. R., Campello, R. J. G. B., Freitas, A. A., & de Carvalho, A. C. P. L. F. (2009): A Survey of Evolutionary Algorithms for Clustering. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **39**, 133–155.
- [5] Hruschka, E. R. & Ebecken, N. F. F. (2003): A Genetic Algorithm for Cluster Analysis. *Intelligent Data Analysis* **7**, 15–25.
- [6] José-García, A. & Gómez-Flores, W. (2016): Automatic Clustering Using Nature-Inspired Metaheuristics: A Survey. *Applied Soft Computing* **41**, 192–213.
- [7] Murthy, C. A. & Chowdhury, N. (1996): In Search of Optimal Clusters Using Genetic Algorithms. *Pattern Recognition Letters* **17**, 825–832.
- [8] Nanda, S. J. & Panda, G. (2014): A Survey on Nature Inspired Metaheuristic Algorithms for Partitional Clustering. *Swarm and Evolutionary Computation* **16**, 1–18.
- [9] Sokol, O. & Holý, V. (2020): Computational Aspects of Product Clustering Based on Transaction Incidence. In: *Proceedings of the 20th International Conference Quantitative Methods in Economics*. Letra Edu, Púchov, 312–318.
- [10] Sokol, O. & Holý, V. (2020): The Role of Shopping Mission in Retail Customer Segmentation. *International Journal of Market Research*.
- [11] Sokol, O., Holý, V., & Cipra, T. (2021): Customer and Product Clustering in Retail Business. In: *Proceedings of the 7th World Conference on Soft Computing*. Springer, Baku, 529–537.
- [12] Warren Liao, T. (2005): Clustering of Time Series Data - A Survey. *Pattern Recognition* **38**, 1857–1874.
- [13] Xiong, Y. & Yeung, D.-Y. (2002): Mixtures of ARMA Models for Model-Based Time Series Clustering. In: *Proceedings of the 2nd IEEE International Conference on Data Mining*. IEEE, Maebashi City, 717–720.
- [14] Xiong, Y. & Yeung, D.-Y. (2004): Time Series Clustering with ARMA Mixtures. *Pattern Recognition* **37**, 1675–1689.

Enumerative Core of Pólya's Theorems on Random Walk and the Role of Generating Functions in their Proofs

Richard Horský

Abstract. The theorems of George Pólya on random walks are a popular topic in the theory of probability and have been a great inspiration for next research and for many other results. In the proofs of them we use the facts from the theory of generating functions. We concentrate on the enumerative core of wanderer's problem.

We derive some effective estimate for the case of recurrent random walk in two dimensions. By means of them we estimate waiting time for the return of an aimless wanderer to Times Square with probability at least 90%.

Keywords: random walk, $2d$ -regular and vertex transitive graph, generating function power series

JEL Classification: C02

AMS Classification: 05C30

1 Introduction

Random walk (promenade au hazard, Irrfahrt) is a concept which was introduced by G. Pólya in 1921 (see [8]). It is a quite simple process: we jump from one point to another adjacent one in \mathbb{Z}^d and ask the question whether we may reach a predetermined point? To express otherwise in our real life imagine a wanderer moving aimlessly and randomly in a rectangular net of streets, as for example one in the city of New York. What is the probability of his return to the start point? Pólya shows that it depends on the dimension d . If this dimension is equal or less than 2 we can reach the starting (or predetermined) point almost surely (the case of the wanderer in New York) but for the dimension at least 3 we never return with a positive probability. In the literature the designation recurrent walk and transient walk resp. is usually used. The Pólya's theorems may be now expressed in brief form: *The simple random walk on the 2-dimensional grid is recurrent but on the 3-dimensional grid it is transient.*

Our presentation is based on W. Feller's classics [4]. A few more references are P. Billingsley [2], K. Lange [5], D.A. Levin and Y. Peres [6], J. Novak [7] and Rogers [10]. This list could be much extended.

In this section we remind some notions from the theory of graphs and some facts concerning to generating functions and power series which are necessary for the next argumentation. The real numbers (scalars) will be denoted by Greek letters while the vertices (d -tuples), as well as integer parameters (dimension, length etc.) by Latin letters.

1.1 The grid (graph) \mathbb{Z}^d and some combinatoric enumeration

Suppose $G = (V; E)$ is a finite or infinite unoriented graph, a vertex $u \in V$, and a number $n \in \mathbb{N}_0$ (the domain of all non-negative integers). We denote by

$$d_n = d_n(u, G) \in \mathbb{N}_0, \text{ resp. by } l_n = l_n(u, G) \in \mathbb{N}_0, \quad (1)$$

the number of all walks $W = (u_0, u_1, \dots, u_n)$ in the graph G with $u_0 = u$ as the starting point and of the length $n \in \mathbb{N}_0$ (here is of course $\{u_{i-1}, u_i\} \in E$ for any $i = 1, \dots, n$), resp. those walks of length $n \in \mathbb{N}_0$ that revisit the starting point: $u_0 = u_i = u$ for some $i = 1, \dots, n$, so called *recurrent walks*.

The graph $G = (V; E)$ is called *vertex-transitive graph* if for any pair of its vertices $u, v \in V$ there is a graph automorphism $f: G \rightarrow G$ such that $f(u) = v$. The example of such a graph is the path $P = (\mathbb{Z}; \{\{k; k+1\}; k \in \mathbb{Z}\})$. In vertex-transitive graphs $d_n(v, G) = d_n(w, G)$, resp. $l_n(v, G) = l_n(w, G)$ for any $v, w \in V$ and $n \in \mathbb{N}_0$. Loosely speaking the choice of the starting point is in such a graph irrelevant.

Another important property of the graph is its *m-regularity*: the graph is called *m-regular graph* if all of its vertices have the same number m of neighbours. It is obvious that $d_n(u, G) = m^n$ in such a graph. For example the path $P = (\mathbb{Z}; \{\{k; k+1\}; k \in \mathbb{Z}\})$ is a 2-regular graph with $d_n(u, G) = 2^n$. This path is a special case of the graph

$$G_d = (\mathbb{Z}^d; E), \quad (2)$$

where $d \in \mathbb{N}$ (the set of all positive integers), \mathbb{Z}^d is the set of all d -tuples of integers, i.e. a vertex is of the form $u = (\alpha_1, \alpha_2, \dots, \alpha_d)$, $\alpha_j \in \mathbb{Z}$, $j = 1, \dots, d$ and $\{u; v\} \in E$ if and only if $\|u - v\|_1 = 1$, which means that

$$u - v \in \{(\pm 1, 0, 0, \dots, 0); (0, \pm 1, 0, \dots, 0); \dots; (0, \dots, 0, 0, \pm 1)\}.$$

One can go from u to v and back only by a unit step in the direction of one of the d coordinate axes. The graph (2) is obviously $2d$ -regular. At the same time this graph G_d is vertex-transitive; the automorphism sending a vertex u to a vertex v is simply the shift $f(x) = x + v - u$. It implies that for any dimension d , for arbitrary chosen vertices $u, v \in \mathbb{Z}^d$ and any length $n \in \mathbb{N}_0$ one has

$$d_n(u, G_d) = (2d)^n, \text{ and } l_n(u, G_d) = l_n(v, G_d).$$

Note that for the latter equality the vertex-transitivity is required and the $2d$ -regularity is not sufficing alone. The given circumstances allow us to choose the origin $o = (0, 0, \dots, 0) \in \mathbb{Z}^d$ as the starting point for our walks and we will assume it in what follows.

Finally, we define two more terms: by the symbol b_n we denote the number of all walks $W = (u_0, u_1, \dots, u_n)$ in the graph (2) with $u_0 = u_n = o$ (closed cycles) and by c_n the number of those closed cycles, where the origin occurs only in the first and the last term, i.e. $u_j \neq o$ for any $0 < j < n$ for positive n (Hamiltonian cycle) and we set $c_0 = 0$. By the way, these numbers do not depend on the choice of the starting point of the walk thanks to the vertex-transitivity of (2). We can observe that the following inequalities hold:

$$l_n \leq d_n = (2d)^n, \text{ and } c_n \leq b_n \leq d_n.$$

By pigeonholing the walks counted by l_n by their first return to the origin o at the step j we obtain for any $n \in \mathbb{N}_0$

$$l_n = \sum_{j=0}^n c_j d_{n-j} = \sum_{j=0}^n c_j (2d)^{n-j} \text{ and } \frac{l_n}{d_n} = \frac{1}{(2d)^n} \sum_{j=0}^n c_j (2d)^{n-j} = \sum_{j=0}^n \frac{c_j}{(2d)^j} \leq 1.$$

Let us realize that our main goal is to analyze the limit of the sequence $\frac{l_n}{d_n}$ and we will carry out this by the analyzing of the series

$$\sum_{j=0}^{\infty} \frac{c_j}{(2d)^j}. \quad (3)$$

It is obvious that the series (3) is convergent for any $d \in \mathbb{N}$ because it is the series with non-negative terms and bounded sequence of partial sums. The question is only whether the sum is equal or less than 1. In the section 2 we show that the situation is changed if we go from the value $d = 2$ to the value $d = 3$.

1.2 The generating functions, classical Abel's theorem and the power series with non-negative coefficients

The concepts of the generating function and the power series is well-known and it is not necessary to give the appropriate definitions here. They are widely used in various branches of mathematics and statistics (e.g. probability generating function or moment generating function in [3]). We will concentrate to certain detail in their use in the proofs of Pólya's theorems on random walk.

The articles [7] or [10] invoke the classical Abel's theorem. This theorem is dated to 1826 but firstly published till 1841, see [1].

Theorem 1 (Abel [1]). *If a power series $f(z) = \sum_{n=0}^{\infty} \alpha_n z^n$ with complex coefficients α_n converges for $|z| < 1$ and if the series $\sum_{n=0}^{\infty} \alpha_n$ converges to a sum $s \in \mathbb{C}$ (the domain of all complex numbers) then $\lim_{x \rightarrow 1^-} f(x) = s$, where the limit is taken along the real line from the left to 1.*

The general power series may have another radius of convergence, say $0 < R < +\infty$ and the center of the circle of its convergence need not be the origin, but simple linear transformation allows us to consider the center at the origin and the radius equal to 1. However, we will see that another easier result on power series, Proposition 1 below, is in fact sufficient in the proofs.

Proposition 1 (weak Abel's theorem). *If a power series $f(z) = \sum_{n=0}^{\infty} \alpha_n z^n$ has non-negative coefficients and converges for any $x \in [0; 1)$, then*

$$\lim_{x \rightarrow 1^-} f(x) = \sum_{n=0}^{\infty} \alpha_n \quad (4)$$

no matter whether the limit and the sum are finite or infinite.

Proof. For arbitrary $N \in \mathbb{N}$ and $x \in [0;1)$ we have

$$\sum_{n=0}^N \alpha_n = \lim_{x \rightarrow 1^-} \sum_{n=0}^N \alpha_n x^n \leq \lim_{x \rightarrow 1^-} \sum_{n=0}^{\infty} \alpha_n x^n \leq \sum_{n=0}^{\infty} \alpha_n.$$

The first equality is by continuity of the polynomial (partial sum in the power series), all limits and infinite summations are defined (with possible value $+\infty$) by monotonicity. The two inequalities follow from the non-negativity of the coefficients α_n . We send $N \rightarrow \infty$ and obtain (4).

This is another show of the strength of the non-negativity which implies monotonicity and hence the existence of limits and infinite sums is guaranteed.

2 Wandering in the graphs G_2 and G_3

In this section we give the precise formulations and proofs of the well-known Pólya's theorems on random walk. In the proofs we will see the precise role of power series (generating functions) and the fact that only weak Abel's theorem is employed rather than the general version of this theorem. We restrict ourselves on the cases $d = 2$ and $d = 3$ respectively for the step from the dimension 2 to the dimension 3 is the step where the quality changes.

2.1 Wandering in G_2

Theorem 2 (wandering in G_2). *Take the origin $o = (0; 0)$ as the initial vertex in $G_2 = (\mathbb{Z}^2; E)$. Then*

$$\lim_{n \rightarrow \infty} \frac{l_n(o, G_2)}{d_n(o, G_2)} = \lim_{n \rightarrow \infty} \frac{l_n}{d_n} = \lim_{n \rightarrow \infty} \frac{l_n}{4^n} = 1. \quad (5)$$

In other words: random walk in G_2 returns to the start point with the probability 1.

Proof. The symbols in (5) have the meaning from (1). As it was shown in the subsection 1.1 the fractions $\frac{l_n}{d_n}$ are the partial sums of the series (3), which is now of the form $\sum_{j=0}^{\infty} \frac{c_j}{4^j}$. It means that we have to prove that

$$\sum_{j=0}^{\infty} \frac{c_j}{4^j} = 1. \quad (6)$$

We will consider these generating functions

$$B(x) = \sum_{n=0}^{\infty} \frac{b_n}{4^n} x^n \text{ and } C(x) = \sum_{n=0}^{\infty} \frac{c_n}{4^n} x^n. \quad (7)$$

If we take the series (7) as formal power series it can be seen by a splitting a walk counted by b_n in its $k + 1$ returns to o into k segments of lengths n_1, n_2, \dots, n_k ; $n_1 + n_2 + \dots + n_k = n$, counted by $c_{n_1}, c_{n_2}, \dots, c_{n_k}$ that there is the relation

$$B(x) = \sum_{k=0}^{\infty} C(x)^k = \frac{1}{1-C(x)}. \quad (8)$$

However, both series in (7) have the radius of convergence at least 1 and hence the generating functions in (7) are the real functions defined certainly for $x \in [0;1)$. To show that (6) holds it suffices to prove with respect to (8) that

$$\lim_{x \rightarrow 1^-} B(x) = +\infty. \quad (9)$$

To prove (9) we use the Proposition 1, i.e. we prove that $\sum_{j=0}^{\infty} \frac{b_j}{4^j} = +\infty$. We do it by computing b_n . It is obvious that $b_n = 0$ for odd n . For even lengths

$$b_{2n} = \sum_{j=0}^n \frac{(2n)!}{j!j!(n-j)!(n-j)!} = \binom{2n}{n} \sum_{j=0}^n \binom{n}{j}^2 = \binom{2n}{n}^2.$$

The first equality follows by considering all positions of j steps of the walk to the right, which force the same number j of steps to the left and the same number $n - j$ steps up and down. The possibilities are counted by the multinomial coefficient $\binom{2n}{j, j, n - j, n - j}$. The last equality follows from the identity $\sum_{j=0}^n \binom{n}{j}^2 = \binom{2n}{n}$.

The Stirling's asymptotic formula yields the asymptotics $\binom{2n}{n} \sim K \frac{4^n}{\sqrt{n}}$ for $n \rightarrow \infty$ and a constant K . Hence we obtain $\frac{b_{2n}}{4^{2n}} \sim K^2 \frac{1}{n}$ which implies

$$\lim_{x \rightarrow 1^-} B(x) = \sum_{j=0}^{\infty} \frac{b_j}{4^j} = \sum_{n=0}^{\infty} \binom{2n}{n}^2 4^{-2n} = +\infty$$

since the harmonic series is divergent. The claim (6) is now the corollary of the Proposition 1:

$$\lim_{x \rightarrow 1^-} C(x) = \sum_{j=0}^{\infty} \frac{c_j}{4^j} = 1.$$

2.2 Wandering in G_3

Theorem 3 (wandering in G_3). *We start wandering at the origin $o = (0,0,0)$ in the graph $G_3 = (\mathbb{Z}^3; E)$. Then*

$$\lim_{n \rightarrow \infty} \frac{l_n(o, G_3)}{d_n(o, G_3)} = \lim_{n \rightarrow \infty} \frac{l_n}{d_n} = \lim_{n \rightarrow \infty} \frac{l_n}{6^n} < 1. \quad (10)$$

In other words: random walk in G_3 returns to the start point with the probability less than 1 and it disappears in infinity without return with a positive probability.

Proof. The symbols in (10) have again the meaning from (1). We consider similarly as in the proof of Theorem 1. The form of generating functions is now $B(x) = \sum_{n=0}^{\infty} \frac{b_n}{6^n} x^n$ and $C(x) = \sum_{n=0}^{\infty} \frac{c_n}{6^n} x^n$. Since the relation (8) holds it is sufficient to show that

$$\lim_{x \rightarrow 1^-} B(x) = \sum_{j=0}^{\infty} \frac{b_j}{6^j} < +\infty. \quad (11)$$

The equality in (11) follows again from Proposition 1. We have clearly $b_n = 0$ for odd n . For even members we find an upper bound:

$$\begin{aligned} \frac{b_{2n}}{6^{2n}} &= \frac{1}{6^{2n}} \sum_{\substack{j+k \leq n \\ j, k \in \mathbb{N}_0}} \frac{(2n)!}{j!j!k!(n-j-k)!(n-j-k)!} = \binom{2n}{n} 4^{-n} \sum_{\substack{j+k \leq n \\ j, k \in \mathbb{N}_0}} \left(3^{-n} \binom{n}{j, k, n-j-k} \right)^2 \leq \\ &\leq \binom{2n}{n} 4^{-n} \max_{\substack{x, y, z \in \mathbb{N}_0 \\ x+y+z=n}} 3^{-n} \binom{n}{x, y, z} = \binom{2n}{n} 12^{-n} \binom{n}{x_0, y_0, z_0}, \end{aligned} \quad (12)$$

where $(x_0, y_0, z_0) = \begin{cases} (m, m, m), & n = 3m \\ (m+1, m, m), & n = 3m+1 \\ (m+1, m+1, m), & n = 3m+2 \end{cases}, m \in \mathbb{N}_0$. In the first expression in (12) we counted as in

the proof of Theorem 1, j is the number of steps in the walk to the right, k is the number of steps up, and $n - j - k$ the number of steps back. The second expression is an algebraic rearrangement. In the third expression we used the fact that if $\alpha_1, \alpha_2, \dots, \alpha_p$ are non-negative real numbers satisfying $\sum_{i=1}^p \alpha_i = 1$, then $\sum_{i=1}^p \alpha_i^2 \leq \max_{1 \leq i \leq p} \alpha_i$. In our case we have $\alpha_i = 3^{-n} \binom{n}{j, k, n-k}$ and $3^n = (1+1+1)^n = \sum_{\substack{j+k \leq n \\ j, k \in \mathbb{N}_0}} \binom{n}{j, k, n-j-k}$. In the end we found the maximum value of the trinomial coefficients by the fact that $p!q! > (p-1)!(q+1)!$ for $p \geq q+2$.

By the Stirling's formula for the factorial we have estimates with constants K, L

$$\binom{2n}{n} < K \frac{4^n}{\sqrt{n}}, \quad \binom{n}{x_0, y_0, z_0} < L \frac{3^n}{n}. \quad (13)$$

Using the inequalities in (13) we obtain the final estimate for the even members of the series in (11)

$$\frac{b_{2n}}{6^{2n}} < K \frac{1}{\sqrt{n}} L \frac{1}{n} = C n^{-\frac{3}{2}}.$$

The inequality in (11) follows from the fact that

$$\sum_{j=0}^{\infty} \frac{b_j}{6^j} = \sum_{n=0}^{\infty} \frac{b_{2n}}{6^{2n}} < C \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} < +\infty.$$

3 Estimate for wandering in two dimensions based on more effective form of Proposition 1

The formulation of the **Theorem 2** (wandering in G_2) does not provide a more detailed look at the convergence of fractions in (5). It says nothing about the fraction $l_n/4^n$ for concrete n . To remedy this deficiency we formulate a proposition which contains an effective lower bound for the sequence in (5). But first we have to focus on the Proposition 1 which in the present form is not sufficient for our purpose.

3.1 Some estimate for the power series with non-negative terms

Suppose that we have two power series $B(x) = \sum_{n=0}^{\infty} \beta_n x^n$ and $C(x) = \sum_{n=0}^{\infty} \gamma_n x^n$, $\gamma_0 = 0$, $\gamma_n \leq 1$, with non-negative coefficients satisfying the relation (8) as the formal power series, convergent for any $x \in [0; 1)$. Further suppose that $\beta_{2n} > c/n$ for any $n \in \mathbb{N}$ and a positive constant c . It implies that $B(1) := \sum_{n=0}^{\infty} \beta_n = +\infty$. By Proposition 1 $\lim_{x \rightarrow 1^-} B(x) = +\infty$ and thus $\lim_{x \rightarrow 1^-} C(x) = 1$ by (8). Hence $C(1) := \sum_{n=0}^{\infty} \gamma_n = 1$ again by Proposition 1.

We have also for any $x \in (0; 1)$ the inequality

$$B(x) = \sum_{n=0}^{+\infty} \beta_n x^n > c \sum_{n=0}^{+\infty} \frac{1}{n} x^{2n} = c \ln \frac{1}{1-x^2}.$$

Finally for any $N \in \mathbb{N}$ and $x \in (1/2; 1)$, with respect to the assumptions $\gamma_n \leq 1$ and $C(x) = 1 - \frac{1}{B(x)}$ we get

$$\begin{aligned} \sum_{n=0}^N \gamma_n &> \sum_{n=0}^N \gamma_n x^n = \sum_{n=0}^{+\infty} \gamma_n x^n - \sum_{n=N+1}^{+\infty} \gamma_n x^n \geq 1 - \frac{1}{B(x)} - \frac{x^{N+1}}{1-x} > \\ &> 1 + \frac{1}{c \ln(1-x^2)} - \frac{x^{N+1}}{1-x} > 1 - \frac{1}{c \ln \frac{1}{1-x} - c \ln 2} - \frac{x^{N+1}}{1-x}. \end{aligned} \quad (14)$$

3.2 The lower estimate of the speed of the convergence for the wandering in G_2

We use the estimate (14) to obtain a lower bound for the sequence of fractions in **Theorem 2** (wandering in G_2).

Proposition 2. For any $N \in \mathbb{N}, N > 2$,

$$1 > \frac{l_N(o, G_2)}{4^N} > 1 - \frac{1}{0.2 \ln N - 0.16} - \frac{N^{8/9}}{\exp N^{1/9}}. \quad (15)$$

Proof. The symbols l_n, b_n, c_n are defined in the subsection 1.1. We set $\gamma_n = \frac{c_n}{4^n}, \beta_n = \frac{b_n}{4^n}$ in the previous subsection. All the considerations in the previous section work and the function B and C are exactly those in (7). We deduce the required lower bound on β_n from the estimates

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n \exp \frac{1}{12n+1} < n! < \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \exp \frac{1}{12n}$$

which is due to H. Robins [9] and hold for any $n \in \mathbb{N}$. Hence we obtain the estimate

$$\beta_{2n} = 4^{-2n} \frac{(2n)!^2}{n!^4} > \frac{\exp \frac{2}{24n+1}}{\pi n \exp \frac{1}{3n}} > \frac{1}{\pi \exp \frac{1}{3}}$$

We round the constant $\frac{1}{\pi \exp \frac{1}{3}}$ and get $\beta_{2n} > \frac{0.228}{n}$. If we set $x = x(N) = 1 - N^{-\frac{8}{9}} \in [0;1)$, $N \in \mathbb{N}$, then for $N > 2$ we obtain $x \in (\frac{1}{2}; 1)$, $\ln \frac{1}{1-x} = \frac{8}{9} \ln N$ and with $c = 0.228$ we get $c \ln \frac{1}{1-x} = \frac{8}{9} c \ln N > 0.2 \ln N$ and $c \ln 2 < 0.16$. For the last fraction in (14) we obtain the inequality $\frac{x^{N+1}}{1-x} < \frac{N^{8/9}}{\exp N^{1/9}}$. In this way the required inequality (15) follows from the inequality (14).

3.3 A wanderer on Manhattan

How long do we have to wait on Times Square for an aimless wanderer (starting at the square) to achieve the return probability more than 90%? The net of streets and avenues is not exactly square but rectangular grid. However, we replace it by the square grid with side length 200 meters. We assume that the wanderer can walk this distance in 2 minutes.

Proposition 3. *Under the assumption above, the aimless wanderer returns to Times Square with probability more than 90% after*

$$\frac{e^{[51]}}{720} < 5.363 * 10^{16} \text{ years.}$$

Proof. We set $N = [e^{51}]$. Then the first subtracted term in (15) is less than $(0.2 * 51 - 0.16)^{-1} < 0.0997$ and the second one is quite negligible, less than e^{-90} . Thus $\frac{1}{4^N} > 0.9$.

References

- [1] Abel, N. H. (1826): Untersuchungen über die Reihe: $1 + mx + \frac{m(m-1)}{1.2} x^2 + \frac{m(m-1)(m-2)}{1.2.3} x^3 + \dots$, Journal für die reine und angewandte Mathematik 1, 311-339.
- [2] Bilingsley, P. (1995). *Probability and measure*. New York: John Wiley & Sons Inc.
- [3] Dhrymes, P.J. (1980): *Distributed Lags. Problems of Estimation and Formulation*. North Holland, Amsterdam.
- [4] Feller, W. (1968): *An Introduction to Probability Theory and Its Applications, Volume I*, John Wiley & Sons, New York, 3rd edition.
- [5] Lange, K. (2015): Pólya's random walks theorem revisited, *Amer. Math. Monthly* 122, 1005-1007.
- [6] Levin, D.A. and Peres, Y. (2010): Pólya's theorem on random walks via Pólya's urn, *Amer. Math. Monthly* 117, 220-231.
- [7] Novak J. (2014): Pólya's random walk theorem, *Amer. Math. Monthly* 121, 711-716.
- [8] Pólya, G. (1921): Über eine Aufgabe der Wahrscheinlichkeitsrechnung betre_end die Irrfahrt im Strassen-netz, *Math. Ann.* 84 (1921), 149-160.
- [9] Robins, H. (1955): A remark on Stirling's formula, *Amer. Math. Monthly* 6, 26-29.
- [10] Rogers, K. (2017): Transient and recurrent random walks on integer lattices. 10 pages

Numerical Valuation of the Investment Project with Expansion Options Based on the PDE Approach

Jiří Hozman¹, Tomáš Tichý²

Abstract. Compared to the standard DCF methodology, the real options approach provides a solution to optimal investment decisions that captures the value of flexibilities embedded in a project. In this paper we focus on one specific kind of investment decisions — an option to expand.

Assuming values of both the project and the embedded option are determined in terms of time and underlying output price, driven by a relevant stochastic process, one can unify the PDE approach to describe the development of values of the project and options. More precisely, the link is realized through a payoff function enforced at a fixed time. As a result, we obtain a system of relevant governing equations of the Black-Scholes type.

Since explicit formulae are known for this type of PDE problem only in specific cases, one must turn to some approximation methods. With reference to the results obtained in valuing financial options, we apply the discontinuous Galerkin method to solve the relevant governing equations. The obtained numerical scheme is applied to a simple illustrative expansion decision problem.

Keywords: real options valuation, project value, option to expand, Black-Scholes equation; discontinuous Galerkin method, numerical solution

JEL Classification: C44, G13

AMS Classification: 65M60, 35Q91, 91G60

1 Introduction

Capital budgeting is an essential discipline of corporate finance and basically can be viewed as the planning process that aims to increase the value of the firm/project by proper investment decisions within a long time horizon. Hence, valuing the profitability of investment projects plays a particularly important role here.

More than four decades ago, the cornerstone of modern investment theory was built, linking valuation of corporate investment opportunities as pricing of financial options on real assets, see the pioneering paper by Myers [12]. Due to the analogy with an option on financial asset, the methodology has become known as real options approach that interprets the flexibility value as the option premium. Since then, a large number of various solution techniques have been developed [11], from a simulation approach, over dynamic programming to contingent claims analysis, which compares the change in option/project values with the change in the value of a suitably constructed portfolio of trading assets within the relevant partial differential equation (PDE), see [2].

In this short contribution we extend our previous results on the topic of numerical pricing of financial option contracts using discontinuous Galerkin (DG) method, see [7], [8] and [9], to valuation of implicit flexibility in the investment project. We proceed as follows – in Section 2 the relevant PDE model is formulated, while in Section 3 a numerical valuation scheme is presented. Finally, in Section 4 a simple numerical experiment related to reference data is provided.

2 PDE Model for Valuation of the Embedded Option

In this paper we concentrate on valuing the flexibility of an investment project, i.e. the real option value embedded in that project. At first it is necessary to describe the value of the project itself and then we are able to find the value of its flexibility by solving the relevant PDEs that link both option and project values, see inspiring ideas in [10]. Let us consider that fluctuations in project values (and also in real option prices) are tracked back to uncertainty via the underlying output price P that evolves in time t according to the following stochastic differential equation (proposed in [3]):

$$dP(t) = (r - \delta)P(t)dt + \sigma P(t)dW(t), \quad P(0) > 0, \quad (1)$$

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

where $r > 0$ is the risk-free interest rate, $\delta > 0$ is the mean convenience yield on holding one unit of the output, $W(t)$ is a standard Brownian motion and $\sigma > 0$ is the volatility of the output price.

Let $V_0(P, t)$ and $V_1(P, t)$, for current price P and time t , denote the value of the project having no options and with the embedded option allowing the particular action (e.g., expansion), respectively. We also assume that the possibility of a single decision-making is related to a prespecified time T only, i.e., a real option is held in isolation by a single firm and it is of a European-style. Let $T^* > T$ be the maximal life-time of both projects and $\varphi_0(P, t)$ and $\varphi_1(P, t)$ represent (after-tax) cash flow of the corresponding project (i.e., without and with the embedded option). Intuitively, from the definitions above we expect for all $P \geq 0$ that $V_1(P, t) \geq V_0(P, t)$, if $t \in [0, T)$; $V_1(P, T^*) = V_0(P, T^*) = 0$ and $\varphi_1(P, t) = \varphi_0(P, t)$, if $t \in [0, T)$.

In line with the contingent claims framework [1], namely assuming no arbitrage opportunities for trading in the real and financial assets, and using the delta-hedging techniques, it can be shown that value functions V_0 and V_1 at times $t \in [T, T^*)$, i.e., between the change of operating and the project life-time, are characterized in general as solutions of a couple of deterministic backward PDEs (see [4]):

$$\frac{\partial V_i}{\partial t} + \underbrace{\frac{1}{2}\sigma^2 P^2 \frac{\partial^2 V_i}{\partial P^2} + (r - \delta)P \frac{\partial V_i}{\partial P} - rV_i}_{\mathcal{L}_{BS}(V_i)} = -\varphi_i, \quad 0 < P < +\infty, \quad T \leq t < T^*, \quad i = 0, 1, \quad (2)$$

with the terminal states $V_i(P, T^*) = 0, P > 0, i = 0, 1$.

Further, for all time instants $t \in [0, T)$, the project value $V_1(P, t)$ is equal to the project value $V_0(P, t)$ increased by the true value of the flexibility of the given investment opportunity, see [14]. In view of this fact, it is possible to track values of both projects and the embedded option value simultaneously within one timeline on $[0, T)$, that are linked through the function $\Pi(V_0, V_1)$ enforced at the expiry date T . More precisely, let $F(P, t)$ denote the value of the flexibility at the current price P and actual time $t \in [0, T)$, then the value function F satisfies the following PDE (see [4]):

$$\frac{\partial F}{\partial t} + \mathcal{L}_{BS}(F) = 0, \quad 0 < P < +\infty, \quad 0 \leq t < T, \quad \text{with } F(P, T) = \Pi(V_0(P, T), V_1(P, T)), \quad (3)$$

where \mathcal{L}_{BS} is the second order linear differential operator of the Black-Scholes (BS) type defined in (2) and Π plays the role of a payoff function, the specific form of which depends on the type of flexibility provided by the particular real option, see Section 4 for the case of an option to expand.

In summary, determining the value of flexibility of an investment project at the present time $t = 0$ consists of two consecutive problems. First, we solve a pair of PDEs (2) with homogeneous terminal conditions to obtain the project values at $t = T$ that we use in the construction of the terminal value of the embedded flexibility at $t = T$. Consequently, we solve the problem (3) to obtain the present value of flexibility.

3 Numerical Valuation

Since analytical formulae for the BS type PDEs are available only either in the simplest cases or under very strong limitations, an application of modern numerical methods takes a crucial part for real options valuation. In our study, we employ the DG method, successfully used also in the field of financial options pricing (see, e.g., [8] and [9]), to improve the numerical valuation process. We proceed as follows. At first, we localize the governing equations to a bounded spatial domain and discuss the choice of suitable boundary conditions. Next, we apply the standard discretization steps and present the resulting numerical scheme.

3.1 Localization and boundary conditions

The proposed valuation methodology, related to numerical solving of terminal-value problems (2) and (3), requires their localization to a bounded interval $\Omega = (0, P_{\max})$, where $P_{\max} \gg 0$ is the maximal sufficient value of the underlying output price. Formally, we restrict the governing equations and the relevant terminal conditions to the bounded domain Ω . Therefore, we have to impose the project values at both endpoints $P = 0$ and $P = P_{\max}$. As the prescribed values the estimations based on the net present value approach are performed, see [10]. Since the cash flows of the projects are $\varphi_i, i = 0, 1$, then the present values of the projects $V_i, i = 0, 1$, at endpoints of Ω are defined as

$$V_i(0, t) = \int_t^{T^*} \varphi_i(0, \xi) e^{-r(\xi-t)} d\xi, \quad V_i(P_{\max}, t) = \int_t^{T^*} \varphi_i(P_{\max}, \xi) e^{-r(\xi-t)} d\xi, \quad t \in [T, T^*), \quad i = 0, 1. \quad (4)$$

In the case of flexibility (real option) values, the choice of boundary conditions is a more delicate issue, which in principle depends on the type of flexibility. For the embedded expansion option exercisable at $t = T$, it is clear that its exercising add negative value to the investment when $P = 0$ and thus the flexibility has zero value on the left boundary of Ω . Furthermore, to ensure the compatibility of the flexibility function F with the payoff function Π also at the right boundary point $P = P_{\max}$, we prescribe here the discounted values of the payoff function. In accordance with the above, we prescribe Dirichlet boundary conditions in the form

$$F(0, t) = 0, \quad F(P_{\max}, t) = e^{-r(T-t)} \Pi(V_0(P_{\max}, t), V_1(P_{\max}, t)), \quad t \in [0, T]. \quad (5)$$

Finally, let us mention that governing equations (2) and (3) are closely related to the class of convection-diffusion equations, which exhibits a hyperbolic and parabolic behaviour in dependency on a convection/diffusion ratio. Therefore, the proposed numerical schemes for solving such problems have to take these properties into account.

3.2 Discontinuous Galerkin scheme

We recall a numerical scheme based on the DG method (see [13] for a complete overview) applied to European vanilla option pricing in [7] and modify it to the numerical valuation of the real options considered. The approximate solutions representing project as well as real option values are sought in the finite dimensional space S_h^p consisting from piecewise polynomial, generally discontinuous, functions of the p -th order defined over the partition of the domain Ω with the assigned mesh size h .

Similarly as in [7], the semi-discrete functions $u_h^{(i)} = u_h^{(i)}(t) \in S_h^p, i = 0, 1$, related to the values of the projects $V_i, i = 0, 1$, are represented by two systems of the ordinary differential equations (ODEs), i.e.,

$$\frac{d}{dt} (u_h^{(i)}, v_h) + \mathcal{A}_h (u_h^{(i)}, v_h) = \ell_h^{(i)}(v_h)(t) - (\varphi_i(t), v_h) \quad \forall v_h \in S_h^p, \forall t \in [T, T^*), \quad i = 0, 1, \quad (6)$$

with the terminal states $u_h^{(i)}(T^*) = 0, i = 0, 1$. The notation (\cdot, \cdot) denotes the inner product in $L^2(\Omega)$ and the bilinear form $\mathcal{A}_h(\cdot, \cdot)$ stands for the DG semi-discrete variant of the spatial partial differential operator \mathcal{L}_{BS} from (2) accompanied with penalties and stabilizations. Further, the linear form $\ell_h^{(i)}(\cdot)(t)$ contains terms arising from boundary conditions (4) corresponding to the particular project value V_i . For the detailed derivation of the above-mentioned forms we refer the interested reader to [7].

Next, our aim is to construct the solution $w_h = w_h(t) \in S_h^p$ related to the value of the flexibility function F . In the same manner, the DG discretization in spatial coordinates leads to a system of the ODEs for unknown function w_h , i.e.,

$$\frac{d}{dt} (w_h, v_h) + \mathcal{A}_h (w_h, v_h) = \ell_h(v_h)(t) \quad \forall v_h \in S_h^p, \forall t \in [0, T], \quad (7)$$

where the terminal state $w_h(T)$ is given by the payoff function Π , the bilinear form \mathcal{A}_h remains the same and the linear form $\ell_h(\cdot)(t)$ contains terms arising from boundary conditions (5).

Consequently, we realize the discretization of (6) and (7) in time by an implicit Euler scheme for the equidistant time partition $T^* = t_0 > t_1 > \dots > t_R = T > t_{R+1} > \dots > t_M = 0$ with the time step $\tau = T^*/M$. Denote $u_{h,m}^{(i)} \in S_h^p, i = 0, 1$, the approximation of the solutions $u_h^{(i)}(t), i = 0, 1$, from (6) at time level $t_m \in [T, T^*], m = 0, \dots, R$. Similarly, we define the DG approximate solution of problem (7) as functions $w_h^m \approx w_h(t_m), t_m \in [0, T], m = R, \dots, M$. Let $u_{h,0}^{(0)} = u_{h,0}^{(1)} = 0$ be the initial states, then $w_h^M \approx w_h(0)$ is computed in the following three steps (note that we use the backward time running, i.e., $t_{m+1} - t_m = -\tau$):

$$\begin{aligned} (u_{h,m+1}^{(i)}, v_h) - \tau \mathcal{A}_h (u_{h,m+1}^{(i)}, v_h) &= (u_{h,m}^{(i)}, v_h) - \tau \ell_h^{(i)}(v_h)(t_{m+1}) + \tau (\varphi_i(t_{m+1}), v_h) \\ &\forall v_h \in S_h^p, \quad m = 0, 1, \dots, R-1, \quad i = 0, 1, \end{aligned} \quad (8)$$

$$(w_h^R, v_h) = (\Pi(u_{h,R}^{(0)}, u_{h,R}^{(1)}), v_h) \quad \forall v_h \in S_h^p, \quad (9)$$

$$(w_h^{m+1}, v_h) - \tau \mathcal{A}_h (w_h^{m+1}, v_h) = (w_h^m, v_h) - \tau \ell_h(v_h)(t_{m+1}) \quad \forall v_h \in S_h^p, \quad m = R, \dots, M-1, \quad (10)$$

where the starting data (at expiration date) w_h^R are given as S_h^p -approximation of the payoff function Π depending on states $u_{h,R}^{(i)}, i = 0, 1$, see (9). Finally, note that the equations (8) and (10) result into a sequence of systems of linear algebraic equations with sparse matrices that uniquely determine the relevant solutions on the corresponding time levels, see [8].

4 Numerical Experiment: Option to Expand

In this section, to briefly illustrate capabilities of the numerical scheme introduced above, we present numerical experiments on valuing an option to expand an investment project in the mining industry. The proposed valuation procedure is implemented in the solver Freefem++, incorporating GMRES as a solver for non-symmetric sparse systems, for more details, see [6].

As in [10] we consider iron ore mine, value of which is given by the project value $V_0(P, t)$. Concurrently, we consider the mining company adopting the embedded option $F(P, t)$ for investment in expansion in the mining project $V_1(P, t)$ for $t \in [0, T)$. This expansion option is exercisable at T and requires the amount $\mathcal{K} > 0$ for expansion. Thus the value of $F(P, T)$ is positive when $V_1(P, T) > V_0(P, T) + \mathcal{K}$ and otherwise has zero value. In line with this the payoff function corresponds to the European vanilla call option $F = V_1 - V_0$ with strike \mathcal{K} and it is defined as follows

$$\Pi(V_0(P, T), V_1(P, T)) = \max(V_1(P, T) - V_0(P, T) - \mathcal{K}, 0), \quad P \in (0, P_{\max}). \quad (11)$$

Further, let Q denotes the total reserve of the iron ore mine in thousands of million dry metric tonnes (dmt) and $q_i(t)$, $i = 0, 1$, be the iron ore production rates (in thousands of million dmt per year) associated with projects V_i , $i = 0, 1$. The life-times T_0^* and T_1^* (in years) of both projects are defined as the mine is operated at a rate q_i and satisfy the relationship

$$Q = \int_0^{T_0^*} q_0(\xi) d\xi = \int_0^{T_1^*} q_1(\xi) d\xi. \quad (12)$$

Obviously, for the expansion options, we expect $T^* = T_0^* > T_1^* \gg T$ and assume that production rates satisfy

$$q_1(t) = \begin{cases} q_0(t), & \text{if } t \in [0, T), \\ \kappa \cdot q_0(t), & \text{if } t \in [T, T_1^*), \\ 0, & \text{if } t \in [T_1^*, T^*], \end{cases} \quad (13)$$

where the factor $\kappa > 1$ represents the expansion rate. Then the after-tax cash flow (related to the relevant project) is given according to [5] by

$$\varphi_i(P, t) = q_i(t) \left((1 - D)P - c(t) \right) (1 - B), \quad P \in [0, P_{\max}], \quad t \in [0, T^*], \quad i = 0, 1, \quad (14)$$

where $c(t)$ is the average cash cost rate of iron ore production per dmt, D is the rate of state royalties and B the income tax rate.

The numerical experiments are performed on the reference data from [10], where the expansion options are evaluated using an upwind finite difference scheme. In all cases, we consider the following project and market data:

$$\begin{aligned} Q = 10, \quad q_0(t) &= \frac{1}{10} e^{0.007t}, \quad D = 0.05, \quad B = 0.30, \quad c(t) = 35e^{0.005t}, \\ T = 2, \quad \mathcal{K} &= 10(\kappa - 1), \quad r = 0.06, \quad \delta = 0.02, \end{aligned} \quad (15)$$

which are the representatives of parameter values of practical significance. More precisely, the value of \mathcal{K} is given in thousands of million USD and $c(0) = 35$ USD per dmt is based on prices from 01/2007.

For the purpose of a broader comparison of the obtained results with the reference ones from [10], we set $P_{\max} = 100$ USD and compute the piecewise linear solutions on the fixed uniformly partitioned grid with the unit mesh size, i.e., $p = 1$ and $h = 1$. Using (12) and (15), easy calculation leads to $T^* \doteq 75.8$ and we determine T_1^* in a similar way for (13) and various factors κ . The time step is set in consistency with [10] as $\tau = 0.02$.

First, we investigate the behaviour of the option values with the fixed volatility $\sigma = 0.3$ and for various expansion rates $\kappa \in \{2, 3, 4\}$. Figure 1 captures the development of the DG approximations of option values (in thousands of million USD) in the whole space-time domain $\Omega \times (0, T)$ for the particular scenario. One can easily observe that the surface plot is similar to the conventional financial European call option with the relevant Black-Scholes model parameters. Also the payoff function is well resolved in Figure 1 at $t = 2$. Figure 2 (left) records flexibility values at present time $t = 0$ for all three scenarios together with the comparative results from [10], evaluated at three underlying reference prices $P_{\text{ref}} \in \{40, 50, 60\}$. Without surprise, one can easily observe that piecewise linear DG approximations match well the reference values and give fairly the same results as the upwind finite difference methods. More precisely, Figure 2 (left) illustrates that the value of flexibility F is an increasing function of κ .

Secondly, we study the influence of volatility σ to the option values under the fixed expansion rate $\kappa = 2$. The flexibility values at present time $t = 0$ for $\sigma \in \{0.1, 0.35, 0.7\}$ and reference values at $P_{\text{ref}} = 40$ are depicted in Figure 2 (right). According to these graphs, we can deduce that flexibility values increase (for all $P \in \Omega$) with higher values of σ for high volatile cases, i.e., for $\sigma \in (0.35, 0.7)$. On the other hand, for less volatile scenarios, $\sigma \in (0.1, 0.35)$ this rule holds only when the output price P is smaller than some critical values. In the remaining region, where the output price P is greater than some critical values the situation is quite opposite, i.e., the flexibility values decrease with respect to increasing σ . From this point of view, we come to the same conclusions as in the paper [10] and thus the presented DG approach shows one part of its promising potential.

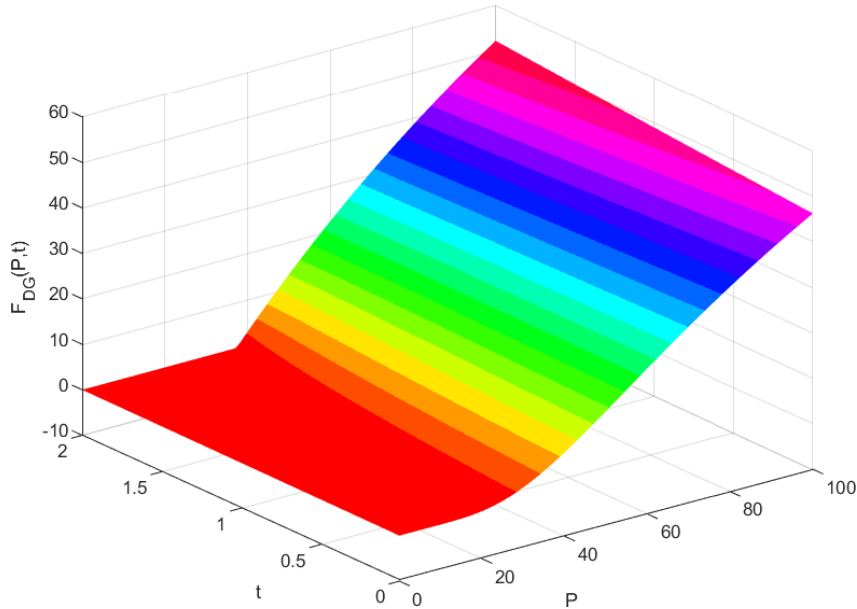


Figure 1 Surface plot of piecewise linear approximations of embedded option values in the space-time domain $\Omega \times (0, T)$ for $\kappa = 2$ and $\sigma = 0.3$

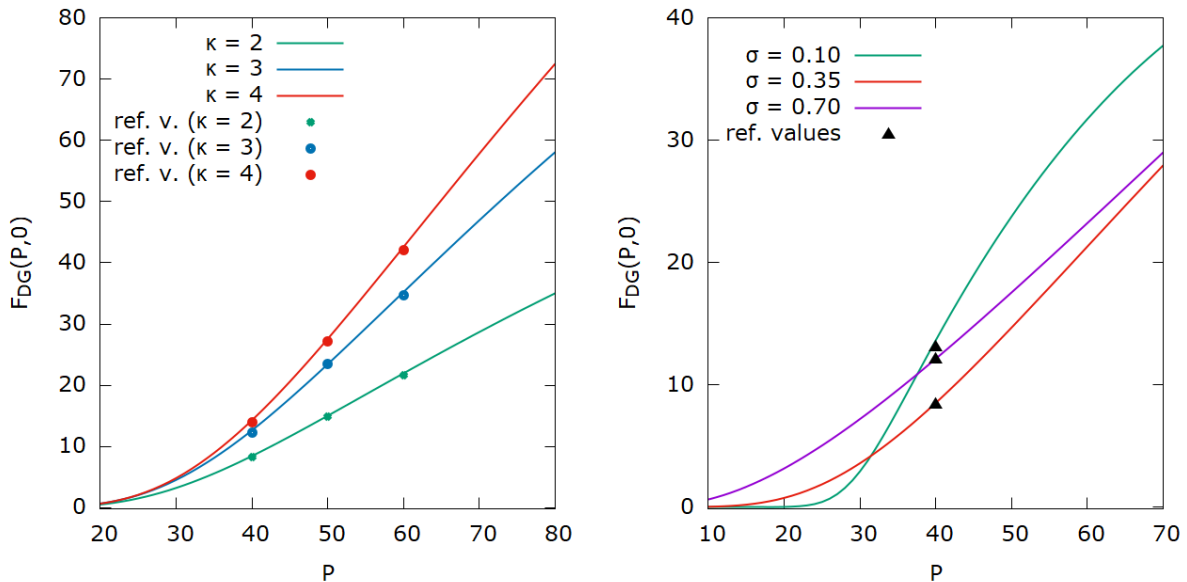


Figure 2 Approximate embedded option values (in thousands of million USD) at $t = 0$ for different scenarios with fixed $\sigma = 0.3$ (left) and with fixed $\kappa = 2$ (right)

5 Conclusion

Pricing of real options is very challenging and no less important part of corporate finance. In this paper we have recalled PDE models to valuation of investment projects and embedded flexibility to expansion, and presented numerical scheme based on the DG method to solve them. The presented numerical experiments, arising from practical issues in the iron ore mining industry, evaluate the European-style expansion option under various scenarios given by the parameters of expansion rate and volatility. As a result, they enable to clarify the possible investment decision to expand in the mining project that takes into account fluctuations in output prices as well as different rates of expansion. The results obtained provide very good similarity to reference ones and the future work should be addressed to extend the DG approach to other common types of real options such as the option to defer, the option to abandon, the option to switch, etc. Last but not least the key benefit of this approach lies in easier incorporation of the American-style nature of real options (i.e. exercising at any time before the final maturity day).

Acknowledgements

The first author acknowledges support provided within the project No. PURE-2020-4003 financed by TU Liberec. The second author was supported through SP2021/15, an SGS research project of VSB-TU Ostrava.

References

- [1] Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81, 637–659.
- [2] Brennan, M.J. & Schwartz E.S. (1977). The Valuation of American Put Options. *Journal of Finance*, 32, 449–462.
- [3] Cortazar, G. & Schwartz, E.S. & Casassus, J. (2001). Optimal exploration investments under price and geological-technical uncertainty: a real options model. *R&D Management*, 31, 181–189.
- [4] Dixit, A. & Pindyck, R. (1994). *Investment Under Uncertainty*. Princeton: Princeton University Press.
- [5] Haque, M. & Topal, E. & Lilford, E. (2014). A numerical study for a mining project using real options valuation under commodity price uncertainty. *Resources Policy*, 39, 115–123.
- [6] Hecht, F. (2012). New development in FreeFem++. *Journal of Numerical Mathematics*, 20, 251–265.
- [7] Hozman, J., et al. (2018). *Robust Numerical Schemes for Pricing of Selected Options*. SAEI, Vol. 57., Ostrava: VŠB-TU Ostrava.
- [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*, 43, 7726–7746.
- [10] Li, N. & Wang, S. (2019). Pricing options on investment project expansions under commodity price uncertainty. *Journal of Industrial & Management Optimization*, 15, 261–273.
- [11] Mun, J. (2002). *Real Options Analysis: Tools and Techniques for Valuing Strategic Investments and Decisions*. John Wiley & Sons, Inc., Hoboken.
- [12] Myers, S.C. (1977). Determinants of corporate borrowing. *Journal of Financial Economics*, 5, 147–175.
- [13] Rivière, B. (2008). *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*. SIAM, Philadelphia.
- [14] Trigeorgis, L. (1993). Real options and interactions with financial flexibility. *Financial Management*, 22, 202–224.

Housing Submarkets: The case of the Prague Housing Markets

Petr Hrobař¹, Vladimír Holý²

Abstract. In this study, we identify and investigate the housing submarkets in the Prague flat market. We overview and utilize the model frameworks for doing so. In order to analyze the variability of price determinants in space as well as to model the spatial heterogeneity, the Geographically Weighted regression model is used. Then, using data dimensionality reduction techniques combined with the clustering techniques, the housing submarkets are identified and interpolated. Lastly, the higher level of homogeneity within each submarket clusters is investigated and analyzed. We confirm that the housing submarkets exist in the Prague flat estate market.

Keywords: Spatial Statistics, Spatial Data, Real Estate, GWR, Submarkets

JEL Classification: C33, C38, R31

AMS Classification: 90C15

1 Introduction

The real estate sector, particularly in Prague, disposes of a fairly high level of heterogeneity. It is known that the real estate sector is of spatial nature. Thus, empirical frameworks must take into consideration both of these phenomena. The spatial nature of the data combined with certain levels of heterogeneity. The spatial clusters, the housing submarkets, need to be identified and analyzed separately to model the heterogeneity. To determine each individual housing submarkets, within the area of Prague, the spatial statistic frameworks, combined with the statistical learning techniques, are utilized.

The main inspiration for our study is the research of [4], where authors identify and analyze the housing submarkets and common price determinants effects therein in the housing market of Warsaw (Poland). As far as an analysis of the housing market of Prague goes, a few interesting studies have been conducted over the past years. Namely, the study [5] from 2016, where the author identifies the "living premium" locations by interpolating the residuals from the linear regression model, assuming that the residuals are an estimate of the unobserved compound effects of the location. We follow both of the mentioned studies as well as our own contribution [3] and identify the housing submarkets over the area of the Prague housing market using only the flat estates.

The entire study is structured as follows: In Section 2, the methodological frameworks and steps used for the submarkets identification are presented, in Section 3, we introduce and describe the dataset used for modeling purposes. Section 4 then uses all methodology frameworks described and identifies the housing submarkets in the flat estate market in Prague. Lastly, Section 5 concludes our study.

2 Methodology and Frameworks

This section summarises the utilized frameworks for identifying the housing submarkets. We provide a fundamental description of the used methods and corresponding literature. Firstly, the Geographically Weighted Regression (GWR for short) is outlined. As a next step, we proceed with describing the submarket identification workflow and all techniques required for such a step.

2.1 Geographically Weighted Regression

The GWR model is an adaptation and modification of the weighted regression model. An idea is to estimate not one global model but rather a series of smaller local models. We estimate the value at a given point in space for one particular spatial unit (in our case, one real estate) based only on its immediate neighborhood units. To identify the closest neighbors, the distance between them is measured. Therefore, the coordinates for each spatial unit are required. The general GWR model, using the example of [4], can be written as follows:

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

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)x_{ik} + \varepsilon_i ; i = 1, 2, \dots, n, \quad (1)$$

where β_0 is the intercept, (u_i, v_i) are the coordinates of i -th spatial unit. $\beta_k(u_i, v_i)$ is the k -th regression parameter (it is a set of numbers) and ε_i is the random error term.

The hedonic price model, which is described in section 3, is then estimated using the GWR model. This allows analyzing the variability of common price determinants over the space. By the definition of GWR, each individual spatial unit i obtains its own estimate of the model coefficients, and therefore the GWR models the spatial heterogeneity of the data simultaneously [4]. In order to obtain the proper neighborhood structure for the local regressions, the Gaussian distance-decaying kernel is used.

The Gaussian Kernel function can be expressed as [4]:

$$K(z) = \exp(-z^2/2), \quad (2)$$

where z is constructed as:

$$z = \begin{cases} (u_i - v_i)/h, & \text{for } u_i - v_i \leq h \\ 0, & \text{for } u_i - v_i > h, \end{cases}$$

where h denotes the selected bandwidth size, which was selected by the LOOCV criterion (Leave-one-out-cross-validation) [4]. Once the proper bandwidth size for each spatial unit (thus we are using adaptive bandwidth size) is obtained and the Gaussian Kernel is used, it yields the final \mathbf{W} matrix. This \mathbf{W} matrix is a matrix that captures the proper observation points to be used in the kernel function and gives them proper weight by pasting its distance to the reference points into the kernel function.

Therefore, the GWR estimation is equivalent to the weighted regression estimation, which can be, in the matrix form, represented as follows [7]:

$$\hat{\beta}_i(u_i, v_i) = [\mathbf{X}'\mathbf{W}(u_i, v_i)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}(u_i, v_i)y, \quad (3)$$

where \mathbf{X} is a design matrix of the regression and the y is a vector of the dependent variable. For more technical details of the estimation techniques and details regarding the estimation of the covariance matrix and associated models diagnostics refer to [4] and [7].

A unique aspect of the GWR model is the fact that the model yields different coefficient parameter values for each observation e.g. the row of the design matrix of the regression \mathbf{X} and thus the spatial units of more equivalent coefficients can be clustered together. This essentially summarises the concept of the housing submarkets, which are more deeply discussed in the following section.

2.2 The Housing Submarkets

In our modeling frameworks, we define the housing submarkets as the cluster within which the real estates dispose of a considerably higher level of homogeneity. Many recent studies seem to agree that the housing submarkets can be identified by clustering the coefficients of the GWR model [4]. As far as the clustering of the coefficients goes, the k -means algorithm [2] is commonly used. Before that, however, as we are working with quite a large number of housing features, some form of dimensionality reduction is required. This reduction is performed by the famous Principal Component Analysis (PCA) estimated via the Singular Value Decomposition (SVD), see [2].

The complete analytical framework of constructing the housing submarkets is summarized and illustrated in Figure 1. Firstly, the design matrix of the regression \mathbf{X} is constructed (Including the selection of proper model form as well as dummy coding of categorical variables). Secondly, the parameters regarding the GWR model are determined (the bandwidth sizes) and the GWR model is estimated. Then the GWR coefficient matrix is scaled using the z -standardization, and the SVD of the matrix obtained. For the interpretation of the main sources of variability between the housing submarkets, the first four columns of the \mathbf{V} matrix from the SVD can be inspected, which is not done in this study. Then, the \mathbf{U} matrix from the SVD is taken, and the first eight columns of the matrix \mathbf{U} , which explain more than 90 percent variability in the coefficient matrix, are clustered, using the k -means. The

number of k was determined via common sense combined with the *Total Sum of Squares* loss, for different values of k .

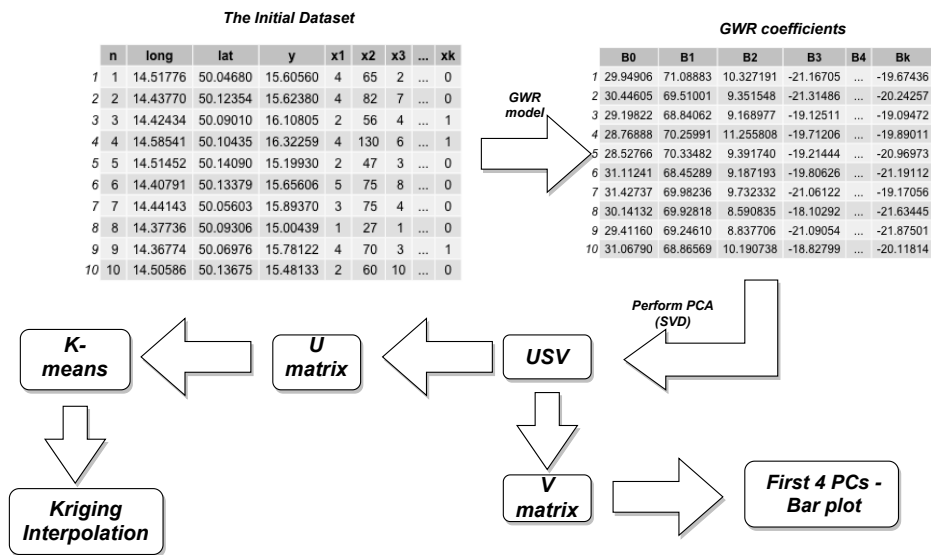


Figure 1 Housing Submarkets estimation Framework

Then, the last step includes, in order to have submarkets distribution continuously, the kriging interpolation of the values. The kriging technique is a form of a spatial prediction, which under certain assumptions is guaranteed to have *Best Linear Unbiased Predictions*. For more technical details and an overview of kriging technique see [1].

3 Dataset and Model Structural Form

The Data used for the purposes of our analyses were retrieved from the real estate webpage <https://www.sreality.cz/>. This server contains multiple estate advertisements and is assumed to be a credible representation of the flats market within the area of study. From the time period from February 2021 to May 2021, more than 2500 real estate advertisements were collected. However, some advertisements were not suitable for the purposes of our analysis as some forms of human errors or missing listed prices were present. For example, observations with the listed price of 0 CZK and 1 CZK had to be withdrawn from the dataset. The hedonic price equation is constructed based on the variables discussed in this section.

The dependent variable of interest is the price of the estate in CZK. Following many studies, we used the logarithmic transformation as this allows for much better interpretability and may help with the absence of normality of the distribution. Therefore, the $\log(\text{price})$ is our dependent variable.

The first independent variable is the *Meters*. This variable is commonly found to be highly correlated with the *price* and has very strong explanatory power. For the second independent variable, we use the variable *Room*, which is indicating the number of rooms the real estate disposes of. Another, yet important variable is the *Floor*, which indicates the vertical position within the housing unit. Taking the inspiration from the study of [4], we also use derived variables indicating the *zero* and *top* floors. Additionally, variables regarding the building condition were also utilized, combined with other variables regarding the type of the building i.e. the *Concrete* and *Brick*.

A very important key price determinant which many studies do not seem to be utilizing is the type of ownership. The two main types of ownerships are common, i.e. the *Private* and the *Cooperative*. It is naturally assumed that the *Private* type of the ownership is perceived exceptionally grandiosely, as there are usually certain legislative limitations associated with the *Cooperative* type of ownership. Lastly, another additional estate features such as the *Garage*, which indicates that an estate disposes of a garage, the *Balcony* indicating the presence of Balcony and/or Terrace and, lastly, the *Kitchenette* are also modeled. The data contain all described variables for 2 314 real estates and thus $n = 2314$. Additionally, since we are interested in spatial modeling, coordinates, i.e. longitude and latitude, were also kept.

All of these presented variables were used for our hedonic model and thus the following hedonic (*log*) price equation is estimated:

$$\begin{aligned} \log(\text{price}) = & \beta_0 + \beta_1 \text{Meters} + \beta_2 \text{Room} + \beta_3 \text{Floor} + \beta_4 \text{Floor zero} + \beta_5 \text{Floor top} + \\ & \beta_6 \text{After reconstruction} + \beta_7 \text{Very good} + \beta_8 \text{Concrete} + \beta_9 \text{Private} + \\ & \beta_{10} \text{Kitchenette} + \beta_{11} \text{Balcony} + \beta_{12} \text{Garage} + \beta_{13} \text{New building} \times \text{Brick} + \\ & \beta_{14} \text{New building} \times \text{Concrete} + \varepsilon. \end{aligned} \quad (4)$$

When reporting the coefficients of the GWR, it is not very suitable to report every single coefficient estimate for every single spatial unit. Thus, we report certain summary statistics of the model, and the estimated GWR model coefficients are presented in Table 1.

Table 1 GWR coefficient summary: Capital city Prague

key	Min	1st Qu.	Median	3st Qu.	Max	Global	Frac Positive
Meters	0.002	0.009	0.010	0.011	0.014	0.010	1
Room	-0.005	-0.001	0.002	0.026	0.216	0.031	0.669
Floor	-0.029	-0.005	0.003	0.008	0.033	0.002	0.620
Floor zero	-0.286	-0.127	-0.073	-0.003	0.109	-0.066	0.241
Floor Top	-0.199	-0.044	-0.006	0.029	0.125	-0.009	0.454
Building type: Concrete	-0.610	-0.198	-0.133	-0.088	0.258	-0.153	0.026
Condition: Very good	-0.235	0.025	0.050	0.078	0.294	0.049	0.889
Condition: After reconstruction	-0.121	0.034	0.069	0.106	0.312	0.071	0.940
Private	-0.095	0.096	0.165	0.255	0.624	0.183	0.988
Kitchenette	-0.119	-0.003	0.032	0.066	0.227	0.041	0.729
Balcony/Terrace	-0.099	0.028	0.053	0.080	0.279	0.056	0.932
Garage	-0.081	-0.003	0.025	0.085	0.314	0.043	0.725
Concrete x New Estate	-0.266	0.119	0.198	0.268	0.566	0.188	0.935
Brick x New Estate	-0.279	0.008	0.065	0.113	0.449	0.061	0.780
Constant	13.908	14.497	14.563	14.689	14.963	14.570	1

When investigating the GWR coefficients, a few fairly interesting observations can be made. Firstly, most of the regressors have a relatively large fraction of positive values the only exceptions being the *Floor Zero* and *Concrete*. The negative effect on the price is fairly expected from those characteristics as they have mostly negative perceptions associated with them. On the other hand, the characteristics such as *After reconstruction*, *Private* and *New Estate* have fairly large fraction of positive values and relatively high median and global effects.

As the baseline, the simple linear regression model was also used to estimate the hedonic equations (4). In order to compare both models, various metrics such as e.g. AIC and R^2_{pse} , calculated as $\text{corr}(y, \hat{y})^2$, can be compared. The R^2_{pse} for the OLS model is 0.61, where for the GWR model is 0.92. Similarly, the AIC seems to be in favor of the GWR model, with values $AIC_{GWR} = 128$ and $AIC_{OLS} = 376.37$. Clearly, the GWR model provides a much better model fit and thus better statistical inference. Moreover, as an extra feature of GWR, the housing submarkets can be identified. This would not be the case of the OLS model, as the model does not model the spatial heterogeneity nor does it allow the coefficient estimates to vary in the space.

4 Housing Submarkets

In order to identify the housing submarkets in Prague, the GWR coefficient matrix is centered, and then, using the methodology summarised in the Figure 1, the SVD matrix factorization is performed. Firstly, the main sources of variability can be investigated, using the \mathbf{V} matrix from the SVD (see e.g. [6]). Then, in order to identify the housing submarkets, the k -means clustering algorithm is applied on the \mathbf{U} matrix of the SVD¹. The selected parameter k was determined via the loss function and common sense approach discussed above and equals to ten. The final interpolated housing submarkets in Prague can be closely observed in the Figure 2.

¹ We clustered the first eight columns of the matrix, as they capture more than 90 percent of the variability.

Housing Submarkets: Spatial Distribution Capital city Prague

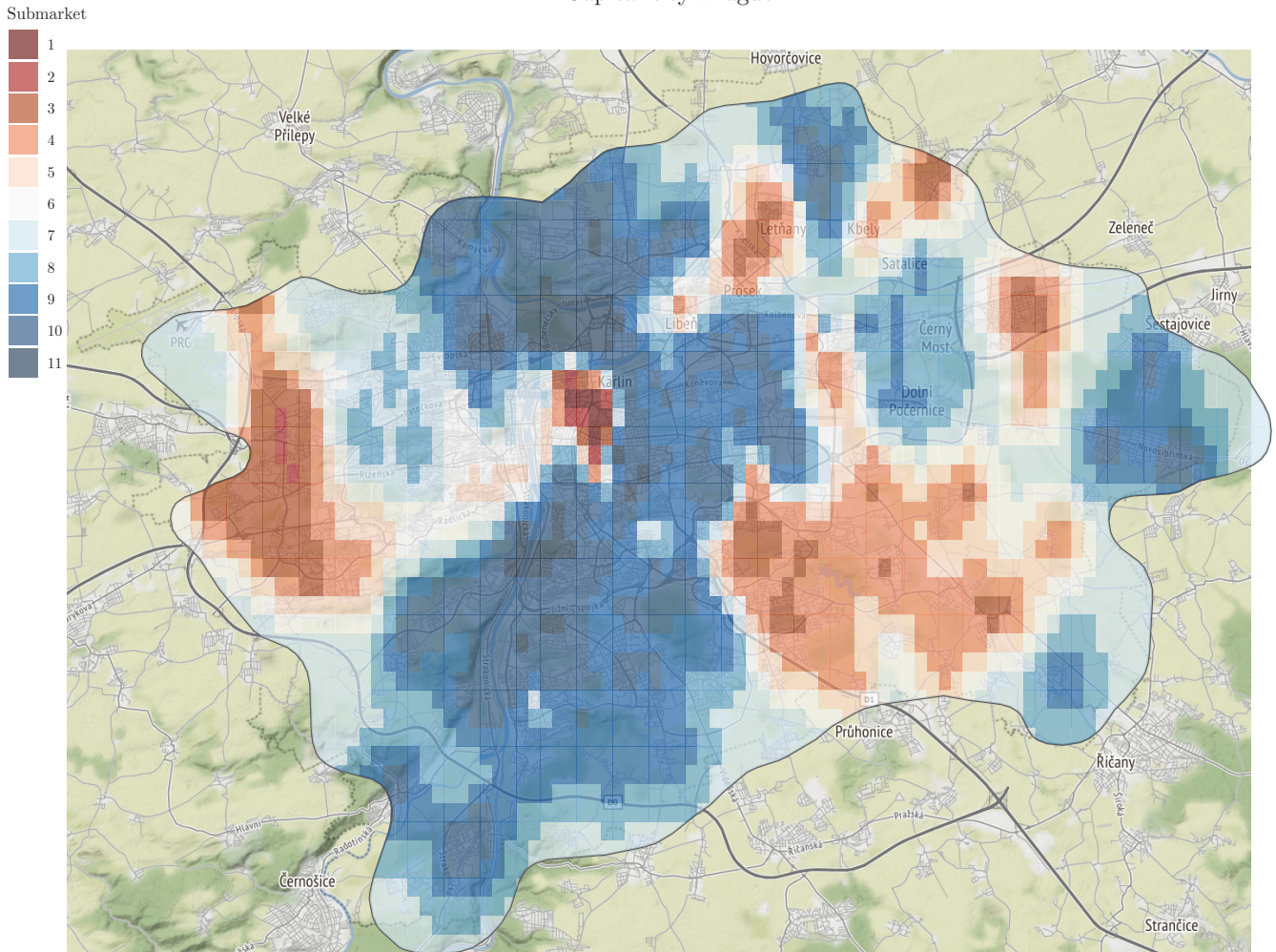


Figure 2 Housing submarkets of Prague ($n = 10$)

Once the housing submarkets are identified, it is assumed that the level of homogeneity within each cluster is considerably higher as opposed to the levels between the clusters. Particularly, the effects of the key price determinants, which are modeled in the hedonic equation (4), are assumed to be extensively similar within every single housing submarket. Moreover, as an extra feature of the described framework of identifying the housing submarket, assuming that the real estate data are collected over multiple time periods (say, once a year for the past five years), the housing submarkets allow for the analysis and evaluation of the key price determinants over time, modeling both phenomena of spatial heterogeneity as well as time effects.

First and most importantly, inspecting the Figure 3, we can conclude that the effect of price determinants does vary in space. Additionally, a few interesting conclusions can be made. The effect of *Concrete* building type has quite a large negative effect on the price, which is naturally expected. Interestingly enough, the *Garage* seems to be perceived as an extravagant feature, especially in the historical parts of the city. Moreover, the effects of *Meters* and *After a reconstruction* tend to vary quite vigorously over the analyzed area.

Last but not least interesting conclusion which can be made is the one submarket *n.1*, which is associated with the very historical part of the city. Even though this submarket is quite small, a very large heterogeneity compared to the other clusters is present.

5 Conclusion

The spatial statistics techniques (the GWR model and Kriging interpolation) were combined with the techniques of statistical learning i.e. the PCA and the *k*-means methods in order to identify the housing submarkets in Prague.

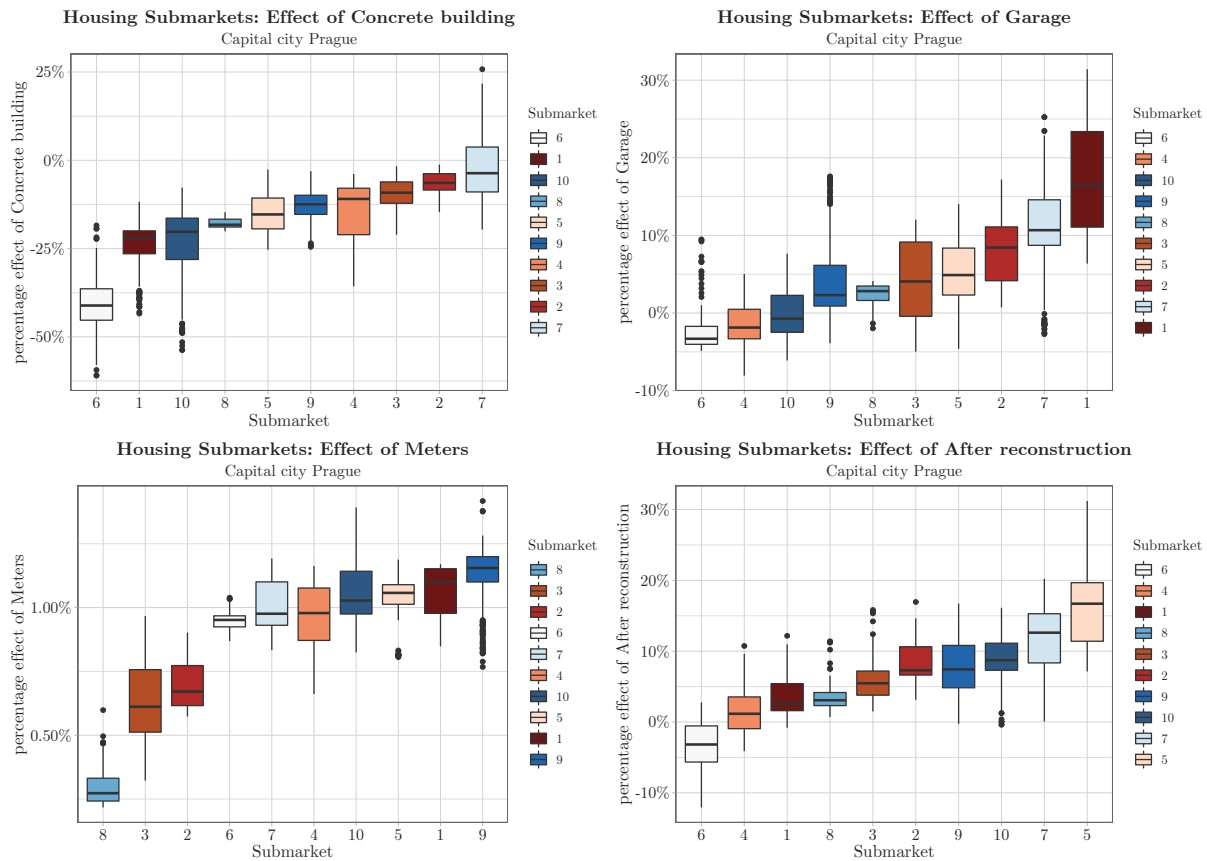


Figure 3 Distributions of effect on the price for selected price determinants

To model the spatial heterogeneity and to explore the effect of key price determinants over space, the GWR model proved to be a very suitable model to utilize. In future researchers, one can explore the effect of time, which is without any doubt also a key factor for the hedonic price models. The framework introduced in this study is very suitable for both one time period as well as multiple time period comparisons.

Acknowledgements

This research was supported by the Internal Grant Agency of the Prague University of Economics and Business under project F4/27/2020.

References

- [1] Bivand, Roger S and Pebesma, Edzer J and Gomez-Rubio, Virgilio and Pebesma, Edzer Jan. (2013). *Applied spatial data analysis*, Springer.
- [2] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- [3] Hrobař, Petr., Holý, Vladimír. (2020). *Spatial Analysis of the Flat Market in Prague*. In *International Conference on Mathematical Methods in Economics 2020 (MME 2020)*. Brno: Mendel University, s. 193–199. ISBN 978-80-7509-734-7.
- [4] Kopczewska, Katarzyna and Ćwiakowski, Piotr. (2021). *Spatio-temporal stability of housing submarkets. Tracking spatial location of clusters of geographically weighted regression estimates of price determinants*. Land Use Policy, 103
- [5] 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
- [6] Robert, Christian. (2014). *Machine learning, a probabilistic perspective*, Taylor & Francis
- [7] Zhou, Qianling and Wang, Changxin and Fang, Shijiao. (2019). *Application of geographically weighted regression (GWR) in the analysis of the cause of haze pollution in China*. Atmospheric Pollution 10-3

Forecasting Czech unemployment rate using dimensional reduction approach

Filip Hron¹, Lukáš Frýd²

Abstract.

We compare prediction power within ARMA, vector autoregression (VAR) and factor augmented VAR model. The prediction is done for the Czech unemployment rate. We show, that the few unobserved common factors estimated by PCA are effective to predict different horizon of the series and outperformed traditional VAR approach.

Keywords: dimensional reduction, unemployment rate, factor augmented VAR, high dimensional time series

JEL Classification: C50, O40

AMS Classification: 62P20

1 Introduction

The curse of dimensionality is a problem we face in trying to understand the behaviour of complex systems, such as the national economy. The monthly and quarterly frequency of data collection significantly limits the estimation of a larger model. In the case of a popular vector autoregression, the number of coefficients increase quadratically based on the number of included variables in the model and their lagged values. The limitation to a few variables significantly affects the analysis of the transmission of information through the economy as well as its predictive power.

One possible solution to the curse of dimensionality problem is dimension reduction. Geweke [4], assumes, that a complex system is driven by a few unobserved common factors. These factors have a vector autoregressive behaviour and they are a combination of a exogenous variables. The first well-known empirical study by Sargent and Sims [2] establish and show, that the two dynamic factors are relevant predictors for explaining a substantial variance of the main U.S. macroeconomic indicators with the quarterly period. Furthermore, Stock & Watson [3] highlight the usability of the dynamic factor models in forecasting. Their study supports the assumption, that the few factors estimated by the principal components have robust forecasting properties.

Finally, Bernanke *et al.* [5] extend existing methodology of the dynamic factor modelling by the vector autoregression process (FAVAR). They propose vector autoregression estimator based on the principal component analysis capable shrinks over the one hundred macroeconomic variables into the three respectively five factors. Hence, FAVAR process is represented as a combination of a directly observable relatively small number of the variables and the few unobserved common factors. Moreover, the factor augmented VAR approach allows calculate effects of many observable variables represented by factors.

In this paper, we evaluate the predictive power of the ARMA, VAR and FAVAR models. Specifically, we focus on the forecasting of the Czech unemployment rate during the period 2000 to 2020.

2 Data and methodology

2.1 Data

The data set consists of the time series from 2000 until the end of 2020 with quarterly period. All time series are seasonally adjusted and transformed into the stationary process. Moreover, we standardize each series (*Z-score*) for the proper use of the *principal components* approach (PCA).

In the study is used predominantly Czech predictors. The assumption is the domestic indicators has the most significant impact on the behavior of the unemployment rate. Furthermore, the variable selection mostly depends on the overall availability of the series during the investigated period. On the other hand, we are aware, that German

¹ Department of Econometrics, University of Economics in Prague, Winston Churchill Square 4, 13067 Prague, Czech Republic, hrof01@vse.cz

² Department of Econometrics, University of Economics in Prague, Winston Churchill Square 4, 13067 Prague, Czech Republic, lukas.fryd@vse.cz

economy had a large impact on our country, too. Therefore, we include the German gross domestic product. The series are listed in the table below:

Variable
General unemployment rate
Actual individual consumption
Economically inactive
Gross value added
Imports of goods and services
Gross domestic product
Industry production
Public 15+
Wages and salaries
Gross household saving rate
Customer price index
Gross domestic product (<i>Germany</i>)

Table 1 Overall data set for forecasting

2.2 Factor-augmented vector autoregression approach

The curse of dimensionality problem can be solved for example by shrink the number of coefficients based on the proper prior beliefs (*Bayesian VAR*) or reduce the space of the used predictors - dimension reduction. In general, it is assumed, that the economy is influenced and driven by the relatively small number of the common unobserved factors. These factors can be understood as invisible hand or the moves on the demand side.

If we consider the static factor model firstly introduced by Geweke [4], we assume that, these common factors affecting all the time series only contemporaneously and the process follows:

$$X_t = \Lambda F_t + \epsilon_t, \quad (1)$$

where vector X_t $N \times 1$ represents the observable variables, vector F_t $K \times 1$ represents the unobservable factors and ϵ_t represents uncorrelated idiosyncratic errors. In addition, we assumed, that the number of the common factors is relatively smaller than the observable series ($K \ll N$).

Next, Geweke introduce dynamic form of the factor models. The equation 1 is extended by lagged values of F_t with lag polynomials $\Lambda(L)$ of q order.

$$X_t = \Lambda(L)F_t + e_t, \quad (2)$$

Stock & Watson [3] points out the usability of dynamic factor models. Authors demonstrate, that a few common factors estimated by PCA are great predictors for forecasting various time series.

Based on the standard VAR model and dynamic factor models Bernanke *et al.* [5] define the *Factor-augmented vector autoregression* approach as follows:

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Lambda(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + \varepsilon_t, \quad (3)$$

where $M \times 1$ vector Y_t contains the directly observable factors, $K \times 1$ vector F_t represents the unobserved common factors similarly to the equations 1 and 2. These factors are estimated from the data set X_t , which is a relatively large macroeconomic informational set of N observable variables. In the study, the Y_t represents the Czech unemployment rate and the factors F_t corresponds to the shrinkage of the other mentioned predictors into the few common components. Similarly to the dynamic factor models the $\Lambda(L)$ indicates the polynomial lag operator [5].

The vector X_t can be expressed as a combination of F_t and Y_t follows:

$$X_t = \lambda^F F_t + \lambda^Y Y_t + e_t, \quad (4)$$

where λ^F and λ^Y represents the factor ($N \times K$) and the directly observable variables ($N \times M$) loading matrices [5]. Lastly, e_t represents the idiosyncratic specific error components.

Bernanke et al. [5] introduce two estimators. The first one based on the principal component analysis, *two-step principal components* and next estimator based on the bayesian approach. In this paper we utilize only the *two-step principal components* approach. The factors can be estimated based on the expression below:

$$\hat{F}_t = \hat{C}_t - \hat{\beta}_y Y_t, \quad (5)$$

where \hat{C}_t represents the estimation of $C_t = (F_t', Y_t')$ by $K + M$ principal components of Y_t and X_t . The space spanned by C_t do not take into account the fact, the Y_t is observable variable. However, Stock & Watson [7] underline and highlight, that with eligible large N and the number of components P , when $P \gg K$, the principal components precisely recover the spanned space. Consequently, the factors F_t are estimated from space described by \hat{C}_t reduced by effect of Y_t , in the first step. In the second step, standard VAR model is used.

2.3 Forecast evaluation

We use 3 well known indicators to properly evaluate the forecast power of each approaches: Root Mean Square error, Mean Absolute Percentage error and Median Absolute Percentage error. These benchmarks follows:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - p_t)^2}, \quad (6)$$

$$MAPE = \frac{1}{T^{MAPE}} \sum_{t=1}^{T^{MAPE}} \left| \frac{y_t - p_t}{y_t} \right|, \quad (7)$$

$$MdAPE = median \left(\left| \frac{y_t - p_t}{y_t} \right| \right), \quad (8)$$

where y_t represents actual value and p_t predicted.

3 Results

The standard VAR process follows a combination of the unemployment rate, inflation and gross domestic product. The inflation is represented by consumer price index. In addition, the best lag order for the FAVAR and VAR is set to $p = 4$. According to the Bernanke et al. [5], the optimal number of used factors can be estimated based on the informational criteria. However, authors also points out, that the number of factors can be determined *ad-hoc*. We conclude, the methodology works well using only 3 factors. These factors described almost 60% of the system variance during the current period. We assume, the described variance increases with the lagged values.

We provide in-sample and out-of-sample predictions. The forecast horizons are from one to tree years starting by year 2018. The in-sample prediction are displayed in table 2.

Est	RMSE	MAPE	MdAPE
VAR(4)	0.0894	0.1034	0.1013
FAVAR (4)	0.0518	0.0648	0.0577
ARMA(4,4)	0.1012	0.1162	0.1034

Table 2 Prediction of each approach and their benchmarks on the period 2000 - 2017

Next, we highlight the fit of each model. The figure 3 shows, that FAVAR approach fits well the period of the Great Recession in 2008 followed by full economy recovery. The benchmarks show, the FAVAR is significantly better than other approaches in in-sample predictions. The FAVAR produce only around 5% prediction accuracy error respectively to the measures. On the other hand, the standard VAR and ARMA methodology have average 10% errors.

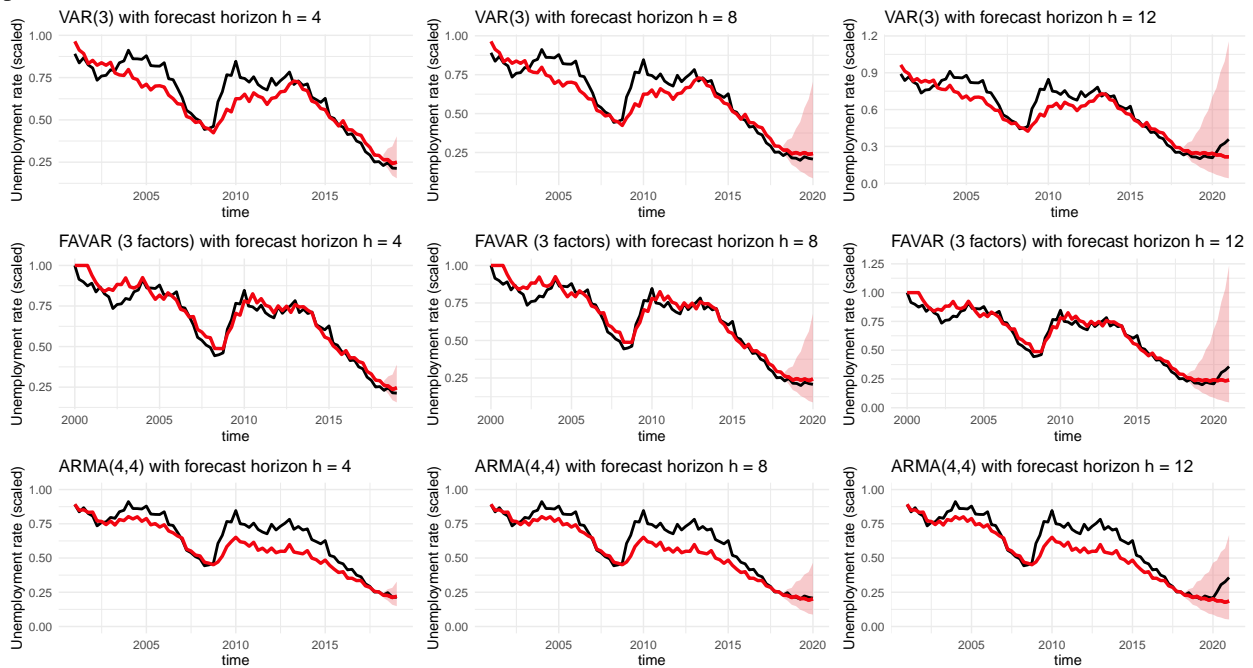
Out-of-sample predictions are slightly different. The FAVAR is definitely better then the standard VAR approach with included inflation and gross domestic product, see table 3. The included information in the estimated factors definitely help to produce better forecasts. However, ARMA model outperforms the the others models significantly in first two horizons. The ARMA model produces more accured forecast. More specifically, ARMA produce forecast 1 year ahead with only $RMSE = 1.1\%$, compared to 3% RMSE measure of the FAVAR and VAR.

In addition, MAPE measure for ARMA with $h = 8$ is equal to 4.6%, compared to $MAPE_{FAVAR} = 12\%$ and $MAPE_{VAR} = 14.16\%$, which is significantly higher. On the other hand, based on the RMSE and MAPE with the horizon $h = 12$ the FAVAR outperformed ARMA model. These measurements differ around 2 percentage points. Forecasting Czech unemployment rate is slightly difficult questions during period 2018 and 2020 because of its relatively small variance. Furthermore, the COVID situation represents relatively great shock in the Czech labor market. Moreover, it is appropriate to use more forecast accuracy measures. On the one hand ARMA holds slightly similar values for the measures, on the other hand, measures for VAR and FAVAR are considerably different.

Est	h	RMSE	MAPE	MdAPE
VAR(4)	4	0.0299	0.1297	0.1376
FAVAR (4)	4	0.0254	0.1064	0.1159
ARMA(4,4)	4	0.0110	0.0352	0.0248
VAR(4)	8	0.0313	0.1416	0.1376
FAVAR (4)	8	0.0272	0.1200	0.1159
ARMA(4,4)	8	0.0126	0.0460	0.0320
VAR(4)	12	0.0622	0.1837	0.1607
FAVAR (4)	12	0.0521	0.1539	0.1452
ARMA(4,4)	12	0.0772	0.1627	0.0819

Table 3 Unemployment rate forecast with the different horizons (4,8,12)

Figure 1 Actual (black) versus predicted (red) values of scaled Unemployment rate. Different forecast horizons provided.



4 Conclusion

In the study, we utilize Factor-augmented vector autoregression approach to produce prediction of the Czech unemployment rate. The methodology supports the assumption, that the economy is driven by a few unobserved common factors. These factors can be estimated from the large informational set of macroeconomic variables. We show, that three unobserved factors extracted from the twelve macroeconomics time series significantly improve long-term (3 years) unemployment rate forecast. On the other hand, ARMA model has better forecast accuracy for short-term period (1-2 years). Moreover, FAVAR model captures unemployment rate during Great Recession in 2008 better, than VAR and ARMA model.

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] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
- [2] Sargent, T. J., & Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research*, 1, 145-168.
- [3] Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460), 1167-1179.
- [4] Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*.
- [5] Bernanke, B. S., Boivin, J., & Elias, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of economics*, 120(1), 387-422.
- [6] Lombardi, M. J., Osbat, C., & Schnatz, B. (2012). Global commodity cycles and linkages: a FAVAR approach. *Empirical Economics*, 43(2), 651-670.
- [7] Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460), 1167-1179.

Health index for the Czech districts calculated via methods of multicriteria evaluation of alternatives

Dana Hübelová¹, Beatrice-Elena Chromková Manea², Alice Kozumplíková³, Martina Kuncová⁴, Hana Vojáčková⁵

Abstract. The aim of this paper is to calculate the health index for each of the 77 Czech districts to identify the best and worst districts from this perspective. The health index is constructed as a composite indicator of 8 different areas covering 60 criteria. The index reduces the size of the original set of indicators, which allows its clear and easy interpretation, including its spatial differentiation. Thanks to this, the results can be a suitable basis for decision-making by state administration and self-government bodies or other organizations dealing with the issue of population health. The data come from publicly available databases of the Czech Statistical Office, the Institute of Health Information and Statistics of the Czech Republic, the Ministry of Labor and Social Affairs of the Czech Republic, the Czech Hydrometeorological Institute and the Czech Panel Survey of Households. For the health index construction, the methodology using selected methods of multicriteria evaluation of alternatives is described. With regard to the available data, methods using numerical input, weights of criteria and determining the complete order of alternatives were analyzed and used.

Keywords: multicriteria comparison, health index, health inequality, Czech districts

JEL Classification: C44, I14

AMS Classification: 90B50, 91B06

1 Introduction

Health is a concept of a complex nature that raises many theoretical (concerning the definition and conceptualization of health) and empirical (as to the way of how we operationalize, measure, and analyse health status) questions. The World Health Organization (WHO) defines health as "... an overall state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity." It is also "... a source of daily life, not a goal in life"[26]. The health of the population is considered an essential indicator of the development and competitiveness of regions. Health quality indicates the state and links between social, economic, demographic, environmental, but also political processes [12]. Determinants of health then represent indicators that influence the presence and development of risk factors for the disease [3]. Health inequalities are unequal differences resulting from inequalities in a number of determinants of different natures, which, due to their uneven distribution, are considered one of the main causes of creating and maintaining health inequalities [13]. Any measurable aspect of health that varies among individuals or according to socially relevant groups in a population are regarded to be health inequalities. Ideally, everyone should have the same chances to reach his or her full health potential [4]. Although health inequalities are uneven, they can be prevented, as they are the result of inappropriate public policies or unhealthy lifestyles [25]. Health inequalities are apparent in different groups of the population, according to age, gender, ethnicity, socio-economic status or place of residence, etc., resulting in the exposure of individuals and groups to spatially distributed health risks. Spatial epidemiology focuses its activity on the evaluation of territorial aspects of health inequalities based on the interpretation of geographical contexts. When using secondary data, acceptable explanations of spatial differentiation of determinants of health inequalities can be found, while the disadvantages (time and economic complexity) of case and control studies, resp. cohort studies are eliminated. The direct applicability of the method of descriptive epidemiology has been proved, for example, in the creation of the Atlas of Cancer Incidence in England and Wales [24] or the US Mortality Atlas [2]. Spatial data in relation to health characteristics are also presented by the Eurostat atlases [8], [10], which use various

¹ Mendel University in Brno, Faculty of Regional Development and International Studies, Department of Social Studies, třída Generála Píky 2005/7, 613 00 Brno, Czech Republic, hubelova@mendelu.cz

² Mendel University in Brno, Faculty of Regional Development and International Studies, Department of Social Studies, třída Generála Píky 2005/7, 613 00 Brno, Czech Republic, chromkov@mendelu.cz

³ Mendel University in Brno, Faculty of Regional Development and International Studies, Department of Environmental Sciences and Natural Resources, třída Generála Píky 2005/7, 613 00 Brno, Czech Republic, alice.kozumplikova@mendelu.cz

⁴ College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 586 01 Jihlava, Czech Republic, kuncova@vspj.cz

⁵ College of Polytechnics Jihlava, Department of Technical Studies, Tolstého 16, 586 01 Jihlava, Czech Republic, hana.vojackova@vspj.cz

calculations and cartographic outputs to describe in detail the situation in European countries on the basis of selected causes of avoidable mortality and taking into account the cultural and social context of the country. Another important atlas concerning the mortality and health situation of the population is the Atlas of Health in Europe [27]. This atlas contains basic statistics on the health status of the population in the period 1980-2001. The data were collected, verified, and processed in a uniform way for a better comparison. The second edition of the 2008 atlas provided a summary of current data from 53 European countries (1980-2006) to effectively address the new public health challenges [28].

Outputs using similar methods on data from the Czech Republic have been published since the 1990s when researchers from the Institute of Health Information and Statistics prepared the Atlas of Avoidable Deaths in Central and Eastern Europe, similar to the already published atlas at that time for Western European countries [14]. An Atlas of Morbidity and Mortality was also created by the same researchers, where the health status and mortality in the EU15 countries and in other acceding countries were compared. The Atlas of Socio-Spatial Differentiation [23] and the context of social and health services presented in the Atlas of Long-Term Care of the Czech Republic [29] deal with the social aspects of development. As these resources do not cover more than one health-related area, the aim of this article is to fill this gap and to calculate the health index for each of the 77 Czech districts to identify the best and worst districts using selected methods of multicriteria evaluation of alternatives. The health index is constructed as a composite indicator of 8 different areas covering 60 criteria. The index reduces the size of the original set of indicators, which allows its clear and easy interpretation, including its spatial differentiation.

2 Data and methodology

2.1 Data

The data come from publicly available databases of the Czech Statistical Office [7], the Institute of Health Information and Statistics of the Czech Republic [17], the Ministry of Social Affairs of the Czech Republic [21], the Czech Hydrometeorological Institute [5] and the Czech Panel Survey of Households [6]. They cover 60 criteria in 8 different areas (Table 1) and 77 Czech districts. The selection of the areas and criteria was inspired by the Euro-Healthy project [8], which created an index for the health status of the population in EU countries at their regional level NUTS2 and metropolitan areas. As in the Euro-Healthy project, we have compiled a comprehensive health index, which is composed of sub-indices for various individual dimensions of health. Unlike the "Euro-Healthy" project, our results differ in their regional dimension, which allows us to specify in more detail the individual determinants of health inequalities in the Czech Republic at the level of districts and selected municipalities. It allows us to more objectively identify the regional differentiations in their geographical, settlement, economic, social, environmental differences. Most of the 60 criteria are of minimization type, only 16 of them are of the maximization type. The choice of the areas and criteria is based on previous research (see for example [15]) and on data availability for the chosen topic.

No.	Area	Number of criteria
1	Economic conditions and social protection	14
2	Education	2
3	Demographical changes	4
4	Environmental conditions	6
5	Individual conditions	3
6	Safety in road transport and crime	5
7	Sources of health and social care	5
8	Health status	24

Table 1 Basic description of compared areas

2.2 Methodology

The original idea of the research was to determine the health index for each district in the Czech Republic on the basis of the 60 criteria described above, divided into 8 areas. Methods of multicriteria evaluation of alternatives are suitable for this type of analysis as described in [18]. These methods have been developed to help the decision-maker to find the best alternative or the ranking of alternatives [11]. To find the ranking of alternatives (a_1, a_2, \dots, a_p) via numerical criteria (f_1, f_2, \dots, f_k) when equal preferences are set for the criteria, the methods using criteria weights which result in a complete order of alternatives can be applied [1]. The methodology for this paper is based on two phases: in the first one, the 14 criteria of the 1.area (Economic conditions and social protection) were

used as a sample to assess the suitability of the application of individual methods; in the second phase, the health index calculation procedure based on 3 steps was suggested. In the first step, the evaluation of districts was obtained according to each area separately (with the equal weights of criteria within the areas). In the second step the same method was used for the complete evaluation in 8 areas together. This result could be taken as the health index of each district but for the graphical representation in the map of the Czech Republic, the results were divided into 6 clusters in the third step.

Due to the nature of the input data, methods using quantitative evaluation of alternatives were tested to select the one that will correspond to the expected evaluation of districts and can be easily programmed to create the graphical output. The health index can therefore be estimated as a utility based on the mentioned 8 areas - then methods for maximizing the utility function are offered for calculations, i.e. WSM (Weighted Sum Method), SAW (Simple Additive Weighting), WSA (Weighted Sum Approach), UFA (Utility Function Approach) [1]. The second possibility is to create the benefit as a relative distance from the ideal and nadir alternative, which is a typical procedure for the TOPSIS method. As the aim is to divide the districts into color clusters based on the index, and the utilities of each criterion should be obtained from the method and not set by decision-maker (as in UFA), we decided first to test WSM, SAW, WSA and TOPSIS and then to choose one of them for the next calculations.

2.3 WSM, SAW and WSA methods

WSM, SAW and WSA belong to the category of Multi-criteria decision making (MCDM) where the principle of utility maximization is used [1]. For all of them the formula (1), for the calculation of the final utility $u(a_i)$ for each alternative i , is common, the only difference is the calculation of normalized values r_{ij} that express the transformation of data into 0-1 scale (if w_j are the criteria weights).

$$u(a_i) = \sum_{j=1}^k w_j r_{ij}, \quad \forall i = 1, \dots, p \quad (1)$$

WSM is one of the simple methods that does not use any criteria normalization function and for the calculation of the final utility of the alternatives the r_{ij} is equal to real data a_{ij} . SAW normalizes real data and several normalization formulas can be used, we calculate r_{ij} as in formula (2) for the maximization type of criterion or (3) for minimization type, where $\min(a_{ij})=0$ in both formulas. WSA is a case of SAW method where for the normalization formulas (2) and (3) are used with $\min(a_{ij})$ taken from the real data.

$$r_{ij} = \frac{a_{ij} - \min_i a_{ij}}{\max_i a_{ij} - \min_i a_{ij}} \quad (2)$$

$$r_{ij} = \frac{\max_i a_{ij} - a_{ij}}{\max_i a_{ij} - \min_i a_{ij}} \quad (3)$$

2.4 TOPSIS method

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 nadir alternative. The steps of this method can be described as follows [11]: normalise the decision matrix according to Euclidean metric (4), calculate the weighted decision matrix $W = (w_{ij}) = v_j \cdot r_{ij}$, and identify vectors of the hypothetical ideal H and nadir 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 nadir alternatives over each attribute using formulas (5) and finally order alternatives by maximizing ratio c_i which is equal to the ratio of d_i^- and the sum of d_i^+ and d_i^- .

$$r_{ij} = \frac{y_{ij}}{\sqrt{\sum_{i=1}^p y_{ij}^2}}, \quad \forall i = 1, \dots, p, \quad j = 1, \dots, k, \quad (4)$$

$$d_i^+ = \sqrt{\sum_{j=1}^n (w_{ij} - H_j)^2} \quad \text{and} \quad d_i^- = \sqrt{\sum_{j=1}^n (w_{ij} - D_j)^2}, \quad \forall i = 1, \dots, p, \quad (5)$$

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 (TOPSIS-classic) for transformation but it was proved that the inverse values (TOPSIS-inverse) could be better for TOPSIS method with more minimization criteria [19], that is why we use both kinds of transformation.

3 Results and discussion

Before the final health index construction, it was necessary to select the appropriate method. The 1.area with 14 criteria, where 4 are maximizing and 10 are minimizing, was taken as a sample to select the method. Since the input values are both in different units and in different scales, it is not appropriate to use the WSM method. The results (utilities for SAW and WSA, and relative distance for TOPSIS) are shown on Figure 1. It is clear that there are indeed differences in the TOPSIS method for different transformations, and it is visible that the WSA method better spreads the districts in terms of the resulting utility on a scale of 0-1. Therefore, WSA was used as an input for the 2.phase of the index calculations. We note that the correlation of the results between WSA and other methods was in all cases higher than 0.84, so the results of all tested methods are very similar.

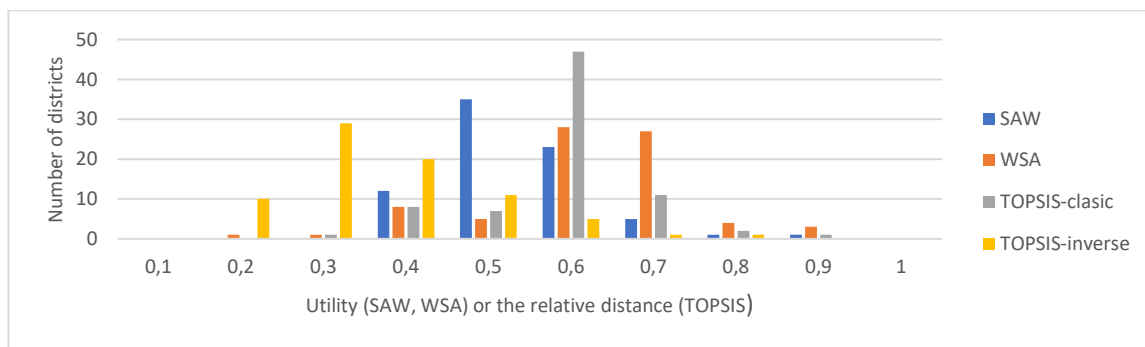


Figure 1 First phase results - number of districts by final utility (WSA) or relative distance (TOPSIS)

The second phase was based on WSA method. Table 2 illustrates the first step in the health index calculations, it means the scores (utilities) of each area of measurement for the first 10 districts (alphabetically ordered – there is no space to show all 77 districts). The higher the value, the better a district is situated on the index (area).

District	1.area	2.area	3.area	4.area	5.area	6.area	7.area	8.area
Benešov	0.65990	0.37688	0.49608	0.58251	0.40297	0.61525	0.39019	0.59461
Beroun	0.60963	0.43044	0.47313	0.50299	0.44326	0.75968	0.46559	0.60469
Blansko	0.59635	0.43331	0.50368	0.61504	0.16543	0.81784	0.35800	0.71159
Brno-město	0.52210	0.91188	0.57322	0.47588	0.57080	0.71333	0.80041	0.69450
Brno-venkov	0.62155	0.46245	0.49281	0.57769	0.47073	0.76993	0.20837	0.73282
Bruntál	0.32968	0.16065	0.53924	0.70661	0.32183	0.60676	0.43221	0.37165
Břeclav	0.61993	0.19019	0.45514	0.47287	0.54274	0.73828	0.41117	0.65755
Česká Lípa	0.55099	0.20507	0.65403	0.66022	0.30179	0.40803	0.24782	0.43539
České Budějovice	0.62144	0.56948	0.57650	0.73994	0.48904	0.74659	0.58184	0.62593
Český Krumlov	0.58995	0.18394	0.50539	0.93554	0.58595	0.61723	0.39332	0.45265

Table 2 Demonstration of the first step of calculations using the WSA method (example of 10 districts)

Rank	1	2	3	4	5	6	7	...	73	74	75	76	77
District	Praha-záp.	Brno-město	České Budějovice	Praha-východ	Plzeň-město	Jihlava	Praha	...	Most	Bruntál	Karviná	Jeseník	Tachov
H.Index	0.696	0.690	0.637	0.611	0.594	0.582	0.581	...	0.359	0.348	0.344	0.340	0.329

Table 3 Demonstration of the second step of calculations via WSA (example of best and worst districts)

We can observe differences both within the same district (among various areas measured) and between districts (within the same area). For example, within the Brno-město district the 2. area (Education) has the highest value of the index (0.91188) as compared to the other areas, and also the highest level on the same area (e.g. 2. area) among presented districts (0.91188 compared to 0.16065 for Bruntal district).

In the second step, WSA method was used again for the final health index calculation, where all areas were supposed to have the same weight. Table 3 describes the best and the worst districts. As we assumed, big cities like Prague and Brno belong to healthier areas (mainly for the good results in 1.area – Economic conditions and social protection; 2.area – Education; 5.area – Individual conditions and 8.area – Health status), while northwest

Bohemia and north Moravia belong to the less healthy ones (for example the last district, Tachov, has below-average results in all areas except 4.area – Environmental conditions; Jeseník or Bruntál has low values also in the 4.area and in the 6.area – Safety in road transport and crime). In the last (third) step, the obtained health index values were divided into 6 clusters according to the size of the resulting index (1.cluster cover districts with the values higher than 0.65, the last, 6.cluster, cover districts with the value lower than 0.45), and then the clusters were incorporated into the map of the Czech Republic for better visualization of the results (see Figure 2).

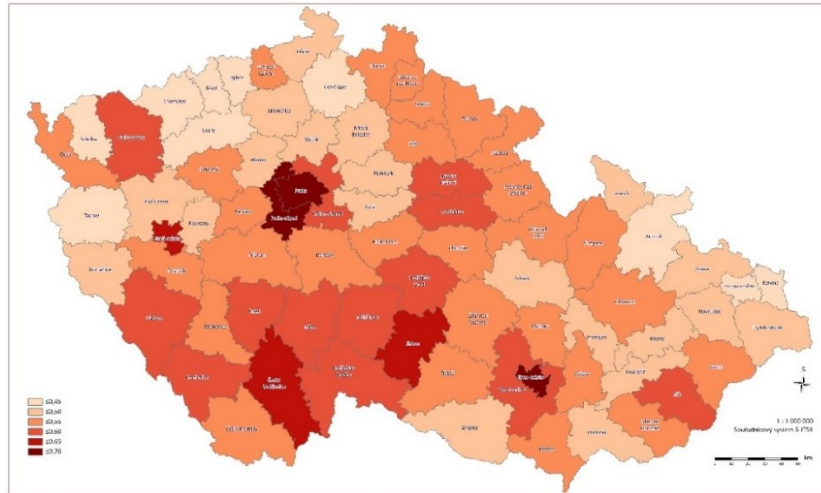


Figure 2 Districts according to the clusters of the health index

4 Conclusion

For the optimal development of the regions, it is important to create and implement strategies that will lead to ensuring competitiveness and sustainable regional development, for which the optimal quality of population health is one of the conditions. A satisfactory health status of the population directly affects the economic level of the region, and it has a positive effect on the social development and demographic stability at the regional level [16]. Not only due to the current pandemic situation (COVID-19), it is obvious that an increased morbidity and lower quality of health in the population puts more strain on the public budget and the use of public goods and services (e.g. social and health expenditures). A healthy region and sustainable regional development therefore require a healthy population, a high-quality health system, individual responsibility for own health, and proper prevention[22]. The aim of this article was to propose a three-step procedure using the WSA method for calculating the health index for each of the 77 Czech districts and finally put it in the map as clusters. The results can be a suitable basis for decision-making by state administration and self-government bodies or other organizations dealing with the issue of population health. The proposed procedure also allows the setting of weights of criteria and areas according to requirements of the decision-maker.

Acknowledgements

The paper was supported by the contribution of long-term institutional support of research activities by the College of Polytechnics Jihlava and by project TL03000202 Health Inequalities in the Czech Republic: Importance and Relationship of Health Determinants of Population in Territorial Disparities is co-financed with the state support of the Technology Agency of the Czech Republic in the Program ÉTA.

References

- [1] Alinezhad, A. & Khalili, J. (2019). *New Methods and Applications in Multiple Attribute Decision Making (MADM)*. Springer Nature Switzerland AG.
- [2] Bell, B. S., Hoskins, R. E., Pickle, L. W. et al. (2006). Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. *Int J Health Geogr* 5, 49. <https://doi.org/10.1186/1476-072X-5-49>
- [3] Berman, J. D., Fann, N., Hollingsworth, J. W. et al. (2012) Health Benefits from Large-Scale Ozone Reduction in the United States. *Environmental Health Perspectives*, 120(10): 1404–1410.

- [4] CSDH. (2008). *Closing the gap in a generation: Health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health*. WHO, Geneva.
- [5] Czech Hydrometeorological Institut (2020). National Geoportal INSPIRE. <http://geoportal.gov.cz>
- [6] Czech Statistical Office (2020). *Population Censuses*. <https://www.czso.cz/csu/czso/population-censuses>
- [7] Czech Statistical Office. (2020). *Demografická ročenka okresů*. <https://www.czso.cz/csu/czso/demograficka-rocenka-okresu-2010-az-2019>
- [8] EC. 2020. Shaping EUROpean policies to promote HEALTH equity. <https://cordis.europa.eu/project/id/643398/reporting>
- [9] Eurostat. (2002). *Health statistics: atlas on mortality in the European Union: data 1994–96*. ed. Luxembourg: Office for Official Publications of the European Communities.
- [10] Eurostat. (2009). *Health statistics: atlas on mortality in the European Union*. ed. Luxembourg: Office for Official Publications of the European Communities.
- [11] Figueira, J., Greco, S. & Ehrgott, M. (2005). *Multiple Criteria Decision Analysis – State of the Art Surveys*.
- [12] Fraser, S. D. S. & George, S. (2015). Perspectives on Differing Health Outcomes by City: Accounting for Glasgow’s Excess Mortality. *Risk Management and Healthcare Policy* 8: 99–110.
- [13] Graham, H. (2004). Social determinants and their unequal distribution: clarifying policy understandings. *The Milbank quarterly*, 82(1): 101–124.
- [14] Holland, W. W. (1993). European Community Atlas of Avoidable Death. 2nd edition, vol. 2. *Commission of the European Communities Health Services Research Series no. 9*. Oxford: Oxford University Press.
- [15] Hübelová, D., Chromková-Manea, B. & Kozumplíková, A. (2021). Zdraví a jeho sociální, ekonomické a environmentální determinanty: teoretické a empirické vymezení. *Sociológia*. 53 (2), 119-146.
- [16] Hübelová, D., Ptáček, P. & Šlechtová, T. (2021). Demographic and socio-economic factors influencing health inequalities in the Czech Republic. *GeoSpace*. 15 (1) in press.
- [17] Institute of Health Information and Statistics of the Czech Republic. (2020). *Mortalita*. <https://reporting.uzis.cz/cr/index.php?pg=statisticke-vystupy--mortalita>
- [18] Ishizaka, A. & Nemery, P. (2013). *Multi-Criteria Decision Analysis*. UK: John Wiley & Sons, Ltd.
- [19] Kuncová, M. & Sekničková, J. (2020). Influence of the Different Transformation of the Minimization Criteria on the Result – the Case of WSA, TOPSIS and ARAS Methods. In: International Conference on Mathematical Methods in Economics 2020 (MME 2020) [online]. Brno, 09.09.2020 – 11.09.2020. Brno: Mendel University, 2020, pp. 332–338.
- [20] Mardani, A., Jusoh, A., Nor, K.MD., 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-Ekonomika Istraživanja*, 28(1), 516-571.
- [21] Ministry of Labor and Social Affairs of the Czech Republic. (2020). *Statistiky o trhu práce*. <https://data.mpsv.cz/web/data/statistiky> New York: Springer Science + Business Media Inc.
- [22] Novotná, H., Ponikelský, P., Slepíčka, A. & Šafařík, F. (2011). *Společnost a životní prostředí v regionálním rozvoji*. Praha: Vysoká škola regionálního rozvoje.
- [23] Ouředníček, M., Temelová, J. & Pospíšilová, L. (eds.). (2011). *Atlas sociálně prostorové diferenciace*. Praha: Karolinum.
- [24] Silva dos S. & Swerdlow, A., J. (1993). Thyroid cancer epidemiology in England and Wales: time trends and geographical distribution. *British Journal of Cancer*, 67: 330–340.
- [25] Whitehead, M. & Dahlgren, G. (2007). *Concepts and principles for tackling social inequities in health: Levelling up Part 1*. Copenhagen: WHO Regional Office for Europe. http://www.euro.who.int/__data/assets/pdf_file/0010/74737/E89383.pdf
- [26] WHO, (2019): Annual Report 2018. *Promoting Access to Safe, Effective, Quality and affordable essential medical products for all*. WHO, Geneva. <https://apps.who.int/iris/bitstream/handle/10665/324765/WHO-MVP-EMP-2019.03-eng.pdf>
- [27] WHO. (2003). *Atlas of health in Europe*. WHO, Copenhagen, Denmark: Regional Office for Europe, 1 atlas (vii, 112 p.).
- [28] WHO. (2008). *Atlas of health in Europe. 2nd ed., 2008*. WHO, Copenhagen, Denmark: Regional Office for Europe, 1 atlas (vii, 126 p.).
- [29] Wija, P., Bareš, P. & Žofka, J. (2019). *Atlas dlouhodobé péče*. Praha: Institut pro sociální politiku a výzkum, z. s.

Modelling of PX Stock Returns during Calm and Crisis Periods: A Markov Switching Approach

Michaela Chocholatá¹

Abstract. This paper deals with the modelling of the Czech stock market characterized by the weekly PX stock returns based on the Markov switching (MSW) approach. The analysed period spanning from April 8, 2007 to February 7, 2021 comprises both the “normal” calm and “turbulent” crisis periods. The two-regime MSW model thus enables to capture and specify the periods of bull and bear markets characterized by positive mean return-low volatility and negative mean return-high volatility periods, respectively. The analysis was enriched by consideration of the price/return – trading volume relationship, but it led neither to recognizable changes in values of estimated parameters nor to substantial differences in transition matrices. The presented results clearly confirmed the existence of several turbulent periods reflecting the worldwide financial and economic situation indicating the occurrence of crisis period with the probability of almost 0.14. However, the “normal” calm behaviour of returns occurred much more often and was proved to be more persistent in comparison to “turbulent” crisis period.

Keywords: PX index, stock returns, Markov switching model, bull and bear market

JEL Classification: G10, C22, C58

AMS Classification: 62M05, 62P20, 91B84

1 Introduction

The PX index is the official index of the Prague Stock Exchange and consists of the most actively traded blue chips of the Prague Stock Exchange [17]. The PX index is a free-floating price-weighted index calculated in CZK, the dividend yields are not included in its calculation. It was calculated for the first time on March 20, 2006 replacing the PX 50 and PX-D indices. The PX index continues in the development of the PX 50 index adopting its historical values. The PX 50 index, composed of 50 issues, was launched at April 5, 1994 with the opening value fixed at 1000 points. The number of basic issues of PX index has become variable since December 2001. The PX index is reviewed on a quarterly basis in order to maintain index quality [19]. Although nowadays the PX index contains issues of only 12 companies, it can serve as an indicator of the Czech economics [16].

In general, stock indices exhibit upward and downward trends reflecting impacts of various types of shocks and crises. Considering the ongoing Covid-19 pandemic, plenty of studies have been published to capture its effects on stock market returns across different countries and areas (see e.g., [2] and [13]). Very popular seems to be, to use the Markov switching (MSW) model enabling to distinguish different states of the world – calm periods with the normal behaviour of stock returns (bull market) and crisis periods with dramatically changing behaviour of stock returns characterised by high volatility and falling stock prices (bear market), see e.g., [7]. The MSW model for the Czech PX stock returns data was estimated e.g., by [12] and [14].

It is clear that the time-varying behaviour of stock returns is a natural part of trading and is linked to the arrival of new information and the subsequent reaction of market participants. Based on the new information, expectations regarding future market prices are being revised and this will also be reflected in the trading volumes. The analysis of stock returns can be thus further extended by consideration of the trading volume variable. Various approaches are known to analyse the relationship between stock prices/returns and trading volume – see e.g. [9] and [15]. Lamoureux and Lastrapes [11] supposed that volatility and trading volume are simultaneously and positively correlated, as they are a function of a stochastic variable defined as an information flow. This means that the arrival of new information on the market will simultaneously cause a change in volatility and trading volumes. However, as pointed out by Kalotychou and Staikouras [8], such a model does not explicitly exclude the possibility of different lags in price-volume relationship. The study of Wu [15] implements the Markov switching model for weekly returns of the Taiwan stock exchange index, including period January 2000 - June 2015, to study the relationship

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.

between stock returns and trading volume. Presented results documented different performance across regimes and analysed industries.

This paper attempts to capture the dynamic behaviour of the Czech stock market characterized by the weekly values of the PX stock index during the period April 8, 2007 – February 7, 2021. To identify the bull and bear regimes, the two-regime MSW model is used. The analysis is enriched with the consideration of the trading volume variable to study its impact on stock returns across analysed period.

The rest of the paper is organized as follows. Section 2 is devoted to methodology issues including the MSW model and the stock return – trading volume relationship, section 3 presents the data and empirical estimation results and section 4 concludes.

2 Methodology

Since the behaviour of financial time series can change quite dramatically during some periods of time, in recent years several time series models have been developed to capture the occurrence of different regimes (states) generated by a stochastic process². Literature, see e.g. [5], generally distinguishes two categories of regime-switching models – models with regimes determined by observable variables and models with regimes determined by unobservable variables. With regard to the empirical part of the paper, we will concentrate on the second category of models which supposes that the occurrence of the particular regime is determined by an unobservable stochastic process usually denoted as s_t . The most famous model in this category, the Markov switching model (MSW) in which the transitions between regimes are governed by a Markov process, was popularized by Hamilton [6]. The basic MSW model distinguishes only two states of the world corresponding to calm “positive mean-low volatility” periods and turbulent “negative mean-high volatility” periods, also known as bull and bear markets ([1], [3], [10], [14]). In case of two regimes, the variable s_t takes on two values, 1 and 2, i.e. if $s_t = 1$ the process is in regime 1 at time t and if $s_t = 2$, the process is in regime 2 at time t [4].

The dynamic behaviour of stock returns r_t can be in general described by a linear AR(p) model in both regimes [5]:

$$r_t = \phi_{0,s_t} + \phi_{1,s_t}r_{t-1} + \dots + \phi_{p,s_t}r_{t-p} + \varepsilon_t, \quad s_t = 1,2, \quad \varepsilon_t \sim N(0, \sigma_{s_t}^2) \quad (1)$$

where s_t is a random variable governed by a first-order Markov process. It means, that the regime s_t at any time t only depends on the regime at time $t - 1$, i.e. s_{t-1} . Transition probabilities of moving from one regime to the other are specified as follows:

$$P\{s_t = j | s_{t-1} = i\} = p_{ij} \quad (2)$$

$\{p_{ij}\}_{i,j=1,2}$ thus, denotes the transition probability for the two-regime Markov chain, i.e. probability that the regime i (at time $t - 1$) will be followed by the regime j (at time t). Since at any time, the variable should be in one of the two considered regimes, it should hold that

$$p_{i1} + p_{i2} = 1, \quad i = 1, 2 \quad (3)$$

The unconditional probabilities, i.e. probabilities that the process is in each of the regimes, are in case of two-regime MSW model as follows [5]:

$$P(s_t = 1) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}} \quad (4)$$

$$P(s_t = 2) = \frac{1 - p_{11}}{2 - p_{11} - p_{22}} \quad (5)$$

To capture the impact of change in the trading volume onto the stock returns behaviour, model (1) can be modified as follows:

² In general, financial analysts and researchers usually analyse the log return series (i.e. continuously compounded returns) instead of prices (which are usually non-stationary) using various approaches. This paper also applies this definition of returns - the log return series are denoted as r_t .

$$r_t = \phi_{0,s_t} + \phi_{1,s_t}r_{t-1} + \dots + \phi_{p,s_t}r_{t-p} + \alpha_{s_t}vol_{t-1} + \varepsilon_t, \quad s_t = 1,2, \quad \varepsilon_t \sim N(0, \sigma_{s_t}^2) \quad (6)$$

where vol_t denotes changes in trading volume (calculated as the first-difference of trading volume expressed in natural logarithm) and α_{s_t} represents corresponding parameter in individual regimes. Model (6) thus enables to capture different behaviour of returns enabling to estimate regime-specific parameters for mean, AR model, trading volume variable as well as changes in the variance between individual regimes. Markov switching model is estimated using the maximum likelihood techniques (for details see e.g., [5]).

3 Data and empirical results

The analysed data set consists of weekly data of the PX stock price index and corresponding trading volume spanning from April 8, 2007 to February 7, 2021 (i.e. 723 observations)³. The focus will be on PX log returns (i.e. continuously compounded returns) r_t which were calculated using $r_t = \ln(P_t) - \ln(P_{t-1})$, where symbol P_t denotes the closing price of the PX index on time t . Changes in the trading volume vol_t are calculated in a similar way, i.e. $vol_t = \ln(TV_t) - \ln(TV_{t-1})$ where TV_t is the trading volume on time t . The data were retrieved from the web-page [18], all the calculations were carried out in software EViews.

The weekly data of PX stock prices together with PX returns as well as changes in the trading volume are displayed in Figure 1. The data show that historical maximum of the PX index in October 2007 (1936 points) was followed by a sharp downward movement caused by the world financial crisis with minimum of 628 points reached in February 2009. After the stabilization of the financial crisis, the development on the Czech stock market in 2010 was already showing signs of stabilization. In 2011, however, the global recovery gradually slowed down and the economic situation was mainly affected by the spread of the sovereign debt crisis in the euro area. This growing uncertainty was also reflected in financial market developments. However, during the next period of 2012 – 2019, the PX index oscillated around its original opening value of 1000 points. As a result of a global Covid-19 pandemic outbreak, the PX index took a declining trend with a minimum of 690 points in March 2020 and some other swings in PX index values followed during the next months, as well. Corresponding PX returns show the evidence of volatility clustering since the large price movements tend to be followed by other large price movements and vice versa. High volatilities are visible especially during the financial crisis period in 2008 and during the ongoing Covid-19 pandemics. Changes in trading volume show clearly time-varying character during the whole analysed period.

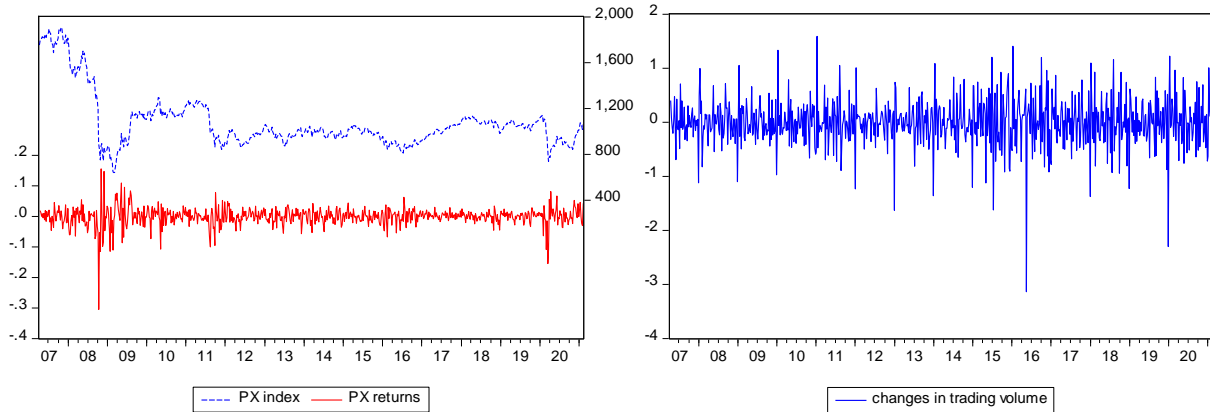


Figure 1 Weekly data of PX stock prices, PX stock returns (left) and changes in the trading volume (right)

The first four moments of the weekly PX returns and changes in the trading volume are shown in Table 1. The weekly percentage PX return was slightly negative (-0.1%) with volatility of 3.1%, The distribution is negatively skewed and clearly leptokurtic. Changes in the trading volume showed zero mean, volatility of 4.45% and negatively skewed distribution with higher kurtosis compared to normal distribution. The non-normality of the distributions for both series is also indicated by the values of the Jarque-Bera test statistics.

³ The period under consideration was determined by the availability of the trading volume data. Weekly data were used to eliminate the day-of-the-week effects.

	Mean	Std. Dev.	Skewness	Kurtosis	Jarque-Bera
r_t	-0.001	0.031	-1.635	20.023	9039.515
vol_t	0.000	0.445	-0.640	7.890	768.6752

Table 1 The first four moments of PX returns (r_t) and changes in the trading volume (vol_t) together with the Jarque-Bera test statistics

To capture the changing behaviour of PX returns during the analysed period, the two-regime MSW models (1) and (6) were estimated however considering the regime-invariant AR(p) process (Model 1) and regime-invariant volume variable (Model 2)⁴. Based on the Ljung-Box Q-statistics, four lags were used to explain the returns behaviour⁵. The estimated parameters – regime specific mean and variance, as well as regime-invariant AR(4) parameters and volume parameter α are presented in Table 2.

	Parameter	Model 1	Model 2
Regime 1	c_1	-0.011*	-0.011*
	$\log(\sigma_1)$	-2.747***	-2.750***
Regime 2	c_2	0.001	0.001
	$\log(\sigma_2)$	-3.955***	-3.959***
Common	ϕ_1	0.029	0.026
	ϕ_2	0.003	0.008
	ϕ_3	-0.087**	-0.084**
	ϕ_4	0.049	0.049
	α	-	-0.003*
Transition matrix parameters	p_{11}	2.432***	2.421***
	p_{21}	-4.341***	-4.316***

Table 2 Estimation results of the two-regime Markov switching model (Model 1 – without trading volume, Model 2- with trading volume). Note: Symbols ***, ** and * denote rejection of the null hypothesis at the 1%, 5 % and 10% significance level, respectively.

The statistical significance of estimated parameters indicates that not all estimated parameters were statistically significant. The estimated parameters for both models (Model 1 and Model 2) were very similar in values and signs whereas the changes in the trading volume had surprisingly negative impact on returns. Regime 1 corresponds to a bear market since it can be characterized by negative mean return and high volatility (0.064); whereas regime 2, known as a bull market, shows positive mean return and low volatility (0.019). The transition probabilities and expected durations gathered in Table 3 imply that the probabilities of staying in regime 1 and regime 2 are 0.919 and 0.987, respectively (for Model 1) and very similar for Model 2 with values of 0.918 and 0.987, respectively. Regime 2 characterizing the “normal” calm behaviour of returns is thus more persistent than turbulent regime 1. The transition probability from the high volatility regime 1 to the low volatility regime 2 of 0.081⁶ is higher in comparison to transition probability of 0.013 from regime 2 to regime 1. The corresponding expected durations in a regime are in case of Model 1 approximately 12.384 and 77.766 weeks, respectively and 12.262 and 75.859 weeks, respectively for Model 2.

	Model 1			Model 2		
		1	2		1	2
Constant transition probabilities	1	0.919248	0.080752	1	0.918451	0.081549
	2	0.012859	0.987141	2	0.013182	0.986818
Constant expected durations		1	2		1	2
		12.38354	77.76609		12.26249	75.85928

Table 3 Transition probabilities and expected durations (Model 1 – without trading volume, Model 2- with trading volume).

⁴ Since the restriction that the parameters ($\phi_1, \dots, \phi_p, \alpha$) are the same across both regimes could not be rejected, it was supposed that they are regime-invariant.

⁵ The results are available from the author upon request.

⁶ Value 0.081 corresponds to Model 1, the probability attributable to Model 2 is 0.082.

Furthermore, another important information is given by probabilities with which each regime occurs at each time t [5]. The MSW smoothed regime probabilities of being in regime 1 at time t are shown in Figure 2. The presented results for Model 1 and Model 2 show high probabilities (more than 0.5) of being in regime 1 (bear market) for 90 and 91 observations, respectively. Based on this result we are able to identify 5 periods of turbulence⁷ for the data under consideration:

- January 6, 2008 – January 27, 2008 - deep fall in shares was caused by panic concerning the subprime mortgage crisis in the USA and hit stock markets worldwide,
- September 7, 2008 – August 8, 2009 – global financial crisis period,
- May 2, 2010 – May 23, 2010 – flash crash attributed to automated algorithmic trades,
- July 31, 2011 – December 25, 2011 - outbreak of the European debt crisis and its possible effects,
- February 2, 2020 – April 12, 2020 – outbreak and spread of the Covid-19 pandemic.

The MSW model seems to be highly adequate to specify in which regime the PX return should be in, since the probability in Figure 2 tends to spend most of time either close to 0 or close to 1 [4]. Valuable information is furthermore given by unconditional probabilities calculated according to (4) and (5). Using Model 1, the process is in regime 1 with the probability of 0.137 and in regime 2 with probability of 0.863. Considering Model 2 with trading volume included, the probability of being in each of the regimes is 0.139 and 0.861, respectively. Indeed, the smoothed regime probabilities captured in Figure 2 clearly confirm that regime 2 occurs much more often in comparison to regime 1.

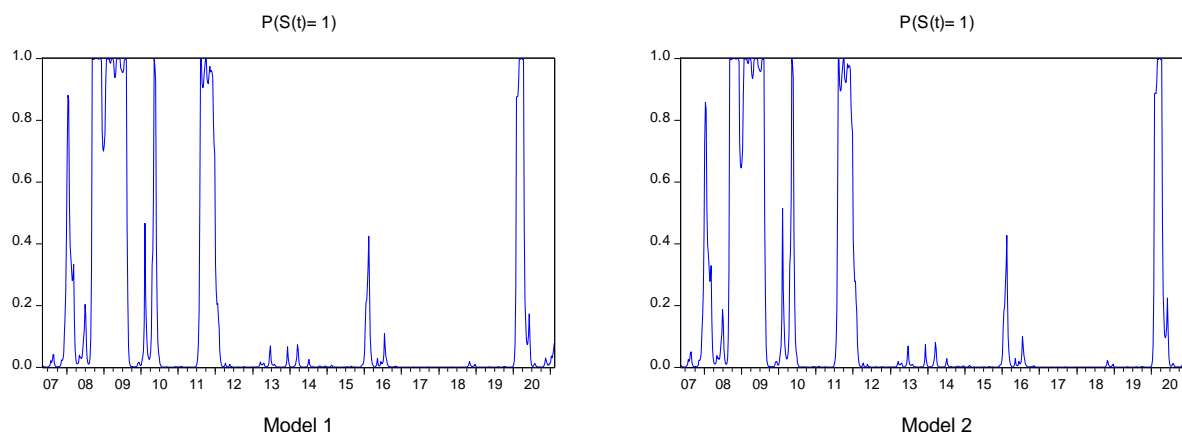


Figure 2 MSW smoothed regime probabilities for the regime 1 (Model 1 – without trading volume, Model 2- with trading volume).

4 Conclusion

This paper employed the MSW model to specify when the Czech stock market characterized by the PX returns entered the periods of high volatility. Considering the dynamic behaviour of the weekly PX returns during the period spanning from April 8, 2007 to February 7, 2021 it was found out that the identified turbulent periods could be attributed to the well-known crises' periods (subprime mortgage crisis in the USA, global financial crisis period, flash crash attributed to automated algorithmic trades, outbreak of the European debt crisis and its possible effects, outbreak and spread of the Covid-19 pandemic). The analysis was enriched by the consideration of changes in the trading volume, but it was proved that it led neither to recognizable changes in values of estimated parameters nor to substantial differences in transition matrices. The presented results show that the MSW approach can successfully describe the weekly PX returns which were drawn from a mixture of two normal distributions. The transition process is coupled with the bull-bear market switches directed by clearly time-varying probabilities of being in one of the regimes. Furthermore, the results confirmed that the “normal” calm behaviour of returns is more persistent than the turbulent one. To study the dynamic behaviour of the Czech stock market in relation to other Central European stock markets and/or the US stock market in order to investigate the potential contagion effects will be a challenging issue for the future research.

⁷ In case of Model 2, the week of February 7, 2010 was in addition identified to be more likely in regime 1 with corresponding smoothed probability of 0.514.

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. 1/0211/21, "Econometric Analysis of Macroeconomic Impacts of Pandemics in the World with Emphasis on the Development of EU Economies and Especially the Slovak Economy".

References

- [1] Almonares, R.A.L. (2019). Markov Switching Model of Philippine Stock Market Volatility. *DLSU Business & Economics Review*, 29 (1), 24–30.
- [2] Baker, S. R., Bloom, N., Davis, S.J., Kost, K.J., Sammon, M.C. & Viratyosin, T. (2020). The unprecedented stock market impact of Covid-19. [Online]. Available at: https://www.nber.org/system/files/working_papers/w26945/w26945.pdf [cited 2021-01-24].
- [3] Bialkowski, J. (2004). Modelling returns on stock indices for western and central European stock exchanges - A Markov switching approach. *South Eastern Europe Journal of Economics*, 2, 81–100.
- [4] Brooks, C. (2008). *Introductory Econometrics for Finance*. 2nd ed. Cambridge: Cambridge University Press.
- [5] Franses, P.H. & van Dijk, D. (2000). *Non-linear time series models in empirical finance*. Cambridge: Cambridge University Press.
- [6] Hamilton, J.D. (1989). A New approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57 (2), 357–384.
- [7] Just, M. & Echaust, K. (2020). Stock market returns, volatility, correlation and liquidity during the COVID-19 crisis: Evidence from the Markov switching approach. *Finance Research Letters*, 37 (2020), 101775.
- [8] Kalotychou, E. & Staikouras, S.K. (2009). An Overview of the Issues Surrounding Stock Market Volatility. In Gregoriou, G. N.: *Stock Market Volatility*. New York: CRC Press, Taylor & Francis Group, 3–29.
- [9] Karpoff, J. M. (1987). The Relation between Price Changes and Trading Volume: A Survey. *Journal of Financial and Quantitative Analysis*, 22, 109–126.
- [10] Koško, M. (2006). An Application of Markov-switching Model to stock Returns Analysis. *Dynamic Econometric Models*, 7, 259–268.
- [11] Lamoureux, C. & Lastrapes, N. (1990). Heteroscedasticity in stock return data: volume versus GARCH Effects. *The Journal of Finance*, XLV (1), 221–229.
- [12] Linne, T. (2002). A Markov Switching Model of Stock Returns: An Application to the Emerging Markets in Central and Eastern Europe. In: Charemza, W.W. & Strzała, K. (eds). *East European Transition and EU Enlargement. Contributions to Economics*. Heidelberg: Physica, 371–384.
- [13] Liu, H.Y., Manzoor, A., Wang, C.Y., Zhang, L. & Manzoor, Z. (2020). The COVID-19 Outbreak and Affected Countries Stock Markets Response. *International Journal of Environmental Research and Public Health*, 17(8), 2800.
- [14] Reiff, M. (2018). *Analýza parametrov skrytého Markovovho modelu finančných časových radov*. Bratislava, Edícia: Habilitačné a inauguračné prednášky.
- [15] Wu, J.-T. (2016). Stock Return and Trading Volume - An Application of the Markov Switching Model. *Asian Business Research*, 1(1), 55–69.
- [16] <https://finex.cz/index/px-index/> [cited 2021-02-14].
- [17] <https://www.pse.cz/en/indices/description-of-indices> [cited 2021-02-14].
- [18] <https://stooq.com/q/d/?s=^px> [cited 2021-02-14].
- [19] <https://www.wienerborse.at/indizes/index-kooperationen/prague-stock-exchange/px-px-tr-rules/> [cited 2021-02-14].

The Performance Assessment of Different Types of Investment Funds Using Markowitz Portfolio Theory

Zuzana Chvátalová¹, Oldřich Trenz², Jitka Sládková³

Abstract. The current scientific state of knowledge deals with the Markowitz Mean-Variance model, which modifies various identified aspects (most often including risk, infections, cardinality problems and others). They modify the initial algorithm and then apply it to selected funds. Our aim was not to modify the Markowitz model, but to map the performance of individual types of funds using computer technology. The original intention was to compare the performance of social responsible investments (SRI) funds with others and to find out if they are economically interesting for the investors. Due to a lack of consistent data, the research team was only able to assess bond, equity and mixed assets funds. Five-year periods were chosen that showed consistent results (compared to one-year and research results). The data obtained was limited to a minimum of 5 funds in the portfolio, a minimum proportion of 5% for the fund in the portfolio, and 1000 possible variations of portfolios for the given type of fund were always created. These limits were applied to simulate the possible real decision-making of the investor on the financial market. In total, 550 Markowitz Bullets were examined, from which 8 key conclusions were drawn.

Keywords: economic and mathematical modelling, investment portfolio management, Markowitz theory.

JEL Classification: G11

AMS Classification: 91B28

1 Introduction

The idea to reduce risk by investing in more assets by diversification was used intuitively since antiquity. How to diversify any portfolio to maximize returns with the required level of risk is a question for all of investors. This question is more relevant every day, as the increase in capital globally has a growing trend, and the amount of free money in circulation also increases. The classical model for portfolio optimization is the Markowitz mean-variance model [15], [10]. Markowitz in his theory practically shows how diversification works. The return on assets is not perfectly dependent. As a result, a multi-asset portfolio has a better return on risk than individual assets alone. It is common knowledge that equity funds are riskier (but with higher returns) than bond funds. Roebbers at all [14] states the Markowitz model is theoretically very strong but has received a lot of criticism because the setting is not realistic; thus, it is important to extend the simple Markowitz model with cardinality and quantity constraints (see [4]). It is desirable to try to capture the imperfections of the Markowitz model, such as cardinality and quantity constraints [15]. On the other hand, it is the simplicity of the model that, together with high-performance computing and the knowledge of historical performance of investment funds, can contribute to a better orientation of potential investors and to their questions: where and how to diversify their portfolio to achieve the highest possible return [10].

The paper aims to introduce to map the performance of individual types of funds using computer technology.

¹ 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.

² Oldřich Trenz, Mendel University in Brno/ Faculty of Business and Economics/Department of Informatics, Zemědělská 1, 613 00 Brno, Czech Republic, oldrich.trenz@mendelu.cz.

³ Jitka Sládková, Sting Academy/ Department of Economics and Management, M. Horákové 2031/1, 602 00, Brno, Czech Republic, sladkova.jit@gmail.com.

2 Material and Methods

2.1 Research Data

It has been created groups of different portfolios. The main criteria were the type of funds and the country in which they are traded. As national economic policies affect the securities markets, the goal was set to find out in which types of funds in which country it had a greater potential to invest. It would not be possible to approach this ambitious goal without using computer algorithms.

The data was extracted from the Morningstar web portal [12]. Most funds reported their performance in EUR (the other most often in USD or GBP). However, the currency was not decisive as only the daily percentage changes in performance were monitored

- The total amount of single data: 7 152 369
- The total number of funds: 2284
- It was created 116 categories in total, but only 75 of them and enough funds for our requirements to compute the Markowitz portfolios (this means at least 5 funds per category)
- In total the 576 Markowitz distribution was calculated, this gives us 576,000 portfolios we had reviewed.
- The period from 1.8.2008 to 29.10.2018 was covered
- The data has 111 MB

The following fund categories were selected to observe: alternatives; equity; bonds; mixed assets; commodity; money markets.

1-year and 5-year time series were selected between years 2008 and 2018 (different combinations) in different fund category (not mixed categories). However, the one-year outputs are significantly biased compared to the five-year ones, therefore the long-term ones were subjected to further examination, from more than 20 different countries. Unfortunately, due to the absence of some time series, only equity funds, bonds and mixed assets were examined.

Based on the obtained data, 1630 possible Markowitz portfolios were created, differing in the date, place and type of investment funds. However, only 576 of them had at least 5 different funds (less funds in portfolio was not taken in account). The others were discarded due to the absence of a continuous data series, which could then be worked so that the data could be compared comparatively.

Although we research team worked with a huge amount of data at the beginning, the final research shows only partial conclusions from selected countries and periods. The reason is incomplete time series of performances of individual funds. For this reason, outputs separately from measured countries were examined. In addition, it has been shown that almost all calculations regarding annual values are greatly distorted by extreme fluctuations in individual portfolios. For this reason, only 5-year periods have been researched

Type of Funds	Period	Countries Examined
Bonds	2009–2014	Austria, Belgium, Czech Republic, Estonia, Hungary, Chile, Latvia, Liechtenstein, Netherlands, Peru, Poland, Singapore, Switzerland
	2010–2015	Austria, Belgium, Czech Republic, Estonia, Hungary, Chile, Latvia, Liechtenstein, Netherlands, Peru, Poland, Singapore, Switzerland
	2011–2016	Austria, Belgium, Czech Republic, Estonia, Hungary, Chile, Latvia, Liechtenstein, Malta, Netherlands, Peru, Poland, Singapore, Switzerland
	2012–2017	Austria, Belgium, Czech Republic, Estonia, Hungary, Chile, Latvia, Liechtenstein, Malta, Netherlands, Peru, Poland, Singapore, Switzerland
	2013–2018	Austria, Belgium, Czech Republic, Estonia, Hungary, Chile, Latvia, Liechtenstein, Malta, Netherlands, Peru, Poland, Singapore, Switzerland
Equity	2009–2014	Bahrain, Bulgaria, Estonia, Finland, Hungary, Greece, Lithuania, Netherlands, Norway, Peru, Slovakia, Switzerland
	2010–2015	Bahrain, Bulgaria, Estonia, Finland, Hungary, Greece, Liechtenstein, Lithuania, Netherlands, Norway, Peru, Singapore, Slovakia, Switzerland
	2011–2016	Bahrain, Bulgaria, Estonia, Finland, Hungary, Greece, Liechtenstein, Lithuania, Netherlands, Norway, Peru, Singapore, Slovakia, Switzerland
	2012–2017	Bahrain, Bulgaria, Estonia, Finland, Hungary, Greece, Liechtenstein, Lithuania, Netherlands, Norway, Peru, Singapore, Slovakia, Switzerland

	2013–2018	Bahrain, Bulgaria, Estonia, Finland, Hungary, Greece, Liechtenstein, Lithuania, Netherlands, Norway, Peru, Singapore, Slovakia, Switzerland
Mixed Assets	2009–2014	Austria, France, Greece, Italy, Spain, United Kingdom
	2010–2015	Austria, France, Greece, Italy, Spain, United Kingdom
	2011–2016	Austria, France, Greece, Italy, Spain, United Kingdom
	2012–2017	Austria, France, Greece, Italy, Spain, United Kingdom
	2013–2018	Austria, France, Greece, Italy, Spain, United Kingdom

Table 1 The final examined data shows the table below.

2.2 Portfolio Theory and Modern Investment Theory

Portfolio theory is an economic discipline examining a suitable portfolio structure. According to Haugen's "Modern Investment Theory", delays in responding to new information are minimal in efficient markets [7]. Other important assumptions effective market is the random behavior of changes in market prices. The fundamentals of portfolio theory were described by J. Hicks in "Application of Mathematical Methods to the Theory of Risk " in 1934. Hicks [8] suggests that investors make decisions in investment decisions using statistical characteristics of the probability distribution of investment returns. The Portfolio Theory was established by Harry Markowitz in 1952. His findings called "Portfolio Selection" (1952) were published in The Journal of Finance. Seven years later, in 1959, the key elements of Modern Portfolio Theory (MPT) were published in his book Portfolio Selection: Efficient Diversification (1959) [11],[10], [2].

Shortly let's mention the basic idea: An important prerequisite for Markowitz's portfolio theory (Markowitz's Mean-Variance Model) is that securities yields can be viewed as random variables. Expected return on the portfolio can be calculated as a weighted average of the expected return on each asset in the portfolio [1]. The expected portfolio risk is measured using the standard deviation. The investors behave rationally [3].

By varying the possible portfolios, we can get a graphical output, forming the so-called Markowitz Bullet⁴ and depicting the so-called Effective Frontier (Effective Set). Knowing the above, and having the opportunity to simulate random selections of different portfolios, according to Markowitz's theory, it is possible to determine which of the fund types are more acceptable for a potential investment [6], [17]. The investors seek for the Efficient Frontier (Efficient Set), more in [16], [15]. Other facts concerning the Markowitz model (e.g. convex quadratic programming model, cardinality constraint, effect of genetic algorithms etc.) are discussed in a number of publications as Chang at all [5] and [3].

2.3 Algorithms used

Two open-source programs were used, namely the Jupyter notebook and the Pandas. The input for the calculation is the time series of outputs of individual funds by days in different time periods. Subsequently, the daily percentage change is calculated.

```
DataFrame.pct_change(self, periods=1, fill_method='pad', limit=None, freq=None, **kwargs) [9]
```

Percentage change between the current and a prior element.

Computes the percentage change from the immediately previous row by default. This is useful in comparing the percentage of change in a time series of elements. Example: [90, 91, 85] -> [0.011111, -0.065934].

Then the Monte Carlo simulation to run 1000s of runs of different randomly generated weights for the individual stocks were used. Finally, the calculation of the expected return, expected volatility and Sharpe Ratio for each of the randomly generated portfolios were done.

The limit for the calculation was always 1000 random distributions for a given portfolio, with the minimum share per fund was set at 5%. This value was determined on the basis of a realistic estimate of the potential investor's behavior. The goal is not to find the ideal machine combinations for portfolios that are far from human reasoning. The aim is to simulate possible investor decisions that could potentially be made in given securities markets.

The script was adopted from [13], see below:

⁴ Note: The Markowitz Bullet is related to shaping (drawing) the Markowitz efficient frontier. This is well and detailed described, for example, in the text by Ian Rayner, 2019 (<https://www.raynergobran.com/2019/02/the-markowitz-bullet-a-guided-tour/>), Ian Rayner is originator of Rayner Gobran LLC (<https://www.raynergobran.com/>).

```

returns = data.pct_change()
#calculate mean daily return and covariance of daily returns
mean_daily_returns = returns.mean()
cov_matrix = returns.cov()

#set up array to hold results
results = np.zeros((3,num_portfolios))

for i in iter(range(num_portfolios)):
    #select random weights for portfolio holdings
    weights = np.random.random(len(data.columns))
    #rebalance weights to sum to 1
    weights /= np.sum(weights)

    #calculate portfolio return and volatility
    portfolio_return = np.sum(mean_daily_returns * weights) * 252
    portfolio_std_dev = np.sqrt(np.dot(weights.T,np.dot(cov_matrix, weights))) * np.

    #store results in results array
    results[0,i] = portfolio_return
    results[1,i] = portfolio_std_dev
    #store Sharpe Ratio (return / volatility) - risk free rate element excluded for
    results[2,i] = results[0,i] / results[1,i]

#convert results array to Pandas DataFrame
results_frame = pd.DataFrame(results.T,columns=['ret','stdev','sharpe'])

```

3 Results

The outcome of our research is a graphical representation of sets of Markowitz portfolios. Each output is limited either by time – by a year or by a 5-year period. Individual types of funds that are traded in individual countries have been examined.

Findings: Every investor is looking for the best way to invest. In today's globalized world, where it is possible to invest in different countries, there is a common question of where to invest (in what region) and in what type of investment funds. This is what our research focused on. The types of funds traded in selected countries with the best results are presented below (due to the scope of the paper, we present only selected ones and in brief). It should be added that we based our research on the 5-year development of individual funds, their percentage changes, and these values were averaged to six decimal places. This gives only a summary picture of the performance of each type of investment fund by country, not an examination of the sub-development of selected funds (in the order of units), as it is the case with most research conducted in the Markowitz model. It is the combination of return and volatility (the C value, the color scale shown to the right of the graph) shows what values a potential investor may be in.

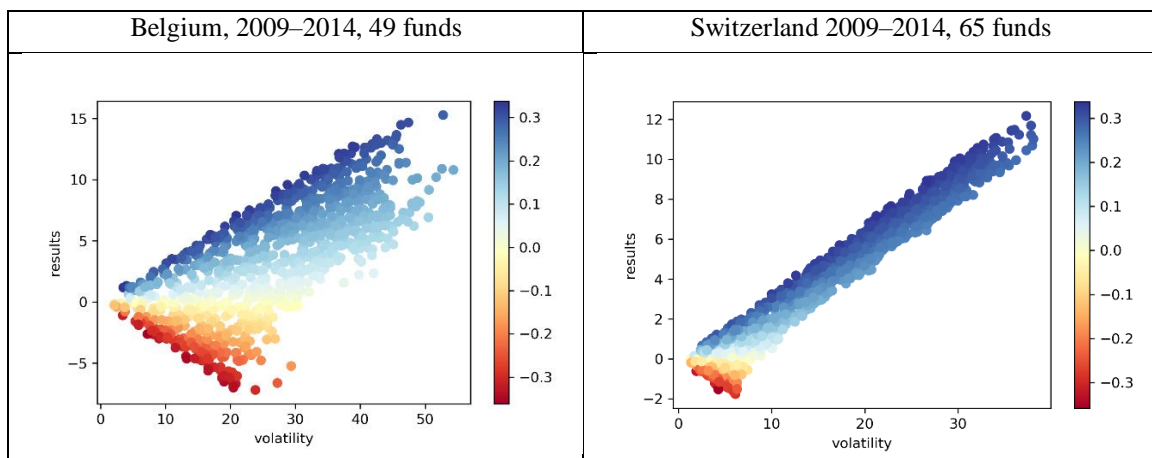


Figure 1 5- years Performance Bond Funds Markowitz Portfolios Results

Mixed Assets funds are available mainly in in the developed economies of Western Europe, where they are traditional investment funds.

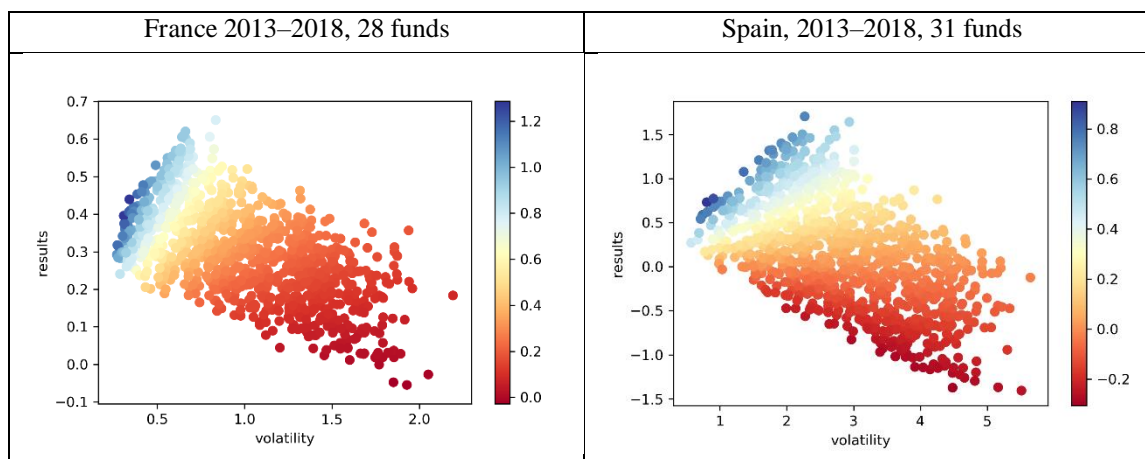


Figure 2 5- years Performance Mixed Assets Funds Markowitz Portfolios Results

The limitation of this model is, that not all the funds are covered. Only the funds with daily 5 years performance was considered. Furthermore, it was calculated with average values. On the other hand, the research team want to get there research as close to praxis as possible. The assumption was that the investor follows the historical development of individual markets and funds. Our potential investor only monitors stable funds, i.e. those that publish their results daily. If we wanted to create a basis for the decision-making of a risk-seeking investor, we would have to include those funds that do not have a 5-year history. However, this has not been addressed in the research.

After studying all the outputs, the research team made the following key conclusions: Annual performances often reach extreme values; 5-year performances show more stable indicators Mixed Assets funds are able to generate higher value but are riskier; Developed markets are less loss-making than young stock markets (non-G20); The effective set and typical Markowitz Bullet can be seen for portfolios containing more than 20 and 30 different funds, respectively; For under-developed countries under 20 funds, mainly the convex downward part of the Markowitz Bullet shows; In developed countries with more than 20 funds, there is a dominant upward concave part of the Markowitz Bullet.

4 Conclusions

A differentiation of the portfolio is a common strategy of every investor. How to break down portfolio diversification into what types of assets, in what proportion and in which country is a key question of success in the financial markets. The theory states that the financial markets, in terms of their awareness, are closest to perfectly competitive markets. In general, the historical values of the various types of investment funds are well known and readily available, and they can be used to estimate their future development. Of course, there is no internal information on the actual condition of the funds, which is an unavoidable investment risk. Nevertheless, the statistics help to estimate very closely the future development of investment funds based on historical data. This fact, together with Markowitz's model theory, was the main starting point of our investigation. In the initial phase of our research, we focused primarily on funds such as alternatives, equity, bonds, mixed assets, commodities, and money markets.

1620 possible Markowitz portfolios were created using advanced computational tools, each containing 1,000 possible investment portfolio layouts. The limitation of our investigation was mostly the availability of complete time series of examined funds. We used daily percentage changes in all available funds by country and fund type. From the point of view of the availability of the data we required 5 years of daily values (different periods were examined, always for 5 years since 2009), we were able to examine only bonds, equity and mixed assets funds. The whole research was aimed at maximum practicality and simplicity of the model. Therefore, in addition to the 5-year daily values of individual funds (averaged percentage changes, rounded to 6 decimal places), we also based on other limitations. Those were minimum 5% investment in one fund, a portfolio containing at least 5 funds, and exactly 1,000 portfolio combinations. This limits mean that we expect a more conservative type of investor based on well-established and historically known investment funds. Steady yields, with less frequent negative values, have been shown by bond funds, while equity funds are generally more dispersed. Mixed Assets have known values in Western European countries.

Despite the initial ambitions of the research team, a large number of Markowitz Bullets have been generated, namely 550, from which 8 key conclusions have been drawn. In the future, we would like to examine the general performance of funds by type. We want to focus on examining the economic efficiency of SRI funds compared to other types. (That was also our original research intention.) Research has shown that to this end it will be necessary to better define the period of time for the relevant data, only 3 years (instead of 5 years), and we also recommend creating a rather triple number instead of 1000 Markowitz combinations.

Acknowledgements

This paper was supported by grant FP-S-20-6376 of the Internal Grant Agency at Brno University of Technology.

References

- [1] Anagnostopoulos, K. P., & Mamanis, G. (2011). The mean–variance cardinality constrained portfolio optimization problem: an experimental evaluation of five multiobjective evolutionary algorithms. *Expert Systems with Applications*, 38(11), 14208–14217.
- [2] Baule, R. Korn, O. Kuntz, L. (2019) Markowitz with regret. *Journal of Economic Dynamics and Control*, 103, 1–24. ISSN 0165-1889, <https://doi.org/10.1016/j.jedc.2018.09.012>. [cit. 2020-12-21] Available at: <http://www.sciencedirect.com/science/article/pii/S0165188919300600>.
- [3] Bienstock, D. (1996). Computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming*, 74(2), 121–140.
- [4] Cesarone, F., Scozzari, A., & Tardella, F. (2013). A new method for mean-variance portfolio optimization with cardinality constraints. *Annals of Operations Research*, 205(1), 213–234.
- [5] Chang, T.-J., Meade, N., Beasley, J. E., & Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, 27(13), 1271–1302.
- [6] Elton, E. J. (2007) *Modern portfolio theory and investment analysis*. 7th ed. New York: John Wiley & Sons, 728 p. ISBN 978-0-470-05082-8.
- [7] Haugen, H. (2000) *Modern Investment Theory*. Pearson, 5th edition, 680 p. ISBN 978-0130191700.
- [8] Hicks, J. R. (1934) *Application of Mathematical Methods to the Theory of Risk*, *Econometrica*.
- [9] Jupyter – Python. (2020) [cit. 2020-12-21] Available at: <https://github.com/pandas-dev/pandas/blob/v0.25.3/pandas/core/generic.py#L10421-L10435>.
- [10] Markowitz model, *The Engineering Economist*, DOI: 10.1080/0013791X.2019.1636439.
- [11] Markowitz, H. (1959) *Portfolio Selection: Efficient Diversification of Investment*. John Wiley & Sons, New York.
- [12] Morning Star. (2020). [cit. 2020-12-21] Available at: <https://www.morningstar.co.uk/>.
- [13] Python Data Analysis Library. (2020) [cit. 2020-12-21] Available at: <https://pandas.pydata.org/>.
- [14] Python for finance. (2020) *Investment portfolio optimization with Python*. [cit. 2020-12-21] Available at: <https://www.pythonforfinance.net/2017/01/21/investment-portfolio-optimisation-with-python>.
- [15] Roebers, L.M., Selvi, A., Vera, J. C. (2019) Using column generation to solve extensions to the Markowitz model, *The Engineering Economist*, 64:3, 275-288, DOI: 10.1080/0013791X.2019.1636439.
- [16] Streichert, F., Ulmer, H., & Zell, A. (2004). Evolutionary algorithms and the cardinality constrained portfolio optimization problem. *Operations Research Proceedings*, 2003, 253–260.
- [17] Way, R., Lafond, F., Lillo, F., Panchenko, V., Farmer, J.D. (2019) Wright meets Markowitz: How standard portfolio theory changes when assets are technologies following experience curves, *Journal of Economic Dynamics and Control*, 101, 211–238, ISSN 0165-1889, <https://doi.org/10.1016/j.jedc.2018.10.006>. [cit. 2020-12-21] Available at: <http://www.sciencedirect.com/science/article/pii/S0165188919300181>.

SBM models in data envelopment analysis: A comparative study

Josef Jablonský¹

Abstract. Data envelopment analysis (DEA) is a traditional modelling tool for relative efficiency and performance evaluation of a set of decision-making units. One of the classes of DEA models measures the level of efficiency using slack variables – slacks-based measure (SBM) models. There have been proposed various models of this class for measuring efficiency and super-efficiency by various authors in the past. The paper compares the typical representatives of SBM efficiency and super-efficiency models and discusses their properties. Typical SBM models do not consider integer inputs and/or outputs. The paper extends typical SBM models to include integer conditions and analyses the differences in their results compared to the non-integer models.

Keywords: Data envelopment analysis, slacks-based measure, SBM model, efficiency, integer programming

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

DEA models were introduced by Charnes et al. (1978) as a tool for relative efficiency and performance evaluation of the set of decision-making units (DMUs). The model evaluates inputs (to be minimized in the typical case) and outputs (to be maximized) of the DMU under evaluation within the production possibility set defined by the linear combination of the DMUs in the set. The envelopment form of their input-oriented model is formulated as follows:

$$\begin{aligned}
 &\text{Minimize} && \theta_q \\
 &\text{subject to} && \sum_{i=1}^n x_{ij} \lambda_i + s_j^- = \theta_q x_{qj}, && j = 1, \dots, m, \\
 & && \sum_{i=1}^n y_{ik} \lambda_i - s_k^+ = y_{qk}, && k = 1, \dots, r, \\
 & && \lambda_i \geq 0, s_j^- \geq 0, s_k^+ \geq 0, && i = 1, \dots, n, j = 1, \dots, m, k = 1, \dots, r,
 \end{aligned} \tag{1}$$

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} , $i = 1, \dots, n$, $k = 1, \dots, r$, λ_i , $i = 1, \dots, n$, are the weights of the DMUs, s_j^- , $j = 1, \dots, m$, s_k^+ , $k = 1, \dots, r$, are slack variables, and DMU _{q} is the unit under evaluation. The optimal value $\theta_q = 1$ indicates that this unit is on the efficient frontier, and it is at least weakly efficient. The value $\theta_q < 1$ indicates inefficiency. It is a radial measure of efficiency because all inputs of the model must be reduced by θ_q to reach the efficient frontier.

Since 1978, when the first DEA model was formulated, many modifications of the traditional model (1) have been proposed by various authors. Very popular is the group of models that measure the level of efficiency using slack variables only. This paper aims to review the main representatives of slacks-based DEA models and discuss their properties.

Section 2 contains the formulation of typical slacks-based models. Section 3 deals with slacks-based super-efficiency models that allow ranking of DMUs identified as efficient by traditional models. Both sections 2 and 3 show the extension of the models by integer constraints and discuss how these constraints affect the results. The next section of the paper contains numerical illustration. Finally, the study concludes by discussion and possible research directions.

¹ Prague University of Economics and Business, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, Praha 3, Czech Republic, e-mail: jablon@vse.cz

2 Slacks-based DEA models

The radial input-oriented model (1) considers proportional input reductions while keeping the current level of outputs. On the contrary, the radial output-oriented models increase outputs by considering the current level of inputs. Slacks-based DEA models consider in efficiency evaluation both input and output variables simultaneously. Charnes et al. (1985) formulated the following model that is often denoted as the additive DEA model:

$$\text{Maximize} \quad \sum_{j=1}^m s_j^- + \sum_{k=1}^r s_k^+ \quad (2)$$

$$\begin{aligned} \text{subject to} \quad & \sum_{i=1}^n x_{ij} \lambda_i + s_j^- = x_{qj}, & j = 1, \dots, m, \\ & \sum_{i=1}^n y_{ik} \lambda_i - s_k^+ = y_{qk}, & k = 1, \dots, r, \\ & \lambda_i \geq 0, s_j^- \geq 0, s_k^+ \geq 0, & i = 1, \dots, n, j = 1, \dots, m, k = 1, \dots, r, \end{aligned} \quad (3)$$

The model (2) - (3) returns the optimal objective function value equal to 0 for the efficient DMUs, and the value greater than 0 for the inefficient units. The main drawback of model (2) - (3) is the incomparability of slacks' units included in the objective function. The value of the objective function cannot be explained in any way. Moreover, the efficiency score (optimal value of the objective function) is not invariant on the scale used. A possible solution is to normalize the input/output values by dividing by their maximum. Another possibility is to use a weighted slacks-based model proposed by (Ali et al., 1995). This model has the same set of constraints (3), but its objective function is the following:

$$\text{Maximize} \quad \sum_{j=1}^m w_j^- s_j^- + \sum_{k=1}^r w_k^+ s_k^+ \quad (4)$$

where w_j^-, w_k^+ are the positive weights of the slacks. The problem of this model may be the appropriate setting of the weights, and again, the efficiency score can just hardly be explained to decision-makers. Both unweighted and weighted additive models can be extended by the constraints that ensure the assumptions about the returns to scale – the sum of lambda variables is unrestricted/equal to 1/greater or equal than 1/lower or equal than 1 for the assumption of constant/variable/non-decreasing/non-increasing returns to scale.

The main drawbacks of the previous two models are solved by the model introduced by Tone (2001). This model is usually known as SBM (slacks-based measure) model. It belongs to one of the very often used DEA models at all. The SBM model is based on minimization of all slacks, but its objective function is formulated as follows:

$$\text{Minimize} \quad \rho_q = \frac{1 - \frac{1}{m} \sum_{j=1}^m (s_j^- / x_{qj})}{1 + \frac{1}{r} \sum_{k=1}^r (s_k^+ / y_{qk})}. \quad (5)$$

The set of constraints of the SBM model is the same as in additive models. The model is not linear but can be moved to a linear model easily. SBM model returns objective function $\rho_q = 1$ for efficient units and $\rho_q < 1$ for the inefficient ones. The advantage of this model in comparison with the additive models is in the following:

- Efficiency scores are invariant on the measurement scales of inputs and or outputs,
- Efficiency scores are decreasing functions of all slacks variables, i.e. increasing/decreasing of any input/output leads to a decreasing of the efficiency score of the unit under evaluation.

The model (2), (5) is not linear in its objective function but can easily be linearized by Charnes-Cooper transformation. The unit under evaluation is efficient if $\rho_q = 1$; lower values indicate inefficiency. As in the previous case, the SBM model can be extended by returns to scale constraints.

In some cases, the inputs and outputs in DEA models may be defined as integers. So, it is necessary considering integer constraints in DEA models, i.e. the following expressions in the mathematical formulation of the models must be an integer:

$$\begin{aligned} \sum_{i=1}^n x_{ij} \lambda_i, & \quad j = 1, \dots, m, \\ \sum_{i=1}^n y_{ik} \lambda_i, & \quad k = 1, \dots, r. \end{aligned} \quad (6)$$

In the input-oriented version of the CCR model (1), the set of constraints is modified as follows:

$$\begin{aligned} x'_j + s_j^- &= \theta_q x_{qj}, & j = 1, \dots, m, \\ \sum_{i=1}^n x_{ij} \lambda_i &= x'_j, & j = 1, \dots, m, \\ \sum_{i=1}^n y_{ik} \lambda_i - s_k^+ &= y_{qk}, & k = 1, \dots, r, \\ \lambda_i \geq 0, s_j^- \geq 0, & & i = 1, \dots, n, j = 1, \dots, m, \\ x'_j \geq 0, \text{integer}, & & j = 1, \dots, m, \\ s_k^+ \geq 0, \text{integer} & & k = 1, \dots, r, \end{aligned} \quad (7)$$

An interesting formulation of an SBM-based model was published in (Tone, 2016). This model leads to finding the closest virtual unit on the efficient frontier by maximizing the objective function (5) instead of minimizing the traditional SBM model. The new model is denoted as the SBM-max model, and its objective function is optimized over the feasible set that is the same as the efficient frontier derived using the classical SBM model (sometimes called the SBM-min model). However, the feasible set is non-convex in typical cases. That is why the solution of the SBM-max model cannot be derived easily. Tone (2016) proposes an iterative algorithm for the approximation of this optimal solution. The efficiency score derived by the SBM-max model is always greater or equal to the SBM efficiency score.

Additive and weighted additive models (2) - (3) and (3) - (4) can be extended by integer constraints easily by considering all slacks as integer values. However, in the SBM model (3), (5) is the situation with integer constraints more complex, and their adding leads to solving an integer non-linear program. The same holds for the SBM-max model that is even more computationally complex.

3 Slacks-based super-efficiency DEA models

Traditional DEA models, including all models presented in the previous section, cannot distinguish among efficient units as they have identical efficiency scores. Therefore, many approaches how to rank efficient units have been proposed in the past. In this section, we will summarize the most important models that are based on measuring super-efficiency using slacks.

The first model of this category was introduced by Tone (2002). This model is widely applied in a variety of studies. This model removes the unit under the evaluation from the dataset and looks for a unit DMU* with inputs x_j^* , $j = 1, \dots, m$, and outputs y_k^* , $k = 1, \dots, r$ that is efficient in the SBM model (3), (5) after this removal. The super-efficiency measure is the distance of the unit under evaluation, and the DMU* measure using slack variables. The model is as follows:

$$\text{Minimize} \quad \frac{\frac{1}{m} \sum_{j=1}^m x_j^* / x_{qj}}{\frac{1}{r} \sum_{k=1}^r y_k^* / y_{qk}} \quad (8)$$

$$\begin{aligned} \text{subject to} \quad & \sum_{i=1, i \neq q}^n x_{ij} \lambda_i + s_j^- = x_j^*, & j = 1, \dots, m, \\ & \sum_{i=1, i \neq q}^n y_{ik} \lambda_i - s_k^+ = y_k^*, & k = 1, \dots, r, \\ & x_{qj} \leq x_j^*, & j = 1, \dots, m, \end{aligned} \quad (9)$$

$$\begin{aligned}
y_{qk} &\leq y_k^*, & k &= 1, \dots, r, \\
\lambda_i \geq 0, s_j^- \geq 0, s_k^+ &\geq 0, & i &= 1, \dots, n, j = 1, \dots, m, k = 1, \dots, r,
\end{aligned}$$

The objective function (8) is always greater or equal than 1 – it is equal to 1 for SBM inefficient DMUs and greater than 1 for SBM efficient ones. That is why this model cannot be used to obtain a complete ranking of all DMUs but just for the units identified as efficient by the SBM model. The objective function is not linear but can be transformed into a linear program, similarly as in the SBM model. Except for the linearized version of his model, its input- and output-oriented versions are often used.

Two additive super-efficiency models were formulated in (Du et al., 2010). Both models have the same set of constraints and differ slightly in the objective function only. We will work with the second of the two models. The procedure works in two steps. The first step consists of solving an optimization model. In the second step, the results of the optimization model are used for deriving the super-efficiency measure. The optimization model is the following:

$$\text{Minimize} \quad \frac{1}{m+r} \left(\sum_{j=1}^m \frac{s_j^-}{x_{qj}} + \sum_{k=1}^r \frac{s_k^+}{y_{qk}} \right) \quad (10)$$

$$\begin{aligned}
\text{subject to} \quad & \sum_{i=1, i \neq q}^n x_{ij} \lambda_i - s_j^- \leq x_{qj}, & j &= 1, \dots, m, \\
& \sum_{i=1, i \neq q}^n y_{ik} \lambda_i + s_k^+ \geq y_{qk}, & k &= 1, \dots, r, \\
& \lambda_i \geq 0, s_j^- \geq 0, s_k^+ \geq 0, & i &= 1, \dots, n, j = 1, \dots, m, k = 1, \dots, r.
\end{aligned} \quad (11)$$

Let $s_j^-^*$, $j = 1, \dots, m$, and $s_k^+^*$, $k = 1, \dots, r$, be the optimal values slacks in solving model (10) - (11). Then, the super-efficiency measure δ_q for the unit under evaluation is calculated using the following formula:

$$\delta_q = \frac{1 + \frac{1}{m} \sum_{j=1}^m \frac{s_j^-^*}{x_{qj}}}{1 - \frac{1}{r} \sum_{k=1}^r \frac{s_k^+^*}{y_{qk}}} \quad (12)$$

This measure is always greater or equal to 1. The disadvantage of this approach is that the optimization model itself does not directly return the super-efficiency score, but it must be calculated using (12) in the second step.

Jablonský (2012) introduced a super-efficiency model based on the goal programming methodology. Its formulation follows:

$$\text{Minimize} \quad 1 + tD + (1-t) \left(\sum_{j=1}^m [s_{j1}^+ / x_{qj}] + \sum_{k=1}^r [s_{k2}^- / y_{qk}] \right), \quad (13)$$

$$\begin{aligned}
\text{subject to} \quad & \sum_{i=1, i \neq q}^n x_{ij} \lambda_i + s_{j1}^- - s_{j1}^+ = x_{qj}, & j &= 1, \dots, m, \\
& \sum_{i=1, i \neq q}^n y_{ik} \lambda_i + s_{k2}^- - s_{k2}^+ = y_{qk}, & k &= 1, \dots, r,
\end{aligned} \quad (14)$$

$$\begin{aligned}
s_{j1}^+ &\leq Dx_{qj}, s_{k2}^- \leq Dy_{qk} & j &= 1, \dots, m, k = 1, \dots, r, \\
s_{j1}^-, s_{j1}^+, s_{k2}^-, s_{k2}^+ &\geq 0, \lambda_i \geq 0, & j &= 1, \dots, m, k = 1, \dots, r, i = 1, \dots, n, \\
t &\in \langle 0, 1 \rangle.
\end{aligned}$$

where $s_{j1}^-, s_{j1}^+, s_{k2}^-, s_{k2}^+$ are negative and positive deviational variables from the inputs and outputs of the DMU_q. The model considers the undesirable deviations only (negative deviations for inputs and positive for outputs) in the objective function. D is the maximum relative deviation, and t is the parameter. Its value 0 leads to minimizing the sum of relative deviations, the value 1 to minimize the maximum deviation. The optimal objective function of the model is always greater than 1 for the units being SBM efficient.

If some of the inputs or outputs must be integers, the super-efficiency models (10) - (12) and (13) - (14) may be applied after a simple modification as presented in the previous section of the paper. However, Tone's super-efficiency SBM model (8) - (9) requires solving a quite complex integer non-linear optimization problem, and it is hardly applicable in practice.

4 A numerical example

The results of all presented models will be illustrated on a small example (12 DMUs, two inputs and two outputs) without any economic background. The dataset is presented in Table 1. This table also contains the results (efficiency scores) of the traditional CCR input-oriented model and the following SBM-based models:

- (Unweighted) additive model (2) - (3) – ADD.
- The weighted additive model with the weights equal to the reciprocal values of maximums of each input and output – ADDW.
- Tone's SBM model (3), (5).
- SBM-max model.

Table 1 A dataset and efficiency scores for the illustrative example

DMUs	I1	I2	O1	O2	CCR	ADD	ADDW	SBM	SBMmax
DMU1	5	42	126	10	1.000	0.00	0.000	1.000	1.000
DMU2	14	84	184	16	0.742	79.71	0.843	0.632	0.737
DMU3	7	49	150	12	1.000	0.00	0.000	1.000	1.000
DMU4	3	27	82	6	1.000	0.00	0.000	1.000	1.000
DMU5	18	98	201	9	0.664	118.00	1.694	0.351	0.381
DMU6	9	72	96	11	0.527	129.14	1.030	0.506	0.612
DMU7	6	52	101	18	1.000	0.00	0.000	1.000	1.000
DMU8	3	36	73	3	0.890	21.00	0.309	0.530	0.693
DMU9	10	72	144	7	0.648	86.62	1.065	0.407	0.556
DMU10	6	46	142	8	1.000	0.00	0.000	1.000	1.000
DMU11	5	28	85	7	1.000	0.00	0.000	1.000	1.000
DMU12	6	38	63	4	0.538	59.20	0.652	0.388	0.551

In our example, all models identify 6 DMUs as efficient, and the same number is inefficient. The efficiency scores of both weighted and unweighted additive models can be just hardly explained. It is a simple sum of slacks, and the higher values correspond to the less efficient units. The efficiency scores obtained by the SBM model are always less than or equal to the scores produced by the CCR model – this property was proved in (Tone, 2001). SBM-max model finds a unit on the efficient frontier closer to the unit under evaluation, i.e. the efficiency score calculated by this concept is always greater than or equal to the SBM efficiency score and can be (but need not be) greater than the CCR efficiency score. Note that the ranking of inefficient units produced by all models is not the same even though all models (except CCR) are based on the same principle. We have tested how the results are influenced by adding integer constraints. The efficiency scores of integer models always lead to only slightly better results than the data in Table 1, but the improvement was insignificant.

Table 2 presents the results of several super-efficiency DEA models described in the previous section of the paper. Except for all SBM-based super-efficiency models, for comparison purposes, we added results of the radial super-efficiency model derived from CCR model (Andersen and Petersen model - AP). The included SBM-based models are the following:

- Tone's super SBM model (8) - (9) – SSBM.
- The model (10) - (12) proposed in (Du et al., 2010) – SDU.
- Goal programming model (13) - (14) for the values of parameter $t = 0$ and $t = 1$ – SBMG0 and SBMG1.

Table 2 Super-efficiency scores

DMUs	AP	SSBM	SDU	SBMG0	SBMG1
DMU1	1.0097	1.0058	1.0117	1.0117	1.0213
DMU3	1.0142	1.0074	1.0149	1.0149	1.0308
DMU4	1.0847	1.0406	1.0847	1.0780	1.0826
DMU7	1.5000	1.2000	1.5000	1.3333	1.7739
DMU10	1.0115	1.0057	1.0115	1.0114	1.0210
DMU11	1.0054	1.0027	1.0054	1.0054	1.0108

The aim of this paper is not a detailed analysis of the results of super-efficiency models but rather a review of available SBM-based models. Interestingly, the models (Du et al., 2010) and Jablonsky (2012) return identical super-efficiency scores for some DMUs. The second of these two models solves just one optimization problem, whereas the first model works in two stages. Therefore, it is less convenient and computationally demanding. All presented models rank as the best DMU7 followed by DMU4. The ranking of the remaining four SBM efficient units is not the same as shown in Table 2.

5 Conclusions

This paper aimed to overview DEA models that measure the level of efficiency and/or super-efficiency using slack variables. This group of models is growing in popularity among researchers and is often used for comparison purposes as a complement to traditional radial DEA models.

The illustrative example shows that the results of the presented models are not consistent with each other. Future research can include analyzing the mutual relations of the SBM-based models and their relation to other DEA models. Also, an analysis of integer and continuous SBM-based models is an open task. Even though various SBM-based models for network systems and multi-period analyses were proposed, the research in this field is open to new ideas.

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] Ali, A.I., Lerne, C.S. & Seiford, L.M. (1995). Components of efficiency evaluation in data envelopment analysis. *European Journal of Operational Research*, 80(2): 462–473.
- [2] Charnes, A., Cooper, W.W., Golany, B., Seiford, L.M. & Stutz, J. (1985). Foundations of data envelopment analysis for Pareto-Koopman's efficient empirical production functions. *Journal of Econometrics*, 30(1-2): 1-17.
- [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] Du, J., Liang, L. & Zhu, J. (2010). A slacks-based measure of super-efficiency in data envelopment analysis: A comment. *European Journal of Operational Research*, 204(3): 694–697.
- [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] Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, 130(3), 498–509.
- [7] Tone, K. (2002). A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, 143(1), 32–41.
- [8] Tone, K. (2016). Data envelopment analysis as a Kaizen tool: SBM variations revisited. *Bulletin of Mathematical Sciences and Applications*, 16, 49–61.

Swap Heuristics for Emergency System Design with Multiple Facility Location

Jaroslav Janáček¹, Marek Kvet²

Abstract. Appropriate deployment of facilities in a node set of a road network is a core of efficient emergency system design. The previously suggested models based on the original weighted p -median problem did not reflect limited capacity of the deployed service centers. In addition, the original approaches assumed that only one facility can be located at a given road network node. This paper introduces a new formulation of the problem, where the temporary inaccessibility of the nearest service facility is taken into account and furthermore, a service center can be equipped with more than one facility. To be able to solve large instances of the generalized problem, a swap heuristics was suggested and a strategy combining the best admissible and first admissible strategies was studied. The hypothesis that even a simple swap heuristics is able to reach a near-to-optimal solution of the complex location problem was verified by the documented computational study.

Keywords: location problems, emergency medical service system, multiple facility location, swap heuristics

JEL Classification: C44

AMS Classification: 90C06, 90C10, 90C27

1 Introduction

Advanced knowledge of mathematical modelling and optimization is an essential property of professional top managers responsible for efficient usage of money and other shared public resources – people, vehicles, technical equipment, etc. Another competitive advantage of the best leaders consists in the ability to find a suitable solving tool for any problem. Different quantitative methods developed by the operations researchers and other IT experts can significantly help us in making the right decisions or in choosing the best admissible solution from the set of all alternatives. Due to a wide range of possible applications of mentioned mathematical methods not only in Economics, but also in many other fields, we concentrate our effort on developing an effective and fast solving tool for a special class of problems used to optimize the urgent pre-hospital healthcare system.

The studied challenge originates from the weighted p -median problem, the idea of which consists in choosing given exact number p of service center locations from a specific set of candidates in order to optimize the quality criterion of system design [1, 2, 3, 7, 13, 14]. Even if this kind of problem is solvable quite well either by various exact [5, 8, 15] or heuristic methods [4, 6, 17], the mathematical model itself does not take into account several important aspects of the real system especially when the Emergency Medical Service (EMS) system is considered. Therefore, the original formulation needs some extension.

The first disadvantage of common weighted p -median problem can be expressed as disregard for stochastic behavior of real EMS system and randomly occurring demands for service. When a new emergency occurs, the nearest located center from the demand point may not have sufficient capacity to cover the demand. In such a case, the request is assigned to the closest available crew, which does not have to be the nearest one. This model extension can be achieved by the concept of so-called *generalized disutility*, which allows us to model providing the service from more centers [10, 11, 12].

Another weakness of the original model lies in the fact that it does not allow to locate more facilities in the same network node. It means that no candidate for service center may get equipped with more than one ambulance vehicle or other resource. On the other hand, if we look at some bigger cities, it is common that there are more service centers spread over the territory. For example, there are four EMS stations located in Žilina. Thus, the associated facility location problem should comply with multiple facility location.

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

Obviously, adjusting the original model to these new requirements can make the problem harder for effective and fast solving. Therefore, the main research idea of this paper consists in suggesting a swap heuristics, which would be able to comply with multiple facility location in one network node. The background of this heuristic approach was introduced for the original weighted p -median problem in [16]. It is based on processing the set of feasible solutions of the problem formed by a special set called *uniformly deployed set*. The biggest advantage of the uniformly deployed set of solutions lies in its constructing, which is completely independent from the solved problem and used objective function [9].

To study the basic characteristics of suggested heuristic approach (computational time demands and the resulting solution accuracy), the numerical experiments with real-world middle-sized problem instances have been performed and the obtained results are discussed in a separate section.

2 P-Facility Location Problem with Multiple Facility Location

A p -location problem formulation is often used in connection with public service system designing, where the designed system services randomly occurring demands from a group of service centers, which have limited ability to yield service. The designed system provides the service to people, who are concentrated at n dwelling places of a serviced region. It is assumed that a dwelling place $j = 1, \dots, n$ will generate demands for service with frequency b_j . The system services a demand emerged at the dwelling place j by a facility, which is located at a service center location i . The destined facility, e.g. servicing vehicle, has to travel from location i to the location j , then it satisfies the demand and returns back to the service center. When employed by the service, the facility is unable to service other demands. That is why, a currently emerged demand cannot be satisfied ever from the closest service center, but it is covered by the first available facility in general. To describe this mode of facility operating in an associated model, the probability values p_1, \dots, p_r are introduced to express by the value p_k a number of cases, when a demand is satisfied from the k -th nearest facility due to the fact that this facility is the nearest available one [10, 11, 12].

An original public service system design problem is usually formulated as a choice of p service center locations from a set of m possible service center locations so that the expected time distance from a demand location to the nearest available facility location is minimal. This original problem has been broadly studied and successfully solved by many authors [1, 2, 3, 4, 5, 6, 7, 13, 14], but their approaches assumed that exactly p service centers were to be deployed. As each possible service center location usually corresponds to one road network node, the assumption acceptance excludes situations, when more than one facility is established at one service center location. If we introduce symbol d_{ij} to denote a network time-distance between locations i and j , then a combinatorial model of (1) can describe the original problem.

$$\min \left\{ \sum_{j=1}^n b_j \sum_{k=1}^r q_k d_{v(P,j,k),j}, P \subset \{1, \dots, m\}, |P| = p \right\} \quad (1)$$

For given demand location j , the ranking function $v(P, j, k)$ returns the k -th nearest service center location from the set P . A much more complex problem is faced, when the case of service centers equipped with more than one facility is admitted. To model the case, a mapping $y: P \rightarrow \{1, \dots, p\}$ is introduced to express the number $y(i)$ of facilities assigned to a service center $i \in P$. Then the model (1) can be adapted to the studied case in the form of (2).

$$\min \left\{ \sum_{j=1}^n b_j \sum_{k=1}^r q_k d_{w(P,y,j,k),j}, P \subset \{1, \dots, m\}, y \in \{1, \dots, p\}^P, \sum_{i \in P} y(i) = p \right\} \quad (2)$$

The models (1) and (2) are very similar, but the ranking function $w(P, y, j, k)$ is much more complex than $v(P, j, k)$, what can be demonstrated by the following algorithm, which maps the quadruple $[P, y, j, k]$ to an element of P .

$w(P, y, j, k)$

0. Order the elements of P into a sequence $i(1), \dots, i(|P|)$ so that the following inequalities hold: $d_{i(1),j} \leq d_{i(2),j} \leq \dots \leq d_{i(|P|),j}$.
1. Find t such that it satisfies: $y(i(1)) + \dots + y(i(t-1)) < k \leq y(i(1)) + \dots + y(i(t))$.
2. Export $i(t)$ as the result of $w(P, y, j, k)$.

Excellent results [16] of swap heuristics applied to the problem (1) evoked a hypothesis that a similar approach may be successful, when applied to problem (2) after a necessary generalization. In the further sections, we will verify this hypothesis.

3 Extended Swap Operation for Multiple Facility Location

A standard swap operation is a frequently used way of changing a solution, which is determined by a simple subset of the elements of some finite superset. As concerns the problem (1), a solution is given by a subset P of cardinality p , where the elements of P are chosen from the domain $D = \{1, \dots, m\}$ of possible service center locations. The swap operation consists of replacing one element $i \in P$ with another element $j \in D - P$. Using the swap operation, a neighborhood of the solution P can be defined. The neighborhood consists of all solutions, which can be obtained by one swap operation applied to the solution P . Obviously, the neighborhood has exactly $|P| \cdot |D - P| = p \cdot (m - p)$ elements.

Contrary to the above standard case, a solution of the studied problem (2) is determined by a subset $P \subseteq D$ satisfying $|P| \leq p$ and by a mapping $y: P \rightarrow \{1, \dots, p\}$, which distributes exactly p facilities to locations from P so that at least one facility is assigned to each location. The integer positive value $y(i)$ gives the number of facilities located at the service center location i . In this case, the swap operation is performed by moving a facility from its original center location to another possible center location.

As there are $m - 1$ possible facility destinations for a facility originally located at the location $i \in P$, the neighborhood cardinality is $|P| \cdot (|D| - 1) = |P| \cdot (m - 1)$.

Based on the above explanation, we introduce the swap operation $Swap(P, y, i, j)$ for $i \in P$ and $j \in D - \{i\}$ defined by the following commands:

Swap(P, y, i, j)

If $j \in D - P$ then $P = P \cup \{j\}$, $y(j) = 1$,

else $y(j) = y(j) + 1$.

Set $y(i) = y(i) - 1$.

If $y(i) = 0$, then $P = P - \{i\}$.

Return the pair P, y .

4 Mixed Strategy of the Neighborhood Search

A perturbation minimizing heuristic based on the neighborhood search operates according to a very simple scheme, which starts from a feasible solution declared as the current one, and searches the neighborhood of the current solution up to the moment, when a better solution is found. Then, this better solution is declared as the new current solution and new neighborhood search is performed. The optimization process terminates, when no better solution is found in the neighborhood of the current solution.

Most of these perturbation heuristics use one of two strategies, which differ in a rule, under which the current solution is updated. The strategy "first admissible" searches the current neighborhood until such solution is found, which is better than the current one. Then, the neighborhood searching is stopped and the current solution is updated.

The strategy "best admissible" always completes the current neighborhood search and then compares the best-found solution of the neighborhood to the current solution. If the best solution of the neighborhood is better than the current one, then the current solution is updated.

In connection with these two strategies, the question arises, which of them is better for the solved problem. To answer the question for the case of p -location problem with multiple facility location, we suggested the following mixed strategy. The termination rule of the current neighborhood search is controlled by setting of two parameters *maxNos* and *Threshold*. The value of *Threshold* gives the minimal relative decrease of objective function of an inspected solution in comparison to the current solution, at which the inspected solution is considered admissible. The integer value of parameter *maxNos* gives the number of admissible solutions, which must be found during the neighborhood inspection to terminate the search. The control parameters enable to apply the first admissible strategy, when *maxNos* = 1 and *Threshold* = 0. The best admissible strategy is followed, when, for example, *maxNos* $\geq p \cdot (m - 1)$ and *Threshold* = 0. In the mixed strategy, the best-found admissible solution is used to update the current solution. The proposed swap heuristic with mixed strategy performs according to the following steps:

SwapMS(P, y, maxNos, Threshold)

0. {Initialization of the best-found solution}
Set $f^* = f(P, y)$, $P^* = P$, $y^* = y$.
1. {Initialization of the neighborhood search}
Initialize list L of all pairs $[i, j] \in P \times (D - \{i\})$ and set $Nos = 0$.
Perform step 2 and having performed step 2, continue with step 3.
2. {The neighborhood search}
While $Nos < maxNos$ and $L \neq \emptyset$ do
 Withdraw a pair $[i, j]$ from L .
 If $f(\text{Swap}(P^*, y^*, i, j)) < f(P, y)$, then $[P, y] = \text{Swap}(P^*, y^*, i, j)$.
 If $f(\text{Swap}(P^*, y^*, i, j)) < (1 - \text{Threshold}) \cdot f^*$, then $Nos = Nos + 1$.
end-do.
3. {}
If $f(P, y) < f^*$, then update $f^* = f(P, y)$, $P^* = P$, $y^* = y$ and go to step 1, otherwise terminate and return P^* , y^* and f^* as the best-found solution and its objective function value.

5 Numerical Experiments

The further presented numerical experiments were suggested and performed to find, whether the swap heuristic is able to solve the newly introduced p -location problem with multiple facility location with the same efficiency as the former swap heuristics intended for standard p -location problems reported originally in [16]. In addition, the experiments were also proposed to decide between the best admissible and first admissible strategies as concerns accuracy of the results and computational time consumption.

Each relevant computational study aimed at verifying the performance characteristics of heuristic solving approaches requires a sufficient set of benchmarks. For this purpose, we make use of common problem instances used in our previous research [9, 10, 11, 15, 16]. These benchmarks correspond to existing EMS system operated by private agencies in eight regions according to the territorial and administrative arrangement of Slovakia.

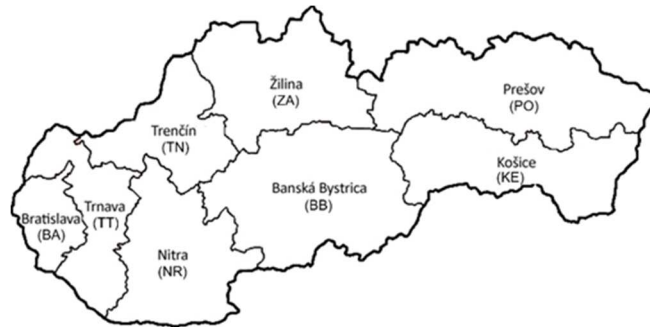


Fig. 1 Administrative organization of Slovakia

The $\{d_{ij}\}$ matrices correspond to the geographical distances in the associated road network. In each self-governing region depicted in Fig. 1, given number p of EMS stations is spread over the territory in order to satisfy demands of clients located in the network nodes. Each client location is connected with the coefficient b_j , the value of which was set according to the number of inhabitants sharing the location rounded up to hundreds. The sizes of the individual benchmarks are reported in the left par of Table 1. The generalized disutility objective function was computed for $r=3$. The coefficients q_k for $k=1 \dots r$ have been obtained from statistics presented in [12] and their values are: $q_1 = 77.063$, $q_2 = 16.476$ and $q_3 = 100 - q_1 - q_2$. The experiments reported in this Section were performed on a common PC equipped with the Intel® Core™ i7-3610QM CPU@2.30 GHz processor and 8 GB RAM. The algorithms were implemented in Java programming language making use of the NetBeans IDE 8.2 environment. The following Table 1 contains the basic characteristics of used problem instances together with the objective function values $OptSol$ of the optimal solutions, which were obtained from [8].

Table 1 Basic characteristics of used benchmarks and the optimal objective function values

Region	BA	BB	KE	NR	PO	TN	TT	ZA
m	87	515	460	350	664	276	249	315
p	25	46	38	36	44	26	22	36
$OptSol$	18450	38008	40711	40987	46884	31260	36401	36929

An individual experiment with the proposed heuristic follows from the fact, that the swap operation needs a current starting solution, the neighborhood of which is inspected. The starting solutions may be taken from a special list called a *uniformly deployed set* of solutions (UDS). Such a set can represent maximally diversified population and may play an important role in all metaheuristics, where population diversity have to be maintained. The biggest advantage of the UDS consists in the fact that it can be constructed independently on used objective function. Just the cardinality of the service center candidates set and the number of facilities to be located are necessary. The construction of a UDS and its possible usage were discussed in [9, 10, 16]. Furthermore, we can use another strong side of the UDS. As the maximal UDS can be formed regardless of the location numbering, each permutation of subscripts of locations gives a different maximal uniformly deployed set. This property enables us to run the heuristics repeatedly with different input set of solutions to be explored. Therefore, the results reported in the following tables represent the average of ten runs with different input sets.

The experiments were performed in two batches. In the first one, the parameter *Threshold* was fixed at zero and performance of *SwapMS* was studied for different values of parameter *maxNos*. In Table 2, the average results for *maxNos* = $p.(m-1)$, 75, 25 and 1 are presented. It can be noticed that the first case corresponds to the standard best admissible strategy and the last one performs according to the first admissible strategy. For each setting of the method parameter, we report the objective function value *ObjF* and the computational time in seconds denoted by *CT*. Each row of the table correspond to one used problem instance. The left part of the table contains the results reported in Table 1 to make the evaluation of the accuracy simpler.

Table 2 Results of numerical experiments – part I

Region	<i>OptSol</i>	<i>maxNos</i> = $p.(m-1)$		<i>maxNos</i> =75		<i>maxNos</i> =25		<i>maxNos</i> =1	
		<i>ObjF</i>	<i>CT</i>	<i>ObjF</i>	<i>CT</i>	<i>ObjF</i>	<i>CT</i>	<i>ObjF</i>	<i>CT</i>
BA	18450	18450	0.07	18450	0.05	18450	0.04	18450	0.11
BB	38008	38038	9.93	38032	8.09	38012	9.57	38060	34.40
KE	40711	40752	5.58	40740	4.17	40764	4.61	40726	15.01
NR	40987	41024	2.68	41011	2.08	41011	2.35	40994	7.55
PO	46884	46950	14.50	46975	11.41	46957	13.56	46953	56.33
TN	31260	31260	0.95	31260	0.71	31260	0.84	31260	3.05
TT	36401	36419	0.57	36412	0.43	36410	0.43	36420	1.28
ZA	36929	36943	2.22	36944	1.52	36937	1.78	36930	6.73

The second batch of experiments was aimed at the mixed strategy behavior for the different values of the *Threshold*. The parameter *maxNos* was fixed at the value of one, what for *Threshold* = 0 corresponds to the first admissible strategy, but the parameter was varied in this batch starting from the value of 0.005. It was found that the relevant values of the parameter lie in the interval [0, 0.035]. The results for *Threshold* = 0.005, 0.015, 0.025 and 0.035 are plotted in Table 3, which has the same structure as Table 2.

Table 3 Results of numerical experiments – part II

Region	<i>OptSol</i>	<i>Threshold</i> = 0.005		<i>Threshold</i> = 0.015		<i>Threshold</i> = 0.025		<i>Threshold</i> = 0.035	
		<i>ObjF</i>	<i>CT</i>	<i>ObjF</i>	<i>CT</i>	<i>ObjF</i>	<i>CT</i>	<i>ObjF</i>	<i>CT</i>
BA	18450	18450	0.06	18450	0.04	18450	0.05	18450	0.05
BB	38008	38048	9.39	38037	7.95	38036	8.92	38031	9.12
KE	40711	40784	4.84	40724	4.55	40743	4.85	40745	5.09
NR	40987	41008	2.47	41045	2.06	41015	2.36	41026	2.42
PO	46884	46958	13.80	46934	11.51	46951	12.68	46942	13.75
TN	31260	31260	1.07	31260	0.78	31260	0.74	31260	0.79
TT	36401	36413	0.53	36414	0.43	36418	0.45	36421	0.48
ZA	36929	36930	1.92	36957	1.61	36950	1.85	36943	1.96

6 Conclusions

This paper brought a new formulation of the former weighted p -median problem, which is able to comply with multiple facility location at the same network node. Special attention was paid to a swap heuristics, in which a strategy combining the best admissible and the first admissible principle was studied.

To verify the hypothesis that even a simple swap heuristics is able to reach a near-to-optimal solution of the complex location problem, a computational study with real-world middle-sized problem instances was performed. As a source of input solutions for inspecting their neighborhood, a uniformly deployed set was taken

and its useful features enabled us to perform more experiments with different input solutions. Since the suggested heuristic algorithm performance can be influenced by two parameters, we studied their impact on the solution accuracy and computational time demands.

The obtained results proved that the suggested swap heuristic method achieves the optimal solution or a near-to-optimal solution very fast and therefore, it is suitable for practical usage not only for EMS system designing, but in much wider spectrum of applications.

Future research in this field could be aimed at other heuristic or metaheuristic methods for p -location problem with multiple facility location and generalized objective function.

Acknowledgements

This work was supported by the research grants VEGA 1/0089/19 “Data analysis methods and decisions support tools for service systems supporting electric vehicles”, VEGA 1/0689/19 “Optimal design and economically efficient charging infrastructure deployment for electric buses in public transportation of smart cities”, and VEGA 1/0216/21 “Design of emergency systems with conflicting criteria using artificial intelligence tools”. This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-19-0441.

References

- [1] Avella, P., Sassano, A. & Vasil'ev, 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. (2021). Multiple p -Facility Location Problem with Randomly Emerging Demands. In: Strategic Management and its Support by Information Systems 2021: 14th International Conference, Technical University of Ostrava, in print
- [9] Janáček, J. & Kvet, M. (2020). Uniform Deployment of the p -Location Problem Solutions. In: Operations Research Proceedings 2019: Selected Papers of the Annual International Conference of the German Operations Research Society, Dresden, Germany, September 4-6, 2019: Springer, pp. 315-321.
- [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, E. & 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. (2020). Usage of Uniform Deployment for Heuristic Design of Emergency System. In: Operations Research Proceedings 2019: Selected Papers of the Annual International Conference of the German Operations Research Society, Dresden, Germany, September 4-6, 2019: Springer, pp. 309-314.
- [17] Rybičková A., Mocková D. & Teichmann D. (2019). Genetic Algorithm for the Continuous Location-Routing Problem, *Neural Network World* 29(3), pp. 173-187.

The minimal network of hospitals in terms of transportation accessibility

Ľudmila Jánošíková¹, Peter Jankovič²

Abstract. The paper addresses the problem of designing a minimal network of hospitals over the state territory. The analysis was ordered by the Ministry of Health of the Slovak Republic during the preparation of the national Recovery and Resilience Plan. The goal is to propose the locations of the hospitals so that the total travel time of citizens to the hospital is as small as possible and the given maximum size of the hospital's catchment area is respected. Three scenarios that differ in the maximum size of the catchment area are investigated. This parameter was set to 100, 150 and 200 thousand people, respectively. All municipalities are considered as candidates for hospital location. Supposing individual transportation mode, fastest routes among all municipalities are calculated using the digital road network from the OpenStreetMap database. The problem is formulated as the capacitated p -median location problem and solved using an efficient matheuristic method. The results are statistically evaluated and visualized in the map.

Keywords: discrete location, capacitated p -median problem, matheuristic

JEL Classification: C61

AMS Classification: 90B90, 90C10

1 Introduction

The Recovery and Resilience Plan of the Slovak Republic in Component 11 presents a reform of acute and inpatient health care. The goal is to create a modern, accessible and effective network of hospitals that ensures good-quality health care for all citizens. During the preparation of this plan we were addressed by the Ministry of Health to conduct a preliminary study with the aim to propose a minimal network of hospitals in the country. The optimization criterion was defined as transportation accessibility by private cars. The only constraint included the size of the catchment area where three scenarios were supposed. The Ministry defined the size of the catchment area to range from 100,000 to 200,000 people with the increment of 50,000. All municipalities were supposed as candidate locations for new hospitals.

The problem can be formulated as the capacitated p -median problem. This problem seeks optimal locations of p facilities serving customers with defined demands. The sum of travel times between facilities and customers weighted by customers' demand is minimized. Each customer must be assigned to exactly one facility. Each facility can serve only a limited amount of demand. The problem is known to be NP-hard. It means that real-world instances cannot be solved to optimality due to time and computer memory limitations. Heuristic and metaheuristic methods are developed instead that produce good solutions in an acceptable amount of time. Matheuristics represent a new branch in the development of heuristics that has been intensively investigated in the last decade [6]. Matheuristics utilize a mathematical programming formulation of the problem. The solution procedure combines a general heuristic or metaheuristic approach with a mathematical programming solver. Recently developed matheuristics include two successful algorithms, namely the LOPT algorithm [5, 9] and kernel search [2, 3].

2 Problem formulation

The aforementioned problem of designing a minimal network of hospitals can be formulated as the capacitated p -median problem [1, 7] using the following interpretation.

¹ University of Žilina, Faculty of Management Science and Informatics, Univerzitná 1, 010 26 Žilina, Slovak Republic, Ludmila.Janosikova@fri.uniza.sk.

² University of Žilina, Faculty of Management Science and Informatics, Univerzitná 1, 010 26 Žilina, Slovak Republic, Peter.Jankovic@fri.uniza.sk.

We face a public service system where municipalities represent customers. The demand of a customer is the municipality's population. At the same time, municipalities are considered as candidates for hospital locations. The municipalities include 2928 villages, towns and boroughs of two largest cities (17 boroughs in Bratislava and 22 boroughs in Košice). The population data is taken from the database published by the Statistical Office of the Slovak Republic. In this study we use data valid for March 2020, when the population of Slovakia was 5,457,926.

The hierarchy of the health care provision is not considered, it means all hospitals are assumed to provide the same types of services and have the same capacity. The capacity of a hospital is defined in the problem setting as the size of the catchment area. The number of hospitals to be located is calculated by dividing the total population by the hospital capacity. This way we get $p = 55, 37, \text{ and } 28$ hospitals with the capacity of 100, 150, and 200 thousand people, respectively.

The output of the location model is the deployment of hospitals and the list of municipalities that create the catchment area of each hospital. However, with the first two scenarios it is difficult to distribute municipalities into compact catchment areas, because hospitals have only a small spare capacity. For example, if the capacity is 100,000 people, 55 hospitals can serve 5,500,000 people. The total spare capacity is $5,500,000 - 5,457,926 = 42,074$ people. The average spare capacity is approximately 765 people per one hospital. Subject to such tight constraints it is impossible to find reasonable small and compact catchment areas. The result is that the catchment areas contain also distant municipalities and do not cover a coherent territory. That is why we decided to increase the given capacity in the first two settings by 5000 people resulting into the capacity of 105,000 and 155,000 people, respectively.

The towns, boroughs and villages are represented by the nodes on the road network that are closest to the centre of the municipality. This way the calculation of travel times can be based on real network distances. The digital road network was downloaded from the OpenStreetMap database [8], which is a freely available source of geographical data. The travel times are related to deterministic speed of vehicles. The speed depends on the quality of the road and its location inside or outside built-up area. The average speeds of vehicles with regard to the road category were derived from the average speeds of ambulances [4] and adjusted by the Ministry (Table 1).

Road category	Average speed in rural areas (kilometres per hour)	Average speed in residential areas (kilometres per hour)
Motorway	100	90
Expressway	100	90
Important national road	73	39
National road	57	34
Local road	49	31
Residential road		8
Minor road	5	5

Table 1 Average speed of vehicles

A mathematical programming model of the problem can be formulated using the following notation:

- I the set of candidate locations
- J the set of municipalities
- p the number of hospitals to be located
- t_{ij} the shortest travel time from node i to node j
- b_j the number of inhabitants of municipality $j \in J$
- Q the capacity limit of a hospital

The decision variables include location variables y_i and allocation variables x_{ij} . Variable y_i decides whether a hospital is located at node $i \in I$. The assignment of municipality j to the catchment area of a hospital located at node i is modeled by binary variable x_{ij} . The model of the weighted p -median problem can be written as:

$$\text{minimize} \quad \sum_{i \in I} \sum_{j \in J} t_{ij} b_j x_{ij} \quad (1)$$

$$\text{subject to} \quad \sum_{i \in I} x_{ij} = 1 \quad \text{for } j \in J \quad (2)$$

$$x_{ij} \leq y_i \quad \text{for } i \in I, j \in J \quad (3)$$

$$\sum_{j \in J} b_j x_{ij} \leq Q \quad \text{for } i \in I \quad (4)$$

$$\sum_{i \in I} y_i = p \quad (5)$$

$$x_{ij}, y_i \in \{0,1\} \quad \text{for } i \in I, j \in J \quad (6)$$

The objective function (1) minimizes the total travel time of all inhabitants to their hospital. The average travel time can be calculated easily by dividing the total travel time by the whole population. Constraints (2) ensure that every municipality j will be assigned to exactly one hospital i . Constraints (3) say that if a municipality j is assigned to a node i , then there must be a hospital located at the node i . Constraints (4) ensure that the total number of inhabitants served by a hospital does not exceed the given size of the catchment area. Constraint (5) limits the total number of hospitals that can be sited. The remaining obligatory constraints (6) specify binary variables.

3 Solution method

To solve the described problem, we use an efficient heuristic method proposed by Taillard [10, 11]. In [10] the method is referred to as a local optimization method (LOPT).

The principle of the method is very simple: if the problem cannot be optimized as a whole, optimize it in parts. The method can be used for every problem that can be divided into subproblems. Then every improvement of the subproblem corresponds to an improvement of the solution of the whole problem. The capacitated p -median problem can be decomposed in such a way that several medians together with their attraction zones create a subproblem. Thus LOPT can be applied very well.

LOPT starts from an initial location of p centres. All centres are denoted as temporary and inserted into set C . Then a centre from C is randomly selected. This selected centre, a few of its closest centres, and the customers allocated to them in the current solution create a subproblem, which is again the location problem but smaller than the original one. The subproblem is optimized using either exact, heuristic or metaheuristic method. If an improved solution is found, all the centres located in the subproblem remain temporary, otherwise the first centre is removed from C . The process stops when C is empty.

The algorithm of the LOPT method can be described in a more formal way [10]:

1. Input: initial position of the p centres, parameter r .
2. Set $C = \{1, \dots, p\}$
3. While $C \neq \emptyset$, repeat the following steps:
 - a) Randomly select a centre $i \in C$.
 - b) Let S be the subset of the r closest centres to i ($i \in S$).
 - c) Consider the subproblem constructed with the entities allocated to the centres of S and optimize this subproblem with r centres.
 - d) If no improved solution has been found at step 3c, set $C = C \setminus \{i\}$, else set $C = C \cup S$.

If the location problem is formalized as a mathematical programming problem, then an IP solver can be used to calculate the initial solution at Step 1, as well as to solve subproblems at Step 3c. The resulting procedure is then a matheuristic [9].

4 Results and discussion

The LOPT method was implemented in Java language with solver Gurobi 9.0.3 library. The algorithm was run for three different sizes of the catchment area. The results are in Table 2. For every scenario the following indicators are evaluated:

- statistics for the size of the catchment area (mean, maximum, and standard deviation);
- accessibility indicators, namely
 - average travel time of an inhabitant to the assigned hospital;
 - maximum travel time for 80% of inhabitants;
 - maximum travel time for 100% of inhabitants.

Indicator	Scenario 1	Scenario 2	Scenario 3
Input size of the catchment area (people)	105,000	155,000	200,000
Number of hospitals	55	37	28
Maximum size of the catchment area (people)	105,000	155,000	199,999
Mean size of the catchment area (people)	99,235	147,512	194,926
St. dev. of the size of the catchment area (people)	11,705	11,391	12,107
Average travel time (min)	12.0	14.2	16.8
Maximum travel time for 80% of inhabitants (min)	18.8	22.4	27.1
Maximum travel time for 100% of inhabitants (min)	75.3	84.1	78.5

Table 2 Indicators of the hospital network with a limited size of the catchment area

One can see that regardless the size of the catchment area all people can reach their hospital in 90 minutes. The maximum travel time for 80% of population as well as the average travel time increases with the decreasing number of hospitals. The accessibility for 80% of inhabitants ranges from approximately 18 to 27 minutes. The hospitals and their catchment areas for scenario 3 are displayed in Fig. 1. Even in this case with a quite big spare capacity some hospitals do not have a coherent catchment area, e.g. Vrútky.

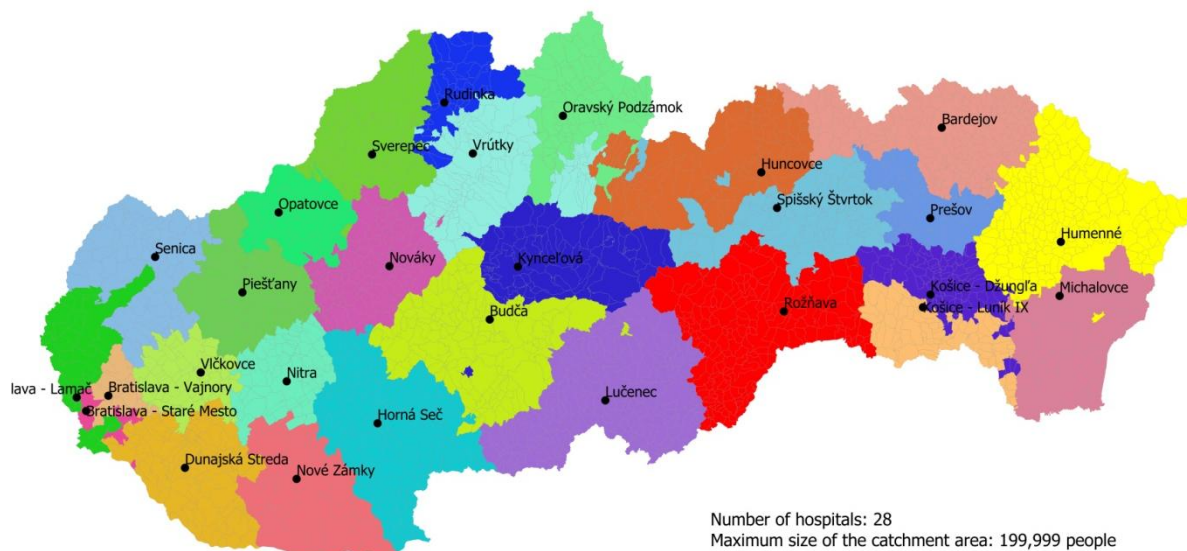


Figure 1 Hospitals with a limited size of the catchment area, scenario 3

The proposed catchment areas might be modified when implemented in practice, at least the most scattered ones. However, unless people are forced into maintaining the catchment areas by administration rules, they will probably travel to their closest hospital. In such a case the catchment areas will be coherent (Fig. 2) but not the same size. Table 3 presents the indicators obtained when the positions of the hospitals are optimized by the capacitated p -median model but the municipalities are assigned to the closest hospital. As was expected, the maximum size of the catchment area as well as the standard deviation increases significantly. The average travel time decreases but only by less than a minute. The most noticeable improvement can be observed in accessibility for most distant villages. Now, everybody can reach the closest hospital in approximately an hour.

The model places approximately 60% of hospitals in small villages nearby the towns where hospitals are nowadays. Such a solution might not be acceptable by decision makers. That is why we performed another experiment in which candidate locations are limited to 63 locations of present hospitals. The results are in Table 4. According to the expectations, there are bigger differences among catchment areas in comparison with the previous hospital distribution. The average and 80% accessibility slightly worsened. The only improvement is in the longest travel time that is by 2 minutes better now with scenarios 2 and 3.

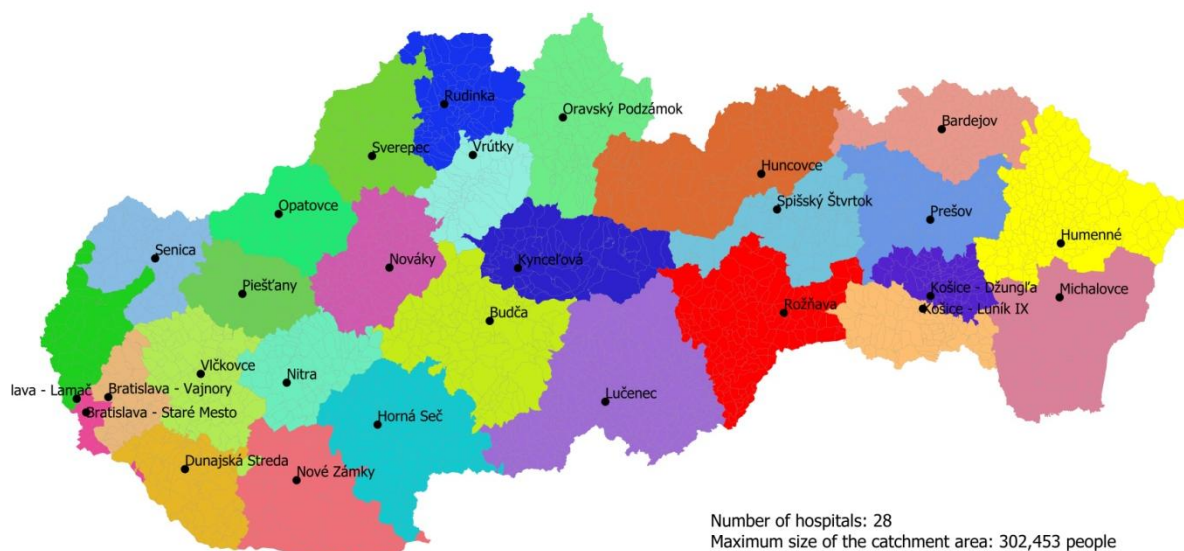


Figure 2 Municipalities are assigned to the closest hospital, scenario 3

Indicator	Scenario 1	Scenario 2	Scenario 3
Input size of the catchment area (people)	105,000	155,000	200,000
Number of hospitals	55	37	28
Maximum size of the catchment area (people)	192,719	258,335	302,453
St. dev. of the size of the catchment area (people)	31,100	43,556	53,800
Hospital with the maximum catchment area	Bratislava - Nové Mesto	Prešov	Bratislava - Staré Mesto
Average travel time (min)	11.1	13.5	16.0
Maximum travel time for 80% of inhabitants (min)	17.2	20.9	25.7
Maximum travel time for 100% of inhabitants (min)	54.7	63.5	63.5

Table 3 Indicators of the hospital network, municipalities are assigned to the closest hospital

Indicator	Scenario 1	Scenario 2	Scenario 3
Input size of the catchment area (people)	105,000	155,000	200,000
Number of hospitals	55	37	28
Maximum size of the catchment area (people)	254,546	270,400	306,685
St. dev. of the size of the catchment area (people)	44,256	49,425	54,389
Hospital with the maximum catchment area	Bratislava - Ružinov	Prešov	Trnava
Average travel time (min)	11.6	14.1	16.6
Maximum travel time for 80% of inhabitants (min)	18.5	22.3	26.4
Maximum travel time for 100% of inhabitants (min)	58.7	61.6	61.6

Table 4 Indicators of the hospital network, candidates are current hospitals, municipalities are assigned to the closest hospital

5 Conclusions

The paper deals with designing a minimal network of hospitals given the maximum size of catchments areas as a parameter. The problem is formulated as the capacitated p -median problem. Every municipality is allocated to exactly one hospital. The optimization criterion is the total (or average) travel time of an inhabitant to the hospital in his/her catchment area. The problem is solved using an efficient matheuristic algorithm. Two different sets of candidates for hospital locations are examined. The average computational time with the complete set of 2928 candidates was more than 9 hours on a personal computer equipped with the Intel Core i7 1.60 GHz processor. An instance with 63 candidates was solved in 55 minutes on average. Although the computational time was rather long, especially in the case of the complete candidate set, the solutions of the strategic nature are worth of extra time. The model together with the solution algorithm represent a useful decision support tool. By setting parameters of the model to different values we can generate alternative solutions that can be further evaluated by decision makers.

Acknowledgements

This research was supported by the Slovak Research and Development Agency under the project APVV-19-0441 “Allocation of limited resources to public service systems with conflicting quality criteria” and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences under the project VEGA 1/0216/21 “Designing of emergency systems with conflicting criteria using tools of artificial intelligence”.

References

- [1] Czimmermann, P. (2016). Location problems in transportation networks. *Communications – Scientific Letters of the University of Zilina*, 18(3), 50–53.
- [2] Guastaroba, G., Savelsbergh, M. & Speranza, M. G. (2017). Adaptive Kernel Search: A heuristic for solving Mixed Integer linear Programs. *European Journal Operational Research*, 263(3), 789–804.
- [3] Jánošíková, L. & Jankovič, P. (2018). Emergency medical system design using kernel search. In *Proceedings of the 2018 IEEE Workshop on Complexity in Engineering (COMPENG)*. Firenze: Sezione Italia dell'IEEE.
- [4] Jánošíková, L., Kvet, M., Jankovič, P. & Gábrišová, L. (2019). An optimization and simulation approach to emergency stations relocation. *Central European Journal of Operations Research*, 27(3), 2019, 737–758
- [5] Jánošíková, L. & Žarnay, M. (2014). Location of Emergency Stations as the Capacitated p -median Problem. In *Proceedings of the International Scientific Conference Quantitative Methods in Economics – Multiple Criteria Decision Making XVII* (pp. 116–122). Bratislava: Ekonóm.
- [6] Maniezzo, V., Boschetti, M. A. & Stützle T. (2021). *Matheuristics: Algorithms and Implementations*. Springer.
- [7] Marianov, V. & Serra, D. (2004). Location problems in the public sector. In Z. Drezner & H. W. Hamacher (Eds.), *Facility Location: Applications and Theory* (pp. 119–150). Berlin: Springer.
- [8] OpenStreetMap database. <https://www.openstreetmap.org>. Accessed 16 April 2019.
- [9] Stefanello, F., de Araújo, O. C. B. & Müller, F. M. (2015). Matheuristics for the capacitated p -median problem. *International Transactions in Operational Research*, 22, 149–167.
- [10] Taillard, É. D. (2003). Heuristic methods for large centroid clustering problems. *Journal of Heuristics*, 9, 51–73.
- [11] Taillard, É. D. & Voß, S. (2002). POPMUSIC: Partial Optimization Metaheuristic Under Special Intensification Conditions. In C. C. Ribeiro & P. Hansen (Eds.), *Essay and Survey in Metaheuristics* (pp. 613–629). Boston: Springer.

Multifractal approaches in econometrics and fractal-inspired robust regression

Jan Kalina¹

Abstract. While the mainstream economic theory is based on the concept of general economic equilibrium, the economies throughout the world have recently been facing serious transformations and challenges. Thus, instead of a convergence to equilibrium, the economies can be regarded as unstable, turbulent or chaotic with properties characteristic for fractal or multifractal processes. This paper starts with a discussion of recent data analysis tools inspired by fractal or multifractal concepts. We pay special attention to available data analysis tools based on reciprocal weights assigned to individual observations; these are inspired by an assumed fractal structure of multivariate data. As an extension, we consider here a novel version of the least weighted squares estimator of parameters for the linear regression model, which exploits reciprocal weights. Finally, we perform a statistical analysis of 31 datasets with economic motivation and compare the performance of the least weighted squares estimator with various weights. It turns out that the reciprocal weights, inspired by the fractal theory, are not superior to other choices of weights; in fact, the best prediction results are obtained with trimmed linear weights.

Keywords: chaos in economics, fractal market hypothesis, reciprocal weights, robust regression, prediction.

JEL classification: C14

AMS classification: 62F35

1 Introduction

While the mainstream economic theory is based on the concept of general economic equilibrium, the economies throughout the world have recently been facing serious transformations and challenges. The economies can thus be characterized as diverging to disequilibrium (non-equilibrium), instability, or chaos. Already before the COVID-19 pandemic, the economies were described as turbulent or chaotic [7] with negative consequences for economists, managers or politicians. At the same time, the complexity (dimensionality) of economic systems and of the whole society increases as well together with their increasing vulnerability (sensitivity, non-robustness) to small changes [15]. The main reasons include disruptive events such as the economic recession as a consequence of the pandemic, but also globalization, negative interest rates, high noise level in capital markets in [17], or uncertain (indecisive) economic policy [20]. Investors must process very large amounts of information [1].

In general, analyzing economic data should naturally reflect current socio-economic processes. To be specific, we believe that exploratory methods of data analysis will be more useful than confirmatory (i.e. hypothesis testing) for modeling economic data in non-equilibrium conditions. Further, we believe in an increasing importance of non-stationary time series methods or nonparametric methods (not relying on the Gaussian distribution of errors). As a result of the availability of Big Data, bigger uncertainty will complicate their analysis; at the same time, the availability of prior knowledge will be increasing, which allows exploiting Bayesian statistical methods. Another important aspect is an increasing need for using robust statistical procedures, which are resistant with respect to various forms of data contamination (e.g. by outliers or measurement errors [4]). Also for the future, we expect that a chaotic (i.e. multifractal) structure of economic data will be increasing [16].

Chaos theory represents as a perspective paradigm or theoretical framework able to reflect the latest economic development modeling [8], which has however acquired only little attention in the mainstream economics. Chaos theory regards chaotic objects as having either a fractal or a multifractal nature. Often, economic data especially (but not only) with a high level of complexity and/or nonlinearity may be regarded as having a fractal or multifractal structure. Thus, it is natural to ask if common mathematical and statistical tools habitually applied to economic modeling are suitable at all. So far, econometricians only very rarely incline to using tools inspired by fractal or multifractal structure of data, perhaps with the exception of time series related to financial markets. Indeed, fractal-based tools turn out to be able to grasp difficulties of time series analysis by standard statistical tools. However, there remains a gap of fractal-inspired tools for other data analysis tasks including also linear regression.

¹The Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 2, Prague 8, Czech Republic & Charles University, Faculty of Mathematics and Physics, Sokolovská 83, 186 75 Prague 8, Czech Republic, kalina@cs.cas.cz

Section 2 of this paper discusses recent trends in the analysis of economic data related to chaos theory, fractals and multifractals, including methods exploiting reciprocal weights implicitly assigned to individual observations, and their economic applications. Section 3 presents a linear regression analysis of 31 datasets with an economic motivation with the aim to reveal whether reciprocal weights implicitly assigned to individual observations are superior to other choices of weights. Section 4 concludes the paper.

2 Fractals and multifractals in data analysis

Economic data may look like arising from a fractal process. Such idea may be true not only for big or complex data, for which fractals definitely represent a suitable approach [18], but actually also for common economic data, which are available in a classical setting with the number of observations n exceeding the number of variables p . Data analysis tools stemming from the idea of fractals or multifractals rely on the assumption (or empirical experience) that real data sometimes (or perhaps often) possess some form of scaling. This means that it is realistic to assume that the distances between points are governed by a certain scaling law [2].

In general, fractal structures can be described as self-similar, self-developed and self-organized; in other words, one assumes their high organization automatically appears from chaos. Such fractals are inspired from natural sciences, e.g. from the structure of the DNA or from the Darwinian evolution. Fractals were theoretically investigated by Benoît Mandelbrot (1924–2010), the father of fractal geometry, or by the physicist Ilya Prigogine (1917–2003) in the context of nonlinear dynamic systems. In economics, the properties of fractals are appealing e.g. for modeling of self-regulatory economic mechanisms [21]. Fractal-based modeling of economic data considers the data to represent a fractal structure, i.e. the data are assumed to bear a scaling property with a certain (possibly unknown) single value of the scaling exponent, which represents a parameter uniquely characterizing the scaling. However, if the scaling is not the same (homogeneous) across the whole data space, the concept of fractals is extended to multifractals. Data with a multifractal structure are characterized by the dependence of the scaling exponent on the position of a particular point in the data space.

The stock market volatility has been repeatedly explicitly claimed to have a multifractal nature, especially by proponents of the fractal market hypothesis [4]. As a consequence, fractal-based tools have found interesting applications in the analysis of financial time series. In R software, a specialized package DChaos is available for a chaotic (multifractal) time series analysis for financial applications. Let us now recall at least a few of recent applications in the field of financial time series. Multifractal detrended fluctuation analysis was used in [5] to analyze time series of freight rates (prices) of bulk ships (the largest cargo ships); there, the pre-processing involved the estimation of local scale exponents. A detailed overview of multifractal time series analysis applications to financial markets was presented in [10]; the paper described detrended fluctuation approaches or tools based on the wavelet transform as the main classes of multifractal methods in the field. The work [22] used a self-similarity index for detecting volatility clusters (i.e. clusters with a high volatility) in series of asset returns. A specific topic in the multifractal analysis of time series is related to predicting extremely rare events, i.e. events with an extreme risk. Fractals turned out to be successful for rare disruptive events in the times of the global economic crisis in 2009 [23]. Fractal-based algorithms were successfully used by trading robots developed in [4]. In addition, the multifractal approach for modeling rare events of [24] was empirically shown to be more appropriate for outliers (anomalies) compared to classical data analysis relying on the concept of statistical regularity.

The fractal (or multifractal) structure of data may be directly exploited not only for time series, but also within fractal-based statistical tools for a variety of other tasks of data analysis. Such tasks include regression modeling, classification analysis, cluster analysis, dimensionality reduction [25], or quantile estimation. A regression model for the information flow related to tourism was modeled by means of fractals and correlation dimensions in [27]. An example of estimates of extreme quantiles is the work [9] based on the fractal-inspired power law. Nevertheless, none of the fractal-based approaches mentioned in this section is supported by any theoretical argument; theoretical models for multifractal modeling of financial returns have been described only rarely [3]. Still, we can say in general that individual fractal-based methods rely on assumptions, which are usually not explicitly formulated in research papers. In addition, there remains a gap of theoretical results related to the asymptotic performance (convergence) of the methods.

2.1 Statistical methods with reciprocal weights

This section is devoted to statistical methods with reciprocal weights inspired by fractals. Let us consider magnitudes of reciprocal weights defined as $1, 1/2, 1/3, \dots, 1/n$; these are assigned to each of the total number n of available observations in an implicit way, i.e. in the course of the computation; the user does not know before analyzing the data, which observation will obtain which weight.

A prominent example of a weighted classification method with reciprocal weights is the IINC classifier (Inverted Indexes of Neighbors Classifier) of [11]. IINC is an extension of the k -nearest neighbor (k -NN) classifier, where the

latter can be characterized as a simple (or simplistic) but powerful method for assigning (say) p -variate observations to one of two given groups, based on a training dataset of observations from both the groups. Let us describe the IINC classifier for a given observation $x \in \mathbb{R}^p$ not present in the training dataset. First, the method evaluates the distances between x and each of the observations from the first group. The distances are ranked, i.e. the smallest distance obtains the rank equal to 1, the next obtains the rank equal to 2 etc. The IINC classifier sums reciprocal values (i.e. inverted values) of these weights. In an analogous way, the sum of reciprocal ranks is obtained based on distances between x and each of the observations from the second group. Finally, x is assigned to such of the two groups, for which the sum is smaller. However, to the best of our knowledge, data analysis methods based on implicit reciprocal weights have not been applied to economic data.

Implicitly assigned reciprocal weights are related with fractals through the Zipf's law. This original law was formulated by George K. Zipf (1902–1950) for the context of linguistics. If simplified, it can be described as a distribution of ranks related to the occurrence of individual words in a given language; Zipf observed the most common word to be twice as common as the second most common word, three times more common than the third most common word etc. The reciprocal weights thus evaluate the contribution of a given word relatively to the importance (occurrence) of the most common word. Such experimental law, later generalized to the Zipf-Mandelbrot law, has been empirically observed in a variety of some other applications [19]. Basically, we may take as an assumption (rather than a generally valid law) that ranks of distances evaluated for pairs of observations in a multivariate data space are distributed according to the Zips's law. A deeper justification of reciprocally weighted ranks was given in [11] by means of several fractal-related concepts including correlation integral, correlation dimension, or distribution mapping exponent (see also [6]).

3 Least weighted squares with reciprocal weights

We will now recall the appealing properties of the least weighted squares (LWS) estimator [26, 14] for linear regression and then propose its new version inspired by fractals, i.e. a version of the LWS with implicitly assigned reciprocal weights. After that, we investigate its performance over 31 datasets with an economic motivation.

If the LWS estimator is used with suitable continuous weights, it is able to combine the following desirable properties (see [26] for a discussion):

- High efficiency for non-contaminated data with normally distributed errors;
- High robustness with respect to the breakdown point (i.e. high resistance to severe outliers);
- Low global sensitivity to the influence of an individual outlier (i.e. in terms of maximum of the influence function);
- Low local sensitivity to small changes (perturbations) of the data; see Section 4.9 of [13].

Let us assume the standard linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + e_i, \quad i = 1, \dots, n. \quad (1)$$

Let $u_i(b)$ denote the residual of the i -th measurement based on a given estimate $b = (b_0, b_1, \dots, b_p)^T \in \mathbb{R}^{p+1}$ of β , i.e.

$$u_i(b) = Y_i - b_0 - b_1 X_{i1} - \cdots - b_p X_{ip}, \quad i = 1, \dots, n. \quad (2)$$

We denote the ranks of squared residuals by $R_1(b), \dots, R_n(b)$, i.e. with $R_i(b)$ denoting the rank of $u_i^2(b)$ among $u_1^2(b), \dots, u_n^2(b)$. Just like the (classical) k -nearest neighbor classifier was replaced by the IINC classifier in [11], the least squares estimator can be replaced by an implicitly weighted version with reciprocal weights.

We define the LWS-R estimator (i.e. least weighted squares with reciprocal weights) as

$$b_{LWS} = \arg \min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^n \frac{1}{R_i(b)} u_i^2(b). \quad (3)$$

It is natural to perceive (3) as a special case of the LWS estimator, which may be expressed as

$$b_{LWS} = \arg \min_{b \in \mathbb{R}^{p+1}} \sum_{i=1}^n \psi \left(\frac{R_i(b)}{n} \right) u_i^2(b), \quad (4)$$

with a selected continuous weight function $\psi : [0, 1] \rightarrow [0, 1]$ with $\psi(0) = 1$ and $\psi(1) = 0$. The LWS-R estimator corresponds to a reciprocal weight function as in [11], i.e. the magnitudes of the weights $1, 1/2, 1/3, \dots, 1/n$ are assigned to individual observations after a permutation, which is determined only implicitly (in the course of computing the estimator).

For the sake of comparisons, we consider three other weight functions. Linear weights are defined by

$$\psi_L(t) = 1 - t, \quad t \in (0, 1), \quad (5)$$

Table 1 The 31 datasets together with their basic characteristics (n and p). Ranks corresponding to the mean prediction errors evaluated in (1) in a 10-fold cross validation for various versions of the LWS estimator are presented here. The novel LWS-R method (3) is presented in the last column.

Index	Dataset	Response variable	n	p	Version of LWS			
					(5)	(6)	(7)	(3)
1	Aircraft	Cost	23	5	1	3	4	2
2	Ammonia	Unprocessed percentage	21	4	2	4	1	3
3	Auto MPG	Miles per gallon	392	5	2	3	4	1
4	Boston housing	Crime rate	506	6	4	3	2	1
5	Building	Electricity consumption	4208	7	3	4	1	2
6	California housing	Median house price	20 640	9	2	1	3	4
7	Cirrhosis	Death rate	46	5	3	1	4	2
8	Coleman	Test score	20		4	2	1	3
9	Concrete compression	Concrete compression						
	strength	strength	1030	7	2	4	3	1
10	Delivery	Delivery time	25	3	1	2	4	3
11	Education	Education expenditures	50	4	1	4	3	2
12	Electricity	Output	16	4	2	1	4	3
13	Employment	# of employed people	16	7	3	1	2	4
14	Engel	Food expenditures	235	2	4	3	1	2
15	Furniture	Log relative wage	11	2	4	1	2	3
16	Houseprices	Selling price	28	6	3	4	1	2
17	Imports	Level of imports	18	4	2	4	1	3
18	Investment	Investment	22	2	2	1	3	4
19	Istanbul stock exchange	Istanbul index	536	8	3	2	1	4
20	Kootenay	Newgate	13	2	1	3	4	2
21	Livestock	Expenses	19	5	4	2	1	3
22	Machine	PRP	209	7	4	2	3	1
23	Murders	# of murders	20	4	3	4	2	1
24	NOx emissions	LNOx	8088	4	2	4	1	3
25	Octane	Octane rating	82	5	2	3	1	4
26	Pasture	Pasture rental price	67	4	2	3	1	4
27	Pension	Reserves	18	2	1	2	3	4
28	Petrol	Consumption	48	5	1	2	4	3
29	Stars CYG	Log temperature	47	2	3	1	4	2
30	Travel and tourism	TSI	141	13	3	2	4	1
31	Wood	Wood gravity	20	6	4	1	2	3

trimmed linear weights generated by the weight function

$$\psi_{TL}(t) = \left(1 - \frac{t}{\tau}\right) \cdot \mathbb{1}[t < \tau], \quad t \in (0, 1), \quad (6)$$

where $\mathbb{1}[\cdot]$ denotes an indicator function, and weights generated by the error function

$$\psi_E(t) = 1 - \frac{2}{\sqrt{\pi}} \int_0^t \exp\{-x^2\} dx, \quad t \in (0, 1). \quad (7)$$

The (normalized) ranks $R_i(b)/n$ for $i = 1, \dots, n$ play the role of $t \in (0, 1)$ within (4).

3.1 Numerical experiments

We performed a numerical study over 31 datasets with economic motivation with the aim to compare reciprocal weights within the LWS estimator in linear regression with other choices of weights. The datasets were carefully selected so that the linear model is meaningful for them. To describe the experiments, each of 4 versions of the LWS estimator is computed for each of the 31 datasets. In a 10-fold cross validation, the mean square error as the most basic measure of prediction ability is evaluated for each situation. Based on the results, we computed the ranks corresponding to the mean prediction errors (MSE) of various linear regression estimators.

Let us first present aggregated results over the 31 datasets. The LWS estimator with trimmed linear weights (7) turns out to be the best among the 4 versions of the LWS. Particularly, (5) yields the minimal prediction error for 6 datasets (19 % of the datasets), (6) for 8 datasets (26 % of the datasets), (7) for 11 datasets (35 % of the datasets), and LWS-R (3) for 6 datasets (19 % of the estimators).

In addition to the aggregated results, let us discuss also the results for individual datasets. These are presented in Table 1. Let us now explain the presented results on a particular example considering the first dataset denoted as Aircraft; the LWS estimator with the weight function (5) turns out to be the best for this dataset, (3) is the next, (6) comes as the third, and the weight function (7) yields the worst (largest) value of MSE among the four versions of the LWS for this dataset.

4 Conclusions

As we can currently experience disruptive or unexpected patterns in the economies around the world (not only) as a consequence of the COVID-19 pandemic, it is natural to think about the perspective of methods of chaos theory and/or fractals within econometric data analysis. This paper gives an overview of some recent fractal-based tools for the analysis of economic data with a (multi)fractal structure; recent applications turn out to appear especially in the context of financial time series and only rarely in other tasks for other than temporal data. Further, we realized that a data analysis approach based on reciprocal weights assigned to individual observations can be interpreted as a method inspired by fractals [11]. We arrive at introducing implicit reciprocal weights to individual observations in the context of the LWS-R estimator in the linear regression model.

Our numerical study over 31 datasets in the linear regression model is especially focused on the performance of the novel LWS-R method. The results however reveal that the LWS-R estimator is able to overcome other versions of the LWS estimator (with other choices of weights) only for a small percentage of the datasets. Nevertheless, we recommend to repeat such study for data with a larger number of variables. Without surprise, no weights turn out to be uniformly the best across all datasets; trimmed linear weights are the best in more datasets than any other weights. Optimality of weights can be derived only under specific assumptions on the distribution of the distances; such optimality results were derived for nonparametric hypothesis tests for multivariate (possibly high-dimensional) data in [12]. It is natural to plan future research of the performance of estimators based on reciprocal weights in other tasks, such as in nonlinear regression or in the task to estimate expectation and scatter in the multivariate model, i.e. by means of the minimum weighted covariance determinant (MWCD) estimator, which is based on robust Mahalanobis distances.

Acknowledgements

The work was supported by the Czech Science Foundation projects GA21-19311S (*Information flow and equilibrium in financial markets*). The author is grateful to Lubomír Soukup and Marcel Jiřina for discussion.

References

- [1] Biais, B., Foucault, T., & Moinas, S. (2015). Equilibrium fast trading. *Journal of Financial Economics*, 116, 292–313.

- [2] Briggs, J. (2015). *Fractals: The patterns of chaos. Discovering a new aesthetic of art, science, and nature*. Echo Point Book & Media, Brattleboro.
- [3] Calvet, L.E. & Fisher, A.J. (2013). Extreme risk and fractal regularity in finance. *Contemporary Mathematics*, 601, 65–94.
- [4] Caporale, G.M., Gil-Alana, L., Plastun, A., & Makarenko, I. (2016). Intraday anomalies and market efficiency: A trading robot analysis. *Computational Economics*, 47, 275–295.
- [5] Chen, F., Tian, K., Ding, X., et al. (2017). Multifractal characteristics in maritime economics volatility. *International Journal of Transport Economics*, 44(3), 365–380.
- [6] Dhifaoui, Z. (2016). Robust to noise and outliers estimator of correlation dimension. *Chaos, Solitons & Fractals*, 93, 169–174.
- [7] Faggini, M. & Parziale, A. (2016). More than 20 years of chaos in economics. *Mind & Society*, 15, 53–69.
- [8] Faggini, M., Bruno, B., & Parziale, A. (2019). Does chaos matter in financial time series analysis? *International Journal of Economics and Financial Issues*, 9(4), 18–24.
- [9] Higgins, D.M. (2017). Residential property market performance and extreme risk measures. *Pacific Rim Property Research Journal*, 23, 1–13.
- [10] Jiang, Z.Q., Xie, W.J., Zhou, W.X., & Sornette, D. (2019). Multifractal analysis of financial markets. *Reports on Progress in Physics*, 82(12), Article 125901.
- [11] Jiřina, M. & Jiřina, M. (2015). Classification using Zipfian kernel. *Journal of Classification*, 32, 305–326.
- [12] Jurečková, J. & Kalina, J. (2012). Nonparametric multivariate rank tests and their unbiasedness. *Bernoulli*, 18(1), 229–251.
- [13] Jurečková, J., Pícek, J., & Schindler, M. (2019). *Robust statistical methods with R*. 2nd edn. Boca Raton: CRC Press.
- [14] Kalina, J. (2012). Highly robust statistical methods in medical image analysis. *Biocybernetics and Biomedical Engineering*, 32(2), 3–16.
- [15] Kalina, J. (2014). On robust information extraction from high-dimensional data. *Serbian Journal of Management*, 9(1), 131–144.
- [16] Kalina, J. (2021). Managerial decision support in the post-COVID-19 era: Towards information-based management. In L.C. Carvalho, L. Reis, & C. Silveira (Eds.), *Handbook of Research on Entrepreneurship, Innovation, Sustainability, and ICTs in the Post-COVID-19 Era* (pp. 225–241). Hershey: IGI Global.
- [17] Klioutchnikov, I., Sigova, M., & Beizerov, N. (2017). Chaos theory in finance. *Procedia Computer Science*, 119, 368–375.
- [18] Lahmiri, S. & Bekiros, S. (2020). Big data analytics using multi-fractal wavelet leaders in high-frequency Bitcoin markets. *Chaos, Solitons & Fractals*, 131, Article 109472.
- [19] Latif, N., Pečarić, D. & Pečarić, J. (2017). Majorization, Csiszár divergence and Zipf-Mandelbrot law. *Journal of Inequalities and Applications*, 2017, Article 197.
- [20] Liu, Z., Ye, Y., Ma, F., & Liu, J. (2017). Can economic policy uncertainty help to forecast the volatility: A multifractal perspective. *Physica A*, 482, 181–188.
- [21] Redko, V.G. & Sokhova, Z.B. (2017). Processes of self-organization in the community of investors and producers. *Studies in Computational Intelligence*, 736, 163–169.
- [22] Segovia, J.E.T., Fernández-Martínez, M., & Sánchez-Granero, M.A. (2019). A novel approach to detect volatility clusters in financial time series. *Physica A*, 535, Article 122452.
- [23] Siokis, F.M. (2014). European economies in crisis: A multifractal analysis of disruptive economic events and the effects of financial assistance. *Physica A*, 395, 283–292.
- [24] Stanley, H.E., Gabaix, X., Gopikrishnan, P., & Plerou, V. (2007). Economic fluctuations and statistical physics: Quantifying extremely rare and less rare events in finance. *Physica A*, 382(1), 286–301.
- [25] Traina, C., Traina, A., Wu, L., & Faloutsos, C. (2010). Fast feature selection using fractal dimension—Ten years later. *Journal of Information and Data Management*, 1(1), 17–20.
- [26] Víšek, J.Á. (2011). Consistency of the least weighted squares under heteroscedasticity. *Kybernetika*, 47, 179–206.
- [27] Yan, Q., Su, M., Wu, Y., & Wang, X. (2019). Economic efficiency evaluation of coastal tourism cities based on fractal theory. *Journal of Coastal Research*, 93, 836–842.

LTPD variables inspection plans and effect of wrong process average estimates

Nikola Kaspříková¹

Abstract. The lot tolerance proportion defective acceptance sampling plans were designed by Dodge and Romig to minimize the mean number of items inspected per lot of the process average quality when the remainder of rejected lots is inspected (rectifying plans). It has been shown that for the attributes sampling plans the mean number of items inspected per lot of the process average quality increases both when the true process average is greater and when the true process average is smaller than the estimated value of this parameter. 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 measured by the mean number of items inspected per lot of the process average quality. The rectifying lot tolerance proportion defective sampling plans are calculated and evaluated using an R software extension package.

Keywords: acceptance sampling, inspection cost, LTPD, AOQL

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

The lot tolerance proportion defective (LTPD) acceptance sampling plans were designed by Dodge and Romig to minimize the mean number of items inspected per lot of the process average quality when the remainder of rejected lots is inspected (rectifying plans). 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 LTPD plans for inspection by variables and attributes have been introduced in [7], 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 [6]. 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, see the analysis provided in [8]. The calculation of the LTPD variables sampling plans is implemented in the R extension package [6], providing both operating characteristics shown in [3] and [4]. Furthermore, the package covers the LTPD variables plans which are using the exponentially weighted moving average (EWMA) statistic in the inspection procedure to reflect the recent development in acceptance sampling plans design, for more details and references see [6].

The economic performance of the Dodge and Romig plans is based on good estimate of the process average quality, as discussed in [2] for the case of attribute sampling plans. The recent paper [5] showed the effects of wrong guess of the process average quality on the Average Outgoing Quality Limit (AOQL) plans for the inspection by variables. This paper considers the LTPD plans proposed in [8] 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 Dodge-Romig LTPD sampling plans for the inspection by attributes (see [1]) is recalled. Then we recall the design of the the LTPD variables sampling plans as shown in [8]. 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 then shown.

The calculation and economic evaluation of the plans is done using the free [6] software which has been published on the Comprehensive R Archive Network.

¹ Prague University of Business and Economics, Department of Mathematics,
Nám. W. Churchilla 4, Praha, Czech Republic

2 LTPD 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 [1]) 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

$$I_s = N - (N - n) \cdot L(\bar{p}; n; c) \quad (1)$$

under the condition

$$L(p_t; n; c) \leq \beta, \quad (2)$$

where $L(p, n, c)$ is the operating characteristic (the probability of accepting a submitted lot with the proportion defective p when using the plan (n, c) for acceptance sampling),

N is the number of the items in the lot,

\bar{p} is the process average quality,

p_t is the lot tolerance proportion defective ($P_t = 100p_t$ is the lot tolerance per cent defective, denoted LTPD),

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).

Condition (2) provides a guarantee for the consumer that the lots of the unsatisfactory quality level, with the proportion defective p_t are going to be accepted only with the specified probability β at most (the so-called consumer's risk). The standard value 0.1 is used for the consumer's risk in [1].

3 LTPD sampling plans for the inspection by variables

As a more economic alternative to the sampling plans for the inspection by attributes, the LTPD plans for the inspection by variables were designed in [7]. The LTPD plans were designed under the following assumptions: The measurements of a single quality characteristic X are independent and identically distributed normal random variables with unknown parameters μ and σ^2 . 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.

Select a random sample of n items in the lot, calculate sample mean \bar{x} and sample standard deviation s and then accept the lot if

$$\frac{U - \bar{x}}{s} \geq k \quad \text{or} \quad \frac{\bar{x} - S}{s} \geq k. \quad (3)$$

The exact operating characteristic for this case is (see the approximative and the exact operating characteristic in [3] and [4])

$$L(p, n, k) = \int_{k\sqrt{n}}^{\infty} g(t, n-1, u_{1-p}\sqrt{n}) dt, \quad (4)$$

where $g(t, n-1, u_{1-p}\sqrt{n})$ is probability density of the noncentral t distribution with $(n-1)$ degrees of freedom and noncentrality parameter $u_{1-p}\sqrt{n}$, where u_{1-p} is $(1-p) \cdot 100\%$ quantile 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 (4) is used as the operating characteristic.

The optimal economic characteristics shall mean that the mean inspection cost per lot of the process average quality is minimized

$$I_{ms} = N - (N - n) \cdot L(\bar{p}; n; k). \quad (5)$$

4 Economic evaluation of the plans

Let's calculate the LTPD acceptance sampling plan for sampling inspection by variables if the standard deviation of the quality characteristic is unknown in a case study below. 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 with $N = 1000$ items is considered in the acceptance procedure. The lot tolerance proportion defective is given to be $p_t = 0.01$, and the process average quality is $\bar{p} = 0.001$. Find the LTPD acceptance sampling plan for sampling inspection by variables when remainder of rejected lots is inspected.

The plan can be calculated using the functions available in the LTPDvar package for the R software [9], see the documentation of the package for a more detailed description.

The solution is $n = 85$, $k = 2.627151$. The mean inspection cost per lot of the process average quality for this plan is 104.67.

The values of the input parameters influence the resulting sampling plan and its economic characteristics. The LTPD 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} .

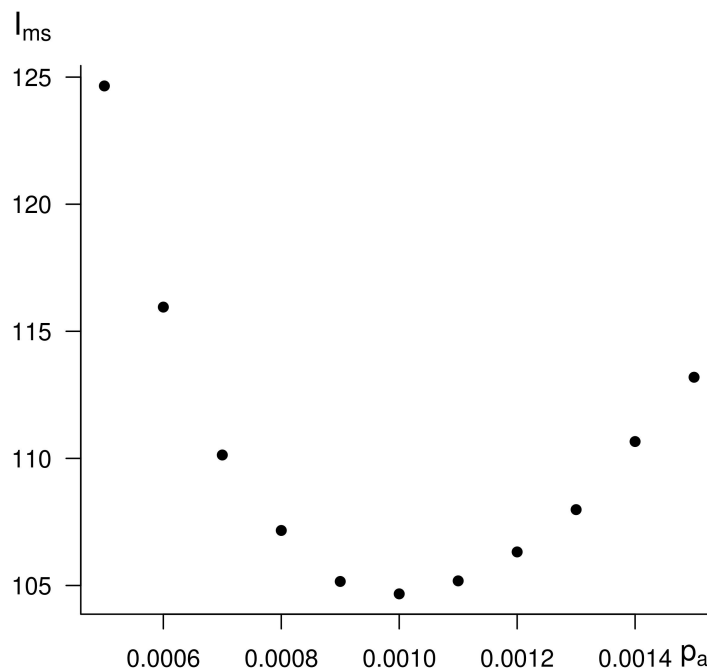


Figure 1 Inspection cost for plans constructed for various \bar{p} guess p_a , real $\bar{p} = 0.001$

Example 2. Let us change some of the input parameter values from Example 1 and consider now the lot size $N = 5000$, $\bar{p} = 0.005$.

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.

It may be observed that the outcomes are very similar to those obtained in [5] for AOQL variables sampling plans and both are in line with results shown in [2] for the corresponding Dodge-Romig attribute sampling plans.

p_a	n	k	I_{ms}
0.0005	62	2.686609	124.65
0.0006	67	2.67084	115.95
0.0007	72	2.656909	110.13
0.0008	76	2.646865	107.16
0.0009	81	2.635469	105.16
0.001	85	2.627151	104.67
0.0011	90	2.617612	105.18
0.0012	94	2.610582	106.32
0.0013	98	2.604022	107.98
0.0014	103	2.596409	110.66
0.0015	107	2.590737	113.19

Table 1 LTPD plans for process average quality guess between 0.0005 and 0.0015, real $\bar{p} = 0.001$

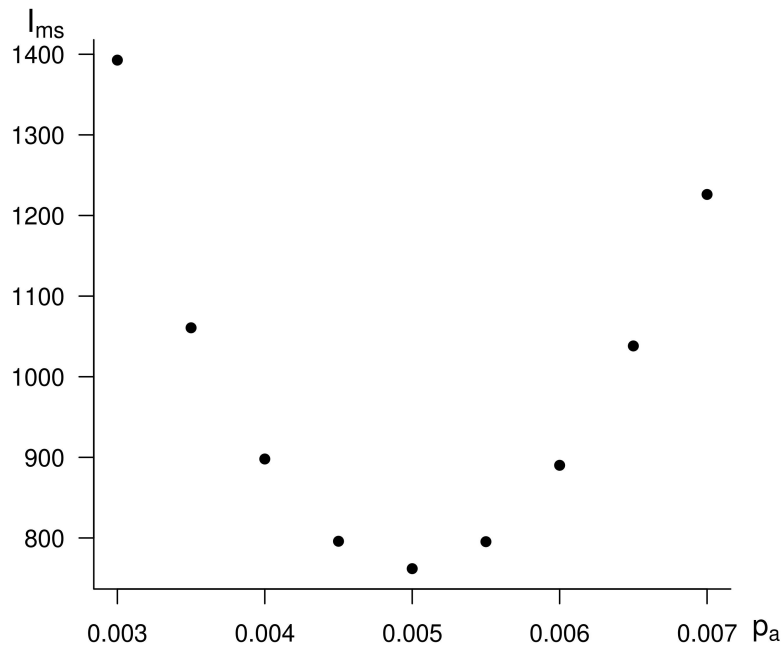


Figure 2 Inspection cost for plans constructed for various p_a , real $\bar{p} = 0.005$

p_a	n	k	I_{ms}
0.003	286	2.481788	1278.48
0.0035	345	2.467027	1060.68
0.004	414	2.454095	897.95
0.0045	497	2.442386	795.91
0.005	596	2.431858	761.91
0.0055	715	2.422303	795.37
0.006	857	2.413686	890.14
0.0065	1027	2.405877	1038.19
0.007	1223	2.399022	1226.11

Table 2 LTPD plans for process average quality guess between 0.003 and 0.007, real $\bar{p} = 0.005$

5 Conclusion

This paper addressed rectifying LTPD sampling plans for the inspection by variables 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 was 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. Based on the results of the observations in the case study in this paper it seems that it is better to overestimate the process average quality than to underestimate it. Nevertheless the results of this paper are just observations in numerical experiments and it would be interesting to find some results in analytic form in future research.

Acknowledgements

This paper has been produced with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, Prague University of Business and Economics.

References

- [1] Dodge, H., F., and Romig, H. G. (1998). *Sampling Inspection Tables: Single and Double Sampling*. New York: John Wiley.
- [2] Hald, A. (1981). *Statistical theory of sampling inspection by attributes*. New York: Academic Press.
- [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] Kaspříková, N. (2020). Remarks on economic characteristics of rectifying AOQL plans by variables. In *International Conference on Mathematical Methods in Economics 2020* (pp. 253–258). Brno : Mendel University.
- [6] 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>.
- [7] Klůfa, J. (1994). Acceptance sampling by variables when the remainder of rejected lots is inspected, *Statistical Papers*, 35, 337 – 349.
- [8] Klůfa, J. (2010). Exact calculation of the Dodge-Romig LTPD single sampling plans for inspection by variables, *Statistical Papers*, 51(2), 297-305.
- [9] R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>

Flexible Job Shop Schedule generation in Evolution Algorithm with Differential Evolution hybridisation

František Koblasa¹, Miroslav Vavroušek²

Abstract. Flexible Job Shop Scheduling becomes an emerging scheduling problem due to its nature to model constraints in holistic manufacturing systems. Its flexibility during sequencing and assigning tasks is typical for most smart factories, cyber-physical systems, new systems in distribution and procurement. There are many ways to deal with these scheduling problems, and population-based heuristics are the most common and thriving. Evolution Algorithms are the most popular as most general and practically used in many optimisation areas, while Differential Evolution principles are considered the most successful.

This paper addresses the problem representation by chromosome and schedule generation to be suitable for hybridising Evolution Algorithm optimisation with Differential Evolution principles. Semi-active, Active and Non-Delay schedules are experimentally compared on benchmark models to find their suitability to be represented by one or two chromatids chromosome. Subsequently, several Differential Evolution strategies are tested and discussed to find their suitability to be implemented as a mutation operator in the Random key-based Evolution Algorithm.

Keywords: Flexible Job Shop, Scheduling, Chromosome Representation, Evolution Algorithm, Differential Evolution.

JEL Classification: C63, L23

AMS Classification: 90C59

1 Introduction

Job shop problems reflect decision-making during production planning, and their complexity rises over time as emerging technologies and production principles are developed at a high pace. Where classical Job Shop (JSP) reflects manufacturing systems suited for customer-oriented products, Flexible Job shops (FJSP) follow constraints given by flexibility of manufacturing devices [20], ability to simulate manufacturing[13] and maintenance[5] processes with the aid of Digital Twins [18]. There is a wide variety of heuristic and metaheuristic methods to solve FJSP, beginning with fast dispatching rules over searching strategies to population-based biomathematics inspired heuristics.

Population-based metaheuristics [11] are long-time emerging techniques and the most commonly used in academic optimisation for their ease of use for various decision problems. The majority of them follow the basic pattern of Evolution, where individuals (solutions) are selected to be further modified by reproduction-recombination operations (crossover, mutation), after which old and new solutions compete or collaborate to create a new generation of better solutions. Despite a significant effort to specify them and present them as a different class of algorithms, they differ only in focus on a particular evolution operator. Genetic Algorithms [25] following a strict pattern of surviving of the fittest and recombination by crossover and mutation are often bound in literature with Boolean representation. Evolution Strategies [2] are using Real number representation of problem and mutation as recombination while focusing on the strategy of building and substituting old and new population. Differential evolution [3], with better parent-child selection and elimination strategy, uses distance in representation as a tool to recombine (evolve) individuals into the new population of successors. It is possible to get a swarm optimisation algorithm by using differential evolution while considering the same individual moving in time and space as two individual succeeding each other and keeping in mind the history of its motion (place, speed and direction). This article is not focusing on developing a new Evolution Algorithm (EA) inspired by an animal (Lion [28], Dolphin [26], Squirrel [10], Moth[17] or Bat [27] etc.) or universe (Blackhole[7], Lightning [21], Gravity [8]) behaviour, but

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.

rather on mechanical nature of itself optimisation. We propose hybridisation of the Evolution Algorithm by making new individuals with differential evolution recombination operators, which is not common in combinatorial optimisation.

Firstly, differences between classical Simple Genetic (Evolution) Algorithm (SGA) and Differential Evolution (DE) are explained and defined while pointing out its suitability and difficulties while implementing a typical level of process optimisation of DE on combinatorial problems as scheduling. Most common DE operators are explained to be further tested on benchmark problems. Secondly, the test problem class of FJSP problems is explained together with solution construction influenced by problem representation and schedule type generation. Different results (makespan and generation of convergence) given by various schedule generation and DE mutations are analysed to find the best suitable for SGA hybridisation. dGA, which follows SGA steps while using DE mutation operator, is proposed.

2 Evolution Algorithm and Differential Evolution

DE follows the general approach of EA while keeping a specific approach to select parents, reproduction cycle and creating new generation (see Figure 1). DE makes new individuals base on crossover with every individual in the population. So while DE uses every individual, EA can use a specific strategy to select new parents [15]; thus, not every individual in the population has an opportunity to reproduce. EA reproduction is made by two (rarely by more) parents recombination during crossover and mutation based on individual gene exchanges. DE first generates mutants (1-10) base on several individuals from the population and then recombine them with the parent individuals (current population) during a crossover. Finally, elimination of DE differs in rigid “parent vs child” elimination strategy where offspring substitutes parent if the value of the objective function is better [14]. In EA, parent and child can coexist in the same generation if they succeed in the elimination step (elitist in this case).

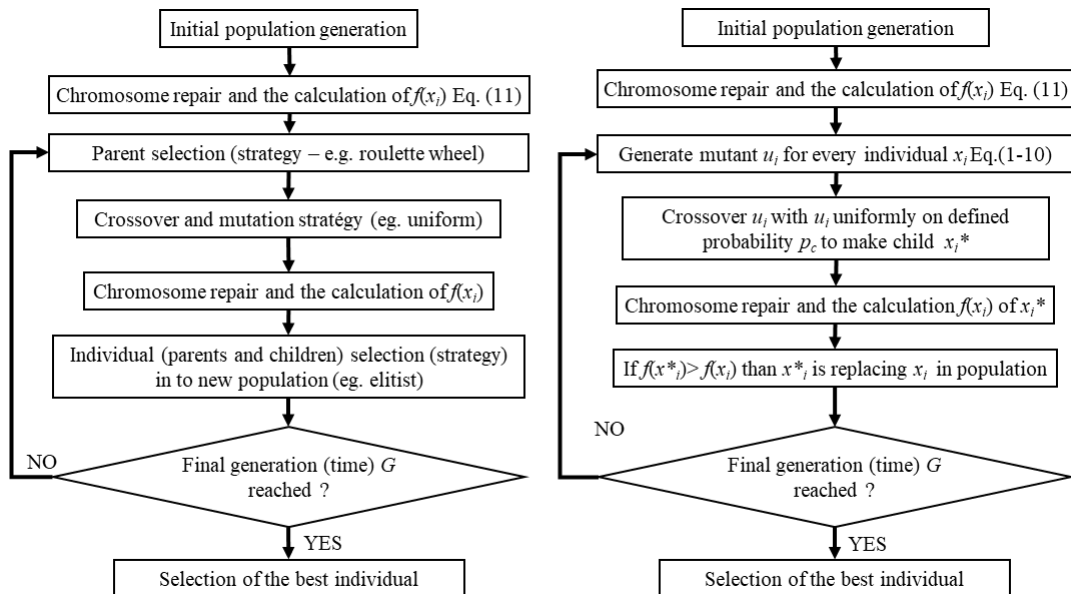


Figure 1 Simple Evolution and Differential Evolution comparison

There are various approaches to generate mutant. The “simplest” way is to generate a gene base on randomly selected individuals (1), (2), (9). The class of „faster convergence to the local extreme“ is including the gene of the fittest (3), (4), (7), (8). Some are taking beside best individuals also genes of the individual x_i to perform crossover with (5), (6). The last group is those who manipulate with randomly selected individuals (9) by sorting them by fittest $f(x)$ or performing crossover as part of mutant creation (10). Tested mutation in this paper:

$$\text{R/1} \quad u = r_1 + F(r_2 - r_3) \quad [23] \quad (1)$$

$$\text{R/2} \quad u = r_1 + F(r_2 - r_3) + F(r_4 - r_5) \quad [23] \quad (2)$$

$$\text{B/1} \quad u = x_{best} + F(r_2 - r_3) \quad [23] \quad (3)$$

$$\text{B/2} \quad u = x_{best} + F(r_2 - r_3) + F(r_4 - r_5) \quad [23] \quad (4)$$

$$\text{C2b/1} \quad u = x_i + F(x_{best} - x_i) + F(r_1 - r_2) \quad [4] \quad (5)$$

$$\text{C2b/2} \quad u = x_i + F(x_{best} - x_i) + F(r_1 - r_2) + F(r_2 - r_3) \quad [4] \quad (6)$$

$$\text{R2b/1} \quad u = r_1 + F(x_{best} - r_1) + F(r_2 - r_3) \quad [19, 23] \quad (7)$$

$$\text{R2b/2} \quad u = r_1 + F(x_{best} - r_1) + F(r_2 - r_3) + F(r_4 - r_5) \quad [23] \quad (8)$$

$$\text{Rr1} \quad u = r_1 + F(r_2 - r_3); f(r_1) > f(r_2) > f(r_3) \quad [12] \quad (9)$$

$$\text{C2r} \quad u = x_i + \text{rand}(0,1)(r_1 - x_i) + F(r_2 - r_3) \quad [16] \quad (10)$$

where u is a so-called mutant, r_i is a randomly selected individual from the population, x_i is an individual whom crossover with mutant will be performed, x_{best} is an individual with the best value of $f(x)$, $F = \langle 0; 1 \rangle$ is scaling factor. The following chapter describes the optimisation problem used to compare DE strategies (1-10) in combinatorial optimisation.

3 Flexible Job Shop scheduling problem

Flexible Job Shop (FJS) scheduling problem is a well-known NP – Hard combinatorial problem [6], which besides classical sequencing common to scheduling problems, also includes machine job assigning. This research uses the most benchmarked FJS objective function of minimising makespan [9] $f(x) = C_{max}$ (11). Equations (12) and (13) are describing the operating sequence constraint [24]. Constraint (14) guarantees machine allocation so that each operation can be processed only on one machine (unlike distributed FJS) from the machine set at one time. The constraints (15) and (16) are non-negative or 0 – 1 binary variables, which are restrictions on decision variables.

$$C_{max} = \max\{c_{in_i}\} \quad (11)$$

$$\text{s.t. } c_{ik} - c_{i(k-1)} \geq t_{ikj} x_{ikj}, k = 2, \dots, n_i, \forall i, j \quad (12)$$

$$[(c_{hg} - c_{ik} - t_{hgj}) x_{hgj}] \geq 0 \vee [(c_{ik} - c_{hg} - t_{ikj}) x_{ijk} \geq 0] \forall i, j, g, h \quad (13)$$

$$\sum_{x_{ikj} \in A_{ik}} x_{ikj} = 1 \forall i, k, j \quad (14)$$

$$c_{ik} \geq 0, \forall i, k \quad (15)$$

$$x_{ikj} \in \{0, 1\}, \forall i, k, j \quad (16)$$

Job indexes are noted as i and h , $i, h = 1, 2, \dots, n$; j is machine index, $j = 1, 2, \dots, m$; k, g are operation indexes, $k, g = 1, 2, \dots, n_i$ (where n_i is the total number of operations of job i). Processing time of k^{th} operation of job i is t_{ikj} which leads to completion time c_{ik} of operation O_{ik} . Logic variable x_{ikj} stand for machine j is selected for O_{ik} . A_{ik} is a machine set covering operation variants.

The constructive algorithm for FJSP is scheduling operation (a – for active (AS) and b - for non-delay schedules (ND) in step $t = \{1, n\}$, where n is the total number of operations:

1. Creating list V_t of schedulable O_{ik} operations, including all machine variants.
2. Conflict set creation
 - a) Find possible earliest ending time $f_t^* = \min_{O_{k \text{ in } V_t} \{c_{ik}\}}$ and machine M^* on which c_{ik} occurs
 - b) Find possible earliest starting time $s_t^* = \min_{O_{k \text{ in } V_t} \{s_{ik}\}}$ and machine M^* on which c_{ik} occurs.
3. Choose optimal operation which requires M^* and:
 - a) its starting time $c_{i(k-1)} < c_{ik}$
 - b) its starting time $s_t^* = s_{ik}$
4. Continue until there is O_{ik} unscheduled

It is expected to get different quality of results of makespan (11) by different schedule generations as AS guarantees search in the neighbourhood where the optimal solution lies. It is impossible to construct another schedule by changing the processing order on the machines and having at least one job/operation finishing earlier and no job/operation finishing later. The subset of AS are ND schedules that can find the optimal solution. However, it is not guaranteed as it is constructed to no machine is kept idle while a job/an operation is waiting for processing. ND is included in the original comparison as most used in practice. Semi-active schedules are often used without recognition [29] (as well as active [22]) in scientific articles while solving FJSP by pair chromosome where the first represents job sequence, the second machine assignment. Semiactive schedules (SA) are still widely used as they allow optimisation of job assignment while having larger search space than AS or ND. In SA no job/operation can be finishing earlier without changing the order of processing on any one of the machines,

4 Random key-based SGA – dGA - DE

Testing efficiency of Differential approaches to manipulate with genes (1) in EA is done by comparison of classical SGA and DE with hybrid dGA (Table 1) with a sequence of operators given by Figure 1 with the following setup:

- Generations – up to the moment of population convergence or 200 generations without improvement of $f(x)$; the goal is to find how fast the algorithm converges.
- Crossover – uniform crossover with probability $p_c = 0.85$ that new individual will share gene of 2nd parent. SGA uses classical crossover, DE and dGA use mutation-crossover mechanism (1-10).
- Population size $N = 2JM + 100$, where J is the number of jobs and M is the number of machines.
- Representation - Random key (RK) representation [1] is used for sequencing and new machine representation based on RK is introduced to substitute the usual Integer based representation of machine assignment so it can be optimised by DE mutant generation.

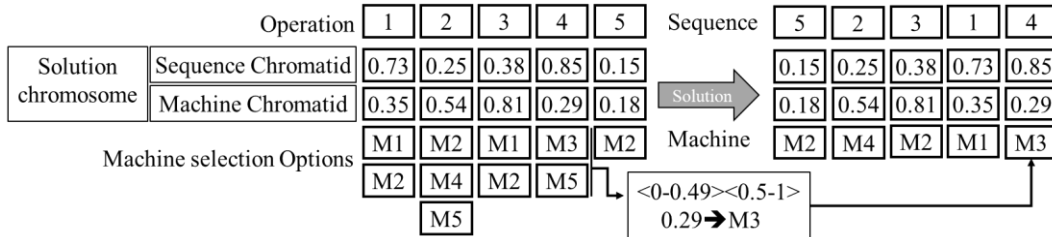


Figure 2 Sequencing and machine selection (assigning) by RK chromosome

- Parent selection – SGA and dGA – Roulette wheel. In DE all individuals in the population are selected.
- Elimination – SGA and dGA use the Elitist strategy surviving of the fittest N , while DE uses parent-offspring tournament selection.
- Schedule generation – All algorithms are tested with SA, AS, ND schedule generations. AS and ND are not dealing with job assigning by chromosome as it is done by the earliest possibly finished operation (step 2) while SA strictly uses chromosome information to reduce FJSP to JSP.

Model	JxM	LB	Best	Schedule	SGA		DE		dGA		SGA		DE		dGA	
					f(x)b	f(x)a	f(x)b	f(x)a	f(x)b	f(x)a	C(x)b	C(x)a	C(x)b	C(x)a	C(x)b	C(x)a
Mk01	10x6	36	40	SA	40	41.8	40	41.5	40	41.6	80	56	907.0	522.3	1052.0	530.3
				A	41	41.9	41	41.8	40	41.5	176	111	622.0	342.7	521.0	356.3
				ND	41	42	41	41.8	41	41.9	120	71.8	409.0	251.8	422.0	295.8
Mk02	10x6	24	26	SA	29	29.7	36	37.7	29	30.5	163	131	919.0	387.2	1046.0	794.0
				A	31	33.3	33	35.5	31	33.7	402	275	787.0	312.8	985.0	447.7
				ND	32	33.3	32	33.3	31	32.6	310	201	566.0	281.3	597.0	394.4
Mk03	15x8	204	204	SA	204	204	211	218.5	204	204	89	80.6	1195.0	620.0	1577.0	840.3
				A	204	204	204	204.4	204	204	449	305	664.0	448.6	579.0	376.9
				ND	204	204	204	204	204	204	243	223	330.0	250.0	447.0	290.8
Mk04	15x8	48	60	SA	62	65.2	66	68.8	62	66.3	246	151	1119.0	545.5	997.0	697.5
				A	67	69.3	68	73	67	70.6	524	323	703.0	411.1	780.0	489.5
				ND	66	66.9	67	67.3	66	66.8	353	242	757.0	335.3	696.0	432.0
Mk05	15x4	168	172	SA	173	176.1	180	182.3	173	176.6	282	164	985.0	624.6	1236.0	693.4
				A	182	185.1	187	190.1	183	187.3	727	497	747.0	464.3	894.0	514.2
				ND	177	179.3	179	181.3	177	179	626	452	677.0	415.1	955.0	542.3
Mk06	10x15	33	57	SA	65	68.9	87	99.7	65	67.3	271	224	1497.0	819.8	1833.0	1333.6
				A	78	81.4	81	84.6	79	83.2	538	434	691.0	342.5	860.0	460.0
				ND	74	79	79	81.9	77	81.4	572	365	657.0	346.6	891.0	418.3
Mk07	20x5	133	139	SA	144	147	157	168.1	144	147.2	241	164	1423.0	527.0	1826.0	1132.8
				A	164	173.4	175	180.9	173	177.5	650	476	843.0	344.4	792.0	471.9
				ND	164	172.6	172	177.1	169	175.5	755	474	770.0	333.2	865.0	483.6
Mk08	20x10	523	523	SA	523	523	523	523	523	523	89	78.3	951.0	514.0	1069.0	643.3
				A	523	523	523	525.1	523	523.5	412	324	892.0	540.9	827.0	517.4
				ND	523	523	523	523	523	523	50	46.4	231.0	211.9	228.0	210.6
Mk09	20x10	299	307	SA	307	316.2	369	388	311	330.7	320	245	1475.0	692.1	1916.0	1322.2
				A	373	385.8	385	395	375	385.1	725	550	818.0	433.5	819.0	511.1
				ND	321	328.7	327	336.6	322	327.7	810	586	898.0	557.2	1112.0	629.8
Mk10	20x15	165	197	SA	219	226.3	323	342.8	278	293.4	475	401	936.0	697.3	1923.0	1226.9
				A	298	304.4	306	314.3	298	304.3	614	446	869.0	313.1	825.0	502.6
				ND	259	263.1	260	264.9	255	260.8	549	301	754.0	150.5	841.0	447.3

Table 1 Test results of MK1-MK10 problem by SGA, DE and dGA evolution algorithms

Average best $f(x)_a$ (marked yellow in Table 1) in Table 1 shows the best average of particular best mutation (for DE and dGA), not the average of all mutations. It is not possible to show detailed results as the size of this paper is limited. Results considering the objective function of makespan (total competition time i.e. production lead time) (11) show that best average results $f(x)_a$ based on 10 replications are obtained with Semi-active schedules. The only exceptions are problems of MK9-10 (Table 1 - expected better but worse result is orange) where DE and dGA non-delay schedules give us better results. That means we can obtain regularly better results during optimisation by manipulating not only the sequence of operations as in Active and Non-delay schedules. Semi-active schedules offer us bigger flexibility in multicriteria optimisation, taking into account also other possible objective functions based on analysing machines (Utilisation, ROI, OEE).

We can assume that Semi-active schedules using dGA and DE algorithms give us the best results of $C(x)$, which is representing the generation of population convergence to one solution (so further optimisation is not possible). On the other hand, active schedules are the ones with the slowest convergence in the case of SGA (highest $C(x)$).

The most desired algorithm property of late convergence (high $C(x)$) shows superiority of dGA as it converges six times later than SGA in average and 1.38 later than DE. It has to be pointed out that dGA has not converged in most cases and experiments were ended after 200 generations without improvement of $f(x)$. The main influence on $C(x)$ has probably itself differential approach to generate new individuals. Another possible influence which can be the synergy of differential evolution recombination and Elitist elimination is unlike; however, it would be necessary to do the experiment in which SGA would use parent versus child elimination to prove this hypothesis.

Comparing SGA, DE and dGA in terms of $f(x)_b$ with theoretical lower bound (LB) and the best known solutions, it can be assumed that SGA has not significantly but still better results. All of the tested algorithms have found optimum in the case of MK03 and MK08 ($f(x)_a = f(x)_b$).

One of the important goals was to find out if there is some dominant DE mutation operator (1-10). Results have shown that none of the mutations shows significantly better $f(x)_a$ or $f(x)_b$. However, R1 (1) and C2b/1 (5) were the most frequent mutations finding the best and average best of 10 replications results in both DE and dGA.

5 Conclusion

Our research found that Semi-active schedules generate the best results except for MK09-10. This was not expected as Active schedules generation searches in the neighbourhood which expect to have an optimal schedule. Proposed hybridisation of SGA by DE mutation as recombination operator shows that results of $f(x)$ are not much different. However, dGA and DE converges much slower than SGA, which opens possibilities in long run optimisations. None of DE mutations has shown dominance in both convergence and $f(x)$; however, R1 and C2b/1 with SA schedules were the most frequent in getting the best results.

Further research will focus on improving the proposed dGA in job assigning part of the problem. It has strong potential in general optimisation thanks to not significantly worst results in combinatorial optimisation and known good results of DE in processing optimisation.

Acknowledgements

This work was supported by the Student Grant Competition of the Technical University of Liberec under the project Optimisation of manufacturing systems, 3D technologies and automation No. SGS-2019-5011

References

- [1] Bean, J. C. (1994). Genetic algorithms and random keys for sequencing and optimisation. *ORSA journal on computing*, 6(2), 154–160.
- [2] Beyer, H.-G., & Schwefel, H.-P. (2002). Evolution strategies – A comprehensive introduction. *Natural Computing*, 1(1), 3–52. <https://doi.org/10.1023/A:1015059928466>
- [3] Bilal, Pant, M., Zaheer, H., Garcia-Hernandez, L., & Abraham, A. (2020). Differential Evolution: A review of more than two decades of research. *Engineering Applications of Artificial Intelligence*, 90, 103479. <https://doi.org/10.1016/j.engappai.2020.103479>
- [4] Das, S., & Suganthan, P. N. (2010). Differential evolution: A survey of the state-of-the-art. *IEEE transactions on evolutionary computation*, 15(1), 4–31.

- [5] Fusko, M., Rakyta, M., Krajcovic, M., Dulina, L., Gaso, M., & Grznar, P. (2018). Basics of Designing Maintenance Processes in Industry 4.0. *MM Science Journal*, 2018, 2252–2259. https://doi.org/10.17973/MMSJ.2018_03_2017104
- [6] Garey, M. R., Johnson, D. S., & Sethi, R. (1976). Complexity of flow shop and job shop scheduling. *Mathematics of Operations Research*, 1(2), 117–129. Scopus. <https://doi.org/10.1287/moor.1.2.117>
- [7] Hatamlou, A. (2013). Black hole: A new heuristic optimisation approach for data clustering. *Information sciences*, 222, 175–184.
- [8] He, S., Zhu, L., Wang, L., Yu, L., & Yao, C. (2019). A modified gravitational search algorithm for function optimisation. *IEEE Access*, 7, 5984–5993.
- [9] Chaudhry, I. A., & Khan, A. A. (2016). A research survey: Review of flexible job shop scheduling techniques. *International Transactions in Operational Research*, 23(3), 551–591. <https://doi.org/10.1111/itor.12199>
- [10] Jain, M., Singh, V., & Rani, A. (2019). A novel nature-inspired algorithm for optimisation: Squirrel search algorithm. *Swarm and evolutionary computation*, 44, 148–175.
- [11] Jedrzejowicz, P. (2019). Current Trends in the Population-Based Optimisation. In N. T. Nguyen, R. Chbeir, E. Exposito, P. Anioté, & B. Trawiński (Ed.), *Computational Collective Intelligence* (s. 523–534). Springer International Publishing. https://doi.org/10.1007/978-3-030-28377-3_43
- [12] Kaelo, P., & Ali, M. M. (2006). Some variants of the controlled random search algorithm for global optimisation. *Journal of optimisation theory and applications*, 130(2), 253–264.
- [13] Kliment, M., Trebuna, P., Pekarčíková, M., Straka, M., Trojan, J., & Duda, R. (2020). Production Efficiency Evaluation and Products' Quality Improvement Using Simulation. *International Journal of Simulation Modelling*, 19(3), 470–481. <https://doi.org/10.2507/IJSIMM19-3-528>
- [14] Koblasa, F., Králiková, R., & Votrubec, R. (2020). Influence of EA control parameters to optimisation process of FJSSP problem. *International Journal of Simulation Modelling*, 19(3), 387–398. <https://doi.org/10.2507/IJSIMM19-3-519>
- [15] Koblasa, F., Vavroušek, M., & Manlig, F. (2020). Selection Strategies in Evolution Algorithms and Biased Selection with Incest Control. *38th International Conference on Mathematical Methods in Economics 2020*.
- [16] Mallipeddi, R., Suganthan, P. N., Pan, Q.-K., & Tasgetiren, M. F. (2011). Differential evolution algorithm with ensemble of parameters and mutation strategies. *Applied soft computing*, 11(2), 1679–1696.
- [17] Mirjalili, S. (2015). Moth-flame optimisation algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-Based Systems*, 89, 228–249. <https://doi.org/10.1016/j.knosys.2015.07.006>
- [18] Pekarčíková, M., Trebuňa, P., Kliment, M., Edl, M., & Rosocha, L. (2020). Transformation the Logistics to Digital Logistics: Theoretical approach. *Acta Logistica*, 7(4), 217–223. <https://doi.org/10.22306/al.v7i4.174>
- [19] Qin, A. K., Huang, V. L., & Suganthan, P. N. (2008). Differential evolution algorithm with strategy adaptation for global numerical optimisation. *IEEE transactions on Evolutionary Computation*, 13(2), 398–417.
- [20] Sevic, M., & Keller, P. (2019). Design of Cnc Milling Machine as a Base of Industry 4.0 Enterprise. *Mm Science Journal*, 2019, 3555–3560. https://doi.org/10.17973/MMSJ.2019_12_2019042
- [21] Shareef, H., Ibrahim, A. A., & Mutlag, A. H. (2015). Lightning search algorithm. *Applied Soft Computing*, 36, 315–333.
- [22] Sriboonchandr, P., Kriengkarakot, N., & Kriengkarakot, P. (2019). Improved Differential Evolution Algorithm for Flexible Job Shop Scheduling Problems. *Mathematical and Computational Applications*, 24(3), 80. <https://doi.org/10.3390/mca24030080>
- [23] Storn, R., & Price, K. V. (1997). Differential Evolution—a simple and efficient heuristic for global optimisation over continuous spaces—J. of Global Optimisation. *Journal of*, 11.
- [24] Sun, L., Lin, L., Wang, Y., Gen, M., & Kawakami, H. (2015). A Bayesian Optimisation-based Evolutionary Algorithm for Flexible Job Shop Scheduling. *Procedia Computer Science*, 61, 521–526. <https://doi.org/10.1016/j.procs.2015.09.207>
- [25] Vose, M. D. (1999). *The simple genetic algorithm: Foundations and theory*. MIT press.
- [26] Wu, T., Yao, M., & Yang, J. (2016). Dolphin swarm algorithm. *Frontiers of Information Technology & Electronic Engineering*, 17(8), 717–729.
- [27] Yang, X.-S. (2010). A new metaheuristic Bat-inspired Algorithm. *Studies in Computational Intelligence*, 284, 65–74. Scopus. https://doi.org/10.1007/978-3-642-12538-6_6
- [28] Yazdani, M., & Jolai, F. (2016). Lion Optimization Algorithm (LOA): A nature-inspired metaheuristic algorithm. *Journal of Computational Design and Engineering*, 3(1), 24–36. <https://doi.org/10.1016/j.jcde.2015.06.003>
- [29] Zhang, G., Hu, Y., Sun, J., & Zhang, W. (2020). An improved genetic algorithm for the flexible job shop scheduling problem with multiple time constraints. *Swarm and Evolutionary Computation*, 54, 100664. <https://doi.org/10.1016/j.swevo.2020.100664>

Distortion risk measures in portfolio optimization

Miloš Kopa¹, Juraj Zelman²

Abstract. The paper deals with mean-risk problems where the risk is modeled by a distortion measure. This measure could be seen as a generalization of Conditional Value-at-Risk or Expected shortfall. If the associated distortion function is concave the measure is coherent. We analyze several distortion measures for different choices of a concave distortion function. First, assuming a discrete distribution of returns, we identify the efficient frontier. Then we compute the portfolio maximizing reward-risk ratio. Finally, we compare the results for various distortion measures among each other.

Keywords: portfolio optimization, distortion risk measure, efficient frontier, performance ratio

JEL Classification: D81, G11

AMS Classification: 91B16, 91B30

1 Introduction

Historically, distortion risk measures have their roots in the dual theory of choice under uncertainty proposed by [13] and were later developed by the axiomatic approach in [11]. The idea behind the distortion risk measure is the transformation of the given probability measure in order to quantify the tail risk more accurately and therefore give more weight to higher risk events. The motivation for distorting a probability measure arose from numerous studies on risk perception, such as the work [4], who observed that people evaluate risk as a non-linear distorted function rather than a linear function of the probabilities. Originally, distortion risk measures found their application in the insurance problems. For example, [10] presented an approach to insurance pricing using the proportional hazards transform. However, due to the relation between insurance and investment risks, distortion risk measures started to be also used in the investment context and portfolio selection problems (see for example [9]). Perhaps interesting could be a relation to stochastic dominance which is an attractive tool for random returns comparisons in various applications, see e.g. [7], [3], or [5] for recent applications of stochastic dominance in pension fund management.

The remainder of this paper is structured as follows. Section 2 presents a notation and basic properties of the distortion risk measures. It is followed by a formulation of reward-risk ratio model based on distortion measures of risk in Section 3. Empirical study is presented in Section 4 and the paper is concluded in Section 5.

2 Distortion risk measures

In the whole text, we assume that \mathcal{X} is a set of random variables on a probability space (Ω, \mathcal{F}, P) . A random variable $X \in \mathcal{X}$ represents a loss random variable (typically, positive values are associated with losses and negative values represent gains) of some financial asset over a time interval of length $T \in \mathbb{R}_+$.

Definition 1. ([2]) Suppose that $g : [0, 1] \rightarrow [0, 1]$ is a non-decreasing function such that $g(0) = 0$ and $g(1) = 1$ (also known as the **distortion function**) and $X \in \mathcal{X}$ with a distribution function $F_X(x)$. Then, the **distortion risk measure** associated with the distortion function g is defined as

$$\rho_g(X) = - \int_{-\infty}^0 [1 - g(1 - F_X(x))] dx + \int_0^{\infty} g(1 - F_X(x)) dx,$$

provided that at least one of the integrals is finite.

When we define the **decumulative distribution function** (also known as the **survival function**) $S_X(x) = 1 - F_X(x) = P(X > x)$ and we use it instead of the distribution function, we obtain

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

² Charles University, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, zelman.juraj@gmail.com

$$\rho_g(X) = - \int_{-\infty}^0 [1 - g(S_X(x))] dx + \int_0^{\infty} g(S_X(x)) dx.$$

The interpretation of this definition is that the distortion measure represents the expectation of a new random variable with re-weighted probabilities. In some cases, such as problems related to insurance or capital requirements, it is appropriate to assume that the random variable $X \in \mathcal{X}$ is non-negative. In this case, when $X \in \mathcal{X}$ is a non-negative random variable, then ρ_g reduces to

$$\rho_g(X) = \int_0^{\infty} g(S_X(x)) dx.$$

The class of distortion risk measures is prospective, because distortion measures, in the general case, fulfill the conditions of monotonicity, positive homogeneity and translation invariance.

Theorem 1. ([8]) (Monotonicity) Suppose that $X, Y \in \mathcal{X}$ and $X \leq Y$. Then $\rho_g(X) \leq \rho_g(Y)$.

(Positive homogeneity) For a distortion risk measure ρ_g , $X \in \mathcal{X}$ and $\lambda \geq 0$: $\rho_g(\lambda X) = \lambda \rho_g(X)$.

(Translation invariance) For a distortion risk measure ρ_g and $X \in \mathcal{X}$ it holds that $\forall c \in \mathbb{R} : \rho_g(X + c) = \rho_g(X) + c$.

Theorem 2. ([12]) The distortion risk measure $\rho_g(X)$ is sub-additive

$$\rho_g(X + Y) \leq \rho_g(X) + \rho_g(Y),$$

if and only if g is a concave distortion function.

Summarizing, a distortion risk measure $\rho_g(X)$ is coherent iff g is a concave distortion function.

In the following example, we present the representations of risk measures Value-at-Risk (VaR) and Expected Shortfall (ES) as distortion risk measures.

Example 1. Suppose that $X \in \mathcal{X}$, $\alpha \in (0, 1)$. If

$$g(x) = \begin{cases} 0 & \text{if } 0 \leq x < 1 - \alpha \\ 1 & \text{if } 1 - \alpha \leq x \leq 1. \end{cases}$$

then $VaR_\alpha(X) = \rho_g(X)$ and if

$$g(x) = \min\left(\frac{x}{1 - \alpha}, 1\right), \text{ where } x \in [0, 1].$$

then, $ES_\alpha(X) = \rho_g(X)$.

Another example of distortion risk measure includes the **Proportional Hazard (PH) transform** proposed by [10] as a new risk-adjusted premium for insurance risk pricing. This measure has a distortion function

$$g(x) = x^{1/\gamma}, \quad x \in [0, 1], \gamma \geq 1. \quad (1)$$

Consequently, we define the **PH-transform measure** as:

$$\rho_{PH}(X) = \int_0^{\infty} S_X(x)^{1/\gamma} dx, \quad \gamma \geq 1,$$

where $S_X(x) = 1 - F_X(x)$ is defined as previously.

As we can see from the definition of the distortion function g of the PH transform, this function is concave and therefore, the PH-transform measure satisfies the sub-additivity property. As [10] mentions, this is an important property as it does not provide any advantage to policy-holders when splitting the risk of their positions into pieces.

Another well known examples of distortion functions which generate coherent risk measures include:

- The **Wang transform** ([11])

$$g_\lambda(x) = \Phi(\Phi^{-1}(x) + \lambda) \quad \text{for } x \in [0, 1], \lambda \geq 0,$$

where Φ is the standard normal distribution function.

- The **MINVAR** distortion function ([1])

$$g(x) = 1 - (1 - x)^{1+\lambda} \quad \text{for } x \in [0, 1], \lambda \geq 0. \quad (2)$$

- The **MINMAXVAR** distortion function ([1])

$$g(x) = 1 - (1 - x^{1/(1+\lambda)})^{1+\lambda} \quad \text{for } x \in [0, 1], \lambda \geq 0.$$

3 Reward-risk ratio

Suppose that we have a discrete real random variable Y , representing losses (in percent), with possible values $y_1, \dots, y_m \in \mathbb{R}$, where $y_1 \leq y_2 \leq \dots \leq y_m$. As we need to separate these values to negative and non-negative, assume that the index $k \in \{0, \dots, m\}$ is such that values y_1, y_2, \dots, y_k are negative and y_{k+1}, \dots, y_m are non-negative (where for $k = 0$ we understand that all values are non-negative and for $k = m$ are all negative). For the simplicity, we assume that $\forall i \in \{1, \dots, m\} : P(Y = y_i) = \frac{1}{m}$. Then, we know that its cumulative distribution function is $F_Y(y) = \frac{1}{m} \sum_{i=1}^m \mathbf{1}_{\{y_i \leq y\}}$, where $\mathbf{1}_A$ denotes an indicator function of a set A . This means that $F_Y(y)$ is constant on intervals $(-\infty, y_1), [y_1, y_2), \dots, [y_m, \infty)$. Thus, from Definition 1 of a distortion measure ρ_g , we can derive that

$$\begin{aligned} \rho_g(Y) &= - \sum_{i=1}^{k-1} (y_{i+1} - y_i) \left[1 - g\left(1 - \frac{i}{m}\right) \right] + y_k \left[1 - g\left(1 - \frac{k}{m}\right) \right] + \\ &+ y_{k+1} g\left(1 - \frac{k}{m}\right) + \sum_{i=k+1}^{m-1} (y_{i+1} - y_i) g\left(1 - \frac{i}{m}\right) = y_1 + \sum_{i=1}^{m-1} (y_{i+1} - y_i) g\left(1 - \frac{i}{m}\right). \end{aligned}$$

Therefore, to compute distortion risk measure ρ_g for a discrete random variable Y , it is sufficient to have all the possible values y_i , where $i \in \{1, \dots, m\}$, ordered. We do not need to differentiate between non-negative and negative values. Now we can focus on the formulation of the reward-risk optimization problem.

Assume that we have $m \in \mathbb{N}$ time periods (e.g. weeks) numbered $1, \dots, m$ and $n \in \mathbb{N}$ financial assets $1, \dots, n$. Let $l = (l_{ij})_{i=1, j=1}^{n, m} \in \mathbb{R}^{n \times m}$, $m, n \in \mathbb{N}$ be a matrix, where l_{ij} represents a concrete realization of a loss of i -th financial asset at time j . Suppose that $w = (w_1, \dots, w_n)^T \in \mathbb{R}^n$ denotes weights of a portfolio associated to our financial assets such that $w^T e = 1$ and $\forall i \in \{1, \dots, n\} : w_i \geq 0$ (we do not allow short sales).

For a given vector of weights w , we can calculate a vector of gross losses for this portfolio as $\tilde{l}_p = (w^T \tilde{l})^T \in \mathbb{R}^m$ where loss matrix l is substituted by gross losses \tilde{l} by adding one (e.g. the value 1,1 represents 10% loss and value 0,9 represents 10% return). Equivalently the j -th position of the vector \tilde{l}_p is equal to re-weighted sum of assets' gross losses at time j or $\sum_{i=1}^n w_i \tilde{l}_{ij}$. However, as we see from the previous part, where we derived the formula for distortion measure, to calculate values of risk measure for different portfolios, we need to first re-order the values of \tilde{l}_p . Therefore, in our optimization problem we need to define a permutation matrix $P = (p_{i,j})_{i=1, j=1}^{m, m}$ consisting of 0 and 1 such that the sum in every row and column is equal to 1. Then, we can define a new vector $\tilde{y} = (\tilde{y}_1, \dots, \tilde{y}_m) \in \mathbb{R}^m$ such that it has the same values as \tilde{l}_p , but its values are ordered from the lowest to the highest. Finally, \tilde{Y}_- denotes gross return of \tilde{Y} (representing gross loss).

If we define a variable R representing the reciprocal value of a distortion reward-risk ratio (minimization over a reciprocal value of a reward-risk ratio is equivalent to maximization of a reward-risk ratio), we can formulate the distortion reward-risk optimization problem as

$$\begin{aligned}
& \underset{w}{\text{minimize}} && R \\
& \text{subject to} && \rho_g(\tilde{Y}) = \mu(\tilde{Y}_-) \times R \\
& && \tilde{l}_p = (w^T \tilde{l})^T \\
& && P \tilde{l}_p = \tilde{y}, \text{ where } P = (p_{i,j})_{i=1,j=1}^{m,m} \\
& && \sum_{i=1}^m p_{ij} = 1 \quad \forall j \in \{1, \dots, m\} \\
& && \sum_{j=1}^m p_{ij} = 1 \quad \forall i \in \{1, \dots, m\} \\
& && p_{ij} \in \{0, 1\} \quad \forall i, j \in \{1, \dots, m\} \\
& && \tilde{y}_1 \leq \tilde{y}_2 \leq \dots \leq \tilde{y}_m \\
& && w^T e = 1 \\
& && w \geq 0.
\end{aligned} \tag{3}$$

4 Empirical study

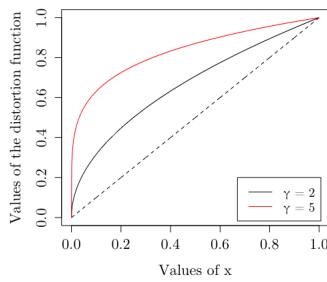
To demonstrate our model, we selected ten stocks (A1 - A10), which are traded at stock exchanges NYSE and Nasdaq, see Table 1. We restrict to a smaller sample of weekly adjusted closing prices ranging from 2020-12-21 to 2021-02-22. A smaller sample was selected due to the computational complexity of our model, which leads to a non-linear mixed-integer optimization problem.

Asset	Company	Ticker	GICS Sector
A1	Microsoft Corp.	MSFT	Information Technology
A2	Intel Corp.	INTC	
A3	Goldman Sachs Group	GS	Financials
A4	BlackRock	BLK	
A5	Alphabet Inc.	GOOGL	Communication Services
A6	AT&T Inc.	T	
A7	Amazon.com, Inc.	AMZN	Consumer Discretionary
A8	Johnson & Johnson	JNJ	Health Care
A9	General Electric	GE	Industrials
A10	Exxon Mobil Corp.	XOM	Energy

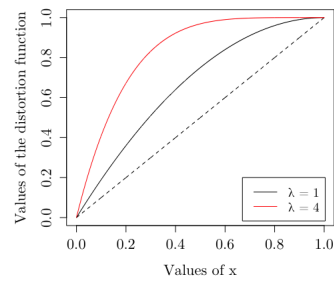
Table 1: Selected assets and their corresponding GICS sectors

In our implementation, we focused on two distortion risk measures. The Proportional Hazard transform (defined in (1)) for two different parameters $\gamma = 2$ and $\gamma = 5$ and the MINVAR distortion risk measure (defined in (2)) for two parameters $\lambda = 1$ and $\lambda = 4$. For better illustration of the position of the portfolio with the highest reward-risk ratio, we present it with resulting efficient frontiers in Figures 2a and 2b and Tables 2 and 3 with allocations of the optimal portfolios.

As can be seen in Figure 2a, different choices of parameter γ does not only affect the position of the efficient frontiers but influences their shape as well. This is the result of the shapes of Proportional Hazard functions depicted in Figure 1a. As we can see, these functions assign higher values especially to lower values of x . Thus, the corresponding risk measure assigns higher probabilities to realizations with the highest losses. This effect is noticeable especially on the portfolios beyond the highest reward-risk ratio portfolio, where risks grow significantly faster than in the previous part of the efficient frontier. Therefore, different choices of parameters allow us to model various levels of risk perception and to construct optimal portfolios with respect to these levels. Moreover, as can be seen from Table 2, the optimal portfolios with the lowest risk and the highest reward-risk ratio differ significantly. Not only with respect to their values of risk but regarding their allocations as well. Similar results are obtained for the MINVAR distortion function. In this case, different choices of parameter λ do not only lead to different values of risk but also to different allocations of optimal portfolios. These differences can be noticed from Table 3. The effect on the shapes of efficient frontiers and their positions is depicted in Figure 2b. As we can see, the



(a) Proportional Hazard transform



(b) MINVAR distortion function

Figure 1: Selected distortion measures for different parameters

Proportional Hazard transform, $\gamma = 2$							
Return	A1	A2	A3	A10	Risk	RRR	Optimum
1,93%	0,386	0,310	0,024	0,280	0,992774	1,026696	Min Risk
2,68%	0	0,860	0	0,140	0,993617	1,033354	Max RRR

Proportional Hazard transform, $\gamma = 5$							
Return	A1	A2	A9	A10	Risk	RRR	Optimum
1,28%	0,537	0,071	0,294	0,098	0,99964	1,013188	Min Risk
2,54%	0,071	0,759	0	0,170	1,009303	1,015921	Max RRR

Table 2: Optimal portfolios with respect to the Proportional Hazard transform with corresponding mean returns, risks and reward-risk ratios (RRR).

shapes of MINVAR distortion functions from 1b are translated into the shapes of efficient frontiers. Therefore, in comparison to the PH measure, we also obtain different allocations of optimal reward-risk portfolios.

MINVAR distortion function, $\lambda = 1$							
Return	A1	A2	A3	A10	Risk	RRR	Optimum
1,93%	0,399	0,264	0,187	0,150	0,993088	1,026426	Min Risk
2,82%	0	0,401	0,599	0	0,994221	1,034163	Max RRR

MINVAR distortion function, $\lambda = 4$								
Return	A1	A2	A3	A9	A10	Risk	RRR	Optimum
1,32%	0,471	0,155	0	0,374	0	1,0021	1,0111	Min Risk
1,90%	0,421	0,169	0,211	0	0,200	1,0047	1,0142	Max RRR

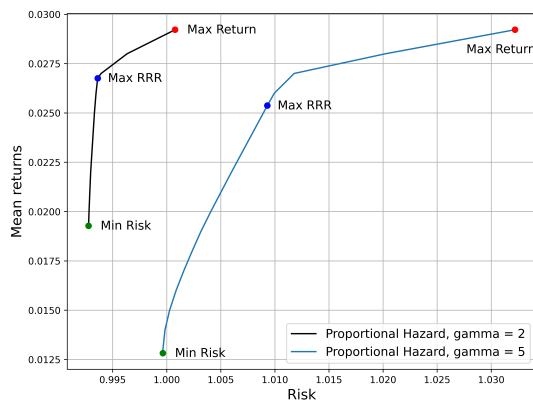
Table 3: Optimal portfolios with respect to the MINVAR distortion function with corresponding mean returns, risks and reward-risk ratios (RRR).

5 Conclusions

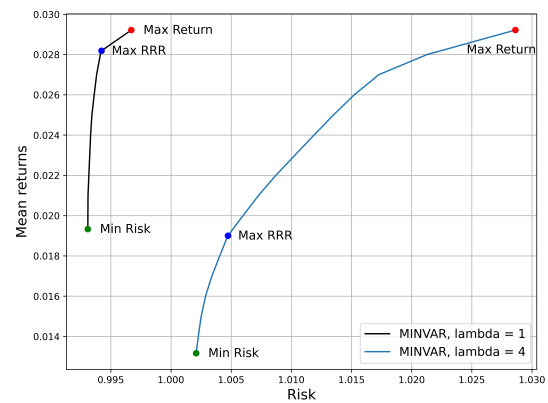
The paper presents a tractable approach to portfolio optimization using distortion risk measures which could be seen as generalizations of Value-at-Risk and Expected Shortfall. In the empirical study, two different formulations (mean-risk and risk-reward ratio) and four different distortion measures are considered. The corresponding efficient frontiers and reward-risk maximizing portfolios are compared. Although the paper presents only static model, the distortion measures could be similarly applied to multistage models with exogenous [14], [15] or endogenous randomness [6].

Acknowledgements

The paper was supported by the grant No. 19-28231X of the Czech Science Foundation.



(a) Proportional Hazard transform



(b) MINVAR distortion function

Figure 2: The efficient frontiers. Portfolios with the highest return, the highest reward-risk ratio and the lowest risk are highlighted

References

- [1] Cherny, A. and Madan, D. (2009). New measures for performance evaluation. *The Review of Financial Studies*, 22, 2571–2606.
- [2] Dhaene, J., Kukush, A., Linders, D. and Tang, Q. (2012). Remarks on quantiles and distortion risk measures. *European Actuarial Journal*, 2, 319–328.
- [3] Kabasinkas, A., Sutiene, K., Kopa, M., Luksys, K. and Bagdonas, K. (2020). Dominance-Based Decision Rules for Pension Fund Selection under Different Distributional Assumptions. *Mathematics*, 8, n.719.
- [4] Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47, 263–292.
- [5] Kopa, M., Kabasinkas, A. and Sutiene, K. (2021). A stochastic dominance approach to pension-fund selection, *IMA Journal of Management Mathematics*, <https://doi.org/10.1093/imaman/dpab002>.
- [6] Kopa, M., and Rusý, T. (2021). A decision-dependent randomness stochastic program for asset–liability management model with a pricing decision. *Annals of Operations Research*, 299, 241–271.
- [7] Moriggia, V., Kopa, M. and Vitali, S. (2019). Pension fund management with hedging derivatives, stochastic dominance and nodal contamination. *Omega*, 87, 127–141.
- [8] Sereda, E. N., Bronshtein, E. M., Rachev, S. T., Fabozzi, F. J., Sun, W. and Stoyanov, S. V. (2010). Distortion risk measures in portfolio optimization *Handbook of portfolio construction*, pp. 649–673.
- [9] Van der Hoek, J. and Sherris, M. (2001). A class of non-expected utility risk measures and implications for asset allocations. *Insurance: Mathematics and Economics*, 28, 69–82.
- [10] Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance. Mathematics and Economics*, 17, 43–54.
- [11] Wang, S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of risk and insurance*, 67, 15–36.
- [12] Wirch, J. and Hardy, M. (1999). A synthesis of risk measures for capital adequacy *Insurance: mathematics and economics*, 25, 337–347.
- [13] Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society*, 55, 95–115.
- [14] Vitali, S., Moriggia, V. and Kopa, M. (2017). Optimal pension fund composition for an Italian private pension plan sponsor. *Computational Management Science*, 14, 135–160.
- [15] Zapletal, F., Šmíd, M. and Kopa, M. (2020). Multi-stage emissions management of a steel company. *Annals of Operations Research*, 292, 735–751.

The goal programming approach to investment portfolio selection during the COVID-19 pandemic

Donata Kopańska-Bródka¹, Renata Dudzińska-Baryła²,
Ewa Michalska³

Abstract. The coronavirus pandemic has an impact on almost every field of our lives, including investments. More than a year has passed since the outbreak of the pandemic. Therefore, we have enough data to analyse the effects of the coronavirus crisis. The authors of this paper present the analysis of the impact of COVID-19 pandemic on the investment portfolio of risky assets that are traded on the various markets, such as stock market, currency market and commodity market. Such components are supposed to be statistically independent. Our results show how portfolio responds to coronavirus pandemic and how this response changes over time with regard to investors with various preferences. The preferences at the moments of distribution of the portfolio's rate of return are expressed in the scenario form. In our study we apply the polynomial goal programming model to construct the investment portfolio in each month, taking into account the period from the pandemic outbreak.

Keywords: goal programming, investor's preferences, investment portfolio, mean-variance-skewness portfolio, COVID-19

JEL Classification: G11

AMS Classification: 91G10

1 Introduction

Historically, epidemics, both global and regional, are most often accompanied by studies and empirical analyses regarding their impact on the economy and human behaviour. Various restrictions, limitations and sanitary regimes introduced during a pandemic distract decision-makers and the decisions they make are often motivated by other factors (emotional factors) instead of rationality. The outbreak of a pandemic can be a shock to the global economy but its economic consequences vary widely and some industries become financial beneficiaries. The coronavirus pandemic spreading globally from 2020 is seen as a real threat to the world economy, financial markets and commodity markets.

The impact of the COVID-19 pandemic on economies and financial markets was already noticed in the first months of the global spread of the virus. The paper [12] shows empirical studies supporting the thesis of the immediate reaction of global financial markets to the unexpected outbreak of epidemic. Because of the pandemic, some international institutions such as the International Monetary Fund (IMF) and the Organisation for Economic Co-operation and Development (OECD) made radical reductions in global economic growth forecasts [12]. A research regarding 30 countries conducted by Fernandez [11] showed that in 2020 there will be a decline in GDP of 2.8% on average. The quotations of most financial market indices reached very low levels in March 2020 [3]. The link between pandemic outbreak and stock market risk was examined in the paper of Zhang, Hu and Ji [24], among others. The studies on the economic impact of previous epidemics such as SARS in 2003 in Taiwan [4] and Ebola in 2014 [7] also shown that these outbreaks had a significant impact on financial markets.

At the end of the first quarter of 2020, global stock market indices such as the Dow Jones Industrial Average (USA), the SSE Composite Index (China) or the Euronex 100 (Europe) reacted with strong declines. On the US market, the three main stock market indices DJIA, NASDAQ and S&P 500 recorded the biggest falls, with declines of 37.1%, 30.1% and 31.9%, respectively. Such a strong financial market reaction resulted in a decline of more than 10 trillion USD in trading value [9]. Moreover, on 23 March 2020, the global share market index (MSCI ACWI) hit its lowest level and declined by 32% compared to 2 January 2020. Also in March 2020, the major stock exchanges around the world experienced the largest declines but already five months later, despite the pandemic, there was a rebound leading to high quotation levels. Strong falls were again seen at the turn of September and October as well as October and November 2020.

¹ University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, Poland, donate.kopanska-brodka@ue.katowice.pl.

² University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, Poland, renata.dudzinska-baryla@ue.katowice.pl.

³ University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, Poland, ewa.michalska@ue.katowice.pl.

Since the beginning of the global spread of the coronavirus in 2020, there was no significant impact of the pandemic on efficiency on the forex market. Changes in quotations in the area of the three currencies of interest: EUR, USD and PLN were not as abrupt as they were on the share markets and the upward trend of EUR quotes against USD and PLN was stable throughout the period.

The reaction of the cryptocurrency market at the start of the pandemic was similar to that seen on global stock markets. The price of BITCOIN reached its lowest level since April 2019 on 12 March 2020, thus violating the myth of cryptocurrencies as a “safe haven” for times of crises and stock market crashes [2].

The rapid spread of confirmed cases of coronavirus infection significantly affected the demand and supply of commodities thus contributing to declines in commodity trading [19]. Since January 2020, prices of major commodities have indicated a downward trend. The demand for oil collapsed, which contributed to the fall in its price on the stock exchanges [23]. Following the WHO announcement of the COVID-19 pandemic in March 2020, the commodity market also experienced short-term significant fluctuations in precious metal prices but the following months already showed an upward trend. Due to the uncertainty surrounding WHO reports on the global spread of the coronavirus, gold prices continued to rise until March 2020. A significant but short-term reduction in the gold price occurred on 16 March 2020 when the amount of \$1451.5 was paid for an ounce of gold.

The impact of the epidemic can be transferred to financial markets through various channels. Not only through a decrease in the number of economic actors, high level of market linkages or financial integration but also through investors' decisions on the structure of their investment portfolio. According to Ramelli and Wagner [20], the investors are paying more attention to the economic and financial impact of the COVID-19 pandemic.

The problems addressed in this article focus on the decisions of the investor with fixed preferences regarding the parameters of rate of return of an optimal portfolio whose components are assets from different independent markets. The potential components of the investment portfolio are three assets treated as a “safe haven” (gold, currency and cryptocurrency), oil which is an important raw material determining the development of the global economy and investments in shares of companies on the domestic and foreign markets (represented by stock market indices). Optimal portfolios are constructed using methods of polynomial goal programming for different periods of development of the COVID-19 pandemic, while the reaction of the decision-maker to various events accompanying the pandemic is described by the structure of the obtained portfolios.

2 Polynomial goal programming models in investment decisions

In the classical portfolio selection problem proposed by Markowitz [16], we minimize the portfolio variance for a fixed level of rate of return. Scott and Horvath [22] showed that if the assumptions of the Markowitz model are not satisfied (i.e. the distribution of random rates of return is asymmetric or the utility function is not a quadratic function), the evaluation of investments should be based on central moments of at least the third or fourth order.

Portfolio selection models which take into account higher order moments are classified as multi-criteria optimisation problems. Various techniques for solving such problems are considered in the literature, including the approximation of expected utility with higher order moments proposed on the grounds of utility theory or reduction of the problem to a goal programming approach with a linear or nonlinear criterion function [1, 8, 13, 15, 21]. Lai [14] was a precursor who used polynomial goal programming in optimal portfolio selection. This approach had many imitators [5, 6] but the assumptions they make about the model could lead to solutions that are not feasible from a practical point of view.

The studies that have been conducted for years on distributions of random rates of return confirm the failure of the basic assumptions of the Markowitz model concerning the normality of the distribution of random rates of return [10, 18]. Thus, there is a growing interest in models that are extensions of the classical two-criteria model of optimal portfolio selection. The literature of the last decade abounds in modifications of the Markowitz model involving the inclusion of higher order central moments as additional criteria as well as in methods for solving the resulting problem. A multi-criteria model of selecting a share portfolio (without short selling) taking into account the expected value E_p , variance V_p and the third central moment as a skewness measure S_p is formulated as follows:

$$\begin{aligned}
& \max(E_p) \\
& \min(V_p) \\
& \max(S_p) \\
& \sum_{i=1}^N x_i = 1 \\
& x_i \geq 0, i = 1, \dots, N
\end{aligned} \tag{1}$$

where the quantities x_i for $i = 1, \dots, N$ denote the shares in the portfolio.

The preference for the value of parameters of the rate of return portfolio distribution occurring in model (1), i.e. the expected value E_p , variance V_p and third central moment S_p are justified by expected utility theory [17, 8]. The E_p maximisation represents the preference for higher expected benefits, V_p minimisation corresponds to risk aversion, S_p maximisation on the other hand refers to a preference for a positive skewness of the portfolio distribution that guarantees a lower probability of very low portfolio rates of return.

Model (1) is a multi-criteria non-linear problem and its solution depends on the assumptions made and the choice of method. The simplest technique for solving such a problem is reduction to the problem with a single objective function. Many solutions of this type are proposed in the literature, one of the possibilities is the use of goal programming. The formal model of selecting an optimal stock portfolio using goal programming is as follows:

$$\begin{aligned}
& \min(z(d)) \\
& E_p + de = E_o \\
& V_p - dv = V_o \\
& S_p + ds = S_o \\
& \sum_{i=1}^N x_i = 1 \\
& x_i \geq 0, i = 1, \dots, N \\
& de, dv, ds \geq 0
\end{aligned} \tag{2}$$

The minimised criterion function $z(d)$ is defined as a function of deviations that depends additionally on the investor's preferences with respect to the moments of the distribution. These preferences are expressed by the ranks (α, β, γ) assigned to undesired deviations (de, dv, ds) from the aspiration levels regarding the portfolio distribution parameters. The desired levels (E_o, V_o, S_o) can be assumed or determined as optimal solutions of other models. Among the forms of the $z(d)$ function considered in the literature, there is the polynomial form⁴. This can be the polynomial form of the absolute deviations

$$z(d) = (de)^\alpha + (dv)^\beta + (ds)^\gamma \tag{3}$$

or polynomial form of the relative deviations

$$z(d) = \left| \frac{de}{E_o} \right|^\alpha + \left| \frac{dv}{V_o} \right|^\beta + \left| \frac{ds}{S_o} \right|^\gamma \tag{4}$$

The rank triple of (α, β, γ) represents the structure of the decision-maker's preferences regarding the distribution parameters.

The procedure for determining the optimal multi-criteria portfolio proposed in this work consists of two stages and the potential components of the portfolio are selected assets whose rates of return are statistically independent random variables. In the first stage the reference values of portfolio parameters (E_o, V_o, S_o) are determined. The expected rate of return E_o is equal to the maximum value from among the expected rates of return of all assets considered in a given period, the variance V_o is equal to the variance of the global minimum risk portfolio, and the skewness S_o value is equal to the maximum value from among the skewness of all assets in the given period. In the second stage, the previously determined values (E_o, V_o, S_o) are used as reference values (aspiration levels) in the polynomial goal programming model of the form:

⁴ The polynomial model proposed by Lai [14] considers the minimisation of deviations from aspiration levels defined only for the expected value and skewness of the portfolio, while the variance of the optimal portfolio satisfies rigid constraints and takes the value of one.

$$\begin{aligned}
& \min \left(\left(\frac{de}{|E_o|} \right)^\alpha + \left(\frac{dv}{V_o} \right)^\beta + \left(\frac{ds}{|S_o|} \right)^\gamma \right) \\
& \quad E_p + de = E_o \\
& \quad V_p - dv = V_o \\
& \quad S_p + ds = S_o \\
& \quad \sum_{i=1}^N x_i = 1 \\
& \quad x_i \geq 0, i = 1, \dots, N \\
& \quad de, dv, ds \geq 0
\end{aligned} \tag{5}$$

where the values x_i for $i = 1, \dots, N$ denote the shares of assets in the portfolio and the values de, dv, ds represent deviations from desired values. The parameters of a multi-criteria portfolio of assets that are independent random variables are determined as follows:

- expected value: $E_p = \sum_{i=1}^N (x_i \cdot E(R_i))$,
- variance: $V_p = \sum_{i=1}^N ((x_i)^2 \cdot V_i)$,
- skewness: $S_p = \sum_{i=1}^N ((x_i)^3 \cdot S_i)$,

where $E(R_i)$ denotes the expected value of the i th asset, V_i its variance, and S_i the third central moment treated as a measure of skewness.

The rank triple (α, β, γ) describes the considered scenario of preferences with respect to the (E_p, V_p, S_p) parameters, whereas $\alpha, \beta, \gamma \in \{1; 2; 3\}$. For example, the scenario $(2, 1, 3)$ corresponds to the situation when $(V_p > E_p > S_p)$ which means that achieving the aspiration level for the portfolio variance is more preferred than achieving the aspiration level for the expected value and skewness.⁵

3 Portfolio structure during the COVID-19 pandemic – empirical study

The aim of our studies is to analyse the structure of optimal portfolios determined for investors with different preferences regarding expected value, variance and skewness in subsequent months of the COVID-19 pandemic. Potential components of portfolios were selected from various independent markets: cryptocurrencies (BITCOIN quoted on BitStamp), commodities (GOLD, OIL), US stock exchange (DJIA index), currencies (USD/EUR quoted on Forex), Polish stock exchange (WIG index). All assets are quoted in USD. Additionally, for the DJIA index one index point corresponds to 1 USD and WIG index quotations (one index point corresponds to 1 PLN) are expressed in USD using the average USD/PLN exchange rate provided by the National Bank of Poland. All quotations are from www.biznesradar.pl.

In order to capture changes in the structure of the optimal portfolios over successive periods of the spread of the coronavirus pandemic, the analysis covered the period from October 2019 to March 2021 in which 16 three-month sub-periods were identified. The first portfolios (for the preference scenarios considered) were determined as at 1 January 2020 on the basis of the logarithmic daily rates of return from the last quarter of 2020. These were therefore portfolios corresponding to the pre-pandemic period. Subsequent portfolios were determined on a monthly basis (on the first day of the month) based on data from the preceding three months (58-66 observations). The last group of portfolios was calculated on 1 April 2021.

In each period, the obtained optimal portfolios are a solution to the polynomial goal programming model (5). Due to non-linearity of this model, calculations were made in the SAS software using the NLP solver and self-prepared programs. The desired values of parameters of the distribution of rates of return of the portfolio were determined (separately for each of the sixteen sub-periods) according to the previously presented method of determining the reference values (E_o, V_o, S_o) . The investors' preferences with respect to the parameters of the distribution of the portfolio return rate were modelled by means of an ordered rank triple (α, β, γ) . If for investors the most important thing is the expected value of the portfolio's rate of return then their preferences are represented by rank triples $(1, 2, 3)$, $(1, 3, 2)$, $(1, 2, 2)$. If the variance of the portfolio rates of return is the most important for investors, their preferences are represented by the rank triples $(2, 1, 3)$, $(3, 1, 2)$, $(2, 1, 2)$. On the other hand, if for investors the

⁵ The symbol " $>$ " denotes the relation of preferences to moments of the portfolio distribution.

skewness of the distribution of portfolio rates of return is the most important then preferences are represented by the rank triples (2, 3, 1), (3, 2, 1), (2, 2, 1).

The obtained optimal portfolios were analysed both in terms of changes in their structure and the dynamics of changes in the shares of individual assets in subsequent months of the pandemic. Figure 1 shows the structure of portfolios determined for 16 sub-periods and two selected preference scenarios (1, 2, 3) and (2, 1, 3)⁶. The dominant asset in all portfolios regardless of preference and sub-period is the EUR currency. The announcement of a coronavirus pandemic by WHO on 11 March 2020 causes panic among investors and the abandonment of riskier assets. As of 25 March 2020, all the European Economic Area countries and more than 150 countries worldwide had been affected. For sub-periods 4 and 5 (covering data from January to April 2020), the optimal portfolios include only two components: EUR and GOLD which is traditionally seen as a safe investment. The shares of EUR and GOLD in the portfolios in all 16 sub-periods under the assumption that the expected rate of return on the portfolio is preferred over the variance and skewness (scenario (1, 2, 3)) are presented in Figure 1a. In the period from May to mid-September, the infection curve in Poland flattened and the daily number of positive tests remained at the level of several hundred. Information in the media about the end of the first wave of the pandemic and relaxation of restrictions during the holiday period caused people to return to normality also in the area of investments – in sub-periods 8, 9 and 10 (covering data from May to September 2020) we observe a situation analogous to that before the pandemic, i.e. full diversification of portfolios. A similar situation is observed in periods 13-15, but this time the reasons can be attributed to the start of mass vaccination and the prospect of a return to “normality”.

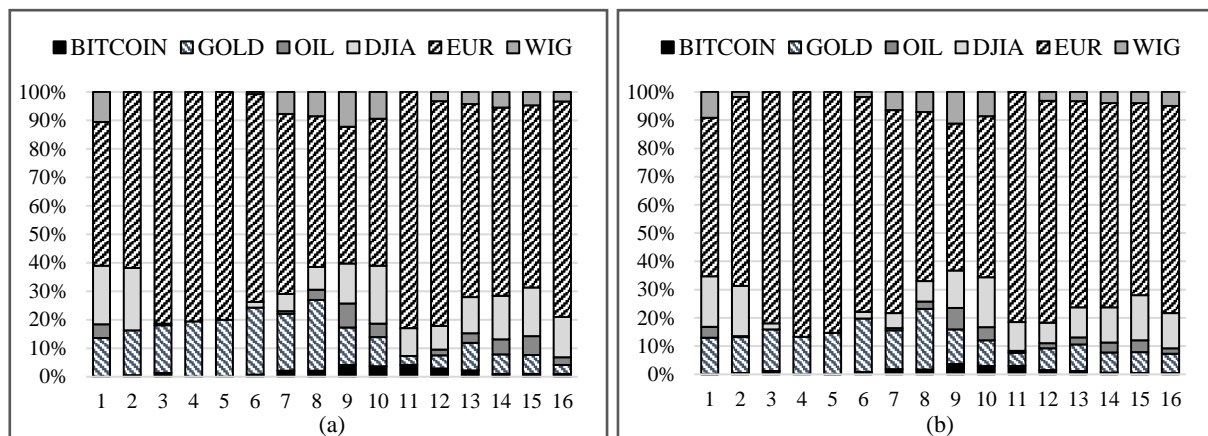


Figure 1 Structure of optimal portfolios: (a) portfolios for preference scenario (1, 2, 3), (b) portfolios for preference scenario (2, 1, 3)

The main asset constantly present in portfolios is EUR. During the first wave of the pandemic (Figure 2), we observe an increase in the share of the European currency in portfolios, up to 80%. During the holiday season EUR is no longer as attractive as before and its share in portfolios returned to pre-pandemic levels (around 50%). The next wave of the pandemic brings a renewed increase in the share of EUR in portfolios.

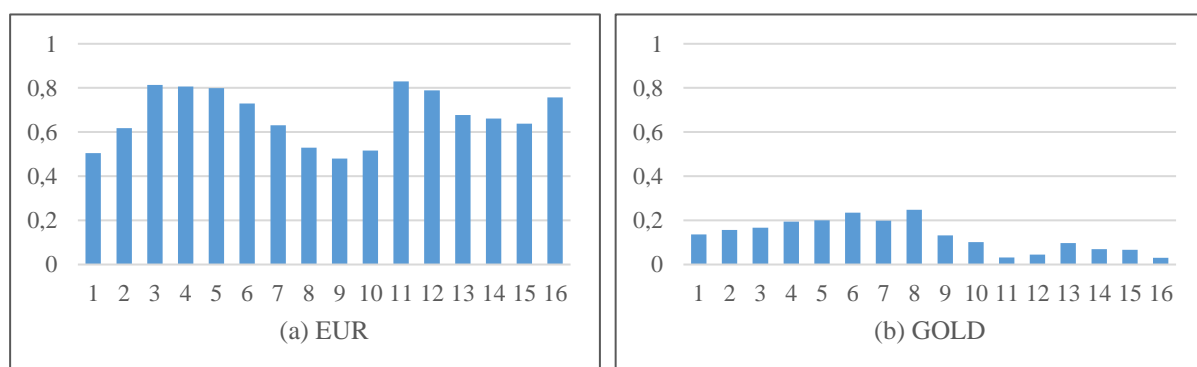


Figure 2 EUR and GOLD shares in portfolios for the preference scenario (1, 2, 3)

The investment in BITCOIN comes as a surprise – the strongly rising quotations of this cryptocurrency as of December 2020 did not translate into significant shares in portfolio (Figure 2). Throughout the period, regardless

⁶ For the other preference scenarios, the portfolio structures are similar.

of the pandemic situation in the country and the world, the shares of BITCOIN in optimal investment portfolios are negligible and do not exceed 7%. The cryptocurrency, compared to the EUR currency dominating the portfolios in different sub-periods of the ongoing COVID-19 pandemic and different scenarios of investor preferences regarding profit, risk and skewness, is not a safe investment. The quotation charts of EUR currency and BITCOIN cryptocurrency are presented in Figure 3.

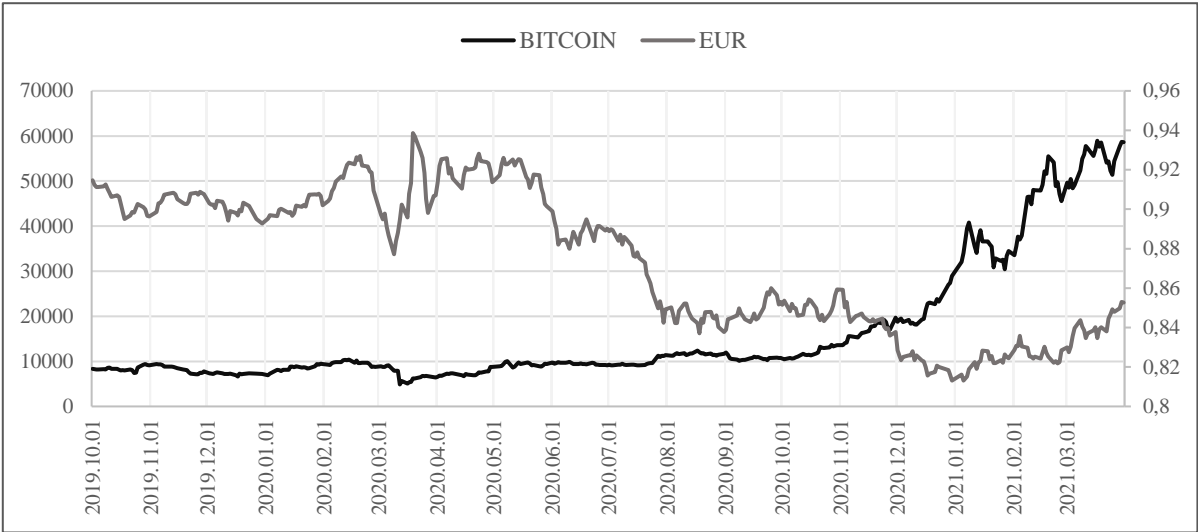


Figure 3 Quotations of BITCOIN and EUR in period October 2019-March 2021

The shares of OIL in portfolios do not exceed 10% throughout the period under review. Regardless of the investor preference scenario and the pandemic period, the asset are perceived as unattractive. Only during periods of easing restrictions (periods 7-10) and positive investor approach due to mass vaccinations (periods 12-16), the shares of OIL in portfolios are non-zero. A similar situation is observed for domestic investment (WIG), for which the highest shares, similar to those before the pandemic but not exceeding 14%, are observed during periods of withdrawal of restrictions imposed on society and an upward trend in financial markets (periods 7-10), as illustrated by the chart in Figure 4.

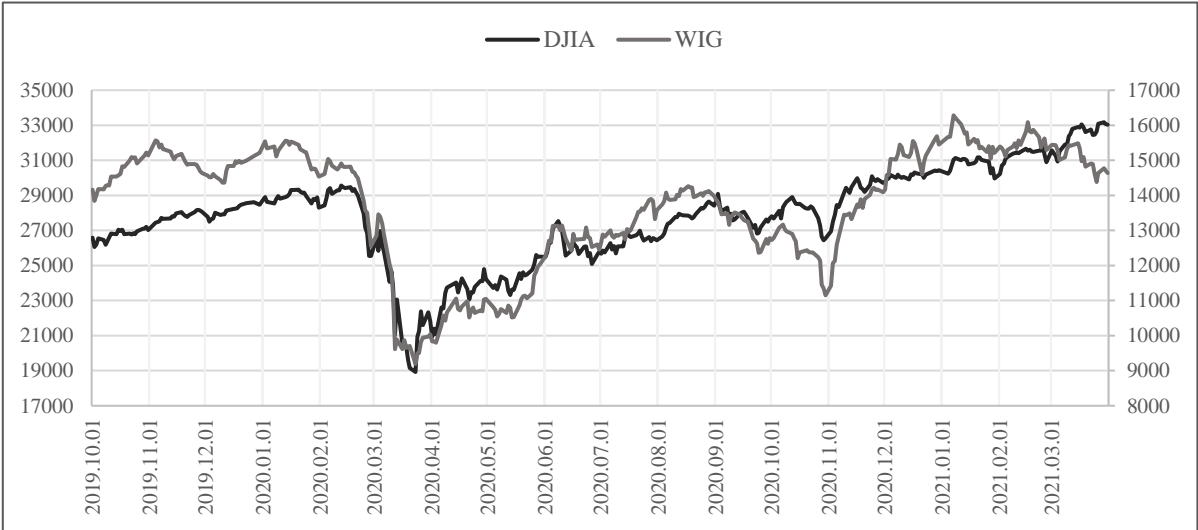


Figure 4 Quotations of DJIA and WIG in period October 2019-March 2021

The US market started to react to the pandemic situation in China already in the first months of 2020 which is reflected in the rapidly declining DJIA index quotations (Figure 4). In portfolios determined for 1 March 2020, the shares of this index are zero. In the following months the situation stabilises. The shares increase and are no less than 7%, despite the successive waves of the pandemic.

The events surrounding the spread of the COVID-19 coronavirus and its mutation generate fear and uncertainty, i.e. conditions in which gold always appreciates in value – gold retains the purchasing value of capital over time. Since August 2020 we have been observing a decline in gold prices, which may be explained by a familiarisation

with the prevailing situation and a reduction in investments in "safe" gold. Consequently, the share of GOLD in portfolios are also declining.

4 Conclusions

The rapidly developing coronavirus pandemic in 2020 has not only become a threat to the proper functioning of financial institutions but has also affected the investment decisions made by individuals with a subjective preference regarding selection criteria. In this paper the authors investigate how COVID-19 affects the structure of an optimal investment portfolio renewed monthly. Portfolios with components from different markets were analysed, starting with a portfolio determined on the basis of data from the quarter preceding the COVID-19 outbreak through the subsequent months of the pandemic. On the basis of the results obtained, the structure of the optimal portfolios was found to depend on events related to the spread of the pandemic. During the period of relaxation of restrictions, the shares of assets were similar to those before the pandemic. The scenarios considered in the proposed model regarding preferences over expected value, variance and skewness did not affect the structure of the optimal portfolios.

Currency and gold have proven to be a safe haven for equity investors amid the economic and financial market turmoil associated with the COVID-19 pandemic at various stages of its development. The allocation of BITCOIN with zero or marginal share in optimal portfolios supports the view that this cryptocurrency is not a "safe haven" for investors during the pandemic. Lack of investments in domestic and US-listed stocks during periods of pandemic outbreak turbulence regardless of the preference scenario underline the investor distrust towards these securities.

References

- [1] Arditti, F.D. & Levy, H. (1975). Portfolio Efficiency Analysis in Three Moments: The Multiperiod Case. *Journal of Finance*, 30(3), 797–809.
- [2] Baur, D.G. & Hoang, L.T. (2021). A crypto safe haven against Bitcoin. *Finance Research Letters*, 38, 101431, <https://doi.org/10.1016/j.frl.2020.101431>.
- [3] Cheema, M., Faff, R. & Szulczyk, K. (2020). The influence of the COVID-19 pandemic on safe haven assets. <https://voxeu.org/article/influence-covid-19-pandemic> (access 17.04.2021).
- [4] Chen, M.P., Lee, C.C., Lin, Y. H. & Chen, W.Y. (2018). Did the SARS epidemic weaken the integration of Asian stock markets? Evidence from smooth time-varying cointegration analysis. *Economic Research – Ekonomska Istraživanja*, 31(1), 908-926, <https://doi.org/10.1080/1331677X.2018.1456354>.
- [5] Chunnachinda, P., Dandapani, K., Hamid, S. & Prakash, A.J. (1997). Portfolio Selection and Skewness: Evidence from International Stock Markets. *Journal of Banking and Finance*, 21, 143–167.
- [6] Davies, R.J., Kat, H.M. & Lu, S. (2009). Fund of Hedge Funds Portfolio Selection: A Multiple-Objective Approach. *Journal of Derivatives and Hedge Funds*, 15(2), 91–115.
- [7] Del Giudice, A. & Paltrinieri, A. (2017). The impact of the Arab Spring and the Ebola outbreak on African equity mutual fund investor decisions. *Research in International Business and Finance*, 41, 600–612.
- [8] Eichner, T. & Wagener, A. (2011). Increases in skewness and three-moment preferences. *Mathematical Social Sciences*, 61(2), 109–113.
- [9] Gao, X., Ren, Y. & Umar, M. (2021). To what extent does COVID-19 drive stock market volatility? A comparison between the U.S. and China. *Economic Research – Ekonomska Istraživanja*, <https://doi.org/10.1080/1331677X.2021.1906730>.
- [10] Fama, E.F. (1965). The Behavior of Stock Market Prices. *Journal of Business*, 38(1), 34–105.
- [11] Fernandez, N. (2020). Economic effects of coronavirus outbreak (COVID-19) on the world economy. IESE Business School Working Paper No. WP-1240-E, <http://dx.doi.org/10.2139/ssrn.3557504>.
- [12] Khatatbeh, I.N., Bani Hani, M. & Abu-Alfoul, M.N.(2020). The Impact of COVID-19 Pandemic on Global Stock Markets. *International Journal of Economics and Business Administration*, VIII(4), 505–514.
- [13] Kopańska-Bródka, D., Dudzińska-Baryła, R. & Michalska, E. (2019). The Investor's Preferences in the Portfolio Selection Problem Based on the Goal Programming Approach. In: W. Tarczyński & K. Nermend (Eds.) *Effective Investments on Capital Markets. Series: Springer Proceedings in Business and Economics* (pp. 151–163). Springer.
- [14] Lai, T.Y. (1991). Portfolio Selection with Skewness: A Multiple-Objective Approach. *Review of Quantitative Finance and Accounting*, 1, 293–305.
- [15] Levy, H. (1969). A utility function depending on the first three moments. *Journal of Finance*, 24, 715–19.
- [16] Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, 7(1), 77–91.

- [17] Menezes, C., Geiss, C. & Tressler, J. (1980). Increasing downside risk. *American Economic Review*, 70(5), 921–932.
- [18] Piasecki, K. & Tomasik, E. (2013). *Rozkład stóp zwrotu z instrumentów polskiego rynku kapitałowego*. Kraków-Warszawa: Edu-Libri (in Polish).
- [19] Rajput, H., Changotra, R., Rajput, P., Gautam, S., Gollakota, A.R.K. & Arora A.S. (2021). A shock like no other: coronavirus rattles commodity markets. *Environment, Development and Sustainability*, 23, 6564–6575, <https://doi.org/10.1007/s10668-020-00934-4>.
- [20] Ramelli, S. & Wagner, A.F. (2020). Feverish Stock Price Reactions to COVID-19. *Review of Corporate Finance Studies*, 9(3), 622–655, <https://dx.doi.org/10.2139/ssrn.3550274>.
- [21] Samuelson, P.A. (1970). The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means, Variances and Higher Moments. *Review of Economic Studies*, 37(4), 537–542.
- [22] Scott, R. & Horvath, P. (1980). On the Direction of Preference for Moments of Higher Order than the Variance. *Journal of Finance*, 35, 915–919.
- [23] Sharif, A., Aloui, C. & Yarovaya, L. (2020). COVID-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the US economy: Fresh evidence from the wavelet-based approach. *International Review of Financial Analysis*, 70, <https://doi.org/10.1016/j.irfa.2020.101496>.
- [24] Zhang, D., Hu, M. & Ji, Q. (2020). Financial markets under the global pandemic of COVID-19. *Finance Research Letters*, 36, <https://doi.org/10.1016/j.frl.2020.101528>.

Robust First Order Stochastic Dominance in Portfolio Optimization

Karel Kozmík¹

Abstract. We use modern approach of stochastic dominance in portfolio optimization, where we want the portfolio to dominate a benchmark. Since the distribution of returns is often just estimated from data, we look for the worst distribution that differs from empirical distribution at maximum by a predefined value. First, we define in what sense the distribution is the worst for the first order stochastic dominance. We derive a robust stochastic dominance test for the first order stochastic dominance and find the worst-case distribution as the optimal solution of a non-linear maximization problem. We apply the derived optimization programs to real life data, specifically to returns of assets captured by Dow Jones Industrial Average, and we analyze the problems in detail using optimal solutions of the optimization programs with multiple setups.

Keywords: portfolio optimization, stochastic dominance, robustness

JEL Classification: G11

AMS Classification: 91G10

1 Introduction

The problem of portfolio optimization is a typical problem in economics and finance. Modern methods that deal with portfolio optimization include stochastic dominance, e.g. [7] and [4]. The concept of stochastic dominance allows us to compare two random variables, which in this case represent the return of our final portfolio and a benchmark portfolio. In this work we explore the resistance of the optimal portfolio to the changes of distribution of the returns. To get the optimal portfolio, historical observations are usually used, which represent the empirical distribution.

To present the motivation: let us suppose that investor's preferences can be represented by a utility function u (non-decreasing function of money) and he makes choices based on the expected utility from the investment $E u(R)$, where random variable R denotes the returns of the investment. Then we are able to order possible portfolios based on their expected utility. But when we do not know the utility function or we want the portfolio to be acceptable for a large amount of people (for example pension funds etc.), we can use the concept of stochastic dominance.

In this work, we firstly present the concept on stochastic dominance and define the distributional robust version of the first order stochastic dominance. Then we present the Wasserstein distance and based on it, we derive a reformulation of the robust stochastic dominance conditions of the first order. We also derive a program to find the worst case distribution. The derived programs are then tested using real life data in the empirical analysis section.

2 Robust stochastic dominance

Assumptions and notation

Let us have a portfolio of N stocks with weights $\mathbf{w} = (w_1, \dots, w_N)$, $W = \mathbf{w}$, where $\sum_{i=1}^N w_i = 1$, $w_i \geq 0$, $i = 1, \dots, N$. The no short selling assumption is not really needed but was used when deriving the tested portfolios. In this work, we often talk about random returns, which are often represented by a random variable and about observed returns, which are represented by scenarios (denoted r_{is} - return of asset i in scenario s). Benchmark portfolio weights are denoted $\boldsymbol{\tau}$.

2.1 Theoretical background

Let us define the first order stochastic dominance (FSD) in accordance with [6].

Definition 1. First we define set \mathcal{U} as the set of all utility functions. Let X and Y be random variables, we say that X first order stochastically dominates Y ($X \succeq_{FSD} Y$) if

$$E[u(X)] \geq E[u(Y)] \quad \forall u \in \mathcal{U}.$$

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

We also state the equivalent conditions for the FSD as in [6]. Proof can be found for example in [3].

Theorem 1. Let X_1 and X_2 be a random variable and F_{X_1} and F_{X_2} their distribution functions, then:

(i)

$$X_1 \geq_{FSD} X_2 \iff F_{X_1}(x) \leq F_{X_2}(x), \forall x \in \mathbb{R}$$

Now define the robust stochastic dominance in accordance with [1].

Definition 2. We say that a random variable X dominates robustly a random variable Y in the first order over a set of probability measures \mathcal{Q} ($X \geq_{FSD}^{\mathcal{Q}} Y$) if

$$E_P[u(X)] \geq E_P[u(Y)] \quad \forall u \in \mathcal{U} \quad \forall P \in \mathcal{Q}. \quad (1)$$

We understand X and Y as random variables denoting return of some portfolios, which consist of stocks, and that the joint distribution of the random returns of the underlying stocks is defined by the distribution P , which we allow to change slightly.

The set \mathcal{Q} was not specified, we want our portfolio to be prepared for slight changes in the distribution. For this purpose, we select a suitable measure of distance between two distributions and we bound the change of the distribution by a constant, which defines the set \mathcal{Q} .

2.2 The Wasserstein distance in discrete framework

We use the definition and computation procedures from [8]. We want to define a distance between two probability distributions on \mathbb{R}^N (N is the number of stocks). We adjust the general Wasserstein distance definition to our case, where both distributions have the same number of atoms.

Definition 3. Let us have two discrete distributions with finite support. P_1 attains values x_1, \dots, x_T with probabilities p_1, \dots, p_T and P_2 attains values y_1, \dots, y_T with probabilities q_1, \dots, q_T . The Wasserstein distance of order r ($r \geq 1$) corresponds to solving the following linear program:

$$\begin{aligned} & \min_{\xi_{ts}} \sum_{t=1}^T \sum_{s=1}^T \xi_{ts} d_{ts}^r \\ & \text{subject to} \quad \sum_{s=1}^T \xi_{ts} = p_t, \quad t = 1, \dots, T \\ & \quad \quad \quad \sum_{t=1}^T \xi_{ts} = q_s, \quad s = 1, \dots, T \\ & \quad \quad \quad \xi_{ts} \geq 0, \quad t, s = 1, \dots, T. \end{aligned} \quad (2)$$

where $d_{ts} = d(x_t, y_s)$ denotes a distance between the points x_t and y_s .

For our purposes, we need the flexibility both in values and in probabilities. The first distribution represents our estimate of the distribution based on the observations $r_{it}^0, i = 1, \dots, N; t = 1, \dots, T$ and $p_t = 1/T, t = 1, \dots, T$, and the second represents the changed distribution for robustness purposes. As for the value of r $r = 1$ is usually chosen.

2.3 Robust first order stochastic dominance

In theorem 1 we did state equivalent conditions for the first order stochastic dominance. We can extend this for the first order robust stochastic dominance $X \geq_{FSD}^{\mathcal{Q}} Y$:

$$X \geq_{FSD}^{\mathcal{Q}} Y \iff \sup_{P \in \mathcal{Q}, x \in \mathbb{R}} F_{X,P}(x) - F_{Y,P}(x) \leq 0. \quad (3)$$

The condition is understood in the way that the distribution functions of both portfolios depend on a distribution of underlying assets which can vary and is defined by P (we stress this dependence in the subscript of F).

Now we can think of the worst case distribution for a given x as the one when in which the supremum is attained or some limit probability distribution if it is not attained.

Definition 4. We say that distribution P is the worst-case distribution for the FSD from Q if the supremum of the LHS in problem (3) is attained for this P for some $x \in \mathbb{R}$.

Now let us have N stocks/assets and let the random return rates have a discrete joint distribution P with realizations $r_{jt}, t = 1, \dots, T; j = 1, \dots, N$, attained with probabilities $p_t, t = 1, 2, \dots, T$. Let our portfolio have weights \mathbf{w} and our benchmark portfolio has weights $\boldsymbol{\tau}$. Then we can rewrite the difference of the distribution functions in x as:

$$\sum_{t=1}^T p_t \mathbb{I}_{[\sum_{i=1}^N w_i r_{it} \leq x]} - \sum_{t=1}^T p_t \mathbb{I}_{[\sum_{i=1}^N \tau_i r_{it} \leq x]},$$

where \mathbb{I} represent indicator whether the return of portfolio is smaller than x or not. We also know that a distribution function is non-decreasing which means it is enough to require it only for the points where our portfolio, defined by \mathbf{w} , has probability atoms. Using again the Wasserstein distance we try to evaluate the supremum by maximizing:

$$\begin{aligned} & \max_{\xi_{ts}, r_{it}, p_t, k} \sum_{t=1}^T p_t \mathbb{I}_{[\sum_{i=1}^N w_i r_{it} \leq x_k]} - \sum_{t=1}^T p_t \mathbb{I}_{[\sum_{i=1}^N \tau_i r_{it} \leq x_k]} \\ & \text{subject to } x_k = \sum_{i=1}^N w_i r_{ik}, \quad k = 1, \dots, T \\ & \sum_{t=1}^T \sum_{s=1}^T \xi_{ts} \|\mathbf{r}_t^0 - \mathbf{r}_s\| \leq \epsilon \quad (4) \\ & \sum_{t=1}^T \xi_{ts} = \frac{1}{T}, \quad t = 1, \dots, T \\ & \sum_{t=1}^T \xi_{ts} = p_s, \quad s = 1, \dots, T \\ & r_{it} \geq -1, \xi_{ts} \geq 0, p_t \geq 0, \quad i = 1, \dots, N; t, s = 1, \dots, T. \end{aligned}$$

Now we use the big M to rewrite the indicators using binary variables. We deal with the problem $\max \mathbb{I}_{[\sum_{i=1}^N w_i r_{it} \leq x_k]}$, we reformulate it using u_{tk} representing the indicator:

$$\begin{aligned} & \max u_{tk} \\ & \text{subject to } \sum_{i=1}^N w_i r_{it} \leq x_k + (1 - u_{tk})M, \quad t, k = 1, \dots, T \quad (5) \\ & u_{tk} \in \{0, 1\}, \quad t, k = 1, \dots, T, \end{aligned}$$

where M is sufficiently large constant. For the case of the second indicator, there is minus in front of it, which makes it much more difficult to handle. We need to use the inverse inequality, which in this case is sharp inequality. We get a reformulation for $\max -\mathbb{I}_{[\sum_{i=1}^N \tau_i r_{it} \leq x_k]}$:

$$\begin{aligned} & \max -v_{tk} \\ & \text{subject to } \sum_{i=1}^N \tau_i r_{it} > x_k - v_{tk}M, \quad t, k = 1, \dots, T \quad (6) \\ & v_{tk} \in \{0, 1\}, \quad t, k = 1, \dots, T. \end{aligned}$$

The problem is that for software implementation, sharp inequality cannot be used, so we approximate it by adding a very small term, for example 10^{-5} (we represent it as $p_t \geq m$, where m (margin) is a small positive constant).

So we get a reformulated constraint $\sum_{i=1}^N \tau_i r_{it} \geq m + x_k - v_{tk}M, t, k = 1, \dots, T$.

We use Euclidean norm squared and objective function and the first constraint are merged for implementation purposes. By using a sharp inequality, then maximum might not be attained, so we approximate it. m basically sets the minimal recognizable difference between the returns of our and the benchmark portfolios. By setting m to zero, we would consider the same returns (imagine $x_t = y_s$ for some s, t) actually being higher for the benchmark

(even though they are the same), this is the limit version for $m \rightarrow 0+$. We can do this and bear in mind that the returns of benchmark that are the same as some returns of our portfolio actually mean that they are infinitesimally higher for the actual worst case distribution.

$$\begin{aligned}
& \max_{\xi_{ts}, r_{it}, p_t, u_{tk}, v_{tk}, b_k} \sum_{k=1}^T \left(\sum_{t=1}^T p_t (u_{tk} - v_{tk}) \right) \cdot b_k \\
& \text{subject to } x_k = \sum_{i=1}^N w_i r_{ik}, \quad k = 1, \dots, T \\
& \sum_{t=1}^T \sum_{s=1}^T \xi_{ts} \sum_{i=1}^N (r_{it}^0 - r_{is})^2 \leq \epsilon \\
& \sum_{s=1}^T \xi_{ts} = \frac{1}{T}, \quad t = 1, \dots, T \\
& \sum_{t=1}^T \xi_{ts} = p_s, \quad s = 1, \dots, T \\
& \sum_{i=1}^N w_i r_{it} \leq x_k + (1 - u_{tk})M, \quad t, k = 1, \dots, T \\
& \sum_{i=1}^N \tau_i r_{it} \geq m + x_k - v_{tk}M, \quad t, k = 1, \dots, T \\
& u_{tk} \in \{0, 1\}, v_{tk} \in \{0, 1\} \quad t, k = 1, \dots, T \\
& r_{it} \geq -1, \xi_{ts} \geq 0, b_k \geq 0, p_t \geq 0, \quad i = 1, \dots, N; t, s, k = 1, \dots, T \\
& \sum_{k=1}^T b_k = 1.
\end{aligned} \tag{7}$$

We can understand the optimal values of r_{it} and p_t defining a distribution as the worst case distribution in the sense of definition 4. We formulate the derived results in the following theorem.

Theorem 2. *Let X and Y be random variables denoting returns of a portfolio defined by weights w and τ . Let us have observed historical returns $r_{it}^0, i = 1, \dots, N; t = 1, \dots, T$. Let us have Q a set of probability measures defined on \mathbb{R}^N with T atoms: determined by returns $r_{it}, i = 1, \dots, N; t = 1, \dots, T$ and probabilities $p_t, t = 1, \dots, T$ defined as a neighborhood of the empirical distribution. Let the neighborhood be defined with the use of the Wasserstein distance and let the distance on \mathbb{R}^N be defined as Euclidean norm squared, i.e. let $x, y \in \mathbb{R}^N$, then $d(x, y) = \sum_{i=1}^N (x_i - y_i)^2$. Then X dominates robustly Y in the first order over the set of probability measures Q ($X \geq_{FSD}^Q Y$) if and only if there exists a right open neighborhood of 0 such that for each value m from this neighborhood the optimal value of the problem (7) is less than or equal to zero.*

We can also use the program with $m = 0$ and then optimal value being less than or equal to zero is a sufficient condition for the robust stochastic dominance. By setting $m = 0$ we enlarge the set of feasible distributions so if robust FSD holds for even larger set, then it certainly holds for the smaller set. Also if we find m for which there is a feasible solution and the objective value is positive, then the robust stochastic dominance cannot hold.

We can generalize the test for probabilities different from $1/T$. Let us have general probabilities $p_t^0, t = 1, \dots, T$ satisfying $\sum_{t=1}^T p_t^0 = 1$, then if we look at the program, the only place we use the observed distribution is in the Wasserstein distance, so we can easily generalize the program by replacing the constraint $\sum_{s=1}^T \xi_{ts} = \frac{1}{T}, t = 1, \dots, T$ by the constraint $\sum_{s=1}^T \xi_{ts} = p_t^0, t = 1, \dots, T$.

3 Empirical analysis

3.1 Data

We used stock prices of assets covered by the Dow Jones Industrial Average (DJIA) index, which consists of 30 largest and most traded American companies. The dataset used consists of observations from April 2008 to July

2018. Quarterly returns were used and because the complexity of the problem was very high, we used only first 5 observations for the final experiments. In the dataset, Apple has the highest mean, followed by V (Visa) and UNH (United Health). The lowest mean return had XOM (Exxon Mobil). For the optimization, GAMS software ([2]) was used.

3.2 Worst case distribution

In this section, we analyzed the worst case distribution using (7) for the 2 following portfolios. The first maximizes mean return and stochastically dominates the benchmark portfolio in the first order. How to achieve the portfolio can be found in [5], the portfolio contained V (0.354), MCD (0.391), AAPL (0.255), where the numbers in brackets represent weights. The second one is a benchmark portfolio with weights $1/N$ strategy - in our notation $\tau_i = 1/30, i = 1, \dots, 30$.

We test whether this portfolio also robustly dominates the benchmark and observe the worst case distribution. We analyzed the development of the worst case distribution in dependence on the value of ϵ , which defines the radius of the neighborhood of the empirical distribution. Investor with such portfolio wants to know which is the worst possible distribution for him.

The robust FSD test given by (7) faces great challenges given that the problem is not only non-linear and non-convex, but also mixed integer. We used BONMIN solver for mixed integer non-linear programming. Using more than 5 scenarios caused computational problems, solver converging to an infeasible solution. We chose the values of ϵ as 0 for the first case and then starting with 0.001, we doubled the value of epsilon, ending with 0.128, when the when the difference of the distribution functions was maximal (value 1). The computation took around 2 minutes for all the considered values of ϵ (PC with 16 GB RAM and Intel core i5 6500 Skylake).

The constraint for the distance of distributions was fulfilled as equality, the value was slightly lower on the left side of the equality for the last value of ϵ ($0.127 < 0.128$). This does makes sense, there was no option for the objective to further improve as it was already 1. The probabilities were changed in the worst case scenario for the lower values of ϵ , because the change of probability could make the difference between cumulative distribution functions larger. As the program was able to make the difference close to 1, there was no need to adjust the probabilities as all the values of benchmark had to be larger than the values of our portfolio.

We can see multiple cumulative distribution functions plotted in fig. 1. For $\epsilon = 0$ it is the problem under the empirical distribution, then with the increasing ϵ we can see the distribution function plotted under the worst case distribution. We can see the rise of the CDF for our portfolio in one part of the graphs, because the FSD condition is strictly the difference between CDFs. The fact that the lowest return is actually where the difference is the highest is connected to the fact that the portfolio was created such that the conditions are fulfilled as an equality on the left tail (lower returns) so it is easier (in terms of distance of distributions) to violate those conditions.

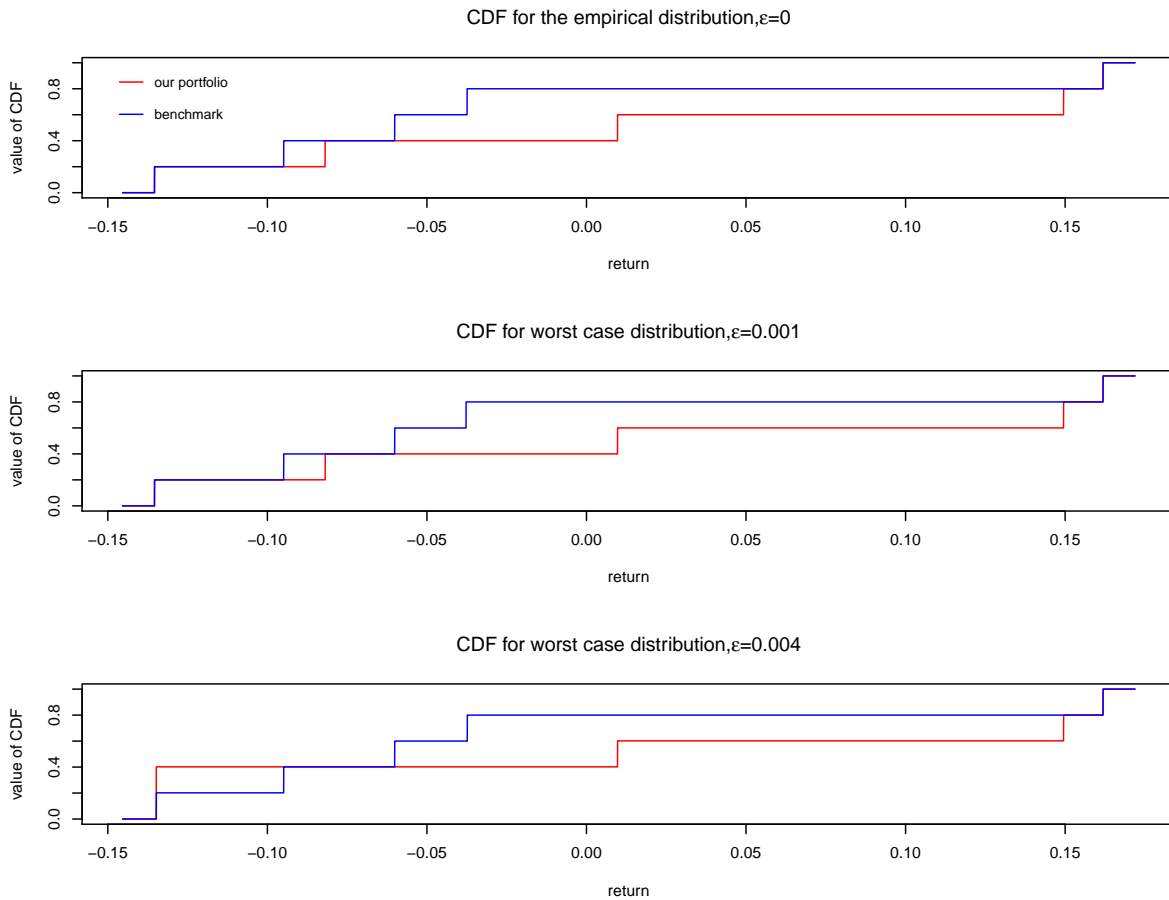
Even though it is not visible from the graph, the FSD condition is violated for $\epsilon = 0.001$, the difference in the returns is so small it cannot be captured by the graph, the critical points are return -0.13538 for our portfolio and -0.13536 for the benchmark. This is the m (margin) we used for the computation, this difference can be made arbitrarily small and the FSD conditions would be violated. The situation is the same for other depicted ϵ , for example for $\epsilon = 0.064$ the returns are -0.10312 for our portfolio and -0.10311 for the benchmark.

4 Conclusion

In this work we introduced the concept of stochastic, robust optimization and the Wasserstein distance. We defined the worst case distribution for the first order stochastic dominance. Based on the equivalent conditions, we derived test of robust stochastic dominance of the first order and program to find the worst case distribution.

In the empirical part, we tested all the derived program on real life data, specifically on returns of assets captured by Dow Jones Industrial Average. We analyzed the development of the worst case distribution with increasing value of the radius of the neighborhood around the empirical distribution. The results for all the tests were confronted with intuition and expectations. The main features were captured graphically and by running the programs with multiple set ups, we were able to understand the behavior in detail. All of the tests showed some level of numerical instability because the programs are non-linear, non-convex and mixed integer. Even though the programs faced numerical challenges, we were able to get results, which made good sense, followed the definitions and used the strictest parts of the dominance conditions.

Figure 1 CDFs for empirical distribution and for the worst case distribution for different values of ϵ



Acknowledgements

This work was partially supported by Czech Science Foundation (grant 19-28231X) and SVV project of Charles University n. 260580.

References

- [1] Dentcheva, D., & Ruszczyński, A. (2010). Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming*, 123(1), 85-100.
- [2] GAMS Development Corporation. General Algebraic Modeling System (GAMS) Release 24.7.4. Washington, DC, USA, "<http://www.gams.com>".
- [3] Hanoch, G., & Levy, H. (1969). The efficiency analysis of choices involving risk. *The Review of Economic Studies*, 36(3), 335-346.
- [4] Kopa, M., Kabašinskas, A., & Štutienė, K. (2021). A stochastic dominance approach to pension-fund selection. *IMA Journal of Management Mathematics*.
- [5] Kuosmanen, T. (2004). Efficient diversification according to stochastic dominance criteria. *Management Science*, 50(10), 1390-1406.
- [6] Levy, H. (2015). *Stochastic dominance: Investment decision making under uncertainty*. Springer.
- [7] Moriggia, V., Kopa, M., & Vitali, S. (2019). Pension fund management with hedging derivatives, stochastic dominance and nodal contamination. *Omega*, 87, 127-141.
- [8] Pflug, G. C., & Pichler, A. (2014). *Multistage stochastic optimization*. Switzerland: Springer International Publishing.

System Dynamic Model of Beehive Trophic Activity

Kratochvilová Hana¹, Rydval Jan², Bartoška Jan³, Chamrada Daniel⁴

Abstract. The article proposes the system dynamic model of beehive trophic activity, including the influence of weather and daily rhythm. The proposed model is based on a stock and flow diagram (SFD). By designing a SFD model, the authors define the dynamic behavior of key aspects of bee swarm prosperity, distinguishing between the internal and external activity of bees during continuous consumption of stocks in the beehive. The article will also outline the procedure on how to estimate the vitality of bee swarm in the long term, because due to frequent bee mortality, bee vitality has become one of the key topics in the field of nature conservation and agricultural production. The results presented in the paper are intended to help verify the hive weights and data collected from field research. 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.

Keywords: Bee Breeding; Honey and Nectar Stocks; Landscape Fertility; Mathematical model; Temperature, Weight and Humidity of Beehive; Stock and Flow Diagram; System Dynamic Model.

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

International academic and professional literature does not provide very extensive information on the issue of beehive trophic activity. Only fragments that deal with terms as fertility or trophies can be traced, and only a small number of authors focuses on sophisticated mathematical models, so any relevant informations are very rare. Nevertheless, the vitality of bee colonies can be described as an important determinant in the area of nature protection and agricultural production at the global level, which are topics that are often mentioned in many countries around the world.

The term "landscape fertility" itself has been present in the literature in the observed context since the beginning of the 21st century. It usually deals with the explanation of this concept [8] in the biological and environmental context and also [13] in connection with fertilization, biochemistry, urban management and the functions of the natural environment. [13] builds presented ideas on several sources of earlier literature, in which the effects of soil chemical composition on landscape fertility are most often discussed. [1] then focuses directly on the behavior of bee colonies and in the proposed mathematical model compares the usability of the landscape fertility in relation to the behavior and performance of bees. His research work examines and evaluates the effect of pollen on the of the honey bee colonies dynamics in the context of the possible impact of adverse events (e.g. pesticides, parasites, nutritional stress).

The use of system dynamics models in living nature is rather a new scientific discipline, that develops especially due to computer technology since 1990s. [10] presents a model that focuses on the analysis of the growing mortality of bee colonies in the observed colonies. In his study, he creates a dynamic model to identify key factors that have a major impact on the growth and survival of hives. Its analysis is based on a three-year follow-up, on the basis of which a simulation of possible population growth or decline is created. The author concludes that the fluctuations were caused by the high sensitivity of the bee colony to the composition of food sources and sharp changes in temperature during the changing seasons. [4] addressed the issue of modeling the dynamics of biological systems in his article. Its outputs can be used as a basis for the general application of system dynamics models in this area. The author presents the possibility of using selected types of organic, genetic, population or biochemical models. To display the dynamic complexity of system and understand animal behavior, Krejci et al. [7] uses a system dynamics model, especially a stock and flow diagram to capture a livestock system behavior.

¹ Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, ha.kratochvilova@gmail.com.

² Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, rydval@pef.czu.cz.

³ Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, bartoska@pef.czu.cz.

⁴ Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, chamrada@pef.czu.cz.

The aim of this paper is to create and present a model of trophic activity of bees, including the influence of weather and circadian rhythms on bee activity. This model will provide an idea of the short-term trophic activity of bees, including honey production, including the effect of outdoor temperature and the length of sunlight. The model will be designed by using the basic elements of system dynamics and it will be verified using the data obtained from beehive weights and data collected from field research.

2 Material and methods

2.1 System Dynamics and Stock and Flow Diagram

System Dynamics is a discipline that uses modeling and computer simulation to analyze, understand, and improve complex dynamic systems [5], [11] and [14]. The main idea of system dynamics is, that the system behavior is determined mainly by its own structure, structure elements and by the interconnections between them [9], [11]. System dynamics methodology is based on the feedback concepts of control theory [5], the principles of cognitive limitations, mental modelling [6] and bounded rationality [12]. System dynamics is an appropriate technique to handle complex systems, to understand them and to improve system thinking and system learning. To describe and define a system using system dynamics, a Causal Loop Diagram (CLD) is used firstly, and subsequently, a Stock and Flow Diagram (SFD) is created to enable mathematical modelling of the system. The basic building blocks used in the SFD with icons and their interpretation are shown in Table 1.





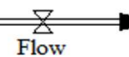




Symbol	Mathematics	Interpretation
	$\partial Y / \partial X > 0$ In the case of accumulations, $Y = \int_{t_0}^t (X + \dots) dt + Y_{t_0}$	All else equal, if X increases (decreases), then Y increases (decreases) above (below) what it would have been. In the case of accumulations, X adds to Y.
	$\partial Y / \partial X < 0$ In the case of accumulations, $Y = \int_{t_0}^t (-X + \dots) dt + Y_{t_0}$	All else equal, if X increases (decreases), then Y decreases (increases) below (above) what it would have been. In the case of accumulations, X subtracts from Y.
		Delay mark
	$Stock(t) = \int_{t_0}^t (Inflow - Outflow) dt + Stock(t_0)$	Stock variable
		Flow variable
		Stocks outside model boundary
		Variable
		Causal loop Reinforcing (+)
		Causal loop Balancing (-)

Table 1 Symbols of the stock and flow diagram (based on [14])

The Stock and Flow diagram represents the structure of the system in terms of stock and flow, and usually follows the Causal Loop Diagram. The stock represents the state (or condition) of thy system and the flow is changed by decisions based on the condition of the presented system [2]. The SFD is the structure of the system and can be simulated to generate the dynamic behavior of the presented system [14], it is represented by integral finite difference equations involving the variables of the presented feedback loop structure of the presented system and simulates its dynamic behavior.

When a model is developed and represented by the SFD, model validation is conducted to develop confidence in the model. The validity and usefulness of a dynamic model should be assessed. Testing means comparing the model with empirical reality for accepting or rejecting the model, and validation means to create confidence in the usefulness of the created model. Sterman [14] lists many different tests for testing the created model. For example, the Test of Model Structure and Dimensional Consistency are the most basic tests. Dimensional Consistency; it should be always specified the units of measure for each variable, of course with real world meaning. The Behavior Reproduction Test should assess a model's ability to reproduce the dynamic behavior of a real system. The Mean Absolute Percent Error (MAPE) is one of the commonly used tools for measuring the average error between the simulated and actual real values.

$$MAPE = \frac{1}{n} \sum \frac{|X_m - X_d|}{X_d}; \quad (\text{multiply by 100 for \%}) \quad (1)$$

where: X_d represents the data series, X_m represents the model outputs and n is the length of the data series.

2.2 Beehive weighing-machines on the *Včelstva Online* web portal

Beehive weighing-machine is an autonomic electronic mechanism under the beehive. It reads weight, inside temperature, outside temperature, and humidity. Beekeepers use a beehive weighing-machine for online observation of bee colony condition at the station of bee colony. The authors of this paper are involved in a research project aimed at the operation and development of online beekeeping web portal *Včelstva online* (<https://vcelstva.czu.cz/>) using bee scales for beekeepers from the public. (No. 2019B0001, Internal Grant Agency, FEM CULS Prague). This paper follows this previous research [3].

The *Včelstva Online* web 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 [3]: collecting phenological records, collecting data from beehives (weight, inside temperature, outside temperature, humidity). Records from the beehive weighing-machine and phenological records from beekeepers are collected on the web portal - paired data can be used for monitoring the landscape fertility.

3 Results and Discussion

3.1 System Dynamic Model of Beehive Trophic Activity

The SFD of the trophic activity of bees was created based on CLD, results of our previous research [3] and continuous results from the *Včelstva Online* web portal (the research project No. 2019B0001, Internal Grant Agency, FEM CULS Prague). This model represents the bees' behavior during a nectar collection season in the spring and it simulates the bees' behavior for one week. The key point of this simplified model of the beehive trophic activity (Figure 1) is the weight of the beehive, which consists of a stock of honey and sugar solution, the weight of bees, and the weight of beehive construction. The weight of beehive is related to another monitored variables - inside temperature of beehive, outside temperature outside of the beehive and beehive's humidity. These variables are collected by beehive weighing-machines. These variables are closely connected and indirectly describe the trophic activity of bees. (Figure 1). The SFD of the trophic activity of bees represents simplification and abridgment of biologic process, which takes place in a beehive with distinction of day and night time in the time during the main beekeeping season.

Very important in the model presented by the SFD are two main feedback loops. The reinforcing feedback loop (in the model displayed by R) describes nectar collection and honey production. With the suitable outside temperatures, healthy bees are physically active in the landscape, collect nectar and process it into honey. The bees collect nectar in the area until the stock of nectar in the landscape is used up, which describes the balancing loop (B). The health of the bees is affected not only by the stock of honey, but also by humidity inside the beehive and by a current temperature. The bees tend to balance a difference between outside and inside temperature in order to preserve a usual (optimal) inside temperature. In summer they cool themselves by increasing the humidity inside the hive and when it is cold outside, they are moving (physically active inside the beehive) to raise the inside temperature. Increased physical activity causes excessive honey consumption, which can decrease the vitality of bees. When their vitality decreases, their death rate increases. To help them survive winter, the beekeeper can feed them with sugar solution, which supplements the honey.

The honey production is affected by the landscape fertility, which is determined by the species diversity of plants, vegetation and crops in the landscape, stock of nectar and stock of pollen. The production is also influenced by the physical activity of bees in the landscape, which is caused by warm outside temperature and light through a day. Cold outside temperature causes low physical activity of bees in the landscape and thus low nectar collection and honey production.

In this simplified model representing the behavior of bees, a Test of Model Structure was performed by comparing the structure of the model with knowledge of the real system, simplified for the situation represented by the SFD. The Dimensional Consistency was also conducted. After inserting the real data of outside temperature into the model, the development of the weight values was determined and in comparison with the real data, MAPE = 5.49

% was calculated according to formula 1. The highest deviation from the real data was 13.80 %. The whole model including simulation runs was developed in Vensim software (industrial strength simulation software for improving the performance of real systems).

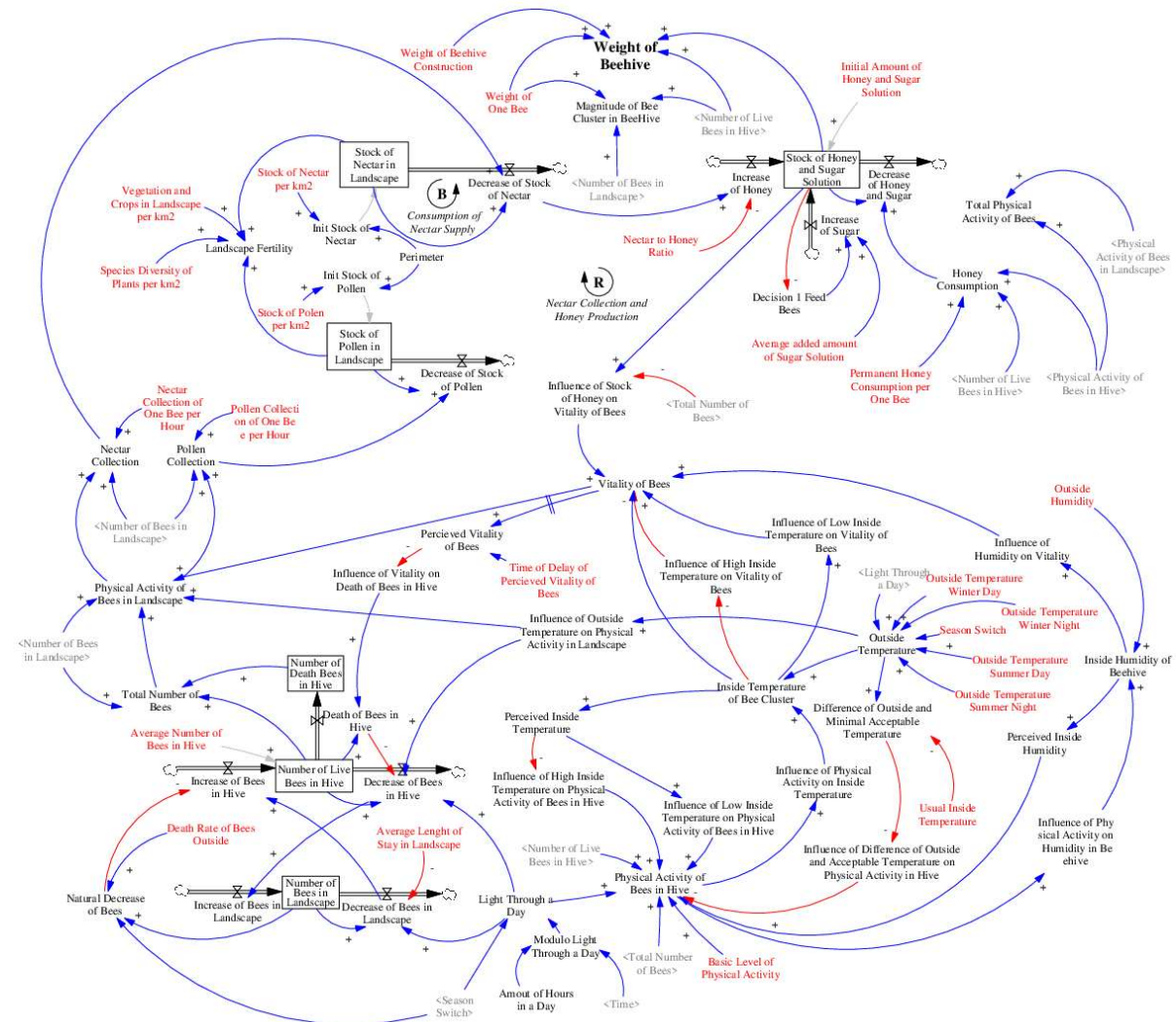


Figure 1 Stock and Flow Diagram - Beehive Trophic Activity

3.2 Trophic Activity of Bees

For beekeepers is very important to monitor the trophic activity of bees (i.e. activities aimed at procuring food supplies for bees) continuously for procuring prospering bee breeding. The trophic activity of bees during the season consists of gathering food supplies from the beehive vicinity (radius 5 km) in the form of nectar and pollen. Bees make the gathering of food supplies every day in relation to outside temperature, when they leave the beehive (the weight of beehive decreasing) and with the time lag bees return (the weight of beehive increasing). Mainly, the observation of the outside temperature and inside temperature of the beehive is crucial because the fluctuation of temperature could endangered bees. The too low or too high outside temperature could cause impossible or decreased gathering of food supplies, too low or to high inside temperature could be endangered beehive (potential death of bees in the consequence of freeze or overheating).

The trophic activity of bees, especially the physical activity of bees in the landscape, is influenced by outdoor weather, especially outside temperature, and the length of sunlight. The physical activity of bees in the landscape is directly affected by the outside temperature. If the outside temperature is low, the physical activity of bees also decreases, and if the outside temperature is too low (less than 12°C), the bees do not fly outside at all. The real data were used to verify the functionality of the model by conducting one simulation run (the data describe weekly observations, i.e. 168 hours; the one-week period is the usual period of monitoring beehives on the *Včelstva Online* web portal; the weekly collection of the observations took place at a selected bee habitats during the beginning of the beekeeping season). After real data entry of the outside temperature to the model, the model simulates the

physical activity of bees in landscape (Figure 2). Based on the model outputs, it is evident that at lower temperatures there is lower activity of bees. To simplify, only outdoor temperature was included as a weather factor in this model. But of course, other weather factors such as rainfall, sunshine intensity and others also influence the activity of the bee colony. For simplicity's sake, the model does not operate with the inside humidity of the hive as a stock variable. This simplification is expected to be removed in the further development of the model.

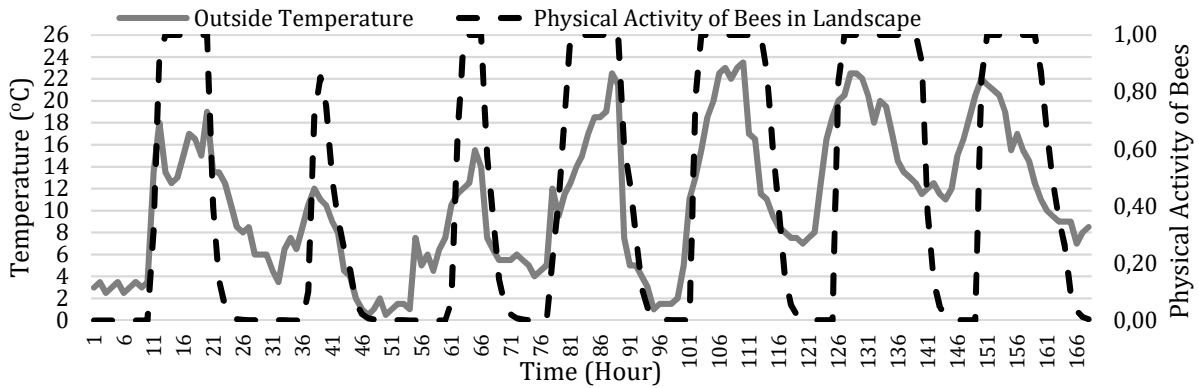


Figure 2 Physical Activity of Bees

Of course, the collection of pollen and nectar needed for honey production also depends on the physical activity of the bees in the landscape. If the environment is cold outside, then with lower physical activity of the bees there is a reduced collection of nectar and thus a lower production of honey. Even this situation can be simulated by the presented model and it is clear from the graph (Figure 3) that in cold weather the production of honey is very low, while at higher outside temperatures the production of honey is higher.

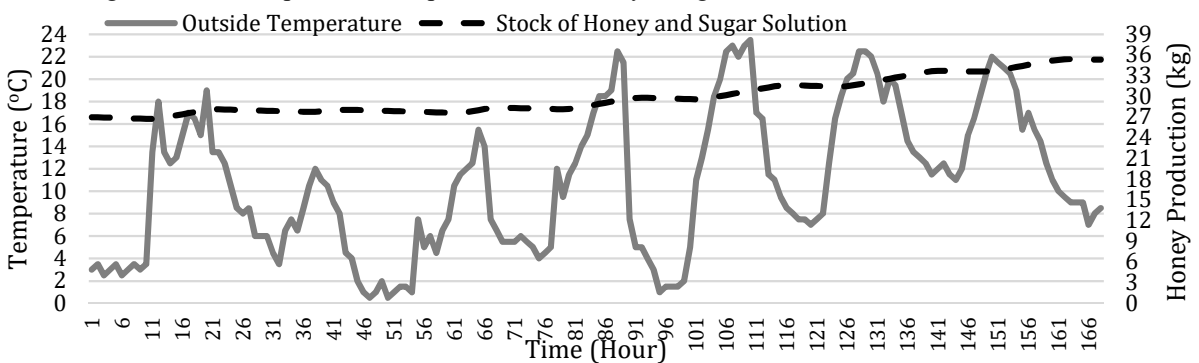


Figure 3 Honey Production of Bees

The outside temperature does not only affect the physical activity of the bees in the landscape, but also in the beehive. It also affects the activity of bees inside the beehive. If the outside temperature is too low, the activity of the bees inside the beehive increases, thus also the inside temperature increases to be as close as possible to their conformal zone around 24°C. In fact, no matter how the temperature fluctuates outside, the average internal temperature of the beehive should be around this conformal temperature.

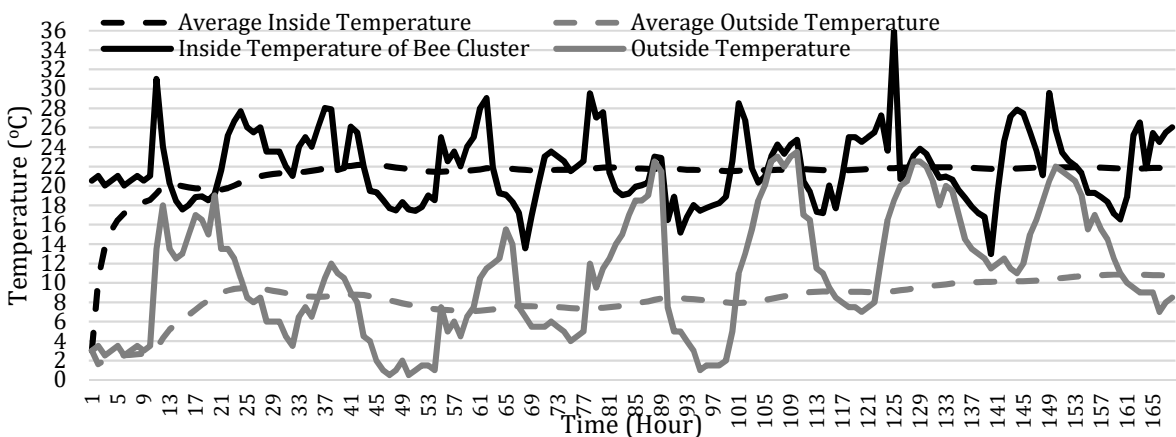


Figure 4 Outside and Inside Temperature of Beehive

Measuring the weight and temperature of beehives has been a common research and beekeeping practice for many decades. The contribution of followed research consists of online collected data due to the web portal *Včelstva Online* and in proposing SFD for the trophic activity of bees. Although, the presented real data, which was used for model verification, could be interpreted in a much wider context. Their course gives evidence of the function and usability of the model for potential prediction of the trophic activity of bees at the real beehive station.

4 Conclusion

The paper deals with creation and application of System Dynamic Model in the form of Stock and Flow Diagram. The proposal follows previous works of authors, especially the model in the form of Causal Loop Diagram [3]. The SFD of the trophic activity of bees is designed and verified on the basis of real data and knowledge obtained from long-term research of authors of Citizen Science type (beekeepers from the public could use beehive weighing-machines and web portal *Včelstva Online*). The results provide a partial point of view that applies to the field of breeding and protection of bees in the case of prediction of elementary parameters of beehive of the beehive station. The presented SFD is possible to use as a basic concept for further development of web portal *Včelstva Online* (e.g. for prediction of beehive condition for beekeepers regards to expected weight) or for further research in the field of fertility landscape for bees – for different types of landscape can be expected different trophic activity of bees. The result is necessary to put to professional discussion and further verification against real data. Plans for further extensions and development of the model are mainly to include other weather conditions influencing bees' activity, such as rainfall, sunshine intensity, possibly windiness, and scaling of colonies by population size, among others. Furthermore, modelling the effect of nectar and pollen supply in the surrounding area could prove helpful.

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, Czech University of Life Sciences Prague.

References

- [1] Bagheri, S. (2019). A mathematical model of honey bee colony dynamics to predict the effect of pollen on colony failure. *PLoS ONE* 14(11): e0225632.
- [2] Bala, B.K., Arshad, F.M. & Noh, K.M. (2017). *System Dynamics: Modelling and Simulation*. Springer Texts in Business and Economics, ISBN 978-981-10-2043-8.
- [3] Bartoška, J., Šubrt, T., Rydval, J., Kazda, J., Stejskalová, M. (2020). System Dynamic Conceptual Model for Landscape Fertility of Bees. In: *Proceedings of the 38th International Conference On Mathematical Methods in Economics*. Brno: Mendel University. ISBN 978-80-7509-734-7.
- [4] Bruce (2014). *Modeling Dynamic Biological Systems*. Springer International Publishing, 2nd Edition. ISBN: 978-3-319-05614-2.
- [5] Forrester, J.W. (1968). *Principles of system dynamics*. MIT Press, Cambridge, MA.
- [6] Kim, D.H. & Senge, P.M. (1994). Putting systems thinking into practice, *Syst. Dyn. Rev.*, 10 (2-3), 277-290.
- [7] Krejci, I., Moulis, P., Pitrova, J., Ticha, I., Pilar, L. and Rydval, J. (2019). Traps and Opportunities of Czech Small-Scale Beef Cattle Farming. *Sustainability*. 11 (15), ISSN 2071-1050, DOI: 10.3390/su11154245.
- [8] Lechmere-Oertel, R. G. (2005). Landscape dysfunction and reduced spatial heterogeneity in soil resources and fertility in semi-arid succulent thicket, *Austral Ecology*. 30 (6) 615-624.
- [9] Meadows, D.H. (2008). *Thinking in Systems. A Primer*. Chelsea Green Publishing Company.
- [10] Russel (2013). Dynamic modelling of honey bee (*Apis mellifera*) colony growth and failure. *Ecological Modelling* 265 (2013) 158-169.
- [11] Senge, P.M. (2006). *Fifth Discipline: The Art and Practice of the Learning Organization*. Random House Business Books.
- [12] Simon, H.A. (1979). Rational decision making in business organizations, *Am. Econ. Rev.* 69 (4), 493-513.
- [13] Smith, K. T. (2010). *Biogeochemistry and landscape fertility*. The Landsculptor, February: 43-45.
- [14] Sterman, J.D. (2000). *Business Dynamics: System Thinking and Modelling for a Complex World*. Irgwin/McGraw-Hill, Boston.

An Analysis of Dependence between German and V4 Countries Stock Market

Radmila Krkošková¹

Abstract. The topic of relations between individual markets has been frequently discussed recently. Especially on the stock markets, we can watch a tendency of the more developed markets to affect developments on the less developed markets. This is also valid for the V4 stock markets, where it is potential to anticipate a strong influence of the German stock market. There has been used the Granger causality. Quarterly data for the period from 2005/Q1 to 2020/Q4 was used for the analysis. This period has been selected because all of the V4 countries have been members of the European Union since 2004. The EViews software version 11 was used for the calculations. Variables used in this research are the stock exchange indices of the countries. The PX, SAX, BUX, WIG 20 and DAX stock indices are considered to be the crucial representatives of individual stock markets in this work. The results show that German DAX stock index was Granger-causing the development of the Czech (PX), Hungarian (BUX) and Polish (WIG 20) stock indices.

Keywords: ADF test, Granger causality, stock market, V4

JEL Classification: C19, F65, G15

AMS Classification: 62P20, 91B28

1 Introduction

The Visegrad Four (V4), an informal grouping of the Czech Republic, Slovakia, Poland and Hungary, commemorates 30 years since its inception. According to recent research the biggest economic leap in three decades was made by Poland, but the economic leader of the countries is still the Czechia. Some analysts also believe that today Visegrad has a rather political significance that outweighs the economic one.

As far as the economic level is concerned, the V4 countries have, above all, a very strong individual economic connection with Germany, both in terms of investment and foreign trade. Economic flows within the V4 region are weaker than between the individual V4 countries and Germany. For example, the share of foreign trade in goods in 2020 in the case of the Czech Republic with the V4 countries was less than two-thirds of what it did with Germany. Although the starting positions of individual states were relatively different in the early 1990s and the elements of economic transformation also varied considerably, all four countries managed to compare the values of macroeconomic indicators in thirty years. Selected macroeconomic indicators of the V4 countries and Germany are shown in Table 1.

Indicator	GDP per capita, PPP (U.S.dollars)		Unemployment rate (%)		Inflation (%)	
	1991	2019	1991	2020	1991	2020
Country/Year						
Czech Republic	20962	40836	4.1	3.1	56.7	3.2
Slovakia	11728	31966	11.8	7	61.2	1.6
Hungary	16477	32644	12.3	4.3	35	2.4
Poland	10517	33222	11.8	3.3	59.4	3.4
Germany	38360	53785	5.3	4.3	4	0.4

Table 1 Selected macroeconomic indicators

Table 1 shows that in 2020 Slovakia reached the highest value of unemployment rate and it was 7% whereas values of other countries were lower. As for inflation, Poland has 3.4%, and it is the highest value from those countries. Germany has the highest value of GDP per capita.

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

2 Literature Review and Data

2.1 Literature Review

From the point of view of the markets of the V4 countries (Czech Republic, Hungary, Poland, Slovakia), it is possible to assume a significant influence of the German stock market in particular. This is due not only to geographical proximity but also to strong economic ties, as shown by several studies (Baláž and Hamara, [3]; Elekdag et al., [5]; Komárek et al., [12]; Taušer et al., [20]). In these countries, German capital is strongly active in the form of foreign direct investment, and so this correlation is logical.

Developments in the stock markets of Central and Eastern Europe (CEE) are a topic that has been addressed by many authors. Hegerty [10] deals mainly with the effects of oil price volatility on their stock markets. He concluded that the impacts vary significantly, depending on the level of the country's economic level. Ison and Hudson [11] addressed the question of whether it is possible to predict developments in the stock market markets of the CEE region by analyzing previous price movements. Árendáš and Chovancová [1] found that the stock markets of the V4 countries, with the exception of the Slovak stock market, tend to perform significantly better in the winter than in the summer half of the year. In the case of the Slovak stock market, the difference between the two half-years is negligible. Árendáš et al., [2] describe the influence of German stock market on stock market of V4 countries. Pietrzak et al. [16] and Cevik et al. [4] examined the interdependencies between the individual stock markets of the CEE countries. Reboredo et al. [17] examined the dependencies between the stock markets of the Czech Republic, Hungary, Poland and Romania. They found that there was a strong positive dependence between the stock markets of the Czech Republic, Hungary and Poland. The relationship between these three stock markets and the Romanian stock market is significantly weaker. The effects of the global financial crisis on the stock markets of the CEE region were also addressed by Olbrys and Majewska [15], while Vychytilová [21] focused mainly on the post-crisis development of the stock markets of the V4 countries.

2.2 Data and Methods

The aim of this paper is to examine the existence of a causal relationship between the German stock market and the stock markets of the V4 countries in the period 2005-2020. Quarterly data for the period from 2005/Q1 to 2020/Q4 was used for the calculations. All values were seasonally adjusted and were considered in logarithmic terms. The EViews software version 11 was used for the calculations. Time series were obtained from the Bloomberg database.

The German stock market is represented by the DAX stock index, which includes the 30 most important German joint stock companies. The stock markets of the V4 countries are represented by their main stock indices, namely: BUX (Hungary), PX (Czech Republic), SAX (Slovakia) and WIG 20 (Poland).

There is tested following hypothesis: There is a Granger causality between the individual stock markets of the V4 countries and the German stock market, when the German stock market represented by the DAX stock index influences the development of the stock markets of the V4 countries, represented by the PX, BUX, WIG 20 and SAX stock indices.

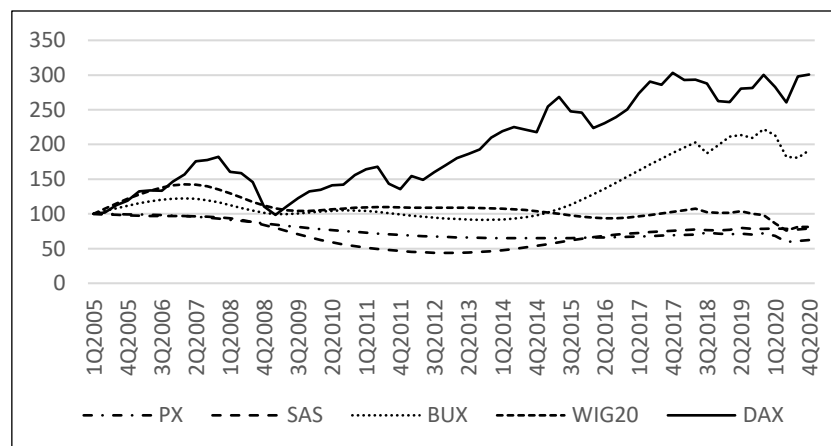


Figure 1 Development values of stock indices

The development of the values of individual stock indices is shown in Figure 1. At the beginning of the period from 2005 to 2007, the values of all five stock indices were increasing. During the crisis years of 2008 and 2009, all stock indices declined. During the post-crisis period, it is possible to observe, above all, the long-term stagnation of the Polish and Czech stock markets. Although the Hungarian economy and the banking sector have been hit hard by the global financial crisis, a very sharp rise in the value of the BUX index has been observed since 2014.

The main steps of the research are: stationarity test; if the levels values of the series are stationary we should use VAR model and if the differences of the variable are stationary we test cointegration; with regard to the results we use VAR model (if there is no cointegration relationship) or VECM model (if there is cointegration relationship); and finally, it is tested the Granger causality.

The causal dependence of the development of the values of the German stock index and the stock indices of the V4 region is tested using Granger causality. Granger [6] based causal dependency testing on the use of the following VAR model:

$$X_t = \sum_{j=1}^m a_j X_{t-j} + \sum_{j=1}^m b_j Y_{t-j} + \varepsilon_t \quad (1)$$

$$Y_t = \sum_{j=1}^m c_j X_{t-j} + \sum_{j=1}^m d_j Y_{t-j} + \delta_t \quad (2)$$

The Granger causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another [6]. Using the term "causality" is a misnomer, as Granger-causality is better described as "precedence" [14], or, as Granger later claimed "temporally related" [7]. Rather than testing whether Y causes X , the Granger causality tests whether Y forecasts X , [8]. It is necessary to work with the stationary time series.

3 Data Analysis

The modelling structure is similar for all the countries studied and consists of the following steps: testing the presence of unit roots, the VAR model estimation, response impulse analysis, and Granger causality. The similar procedure is listed by Krkošková [13]. All values were *seasonally adjusted* and were considered *in logarithmic terms*.

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 last column includes the result of testing: N = non-stationary (H0 not rejected), S = stationary (H0 rejected).

Variable	n/c/c+t	T-stat	Signif.	Result	Variable	n/c/c+t	T-stat	Signif.	Result
<i>PX</i>	c	0.629	0.989	N	D(<i>PX</i>)	c	-2.948	0.046	S
<i>SAX</i>	n	-1.191	0.211	N	D(<i>SAX</i>)	n	-2.984	0.044	S
<i>BUX</i>	n	0.695	0.862	N	D(<i>BUX</i>)	n	-1.933	0.049	S
<i>WIG 20</i>	c	0.229	0.972	N	D(<i>WIG 20</i>)	c	-2.946	0.046	S
<i>DAX</i>	c	-1.349	0.601	N	D(<i>DAX</i>)	c	-6.036	0.000	S

Table 2 Testing the unit root of the variables in levels and their first differences

The time series are non-stationary and co-integrated. But calculations showed that no cointegration relationship was demonstrated here. Therefore, it has been used the VAR model. Figure 2 shows the impulse response functions. This article deals with the response of the stock indices of V4 countries to shock in the change in the *DAX* index growth rate.

For the Czech Republic there is an immediate reaction in both the *DAX* and the *PX* variable. The *PX* shock is absorbed faster than the *DAX* shock. The impulse response functions are similar for Slovakia and Hungary. There are immediate reactions in both variables, the effect of which persists for several periods. In the case of Poland, we can see that shocks both variables are absorbed during four periods.

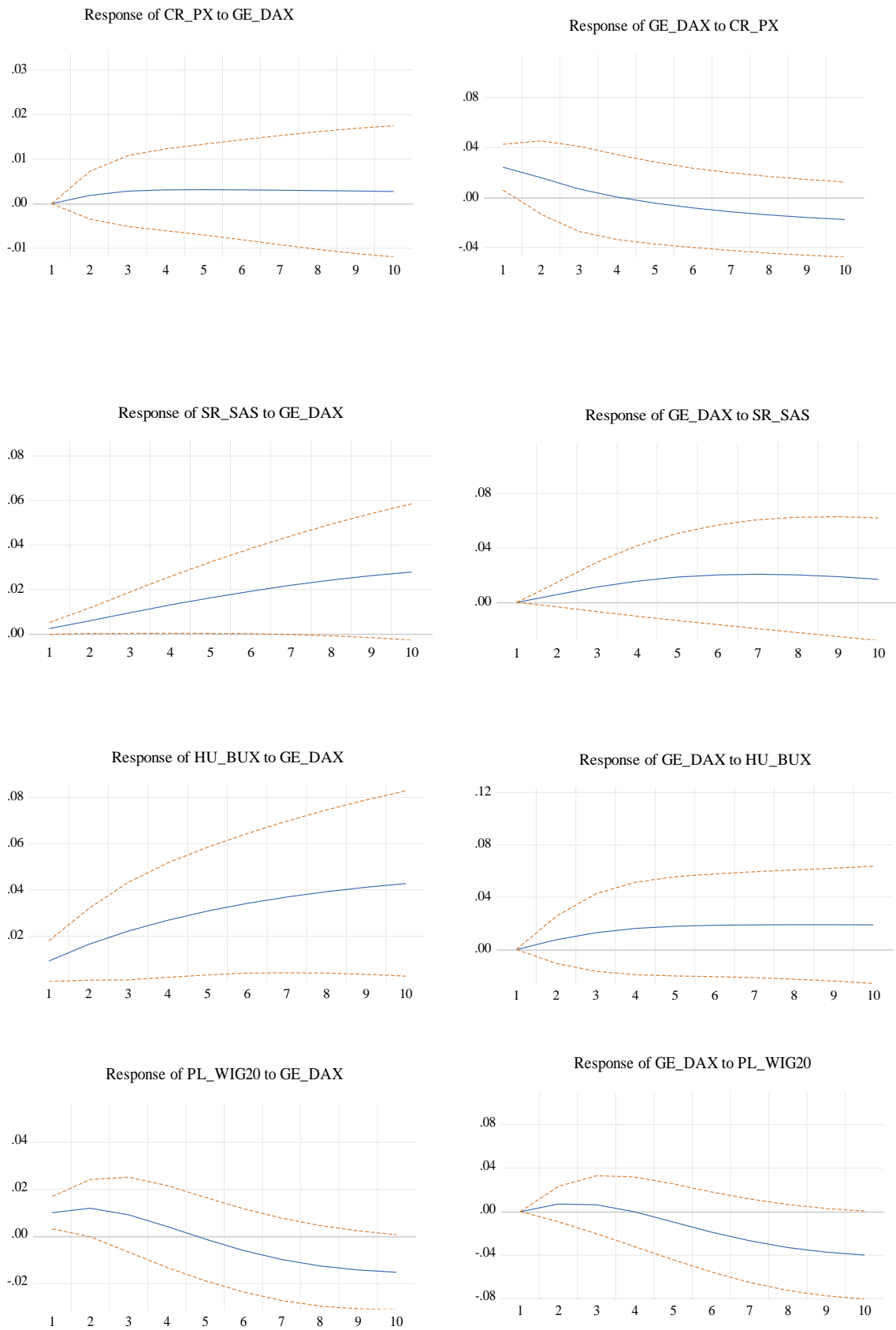


Figure 2 Response to Cholesky One S.D. Innovations

This part deals with the testing of short-term relationships (Granger causality). The results of the series 1 delay test are shown in Table 3.

Null Hypothesis:	F-Statistic	p-value	Result for $\alpha = 0.1$
D(PX) does not Granger Cause D(DAX)	6.7	0.091	TO REJECT
D(DAX) does not Granger Cause D(PX)	50.6	0.000	TO REJECT
D(SAX) does not Granger Cause D(DAX)	0,25	0.861	NOT TO REJECT
D(DAX) does not Granger Cause D(SAX)	3,1	0.213	NOT TO REJECT
D(BUX) does not Granger Cause D(DAX)	14.8	0.421	NOT TO REJECT
D(DAX) does not Granger Cause D(BUX)	35.4	0.001	TO REJECT
D(WIG20) does not Granger Cause D(DAX)	0.92	0.872	NOT TO REJECT
D(DAX) does not Granger Cause D(WIG20)	12.3	0.021	TO REJECT

Table 3 Pairwise Granger causality test (Lag 1)

The results of the Granger causality test for individual time series pairs are shown in Table 3. The results for the period 2005–2020 confirm that the German stock market affects the stock markets of the V4 countries. At the significance level $\alpha = 0.05$, it is possible to observe Granger causality, when the development of the value of the German stock index *DAX* affects the Hungarian *BUX*, as well as the Czech *PX* and Polish *WIG 20*. The only exception is the Slovak stock index *SAX*, for which no statistically significant the impact of the German *DAX* index. The relationship between the *PX* and *DAX* indices is also interesting, where the results suggest that at the significance level $\alpha = 0.1$ there is Granger causality in the direction from the Czech to the German stock market.

4 Conclusion

According to recent research Germany has become the first export and import partners in each Visegrád country. German trade represents the largest share in the Czech Republic and the smallest one in Slovakia. It is not surprising that the smaller and less developed countries are dependent on German economy but this dependency is mutual. It is natural that there could be found relationship between stock market of Visegrád countries and Germany.

The results confirm that the hypothesis (there is a Granger causality between the individual stock markets of the V4 countries and the German stock market, when the German stock market represented by the *DAX* stock index affects developments in the stock markets of the V4 countries represented by *PX*, *BUX*, *WIG 20* and *SAX* stock indices) applies only to the Czech, Hungarian and Polish stock markets.

The results show that for the period 2005–2020 there is a Granger causality between the development of the stock markets of the Czech Republic, Poland, and Hungary and the development of the German stock market. This finding is consistent with Harkmann's conclusions [9], which found out a steady increase in the correlation between Western European and Eastern European stock markets. The influence of Germany's macroeconomic indicators on the stock markets of the V4 countries are also confirmed by a study by Samitas and Kenourgios [18]. They found that the stock markets of the V4 countries are influenced by German macroeconomic indicators, while the impact of US macroeconomic indicators is not significant. On the contrary, the Slovak stock market does not depend on the German stock market. This situation can be explained mainly by the low degree of development of the Slovak stock market and its extremely low level of liquidity. A similar conclusion was reached by Stoica et al. [19]. They found that the stock markets of the CEE region were responding to developments in the stock markets of Germany and France. However, they stated that Slovakia is the only one of the six countries examined (Bulgaria, the Czech Republic, Hungary, Poland, Romania, Slovakia), whose stock market is developing independently. The Slovak stock market behaves differently compared to other stock markets in the region.

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 2021.

References

- [1] Árendáš, P. & Chovancová, B. (2016). Central and Eastern European Share Markets and the Halloween Effect. *Montenegrin Journal of Economics*, 12(2), 61–71.
- [2] Árendáš, P., Chovancová, B. & Pavelka, L. (2020). Vplyv nemeckého akciového trhu na akciové trhy krajín V4. *Politická ekonomie*, 68(5), 554–568.
- [3] Baláž, P. & Hamara, A. (2016). Analýza závislosti exportu SR na vývoji ekonomiky SRN. *Politická ekonomie*, 64(5), 573–590.
- [4] Cevik, E. I., Korkmaz, T. K. & Cevik, E. (2017). Testing Causal Relation among Central and Eastern European Equity Markets: Evidence from Asymmetric Causality Test. *Economic Research – Ekonomika Istrazivanja*, 30(1), 381–393.
- [5] Elekdag, S., Muir, D. & Wu, Y. (2015). Trade Linkages, Balance Sheets, and Spillovers: The Germany-Central European Supply Chain. *Journal of Policy Modeling*, 37(2), 374–387.
- [6] Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3), 424–438.
- [7] Granger, C.W.J. & Newbold, P. (1977). *Forecasting Economic Time Series*. New York: Academic Press.
- [8] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- [9] Harkmann, K. (2014). Stock Market Contagion from Western Europe to Central and Eastern Europe During the Crisis Years 2008–2012. *Eastern European Economics*, 52(3), 55–65.
- [10] Hegerty, S. W. (2015). Oil-Price Volatility and Macroeconomic Spillovers in Central and Eastern Europe: Evidence from a Multivariate GARCH Model. *Zagreb International Review of Economics & Business*, 18(2), 31–44.
- [11] Ison, L. & Hudson, R. (2017). Stock Predictability and Preceding Stock Price Changes – Evidence from Central and Eastern European Markets. *Economics Bulletin*, 37(2), 733–740.
- [12] Komárek, L., Motl, M. & Novotný, F., et al. (2012). Německá ekonomická “lokomotiva” a česká ekonomika. *Politická ekonomie*, 60(4), 442–458.
- [13] Krkošková, R. (2020). Impact of Stock Markets on the Economy in V4 Countries. *E&M Economics and Management*, 23(3), 138–154.
- [14] Leamer, E. E. (1985). Vector Autoregressions for Causal Inference? *Carnegie-Rochester Conference Series on Public Policy*, 22(1), 255–304.
- [15] Olbrys, J. & Majewska, E. (2016). Crisis Periods and Contagion Effects in the CEE Stock Markets: the Influence of the 2007 US Subprime Crisis. *International Journal of Computational Economics and Econometrics*, 6(2), 124–137.
- [16] Pietrzak, M. B., Faldzinski, M. & Balcerzak, A. P., et al. (2017). Short-term Shocks and Long-term Relationships of Interdependencies among Central European Capital Markets. *Economics & Sociology*, 10(1), 61–77.
- [17] Reboredo, J. C., Tiwari, A. K. & Albulescu, C. T. (2015). An Analysis of Dependence between Central and Eastern European Stock Markets. *Economic Systems*, 39(3), 474–490.
- [18] Samitas, A. G. & Kenourgios, D. F. (2007). Macroeconomic Factors’ Influence on “New” European Countries’ Stock Returns: the Case of Four Transition Economies. *International Journal of Financial Services Management*, 2(1/2), 34–49.
- [19] Stoica, O., Perry, M. J. & Mehdian, S. (2005). An Empirical Analysis of the Diffusion of Information Across Stock Markets of Central and Eastern Europe. *Prague Economic Papers*, 24(2), 192–210.
- [20] Taušer, J., Artlová, M. & Žamborský, P. (2015). Czech Exports and German GDP: A Closer Look. *Prague Economic Papers*, 24(1), 17–37.
- [21] Vychytilová, J. (2018). Stock Market Development Beyond the GFC: The Case of V4 Countries. *Journal of Competitiveness*, 10(2), 149–163.

Evaluation of the Construction Sector: a Data Envelopment Analysis Approach

Markéta Křetínská¹, Michaela Staňková²

Abstract. This article deals with the evaluation of the efficiency of the construction sector at the level of individual EU countries. The construction sector is an integral part of the economy in every country, and has a considerable impact on the economy. Efficiency is calculated using a non-parametric deterministic method of data envelopment analysis. Radial input-oriented models are constructed separately for individual years. The gross output and value added variables are used as the output variables. The labor and capital factors (i.e., inputs) are expressed using the total hours worked and nominal gross fixed capital formation variables. The data used are contained in the EU KLEMS database.

The results show that countries such as Austria and Belgium are efficient throughout the period under review. The efficiency of the Austrian construction industry is largely determined not only by a high increase in production volume (output of construction, construction of civil engineering and construction of buildings), but also by an increase in labor productivity in this sector which is above the European average.

Keywords: construction, data envelopment analysis, efficiency, linear programming

JEL Classification: C44, D24

AMS Classification: 90B50, 90C08

1 Introduction

Efficiency evaluation is currently a much-discussed topic. In research into it, attention is focused on the resources used and the outputs produced. The issue of efficiency evaluation is closely linked to the growing effort not to waste especially limited resources. Competition forces companies to constantly push their limits as well as to minimize the amount of resources used and of course their cost. Therefore, efficiency is an important aspect for the success of every company.

In the field of efficiency evaluation from a non-parametric linear programming approach, the Data Envelopment Analysis (DEA) method dominates. It is a deterministic method which measures relative efficiency derived mainly from the technologies used by units. DEA is based on Farrell's idea [5], which described in detail procedures for measuring productivity in another way than by using indices. The CCR model [8] was introduced in 1978 and the BCC model [2] later in 1984. These models are considered the basic models of DEA. DEA measures the efficiency of each decision-making unit (DMU) relatively to another DMUs. Current studies focusing on measures of efficiency via DEA method are: [6], [12], [13], [14] and [15].

Efficiency is very important in every sector of the economy, including the construction sector. The construction sector plays an important role in every country of the European Union as well as throughout the world. Construction forms an integral part of the national and international economy. The construction industry accounts for 6% of global GDP and employs about 100 million people worldwide. Within the EU, construction generates about 9% of gross domestic product and provides 18 million jobs.

Construction relates to many other sectors across the economy [10]. It involves many processes such as raw materials extraction, supply services, production processes, distribution of construction products up to design, management and control of workers, demolition and recycling of construction and demolition waste [1]. The construction of buildings such as housing for the private and public sectors, infrastructure and constructions output have a major impact on the environmental and on quality of life. For example, [9] confirms the important

¹ Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, xkretins@node.mendelu.cz.

² Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, michaela.stankova@mendelu.cz.

role of the construction sector and shows that most people in USA spend about 90% of their time indoors. The construction sector is the biggest consumer of steel in the world. Buildings consume over 20% of energy and 40% of electricity worldwide. In the EU the consumption of energy is about 36%. Yet the construction sector is still poorly explored.

[7] shows that most companies in Russia are using recycled construction and demolition waste in very small quantities. Companies are thus losing a large amount of high-quality materials. By recycling materials, they can minimize the cost of extracting natural resources and reduce the burden on the environment, which plays an increasingly important role these days. With efficient recycling, companies can even make a profit from processing waste materials. [11] surveyed UK companies and found that most of them are having problems with time delays, cost limit over-runs and have various errors in the quality of construction. [15] examines the efficiency trends within the building projects sector in the Czech Republic between 2013 and 2015. They had financial data for each company in the construction sector, and they found that most companies are inefficient in their activities. In more detailed analyses of efficiency over time they found out that some companies changed dramatically in terms of efficiency between the dates monitored.

The main aim of this article is to evaluate the development of the efficiency of the whole construction sector in European Union (EU 28) and compare efficiencies between countries. This is an attempt to follow up on publication from [15]. In contrast to their article, in this study we will work with aggregated data for the whole construction sector of each EU country.

2 Material and Methods

The data file is formed from selected indicators of EU countries from the EU KLEMS database. The chosen input variables are total hours worked by employees (HEMPE) in thousands, and nominal gross fixed capital formation (GFCF) in millions of euros. The output variables are gross output in current prices (GO) and gross value added at current basic prices (VA); both are stated in millions of euros. Due to the fact that the EU KLEMS database provides financial data in national currencies, all values have been converted to euros at current annual exchange rates in order to compare the efficiency of individual countries. Annual data are examined from 2015 to 2017. This period is the most currently available.

Because of missing variable values, we have had to remove countries like Croatia, Ireland, and Poland. In 2015 and 2016 there are 25 countries. In 2017, the number of available data points decreased further and therefore in this period we were forced to omit Cyprus, Estonia, Latvia, Portugal, Romania, Spain, Sweden and UK. The basic characteristics of the selected variables are given in Table 1. The average hours worked by employees is growing during the period under review. By contrast, the median and average of gross output is decreasing. The maximum values of nominal gross fixed capital formation were always reached in the UK. The lowest values for added value are typical in Malta throughout the observed period.

Variable	Type	Period	Min	Max	Average	Median
Hours worked by employees	Input	2015	14 941	2 852 000	654 284	297 756
		2016	15 585	2 889 000	663 744	299 264
		2017	15 515	2 937 000	679 140	305 100
Nominal gross fixed capital formation	Input	2015	3	32 182	2 889	825
		2016	13	27 747	2 882	894
		2017	31	28 297	3 696	1 167
Gross output	Output	2015	1 077	362 113	68 398	27 510
		2016	1 081	336 399	68 781	27 332
		2017	1 201	312 605	61 129	18 782
Value added	Output	2015	328	140 897	26 504	8 513
		2016	327	133 905	26 841	8 694
		2017	359	144 300	23 873	7 974

Table 1 Basic characteristics of selected variables. Hours worked by employees in thousands, nominal gross fixed capital formation, gross output and value added in millions of euros.

The DEA models were constructed separately for particular years. An input-oriented CCR model was used to calculate the efficiency:

$$\begin{aligned}
 \max \quad & E_H = \sum_{j=1}^n v_{jH} y_{jH}, \\
 \text{subject to} \quad & \sum_{i=1}^m u_{iH} x_{iH} = 1, \\
 & - \sum_{i=1}^m u_{iH} x_{ik} + \sum_{j=1}^n v_{jH} y_{jk} \leq 0, \forall k = 1, 2, \dots, p, \\
 & u_{iH} \geq \varepsilon, \forall i = 1, 2, \dots, m, \\
 & v_{jH} \geq \varepsilon, \forall j = 1, 2, \dots, n,
 \end{aligned}$$

where the model is constructed for unit H , which is 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.

The Malmquist productivity index (MI) is used to calculate the change in efficiency over time. This index allows us to monitor separately the change in technical efficiency (the so-called catch-up effect) and changes in the production frontier (frontier-shift). Values greater than one in the index itself or its sub-components mean an improvement in the given area, values less than one mean a deterioration. The MI can be defined as the geometric mean of two efficiency ratios, where one is the efficiency change measures by the period 1 technology and the other is the efficiency change measured by the period 2 technology:

$$MI = \left[\frac{E^1((x_H, y_H)^2)}{E^1((x_H, y_H)^1)} * \frac{E^2((x_H, y_H)^2)}{E^2((x_H, y_H)^1)} \right]^{1/2}.$$

To calculate the MI, it is therefore necessary to solve four linear programs (i.e. four CCR models in this paper), which correspond to the four terms that make up the MI. Technical details about all the above stated procedures can be found in [3]. The calculations are performed in the MATLAB R2021a computational system and in DEA SolverPro version 15.

3 Results

The results of the efficiency scores of individual countries in particular years are recorded in Figure 1. In 2015, there are four fully efficient countries: Austria, Belgium, Netherlands and Spain. In 2016, Denmark joined this group. All these results are supported by the European Commission's reports [4]. Austria has earned its privileged position mainly on above-average labor productivity with regard to the size of the working-age population. Furthermore, it has taken several initiatives to support entrepreneurship and, in particular, start-ups. The fact that the production index in the sector increased by 20% (the index is only available for the total period from 2010 to 2019) also corresponds to the full efficiency of Austria, and an increase of 15% in the number of enterprises was also observed here. GDP and volume index of production are increasing in Belgium over the period. Efficiency has also been influenced by the increasing investment ratio and mean equalized net income followed with total population living in cities. In Netherlands the volume index of production since 2013 and the number of enterprises has significantly increased. The Netherlands is among the most productive economies in the EU, with a 29% higher labor productivity rate than the EU average. The unemployment rate is 2.2% and it is below the EU average of 6%. The Netherlands is also 25th out of 190 countries in terms of starting up a business. Denmark is a leader in eco-innovation and sustainable construction. The Danish government is systematically trying to strengthen productivity and employment in the building sector.

In addition to Denmark, France, Germany, Italy, Sweden, Slovakia and the UK are among the countries with high efficiency scores during the first two years. By contrast, countries such as Bulgaria, Hungary, Romania and Lithuania have the lowest efficiency scores. In terms of absolute efficiency values, Bulgaria lags the most. In the case of this country, several problem areas can be identified. Examples include: the continuing shortage of a skilled and professional workforce, the business environment remains heavily regulated, the lower digital skills of its workforce, labor productivity significantly below the EU average, late payments and a large budget deficit [4].

In 2017, we have data on only 17 countries and the results of efficiency in this period therefore represent a less comprehensive view of the surveyed sector. However, despite this problem with 2017 data availability, Austria, Belgium and Netherlands are among the top countries. Unfortunately, the last of the fully efficient countries from previous years (Spain) did not have data available. Even in the last monitored year, Bulgaria is in the last position in terms of the order derived from the resulting efficiency. However, Hungary, for example, improved, skipping Greece, which took the second worst place in terms of efficiency in 2017. A detailed view of efficiency changes over time was performed using the Malmquist index. The changes in the efficiency of individual countries calculated on the basis of the breakdown of the Malmquist index are presented in Figure 2. As it is not possible to use an unbalanced data set when calculating the Malmquist index, the index was calculated only for 17 countries that have values available for the entire reference period.

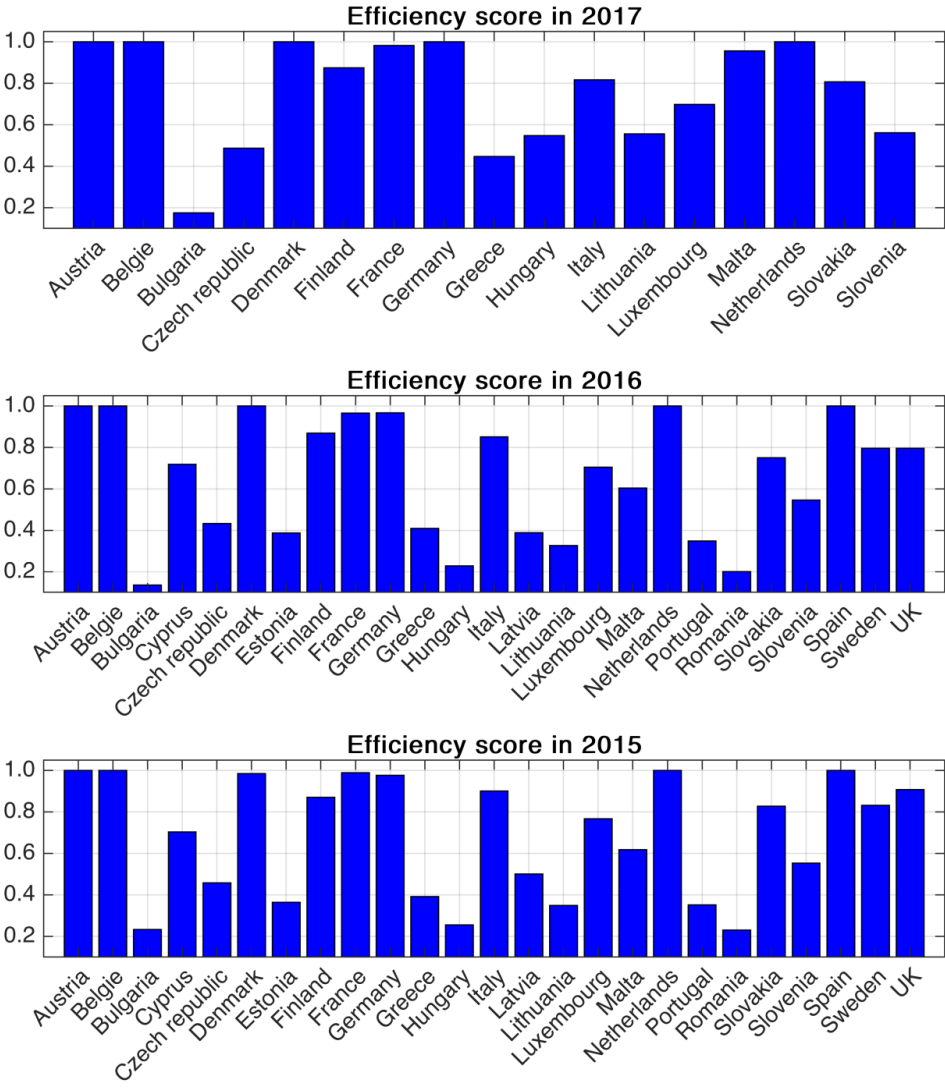


Figure 1 Efficiency scores of individual countries in particular years

Countries that have values higher than 1 in Figure 2 were able to increase their efficiency over time. Conversely, for countries with values below 1, efficiency has decreased over time. A value of 1 means that there have been no changes in efficiency in the case of a given country. It was found that the largest efficiency gains from 2015 to 2016 were recorded in Bulgaria and the Netherlands. The government of Bulgaria invested a large amount to renovate and modernize new highways, railway, and metro in this period. Also, the European Structural Invest Fund (ESIF) provided Bulgaria with development finance.

When comparing the changes from 2015 to 2016, efficiency dropped dramatically in the case of Lithuania. On the other hand, between 2016 and 2017 Lithuania is one of the countries with the biggest increase in efficiency. In 2016 the consumer confidence indicator in Lithuania was -7.6 which is below the EU average of -6.3. This fact reflects the continuous risk perception in the construction sector. However, all confidence indicators have generally improved later. After 2016, Lithuania also improved in terms of dealing with construction permits and

it is considered as a moderate innovator [4]. A similar development in efficiency as for Lithuania has been observed in the Czech Republic.

When monitoring the efficiency changes between 2016 and 2017 (upper part of Figure 2), it can be stated that in this period there were no such dramatic changes as in the previous period. In general, it can be stated that the efficiency of countries increased more often than decreased. The median and average value of the change in efficiency is equal to 1.01 in this period.

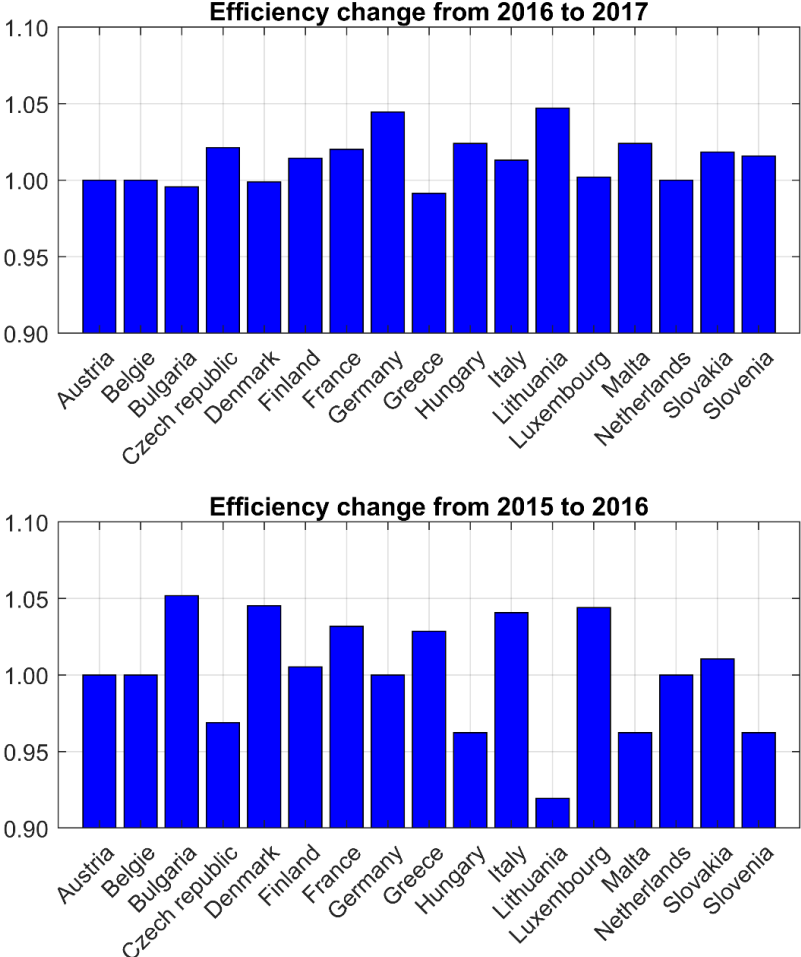


Figure 2 Change in the efficiency of individual countries based on the decomposition of the Malmquist index

4 Discussion

When evaluating the efficiency of the construction sector, it was found that most countries were identified as inefficient. Only around seven countries maintained a high efficiency score during the whole period. The construction sector is currently still a relatively little explored area and there are very few studies focusing on the efficiency of this sector. However, it can be stated that the results of this study are in agreement with the analyzes performed on the microdata in the study [15]. Study [15] showed that companies within the Czech construction industry were not doing their best during 2013 to 2015. Even in the case of our evaluation at the level of individual EU countries, the Czech Republic is among the ten worst countries in 2015 (as in 2016 and 2017). The reliability of the performed analyses is also supported by the results of study [11], where multiple problems were identified in the case of the UK. Although it was not possible to fully analyse the UK due to the unavailability of data, from 2015 to 2016 the UK was always classified as an inefficient country.

Due to the poor availability of data for 2017, it is necessary to take the resulting values of efficiency in this period only as a basic view of the issue. For a more detailed description of construction sector, it would be better to do a deeper survey in the construction sector by extending the timeline and including all EU countries. Using a longer period of time, the plausibility of DEA models could be verified using competitive (parametric) methods.

5 Conclusion

In this article, we focused on evaluating the efficiency of the construction sector at the level of individual EU countries. The results showed that Austria, Belgium, Denmark, the Netherlands and Spain are among the most powerful countries. By contrast, Bulgaria lags behind the others the most. Based on the analysis of changes in efficiency, it can be stated that over time, changes (either in the positive or negative direction) in efficiency have decreased. In general, it can be said that the selected countries have succeeded in improving the efficiency of the sector over the years. Governments are striving for higher company competitiveness and sustainable and ecological development of the construction sector. Among other things, the pressure to invest in equipment innovation and modernization is increasing. However, the importance of construction in the economy is still overlooked from a research perspective.

Acknowledgements

This article was supported by the grant No PEF/TP/2021002 of the IGA PEF MENDELU Grant Agency .

References

- [1] Ahmad, T. & Thaeheem, J. M. (2018) Economic sustainability assessment of residential buildings: A dedicated assessment framework and implications for BIM. *Sustainable Cities and Society*, 38, 476–491.
- [2] Banker, R. D., Charnes, A., Cooper, W. W. (1984). The Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, 30(9), 1078–1092.
- [3] Cooper, W. W., Seiford, L.M. & Tone, K. (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA/Solver Software*. New York: Springer Science & Business Media.
- [4] European Commission. (2021). *Country fact sheets* [online]. Available at: https://ec.europa.eu/growth/sectors/construction/observatory/country-fact-sheets_en. [Accessed: 29 January 2021].
- [5] Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 120(3), 253–290.
- [6] Gaebert, T. & Staňková, M. (2020). Efficiency Development in the German Pharmaceutical Market. *Acta Universitatis agriculturae et silviculturae Mendelianae Brunensis*, 68(5), 877–884.
- [7] Galitskova, Y. & Mikhasek, A. (2017). Efficiency of construction waste recycling. *MATEC Web of Conferences*, 117(47), 00055.
- [8] Charnes, A., Cooper, W. W. & Rhodes, E. (1978). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 2(6), 429–444.
- [9] Klepeis, N.E. et al. (2001). The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. *Journal of Exposure Analysis and Environmental Epidemiology*, 11(3), 231–252.
- [10] Langston, C. (2014). Construction efficiency: A tale of two developed countries. *Engineering, Construction & Architectural Management*, 21(3), 320–335.
- [11] Meng, X. (2012). The effect of relationship management on project performance in construction. *International Journal of Project Management*, 30(2), 188–198.
- [12] Staňková, M. (2020). Efficiency comparison and efficiency development of the metallurgical industry in the EU: Parametric and non-parametric approaches. *Acta Universitatis agriculturae et silviculturae Mendelianae Brunensis*, 68(4), 765–774.
- [13] Staňková, M. & Hampel, D. (2020). Efficiency Assessment of the UK Travel Agency Companies – Data Envelopment Analysis Approach. *Mathematical Methods in Economics 2020: Conference Proceedings*. Brno: Mendelova univerzita v Brně, 550–556.
- [14] 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.
- [15] 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.

The path-relinking based search in unit lattice of m -dimensional simplex

Marek Kvet¹, Jaroslav Janáček²

Abstract. Path-relinking based searches proved to be very powerful optimization tools, when applied to the problems, solutions of which can be represented as a subset of m -dimensional unit hypercube vertices. The efficiency of the searching strategies can be considerably improved, if the starting set of solutions is uniformly deployed over the set of feasible hypercube vertices. The corresponding p -location problems solvable by the original path-relinking based methods assume that exactly p locations must be chosen from a set of m possible places. This way, each solution can be described by a zero-one m -dimensional vector, which coincides to a hypercube vertex. If the p -location problem is generalized so that its formulation admits to locate more than one facility at one place, then the set of feasible p -facility location solutions extends from the subset of hypercube vertices to a subset of integer point belonging to a facet of the simplex. Within this contribution, we suggest a path-relinking method, which is able to cope with the generalized set of feasible solutions of the p -facility location problem with multiple locating. Furthermore, we provide the generalized path-relinking search with an extension of a uniformly deployed set to obtain a starting set suitable for a search in a simplex facet. The appended computational study proves that the generalized path-relinking based search can reach a near-to-optimal solution of the real-world location problem.

Keywords: location problems, multiple facility location, path-relinking method, generalized set of solutions

JEL Classification: C44

AMS Classification: 90C06, 90C10, 90C27

1 Introduction

The recent model of the Emergency Medical Service (EMS) system design problem reflects two substantial generalizations. The first of them takes into account the fact that demands for service emerge randomly and the ability of the service centers are limited. Consequently, the generalized approach has to model the cases, when a current demand cannot be served from the geographically closest service center, but from the nearest available one. This generalization is based on introducing a series q_1, q_2, \dots, q_r of probability values, where q_k is the probability of the case, in which the k -th nearest center to a demand location j is the first available one, because the nearer centers are busy with servicing previously occurred demands [6, 7, 10, 13].

The second generalization of originally binary mathematical programming model is connected with the usually used data model of a serviced region. The serviced region is described by a road network graph with finite set of nodes, at which demands for service occur and the service centers are located at. The original binary models enable only to choose a given network node for a center location or leave it unused [1, 3, 11, 12]. Here, it is not possible to equip a service center with more than one facility, such as an ambulance vehicle, without having to create a more complex model with an enormous number of allocation variables and capacity constraints. The first generalization of the emergency system model enabled to perform the second model generalization, in which more than one facility can be assigned to the same network node in order to locate a service center [5].

Obviously, the more complex the associated mathematical models are, the more difficult is the challenge to find the optimal solution of the problem. Common exact methods [1, 3] usually suffer from unpredictable time demands, because a big part of total computational time is spent by verifying the optimality of the best-found solution. Therefore, emphasis is currently being placed on the development of effective heuristic and metaheuristic methods [2, 4, 14, 15]. Among all the directions that research in this area takes, we will limit ourselves only to

¹ University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovakia, marek.kvet@fri.uniza.sk

² University of Žilina, Faculty of Management Science and Informatics, Univerzitná 8215/1, 010 26 Žilina, Slovakia, jaroslav.janacek@fri.uniza.sk

the algorithms based on processing a population of feasible solutions. In this contribution, we suggest a path-relinking method, which is able to cope with the generalized set of feasible solutions of the p -facility location problem with multiple center locating. Our suggestions are experimentally verified using real data obtained from existing EMS system operating on the road network of Slovakia.

2 Generalized Emergency System Design Model

The emergency system design problem can be formulated as a task to deploy p facilities in a set of network nodes to minimize average response time of the system to the users' demands. It is assumed that the demands can emerge only at nodes $1, \dots, n$ of the network and the frequency of the randomly occurring demands at the node j is estimated by a constant b_j . Let a set I of m network nodes be specified as a set of possible service center locations. The network time-distance necessary for traversing the network from possible center location $i \in I$ to a demand location j will be referred as d_{ij} . It is also assumed that the probability values q_1, \dots, q_r are at disposal. The substantial decisions on the facility deployment will be modelled by a vector $\mathbf{y} \in \{Z^+\}^m$ of nonnegative integer components, where y_i for $i = 1, \dots, m$ gives the number of facilities located at the possible service center location i . After these preliminaries, the following expressions (1) and (2) can formulate the combinatorial model of the generalized EMS system design problem.

$$\text{Minimize } f(\mathbf{y}) = \sum_{j=1}^n b_j \sum_{k=1}^r q_k d_{e(\mathbf{y}, k, j), j} \quad (1)$$

$$\text{Subject to: } \mathbf{y} \in \mathbf{Y} = \left\{ \mathbf{x} \in \{0, \dots, p\}^m, \sum_{i=1}^m x_i = p \right\} \quad (2)$$

The mapping $e: \{0, \dots, p\}^m \times \{1, \dots, r\} \times \{1, \dots, n\} \rightarrow \{1, \dots, m\}$ is defined according to the following course of actions.

$e(\mathbf{y}, k, j)$

0. Order subscripts $i \in K = \{i=1, \dots, m: y(i) \geq 1\}$ into a sequence $i(1), \dots, i(|K|)$ according to increasing distances of the node i from the node j , i.e. $d_{i(1), j} \leq d_{i(2), j} \leq \dots \leq d_{i(|K|), j}$.
1. Define $e(\mathbf{y}, k, j) = i(t)$ so that the following inequalities hold for t : $y(i(1)) + \dots + y(i(t-1)) < k \leq y(i(1)) + \dots + y(i(t))$.

We have to point out the substantial differences between the set of all feasible solutions of the generalized problem (1)-(2) and the standard binary formulation (1) and (3), where the set of all feasible solutions \mathbf{Y} is described by (3).

$$\mathbf{Y} = \left\{ \mathbf{x} \in \{0, 1\}^m, \sum_{i=1}^m x_i = p \right\} \quad (3)$$

When comparing two sets of solutions, it can be found that the set (3) corresponds to a sub-set of vertices of an m -dimensional unit hypercube. In addition, the vertices also belong to a facet of m -dimensional simplex, which is determined by the condition that exactly p components of the vector associated with the solution equal to p . The set of integer points from the intersection of the m -dimensional hypercube and the simplex facet has special characteristics, which enabled us to suggest efficient heuristics, which were based on a special kind of the path-relinking method [4, 7, 9, 14].

Contrary to the set (3), the set (2) contains all integer points of m -dimensional space, which belong to the simplex facet and, of course, to the non-negative subspace. Usage of the above mentioned heuristics for the search in the integer lattice of the simplex facet demands for a substantial adjustment of the original path-relinking method. Then, it is questionable, whether the adjusted method will perform as well as the original one. The remainder of this contribution is focused on answering the question.

3 Generalization of the Path-Relinking Method for Integer Programming Problem

The original version of the path-relinking method for p -location problems was defined in m -dimensional unit hypercube of input solutions. The neighborhood members could be reached from the current solution by a simple exchange of one center location of the current solution with another location, which does not belong to the current solution.

The newly suggested version of the path-relinking method will also inspect solutions, Manhattan distance of which from a current solution equals to two. Nevertheless, the shortest path between the input solutions generally does not lie in the surface of m -dimensional hypercube. It consists of nodes of integer lattice nested in a facet of an m -dimensional simplex. While a move along the shortest path in the original path-relinking method corresponds with a simple exchange of positions of located centers, the move in the integer lattice of a simplex facet is a more complex operation. Contrary to the simple exchange of locations, the move can be performed either by the location swap or by withdrawing a facility from a center and relocating the facility at a possible center location, where another facility has been located. This more complex operation results in the modification of the path-relinking method that led to the *PathRelinkingGP* procedure.

The input solutions for this procedure are presented by sets L^1 and L^2 of service center locations and by mappings $y^1: L^1 \rightarrow \{1, \dots, p\}$ and $y^2: L^2 \rightarrow \{1, \dots, p\}$, which give the number of facilities located at individual centers of the considered solution. The procedure is described below.

PathRelinkingGP(L^1, y^1, L^2, y^2)

0. Initialize L^*, y^* by $\operatorname{argmin}\{f(L^1, y^1), f(L^2, y^2)\}$ and initialize the following sequence of location subsets: $L^- = \{i \in L^1 \cap L^2: y^1(i) = y^2(i)\}$, $L^+ = \{i \in L^1 \cap L^2: y^1(i) < y^2(i)\}$, $L = \{i \in L^1 \cap L^2: y^1(i) > y^2(i)\}$, $L^n = L^1 - L^2$ and $L^e = L^2 - L^1$.
1. If $M(L^1, y^1, L^2, y^2) > 2$ perform step 2, else return L^*, y^* and stop the run of procedure.
2. Find $i^*, j^* = \operatorname{argmin}\{f(\operatorname{Relocation}(L^1, y^1, i, j)): i \in L^- \cup L^n, j \in L^+ \cup L^e\}$ and update $L^1, y^1 = \operatorname{Relocation}(L^1, y^1, i^*, j^*)$. Update L^*, y^* by $\operatorname{argmin}\{f(L^1, y^1), f(L^*, y^*)\}$. Swap L^1, y^1 with L^2, y^2 . Redefine $L^- = \{i \in L^1 \cap L^2: y^1(i) = y^2(i)\}$, $L^+ = \{i \in L^1 \cap L^2: y^1(i) < y^2(i)\}$, $L = \{i \in L^1 \cap L^2: y^1(i) > y^2(i)\}$, $L^n = L^1 - L^2$ and $L^e = L^2 - L^1$. Go to step 1.

Comments: Function $\operatorname{Relocation}(L, y, i, j)$ returns solution L, y changed according to the following rule.

$\operatorname{Relocation}(L, y, i, j)$

$y(i) = y(i) - 1$; if $y(i) = 0$, then $L = L - \{i\}$.

If $j \in L^e$, then $L = L \cup \{j\}$ and $y(j) = 1$, else $y(j) = y(j) + 1$.

Return L, y .

The symbol $M(L^1, y^1, L^2, y^2)$ stands for Manhattan distance of solutions L^1, y^1 and L^2, y^2 computed by (4).

$$M(L^1, y^1, L^2, y^2) = \sum_{i \in L^-} (y^2(i) - y^1(i)) + \sum_{i \in L^+} (y^1(i) - y^2(i)) + \sum_{i \in L} y^1(i) + \sum_{i \in L} y^2(i) \quad (4)$$

It must be noted, that the complexity of obtaining i^*, j^* in the step 2 is $|L^- \cup L^n| \times |L^+ \cup L^e|$, while the complexity of associated step of the original path-relinking method is only $|L^n| \times |L^e|$. This may influence efficiency of the approaches based on usage of path-relinking method, when applied to the p -location problems with multiple facility location.

4 One-To-All Strategy with Extension

One-to-all strategy proved to be very efficient way of search, when used for original emergency service system design modelled by means of the binary linear programming [8]. This original approach starts with so-called uniformly deployed set of feasible solutions, which was considered an initial swarm of particles. The solution with minimal objective function value was used as a leader of the particle swarm optimization process. Contrary to the other particle swarm optimization strategies, positions of particles were not be changed with exception of the leader's position, which was updated after each inspection of the shortest path connecting the current leader position to a particle position. The inspection of the path was performed by the original path-relinking method adjusted for the problem (1), (3). A direct application of the above strategy together with usage of the uniformly

deployed set to the p -location problem with multiple facility location is not possible due to the inborn property of the path-relinking method. The method is unable to inspect any solution lying outside the sub-space restricted by the different components of the input solutions.

To employ benefit of uniformly deployed sets, an extension of them has been suggested to enable the search in the neighborhood of a promising solution outside the m -dimensional unit hypercube. We came from the obvious preposition that the model (1), (2) enables to locate at most r facilities to a one service center location. Furthermore, we did a heuristic assumption that the most important possible service center locations are those, which have the biggest population. Then, the extension algorithm, which can adjust solutions of a uniformly deployed set for the one-to-all strategy solving the p -location problem with multiple facility location, can be constituted in the following way.

Input of the algorithm is represented by parameter $noPref$ and a uniformly deployed set $S = \{P^1, P^2, \dots, P^{|S} \}$ of p -tuples of service center locations. Each p -tuple $P^i \in S$ corresponds to one feasible solution of the problem (1), (3), i.e. it is a vertex of an m -dimensional unit hypercube. The denotation of “uniformly deployed” means that Hamming distance between each pair of vertices of S is bigger than a given threshold, the value of which is usually near to the value $2p$. The output of the algorithm is a set E of feasible solutions W^i, y^i of the problem (1), (2), where some of the solutions lie outside the unit hypercube.

Extension(S, noPref)

0. Compute $noO(j)$ for each possible center location $j=1, \dots, m$, where $noO(j)$ is the number of occurrences of j in elements of S . Determine list $LPref$ of $noPref$ maximally populated possible center locations.
1. Process p -tuples $P^1, P^2, \dots, P^{|S} \}$ according to step 2 and, after having processed all P^i , go to step 3.
2. Find location $k \in P^i$, which belongs to $LPref$. If there is no such k , then define $W^i = P^i$ and $y^i(j) = 1$ for each $j \in W^i$. Otherwise, initialize $W^i = \{k\}$ and $y^i(k) = 1$ and perform the following decisions for each other element $j \in P^i - \{k\}$: If $y^i(k) < r$ and $noO(j) > 1$, then set $y^i(k) = y^i(k) + 1$ and $noO(j) = noO(j) - 1$, otherwise set $W^i = W^i \cup \{j\}$ and $y^i(j) = 1$.
3. Return $E = \{W^1, y^1, \dots, W^{|S} \}, y^{|S} \}$.

5 Computational Study

The further reported numerical experiments were suggested to prove or disprove the hypothesis that the path-relinking method adjusted for processing solutions of the p -facility location problem with multiple facility location is able to reach as excellent results as the algorithms destined for problems defined only on a unit hypercube in m -dimensional space.

Used benchmarks were derived from real emergency health care system, which operates in eight regions of Slovak Republic. For each self-governing region, i.e. Bratislava (BA), Banská Bystrica (BB), Košice (KE), Nitra (NR), Prešov (PO), Trenčín (TN), Trnava (TT) and Žilina (ZA), all cities and villages with corresponding number of inhabitants b_j were taken into account. The coefficients b_j were rounded to hundreds. The set of communities represents both the set J of users' locations and the set I of possible center locations as well. The coefficients q_k for $k=1 \dots 3$ were set at the values: $q_1 = 0,77063$, $q_2 = 0,16476$ and $q_3 = 1 - q_1 - q_2$. These values were obtained from a simulation model of existing emergency medical system in Slovakia [10]. The optimal solutions of all studied problem instances are available and the associated objective function values can be found in [5]. We report them also in the column of Table 1 denoted by $OptObjF$. Table 1 contains also the problem sizes expressed by the values of m and p respectively.

Table 1 Basic benchmarks characteristics and the optimal objective function values

Region	m	p	$OptObjF$
BA	87	25	18450
BB	515	46	38008
KE	460	38	40711
NR	350	36	40987
PO	664	44	46884
TN	276	26	31260
TT	249	22	36401
ZA	315	36	36929

Since the suggested path-relinking method processes a set of solutions, the uniformly deployed set can serve as a source of necessary data for the algorithm. The process of uniformly deployed set construction and usage are

reported in [8] and [9]. The common property of a uniformly deployed set consists in the fact that an arbitrary permutation of the locations generates a new uniformly deployed set with the same characteristics. We used this property to obtain ten different starting sets for each self-governing region. Therefore, we report the average results of ten runs.

For completeness, the numerical experiments reported in this paper were performed on a notebook equipped with the Intel® Core™ i7 3610QM 2.3 GHz processor and 8 GB of memory. The presented algorithms were implemented in the Java language making use of the NetBeans IDE 8.2 environment.

Let us now focus on performed numerical experiments and on discussing the obtained results. As the usage of the generalized path-relinking method was conditioned by an extension of the input uniformly deployed set, a special attention was devoted to the setting of the parameter *noPref*. This number of the most populated service centers was determined proportionally to the number *m* of the possible service center locations. We report results obtained by performing three series of experiments for various percentages of *m*, which were used for *noPref* determination. The percentages were 1.5%, 10% and 20% subsequently. The results are summarized in the following Table 2, Table 3 and Table 4, which have the same structure. Each row of the table corresponds to one problem instance. The left part of each table contains the optimal objective function value *OptObjF* taken from [5]. This value is reported to make the path-relinking method results quality evaluation more convenient. The results obtained by suggested heuristic approach are reported by the following five values: The column denoted by *ObjF* contains the objective function value of the resulting solution, which was obtained in the computational time in seconds denoted by *CT*. The symbol *noY* denotes the average resulting number of service centers. Finally, *Count_2* and *Count_3* denote the average number of centers, where two or three facilities were located.

Table 2 Results of numerical experiments for *noPref* = 1% of *m*

Region	<i>OptObjF</i>	<i>ObjF</i>	<i>CT</i>	<i>noY</i>	<i>Count_2</i>	<i>Count_3</i>
BA	18450	18752	0.68	24.70	0.30	0.00
BB	38008	38106	52.54	43.40	1.60	0.50
KE	40711	40734	28.80	37.00	1.00	0.00
NR	40987	41061	10.70	34.00	2.00	0.00
PO	46884	47012	65.80	41.10	2.90	0.00
TN	31260	31545	6.31	25.00	1.00	0.00
TT	36401	36780	3.67	21.20	0.80	0.00
ZA	36929	37030	8.78	35.00	1.00	0.00

Table 3 Results of numerical experiments for *noPref* = 10% of *m*

Region	<i>OptObjF</i>	<i>ObjF</i>	<i>CT</i>	<i>noY</i>	<i>Count_2</i>	<i>Count_3</i>
BA	18450	18717	0.69	24.60	0.40	0.00
BB	38008	38193	53.64	44.40	1.40	0.10
KE	40711	40761	29.22	37.30	0.70	0.00
NR	40987	41109	10.78	34.20	1.80	0.00
PO	46884	47132	66.93	42.00	2.00	0.00
TN	31260	31550	6.31	25.00	1.00	0.00
TT	36401	36801	3.66	21.30	0.70	0.00
ZA	36929	37070	8.82	35.20	0.80	0.00

Table 4 Results of numerical experiments for *noPref* = 20% of *m*

Region	<i>OptObjF</i>	<i>ObjF</i>	<i>CT</i>	<i>noY</i>	<i>Count_2</i>	<i>Count_3</i>
BA	18450	18717	0.70	24.60	0.40	0.00
BB	38008	38255	53.96	45.20	0.80	0.00
KE	40711	40809	29.12	37.70	0.30	0.00
NR	40987	41351	10.95	35.20	0.80	0.00
PO	46884	47230	67.53	42.80	1.20	0.00
TN	31260	31556	6.33	25.40	0.60	0.00
TT	36401	36807	3.66	21.40	0.60	0.00
ZA	36929	37133	8.84	35.60	0.40	0.00

6 Conclusions

The main research topic of this paper was focused on extending the path-relinking method to be able to comply with the p -location problem with multiple facility location. Mentioned algorithm adjustment could change its previous characteristics. Therefore, the performance of the algorithm was studied and the obtained results were analyzed from the point of solution accuracy. Furthermore, we provided the readers with the generalized path-relinking search with an extension of a uniformly deployed set to obtain a starting set suitable for a search in a simplex facet. Since the performance and resulting solutions may be affected by algorithm parameter, we performed a case study to study mentioned possible impact.

The obtained results reported in a separate section showed that the generalized path-relinking based search can reach a near-to-optimal solution of the real-world location problems in a acceptably short computational time. Thus, we can conclude that suggested heuristic method combined with the usage of uniformly deployed sets of solutions brings excellent results and may be practically used by the operations researchers and other professionals responsible for decision-making.

Acknowledgements

This work was supported by the research grants VEGA 1/0089/19 “Data analysis methods and decisions support tools for service systems supporting electric vehicles”, VEGA 1/0689/19 “Optimal design and economically efficient charging infrastructure deployment for electric buses in public transportation of smart cities”, and VEGA 1/0216/21 “Design of emergency systems with conflicting criteria using artificial intelligence tools”. This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-19-0441.

References

- [1] Avella, P., Sassano, A. & Vasil'ev, I. (2007). Computational study of large scale p -median problems. *Mathematical Programming*, 109, pp. 89-114.
- [2] 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.
- [3] 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.
- [4] Gendreau, M. & Potvin, J. (2010). *Handbook of Metaheuristics*, Springer Science & Business Media.
- [5] Janáček, J. (2021). Multiple p -Facility Location Problem with Randomly Emerging Demands. In: Strategic Management and its Support by Information Systems 2021, Technical University of Ostrava, in print
- [6] Janáček, J. & Kvet, M. (2021). Efficient incrementing Heuristics for Generalized p -Location Problems. In: Central European Journal of Operations Research, Springer, in print
- [7] Janáček, J. & Kvet, M. (2020). Discrete self-organizing migration algorithm. In: Croatian Operational Research Review 11(2), pp. 241-248.
- [8] Janáček, J. & Kvet, M. (2020). Uniform Deployment of the p -Location Problem Solutions. In: Operations Research Proceedings 2019: Selected Papers of the Annual International Conference of the German Operations Research Society, Dresden, Germany, September 4-6, 2019: Springer, pp. 315-321.
- [9] 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.
- [10] 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.
- [11] Jánošíková, E. & 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.
- [12] Karatas, M. & Yakıca, 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.
- [13] Kvet, M. (2014). Computational Study of Radial Approach to Public Service System Design with Generalized Utility. In *Digital Technologies 2014*, pp. 198-208.
- [14] Kvet, M. & Janáček, J. (2020). Spider network search strategy for p -location problems. In: CINTI 2020: 20th International Symposium on Computational Intelligence and Informatics, Budapešť, 2020, pp. 49-54.
- [15] Rybičková A., Mocková D. & Teichmann D. (2019). Genetic Algorithm for the Continuous Location-Routing Problem, *Neural Network World* 29(3), pp. 173–187.

Portfolio discount factor evaluated by oriented fuzzy numbers

Anna Łyczkowska-Hanćkowiak¹, Krzysztof Piasecki²

Abstract. In financial portfolio management, utilizing oriented fuzzy numbers is more useful than utilizing fuzzy numbers. Moreover, a portfolio analysis based on fuzzy discount factor is simpler than portfolio analysis based on return rate. For this reason, we consider here a discount factor evaluated by oriented fuzzy number. The main goal of our paper is to find an analytical formula describing portfolio expected discount factor as a function of expected discounts factors of portfolio components. In our considerations, we take into account the fact that addition of oriented fuzzy numbers is not associative. Therefore, we propose to calculate separately the weighted sum of positively oriented discount factors and the sum of negatively oriented discount factors. Then the portfolio discount factor is obtained by weighted addition of these sums. Such a procedure for determining the discount factor of the portfolio is justified by economic premises.

Keywords: discount factor, portfolio, oriented fuzzy number

JEL Classification : C44, G11

AMS Classification : 03E72

1. Introduction

Imprecision of financial data is usually modelled by fuzzy numbers (FNs). In [9] it is demonstrated that for a portfolio analysis, oriented FNs are more useful than FNs. Moreover, in the case of financial data imprecision the expected discount factor is a more convenient portfolio analysis tool than the expected return rate [10, 11]. For these reasons, the main aim of our paper is to study portfolio discount factor for the case when portfolio assets are evaluated by trapezoidal oriented FNs.

2. Trapezoidal oriented fuzzy numbers - basic facts

The symbol $\mathcal{F}(\mathbb{R})$ denotes the family of all fuzzy subsets in the real line \mathbb{R} . Fuzzy number (FN) is usually defined as a fuzzy subset of the real line \mathbb{R} . The most general definition of FN was formulated by Dubois and Prade [1]. The set of all FN we denote by the symbol \mathbb{F} . Any FN may be represented in following way

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 $\mu_{\mathcal{L}}(\cdot | a, b, c, d, L_L, R_L) \in [0, 1]^{\mathbb{R}}$ described by the identity

$$\mu_{\mathcal{L}}(x | a, b, c, d, L_L, R_L) = \begin{cases} 0, & x \notin [a, d], \\ L_L(x), & x \in [a, b[, \\ 1, & x \in [b, c], \\ R_L(x), & x \in]c, d], \end{cases} \quad (1)$$

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:

$$\forall_{x \in]a, d[} : \mu_{\mathcal{L}}(x | a, b, c, d, L_L, R_L) > 0. \quad (2)$$

The notion of ordered FN is introduced by Kosiński et al [3, 4]. From formal reasons, the Kosiński's theory is revised in [7]. In revised theory, the notion of ordered FN is narrowed down to the notion of oriented FN (OFN) defined as follows:

¹ WSB University in Poznań, Institute of Economy and Finance, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, Poland; anna.lyczkowska-hanckowiak@wsb.poznan.pl

² WSB University in Poznań, Institute of Economy and Finance, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, Poland; krzysztof.piasecki@wsb.poznan.pl

Definition 1 [7]. For any monotonic sequence $(a, b, c, d) \subset \mathbb{R}$, OFN $\vec{\mathcal{L}}(a, b, c, d, S_L, E_L) = \vec{\mathcal{L}}$ is the pair of orientation $\overline{a, d} = (a, d)$ and FN $\mathcal{L} \in \mathbb{F}$ described by membership function $\mu_{\mathcal{L}}(\cdot | a, b, c, d, S_L, E_L) \in [0, 1]^{\mathbb{R}}$ given by the identity

$$\mu_{\mathcal{L}}(x | a, b, c, d, S_L, E_L) = \begin{cases} 0, & x \notin [a, d] \equiv [d, a], \\ S_L(x), & x \in [a, b[\equiv]b, a], \\ 1, & x \in [b, c] \equiv [c, b], \\ E_L(x), & x \in]c, d] \equiv [d, c[, \end{cases} \quad (3)$$

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 (2)

Remark: The identity (3) additionally describes such modified notation of intervals which is used in the OFN theory. The notation $J \equiv \mathcal{K}$ means that “the interval J may be equivalently replaced by the interval \mathcal{K} ”.

The relationships between FNs, ordered FNs, and OFNs are discussed in detail in [9].

The symbol \mathbb{K} denotes the space of all OFNs. If $a < d$ then OFN $\vec{\mathcal{L}}(a, b, c, d, S_L, E_L)$ has the positive orientation $\overline{a, d}$ which informs us about possibility of an increase in approximated number. If $a > d$, then OFN $\vec{\mathcal{L}}(a, b, c, d, S_L, E_L)$ has the negative orientation $\overline{a, d}$ which informs us about possibility of a decrease in approximated number. If $a = d$, then OFN $\vec{\mathcal{L}}(a, a, a, a, S_L, E_L) = \llbracket a \rrbracket$ describes the real number $a \in \mathbb{R}$.

A special case of OFNs are trapezoidal fuzzy numbers (TrOFNs).

Definition 2. [6] For any monotonic sequence $(a, b, c, d) \subset \mathbb{R}$, TrOFN $\vec{\mathcal{T}r}(a, b, c, d) = \vec{\mathcal{T}r}$ is OFN $\vec{\mathcal{T}r} \in \mathbb{K}$ determined explicitly by its membership functions $\mu_{\mathcal{T}r} \in [0, 1]^{\mathbb{R}}$ as follows

$$\mu_{\mathcal{T}r}(x) = \mu_{\mathcal{T}r}(x | a, b, c, d) = \begin{cases} 0, & x \notin [a, d] \equiv [d, a], \\ \frac{x-a}{b-a}, & x \in [a, b[\equiv]a, b], \\ 1, & x \in [b, c] \equiv [c, b], \\ \frac{x-d}{c-d}, & x \in]c, d] \equiv [c, d]. \end{cases} \quad (4)$$

The symbol $\mathbb{K}_{\mathcal{T}r}$ denotes the space of all TrOFNs. The space of all positively oriented TrOFNs is denoted by the symbol $\mathbb{K}_{\mathcal{T}r}^+$. The space of all negatively oriented TrOFNs we denote by the symbol $\mathbb{K}_{\mathcal{T}r}^-$.

Let symbol $*$ denotes any arithmetic operation defined in \mathbb{R} . By symbol $\boxed{*}$ we denote an extension of arithmetic operation $*$ to \mathbb{K} . Kosiński has defined arithmetic operators on ordered FNs in an intuitive way. The addition and dot product extended to the space \mathbb{K} have a very high level of formal complexity [9]. Therefore, in many applications researchers limit their calculations to arithmetic operations determined on the space $\mathbb{K}_{\mathcal{T}r}$.

In line with the Kosinski’s approach, we can extend basic arithmetic operators to the case of $\mathbb{K}_{\mathcal{T}r}$ in such way that for any pair $(\vec{\mathcal{T}r}(a, b, c, d), \vec{\mathcal{T}r}(p - a, q - b, r - c, s - d)) \in \mathbb{K}_{\mathcal{T}r}^2$ and $\beta \in \mathbb{R}$, arithmetic operations of extended sum \boxplus and dot product \boxtimes are defined as follows [6]:

$$\begin{aligned} \vec{\mathcal{T}r}(a, b, c, d) \boxplus \vec{\mathcal{T}r}(p - a, q - b, r - c, s - d) &= \\ &= \begin{cases} \vec{\mathcal{T}r}(\min\{p, q\}, q, r, \max\{r, s\}), & (q < r) \vee (q = r \wedge p \leq s), \\ \vec{\mathcal{T}r}(\max\{p, q\}, q, r, \min\{r, s\}), & (q > r) \vee (q = r \wedge p > s). \end{cases} \end{aligned} \quad (5)$$

$$\beta \boxtimes \vec{\mathcal{T}r}(a, b, c, d) = \vec{\mathcal{T}r}(\beta \cdot a, \beta \cdot b, \beta \cdot c, \beta \cdot d). \quad (6)$$

In general, the TrOFNs addition is not associative [9]. Moreover, for any pair $(\vec{\mathcal{T}r}(a, b, c, d), \vec{\mathcal{T}r}(e, f, g, h)) \in (\mathbb{K}_{\mathcal{T}r}^+ \cup \mathbb{R})^2 \cup (\mathbb{K}_{\mathcal{T}r}^- \cup \mathbb{R})^2$ we have [9]

$$\vec{\mathcal{T}r}(a, b, c, d) \boxplus \vec{\mathcal{T}r}(e, f, g, h) = \vec{\mathcal{T}r}(a + e, b + f, c + g, d + h). \quad (7)$$

Any monotonic unary operator $G: \mathbb{R} \supset \mathbb{A} \rightarrow \mathbb{R}$ may be extended to TrOFN case in the following way. Using the Kosiński’s approach, we define an extended unary operator $\vec{G}: \mathbb{K}_{\mathcal{T}r} \supset \mathbb{H} \rightarrow \mathbb{K}$ as follows:

$$\vec{\mathcal{L}}(G(a), G(b), G(c), G(d), S_L, E_L) = \vec{G}(\vec{\mathcal{T}r}(a, b, c, d)), \quad (8)$$

where the starting function and the ending function are given by formulas

$$\forall y \in [G(a), G(b)[\quad S_L(y) = \frac{G^{-1}(y) - a}{b - a}, \quad (9)$$

$$\forall y \in]G(c), G(d)] \quad E_L(y) = \frac{G^{-1}(y) - d}{c - d}. \quad (10)$$

3. Expected discount factor

Let us assume that the time horizon $t > 0$ of an investment is fixed. Then, the asset considered here is determined by two values:

- Anticipated future value (FV) V_t ,
- Assessed present value (PV) V_0 .

The basic characteristic of benefits from owning this asset is a simple return rate r_t given by the identity:

$$r_t = \frac{V_t - V_0}{V_0} = \frac{V_t}{V_0} - 1. \quad (11)$$

In [11], it is justified that FV is a random variable $\tilde{V}_t: \Omega \rightarrow \mathbb{R}^+$. The set, Ω , is a set of elementary states, ω , of the financial market. In a classical approach to a return rate estimation, PV is identified with the observed quoted price \tilde{P} . Thus, the return rate is a random variable determined by identity:

$$r_t(\omega) = \frac{\tilde{V}_t(\omega) - \tilde{P}}{\tilde{P}}. \quad (12)$$

Uncertainty risk is a result of a lack of knowledge about the future state of affairs. In practice of financial markets analysis, the uncertainty risk is usually described by the probability distribution of return rate (12) which may be given by a cumulative distribution function $F_r(\cdot | \bar{r}): \mathbb{R} \rightarrow [0, 1]$. We assume that the expected value, \bar{r} , of this distribution exists. Then also the expected discount factor (EDF) \bar{v} exists and is determined by the dependency

$$\bar{v} = (1 + \bar{r})^{-1}. \quad (13)$$

If we take together (11) and (12), then we obtain a following formula describing the return rate

$$r_t = r_t(V_0, \omega) = \frac{\tilde{P} \cdot (1 + r_t(\omega))}{V_0} - 1. \quad (14)$$

It implies that the expected return rate may be expressed in a following way

$$\mathcal{R}(V_0) = \int_{-\infty}^{+\infty} \frac{\tilde{P} \cdot (1 + y)}{V_0} - 1 dF_r(y | \bar{r}) = \frac{\tilde{P} \cdot (1 + \bar{r})}{V_0} - 1. \quad (15)$$

Thanks to that we determine the imprecise EDF $\mathcal{V}: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ as a unary operator transforming PV as follows

$$\mathcal{V}(V_0) = \left(\frac{\tilde{P} \cdot (1 + \bar{r})}{V_0} \right)^{-1} = \frac{\bar{v}}{\tilde{P}} \cdot V_0. \quad (16)$$

If PV is imprecisely estimated then it may be evaluated by TrOFN

$$\overline{PV} = \overline{Tr}(V_s, V_f, V_l, V_e) \quad (17)$$

where the monotonic sequence $(V_s, V_f, \tilde{P}, V_l, V_e)$ is determined in following way

- $[V_s, V_e] \subset \mathbb{R}^+$ is an interval of all possible values of PV,
- $[V_f, V_l] \subset [V_s, V_e]$ is an interval of all prices which do not noticeably differ from a quoted price \tilde{P} .

If we predict a rise in price then PV is described by a positively oriented TrOFN. If we predict a fall in price, then PV is described by a negatively oriented TrOFN.

Then using the Kosinski's approach, we define the imprecise EDF by an extension $\tilde{\mathcal{V}}: \mathbb{K}_{Tr} \rightarrow \mathbb{K}$ of a unary operator (68). The identities (8), (9), (10), and (17) imply that the imprecise EDF $\tilde{\mathcal{V}}(\overline{PV})$ is given by TrOFN

$$\tilde{\mathcal{V}}(\overline{PV}) = \overline{Tr} \left(\frac{V_s \cdot \bar{v}}{\tilde{P}}, \frac{V_f \cdot \bar{v}}{\tilde{P}}, \frac{V_l \cdot \bar{v}}{\tilde{P}}, \frac{V_e \cdot \bar{v}}{\tilde{P}} \right) = \tilde{\mathcal{V}} \left(\overline{Tr}(V_s, V_f, V_l, V_e) \right). \quad (18)$$

In [8] it shown that analogous extension $\tilde{\mathcal{R}}(\overline{PV})$ is such OFN which is not TrOFN. In considered case it causes that for portfolio analysis, imprecise EDF is more useful than imprecise expected return rate.

4. Expected discount factors for portfolio

By a financial portfolio we will understand an arbitrary, finite set of assets. Any asset is determined as fixed security in long position. On the other hand, any portfolio also is a security. Let us consider the case of a multi-asset portfolio π^* , built of securities Y_i . We describe this portfolio as the set $\pi^* = \{Y_i: i = 1, 2, \dots, n\}$. Any security Y_i is characterized by its price $\tilde{P}_i \in \mathbb{R}^+$, by its imprecise PV evaluated by TrOFN

$$\overline{PV}_i = \overline{Tr}(V_s^{(i)}, V_f^{(i)}, V_l^{(i)}, V_e^{(i)}) \quad (19)$$

and by its EDF \bar{v}_i determined by (13). Taking into account all the above, we evaluate any security Y_i by its imprecise EDF

$$\tilde{\mathcal{V}}_i = \overline{Tr}(D_s^{(i)}, D_f^{(i)}, D_l^{(i)}, D_e^{(i)}) = \overline{Tr} \left(V_s^{(i)} \cdot \frac{\bar{v}_i}{\tilde{P}_i}, V_f^{(i)} \cdot \frac{\bar{v}_i}{\tilde{P}_i}, V_l^{(i)} \cdot \frac{\bar{v}_i}{\tilde{P}_i}, V_e^{(i)} \cdot \frac{\bar{v}_i}{\tilde{P}_i} \right). \quad (20)$$

A portfolio PV is always equal to the sum of its components' PVs. In the case where the components' PVs are estimated by TrOFNs, addition is not associative. Then multiple addition depends on the order of the summands. This implies that a portfolio's PV, given as any sum of its components' PVs, is not explicitly determined. Therefore, in the considered case, any method of calculating the portfolio PV should be supplemented with a reasonable method for the ordering of the portfolio components.

We will use a method of ordering the assets proposed and justified in [6]. At the outset, we distinguish the portfolio of rising securities $\pi^+ = \{Y_i \in \pi^*: \overline{PV}_i \in \mathbb{K}_{Tr}^+\}$ and the portfolio of falling securities $\pi^- = \pi^* \setminus \pi^+$. Then, using (7) we calculate the PV of portfolio π^+ , denoted by the symbol \overline{PV}^+ and the PV of portfolio π^- , denoted by the symbol \overline{PV}^- . We have

$$\overline{PV}^+ = \overline{Tr}(V_s^{(+)}, V_f^{(+)}, V_l^{(+)}, V_e^{(+)}) = \overline{Tr}(\sum_{Y_i \in \pi^+} V_s^{(i)}, \sum_{Y_i \in \pi^+} V_f^{(i)}, \sum_{Y_i \in \pi^+} V_l^{(i)}, \sum_{Y_i \in \pi^+} V_e^{(i)}), \quad (21)$$

$$\overline{PV}^- = \overline{Tr}(V_s^{(-)}, V_f^{(-)}, V_l^{(-)}, V_e^{(-)}) = \overline{Tr}(\sum_{Y_i \in \pi^-} V_s^{(i)}, \sum_{Y_i \in \pi^-} V_f^{(i)}, \sum_{Y_i \in \pi^-} V_l^{(i)}, \sum_{Y_i \in \pi^-} V_e^{(i)}). \quad (22)$$

Finally, we calculate the PV of portfolio π^* , denoted by the symbol \overline{PV}^* . We get

$$\overline{PV}^* = \overline{PV}^+ \boxplus \overline{PV}^- = \overline{Tr}(V_s^{(+)}, V_f^{(+)}, V_l^{(+)}, V_e^{(+)}) \boxplus \overline{Tr}(V_s^{(-)}, V_f^{(-)}, V_l^{(-)}, V_e^{(-)}) = \overline{Tr}(V_s^{(*)}, V_f^{(*)}, V_l^{(*)}, V_e^{(*)}). \quad (23)$$

Now we can start calculating EDFs of considered portfolios. The values of portfolios π^+, π^-, π^* are respectively calculated in following way

$$M^+ = \sum_{Y_i \in \pi^+} \check{P}_i, \quad M^- = \sum_{Y_i \in \pi^-} \check{P}_i, \quad M^* = M^+ + M^-. \quad (24)$$

The share p_i^+ of the asset $Y_i \in \pi^+$ in the portfolio π^+ and the share p_i^- of the asset $Y_i \in \pi^-$ in the portfolio π^- are given by formulas

$$p_i^+ = \frac{\check{P}_i}{M^+}, \quad p_i^- = \frac{\check{P}_i}{M^-}. \quad (25)$$

The share q^+ of portfolio π^+ in the portfolio π^* and the share q^- of portfolio π^- in the portfolio π^* are given by formulas

$$q^+ = \frac{M^+}{M^*}, \quad q^- = \frac{M^-}{M^*}. \quad (26)$$

The EDF \bar{v}^+ of portfolio π^+ , the EDF \bar{v}^- of portfolio π^- and the EDF \bar{v}^* of portfolio π^* are calculated as follows

$$\bar{v}^+ = \left(\sum_{Y_i \in \pi^+} \frac{p_i^+}{\bar{v}_i} \right)^{-1}, \quad \bar{v}^- = \left(\sum_{Y_i \in \pi^-} \frac{p_i^-}{\bar{v}_i} \right)^{-1}, \quad \bar{v}^* = \left(\frac{q^+}{\bar{v}^+} + \frac{q^-}{\bar{v}^-} \right)^{-1}. \quad (27)$$

Due results obtained in [8,9,10] and (7) we have that:

- the imprecise EDF \check{V}^+ of portfolio π^+ is given by the formula

$$\begin{aligned} \check{V}^+ &= \overline{Tr}(D_s^{(+)}, D_f^{(+)}, D_l^{(+)}, D_e^{(+)}) = \bar{v}^+ \boxminus \left(\boxplus_{Y_i \in \pi^+} \left(\frac{p_i^+}{\bar{v}_i} \boxminus \check{V}_i^+ \right) \right) = \\ &= \overline{Tr} \left(\sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^+}{\bar{v}_i} \cdot D_s^{(i)}, \sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^+}{\bar{v}_i} \cdot D_f^{(i)}, \sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^+}{\bar{v}_i} \cdot D_l^{(i)}, \sum_{Y_i \in \pi^+} \frac{\bar{v}^+ \cdot p_i^+}{\bar{v}_i} \cdot D_e^{(i)} \right), \end{aligned} \quad (28)$$

- the imprecise EDF \check{V}^- of portfolio π^- is given by the formula

$$\begin{aligned} \check{V}^- &= \overline{Tr}(D_s^{(-)}, D_f^{(-)}, D_l^{(-)}, D_e^{(-)}) = \bar{v}^- \boxminus \left(\boxplus_{Y_i \in \pi^-} \left(\frac{p_i^-}{\bar{v}_i} \boxminus \check{V}_i^- \right) \right) = \\ &= \overline{Tr} \left(\sum_{Y_i \in \pi^-} \frac{\bar{v}^- \cdot p_i^-}{\bar{v}_i} \cdot D_s^{(i)}, \sum_{Y_i \in \pi^-} \frac{\bar{v}^- \cdot p_i^-}{\bar{v}_i} \cdot D_f^{(i)}, \sum_{Y_i \in \pi^-} \frac{\bar{v}^- \cdot p_i^-}{\bar{v}_i} \cdot D_l^{(i)}, \sum_{Y_i \in \pi^-} \frac{\bar{v}^- \cdot p_i^-}{\bar{v}_i} \cdot D_e^{(i)} \right), \end{aligned} \quad (29)$$

- the imprecise EDF \check{V}^* of portfolio π^* is given by the formula

$$\check{V}^* = \overline{Tr}(D_s^{(*)}, D_f^{(*)}, D_l^{(*)}, D_e^{(*)}) = \left(\frac{\bar{v}^+ \cdot q^+}{\bar{v}^+} \boxminus \check{V}^+ \right) \boxplus \left(\frac{\bar{v}^- \cdot q^-}{\bar{v}^-} \boxminus \check{V}^- \right). \quad (30)$$

5. Case study

Our case study builds on data already discussed in a different context [5]. We observe the portfolio π^* composed of company shares included in WIG20 quoted on the Warsaw Stock Exchange (WSE). The portfolio π^* contains the following securities: 1 share Y_1 issued by the stock company ALR, 1 share Y_2 issued by the stock company CCC, 1 share Y_3 issued by stock company CDR, 1 share Y_4 issued by the stock company CPS, 1 share Y_5 issued by stock company DNP, 1 share Y_6 issued by the stock company LTS, and 1 share Y_7 issued by stock company MBK. Based on a session closing on the WSE on January 28, 2020, for each observed share we assess its PV equal to TrOFN \check{P}_i describing its Japanese candle [8]. Shares' PVs, obtained in such a manner, are presented in Table 1. For each portfolio component Y_i , we determine its quoted price \check{P}_i as an initial price on 29.01.2020. All considerations in the paper are run for the quarterly period of the investment time. Therefore, each security Y_i is tentatively characterized by its quarterly EDF \bar{v}_i of portfolio components. Using (30), for each security Y_i we calculate its imprecise EDF \check{V}^* . All these valuations are presented in Table 1.

The shares Y_2, Y_3, Y_4, Y_5 belong to portfolio π^+ of rising securities. The shares Y_1, Y_6, Y_7 belong to portfolio π^- of falling securities. Respectively using (21), (22), and (23) we calculate PVs of portfolios π^+, π^-, π^*

$$\begin{aligned} \overline{PV}^+ &= \overline{Tr}(536.27, 541.10, 546.82, 541.20), \\ \overline{PV}^- &= \overline{Tr}(478.30, 476.70, 467.96, 464.10), \\ \overline{PV}^* &= \overline{Tr}(1017.80, 1017.80, 1014.78, 1005.30). \end{aligned}$$

In the next step, using (27) we calculate EDFs of portfolios π^+, π^-, π^*

$$\bar{v}^+ = 0.866306, \quad \bar{v}^- = 0.93368, \quad \bar{v}^* = 0.896089.$$

Finally, respectively using (28), (29), and (30) we calculate imprecise EDFs of portfolios π^+, π^-, π^*

$$\check{V}^* = \overline{Tr}(0.848537, 0.856179, 0.865230, 0.870578),$$

$$\bar{V}^- = \bar{Tr}(0.955379, 0.952183, 0.934726, 0.927016),$$

$$\bar{V}^* = \bar{Tr}(0.898614, 0.898614, 0.895948, 0.895524).$$

Share	PV	Price	EDF	Imprecise EDF
Y_1	$\bar{Tr}(27.42; 27.30; 27.00; 26.84)$	27.00	0.9744	$\bar{Tr}(0.9896; 0.9852; 0.9744; 0.9686)$
Y_2	$\bar{Tr}(83.35; 88.00; 88.00; 89.65)$	88.00	0.9646	$\bar{Tr}(0.9136; 0.9646; 0.9646; 0.9827)$
Y_3	$\bar{Tr}(271.50; 271.50; 276.30; 276.30)$	277.00	0.8006	$\bar{Tr}(0.7847; 0.7847; 0.7986; 0.7986)$
Y_4	$\bar{Tr}(26.42; 26.60; 27.04; 27.34)$	27.20	0.9439	$\bar{Tr}(0.9168; 0.9231; 0.9384; 0.9488)$
Y_5	$\bar{Tr}(155.00; 155.00; 155.10; 157.30)$	155.30	0.9370	$\bar{Tr}(0.9352; 0.9352; 0.9358; 0.9491)$
Y_6	$\bar{Tr}(83.88; 83.40; 81.16; 80.26)$	81.44	0.9047	$\bar{Tr}(0.9318; 0.9265; 0.9016; 0.8916)$
Y_7	$\bar{Tr}(367.00; 366.00; 359.80; 357.00)$	359.00	0.9369	$\bar{Tr}(0.9578; 0.9552; 0.9390; 0.9317)$

Table 1. Evaluations of portfolio components

Comparing dependences (28) and (29) with dependency (30) raises the following question: whether there are constants α and β satisfying the conditions

$$\alpha \cdot 0.848537 + \beta \cdot 0.955379 = \alpha \cdot D_s^+ + \beta \cdot D_s^- = D_s^* = 0.898614, \quad (31)$$

$$\alpha \cdot 0.856179 + \beta \cdot 0.952183 = \alpha \cdot D_f^+ + \beta \cdot D_f^- = D_f^* = 0.898614, \quad (32)$$

$$\alpha \cdot 0.865230 + \beta \cdot 0.934726 = \alpha \cdot D_l^+ + \beta \cdot D_l^- = D_l^* = 0.895948, \quad (33)$$

$$\alpha \cdot 0.870578 + \beta \cdot 0.927016 = \alpha \cdot D_e^+ + \beta \cdot D_e^- = D_e^* = 0.895524. \quad (34)$$

Using (7) and (30), we get the unique solution $\alpha = 0.557987$ and $\beta = 0.442013$ of the system of equations (32) and (33). Then, by checking equation (31), we get

$$\alpha \cdot D_s^+ + \beta \cdot D_s^- = 0.557987 \cdot 0.848537 + 0.442013 \cdot 0.955379 = 0.895763 \neq 0.898614 = D_s^*.$$

This shows that linear portfolio analysis is not possible for considered here portfolio π^* . It is obvious, that this conclusion can be generalized for any portfolio π^* .

6. Final remarks

Obtained results may provide theoretical foundations for portfolio analysis of securities described with use TrOFNs. Then the criterion of maximization expected return rate is replaced by criterion on minimalization expected discount factor.

The proposed portfolio analysis can be fully used for portfolios π^+ of rising securities and π^- of falling securities. This is sufficient to manage portfolio risk, because only rising securities can get BUY or ACCUMULATE recommendations and only falling securities can get SELL or REDUCE recommendations. In considered case study, we show that the dependence (30) is not linear. Such form of the relationship (30) allows us to use them only for evaluating an already constructed portfolio π^* . Such evaluation may be carried out using the analytical tools described in [9].

All results obtained with use imprecise expected discount factor may be applied as input data for robo-advice systems described in [5].

The obtained results may as well be useful for a future research on the impact of the PV imprecision and orientation on portfolio analysis.

References

- [1] Dubois, D.; Prade, H. (1978) Operations on fuzzy numbers. *International Journal of System Science* 9, 613-629. <https://doi.org/10.1080/00207727808941724>
- [2] Delgado, M.; Vila, M. A.; Voxman, W. (1998) On a canonical representation of fuzzy numbers. *Fuzzy Sets and Systems* 93(1), 125-135. [https://doi.org/10.1016/S0165-0114\(96\)00144-3](https://doi.org/10.1016/S0165-0114(96)00144-3)
- [3] Kosiński, W.; Prokopowicz, P.; Ślęzak, D. (2002) Fuzzy numbers with algebraic operations: algorithmic approach. In *Proc. IIS'2002*; Kłopotek, M., Wierzchoń, S.T., Michalewicz, M., (Eds), Sopot, Poland, Physica Verlag, Heidelberg, 2002, pp. 311-320
- [4] Kosiński, W. (2006) On fuzzy number calculus. *Int. J. Appl. Math. Comput. Sci.*, 16(1), 51-57
- [5] Łyczkowska-Hanćkowiak A. (2020) On Application Oriented Fuzzy Numbers for Imprecise Investment

- Recommendations, *Symmetry* 12(10), <https://doi.org/10.3390/sym12101672>
- [6] Łyczkowska-Hanćkowiak, A. Piasecki, K. (2018) The Present Value of a Portfolio Of Assets With Present Values Determined by Trapezoidal Ordered Fuzzy Number, *Operations Research and Decisions* 28(2), 41-56, <https://doi.org/10.5277/ord180203>
- [7] Piasecki, K. (2018) Revision of the Kosiński's Theory of Ordered Fuzzy Numbers. *Axioms* 7(1), <https://doi.org/10.3390/axioms7010016>
- [8] Piasecki, K.; Łyczkowska-Hanćkowiak, A. (2019) Representation of Japanese Candlesticks by Oriented Fuzzy Numbers. *Econometrics* 8(1), 523. <https://doi.org/10.3390/econometrics8010001>
- [9] Piasecki, K.; Łyczkowska-Hanćkowiak, A. (2021) Oriented Fuzzy Numbers vs. Fuzzy Numbers. *Mathematics* 9(3), 523. <https://doi.org/10.3390/math9050523>
- [10] Piasecki K., Siwek J., (2018a), Two-Asset Portfolio with Triangular Fuzzy Present Values – An Alternative Approach, In T. Choudhry, J. Mizerka (Eds.) *Contemporary Trends in Accounting, Finance and Financial Institutions*, Springer Proceedings in Business and Economics. Springer, Cham, 11-26, https://doi.org/10.1007/978-3-319-72862-9_2
- [11] Piasecki K., Siwek J., (2018b), Multi-asset portfolio with trapezoidal fuzzy present values, *Przegląd Statystyczny/ Statistical Review* LXV(2),183-199.

Comparing TV advertisement in the year 2019 using DEA models

Jan Malý¹, Petra Zýková²

Abstract This paper focuses on the TV advertisement sector in the Czech Republic in the year 2019. The article aims to see how efficiency changes during the year and whether or not it is more suitable to choose less or more TRP for the campaign. This paper analyses the TV advertisement by the Data Envelopment Analysis (DEA). DEA models compute the relative efficiency of decision-making units, which transform multiple inputs into multiple outputs. The BCC input-oriented model (variable return to scale) is used for the analysis. This paper aims to identify optimal level of TRP for various target audiences across five channel mixes and compare some discovered trends with traditional views and metrics from advertisement.

Keywords: TV advertisement, DEA models, efficiency analysis

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

This paper compares commonly used parameters using DEA models during the year 2019 to show how viewership, parameters derived from it and its efficiency change throughout the year.

In the Czech TV, advertising environment consists of two main groups, Media Club and Nova Group, who mediate advertising time. Nova Group [7] is responsible for managing sales of Nova stations. Media Club [6] covers stations of Prima, Barrandov, Óčko and thematical stations under Atmedia. Stations under Česká Televize are not mentioned here because these are limited by [8], so their impact on the advertising environment is low. However, there is one big difference between Nova Group and Media Club. Nova group targets an audience A15 – A54 (both sexes), and Media Club targets an audience A15 – A69 (both sexes).

In TV advertising four main parameters are observed according to [1]: GRP (Gross Rating Point), TRP (Target Rating Point), Affinity, and Reach. GRP is the cumulative percentage of viewership in buying audience. TRP measures cumulative viewership percentage in the target audience. This paper used target audiences for men, women, and all in three age intervals of 18 - 35, 20 - 50 and 40 - 60. Affinity is the ratio of TRP and GRP. Affinity describes how the target audience (TA) watches TV compared to the population represented by GRP. Reach represents the percentage of TA who saw corporate advertising campaign (spot). There was used one input: GRP (Gross Rating Point), and two outputs: Affinity and Reach.

The paper is organised as follows. The following section presents the definition of DEA models generally. Section 3 contains the analysis of TV advertisement in the Czech Republic in 2019. The last section of the article concludes the results and discusses future research.

2 DEA models

DEA models have been first developed by Charnes, Cooper and Rhodes [1] based on the concept introduced by Farrell [4]. 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 applying DEA models. Let us denote $\mathbf{Y} = (y_{rj}, r = 1, \dots, t, j = 1, \dots, n)$ a non-negative matrix of outputs and $\mathbf{X} = (x_{kj}, k = 1, \dots, w, j = 1, \dots, n)$ a non-negative matrix of inputs. The efficiency score of the unit under evaluation DMU j_0 is derived as follows:

¹ Prague University of Economics and Business, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, Czech Republic, malj09@vse.cz

² Prague University of Economics and Business, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, Czech Republic, petra.zykova@vse.cz

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

$$\begin{aligned} \frac{\sum_{r=1}^t u_r y_{rj}}{\sum_{k=1}^w v_k x_{kj}} &\leq 1, \quad j = 1, \dots, n, \\ u_r &\geq \varepsilon, \quad r = 1, \dots, t, \\ v_k &\geq \varepsilon, \quad k = 1, \dots, w, \end{aligned} \quad (1)$$

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

$$U_{j_0} = \sum_{r=1}^t u_r y_{rj_0}$$

$$\sum_{k=1}^w v_k x_{kj_0} = 1,$$

subject to

$$\begin{aligned} \sum_{r=1}^t u_r y_{rj} - \sum_{k=1}^w v_k x_{kj} &\leq 0, \quad j = 1, \dots, n, \\ u_r &\geq \varepsilon, \quad r = 1, \dots, t, \\ v_k &\geq \varepsilon, \quad k = 1, \dots, w. \end{aligned} \quad (2)$$

The model (2) assumes a constant return to scale (CRS). There are other types of return to scale: variable return to scale (VRS), non-increasing return to scale (NIRS) and non-decreasing return to scale (NDRS). For this article is the proper variable return to scale. Thus free variable μ is added to the model (2). The input-oriented model with a variable return to scale (often called the BCC model) is as follows:

Maximise

$$U_{j_0} = \sum_{r=1}^t u_r y_{rj_0} + \mu$$

$$\sum_{k=1}^w v_k x_{kj_0} = 1,$$

subject to

$$\begin{aligned} \sum_{r=1}^t u_r y_{rj} - \sum_{k=1}^w v_k x_{kj} + \mu &\leq 0, \quad j = 1, \dots, n, \\ u_r &\geq \varepsilon, \quad r = 1, \dots, t, \\ v_k &\geq \varepsilon, \quad k = 1, \dots, w. \end{aligned} \quad (3)$$

The model (3) is a multiplicative BCC model. From the model (3) is derived the dual version of the model (3) called envelopment model, which is as follows:

Minimise

$$U_{j_0} = \theta_{j_0} - \varepsilon \left(\sum_{k=1}^w s_k^- - \sum_{r=1}^t s_r^+ \right)$$

$$\sum_{j=1}^n x_{kj} \lambda_j + s_k^- = \theta_{j_0} x_{kj_0}, \quad k = 1, \dots, w,$$

subject to

$$\sum_{j=1}^n y_{rj} \lambda_j - s_r^+ = y_{rj_0}, \quad r = 1, \dots, t, \quad (4)$$

$$\sum_{j=1}^n \lambda_j = 1$$

$$\lambda_j \geq 0, \quad j = 1, \dots, n,$$

$$s_k^- \geq 0, \quad k = 1, \dots, w,$$

$$s_r^+ \geq 0, \quad r = 1, \dots, t,$$

where $\lambda = (\lambda_1, \dots, \lambda_n)$, $\lambda \geq \mathbf{0}$ is a vector of weights assigned to particular DMUs, $\mathbf{s}^- = (s_1^-, \dots, s_w^-)$ and $\mathbf{s}^+ = (s_1^+, \dots, s_t^+)$ are vectors of slack/surplus variables. Efficient units identified by this model have efficiency scores equal to one, and all slack/surplus variables are equal to 0. Inefficient units have efficiency scores lower than 1. The model is not able to rank efficient units because of their identical efficiency scores. The model (4) is used in the calculation further in this paper. All mentioned models are described in [3].

3 TV advertisements

The dataset analysed in this paper was obtained by [5]. Data were collected for each month in the year 2019, and then they are divided into channel mixes based on their share of TRP between Media Club and Nova Group, as shown in Table 1.

Name of channel mix	Share of Nova Group (%)	Share of Media Club (%)
MC (Media Club)	(0, 3)	(100, 97)
NG LOW (Nova Group low)	(3, 35)	(97, 65)
BOTH (Nova Group & Media Club)	(35, 65)	(65, 35)
MC LOW (Media Club low)	(65, 97)	(35, 3)
NG (Nova Group)	(97, 100)	(3, 0)

Table 1 Definition of channel mixes

Dataset was further specified. There are several types of viewership: live viewership, viewership with deferred viewership up to 7-days, and also, viewership with or without guests. In this paper, the data is specified as live viewership + 3-days deferred viewership with guests. As GRP definition the market definition for both TV groups is used, like it is mentioned in chapter 1. TRP is used at TA for men, women, and all in three age intervals of 18-35, 20-50 and 40-60 with Reach 2+.

TV campaigns are planned at all parameters mentioned before (TRP, GRP, Affinity, and Reach). The efficiency obtained from DEA models can be used to select the right combination of these parameters if the campaign is as successful as possible.

The course of the dataset used for the analysis is shown in Figure 1. GRP grows linearly with TRP levels. However, Affinity is almost independent on TRP levels. Instead, Reach is dependent on TRP levels logarithmic. Therefore, the data is analysed at a given level of TRP. The data is divided into nine groups based on TRP levels from 200 TRP to 600 TRP with 50 TRP steps.

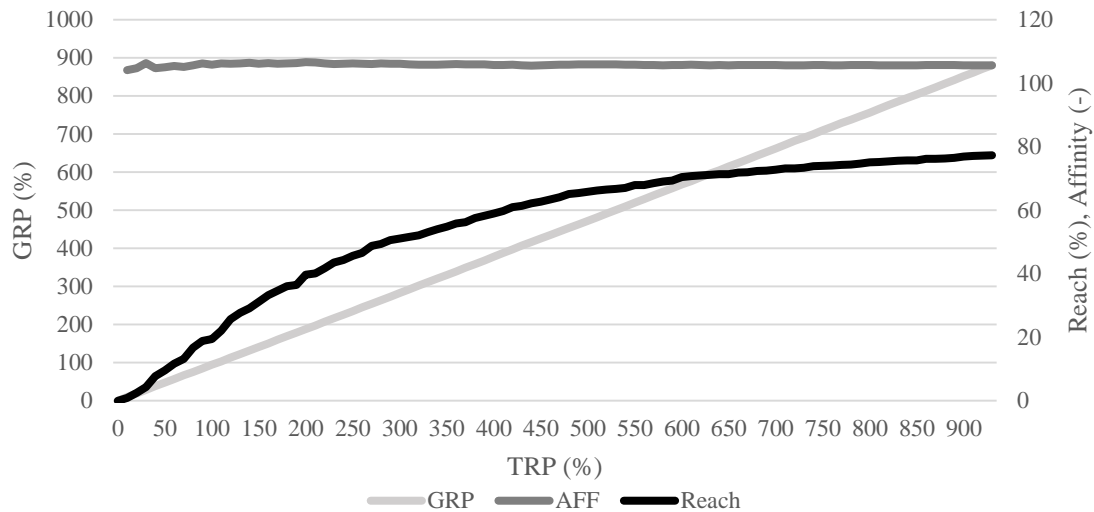


Figure 1 Course of GRP, Affinity and Reach depending on TRP levels

Firstly, the model (4) was computed for the TRP equal to 400 for all channel mixes. The average efficiency scores are shown for every month of the year 2019 according to gender in Figure 2. Figure 2 also shows that the efficiency is higher for men than for women during 2019 at TRP 400. The audience watches TV more (more frequently, for longer period of time, or both) during winter than during summer.

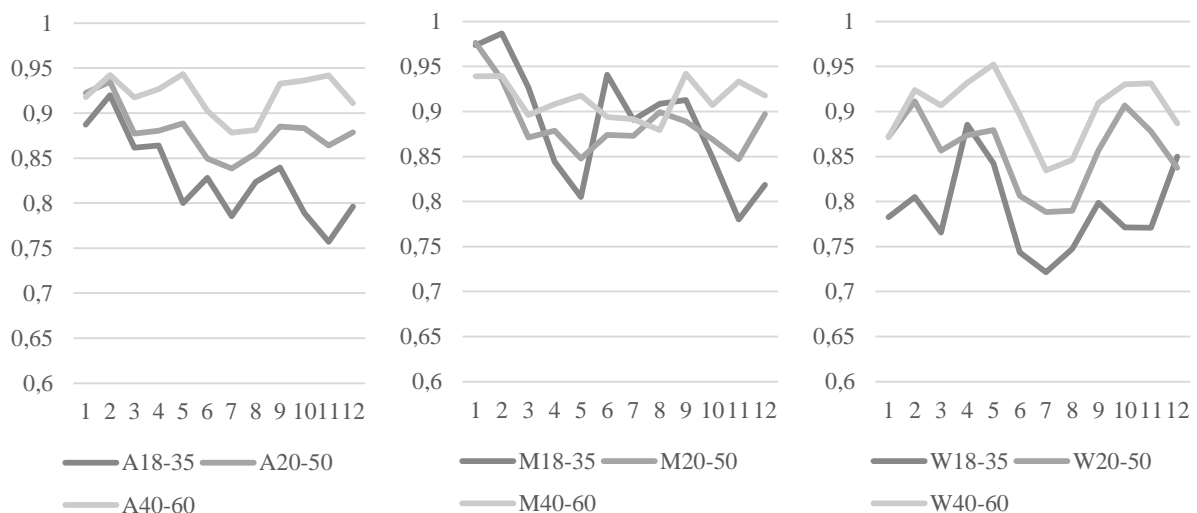


Figure 2 Monthly average efficiency computed by the model (4) for TRP equal to 400 for all channel mixes

Longer days decrease the viewership and for the younger audience even more, as shown in Figure 3, for All target audiences, for male and women audiences the curves look almost the same. The fact that average BCC efficiency is not the lowest in all and male target audiences in summer months do not correspond with Figure 3 which shows that average time spent (ATS) in front of TV is lower in summer months (June, July, August) so lower BCC efficiency would be expected, but this is in line only in women target audiences. Low BCC efficiency in November and December despite of the highest ATS in those months could be caused by high interest of companies in placing their adverts in those times because of Christmas. Relatively high BCC efficiency in winter months is in accordance with Figure 3, where ATS is high and also in those months there is more free space for adverts therefore, companies and advertising agencies can plan more effectively. ATS also shows that despite traditional view that men are watching TV more than women, in reality the situation is reversed. For target audience M20-50 the ATS for year 2019 is 2h18m while for W20-50 it is 2h52m.

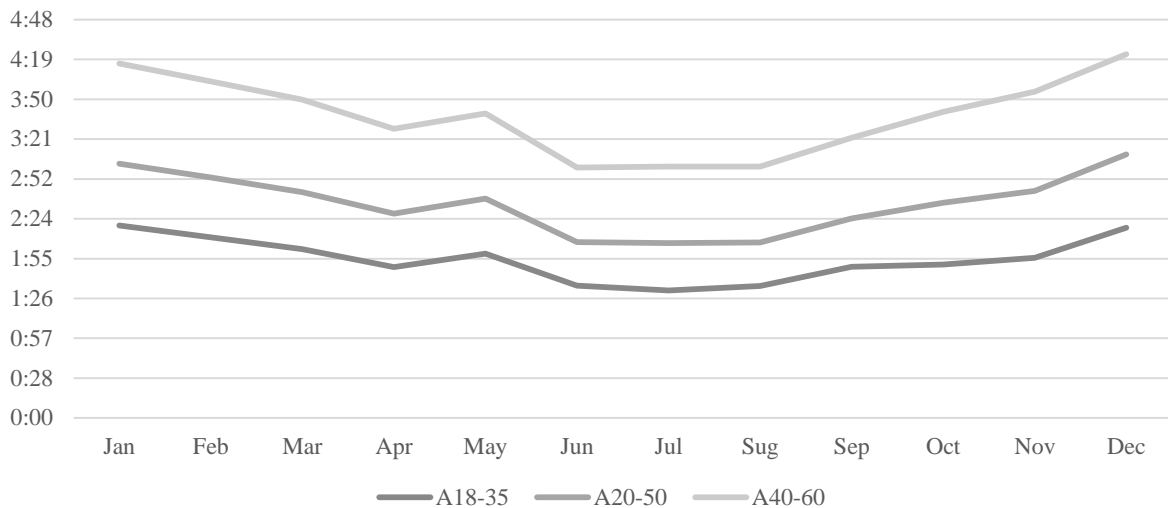


Figure 3 Monthly average time spent in front of TV during year 2019 for All target audiences

Secondly, the model (4) was computed for different TRP levels and a particular target audience (all, men, women). There are relatively significant differences between minimal and maximal efficiency scores across the level of TRP, which indicates that for the specific target audience in a particular channel mix, different amount of TRP for the campaign should be chosen to increase efficiency. There are results for the target audience A18 - 35 in Table 2. An A18 - 35 audience has shown interesting results in efficiency with set levels of TRP, especially the alternation of maximal and minimal efficiency score between the channel mixes. When starting a campaign in channel mixes MC, NG LOW or NG, the target level of TRP is 500 - 600 as it leads to increased or even maximal efficiency. On the other hand, lower levels of TRP significantly decrease efficiency. On the contrary, in channel mixes BOTH and MC LOW, the most suitable TRP levels are 300 - 350 since higher levels of TRP cause efficiency score to drop.

Level of TRP	MC	NG LOW	BOTH	MC LOW	NG
200	0,691	0,732	0,844	0,883	0,926
250	0,686	0,744	0,847	0,877	0,925
300	0,693	0,772	0,873	0,893	0,930
350	0,702	0,789	0,874	0,885	0,934
400	0,702	0,778	0,857	0,876	0,933
450	0,701	0,783	0,851	0,869	0,931
500	0,703	0,780	0,844	0,881	0,941
550	0,699	0,791	0,864	0,890	0,935
600	0,700	0,776	0,821	0,876	0,943

Table 2 Yearly average efficiency scores computed by model (4) for target audience A18 - 35

Conclusions of target audience A20 - 50, in Table 3, are analogous with A18 - 35. Conversely, A40 - 60 shows specific changes, see in Table 3. For instance, in all channel mixes except MC LOW, a bigger campaign from 450 to 600 TRP has a higher efficiency score. Only for MC LOW, it is more advantageous to choose a campaign between 250 - 350 TRP.

Level of TRP	A20-50					A40-60				
	MC	NG LOW	BOTH	MC LOW	NG	MC	NG LOW	BOTH	MC LOW	NG
200	0,753	0,785	0,885	0,923	0,984	0,800	0,863	0,916	0,955	0,962
250	0,755	0,816	0,912	0,942	0,985	0,813	0,882	0,938	0,969	0,969
300	0,749	0,809	0,910	0,935	0,983	0,823	0,886	0,943	0,970	0,972
350	0,753	0,814	0,912	0,937	0,987	0,824	0,887	0,936	0,965	0,967
400	0,755	0,818	0,908	0,932	0,985	0,829	0,888	0,948	0,962	0,971
450	0,758	0,822	0,908	0,937	0,988	0,831	0,894	0,949	0,963	0,971
500	0,756	0,816	0,908	0,924	0,984	0,832	0,888	0,939	0,959	0,970
550	0,758	0,809	0,907	0,920	0,985	0,832	0,889	0,937	0,956	0,970
600	0,758	0,823	0,908	0,923	0,986	0,836	0,891	0,938	0,958	0,970

Table 3 Yearly average efficiency scores computed by model (4) for target audiences A20-50, A40-60

There are results for the target audience M18 - 35 in Table 4. Target audience M18 - 35 shows a completely different situation as any campaign with more than 500 TRP is insufficient. For channel mixes BOTH, MC LOW and NG, the most convenient range of TRP is from 250 to 400, while for MC and NG LOW 400 - 500 TRP is more efficient.

Level of TRP	MC	NG LOW	BOTH	MC LOW	NG
200	0,867	0,856	0,861	0,885	0,905
250	0,883	0,866	0,883	0,902	0,904
300	0,887	0,860	0,874	0,889	0,905
350	0,896	0,858	0,879	0,895	0,905
400	0,903	0,864	0,874	0,881	0,909
450	0,917	0,884	0,868	0,894	0,854
500	0,882	0,898	0,870	0,770	0,843
550	0,867	0,822	0,870	0,713	0,841
600	0,836	0,819	0,841	0,694	0,826

Table 4 Yearly average efficiency scores computed by model (4) for target audience M18-35

In men, audiences M20 - 50, a 450 - 600 TRP campaign in channel mixes MC, BOTH, and NG is more suitable since efficiency is at a peak as is shown in Table 5. Whereas for MC LOW and NG LOW a campaign between 300 and 400 TRP provides higher efficiency. For target audience M40 - 60, in Table 5, the efficiency score decreases together with lower TRP levels and lower share of Media Club. Hence, in the MC channel mix, a 500 - 600 TRP campaign is the most suitable of the other channel mixes, where the most effective campaign's TRP drops by 100 - 150 from the previous channel mix. This trend terminates with channel mix NG in which a 250 - 400 TRP is more efficient.

Level of TRP	M20-50					M40-60				
	MC	NG LOW	BOTH	MC LOW	NG	MC	NG LOW	BOTH	MC LOW	NG
200	0,806	0,834	0,879	0,910	0,939	0,823	0,889	0,930	0,964	0,964
250	0,811	0,839	0,885	0,923	0,944	0,832	0,882	0,936	0,955	0,971
300	0,818	0,849	0,894	0,924	0,951	0,825	0,880	0,922	0,950	0,963
350	0,828	0,853	0,895	0,926	0,956	0,832	0,893	0,935	0,952	0,970
400	0,827	0,854	0,886	0,927	0,946	0,833	0,886	0,937	0,946	0,967
450	0,838	0,848	0,889	0,916	0,954	0,830	0,887	0,932	0,953	0,957
500	0,834	0,849	0,893	0,918	0,956	0,834	0,895	0,933	0,949	0,961
550	0,836	0,847	0,903	0,919	0,965	0,838	0,888	0,928	0,955	0,958
600	0,840	0,846	0,881	0,920	0,957	0,836	0,892	0,933	0,944	0,962

Table 5 Yearly average efficiency scores computed by model (4) for target audiences M20-50, M40-60

W 20 - 50 target audience, in Table 6, follows the same pattern as the M 40-60 audience where the efficiency score is dropping with lower levels of TRP and decreasing MC share, but by lower levels of TRP. Moreover, this trend does not affect channel mix NG as the most suitable TRP levels are between 500 and 600. Generally, TRP of 450 - 600 in W40 - 60 audience shows the highest performance success except for MC LOW channel mix, which has the highest efficiency score in the range of 300 - 350 TRP, see in Table 6.

Level of TRP	W20-50					W40-60				
	MC	NG LOW	BOTH	MC LOW	NG	MC	NG LOW	BOTH	MC LOW	NG
200	0,668	0,735	0,859	0,906	0,970	0,750	0,823	0,880	0,931	0,911
250	0,673	0,745	0,876	0,918	0,970	0,767	0,840	0,909	0,935	0,928
300	0,678	0,779	0,892	0,931	0,976	0,788	0,864	0,932	0,954	0,940
350	0,681	0,770	0,888	0,920	0,975	0,794	0,868	0,936	0,957	0,939
400	0,685	0,779	0,904	0,929	0,976	0,809	0,870	0,933	0,948	0,949
450	0,683	0,777	0,891	0,919	0,975	0,812	0,863	0,928	0,947	0,952
500	0,684	0,788	0,907	0,923	0,981	0,811	0,877	0,937	0,952	0,949
550	0,683	0,782	0,889	0,915	0,981	0,812	0,876	0,933	0,946	0,952
600	0,686	0,782	0,887	0,905	0,980	0,813	0,875	0,940	0,951	0,955

Table 6 Yearly average efficiency scores computed by model (4) for target audiences W20-50, W40-60

There are results for the target audience W18 - 35 in Table 7 where on the other hand, in W18 - 35 audience, TRP of 300 - 350 shows high efficiency through all channel mixes, but the efficiency drops with higher TRP levels, while the decrease stops at 500 - 550 TRP where the efficiency increases again.

Level of TRP	MC	NG LOW	BOTH	MC LOW	NG
200	0,581	0,664	0,798	0,861	0,943
250	0,586	0,717	0,819	0,867	0,933
300	0,591	0,740	0,834	0,888	0,937
350	0,589	0,726	0,837	0,891	0,944
400	0,584	0,747	0,822	0,872	0,927
450	0,576	0,678	0,818	0,859	0,937
500	0,587	0,726	0,801	0,874	0,945
550	0,592	0,731	0,833	0,863	0,953
600	0,578	0,711	0,822	0,848	0,936

Table 7 Yearly average efficiency scores computed by model (4) for target audience W18-35

4 Conclusion

The paper dealt with efficiency analysis based on Data Envelopment Analysis. DEA models can be used in the TV advertisement field only as one of several indicators of campaign evaluation and planning since the evaluation serves only with final data of the specific campaign. That is due to a possibility of misinterpretation caused by evaluating campaign progress with all TRP levels included together in a model. Possibly, a more sophisticated model could operate with data as complex as described.

This paper has provided an insight into the tendencies of men and women of watching TV throughout the year, which could be beneficial for companies' ability to choose the proper periods for their campaigns. Of course, having a campaign as efficient as possible depends on the right combination of both the chosen TRP and channel mix levels for the specific target audience. Nonetheless, there is no general rule that would provide companies with complete certainty that their campaign is going to be successful and efficient.

The paper has shown that BCC efficiency sometimes goes against traditional view, especially in summer months where despite low ATS, relatively great values of efficiency were obtained for all and male audiences, particularly for male audiences was BCC efficiency one of the greatest throughout the year. On other hand, the use of lower levels of ATP has been acknowledged as suitable for younger audiences (18-35) since younger people tend not to watch the TV as frequently, thus aiming for higher TRP is wasteful. In the majority of cases in older target audiences (40-60), the rule "the more the better" has been found suitable. For women target audiences a higher share of Nova Group has been found more efficient in spite of Media Club having more channels that broadcast women-oriented films and shows. All computations were done in the LINGO modelling system. Used data was from [5] a they were collected by Adwind Kite software with the consent of ATO.

Acknowledgements

This work was supported by the Internal Grant Agency of the Faculty of Informatics and Statistics, Prague University of Economics and Business, project F4/29/2020 (*Dynamic data envelopment analysis models in economic decision making*).

References

- [1] *Asociace televizních organizací*. (2021). [Online]. Available: <https://www.ato.cz/tv-vyzkum/terminologie/>
- [2] Charnes, A., Cooper, W. and Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6), p. 429-444.
- [3] Dlouhý, M., Jablonský, J. and Zýková P. (2018). *Analýza obalu dat*. Praha: Professional Publishing.
- [4] Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3): p. 253-290.
- [5] Malý, J. (2020). *Využití DEA modelů v oblasti TV reklamy*, Prague University of Economics and Business.
- [6] *Media Club*. (2021). [Online]. Available: <https://media-club.tv/zastupovana-media-2/>.
- [7] *Nova Group*. (2021). [Online]. Available: <https://www.novagroup.cz/nase-znacky/televize>.
- [8] *Zákon č. 231/2001 Sb.* (2017). [Online]. Available: <https://www.zakonyprolidi.cz/cs/2001-231>.

Efficiency of tertiary education in EU countries

Klára Mašková¹, Veronika Blašková²

Abstract. Modern societies place ever-increasing demands on the level of education of the population. This article focuses on the issue of evaluating the efficiency of tertiary education in EU countries. The evaluation was performed using the data envelopment analysis method. The model was built based on two input and two output variables. Public expenditure on tertiary education and the number of teachers in tertiary education represent the input variables. The employment rate of graduates of tertiary education and the number of graduates in tertiary education represent the outputs. The radial input-oriented model with constant returns to scale was selected to calculate the efficiency in 2016, 2017 and 2018.

The results show that countries such as Austria and Germany are in the lowest positions throughout the period under review and lag far behind other countries. In contrast, some economically weaker countries, such as the Czech Republic, have been identified as fully efficient.

Keywords: data envelopment analysis, efficiency, education, competitiveness, linear programming

JEL Classification: C44, H52, I21

AMS Classification: 90B50, 90C08

1 Introduction

Currently, the situation on the tertiary education market is changing dramatically. This process of change began almost ten years ago, when there was a large increase in the number of educational institutions and thus a boom in private education, but on the other hand, demographic changes (declines) began to show in this market ten years ago. Due to the fight for students, tertiary education is becoming much more accessible than in the past [20]. It is also necessary to highlight the significant decline in the number of traditional and highly motivated students and conversely the influx of various students who have neither the prerequisites nor the motivation to study at university [17]. Another problem which higher education is currently facing is the fact that there is a shortage of students in some sectors and a surplus in others [9]. For example, there is a great demand for graduates in science and technology and in foreign languages [1].

Higher education is expanding a lot today, as mentioned above, and it is therefore necessary to maintain quality. For this reason, it is important to increase spending on education both on the part of the state and on the part of entrepreneurs or the customers for educational services [4]. According to [1] increasing investment in tertiary education can improve the quality of education and the same time increase the number of graduates. For the tertiary level of education the size of the investment is as important as the structure and efficiency of its allocation. It should also be noted that investment in education is influenced by several factors. For example, it could be the size of tuition fees, the demographic structure of the population or the level of the economy [9]. Universities are under great pressure for the above reasons because of a steady decline in public support. Because of this, universities must look for external funding, and provide the best and most interesting teaching and research. Otherwise, they will not survive current global competition [22]. The funding of higher education is itself a much-discussed topic, see [24].

Today, tertiary education gives a great advantage when entering the labour market. Unfortunately, there is the problem that some graduates will not succeed in the labour market, even though the demand for university-educated employees is still growing [18]. According to [21] this failure is mainly associated with the absence of work experience and work habits. Another problem is poor awareness of the value of salaries. Conversely, employers appreciate graduates with a knowledge of foreign languages, communication skills and excellent computer skills. They also appreciate the potential of graduates for further training and self-development. In many

¹ Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, xmaskov1@node.mendelu.cz.

² Mendel University in Brno, Department of Statistics and Operation Analysis, Zemědělská 1, 613 00 Brno, Czech Republic, veronika.blaskova@mendelu.cz.

studies, for example [20], [21], the authors deal with the fact that computer skills are a great advantage for graduates who are just entering the labour market. At the time of the Industry 4.0 technical revolution, demand for good computer skills is becoming more common. However, the shortcomings are complemented by low work responsibility and insufficient active knowledge of a foreign language. The employability of graduates is also currently a much-discussed topic, on which [16] also worked. Graduates of tertiary education in 2016–2018 are included in generation Y. In [15], among other things, it was found that most respondents from generation Y think that they will succeed in the labour market if they focus on the natural or technical sciences, especially the currently evolving field – IT. Even this finding is consistent with the current industrial revolution.

Given the limited possibilities for countries regarding the financing of education and the growing fight for students, the need to evaluate the efficiency of entities in this sector comes to the fore. Within the methods used to evaluate efficiency, the non-parametric method of data envelopment analysis (DEA) appears most often. This method is popular in all areas. It is possible to find applications in healthcare [3], construction [11] or in the evaluation of manufacturing companies [12]. The DEA method is also widely used in the field of education, see [6], [7] and [8]. Thanks to the high homogeneity of the subjects analyzed, it is possible to perform the analysis at the level of whole countries.

The main aim of this article is to evaluate the efficiency of individual EU countries in the field of tertiary education from 2016 to 2018. The efficiency of education can be derived from different perspectives. In this article we solve the efficiency of tertiary education in terms of graduate employability in conjunction with the financing of tertiary education. Countries should strive to maximize the employment of graduates.

2 Material and Methods

Regarding the availability of data and the number of observations, a total of four variables were selected for the efficiency analysis. Public expenditure on tertiary education in millions of euros and the number of teachers in tertiary education represent the input variables. The employment rate (%) of graduates of tertiary education and the number of graduates in tertiary education represent the outputs. The basic characteristics of the data set are in Table 1.

Due to the unavailability of certain data, 6 countries had to be excluded from the analysis, these are Bulgaria, Cyprus, Estonia, Greece, Malta and Latvia. When we look at the values shown in the Table 1, Luxembourg had the lowest public expenditure, number of teachers and the number of absolvents in all years. On the contrary, Germany had the highest public expenditure, the most teachers and graduates. In 2016, the employment rate was the lowest in Slovakia and the highest in Lithuania. In 2017 and 2018 the employment rate was lowest in Italy and still the highest in Lithuania.

Year	Variable	Min	Max	Mean	Lower quartile	Median	Upper quartile
2016	Public expenditure	193.5	29800.0	4561.9	547.9	2280.5	3828.0
	Employment rate	77.3	90.4	83.5	82.2	83.7	85.8
	Number of teachers	0.8	402.4	60.4	14.9	30.5	65.0
	Number of absolvents	6.6	2729.6	682.3	205.3	347.4	770.0
2017	Public expenditure	206.1	30500.0	4671.5	710.0	2253.0	3937.0
	Employment rate	78.2	90.0	84.4	82.9	84.3	86.8
	Number of teachers	1.0	407.1	60.3	14.9	26.6	68.7
	Number of absolvents	6.5	2779.4	688.7	213.8	329.2	813.3
2018	Public expenditure	232.8	31100.0	4818.9	759.2	2344.0	4141.0
	Employment rate	78.7	90.5	85.1	83.4	85.5	88.0
	Number of teachers	1.3	416.2	61.7	15.1	26.7	69.8
	Number of absolvents	6.6	2805.8	694.8	219.6	314.3	824.6

Table 1 Basic descriptive characteristics of the variables used in individual years. Public expenditure on tertiary education in millions of euros, employment rate of graduates of tertiary education in % , number of teachers in tertiary education in thousands, number of graduates in tertiary education in thousands.

The DEA method was chosen to calculate the efficiency. Given the selected variables, the input orientation of the model was selected, and the model was estimated assuming constant returns to scale (i.e. the CCR model) as in [11] and [13]. Based on input matrix (X) and output matrix (Y) in can be calculated efficiency of each county c by solving n times model:

$$\begin{aligned} \min_{\theta, \lambda} \theta & \quad (1) \\ \theta x_c - X\lambda & \geq 0 \\ Y\lambda & \geq y_c \\ \lambda & \geq 0. \end{aligned}$$

Similar to [10] and [13], we used the Malmquist index to calculate the change in efficiency, more precisely its partial component, the so-called catch-up effect. Malmquist index can be defined as the geometric mean of two efficiency ratios, where one is the efficiency change measures by the period 1 technology and the other is the efficiency change measured by the period 2 technology:

$$MI = \left[\frac{\delta^1((x_c, y_c)^2)}{\delta^1((x_c, y_c)^1)} * \frac{\delta^2((x_c, y_c)^2)}{\delta^2((x_c, y_c)^1)} \right]^{1/2}. \quad (2)$$

According to Formula 2, it can be stated that the Malmquist index consists four terms: $\delta^1((x_c, y_c)^2)$, $\delta^2((x_c, y_c)^2)$, $\delta^1((x_c, y_c)^1)$ and $\delta^2((x_c, y_c)^1)$. For each of these terms it is necessary to solve linear programs, and in this paper this calculation is based on the CCR model in Formula 1. Technical details about the DEA method and the Malmquist index can be found in [2]. The calculations are performed in the MATLAB R2021a computational system and in DEA SolverPro, Version 15.

3 Results and Discussion

The following Figure 1 shows the efficiency score in % for each country in individual years (2016 is marked in blue, 2017 in red and 2018 in orange). The average efficiency of tertiary education in 2016 was approximately 74%. The median value was slightly higher, at around 77%, this value meant that half of the countries had their efficiency score above 77%. In the countries of the euro area, the average efficiency was also 74%, in the V4 countries the efficiency was higher, at 86%. The full efficiency score in this year was achieved by five countries, namely Czechia, Ireland, Lithuania, Luxembourg and Romania. On the contrary, the lowest efficiency score can be seen in Germany, Austria and Denmark. Only these three countries had an efficiency score under 50%. On the 50% border we can find Sweden and Spain.

In 2017, five countries again achieved full efficiency, the same states as in 2016. We also calculated the average and median value for this year. The average efficiency of European countries was almost 77%. The median value was almost the same as in 2016, also 77%. In 2017, we could observe that the average and median value were almost the same. In the euro area countries, the average efficiency score was lower, 75%; in the V4 countries the efficiency score was higher again, almost 84%. Only two countries were below 50% efficiency. These were Germany with a score of 34%, and Austria with a score of 38%. The value of 50% was also approached by Sweden with a score of 51% and Spain with a score of 56%

In 2018, only four countries had the full efficiency score, namely Ireland, Lithuania, Luxembourg and Romania. In this year, we observed that the average European efficiency was almost the same as in 2017, approximately 77% However, the median value of this year was higher than the average. Half of the countries were above the 83% efficiency score. In the euro area countries, the average efficiency was approximately at the same level, in the V4 countries the efficiency was higher again, 82%. Among the countries where we observed the lowest efficiency of tertiary education was again Germany with a score of 34%, Austria with a score of 36% and Denmark with a score of 49.5%. Sweden (51%) and Spain (55%) were again at the 50% mark.

We observed that in the V4 countries, the average efficiency was higher than the average for all countries each year. A common feature was also that economically developed European countries such as Germany and Austria failed in the evaluation of the efficiency of tertiary education and were in the last two places in our ranking. On the contrary, some less economically viable countries, such as Romania, finished as a country with full efficiency. The “inefficiency” of Germany is mainly caused by the high number of teachers in contrast to other countries. And this fact in also reflected in the value of public expenditure.

For example [14], their work dealt with the efficiency of tertiary education expenditure in the European Union. Their three efficiency models confirm that we can rank both Romania and Czechia among the most efficient countries. In their work the worst was Estonia, which we had to exclude due to the unavailability of certain data. A similarity with our results can be also found in [23], where Latvia, Lithuania, Romania and Czechia were best placed. Similar results are also confirmed by [19]. In [5] it was also found that more economically developed countries such as Austria are much less efficient than, for example, Hungary.

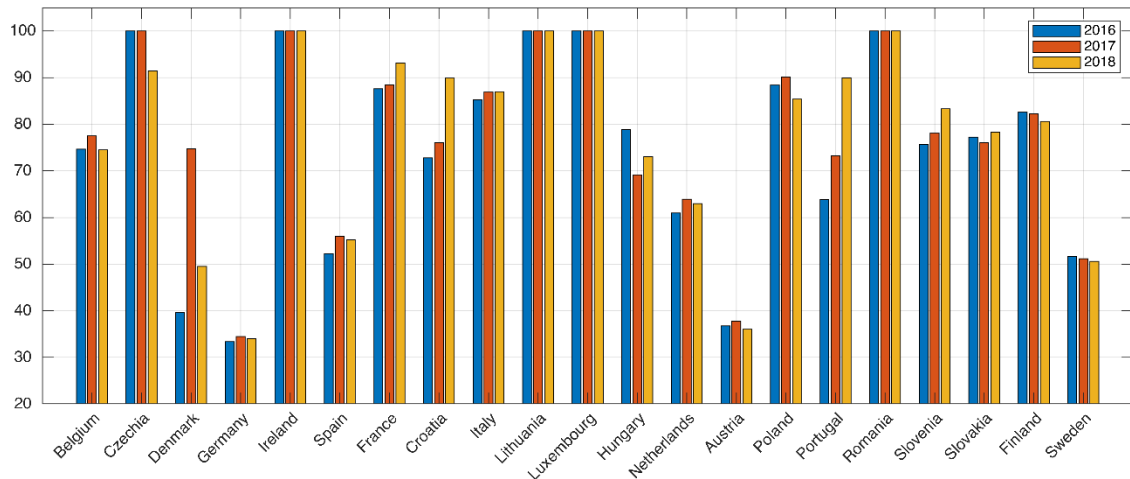


Figure 1 The resulting efficiency score in % for each country in individual years.

The following Figure 2 shows efficiency changes for each country in individual years (change for the period 2016/2017 is marked in blue, change for the period 2017/2018 is marked in red). Figure 2 was compiled in the form of a growth coefficient – values greater than zero mean an increase in technical efficiency and, conversely, values less than zero mean a decrease. If there is no change in efficiency over time, then the value of the coefficient for the given country is equal to zero.

In the first monitored period, we could see the biggest change in Denmark, where the efficiency score increased by almost 90%. This year-on-year change was mainly due to a change on the input side – in this period the number of teachers in Denmark decreased by approximately 46%. The other variables of the model remained almost at the same level (a change of up to 10%). We could also observe an increase in another eleven countries, but this increase was not as dramatic (only between 1 and 15%). Some countries also showed a decrease in efficiency scores, namely Hungary, Slovenia, Slovakia, Finland and Sweden. It was only a slight decrease of about 1%. But only Hungary has declined by 13%. Five countries did not change their efficiency scores; these are the countries with full efficiency in Figure 1.

In the second monitoring period, changes up to 35% were observed. The highest growths were in the case of Portugal and Croatia – almost 23% and 19%. We observed less significant growth, of up to about 7%, in 5 countries. In the case of 4 countries, the efficiency score did not change, these are Ireland, Lithuania, Luxembourg and Romania – again countries that were fully efficient in Figure 1. We identified falls in the efficiency score in 10 countries. The biggest fall was recorded in Denmark, a drop of almost 35%. In Denmark in 2017, there was a reform of tertiary education. This year-on-year decrease was again caused by a significant change on the input side – in this period, on the contrary, the number of teachers increased by almost 55%. The other variables of the model remained almost the same, these were only minimal changes. The second largest decline was observed in Czechia at around 9%. Other declines were up to about 5%.

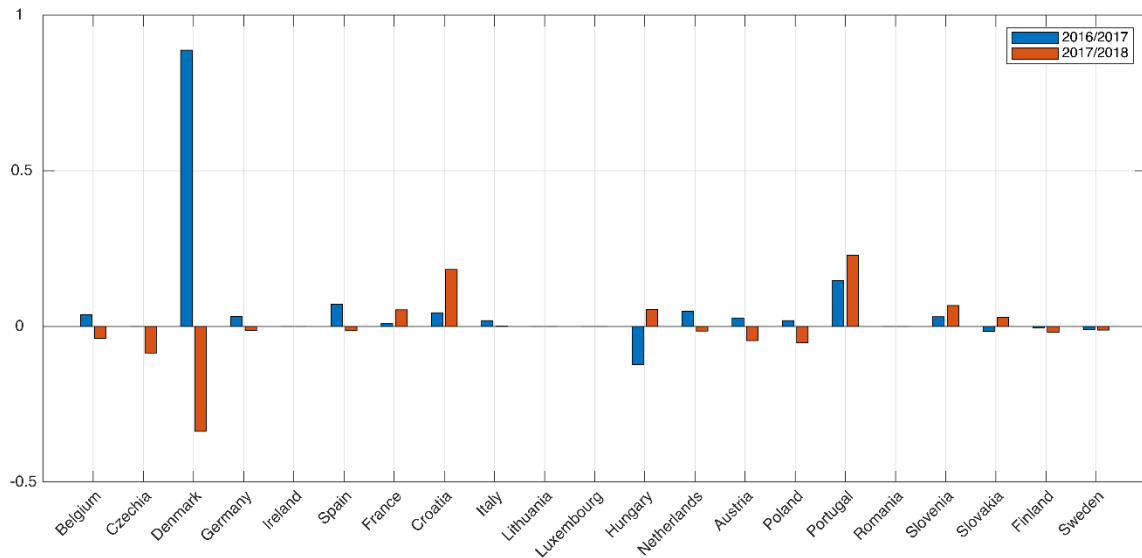


Figure 2 Efficiency increment of individual countries in individual years

4 Conclusion

The results of this article show that some economically stronger countries, such as Germany and Austria, do not perform well in tertiary education compared to other economically weaker countries, such as the Czech Republic. Among the best countries in terms of efficiency derived on the basis of graduate employability in conjunction with the financing of tertiary education, in addition to the already mentioned Czech Republic, were also, for example, Ireland and Lithuania. It was also found that Denmark underwent significant changes in education during the period considered. For other countries, changes in efficiency over time have not been so dramatic.

Acknowledgements

This article was supported by grant No. PEF/TP/2021003 of the Grant Agency IGA PEF MENDELU.

References

- [1] Aristovnik, A. & Obadić, A. (2011). The Funding and Efficiency of Higher Education in Croatia and Slovenia; A Nonparametric Comparison. *Proceedings of the International Scientific Conference, Juraj Dobrila University of Pula*. 218–244.
- [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] Gaebert, T. & Staňková, M. (2020). Efficiency Development in the German Pharmaceutical Market. *Acta Universitatis agriculturae et silviculturae Mendelianae Brunensis*, 68(5), 877–884.
- [4] Glushak, N., Katkow, Y., Glushak, O., Katkova, E., Kovaleva, N. (2015). Contemporary Economic Aspects of Education Quality Management at the University. *Procedia – Social and Behavioral Sciences*, 214, 252–260.
- [5] Jelic, O. & Kedzo, M. (2018). Efficiency vs effectiveness: an analysis of tertiary education across Europe. *Public sector economics*, 42(4), 381–414.
- [6] López-Torres, L. & Prior, D. (2020). Long-term efficiency of public service provision in a context of budget restrictions. An application to the education sector. *Socio-Economic Planning Sciences*, 100946.
- [7] Melo-Becerra, L. A., Hahn-De-Castro, L. H., Ariza, D. A. & Carmona, C. O. (2020). Efficiency of local public education in a decentralized context. *International Journal of Educational Development*, 76, 102194.
- [8] Mikusova, P. (2020). The Efficiency of Public Higher Education Institutions: A Meta-Analysis. *Ekonomický časopis*, 68(9), 963–977.
- [9] Rozborilová, D. (2018). Investments in Tertiary Education in the Context of Labor Market needs at the Beginning of the 21st Century. *International Review of Research in Emerging Markets*, 4(1), 1248–1264.

- [10] Staňková, M. (2020). Efficiency comparison and efficiency development of the metallurgical industry in the EU: Parametric and non-parametric approaches. *Acta Universitatis agriculturae et silviculturae Mendeliana Brunensis*, 68(4), 765–774.
- [11] 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.
- [12] 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.
- [13] Staňková, M. & Hampel, D. (2020). Efficiency Assessment of the UK Travel Agency Companies – Data Envelopment Analysis Approach. *Mathematical Methods in Economics 2020: Conference Proceedings*. Brno: Mendelova univerzita v Brně, 550–556.
- [14] Stefanova, K. & Velichkov, N. (2020). Analysis of the Efficiency of Tertiary Education Expenditure in European Union Member States from Central and Eastern Europe: An Efficiency Frontier Approach. *South-Eastern Europe Journal of Economics* 1, 115–128.
- [15] Stojanová, H., Blašková, V., Tomšík, P. & Tesařová, E. (2015). Specification and Characteristic of Generation Y in the Sphere of Work Attitude. *DIEM 2015: Innovation, Leadership & Entrepreneurship*, 565–579.
- [16] Stojanová, H. & Blašková, V. (2014). The significance of the chosen field of study, depending on the difficulty of finding a job. *INTED2014 Proceedings*. Valencia, Spain: IATED, 4002–4012.
- [17] Šmejkalová, J. (2016). Proč se zabývat kvalitou vysokoškolského vzdělávání? *Ekonomické listy* (2), 42–51.
- [18] Šnýdrová, M., Šnýdrová, I., Vnoučková, L. (2017). Vnímání příčin uplatnitelnosti absolventů a jejich závislosti a specifika u dílčích skupin. *Ekonomické listy* 8(2), 40–55.
- [19] Šonje, A., Deskar-Škrbić, M. & Šonje, V. (2018). Efficiency of public expenditure on education: comparing Croatia with other NMS. *International Technology, Education and Development Conference Valencia, Spain*, 12, 1-14.
- [20] Trčka, L. (2014). Vzdělávací procesy – výzkum potřeb a zkušeností zaměstnavatelů absolventů. *Trendy ekonomiky a managementu*, 8(19), 63–69.
- [21] Úlovec, M. (2014). *Potřeby zaměstnavatelů a připravenost absolventů škol – komparační analýza*. Praha: NÚV.
- [22] Wolszczak, J. (2014). An evaluation and explanation of (in)efficiency in higher education institutions in Europe and the U.S. with the application of two-stage semi-parametric DEA. *UC Berkeley; Institute for Research on Labor and Employment*.
- [23] Yotova, L. & Stefanova, K. (2017). Efficiency of Tertiary Education Expenditure in CEE Countries: Data Envelopment Analysis. *Economic Alternatives*, 3, 352–364.
- [24] Zámková, M. & Blašková, V. (2013). Comparing the views on tuition fee introduction of Brno university students. *Efficiency and Responsibility in Education 2013*. Praha: CULS, 671–679.

Application of robust efficiency evaluation method on the Czech life and non-life insurance markets

Markéta Matulová¹, Lucia Kubincová²

Abstract. The paper presents the analysis of more than forty insurance companies operating in the Czech Republic in the period 2004–2018. The performance of the companies is evaluated by an universal robust Data Envelopment Analysis. The specification of the model applied in our study includes six variables as the inputs for the analysis: the number of employees and intermediaries, operating costs for the life and non-life segment, equity and total liabilities. As outputs, we use three variables: the financial placement and earned premiums for life and non-life insurance. We compare the efficiency of the companies operating in the individual insurance markets using nonparametric statistical tests. We also try to monitor the dynamics of the performance of individual companies.

Keywords: efficiency, life insurance, non-life insurance, robust DEA

JEL Classification: G22, C67

AMS Classification: 90C05

1 Introduction

Since the re-establishment of the Czech insurance market in the early 1990s, the the insurance sector has undergone a number of transformations and changes. Legislative changes such as the integration of European legislation and the economic situation in the country (the impact of the economic crisis at the end of 2009) are just one of many factors that directly affect the activity of the insurance market. In 1991, Act No. 185/1991 on insurance was adopted, opening up the Czech insurance market. In the following years, a number of insurance companies were established and entered the Czech insurance market. By the end of the 20th century, the market had taken shape, a number of laws and legislative regulations. The requirements for the entry of insurance companies into the market were tightened (minimal value of capital for insurance companies, control of owners) and other measures and broader powers of supervisory authorities in the event of insolvency of insurance companies were set. In 1999 Act No 168/1999 on public liability insurance was adopted which abolished the monopoly of Česká pojišťovna on this product. With the accession to the EU in 2004, insurance companies operating in other European countries were able to enter the Czech insurance market. In 2005, the Act on Supplementary Supervision of Banks and Insurance Companies was adopted. In 2006, the Czech insurance market gained the highest profit after tax since the beginning of its existence. Life insurance growth has resumed, especially for investment products. At the end of 2008, both the global and the Czech economy began to feel the effects of of the economic crisis, with a decline in the performance of the insurance market in the following years. The year 2010 is characterized by a number of natural catastrophes, which was reflected in particular in the high cost of non-life insurance claims. In 2011, the insurance market experienced a slight decline. Negative trends were present even in the life insurance market, which has been relatively stable in recent years thanks to the rise in average rates. However, the number of life insurance contracts was declining. The non-life insurance market continued to stagnate. The crisis has reduced rates for the key sector - car insurance - to the minimum possible level. However, non-life insurance continued to lag behind life insurance in 2013. High claims costs caused by natural catastrophes contributed significantly to this effect. In 2014, we are seeing modest growth in the insurance market after years of stagnation. Life insurance is experiencing losses in premiums, driven by the continuing trend of cancellations of insurance contracts as a consequence of the change in the recognition of tax benefits. This trend is peaking at the end of the year 2014. The year 2015 was affected by the implementation of solvency rules and preparatory processes of the long-discussed European Solvency II Directive. The trend from 2014 continues, with the non-life segment growing. The growth of the non-life segment and the stagnation of the life sector continued in 2017 and 2018. Act No. 170/2018 Coll. had a significant impact on the insurance sector, on the distribution of insurance and reinsurance, which implements the so-called "Directive IDD" of the European Council and Parliament. For insurance intermediaries and independent liquidators, it introduces changes in the structure and imposes new requirements for business authorizations in the insurance sector. The insurance market has grown in 2019, primarily driven by the non-life insurance sector. The

¹ Masaryk University, Faculty of Economics and Administration, Lipová 41a, 602 00 Brno, Czech Republic, Marketa.Matulova@econ.muni.cz

² Slovenská Sporiteľňa, Tomášikova 48 832 37 Bratislava, Slovakia, 451791@mail.muni.cz

ratio of segments in this year's has tilted even more strongly towards the non-life insurance market, but this has clearly been influenced by that year's change in risk classification.

We present selected indicators of the insurance market in a graphical form in Figure 1. The data are obtained from the website of the Czech National Bank from the public database ARAD [1]. The coloured bars represent the development of the number of insurance companies in the Czech market. Number of insurance companies stabilised in 2007 after the reopening of the market and EU accession, since when it has been oscillating around value of 53. The largest share in the number of insurers on the market is represented by the entities specializing in a particular non-life insurance sector. We can also observe a black line representing annual insurability values, which is the ratio of gross premiums written and gross domestic product in a given year. It is an important indicator of the quality of the insurance market. In this context, we see that the highest insurability values are achieved by Czech insurance market for the pre-crisis period (2003 and 2004) and in the post-crisis period, which is mainly due to the decline in GDP in these years. The comparison of the insurability of the Czech insurance market with the European average is rather difficult and not entirely adequate. In Western European countries, commercial insurers also participate in social, health or pension insurance in different ways, which considerably changes the input values for this indicator. This is also related to the different structure of market structure, where for Western Europe or the USA the ratio of life insurance to the non-life segment is 40:60. For the Czech Republic, this ratio has long been reversed and in recent years, the dominance of the non-life insurance market has even increased to 70:30.

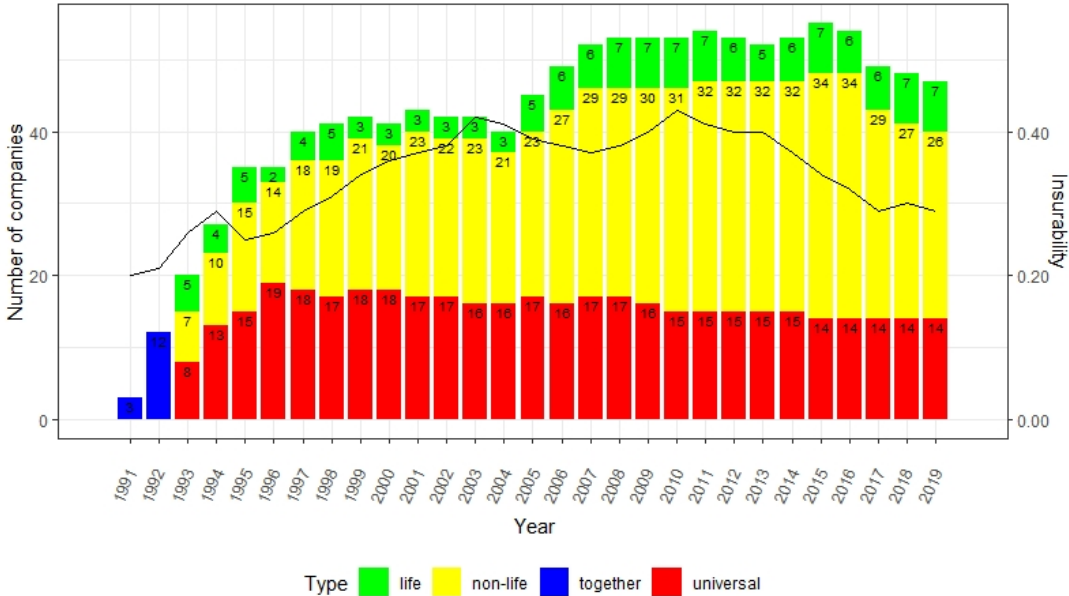


Figure 1 Number of insurance companies and insurability

Important indicators of the insurance market include the level of premiums and the level of claims costs. In Figure 2, we plot the development of gross written premiums and the value of gross claims costs separately for the life and non-life segment.

The values depicted in Figure 2 correspond to the market developments described at the beginning of this chapter. The annual aggregate values of non-life premiums exceed each year the amount of in the life sector. We can see that the gap has widened in recent years, as evidenced by the 70:30 segment comparison for non-life insurance mentioned above. In the post-crisis period, we observe a stagnation of premiums in both segments. Since 2015, we identify a decline in the life segment, with both premiums and the cost of claims. This development is related to the increasing trend of cancellations in the life sector. We observe the highest values of life claims costs in 2014, which is a result of the peak in the life insurance market. We are seeing a huge number of early redemption of policies due to the removal of tax benefits for life insurance products from this year onwards.

2 Literature review

There is an increasing interest in assessing efficiency of the insurance sector and hence the amount of published literature is growing as well.

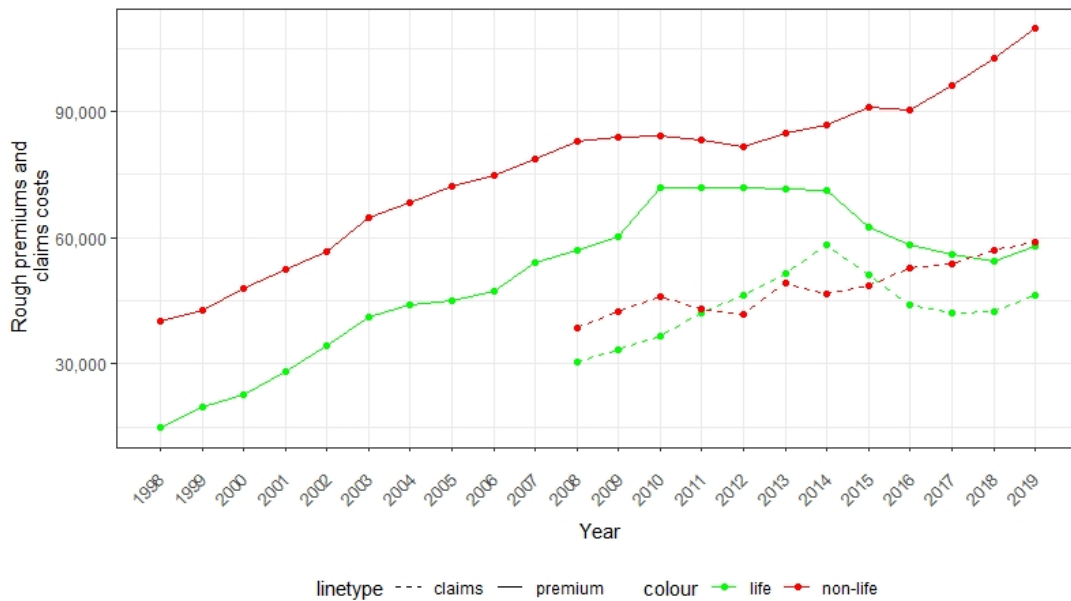


Figure 2 Rough premiums and claims costs

The often cited article [6] deals with the comparison of the efficiency of insurance companies among 36 countries around the world. It covers both life and non-life insurance companies over the time span 2002-2006. The authors used both the non-parametric DEA approach and the parametric SFA approach. They identify great differences between the countries, with higher efficiency achieved by countries in Europe and Asia - with the best scores for Japan and Denmark. The calculated efficiencies were further examined on the basis of the explanatory variables to identify firm- or country-specific effects.

As part of their work, Cummins and Weiss [4] conducted a review of studies on insurance market efficiency. It consisted of 74 studies published between 1983 and 2000 and 37 articles published in leading journals between 2000 and 2011. In 59.5% of the studies reviewed, DEA was applied. The authors state that for the insurance sector, non-parametric methods are an appropriate choice as a tool for efficiency analysis. This is related to the nature of the insurance sector, which does not produce products but offers services, therefore the use of parametric methods with specific assumptions on the shape of the production function are less appropriate choice. Part of the paper addresses the issue of selection of inputs and outputs. The authors present the main principles for measuring outputs of financial institutions, namely the intermediation approach, the so-called user-cost approach and the value-added approach. The last one, the value-added approach, is considered to be the most appropriate and therefore the most used. Authors state that the most common and also the most appropriate option for outputs are insurance claims together with changes in reserves separately for non-life and life insurance and the value of investments. In the analysed studies, a frequent choice of output was also written or received premiums, a choice not supported by Cummins and Weiss. Considering inputs, labor and various forms of capital dominated. Labor is usually expressed as the number of employees, the number of hours, or, if data are not available, average wage per worker in the industry is used for the calculations. Some studies distinguished between employees and intermediaries, and capital was sometimes divided into physical, debt (total) and equity. Other authors used material (which often included physical capital) as input as well.

Newer article of Wise [14] compares 6 different approaches to the estimation of the effective frontier in the evaluation of life insurance companies. He provides information on the approaches used in 190 empirical studies and finds that the SFA approach is used less frequently than DEA from year to year. The difference is significant especially in recent years.

Another survey that analyzes insurance market efficiency was written by Kaffash [11]. It includes 132 articles dealing with the assessment of the insurance market efficiency using DEA exclusively, covering the period from 1993 to 2018. The paper provides an overview of the choice of variables and DEA models used in measuring efficiency of the insurance sector. The most widely used approach in the choice of variables to model proved to be the value-added approach (applied by 68% of the studies). The following inputs dominated - labor/number of workers (60.72%), capital/debt capital (49.18%), equity (37.70%), and materials and business services (32.79%).

For outputs the most frequent choice was insurance premiums (50.82%), personal insurance costs separately for life and non-life insurance (22.95%), incurred losses (22.13%) and investment income (21.31%).

When conducting the literature search, we looked for studies that evaluated the effectiveness of using DEA for the Czech insurance market. We found several studies. The first of them, [9], estimates the efficiency of life insurance markets for Poland and the Czech Republic. The aim of the study is to compare the efficiency of 17 Czech and 26 Polish life insurance companies for the year 2014.

Furthermore, several studies are comparing Czech and Slovak insurance markets. The study [7] analyzes universal insurance companies from the Czech Republic and Slovakia using the data sample covering year 2007. The following inputs were used for this study - costs of indemnity and operating costs. Outputs included earned written premiums and other income. The level of technical efficiency was calculated on the basis of an input-oriented model with variable returns to scale. When comparing the two countries, the Czech Republic had a higher average of effective scores and less variability. Another conclusion of this paper is that both markets are strongly influenced by the so-called giants, i.e. several large insurance companies. Česká pojišťovna was rated the most efficient of the two countries.

The aim of the article [8] was to compare the effectiveness of Slovak and Czech insurance market using network DEA, i.e. a multi-stage DEA model. The study used operating costs and claims incurred as inputs to the first phase that determined the efficiency with respect to minimizing costs. Written premiums were used as the output of the first phase and at the same time as the input entering the second phase. The second phase evaluated the efficiency based on profit maximization, the outputs being a income from financial investments.

Masárová et al., [13] evaluated the development of the Slovak and Czech insurance market in the period 2004-2014. They state that life insurance dominates on the Slovak market whereas non-life insurance dominates in the Czech Republic (on the basis of gross written premiums). For both countries, the mean value of insurance (gross written premium capita) is still below the European Union average, which is interpreted by the authors as the growth potential of these markets.

3 Methodology

Many of the studies mentioned in the previous chapter deal with the appropriate choice of inputs and outputs to of the production process in the insurance industry. When using financial data, we have to deal with the variables with both positive and negative values, which is not possible in traditional DEA model. The issue of negative values of inputs or outputs for DEA models is described by Lovell et al., [12]. His study deals with models invariant to transformation of inputs and outputs. Biener et al., [2] applied transformation to get non-negative values on the Swiss insurance data.

We decided to use the following robust procedure for computing DEA scores introduced by Hladík [10]. The unit under evaluation DMU_0 is assigned the efficiency score $r = 1 + \delta^*$, where δ^* is the optimal solution of the linear program

$$\delta^* = \max \delta \quad \text{subject to} \quad y_0^T u - v_0 \geq 1 + \delta, \quad x_0^T v \leq 1 - \delta, \quad Yu - Xv - 1v_0 \leq 0, \quad u, v \geq 0 \quad (1)$$

where

- $x_0 \in \mathbb{R}^{n_1}$ is the input nonnegative vector for DMU_0 ,
- $y_0 \in \mathbb{R}^{n_2}$ is the output nonnegative vector for DMU_0 ,
- $X \in \mathbb{R}^{m \times n_1}$ is the input nonnegative matrix for the other DMUs, in particular, the i th row of X is the input vector for the i th DMU
- $Y \in \mathbb{R}^{m \times n_2}$ is the output nonnegative matrix for the other DMUs, in particular, the i th row of Y is the output vector for the i th DMU,
- u and v are vectors of variables representing output and input weights, respectively,
- 1 is a vector of ones with convenient dimension.

According to [10], $r \in [0, 2]$ and DMU_0 is efficient if and only if $r \geq 1$, otherwise $r < 1$. In contrast to the classical DEA approach, the proposed score r can be also used for comparing DMU's from different models, because it is naturally normalized and universal score. The author also proves that proposed method does not change the scores of inefficient units obtained by classical DEA approach. Moreover, for efficient units, the score r shows how much the efficient DMU's are close to inefficiency. Below we show another useful property of the score r .

Theorem 1. *The score r defined by the equation (1) is translation invariant with respect to outputs.*

Proof. Lets assume that i -th output is enlarged by a constant k , so that we have a new matrix \tilde{Y} with the i -th column equal to $(y_{1i} + k, \dots, y_{mi} + k)^\top$ and other columns the same as in the original matrix. The formulation of the model 1 for the new output matrix is then following:

$$\delta^* = \max \delta \quad \text{subject to} \quad y_0^\top u + ku_i - v_0 \geq 1 + \delta, \quad x_0^\top v \leq 1 - \delta, \quad Yu - Xv + 1ku_i - 1v_0 \leq 0, \quad u, v \geq 0 \quad (2)$$

The constraints can be rewritten using the substitution $\tilde{v}_0 = v_0 - ku_i$ into

$$y_0^\top u - \tilde{v}_0 \geq 1 + \delta, \quad x_0^\top v \leq 1 - \delta, \quad Yu - Xv - 1\tilde{v}_0 \leq 0, \quad u, v \geq 0 \quad (3)$$

which gives us the same feasible set as in the model 1 as the variable v_0 is of an arbitrary sign. \square

4 Data

We analyzed the data available from the annual reports of the association Czech Insurance Association members, ČAP. The years 2004-2018 are captured in our data sample. For each year, we record the overall results for that year, the balance sheet and profit and loss account, and also the amount of written premiums according to the ČAP methodology. We have information on 43 different insurance companies operating on the Czech market, which is not a complete sample. Despite the possibilities of supplementing the missing data from the annual reports of individual insurance companies, we kept only the data published by the ČAP organization. The reason is that the share of members ČAP insurance companies on the total written premium in the Czech Republic is about 97%. We can therefore say that these data sufficiently represent the Czech insurance market. Our dataset is incomplete not only because of the existence of ČAP non-members, but also because some insurance companies left the Czech market or entered it only after 2004. In addition to this we have to adjust the sample due to the mergers of some insurance companies. So we had to apply various transformations and selection criteria (especially deflating the financial data and translation of negative outputs) to comply with the specifics of the methodology as described in the literature. That allowed us to use selected methods of estimating efficiency on our modified datasets. In our DEA model, we use an intertemporal production frontier approach that assumes, that the technology is the same for all periods. Thus, all units are evaluated for all periods under one DEA model.

The life products are offered exclusively by 6 life insurance companies and 18 universal life insurers in our dataset. The most commonly offered products of life insurance is insurance with investment fund, which account for almost half of of life insurance premiums written, followed by supplementary insurance (insurance accident and sickness) and insurance for the case death. The non-life segment consists of a total of 15 non-life and 18 universal insurance companies. The structure of the non-life insurance market covers motor vehicle liability insurance, property insurance, accident insurance, general liability insurance, etc.

5 Results

We evaluated the data by the method described above and present the average efficiencies of the companies in Table 1. As mentioned in the comment to the model (1), efficient are those units with the score $r \geq 1$. The companies with the mean scores above one are highlighted.

After the computation of the scores, we have further analyzed the efficiencies of insurance companies to test the influence of specific factors on the companies performance. In the survey [11], the authors provide a fairly exhaustive list of factors influencing efficiency, among the frequently analyzed factors are: organizational structure, corporate governance, mergers and acquisitions, deregulation and many others. Since the Czech insurance market is relatively homogeneous in terms of the origin of capital our work did not deal with the analysis of the impact of the organizational structure. We also do not identify a sufficient number of mergers and acquisitions to ascertain the degree of influence of efficiency by this factor. Thus, as a part of the secondary analysis, we address the impact of specialization of insurance business on efficiency.

For the life and non-life segments, we compared results of specialized and universal insurance companies. We analyze whether the larger universal insurance companies achieve higher efficiency rates than narrowly specialized smaller entities within the life and non-life segment. We use nonparametric Mann-Whitney U-test comparing medians of efficiencies in the individual groups. Table 2 shows the values of the test statistic and the p-value for given comparisons. For life insurance, we reject the null hypothesis, which implies the conclusion that pure life insurers are more efficient than universal life insurers. For the non-life insurance segment, we do not reject the hypothesis, that universal insurers are more efficient than pure non-life insurers and therefore no clear conclusions can be drawn for the non-life segment. We also test the hypothesis of a higher efficiency rate of the non-life segment

company	type	mean	std	company	type	mean	std
aegon	L	0.9636	0.0569	egap	N-L	1.4104	0.5386
aig	N-L	0.9269	0.0108	euler	N-L	1.3353	0.5756
allianz	U	0.9999	0.0276	gerling	N-L	1.0603	0.1104
amcico metlife	U	0.9270	0.1315	gp	U	0.9558	0.0523
aviva	L	1.0090	0.3723	halali	N-L	1.0480	0.0239
axa	N-L	1.0204	0.3650	hdi	N-L	0.9863	0.0513
axa zp	U	0.9594	0.0788	hvp	U	0.8887	0.0247
basler	L	1.0184	0.0176	ing nn	L	1.1161	0.2314
BNP cardif	U	1.0823	0.2033	koop	U	1.1259	0.2844
ckp	N-L	1.0593	0.1526	kp	U	1.1044	0.2209
colonnade	N-L	0.9858	0.1376	maxima	U	1.0707	0.4398
cp	U	1.1825	0.3426	pcs	U	1.0953	0.2437
cp zdravi	N-L	0.9733	0.0411	pvzp	U	0.8838	0.0471
cpp	U	0.9320	0.0480	slavia	U	0.9116	0.0666
csobp	U	0.9946	0.0525	triglav direct	N-L	1.0936	0.3246
d.a.s.	N-L	0.935	0.0378	uniqa	U	0.9081	0.0396
direct	N-L	0.7852	0.1368	victoria ergo	U	0.8206	0.0337
dr leben	L	1.0670	0.0677	wust	N-L	0.9792	0.0530
ecp erv	N-L	0.9975	0.0966	wust pob	U	1.4704	0.7489
				wust zp	L	0.9450	0.0297

Table 1 Efficiencies of insurance companies (for their full names see [3])

relative to the life segment. The test did not reject the null hypothesis and thus we cannot say that the leading position of the non-life of the Czech insurance market is accompanied by higher efficiency values compared to life insurance.

life vs.	universal	non-life	universal	life vs.	non-life
M-W U	p-value	M-W U	p-value	M-W U	p-value
4871	0.01256	13688	0.2158	46395	0.7982

Table 2 Mann-Whitney U-test for the comparisons

6 Conclusion

In our work, we have dealt with the evaluation of efficiency of decision making units in life and non-life segment of the insurance market. The results of the efficiency evaluation showed that a high level of efficiency is achieved by both smaller, specialised insurers and large market players. Similar conclusions though applied only to the life insurance segment, were presented by Diacon et al. in 2002 [5]. In the evaluation of the segments of the Czech insurance market by the ČAP or the Czech national bank, the non-life segment has been predominant, especially in recent years. Due to this fact, we therefore assumed a higher level of efficiency of the non-life segment. However, there was no statistically significant difference between the efficiency of the life and non-life segments. The results of the assessment of the individual segments even indicate that in the non-life insurance market, no single insurer performs efficiently through all the years. The practical reason why the outputs of insurance companies do not sufficiently cover their inputs may be due to the high competition in the non-life insurance market or low returns on investment activity. For the life segment, the results of the analysis suggest that better results are acquired by insurers specialized in the life segment compared to universal life insurers. The opposite results are suggested by the study of Grmanová and Pukala, [9] which assesses the life insurance market in Poland and the Czech Republic for the year 2018. Also their analysis for the year 2015, [8], suggests that universal insurance companies are the ones that are more efficient in the life insurance market. However, their study uses a different approach and model parameters, moreover, it assesses different dataset for only one year, which may explain the differences in the

conclusions of the papers. The contribution of our paper may be seen in the covering of a broader time span and the use of a robust universal DEA model for the evaluation of efficiencies.

References

- [1] Czech National Bank database ARAD [online] [cit. 2020-07-18] Available from: https://www.cnb.cz/cs/statistika/menova_bankovni_stat/bankovni-statistika/
- [2] Biener, C., Eling, M. & Wirfs, J. H. (2016). The determinants of efficiency and productivity in the swiss insurance industry. *European journal of operational research*, 248, (2), 703–714
- [3] ČAP. Česká asociace pojišťoven: individuální výsledky členů 2006-2019 [online] [cit. 2020-07-18]. Available from: <https://www.cap.cz/statistiky-prognozy-analyzy/individualni-vysledky-clenu>
- [4] Cummins, J.D. & Weiss, M. A. (2013). *Handbook of insurance*, Springer
- [5] Diacon, S.R., Starkey, K. & O'Brien, C., (2002). Size and Efficiency in European Longterm Insurance Companies: An International Comparison. *The Geneva Papers on Risk and Insurance. Issues and Practice* 27, (3) [online] [cit. 2020-10-24].
- [6] Eling, M. & Luhnen, M. (2010). Efficiency in the international insurance industry: a cross-country comparison. *Journal of banking and finance*, 34 (7), 1497–1509
- [7] Grmanová, E. & Jablonský, J. (2009). Analýza efektivnosti slovenských a českých poisťovní pomocou modelov analýzy obalu dát. *Ekonomický časopis*, 57 (9), 857–869
- [8] Grmanova, E. (2015). Efficiency in two-stage data envelopment analysis: an application to insurance companies. In: Kajurova, V., Krajicek, J. (Eds.), *Proceedings of the 12th international scientific conference* 158–165.
- [9] Grmanová, E. & Pukala, R., (2018). Efficiency of insurance companies in the Czech Republic and Poland. *Oeconomia copernicana*, 9 (1), 71–85
- [10] Hladík, M., (2019). Universal efficiency scores in data envelopment analysis based on a robust approach. *Expert systems with applications*, 122, 242–252.
- [11] Kaffash, S., Azizi, R., Huang, Y. & Zhu, J. (2020). A survey of data envelopment analysis applications in the insurance industry 1993–2018. *European journal of operational research*, 284 (3), 801–813
- [12] Lovell, C.A. & Knox, P. T. (1995). Units invariant and translation invariant dea models. *Operations research letters*, 18 (3), 147–151
- [13] Masárová, J. & Koišová, E. (2016). Insurance market in Slovak Republic and Czech Republic. *International multidisciplinary scientific conference on social sciences*, 353–360
- [14] Wise, W. (2017). A survey of life insurance efficiency papers: methods, pros & cons, trends. *Accounting*, 3 (3), 137–170

Stochastic reference point in the evaluation of risky decision alternatives

Ewa Michalska¹, Renata Dudzińska-Baryła²

Abstract. In the first-generation (PT) and second-generation (CPT) prospect theory introduced by Kahneman and Tversky, a non-stochastic reference point is assumed in the evaluation of risky alternatives. In the third-generation prospect theory, Schmidt, Starmer, and Sugden took into consideration, besides well-known elements of the prospect theory, the uncertainty of the reference point. In this paper, we propose a method for the evaluation of risky decision alternatives using a random variable with a discrete distribution (another risky alternative) as a stochastic reference value. We provide the examples on the basis of which we demonstrate that the ordering of decision alternatives can depend on whether full information on the distribution is taken into account or not. The change of ordering (and also preferred alternative) can be observed regardless of considered criteria, subjective or objective, for example, stochastic dominance, Omega ratio, Omega-PT ratio, PT, and CPT.

Keywords: stochastic reference point, ordering of alternatives, prospect theory, Omega ratio, stochastic dominance, threshold level

JEL Classification: G40

AMS Classification: 91B06

1 Introduction

The concept of a benchmark is a well-known concept in decision-making theory and practice. Benchmark (also known as reference value, reference point) means any value against which a decision alternative is evaluated. The reference value expresses the individual preferences of the decision-maker, such as e.g. the preferred portfolio rate of return or the desired outcome of a decision. However, the assessment of a risky decision alternative can depend not only on the adopted reference value but also on the measure used (objective or subjective). In the theories considered today, it is assumed that emotions and subjective perception of information are inherent in decision-making process. Therefore, decision support methods used in practice increasingly often take into account behavioural elements (including reference value), thus becoming part of the stream of modern economics, management and finance. These include the PT [4] or CPT [9] evaluation proposed in the prospect theory as well as performance ratios such as the omega ratio [7] or omega-PT [6].

In the third generation prospect theory proposed in 2008 by Schmidt, Stramer and Sugden, the uncertainty of the reference value is taken into account in addition to the basic assumptions of prospect theory [8]. In decision-making practice, however, an approach is used in which the reference point being a random variable is replaced by a fixed value e.g. the expected value of a random variable or its certainty equivalent. This approach leads to a loss of information on the distribution, which may cause a reordering of the preferred decision alternatives.

The aim of this paper is to propose a method of taking into account (in the evaluation of risky decision alternatives) a reference value that is a random variable with a discrete distribution, which does not lead to a loss of information about the distribution. The paper also shows (through a case study method) that the order of preferred decision alternatives can depend on whether the reference value is a random variable or a fixed value (the expected value of a random variable). The order is determined based on both objective and subjective evaluation criteria.

2 Reference value in decision making

Risky decisions occur when the outcomes of decisions are uncertain and the decision problem is to choose one of many alternatives. In the normative approach, when the probability distribution of decision outcomes is known, a choice rule based on the concept of expected value or expected utility is used. In 1944, von Neumann and Morgenstern provided the axioms of utility theory and showed that the concept is consistent with the preferences of

¹ University of Economics in Katowice, ul. 1 Maja 50, 40-287 Katowice, Poland, ewa.michalska@ue.katowice.pl.

² University of Economics in Katowice, ul. 1 Maja 50, 40-287 Katowice, Poland, renata.dudzinska-baryla@ue.katowice.pl.

a rational decision-maker [10]. However, further empirical studies on decision-making criteria revealed that decision-makers do not always use the principles of expected utility theory.

The next step in studies on decision making under risk was the paper published in 1979 by Kahneman and Tversky [4]. In their concept called prospect theory, attention is paid to the context of the decision, and the possible outcomes of the decision are related to an assumed reference point (a fixed value, e.g. a state of wealth) and expressed as relative gains and losses. The decision-maker subjectively evaluates gains and losses (relative outcomes), showing loss aversion and risk aversion in the face of gains and risk seeking in the face of losses. The probabilities of relative outcomes are also subjectively evaluated.

In the third generation prospect theory proposed in 2008 by Schmidt, Starmer and Sugden, in addition to the characteristic elements present in the previous generation prospect theory, the randomness of the reference value, which can be another risky decision alternative, is assumed [8]. The need to take into account the randomness of the reference value has empirical and theoretical justification. In the evaluation of risky decision alternatives (e.g. lotteries, stocks, investment funds) the adoption of a fixed value of the reference point does not always correspond to reality. In practice, the role of the benchmark is often played by financial instruments that are random in nature. For investment funds, the benchmark portfolio can be, for example, a stock market index [1].

Under risk conditions, the decision alternative X is a random variable of the form of

$$X = ((x_1; p_1), (x_2; p_2), \dots, (x_n; p_n))$$

where: x_i means the absolute outcome, and p_i the corresponding probability for $i = 1, 2, \dots, n$. Similarly, the random reference alternative L is a random variable

$$L = ((l_1; h_1), (l_2; h_2), \dots, (l_m; h_m))$$

where: l_j means reference value, and h_j the corresponding probability for $j = 1, 2, \dots, m$.

Taking into account the reference alternative L by replacing it with the expected value $E(L)$, used in decision-making practice (in assessing the risky decision alternative X against the another risky decision alternative L), leads to a decision alternative of the form of

$$X - E(L) = ((x_1 - E(L); p_1), (x_2 - E(L); p_2), \dots, (x_n - E(L); p_n))$$

where: $(x_i - E(L))$ denotes relative outcome and p_i the corresponding probability for $i = 1, 2, \dots, n$. The consequence of such a procedure is the loss of information on the distribution of the random variable L , nevertheless it is widely used e.g. in various types of performance ratios [3, 5, 7].

In the approach proposed in this paper, we assume that the random variables X and L are independent. In assessing the risky decision alternative X against the another risky decision alternative L we take into consideration the relative outcomes $z_k = x_i - l_j$ and the corresponding probabilities of the form of $q_k = p_i \cdot h_j$ for $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m$, $k = 1, 2, \dots, n \cdot m$. These elements form a new random decision alternative

$$Z = X - L = ((z_1; q_1), (z_2; q_2), \dots, (z_{n \cdot m}; q_{n \cdot m})),$$

whose ordered outcomes of $z_1 \leq z_2 \leq \dots \leq z_{n \cdot m}$ represent gains or losses relative to the reference values.

3 Representation of a random reference alternative in evaluation of decision alternatives

Taking into account the randomness of the reference point makes it possible to model the decision-making situations in which instability of the decision-maker's revealed preferences is observed. The manner in which we take into account the random reference alternative in the evaluation of risky decision alternatives (random variable or fixed value) can influence the order of the preferred decision alternatives. In examples illustrating this fact, the risky decision alternatives were evaluated on the basis of selected objective and subjective criteria. For the decision alternatives X and Y and a random reference alternative represented by the random variable L the preference relation " $>_L$ " was determined on the basis of the following criteria:

Criterion 1. Maximisation of the omega performance ratio

$$X >_L Y \Leftrightarrow \Omega_X(L) > \Omega_Y(L)$$

where

$$\Omega(L) = \frac{\sum_{k, z_k \geq 0} z_k \cdot q_k}{\sum_{k, z_k < 0} z_k \cdot q_k}$$

Criterion 2. Maximisation of PT value

$$X \succ_L Y \Leftrightarrow PT_X(L) > PT_Y(L)$$

where

$$PT(L) = \sum_{k, z_k < 0} v(z_k) \cdot w(q_k) + \sum_{k, z_k \geq 0} v(z_k) \cdot w(q_k)$$

Criterion 3. Maximisation of CPT value

$$X \succ_L Y \Leftrightarrow CPT_X(L) > CPT_Y(L)$$

where

$$CPT(L) = \sum_{k, z_k < 0} v(z_k) \cdot W_k^- + \sum_{k, z_k \geq 0} v(z_k) \cdot W_k^+$$

Criterion 4. Maximisation of the omega-PT ratio

$$X \succ_L Y \Leftrightarrow \Omega PT_X(L) > \Omega PT_Y(L)$$

where

$$\Omega PT(L) = \frac{\sum_{k, z_k \geq 0} v(z_k) \cdot w(q_k)}{\sum_{k, z_k < 0} v(z_k) \cdot w(q_k)}$$

In criteria 1-4, the z_k values are relative results and q_k the corresponding probabilities. In turn, the $v(\cdot)$ function is a Kahneman-Tversky value function of the form of

$$v(z_k) = \begin{cases} z_k^{0,88} & z_k \geq 0 \\ -2,25(-z_k)^{0,88} & z_k < 0 \end{cases}$$

whereas $v(0) = 0$ and the function $w(\cdot)$ is a probability weighting function of the form

$$w(q_k) = \frac{(q_k)^\gamma}{((q_k)^\gamma + (1 - q_k)^\gamma)^{1/\gamma}}$$

where $\gamma = 0,61$ for gains and $\gamma = 0,69$ for losses, moreover $w(0) = 0$, $w(1) = 1$. Symbols W_i^- , W_i^+ mean weights depending on the evaluation of cumulative probabilities [2, 9].

In an analogous way, based on criteria 1-4, the " $\succ_{E(L)}$ " preference relation for the decision alternatives X and Y and a random reference alternative represented by the expected value $E(L)$ is defined.

Example 1. Suppose that we have two random decision alternatives:

$$X = ((20; 0.1), (40; 0.7), (60; 0.2))$$

$$Y = ((10; 0.2), (50; 0.4), (110; 0.4))$$

and the reference alternative $L = ((30; 0.9), (50; 0.1))$, whose expected value is $E(L) = 32$. Moreover, $\Omega_X(L) = 6.26$, $\Omega_Y(L) = 8.73$, $\Omega_X(E(L)) = 9.33$, $\Omega_Y(E(L)) = 8.73$.

Based on the calculated values of the omega ratio (criterion 1) for the decision alternatives X and Y the following relationships were obtained $\Omega_X(L) < \Omega_Y(L)$ and $\Omega_X(E(L)) > \Omega_Y(E(L))$. Thus, if in the evaluation of decision-making alternatives we consider the reference alternative L as a random variable, then $Y \succ_{\Omega(L)} X$, and if we replace it by the expected value $E(L)$, then $X \succ_{\Omega(E(L))} Y$.

Example 2. Consider random decision alternatives:

$$X = ((10; 0.2), (70; 0.5), (90; 0.3))$$

$$Y = ((10; 0.1), (60; 0.6), (90; 0.3))$$

and the reference alternative $L = ((20; 0.5), (50; 0.5))$, whose expected value is $E(L) = 35$. Moreover, $PT_X(L) = 15.78$, $PT_Y(L) = 13.77$, $PT_X(E(L)) = 10.61$, $PT_Y(E(L)) = 12.37$.

Based on the calculated PT values (criterion 2), it was stated that $PT_X(L) > PT_Y(L)$ and $PT_X(E(L)) < PT_Y(E(L))$. Therefore, if in the evaluation of the decision alternatives X and Y we consider the reference alternative L as a random variable, then $X \succ_L Y$ but if we replace it by the expected value $E(L)$, then $Y \succ_{E(L)} X$.

Example 3. Suppose that we have two random decision alternatives:

$$X = ((20; 0.5), (70; 0.2), (110; 0.3))$$

$$Y = ((10; 0.1), (50; 0.5), (70; 0.4))$$

and the reference alternative $L = ((30; 0.9), (50; 0.1))$ whose expected value is $E(L) = 32$. Moreover, $CPT_X(L) = 6.54$, $CPT_Y(L) = 6.84$, $CPT_X(E(L)) = 8.14$, $CPT_Y(E(L)) = 7.62$.

On the basis of the CPT values (criterion 3) calculated for the decision alternatives X and Y , it was stated that $CPT_X(L) < CPT_Y(L)$ and $CPT_X(E(L)) > CPT_Y(E(L))$. Thus, if in the evaluation of the decision alternatives we take into consideration the reference alternative L as a random variable, then $Y \succ_L X$, and if we replace it by the expected value $E(L)$, then $X \succ_{E(L)} Y$.

Example 4. We have two random decision alternatives:

$$X = ((20; 0.5), (60; 0.4), (70; 0.1))$$

$$Y = ((10; 0.3), (50; 0.2), (70; 0.5))$$

and the reference alternative $L = ((30; 0.9), (60; 0.1))$ whose expected value is $E(L) = 33$. Moreover, $\Omega PT_X(L) = 0.87$, $\Omega PT_Y(L) = 0.89$, $\Omega PT_X(E(L)) = 1.15$, $\Omega PT_Y(E(L)) = 1.14$.

The values of the omega-PT ratio (criterion 4) calculated for the decision alternatives X and Y lead to the following inequalities: $\Omega PT_X(L) < \Omega PT_Y(L)$ and $\Omega PT_X(E(L)) > \Omega PT_Y(E(L))$. If in the evaluation of the decision alternatives we take into consideration the reference alternative L as a random variable, then $Y \succ_L X$ and if we replace it by the expected value $E(L)$, we have $X \succ_{E(L)} Y$.

The PT, CPT and omega and omega-PT performance ratios used in the examples are decreasing functions of the reference value. An increase in the reference value results in a decrease in the value of the ratio. For the decision alternative X and two random reference alternatives $L1$ and $L2$, such that $E(L1) > E(L2)$, the following inequalities occur:

$$\Omega_X(E(L1)) < \Omega_X(E(L2))$$

$$\begin{aligned}
PT_X(E(L1)) &< PT_X(E(L2)) \\
CPT_X(E(L1)) &< CPT_X(E(L2)) \\
\Omega PT_X(E(L1)) &< \Omega PT_X(E(L2)).
\end{aligned}$$

The analogous property is observed for the omega ratio and CPT value, when (in the evaluation of decision alternatives) the reference alternatives are random variables ordered based on the first-order stochastic dominance (FSD). For the decision alternative X and two random reference alternatives $L1$ and $L2$, such that $L1 \succ_{FSD} L2$, the following inequalities occur:

$$\begin{aligned}
\Omega_X(L1) &< \Omega_X(L2) \\
CPT_X(L1) &< CPT_X(L2)
\end{aligned}$$

These properties are illustrated by the following example.

Example 5. Let us consider two random decision alternatives $X = ((10; 0.2), (70; 0.5), (90; 0.3))$ and $Y = ((10; 0.1), (60; 0.6), (90; 0.3))$ and three reference alternatives: $L1 = ((25; 0.3), (55; 0.7))$, $L2 = ((20; 0.5), (50; 0.5))$ and $L3 = ((19; 0.6), (49; 0.4))$, whose expected values are $E(L1) = 46$, $E(L2) = 35$, $E(L3) = 31$. The reference alternatives $L1$, $L2$ and $L3$ were ordered according to the first-order stochastic dominance and the following preference relations were obtained $L1 \succ_{FSD} L2 \succ_{FSD} L3$. Moreover, their expected values satisfy the condition $E(L1) > E(L2) > E(L3)$. Table 1 shows the values of PT, CPT and omega and omega-PT performance ratios for different forms of reference value i.e. random variable and expected value of random variable.

Criterion	$(X - L)$ for			$(Y - L)$ for		
	$L = L1$	$L = L2$	$L = L3$	$L = L1$	$L = L2$	$L = L3$
Omega	3.50	6.80	8.86	6.00	12.60	16.71
CPT	0.33	6.91	9.24	3.95	10.24	12.45

Table 1 The evaluations of decision alternatives X and Y for random reference alternative L represented by random variables $L1$, $L2$, $L3$

Criterion	$(X - E(L))$ for			$(Y - E(L))$ for		
	$L = L1$	$L = L2$	$L = L3$	$L = L1$	$L = L2$	$L = L3$
Omega	3.50	6.80	8.86	6.00	12.60	16.71
CPT	0.09	7.60	10.35	3.94	11.00	13.55

Table 2 The evaluations of decision alternatives X and Y for random reference alternative L represented by the expected values of random variables $L1$, $L2$, $L3$

Regardless of how the random reference alternative is represented (random variable or the expected value of random variable) in the evaluation of random decision alternatives, all considered evaluations are decreasing.

The regularities presented in the examples were confirmed in simulation studies performed for the three-outcome decision and reference alternatives. Random reference alternatives satisfying the first-order stochastic dominance (FSD) and random decision alternatives were generated (60000 triples $(L1, L2, X)$). For all replications the considered relations are satisfied.

4 Conclusions

In recent years, there has been growing interest in decision support methods that reflect the behavioural inclinations that characterise the decision-maker. New measures and ratios that fulfil these requirements are constantly being

proposed in the literature, and much attention is devoted to the study of their properties. In this paper we consider the selected evaluations/ratios (PT, CPT, Ω , Ω PT) that allow to take into account both a reference value and full information about the distribution of evaluated random decision alternatives. Taking into account the aforementioned decision alternative evaluations as functions of the reference point represented by the expected value, it was found that the increase in the value of the reference point causes a decrease in the value of the random decision alternative evaluations. By analysing the relationship between the evaluations of the random decision alternatives, we can also conclude that after changing the value of the reference point, a different decision alternative than before can be preferred. Moreover, the reordering of the preferred decision alternatives can occur repeatedly with the change of the reference point. The specified properties are confirmed in the source literature [2, 6].

Analogous properties were observed in the examples analysed for the omega ratio and CPT evaluation, taking into account the considered decision alternative evaluations as functions of the reference point represented by a random variable. Their generalisation requires evidence or more extensive simulation studies. Nevertheless, the presented examples are sufficient to conclude that the manner for taking into account the random decision alternative (random variable or the expectation value of random variable) in the evaluation of risky decision alternatives influences their ordering. This should be taken into consideration when selecting a representation of the random reference alternative.

References

- [1] Borkowski, K. (2011). *Teoria i praktyka benchmarków*. Warszawa: Wydawnictwo Diffin (in Polish).
- [2] Dudzińska-Baryła, R. (2019). *Subiektywna ocena inwestycji giełdowych – ujęcie ilościowe*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach (in Polish).
- [3] Elton, E.J. & Gruber, M.J. (2010). *Investments and Portfolio Performance*. World Scientific.
- [4] Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47(2), 263–291.
- [5] Michalska, E. & Kopańska-Bródka, D. (2015). The Omega Function for Continuous Distribution. In: D. Martinčík, J. Ircingová & P. Janeček (Eds.), *Conference Proceedings, 33rd International Conference Mathematical Methods in Economics* (pp. 543–548). Plzeň: University of West Bohemia.
- [6] Michalska, E. (2018). *Obiektywna a subiektywna ocena efektywności ryzykownych wariantów decyzyjnych*, Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach (in Polish).
- [7] Shadwick, W. & Keating, C. (2002). A Universal Performance Measure. *Journal of Performance Measurement*, 6(3), 59–84.
- [8] Schmidt, U., Starmer, Ch. & Sugden, R. (2008). Third-generation prospect theory. *Journal of Risk and Uncertainty*, 36(3), 203–223.
- [9] Tversky, A. & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323.
- [10] von Neumann, J. & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Structure of the threshold digraphs of convex and concave Monge matrices in fuzzy algebra

Monika Molnárová¹

Abstract. Properties of threshold digraphs of both convex and concave Monge matrices over fuzzy algebra (max-min algebra) are studied. The threshold digraphs of a concave Monge matrix are shown to have a block structure corresponding to the strongly connected components of the digraph with loops on every node and cycles of length two connecting each pair of consecutive nodes in a non-trivial strongly connected component. An algorithm with polynomial computational complexity to determine the strongly connected components is presented. In contrast to the concave Monge matrices all loops lie in the same strongly connected component in the threshold digraphs of a convex Monge matrix. A cycle with odd length guarantees the existence of a loop in the threshold digraph. The existence of a common cycle of length two for minimal and maximal node in a non-trivial strongly connected component is proved.

Keywords: fuzzy algebra, convex Monge matrix, concave Monge matrix, threshold digraph

JEL Classification: C02

AMS Classification: 08A72, 90B35, 90C47

1 Introduction

To model discrete dynamic systems (DDS) using the concept of extremal algebras is a frequent method for solving optimization problems in many diverse areas ([8], [9]). We have studied Monge matrices, their structural properties and algorithms solving many problems related to Monge matrices in [1]. Results concerning robustness of matrices were presented in [5], [6], [7].

The aim of this paper is to present properties of the threshold digraphs of convex and concave Monge matrices over max-min algebra. Obtained results can be useful for investigating properties of Monge matrices related to problems as periodicity or robustness and consequently for deriving efficient algorithms.

We briefly outline the content and main results of the paper. Section 2 provides the necessary preliminaries on max-min algebra and the notion of a convex Monge matrix and a concave Monge matrix is introduced. In Section 3, results concerning properties of threshold digraphs of convex Monge matrices is presented. The key result of this section is Theorem 3, which proves the existence of a common cycle of length two for minimal and maximal node in a non-trivial strongly connected component. In Section 4, results concerning properties of threshold digraphs of concave Monge matrices is presented. The key results of this section are Theorem 4, which proves the existence of the loop on every node in a non-trivial strongly connected component, Theorem 6, which proves the existence of cycles of length two connecting each pair of consecutive nodes in a non-trivial strongly connected component, and Theorem 7, which presents the polynomial algorithm for finding the strongly connected components.

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. 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., $N = \{1, 2, \dots, n\}$.

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 to \oplus , \otimes) as matrix operations over any field.

¹ Technical University of Košice, Department of Mathematics and Theoretical Informatics, B. Němcovej 32, 04200 Košice, Slovakia, Monika.Molnarova@tuke.sk

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 $E' = \{(i, j); i, j \in N, a_{ij} \geq h\}$. 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} . By $\text{SCC}^*(G)$ we denote the set of all non-trivial strongly connected components of G .

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

$$\begin{aligned} a_{ij} \otimes a_{kl} &\leq a_{il} \otimes a_{kj} && \text{for all } i < k, j < l \\ (a_{ij} \otimes a_{kl} &\geq a_{il} \otimes a_{kj} && \text{for all } i < k, j < l). \end{aligned}$$

Obviously it is enough to consider thresholds $h \in H = \{a_{ij}; i, j \in N\}$ to get all threshold digraphs corresponding to the matrix A .

3 Properties of threshold digraphs of convex Monge matrices

In this section we present results related to structure of threshold digraphs of convex Monge matrices. Every cycle in a non-trivial strongly connected component is a concatenation of cycles of length one and two. Consequently there is a node with loop in the cycle of odd length. Moreover, all loops in a threshold digraph lie in the same non-trivial strongly connected component. In addition, we prove the existence of a common cycle for minimal and maximal node in a non-trivial strongly connected component of a threshold digraph.

Theorem 1. [3] Let $A \in B(n, n)$ be a convex Monge matrix. Let $\mathcal{K} \in \text{SCC}^*(G(A, h))$ for $h \in H$. 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.

Corollary 1. [3] Let $A \in B(n, n)$ be a convex Monge matrix. Let $\mathcal{K} \in \text{SCC}^*(G(A, h))$ for $h \in H$. 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.

Theorem 2. [2] Let $A \in B(n, n)$ be a convex Monge matrix. Let for $i, k \in N$ be the loops (i, i) and (k, k) in the digraph $G(A, h)$ for $h \in H$. Then the nodes i and k are in the same non-trivial strongly connected component \mathcal{K} of $G(A, h)$.

Theorem 3. Let $A \in B(n, n)$ be a convex Monge matrix. Let $\mathcal{K} \in \text{SCC}^*(G(A, h))$ for $h \in H$ be generated by the node set $N_{\mathcal{K}}$. Then there is cycle (t, u, t) for $t = \min N_{\mathcal{K}}$ and $u = \max N_{\mathcal{K}}$ in \mathcal{K} .

Proof. Since \mathcal{K} is a non-trivial strongly connected component of $G(A, h)$ for $u = t$ the statement trivially follows. Let $t < u$. Let us assume there is no arc (t, u) in \mathcal{K} , i.e. $a_{tu} < h$, and let $l \in N_{\mathcal{K}}$ be the maximal index for which $a_{tl} \geq h$. Hence $l < u$. Since $u \in N_{\mathcal{K}}$, there exists $k \in N_{\mathcal{K}} \setminus \{t, u\}$ with $a_{ku} \geq h$. Using the Monge property of the matrix A holds

$$h \leq a_{tl} \otimes a_{ku} \leq a_{tu} \otimes a_{kl}.$$

Consequently $a_{tu} \geq h$, i.e. there is an arc (t, u) in \mathcal{K} what is a contradiction. Now, let us assume there is no arc (u, t) in \mathcal{K} , i.e. $a_{ut} < h$, and let $k \in N_{\mathcal{K}}$ be the minimal index for which $a_{uk} \geq h$. Hence $k > t$. Since $t \in N_{\mathcal{K}}$, there exists $l \in N_{\mathcal{K}} \setminus \{t, u\}$ with $a_{lt} \geq h$. By the Monge property of the matrix A holds

$$h \leq a_{lt} \otimes a_{uk} \leq a_{ut} \otimes a_{lk}.$$

Consequently $a_{ut} \geq h$, i.e. there is an arc (u, t) in \mathcal{K} what is a contradiction. Hence there is a cycle (t, u, t) in \mathcal{K} . \square

Example 1. Let us check the structure of the threshold digraphs of the given convex Monge matrix $A \in B(8, 8)$ for $B = [0, 3]$

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 2 & 3 & 1 \\ 0 & 0 & 0 & 0 & 2 & 3 & 2 & 1 \\ 0 & 0 & 0 & 3 & 3 & 2 & 2 & 0 \\ 0 & 0 & 3 & 3 & 3 & 0 & 0 & 0 \\ 0 & 3 & 3 & 2 & 0 & 0 & 0 & 0 \\ 0 & 3 & 2 & 1 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

There is one non-trivial strongly connected component in $G(A, 1)$, there are two non-trivial strongly connected components in $G(A, 2)$ and there are three non-trivial and two trivial strongly connected components in $G(A, 3)$. Due to Theorem 1 every cycle in a non-trivial strongly connected component is a concatenation of cycles of length one and two. According to Theorem 2 all loops of $G(A, h)$ for $h \in H$ lie in the same non-trivial strongly connected component. Moreover, by Theorem 3 there is a common cycle of length two for minimal and maximal node in all non-trivial strongly connected components (see Figure 1).

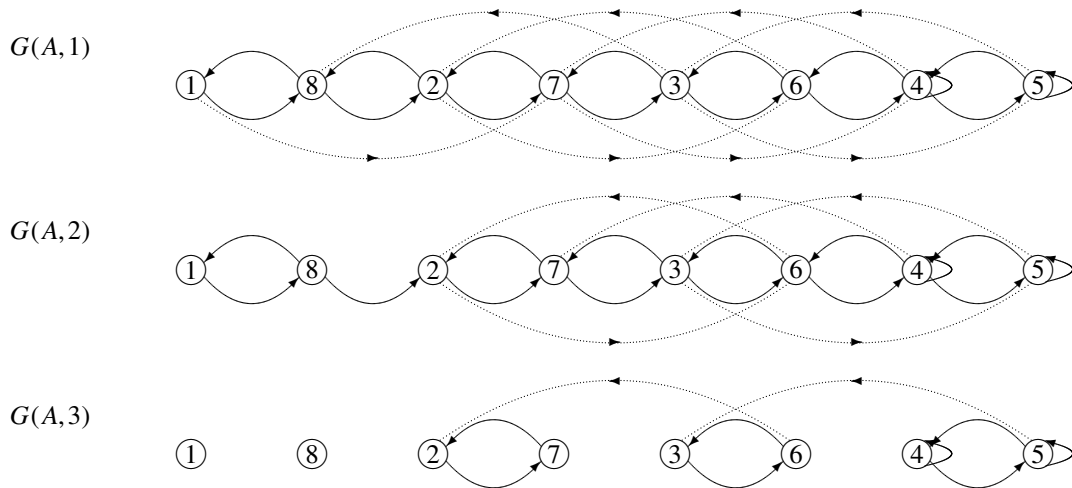


Figure 1 Strongly connected components in threshold digraphs of a convex Monge matrix

4 Properties of threshold digraphs of concave Monge matrices

In this section we present results related to structure of threshold digraphs of concave Monge matrices. We prove the existence of loops on every node in a non-trivial strongly connected component. We show that the nodes of a non-trivial strongly connected component are numbered by a sequence of consecutive natural numbers. Consequently a concave Monge fuzzy matrix has a block form, where the diagonal blocks represent the strongly connected components. Moreover, we prove the existence of cycles of length two for each pair of consecutive nodes in a non-trivial strongly connected component. Finally, we present an algorithm with computational complexity $O(n^3)$ for finding all non-trivial strongly connected components of $G(A, h)$ for all $h \in H$.

Theorem 4. 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 .

Proof. Let c be a cycle in $G(A, h)$ for $h \in H$. The proof proceeds by mathematical recursion on $\ell(c)$. If $\ell(c) = 1$, then the assertion is trivially fulfilled. Let us assume that $\ell(c) = k > 1$ and the assertion is true for all cycles of length less than k .

Let $c_1 = (i_0, i_1, \dots, i_{k-1}, i_k)$ with $i_0 = i_k$ and let $i_s = \max\{i_0, i_1, \dots, i_{k-1}\}$. If $i_{s-1} = i_s$, then the arc $i_{s-1} \rightarrow i_s$ is a loop and clipping out this arc from the cycle c we get a cycle c' of length $\ell(c') < k$ and the assertion follows from the recursion assumption. By analogy we can deal with the case when $i_{s+1} = i_s$.

Thus, we may assume that the nodes i_{s-1} and i_{s+1} are different from i_s . Since i_s is maximal it follows that $i_{s-1} < i_s$ and $i_{s+1} < i_s$ and we can use the Monge property of the matrix A

$$a_{i_{s-1}i_s} \otimes a_{i_s i_{s+1}} \leq a_{i_{s-1}i_{s+1}} \otimes a_{i_s i_s}.$$

Consequently for the cycle c in the threshold digraph $G(A, h)$ for $h \in H$ holds

$$\begin{aligned} h &\leq a_{i_0 i_1} \otimes a_{i_1 i_2} \otimes \dots \otimes a_{i_{s-1} i_s} \otimes a_{i_s i_{s+1}} \otimes \dots \otimes a_{i_{k-1} i_k} \leq \\ &\leq a_{i_0 i_1} \otimes a_{i_1 i_2} \otimes \dots \otimes a_{i_{s-1} i_{s+1}} \otimes a_{i_s i_s} \otimes \dots \otimes a_{i_{k-1} i_k}. \end{aligned} \quad (1)$$

Then $a_{i_s i_s}$ represents a loop on the node i_s in $G(A, h)$. Moreover, by (1) we can clip out the arcs (i_{s-1}, i_s) and (i_s, i_{s+1}) from the cycle c , replacing them by arc (i_{s-1}, i_{s+1}) and we get a cycle c' of length $\ell(c') < k$ and the assertion follows from the recursion assumption. \square

Corollary 2. Let $A \in B(n, n)$ be a concave Monge matrix. Let $\mathcal{K} \in \text{SCC}^*(G(A, h))$ for $h \in H$. Then there is a loop on every node in \mathcal{K} .

Proof. By definition of a non-trivial strongly connected component there exists a cycle connecting all nodes of the component \mathcal{K} . The assertion follows by Theorem 4. \square

Corollary 3. Let $A \in B(n, n)$ be a concave Monge matrix. Element $a_{kk} < h$ represents a trivial strongly connected component of $G(A, h)$.

Proof. $a_{kk} < h$ implies that there is no loop on the node k in $G(A, h)$. Hence, by Corollary 2 there is no cycle in $G(A, h)$ containing node k . \square

Now, we can reformulate Corollary 2.

Corollary 4. 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 c can split in \mathcal{K} into finite number of cycles of length one.

Theorem 5. [4] Let $A \in B(n, n)$ be a concave Monge matrix. Let $c_1 = (i_0, i_1, \dots, i_k)$ with $i_0 = i_k$ and $c_2 = (j_0, j_1, \dots, j_l)$ with $j_0 = j_l$ be cycles in different non-trivial strongly connected components in $G(A, h)$ for $h \in H$. Let $i_s = \min\{i_0, i_1, \dots, i_{k-1}\}$ and $i_t = \max\{i_0, i_1, \dots, i_{k-1}\}$. Then exactly one of the conditions holds

- (i) $j_m < i_s$ for all $m \in \{0, 1, \dots, l\}$,
- (ii) $j_m > i_t$ for all $m \in \{0, 1, \dots, l\}$.

Corollary 5. Let $A \in B(n, n)$ be a concave Monge matrix. Let \mathcal{K}_1 and \mathcal{K}_2 be two different non-trivial strongly connected components of $G(A, h)$ for $h \in H$ generated by the node set $N_{\mathcal{K}_1} = \{i_1, \dots, i_k\}$ and $N_{\mathcal{K}_2} = \{j_1, \dots, j_l\}$, respectively. Let $i_s = \min\{i_1, i_2, \dots, i_k\}$ and $i_t = \max\{i_1, i_2, \dots, i_k\}$. Then exactly one of the conditions holds

- (i) $j_m < i_s$ for all $m \in \{1, 2, \dots, l\}$,
- (ii) $j_m > i_t$ for all $m \in \{1, 2, \dots, l\}$.

Proof. By definition of a non-trivial strongly connected component there exists a cycle connecting all nodes of the component \mathcal{K} . The assertion follows by Theorem 5. \square

Corollary 6. Let $A \in B(n, n)$ be a concave Monge matrix. Then A has a block form in which the diagonal blocks represent the strongly connected components of $G(A, h)$ for $h \in H$.

Proof. The assertion follows by Corollary 5. \square

Theorem 6. Let $A \in B(n, n)$ be a concave Monge matrix. Let i and $i + 1$ be two nodes in a strongly connected component \mathcal{K} of $G(A, h)$ for $h \in H$. Then \mathcal{K} contains the cycle $(i, i + 1, i)$.

Proof. Let \mathcal{K} be a non-trivial strongly connected component of $G(A, h)$ for $h \in H$ generated by the node set $N_{\mathcal{K}}$. Due to Corollary 5 we can assume $N_{\mathcal{K}} = \{m, m + 1, \dots, n\}$, for $m < n$. Let $i, i + 1 \in N_{\mathcal{K}}$. There are two possibilities. First, $N_{\mathcal{K}} = \{i, i + 1\}$, then the assertion follows by strong connectivity of \mathcal{K} . Second, $N_{\mathcal{K}} \neq \{i, i + 1\}$. Hence we can define $N_{\mathcal{K}} = N_{\mathcal{K}}^1 \cup \{i\} \cup N_{\mathcal{K}}^2$, with $N_{\mathcal{K}}^1 = \{m, m + 1, \dots, i - 1\}$ and $N_{\mathcal{K}}^2 = \{i + 1, i + 2, \dots, n\}$. We have two possibilities again.

First, $N_{\mathcal{K}}^1 \neq \emptyset$. The strong connectivity of the component \mathcal{K} implies the existence of arcs (j, k) and (r, l) in \mathcal{K} for $j, l \in N_{\mathcal{K}}^1 \cup \{i\}$ and $r, k \in N_{\mathcal{K}}^2$. Consequently $h \leq a_{jk}$ and $h \leq a_{rl}$. By the Monge property of the matrix A holds

$$h \leq a_{jk} \otimes a_{rl} \leq a_{jl} \otimes a_{rk}.$$

Moreover, by Corollary 2 there is a loop on every node in a strongly connected component of $G(A, h)$ for $h \in H$. Thus $h \leq a_{ii}$ and $h \leq a_{i+1i+1}$. Using by iteration the Monge property of the matrix A and the above inequalities we get gradually the following results

$$h \leq a_{jk} \otimes a_{ii} \leq a_{ji} \otimes a_{ik}$$

$$h \leq a_{rl} \otimes a_{ii} \leq a_{il} \otimes a_{ri}$$

$$h \leq a_{ri} \otimes a_{i+1i+1} \leq a_{i+1i} \otimes a_{ri+1}. \quad (2)$$

$$h \leq a_{ik} \otimes a_{i+1i+1} \leq a_{ii+1} \otimes a_{i+1k}. \quad (3)$$

Consequently by (2) is $h \leq a_{i+1i}$ and by (3) is $h \leq a_{ii+1}$. Hence there is the arc $(i, i+1)$ and the arc $(i+1, i)$ as well in the component \mathcal{K} of $G(A, h)$ for $h \in H$.

Second, $N_{\mathcal{K}}^1 = \emptyset$. The strong connectivity of the component \mathcal{K} implies the existence of arcs (i, j) and (k, i) in \mathcal{K} for $j, k \in N_{\mathcal{K}}^2$. Consequently $h \leq a_{ij}$ and $h \leq a_{ki}$. By the Monge property of the matrix A holds

$$h \leq a_{ij} \otimes a_{ki} \leq a_{ii} \otimes a_{kj}.$$

Similarly to first case we can use the existence of the loop on node $i+1$. Thus $h \leq a_{i+1i+1}$. Using the Monge property of the matrix A , assumptions and the above inequality we get the following results

$$h \leq a_{ij} \otimes a_{i+1i+1} \leq a_{ii+1} \otimes a_{i+1j}. \quad (4)$$

$$h \leq a_{ki} \otimes a_{i+1i+1} \leq a_{i+1i} \otimes a_{ki+1}. \quad (5)$$

Consequently by (4) is $h \leq a_{ii+1}$ and by (5) is $h \leq a_{i+1i}$. Hence there is the arc $(i, i+1)$ and the arc $(i+1, i)$ as well in the component \mathcal{K} of $G(A, h)$ for $h \in H$. \square

Theorem 7. *There is an algorithm with computational complexity $O(n^3)$ for computing the strongly connected components of $G(A, h)$ for all $h \in H$.*

Proof. The computation of all strongly connected components of a threshold digraph $G(A, h)$ for $h \in H$ is based on Corollary 3, Corollary 5 and Theorem 6. It is enough to verify the conditions $a_{ii+1} \geq h$ and $a_{i+1i} \geq h$ to determine the non-trivial strongly connected components and to verify the condition $a_{ii} \leq h$ to determine a trivial strongly connected component. This part takes $O(n)$ time. Since the number of possible inputs of h , i.e. the number of threshold digraphs $G(A, h)$, is given by the number of matrix inputs, namely n^2 , is the total computational complexity $O(n^3)$. \square

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 $B = [0, 3]$

$$A = \begin{pmatrix} 3 & 3 & 2 & 0 & 0 & 0 & 0 & 0 \\ 3 & 3 & 3 & 1 & 0 & 0 & 0 & 0 \\ 1 & 3 & 3 & 3 & 0 & 0 & 0 & 0 \\ 0 & 2 & 3 & 3 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 2 & 2 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 2 & 3 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 1 & 3 & 3 \\ 0 & 0 & 0 & 0 & 0 & 1 & 3 & 3 \end{pmatrix}.$$

Due to Corollary 6 the matrix A has a block form in which the diagonal blocks represent the strongly connected components of $G(A, h)$ for $h \in H$. According to the algorithm described in the proof of Theorem 7 it is enough to check for $a_{ii} \geq h$ the consecutive cycles of length two till at least one of the inequalities $a_{ii+1} \geq h$ and $a_{i+1i} \geq h$

does not hold, i.e. the node set of a non-trivial strongly connected component is completed and the node $i + 1$ belongs to the following component. For $a_{ii} < h$ is the component trivial. Hence there are two non-trivial strongly connected components in $G(A, 1)$ (the corresponding blocks are bounded by single lines in A), there are three non-trivial strongly connected components in $G(A, 2)$ (the corresponding blocks are bounded by single and double lines in A) and there are three non-trivial and one trivial strongly connected component (due to Corollary 3 since $a_{55} < 3$) in $G(A, 3)$ (the corresponding blocks are bounded by single, double and triple lines in A) (see Figure 2). Moreover, due to Corollary 2 there is a loop on every node in all non-trivial strongly connected components of $G(A, h)$ for $h \in H$.

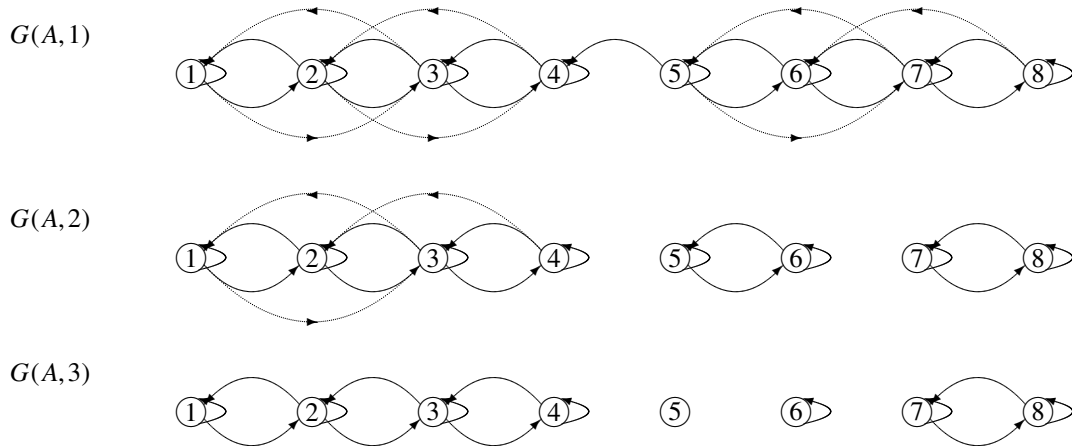


Figure 2 Strongly connected components in threshold digraphs of a concave Monge matrix

5 Conclusion

We have studied structure of threshold digraphs of convex and concave Monge matrices in this paper. Obtained results can be helpful to find efficient algorithms for checking possible and universal robustness of interval Monge matrices over max-min algebra.

The robustness of a matrix A is related to the solution of the eigenproblem $A \otimes x(r) = x(r)$, where an eigenvector $x(r)$ represents the steady state. This can reflect the following economic application. Several projects characterized by different properties should be evaluated in a company. The level of each property i is described by value x_i , influenced by all properties x_j . The influence is represented by a factor a_{ij} .

References

- [1] Burkard, R. E., Klinz, B. and Rudolf, R.: Perspectives of Monge properties in optimization, *DAM*, Volume **70** (1996), 95–161.
- [2] Molnárová, M.: Robustness of Monge matrices in fuzzy algebra, In: *Proceedings of 32nd Int. Conference Mathematical Methods in Economics 2014*, Olomouc, 2014, 679–684.
- [3] Molnárová, M.: Periodicity of convex and concave Monge matrices in max-min algebra, In: *Proceedings of 38th Int. Conference Mathematical Methods in Economics 2020*, Brno, 2020, 377–382.
- [4] Molnárová, M.: Convex and concave Monge matrices in fuzzy algebra - comparison with respect to robustness, (preprint).
- [5] Myšková, H. and Plavka, J.: X-robustness of interval circulant matrices in fuzzy algebra, *Linear Algebra and its Applications*, Volume **438** (2013), 2757–2769.
- [6] Myšková, H. and Plavka, J.: The robustness of interval matrices in max-plus algebra, *Linear Algebra and its Applications*, Volume **445** (2013), 85–102.
- [7] Plavka, J.: The weak robustness of interval matrices in max-plus algebra, *DAM*, Volume **173** (2014), 92–101.
- [8] 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.*, Volume **30** (2020), 25–54.
- [9] Zimmermann, H.J.: Fuzzy Set Theory And Its Applications; Springer Science and Business Media, Berlin, Germany, 2011.

Weak Solvability of Max-plus Matrix Equations

Helena Myšková¹

Abstract. Max-plus algebra is an algebraic structure, in which classical addition and multiplication are replaced by maximum and addition, respectively. Behavior of discrete event systems, in which the individual components move from event to event rather than varying continuously through time, is often described by systems of max-plus linear equations or matrix equations. Discrete dynamic systems can be studied using max-plus matrix operations. It often happens that a max-plus matrix equation with exact data is unsolvable. Therefore, we replace matrix elements with intervals of possible values. In this way, we obtain an interval max-plus matrix equation. Several types of solvability of interval max-plus matrix equations have been studied yet. In this paper, we prove the necessary and sufficient condition for an interval max-plus matrix equation to be strongly tolerance solvable and provide an algorithm for checking whether the given interval matrix equation is strongly tolerance solvable.

Keywords: max-plus algebra, interval matrix, matrix equation, weak solvability

JEL Classification: C02

AMS Classification: 15A18; 15A80; 65G30

1 Motivation

Behaviour of discrete event systems, in which the individual components move from event to event rather than varying continuously through time, is often described by systems of linear equations or by matrix equations. Discrete dynamic systems and related algebraic structures were studied using max-plus matrix operations in [1, 2]. In the last decades, significant effort has been developed to study systems of max-plus linear equations in the form $A \otimes x = b$, where A is a matrix, b and x are vectors of compatible dimensions. Systems of linear equations over max-plus algebra are used in several branches of applied mathematics. Among interesting real-life applications let us mention e.g. a large scale model of Dutch railway network or synchronizing traffic lights in Delfts [8]. In the last two decades, interval systems of the form $A \otimes x = b$ have been studied, for details see [1, 3, 4].

In this paper, we shall deal with interval max-plus matrix equations of the form $A \otimes X = B$, where A, B are given interval matrices of suitable sizes a X is an unknown matrix. Several solvability concepts have been studied in [6], [7]. The following example is a slightly modified version of an example given in [6].

Example 1. Let us consider a situation, in which passengers from places P_1, P_2, \dots, P_m want to transfer to holiday destinations D_1, D_2, \dots, D_r . Suppose that the air transport to the destination D_k is provided by the airport terminal T_k . Different transportation means provide transporting passengers from places P_1, P_2, \dots, P_m to airport terminals T_1, T_2, \dots, T_r . We assume that the connection between P_i ($i \in N$, $N = \{1, 2, \dots, n\}$) and T_k ($k \in R$, $R = \{1, 2, \dots, r\}$) is possible only via one of the check points Q_1, Q_2, \dots, Q_n . Denote by a_{ij} the times needed for transportation and carrying out the formalities on the connection from P_i to Q_j ($j \in N$, $N = \{1, 2, \dots, n\}$). If there is no connection from P_i to Q_j we put $a_{ij} = -\infty$. If the time needed for transportation from place Q_j to terminal T_k is x_{jk} , then the time needed for transportation from P_i to T_k via Q_j is equal to $a_{ij} + x_{jk} + c_{ik}$.

Assume that b_{ik} is the time available to passengers traveling from place P_i to destination D_k to move to terminal T_k . Our task is to choose the appropriate times x_{jk} , $j \in N, k \in R$ so that all passengers get to their airport terminal on time, i.e.,

$$\max_{j \in N} \{a_{ij} + x_{jk}\} = b_{ik}. \quad (1)$$

2 Preliminaries

Max-plus algebra is the triple $(\overline{\mathbb{R}}, \oplus, \otimes)$, where

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}, \quad a \oplus b = \max\{a, b\} \text{ and } a \otimes b = a + b.$$

¹ Technical University in Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, Némcevej 32, 042 00 Košice, Slovak Republic, helena.myskova@tuke.sk

The set of all $m \times n$ matrices over $\overline{\mathbb{R}}$ is denoted by $\overline{\mathbb{R}}(m, n)$ and the set of all column n -vectors over $\overline{\mathbb{R}}$ by $\overline{\mathbb{R}}(n)$. Operations \oplus and \otimes are extended to matrices and vectors in the same way as in the classical algebra. We will consider the ordering \leq on the sets $\overline{\mathbb{R}}(m, n)$ and $\overline{\mathbb{R}}(n)$ defined as follows:

- for $A, C \in \overline{\mathbb{R}}(m, n)$: $A \leq C$ if $a_{ij} \leq c_{ij}$ for each $i \in M$ and each $j \in N$,
- for $x, y \in \overline{\mathbb{R}}(n)$: $x \leq y$ if $x_j \leq y_j$ for each $j \in N$.

We will use the *monotonicity of* \otimes , which means that the inequalities $A \leq C$, $B \leq D$ imply $A \otimes B \leq C \otimes D$.

For given matrix $A \in \overline{\mathbb{R}}(m, n)$ and vector $b \in \overline{\mathbb{R}}(m)$ a system of linear max-plus equation can be written in the form

$$A \otimes x = b. \quad (2)$$

We define the vector $x^*(A, b)$, called a principal solution of (2) as follows:

$$x_j^*(A, b) = \min_{i \in M} \{b_i - a_{ij}\}, \quad j \in N. \quad (3)$$

The main significance of the principal solution is that (2) is solvable if and only if $x^*(A, b)$ is its solution.

Let us replace a vector b in (2) by a matrix $B \in \overline{\mathbb{R}}(m, r)$ and an unknown vector x by an unknown matrix $X \in \overline{\mathbb{R}}(n, r)$. We get the matrix equation of the form

$$A \otimes X = B. \quad (4)$$

It is easy to see that (4) is equivalent to the following r equations:

$$A \otimes X_k = B_k$$

for each $k \in R$, where X_k (B_k) is the k -th column of X (B). Consequently the principal solution of (4) is the matrix $X^*(A, B) \in \overline{\mathbb{R}}(n, r)$ with elements

$$x_{jk}^*(A, B) = x_j^*(A, B_k). \quad (5)$$

Theorem 1. [6] Let $A \in \overline{\mathbb{R}}(m, n)$ and $B \in \overline{\mathbb{R}}(m, r)$ be given.

- If $A \otimes X = B$ for $X \in \overline{\mathbb{R}}(n, r)$, then $X \leq X^*(A, B)$.
- $A \otimes X^*(A, B) \leq B$.
- The matrix equation $A \otimes X = B$ is solvable if and only if the matrix $X^*(A, B)$ is its solution.

Lemma 2. [6] Let $A, A^{(1)}, A^{(2)} \in \overline{\mathbb{R}}(m, n)$ and $B, B^{(1)}, B^{(2)} \in \overline{\mathbb{R}}(m, r)$. The following assertions hold:

- If $A^{(2)} \leq A^{(1)}$ then $X^*(A^{(1)}, B) \leq X^*(A^{(2)}, B)$.
- If $B^{(1)} \leq B^{(2)}$ then $X^*(A, B^{(1)}) \leq X^*(A, B^{(2)})$.

Lemma 3. [6] Let $A^{(1)}, A^{(2)}, B^{(1)}, B^{(2)}$ be matrices of compatible sizes. The system of matrix inequalities of the form $A^{(1)} \otimes X \leq B^{(1)}$, $A^{(2)} \otimes X \geq B^{(2)}$ is solvable if and only if

$$A^{(2)} \otimes X^*(A^{(1)}, B^{(1)}) \geq B^{(2)}. \quad (6)$$

3 Weak solvability of interval matrix equations

A certain disadvantage of the necessary and sufficient condition for the solvability of (4) given in Theorem 1 (iii) stems from the fact that it only indicates the existence or non-existence of the solution but does not indicate any action to be taken to increase the degree of solvability. However, it happens quite often in modelling real situations that the obtained system turns out to be unsolvable. One of possible methods of restoring the solvability is to replace the exact input values by intervals of possible values. In practice, the travelling times a_{ij} may depend on external conditions, so they are from intervals of possible values. Due to this fact, we will require the transportation times from P_i to D_k to be from a given intervals of possible values.

Similarly to [3, 4, 9] we define

$$A = [\underline{A}, \overline{A}] = \left\{ A \in \overline{\mathbb{R}}(m, n); \underline{A} \leq A \leq \overline{A} \right\} \text{ and } B = [\underline{B}, \overline{B}] = \left\{ B \in \overline{\mathbb{R}}(m, r); \underline{B} \leq B \leq \overline{B} \right\}.$$

Denote by

$$A \otimes X = B \quad (7)$$

the set of all matrix equations of the form (4) such that $A \in \underline{A}$ and $B \in \underline{B}$. We call equation (7) an *interval max-plus matrix equation*.

Definition 1. Interval matrix equation (7) is *weakly solvable* if there exist $B \in \mathbf{B}$ and $A \in \mathbf{A}$ such that the equation $A \otimes X = B$ is solvable.

Theorem 4. *If an interval system (2) is weakly solvable then*

$$\overline{A} \otimes X^*(\underline{A}, \overline{B}) \geq \underline{B}. \quad (8)$$

Proof. Suppose that inequality (8) is not satisfied. Then, according to Lemma 3, for each $X \in \overline{\mathbb{R}}(n, r)$ at least one of the inequalities $\underline{A} \otimes X \leq \overline{B}$, $\overline{A} \otimes X \geq \underline{B}$ is not satisfied. Let $X \in \overline{\mathbb{R}}(n, r)$ be arbitrary but fixed.

In the first case, there exists $i \in M$, $k \in R$ such that $[\underline{A} \otimes X]_{ik} > \overline{b}_{ik}$. Then for each $A \in \mathbf{A}$ and for each $B \in \mathbf{B}$ we have $[A \otimes X]_{ik} > b_{ik}$, so $A \otimes X \notin \mathbf{B}$.

In the second case, there exists $i \in M$, $k \in R$ such that $[\overline{A} \otimes X]_{ik} < \underline{b}_{ik}$. Then for each $A \in \mathbf{A}$ and for each $B \in \mathbf{B}$ we have $[A \otimes X]_{ik} < b_{ik}$, so $A \otimes X \notin \mathbf{B}$.

Hence there is no $X \in \overline{\mathbb{R}}(m, r)$ such that $A \otimes X \in \mathbf{B}$ for some $A \in \mathbf{A}$, so $A \otimes X = \mathbf{B}$ is not weakly solvable. \square

Theorem 4 gives a necessary, but not sufficient condition for the weak solvability. The opposite implication does not hold because of the multiplication of two matrices with all sizes greater than one is not continuous. This is why to check for weak solvability we shall use the procedure which finds the matrices A and B such that the matrix equation $A \otimes X = B$ is solvable.

In this paper we will deal with a special case of interval matrix equations with constant right-hand side $\underline{B} = \overline{B} = B$. If any of the equations $\underline{A} \otimes X^*(\underline{A}, B) = B$ or $\overline{A} \otimes X^*(\overline{A}, B) = B$ apply, (7) is weakly solvable. So it only makes sense to deal with the case when $\underline{A} \otimes X^*(\underline{A}, B) \neq B$ and $\overline{A} \otimes X^*(\overline{A}, B) \neq B$.

We shall create the sequence of matrices $\{A^{(l)}\}_{l=1}^T$ such that $A^{(T)}$ is such that the matrix equation $A^{(T)} \otimes X = B$ is solvable in the case that (7) is weakly solvable. In which follows we show the procedure of creating the sequence $\{A^{(l)}\}_{l=1}^{(T)}$ and an auxiliary sequence $\{C^{(l)}\}_{l=0}^{(T)}$.

Let us denote $C^{(0)} = \underline{A}$. Let $l \geq 1$. If $C^{(l-1)}$ is known, define the matrices $A^{(l)}(B_k)$ for each $k \in R$ as follows:

$$a_{ij}^{(l)}(B_k) = \min\{b_{ik} - x_j^*(C^{(l-1)}, B_k), \overline{a}_{ij}\} \quad (9)$$

Lemma 5. Let $A \otimes X = B$ be an interval matrix equation with a constant right-hand side.

- (i) $x_j^*(A^{(l)}(B_k), B_k) = x_j^*(C^{(l-1)}, B_k)$ for any $k \in R$.
- (ii) Let $A \in \mathbf{A}$ be such that $A \geq C^{(l-1)}$. Then $x_j^*(A, B_k) = x_j^*(C^{(l-1)}, B_k)$ if and only if $A \leq A^{(l)}(B_k)$.

Proof. (i): Let $j \in N$, $k \in R$ be arbitrary, but fixed. Let $M_1 = \{i \in M : b_{ik} - x_j^*(C^{(l-1)}, B_k) \leq \overline{a}_{ij}\}$. We obtain

$$x_j^*(A^{(l)}, B_k) = \min_{i \in M} \{b_{ik} - a_{ij}^{(l)}(B_k)\} = \min_{i \in M_1} \{b_{ik} - b_{ik} + x_j^*(C^{(l-1)}, B_k)\}, \min_{i \notin M_1} \{b_{ik} - \overline{a}_{ij}\} = b_{ik}$$

because of $M_1 \neq \emptyset$ and $b_{ik} - \overline{a}_{ij} > x_j^*(C^{(l-1)}, B_k)$ for any $i \notin M_1$.

(ii): Suppose that $A \not\leq A^{(l)}(B_k)$. Then there exist $i \in M$, $j \in N$ such that $a_{ij} > b_{ik} - x_j^*(C^{(l-1)}, B_k)$ which implies $x_j^*(C^{(l-1)}, B_k) > b_{ik} - a_{ij} \geq x_j^*(A, B_k)$. For the converse implication suppose that $x_j^*(A, B_k) < x_j^*(C^{(l-1)}, B_k)$ for some $j \in N$. Then there exists $i \in M$ such that $b_{ik} - a_{ij} < x_j^*(C^{(l-1)}, B_k)$. Then $a_{ij} > b_{ik} - x_j^*(C^{(l-1)}, B_k) = a_{ij}^{(l)}(B_k)$. \square

For a given $k \in R$, $A^{(l)}(B_k)$ represents the greatest matrix with the principal solution corresponding to B_k equal to $X^*(C^{(l-1)}, B_k)$. In general, for $k \neq t$ we have $A^{(l)}(B_k) \neq A^{(l)}(B_t)$. To find the greatest matrix $A \in \mathbf{A}$ such that $X^*(A, B) = X^*(C^{(l)}, B)$, we define the matrix $A^{(l)}$ as follows:

$$A^{(l)} = \min_{k \in R} A^{(l)}(B_k) \quad (10)$$

Lemma 6. Let $A \in \mathbf{A}$ be such that $A \geq C^{(l-1)}$. Then $X^*(A, B) = X^*(C^{(l)}, B)$ if and only if $A \leq A^{(l+1)}$.

Proof. The equality $X^*(A, B) = X^*(C^{(l)}, B)$ is equivalent to $x_j^*(A, B_k) = x_j^*(C^{(l)}, B_k)$ for each $k \in R$. According to Lemma 5 it is equivalent to $A \leq A^{(l+1)}(B_k)$ for each $k \in R$ or equivalently $A \leq \min_{k \in R} A^{(l+1)}(B_k) = A^{(l+1)}$. \square

Suppose that for a fixed $l \in \mathbb{N}$ the equality $A^{(l)} \otimes X^*(A^{(l)}, B) = B$ is not satisfied. Denote

$$U^{(l)} = \{(i, k) \in M \times R : [A^{(l)} \otimes X^*(A^{(l)}, B)]_{ik} < b_{ik}\}.$$

Let $(r, t) \in U^{(l)}$. We shall give the procedure of creating the matrix $C^{(l)}$ such that $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rt} = b_{rt}$ and $[A^{(l)} \otimes X^*(A^{(l)}, B)]_{ik} = b_{ik}$ implies $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{ik} = b_{ik}$ if such matrix exists.

Let $u \in N$ be arbitrary but fixed. Define the matrix $C^{(l)}$ as follows:

$$c_{ij}^{(l)} = \begin{cases} \min\{b_{it} - x_j^*(A^{(l)}, B_t), \bar{a}_{ij}\} & \text{for } i = r, j = u \\ a_{ij}^{(l)} & \text{otherwise.} \end{cases} \quad (11)$$

We can see that $x_j^*(A^{(l)}, B_k) = x_j^*(C^{(l)}, B_k)$ for each $k \in R, j \in N - \{u\}$, but $x_u^*(A^{(l)}, B_k) \neq x_u^*(C^{(l)}, B_k)$ in general. Denote

$$R_u^{(l)} = \{k \in R : x_u^*(A^{(l)}, B_k) = x_u^*(C^{(l)}, B_k)\}$$

Lemma 7. $t \in R_u^{(l)}$ if and only if $b_{rt} - x_u^*(A^{(l)}, B_t) \leq a_{ru}^{(l)}(B_t)$.

Proof. Since $x^*(A^{(l)}, B_t) = x^*(C^{(l-1)}, B_t)$, the equality $x_u^*(A^{(l)}, B_t) = x_u^*(C^{(l)}, B_t)$ is equivalent to $x^*(C^{(l-1)}, B_t) = x^*(C^{(l)}, B_t)$. According to Lemma 5, this is equivalent to $C^{(l)} \leq A^{(l)}(B_t)$. For $(i, j) \neq (r, u)$ the inequality $c_{ij}^{(l)} = a_{ij}^{(l)} \leq a_{ij}^{(l)}(B_t)$ trivially holds and therefore it is sufficient for the inequality $b_{rt} - x_u^*(A^{(l)}, B_t) \leq a_{ru}^{(l)}(B_t)$ to apply. □

Denote $M_i^{(l)}(B_k) = \{j \in N : x_j^*(A^{(l)}, B_k) = b_{ik} - a_{ij}^{(l)}\}$. It is easy to see that $[A^{(l)} \otimes X^*(A^{(l)}, B)]_{ik} = b_{ik}$ if and only if $M_i^{(l)}(B_k) \neq \emptyset$.

Lemma 8. Suppose that $(r, t) \in U^{(l)}$ and there exists $u \in N$ which satisfies the following conditions:

- $t \in R_u^{(l)}$, i. e., $b_{rt} - x_u^*(A^{(l)}, B_t) \leq a_{ru}^{(l)}(B_t)$
- if $k \in R$ is such that $k \notin R_u^{(l)}$ then for each $i \in M, i \neq r$ the inequality $M_i^{(l)}(B_k) \neq \emptyset$ implies $M_i^{(l)}(B_k) - \{u\} \neq \emptyset$

Then the following is true:

- (i) $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rt} = b_{rt}$;
- (ii) $[A^{(l)} \otimes X^*(A^{(l)}, B)]_{ik} = b_{ik}$ implies $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{ik} = b_{ik}$;
- (iii) $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rk} = b_{rk}$ for each $k \in R, k \notin R_u^{(l)}$;

Proof. (i)

$$[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rt} = \max_{j \in N} \{c_{rj}^{(l)} + x_{jt}^*(C^{(l)}, B)\} \geq c_{ru}^{(l)} + x_u^*(C^{(l)}, B_t) =$$

$$b_{rt} - x_u^*(A^{(l)}, B_t) + x_u^*(C^{(l)}, B_t) = b_{rt} - x_u^*(A^{(l)}, B_t) + x_u^*(A^{(l)}, B_t) = b_{rt}$$

Together with the inequality $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rt} \leq b_{rt}$ we obtain $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rt} = b_{rt}$.

(ii) If $[A^{(l)} \otimes X^*(A^{(l)}, B)]_{ik} = b_{ik}$ and $X^*(A^{(l)}, B_k) = X^*(C^{(l)}, B_k)$, then $c_{ij}^{(l)} + x_j^*(C^{(l)}, B_k) = a_{ij}^{(l)} + x_j^*(A^{(l)}, B_k)$ holds for each $j \in N$ and the assertion trivially follows.

If $x_u^*(C^{(l)}, B_k) < x_u^*(A^{(l)}, B_k)$ then the assumption $M_i^{(l)}(B_k) - \{u\} \neq \emptyset$ implies that there exists $j \in N, j \neq u$ such that $x_j^*(A^{(l)}, B_k) = b_{ik} - a_{ij}^{(l)}$ which implies

$$c_{ij}^{(l)} + x_j^*(C^{(l)}, B_k) = a_{ij}^{(l)} + x_j^*(A^{(l)}, B_k) = b_{ik}$$

because of $x_j^*(C^{(l)}, B_k) = x_j^*(A^{(l)}, B_k)$.

(iii) For $k \notin R_u^{(l)}$ we obtain $x_u^*(C^{(l)}, B_k) = b_{rk} - c_{ru}^{(l)}$ which implies

$$[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rk} = \max_{j \in N} \{c_{rj}^{(l)} + x_j^*(C^{(l)}, B_k)\} \geq c_{ru}^{(l)} + x_u^*(C^{(l)}, B_k) = c_{ru}^{(l)} + b_{rk} - c_{ru}^{(l)} = b_{rk}.$$

Similarly as in (i) we obtain $[C^{(l)} \otimes X^*(C^{(l)}, B)]_{rk} = b_{rk}$. □

Lemma 8 provides an algorithm for finding a matrix $A \in \mathbf{A}$ such that $A \otimes X = B$ is solvable, as you can see in the following example.

Example 2. Let us have

$$A = \begin{pmatrix} [9, 10] & [4, 10] & [6, 7] \\ [8, 15] & [2, 2] & [8, 10] \\ [6, 8] & [2, 12] & [6, 7] \end{pmatrix}, \quad B = \begin{pmatrix} 10 & 7 & 13 \\ 12 & 9 & 12 \\ 13 & 6 & 10 \end{pmatrix}.$$

We use the procedure based on Lemma 8 to decide about the weak solvability of an interval matrix equation $A \otimes X = B$.

We can easily show that $\underline{A} \otimes X^*(\underline{A}, B) \neq B$ and $\overline{A} \otimes X^*(\overline{A}, B) \neq B$. We are looking for the matrix $A \in \mathbf{A}$ such that $A \otimes X^*(A, B) = B$. For $C^{(0)} = \underline{A}$ we have

$$C^{(0)} \otimes X^*(C^{(0)}, B) = \begin{pmatrix} 9 & 4 & 6 \\ 8 & 2 & 8 \\ 6 & 2 & 6 \end{pmatrix} \otimes \begin{pmatrix} 1 & -2 & 4 \\ 6 & 3 & 8 \\ 4 & 0 & 4 \end{pmatrix} = \begin{pmatrix} 10 & 7 & 13 \\ 12 & \mathbf{8} & 12 \\ \mathbf{10} & 6 & 10 \end{pmatrix}.$$

We compute the matrix $A^{(1)}$, the greatest matrix with the principal matrix solution equal to $X^*(C^{(0)}, B)$. By formula (9) and (10) we obtain

$$A^{(1)}(B_1) = \begin{pmatrix} 9 & 4 & 6 \\ 11 & 2 & 8 \\ 8 & 7 & 7 \end{pmatrix}, \quad A^{(1)}(B_2) = \begin{pmatrix} 9 & 4 & 7 \\ 11 & 2 & 9 \\ 8 & 3 & 6 \end{pmatrix}, \quad A^{(1)}(B_3) = \begin{pmatrix} 9 & 5 & 7 \\ 8 & 2 & 8 \\ 6 & 2 & 6 \end{pmatrix}, \quad A^{(1)} = \begin{pmatrix} 9 & 4 & 6 \\ 8 & 2 & 8 \\ 6 & 2 & 6 \end{pmatrix}.$$

We check whether the equality $A^{(1)} \otimes X^*(A^{(1)}, B) = B$ is satisfied. We obtain

$$A^{(1)} \otimes X^*(A^{(1)}, B) = \begin{pmatrix} 10 & 7 & 13 \\ 12 & \mathbf{8} & 12 \\ \mathbf{10} & 6 & 10 \end{pmatrix}.$$

We have $U^{(1)} = \{(2, 2), (3, 1)\}$. Let us take $r = 2, t = 2$. For $u = 3$, we have

- $t \in N_3^{(1)} = \{2\}$, i.e., $b_{22} - x_3^*(A^{(1)}, B_2) = 9 - 0 = 9 = a_{23}^{(1)}(B_3)$
- For $k = 1$ we have $M_1^{(1)}(B_1) = \{1, 2, 3\}$, so $M_1^{(1)}(B_1) - \{3\} \neq \emptyset$. ($M_3^{(1)}(B_1) = \emptyset$)
For $k = 3$ we have $M_1^{(1)}(B_3) = \{1\}$, so $M_1^{(1)}(B_3) - \{3\} \neq \emptyset$, $M_3^{(1)}(B_3) = \{1, 2, 3\}$, so $M_3^{(1)}(B_3) - \{3\} \neq \emptyset$.

Since the assumptions of Lemma 8 are satisfied, we can compute the matrix $C^{(1)}$ and verify that the assertion of this Lemma is true. We obtain

$$C^{(1)} \otimes X^*(C^{(1)}, B) = \begin{pmatrix} 9 & 4 & 6 \\ 8 & 2 & \mathbf{9} \\ 6 & 2 & 6 \end{pmatrix} \otimes \begin{pmatrix} 1 & -2 & 4 \\ 6 & 3 & 8 \\ \mathbf{3} & 0 & \mathbf{3} \end{pmatrix} = \begin{pmatrix} 10 & 7 & 13 \\ 12 & \mathbf{9} & 12 \\ \mathbf{9} & 6 & 10 \end{pmatrix}.$$

We compute the matrices $A^{(2)}(B_k)$ for $k = 1, 2, 3$ by formula (9) and the matrix $A^{(2)}$ by (10). We obtain

$$A^{(2)}(B_1) = \begin{pmatrix} 9 & 4 & 7 \\ 11 & 2 & 9 \\ 8 & 7 & 7 \end{pmatrix}, \quad A^{(2)}(B_2) = \begin{pmatrix} 9 & 4 & 7 \\ 11 & 2 & 9 \\ 8 & 3 & 6 \end{pmatrix}, \quad A^{(2)}(B_3) = \begin{pmatrix} 9 & 5 & 7 \\ 8 & 2 & 9 \\ 6 & 2 & 7 \end{pmatrix}, \quad A^{(2)} = \begin{pmatrix} 9 & 4 & 7 \\ 8 & 2 & 9 \\ 6 & 2 & 6 \end{pmatrix}.$$

We check whether the equality $A^{(2)} \otimes X^*(A^{(2)}, B) = B$ is satisfied. We obtain

$$A^{(2)} \otimes X^*(A^{(2)}, B) = \begin{pmatrix} 10 & 7 & \mathbf{13} \\ 12 & \mathbf{9} & 12 \\ \mathbf{9} & 6 & 10 \end{pmatrix}.$$

Since $U^{(2)} = \{(3, 1)\}$ we have $r = 3, t = 1$. For $u = 2$ we the following assertions hold:

- $N_2^{(1)} = \{1\}$, i.e., $b_{31} - x_2^*(A^{(2)}, B_1) = 13 - 6 = 7 = a_{32}^{(2)}(B_1)$
- For $k = 2$ we have $M_1^{(2)}(B_2) = \{1, 2\}$, so $M_1^{(2)}(B_2) - \{2\} \neq \emptyset$, $M_2^{(2)}(B_2) = \{3\}$, so $M_2^{(2)}(B_2) - \{2\} \neq \emptyset$.
For $k = 3$ we have $M_1^{(1)}(B_3) = \{1\}$, so $M_1^{(2)}(B_3) - \{2\} \neq \emptyset$, $M_2^{(2)}(B_3) = \{1, 3\}$, so $M_3^{(2)}(B_3) - \{2\} \neq \emptyset$.

Since the assumptions of Lemma 8 are satisfied, we can compute the matrix $C^{(2)}$ and verify that the assertion of this Lemma is true. We obtain

$$C^{(2)} \otimes X^*(C^{(2)}, B) = \begin{pmatrix} 9 & 4 & 6 \\ 8 & 2 & 9 \\ 6 & 7 & 6 \end{pmatrix} \otimes \begin{pmatrix} 1 & -2 & 4 \\ 6 & -1 & 3 \\ 3 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 10 & 7 & \mathbf{13} \\ 12 & \mathbf{9} & 12 \\ \mathbf{13} & 6 & 10 \end{pmatrix}.$$

Since $C^{(2)} \otimes X^*(C^{(2)}, B) = B$, it is not necessary to compute the matrix $A^{(3)}$. There exists a matrix A , namely $A = C^{(2)}$ such that $A \otimes X^*(A, B) = B$, so the given interval matrix equation is weakly solvable.

Remark 1. The problem is that the choice of the order of the elements in which we achieve equality is currently chaotic. It is possible that for a given set $U^{(l)}$ the above procedure does not lead to a solution even though a solution exists. Then it is necessary to go back one step to the set $U^{(l-1)}$ and select another element (r, t) in which we achieve equality. Thus, the algorithm is generally exponential.

3.1 Conclusion

Max-plus algebra is a useful tool for describing real situation in the traffic, economy and industry. In this paper, we dealt with the solvability of interval matrix equations in max-plus algebra. Returning to Example 1, the weak solvability of interval matrix equation with constant right-hand side means that for given total transportation times b_{ik} there exist possible times a_{ij} from the given intervals for which the total transportation time b_{ik} can be achieved. In Example 1, the values a_{ij}, x_{jl} represent the transportation times. Another possibility is the use of interval matrix equations in economics as follows. Suppose the given selling price is b_{ik} of the product P_i in the store T_k . We are looking for a price a_{ij} from the given interval, at which the manufacturer will offer the product P_i to the intermediary Q_j , so that it is possible to reach the required selling price after adding the margin of the broker x_{jk} . In practice, the selling prices of b_{ik} are also from certain intervals rather than fixed numbers. Therefore, it would be appropriate if the matrix B were also interval. The study of weak solvability for the case that the right side of the equation is an interval matrix is our main goal for the future. Another challenge is to modify the algorithm so that it is polynomial.

References

- [1] Cuninghame-Green, R. A. (1979). *Minimax Algebra*. Lecture notes in Economics and Mathematical systems. Berlin: Springer.
- [2] Gavalec, M. & Plavka, J. (2010). Monotone interval eigenproblem in max-plus algebra. *Kybernetika*, 46, 387–396.
- [3] Myšková, H. (2005). Interval systems of max-separable linear equations. *Lin. Algebra Appl.*, 403, 263–272.
- [4] Myšková, H. (2006). Control solvability of interval systems of max-separable linear equations. *Lin. Algebra Appl.*, 416, 215–223.
- [5] Myšková, H. (2012). On an algorithm for testing T4 solvability of fuzzy interval systems. *Kybernetika*, 48, 924–938.
- [6] Myšková, H. (2016). Interval max-plus matrix equations. *Lin. Algebra Appl.*, 492, 111–127.
- [7] Myšková, H. (2018). Universal solvability of interval max-plus matrix equations. *Discrete Appl. Math.*, 239, 165–173.
- [8] Olsder, G. J. et al. (1998). Course Notes: Max-algebra Approach to Discrete Event Systems. *Algebres Max-Plus et Applications an Informatique et Automatique*, INRIA, 1998, 147–196.
- [9] Plavka, J. (2012): On the $O(n^3)$ algorithm for checking the strong robustness of interval fuzzy matrices. *Discrete Appl. Math.* 160, 640–647.
- [10] Zimmermann, K. (1976). *Extremální algebra*. Praha: Ekonomicko-matematická laboratoř Ekonomického ústavu ČSAV.

The Impact of Technical Analysis and Stochastic Dominance Rules in Portfolio Process

David Neděla¹

Abstract. During the last decades, modern portfolio theory has become one of the most applied portfolio approaches by investors. However, this theory can be regarded as a pillar from which in recent years has been derived and adapted a large number of portfolio approaches. The possible approach is to combine a general portfolio model with the discipline of the financial area to find a more suitable strategy for the investment process. This paper is focused on the impact analysis of several technical analysis indicators and stochastic dominance approach in the portfolio formation process in various markets during different time horizons capturing different market conditions. Two strategies of implementation, technical analysis rules and stochastic dominance rule, in the portfolio creation process are considered. Strategy 1 aims at eliminating the whole market systemic risk with the alternative of investing in a risk-free asset. In contrast, the second strategy focuses on the use of assets meeting particular rules. It was evident from the results that using strategy 1 to eliminate systemic risk during the crisis reduced the risk of the portfolio with similar profitability. Oppositely, strategy 2 is more effective in a period with a growing global economy.

Keywords: technical analysis, portfolio model, moving average, stochastic dominance

JEL Classification: G11, G15, G20

AMS Classification: 90C05, 90C39

1 Introduction

Modern portfolio theory has become an essential approach in portfolio decision process, see [8]. However, this theory can be regarded as a pillar from which in recent years has been derived and adapted a large number of similar portfolio models. Investors are not guided in the investment strategy only by the general portfolio model, see [7]. Feasible approach can be a combination of a particular portfolio model with other disciplines of the financial area to find a more suitable strategy for investment making. In the literature can be found the possibility of including technical analysis rules or stochastic dominance (SD) rules in the portfolio theory, see [5], [9], or [12]. Some researchers dealt with the question of the advantage of the technical analysis rules for predicting the future asset price development. The moving average and indicators derived therefrom (alarms) are typical and most used rules, see [6]. However, Taylor [17] does not consider it appropriate to apply technical analysis rules to predict future price evolution. In contrast, some authors take the opposite view, see [6] or [7]. In portfolio optimization, two methods are frequently used for modeling the choice among uncertain prospects: SD and mean–risk approaches, see [12]. The advantage of the SD approach is mentioned as well in [12]. In the portfolio process, the SD of the first three orders play a crucial role, due to the relation of different investor utility functions. In [9], the SD approach is applied in the asset selection process.

The paper aims at analyzing the impact of technical indicator rules and the stochastic dominance rule in the portfolio decision process in different markets during two investment horizons capturing different market conditions. The motivation for this analysis is to examine the impact of rules from another financial area contained in the portfolio creation strategy.

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 description of technical analysis indicators and SD method is provided. Characterization of performance measures and 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 concluded.

2 Technical Indicators and Stochastic Dominance Trading Rules

We can distinguish a large number of technical analysis indicators. Therefore, only 4 frequently applied indicators by practitioners are selected, see [1], [6], or [10]. The first one is the moving average (MA), which is a general and

¹ VŠB – Technical University of Ostrava, Department of Finance, Sokolská tř. 33, Ostrava 70200, Czech Republic, david.nedela@vsb.cz

often used indicator in technical analysis, see [6] or [10]. The equation of $MA_{T,n}(x)$ calculation is following:

$$MA_{T,n}(x) = \frac{\sum_{i=0}^{n-1} x_{T-i}}{n}, \quad (1)$$

where x_T represents the asset price at time T and n is the selected length of the analyzed period.

The second selected indicator is the exponential moving average (EMA). This indicator is based on a weighted moving average where the highest importance is placed on actual prices. When the EMA is calculated for the first time, the initial value of EMA is the simple MA calculated by formula (1). The formula for calculating $EMA_{T,n}(x)$ for the weighted factor $k = \frac{2}{n+1}$ is as follows:

$$EMA_{T,n}(x) = EMA_{T-1,n}(x) + k \cdot [x_T - EMA_{T-1,n}(x)]. \quad (2)$$

One of the simplest momentum indicators is the moving average convergence divergence (MACD). The MACD equation is defined by subtracting the long-term EMA (N periods) from the short-term EMA (n periods). Mathematical expression is $MACD_{T,n,N}(x) = EMA_{T,n}(x) - EMA_{T,N}(x)$, where $EMA_{T,n}(x)$ is calculated by equation (2). In the decision process, a signal curve (usually $EMA_{T,9}(x)$) or the zero horizontal line are needed.

The relative strength index (RSI) is the last selected technical analysis indicator. Firstly, an up change UC and a down change DC are calculated for each day, see [1]. We can define a relative strength (RS) as the ratio between the n -day EMA of the UC time series and the n -day EMA of the DC time series. Usually, 14 or 9 days EMA are considered in practise. The RS is calculated as $RS = \frac{EMA_n(UC)}{EMA_n(DC)}$, where $EMA_n(x)$ is calculated by equation (2). Therefore, the RSI is calculated as $RSI = 100 - \frac{100}{(1+RS)}$.

Consequently, the signals for investing (buy/hold and sell) are defined by simple rules. If MA is considered as a technical analysis indicator, the signals are defined as following:

$$Signal \begin{cases} MA_{T,n}(x) \geq MA_{T,N}(x) & \text{for buy signal} \\ MA_{T,n}(x) < MA_{T,N}(x) & \text{for sell signal} \end{cases}, \quad (3)$$

where n and N are selected lengths of particular time periods. Instead of MA, the EMA, defined by equation (2), can be substituted. When MACD is considered, the signal rules are slightly different. If the signal curve is represented by $EMA_{T,9}(x)$, the formulation is as follows:

$$Signal \begin{cases} MACD_{T,n,N}(x) > EMA_{T,9}(x) \wedge MACD_{T-1,n,N}(x) \leq EMA_{T-1,9}(x) & \text{for buy signal} \\ MACD_{T,n,N}(x) < EMA_{T,9}(x) \wedge MACD_{T-1,n,N}(x) \geq EMA_{T-1,9}(x) & \text{for sell signal} \end{cases}. \quad (4)$$

Finally, the RSI rules are defined as:

$$Signal \begin{cases} RSI(x) \in (0, 65) & \text{for buy signal} \\ otherwise & \text{for sell signal} \end{cases}. \quad (5)$$

SD allows to compare different random variables such as asset returns or different risk indicators by their distribution function, see [12]. SD rules of the order of one to four are particularly interesting because they cumulatively impose standard assumptions of risk aversion, prudence, and restraint, which are necessary conditions for standard risk aversion, see [4]. Application of SD in portfolio process is provided in [2]. The FSD can be defined as followed: a random variable A first order stochastically dominates a random variable B , written $A >_{FSD} B$, if for any z applies $\Pr(A > z) \geq \Pr(B > z)$, where $\forall z \in \mathbb{R}$ and there is at least one z for which a strong inequality applies.

Let define the cumulative distribution function as $F_A(x) = \int_{-\infty}^x f(z)dz$. Given the previous expression, a random variable A second order stochastically dominates a random variable B , written $A >_{SSD} B$, if for any z applies

$$\int_{-\infty}^x F_A(z)dz \leq \int_{-\infty}^x F_B(z)dz, \quad (6)$$

where there is at least one z for which a strong inequality applies.

Asset selection rule based on SD approach is defined as: assets that dominate at least one asset by formula (6) and concurrently are not dominated by other assets, are selected. Mathematically, the task can be expressed as:

$$Signal \begin{cases} \text{if } x_i >_{SSD} x_j \wedge x_i \text{ is SSD non-dominated, where } i \neq j & \text{for buy signal} \\ otherwise & \text{for sell signal} \end{cases}. \quad (7)$$

3 Portfolio Performance and Models

For measuring portfolio performance, several approaches can be used. The compiled portfolio can be assessed not only on the return basis but also the risk associated with the investment must be considered. Let $\mathbf{x} = [x_1, x_2, \dots, x_z]$ is a vector of asset weights and $\mathbf{r} = [r_1, r_2, \dots, r_z]$ is a vector of gross returns, the expected return of the portfolio is an equal weighted average of the asset's expected return formulated as $E(\mathbf{x}'\mathbf{r})$. Variance of a portfolio σ_p^2 is defined as $\mathbf{x}'\mathbf{Q}\mathbf{x}$, where \mathbf{Q} is a covariance matrix of assets. The portfolio standard deviation is determined as $\sigma_p = \sqrt{\sigma_p^2}$ and semivariance below the mean is expressed as $SV_p = E\left((\mathbf{r}_m - \mathbf{x}'\mathbf{r})_+^2\right)$, where the function $(\tau)_+^2 = (\max(\tau, 0))^2$ and $\mathbf{r}_m = \frac{1}{z} \sum_{i=1}^z r_i$. The riskiness is $VarR_\alpha(X) = F_X^{(-1)}(\alpha) = -\min\{x | \Pr(X \leq x) \geq \alpha\}$, where X is a random return variable and F_X is its distribution function, then $F_X(\mu) = \Pr(X \leq \mu)$, see [14]. The equation for $CVaR_\alpha(X)$ calculation is $CVaR_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha VarR_\alpha(X) d\alpha$.

The performance ratios are more appropriate for examination, due to combining the asset's excess expected return and its risk. Sharpe ratio (SR) is one of the most used ratios, see [7] or [15]. The calculation is $SR = \frac{E(\mathbf{x}'\mathbf{r} - \mathbf{r}_f)}{(\mathbf{x}'\mathbf{Q}\mathbf{x})^{\frac{1}{2}}}$, where r_f is riskless return (or benchmark return). Subsequent ratio used to measure portfolio performance is Rachev Ratio (RR), see [13]. The calculation is $RR = \frac{CVaR_\beta(\mathbf{r}_f - \mathbf{x}'\mathbf{r})}{CVaR_\alpha(\mathbf{x}'\mathbf{r} - \mathbf{r}_f)}$. Sortino Ratio (SoR) is a modified variation of SR, where the standard deviation as a measure of risk is substituted with the downward or negative deviation, see [16]. It is formally defined as follows:

$$SoR = \frac{E(\mathbf{x}'\mathbf{r} - \mathbf{r}_f)}{E\left((\mathbf{r}_f - \mathbf{x}'\mathbf{r})_+^2\right)^{\frac{1}{2}}}, \quad (8)$$

where the function $(\tau)_+^2 = (\max(\tau, 0))^2$. Another variation is to measure the portfolio excess return with the maximum drawdown, see [18]. This ratio is called Calmar Ratio (CalR) and the calculation is as follows:

$$CalR = \frac{E(\mathbf{x}'\mathbf{r} - \mathbf{r}_f)}{\max_{t=1, \dots, T} dd_t(\mathbf{x}'\mathbf{r})}, \quad (9)$$

where $dd_t(\mathbf{x}'\mathbf{r}) = \max_{s=1, \dots, t} w_s(\mathbf{x}'\mathbf{r}) - w_t(\mathbf{x}'\mathbf{r})$ provided that for the calculation of $w_s(\mathbf{x}'\mathbf{r}) = \sum_{i=1}^t \mathbf{x}'\mathbf{r}_s - \mathbf{r}_f t$.

In the portfolio process, several approaches can be distinguished and applied. Risk minimization model is one of many. A portfolio with a minimum of risk can be defined as the following optimization problem:

$$\begin{aligned} \min \Psi(\mathbf{x}'\mathbf{r}) \\ \mathbf{x}'\mathbf{e} = 1 \\ x_i \geq 0; i = 1, \dots, z, \end{aligned} \quad (10)$$

where $\Psi(\mathbf{x}'\mathbf{r})$ is a risk indicator (such as σ_p^2 , $VarR_p$, or $CVaR_p$) and \mathbf{e} is an z -column unit vector with all values being equal to one. Due to many performance indicators of the portfolio are identified and compared, it is as well possible to optimize the portfolio based on these indicators, see [13]. In general, this problem can be written by the following equation:

$$\begin{aligned} \max \Gamma(\mathbf{x}'\mathbf{r}) \\ \mathbf{x}'\mathbf{e} = 1 \\ x_i \geq 0; i = 1, \dots, z, \end{aligned} \quad (11)$$

where $\Gamma(\mathbf{x}'\mathbf{r})$ indicates one of the performance measures (such as SR , RR , $STARR$, SoR , or CR).

4 Data Description and Empirical Application of Selected Approaches

For the empirical analysis, the data set, containing daily adjusted close prices of stocks included in the major world indices, is used. Specifically, the set of indices is contained by FTSE 100 and NASDAQ-100 indices traded on the UK and US stock markets. Since the comparison is performed at different time periods, two periods are selected, containing both the crisis period (2005-2010) and the "boom" period (2014-2019). The investment itself starts in the second year due to decision-making on a one-year time series. The risk-free rate, needed for the calculation of performance measures or as an alternative investment instrument, 10 year government bond returns is used. In all indices, several stock time series are not included in this analysis due to the incomplete data in the analysis period.

The applicability of alarm portfolio approaches is motivated by the work of [3], [5], or [6], in which the authors mention several ways how to use the alarm in the decision-making and investment process. Therefore, the whole empirical process can be divided into several steps. In the first step, the supplementary matrix $M_{i,T}$ during the investment period T is defined. The number of columns in the matrix corresponds to the number of particular assets x_i , where $i \in \langle 1, z \rangle$. The matrix data are found throughout the investment horizon according to the alarm rules of the technical indicators and SSD by equations (3), (4), (5), and (7). Due to the requirement of two length periods that are necessary in MA and EMA calculation, several combinations of $(n, N) = (5, 100), (5, 150), (10, 100)$ are considered. The recommended values (12, 26, 9) are used when the MACD indicator is applied and only nondominated assets are selected by SSD approach. The values obtained in the matrix are calculated as follows:

$$M_{i,T} \begin{cases} 1 & \text{if buy alarm in time } t \text{ applies} \\ 0 & \text{if sell alarm in time } t \text{ applies} \end{cases} \quad (12)$$

After assembling the matrix, we can proceed to the second step, where two basic strategies for using the alarm are distinguished. If applying strategy 1 (S1), the alarm rules are used to predict the interval of systemic risk during the investment horizon, see [3]. If the proportion of assets satisfying rule $M_T = \frac{\sum_{i=1}^z \mathbb{1}_{[M_{i,T}=1]}}{z} \geq \omega$, where ω is the threshold parameter of systemic risk, it is assumed that the probability of systemic risk loss in the market is low. Otherwise, the investment in a risk-free asset is preferred. In [5] is mentioned, the amount for the threshold value ω depends on the decision-maker, but in this project is set to a value of 25% for technical analysis rules and 11% for SSD rule. Strategy 2 (S2) differs from the previous one in that the investor should only invest in specific assets that meet the rules in a given interval. The rules that must be met are the same as for S1. Furthermore, the asset weights in a portfolio are determined according to the selected portfolio models, where the weights are calculated based on one-year rolling window. Finally, the portfolio value and performance indicators are calculated.

For the purposes of the analysis, a 20 days re-balancing interval is considered. The initial value of all investments in the portfolio W_0 is set as 1 currency unit. It is assumed that the value of the particular weight $x_i \geq 0$ and $x_i \in \langle 0, 1 \rangle$, therefore a short sale of shares and weight limitation are not assumed. Two approaches were used to calculate the portfolio weights: the minimum risk approach defined by equation (10) and the maximum performance ratio approach defined by equation (11). Due to the suitability of particular models in the individual markets analyzed in previous research [11], specific portfolio models vary depending on the market. The particular results are depicted in Tables 1 and 2.

From the results in Table 1, it is evident that using S1, which reduces the systemic risk, has a positive impact mainly in the period with crisis. While using technical analysis rules, the level of risk expressed by σ or VaR in the minimization risk model is lower compared to the general portfolio model in both markets. With the right choice of technical indicator, the profitability of the portfolio could have been higher. In the situation where the SSD alarm is considered in min risk model, the level of risk is essentially the same, but higher profitability is generated with a higher value of W , SR , or RR . The same conclusion is evident for the max performance ratio portfolio models. However, only a few technical indicators are suitable for applying, but it is not possible generally determining individual ones. When analyzing the min risk model during the period 2015-2019, the W of portfolios and overall portfolio performance (SR , RR) are slightly higher, however, this conclusion does not apply to SSD. Due to the properties of the model, a rapidly higher profitability cannot be expected. When examining the effect on σ or VaR , there is no significant decrease compared to the previous period. Regarding the combination of the S1 with the maximizing performance ratio model, this combination does not generate sufficient results compared to the general model, both in terms of return and risk. Generally, it can be concluded that the application of S1 is more suitable mainly during an economic crisis due to the risk-free investment at the time of the market downturn.

When we focus on the results of S2 in Table 2, the conclusions are different. Oppositely to the previous strategy, results do not show the advantage of using S2 in a situation, where there was a crisis in part of the investment period, suggesting that the alarm strategy implementation is not always beneficial but could be inefficient likewise. The previous ones apply mainly for max performance ratio models. However, when the min risk model is selected especially on UK market, the slight advantage of trading rules can be achieved when comparing SR or RR . It can be determined that this mainly meets the SSD rule or EMA indicator. Looking at the second investment period, there is already reflected positive impact of portfolio. In general, the SSD rule is not appropriate for the min risk model, but according to the second portfolio approach, the results of W , $E(R)$, or SR are the preferable. Otherwise, if the investor in US market decides to apply the rules of technical analysis, an improvement in return and performance is visible for most portfolios. Although the same is not true for the UK market, where the ratio of more favourable portfolios to less favourable ones is lower, even more profitable portfolios can be seen than using the general model. Generally, more profitable portfolios in periods with economic growth are achieved by applying this strategy. Therefore, it can be summarized that the suitability of strategies for particular periods is different.

2006-2010							2015-2019						
US Market													
Alarm	W	E(R)	σ	VaR	SR	RR	Alarm	W	E(R)	σ	VaR	SR	RR
minCaR Model													
MA(1,50)	1.6459	0.0004	0.0085	0.0133	0.0276	0.9328	MA(10,200)	1.4333	0.0003	0.0066	0.0115	0.0280	0.8419
MA(10,50)	1.3492	0.0002	0.0079	0.0128	0.0124	0.8943	MA(15,200)	1.4866	0.0003	0.0068	0.0117	0.0310	0.8564
MA(15,200)	1.7829	0.0004	0.0080	0.0125	0.0354	0.9297	EMA(5,50)	1.5965	0.0003	0.0061	0.0105	0.0415	0.8909
EMA(1,50)	2.0860	0.0005	0.0078	0.0125	0.0497	1.0324	EMA(15,200)	1.5529	0.0003	0.0067	0.0115	0.0358	0.8425
MACD	2.0918	0.0005	0.0085	0.0126	0.0465	1.0442	MACD	1.5754	0.0003	0.0053	0.0089	0.0456	0.8635
RSI	1.9823	0.0005	0.0108	0.0154	0.0351	0.9933	RSI	1.5947	0.0003	0.0068	0.0117	0.0377	0.8597
SSD	1.9059	0.0005	0.0107	0.0151	0.0330	0.9989	SSD	1.4207	0.0003	0.0063	0.0107	0.0282	0.8481
General	1.7704	0.0004	0.0108	0.0153	0.0196	0.9966	General	1.5493	0.0003	0.0068	0.0116	0.0351	0.8477
maxCaR Model													
MA(1,50)	1.8956	0.0006	0.0196	0.0299	0.0248	1.0206	MA(10,200)	3.2600	0.0009	0.0181	0.0286	0.0483	1.0183
MA(10,50)	1.8729	0.0006	0.0170	0.0268	0.0252	1.0469	MA(15,200)	3.7446	0.0010	0.0183	0.0286	0.0528	1.0400
MA(15,200)	2.2133	0.0007	0.0196	0.0289	0.0300	1.0455	EMA(5,50)	4.3406	0.0011	0.0173	0.0262	0.0607	1.0458
EMA(1,50)	2.7120	0.0008	0.0188	0.0282	0.0376	1.0800	EMA(15,200)	3.7598	0.0010	0.0182	0.0286	0.0532	1.0279
MACD	2.2404	0.0007	0.0193	0.0278	0.0306	1.1012	MACD	3.8886	0.0010	0.0159	0.0205	0.0597	1.0644
RSI	2.2246	0.0008	0.0220	0.0320	0.0291	1.0710	RSI	4.0560	0.0011	0.0183	0.0286	0.0557	1.0400
SSD	2.1959	0.0008	0.0217	0.0319	0.0288	1.0776	SSD	3.2338	0.0009	0.0174	0.0278	0.0492	1.0266
General	2.2345	0.0008	0.0220	0.0324	0.0293	1.0705	General	3.9146	0.0011	0.0183	0.0286	0.0544	1.0401
UK Market													
Alarm	W	E(R)	σ	VaR	SR	RR	Alarm	W	E(R)	σ	VaR	SR	RR
minCaR Model													
MA(1,50)	1.7799	0.0004	0.0088	0.0119	0.0319	1.1637	MA(10,200)	2.6677	0.0007	0.0070	0.0101	0.0905	1.2321
MA(10,50)	1.6083	0.0003	0.0084	0.0117	0.0251	1.1249	MA(15,200)	2.6677	0.0007	0.0070	0.0101	0.0905	1.2320
MA(15,200)	1.6554	0.0004	0.0081	0.0114	0.0280	1.0467	EMA(5,50)	2.6553	0.0007	0.0072	0.0103	0.0879	1.2159
EMA(1,50)	1.5019	0.0003	0.0092	0.0127	0.0188	1.0839	EMA(15,200)	2.7842	0.0007	0.0071	0.0102	0.0932	1.2420
MACD	1.5921	0.0003	0.0091	0.0124	0.0231	1.1159	MACD	2.0471	0.0005	0.0066	0.0098	0.0693	1.2149
RSI	1.7066	0.0004	0.0112	0.0154	0.0248	1.0736	RSI	2.9291	0.0007	0.0074	0.0105	0.0947	1.2218
SSD	1.9188	0.0005	0.0108	0.0143	0.0324	1.0920	SSD	1.7110	0.0004	0.0063	0.0093	0.0535	1.1513
General	1.6810	0.0004	0.0111	0.0154	0.0240	1.0758	General	2.7720	0.0007	0.0077	0.0105	0.0919	1.1564
maxSoR Model													
MA(1,50)	10.3996	0.0018	0.0246	0.0191	0.0669	1.7206	MA(10,200)	3.8595	0.0009	0.0107	0.0147	0.0853	1.2042
MA(10,50)	8.6427	0.0016	0.0231	0.0190	0.0648	1.6401	MA(15,200)	3.8595	0.0009	0.0107	0.0147	0.0853	1.2042
MA(15,200)	5.3451	0.0013	0.0223	0.0176	0.0521	1.5045	EMA(5,50)	4.0015	0.0010	0.0109	0.0151	0.0863	1.1914
EMA(1,50)	6.8374	0.0015	0.0248	0.0204	0.0555	1.6148	EMA(15,200)	3.9617	0.0010	0.0109	0.0154	0.0853	1.1862
MACD	7.7545	0.0016	0.0247	0.0193	0.0590	1.6269	MACD	2.8419	0.0007	0.0103	0.0142	0.0686	1.1713
RSI	8.6987	0.0017	0.0257	0.0216	0.0606	1.5711	RSI	4.1371	0.0010	0.0113	0.0158	0.0855	1.1502
SSD	9.8185	0.0018	0.0255	0.0213	0.0640	1.6119	SSD	2.6883	0.0007	0.0097	0.0127	0.0682	1.1421
General	8.6986	0.0017	0.0257	0.0216	0.0606	1.5711	General	4.1371	0.0010	0.0113	0.0158	0.0855	1.1502

Table 1 Results of S1 with particular alarm on different markets

2006-2010							2015-2019						
US Market													
Alarm	W	E(R)	σ	VaR	SR	RR	Alarm	W	E(R)	σ	VaR	SR	RR
minCaR Model													
MA(1,50)	1.5310	0.0004	0.0125	0.0175	0.0183	1.0230	MA(1,50)	2.0081	0.0005	0.0079	0.0128	0.0529	0.9534
MA(10,150)	1.6377	0.0004	0.0116	0.0169	0.0226	1.0249	MA(5,50)	1.9600	0.0005	0.0074	0.0126	0.0536	0.9489
EMA(1,50)	1.6520	0.0004	0.0104	0.0157	0.0246	0.9739	MA(15,50)	1.9624	0.0005	0.0072	0.0117	0.0551	0.9325
EMA(1,100)	1.7618	0.0005	0.0124	0.0163	0.0258	1.0392	EMA(1,50)	2.1464	0.0005	0.0077	0.0116	0.0598	0.9463
MACD	1.7388	0.0004	0.0127	0.0175	0.0248	0.9930	MACD	1.5157	0.0003	0.0076	0.0129	0.0302	0.8434
RSI	1.4515	0.0003	0.0110	0.0157	0.0160	0.9841	RSI	1.7359	0.0004	0.0070	0.0114	0.0448	0.8747
SSD	1.4425	0.0003	0.0108	0.0150	0.0157	0.9361	SSD	1.2513	0.0002	0.0073	0.0116	0.0137	0.8279
General	1.7704	0.0004	0.0108	0.0153	0.0196	0.9966	General	1.5493	0.0003	0.0068	0.0116	0.0351	0.8477
maxCaR Model													
MA(1,50)	1.7927	0.0006	0.0211	0.0323	0.0226	1.0325	MA(1,50)	4.6298	0.0012	0.0167	0.0255	0.0647	1.0810
MA(10,150)	2.3024	0.0008	0.0212	0.0324	0.0305	1.0445	MA(5,50)	4.1012	0.0011	0.0153	0.0240	0.0639	1.0048
EMA(1,50)	1.7935	0.0006	0.0194	0.0305	0.0229	1.0051	MA(15,50)	4.0073	0.0010	0.0161	0.0244	0.0604	1.0662
EMA(1,100)	1.9102	0.0007	0.0217	0.0325	0.0246	1.0245	EMA(1,50)	4.6451	0.0012	0.0173	0.0241	0.0633	1.0383
MACD	2.8280	0.0009	0.0217	0.0309	0.0365	1.1344	MACD	3.5350	0.0009	0.0145	0.0235	0.0599	0.9419
RSI	1.1545	0.0003	0.0222	0.0340	0.0095	1.0224	RSI	2.9108	0.0009	0.0184	0.0262	0.0438	0.9379
SSD	1.6047	0.0005	0.0212	0.0344	0.0192	1.0608	SSD	6.8823	0.0015	0.0187	0.0271	0.0738	1.0467
General	2.2345	0.0008	0.0220	0.0324	0.0293	1.0705	General	3.9146	0.0011	0.0183	0.0286	0.0544	1.0401
UK Market													
Alarm	W	E(R)	σ	VaR	SR	RR	Alarm	W	E(R)	σ	VaR	SR	RR
minCaR Model													
MA(1,50)	1.9604	0.0005	0.0121	0.0155	0.0314	1.0313	MA(1,50)	3.1694	0.0008	0.0077	0.0108	0.0983	1.1521
MA(10,150)	1.6040	0.0004	0.0127	0.0161	0.0201	1.0042	MA(5,50)	2.6098	0.0007	0.0075	0.0110	0.0839	1.1323
EMA(1,50)	2.1751	0.0006	0.0119	0.0155	0.0374	1.0444	MA(15,50)	2.3810	0.0006	0.0079	0.0109	0.0718	1.1162
EMA(1,100)	1.6624	0.0004	0.0122	0.0160	0.0223	1.0112	EMA(1,50)	3.4655	0.0008	0.0077	0.0105	0.1064	1.1946
MACD	1.9751	0.0005	0.0114	0.0172	0.0330	1.0189	MACD	2.2345	0.0006	0.0072	0.0110	0.0727	1.1458
RSI	1.7127	0.0004	0.0109	0.0154	0.0254	1.0918	RSI	2.7779	0.0007	0.0084	0.0112	0.0802	1.2453
SSD	2.2809	0.0006	0.0112	0.0142	0.0417	1.0522	SSD	2.7868	0.0007	0.0086	0.0124	0.0789	1.0786
General	1.6810	0.0004	0.0111	0.0154	0.0240	1.0758	General	2.7720	0.0007	0.0072	0.0105	0.0919	1.1564
maxSoR Model													
MA(1,50)	6.3091	0.0015	0.0256	0.0223	0.0527	1.4111	MA(1,50)	3.7382	0.0009	0.0118	0.0158	0.0763	1.1918
MA(10,150)	6.9036	0.0015	0.0257	0.0231	0.0548	1.3988	MA(5,50)	3.2804	0.0009	0.0117	0.0155	0.0697	1.1668
EMA(1,50)	6.5512	0.0015	0.0254	0.0224	0.0538	1.4406	MA(15,50)	3.3731	0.0009	0.0119	0.0157	0.0703	1.1565
EMA(1,100)	7.1727	0.0016	0.0261	0.0240	0.0553	1.4181	EMA(1,50)	4.0549	0.0010	0.0118	0.0155	0.0813	1.1962
MACD	3.3111	0.0009	0.0180	0.0238	0.0451	1.0928	MACD	2.1340	0.0006	0.0107	0.0157	0.0490	0.9866
RSI	6.2773	0.0015	0.0245	0.0226	0.0538	1.3953	RSI	3.6436	0.0009	0.0137	0.0161	0.0661	1.1928
SSD	8.9122	0.0017	0.0260	0.0226	0.0608	1.5592	SSD	4.8422	0.0011	0.0125	0.0171	0.0864	1.2018
General	8.6987	0.0017	0.0257	0.0216	0.0606	1.5711	General	4.1371	0.0010	0.0113	0.0158	0.0855	1.1502

Table 2 Results of S2 with particular alarm on different markets

5 Conclusion

The objection of the project was to analyze the impact of several technical analysis rules and the SD rule included in the portfolio strategy within different world markets during two investment horizons capturing different market conditions. For implementation in the portfolio creation process, two general strategies (S1 and S2) were considered, where S1 aimed to reduce systemic risk with alternative risk-free investment. In contrast, S2 selected assets that complied with the rules, without the possibility of a risk-free investment. From the results, it was evident that using Strategy 1 with selected technical analysis indicators to find systemic risk during the crisis period helped to reduce the risk of the portfolio with similar level of the profitability. In general, the conclusion of Kouaissah [5], that an alarm rule is a suitable tool for predicting future market risk, can be confirmed for this period. In contrast, Strategy 2 had the opposite effect, meaning an increase of the profitability with an unchanged level of risk in specific situations. In connection with this strategy, sufficient results were achieved with the SSD rule as well. During a period with a growing global economy, the use of strategy 1 for both models was not profitable, which meant that the strategy implementation was not always beneficial but inefficient.

Acknowledgements

Author greatly acknowledged support through the Czech Science Foundation (GACR) under project GA20-16764S, SGS research project SP2021/15 of VSB-TU Ostrava, and Moravian-Silesian region by the project RRC/02/2020.

References

- [1] Anderson, B. & Li, S. (2015). An investigation of the relative strength index. *Banks and Bank Systems*, 10(1), 92–96.
- [2] Dupačová, J. & Kopa, M. (2014). Robustness of optimal portfolios under risk and stochastic dominance constraints. *European Journal of Operational Research*, 234(2), 434–441.
- [3] Giacometti, R., Ortobelli, S. & Tichý, T. (2015). Portfolio selection with uncertainty measures consistent with additive shifts. *Prague Economic Papers*, 24(1), 3–16.
- [4] Kimball, M.S. (1993). Standard Risk Aversion. *Econometrica*, 61(3), 589–611.
- [5] Kouaissah, N. & Hocine, A. (2020). Forecasting systemic risk in portfolio selection: The role of technical trading rules. *Journal of Forecasting*, 1–22.
- [6] Kouaissah, N., Orlandini, D., Ortobelli, S. & Tichý, T. (2019). Theoretical and practical motivations for the use of the moving average rule in the stock market. *IMA Journal of Management Mathematics*, 31(1), 117–138.
- [7] Kouaissah, N., Ortobelli, S. & Tichý, T. (2018). *Portfolio Theory and Conditional Expectations: Selected Models and Applications*. SAEI, vol. 59, VŠB-TU Ostrava.
- [8] Markowitz, H.M. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77–91.
- [9] McNamara, J.R. (1998). Portfolio Selection Using Stochastic Dominance Criteria. *Decision Sciences*, 29(4), 785–801.
- [10] Mills, T.C. (1997). Technical Analysis and the London Stock Exchange: Testing Trading Rules Using the FT30. *International Journal of Finance & Economics*, 2(4), 319–331.
- [11] Neděla, D. (2020). *On the Impact of Technical Analysis Rules in Selected Portfolio Approaches Under Different Market Conditions*. Subject Project. VSB–TU Ostrava.
- [12] Ogryczak, W. & Ruszczyński, A. (2001). On consistency of stochastic dominance and mean–semideviation models. *Mathematical Programming*, 89(2), 217–232.
- [13] Rachev, S.T., Stoyanov, S.V. & Fabozzi, F.J. (2008). *Advanced stochastic models, risk assessment and portfolio optimization: The ideal risk, uncertainty and performance measures*. New York: Wiley Finance.
- [14] Rockafellar, R.T. & Uryasev, S.P. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7), 1443–1471.
- [15] Sharpe, W.F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, 21(1), 49–58.
- [16] Sortino, F.A. & Price, L.N. (1994). Performance measurement in a downside risk framework. *Journal of Investing*, 3(3), 59–65.
- [17] Taylor, N. (2014). The rise and fall of technical trading rule success. *Journal of Banking & Finance*, 40(1), 286–302.
- [18] Young, T.W. (1991). Calmar ratio: A smoother tool. *Futures*, 20(1), 40.

Information Retrieval System for IT Service Desk for Production Line Workers

Dana Nejedlová¹, Michal Dostál²

Abstract. The information retrieval system gets answers to its users' questions. As a part of an IT service desk available for the workers at the production line it recognizes a spoken request and selects a proper answer or action for this request. Speech recognition done at the beginning of the process and presentation of the selected answer done in the end of the process is solved using freely available modules for speech recognition and text to speech. The subject of the presented research is the process of selecting answers to recognized questions. The software solving this process is done by the authors of this contribution. The main result of the research is the comparison of two approaches of selecting answers to questions, namely decision tree and neural network. The input of both approaches is the set of keywords recognized in the user's speech. Due to the lack of real data from a real production line, the training set of keywords and correct answers is generated artificially. The result of the research shows that the neural solution is advantageous when the number of possible answers and the number of keywords in questions is high.

Keywords: information retrieval, decision tree, neural network

JEL Classification: C61, D83

AMS Classification: 68T30

1 Introduction

Automatic processing of natural language is becoming more important as more information written or spoken is communicated through the internet and other electronic media. During recent years important improvements in this task have been achieved thanks mainly to the field of deep learning. Some older approaches to solving this task have also proved to meet all requirements imposed by practical usages.

The aim of this contribution is to present two types of information retrieval system and compare the traditional approach to making such a system using decision tree and a new solution using neural network. Information retrieval (IR), as defined in [8], is a task of finding information that answers a given query (i.e. question). IR system dealt with in this contribution is intended to be used by production line workers who need help from IT service desk. The user interface of this system uses a microphone and a speech recognition module that outputs query in the form of the sequence of words. This query is processed by decision tree or neural network that select some reaction to the query from the set of possible answers or actions. This selection can also be called a classification of query, and because the query has been converted to text, we can call it text classification task.

One of the earliest works about solving text classification by decision trees is by Lewis and Ringuette [6] from 1994. Early works describing solving the same task by neural networks are by Wiener et al. [14] and Schütze et al. [12], both from 1995. The other efficient text classification method based on the theory of probability, Naïve Bayes, was published in 1960 by Maron and Kuhns [9]. The queries and the documents to be retrieved in early neural network and probabilistic solutions were represented by vectors which had as many elements as was the size of dictionary, and the values of their elements have been derived from the frequency of each word in the represented document. The process of finding the category for the query involved the dimensionality reduction of the vector space by statistical techniques closely related to Principal Component Analysis (PCA). The way of describing the probability of occurrence of one word on condition that two other words precede it in the text, i.e. the so-called trigram language model, was represented as the table of (size of dictionary)³ values and the typical size of dictionary, i.e. the number of different words in the language, is 100 000. This meant that the system that

¹ Technical University of Liberec, Faculty of Economics, Department of Informatics, Studentská 2, 461 17 Liberec, Czech Republic, dana.nejedlova@tul.cz.

² Technical University of Liberec, Faculty of Economics, Department of Informatics, Studentská 2, 461 17 Liberec, Czech Republic, michal.dostal1@tul.cz.

predicted the most probable next word in e.g. speech recognition task had to work with a table of 10^{15} values. Youshua Bengio et al. [1] showed in 2003 directions to escape this curse of dimensionality by representing each word by a vector typically not longer than 100 elements which describe its relationship with other words using the so-called distributed representation. These vectors can be trained from texts by neural networks and today they are called **word embeddings**. Recently, the fast development in the field of deep learning with clusters of computers and large corpora of text and multimedia content found on the internet lead to such combinations of trained word embeddings and neural networks for their processing that enables to work with text and some multimedia in such a way that borders on the level of human understanding.

Besides typical tasks of natural language processing, like speech recognition, question answering, translation, semantic analysis, text summarization, text generation, and sentiment analysis, some more specialized applications of contemporary speech processing are developed. Examples of economic applications include training of word embeddings from texts specialized on particular problem domain, e.g. the oil and gas industry [3], customer segmentation based on their queries and descriptions of items they have positively rated from which word embeddings have been trained [2], description of the evolution of understanding of a particular economic concept from literature about this concept from various time epochs [7], text summarization of articles that aids access to medical evidence [11], identification of similar companies based on a corpus of financial news articles [5], and estimation of uncertainty about the financial health of companies from their annual reports [13]. The aspects of use of virtual assistants in various kinds of businesses are discussed in [10].

2 Data Preparation

Both forms of information retrieval system presented in this contribution have the same input and output. The input is a question in the form of a set of keywords. The output is a single answer. The set of keywords is not ordered, and each instance of its member is contained in it only once. A single keyword in real-life application may be either a single word or a phrase. If phrases are included, then each phrase has its own identification number (ID). If distinct words forming a phrase are the same as keywords found elsewhere in the data, then each of them is given another ID.

Keywords characterize the answer and in natural language they are accompanied by function words, defined in [8], that carry information about relationships among keywords and are present repeatedly in texts on many different topics. The rule-based system described in Section 2.1 has manually designed a set of keywords which does not contain function words. Data for the neural network, preparation of which is described in Section 2.2, contain function words, and one of the aims of this research is to test whether neural network can correctly classify questions that contain also function words without informing it what words are keywords and what words are function words.

2.1 Rule-Based Expert System

The types of keywords detected by the system are grouped by corresponding areas of ITIL (Information Technology Infrastructure Library), by which the IT services are built and managed. The keywords are in lemmatized form, so that the system is able to detect them in all their grammatical forms. To this day there are over 30 keywords that are detected by the system. The set of possible answers develops with the needs of the IT services department and is based on the experiences with communication with the users, i.e. the workers on the production line.

2.2 Neural Network

The words or phrases found in real-life application are represented by single letters of English alphabet in the data for the neural network. Due to the lack of real data from a real production line, which is the supposed environment where our information retrieval system should be used, our data are generated artificially according to model (1).

$$A \wedge (B \vee C) \Rightarrow Answer_1 \quad (1)$$

Model (1) reads as “If set of keywords is composed of word *A* and word *B* or *C*, then the output of the system should be *Answer_1*.” In logic such “if *X* then *Y*” rule can be represented by logical operator of implication “ $X \Rightarrow Y$ ” which can be read as “*X* implies *Y*.” In this implication *X* is called *antecedent* and *Y* is called *consequent*. Model (1) can be split into two implications (2) and (3) meaning that a single answer can be retrieved for more than a single set of keywords.

$$A \wedge B \Rightarrow Answer_1 \quad (2)$$

$$A \wedge C \Rightarrow Answer_1 \quad (3)$$

There is a possibility in real-life data that for a single set of keywords more than one answer is correct, which may be caused by some circumstances like word order, intonation, and processes around production line not captured in recognized words shown in implications (4) and (5):

$$A \wedge B \Rightarrow Answer_1 \quad (4)$$

$$A \wedge B \Rightarrow Answer_2 \quad (5)$$

Such data cannot be processed without error and the solution of this phenomenon would involve inclusion of more information (e.g. word order) into the input of the information retrieval system. Our artificial data are designed in such a way, so that there are no ambiguities of the type (4) and (5).

We have manually designed 52 sets of 1 to 4 keywords leading to 27 different answers. The same keywords are present in different sets and some different sets lead to the same answer.

There are 18 keywords in our data represented by letters *A* to *R*. Keywords in speech processing should be recognized from all other words. We augment our sets of keywords by zero to 4 different function words which are represented by 8 letters *S* to *Z*.

In rule-based systems the function words should be discarded from the set of keywords by statistical processing of word frequencies in large text corpora or manually when only a small amount of text is available. Neural network should recognize keywords from functional words when it learns large enough amount of data with mixed keywords and functional words, because it will see some repeating pairs of the same set of keywords and the same answer but no or little number of repeating pairs of the set of functional words and the same answer.

All possible combinations of zero to 4 function words have been added to each of our 52 sets of keywords with answers. In such a way each set of keywords has acquired 163 instances with different set of additional function words. All sets of words that form the antecedent of all 8476 (= 52 x 163) implications are connected by logical operator AND represented as \wedge in our models (1) to (5). All data can be after this augmentation divided into training (52 x 81 implications) set and test set (52 x 82 implications) each containing the same 52 sets of keywords but differing in the additional set of function words.

3 Traditional Solution – Rule-Based Expert System

The rule-based expert system is written by the second author of this paper in Python programming language and is using open source libraries for artificial intelligence capabilities such as speech recognition and text-to-speech. This system has a three-layer architecture. There is a presentation, application, and data layer in it.

The presentation layer (the front-end part of this application) contains a communication interface with a microphone attached to speech recognizer for user input and text to speech module attached to a speaker for the acoustic output. The application layer of the system processes information provided by the speech recognizer. This layer forms the main part of the application. The logic of the decision tree, the heart of this expert system, is programmed here. The data layer (the back-end part of this application) contains the database with all important information needed for resolving the queries placed by the users once they are classified by the decision tree.

The nodes of the decision tree represent the attributes (in our case a keyword or more keywords present in the spoken statement or question). The branches then represent decision rules, which determine what answer should be replied to the user. As usual, the leaf nodes represent the outcome (in our case the answer). The decision tree is developed manually based on the needed dialogue flow. For example, one branch of the decision tree is executed if the user input contains a keyword “password”. The user is asked for his or her employee ID and other authentication information in the subsequent nodes on this branch. The branches of the decision tree were developed gradually based on the topics that were selected as desired services to be delivered by the system to its users.

The question answering process of this expert system begins with calling a special telephone number allocated for this specific use case. The phone call is directed to the communication interface of the system which invites the user to place a question or a request. The expert system then evaluates the input and tests for specific keywords contained in it. This will initiate the execution of the decision tree. Based on the specific keywords, the

system chooses a path (branch of the decision tree) and either provides an answer or asks the user a follow-up question. When the system asks a follow-up question, its purpose is mainly to get more precise information. For example, if the employee asks for a piece of information or a task that needs authentication, the system asks him or her for his or her credentials to ensure that he or she is authorized to get that information or some process (e.g. password change, contact telephone number change, etc.) could be initiated for him or her.

4 Connectionist Solution – Neural Network

Deep feedforward neural network with five layers of neurons, i.e. with three hidden layers, has been programmed by the first author of this paper in C language and is freely available together with the analysis of results on the link <https://owncloud.cesnet.cz/index.php/s/HV3mQrhruV2cc1>. Neurons in adjacent layers are fully connected forming a matrix of weights which are its parameters. To all but the first layer of neurons one extra connection with a trainable weight leads from a small nonzero constant input forming the so-called bias term which is added to the rest of each neuron's input, and this input is a dot product of the output of preceding layer and weights on the connections from all neurons on the preceding layer to the neuron on the current layer. The output of each neuron on all but the first layer is the result of the activation function. Neurons on the first layer just pass forward the input data to the network. The activation function on all hidden layers is tanh and on the output layer is sigmoid. The only difference from a standard backpropagation algorithm is a constant called the error propagation amplifier, which is a number larger than one that is multiplied with the error propagated backwards through the network without which this error would be dampened in its path through the network having more than a single hidden layer. Initial weights of the network and initial values of elements of word embeddings are set to random values from the interval of (-0.5, +0.5). The network is trained according to the following algorithm:

1. The input of the network is a single word embedding.
2. The network computes the output of this word embedding on its output layer.
3. The network computes squares of differences of its activations on the output layer from the correct values of the output which is one of 27 possible answers, mentioned in Section 2.2, encoded as a vector of 26 zeros and a single value of one on a position characterizing this answer. Such a representation is called one-hot vector in [4]. These squares of differences are errors on neurons on the output layer.
4. The network computes errors on neurons on all its other layers in such a way that each but the output layer computes its error from the error on the next layer.
5. The error on the first (input) layer is used to update the values of the input word embedding.
6. All weights of the network are updated according to the errors computed on the layers that they lead to.

The process of learning a single implication, defined in Section 2.2, consists of the above mentioned 6 points applied successively to all word embeddings forming the antecedent of the implication. The network repeats learning all implications until it reaches its maximal accuracy of the prediction of all consequents from all antecedents. The implications in the training set are processed one after another. After the processing of this set the neural network outputs the number of correctly classified implications into consequents (i.e. answers) in the training set and test set separately.

To get these numbers of correctly classified implications the network at first computes the sum of activations of the output neurons for all 26 kinds of word embeddings which are English letters A to Z, see Section 2.2, and the result is a vector called **sum_of_output** with as many elements as is the number of neurons in the output layer. Then the network computes the vector of the same number of elements with the sum of the output for all word embeddings present in the set for a given implication called **sum_of_output_in_implication** according to formula (6).

$$\mathbf{sum_of_output_in_implication} = \sum_{\text{words in implication}} 2 \cdot \mathbf{output_of_word} \quad (6)$$

Then the index of the answer to which the network classifies the set of words in the implication is the index of maximal value in the vector (7).

$$\mathbf{sum_of_output_in_implication} - \mathbf{sum_of_output} \quad (7)$$

The value of elements in vector (7) is equal to the sum of positive values of the output of words in the implication and the negative values of the output of all other words. In this way the network can discriminate between such sets of words where one set is the subset of the other set.

5 Experimental Results

This section presents the quality (in terms of algorithmic complexity and accuracy of its response) of information retrieval system in the form of a rule-based expert system, see Section 3, and in the form of a neural network, see Section 4.

5.1 Rule-Based Expert System

The algorithmic complexity of the already compiled decision tree is directly proportional to the average length of its branches leading from the initial node that processes the first recognized keyword to the final node with the answer.

Because of the fact that the rule-based expert system does not run in real production line, we cannot statistically determine its rate of successful use cases. If all words are correctly recognized by the speech recognition module and the user uses the right words for the answer encoded in the decision tree than the system is always successful.

5.2 Neural Network

The algorithmic complexity of a trained neural network is directly proportional to the number of its weights multiplied with the number of words in the question. The fully connected network with 5 layers with l_1 to l_5 number of neurons respectively has number of weights determined by formula (8).

$$\text{number of weights} = (l_1 + 1) \cdot l_2 + (l_2 + 1) \cdot l_3 + (l_3 + 1) \cdot l_4 + (l_4 + 1) \cdot l_5 \quad (8)$$

The best results on data described in Section 2.2 have been achieved with the network with 10 neurons on the input layer, which means that the embeddings were represented by vectors of 10 values. Each of all three hidden layers had 20 neurons and the output layer had 27 neurons, so that the network could classify word embeddings into 27 different answers. The learning rate parameter was 0.0001 and the error propagation amplifier was 1.1.

The network has been cycling through the training set and the number of correctly classified implications had both in the training set and the test set its highest values after 21000 cycles. These values were 3762 for the training set and 3800 for the test set. The reason why the number of correctly classified implications in the test set was higher than at the training set was the way of computing the class according to formulas (6) and (7) which are different from the result of standard minimization of the sum of square error.

Another experiment has been conducted with the network that learned implications defined only by the keywords, as if the function words were filtered from the input using some external knowledge, as is the case in our rule-based expert system.

Analysis of erroneously classified implications has shown that misclassified implications in the test set were almost the same as those in the training set as well as in the results of different experiments with neural network being trained from the state of random weights and random word embeddings. These errors are related to the formulas (6) and (7), because for some implications mainly with keywords designed as synonyms of words in other implications with the same answer only the **sum_of_output_in_implication** would be more appropriate.

The accuracy of classification as the percentage of correctly classified implications in the number of all implications is shown in Table 1.

Set	Number of antecedents	Number of correctly classified antecedents	Accuracy
Training	4212	3762	89.32%
Test	4264	3800	89.12%
Keywords only	52	47	90.38%

Table 1 Accuracy of neural network

The analysis of variance of random projections of resulting word embeddings shows that words that had a relatively large number of answers associated with them have lower variance than words in implications having a relatively small number of answers as their consequent. In such a way function words could be statistically discriminated from keywords.

6 Conclusion

In this work, two approaches to the task of information retrieval have been compared: the decision tree and the neural network. The decision tree has been compiled manually from manually designed set of keywords. The data for the neural network have been compiled manually as well but function words have been added to keywords. Experimental results have shown that the trained word embeddings could be statistically processed to discern keywords from function words.

The main advantage of our decision tree is its 100% accuracy when it is properly used. Its main disadvantage is the need of manual work with words and manual addition of branches for new answers. The main advantage of neural network is its ability to work with speech that contains also function words when it is trained on enough data. Its main disadvantage is its non-zero error rate even for a relatively small amount of data and the fact that it cannot recognize utterly wrong answers while the decision tree has branches for such combinations of words.

In the next phase of our research, we will seek to find proper uses of both systems in real production settings.

One of the aims of this paper is to draw attention to the emerging way of economic data analysis in which economic objects would be represented by embeddings. Analysis applied to text, like prediction of next word or finding the relationships between words where words or other units of the text are represented as embeddings, as it is shown in this paper, can be likewise done with various participants and phenomena of economic processes, especially from big data produced by the internet and other information and communications technology.

Acknowledgements

This work is supported by internal grant of the Faculty of Economics of the Technical University of Liberec.

References

- [1] Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, Ch. (2003). A Neural Probabilistic Language Model. *Journal of Machine Learning Research*, 3(Feb), 1137–1155.
- [2] Boratto, L., Carta, S., Fenu, G., & Saia, R. (2016). Using neural word embeddings to model user behavior and detect user segments. *Knowledge-Based Systems*, 108, 5–14. DOI: [10.1016/j.knosys.2016.05.002](https://doi.org/10.1016/j.knosys.2016.05.002)
- [3] Gomes, D. da S., M., Cordeiro, F. C., Consoli, B. S., Santos, N. L., Moreira, V. P., Vieira, R., Moraes, S., & Evsukoff, A. G. (2021). Portuguese word embeddings for the oil and gas industry: Development and evaluation. *Computers in Industry*, 124, 1–14. DOI: [10.1016/j.compind.2020.103347](https://doi.org/10.1016/j.compind.2020.103347)
- [4] Jurafsky, D. & Martin, J. H. (2020). *Speech and Language Processing*. Third Edition draft. <https://web.stanford.edu/~jurafsky/slp3/>
- [5] Kee, T. (2019). Peer Firm Identification Using Word Embeddings. In *2019 IEEE International Conference on Big Data (Big Data)*, (pp. 5536– 5543). Los Angeles, CA, USA. DOI: [10.1109/BigData47090.2019.9006438](https://doi.org/10.1109/BigData47090.2019.9006438)
- [6] Lewis, D. D. & Ringuette, M. (1994). A comparison of two learning algorithms for text categorization. In *Proc. SDAIR 94*, (pp. 81–93). Las Vegas, NV.
- [7] Mahanty, S., Boons, F., Handl, J., & Batista-Navarro, R. T. (2019). Understanding the Evolution of Circular Economy through Language Change. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, (pp. 250–253). Florence, Italy.
- [8] Manning, Ch. D. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts, London, England: The MIT Press
- [9] Maron, M. E. & Kuhns, J. L. (1960). On relevance, probabilistic indexing, and information retrieval. *Journal of the ACM*. 7(3), 216–244. DOI: [10.1145/321033.321035](https://doi.org/10.1145/321033.321035)
- [10] Quarteroni, S. (2018). Natural Language Processing for Industrial Applications. *Informatik-Spektrum*, 41, 105–112. DOI: [10.1007/s00287-018-1094-1](https://doi.org/10.1007/s00287-018-1094-1)
- [11] Sarker, A, Yang, Y-C, Al-Garadi, M. A., & Abbas, A. (2020). A Light-Weight Text Summarization System for Fast Access to Medical Evidence. *Frontiers in Digital Health*, 2. DOI: [10.3389/fdgth.2020.585559](https://doi.org/10.3389/fdgth.2020.585559)
- [12] Schütze, H., Hull, D. A., & Pedersen, J. O. (1995). A comparison of classifiers and document representations for the routing problem. In *SIGIR '95*, (pp. 229–237). DOI: [10.1145/215206.215365](https://doi.org/10.1145/215206.215365)
- [13] Theil, Ch. K., Štajner, S., & Stuckenschmidt, H. (2020). Explaining Financial Uncertainty through Specialized Word Embeddings. *ACM/IMS Transactions on Data Science*. 1(1). DOI: [10.1145/3343039](https://doi.org/10.1145/3343039)
- [14] Wiener, E., Pedersen, J., & Weigend, A. (1995). A neural network approach to topic spotting. In *Proc. SDAIR 95*, (pp. 317–332). Las Vegas, NV.

Permanent Income Hypothesis with the Aspect of Crises. Case of V4 Economies

Václava Pánková

Abstract. Some households manage their consumption according to the permanent income hypothesis (PIH) and others consume their actual income without any apparent long-run conception. The quantification of a percentual share of both groups bring an important macroeconomic information about a considerable part of GDP.

The theory of PIH allowing for computing the share of both consumption sorts is described and applied to V4 (Czech Republic, Hungary, Poland, Slovakia) economies. The influence of financial crisis in the end of first decade and a starting period of covid crisis is formulated by the help of a multiplicative dummy variable. The share of households consuming according to actual income is from two thirds to three quarters and uses to increase significantly due to the impact of crises.

Keywords: PIH, cointegration, seemingly unrelated regression

JEL Classification: E2, C82, C51

AMS Classification: 62J05

1 Introduction

Consumption is an important entity both for the households and for the policymakers. As a considerable part of GDP (represents 50 – 70 % of spending in most economies) it is a subject of theoretical studies and theories. In 1957, M. Friedman defined in [4] a concept of permanent income hypothesis (PIH) based on an assumption that the economic subject consumes according to its expected rather than actual income. Permanent income and consumption represent a concept which brings a long – run information because of its nature. It is a theory of consumer spending which states that people will spend money at a level consistent with their expected long term average income. Consumers do not care about the past, they only care about the present and future. The consumed proportion of the expected income may be influenced by the shocks which are either transitory (e.g. short-run changes of taxes, super-gross wage cancelled for two years only) or permanent (e.g. changes of the social system, possible retirement reform). Consumers will react differently if a shock is permanent rather than transitory.

In case of PIH, consumption functions should not be formulated in terms of consumption expenditures and disposable income but in terms of permanent and transitory income and consumption which are not observable. The permanency phenomenon was studied by using different complementary theories: Adaptive expectation is applied e.g. in [2]. Based on rational expectation hypotheses it is elaborated by Hall [5] and Sargent [8], both approaches harmonized by Flavin [3]. The concept of rational inattention was formulated by Sims [9] changing the quality of the topic by an addition of further aspects. Validation of permanent income hypothesis made by the help of some of the mentioned approaches can bring eventual implications for policy makers.

2 Material and Methodology

Though the theory never was called into question, empirical data often used not to validate the PIH even in the most developed economies. A convenient analytical apparatus occurred due to articles of Hall [5], Campbell and Mankiw [1] and Flavin [3]. A realistic assumption is incorporated that only a part of households uses a permanent income and a percentage of it can be estimated.

One group of consumers reflects the actual disposable income as $C_{1t} = Y_{1t}$ and the other group consumes according to the permanent income $C_{2t} = Y_{2t}^P$. Together the income is

$$Y_t = Y_{1t} + Y_{2t}^P = \omega Y_t + (1 - \omega) Y_t \quad , \quad 0 \leq \omega \leq 1 \quad (1)$$

ω explaining the complementarity. Further we have $\Delta C_{1t} = \omega \Delta Y_{1t}$, $\Delta C_{2t} = (1 - \omega) \Delta Y_{2t}$.

¹ VŠE, FIS, Dept. of Econometrics, 130 00 Praha, nám. Winstona Churchilla 4, pankova@vse.cz

In [3] it is demonstrated that it is also $\Delta C_{2t} = \alpha + (1 - \omega)\varepsilon_t$ with ε_t representing innovation of actual income which gives an impact to a revision of the permanent income.

Changes of aggregate consumption, actual as well as permanent, are given by

$$\Delta C_t = \Delta C_{1t} + \Delta C_{2t} = \alpha + \omega\Delta Y_t + (1 - \omega)\varepsilon_t \quad . \quad (2)$$

Evidently, we only need the increments of current values of consumption and income. On the base of (2), PIH can be formulated as $H_0: \hat{\omega} = 0$. In case of rejection of the null hypothesis, consumption tracks income closely. Empirical analyses use to reject the hypothesis; developed economies tend to smaller values of $\hat{\omega}$.

Both actual and permanent consumption use to change under an impact of crises. In general, it is observed that consumption decreases more than income, on the other hand, savings use to increase. An Asian economic crisis took place at the last 90-ties. It arrived unexpectedly and its consequences were studied afterwards. A convenient approach using relevant dummy variables is described e.g. in [6]. In this millennium, in 2008 – 2009 was the undeniable financial crisis with relevant economic consequences. The other crisis started in 2020 in connection with the Covid-19 disease and is not yet over. The former and the beginning of the latter crisis are covered by our data.

As we believe that the share of PIH households may decrease due to a crisis, (2) was transformed into

$$\Delta C_t = \alpha + \omega\Delta Y_t + \delta D\Delta Y_t + (1 - \omega)\varepsilon_t \quad . \quad (3)$$

D is a multiplicative dummy variable, $D = 1$ during the periods of crises. The multiplicative dummy shows

$$\Delta C_t \approx (\omega + \delta D)\Delta Y_t = \begin{cases} \omega\Delta Y_t & \text{if } D = 0 \\ (\omega + \delta)\Delta Y_t & \text{if } D = 1 \end{cases}$$

δ is anticipated to be positive. By δ we do not measure the changes in consumption but the changes in the PIH versus non-PIH distribution.

3 Results

Model (3) is applied to V4 economies (Czech Republic, Hungary, Poland, Slovakia).

Quarterly data (source: Eurostat, seasonally unadjusted) cover the period from 1996Q1 to 2020Q2. Consumption C and GDP Y are distinguished by suffixes -cz for Czechia, -hu Hungary, -pl Poland and -sk Slovakia. GDP is used as a proxy for aggregate income. First, the series were transformed using the HP (Hodrick – Prescott) filter. Further, the ADF (Augmented Dickey Fuller) test was performed to study stationarity. Relevant t-statistics are shown in Table 1. Critical value is -2.894 at 5 % level Evidently, all the series are I(1). Using the differences in equation (3) we are sure to avoid a spurious regression.

	t-statistic /series/	t-statistic /first differences/
Ccz	-1.732	-19.828
Chu	-1.936	-19.801
Cpl	-1.853	-19.739
Csk	+4.997	-17.814
Ycz	-1.725	-19.786
Yhu	-1.985	-19.784
Ypl	-1.913	-19.750
Ysk	+1.272	-18.949

Table 1 t-statistics of ADF test

All the data is measured in local currencies; as for the Slovakia both variables are in Euro during all the followed period. The transformation from Slovak crown to Euro of the data from the pre-Euro period was done by Eurostat.

The non-homogeneity in currencies is the reason why the seemingly unrelated regression SUR was used instead of a panel technique. Such estimates are more efficient (in comparison to OLS applied to single equations) if the error terms are correlated between the equations. It respects single equations (one for each country) but allow for an assumption of common disturbance impacts what in V4 can very well be reasoned due to the rather long common history and very similar economic development.

The substantial part of computation (Eviews) is presented as Table 2 with the parameters re-named for a better orientation.

Estimation Method: Seemingly Unrelated Regression
Included observations: 392

	Coefficient	Std. Error	t-Statistic	Prob.
Alpha-cz	3.516108	25.83382	0.136105	0.8918
Omega-cz	0.657512	0.004777	137.6365	0.0000
Delta-cz	0.179703	0.020425	8.798326	0.0000
Alpha-hu	-320.5660	275.3116	-1.164375	0.2445
Omega-hu	0.709742	0.004892	145.0720	0.0000
Delta-hu	-0.003012	0.022498	-0.133898	0.8935
Alpha-pl	12.18713	10.17939	1.197236	0.2314
Omega-pl	0.731144	0.004306	169.7962	0.0000
Delta-pl	0.040384	0.010869	3.715496	0.0002
Alpha-sk	0.275806	0.776889	0.355014	0.7226
Omega-sk	0.765118	0.008276	92.44940	0.0000
Delta-sk	0.125075	0.027235	4.592348	0.0000

Equation: DCCZ_HP = Alpha-cz + Omega-cz*DYCZ_HP + Delta-cz*D1Y

Observations: 391

R-squared 0.980779

Equation: DCHU_HP = Alpha-hu + Omega-hu*DYHU_HP + Delta-hu*D2Y

Observations: 391

R-squared 0.982175

Equation: DCPL_HP = Alpha-pl + Omega-pl*DYPL_HP + Delta-pl*D3Y

Observations: 391

R-squared 0.988145

Equation: DCSK_HP = Alpha-sk + Omega-sk*DYSK_HP + Delta-sk*D4Y

Observations: 391

R-squared 0.959429

Table 2 The Eviews output, parameters of model (3)

The share of those who consume according to their permanent income (given by $1 - \omega$) and a change expected under an influence of a crisis (given by δ) is presented in Table 3.

	Percentage of PIH households	Influence of crises
Czech Republic	34.25	-17.97
Hungary	19.03	no apparent change
Poland	26.86	- 4.03
Slovakia	23.49	-12.50

Table 3 The share of PIH households

4 Discussion and Conclusions

Consumption following a permanent income is a theoretical concept the confirmation or non-confirmation of which brings a consequence to an eventual forecast of future aggregate consumption. An econometric approach for testing the validity of permanency is known but empirical data often used not to validate the PIH. A more realistic assumption incorporates an idea that only a part of households uses a permanent income and a percentage of it can be estimated. The other group of households are those whose consumption tracks their actual income closely.

In the article 1, the economic idea and basic econometric theory of the relevant approach are recapitulated. Model is formulated and extended by a variable allowing for a quantification of a possible crises impact on the proportion of both groups of consumers.

Article 2 delivers an application to V4 economies. It starts by a routine time – series analysis of the data. Econometric estimation is made by the help of the method SUR which allows for profiting from the fact that economic, social and historical background of countries comprised leads to correlated disturbances and thus to more efficient estimates. A brief economic interpretation of the results follows.

We conclude that in all the countries of V4 a consumption tracks income very closely. The share of households consuming according to actual income is from two thirds to three quarters and uses to increase significantly due to the impact of crises. The largest part of PIH consumers is detected in CR but the group seems not to be very much stable; the impact of crises decreases it approximately by one half. Very similar influence of crises is apparent in case of Slovakia. In both countries, the permanent consumers form a group which is apparently very fragile and threatened by serious economic shocks. The PIH group in Poland is much more stable and the Hungarian PIH consumers form though a smaller however a robust part of the households of the country.

In [7] the author tried to answer the same question by using the same data but another computational method. The results for Czech Republic, Poland and Slovakia are almost equal, in case of Hungary there is a ten-percent fluctuation in favour of the ad hoc consumption.

The detailed knowledge of a consumption pattern can bring important implications for policy makers as well as for the producers and traders.

Acknowledgements

The financial support of IGA F4/34/2020 is gratefully acknowledged.

References

- [1] Campbell, J. Y. & Mankiw, N. G. (1990). Permanent Income, Current Income and Consumption, *Journal of Business & Economic Statistics*, Vol. 8 No. 3 (pp 265 – 279).
- [2] Dougherty, C. (2016) Introduction to econometrics, Oxford University Press.
- [3] Flavin, M. A. (1981). The Adjustment of Consumption to Changing Expectations about Future Income, *The Journal of Political Economy*, Vol. 89 No. 5 (pp 974 – 1009)
- [4] Friedman, M. A. (1957). *A Theory of Consumption Function*. Princeton Univ. Press
- [5] Hall, R. E. (1978). Stochastic Implications of the Life-Cycle-Permanent Income Hypothesis: Theory and Evidence. *Journal of Political economy*, Vol. 86 No. 6 (pp 971 – 987).
- [6] Kuan-Min, W. (2011). Does the Permanent Income Hypothesis Exist in 10 Asian Countries? *E+M Ekonomie a Management*, Vol. 14 No 4 (pp 1212 – 3609).
- [7] Pánková, V. (2021). Permanentní spotřeba v zemích V4. Competition 2021, Proceedings of conference VŠP (will be issued December 2021), accepted
- [8] Sargent, T. J. (1978) Rational Expectations, Econometric Exogeneity, and Consumption. *Journal of Political Economy*, Vol. 86, No. 4 (pp 673-700).
- [9] Sims, C. A. (2002). Implications of rational inattention, <http://pages.stern.nyu.edu/~dbackus/Exotic/1Robustness/Sims%20inattention%20JME%2003.pdf>

Multi Vehicle Routing Problem Depending on Vehicle Load

Juraj Pekár¹, Zuzana Čičková², Ivan Brezina³

Abstract: Nowadays, the emphasis in the transport planning is on its efficiency in connection with the ever-increasing environmental aspects. The related problems are solved in the field of logistics. In general, various optimization models aimed at minimizing the total distance traveled, minimizing the driving time of the vehicle, fuel consumption, CO₂ emissions or to meet other objectives have been developed. The known models of multiple vehicles enable to coordinate the deployment of several types of vehicles. This paper focuses on the problem of tracing several types of vehicles, which minimizes the fuel consumption of used vehicles depending on the length of the route, but also on their load. This idea is presented through a constructed multi vehicle routing problem depending on vehicle load (MVRPVL). The difference in the distribution compared to the multi vehicle routing problem illustrates the solving of the given problem in Slovakia.

Keywords: Multi Vehicle Routing Problem, Vehicle Load, Mathematical Model

JEL Classification: C02, C61

AMS Classification: 90C11, 90B06

1 Introduction

A variety of vehicle routing problems optimization models aimed at minimizing the total traveled distance (or the other related costs) are commonly known. Most models are based on capacitated vehicle routing problem (CVRP), which designs optimal set of routes of vehicles from depot (suppose unlimited number of the same vehicles in the depot) aimed to serve a set of customers with a known demand, where each vehicle travels exactly one route so that the demand of customers must to be met in full by exactly one vehicle and vehicle's capacity must not be exceeded. It is assuming the known lowest cost (usually distance) between depot and all of the customers, as well as between each pairs of customers ([1], [3], [4], [5], [7], [9]) and its application can led to significant cost savings.

However, when analyzing the transport costs fuel consumption, it is clearly observed that they are related to fuel consumption. That means that the traveled distance is not the only relevant factor and undoubtedly also vehicle load has significant impact on the consumption. This idea was introduced in [3], where the authors presented capacitated vehicle routing problem depending on vehicle route. The paper was aimed on a model that minimizes the fuel consumption, depending on the length of the traveled route and also on the vehicle load. Let's extend this idea and present *multi vehicle routing problem depending on vehicle load* (MVRPVL). This paper is divided into following interrelated parts: In the first part of second section we present presuppositions and a mathematical model of MVRPVL. At first, we present mixed integer non-linear formulation of the problem and then its modification to mixed integer programming model that contains linear equations and linear objective. The third part is devoted to an illustrative example, where we calculate optimum route based on CVRP as well as on a CVRPVL. The results illustrate the difference using both of approaches, while we also report an achieved decrease in the fuel consumption.

The presented approach is an extension of classical routing problems with the aim of reducing CO₂ emissions while minimizing fuel consumption. The result of MVRPVL is the delivery of a certain commodity from one service center to individual customers, which require the delivery of known quantities of commodity, provided that a transport with different vehicles with limited capacity. However, the modified objective function reflects not only the shortest distances, which represent the evaluation between nodes in the transport network, but also the set of considered vehicles with known and different fuel consumption per unit of distance and also additional vehicle fuel consumption per tonne of payload. This is related to payload weight of the vehicle to the customer served.

¹ Department of Operations Research and Econometrics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemska cesta 1, 852 35 Bratislava, e-mail: juraj.pekar@euba.sk.

² Department of Operations Research and Econometrics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemska cesta 1, 852 35 Bratislava, e-mail: zuzana.cickova@euba.sk

³ Department of Operations Research and Econometrics, Faculty of Economic Informatics, University of Economics in Bratislava, Dolnozemska cesta 1, 852 35 Bratislava, e-mail: ivan.brezina@euba.sk.

2 Multi Vehicle Routing Problem Depending on Vehicle Load

The mathematical formulation of multi vehicle routing problem depending on vehicle load (MVRPVL) is based on Miller-Tucker-Zemlin's formulation of the traveling salesman problem ([8]). Let's use the following notation: Let $N = \{1, 2, \dots, n\}$ be the set of served nodes (customers) and let $N_0 = N \cup \{0\}$ be a set of nodes that represents the customers as well as the origin (depot). Certain demand q_i , $i \in N$ is associated with each customer. Further on there exists a matrix $D_{(n+1) \times (n+1)}$ associated with pairs $i, j \in N_0$, $i \neq j$ that represents the minimum distances between all the pairs of nodes (customers and the depot). There are different types of vehicles in unlimited numbers in the depot, which are represented by the set $H = \{1, 2, \dots, h\}$. Each vehicle has a certain capacity, designated as g_k , $k \in H$. All the customer's demands have to be met from the depot in full and in such a way that the distribution is performed by exactly one of the vehicles. We implicitly assume that $q_i \leq \min_k q_k$, $i \in N$, $k \in H$, i.e. the demand of each customer does not exceed the capacity at least one of the vehicles. Now consider parameters associated with the fuel consumption, that can differ according to vehicle type. Consider parameter a_k and b_k , which are related to consumption of the k -th vehicle, $k \in H$, when the parameter a_k represents the consumption per unit distance and the parameter b_k be parameter represents the increase in consumption for one unit per unit distance.

We will use binary variables x_{ijk} ($i, j \in N_0$, $i \neq j$, $k \in H$) that determine if the node i precedes node j in the route of k -th vehicle (if yes: $x_{ijk} = 1$ and if not: $x_{ijk} = 0$) in the case of distribution. In the case of collection those variables determine if node j precedes node i in the route of k -th vehicle. Further on, we will use variables u_i , $i \in N$. When suppose that the goods are collected, those variables represent cumulative load of vehicle on its one particular route (to i -th customer). On the other hand, they represent cumulative load of vehicle (from i -th customer) on its one particular route (note the meaning of binary variables x_{ij} differ depending on problem type (collection or distribution)).

Unlike the classical CVRP [11], the goal is to find out such distribution, which minimizes whole vehicle's fuel consumption (not minimizing the total distance).

Let us recap the variables and parameters once again:

Sets and parameters: n – number of customers (served nodes),

$N = \{1, 2, \dots, n\}$ – set of customers (served nodes),

$N_0 = N \cup \{0\}$ – set of customers and the depot,

d_{ij} , $i, j \in N_0$, $i \neq j$ – shortest distance between nodes i to node j ,

q_i , $i \in N$ ($q_0 = 0$) – demand of i -th customer, zero demand in the depot

g_k – capacity of k -th vehicle, $k \in H$,

a_k – consumption of the k -th vehicle per unit distance, $k \in H$,

b_k – increase in consumption for one unit of k -th vehicle vehicle load per unit distance, $k \in H$.

Variables: $x_{ijk} \in \{0, 1\}$, $i, j \in N_0$, $i \neq j$, $k \in H$ representing if the node i precedes node j on the route of k -th vehicle,

$u_{ik} \geq 0$, $i \in N$, $k \in H$, $u_{0k} = 0$, representing vehicle load to i -th node (including) in the case of goods collection.

Mathematical model:

$$\min f(\mathbf{X}) = \sum_{i \in N_0} \sum_{j \in N} \sum_{k \in H} (d_{ij} (a_k x_{ijk} + b_k u_{jk} x_{ijk})) \quad (1)$$

$$\sum_{i \in N_0} \sum_{k \in H} x_{ijk} = 1, \quad j \in N, \quad i \neq j \quad (2)$$

$$\sum_{j \in N_0} \sum_{k \in H} x_{ijk} = 1, \quad i \in N, \quad i \neq j \quad (3)$$

$$\sum_{i \in N_0} x_{ijk} = \sum_{i \in N_0} x_{jik}, \quad j \in N, \quad k \in H, \quad i \neq j \quad (4)$$

$$x_{ijk} (u_{ik} + q_i - u_{jk}) = 0, \quad i \in N_0, \quad j \in N, \quad i \neq j, \quad k \in H \quad (5)$$

$$q_i \sum_{j \in N_0} x_{ijk} \leq u_{ik} \leq g_k, \quad i \in N, \quad k \in H \quad (6)$$

Objective function (1) determines whole fuel consumption: minimizing the first part of the sum enables modeling the fuel consumption depending on the length of the route and the second part enables modeling increased consumption depending on the vehicles' load. Equations (2) and (3) ensure that each customer (except the depot) is visited exactly ones. Equation (4) ensures the connectivity of the route. Equations (5) prevent the formation of such sub-cycles, which do not contain the depot, and they enable calculating values of variables u_i , $i \in N$ (together with equations (6)).

The model of MVRPVL (1)-(6) contains also nonlinear equations (1) and (5). This may complicate the possibility of solving of related problems, because they belongs to mixed integer non-linear programming problems (MINLP). The better way to solve such problems is to use mixed integer programming formulation (MIP) that includes only the linear constraints [3]. Let us provide that. Further on, we will use the non-negative variables $n_{ijk} \geq 0$, $i, j \in N_0, k \in H, i \neq j$ that represent the current load of the k -th vehicle on the edge (i, j) .

Now we can rewrite the objective (1) in the form:

$$\min f(\mathbf{X}) = \sum_{i \in N_0} \sum_{j \in N} \sum_{k \in H} \left(d_{ij} (a_k x_{ijk} + b_k n_{ijk}) \right) \quad (7)$$

where the variables n_{ijk} are calculated as:

$$u_{ik} \leq n_{ijk} + (1 - x_{ijk}) g_k, \quad i, j \in N_0, \quad i \neq j, \quad k \in H \quad (8)$$

The equation (8) enables calculation u_{ik} and n_{ijk} according to the value of parameter g_k (capacity of the k -th vehicle) when the variable $x_{ijk} = 1$ and it also ensures feasibility of solution in the case that $x_{ijk} = 0$.

Then the first part of the sum in (7) represents the fuel consumption dependence on the traveled route and its second part enables modeling increased consumption dependence on the vehicle's load.

For the sake of completeness, we also present the linearization of the equations (5), even if it is trivial [3]:

$$u_{ik} + q_i - g_k (1 - x_{ijk}) \leq u_{jk}, \quad i \in N_0, \quad j \in N, \quad i \neq j, \quad k \in H$$

Let us now present an illustrative example containing cities in Slovakia.

3 Illustrative Example

Consider distribution network consisting of origin (0 – Bratislava) from where eight customers need to be served (1 – Banská Bystrica, 2 – Košice, 3 – Nitra, 4 – Prešov, 5 – Trenčín, 6 – Trnava, 7 – Žilina). Values of input parameters were set as follows:

$$n = 8, \quad N = \{1, 2, \dots, 7\}, \quad N_0 = N \cup \{0\} - \text{number and sets of nodes (customers and also origin),}$$

$$\mathbf{q} = (5, 2, 10, 10, 6, 8, 9)^T - \text{vector of customers' demands,}$$

$$g_1 = 24 - \text{capacity of 1}^{\text{st}} \text{ type of vehicles,}$$

$$g_2 = 10 - \text{capacity of 2}^{\text{nd}} \text{ type of vehicles,}$$

$$a_1 = 0.22 - \text{vehicle consumption per unit distance (liter/km),}$$

$$b_1 = 0.007 - \text{increased consumption per unit distance and per unit of load (liter/tonne/km),}$$

$$a_2 = 0.2 - \text{vehicle consumption per unit distance (liter/km),}$$

$$b_2 = 0.005 - \text{increased consumption per unit distance and per unit of load (liter/tonne/km),}$$

$$\mathbf{D} = \{d_{ij}\}, \quad i, j \in N_0, \quad i \neq j - \text{matrix of shortest distances between node } i \text{ to node } j$$

Firstly we solve the problem using aforementioned linearized model of MVRPVL and also using the classical multiple vehicle routing problem (MVRP), see for example [2]. Both models are implemented in

software GAMS on PC with Intel ® Core ™ i7-3770 CPU with a frequency of 3.40 GHz and 8 GB of RAM under MS Windows 8. Results are given in Table 1 (for MVRPVL) and in Table 2 (for MVRP).

$$D = \begin{pmatrix} 0 & 200 & 400 & 100 & 400 & 100 & 50 & 200 \\ 200 & 0 & 200 & 120 & 200 & 150 & 170 & 100 \\ 400 & 200 & 0 & 300 & 40 & 300 & 350 & 230 \\ 100 & 120 & 300 & 0 & 300 & 90 & 50 & 140 \\ 400 & 200 & 40 & 300 & 0 & 300 & 360 & 220 \\ 100 & 150 & 300 & 90 & 300 & 0 & 80 & 70 \\ 50 & 170 & 350 & 50 & 360 & 80 & 0 & 150 \\ 200 & 100 & 230 & 140 & 220 & 70 & 150 & 0 \end{pmatrix}$$

Route	Sequence	Distance	Fuel consumption without load consumption	Fuel consumption depending on load
Route Large vehicle	0-5-7-1-0	470	103.4	127.76
Route 1 Large vehicle	0-3-4-2-0	840	184.8	225.96
Route 2 Small vehicle	0-6-0	100	20	22
Total		1410	308.2	375.72

Table 1 MVRPVL. *Source:* Own compilation.

Route	Sequence	Distance	Fuel consumption without load consumption	Fuel consumption depending on load
Route Large vehicle	0-1-2-4-5-0	840	184.8	263.48
Route 1 Large vehicle	0-3-7-0	440	96.8	118.92
Route 2 Small vehicle	0-6-0	100	20	22
Total		1380	301.6	404.4

Table 2 MVRP. *Source:* Own compilation.

When we use the two different models (MVRPVL and MVRP), the optimal route length changed. In the case of implementing MVRP we obtain the value of the total traveled distance 1380 km. Using model of MVRPVL leads to route length of 1410 km. However, when we calculate the fuel consumption with dependence on vehicle load, the resulting values are those: 404.4 l in the case of MVRP and 375.72 l in the case of MVRPVL. Using MVRPVL model decreased fuel consumption by 6.12 percent. The change can be clearly attributed to rearranging of nodes to fulfill the goal of minimizing the fuel consumption with respect to the vehicle load, which results in different routes structure.

Conclusion

In this paper we present the modification of multiple vehicle routing problem, which we named multi vehicle routing problem depending on vehicle load (MVRPVL). Such model enables to minimize fuel consumption taking into account not only traveled distance but also the weight of the loaded goods and it is aimed to reduction of CO2 emissions taking into account the environmental aspects. The first part of the paper is devoted to presupposition and non-linear mathematical model of MVRPVL and we present its version containing only linear objective as well as linear constraints, which can be easier to solve. In the second part of the paper, we present an illustrative example. The results are compared to classical multiple vehicle routing problem. The results show that the weight of the load largely affects the individual routes and in that way fuel consumption.

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“.

This work was supported by the Slovak research and development agency, grant no. SK-SRB-18-0009 „Optimizing of logistics and transportation processes based on the use of battery operated vehicles and ICT solutions“

References

- [1] Čičková, Z., Brezina, I., Pekár, J.: Solving the routing problems with time windows. In: *Self-organizing migrating algorithm : methodology and implementation* (editors: Donald Davendra, Ivan Zelinka.). Springer International Publishing AG Switzerland, Cham, 2016, 207-236.
- [2] Čičková, Z., Brezina, I., Pekár, J.: Routes design using own and hired vehicles. In: *Mathematical methods in economics 2015*. University of West Bohemia, Cheb, 2015, 105-108.
- [3] Čičková, Z., Brezina, I., Pekár, J.: Capacitated vehicle routing problem depending on vehicle load. In *Mathematical methods in economics 2017*. Hradec Králové : Gaudeamus, 2017, 108-112.
- [4] Eksioğlu, B., Vural, A. V., Reisman, A.: The vehicle routing problem: A taxonomic review. *Computers & Industrial Engineering* **57**(4/2009), 1472–1483.
- [5] Fábry, J., Kořenář, V., Kobzareva, M.: Column Generation Method for the Vehicle Routing Problem – The Case Study. In: *Mathematical Methods in Economics 2011*. PROFESSIONAL PUBLISHING, Praha, 2011, 140–144.
- [6] Golden, B. L., Raghavan, S., Wasil, E. A.: *The Vehicle Routing Problem: Latest Advances and New Challenges*. Springer, New York, 2008.
- [7] Keçeci, B., Altıparmak, F., & Kara, İ.: A mathematical formulation and heuristic approach for the heterogeneous fixed fleet vehicle routing problem with simultaneous pickup and delivery. *Journal of Industrial & Management Optimization*, 17(3/2021), 1069–1100.
- [8] Miller, C.E., Tucker, A.W., Zemlin, R.A.: Integer programming Formulation of Traveling Salesman Problems. *Journal of the ACM* **7**(4/1960), 326-329.
- [9] Pekár, J., Brezina, I., Čičková, Z.: Synchronization of capacitated vehicle routing problem among periods. *Ekonomický časopis* **65** (1/2017), 66-78.
- [10] Pekár, J., Brezina, I., Čičková, Z., Reiff, M.: *Modelovanie rozmiestňovania recyklačných centier*. EKONÓM, Bratislava, 2012.
- [11] Toth, P., Vigo, D.: *The Vehicle Routing Problem (Monographs on Discrete Mathematics and Applications)*. SIAM. Philadelphia, 2002.

Using Parametric Resampling in Process of Portfolio Optimization

Juraj Pekár¹, Mário Pčolár²

Abstract. The results of optimization models using input data in the form of historical observations are often "unstable", as even with small changes in the expected risk or return of the assets, there are significant changes in the portfolio allocation. Such behavior is a consequence of the optimization models themselves whose essence of finding the extrema leads to the problem of maximizing the estimation error. The aim of this paper is to examine the possibility of applying the resampling and the Monte Carlo simulation in order to improve the application performance of the resulting portfolios on data out-of-sample. In the optimization procedure we are using mean absolute deviation model and daily data of the Dow Jones Industrial Average index components. The performance of the portfolios thus created is compared on the data out-of-sample, together with the portfolios obtained in accordance with the historical data and 1/n portfolio.

Keywords: resampling, generalized lambda distribution, MAD, error maximization

JEL Classification: G11, G17

AMS Classification: 90-10

1 Introduction

Portfolio optimization models depend on a large extent on the quality of the quantified (estimated) input data required. These values are traditionally quantified from historical observations and, as they are estimates, they often introduce a significant amount of estimation error into the model. As Mikkel Rasmussen stated on the sidelines of this problem "it could be said that optimization algorithms are simply too strong in terms of the quality of the data that make up the input of a given procedure"[10]. The problem with insufficient quality of estimations, in the case of optimization algorithms looking for extremes, leads to maximization of the estimation error, which can lead to unstable portfolios which, on the one hand are non-intuitive and at worst unusable from an investor's point of view. This lack of use of historical data can be reduced by implementing statistical methods of bootstrapping and resampling together with the Monte Carlo method into optimization models, which should lead to more stable and more diversified portfolios compared to the traditional procedure [4], [7], [10]. The method of resampling within optimization models was reviewed by several authors, especially when using this procedure in the case of the Markowitz model and the use of data on a monthly or weekly basis. Our article contributes to empirical research by analyzing the use of the resampling procedure within the Mean absolute deviation model (MAD) and in the case of using daily data as an input of optimization.

The aim of this paper is to examine the possibility of applying the Monte Carlo method on the basis of a selected model of probability distribution within the model of portfolio selection in space of expected return and mean absolute deviation. The analysis of the data shows that the data have leptokurtic behavior, so the hypothesis of a normal distribution could be rejected. The pseudorandom data used in the Monte Carlo method are generated using a normal distribution as well as a generalized lambda distribution. We have chosen the generalized lambda distribution (GLD) as a model of the distribution of daily returns on the basis of several contributions, pointing to the given distribution as an interesting alternative for the needs of modeling the daily returns of financial assets [1], [2]. The given distribution offers a significantly better quality of fit in comparison with distributions commonly used in practice, such as the Student's t-distribution, the Laplace distribution or the Cauchy distribution. The distribution parameters are estimated using the maximum likelihood method. All calculations contained in this paper were performed in the RStudio environment.

¹ University of Economics in Bratislava, Department of Operations Research and Econometrics, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia, juraj.pekar@euba.sk

² University of Economics in Bratislava, Department of Operations Research and Econometrics, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia, mario.pcolar@euba.sk

2 Generalized lambda distribution

Generalized lambda distribution is a family of distributions which are characterized by a wide range of shapes of distributions that can be acquired by. This is a modification and extension of the original Tukey's lambda distribution. In our paper we consider FMKL parameterization [3]. Defined by a quantile function with four parameters:

$$Q(y) = \lambda_1 + \frac{1}{\lambda_2} \left[\frac{y^{\lambda_3} - 1}{\lambda_3} - \frac{(1-y)^{\lambda_4} - 1}{\lambda_4} \right]$$

where $0 \leq y \leq 1$. The individual parameters of gld describe: λ_1 represents the location parameter, λ_2 represents the scale parameter and λ_3, λ_4 define the shape of the tails of distribution. In the case of $\lambda_3 = \lambda_4$ the distribution is symmetric. The individual parameters of gld are real numbers with a constraint in the case of $\lambda_2 > 0$. K-th moment of FMKL GLD exists only if $\min(\lambda_3, \lambda_4) > -k^{-1}$. We use the method of maximum likelihood estimation to the estimation of parameters.

3 Portfolio selection model in space of expected return and absolute deviation

It is an optimization model of portfolio selection based on the mean absolute deviation risk measure, also called the mean absolute deviation model (MAD) [5]. It is a model of linear programming, which in the case, when the returns are normally distributed, gives the same results as the Markowitz model. The advantage of MAD is that it does not impose a condition on data distribution. The formulation of the MAD model in space of expected return and absolute deviation with the objective function of minimizing the mean absolute deviation has the form:

$$\begin{aligned} \min & \frac{1}{T} \sum_{t \in T} y_t \\ & y_t - \sum_{j \in N} w_j (r_{j,t} - r_j) \geq 0, \quad t \in T \\ & y_t + \sum_{j \in N} w_j (r_{j,t} - r_j) \geq 0, \quad t \in T \\ & \sum_{j \in N} r_j w_j \geq E_p \\ & \sum_{j \in N} w_j = 1 \\ & 0 \leq w_j \leq 1, \quad j \in N \end{aligned}$$

where r_j represents the average expected return of the asset, $r_{j,t}$ represents the return of the asset j in time t , w_j is the weight of asset j , E_p represents the minimum expected return of the investor, y_t represents the absolute deviation in state t , where $t = 1, \dots, T$ represents the number of states and $j = 1, \dots, N$ represents the number of assets.

4 Portfolio resampling

The Resampling method is based on the use of the Monte Carlo simulation method and the use of stochastic interpretation of returns using a probability distribution. The use of this method is based on the idea that the input parameters used in the optimization are derived from historical returns which represent one representation of the stochastic process which describes the realization of returns. The use of the resampling method with portfolio selection models belongs to the heuristic methods [11]. Its main goal is to solve the so-called problem of error maximization and increase portfolio diversification in order to achieve more usable resulting portfolios on data out-of-sample. The use of the resampling method within the optimization process in portfolio selection models was presented in [8]. In our paper, we use a modified process of resampling, through the following steps

1. We estimate the parameters of the selected parametric distribution model of daily returns from the historical returns of individual assets with the T observations. To preserved the correlation between individual assets, we are using the rank of the observations. Individual observations are denoted by an index

from 1 to T, as they are observed in time. Subsequently, we sort the observations of returns of individual assets from the smallest to the largest and keep the order of the indices, in the case of ascending order.

2. Subsequently, we generate a vector of random realizations of returns of length T from the selected parametric distributional model, using the estimated parameters from historical returns for individual assets. Then we sort the generated vector of random realizations of returns in ascending order and assign it a vector of indices obtained from the historical data of this asset. Subsequently, the vector of random realizations of returns is sorted in accordance with the vector of indices. Such a procedure allows us to simulate the correlation links between individual assets using the rank.
3. We will use the generated vectors of random realizations of individual assets within the selected portfolio selection model. We calculate the efficient frontier, and keep the optimal weights for M evenly distributed portfolios located at the efficient frontier, between the minimum risk portfolio and the maximum return portfolio.
4. Repeat steps 2 and 3 many times, then averaging the weights of the portfolios that share the same rank. Then we get resampled frontier.

The resulting portfolios represent portfolios that are more diversified compared to the classical optimization approach, and also such a procedure should help to reduce the problem of error maximization. It is also possible to use a significantly more sophisticated procedure to maintain the correlation structure by generating data from a multivariate distribution using estimated covariance matrix. In the case of using a normal distribution, this procedure is well known; in the case of using a non-normal distribution, it is significantly more complicated. For the reader who is interested in this issue, we recommend an interesting article which deals with the generation of a random vector from a multivariate non-Gaussian distribution [6].

5 Data Results

We perform computational experiments on data about daily adjusted close positions of 30 assets from 01.01.2000 to 31.12.2007. The assets consisted of DJIA components. Daily returns are quantified as continuously compounded returns, i.e. as the first differences of natural logarithms of each series. We will use the series of daily returns to estimation of the parameters of the normal distribution and the parameters of the generalized lambda distribution, which we will be used in the procedure of resampling. To fit the distribution to the empirical data of each asset, we use the maximum likelihood estimation method. In Monte Carlo simulations, we performed 500 simulations, and in portfolio optimization with the MAD model, we generated 50 evenly distributed portfolios at the efficient frontier.

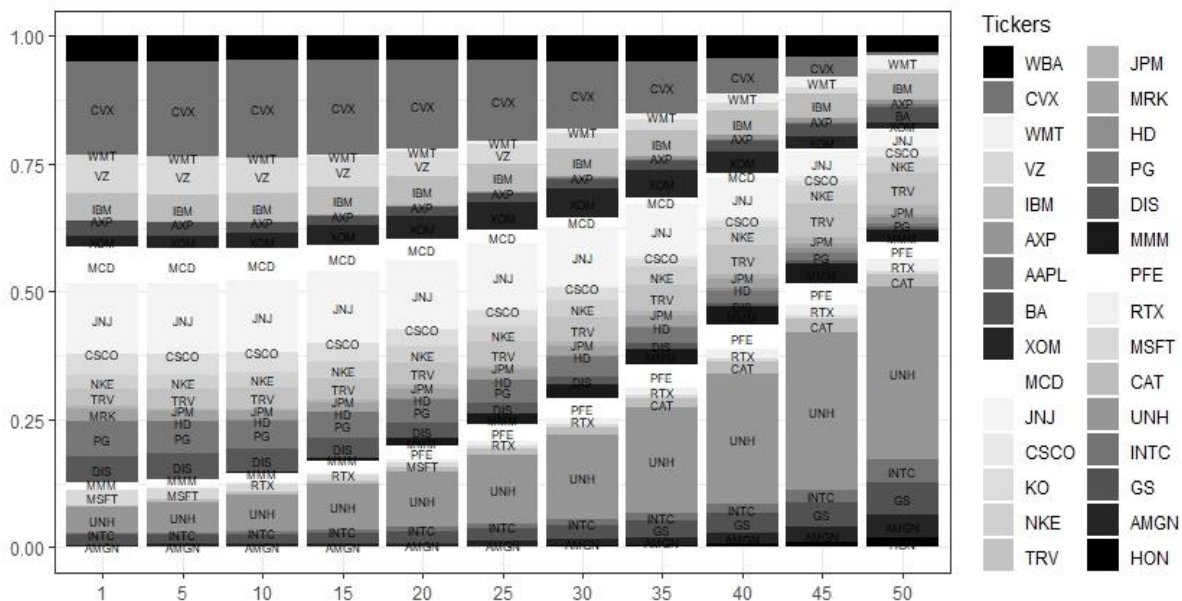


Figure 1 Composition of selected portfolios from resampling method using GLD

Within the section "Data Results" we will compare portfolios generated by the classical approach using estimates from historical data (Hist), portfolios generated using the modified resampling procedure from section 4 with the

normal distribution as a parametric model of daily returns (Res Norm), portfolios generated using the above procedure with the generalized lambda distribution as a model of daily returns (Res GLD) and 1/n portfolio (Even). We used 2 years of data from 01.01.2000 to 31.12.2001 to estimate the parameters of the distributions. From each method, we selected 4 representative portfolios, namely the portfolio with the minimum level of risk, the portfolio with the maximum expected return and two portfolios located evenly distributed on the efficient frontier between them. We will compare the portfolios selected in this way from the point of view of three different investment horizons, namely one-year horizon, two-and-a-half-year horizon and five-year horizon. However, investors cannot sell within this investment horizon. They can sell the assets only during the next year after the investment horizon. To compare individual portfolios, we will quantify the average characteristics of portfolios and selected performance measures for period of one year (always one year after the investment horizon) within which the investor can sell assets. Specifically, for the year 2003, for the period from 01.06.2004 to 01.06.2005 and for the year 2007. In the analysis, we quantify the average return and average risk (using standard deviation and mean absolute deviation as risk measures), from the performance measures, it is Sharpe ratio and modified Sharpe ratio using mean absolute deviation sometimes also called as mean absolute deviation ratio.

Figure 1 and Figure 2 are stacked bar charts, where one column is composed of smaller bars whose height represents the size of the weight of the selected asset. The X-axis represents the rank of the portfolio at efficient frontier and the Y-axis represents cumulated weights of each assets, the sum of which is set to 1 in accordance with the model. In each column are assets ordered the same, there is always a WBA at the top of the bar and HON at the bottom, the remaining assets are arranged in the order in which they are listed (from top: 1st WBA, 2nd CVX and so on...).

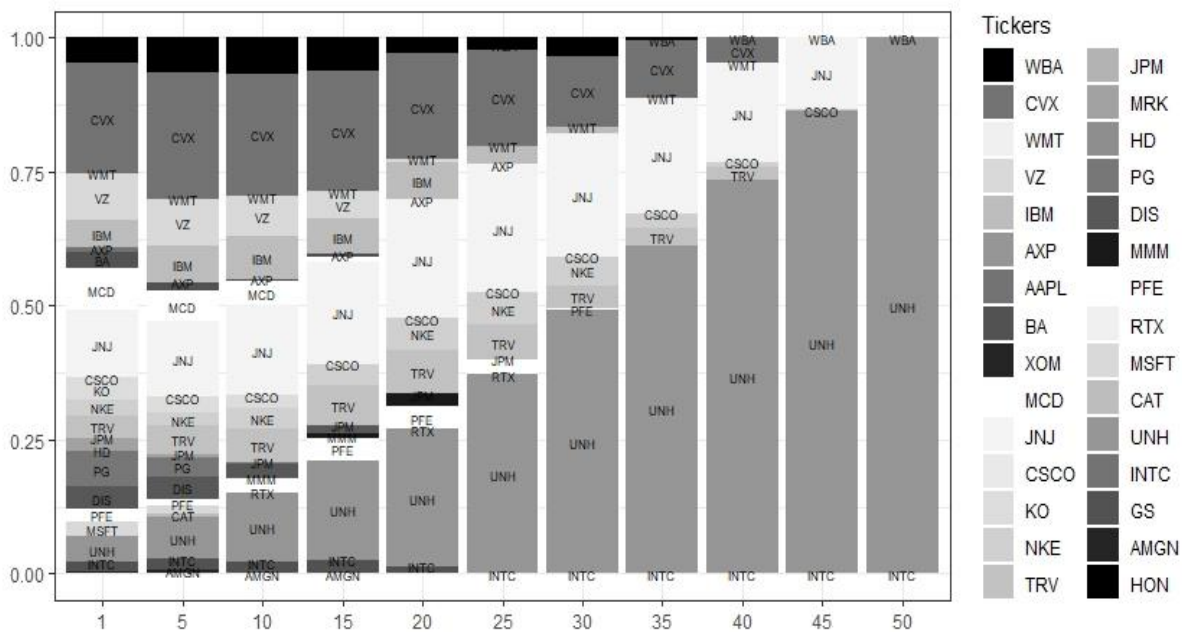


Figure 2 Composition of selected portfolios from MAD using historical data

One of the main differences between the use of the classical approach and the resampling method are more significantly diversified portfolios. In the case of comparing portfolios with a minimal level of risk, the difference in portfolio diversification is not so significant, on the other hand by a gradual shift in ranks to the right, it is clear that portfolios obtained by resampling are much more diversified accordance with portfolios obtained by classical approach. This could be an advantage in comparing the performance of individual portfolios on out-of-sample data

Table 1 shows quantified statistics and performance measures for selected representative portfolios and investment horizons, according to which we compare two resampling procedures used in the portfolio optimization process using the MAD model together with an optimization model using estimates from historical data and evenly distributed portfolio. Selected representative portfolios consist of a portfolio with a minimum risk (rank 1), a portfolio with a maximum return (rank 50) and two other portfolios with an average return (rank 16 and rank 34). The lower-ranked portfolio the lighter shade of gray. The individual representative portfolios are distinguished in the table

by color according to rank. The best values for individual representative portfolios and investment horizons are underlined in the table. When comparing, we do not take into account the values of an evenly distributed portfolio.

In line with the quantified values, it follows that within the one-year and five-year investment horizons, most portfolios obtained using resampling procedure achieve better values compared to portfolios values quantified from historical data. In these investment horizons, comparable values, from portfolios obtained using estimates from historical data, reach only the portfolio with minimal risk (rank 1). In the case of a two-and-a-half-year investment horizon, the portfolios obtained by the classical approach clearly achieve better results. Portfolios obtained through the use of resampling and the modeling of daily returns using the generalized lambda distribution are characterized by the lowest average risk for most portfolios and investment horizons. When comparing a procedure using a normal distribution compared to a procedure using a generalized lambda distribution, on average, portfolios constructed using a normal distribution as a model of daily returns achieve better results from perspective of performance measures. An evenly distributed portfolio is, on average, characterized by a higher degree of risk compared to portfolios acquired through the use of resampling.

		1 Year					2.5 Year					5 Year				
		Mean	Std	Mad	Sharpe Ratio	MAD Ratio	Mean	Std	Mad	Sharpe Ratio	MAD Ratio	Mean	Std	Mad	Sharpe Ratio	MAD Ratio
Hist		0,080%	<u>0,934%</u>	0,848%	8,555%	9,415%	0,049%	0,639%	0,655%	7,673%	7,487%	0,059%	0,838%	<u>0,678%</u>	<u>7,068%</u>	<u>8,725%</u>
		0,084%	0,927%	0,823%	9,105%	10,253%	<u>0,067%</u>	0,711%	0,715%	<u>9,381%</u>	<u>9,332%</u>	0,041%	0,861%	0,688%	4,795%	5,997%
		0,099%	1,149%	1,078%	8,655%	9,222%	<u>0,122%</u>	1,010%	0,855%	<u>12,059%</u>	<u>14,259%</u>	0,036%	0,933%	0,841%	3,900%	4,324%
		0,132%	1,589%	1,325%	8,294%	9,949%	<u>0,158%</u>	1,480%	1,151%	<u>10,667%</u>	<u>13,717%</u>	0,032%	1,270%	1,127%	2,525%	2,843%
Res GLD		0,080%	0,937%	<u>0,847%</u>	8,544%	9,444%	0,050%	<u>0,635%</u>	0,646%	7,857%	7,724%	0,055%	<u>0,835%</u>	0,685%	6,644%	8,099%
		0,086%	<u>0,919%</u>	<u>0,801%</u>	9,368%	<u>10,743%</u>	0,053%	<u>0,654%</u>	<u>0,655%</u>	8,111%	8,105%	<u>0,053%</u>	<u>0,847%</u>	<u>0,644%</u>	<u>6,223%</u>	<u>8,185%</u>
		0,097%	<u>0,941%</u>	<u>0,834%</u>	10,359%	<u>11,685%</u>	0,059%	<u>0,702%</u>	0,706%	8,451%	8,403%	<u>0,040%</u>	0,850%	0,641%	<u>4,759%</u>	<u>6,318%</u>
		0,125%	<u>1,083%</u>	<u>0,986%</u>	11,496%	12,634%	0,069%	<u>0,819%</u>	<u>0,791%</u>	8,393%	8,691%	0,026%	<u>0,921%</u>	<u>0,721%</u>	2,822%	3,600%
Even		0,115%	1,085%	0,978%	10,585%	11,747%	0,041%	0,698%	0,734%	5,889%	5,600%	0,044%	0,927%	0,640%	4,694%	6,802%
		0,115%	1,085%	0,978%	10,585%	11,747%	0,041%	0,698%	0,734%	5,889%	5,600%	0,044%	0,927%	0,640%	4,694%	6,802%
		0,115%	1,085%	0,978%	10,585%	11,747%	0,041%	0,698%	0,734%	5,889%	5,600%	0,044%	0,927%	0,640%	4,694%	6,802%
		0,115%	1,085%	0,978%	10,585%	11,747%	0,041%	0,698%	0,734%	5,889%	5,600%	0,044%	0,927%	0,640%	4,694%	6,802%
Res Norm		0,082%	0,950%	0,858%	<u>8,612%</u>	<u>9,544%</u>	<u>0,054%</u>	0,641%	<u>0,645%</u>	<u>8,364%</u>	<u>8,313%</u>	0,055%	0,846%	0,689%	6,450%	7,925%
		<u>0,088%</u>	0,929%	0,833%	<u>9,437%</u>	10,524%	0,056%	0,656%	0,672%	8,587%	8,393%	0,053%	0,848%	0,662%	6,195%	7,930%
		<u>0,100%</u>	0,959%	0,889%	<u>10,419%</u>	11,237%	0,062%	0,708%	<u>0,683%</u>	8,701%	9,015%	0,040%	<u>0,844%</u>	<u>0,628%</u>	4,688%	6,296%
		<u>0,134%</u>	1,088%	0,988%	<u>12,322%</u>	<u>13,568%</u>	0,081%	0,849%	0,819%	9,557%	9,906%	0,033%	0,932%	0,728%	<u>3,560%</u>	<u>4,559%</u>

Table 1 Quantified statistics and performance ratios for each representative portfolio and investment horizon

6 Conclusion

The article deals with the application of the resampling procedure within the portfolio optimization process. The use of the resampling procedure should help to reduce the effect of the estimation error that is commonly present in portfolio selection models. The aim of applying the procedure described in section 4 is to construct portfolios that perform better on out-of-sample compared to the classical approach using estimates from historical data. Several articles have analyzed the resampling applications using the Markowitz model and monthly or weekly returns data, but little attention has been tribute to other commonly used models in portfolio optimization. The aim of this paper is to contribute to empirical research with analysis of applications of resampling procedure in process of optimization using MAD model and daily data. We applying a modified resampling procedure using rank of data as a tool to maintain the correlation between assets. The application of such procedure allows a simple replacement of the commonly used model of normal distribution by other models that could better capture the characteristics of daily returns. In our paper, we applying procedures using normal distribution and generalized lambda distribution.

We performed a computational experiment on data of daily asset returns from the period from 2000 to 2007, which formed the components of the DJIA index. In our analysis, we quantified average statistics and performance measures over established investment horizons. Consistent with the results obtained, it follows that the application

of the procedure outlined in section 4 using both normal and GLD distributions leads to portfolios that are more diversified. On average, portfolios quantified using procedure of resampling achieved better values of quantified statistics and performance measures over the one-year and five-year investment horizons. Portfolios quantified using a generalized lambda distribution are characterized by the lowest risk on average.

Such results point to the potential in applications of resampling within the portfolio optimization procedure using the MAD model. Achieving more accurate conclusions requires a more detailed analysis over a longer data horizon and the use of more robust techniques. For the future research we also see the space to examine a procedure that uses data generation from a multivariate generalized lambda distribution as mentioned in the article [6].

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] Chalabi, Y., Scott, D. & Wurtz, D. (2009). The Generalized Lambda Distribution as an Alternative Model to Financial Returns. In: *ETH Econohysics Working and White Papers Series*, 1-29.
- [2] Corlu, G. C., Meterelliyoz, M. & Tiniç, M. (2016). Empirical distributions of daily equity index returns: A comparison. *Expert systems with applications*, 54, 170-192.
- [3] Freimer M, Kollia G, Mudholkar, G. S. & Lin C. T. (1988). A study of the generalized Tukey lambda family. *Communications in Statistics-Theory and Methods*, 17, 3547-3567.
- [4] Jorion, P. (1992). Portfolio Optimization in Practice. *Financial Analysts Journal*, 48, 68-74.
- [5] Konno, H. & Yamazaki, H. (1991). Mean-Absolute Deviation Portfolio Optimization and Its Applications to Tokyo Stock Market, *Management Science* 37, 519-531.
- [6] Lange, A., Sohrmann, C., Jancke, R., Haase, J., Cheng, B., Kovac, U. & Asenov, A. (2011) A general approach for multivariate statistical MOSFET compact modeling preserving correlations. *Proceedings of the European 2011* (pp. 163-166). Helsinki: Institute of Electrical and Electronics Engineers
- [7] Mateusz, D. (2021). Sampling methods for investment portfolio formulation procedure at increased market volatility. *Journal of Economics & Management*, 43, 74-93.
- [8] Michaud, O. R. & Michud, O. R. (1998). *Efficient Asset Management*. Boston, MA: Harvard Business School Press
- [9] Radovanov, B. & Marcikić, A. (2012). Usefulness of bootstrapping in portfolio management. *Croatian Operational Research Review*, 3, 68-79.
- [10] Rasmussen, M. (2003). *Quantitative portfolio optimisation, asset allocation and risk management*. New York: Palgrave Macmillan.
- [11] Scherer, B. (2002). Portfolio Resampling: Review and Critique. *Financial Analyst Journal*, 58, 98-109.

Optimal routing order-pickers in a warehouse

Jan Pelikán¹

Abstract. Order picking is an important process in distribution centers. This study is based on the case study of a wholesale warehouse. Each order consists of a number of items to be picked at some location of the warehouse. There are one or more locations for each item. Start point of the pickers is a stand of the picker, end point is a drop-off location. Distances between the locations (nodes of graph) are given, a set of items (with its quantity) for the order, too. A mathematical model is a modification of the travelling salesman problem. In addition, a consolidation of the warehouse is modeled. This problem consists in assemble the same goods from different locations in one location (pallet, shelf). A numerical example is shown in the paper.

Keywords: picker routing, vehicle routing problem, integer programming

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

The topic of this article is focused on picking up items from the wholesale warehouse. Items are placed at pallet racks. The picker is used for collection of the items. The picker moves in aisles between the racks and according to content of an order it picks up the items from the racks. The aim is to achieve the minimal time needed for picking up the requested items. The warehouse can contain in practice tens of thousands items and they can solve daily hundreds of orders. In [1] is solved the problem in which there are three sections in picker and it can be placed items of three different orders in the picker. Therefore, the efficiency of picking up the items increases mainly in case, when these three orders are more or less related to the same item.

This problem is related to the problem of searching three orders with similar structure of requested item, respectively for a daily list of orders to create these triplets. After that there is solved one route of the picker for each triplet of chosen orders. The picker, used for picking up the items, moves between in the lanes between the racks that can be one-way, too. Each route starts and finishes at the same place, where is the item of an order given to further manipulation, pick-up and collection by the customer.

The picker is mostly driven manually and therefore effective routes can represent a major saving of human work and costs. If a great number of items there are in the warehouse it can be used some heuristic method derived from the heuristics for the TSP which are proposed in paper [2].

In this article there will be solved a problem of picking up the items from the warehouse inspired by a case study, where the number of items was not big (it was a shoe warehouse) and where one kind of item could be found at any pallet and at any rack in different places of the warehouse.

Mathematical model of picking up the items from the warehouse is based on these assumptions:

- The warehouse is divided into racks, where are placed pallets with items.
- On one pallet there is mostly placed one kind of item.
- One kind of item (good) - item can be stored on several pallets in different placed racks.
- If the warehouse is refilled, the item of producers is placed into any free racks and so that there is the evidence in the warehouse of what each rack and each pallet consist of, therefore the kind and amount of item.
- The picker moves in aisles between the racks. There is a matrix of distances between the racks and the stand of the picker (in metres or in time of journey of the picker).

Because it is about a small number of items and these items are placed on only one section of picker, the paper is focused on the mathematical modelling. Items of different kinds of goods don't differ too much in a weight and volume. Mathematical model will optimise the route of the picker for one order.

In the second part of the thesis we will focus on reorganisation of the warehouse that consists in a transfer of item of one kind from one pallet to another pallet consisting of the same kind of item. This reorganisation of the item is called consolidation of the warehouse. It should contribute to increase in the efficiency of the picker's routes. It

¹ Prague University of Economics and Business, Department of Econometrics, W. Churchill sq. 4, Prague 3, pelikan@vse.cz.

is supposed that it would take place at time, when it would not be necessary to solve the orders of the items. Both problems and mathematical models are explained in an illustrative example at the end of this thesis.

2 Mathematical model

This chapter proposes the mathematical model for optimisation of the picker route for one order. The number of racks (places in a warehouse, pallets) is $n-1$ which are nodes and node 1 is a stand of a picker which is also a drop-off-location of an item of an order. There is given a matrix of distances between the nodes. These nodes represent the locations of pallets in the warehouse and the picker stand. The quantity of goods on the pallet is Q_{js} where j is a node and s is a kind of item. The order requires amount of s -th item q_s . The wage or size volume of the item could be included into the model but in this case it is restricted on number of pieces only.

Parameters of the model:

- n – number of nodes,
- S – number of kinds of item in a warehouse,
- K – maximal number of routes,
- Q_{js} – number of pieces of s -th item placed on j -th pallet,
- q_j – number of pieces of s -th item of the given order,
- d_{ij} – distance between the i -th and the j -th node,
- W – capacity of the picker.

Variables of the model:

- x_{ijk} – binary, equal to 1 if a picker on k -th route goes from node i to node j ,
- y_{iks} – integer variable represents number of pieces of s -th item being transported on k -th route of the picker from node j ,
- u_{jk} – variable represents number of pieces being transported on k -th route of the picker leaving node j .

Mathematical model

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^K d_{ij} x_{ijk} \rightarrow \min \quad (1)$$

$$\sum_{i=1}^n x_{ijk} = \sum_{i=1}^n x_{jik} \quad j = 1, 2, \dots, n, k = 1, 2, \dots, K \quad (2)$$

$$\sum_{j=1}^n x_{1jk} \leq 1, \quad k = 1, 2, \dots, K \quad (3)$$

$$\sum_{i=1}^n x_{iik} = 0, \quad k = 1, 2, \dots, K \quad (4)$$

$$u_{ik} + \sum_{s=1}^S y_{jks} - W(1 - x_{ijk}) \leq u_{jk}, \quad (5)$$

$$i \neq j, \quad i = 1, 2, \dots, n, \quad j = 2, 3, \dots, n, \quad k = 1, 2, \dots, K, \quad s = 1, 2, \dots, S$$

$$0 \leq u_{jk} \leq W, \quad j = 1, 2, \dots, n, \quad k = 1, 2, \dots, K \quad (6)$$

$$\sum_{k=1}^K y_{jks} \leq Q_{js}, \quad j = 1, 2, \dots, n, s = 1, 2, \dots, S \quad (7)$$

$$\sum_{j=1}^n \sum_{k=1}^K y_{jks} = q_s, \quad s = 1, 2, \dots, S \quad (8)$$

$$\sum_{s=1}^S y_{iks} \leq W \sum_{j=1}^n x_{ijk}, \quad i = 1, 2, \dots, n, k = 1, 2, \dots, K \quad (9)$$

$$x_{ijk} \text{ binary}, y_{iks} \geq 0 \text{ integer}, u_{jk} \geq 0, i = 1, 2, \dots, n \quad (10)$$

Objective function (1) equals to a length of all the routes of the picker. If a picker on k -th route enters j -th node, then it has to leave it (2). The picker can leave the stand maximally once on each route (3). Inequations (5) and (6) prevent the partial cycles and simultaneously they do not allow the transported item to exceed the capacity of the picker. The amount of s -th kind transported from the node j must not exceed the supply of this item in this node (7). Equation (8) guarantees that the total volume of the picked item of s -th kind is equal to the amount requested in the order. Condition (9) means that if a node i is not visited on k -th route, then there cannot be picked any item from it.

Example

Let's have 7 nodes. Node 1 is a stand of picker. In Tab.1 there is a matrix of distances between the nodes. There are 3 kinds of items in a warehouse. These kinds of items are placed in pallets and these places including the number of pieces is in Tab. 2. The order is $q = (70, 60, 20)$.

0	2	2	4	4	6	6
2	0	1	2	2	4	4
2	1	0	2	2	4	4
4	2	2	0	1	2	2
4	2	2	1	0	2	2
6	4	4	2	2	0	1
6	4	4	2	2	1	0

Table 1 Distances matrix

Q_{js}	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$
$s=1$	0	50	0	0	45	0	0
$s=2$	0	0	30	0	0	50	0
$s=3$	0	0	0	80	0	0	60

Table 2 Quantity of items in nodes

Optimal solution:

Route 1: 1-3-4-6-5-1

Route 2: 1-2-1

Length of all the routes = 16

In node 2 there is picked item 1 in the amount of 50.

In node 3 there is picked item 2 in the amount of 30.

In node 4 there is picked item 3 in the amount of 20.

In node 5 there is picked item 1 in the amount of 20.

In node 6 there is picked item 2 in the amount of 30.

3 Model of the warehouse consolidation

If one kind of item can be located on different pallets in different parts of the warehouse, then it is necessary to release the racks of the warehouses after some time. It is a warehouse consolidation. The aim is to gather the content of pallets of one kind of item (if it does not exceed the capacity of a pallet) and therefore to release maximally the area of a warehouse.

Parameters of the model are the same as in the model of the route optimisation.

Variables of the model:

x_{ijs} – binary variable that is equal to 1, if the item s from the pallet i is moved onto the pallet j .

y_j – binary variable that is equal to 1, if there is some item on the pallet j (is not empty),

q_{js} – integer variable – number of pieces of item s on the pallet j after moving of the item s .

Mathematical model

$$\sum_{j=1}^n y_j \rightarrow \min \quad (11)$$

$$q_{js} = Q_{js} + \sum_{i=1}^n Q_{is} x_{ijs} - \sum_{i=1}^n Q_{js} x_{jis}, j=1,2,\dots,n, s=1,2,\dots,S \quad (12)$$

$$\sum_{s=1}^S q_{js} \leq W, j=1,2,\dots,n, s=1,2,\dots,S \quad (13)$$

$$x_{iis} = 0, \quad i=1,2,\dots,n, s=1,2,\dots,S, \quad (14)$$

$$\sum_{j=1}^n x_{ijs} \leq 1, i=1,2,\dots,n, s=1,2,\dots,S \quad (15)$$

$$W y_j \geq \sum_{s=1}^S q_{js}, j=1,2,\dots,n, s=1,2,\dots,S \quad (16)$$

$$x_{ijs}, y_j \text{ binary, } q_{js} \geq 0 \text{ integer, } i, j=1,2,\dots,n, s=1,2,\dots,S \quad (17)$$

The objective function (11) minimises the number of occupied pallets after the moving the item. In equation (12) on the left side there is final amount of the item of the s -th kind in node j after moving the item. On the right side there is the initial amount of pieces increased by the transfers into this node and decreased for the transfers of the s -th item into this node. In (13) the amount of the j -th node is limited by the capacity of this node.

Example

Starting stock of goods in the warehouse is in Tab. 2. The result of the optimal solution are these 3 transfers:

50 pieces of item 1 will transfer from node 2 to node 5.

60 pieces of item 1 will transfer from node 5 do node 5.

50 pieces of item 2 will transfer from node 6 to node 3.

The optimal solution is shown in Tab. 3.

Q_{js}	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$
$s=1$	0	0	0	0	155	0	0
$s=2$	0	0	80	0	0	0	0
$s=3$	0	0	0	80	0	0	0

Table 3 Final stock of goods in nodes

4 Conclusion

This paper deals with order picking in whole sale warehouse. The aim is to obtain optimal routes of the picker to collect requested amount of goods of an order. The time of picker routes is minimized. In the second part of the

paper a reorganisation of the pallets in the warehouse that consists in a transfer of item of one kind of goods from one pallet to another pallet. Mathematical model both problems are proposed numerical example is added.

References

- [1] De Koster, R., Le-Duc, T., Roodbergen, K. (2007). Design and control of warehouse order picking: a literature review. *European Journal of Operational Research*, 182(2), 481-501.
- [2] Matusiak, M., De Koster, R., Kroon, L., Saarinen, J. (2014). A fast simulated annealing method for batching precedence-constrained customer orders in warehouse. *European Journal of Operational Research*, 236, 968-977.

Performance of the CLUTEX Cluster Applying the DEA Window Analysis

Natalie Pelloneová¹

Abstract. Cluster organizations are one of the tools to support regional financial and innovation performance. The creation and development of clusters is supported by the EU Structural Funds. The issue of efficient use of public resources is therefore very important. The article analyses the influence of the existence of a cluster organization in the textile industry on finance performance of its member companies. For research purposes, the companies were divided into two groups. The first group consisted of companies included members of the CLUTEX Cluster which represent an organized cluster. The second group consisted of companies that do business in the same region as the organized cluster, but are not its members. As criteria for assessing financial performance the indicator EVA was used. The technical efficiency of companies in all these groups was examined using DEA Window Analysis for the period of 2009–2019. The aim of the research was to verify the statement that enterprises that create the core of the cluster organization achieved higher financial performance than all other non-member enterprises in the same region doing business in the same industry.

Keywords: window analysis, economic value added, DEA, efficiency, cluster organization

JEL Classification: C10, C67, L25, L67

AMS Classification: 90B90, 90C90

1 Introduction

The presented paper deals with the examination of the impact of the membership of business entities in a cluster organization on their financial performance. The past three decades have witnessed a large wave of interest in establishing cluster organizations, and cluster support has become the predominant strategy for supporting economic development in most foreign countries [4], [7]. Simmie and Sennett [18] define a cluster organization as a large number of interconnected industrial and service companies characterized by a high degree of cooperation that operate under the same market conditions. Cluster organizations offer members a number of specific benefits. These benefits are mainly reflected in the growth of efficiency, productivity, financial performance, innovation activities and increasing competitiveness [10].

The impact of the cluster concept on the performance of member entities is not fully objectively quantified in the conditions of the Czech Republic. In the Czech Republic, the establishment of cluster organizations has been supported since 2004, when clusters began to be supported under the Operational Programme Industry and Entrepreneurship - through the Clusters support programme, under which it was possible to apply for support until 2006. In 2007, this programme was followed by the Cooperation-Clusters support programme within the Business and Innovation Operational Programme, within which it was possible to apply for support until 2013. Since 2014, clusters have been supported by the Business and Innovation Operational Programme for Competitiveness, which lasted until 2020 [11]. Several studies have addressed the question of whether companies in cluster organizations are more financially successful than so-called non-clustered companies, for example Pe'er and Vertinsky [15], Kukalis [8], Ruland [15] and Swann, Prevezer and Stout [19]. The results of these researches show that especially in smaller companies, the profitability of members of the cluster organization is higher than in companies that have decided not to join the cluster organization. Research by Audretsch and Feldman [1] shows that cluster organizations have a positive effect on the performance of all stakeholders, regardless of the size of the business. Following the research by Lei and Huang [9], companies within a cluster organization perform better financially than companies outside the cluster organization.

¹ Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, natalie.pelloneova@tul.cz.

The aim of this paper is to assess and compare the financial performance of companies that are members of the selected cluster organization and companies that operate in the same industry and region of operation of the cluster organization, but are not its member entities. At the same time, this research aims to verify the assumption that the membership of companies in a cluster organization has a positive impact on their financial performance. Thus, the research examines whether companies involved in a cluster organization have different financial performance than companies that operate in the same industry and region, but are not members of any cluster organization. A cluster organization operating in the textile industry was selected for this research.

2 Data Envelopment Analysis and Window DEA

The methodology used in this paper is based on the analysis of data packages. DEA is a nonparametric approach based on linear programming, which has many key advantages. The analysis of data packages is used in practice to evaluate the effectiveness of various units (for example banks, hospitals, universities, transport, and research organizations). In this paper, the evaluated unit is a business entity included in the two research groups examined. Within DEA, efficiency is calculated as the ratio of the weighted sum of outputs and the weighted sum of inputs. This methodology makes it possible to process different types of inputs and outputs together [14]. DEA allows you to work with a wide range of different models. The best known are the Banker, Charnes and Cooper (further BCC) model and the Charnes, Cooper and Rhodes (further CCR) model.

The classical DEA method is suitable for examining efficiency within one time period. DEA with time windows (so-called Window-DEA, WDEA) is used as standard for dynamic problems. This method consists in calculating the DEA efficiency in a given interval of time periods (so-called window), which gradually shifts in time, thus obtaining the trajectory of DEA efficiency units [6]. Thus, it is possible to examine the efficiency not only between individual units, but also for a selected unit in different periods [5].

The window analysis assesses the performance of a DMU over time by treating it as a different entity in each time-period. This method allows for tracking the performance of a unit or DMU over time and provides a better degree of freedom. If a DMU is found to be efficient in one year despite the window in which it is placed, it is likely to be considered strongly efficient compared to its peers [13]. Consider N DMUs ($n = 1, 2, \dots, N$) observed in T ($t = 1, 2, \dots, T$) periods using r inputs to produce s outputs. Let DMU_n^t represent an DMU_n in period t with a r dimensional input vector $x_n^t = (x_n^{1t}, x_n^{2t}, \dots, x_n^{rt})'$ and s dimensional output vector $y = (y_n^{1t}, y_n^{2t}, \dots, y_n^{st})'$. If a window starts at time k ($1 \leq k \leq T$) with window with w ($1 \leq w \leq t - k$), then the metric of input is according to [17] given as (1):

$$x_{kw} = \begin{pmatrix} x_1^k, x_2^k, \dots, x_N^k, x_1^{k+1}, x_2^{k+1}, \dots, x_N^{k+1}, \\ x_1^{k+w}, x_2^{k+w}, \dots, x_N^{k+w} \end{pmatrix}' \quad (1)$$

The metric of outputs is given as (2):

$$y_{kw} = \begin{pmatrix} y_1^k, y_2^k, \dots, y_N^k, y_1^{k+1}, y_2^{k+1}, \dots, y_N^{k+1}, \\ y_1^{k+w}, y_2^{k+w}, \dots, y_N^{k+w} \end{pmatrix}' \quad (2)$$

The CCR model of DEA window problem for DMU_n^t is given by solving the following linear program (3):

$$\begin{aligned} & \min \theta, \\ & \theta' X_t - \lambda' X_{kw} \geq 0 \\ & \text{s. t. } \lambda' Y_{kw} - Y_t \geq 0, \\ & \lambda_n \geq 0 \quad (n = 1, 2, \dots, N \times w). \end{aligned} \quad (3)$$

BCC model formulation can be obtained by add the restriction (4).

$$\sum_{n=1}^n \lambda_n = 1 \quad (4)$$

3 Data and methodology

The presented research was performed on data from 2009–2019. The source of accounting data was the Magnus-Web database [2]. The research was performed on two research sets described below. The research procedure can be divided into the following nine steps.

1. Selection of a suitable cluster organization – a CLUTEX technical textile cluster was selected for this case study. The CLUTEX cluster was established in Liberec in 2006. The CLUTEX cluster is based in Liberec and has the legal form of an association [3]. CLUTEX brings together legal entities doing business in the field of textile and clothing production. The aim of the cluster organization is to create optimal conditions for business and subsequently support its development in the field of research, development and production of technical textiles.

2. Creation of a database of evaluated entities – the performed research was focused on the evaluation of financial performance, therefore all non-business entities were excluded from the analysis. The total number of business entities forming the core of the cluster organization was 26 companies in the first research group. The second research group consisted of companies with the same subjects of activity (i.e. operating in NACE 13200, 13900 and 14100), which are not members of the CLUTEX cluster organization and operate in the same region (i.e. the North-East cohesion region). The total number of these companies in the second research group was about 121 companies in one year.

3. Collection of financial statements – for the above-mentioned companies, it was essential to obtain the necessary data from the financial statements, especially from the balance sheet and profit and loss statement for the years 2009–2019. Unfortunately, not all companies in individual research files complied with the obligation to publish selected data from the balance sheet and profit and loss statement in the collection of documents. The first research group managed to obtain financial statements for 14 companies, the second research group managed to obtain data from 68 companies.

4. Determination of the number of employees – data on the number of employees of companies for the years 2009 to 2019 were also obtained from the MagnusWeb database [2]. If an interval was specified, the middle of the interval was used for further calculation. If the value for the year was not given, the last available figure was used. In case the company stated zero number of employees, one employee was counted (the owner as a person working on his own account).

5. Calculation of economic value added – 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 [12], see relation (5).

Where *ROE* is return on equity, r_e is the alternative cost of equity and *E* is 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 (6). 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.

$$EVA = (ROE - r_e)E \quad (5)$$

$$r_e = r_f + r_{company} + r_{finstr} + r_{finstab} + r_{la} \quad (6)$$

The indicator can be meaningfully determined only for companies with a positive equity value. Therefore, companies with a negative equity value were excluded from the comparison. The first research set of companies to be compared was reduced to 12 companies. The second research set of companies was reduced to 42 companies.

6. Creation of a set of companies that have a complete time series – the research was focused on evaluating financial performance, so the first and the second research set included companies for which it was possible to calculate the EVA indicator for the entire period. The first research set of companies to compare was reduced to 11 companies that had a complete time series for EVA. The second set of research companies was reduced to 14 companies that had a complete time series for EVA.

7. Definition of inputs and outputs for Data Envelopment Analysis – in the next step it was necessary to define inputs and outputs for the needs of data package analysis. Number of employees and long-term invested resources were used as inputs; output was Economic Value Added (furthermore EVA). Long-term invested resources are given by the sum of equity, long-term bonds and long-term bank loans. The above inputs were chosen taking into account previous research and also the low number of DMUs. The values of detailed statistics for both research files for 2019 are shown in Tables 1 and 2.

	Number of employees	Long-term inv. capital	EVA
Mean	165.8	225 250	-21 211.2
Median	82	55 955	-869.3
St. deviation	228.7	378 732	43 374.9
Max	957	1 637 725	50 887.9
Min	10	3 122	-120 476.3

Table 1 Descriptive statistic of inputs and outputs of CLUTEX Cluster in 2019

	Number of employees	Long-term inv. capital	EVA
Mean	179.4	315 002	-12 093.3
Median	91	69 778	-182.8
St. deviation	332.5	547 623	88 212.8
Max	10 124	2 129 734	273 357.3
Min	8	2 256	-490 303.5

Table 2 Descriptive statistic of inputs and outputs of non-member companies in 2019

8. Determination of the DEA window analysis score – efficiency scores were calculated for all companies using the DEA window analysis method. Radial models, input-oriented with assumptions of constant (CCR) and variable returns to scale (BCC), were used. The MaxDEA 7 Ultra software was used to calculate the values of efficiency scores.

9. Comparison of the financial performance of textile companies in both samples – a non-parametric Kolmogorov-Smirnov test was used to compare the financial performance between the two research groups. The non-parametric test was chosen because when using the Shapiro-Wilk test, it was shown that the values of the individual variables do not have a normal distribution. Statistical testing was performed at a significance level of 5% using STATGRAPHICS Centurion XVIII.

4 Research results

The DEA window analysis radial based model was applied to the obtained data. Efficiency of companies was estimated using an input-oriented DEA window analysis model with constant returns to scale (further CRS) and input-oriented DEA window analysis model with variable returns to scale (further VRS). The reason for using both techniques is the fact that the assumption of CRS is accepted only in a situation where all companies operate at the optimum size. This assumption, however, is in practice impossible to fill, so in order to solve this problem we calculate also with VRS. The length of the window was chosen to be three years. The average efficiency scores in the three year rolling periods were determined for each company. The individual efficiency scores were then aggregated according to the affiliation of the member companies to the CLUTEX Cluster or non-member OTHER companies. The result is the average efficiency scores of individual groups of companies, according to both CRS and VRS models.

The results of the DEA efficiency scores under CRS during the period of 2009–2019 for CLUTEX cluster member companies and also for non-member business entities are listed in Table 3. Moving average efficiency is shown in a three-year window. In the period of 2009–2019, the average efficiency of CLUTEX cluster member firms calculated using CCR model ranges from 64% to 83%. The results show that the average inefficiency of CLUTEX cluster member companies in the CCR model ranged from 17% to 36%. In the period of 2009–2019, the average efficiency of non-member companies calculated using CCR model ranges from 62% to 76%. The results show that the average inefficiency of non-member companies in the CCR model ranged from 24% to 38%.

This analysis shows that throughout the period under review, the members of the CLUTEX cluster organization achieved higher efficiency compared to non-member companies in the region in which the cluster organization operates. However, these differences were not significant based on the performed Kolmogorov-Smirnov test.

DMUs	2009– 2011	2010– 2012	2011– 2013	2012– 2014	2013– 2015	2014– 2016	2015– 2017	2016– 2018	2017– 2019	Mean
CLUTEX	0.75	0.73	0.68	0.69	0.64	0.77	0.83	0.79	0.82	0.74
OTHERS	0.70	0.71	0.65	0.63	0.62	0.70	0.74	0.73	0.76	0.69

Table 3 Efficiency of member and non-member companies of the CLUTEX cluster in the CCR model

Table 4 presents the efficiency of CLUTEX cluster member companies and non-member businesses estimated under the VRS. In the period of 2009–2019, the average efficiency of CLUTEX cluster member companies calculated using BCC ranges from 77% to 92%. The results show that the average inefficiency of CLUTEX cluster member companies in the BCC model ranged from 8% to 23%. In the period of 2009–2019, the average efficiency of non-member companies calculated using BCC ranges from 73% to 85%. The results show that the average inefficiency of non-member companies in the BCC model ranged from 15% to 27%.

Looking at the average group efficiency scores in the time windows (see Table 4), it is clear that throughout the period under review, the members of the CLUTEX cluster organization achieved higher efficiency compared to non-member companies in the region in which the cluster organization operates. However, these differences were not significant based on the performed Kolmogorov-Smirnov test. The development of the efficiency showed that the average efficiency was decreasing during the period of 2009–2013. This decrease might have resulted from the financial crisis. After 2012 the average efficiency increased.

DMUs	2009– 2011	2010– 2012	2011– 2013	2012– 2014	2013– 2015	2014– 2016	2015– 2017	2016– 2018	2017– 2019	Mean
CLUTEX	0.85	0.84	0.78	0.79	0.77	0.86	0.89	0.90	0.92	0.84
OTHERS	0.79	0.82	0.75	0.73	0.73	0.84	0.83	0.83	0.85	0.80

Table 4 Efficiency of member and non-member companies of the CLUTEX cluster in BCC model

5 Conclusion

The aim of the research was to evaluate and subsequently compare the financial performance of companies in the textile industry that are members of the CLUTEX cluster organization and non-member companies operating in the same industry and region. For this purpose, 11 member companies of the CLUTEX cluster organization and 14 non-member textile companies from the region for the reference period of 2009–2019 were evaluated. For these years and all companies, data on the economic result for the accounting period were obtained and the values of the EVA indicator were calculated according to the MIT methodology. WDEA analysis was used to make comparisons among companies.

As part of the application of the WDEA method, the average efficiency for the period of 2009–2019 was calculated and it could be stated that the companies in the CLUTEX cluster organization achieve an average score of 0.74 (according to the BCC model 0.84), while the average score of non-member companies is 0.69 (according to the BCC model 0.80). Based on the WDEA application, it can be stated that in comparison with non-member companies, a significant increase in efficiency has been demonstrated for companies in the CLUTEX cluster organization. The companies in the CLUTEX cluster organization also responded better to the effects of the recession in 2012 and 2013. From this it can be concluded that the CLUTEX cluster organization has been associating more successful companies in this industry from the very beginning. Based on the research and the use of WDEA analysis, it can be stated that the existence of a cluster organization has an impact on increasing the efficiency of member companies. However, based on the performed Kolmogorov-Smirnov test, it is not possible to state that these differences were statistically significant.

It is also necessary to draw attention to the limits of research. They are mainly affected by the availability of financial statements, which is worse from year to year. A large number of companies for which it was not possible to obtain complete financial statements or which showed a negative value of equity in a given year for the observed period had to be excluded from the comparison of both research groups. Therefore, only relatively small samples of companies could be examined.

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] Audretsch, D. B. & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review*, 86(3), 630–640.
- [2] Bisnode. (2021). Magnusweb: Komplexní informace o firmách v ČR a SR [online]. Bisnode ČR, Prague, 2021 [cit. 2021-03-08]. Available at: <https://magnusweb.bisnode.cz>.
- [3] Clutex. (2021). O Clutex [online]. Clutex, Liberec, 2021 [cit. 2021-03-25]. Available at: <http://www.clutex.cz/>.
- [4] Fang, L. (2015). Do Clusters Encourage Innovation? A Meta-analysis. *Journal of Planning Literature*, 30(3), 239–260.
- [5] Charnes A., Clark C. T. & Cooper W. W. (1985) A development study of data envelopment analysis in measuring the efficiency of maintenance units in the US air forces. *Annals of Operations Research*, 2, 95–112.
- [6] Klopp, G. (1985). *The Analysis of the Efficiency of Production System with Multiple Inputs and Outputs*. Chicago: University of Illinois at Chicago.
- [7] Kuchiki, A. & Tsuji. M. (2011). *Industrial clusters, upgrading and innovation in East Asia*. Northampton: Edward Elgar.
- [8] Kukalis, S. (2010). Agglomeration Economies and Firm Performance: The Case of Industry Clusters. *Journal of Management*, 36(2), 453–481.
- [9] Lei, H. & Huang, CH. (2014). Geographic clustering, network relationships and competitive advantage: Two industrial clusters in Taiwan. *Management Decision*, 52(5), 852–871.
- [10] Marešová, P. (2012). Evaluating the performance of clusters with focus on Czech Stone Cluster. In D. Špalková & L. Furová (Eds.), *Modern and Current Trends in the Public Sector Research. Proceedings of the 16th International Conference* (pp. 52–60). Brno: Masarykova univerzita.
- [11] MPO. (2021). Dotace a podpora podnikání [online]. Ministry of Industry and Trade, Prague, 2021 [cit. 2021-03-03]. Available at: <https://www.mpo.cz/cz/podnikani/dotace-a-podpora-podnikani/>.
- [12] MPO. (2019). Finanční analýza podnikové sféry za rok 2018. Ministry of Industry and Trade, Prague, 2019.
- [13] Muhammad, A., Tirupathi R. & Farooq, Q. (2018). DEA Window Analysis with slack-based measure of Efficiency in Indian Cement Industry. *Statistics, Optimization & Information Computing*, 6(2), 292–302.
- [14] Ohe, Y. & Peypoch, N. (2016). Efficiency analysis of Japanese Ryokans: A window DEA approach. *Tourism Economics*, 22(6) 1261–1273.
- [15] Peer, A. & Vertinsky, I. (2006). *The Determinants of Survival of De Novo Entrants in Clusters and Dispersal*. Hanover: Tuck School of Business.
- [16] Ruland, W. (2013). Does Cluster Membership Enhance Financial Performance? *iBusiness*, 5(1), 1–11.
- [17] Řepková, I. (2014). Efficiency of the Slovak Commercial Banks Applying the DEA Window Analysis. *International Journal of Economics and Management Engineering*, 8(5), 1342–1347.
- [18] Simmie, J. & Sennet, J. (1999). *Innovation in the London metropolitan region*. Innovative clusters and competitive cities in the UK and Europe. Oxford: Brookes School of Planning.
- [19] Swann, G., Prevezer, M. & Stout, D. K. (1998). *The dynamics of industrial clustering: International comparisons in computing and biotechnology*. Oxford: Oxford University Press.
- [20] Štichhauerová, E. & Žižka, M. (2020). Impact of Cluster Organizations on Financial Performance in Selected Industries: Malmquist Index Approach. In S. Kapounek & H. Vránová (Eds.), *38th International Conference on Mathematical Methods in Economics 2020. Conference Proceedings* (pp. 572–578). Brno: Mendel University in Brno.

New models for a return bus scheduling problem

Štefan Peško¹, Stanislav Palúch², Tomáš Majer³

Abstract. In this paper we study MILP models for real-world situations which can occur in multi-depots bus scheduling problems. Unusual here is an original constraint that ensures the return of buses to home depot via double-index models. Analogical problem has been solved for the multiple vehicle routing problems by three-index models, where the third index represents the depot.

First we formulate a three-index model with multi-depots with given number of buses in depots. Next we formulate a two-index model with one central depot and with some single depots where the drivers' place of residence is. Next, a two-index model is formulated with a fixed number of buses in depots. That model distributes minimum number of buses between depot.

Presented models are verified by computational experiments with an instance of public bus service for city Martin in Slovakia.

Keywords: multi-depot models, bus scheduling, return running boards, public bus service

JEL Classification: 90C10, 90C15

AMS Classification: C02, C61, C65, C685

1 Introduction

The topic discussed in this paper is focused on one specific problem among operational problems confronting the management of public bus service when, in addition to the central bus depot single depots in the drivers' place of residence also make sense.

The basic bus scheduling problem consists of assigning buses to given set of trips in running board such that:

- each trip is performed exactly once,
- each bus must start and end its work day at the same depot,
- the number of buses is as minimal as possible,
- the operate cost is minimal.

This problem has several variations with practical restrictions (on number and types of depots, meal breaks, buses types, length of the running boards, flexible trips) studied in Slovakia by Palúch and his colleagues in [4], [5], [6] and [7]. From the world sources, let's mention at least the papers in [1] and [9].

We will consider the Return Bus Scheduling Problem (RBSP) when the set of trips is given for one type of bus and no practical restrictions are presumed.

- the set of trips is given for one type of bus,
- one central depot with given number of buses,
- several single depots with one bus,
- no practical restrictions are presumed.

2 Notions and definitions

In this paper we will use the terminology introduced by Wren [8]. Given set of regular trips $\mathcal{S} = \{S_1, S_2, \dots, S_n\}$, each trip $S_i, i \in N = \{1, 2, \dots, n\}$ is represented by an arbitrary ordered quadruple

$$S_i = (p_i^d, t_i^d, p_i^a, t_i^a), \quad (1)$$

where

- p_i^d – the departure place,
- t_i^d – the departure time,
- p_i^a – the arrival place,

¹ Univerzita 8215/1, 010 026 Žilina, Slovakia, stefan.pesko@fri.uniza.sk

² Univerzita 8215/1, 010 026 Žilina, Slovakia, stanislav.paluch@fri.uniza.sk

³ Univerzita 8215/1, 010 026 Žilina, Slovakia, tomas.majer@fri.uniza.sk

t_i^a – the arrival time.

The bus models are of the discrete-time rather than that of continuous-time variety. We will suppose that for one day of time schedule $t_i^d, t_i^a \in \{1, 2, \dots, 1440\}$ is in minutes. Note that also all the time dates are in minutes.

Let $\tau(p_i^a, p_j^d)$ be the times of an idle trip from place p_i^a to place p_j^d . Similarly we denote the times of the idle trips from and to home depots d as $\tau(d, p_j^d)$ and $\tau(p_i^a, d)$.

The running board of a bus $\mathcal{T}_i(d)$ from home depot d with m trips from set \mathcal{S} is a sequence

$$S_{i_1} < S_{i_2} < \dots < S_{i_m}, \quad (2)$$

where for all $k \in \{1, 2, \dots, m-1\}$ is

$$t_{i_k} + \tau(p_{i_k}^a, p_{i_{k+1}}^d) < t_{i_{k+1}}. \quad (3)$$

By the operate cost $c(\mathcal{T}_i(d))$ of the running board $\mathcal{T}_i(d)$ we mean the sum of the time of idle trip from the home depot d to the first assigned trip, times of idle trips in running board and the time of return to the home depot from last assigned trip, i.e.

$$c(\mathcal{T}_i(d)) = \tau(d, p_{i_1}^d) + \sum_{k=1}^{m-1} \tau(p_{i_k}^a, p_{i_{k+1}}^d) + \tau(p_{i_{m-1}}^a, d). \quad (4)$$

In conclusion, we will assume that a set of places of home depots $D = \{d_0, d_1, \dots, d_h\}$ is given, where d_0 is a central depot which has q buses and following h single depots have one bus each. It is obvious that $q + h \geq q_{min}$ where q_{min} denotes minimum number of buses for covering given set of trips.

3 Illustrative example

We have an instance of a set $\mathcal{S} = \{S_1, S_2, \dots, S_9\}$ of nine trips given in following table 1. Consider a simple

i	p_i^d	t_i^d	p_i^a	t_i^a
1	1	60	2	70
2	1	60	3	75
3	3	85	2	100
4	1	85	4	100
5	4	85	2	110
6	2	120	1	130
7	3	128	1	150
8	3	128	4	145
9	2	140	1	160

Table 1 Small illustrative instance of trips.

transport network with 4 stops, with a place 1 where a central depot and two busses are located and a place 4 where a single depot is located.

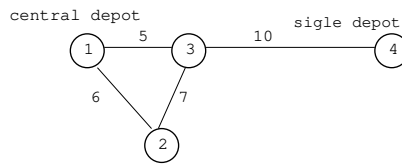


Figure 1 Simple transportation network.

The correspondig minimum lengths of the idle trips (in minutes) between stops are given as a matrix τ

$$\tau = \begin{pmatrix} 0 & 6 & 5 & 15 \\ 6 & 0 & 7 & 17 \\ 5 & 7 & 0 & 10 \\ 15 & 17 & 10 & 0 \end{pmatrix}.$$

The instance of the problem presented in this way allows two solutions that can be seen in table 2. The value in the columns 'bus I' and 'bus II' is the bus number. The home depot for busses 1 and 2 is central depot 1 and for bus 3 it is single depot 4. We can see that in the solution I. the bus 1 starts at depot 1 but ends at depot 4 and vehicle 3

i	p_i^d	t_i^d	p_i^a	t_i^a	bus I.	bus II.
1	1	60	2	70	1	1
2	1	60	3	75	2	2
3	3	85	2	100	1	1
4	1	85	4	100	2	2
5	4	85	2	110	3	3
6	2	120	1	130	3	2
7	3	128	1	150	2	1
8	3	128	4	145	1	3
9	2	140	1	160	3	2

Table 2 Running boards for I. and II. solutions.

opposite. In the solution II. busses start and end at their home depots.

4 Three-index model

This mathematical model is simplified and modified version of the multi depot multiple vehicle type model presented in [1] which is based on a multi-layer network.

First we denote $K = \{0, 1, \dots, h\}$ as a set of indexes of depots. Now we can define vehicle scheduling network $\vec{G}_k = (V, A_k)$ corresponding to depot $d_k, (k \in K)$, which is a directed graph described by set of vertices $V = \{1, 2, \dots, n\}$ and arc $A_k = \{(i, j) \in V \times V : S_i < S_j\} \cup \{(-k, i), (i, -k) : i \in N\}$.

Let c_{ij} be the vehicle cost, time of idle trips, of arc $(i, j) \in A_k$, which represents idle time activities for vehicles from home depot d_k .

Decision three-index variable x_{ij}^k indicates whether an arc (i, j) is used an assigned to the depot d_k or not. Now we can define following problem of a bivalent linear programming (3RBSP):

$$\sum_{k \in K} \sum_{(i,j) \in A_k} c_{ij} x_{ij}^k \rightarrow \min \quad (5)$$

s.t.

$$\sum_{j:(i,j) \in A_k} x_{ij}^k - \sum_{j:(j,i) \in A_k} x_{ji}^k = 0 \quad \forall k \in K, \forall i \in V, \quad (6)$$

$$\sum_{(-k,j) \in A_k} x_{-kj}^k = 1, \quad \forall k \in K - \{0\}, \quad (7)$$

$$\sum_{(0,j) \in A_0} x_{0j}^0 = q, \quad (8)$$

$$\sum_{k \in K} \sum_{j:(i,j) \in A_k} x_{ij}^k = 1, \quad \forall i \in V : i > 0, \quad (9)$$

$$x_{ij}^k \in \{0, 1\} \quad \forall k \in K, \forall (i, j) \in A_k. \quad (10)$$

The objective (5) is to minimize the sum of total vehicles cost. Constraints (6) are flow conservation constraints indicating that the flow into each vertex equals the flow out this vertex. Constraints (7) assure that some trips of bus schedule will be covered by running board from a single depot. Constraints (8) assure that all vehicles from central depot is used. Constraints (9) force coverage all trips from some bus. Constraints (10) are obligatory.

We will further show how depot's index of decision variable can be saved.

5 Reduction of decision variables

In the 3RBSP model we have a decision three-index variable x_{ij}^k indicates whether an arc (i, j) will be selected and will be assigned to the object k , which in our case is one of the depots. If $i \in I, j \in J$ and $k \in K$ then we need in the extreme case $|I| \cdot |J| \cdot |K|$ of 01 variables.

Using other variables $y_i \in R, i \in I \cup J$ and decision variables $z_{ij}, i \in I, j \in J$ we can replace the purpose of the variable x_{ij}^k via following inequalities:

$$M(z_{ij} - 1) \leq y_i - y_j \leq M(1 - z_{ij}), \quad (11)$$

where M is a positive number that is large enough. If $z_{ij} = 1$ then $0 \leq y_i - y_j \leq 0$ and so $y_i = y_j$. If $z_{ij} = 0$ then $M \leq y_i - y_j \leq M$ and there are no restrictions on the variables y_i and y_j .

Note that such a reduction option is relatively universally applicable, but we have not encountered it yet.

6 Two-index model

We can define basic vehicle scheduling network as a directed graph $\vec{G} = (V, A)$ defined by a set of vertices $V = \{-h, \dots, 0, 1, 2, \dots, n\}$ and set of arc $A = \{(i, j) : S_i < S_j\} \cup \bigcup_{i \in N, k \in K} \{(-k, i), (i, -k)\}$. Vehicle cost of arc $(i, j) \in A$ is denoted as c_{ij} again.

We will use decision two-index variable z_{ij} for indication whether an arc (i, j) is used or not. For the assignment trip to home depot we use variable $y_i \in K, i \in V$. If $y_i = -k$ then trip $S_i \in S$ is assigned to depot d_k and we can apply inequalities (11)

Now we can define following problem of mixed linear programming (2RBSP):

$$\sum_{(i,j) \in A} c_{ij} z_{ij} \rightarrow \min \quad (12)$$

s.t.

$$\sum_{j:(i,j) \in A} z_{ij} - \sum_{j:(j,i) \in A} z_{ji} = 0 \quad \forall i \in V, \quad (13)$$

$$\sum_{(-k,j) \in A} z_{-kj} = 1, \quad \forall k \in K - \{0\}, \quad (14)$$

$$\sum_{(0,j) \in A} z_{0j} = q, \quad (15)$$

$$\sum_{j:(i,j) \in A} z_{ij} = 1, \quad \forall i \in V : i > 0, \quad (16)$$

$$h(x_{ij} - 1) \leq y_i - y_j \leq h(1 - z_{ij}) \quad \forall (i, j) \in A, \quad (17)$$

$$y_{-k} = k \quad \forall k \in K, \quad (18)$$

$$z_{ij} \in \{0, 1\} \quad \forall (i, j) \in A, \quad (19)$$

$$y_k \geq 0, \quad \forall k \in V. \quad (20)$$

The objective (12) is again to minimize the sum of total vehicles cost. Constraints (13),(14),(15) and (16) are substantially in line with the constraints (6),(7),(8) and (9). If in the constraints (17) is $z_{ij} = 1$ than trips S_i and S_j are assigned to the same depot by $y_i = y_j$. The fictive trips are assigned corresponding depots by constraints (18). Constraints (19) and (20) are obligatory.

7 Computational result

Mathematical model 3RBSP and 2RBSP was solved for instances of real study for bus company Martin in Slovakia containing maximum 726 trips, for bus network see figure 2. Our experiments were conducted on HP XW6600

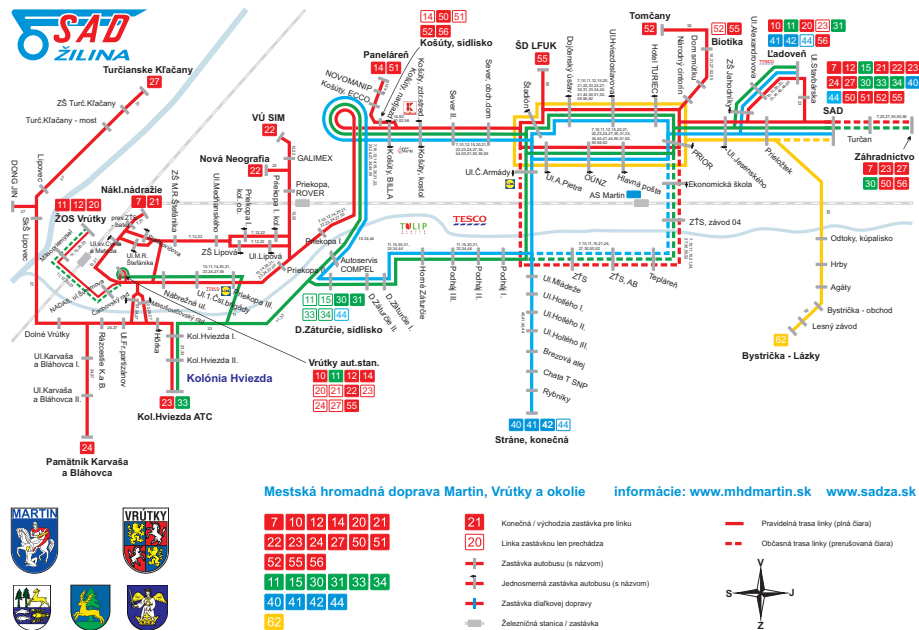


Figure 2 Bus network of city Martin.

Workstation (8-core Xeon 3GHz, RAM 16GB) with OS Linux (Debian/jessie). We also used the Python interface to commercial mathematical programming solver Gurobi. Corresponding computational results are contained in the following tables:

Trips	Depots	Operate cost [min.]	Time [sec.]
666	24; 1	1 303	914.0
666	24; 1, 17	1 269	877.33
666	24; 1, 17, 15	1 232	865.13
666	24; 1, 17, 15, 4	1 228	843.34
180	24; 1	1 209	1.80
180	24; 1, 17	1 084	1.78
180	24; 1, 17, 15	1 054	2.92
180	24; 1, 17, 15, 4	1 050	4.74

Table 3 Computational characteristics for the 3RBSP.

First experiments with tree-index model showed that the reduction of decision 01 variables significantly speeds up calculations. Therefore, we did not even continue to increase the number of connections for the 3RBSP. We continued with the only model 2RBSP, where we required a minimal possible number of buses in the central depot 24. This was calculated using a graph algorithm for maximum flow in the vehicle scheduling network [3].

8 Conclusions

Although it is not surprising that a suitable reduction of conversational variables can have such a significant effect on the computation time, the results surpassed our expectations. Computer experiments have shown that the presented approach can give hope for incorporation a real operational limiting requirements into 2RBSP model [9]

Trips	Depots	Operate cost [min.]	Time [sec.]
726	24; 1	1 356	65.15
726	24; 1, 17	1 319	84.97
726	24; 1, 17, 15	1 299	76.33
726	24; 1, 17, 15, 4	1 295	85.52
720	24; 1	1 925	60.29
720	24; 1, 17	1 304	65.86
720	24; 1, 17, 15	1 299	76.94
720	24; 1, 17, 15, 4	1 295	74.57
666	24; 1	1 303	64.29
666	24; 1, 17	1 269	60.81
666	24; 1, 17, 15	1 232	74.44
666	24; 1, 17, 15, 4	1 228	60.61
180	24; 1	1 209	2.86
180	24; 1, 17	1 084	2.14
180	24; 1, 17, 15	1 054	1.93
180	24; 1, 17, 15, 4	1 050	2.36

Table 4 Computational characteristics for the 2RBSP.

Acknowledgements

The research was supported by the 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] Gintner, V., Kliewer, N., Suhl, L. (2005) Solving large multiple-depot multiple-vehicle-type bus scheduling problem in practice, *OR Spectrum*, 27, Springer-Verlag, pp.507-523, DOI: 10.1007/s00291-005-0207-9
- [2] Gurobi Optimization, Inc. Gurobi Optimizer Reference Manual, 2020, Available at: <https://www.gurobi.com/documentation/9.1/refman/index.html>
- [3] Hagberg, A., Dan Schult, D., Swart, P. (2008) Exploring network structure, dynamics, and function using NetworkX, *Proceedings of the 7th Python in Science Conference (SciPy2008)*, Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp.11-15.
- [4] Palúch, S. (2013) Vehicle and Crew Scheduling problem in Regular Personal Bus Transport – The State of Art in Slovakia. *Pomorstvo, promet in logistika*, 16th international conference on transport science - ICTS 2013, pp. 297-304. Portorož, Slovenija.
- [5] Palúch, S., Peško, S., Majer, T. (2016) An exact solution of the minimum fleet size problem with flexible bus trips, *Quantitative methods in economics, multiple criteria decision making XVIII proceedings of the international scientific conference* 25th-27th May 2016, Vrátna, Slovakia pp. 278-282 [S.l.] Letra Interactive 2016.
- [6] Peško, S., Palúch, S., Majer, T. (2016) A group matching model for a vehicle scheduling problem, *Quantitative methods in economics, multiple criteria decision making XVIII proceedings of the international scientific conference* 25th-27th May 2016, Vrátna, Slovakia pp. 298-302 [S.l.] Letra Interactive 2016.
- [7] Majer, T., Palúch, S., Peško, S. (2016) Algorithms for vehicle and crew scheduling in regular bus transport *18th International Carpathian Control Conference*, ICCO 2017, pp. 300-305 Sinaia, Romania, May 28 - 31, IEEE.
- [8] Wren, A. and Rousseau, J.M. (1995) Bus driver scheduling – an overview, Daduna, J, Branco, I., Pinto Paixao, J. (editors), *Computer-Aided Transit Scheduling*, Springer Verlag, 1995. pp. 173-187.
- [9] Xiaomei Xu, Zhirui Ye, Jin Li, Chao Wang (2018) Solving a Large-Scale Multi-Depot Vehicle Scheduling Problem in Urban Bus Systems, *Mathematical Problems in Engineering*, vol. 2018, Article ID 4868906, 13 pages.

Different ways of extending order scales dedicated to credit risk assessment

Krzysztof Piasecki¹, Aleksandra Wójcicka-Wójtowicz²

Abstract. In our previous work, we have proposed an Extended Order Scale (EOS) dedicated to credit risk assessment. This EOS is linked to Numerical Order Scale (NOS) evaluated by trapezoidal oriented fuzzy numbers. The EOS proposed by us uses two-stage orientation phrases. Such a solution can be too detailed for many bank experts. Therefore, the main aim of our paper is to show some ways of a given EOS simplification to a one-stage or zero stage order scale. Moreover, we present a way of calculating the scoring function which enables to avoid the obstacles related to the lack of associativeness property of addition of trapezoidal oriented fuzzy numbers.

Keywords: order scale, borrowers' scoring, oriented fuzzy numbers

JEL Classification: G21, G24

AMS Classification: 03E72

1 Introduction

Over the years, banks struggled with many obstacles which resulted mainly from insufficient credit assessment of potential borrowers. The main objective of credit risk assessment and management is to keep the credit risk level at a reasonable level along with an increased (or non-decreased) volume of operations. Credit risk experts (also called analysts) consider a wide spectrum of factors which frequently belong to a group of so-called 'soft factors' (qualitative rather than quantitative). Consequently, those factors are expressed by the linguistic scale which in its nature is inaccurate and therefore standard numerical methods do not fully imply in such environment. However, the inaccuracy does not hamper the assessment. Quite the contrary, it allows the experts and decision-makers to incorporate their professional experience and preferences.

Usually the procedure of determining, assessing and constructing the evaluation template is one of the most significant parts in the process of company's assessment and further potential granting of the financial means [2]. The given template must address specific issues and possible options consequently resulting in the acceptable space of any operations. The template also must be unified as each expert, having a specific cultural, professional and educational background, can perceive each factor differently basing on their experience and knowledge and make the decision in different economic and life circumstances. Furthermore, to evaluate individual outcomes multi-criteria methods can be of use, allowing to determine an appropriate scoring function [9].

Due to the imprecision, the oriented fuzzy numbers are a useful tool.

In our previous work [8], we have proposed evaluation scales of linguistic, imprecise phrases used in a process of credit risk evaluation which were basing on preferences of experts - Extended Order Scale (EOS) dedicated to credit risk assessment. This EOS is linked to Numerical Order Scale (NOS) evaluated by trapezoidal oriented fuzzy numbers. The EOS proposed by us uses two-stage orientation phrases. Such a solution can be too detailed for many bank experts. Therefore, the main aim of our paper is to show some ways of a given EOS simplification to a one-stage or zero stage order scale.

Moreover, we emphasise the obstacles related to the lack of associativeness property of addition of trapezoidal oriented fuzzy numbers. We present a way which enables to determine the scoring function and avoid the above-mentioned obstacles.

2 OFN – brief overview

The symbol $\mathcal{F}(\mathbb{R})$ denotes the family of all fuzzy subsets in the real line \mathbb{R} . Fuzzy number (FN) is usually defined as a fuzzy subset of the real line \mathbb{R} . The most general definition of FN was formulated by Dubois and

¹ WSB University in Poznań, Institute of Economy and Finance, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, Poland; krzysztof.piasecki@wsb.poznan.pl

² Department of Operations Research and Mathematical Economics, Poznań University of Economics and Business, al. Niepodległości 10, 61-875 Poznań, Poland; aleksandra.wojcicka@ue.poznan.pl

Prade [1]. The set of all FN we denote by the symbol \mathbb{F} . The notion of ordered FN is introduced by Kosiński et al [4, 5]. From formal reasons, the Kosiński's theory is revised in [6]. In revised theory, the notion of ordered FN is narrowed down to the notion of oriented FN (OFN). On the other hand, arithmetic operations determined for any OFN have a very high level of complexity [7]. For this reason, we restrict our considerations to the case of trapezoidal OFNs (TrOFN) defined as follows:

Definition 1 [6]. For any monotonic sequence $(a, b, c, d) \subset \mathbb{R}$, TrOFN $\overrightarrow{Tr}(a, b, c, d) = \overrightarrow{f}$ is the pair of orientation $\overrightarrow{a, d} = (a, d)$ and FN $\mathcal{J} \in \mathbb{F}$ described by membership function $\mu_{\mathcal{J}}(\cdot | a, b, c, d) \in [0, 1]^{\mathbb{R}}$ given by the identity

$$\mu_{\mathcal{J}}(x) = \mu_{Tr}(x | a, b, c, d) = \begin{cases} 0, & x \notin [a, d] \equiv [d, a], \\ \frac{x-a}{b-a}, & x \in [a, b[\equiv [a, b], \\ 1, & x \in [b, c] \equiv [c, b], \\ \frac{x-d}{c-d}, & x \in]c, d] \equiv [c, d[. \end{cases} \quad (1)$$

Remark: The identity (3) additionally describes such modified notation of intervals which is used in the OFN theory. The notation $\mathcal{J} \equiv \mathcal{K}$ means that "the interval \mathcal{J} may be equivalently replaced by the interval \mathcal{K} ".

The symbol \mathbb{K}_{Tr} denotes the space of all TrOFNs. If $a < d$ then TrOFN $\overrightarrow{Tr}(a, b, c, d)$ has the positive orientation $\overrightarrow{a, d}$ which informs us about possibility of an increase in approximated number. The space of all positively oriented TrOFNs is denoted by the symbol \mathbb{K}_{Tr}^+ . If $a > d$, then OFN $\overrightarrow{Tr}(a, b, c, d)$ has the negative orientation $\overrightarrow{a, d}$ which informs us about possibility of a decrease in approximated number. The space of all negatively oriented TrOFNs we denote by the symbol \mathbb{K}_{Tr}^- . If $a = d$, then OFN $\overrightarrow{Tr}(a, a, a, a) = \llbracket a \rrbracket$ describes the real number $a \in \mathbb{R}$.

Let symbol $*$ denotes any arithmetic operation defined in \mathbb{R} . By symbol $\boxed{*}$ we denote an extension of arithmetic operation $*$ to \mathbb{K}_{Tr} . Kosiński has proposed to define arithmetic operators on \mathbb{K}_{Tr} , in such way that subtraction is the inverse operator to addition. In line with the Kosinski's approach, we can extend basic arithmetic operators to the case of \mathbb{K}_{Tr} , in such way that for any pair $(\overrightarrow{Tr}(a, b, c, d), \overrightarrow{Tr}(p - a, q - b, r - c, s - d)) \in \mathbb{K}_{Tr}^2$ and $\beta \in \mathbb{R}$, arithmetic operations of extended sum \boxplus and dot product \boxtimes are defined as follows [6]:

$$\begin{aligned} \overrightarrow{Tr}(a, b, c, d) \boxplus \overrightarrow{Tr}(p - a, q - b, r - c, s - d) &= \\ &= \begin{cases} \overrightarrow{Tr}(\min\{p, q\}, q, r, \max\{r, s\}), & (q < r) \vee (q = r \wedge p \leq s), \\ \overrightarrow{Tr}(\max\{p, q\}, q, r, \min\{r, s\}), & (q > r) \vee (q = r \wedge p > s). \end{cases} \end{aligned} \quad (2)$$

$$\beta \boxtimes \overrightarrow{Tr}(a, b, c, d) = \overrightarrow{Tr}(\beta \cdot a, \beta \cdot b, \beta \cdot c, \beta \cdot d). \quad (3)$$

In general, the TrOFNs addition is not associative [7]. Moreover, for any pair $(\overrightarrow{Tr}(a, b, c, d), \overrightarrow{Tr}(e, f, g, h)) \in (\mathbb{K}_{Tr}^+ \cup \mathbb{R})^2 \cup (\mathbb{K}_{Tr}^- \cup \mathbb{R})^2$ we have [7]

$$\overrightarrow{Tr}(a, b, c, d) \boxplus \overrightarrow{Tr}(e, f, g, h) = \overrightarrow{Tr}(a + e, b + f, c + g, d + h). \quad (4)$$

3 Order scale dedicated to credit risk assessment

Linguistic decision analysis is implemented in cases of solving decision-making problems under linguistic information [3]. That linguistic information should be ordered due to a given and acceptable scale. The starting point to determine any order scale is to define Tentative Order Scale (TOS) with the use of linguistic variables. TOS is defined as a following sequence:

$$\overline{TOS} = (X_i)_{i=1}^n \quad (5)$$

of linguistic labels X_i . Ordering the linguistic labels is then defined by the natural order determined by the sequence \overline{TOS} . Any TOS can also be enhanced by the intermediate values, which are obtained with the use of perception indicators (PI) given as the sequence

$$\overline{PI} = (Y_j)_{j=-m}^{j=m}. \quad (6)$$

Ordering PI is defined by natural order determined by the sequence \overline{PI} . Cartesian product of sets \overline{TOS} and \overline{PI} forms Extended Order Scale (EOS) determined as the lexicographically ordered set

$$\begin{aligned} \overline{EOS} &= \overline{TOS} \times \overline{PI} = \{(X_i, Y_j); i = \overline{1, n}, j = \overline{-m, m}\} = \\ &= \{Z_{(2 \cdot m + 1) \cdot (i - 1) + m + 1 + j}; i = \overline{1, n}, j = \overline{-m, m}\} = (Z_k)_{k=1}^{n \cdot (2 \cdot m + 1)} \end{aligned} \quad (7)$$

of order labels Z_k . The above determined EOS is called m-stage one. For the convenience of further considerations, TOS and EOS might also be characterised by Numerical Order Scale (NOS). Imprecise understanding of the meaning of order labels results in a fact that NOS should be expressed by a FN. Any triple (TOS, EOS, NOS) is called Complete Order Scale (COS).

The main subject of consideration in this paper will be a two-stage COS proposed in [8] as a tool supporting the process of assessment of potential debtors. The proposed COS contains NOS determined by TrOFNs. This COS will be called by us COS1. COS1 is presented in Table 1.

Table 1. Complete Order Scale COS1

TOS	EOS	Semantic Meaning	NOS
C	C --	much below Bad	$\overrightarrow{Tr}\left(1, 1, \frac{3}{4}, \frac{1}{4}\right)$
	C -	below Bad	$\overrightarrow{Tr}\left(\frac{5}{4}, 1, \frac{3}{4}, \frac{2}{4}\right)$
	C ~	around Bad	$\overrightarrow{Tr}\left(\frac{2}{4}, 1, 1, \frac{6}{4}\right)$
		Bad	$\overrightarrow{Tr}(1, 1, 1, 1)$
	C +	above Bad	$\overrightarrow{Tr}\left(\frac{3}{4}, 1, \frac{5}{4}, \frac{6}{4}\right)$
	C ++	much above Bad	$\overrightarrow{Tr}\left(1, 1, \frac{5}{4}, \frac{7}{4}\right)$
	B	B --	much below Average
B -		below Average	$\overrightarrow{Tr}\left(\frac{9}{4}, 2, \frac{7}{4}, \frac{6}{4}\right)$
B ~		around Average	$\overrightarrow{Tr}\left(\frac{6}{4}, 2, 2, \frac{10}{4}\right)$
		Average	$\overrightarrow{Tr}(2, 2, 2, 2)$
B +		above Average	$\overrightarrow{Tr}\left(\frac{7}{4}, 2, \frac{9}{4}, \frac{10}{4}\right)$
B ++		much above Average	$\overrightarrow{Tr}\left(2, 2, \frac{9}{4}, \frac{11}{4}\right)$
A		A --	much below Good
	A -	below Good	$\overrightarrow{Tr}\left(\frac{13}{4}, 3, \frac{11}{4}, \frac{10}{4}\right)$
	A ~	around Good	$\overrightarrow{Tr}\left(\frac{10}{4}, 3, 3, \frac{14}{4}\right)$
		Good	$\overrightarrow{Tr}(3, 3, 3, 3)$
	A +	above Good	$\overrightarrow{Tr}\left(\frac{11}{4}, 3, \frac{13}{4}, \frac{14}{4}\right)$
	A ++	much above Good	$\overrightarrow{Tr}\left(3, 3, \frac{13}{4}, \frac{15}{4}\right)$

4 Scoring function for borrowers' assessment

Each credit application \mathcal{A} is evaluated by the experts from the point of view of a criteria set $\Phi = \{C_l: l = 1, 2, \dots, p\}$ described in Table 2. The outcome of this assessment is to attribute each credit application \mathcal{A} with the set of partial assessments

$$\Psi(\mathcal{A}) = \{\overrightarrow{Tr}(\mathcal{A}, C_l) = \overrightarrow{Tr}(a_l, b_l, c_l, d_l): l = 1, 2, \dots, p\}. \quad (8)$$

The above attribution bases on an expert method using EOS and NOS presented in Table 1. The level of a scoring function will be determined as an average value of assessment in a set Ψ . However, here we encounter a specific obstacle. As we know, the addition of TrOFNs is not associative. Then multiple addition depends on the order of

the summands. This implies that a scoring function, given as an average sum of assessments $\overrightarrow{Tr}(\mathcal{A}, C_l)$, is not explicitly determined. Assuming a natural order determined by the sequence $(C_l)_{l=1}^p$ would result in the fact that the sets of partial assessments attributed to individual credit applications would vary among one another by the permutation of positively and negatively oriented TrOFNs. This prevents a reliable comparison of scoring evaluations attributed to individual credit applications. Therefore, in the considered case, any method of calculating the scoring function should be supplemented with a reasonable method of ordering portfolio components.

The guarantee of unambiguity of scoring evaluations is assured by the implementation of a following criteria ordering method utilized for each credit application \mathcal{A} separately. At the outset, we distinguish the set

$$\Phi^+(\mathcal{A}) = \{C_l: \overrightarrow{Tr}(\mathcal{A}, C_l) \in \mathbb{K}_{Tr}^+, l = 1, 2, \dots, p\} \quad (9)$$

of criteria evaluated by positively oriented TrOFNs. Then the set $\Phi^-(\mathcal{A}) = \Phi \setminus \Phi^+(\mathcal{A})$ contains all criteria evaluated by negatively oriented TrOFNs.

In the next step, using (4) we calculate a partial sum of scoring function

$$\vec{S}^+(\mathcal{A}) = \boxplus_{C_l \in \Phi^+(\mathcal{A})} \overrightarrow{Tr}(\mathcal{A}, C_l) = \overline{Tr}(\sum_{C_l \in \Phi^+(\mathcal{A})} a_l, \sum_{C_l \in \Phi^+(\mathcal{A})} b_l, \sum_{C_l \in \Phi^+(\mathcal{A})} c_l, \sum_{C_l \in \Phi^+(\mathcal{A})} d_l), \quad (10)$$

$$\vec{S}^-(\mathcal{A}) = \boxplus_{C_l \in \Phi^-(\mathcal{A})} \overrightarrow{Tr}(\mathcal{A}, C_l) = \overline{Tr}(\sum_{C_l \in \Phi^-(\mathcal{A})} a_l, \sum_{C_l \in \Phi^-(\mathcal{A})} b_l, \sum_{C_l \in \Phi^-(\mathcal{A})} c_l, \sum_{C_l \in \Phi^-(\mathcal{A})} d_l). \quad (11)$$

Finally, we calculate the value $\vec{S}(\mathcal{A})$ of scoring function. By (2) and (3), we get

$$\vec{S}(\mathcal{A}) = p^{-1} \boxminus (\vec{S}^+(\mathcal{A}) \boxplus \vec{S}^-(\mathcal{A})). \quad (12)$$

Table 2 presents partial evaluations prepared by one expert using COS1. The table also presents the values of a scoring function for this example.

Table 2. Expert's evaluations with use different Complete Order Scales

No.	Criteria	COS1		COS2		COS3		COS4	
		EOS	NOS	EOS	NOS	EOS	NOS	EOS	NOS
1	prospects of business	C+	$\overrightarrow{Tr}(\frac{3}{4}, 1, \frac{5}{4}, \frac{6}{4})$	C++	$\overrightarrow{Tr}(1, 1, \frac{5}{4}, \frac{7}{4})$	C+	$\overrightarrow{Tr}(\frac{3}{4}, 1, \frac{5}{4}, \frac{6}{4})$	C~	$\overrightarrow{Tr}(\frac{2}{4}, 1, 1, \frac{6}{4})$
2	Board members experience	A++	$\overrightarrow{Tr}(3, 3, \frac{13}{4}, \frac{15}{4})$	A++	$\overrightarrow{Tr}(3, 3, \frac{13}{4}, \frac{15}{4})$	A+	$\overrightarrow{Tr}(\frac{11}{4}, 3, \frac{13}{4}, \frac{14}{4})$	A~	$\overrightarrow{Tr}(\frac{10}{4}, 3, 3, \frac{14}{4})$
3	Chairperson experience	A++	$\overrightarrow{Tr}(3, 3, \frac{13}{4}, \frac{15}{4})$	A++	$\overrightarrow{Tr}(3, 3, \frac{13}{4}, \frac{15}{4})$	A+	$\overrightarrow{Tr}(\frac{11}{4}, 3, \frac{13}{4}, \frac{14}{4})$	A~	$\overrightarrow{Tr}(\frac{10}{4}, 3, 3, \frac{14}{4})$
4	operations range regional	C+	$\overrightarrow{Tr}(\frac{3}{4}, 1, \frac{5}{4}, \frac{6}{4})$	C++	$\overrightarrow{Tr}(1, 1, \frac{5}{4}, \frac{7}{4})$	C+	$\overrightarrow{Tr}(\frac{3}{4}, 1, \frac{5}{4}, \frac{6}{4})$	C~	$\overrightarrow{Tr}(\frac{2}{4}, 1, 1, \frac{6}{4})$
5	operations range international	A--	$\overrightarrow{Tr}(3, 3, \frac{11}{4}, \frac{9}{4})$	A--	$\overrightarrow{Tr}(3, 3, \frac{11}{4}, \frac{9}{4})$	A-	$\overrightarrow{Tr}(\frac{13}{4}, 3, \frac{11}{4}, \frac{10}{4})$	A~	$\overrightarrow{Tr}(\frac{10}{4}, 3, 3, \frac{14}{4})$
6	risk associated with market	B++	$\overrightarrow{Tr}(2, 2, \frac{9}{4}, \frac{11}{4})$	B++	$\overrightarrow{Tr}(2, 2, \frac{9}{4}, \frac{11}{4})$	B+	$\overrightarrow{Tr}(\frac{7}{4}, 2, \frac{9}{4}, \frac{10}{4})$	B~	$\overrightarrow{Tr}(\frac{6}{4}, 2, 2, \frac{10}{4})$
7	risk associated with trade	B+	$\overrightarrow{Tr}(\frac{7}{4}, 2, \frac{9}{4}, \frac{10}{4})$	B++	$\overrightarrow{Tr}(2, 2, \frac{9}{4}, \frac{11}{4})$	B+	$\overrightarrow{Tr}(\frac{7}{4}, 2, \frac{9}{4}, \frac{10}{4})$	B~	$\overrightarrow{Tr}(\frac{6}{4}, 2, 2, \frac{10}{4})$
8	risk associated with suppliers	A-	$\overrightarrow{Tr}(\frac{13}{4}, 3, \frac{11}{4}, \frac{10}{4})$	A--	$\overrightarrow{Tr}(3, 3, \frac{11}{4}, \frac{9}{4})$	A-	$\overrightarrow{Tr}(\frac{13}{4}, 3, \frac{11}{4}, \frac{10}{4})$	A~	$\overrightarrow{Tr}(\frac{10}{4}, 3, 3, \frac{14}{4})$
9	risk associated with customers	A-	$\overrightarrow{Tr}(\frac{13}{4}, 3, \frac{11}{4}, \frac{10}{4})$	A--	$\overrightarrow{Tr}(3, 3, \frac{11}{4}, \frac{9}{4})$	A-	$\overrightarrow{Tr}(\frac{13}{4}, 3, \frac{11}{4}, \frac{10}{4})$	A~	$\overrightarrow{Tr}(\frac{10}{4}, 3, 3, \frac{14}{4})$
10	diversification—products	B~	$\overrightarrow{Tr}(\frac{6}{4}, 2, 2, \frac{10}{4})$	B~	$\overrightarrow{Tr}(\frac{6}{4}, 2, 2, \frac{10}{4})$	B~	$\overrightarrow{Tr}(\frac{6}{4}, 2, 2, \frac{10}{4})$	B~	$\overrightarrow{Tr}(\frac{6}{4}, 2, 2, \frac{10}{4})$
11	diversification—sales markets	C~	$\overrightarrow{Tr}(\frac{2}{4}, 1, 1, \frac{6}{4})$	C~	$\overrightarrow{Tr}(\frac{2}{4}, 1, 1, \frac{6}{4})$	C~	$\overrightarrow{Tr}(\frac{2}{4}, 1, 1, \frac{6}{4})$	C~	$\overrightarrow{Tr}(\frac{2}{4}, 1, 1, \frac{6}{4})$
12	diversification—supply market	B~	$\overrightarrow{Tr}(\frac{6}{4}, 2, 2, \frac{10}{4})$	B~	$\overrightarrow{Tr}(\frac{6}{4}, 2, 2, \frac{10}{4})$	B~	$\overrightarrow{Tr}(\frac{6}{4}, 2, 2, \frac{10}{4})$	B~	$\overrightarrow{Tr}(\frac{6}{4}, 2, 2, \frac{10}{4})$
	Scoring value		$\overline{Tr}(\frac{97}{48}, \frac{104}{48}, \frac{107}{48}, \frac{118}{48})$		$\overline{Tr}(\frac{98}{48}, \frac{104}{48}, \frac{107}{48}, \frac{119}{48})$		$\overline{Tr}(\frac{95}{48}, \frac{104}{48}, \frac{107}{48}, \frac{116}{48})$		$\overline{Tr}(\frac{80}{48}, \frac{104}{48}, \frac{104}{48}, \frac{128}{48})$

5 Simplification of Complete Order Scales

In the experiment presented in [8] the experts used COS1 utilising a set of perception indicators $\{much\ below, below, about, above, much\ above\}$. Experts were presented with a credit application. Each of experts evaluated the credit application themselves. The outcome of the results reached by one of the experts is given in COS1 columns of Table 2. After completion of this experiment the experts agreed that the proposed method (COS1) is too complex to be used in practice.

In this case the construction of a two-stage COS1 allows to offer the experts the following simplified form of COS:

- one-stage COS2 using a set of perception indicators $\{much\ below, about, much\ above\}$,
- one-stage COS3 using a set of perception indicators $\{below, about, above\}$,
- zero-stage COS4 using a set of perception indicators $\{about\}$.

COS2 is derived from COS1 by replacing the indicators $\{below, above\}$ respectively by indicators $\{much\ below, much\ above\}$.

COS3 is derived from COS1 by replacing the indicators $\{much\ below, much\ above\}$ respectively by indicators $\{below, above\}$.

COS4 is derived from COS1 by replacing indicators $\{much\ below, below, about, above, much\ above\}$ by the indicator $\{about\}$. All COS are presented in Table 2.

For cognitive purposes the expert's evaluations were transformed by means of COS1. Using the above determined ways of replacement, the evaluations were transformed into evaluations referring to COS2, COS3 and COS4. All transformed evaluations are shown in Table 2. The last line of Table 2 shows the values of a scoring function calculated for individual types of COS. Those results indicate that the choice of COS has an impact on the value of the scoring function and therefore it can influence the final credit recommendation.

The significant advantage of COS1 is the fact that its implementation allows each expert to choose individually the most suitable type of COS. From the formal point of view, it is acceptable to allow the experts working in one team to use different types of the above determined COS. The only condition is that the expert consistently uses the chosen type of COS.

5 Conclusions

The paper presents a formal structure of COS. The knowledge of that structure allows for the transformation of a given two-stage COS to less complex structures. It was noticed that the change of COS structure influences the value of a scoring function.

In our opinion, a further examination of that influence is an interesting direction of future research which obviously will require collecting a sufficient number of data. The data must be a representative statistical sample.

Application of FNs when defining NOS always leads to the imprecision of a scoring function. A sole phenomenon of imprecision is already broadly presented in literature. However, the study referring to the imprecision of a scoring function should be conducted.

References

- [1] Dubois, D. & Prade, H. (1978) Operations on fuzzy numbers. *International Journal of System Science* 9, 613-629. <https://doi.org/10.1080/00207727808941724>
- [2] Herrera, F., Alonso, S., Chiclana, F. & Herrera-Viedma E. (2009). Computing with words in decision making: foundations, trends and prospects, *Fuzzy Optim. Decis. Making* 8, 337-364.
- [3] Herrera, F. & Herrera-Viedma, E. (2000) Linguistic decision analysis: Steps for solving decision problems under linguistic information. *Fuzzy Sets Syst.*, 115, 67-82, [doi:10.1016/S0165-0114\(99\)00024-X](https://doi.org/10.1016/S0165-0114(99)00024-X)
- [4] Kosiński, W. (2006) On fuzzy number calculus. *Int. J. Appl. Math. Comput. Sc.* 16(1)
- [5] Kosiński, W., Prokopowicz, P. & Ślęzak, D. (2002) Drawback of fuzzy arithmetics – new intuitions and propositions. In: T. Burczyński, W. Cholewa & W. Moczulski (Eds.) *Methods of Artificial Intelligence* (pp. 231-237), Gliwice, Poland
- [6] Piasecki, K. (2018) Revision of the Kosiński's Theory of Ordered Fuzzy Numbers. *Axioms* 7(1), <https://doi.org/10.3390/axioms7010016>

- [7] Piasecki, K. Łyczkowska-Hanćkowiak, A. (2021) Oriented Fuzzy Numbers vs. Fuzzy Numbers. *Mathematics* 9(3), 523. <https://doi.org/10.3390/math9050523>
- [8] Piasecki, K. & Wójcicka-Wójtowicz, A. (2021). Application of the Oriented Fuzzy Numbers in Credit Risk Assessment. *Mathematics* 9, 5, 535. <https://doi.org/10.3390/math9050535>
- [9] Simons, T. & Tripp, T.M. (2003). The Negotiation Checklist. In: R. J. Lewicki, D. M. Saunders, J. W. Minton & B. Barry (Eds.), *Negotiation Reading, Exercises and Cases* (50-63), McGraw-Hill/Irwin, New York.

Interval two-sided (max, min)-linear equations

Ján Plavka¹

Abstract. Practical problems related to scheduling optimization, modeling of fuzzy discrete dynamic systems and fuzzy analysis in which the objective function depends on the operations maximum and minimum, can be formulated and solved in max-min algebra. Systems of discrete events are obviously described by max-min linear equations. In particular, if the system is in a synchronization process, then its state is characterized by a solution of the corresponding two sided linear system. In reality, the entries of matrices and vectors are considered as intervals. The paper deals with the solvability of interval systems of two-sided (max,min)-linear equations depending on the used forall and exists quantifiers and provide the equivalent conditions for a solvability. The results are illustrated by numerical examples.

Keywords: interval solution, solvability, max-min matrix

AMS Classification: 08A72, 90B35, 90C47

1 Introduction

1.1 Motivation

This paper is concerned with a problem solvability of interval two-sided linear equations in max-min linear algebra, which is one of the subareas of tropical mathematics (addition and the multiplication have been formally replaced by the operations of maximum and minimum). Tropical mathematics (also known as idempotent mathematics) can be used in a range of practical problems related to scheduling, optimization, modeling of discrete dynamic systems, graph theory, knowledge engineering, cluster analysis, fuzzy systems and also those related to describing the diagnosis of technical devices [32] or medical diagnosis [31].

Assume that a goods market is represented by an interactive transfer network consisting of m suppliers and one central C . The goods are sent from supplier s_i , $i \in M$ to the central C whereby the explored goods have to be sent back to s_i . Moreover, assume that the connection between s_i and C is only possible via one of n control points p_j , $j \in N$. Further assume that the connections between the s_i and the p_j are one-way connections, and that the capacity of the connection between s_i and p_j is equal to a_{ij} . The control points p_j are connected with C by two-way connections with capacities x_j in both directions. The goods are transfer in goods packets, and any goods packet is transfer over just one connection as an inseparable unit. Thus, the total capacity of the connection between p_i and C is $\max_{j \in N} \{\min\{a_{ij}, x_j\}\}$, whereby the various connections used are included as the maximum of their capacities (not their sum).

The transfer from C to s_i is made over other one-way connections between the control point p_j and s_i with capacities equal to b_{ij} . Since the connections between C and p_j are two-way connections, the total capacity of the connection between C and s_i is equal to $\max_{j \in N} \{\min\{b_{ij}, x_j\}\}$. The task is to achieve a synchronization process in which the maximal capacity of all connections between s_i and C via p_j is equal to the maximal capacity of all connections between C and s_i in the reverse direction, i.e. we look for x_j such that

$$\max_{j \in N} \{\min\{a_{ij}, x_j\}\} = \max_{j \in N} \{\min\{b_{ij}, x_j\}\} \text{ for all } i \in M$$

or in a matrix-vector form $A \otimes x = B \otimes x$, where $a \oplus b = \max\{a, b\}$ and $a \otimes b = \min\{a, b\}$.

The aim of this paper is to characterize a weak solvability of the interval system $A \otimes x = B \otimes x$ and present its equivalent conditions for matrices with inexact (interval) entries since in practice, the values of the matrix entries obtained as data errors are not exact numbers and they are usually contained in some intervals. Interval arithmetic is an efficient way to represent matrices in a guaranteed way on a computer. Moreover, this paper describes polynomial procedure recognizing whether a given reduced interval vector is weak solution of a given interval matrices.

Notice that the paper [30] is related work to this paper where the concepts of interval generalized eigenvectors in fuzzy algebra were introduced and four versions of generalized eigenvectors were studied. The results on interval versions of eigenproblem are given in [5, 6]. The solvability of interval max-min matrix equations is suggested in [26]. The X -robustness of max-min matrices extended to interval vectors using forall–exists quantification of interval entries is presented in [23, 25].

¹ Technical University, Department of Mathematics and Theoretical Informatics, Nemcovej 32, 04200 Košice, Slovakia, Jan.Plavka@tuke.sk

1.2 Preliminaries

Let (\mathbb{B}, \leq) be a bounded linearly ordered set with the least element in \mathbb{B} denoted by O and the greatest one by I . The set of natural numbers is denoted by \mathbb{N} . For given $m, n \in \mathbb{N}$, write $M = \{1, 2, \dots, m\}$ and $N = \{1, 2, \dots, n\}$. The set of $m \times n$ matrices over \mathbb{B} is denoted by $\mathbb{B}(m, n)$, and in particular, the set of $n \times 1$ vectors over \mathbb{B} is denoted by $\mathbb{B}(n)$ and for $\alpha \in \mathbb{B}$ a constant vector is denoted by $\alpha^* = (\alpha, \dots, \alpha)^T$.

A max-min (fuzzy) algebra is a triple $(\mathbb{B}, \oplus, \otimes)$, where $a \oplus b = \max\{a, b\}$ and $a \otimes b = \min\{a, b\}$.

The operations \oplus, \otimes are extended to the matrix-vector algebra over \mathbb{B} by a direct analogy to the conventional linear algebra.

For $A \in \mathbb{B}(m, n)$, $C \in \mathbb{B}(m, n)$ we write $A \leq C$ if $a_{ij} \leq c_{ij}$ for all $i, j \in N$. Similarly, for $x = (x_1, \dots, x_n)^T \in \mathbb{B}(n)$ and $y = (y_1, \dots, y_n)^T \in \mathbb{B}(n)$ we write $x \leq y$ if $x_i \leq y_i$ for each $i \in N$.

Let $L \subseteq \mathbb{B}$. The element $x^\oplus \in L$ is called maximum element of L if $x \leq x^\oplus$ for every $x \in L$.

Note that both operations in fuzzy algebra are idempotent, hence no new numbers are created in the process of generating matrix-vector products.

2 Weak solution

2.1 General case of a solvability

Similarly as in [3]–[6] and [16]–[30] consider an interval vector with bounds $\underline{x}, \bar{x} \in \mathbb{B}(n)$ and interval matrices with bounds $\underline{A}, \bar{A} \in \mathbb{B}(m, n)$, $\underline{B}, \bar{B} \in \mathbb{B}(m, n)$ as follows

$$\begin{aligned} X &= [\underline{x}, \bar{x}] = \{x \in \mathbb{B}(n); \underline{x} \leq x \leq \bar{x}\}, \\ A &= [\underline{A}, \bar{A}] = \{A \in \mathbb{B}(m, n); \underline{A} \leq A \leq \bar{A}\}, \\ B &= [\underline{B}, \bar{B}] = \{B \in \mathbb{B}(m, n); \underline{B} \leq B \leq \bar{B}\}. \end{aligned}$$

For given indices $i \in M, j \in N$ we define generators $\tilde{A}^{(ij)} \in \mathbb{B}(m, n)$, $\tilde{B}^{(ij)} \in \mathbb{B}(m, n)$, $\tilde{x}^{(i)} \in \mathbb{B}(n)$ of A, B, X by putting for every $k \in M, l \in N$

$$\begin{aligned} \tilde{a}_{kl}^{(ij)} &= \begin{cases} \bar{a}_{ij}, & \text{for } k = i, l = j \\ \underline{a}_{kl}, & \text{otherwise} \end{cases}, & \tilde{b}_{kl}^{(ij)} &= \begin{cases} \bar{b}_{ij}, & \text{for } k = i, l = j \\ \underline{b}_{kl}, & \text{otherwise} \end{cases}, \\ \tilde{x}_k^{(i)} &= \begin{cases} \bar{x}_i, & \text{for } k = i \\ \underline{x}_k, & \text{otherwise} \end{cases}. \end{aligned}$$

Lemma 1. [30] Let $x \in \mathbb{B}(n)$ and $A \in \mathbb{B}(m, n)$. Then

- (i) $x \in X$ if and only if $x = \bigoplus_{i \in N} \beta_i \otimes \tilde{x}^{(i)}$ for some values $\beta_i \in \mathbb{B}$ with $\underline{x}_i \leq \beta_i \leq \bar{x}_i$,
- (ii) $A \in A$ if and only if $A = \bigoplus_{i \in M, j \in N} \alpha_{ij} \otimes \tilde{A}^{(ij)}$ for some values $\alpha_{ij} \in \mathbb{B}$ with $\underline{a}_{ij} \leq \alpha_{ij} \leq \bar{a}_{ij}$.

We will consider the following interval solution of the system

$$A \otimes X = B \otimes X \tag{1}$$

depending on the used quantifiers and their order.

Definition 1. If A, B and X are given, then X is called a weak solution of $A \otimes X = B \otimes X$ if

$$(\exists x \in X)(\exists A \in A)(\forall B \in B) A \otimes x = B \otimes x.$$

Theorem 2. Suppose given A, B and X . Then X is a weak solution of $A \otimes X = B \otimes X$ if and only if

$$(\exists x \in X)(\exists A \in A) A \otimes x = \underline{B} \otimes x = \bar{B} \otimes x. \tag{2}$$

Proof. Suppose that there are $x \in X$ and $A \in A$ such that $A \otimes x = \underline{B} \otimes x$ and $A \otimes x = \overline{B} \otimes x$. The implication follows from the formula

$$A \otimes x = \underline{B} \otimes x \leq B \otimes x \leq \overline{B} \otimes x = A \otimes x.$$

The reverse implication trivially follows. \square

Notice that the product $A \otimes x$ can be expressed as a max-min linear combination of generators of A and X according to Lemma 1, i.e. $A = \bigoplus_{i \in M, j \in N} \alpha_{ij} \otimes \bar{A}^{(ij)}$, $x = \bigoplus_{i \in N} \gamma_i \otimes \bar{x}^{(i)}$. After a matrix-vector multiplication $A \otimes x$ we obtain a combination of coefficients γ_k with α_{ij} , or in others words, we get a quadratic part of equality which we do not know to solve. To avoid this type of a complication we shall reduce weak solution of $A \otimes X = B \otimes X$ into not interval form such that, at first, we shall solve the equality $\underline{B} \otimes x = \overline{B} \otimes x$.

Denote $S(\underline{B}, \overline{B}) = \{x \in \mathbb{B}(n); \underline{B} \otimes x = \overline{B} \otimes x\}$.

Theorem 3. Suppose given A , B and X . Then X is a weak XEA-solution of $A \otimes X = B \otimes X$ if and only if

$$(\exists x \in S(\underline{B}, \overline{B}))(\exists A \in A) A \otimes x = \underline{B} \otimes x. \quad (3)$$

Proof. The proof follows from Theorem 2. \square

Theorem 4. Suppose given A , B and X . Then X is a weak solution of $A \otimes X = B \otimes X$ if and only if

$$(\exists x \in S(\underline{B}, \overline{B})) \underline{A} \otimes x \leq \underline{B} \otimes x \leq \overline{A} \otimes x. \quad (4)$$

Proof. Suppose that X is a weak solution of $A \otimes X = B \otimes X$, i.e. there are $x \in X$ and $A \in A$ such that for each $B \in B$ the equation $A \otimes X = B \otimes X$ is fulfilled. By Theorem 3 there are $x \in S(\underline{B}, \overline{B})$ and $A \in A$ such that $A \otimes x = \underline{B} \otimes x (= \overline{B} \otimes x)$, hence we get $\underline{A} \otimes x \leq A \otimes x = \underline{B} \otimes x \leq \overline{A} \otimes x$.

Conversely, assume that $x \in S(\underline{B}, \overline{B})$ satisfies $\underline{A} \otimes x \leq \underline{B} \otimes x = \overline{B} \otimes x \leq \overline{A} \otimes x$. Then there is $A \in A$ such that $A \otimes x = \underline{B} \otimes x \in [\underline{A} \otimes x, \overline{A} \otimes x]$. \square

If a basis of $S(A, B)$ is denoted by $\mathcal{L}(A, B) = \{\ell_1, \dots, \ell_k\}$ then any vector $x \in S(A, B)$ can be expressed as a max-min linear combination of vectors from $\{\ell_1, \dots, \ell_k\}$, i.e. $x = \bigoplus_{i=1}^k \gamma_i \otimes \ell_i$. Now we can reformulate the last theorem.

Theorem 5. Suppose given A , B and X . Then X is a weak solution of $A \otimes X = B \otimes X$ if and only if there is $\gamma_1, \dots, \gamma_k \in \mathbb{B}$ such that

$$\underline{A} \otimes \bigoplus_{i=1}^k \gamma_i \otimes \ell_i \leq \underline{B} \otimes \bigoplus_{i=1}^k \gamma_i \otimes \ell_i \leq \overline{A} \otimes \bigoplus_{i=1}^k \gamma_i \otimes \ell_i, \quad (5)$$

where the set of vectors $\mathcal{L} = \{\ell_1, \dots, \ell_k\}$ is a basis of $S(A, B)$.

Proof. The proof follows from Theorem 4. \square

Inequalities (5) can be rewritten into following form

$$\begin{aligned} \bigoplus_{i=1}^k \underline{A} \otimes \ell_i \otimes \gamma_i &\leq \bigoplus_{i=1}^k \underline{B} \otimes \ell_i \otimes \gamma_i \\ \bigoplus_{i=1}^k \underline{B} \otimes \ell_i \otimes \gamma_i &\leq \bigoplus_{i=1}^k \overline{A} \otimes \ell_i \otimes \gamma_i, \end{aligned} \quad (6)$$

where $\gamma_1, \dots, \gamma_k$ are unknown variables. Observe that (6) gives a hint how to decide whether X is a weak solution of $A \otimes X = B \otimes X$. It suffices to find $\gamma_1, \dots, \gamma_k$ for a given basis $\mathcal{L}(\underline{B}, \overline{B})$.

Note that recognition of the solvability of a system $C \otimes y \leq D \otimes y$ needs $O(rt \cdot \min(r, t))$ time, where $C, D \in \mathbb{B}(r, t)$ (see [9], Algorithm \mathcal{A}).

In the process of decision whether X is or not a weak solution of $A \otimes X = B \otimes X$ is necessary to compute a basis of $S(\underline{B}, \overline{B})$. Unfortunately, the problem of computing minimal (the set of vectors which generate a basis of $S(\underline{B}, \overline{B})$) is long-time open question. The paper [7] describes the whole max-min eigenspace explicitly, what is good starting point for process of obtaining a basis for the case when $\underline{B} = \overline{B} = E$ and $\lambda = O$.

2.2 The restriction of a solvability

Theorem 6. Suppose given $A, B \in \mathbb{B}(m, n)$, $A \leq B$ and $\alpha(A, B) = \bigotimes_{i \in M} \bigoplus_{j \in N} a_{ij}$. Then $x \in S(A, B)$ for any $x \leq \alpha^*(A, B)$.

Proof. Suppose that $A \leq B$, $x \leq \alpha^*(A, B)$ and $x_k = \max_{i \in N} \{x_i\}$. Then we get

$$(A \otimes x)_i = \bigoplus_{j \in N} a_{ij} \otimes x_j = a_{ik} \otimes x_k = x_k = b_{ik} \otimes x_k = \bigoplus_{j \in N} b_{ij} \otimes x_j = (B \otimes x)_i$$

and hence we have that $x \in S(A, B)$. □

Suppose that A, B and X are given. Denote $\{x \in X; O^* \leq x \leq \alpha^*(\underline{B}, \overline{B})\} (= [O^*, \alpha^*(\underline{B}, \overline{B})])$ by Y_X .

Theorem 7. Suppose given A, B and Y_X . Then Y_X is a weak solution of $A \otimes Y_X = B \otimes Y_X$ if and only if there is $y \in Y_X$ such that $\underline{A} \otimes y \leq \underline{B} \otimes y \leq \overline{A} \otimes y$.

Proof. The proof follows from Theorem 4 and Theorem 6. □

For given indices $i \in N$ denote the generators of Y_X by

$$\tilde{y}_k^{(i)} = \begin{cases} \overline{y}_i, & \text{for } k = i \\ \underline{y}_k, & \text{otherwise} \end{cases}.$$

According to Theorem 7 we will look for $y = \bigoplus_{i \in N} \gamma_i \otimes \tilde{y}^{(i)} \in Y_X$ such that

$$\begin{aligned} \bigoplus_{i=1}^k \underline{A} \otimes \tilde{y}^{(i)} \otimes \gamma_i &\leq \bigoplus_{i=1}^k \underline{B} \otimes \tilde{y}^{(i)} \otimes \gamma_i \\ \bigoplus_{i=1}^k \underline{B} \otimes \tilde{y}^{(i)} \otimes \gamma_i &\leq \bigoplus_{i=1}^k \overline{A} \otimes \tilde{y}^{(i)} \otimes \gamma_i, \end{aligned} \quad (7)$$

where $\gamma_1, \dots, \gamma_k, \underline{y}_i \leq \gamma_i \leq \overline{y}_i$ are unknown variables and which can be computed by Algorithm \mathcal{A} (see [9]) in $O(mn)$ time.

Theorem 8. Suppose given A, B, Y_X . Then Y_X is a weak solution of $A \otimes Y_X = B \otimes Y_X$ if and only if the max-min linear system (7) has a solution γ satisfying the condition $\underline{y} \leq \gamma \leq \overline{y}$.

Proof. Suppose that the max-min linear system (7) has a solution γ satisfying the condition $\underline{y} \leq \gamma \leq \overline{y}$. Then the vector $y \in \mathbb{B}(n)$ defined as the max-min linear combination $y = \bigoplus_{k \in N} \gamma_k \otimes \tilde{y}^{(k)}$ belongs to the interval $[\underline{y}, \overline{y}]$ in view of Lemma 1(i). Thus, according to Theorem 7, Y_X is a weak solution of $A \otimes Y_X = B \otimes Y_X$.

If Y_X is a weak solution of $A \otimes Y_X = B \otimes Y_X$, then by Theorem 4 there exist $y \in Y_X$ such that the inequalities $\underline{A} \otimes y \leq \underline{B} \otimes y \leq \overline{A} \otimes y$ holds true. By Lemma 1(i), there exist coefficients $\gamma_i \in \mathbb{B}$, $i \in N$ such that $y = \bigoplus_{i \in N} \gamma_i \otimes \tilde{y}^{(i)}$ and $\underline{y}_i \leq \gamma_i \leq \overline{y}_i$. It is easy to verify that $y \in \mathbb{B}(n, 1)$, where $y_i = \gamma_i$ for every $i \in N$, satisfies the conditions (7). □

Example 1. Suppose that $\mathbb{B} = [0, 10]$ and consider interval matrices A, B and interval vector X which have the forms

$$\underline{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 2 & 1 \end{pmatrix}, \quad \overline{A} = \begin{pmatrix} 2 & 4 \\ 4 & 5 \\ 3 & 1 \end{pmatrix},$$

$$\underline{B} = \begin{pmatrix} 3 & 4 \\ 4 & 5 \\ 3 & 5 \end{pmatrix}, \bar{B} = \begin{pmatrix} 4 & 4 \\ 5 & 8 \\ 7 & 5 \end{pmatrix},$$

$$\underline{x} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \bar{x} = \begin{pmatrix} 4 \\ 8 \end{pmatrix}.$$

Task: Check whether Y_X is a weak solution of $A \otimes Y_X = B \otimes Y_X$.

To construct the system of max-min linear inequalities (7) we need to compute vectors α^* , \underline{y} , \bar{y} and matrices $\underline{A} \otimes \tilde{y}^{(i)}$, $\bar{A} \otimes \tilde{y}^{(i)}$, $\underline{B} \otimes \tilde{y}^{(i)}$ for $i = 1, 2$ which have the following forms:

$$\alpha^* = \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \underline{y} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \bar{y} = \begin{pmatrix} 4 \\ 4 \end{pmatrix},$$

$$(\underline{A} \otimes \tilde{y}^{(1)}, \underline{A} \otimes \tilde{y}^{(2)}) = \begin{pmatrix} 2 & 2 \\ 3 & 4 \\ 2 & 2 \end{pmatrix}, (\bar{A} \otimes \tilde{y}^{(1)}, \bar{A} \otimes \tilde{y}^{(2)}) = \begin{pmatrix} 3 & 4 \\ 4 & 4 \\ 3 & 2 \end{pmatrix}, (\underline{B} \otimes \tilde{y}^{(1)}, \underline{B} \otimes \tilde{y}^{(2)}) = \begin{pmatrix} 3 & 4 \\ 4 & 4 \\ 3 & 4 \end{pmatrix}.$$

Then the system of max-min linear inequalities (7)

$$\begin{pmatrix} \underline{A} \otimes \tilde{y}^{(1)}, \underline{A} \otimes \tilde{y}^{(2)} \\ \underline{B} \otimes \tilde{y}^{(1)}, \underline{B} \otimes \tilde{y}^{(2)} \end{pmatrix} \otimes \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \leq \begin{pmatrix} \underline{B} \otimes \tilde{y}^{(1)}, \underline{B} \otimes \tilde{y}^{(2)} \\ \bar{A} \otimes \tilde{y}^{(1)}, \bar{A} \otimes \tilde{y}^{(2)} \end{pmatrix} \otimes \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \Leftrightarrow$$

$$\begin{pmatrix} 2 & 2 \\ 3 & 4 \\ 2 & 2 \\ 3 & 4 \\ 4 & 4 \\ 3 & 4 \end{pmatrix} \otimes \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \leq \begin{pmatrix} 3 & 4 \\ 4 & 4 \\ 3 & 4 \\ 3 & 4 \\ 4 & 4 \\ 3 & 2 \end{pmatrix} \otimes \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}$$

has the greatest solution $\gamma = (4, 3)^T$ (see [9]) and we can conclude that Y_X is a weak solution of $A \otimes Y_X = B \otimes Y_X$.

3 Conclusion

In this paper, the weak solvability of interval systems of two-sided (max,min)-linear equations depending on the used forall and exists quantifiers is considered and the equivalent conditions for a weak solvability are provided. The obtained results have been illustrated by numerical example.

Acknowledgement

The support of the APVV grant #180373 and VEGA #1/0327/20 is gratefully acknowledged.

References

- [1] A. Di Nola and B. Gerla: Algebras of Łukasiewicz's logic and their semiring reducts. In G. L. Litvinov and V. P. Maslov, editors, *Idempotent mathematics and mathematical physics*, pages 131–144, 2005.
- [2] A. Di Nola and C. Russo: Łukasiewicz transform and its application to compression and reconstruction of digital images. *Information Sci.*, 177:1481–1498, 2007.
- [3] M. Gavalec, J. Plavka: Strong regularity of matrices in general max-min algebra. *Lin. Algebra Appl.* **371** (2003), 241–254.
- [4] M. Gavalec, J. Plavka and H. Tomášková: Interval eigenproblem in max-min algebra *Linear Algebra and its Applications* 440 (2014) 24–33.

- [5] M. Gavalec, J. Plavka and D. Ponce: Tolerance types of interval eigenvectors in max-plus algebra *Information Sciences* 367 (2016) 14–27.
- [6] M. Gavalec, J. Plavka and D. Ponce: Strong tolerance of interval eigenvectors in fuzzy algebra. *Fuzzy Sets and Systems* 369, (2019) 145-156.
- [7] M. Gavalec, Z. Němcová and S. Sergeev: (K,L)-eigenvectors in max-min algebra. *Fuzzy Sets and Systems* 410, (2021) 75–89.
- [8] M. Gavalec: *Periodicity in Extremal Algebra*. Gaudeamus, Hradec Králové 2004.
- [9] M. Gavalec and K. Zimmermann: Solving systems of two-sided (max,min)–linear equations. *Kybernetika* 46 (2010) 405–414.
- [10] J. S. Golan: *Semirings and Their Applications*. Springer, 1999.
- [11] M. Gondran and M. Minoux: *Graphs, Dioids and Semirings: New Models and Algorithms* Springer 2008.
- [12] B. Heidergott, G.-J. Olsder and J. van der Woude: *Max-plus at Work*. Princeton University Press, 2005.
- [13] V. N. Kolokoltsov and V.P. Maslov: *Idempotent analysis and its applications*. Kluwer, Dordrecht, 1997.
- [14] G. L. Litvinov and V.P. Maslov (eds.): *Idempotent Mathematics and Mathematical Physics*, vol. 377 of *Contemporary Mathematics*, AMS, 2005.
- [15] G. L. Litvinov and S. N. Sergeev (eds.): *Tropical and Idempotent Mathematics*, vol. 495 of *Contemporary Mathematics*, AMS, 2009.
- [16] M. Hireš, M. Molnárová, P. Drotár: Robustness of Interval Monge Matrices in Fuzzy Algebra. *Mathematics* 8(4), (2020) 1–16.
- [17] M. Molnárová: Possible and universal robustness of special classes of matrices with inexact data. In: *Mathematical methods in economics – Proceedings of the 36th international conference – Praha (2018)* 348–353.
- [18] M. Molnárová: Fuzzy interval Monge matrices with respect to robustness. *Mathematical methods in economics – Proceedings of the 37th international conference České Budějovice (2019)* 409-414.
- [19] H. Myšková: On an algorithm for testing T4 solvability of fuzzy interval systems *Kybernetika* 48(5) (2012) 924–938.
- [20] H. Myšková: An iterative algorithm for testing solvability of max-min interval systems. *Kybernetika* 48(5) (2012) 879–889.
- [21] H. Myšková and J. Plavka: X-robustness of interval circulant matrices in fuzzy algebra. *Linear Algebra and its Applications* 438 (2013) 2757–2769.
- [22] H. Myšková and J. Plavka: The robustness of interval matrices in max-plus algebra. *Linear Algebra and its Applications* 445 (2013) 85–102.
- [23] H. Myšková and J. Plavka: Reachability of eigenspaces for interval matrices in max-min algebra. *Linear Algebra and its Applications* 578 (2019) 314–333.
- [24] H. Myšková and J. Plavka: X-AE and X-EA robustness of max-min matrices. *Discrete Applied Mathematics* 267 (2019) 142–150.
- [25] H. Myšková and J. Plavka: AE and EA robustness of interval circulant matrices in max-min algebra. *Fuzzy Sets and Systems* 384 (2020) 91–104.
- [26] H. Myšková and J. Plavka: On the solvability of interval max-min matrix equations. *Linear Algebra and its Applications* 590 (2020) 85–96.
- [27] J. Plavka: On the $O(n^3)$ algorithm for checking the strong robustness of interval fuzzy matrices. *Discrete Applied Mathematics* 160 (2012), 640–647.
- [28] J. Plavka: 1-parametric Eigenproblem in max-algebra. *Discrete Applied Mathematics* 150 (2005), 16–28.
- [29] J. Plavka, The weak robustness of interval matrices in max-plus algebra, *Discrete Applied Mathematics* 173 (2014) 92–101.
- [30] J. Plavka and M. Gazda: Generalized eigenproblem of interval max-min (fuzzy) matrices. *Fuzzy Sets and Systems* 410 (2021), 27–44.
- [31] E. Sanchez: Resolution of eigen fuzzy sets equations, *Fuzzy Sets and Systems* 1 (1978), 69–74.
- [32] K. Zimmermann: *Extremální algebra (in Czech)*. Ekon. ústav ČSAV Praha, 1976.

Forecasting of agrarian commodity prices by time series methods

Alena Pozdílková¹, Jaromír Zahrádka², Jaroslav Marek³

Abstract. It is very important for farm management to correctly estimate future demand and agricultural commodity price developments. In the paper, we will use the decomposition of time series into a linear trend component and seasonal component using Fisher's periodogram. Our numerical studies work with time series of selected agrarian commodities on the Commodity Exchange. The theoretical background, assumptions, advantages and disadvantages of methods are also briefly presented and discussed. In the discussion section, the success of the implemented algorithm is measured. We will work with the prices of the period from January 1970 to December 2020 for agrarian commodities from the New York Board of Trade (NYBOT). We will focus on the prediction of the price development of wheat and oat.

Keywords: time series, Commodity Exchange, price of wheat, price of oat

JEL Classification: C32, C44

AMS Classification: 90C15

1 Introduction

Time series in the field of agriculture and their analysis have been studied by many authors. Prediction of future prices of agro commodities is very useful for managerial decisions of farm owners. Decomposition into trend, cyclical, seasonal and error components or smoothing methods or methods based on Box-Jenkins methodology are often used. In this paper, we choose the first approach, when we find the linear trend using the least squares method and determine the cyclical component using Fisher's periodogram.

Mathematical modeling for crop pricing in India was studied by [2], authors also used differential equations to solve this problem. They developed the mathematical model is to evaluate crop valuation considering the cost involved in producing the crop-based products and productivity. Also in India, forecasting of wheat prices was studied by [1], this research is focused on wheat and barley. There are many methods to grasp the issue, for example, the method of hidden periodicities and its application to a 26-day period is studied by [9]. Another important methodology of Bernstein's trigonometric interpolation polynomial is studied by [10]. Commodity price analysis using methods is published in the article [5] or [11], in [8] is based on the Box-Jenkins methodology.

In the Czech Republic, crop exchange has existed since the establishment of the republic in 1918. However, its activity was terminated in the era of communist governments. Data available for processing thus included the period of 1992 to 2007 (Commodity Exchange Brno) and the period of 2007-2019 (Czech Moravian Commodity Exchange Kladno). Therefore, data from the New York Board of Trade (NYBOT) was chosen for processing.

2 Methodology

We will use classical approaches when analyzing time series of agricultural crops. We decompose the time series into a trend and periodic component. All calculations in this article and graphical outputs were performed in MATLAB using our own programmed procedures.

2.1 Decomposition method

Suppose the measurement – a random vector \mathbf{Y} fulfills following decomposition model

¹ University of Pardubice, Department of Mathematics and Physics, Studentská 95, 53210 Pardubice, jaroslav.marek@upce.cz.

² University of Pardubice, Department of Mathematics and Physics, Studentská 95, 53210 Pardubice, jaromir.zahradka@upce.cz.

³ University of Pardubice, Department of Mathematics and Physics, Studentská 95, 53210 Pardubice, alena.pozdilкова@upce.cz.

$$\mathbf{Y} = \mathbf{T} + \mathbf{C} + \varepsilon, \varepsilon \sim \text{i. i. d.}, \quad (1)$$

where \mathbf{T}_t is a trend component and \mathbf{C}_t is a cyclical component at time x_t . We considered a trend component in the linear form

$$\mathbf{T}_t = \beta_0 + \beta_1 x_t,$$

where x is the day from date 1.1.2014 and y is a prize of a commodity. We calculated an ordinary square estimate $\hat{y}_t = \beta_0 + \beta_1 x_t$ of the regression line using well know ordinary squares method and formulas for β_0 and β_1 estimates. Our recent research focused on significant positive trend testing. Thus, we tested the hypothesis $H_0: \beta_1 = 0$ based on test statistics T :

$$T = \frac{|b_1 - \beta_1|}{SE_{\hat{\beta}}} = \frac{b_1}{s} \sqrt{\sum x_i^2 - n\bar{x}^2} \sim t_{n-2}, \quad (2)$$

where

$$s^2 = \hat{\sigma}^2 = \frac{SS_{res}}{n-2}, \quad SS_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 \quad (3)$$

The quality of estimated regression line is given by the index of determination

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}, \quad (4)$$

where

$$SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2. \quad (5)$$

Next, we will search for the periodic component using Fisher's periodogram.

2.2 Periodogram

Given a time series with real values X_1, X_2, \dots, X_n of length n , let us consider the function

$$I(\lambda) = \frac{1}{2\pi n} \left| \sum_{i=1}^n X_t e^{-it\lambda} \right|^2, \quad -\pi \leq \lambda \leq \pi. \quad (6)$$

Function $I(\lambda)$ is called a periodogram of the sequence X_1, X_2, \dots, X_n .

The periodogram $I(\lambda)$ is a key tool in harmonic analysis. If the data contains strong periodic components, this will cause peaks in the periodogram at oscillation frequencies. Thus, it is advantageous to attribute the peaks in the periodogram to the systematic periodic components in the basic mechanism that generated the data. For each fixed n function is $I(\lambda)$ a random variable. Therefore $I(\lambda), -\pi \leq \lambda \leq \pi$ forms a random process. Each process implementation is a continuous function. If the length of the sequence n is small compared to the length of the current period

$$T_j = \frac{2\pi}{\lambda_j}, \quad (7)$$

where $\lambda_1, \dots, \lambda_p$ are mutually different numbers from interval $(-\pi, \pi)$, this period will look more like a trend. On the other hand, a very short period is also not easily recognizable. It is obvious that the shortest detectable periodicity has a length $T = 2$. According to (2.2) it corresponds to the frequency $\lambda = \pi$, which is called Nyquist frequency.

The calculation can be easily performed using a formula

$$I(\lambda) = \frac{1}{4\pi} A^2(\lambda) + \frac{1}{4\pi} B^2(\lambda), \quad (8)$$

where

$$A(\lambda) = \sqrt{2/N} \sum_{t=1}^N X_t \cos x\lambda, \quad B(\lambda) = \sqrt{2/N} \sum_{t=1}^N X_t \sin x\lambda. \quad (9)$$

Fisher's test makes it possible to assess the statistical significance of the peak of the observed periodogram. Fisher's statistic is the ratio of the maximum value of a periodogram to the sum of all periodogram values. [3], [6]

2.3 Sliding windows and spectral decomposition

Consider the data in the sliding window of specified length. First, we adjust the data from the trend component, see section 2.2. In the "classic" application of sliding window, we process the data of the original file indexed from one to the length of the window within the first window. Then we move the window by one index and process the data again, which would have indexes starting with two in the original file. In the following iterations, we move the windows until we reach the end of the time series. Such time series in every window can be explained by significant periods. In Fig. 1 the decomposition into four components is shown. The window can therefore be represented by periods $(\lambda_1^i, \lambda_2^i, \lambda_3^i, \lambda_4^i)$. A blue time series is decomposed into four sine waves with different frequencies and periods. Dashed vertical lines in Figure 1 show the location of the hidden frequencies λ^i .

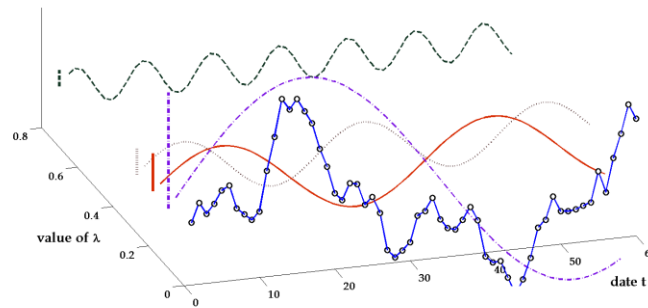


Figure 1 Example of spectral decomposition.

The results obtained by periodogram analysis simply provide information on the periodicity. We model the explained cyclic component by the relation

$$C_t = \sum_{k=1}^k A_k \cos \lambda_k x_t, \quad (10)$$

where A_k are the amplitudes, x_t is time. Estimates of unknown amplitude values can be found again using OLS.

3 Numerical study – results and discussion

In our numerical study, daily data of wheat and oat prices has been often used. Both commodities will be available in the period from January 5, 1970 to January 3, 2020. Therefore, data were available for a continuous period of 50 years. On non-trading days, the data were supplemented with linearly approximated values for wheat and oat between the last day of trading before the break and the first day of trading after the break. The trend and harmonic analysis was performed in a) 4-year window; (b) 16-year window. The beginning of the windows was always elected on 1 September and the end of the window always on 31 August by 4 years, resp. 16 years later.

The price estimate from model (1) was always made from the first day after the end of the window, i.e. from 1 September to 31 August of the following year, i.e. for 1 year.

To analyze the success of the prediction, the predicted prices and the actual prices were compared. When comparing, we will focus on the dates 30. 11., 28. 2. or 29. 2., 31. 5., and 31. 8. I.e. that we compare the estimate with the actual price for 3, 6, 9, respectively 12 months from the end of the selected floating window.

Such predictions will allow us to optimize the date of sale of harvested grain or to optimize the amount of area sown with autumn-sown grains.

The analysis and predictions for wheat and will be done, in the following Figures 2 - Figure 5 can be seen trend and prediction for sliding window of 4 and 16 years.

For both crops, we monitored the ability of the methodology to correctly determine the trend of price developments. Win ratio WR has often used to measure the success of trading strategies on the exchange. Win ratio is the ratio of successful signals to total purchase signals. In 33 sliding windows for 16 years or all 45 sliding windows for 4 years, we predicted the development for the following year using an algorithm. The first window was given by the period 1970 to 2021. First, we determined the relative change in the actual average prices for September of

the current year and the average October prices of the following year. We also determined the relative change in the actual average prices for September of the current year and the average predicted October prices of the following year. If both relative changes had the same sign for a given sliding window, then we considered the estimation methodology to be successful. For wheat November prediction and window of 4-year length, we obtained $WR = 22/45 = 0.49$. For wheat November prediction and window of 16-year length, we obtained $WR = 12/33 = 0.36$. For oat November prediction and window of 4-year length, we obtained $WR = 21/45 = 0.47$. For oat November prediction and window of 16-year length, we obtained $WR = 20/33 = 0.61$. This success of the price prediction is not seemingly visible. However, the technical indicators used for stock market or Forex predictions are able to provide signals to buy with a WR higher than .5, which leads to a profit on trading.

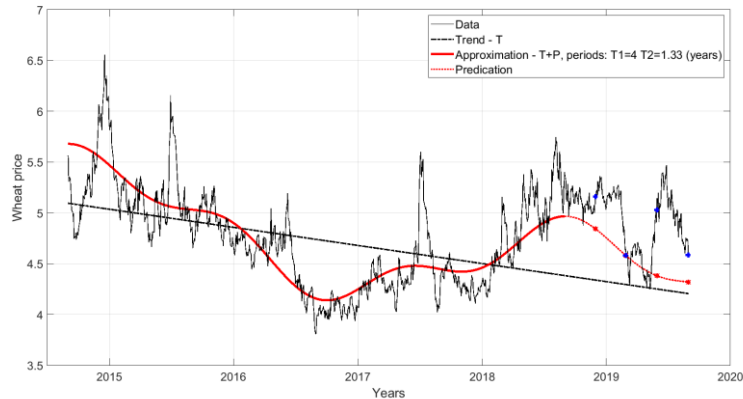


Figure 2 Wheat: fitting of time series with prediction, window 4 years.

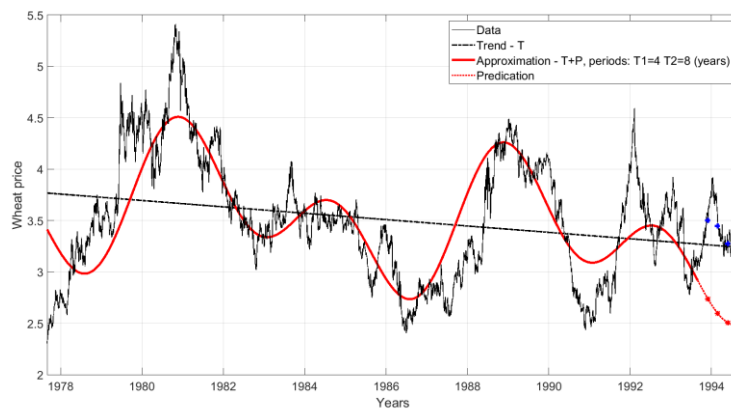


Figure 3 Wheat: fitting of time series with prediction, window 16 years.

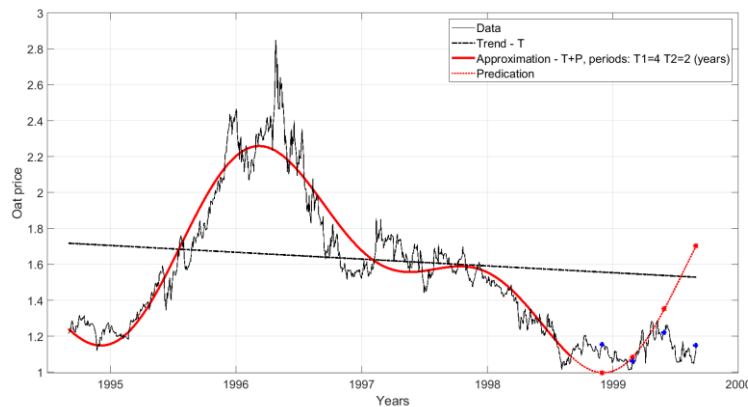


Figure 4 Oat: fitting of time series with prediction, window 4 years.

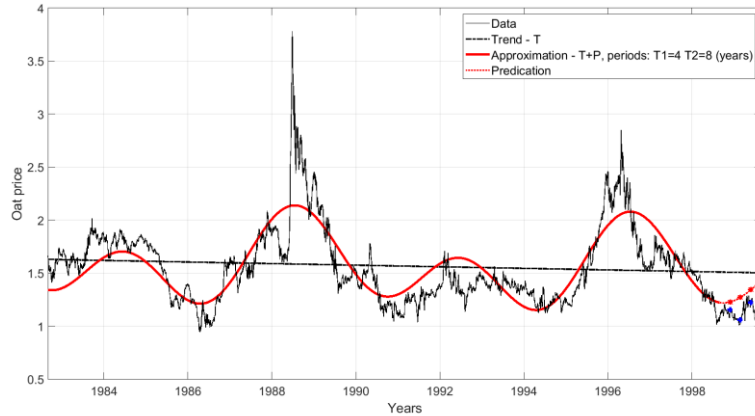


Figure 5 Oat: fitting of time series with prediction, window 16 years.

Following graphs show time series of wheat and oat, sliding window in the graph on the left side is 4 years and sliding window in the graph on the right side is 16 years. It can be seen from the boxplot that the median fluctuates around zero. Predictions for a window length of four years have a larger quartile range for both commodities. For a window of sixteen years, the quartile range did not increase even for the predictions for May and August. We will be focus on oats and wheats. Comparison of average values of absolute values of deviations (in percentages) of prices from those predicted in individual columns. When using a sliding window of 4-year length and sliding window of 16-year length, results are in the second row of Table 1. When using the 4-year length sliding windows, worse predictions and larger deviations for May and August of predicted values from the true values can be seen in the third and fourth column of Table 1. Results in Tab. 1 shows that 16-year predictions for November, February and May have better WR values than 4-year predictions. But WR for the November prediction was higher for a window length of 4 years. Unfortunately, the average WR values were close to .5, so the method did not provide successful predictions.

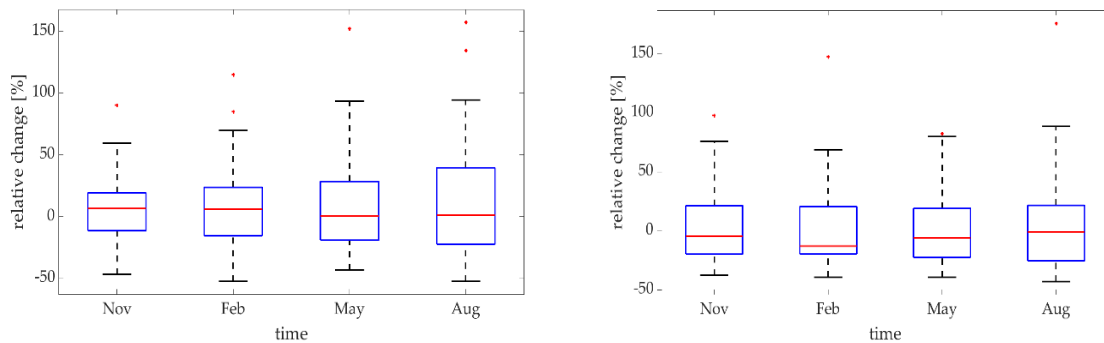


Figure 6 Wheat: Boxplots showing the differences between the actual price and the predicted price, sliding windows 4 years and 16 years

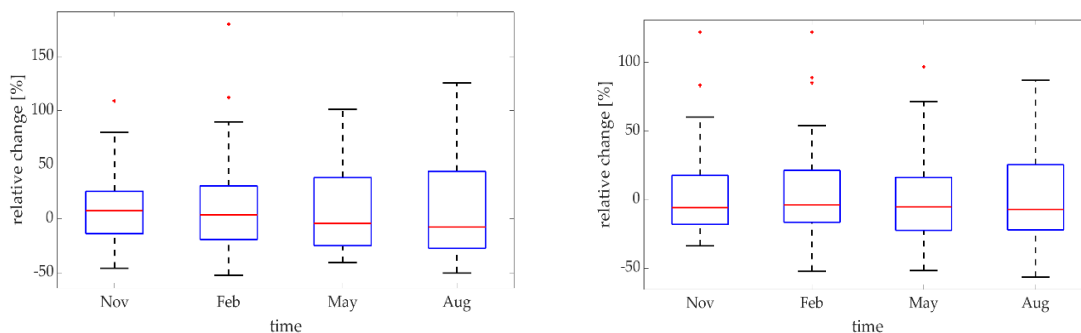


Figure 7 Oat: Boxplots showing the differences between the actual price and the predicted price, sliding windows 4 years and 16 years

Date	November	February	May	August
Wheat 4 years	20,32 % WR=0,49	25,30 % WR=0,38	28,70 % WR=0,44	36,44 % WR=0,40
Wheat 16 years	23,36 % WR=0,36	25,94 % WR=0,42	24,07 % WR=0,58	29,53 % WR=0,55
Oat 4 years	26,69 % WR=0,47	31,04 % WR=0,44	32,56 % WR=0,29	39,18 % WR=0,40
Oat 16 years	25,02 % WR=0,61	27,40 % WR=0,61	24,73 % WR=0,58	26,94 % WR=0,67

Table 1 Average MAPE (Mean Absolute Prediction Error) at the end of November, February, May and August; and WR characteristics

4 Conclusions

The numerical study shows that the algorithm is applicable for a rough prediction of agricultural crop prices. Surely the final decision of the farm manager will never be based solely on this statistical approach. His decision may be changed by information on the fulfillment of demand in the last year. Another important fact for decision making is software which calculates areas for appropriate commodities in order to optimize agricultural resources. However, the method described can help growers improve their economic situation in today's competitive environment. A similar study was conducted by the authors of the article [4], who found differences between the 6-month prediction and the actual price of around 50%. Our results were slightly better, see the results in Table 2.

Acknowledgements

The paper was supported by institutional support of University Pardubice.

References

- [1] Ashwini Darekar, Amarendra Reddy, Amarendra Reddy: Forecasting Wheat Prices in India. *Wheat and Barley Research*. 2018. 10(1): 33-39.
- [2] Chanchal Pramanik, and Srinivasu Pappula: Mathematical Modelling for Crop Pricing based on Market Value of its Products. *Acta Scientifica Agriculture*. 2018; 2(11): 101-105.
- [3] Cipra, Tomáš. Practical guide to financial and insurance mathematics. Ekopres, 2020. ISBN 9788087865675
- [4] Davenport, F., Funk, C. Using time series structural characteristics to analyze grain prices in food insecure countries. *Food Sec.* 7, 1055–1070 (2015). <https://doi.org/10.1007/s12571-015-0490-5>
- [5] Dias, J., Rocha, H. Forecasting Wheat Prices Based on Past Behavior: Comparison of Different Modelling Approaches. *MISRA*, Sanjay,
- [6] Fisher, R. A. Tests of significance in harmonic analysis. *Proc. Roy. Statist. Soc., Ser. A*, 125 (1929), pp. 5459.
- [7] Gervasi, O., Murgante, B. et al., ed. Computational Science and Its Applications – ICCSA 2019 [online]. *Lecture Notes in Computer Science*, Springer International Publishing, 2019, pp. 167-182 [cit. 2021-5-31].
- [8] Ramirez, O.A., Fadiga, M.: Forecasting agricultural commodity prices with asymmetric-error GARCH models. *Journal of Agricultural and Resource Economics*. 71-85 (2003)
- [9] Schuster, A. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism*, 3, 1898, pp. 13-41.
- [10] Xuli, Han: The Trigonometric Polynomial Like Bernstein Polynomial. *Scientific World Journal*, Volume 2014, Article ID 174716, 17 pages. <http://dx.doi.org/10.1155/2014/174716>
- [11] Yang, J., Zhang, J., Leatham, D.J.: Price and volatility transmission in international wheat futures markets. *Annals of Economics and Finance*. 4, 37-50 (2003)

Discrete Time Optimal Control Problems with Infinite Horizon

Pavel Pražák¹, Kateřina Frončková²

Abstract. Optimal control problems are often used in the formulation of many dynamic economic models. A certain specificity of such economic models is that the infinite time horizon is used. This abstract concept is mainly used to represent a very long planning period. Nevertheless, the assumption of the infinite horizon in optimal control problems leads to some complications in their solution. Apart from modifying the definition of the optimal solution, it is necessary to consider the so-called transversality condition. This paper deals with the formulation of the necessary transversality condition for an optimal solution to quite a general infinite horizon discrete time problem. To better characterize the abstract concepts, two solved and illustrative problems are introduced as well.

Keywords: optimal control, dynamic economic models, difference equation, discrete time problems, infinite horizon problems, transversality condition

JEL Classification: C61, O20, O40

AMS Classification: 49K15

1 Introduction

Many economic models are formulated as finite time optimal control problems. Nevertheless, there exist other models, for which it is problematic to find a reason why to use a finite time horizon, [3], [13], [9]. For example, it is difficult to justify that a firm has a finite planning horizon, [2]. This assumption would mean that the management of the firm considers that the firm will exist only within the finite time interval. In this case, a question arises why to ignore profits earned after the given time period. A similar question has to be answered when we consider a finite planning horizon for the economy's optimal path for consumption and capital accumulation. In this case, it is necessary to decide what level of the capital stock is to be left at the end. This information is particularly important for all generations living after the planning period. It is also needed to justify why we ignore the benefits gained by generations alive beyond the planning horizon. Although we know that the infinite time horizon is unrealistic and we may agree with [7] that "In the long run we are all dead," the infinite time horizon can be used to model a very long time period conveniently. In [2], it is noted: "The infinite horizon is an idealization of the fundamental point that the consequences of investment are very-long lived; any short horizon requires some methods of evaluating end-of-period capital stocks, and the only proper evaluation is their value in use in the subsequent future." Typical infinite horizon problems include almost all growth models, e.g. [1], [12], [14], [11] that affirm that the infinite horizon control optimal models are frequently used in economic modelling. Therefore it is worthwhile to know some mathematical issues concerning this problem.

2 Infinite Horizon Control Problems

The dynamic optimization problems that can be found in economic models are usually described by two different classes of variables:

- *State variables* (state vector) $x(t)$ describe state of the economic system in period t . It is considered that $t \in \mathbb{N}_0$ and $x(t) \in X$, where X is an arbitrary nonempty set called the *state space*. Moreover the *initial state* $x(0) \in X$ is usually known and therefore given.
- *Control variables* (control vector) $c(t)$ in period t allow to control the subsequent future state $x(t+1)$ of the economic system. It is considered that there is a non-empty set $\tilde{G}(t, x)$ of all *feasible values* of controls $c(t)$ when the system in period t is $x \in X$.

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, kateřina.fronckova@uhk.cz

It is assumed that the dynamics of the economic system, that is, the evolution of the state $x(t)$ under a given control $c(t)$ in period $t \in \mathbb{N}_0$, is determined by a difference equation

$$x(t+1) = g(t, x(t), c(t)), \quad x(0) = x_0 \quad (1)$$

where $g : \tilde{\Omega} \rightarrow X$ is a map with the domain $\tilde{\Omega} = \{(t, x, c) \mid t \in \mathbb{N}_0, x \in X, c \in \tilde{G}(t, x)\}$. In a dynamical system with control, it is necessary to consider both the sequence of states and sequence of controls together - the term *admissible pair* from the given state $x_0 \in X$ means a state trajectory $(x(t))_{t=0}^{\infty}$ and the control sequence $(c(t))_{t=0}^{\infty}$ if $x(t) \in X$, $c(t) \in \tilde{G}(t, x(t))$ and equation (1) hold for all $t \in \mathbb{N}_0$. Different admissible pairs have different utility and it is useful to be able to compare them and eventually, to find the best one. For this purpose, we suppose that there is an instantaneous *payoff function* $\tilde{F} : \tilde{\Omega} \rightarrow \mathbb{R}$ and consider the *objective function*

$$\sum_{t=0}^{\infty} \tilde{F}(t, x(t), c(t)). \quad (2)$$

Sometimes, it is useful to eliminate control variables. For this purpose, we can consider the following set of all states that can be reached in period $t+1$ if the state in period t is x and if all values of feasible controls are involved

$$G(t, x) = \{g(t, x, c) \mid c \in \tilde{G}(t, x)\}.$$

With this set, the equation (1) can be rewritten as follows

$$x(t+1) \in G(t, x(t)) \quad (3)$$

The state trajectory $(x(t))_{t=0}^{\infty}$ that satisfies (3) and $x(0) = x_0$ is called *admissible path*. It is also necessary to modify the definition of the payoff function. Instead of $\tilde{\Omega}$ consider the set $\Omega = \{(t, x, y) \mid t \in \mathbb{N}_0, x \in X, y \in G(t, x)\}$ as a domain of utility function $F : \Omega \rightarrow \mathbb{R}$ that can be defined by the rule

$$F(t, x, y) = \max\{\tilde{F}(t, x, c) \mid y = g(t, x, c), c \in \tilde{G}(t, x)\}.$$

It means that F is the maximal value of \tilde{F} (provided it exists) if the state in period t is x , the state in period $t+1$ is y and all values of feasible controls c are considered. Now, it is possible to consider the following formulation of the basic dynamic programming problem without an explicit control variable, e.g. [1], [12],

$$\max \sum_{t=0}^{\infty} F(t, x(t), x(t+1))$$

subject to

$$\begin{aligned} x(t+1) &\in G(t, x(t)), \quad t \in \mathbb{N}_0 \\ x(0) &= x_0 \in X. \end{aligned}$$

Moreover, it is assumed that there exists a real value of the sum in objective function for all admissible paths $(x(t))_{t=0}^{\infty}$ or more precisely

$$\lim_{T \rightarrow \infty} \sum_{t=0}^T F(t, x(t), x(t+1)) \in \mathbb{R}. \quad (4)$$

This latter assumption can enable a better formulation of the optimal admissible path. The admissible path $(\hat{x}(t))_{t=0}^{\infty}$ is an optimal if

$$\lim_{T \rightarrow \infty} \sum_{t=0}^T F(t, x(t), x(t+1)) \leq \lim_{T \rightarrow \infty} \sum_{t=0}^T F(t, \hat{x}(t), \hat{x}(t+1))$$

for all admissible paths $(x(t))_{t=0}^{\infty}$. It can be proved, see e.g. [1], [12], that provided

- $X \subset \mathbb{R}^n$ has a non-empty interior,
- $G(t, x) \subset \mathbb{R}^n$ has a non-empty interior for all $x \in X$ and all $t \in \mathbb{N}_0$,
- utility function $F : \Omega \rightarrow \mathbb{R}$ is continuously differentiable with respect to its last two arguments on the interior of $\Omega_t = \{(x, y) \mid x \in X, y \in G(t, x)\} \subset X \times X$
- limit (4) exists for all admissible paths
- $(\hat{x}(t))_{t=0}^{\infty}$ is an interior optimal path

the equation called *Euler equation*

$$F_3'(t, \hat{x}(t), \hat{x}(t+1)) + F_2'(t+1, \hat{x}(t+1), \hat{x}(t+2)) = 0 \quad (5)$$

holds for all $t \in \mathbb{N}_0$. In (5) F_2' and F_3' represent derivative of the function F with respect to its second argument and with respect to its third argument respectively.

3 Transversality Condition

Equation (5) represents a difference equation of the second order with initial state $x_0 \in X \subset \mathbb{R}^n$. It is known, e.g. [4], that the knowledge of the initial condition with n coordinates is not sufficient to find a particular solution to (5). To find the unique solution to this equation, it is necessary to find another condition with n coordinates. Such an additional boundary condition is called *transversality condition*. To formulate this condition one needs to assume more requirements to the given infinite horizon problem. Let us suppose that

- $X \subset \mathbb{R}_+^n$ is a convex and has a non-empty interior,
- $G(t, x) \subset \mathbb{R}^n$ has a non-empty interior for all $x \in X$ and all $t \in \mathbb{N}_0$,
- limit (4) exists for all admissible paths
- utility function $F : \Omega \rightarrow R$ is continuously differentiable with respect to its last two arguments on the interior of $\Omega_t = \{(x, y) \mid x \in X, y \in G(t, x)\} \subset X \times X$
- the set $\Omega_t \subset \mathbb{R}^n \times \mathbb{R}^n$ is a convex and $(0, 0) \in \Omega_t$
- utility function U is concave with respect to $(x, y) \in \Omega_t$ for all $t \in \mathbb{N}_0$
- $(\hat{x}(t))_{t=0}^\infty$ is an interior optimal path such that $F'_3(t, \hat{x}(t), \hat{x}(t+1)) \leq 0$

then Euler equation (5) holds and

$$\lim_{t \rightarrow \infty} F'_3(t, \hat{x}(t), \hat{x}(t+1)) \cdot \hat{x}(t+1) = 0. \quad (6)$$

The proof of this statement is given in [6].

3.1 Halkin's Example

To illustrate some aspect of transversality condition, we introduce Halkin's example, [5] or [3]. This problem is usually formulated as an optimal control problem in continuous time, here we introduce a discrete time version of this problem, eg. [8],

$$\max \sum_{t=0}^{\infty} (1-x(t))u(t)$$

subject to

$$x(t+1) = (1-x(t))u(t), \quad x(0) = 0$$

and $u(t) \in [0, 1]$ for all $t \in \mathbb{N}_0$. Any feasible path satisfies

$$x(t+1) - x(t) = (1-x(t))u(t), \quad t \in \mathbb{N}_0,$$

which means that the objective function can be written as

$$\sum_{t=0}^{\infty} (1-x(t))u(t) = \sum_{t=0}^{\infty} x(t+1) - x(t) = \lim_{T \rightarrow \infty} \sum_{t=0}^T x(t+1) - x(t) = \lim_{T \rightarrow \infty} x(0) + x(T+1) = \lim_{T \rightarrow \infty} x(T+1),$$

because $x(0) = 0$. For the feasible path, the following relation also holds

$$1 - x(t+1) \leq (1-x(t))(1-u(t)) \leq 1-x(t), \quad t \in \mathbb{N}_0.$$

Thus, it means that $x(t) \leq x(t+1)$ and $x(t+1) \leq 1$. From these relations we can see that $x(t+1) \in [x(t), 1]$ and therefore $G(t, x) = [x, 1]$. To summarize, the original problem can be also written as

$$\max \sum_{t=0}^{\infty} x(t+1) - x(t)$$

subject to

$$x(t+1) \in [x(t), 1] \text{ and } x(0) = 0.$$

The above given relations also imply that the path $(x(t))_{t=1}^\infty$ is a non-decreasing sequence with upper bound 1. Thus, for all admissible paths $(x(t))_{t=1}^\infty$ there exist $\lim_{t \rightarrow \infty} x(t) \in [0, 1]$. Consequently, for optimal path $(\hat{x}(t))_{t=0}^\infty$, it follows that

$$\max \sum_{t=0}^{\infty} x(t+1) - x(t) = \sum_{t=0}^{\infty} \hat{x}(t+1) - \hat{x}(t) = \lim_{T \rightarrow \infty} \hat{x}(T+1) = 1.$$

Thus, there is no unique optimal path and any non-decreasing paths $(x(t))_{t=0}^\infty$ such that $\lim_{t \rightarrow \infty} x(t) = 1$ is optimal. For instance, we can consider the sequence given by the rule

$$\hat{x}(t) = 1 - (1/2)^t, \quad t \in \mathbb{N}_0.$$

Let us notice that the utility function of this problem is $F(t, x(t), x(t+1)) = x(t+1) - x(t)$ or $F(t, x, y) = y - x$. It means that $F_3'(t, x, y) = 1$ and therefore for any optimal path is

$$\lim_{t \rightarrow \infty} F_3'(t, \hat{x}(t), \hat{x}(t+1)) \cdot \hat{x}(t+1) = 1 \cdot 1 = 1 \neq 0,$$

which is contrary to the transversality condition (6). Now, it is desirable to look for an assumption of transversality condition that is not met in this example. For instance, it is immediate to find that one of them is $F_3'(t, x, y) > 0$.

3.2 Intertemporal Utility Maximization Problem of a Consumer

The following example is a prototype of the intertemporal utility maximization problem of a consumer, more details can be found in [13]. It is presented here to show the usefulness of the introduced transversality condition (6). Let us consider the problem of an infinitely-lived consumer with instantaneous constant relative risk aversion utility function $u(c) = (c^\gamma - 1)/\gamma$, $\gamma \in (0, 1)$ defined over the consumption c , $c \geq 0$. The consumer considers his future preference over consumption exponentially discounted with the constant factor β , $\beta \in (0, 1)$. He starts his life with a given amount of wealth x_0 , $x_0 \in \mathbb{R}_+$, and receives a constant rate of returns R , $R > 1$, on the difference of his wealth x and his consumption c in the given period. Then the total utility maximization problem of the consumer can be written as

$$\max \sum_{t=0}^{\infty} \beta^t \frac{c(t)^\gamma - 1}{\gamma}$$

subject to

$$x(t+1) = R \cdot (x(t) - c(t)), \quad x(0) = x_0, \quad (7)$$

where $t \in \mathbb{N}_0$. Based on the properties of parameters the maximization problem can be partly simplified as follows

$$\max \sum_{t=0}^{\infty} \beta^t \frac{c(t)^\gamma - 1}{\gamma} = \frac{1}{\gamma} \left[\max \sum_{t=0}^{\infty} \beta^t c(t)^\gamma - \frac{1}{1-\beta} \right].$$

We further consider that debts are not allowed, which means that $c(t) \leq x(t)$. Because $c(t) \geq 0$ for all $t \in \mathbb{N}_0$ it is possible to find that $0 \leq x(t+1) \leq R \cdot x(t)$. Thus, we can write $x(t+1) \in [0, R \cdot x(t)]$. From (7) it is immediately clear that

$$c(t) = x(t) - \frac{1}{R} \cdot x(t+1), \quad t \in \mathbb{N}_0. \quad (8)$$

Now, it is possible to write the original problem in the variational form

$$\max \sum_{t=0}^{\infty} \beta^t \left[x(t) - \frac{1}{R} \cdot x(t+1) \right]^\gamma$$

subject to

$$x(t+1) \in [0, R \cdot x(t)], \quad x(0) = a_0,$$

where $t \in \mathbb{N}_0$. For further analysis, we will add a technical assumption $\beta R^\gamma < 1$. Together with the above given restriction, it means that $R \in (1, \beta^{-1/\gamma})$. To find a candidate for the optimal path Euler equation (5) and transversality condition (6) can be applied. First, it is necessary to verify the assumptions

- $x(t) \in X = [0, \infty)$ has a non-empty interior.
- $G(t, x) = [0, Rx]$ has a nonempty interior for $x > 0$.
- To verify that the objective function is defined well for all admissible paths, we notice that $x(t+1) \leq R \cdot x(t)$ for all $t \in \mathbb{N}_0$. Thus, $x(t) \leq x_0 R^t$. Because $c(t) \geq 0$ and $c(t) \leq x(t)$ we have $0 \leq c(t) \leq x_0 R^t$. If we use these properties in the objective function, we successively gain

$$0 \leq \sum_{t=0}^{\infty} \beta^t c(t)^\gamma \leq \sum_{t=0}^{\infty} \beta^t \cdot (x_0 R^t)^\gamma = x_0^\gamma \sum_{t=0}^{\infty} (\beta \cdot R^\gamma)^t$$

If the assumption $\beta \cdot R^\gamma < 1$ is used, the latter geometric series is convergent. Applying the comparison test for series, it can be directly seen that the objective function is also convergent for all admissible paths $(x(t))_{t=0}^{\infty}$.

- Utility function $F(t, x, y) = \beta^t (x - 1/R \cdot y)^\gamma$ is continuously differentiable with respect to its last two arguments on the interior of $\Omega_t = \{(x, y) \mid x \in [0, \infty), y \in [0, Rx]\}$
- The set Ω_t is a convex and $(0, 0) \in \Omega_t$
- Utility function F is concave with respect to $(x, y) \in \Omega_t$ for all $t \in \mathbb{N}_0$

- Because $F'_3(t, x, y) = \beta^t \gamma [x - 1/R \cdot y]^{\gamma-1} \cdot (-1/R)$ and $x(t) - 1/R \cdot x(t+1) \geq 0$ it holds

$$F'_3(t, x(t), x(t+1)) = \beta^t \gamma \left[x(t) - \frac{1}{R} \cdot x(t+1) \right]^{\gamma-1} \cdot \left(-\frac{1}{R} \right) \leq 0.$$

Euler equation (5) for function F can be applied now. An optimal interior path $(\hat{x}(t))_{t=0}^{\infty}$ satisfies the following difference equation

$$\beta^t \gamma \left[\hat{x}(t) - \frac{1}{R} \hat{x}(t+1) \right]^{\gamma-1} \cdot \left(-\frac{1}{R} \right) + \beta^{t+1} \gamma \left[\hat{x}(t+1) - \frac{1}{R} \hat{x}(t+2) \right]^{\gamma-1} = 0.$$

Because the relation (8) holds, the latter equation can be rewritten for optimal consumption path

$$\hat{c}(t+1) = (\beta R)^{\frac{1}{1-\gamma}} \hat{c}(t).$$

The solution to this recurrent equation that represents a geometric sequence is

$$\hat{c}(t) = c_0 \cdot (\beta R)^{\frac{t}{1-\gamma}}, \quad t \in \mathbb{N}_0, \quad (9)$$

where c_0 is the initial consumption that has to be determined. For this purpose, the transversality condition (6) can be used. First, we find the solution to the stated equation (7). It is possible to find, see [10], [4], that

$$\hat{x}(t) = x_0 R^t + \sum_{j=0}^{t-1} R^{t-1-j} \cdot (-R c(j)) = x_0 R^t - c_0 R^t \sum_{j=0}^{t-1} \left((\beta R^\gamma)^{\frac{1}{1-\gamma}} \right)^j = x_0 R^t - c_0 R^t \frac{(\beta R^\gamma)^{\frac{t}{1-\gamma}} - 1}{(\beta R^\gamma)^{\frac{1}{1-\gamma}} - 1}. \quad (10)$$

To use the transversality condition (6), it is necessary to find the value of the limit given in this condition. If (8), (9), (10) and $\beta R^\gamma < 1$ are applied, it can be shown that

$$\begin{aligned} \lim_{t \rightarrow \infty} \beta^t \gamma \left[\hat{x}(t) - \frac{1}{R} \cdot \hat{x}(t+1) \right]^{\gamma-1} \cdot \left(-\frac{1}{R} \right) \cdot \hat{x}(t+1) &= -\gamma \lim_{t \rightarrow \infty} \beta^t (c(t))^{\gamma-1} \left(\frac{1}{R} \right) \cdot \hat{x}(t+1) \\ &= \left(-\gamma c_0^{\gamma-1} \right) \left(x_0 + \frac{c_0}{(\beta R^\gamma)^{\frac{1}{1-\gamma}} - 1} \right). \end{aligned}$$

In the transversality condition the value of this limit is zero. Because $\gamma \neq 0$ and $c_0 \neq 0$ (otherwise $c(t) = 0$ for all $t \in \mathbb{N}_0$ and the consumer would be starving forever, cf. (9)), it is necessary that the following equation holds

$$x_0 + \frac{c_0}{(\beta R^\gamma)^{\frac{1}{1-\gamma}} - 1} = 0.$$

If this equation is solved for c_0 , we finally gain $c_0 = x_0 \left(1 - (\beta R^\gamma)^{\frac{1}{1-\gamma}} \right)$. If $x_0 > 0$ then $c_0 > 0$ as well. This value can be substituted into (9) and (10) respectively. Then the optimal consumption and interior optimal path of wealth are given by the following geometric sequences

$$\hat{c}(t) = x_0 \left(1 - (\beta R^\gamma)^{\frac{1}{1-\gamma}} \right) \cdot (\beta R)^{\frac{t}{1-\gamma}}, \quad \hat{x}(t) = x_0 (\beta R)^{\frac{t}{1-\gamma}}, \quad t \in \mathbb{N}_0.$$

These rules generate either increasing sequences, if $\beta R > 1$, or decreasing sequences, if $\beta R < 1$. If the value of discount factor β is low, which means that we prefer today's consumption before tomorrow's consumption, and the rate of return R is not very high, the path of optimal consumption and the path of optimal wealth are decreasing. On the other hand, if the value of discount factor β is close to one, which means that we are more indifferent to today's and tomorrow's consumption, and the rate of return is high enough, the path of optimal consumption and the path of optimal wealth are increasing.

4 Conclusion

The necessary condition of transversality for the discrete time version of optimal control over an infinite time horizon was formulated in the paper. It has been shown that this condition enables to specify candidates for an optimal solution of the given problem. However, the validity of this condition is conditioned by the fulfillment of a number

of assumptions. If these assumptions are not met, it is not possible to rely on the introduced transversality condition. This complication was shown in a discrete version of Halkin's example. If the transversality condition does not hold, it is not possible to find any candidates for the optimal solution that is searched. As has been described in the paper, there are quite a lot of assumptions for the validity of transversality condition and moreover, their verification can be arduous and demanding. If, for example, in the problem of optimal consumption, the utility function $u(c) = (c^\gamma - 1)/\gamma$, $\gamma \in (0, 1)$, is replaced by the function $u(c) = \ln c$, it would be difficult to show the convergence of the objective function for all admissible paths. This is why these assumptions are usually not verified in many publications, and this preliminary phase of applying the transversality condition is not accomplished. However, such an inconsistent approach can lead to the wrong selection of the candidates for the optimal solution and cannot be recommended. On the contrary, it is desirable to look for weaker or more general assumptions for the validity of the transversality condition. Another possibility is to reconsider the use of an infinite time horizon in the economic application of discrete time optimal control problems. Both mentioned options involve a number of issues that need to be solved.

Acknowledgements

Support of the Specific Research Project of the Faculty of Informatics and Management of University of Hradec Králové in 2021 is kindly acknowledged.

References

- [1] Acemoglu D. (2009). *Introduction to Modern Economic Growth*, Princeton, Princeton University Press.
- [2] Arrow K. J. and Kurz M. (1970). *Public investment, the rate of return, and optimal fiscal policy*. Baltimore, The Johns Hopkins University Press.
- [3] Carlson D. A. and Haurie A. (1987). *Infinite Horizon Optimal Control, Theory and Applications*, Berlin, Springer-Verlag.
- [4] Elaydi S. (2005). *An Introduction to Difference Equations* (3rd ed.), New-York, Springer-Verlag.
- [5] Halkin H. (1974). Necessary conditions for optimal control problems with infinite horizons. *Econometrica*, 42 (2), 267-272.
- [6] Kamihigashi T. (2002). A simple proof of the necessity of the transversality condition. *Economic Theory*, 20, 427-433.
- [7] Keynes J.M. (1923). *A Tract on Monetary Reform*, Macmillan & Co., London.
- [8] Michel P. (1990). Some clarifications on the transversality condition. *Econometrica* 58, 705-723.
- [9] Pražák P. (2011). On Necessary Transversality Condition for Infinite Horizon Optimal Control Problems, In M. Dlouhý, V. Skočdoplová (Eds.), *Proceedings of 29th International Conference on Mathematical Methods in Economics 2011 - part II* (pp. 575-580). Professional Publishing.
- [10] Pražák P. (2013). *Difference Equations with Economic Applications*, Gaudeamus, Hradec Kralove. (in Czech)
- [11] Pražák P. (2016). Dynamics Model of Firms' Income Tax, In A. Kocourek, M. Vavroušek (Eds.), *Mathematical Methods in Economics 2016* (pp. 705-710). Technical University of Liberec.
- [12] Stokey N. and Lucas R.E., Jr. (1989). *Recursive methods in economic dynamics*. Cambridge, MA: Harvard University Press.
- [13] Sydsæter K. et al. (2008). *Further Mathematics for Economic Analysis* (2nd ed.), Harlow, Prentice Hall.
- [14] Weitzman M. L. (1973). Duality theory for infinite horizon convex models. *Management Science* 19, 783 - 789.

Bankruptcy Problem Under Uncertainty of Claims and Estate

Jaroslav Ramík¹, Milan Vlach²

Abstract. In this paper we focus on real situations where certain perfectly divisible estate has to be divided among claimants who can merely indicate the range of their claims, and the available amount is smaller than the aggregated claim. Funds' allocation of a firm among its divisions, taxation problems, priority problems, distribution of costs of a joint project among the agents involved, various disputes including those generated by inheritance, or by cooperation in joint projects based on restricted willingness to pay, fit into this framework. The corresponding claim of each claimant can vary within a closed interval. For claims and/or estate, intervals are applied whenever the claimants can distinguish the possibility of attaining the amount of estate. Here, a probability interpretation can also be considered, e.g. in taxation problems. We classify the division problems under uncertainty of claims and/or estate and present basic division scheme, which is consistent with the classical bankruptcy proportional rule. Several examples are presented to illustrating particular problems and solution concepts.

Keywords: bankruptcy problem, bankruptcy rule, interval claims and estate

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

The classical bankruptcy problem is described as follows: Several individuals hold claims on a finite resource and the available total amount (total estate) is not enough to fulfill all of the claims. Sometimes, the problem is considered under interval uncertainty of claims and/or the total amount of resource. In some more realistic situations claimants are also likely to declare their claims with vague words just like "about", "around" and so on. The key issue is how to distribute the estate to the claimants. Different from the classical bankruptcy problems which require an exact knowledge of each term, Yager et. al. [6], investigated the uncertainty in the division problems, in which the possible fair proportion of the total estate assigned to each claimant is an interval, eventually, fuzzy interval. Branzei et al. in [2], concentrated on the bankruptcy problems under interval uncertainty of claims. They deal with the uncertainty of claims by compromising the lower and upper bounds of the claim intervals and then considered the deterministic division problems with compromise claims.

This paper is motivated by a generalization of the classical bankruptcy problem, that we also call division problems under uncertainty of claims and estate. In comparison with the former works on this subject, see [2], where only the interval claims have been investigated, here, we consider the interval claims, but also an uncertain (i.e. interval) estate to be divided. Applying interval calculus, we introduce the corresponding division schemes in the spirit of the classical bankruptcy rules.

It is important to consider bankruptcy problems under uncertainty, because in various disputes including inheritance, claimants face uncertainty with regard to their effective rights and as a result, individual claims can be expressed in the form of intervals. In such situations, our model offers flexibility to tackle with resource conflict under uncertainty. In order to get some insight into applications, interpretations and extensions for division problems under uncertainty, the readers can refer to a wide range of papers such as cooperative interval games Branzei et al. [2], and stochastic bankruptcy games Habis et al. [4].

The rest of this paper is organized as follows. In Section 2, we briefly review the classical bankruptcy problems. In Section 3, we propose the bankruptcy rule under interval uncertainty including an illustrating example. In Section 4, some conclusions and future research are mentioned.

¹ Silesian University in Opava, School of Business and Administration in Karvina, Czechia, ramik@opf.slu.cz

² Charles University, Faculty of Mathematics and Physics, Praha, Czechia, milan.vlach@gmail.com

2 Classical bankruptcy problems and games

Bankruptcy problems originate from the situations where several agents claim portions of certain amount of estate and the sum of claims is greater than or equal to the total estate. The estate consists of a given amount of a single (perfectly divisible) good. The bankruptcy problem is then an allocation problem of how to fairly divide the estate among the agents taking into account their claims.

Basic desirable properties of an allocation are that all agents earn a positive part or nothing of the estate and no agent gets more than his claim. An allocation rule that satisfies these two basic properties is called a bankruptcy rule. In this Section, we deal with the so called classical bankruptcy problem, which is defined as follows.

Definition 1. Let $N = \{1, 2, \dots, n\}$, $n \geq 2$, be a set of agents (claimants), $c = (c_1, c_2, \dots, c_n)$ be a nonnegative vector of individual claims, and E be a nonnegative number representing the total estate. A vector $(c_1, \dots, c_n, E) \in (\mathbf{R}_+)^{n+1}$ is called an *instance of the classical bankruptcy problem of n claimants* if the sum of claims is not less than the estate; that is, if

$$\sum_{j \in N} c_j \geq E. \quad (1)$$

Each instance of the classical bankruptcy problem of n claimants (CB-instance) is shortly denoted by (c, E) , and the set of all these instances (that is, the classical bankruptcy problem of n claimants, or shortly CB-problem) is denoted by $\mathcal{B}_{cla}(N)$, or shortly \mathcal{B}_{cla} .

Definition 2. A vector mapping $\mathbf{s} : \mathcal{B}_{cla} \rightarrow (\mathbf{R}_+)^n$, where

$$\mathbf{s}(c, E) = (s_1(c, E), \dots, s_n(c, E)) \text{ for each } (c, E) \in \mathcal{B}_{cla},$$

is called a *bankruptcy rule for \mathcal{B}_{cla}* and the values $s_i(c, E)$, $i \in N$, are called the *claimant shares*, if the following conditions are satisfied for all $i \in N$:

$$s_i(c, E) \leq c_i, \quad (\text{Individual rationality condition}) \quad (2)$$

$$\sum_{j \in N} s_j(c, E) = E. \quad (\text{Efficiency condition}) \quad (3)$$

Proposition 1. The vector mapping $\mathbf{s} : \mathcal{B}_{cla} \rightarrow (\mathbf{R}_+)^n$ whose values, $\mathbf{s}(c, E) = (s_1(c, E), \dots, s_n(c, E))$, are defined for all $i \in N$ by

$$s_i(c, E) = \frac{c_i E}{\sum_{j \in N} c_j} \quad (4)$$

is a bankruptcy rule for CB-problem that satisfies the following two properties:

- (i) If, for some $i, j \in N$, $c_i = c_j$, then $s_i(c, E) = s_j(c, E)$.
- (ii) If $E \leq E'$, $E, E' \in \mathbf{R}_+$, then $s_i(c, E) \leq s_i(c, E')$.

Property (i) is a natural property saying that the claimants with the same claims obtain the same shares of the estate. Property (ii) is a monotonicity of the proportional rule with respect to estate E .

The rule defined in the Proposition (1) is called the *proportional rule (P-rule)*. We now briefly introduce a modification of the propositional rule that is called the *adjusted proportional rule (AP-rule)*.

Let (c, E) be an instance of the CB-problem and let define

$$m(c, E) = (m_1(c, E), m_2(c, E), \dots, m_n(c, E))$$

by

$$m_i(c, E) = \max\{0, E - \sum_{j \in N \setminus \{i\}} c_j\}, \text{ for each } i \in N. \quad (5)$$

Here, $m_i(c, E)$ can be interpreted as the *claim right of i -th claimant*.

Moreover, let

$$E' = E - \sum_{i \in N} m_i(c, E) \text{ and } c' = (c'_1, \dots, c'_n), \text{ where } c'_i = \min\{c_i - m_i(c, E), E'\}, i \in N.$$

The adjusted proportional rule is defined as the vector mapping $\mathbf{P} : \mathcal{B}_{cla} \rightarrow (\mathbf{R}_+)^n$ defined by

$$\mathbf{P}(c, E) = m(c, E) + \mathbf{s}(c', E') \quad (6)$$

where by $\mathbf{s}(c', E')$ we mean the result of application of P-rule (4) to (c', E') .

Example 1. Let us consider 2-claimant problem given by instances (c_1, c_2, E) . Let \mathbf{s} be the mapping that assigns to every instance (c_1, c_2, E) a vector in \mathbf{R}_+^2 according to the following rule:

If $E \leq c_1, c_2$, then $\mathbf{s}(c_1, c_2, E) = (E/2, E/2)$.

If $c_1 \leq E \leq c_2$, then $\mathbf{s}(c_1, c_2, E) = (c_1/2, E - c_1/2)$.

If $c_2 \leq E \leq c_1$, then $\mathbf{s}(c_1, c_2, E) = (c_2/2, E - c_2/2)$.

If $c_1, c_2 \leq E$ then $\mathbf{s}(c_1, c_2, E) = ((E + c_1 - c_2)/2, (E + c_2 - c_1)/2)$.

The reader can easily verify that this mapping is a bankruptcy rule for the problem under consideration. This rule is known under the name the *contested garment rule*, or, *Talmud rule*, see [1].

The bankruptcy rule in this example shows that the bankruptcy problems are strongly connected to the theory of cooperative games. To see it, let us recall that a coalition game with transferable utility is an ordered pair (N, v) , where N is a set of players, and $v: 2^N \rightarrow \mathbf{R}$ is a real-valued function satisfying $v(\emptyset) = 0$. The function v is called the *coalition function of the game*, or the *characteristic function of the game*. The subsets of N are called *coalitions*.

Any instance (c, E) of bankruptcy problem generates a corresponding coalition game (N, v) , called the *bankruptcy game*, whose coalition function is given by

$$v(T) = \max\{0, E - \sum_{i \in N \setminus T} c_i\}, \text{ for any } T \subset N. \quad (7)$$

It turns out that the contested garment rule in Example 1 coincides with the solution of the corresponding coalition game that is known under the name of *nucleolus*.

Solutions of the classical bankruptcy problem are also closely related to the solutions of a variant of Nash's model for cooperative bargaining that is known as the bargaining problem with claims. For details, see [5].

In what follows, we shall deal with the proportional rule and adjusted proportional rule under uncertainty of claims and estate.

3 Interval bankruptcy problem

Sometimes, the problem is considered under interval uncertainty of claims and/or the total amount of resource - estate. We start with some basic concepts of interval arithmetic, then we introduce a bankruptcy problem under interval uncertainty of claims and estate (IB-problem).

Definition 3. Let $I(\mathbf{R})$ be the set of all closed and bounded intervals in the set of real numbers \mathbf{R} . By $I(\mathbf{R})^n$ we denote the set of all n -dimensional vectors with components in $I(\mathbf{R})$. Let $a, b \in I(\mathbf{R})$, with $a = [a^-; a^+]$, $b = [b^-; b^+]$ and $\gamma \geq 0$. Then the interval operations are defined by

$$a + b = [a^- + b^-; a^+ + b^+], \gamma a = [\gamma a^-; \gamma a^+]. \quad (8)$$

In particular, $0 = [0; 0]$. The partial ordering on $I(\mathbf{R})^n$ is defined as follows:

$$a \leq b \text{ if } a^- \leq b^- \text{ and } a^+ \leq b^+. \quad (9)$$

We denote $a = b$, if $a \leq b$ and $b \leq a$. Moreover, we denote $a < b$, if $a \leq b$ and $a \neq b$.

If $a^- = a^+$, then $a = [a^-; a^+]$ is called the *one-point interval*.

Let N be a set of claimants who face (interval) uncertainty regarding claims, and E be an interval representing the total estate that is to be divided among members of N . The claim vector is denoted by $c \in I(\mathbf{R})^n$ with $c_i = [c_i^-; c_i^+] \geq 0, i \in N$, meaning the claim interval of claimant $i \in N$. The total estate, $E = [E^-; E^+] \geq 0$, is also an interval meaning the interval of uncertainty of the known estate to be divided.

Definition 4. Let $N = \{1, 2, \dots, n\}$ be a set of claimants, $c = (c_1, \dots, c_n) \in I(\mathbf{R}_+)^n, c_i = [c_i^-; c_i^+], i \in N$, be an interval vector of claims, and $E = [E^-; E^+] \in I(\mathbf{R}_+)$, be an interval estate. The interval vector $(c_1, \dots, c_n, E) \in I(\mathbf{R}_+)^{n+1}$ is called an *instance of interval bankruptcy problem (IB-problem)*, if the sum of upper bounds of claims is not less than the upper bound of estate, and the sum of lower bounds of claims is not greater than the lower bound of estate, that is, if

$$\sum_{i \in N} c_i^+ \geq E^+ \geq E^- \geq \sum_{i \in N} c_i^-. \quad (10)$$

Each instance of the IB-problem is shortly denoted by (c, E) , and the set of all instances (that is, the interval bankruptcy problem, or shortly IB-problem) is denoted by \mathcal{B}_{int} , or $\mathcal{B}_{int}(N)$.

Definition 5. A vector mapping $\mathbf{s} : \mathcal{B}_{int} \rightarrow I(\mathbf{R}_+)^n$, where

$$\mathbf{s}(c, E) = (s_1(c, E), \dots, s_n(c, E)), \text{ for each } (c, E) \in \mathcal{B}_{int},$$

such that $c = (c_1, \dots, c_n) \in I(\mathbf{R}_+)^n, c_i = [c_i^-; c_i^+], i \in N$, and $E = [E^-; E^+] \in I(\mathbf{R}_+)$, is called a *bankruptcy rule for a IB-problem* and the components of its value are called the *claimant shares for the IB-problem* if the claimant shares for the IB-problem satisfy the following conditions for all $i \in N$:

$$s_i(c, E) = [s_i^-(c, E), s_i^+(c, E)] \subset c_i = [c_i^-, c_i^+], \text{ (Individual rationality.)} \quad (11)$$

$$\sum_{j \in N} s_j(c, E) \subset E = [E^-; E^+]. \quad \text{(Efficiency.)} \quad (12)$$

By Definition 4, the IB-problem can be denoted by a triple (N, c, E) , abbreviated to (c, E) . In particular, when $c_i^- = c_i^+$ for all $i \in N$, the claim interval c_i degenerates into a real number c_i . Moreover, when in addition the total estate $E = E^- = E^+$ also degenerates into a positive number, then we obtain the CB-problem, see e.g. [3]. In case of uncertainty, that is, in an instance of IB-problem, we interpret c_i^- as the lower bound of claim or the least demand of claimant $i \in N$. Similarly, c_i^+ is interpreted as the upper bound of claim, or, the utmost expectation of claimant $i \in N$. For any coalition $T \subset N$, we use the notation

$$c^-(T) = \sum_{i \in T} c_i^-, \quad c^+(T) = \sum_{i \in T} c_i^+. \quad (13)$$

Using (11), (12), we show that the following proposition holds.

Proposition 2. Let $(c, E) \in \mathcal{B}_{int}$ be an instance of IB-problem, $c = (c_1, \dots, c_n) \in I(\mathbf{R}_+)^n, E = [E^-; E^+] \subset \mathbf{R}_+$, and let $\mathbf{s}(c, E) = (s_1(c, E), \dots, s_n(c, E))$ be a bankruptcy rule for an IB-problem \mathcal{B}_{int} satisfying (11) and (12). Then, for each $e \in E$ and each $i \in N$, there exists $s_i(e) \in s_i(c, E)$ such that

$$\sum_{j \in N} s_j(e) = e. \quad (14)$$

and

$$c^-(N) \leq \sum_{j \in N} s_j(e) \leq c^+(N). \quad (15)$$

Evidently, the bankruptcy rule for IB-problem \mathcal{B}_{int} can prescribe for each estate $e \in E$ a specific division of e among n claimants. The following proposition states that under the assumption that the sum of least demands of all claimants is not greater than the lower limit of the estate, E^- , and, the sum of utmost expectations of all claimants is not less than the upper limit of the estate E^+ , then there exists a unique specific division of e satisfying some desirable properties. The following proposition is a parallel version of Proposition 2 for IB-problem.

Proposition 3. Let (c, E) be an instance of IB-problem, that is, $(c, E) \in \mathcal{B}_{int}$. Then $d_i(c, E) = [d_i^-; d_i^+] \in \mathbf{R}^+$ defined for $i \in N$ by

$$d_i^- = c_i^- + (c_i^+ - c_i^-) \frac{E^- - c^-(N)}{c^+(N) - c^-(N)}, \quad (16)$$

$$d_i^+ = c_i^+ - (c_i^+ - c_i^-) \frac{c^+(N) - E^+}{c^+(N) - c^-(N)}, \quad (17)$$

define a bankruptcy rule called the proportional rule (P-rule) for the IB-problem (that is, its claimant shares $s_i(c, E) = [d_i^-; d_i^+]$ satisfy individual rationality condition (11) and efficiency condition (12)) that has the following two properties:

- (i) If for some $i, j \in N, c_i = c_j$, i.e. $[c_i^-; c_i^+] = [c_j^-; c_j^+]$, then $[d_i^-; d_i^+] = [d_j^-; d_j^+]$.
- (ii) If $e \in E, i \in N$, and

$$D_i(e) = c_i^- + (c_i^+ - c_i^-) \frac{e - c^-(N)}{c^+(N) - c^-(N)}, \quad (18)$$

then for $e, e' \in E, e \leq e'$, it holds $D_i(e) \leq D_i(e')$.

Property (i) is a natural property saying that the claimants with the same claims obtain the same shares. Property (ii) is a monotonicity of the P-rule for the IB-problem with respect to estate E .

Example 2. Let (c, E) be an instance of IB-problem defined as follows.

$N = \{1, 2, 3\}$, $c = (c_1, c_2, c_3) \in I(\mathbf{R}_+)^3$, where

$c_1 = [c_1^-; c_1^+] = [10; 35]$, $c_2 = [c_2^-; c_2^+] = [25; 50]$, $c_3 = [c_3^-; c_3^+] = [30; 60]$, $E = [E^-; E^+] = [85; 115]$.

Obviously, we have $c^-(N) = 65$, $c^+(N) = 145$.

Then by formulas (16) and (17) we obtain the claimant shares by the corresponding P-rule:

$d_1(c, E) = [d_1^-; d_1^+] = [16.25; 25.63]$, the mean value is $D_1(M) = D_1(16.25, 25.63) = 20.94$.

$d_2(c, E) = [d_2^-; d_2^+] = [31.25; 40.63]$, the mean value is $D_2(M) = D_2(31.25, 40.63) = 35.94$.

$d_3(c, E) = [d_3^-; d_3^+] = [37.50; 48.75]$, the mean value is $D_3(M) = D_3(37.50, 48.75) = 43.13$.

By the above mentioned P-rule we get $[d_i^-; d_i^+]$, $i \in \{1, 2, 3\}$, being the claimant shares (i.e. "interval solution") of the given IB-problem (c, E) . Moreover, by the mean value $D(M)$ we obtain a "crisp solution" of (c, E) .

Now, we shall propose a more specific form of a bankruptcy rule based on the concepts of minimum claim right m_i^- and maximum claim right m_i^+ of each claimant i . Let (c, E) be an instance of IB-problem. Moreover, for $e \in E = [E^-; E^+]$, $i \in N$, we define the *minimal claim right* m_i^- of e and the *maximal claim right* m_i^+ of e for each claimant $i \in N$, respectively, as follows

$$m_i^-(c, e) = \max\{c_i^-, e - c^+(N \setminus \{i\})\}, m_i^+(c, e) = \min\{c_i^+, e - c^-(N \setminus \{i\})\}. \quad (19)$$

For $T \subset N$, we denote

$$m_T^-(c, e) = \sum_{i \in T} m_i^-(c, e), m_T^+(c, e) = \sum_{i \in T} m_i^+(c, e). \quad (20)$$

In the sequel, if there is no danger of misunderstanding, we leave symbol c in formulas $m_i^-(c, e)$, $m_i^+(c, e)$, etc., writing simply $m_i^-(e)$, $m_i^+(e)$, etc.

Proposition 4. Let $(c, E) \in \mathcal{B}_{int}$, $e \in E$. Then, for $i \in N$, it holds

$$c_i^- \leq m_i^-(e) \leq m_i^+(e) \leq c_i^+. \quad (21)$$

Moreover, if $E^- \leq e \leq e^* \leq E^+$, then

$$m_i^-(e) \leq m_i^+(e) \leq m_i^-(e^*) \leq m_i^+(e^*). \quad (22)$$

Proposition 5. Let (c, E) be an instance of IB-problem. Let $m_N^-(E^-) \leq E^- \leq E^+ \leq m_N^-(E^+)$, and $E = [E^-; E^+]$. Then $s_i(c, E) = [s_i^-; s_i^+] \in I(\mathbf{R}_+)$ defined for $i \in N$ by

$$s_i^- = m_i^-(E^-) + [m_i^+(E^-) - m_i^-(E^-)] \frac{E^- - m_N^-(E^-)}{m_N^+(E^-) - m_N^-(E^-)}, \quad (23)$$

$$s_i^+ = m_i^-(E^+) - [m_i^+(E^+) - m_i^-(E^+)] \frac{E^+ - m_N^-(E^+) - E^+}{m_N^+(E^+) - m_N^-(E^+)}, \quad (24)$$

is a bankruptcy rule called the adjusted proportional rule (AP-rule) for the IB-problem (that is, its claimant shares $s_i(c, E) = [s_i^-; s_i^+]$ satisfy individual rationality condition (I1), efficiency condition (I2)) that satisfies the following condition:

(i) If for some $i, j \in N$, $c_i = c_j$, i.e. $[c_i^-; c_i^+] = [c_j^-; c_j^+]$, then $[s_i^-; s_i^+] = [s_j^-; s_j^+]$.

Property (i) is a natural property saying that the claimants with the same claims obtain the same shares.

Remark 1. Evidently, the classical bankruptcy problem of n claimants, denoted by $\mathcal{B}_{cla}(N)$, is a subset (in an isomorphic sense) of the interval bankruptcy problem of n claimants $\mathcal{B}_{int}(N)$, as follows. Let $(c_1, \dots, c_n, E) \in (\mathbf{R}_+)^{n+1}$ be an instance of the classical bankruptcy problem of n claimants and let $E = [E^-; E^+]$ with $E^- = E^+ > 0$ be a one point interval of the estate. Then interval vector $(d_1, \dots, d_n, E) \in I(\mathbf{R}_+)^{n+1}$, where $d_i = [0; c_i]$, $i \in N$, is clearly an instance of interval bankruptcy problem $\mathcal{B}_{int}(N)$, as (10) holds by (1). Moreover, for all $i \in N$, it holds

$$m_i^-(d, E) = m_i(c, E),$$

$$m_i^+(d, E) = \max\{c_i; E\},$$

where $m_i(c, E)$ is defined by (5).

Example 3. Consider the data from Example 2, i.e., let (c, E) be an instance of IB-problem defined as follows. $N = \{1, 2, 3\}$, $c = (c_1, c_2, c_3) \in I(\mathbf{R}_+)^3$, where $c_1 = [c_1^-; c_1^+] = [10; 35]$, $c_2 = [c_2^-; c_2^+] = [25; 50]$, $c_3 = [c_3^-; c_3^+] = [30; 60]$, $E = [E^-; E^+] = [85; 115]$. By (13) and (19) we calculate

$$\begin{aligned} m_1^-(E^-) &= 10, m_2^-(E^-) = 25, m_3^-(E^-) = 30, \\ m_1^-(E^+) &= 10, m_2^-(E^+) = 25, m_3^-(E^+) = 30, \\ m_1^+(E^-) &= 30, m_2^+(E^-) = 45, m_3^+(E^-) = 50, \\ m_1^+(E^+) &= 35, m_2^+(E^+) = 50, m_3^+(E^+) = 60, \\ m_N^-(E^-) &= 65, m_N^+(E^-) = 125, m_N^-(E^+) = 65, m_N^+(E^+) = 145. \end{aligned}$$

Then by formulas (14) and (15) we obtain the claimant shares of corresponding AP-rule as follows.

$s_1(c, E) = [s_1^-; s_1^+] = [16.67; 25.63]$, the mean value is $D_1(M) = D_1(16.67, 25.63) = 21.15$.

$s_2(c, E) = [s_2^-; s_2^+] = [31.67; 40.63]$, the mean value is $D_2(M) = D_2(31.67, 40.63) = 36.15$.

$s_3(c, E) = [s_3^-; s_3^+] = [36.67; 48.75]$, the mean value is $D_3(M) = D_3(36.67, 48.75) = 42.71$.

By the above mentioned AP-rule we get $[s_i^-; s_i^+]$, $i \in \{1, 2, 3\}$, being the claimant shares (i.e. "interval solution") of the given IB-problem (c, E) . Moreover, by the mean value $D(M)$ we obtain a "crisp solution" of (c, E) .

4 Conclusion

In this paper we considered situations where a perfectly divisible estate had to be divided among claimants who were subjected to uncertainty in the form of closed intervals. The corresponding claim of each claimant could vary within a closed interval. We classified the division problems under uncertainty of claims and presented a basic division scheme, which was consistent with the classical bankruptcy rules. Three examples were presented to illustrate particular problems and solution concepts. Here, we extended the classical bankruptcy problem (CB-problem), and the corresponding proportional rule (P-rule) to IB-problem. The other classical bankruptcy rules could be considered and extended to bankruptcy problems with fuzzy claims and/or estate in the future research.

References

- [1] Aumann R. J., Maschler M. (1985). Game-theoretic analysis of a bankruptcy problem from the Talmud. *Journal of Economic Theory*, 36, 195–213.
- [2] Branzei, R. et al. (2004). How to cope with division problems under interval uncertainty of claims?. *Int. J. Uncertain and Fuzziness*, 12, 191–200.
- [3] Curiel, I. J. et al. (1987). Bankruptcy games. *Z. Op. Res.* 31, A143–A159.
- [4] Habis, H. et al. (2013). Stochastic bankruptcy games. *Int. J. Game Theory* 42, 973–988.
- [5] Thomson, W., Chun, Y. (1992). Bargaining problems with claims. *Mathematical Social Sciences* 24, 19–33.
- [6] Yager, R. R. et al. (2000). Fair division under interval uncertainty. *Int. J. Uncertain. Fuzziness* 8, 611–618.

Acknowledgment.

This research has been supported by GACR project No. 21-03085S.

Searching for a Unique Good Using Imperfect Comparisons

David M. Ramsey ¹

Abstract. A decision maker must choose one of n offers. The reward obtained is the value of the offer accepted minus the costs of making the pairwise comparisons required to make a decision. If pairwise comparisons are perfect, then comparing each successive offer with the currently highest ranked one is optimal. In practice, decision makers may make errors when comparing two offers or be unable to compare them. Under a round robin procedure, the decision maker compares each offer with all of the other offers. The search costs incurred under such a procedure are greater than under the sequential procedure. However, when comparison is imperfect, the round robin procedure is more robust against selecting a sub-optimal offer. Models of these two procedures are presented. Simulations are used to compare these procedures.

Keywords: best choice problem, limited rationality, multi-criteria decision making, consumer decision

JEL Classification: C44

AMS Classification: 90B40

1 Introduction

This paper considers a problem of choosing one offer from a set of n when a decision maker (DM) cannot precisely ascribe a value to an offer or perfectly compare the value of two offers. For example, suppose a DM wishes to buy a flat. Since offers are characterized by multiple attributes, many of which are qualitative, it is impossible for the DM to ascribe a value to an offer based on a single viewing or perfectly compare offers. By assumption, each offer has an underlying value (which is not directly observable to the DM). The goal of the DM is to maximize the expected reward from search, which is taken to be the value of the offer accepted minus the search costs incurred.

Two procedures based on pairwise comparisons are compared. Under the sequential procedure, offers appear in sequence. The one assessed to be the best so far (the current candidate) is compared with each successive offer. After observing all the offers, the current candidate is accepted. This requires $n - 1$ comparisons. Under the round robin procedure, each offer is compared pairwise with all the others. This requires $0.5n(n - 1)$ comparisons. The offers involved in a comparison obtain scores based on the DM's perceived preference. The overall score of an offer is the sum of the corresponding scores. The offer with the highest overall score is selected.

The models presented take into account limitations on perception that affect pairwise comparisons. One limitation might be the context in which an offer is initially viewed. For example, when visiting a flat, the owner's dog might constantly bark. This may lead to systematic underestimation of the attractiveness of the offer in pairwise comparisons. Another limitation is related to how pairwise comparisons are made. By assumption, there is a bias specific to the conditions in which any pairwise comparison is made. Additionally, DMs may express their preferences (based on a pairwise comparison) with different strengths. One DM might always state a clear preference for one offer over another, while another uses a more gradated scale of preferences (see Saaty [19]).

DMs use heuristics adapted to their perceptive abilities, the structure of the information available and the costs of cognitive effort/search (see Simon [20], Todd and Gigerenzer [23], Bobadilla-Suarez and Love [2]). Pairwise comparisons are useful when a DM cannot ascribe values to offers, but can state preferences between pairs of offers. Here, search costs are assumed to be proportional to the number of pairwise comparisons used to select an offer. Suppose that a pairwise comparison is relatively costly, but the results are consistent. The sequential procedure should be preferred over the round robin procedure, as using the latter significantly increases search costs, but has a limited effect on the value of the offer accepted. Now suppose that pairwise comparisons are very cheap, but inconsistent. Using the round robin procedure may lead to the DM avoiding an inappropriate choice, while the search costs are not greatly increased. In such a case, the round robin procedure is preferred.

This research is linked to literature on the use of pairwise comparisons in ranking a set of offers. Much research has been devoted to investigating the consistency of such rankings (see e.g. Kułakowski [11], Kułakowski *et al.* [12], Mazurek [14]) and has obvious applications to how sporting tournaments should be organized (e.g. see Rubenstein

¹ Wrocław University of Science and Technology, Department of Operations Research, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland, david.ramsey@pwr.edu.pl

[17], Cook *et al.* [4], Csató [6]). There has been less research on how such procedures can be used to select a valuable offer at relatively low cost (see e.g. Kawa & Koczkodaj [10], Kułakowski [13], Ramsey [16]). Ryvkin [18]) considers similar processes to the ones used here: the knockout procedure ($n - 1$ comparisons necessary) and the round robin procedure. The Swiss procedure, often used in chess tournaments, is of intermediate complexity. Using this procedure, players are paired with competitors assessed to be of similar strength (see Csató [5]).

The models presented here (particularly the sequential procedure) are also related to the secretary problem. In the classical version (see Gilbert & Mosteller [9]), the DM wishes to maximize the probability of accepting the best offer. Various adaptations of this model have been considered (e.g. Bruss [3], Smith [22]). Skarupski [21] presents a model in which the DM cannot perfectly rank offers, while Georgiou *et al.* [8] consider a model where offers are not always comparable. Relatively little work has been devoted to maximizing the expected value of the offer accepted. Bearden [1] presents such a model (without taking into account search costs), which was adapted by Ferenstein and Krasnosielska [7]. Palley and Kremer [15] extend these models based on experimental evidence.

Section 2 presents models of the two selection procedures. Results based on the simulation of these search procedures are presented in Section 3. Conclusions and directions for future research are given in Section 4.

2 The Model

2.1 General Assumptions

A decision maker (DM) must select one of n offers using pairwise comparisons. These offers are assumed to have an underlying value. Due to limits on perceptive abilities, the DM cannot carry out perfect assessment. These limits are split into three dimensions. A DM's assessment of an offer may have a systematic bias, due to how information about it is presented. Secondly, the conditions in which pairwise comparisons are made may lead to errors in comparisons. Thirdly, a DM cannot assess the difference between the values of offers, but gives an ordinal measure of the strength of such preferences (it is possible that neither offer is preferred to the other).

It is assumed that the standardized value of offer i , V_i , has a normal distribution with mean zero and variance one. The initial appraisal of the value of this offer by the DM, X_i , is given by

$$X_i = \frac{V_i + \epsilon_i}{1 + \sigma_1^2}, \quad (1)$$

where ϵ_i is the error in the appraisal of the value of the i -th offer. This is a systematic error that is incorporated into all the pairwise comparisons involving the i -th offer. It is assumed that ϵ_i has a normal distribution with mean zero and variance σ_1^2 . Hence, the appraisal X_i is standardized, i.e. it has mean zero and variance one.

Pairwise comparisons are based on the DM's assessment of the values of two offers. Assume that

$$Y_{i,j} = \frac{X_i - X_j + \delta_{i,j}}{2 + \sigma_2^2}, \quad (2)$$

where $\delta_{i,j}$ describes the error specific to the comparison of the values of the i -th and j -th offers. It is assumed that $\delta_{i,j}$ has a normal distribution with mean zero and variance σ_2^2 . The value $Y_{i,j}$ is standardized to have a variance of one. Positive values of $Y_{i,j}$ correspond to offer i appearing to have a greater value than offer j . The preferences resulting from such an assessment are modulated by the level of definition of preferences (described below).

Assume that the preferences of a DM when making a pairwise comparison can be described using a Saaty scale. When offer j is not preferred to offer i , then the score $s_{i,j}$ describes the strength of this preference, $s_{i,j} \in \{1, 2, 3, \dots, 9\}$. When $s_{i,j} = 1$, then the DM does not express any preference. The larger the value of $s_{i,j}$, the greater the preference for offer i . It is assumed that $s_{j,i} = 1/s_{i,j}$. The level of definition of preferences is described by the parameter k , where $k \geq 0$. Suppose $Y_{i,j} > 0$, the strength of the preference for i is given by

$$s_{i,j} = \min\{9, 1 + \lfloor 8Y_{i,j}/k \rfloor\}, \quad (3)$$

where $\lfloor x \rfloor$ is the integer part of x . Large values of k correspond to lowly defined preferences. For example, when $k = 0$, then the offer which has the highest value according to a pairwise comparison is always maximally preferred, i.e. $s_{i,j} = 9$. When $k = 2$, for offer i to be preferred to offer j , then $Y_{i,j} \geq 0.25$, i.e. the apparent difference in values must be at least 25% of the standard deviation of these differences. In this case, for offer i to be maximally preferred to offer j , the apparent difference in values must be at least twice this standard deviation.

2.2 The Sequential Procedure

When offers are observed in sequence, the following is a natural procedure for selecting an offer:

1. The first offer observed is set to be the current candidate.
2. Each successive offer to be seen is compared with the current candidate. If a new offer is preferred to the current candidate, then the new offer becomes the current candidate.
3. Once all of the offers have been viewed, the current candidate is the offer selected by the DM.

Note that when a DM neither prefers the current candidate to a new offer nor vice versa, then it is assumed that the current candidate remains unchanged. Although this procedure is natural in the case of sequential search, it can also be applied when offers are observed in parallel. By assumption $n - 1$ pairwise comparisons are made.

2.3 The Round Robin Procedure

Each offer is compared with all of the others. The overall score of offer i , S_i , is the sum of the scores obtained in these comparisons. Hence, $S_i = \sum_{1 \leq j \leq n, j \neq i} s_{i,j}$. The offer with the largest overall score is accepted. If several offers have the same maximum score, then one of them is chosen at random. Using the Saaty procedure, an offer's overall rating is given by the geometric average of the individual scores. However, it is expected that decisions based on the mean score (equivalently, on the sum of scores) will almost always coincide with those based on the geometric average. Also, basing the final decision on the sum of scores is much easier in practice.

2.4 Errors in Perception and the Effectiveness of the Procedure Used

Assume that the reward a DM obtains from search is the value of the offer selected minus the search costs incurred, which are proportional to the number of pairwise comparisons made. The cost of a pairwise comparison is c . When $\sigma_1^2 > 0$ and $\sigma_2^2 = 0$, then the results of pairwise comparisons depend solely on the initial assessments of the individual offers. Hence, the round robin procedure cannot lower the perceptual biases involved in pairwise comparisons. It is expected that the sequential procedure would be preferred to the round robin procedure.

When $\sigma_1^2 = 0$ and $\sigma_2^2 > 0$, then the result of a pairwise comparison depends on the conditions in which it is made. Given the difference in the underlying values of offers, the results of pairwise comparisons are independent. In this case, the round robin procedure would seem to be relatively effective, particularly when c is small, since it is likely to eliminate errors in decision making that result from an error in an individual pairwise comparison.

2.5 Simulations

The results are based on a million simulations of the search procedures (the sequential and round robin procedures are denoted S and R , respectively) for each of the 600 problems defined by a combination of the following set of parameters:

- The total number of offers: $n \in \{5, 10, 20, 30\}$.
- The (inverse) level of discrimination of preferences: $k \in \{0, 0.1, 0.2, 0.5, 1, 2\}$.
- The relative standard deviation of perception bias in initial assessment of an offer: $\sigma_1 \in \{0, 0.1, 0.2, 0.5, 1\}$.
- The relative standard deviation of perception bias in pairwise comparisons of offers: $\sigma_2 \in \{0, 0.1, 0.2, 0.5, 1\}$.

The expected value of the offer accepted using search procedure $s \in \{S, R\}$ for the set of parameters $(n, k, \sigma_1, \sigma_2)$ is estimated as the mean value of the offer accepted in the million corresponding simulations and denoted by $V_s(n, k, \sigma_1, \sigma_2)$. Since the number of pairwise comparisons used by these procedures is fixed, the expected reward from search, $U_s(n, k, \sigma_1, \sigma_2)$ can be estimated for any cost of a pairwise comparison, c , as follows:

$$U_S(n, k, \sigma_1, \sigma_2) = V_S(n, k, \sigma_1, \sigma_2) - cn; \quad U_R(n, k, \sigma_1, \sigma_2) = V_R(n, k, \sigma_1, \sigma_2) - 0.5cn(n - 1). \quad (4)$$

3 Results of the Simulations

3.1 Effects of Errors in Perception

Table 1 illustrates the relation between the mean value of the offer accepted and the standard deviations of the errors in perception when $n = 10$ and $k = 0.5$. As described above, σ_1 describes the error in the initial perception of an offer, which has a systematic effect on the pairwise comparisons, and σ_2 describes the error made in an individual pairwise comparison. The two values given in each cell are the expected values of the offer accepted under the sequential procedure and the round robin procedure, respectively. The following effects are visible:

- The expected value of the offer accepted is decreasing in the standard deviation of both types of error.

- The expected value of the offer accepted is always greater under the round robin procedure.
- When errors result purely from the systematic bias ($\sigma_2 = 0$), then these expected values are almost identical.
- As the errors from individual pairwise comparisons increases (as σ_2 increases), the difference between these expected values increases. The round robin procedure avoids decisions based on a few, fallible comparisons.
- In the case of the sequential procedure, the effects of both type of errors are comparable. The round robin procedure is more robust to errors in individual pairwise comparisons (high values of σ_2).

	$\sigma_2 = 0$	$\sigma_2 = 0.1$	$\sigma_2 = 0.2$	$\sigma_2 = 0.5$	$\sigma_2 = 1$
$\sigma_1 = 0$	(1.5357, 1.5374)	(1.5317, 1.5355)	(1.5215, 1.5315)	(1.4512, 1.5013)	(1.2331, 1.4260)
$\sigma_1 = 0.1$	(1.5277, 1.5315)	(1.5248, 1.5284)	(1.5151, 1.5228)	(1.4446, 1.4925)	(1.2258, 1.4189)
$\sigma_1 = 0.2$	(1.5051, 1.5076)	(1.5025, 1.5058)	(1.4928, 1.5000)	(1.4232, 1.4723)	(1.2090, 1.3979)
$\sigma_1 = 0.5$	(1.3745, 1.3755)	(1.3719, 1.3740)	(1.3621, 1.3692)	(1.2969, 1.3428)	(1.1034, 1.2752)
$\sigma_1 = 1$	(1.0853, 1.0867)	(1.0833, 1.0859)	(1.0759, 1.0831)	(1.0277, 1.0611)	(0.8723, 1.0091)

Table 1 Effect of errors in perception on the mean value of the offer accepted: $n = 10$, $k = 0.5$. First value - sequential procedure, second value - round robin procedure.

3.2 Effect of the Number of Offers

Table 2 illustrates the relation between the mean value of the offer accepted and the number of offers available, n . As n increases, the greater the mean value of the offer accepted. The difference between the mean values of the offer accepted under the two procedures is only noticeable when the error in individual pairwise comparisons (σ_2) is large. This difference is increasing in n . However, note that the reward from search takes into account the search costs (increasing in n for both procedures, but more quickly for the round robin procedure). This is briefly considered in Section 3.4. Due to space limitations, the effect of search costs and the number of offers available on the type of search procedure that should be used will be considered in future research.

	$\sigma_2 = 0$	$\sigma_2 = 0.1$	$\sigma_2 = 0.2$	$\sigma_2 = 0.5$	$\sigma_2 = 1$
$n = 5$	(1.0389, 1.0387)	(1.0353, 1.0379)	(1.0302, 1.0337)	(0.9815, 1.0033)	(0.8484, 0.9223)
$n = 10$	(1.3745, 1.3763)	(1.3719, 1.3731)	(1.3621, 1.3660)	(1.2969, 1.3298)	(1.1035, 1.2498)
$n = 20$	(1.6663, 1.6704)	(1.6637, 1.6681)	(1.6533, 1.6633)	(1.5722, 1.6412)	(1.2998, 1.5944)
$n = 30$	(1.8239, 1.8259)	(1.8196, 1.8253)	(1.8086, 1.8172)	(1.7172, 1.7848)	(1.3921, 1.7410)

Table 2 Effect of the number of offers available on the mean value of the offer accepted: $\sigma_1 = 0.5$, $k = 0.5$. First value - sequential procedure, second value - round robin procedure.

3.3 Effect of the Precision of Expressed Preferences

Table 3 illustrates the relation between the mean value of the offer accepted and the precision of expressed preferences (as described by the parameter k). As k increases, the expressed preferences become less clearly defined. For small values of σ_2 , under the round robin procedure there is no relation between k and the value of the offer accepted. Since the overall ranking of the offers is based on a large number of stable pairwise comparisons, this ranking is robust to changes in the precision of expressed preferences. On the other hand, when the sequential procedure is used, there is a weak negative relation between the value of k and the mean value of the offer accepted, i.e. when the sequential procedure is used and pairwise comparisons are stable, it pays to state a clear preference.

When σ_2 is relatively large (i.e. the results of pairwise comparisons are variable), there is a positive correlation between k and the mean value of the offer accepted under either search procedure. This indicates that when pairwise comparisons are not stable, then it pays a DM to express his/her preferences in a gradated manner (i.e. show a higher level of skepticism to the results of each individual pairwise comparison).

3.4 Overall Comparison of the Procedures

As the round robin procedure involves a more thorough comparison of the offers, the expected value of the offer obtained will be greater than under the sequential procedure. However, for the round robin procedure to be favoured, this difference must be sufficient to overcome the increased search costs. A regression model describing this difference, $d(n, k, \sigma_1, \sigma_2) = V_R(n, k, \sigma_1, \sigma_2) - V_S(n, k, \sigma_1, \sigma_2)$, was constructed using all the results from the simulations by considering the following as possible explanatory variables: $n, n^2, k, \sigma_1, \sigma_2, \sigma_1^2, \sigma_2^2$. Initially, all the

	$\sigma_2 = 0$	$\sigma_2 = 0.1$	$\sigma_2 = 0.2$	$\sigma_2 = 0.5$	$\sigma_2 = 1$
$k = 0$	(1.6703, 1.6702)	(1.6656, 1.6655)	(1.6532, 1.6559)	(1.5618, 1.6062)	(1.2650, 1.5387)
$k = 0.1$	(1.6699, 1.6700)	(1.6664, 1.6677)	(1.6541, 1.6587)	(1.5631, 1.6182)	(1.2715, 1.5563)
$k = 0.2$	(1.6695, 1.6677)	(1.6664, 1.6677)	(1.6552, 1.6597)	(1.5658, 1.6260)	(1.2799, 1.5682)
$k = 0.5$	(1.6663, 1.6704)	(1.6637, 1.6681)	(1.6533, 1.6633)	(1.5722, 1.6412)	(1.2998, 1.5944)
$k = 1$	(1.6584, 1.6699)	(1.6566, 1.6695)	(1.6484, 1.6672)	(1.5765, 1.6528)	(1.3284, 1.6135)
$k = 2$	(1.6302, 1.6696)	(1.6264, 1.6698)	(1.6200, 1.6677)	(1.5712, 1.6573)	(1.3684, 1.6246)

Table 3 Effect of the parameter defining the precision of expressed preferences, k , on the mean value of the offer accepted: $n = 20$, $\sigma_1 = 0.5$. First value - sequential procedure, second value - round robin procedure.

variables were included in the model. At each stage of the analysis, the least significant explanatory variable was removed until all of the remaining explanatory variables were significant at the 5% level. The model derived was

$$d(n, k, \sigma_1, \sigma_2) = 0.006n + 0.015k + 0.215\sigma_2^2 - 0.065 - 0.00008127n^2 - 0.019\sigma_1^2. \quad (5)$$

Since the coefficient of determination, R^2 , is 0.78, the model explains 78% of the variation in the difference between the values of the offer accepted under the two procedures considered. Although the standardized residuals do not follow the normal distribution, they lie between -2.961 and 3.506, indicating that there are no major outliers. The absolute error in the estimation increases as the predicted difference d increases, i.e. in particular, when n and σ_2 are large. These results indicate that this linear model is not ideal, but is useful for qualitative description.

Although the gains from using the round robin procedure are increasing in n (for the range of n used), this gain is slower than linear in n . However, the increase in search costs resulting from using the round robin procedure is greater than linear. Hence, as the number of offers increases, the sequential procedure is likely to become favoured.

As argued in Section 2.4, when the error involved in pairwise comparisons, described by σ_2^2 , increases, then the relative advantage of the round robin procedure over the sequential procedure increases. This is due to the fact that the round robin procedure avoids making decisions based on a small number of comparisons of limited accuracy. However, when the error involved in the initial assessment of the value of an offer, described by σ_1^2 , increases, the relative advantage of the round robin procedure falls slightly. This is due to the fact that such errors introduce a systematic bias into pairwise comparisons which cannot be removed by carrying out a fuller set of comparisons.

Similarly, as the expression of preferences becomes less precise (as k increases), the relative advantage of the round robin procedure increases. Note that under the sequential procedure, if the DM is unable to state which of two offers is preferred, then the present candidate is retained. When the round robin procedure is used, the increased number of comparisons will usually indicate which of two such offers should be preferred.

4 Conclusion

This article has presented models of consumer choice where purchase decisions are based on pairwise comparisons. Such procedures are useful, e.g. when choosing a good characterized by multiple attributes. In such cases, a DM may not be able to ascribe a value to an offer, but can compare the attractiveness of any two offers. The DM has limited perception, which affects the results of the pairwise comparisons made. Firstly, the appraisal of a single offer is subject to a systematic bias, which depends on the conditions in which the offer is presented/first viewed. Secondly, any pairwise comparison is subject to error resulting from the specific conditions under which it is carried out. Thirdly, the DM may gradate his/her expression of preferences based on such comparisons.

When the results of pairwise comparisons are variable, DMs should use a more gradated description of their preferences (i.e. they should be less ready to state very clear preferences). Errors resulting from the specific conditions of individual pairwise comparisons can be mitigated by adopting a round robin procedure (i.e. making all the possible pairwise comparisons) rather than a sequential procedure in which the offer currently assessed to be the best is compared with each successive offer. However, as the number of offers increases, the search costs from adopting the round robin procedure will start to outweigh the gains from making a more thorough comparison.

Future research should investigate other procedures for selecting an offer. For example, a knockout procedure (such as those used in tennis tournaments) might be preferable to the sequential procedure. The number of comparisons is the same, but the effect of where an offer appears in a sequence is mitigated. Other procedures where the number of comparisons lies between $n - 1$ and $0.5n(n - 1)$ might be considered. For example, a procedure similar to the

Swiss system (see Csató [5]) could be used. Using such a procedure, each offer would be subject to a number of comparisons, but an offer is eliminated after several comparisons indicate that it is not attractive.

Acknowledgements

This research is funded by the Polish National Science Centre via grant no. 2018/29/B/HS4/02857, "Logistics, Trade and Consumer Decisions in the Age of the Internet".

References

- [1] Bearden, J. N. (2006). A new secretary problem with rank-based selection and cardinal payoffs. *Journal of Mathematical Psychology*, 50(1), 58–59.
- [2] Bobadilla-Suarez, S. & 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] Bruss, F. T. (2000). Sum the odds to one and stop. *Annals of Probability*, 1384–1391.
- [4] Cook, W. D., Doyle, J., Green, R. & Kress, M. (1996). Ranking players in multiple tournaments. *Computers & Operations Research*, 23(9), 869–880.
- [5] Csató, L. (2017). On the ranking of a Swiss system chess team tournament. *Annals of Operations Research*, 254(1), 17–36.
- [6] Csató, L. (2021). A simulation comparison of tournament designs for the World Men's Handball Championships. *International Transactions in Operational Research*, 28(5), 2377–2401.
- [7] Ferencstein, E. Z. & Krasnosielska, A. (2010). No-information secretary problems with cardinal payoffs and Poisson arrivals. *Statistics & Probability Letters*, 80(3-4), 221–227.
- [8] Georgiou, N., Kuchta, M., Morayne, M. & Niemiec, J. (2008). On a universal best choice algorithm for partially ordered sets. *Random Structures & Algorithms*, 32(3), 263–273.
- [9] Gilbert, J. P. & Mosteller, F. (1966). Recognizing the Maximum of a Sequence. *Springer Series in Statistics*, 61, 355.
- [10] Kawa, A. & Koczkodaj, W. W. (2015). Supplier evaluation process by pairwise comparisons. *Mathematical Problems in Engineering*, <https://doi.org/10.1155/2015/976742>
- [11] Kułakowski, K. (2018). Inconsistency in the ordinal pairwise comparisons method with and without ties. *European Journal of Operational Research*, 270(1), 314–327.
- [12] Kułakowski, K., Mazurek, J., Ramiń, J. & Soltys, M. (2019). When is the condition of order preservation met?. *European Journal of Operational Research*, 277(1), 248–254.
- [13] Kułakowski, K., Szybowski, J. & Tadeusiewicz, R. (2014). Tender with success? The pairwise comparisons approach. *Procedia Computer Science*, 35, 1122–1131.
- [14] Mazurek, J. (2011). Evaluation of ranking similarity in ordinal ranking problems. *Acta Academica Karviniensis*, 2, 119–128.
- [15] Palley, A. B. & Kremer, M. (2014). Sequential search and learning from rank feedback: Theory and experimental evidence. *Management Science*, 60(10), 2525–2542.
- [16] Ramsey, D. M. (2020). A game theoretic model of choosing a valuable good via a short list heuristic. *Mathematics*, 8(2), 199.
- [17] Rubinstein, A. (1980). Ranking the participants in a tournament. *SIAM Journal on Applied Mathematics*, 38(1), 108–111.
- [18] Ryvkin, D. (2010). The selection efficiency of tournaments. *European Journal of Operational Research*, 206(3), 667–675.
- [19] Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. *European Journal of Operational Research*, 48(1), 9–26.
- [20] Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129.
- [21] Skarupski, M. (2020). Secretary Problem with Possible Errors in Observation. *Mathematics*, 8(10), 1639.
- [22] Smith, M. H. (1975). A secretary problem with uncertain employment. *Journal of Applied Probability*, 620–624.
- [23] Todd, P. M. and Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23(5), 727–741.

Dealing with uncertainty by Fuzzy evaluation and robust optimization

Tereza Sedlářová Nehézová¹, Michal Škoda², Helena Brožová³

Abstract. The paper addresses the issue of an emergency supply routes planning in an area affected by an unexpected disaster event. Such event can have extensive harmful effects on the availability of the traffic infrastructure, which is a crucial part of any emergency delivery planning. By the nature of the matter, emergency supply routes planning must cope with a great presence of uncertainty in many aspects. It is demonstrated in this work how the fuzzy evaluation and linear robust optimization can be used to handle the uncertainty and plan emergency supply routes. The fuzzy evaluation is used to describe the consequences of a disaster event on routes. The fuzzy evaluation is based on fuzzy linguistic scales, that allows the decision-maker to use vague terms to describe the current situation and create a relevant uncertainty set later used by the robust optimization model for the route planning itself. The robust optimization model is based on robustified shortest path problem. The presented approach is described in detail with example at the end.

Keywords: Decision making, fuzzy number, fuzzy linguistic scale, robust optimization, uncertainty

JEL Classification:

AMS Classification:

1 Introduction

Emergency management (EM) is characterized with need of fast actions that aims to save lives, minimize property loss, and soon deliver aid as a response to a disastrous event. In such situations, the decision-maker has to deal with large and complex problems, and also with a lack of certain data, that could serve as a basis to plan appropriate reaction [1]. Thus, there is great potential for use of operations research (OR) models and approaches, that allows one to deal with uncertainty. One of such techniques is robust optimization (RO). To name a few cases of use of RO in EM: traffic assignment model for robust logistics plan was proposed [2], model for establishing decontamination facilities [3], and a multi-objective optimization models for choosing emergency facilities locations for rescue operations [4].

When using optimization models, there is a problem with sensitivity to parameter changes [5], [6] which can lead to changes or even infeasibility of the given solution, and such solutions are worthless for decision making. The first effort on behalf of optimization under uncertainty belongs to Dantzig's pioneering work in [7], and later to Soyster in [8], which is first truly robust optimization model providing solutions resistant to data uncertainty. RO models work with uncertainty through hard constraints to maintain the robust solution by restricting the feasible set. Such approach can lead to over-conservative solutions [9], that gives up on optimality as a price of the protection against uncertainty. Bertsimas and Sim [10] introduced an approach called Γ -robustness, where uncertainty is defined by coefficients δ representing deviations in symmetric ranges from deterministic values, and by Γ limiting the maximum number of coefficients that can simultaneously deviate. By this approach, one is able to control the trade-off between optimality and protection against uncertainty – the level of conservatism of the solution.

As mentioned before, emergency management is a challenging area of human endeavor where uncertainty plays one of the main roles. The type of uncertainty and the way of dealing with it have been addressed in many articles, for instance [11]–[13]. In this paper the fuzzy logic introduced by Zadeh [14] will be used as a tool to manage uncertainty because it represents easy way of handling problems with imprecise and incomplete data, has ability to deal with uncertainty and nonlinearity and is easily customizable in natural linguistic terms [15]. Dealing with uncertainty in decision-making problems the fuzzy logic has proven to be very useful [16]. As confirmed [17], fuzzy logic has also been proved as useful in disaster situations decision problems.

¹ Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, nehézova@pef.czu.cz.

² Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, skoda@pef.czu.cz.

³ Czech University of Life Sciences, Kamýcká 129, Prague, Czech Republic, brozova@pef.czu.cz.

The goal of this paper is to show a new approach of finding the most suitable route in crisis situations which is based on Bertimas and Sim's robust transformation of Shortest path linear optimization model and on the fuzzy set logic, mainly the fuzzy linguistic scales, for determination of the uncertainty set necessary for the robust model.

The Materials and methods can be found in the second part of the paper, where the shortest part model and its robust counterpart are presented. Also, the basis of fuzzy logic that is used in this paper, is outlined. Then the practical example of introduced approach is showed in the third part. At the end, there is the conclusion.

2 Materials and methods

2.1 Robust shortest path problem formulation

Let $G = (V, E)$ is an undirected graph, where V is a set of n vertices, and $E = \{(i, j): i, j \in V\}$ denotes set of m edges. For each i , there is denoted incoming edge by $\gamma^+(i)$, and outgoing edge by $\gamma^-(i)$. For each edge $(i, j) \in E$, there is given a non-negative cost c_{ij} . Now, integer formulation of Shortest Path Problem (SPP) from vertex s to vertex t is as follows:

$$\begin{aligned} & \text{Minimize } \sum_{(i,j) \in E} c_{ij} x_{ij} \\ & \text{s.t.} \\ & \sum_{(i,j) \in \gamma^+(i)} x_{ij} - \sum_{(i,j) \in \gamma^-(i)} x_{ji} = \begin{cases} 1 & \text{if } i = s \\ -1 & \text{if } i = t \\ 0 & \text{if } i \neq s, t \end{cases} \quad i \in V \\ & x_{ij} \in \{0,1\}, \forall (i,j) \in E \end{aligned} \quad (1)$$

Where x_{ij} are binary variables, where 1 means that edge (i, j) is used, 0 otherwise, $c_{ij} \geq 0$ are the costs of using an edge (i, j) [18].

If the values of c_{ij} are a priori known, such solution does not take the uncertainty into account. If one is able to define a possible maximal deviation of c_{ij} by positive number δ_{ij}^c , so c_{ij} belongs to the symmetric interval $[c_{ij} - \delta_{ij}^c, c_{ij} + \delta_{ij}^c]$, and also define a number Γ , which denotes the maximum number of coefficients c_{ij} that will deviate at most by δ_{ij}^c then a robust counterpart of SPP (RSPP) can be proposed ([23] and [10]).

$$\begin{aligned} & \text{Minimize } K \\ & \text{s.t.} \\ & - \sum_{(i,j) \in E} c_{ij} x_{ij} + \Gamma z + \sum_{(i,j) \in E} p_{ij} + K \leq 0 \\ & \sum_{(i,j) \in \gamma^+(i)} x_{ij} - \sum_{(i,j) \in \gamma^-(i)} x_{ji} = \begin{cases} 1 & \text{if } i = s \\ -1 & \text{if } i = t \\ 0 & \text{if } i \neq s, t \end{cases}, \quad i \in V \\ & z + p_{ij} \geq \delta_{ij}^c x_{ij}, \forall (i,j) \in U \subseteq E \\ & z \geq 0 \\ & p_{ij} \geq 0, \forall (i,j) \in U \subseteq E \\ & x_{ij} \in \{0,1\}, \forall (i,j) \in E \end{aligned} \quad (2)$$

where U denotes the set of edges that actually can deviate, Γ denotes the maximum number of edges from U which coefficients c_{ij} will deviate at most by δ_{ij}^c , z and p_{ij} are auxiliary variables that connect first and third, respectively fourth constraints. For details, we refer also to [24] and [25], where this robustification approach was shown.

2.2 Fuzzy logic

Decision-makers very often face uncertainty and in consequence are not able to evaluate a quantity by a specific number. In such situation is useful to use fuzzy numbers and fuzzy linguistic scales.

A fuzzy number is a special case of fuzzy sets if it is convex and normal [19]. There are many types of fuzzy numbers, but the most typical example of fuzzy numbers is a triangular fuzzy number. A Triangular fuzzy number

is a number $A = (a_1, a_2, a_3)$ where a_1, a_2, a_3 are real numbers and $a_1 \leq a_2 \leq a_3$ [20]. The product of two triangular fuzzy numbers $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$ is given by the formula (2), if numbers $a_1, a_2, a_3, b_1, b_2, b_3$ are nonnegative.

$$A \times B = (a_1 \times b_1, a_2 \times b_2, a_3 \times b_3) \quad (3)$$

A linguistic variable is defined using a quintuple (X, T, U, G, M) where X is the name of the variable, $T = \{T_1, T_2, \dots, T_n\}$ is the set of terms of X , U is the universe of discourse (generally $[0,1]$), G is a syntactic rule for generating the derived terms, and M is a semantic rule for associating each term with the proper fuzzy set (number) defined on U [21]. In the case a crisp number is required as a result, the defuzzification based on the center of gravity (COG) will be used. For the COG of a triangular fuzzy number $A = (a_1, a_2, a_3)$ the following applies [22]:

$$z = COG(A) = \frac{1}{3} \cdot \frac{a_3^2 - a_1^2 + a_3a_2 - a_2a_1}{a_3 - a_1} \quad (4)$$

3 Results

3.1 The proposed approach application

Our approach of finding the most suitable route in crisis situations is based on Bertimas and Sim's robust transformation of the Shortest path linear optimization model and on the determination of the uncertainty set using proper fuzzy linguistic scale and triangular fuzzy numbers arithmetic.

In the first step, the selection of the criteria by which the emergency situation will be evaluated should be done. In the following practical example of the emergency situation caused by a hurricane the following criteria are used:

n_1 = Hurricane strength

n_2 = Route quality/resistance

n_3 = Risk of impassability

These criteria have to be split into two parts, where the first part called will be used for the calculation of δ_{ij}^c . This will contain criteria n_2 and n_3 in the practical example. For the determination of Γ the remaining criteria will be used, in the practical example it will be the criterion n_1 .

In the second step, the selection of a fuzzy linguistic scale for the evaluation of the criteria selected in the previous step should be done. The fuzzy linguistic scales described in Table 1 will be used in the practical example.

T	Linguistic term	Fuzzy number
T_1	Low (L)/ Small (S)	$Y_1 = (0.05, 0.25, 0.45)$
T_2	Moderate (M)	$Y_2 = (0.30, 0.50, 0.70)$
T_3	High (H)/Big (B)	$Y_3 = (0.55, 0.75, 0.95)$

Table 1 Fuzzy linguistic scale

In the third step, the values of the criteria are obtained for each route using a predetermined fuzzy language scale.

In the fourth step, the determination of the uncertainty set should be done. The first subset of the criteria will be used for the calculation of δ_{ij}^c . δ_{ij}^c will receive the normalized defuzzified value of the product of the criteria fuzzy evaluations for each route. The normalization should be based on the theoretical maximum that can be reached with the relevant criteria;

$$\delta_{ij} = \frac{COG(Y_2^{(i,j)} \times Y_3^{(i,j)})}{COG(Y_2 \times Y_3)} * c_{ij}, (i, j) \in E \quad (5)$$

Γ will receive the normalized defuzzified value of the criterion evaluation of criteria from the second subset multiplied by the number of edges;

$$\Gamma = ceil\left(\frac{COG(Y_1)}{COG(Y_3)}\right) * |E| \quad (6)$$

In the fifth step, the shortest path is calculated with certain cost coefficients c_{ij} using the Shortest path problem optimization model and with uncertain cost using the robust shortest path problem optimization model with deviations created by fuzzy evaluation. The necessary δ_{ij}^c and Γ are received in step four. Then, both results will be compared. The analysis of results with all or selected values of Γ from $[0; |E|]$ also can be done.

3.2 Practical example

We assume a given unoriented graph $G = (V, E)$, $|V| = 10$, $|E| = 13$ (see figure 1), representing a road system among cities and major crossroads, roads are denoted as follows: double line – highway, single line - first-class road, dashed line - local road. It is given the following scenario: City in vertex 10 and surrounding area was struck with a hurricane of category 3 on Saffir-Simpson Scale. The aim is to find the shortest path from vertex 1 where first aid is located to the city in vertex 10.



Figure 1 Given graph with types of roads

The shortest path in the given graph from 1 to 10 without any uncertain coefficients is as follows: 1 – 2 – 8 – 10, with total travel time 2,2 hours.

Calculation for expected scenario

The distance of each route and the linguistic evaluation based on estimation of authors done before the hurricane took place, assigned fuzzy numbers and calculation of δ for each route can be seen in Table 2. The theoretical maximum used for normalization in Table 2 was 0.589.

Route	Travel time (hours)	Route quality/resistance		Distance of the route		$Y_2^{(i,j)} \times Y_3^{(i,j)}$	$\frac{COG(Y_2^{(i,j)} \times Y_3^{(i,j)})}{COG(Y_3 \times Y_3)}$
		Term	Fuzzy Number	Term	Fuzzy Number		
x1_2	0,3	M	(0.30, 0.50, 0.70)	M	(0.30, 0.50, 0.70)	(0.090, 0.250, 0.490)	47.0%
x1_3	0,32	H	(0.05, 0.25, 0.45)	M	(0.30, 0.50, 0.70)	(0.015, 0.125, 0.315)	25.7%
x2_8	1,36	M	(0.30, 0.50, 0.70)	H	(0.55, 0.75, 0.95)	(0.165, 0.375, 0.665)	68.2%
x3_4	0,82	M	(0.30, 0.50, 0.70)	H	(0.55, 0.75, 0.95)	(0.165, 0.375, 0.665)	68.2%
x3_5	0,35	H	(0.05, 0.25, 0.45)	M	(0.30, 0.50, 0.70)	(0.015, 0.125, 0.315)	25.7%
x4_6	1,09	L	(0.55, 0.75, 0.95)	M	(0.30, 0.50, 0.70)	(0.165, 0.375, 0.665)	68.2%
x4_10	0,92	M	(0.30, 0.50, 0.70)	H	(0.55, 0.75, 0.95)	(0.165, 0.375, 0.665)	68.2%
x5_7	0,9	M	(0.30, 0.50, 0.70)	M	(0.30, 0.50, 0.70)	(0.090, 0.250, 0.490)	47.0%
x6_7	0,94	M	(0.30, 0.50, 0.70)	L	(0.05, 0.25, 0.45)	(0.015, 0.125, 0.315)	25.7%
x7_9	0,71	H	(0.05, 0.25, 0.45)	L	(0.05, 0.25, 0.45)	(0.003, 0.063, 0.203)	15.1%
x8_9	0,58	M	(0.30, 0.50, 0.70)	M	(0.30, 0.50, 0.70)	(0.090, 0.250, 0.490)	47.0%
x8_10	0,54	L	(0.55, 0.75, 0.95)	M	(0.30, 0.50, 0.70)	(0.165, 0.375, 0.665)	68.2%
x9_10	0,27	H	(0.05, 0.25, 0.45)	L	(0.05, 0.25, 0.45)	(0.003, 0.063, 0.203)	15.1%

Table 2 Distances and values for the calculation of δ

The calculation of Γ was based on the evaluation of Hurricane strength, which was evaluated as moderate. This linguistic term T_2 has been assigned by fuzzy number (0.30, 0.50, 0.70). Since the theoretical maximum in this criterion is 0.75, the normalized value after defuzzification is 0.667, which means that the final value of Γ is rounded up, thus $\Gamma = \text{ceil}(13 * 0.667) = 9$.

If the fuzzy evaluation of the uncertainty set and robust form of the model is used, then the shortest path with uncertainty considered, is as follows: 1 – 3 – 5 – 7 – 9 – 10, with a total travel time 3,5 hours.

Analysis for possible scenarios

For interest, we calculated a robust version of the model for all possible numbers of edge evaluation changes (Table 3, Figure 2).

It can be seen that from four expected changes the solution is already stable, it is given by the same routes and the same travelling time.

	Γ	Vertices	Travel time
Certainty	0	1 – 2 – 8 – 10	2.2
	1	1 – 3 – 4 – 10	3.0
	2	1 – 3 – 5 – 7 – 9 – 10	3.3
	3	1 – 3 – 5 – 7 – 9 – 10	3.4
	4	1 – 3 – 5 – 7 – 9 – 10	3.5
Uncertainty	9	1 – 3 – 5 – 7 – 9 – 10	3.5
	13	1 – 3 – 5 – 7 – 9 – 10	3.5

Table 3 Results for various values of Γ

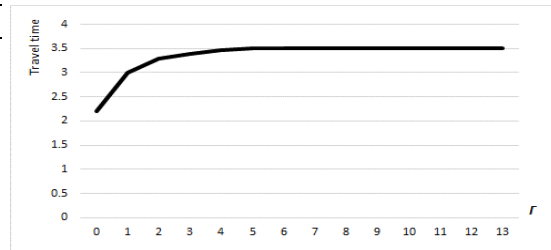


Figure 2 Results for various values of Γ

4 Conclusion

Uncertainty is an integral part of decision-making in emergency management. For a decision-maker is necessary to have a suitable tool that allows him to handle the uncertainty in data so and so the necessary actions based on available data can be taken. We presented an approach based on fuzzy linguistic scales to determine an uncertainty set for route planning in the area affected by the disaster. For the route planning itself, a robust formulation of the shortest path problem based on gamma-robustness was used. The advantage of our approach is such that a decision-maker can use a vague naming of the actual situation and still get relevant outputs from used models, thus can take fast action when detailed data are not available and deliver aid to the affected areas, which is an essential part of the response to any disastrous situation.

Acknowledgements

This paper was supported by Internal Grant Agency (IGA) of FEM CZU, project No. 2021B0003.

References

- [1] R. M. Tomasini and L. N. van Wassenhove, "Pan-american health organization's humanitarian supply management system: de-politicization of the humanitarian supply chain by creating accountability," *Journal of Public Procurement*, vol. 4, no. 3, pp. 437–449, Mar. 2004, doi: 10.1108/JOPP-04-03-2004-B005.
- [2] A. Ben-Tal, B. do Chung, S. R. Mandala, and T. Yao, "Robust optimization for emergency logistics planning: Risk mitigation in humanitarian relief supply chains," *Transportation Research Part B: Methodological*, vol. 45, no. 8, pp. 1177–1189, Sep. 2011, doi: 10.1016/j.trb.2010.09.002.
- [3] Y.-M. Ji and M.-L. Qi, "A robust optimization approach for decontamination planning of emergency planning zone: Facility location and assignment plan," *Socio-Economic Planning Sciences*, vol. 70, p. 100740, Jun. 2020, doi: 10.1016/j.seps.2019.100740.
- [4] J.-R. Feng, W. Gai, and J. Li, "Multi-objective optimization of rescue station selection for emergency logistics management," *Safety Science*, vol. 120, pp. 276–282, Dec. 2019, doi: 10.1016/j.ssci.2019.07.011.
- [5] A. Ben-Tal and A. Nemirovski, "Robust solutions of Linear Programming problems contaminated with uncertain data," *Mathematical Programming*, vol. 88, no. 3, pp. 411–424, 2000, doi: 10.1007/PL00011380.
- [6] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and Applications of Robust Optimization," *SIAM Review*, vol. 53, no. 3, pp. 464–501, Jan. 2011, doi: 10.1137/080734510.
- [7] G. B. Dantzig, "Linear Programming Under Uncertainty," pp. 1–11, 2011, doi: 10.1007/978-1-4419-1642-6.

- [8] A. L. Soyster, "Convex Programming with Set-Inclusive Constraints and Applications to Inexact Linear Programming," *Operations Research*, vol. 21, no. 5, pp. 1154–1157, Oct. 1973, doi: 10.1287/opre.21.5.1154.
- [9] A. Ben-Tal, L. el Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton University Press, 2009.
- [10] D. Bertsimas and M. Sim, "The price of robustness," *Operations Research*, vol. 52, no. 1, pp. 35–53, 2004, doi: 10.1287/opre.1030.0065.
- [11] G. C. Marano and G. Quaranta, "Fuzzy-based robust structural optimization," *International Journal of Solids and Structures*, vol. 45, no. 11–12, pp. 3544–3557, Jun. 2008, doi: 10.1016/j.ijsolstr.2008.02.016.
- [12] W. M. Bulleit, "Uncertainty in Structural Engineering," *Practice Periodical on Structural Design and Construction*, vol. 13, no. 1, pp. 24–30, Feb. 2008, doi: 10.1061/(ASCE)1084-0680(2008)13:1(24).
- [13] E. Nikolaidis, S. Chen, H. Cudney, R. T. Haftka, and R. Rosca, "Comparison of Probability and Possibility for Design Against Catastrophic Failure Under Uncertainty," *Journal of Mechanical Design*, vol. 126, no. 3, pp. 386–394, May 2004, doi: 10.1115/1.1701878.
- [14] L. A. Zadeh, "FUZZY SETS," Nov. 1964. doi: 10.21236/AD0608981.
- [15] U. W. Woju and A. S. Balu, "Fuzzy uncertainty and its applications in reinforced concrete structures," *Journal of Engineering, Design and Technology*, vol. 18, no. 5, pp. 1175–1191, Feb. 2020, doi: 10.1108/JEDT-10-2019-0281.
- [16] D. H. Hong and C.-H. Choi, "Multicriteria fuzzy decision-making problems based on vague set theory," *Fuzzy Sets and Systems*, vol. 114, no. 1, pp. 103–113, Aug. 2000, doi: 10.1016/S0165-0114(98)00271-1.
- [17] B. Vitoriano, J. Montero de Juan, and D. Ruan, *Decision Aid Models for Disaster Management and Emergencies*, vol. 7. Paris: Atlantis Press, 2013.
- [18] L. Taccari, "Integer programming formulations for the elementary shortest path problem," *European Journal of Operational Research*, vol. 252, no. 1, pp. 122–130, Jul. 2016, doi: 10.1016/j.ejor.2016.01.003.
- [19] S. Nahmias, "Fuzzy variables," *Fuzzy Sets and Systems*, vol. 1, no. 2, pp. 97–110, Apr. 1978, doi: 10.1016/0165-0114(78)90011-8.
- [20] George J. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*. Pearson, 1995.
- [21] R. Zhang, P. Yannis, and K. Vassilis, *Fuzzy Control of Queuing Systems*. London: Springer-Verlag, 2005.
- [22] Y.-M. Wang, J.-B. Yang, D.-L. Xu, and K.-S. Chin, "On the centroids of fuzzy numbers," *Fuzzy Sets and Systems*, vol. 157, no. 7, pp. 919–926, Apr. 2006, doi: 10.1016/j.fss.2005.11.006.
- [23] D. Bertsimas and M. Sim, "Robust discrete optimization and network flows," *Mathematical Programming*, vol. 98, no. 1–3, pp. 49–71, Sep. 2003, doi: 10.1007/s10107-003-0396-4.
- [24] T. Nehézová, "Robust optimization approach in travelling salesman problem," in *37th International Conference on Mathematical Methods in Economics 2019*, 2019, pp. 535–540, [Online]. Available: https://mme2019.ef.jcu.cz/files/conference_proceedings.pdf.
- [25] R. Hlavatý and H. Brožová, "Robust optimization approach in transportation problem," no. September, 2017.

The System Dynamics approach to creation of a recovery model of an urban object

Anna Selivanova^{1,2}

Abstract. A mathematical model of recovery of an urban object was proposed, using the System Dynamics methods. The model supposed contamination of surfaces of the object of interest only (or surface activities). Several artificial radionuclides were chosen, according to possible sources of contamination. Conditions in an early phase following environmental emergency releases, nuclear or radiation accidents were assumed.

The mathematical model of decontamination of the selected urban object was partially based on the results of the previous research. Hence, the new model included dosimetry calculations, as well as for the case of recovery of a large grassed area. In order to estimate the costs of each scenario and to simulate behavior of the model over time, the Vensim software was used. Obtained results were compared with available data from the recovery of affected areas after the Fukushima Daiichi Accident or the Chernobyl disaster.

Keywords: decontamination, System Dynamics, simulation, countermeasure, radiation/nuclear accident, emergency release, mathematical modelling, costs

JEL Classification: C63, Q51

AMS Classification: 90B99

1 Introduction

Following large-scale nuclear/radiation accidents, a broad range of miscellaneous decontamination techniques and countermeasures had been developed [10], [16]. The decontamination process of affected areas should be carried out from top surfaces to bottom owing to the prevention of recontamination. Hence, in case of inhabited areas, decontamination of buildings should be performed from roofs to ground surfaces like lawns and paved roads [16].

Based on experience of recovery strategies after nuclear/radiation accidents, some computer programs dedicated to countermeasure methods, e.g. the ERMIN package [4], had been created. However, in order to carry out detailed economical assessments, a mathematical model of recovery of a meadow was proposed for conditions of the Czech Republic, being presented at conferences [11]–[13]. Extending the existing model of the grassed area, a new sub-model of decontamination of parking lots was created. Hence, the modified overall model considered the recovery of the meadow and then recovery of paved areas around it [16], i.e. 4 parking areas altogether.

As well as for the grassed area, the System Dynamics approach was applied to decontamination of parking space. Therefore, a large number of variables and parameters originated in the grassed area recovery model was used in the new sub-model, connecting both sub-models into one complex. For example, initial surface activities for the recreation meadow were employed for parking lots, being corrected according to Andersson et al. [3]. Thereafter, the overall modelling and simulation process was performed in the Vensim software [18].

Considering the subsequent research, the whole mathematical model of recovery of both types of urban objects (the meadow and the adjacent parking areas) could be widened more in future, e.g. by adding decontamination scenarios for buildings or shrubs near the grassed area, using the System Dynamics methods, additional tests and simulations in the Vensim [18].

¹Faculty of Economics and Management, Czech University of Life Sciences Prague, Kamýcká 129, Prague, 165 00, Czech Republic, selivanova@pef.czu.cz;

²National Radiation Protection Institute (SÚRO), Bartoškova 1450/28, Prague, 140 00, Czech Republic.

2 Material and methods

2.1 System Dynamics

Employing the System Dynamics approach, miscellaneous variables and parameters could be mutually connected within difficult complex systems. These systems often demonstrate non-linear behavior over time with typical patterns, or “archetypes”, being observed in very different fields. Applying the System Dynamics methods, these known behavior patterns could be presented as feedback loops [14].

Considering surface deposition of radionuclides after nuclear/radiation accidents, decay of surface activities could be described as a negative (balancing) feedback loop, being an exponential time-dependent process [2], [12]. On the contrary, an exponential growth (e.g. cumulated doses in the investigated case), corresponds to a positive (self-reinforcing) feedback loop [12], [14].

In order to present an integration process of time-dependent variables over selected time intervals, stock and flow diagrams (SFD) were employed, using the Vensim software [18]. For instance, within the modelling, surface activities were presented as stock variables, or boxes in SFD (i.e. definite integrals from mathematical point of view), while rates of their decrease corresponded to flows, or pipes in SFD [12], [14].

2.2 Physics description

The newly created sub-model of decontamination of parking space was linked to the recovery model of the meadow [12], [13]. As well as in the meadow recovery model, ^{134}Cs and ^{137}Cs with the total initial surface activity of 2 MBq m^{-2} , were anticipated. However, according to [3], overall initial surface deposition on asphalt surfaces was equal to half of initial activities on the grassed area, or 1 MBq m^{-2} altogether.

For the baseline scenario presented in this paper, the parking lot area was $56 \times 64 \text{ m}^2$. In calculations, 4 similar parking areas around the meadow were considered. A number of irradiated persons was not changed, being equal to 631 persons [12]. Nevertheless, a perimeter of the whole area, including the meadow and parking space, was extended to 1,856 m.

According to [3], the decline of ^{137}Cs on streets and pavements could be described as a sum of 2 exponential components, corresponding to a fast and slower decrease:

$$A = 0.5 A_{0,grass} e^{-\lambda_r t} (0.7 e^{-\lambda_{w1} t} + 0.3 e^{-\lambda_{w2} t}), \quad (1)$$

where $A_{0,grass}$ is the initial surface activity on the reference grassed area, λ_r is the decay constant, λ_{w1} and λ_{w2} are weathering rates corresponding to half-lives of 120 days and 3 years. However, in the case of decontamination, an additional decontamination rate should be added to the equation [1].

2.3 Radiation impacts

In order to assess radiation impacts on the group of irradiated inhabitants, recalculation of surface activities to dose rates was required, using conversion factors, in accordance with [9]. Afterwards, the modified equation (1) was integrated over the chosen time interval (i.e. 1 year after the overall decontamination process), considering time spent indoor/outdoor and shielding factors of buildings [12]. Hence, based on initial surface activities, effective doses were obtained and then recalculated to collective effective doses. Then, collective doses were converted to a financial equivalent, using a coefficient of 2.5 mln. CZK/Sv [15]. Thereafter, benefits were equal to a financial expression of collective effective doses saved up after the implementation of the countermeasures [12], [15].

2.4 Monte Carlo simulations

Conversion factors mentioned in the paragraph above had been already obtained for diverse surfaces and geometries using Monte Carlo simulations [9]. Nevertheless, published conversion factors significantly differed, depending on surroundings around the detection element. Therefore, due to the actual location of parking lots, own simulations were performed, using the radiation transport Monte Carlo code, MCNP6.1 [5]. In simulations, an asphalt area with real dimensions of parking space was employed. The area was covered by surface planar sources of ^{137}Cs and ^{134}Cs , supposing fresh deposition with a relaxation depth of 1 mm [17]. The detection element, “tally”, was placed at 1 m above the source center. Based on the simulation results, conversion factors of air-kerma rate/surface activity for both radionuclides were obtained. Afterwards, simulated factors were additionally recalculated to effective dose rate/surface activity, using tables in [7].

2.5 Countermeasures

According to [10] or [16], such countermeasures like firehosing, high-pressure water cleaning, surface removal, vacuum sweeping and others could be applicable to paved surfaces and roads. Within the modelling, the high-pressure hosing technique (14 MPa) was selected due to its effectiveness and parameters of the site of interest (parking space). Nevertheless, in future research, other scenarios of decontamination of paved areas will be implemented, considering more detailed conditions of the site. In the overall model, decontamination of the meadow was expected as the first step, being followed by decontamination of parking lots.

2.6 Total costs

Supposing high-pressure hosing, total costs included labor costs (500 CZK h⁻¹), costs of consumables (water and fuel), costs of personal protective equipment (tychems, gloves, boots, etc.), costs of decontamination of workers and vehicles (water) and consumption of fixed capital (pressure washers). A number of workers was equal to 16, considering 2 persons per 1 team [10]. For water and fuel, cost rates were set to 50 CZK m⁻³, resp. 35 CZK l⁻¹.

Additionally, assessments of total costs of recovery of the overall area in the model (grassed meadow and parking space) were carried out. The meadow sub-model was slightly modified for uniformity purposes, however, changes in the final costs of its recovery were not significant. For instance, processes of waste handling were improved [11]. Also, the removal of demarcation after finishing of countermeasures implementation was added with a total cost of roughly 20 thsd. CZK. Afterwards, total costs of both sub-models were summed up and total recovery scenarios for the anticipated area were mutually compared.

3 Results and discussion

3.1 Sub-model of parking space

The sub-model of parking lots included 7 layers with stock and flow diagrams and basic parameters of parking space, cost assessments, dose calculations and a summary of results. The SFD dedicated to the recalculation of the ¹³⁷Cs surface activity to the effective dose could be found in Figure 1. Owing to the weathering process, the surface activity was described as 2 stock variables (boxes), corresponding to the longer and slower component of exponential decline according to equation (1). Thereafter, using additional stock variables (“cumulated decay”), activities were integrated and recalculated to external effective doses. For ¹³⁴Cs, the same approach was employed.

In accordance with the SFD (Figure 1), the decrease of the surface activity was caused by 2 types of processes: natural (radioactive decay and weathering) and artificial (decontamination). Hence, natural processes were joined together as the right outflow, while the left outflow contained the decontamination rate (high-pressure washing) only. As mentioned earlier, the sub-model of the parking space recovery was connected with the meadow sub-model via several common parameters and variables, i.e. initial surface activities, decay constants or costs of items required for decontamination. Hence, variables/parameters used earlier were depicted as titles with angle brackets.

Based on the Monte Carlo simulations, conversion factors were equal to 1.12E-9 mSv h⁻¹ per Bq m⁻² and 2.52E-9 mSv h⁻¹ per Bq m⁻² for ¹³⁷Cs, resp. ¹³⁴Cs. Relative standard deviations of simulations were ≤ 0.11 %. Both factors were added to the SFDs for ¹³⁷Cs and ¹³⁴Cs to convert surface activities to effective doses.

Figure 2 represents the sum of effective doses from ¹³⁷Cs and ¹³⁴Cs deposited on surfaces of parking lots. Owing to the overall model mathematical structure, decontamination of parking areas started after the implementation of the meadow countermeasures [12]. Afterwards, actually 4 scenarios were calculated: demarcation (without any decontamination), grass removal, soil stripping and turf harvesting for the meadow, having been followed by the high-pressure washing of parking lots. The summary of annual effective doses for the meadow and the parking space could be found in Table 1.

For the scenario without any decontamination method, the highest annual effective dose obtained on parking lots did not exceed 2.3 mSv (Table 1), being low due to the rapid decline component of cesium on paved surfaces, while for the meadow annual doses were roughly up to 10.6 mSv [12]. However, according to [3], in the case of parking lots, the weathering process could be less significant and the surface contamination could remain for longer periods. Therefore, corresponding parameters of the sub-model **could be anytime changed** in future, according to actual requirements. For the meadow and the overall site, annual effective doses were > 1 mSv (except the soil stripping) and potentially required additional countermeasures.

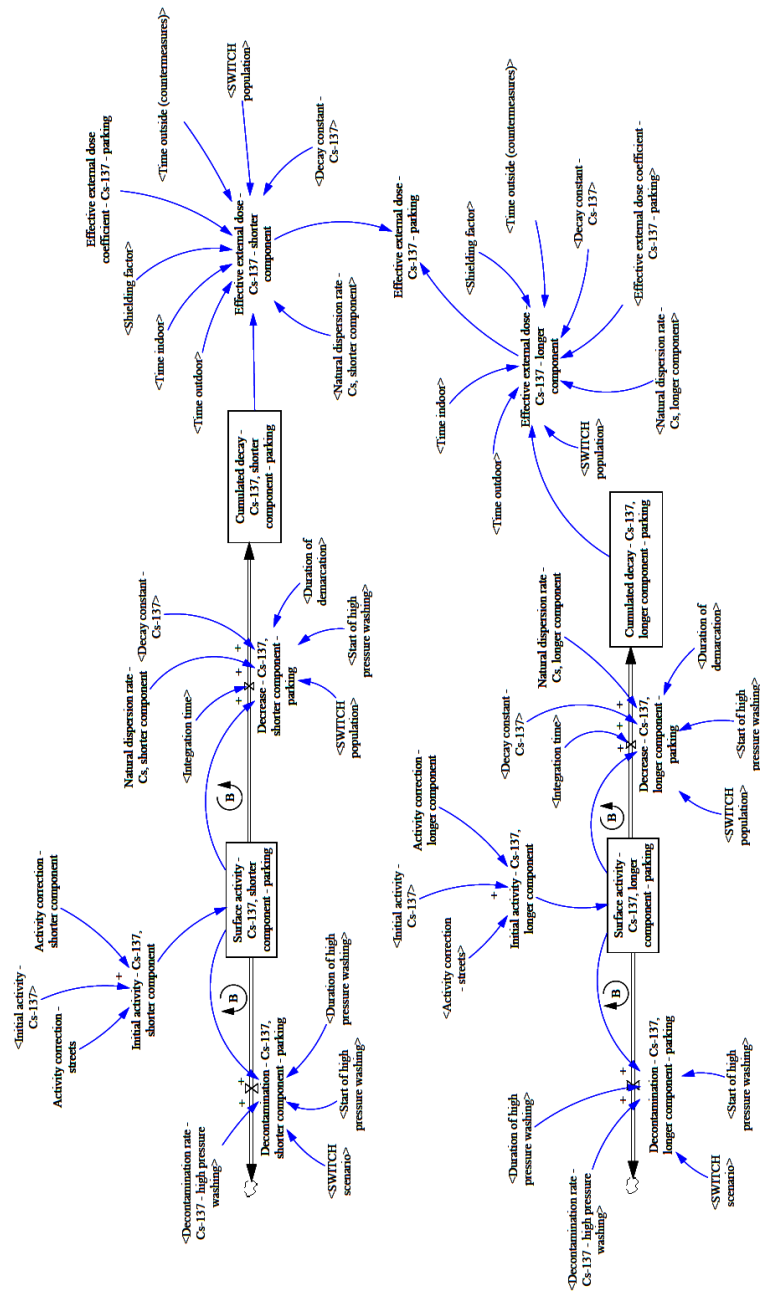


Figure 1 Stock and Flow Diagram of Dose estimation for ¹³⁷Cs deposited on parking space surfaces

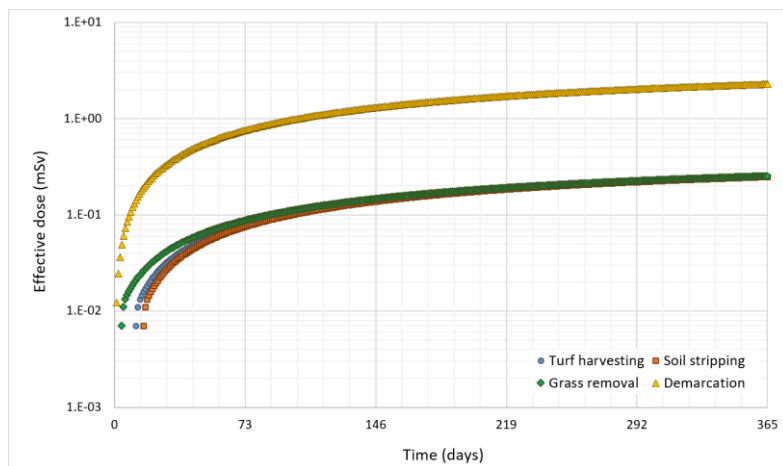


Figure 2 Total annual effective dose for parking lots

Object/Annual effective dose	Demarcation (mSv)	Grass removal (mSv)	Soil stripping (mSv)	Turf harvesting (mSv)
Meadow	10.6	4.1	0.3	1.3
Parking space	2.3		0.3	
Overall site	12.9	4.3	0.6	1.5

Table 1 Annual effective dose from external irradiation by ^{137}Cs and ^{134}Cs deposited on surfaces of the meadow and parking lots, considering 4 scenarios and parameters of the overall model of decontamination

3.2 Cost-Benefit Analysis

The Cost-Benefit Analysis (CBA) is based on the comparison of financial expression of effective doses obtained by the representative persons (Table 1) and costs of recovery [8]. Simulated costs of high-pressure washing could be found in Table 2. Owing to completeness, costs of the meadow decontamination and the overall site (with the subsequent demarcation removal) were added. For parking areas, the recovery cost was roughly 0.6 mln. CZK, while costs for the meadow were in a range of 0.7–3.0 mln. CZK, depending on the countermeasure technique. Considering the overall site, costs were in an interval of 0.7–3.7 mln. CZK.

For the grassed area, costs per 1 m² described in [12] were in a good agreement with the Fukushima clean-up [16]. In case of parking lots, the cost of high-pressure washing per 1 m² was 44 CZK m⁻². In the report [16], a large interval of costs of pressure washing could be found: 150–1,150 JPY m⁻², or 34–276 CZK m⁻² (corrected to 2020). Therefore, the simulated cost of high-pressure washing corresponded with the actual recovery costs in general.

Benefits for the same objects and scenarios could be found in Table 3. Zero benefits corresponded to the reference scenario, or the demarcation only [12]. In comparison with benefits for the meadow recovery (roughly 10–15 mln. CZK), benefits for the parking areas were several times lower (3 mln. CZK), originating in the fast decline of cesium isotopes considered in the physics part of the model (and ideal conditions). However, comparing benefits and costs for the parking space separately, the decontamination of parking lots seemed to be still acceptable.

Nevertheless, the model assumed ^{137}Cs and ^{134}Cs only (due to large experience and many related publications), while an actual number of released radionuclides after radiation/nuclear accidents is expected to be **substantially** higher, including e.g. ^{90}Sr with a half-life of 29 years or hot particles with extremely high activities [6]. Therefore, decontamination of all objects will be certainly required in any case. However, considering the overall site, the total benefits of decontamination of the meadow and the parking space altogether were significantly higher than the total costs of the joint recovery. Hence, the decontamination process of affected areas should be carried out as a complex process, where countermeasures applied to different objects could have significantly higher positive impacts and will be cheaper than separate methods themselves.

Object/Scenario costs	Demarcation (CZK)	Grass removal (CZK)	Soil stripping (CZK)	Turf harvesting (CZK)
Meadow	0.7 mln.	0.8 mln.	4.4 mln.	3.0 mln.
Parking space	0		0.6 mln.	
Overall site	0.7 mln.	1.5 mln.	5.1 mln.	3.7 mln.

Table 2 Costs of decontamination of the meadow, parking space and the overall site (incl. demarcation removal)

Object/Scenario benefits	Demarcation (CZK)	Grass removal (CZK)	Soil stripping (CZK)	Turf harvesting (CZK)
Meadow	0	10.3 mln.	16.3 mln.	14.8 mln.
Parking space	0		3.2 mln.	
Overall site	0	13.5 mln.	19.5 mln.	18.0 mln.

Table 3 Benefits of decontamination of the meadow, parking space and the overall site

4 Conclusion

Employing the System Dynamics approach, the sub-model of parking lots decontamination was proposed, extending the existing model of the meadow recovery. Simulated recovery costs were in a good agreement with the Fukushima experience. Based on simulations, decontamination of the whole affected area led to higher overall benefits than separated decontamination of parking lots. Therefore, the model will be additionally extended to other urban objects in future research.

Acknowledgements

The research was supported by the project 2021B0003 “System approach to cost-benefit analysis of recovery of habitation and adjacent areas after a nuclear or radiation accident” of The Internal Grant Agency of Faculty of Economics and Management CZU Prague (IGA FEM).

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] Albin, S. (2001). *Generic Structures: First-Order Linear Positive Feedback*. static.clexchange.org/ftp/documents/roadmaps/RM4/D-4475-2.pdf. [1 March 2021].
- [3] Andersson, K. G., Roed, J., Fogh, C. L. (2002). Weathering of radiocaesium contamination on urban streets, walls and roofs. *Journal of Environmental Radioactivity*, 62(1), 49–60.
- [4] Charnock, T., Landman, C., Trybushnyi, D., Ievdin, I. (2016). European model for inhabited areas – ERMIN 2. *Radioprotection*, 51(HS1), S23–S25.
- [5] Goorley, T., James, M., Booth, T., Brown, F., Bull, J., Cox, L. J., Durkee, J., Elson, J., Fensin et al. (2012). Initial MCNP6 release overview. *Nuclear Technology*.
- [6] IAEA. (2006). Environmental Consequences of the Chernobyl Accident and their Remediation: Twenty Years of Experience. *Report of the Chernobyl Forum Expert Group ‘Environment.’* Vienna, International Atomic Energy Agency.
- [7] ICRP. (1997). *ICRP Publication 74. Conversion Coefficients for use in Radiological Protection against External Radiation* (Annals of the ICRP). Pergamon.
- [8] ICRP. (2006). *ICRP Publication 101. Assessing Dose of the Representative Person for the Purpose of Radiation Protection of the Public and The Optimisation of Radiological Protection: Broadening the Process* (Annals of the ICRP). Elsevier Ltd.
- [9] Meckbach, R., Jacob, P., Paretzke, H. G. (1988). Gamma Exposures due to Radionuclides Deposited in Urban Environments. Part II: Location Factors for Different Deposition Patterns. *Radiation Protection Dosimetry*.
- [10] Nisbet, A., Brown, J., Cabianca, T., Jones, A. L., Andersson, K. G., Hänninen, R., Ikäheimonen et al. (2010). *Generic handbook for assisting in the management of contaminated inhabited areas in Europe following a radiological emergency* (Version 2).
- [11] Selivanova, A. (2020a). Creation of a Model of Waste Handling Within Recovery After a Nuclear Accident Using the System Dynamics Approach. *ICRP International Conference on Recovery After Nuclear Accidents. Radiological Protection Lessons from Fukushima and Beyond, 1-4.12.2020, on-line*.
- [12] Selivanova, A. (2020b). Sensitivity Analysis of a System Dynamics Decontamination Model. *38th International Conference on Mathematical Methods in Economics (MME 2020) – Conference Proceedings. 9-11.9.2020, Brno, Czech Republic, 523–529*. Mendel University in Brno.
- [13] Selivanova, A., Krejčí, I. (2019). Simulations of decontamination scenarios using the system dynamics approach. *5th NERIS Workshop Proceedings – Key Challenges in the Preparedness, Response and Recovery Phase of a Nuclear or Radiological Emergency. 3-5 April 2019, Roskilde, Denmark, 38–43*.
- [14] Sterman, J. D. (2000). *Business dynamics: systems thinking and modeling for a complex world*. Irwin/McGraw-Hill.
- [15] SÚJB. (2016). *Implementing decree 422/2016 Coll., on Radiation Protection and Security of a Radioactive Source*. Prague, State Office for Nuclear Safety.
- [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] UNSCEAR. (2000). *Sources and Effects of Ionizing Radiation. Official Records of the General Assembly UNSCEAR 2000 Report to the General Assembly, with Scientific Annexes. ANNEX A – Dose Assessment Methodologies.: Vol. I*. United Nations.
- [18] Ventana Systems. (2015). Vensim. *Vensim software*.

Fuzzy discount factor parametrized by logarithmic return rate

Joanna Siwek¹, Krzysztof Piasecki²

Abstract. Investing in financial markets, especially in the highly variable HFT markets, is burdened with risk connected with the quality of information. Expert systems supporting the decision-making process are prone to imprecision that may burden the data. An answer to this problem is modeling the data with attention given not only to the risk of variable future value but also with an emphasis on the imprecision stemming from subjectiveness, used tools and methods, delays, and behavioral aspects. A known method for minimization of the influence of these negative aspects is modeling the present value with fuzzy numbers, while at the same time considering the imprecision by appointing fuzzy discount factors. In the existing literature, we can find solutions that are not very general and require an assumption about simple return rates and their normal probability distribution. The following article presents a more universal model, proving the possible applications of the fuzzy discount factor for a more general case, as well as the case of logarithmic fuzzy return rates.

Keywords: discount factor, logarithmic return rate, fuzzy number

JEL Classification: C44, G11

AMS Classification : 03E72

1. Introduction

Imprecision of financial data is usually modelled by fuzzy numbers (FNs). In this paper we will consider a case when a financial asset is tentatively evaluated by a present value given by a trapezoidal FN (TrFN). If performed financial analysis is based on a simple return rate, then in case of the data being imprecise, the expected discount factor is visibly more convenient portfolio analysis tool than the expected return rate [3, 4]. On the other hand, often the financial analysis is carried out using logarithmic return rates, which stand as an alternative to the simple ones.

For these reasons, the main aim of our paper is to study the discount factor related to logarithmic return rate for the case when the financial asset is evaluated by TrFNs.

2. Trapezoidal fuzzy numbers - basic facts

The symbol $\mathcal{F}(\mathbb{R})$ denotes the family of all fuzzy subsets in the real line \mathbb{R} . A fuzzy number (FN) is usually defined as a fuzzy subset of the real line \mathbb{R} . The most general definition of FN was formulated by Dubois and Prade [1]. The set of all FN we denote by the symbol \mathbb{F} . Any FN may be represented in the following way.

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 $\mu_{\mathcal{L}}(\cdot | a, b, c, d, L_L, R_L) \in [0, 1]^{\mathbb{R}}$ described by the identity:

$$\mu_{\mathcal{L}}(x | a, b, c, d, L_L, R_L) = \begin{cases} 0, & x \notin [a, d], \\ L_L(x), & x \in [a, b[, \\ 1, & x \in [b, c], \\ R_L(x), & x \in]c, d], \end{cases} \quad (1)$$

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 and meeting the condition:

¹ Adam Mickiewicz University, Department of Artificial Intelligence, Faculty of Mathematics and Computer Science Uniwersytetu Poznańskiego 4, 61-614 Poznań, Poland, jsiwiek@amu.edu.pl

² WSB University in Poznań, Institute of Economy and Finance, ul. Powstańców Wielkopolskich 5, 61-895 Poznań, Poland; krzysztof.piasecki@wsb.poznan.pl.

$$\forall_{x \in]a, d[}: \mu_L(x|a, b, c, d, L_L, R_L) > 0. \quad (2)$$

Due to the high complexity of arithmetic operations of FN, in many practical applications researchers limit the use of FNs only to trapezoidal FN (TrFN) defined below.

Definition 1. [5] For any non-decreasing sequence $(a, b, c, d) \subset \mathbb{R}$, TrFN $Tr(a, b, c, d) = \mathcal{T}$ is FN $\mathcal{T} \in \mathbb{F}$ determined explicitly by its membership functions $\mu_{\mathcal{T}} \in [0, 1]^{\mathbb{R}}$ as follows

$$\mu_{\mathcal{T}}(x) = \mu_{Tr}(x|a, b, c, d) = \begin{cases} 0, & x \notin [a, d], \\ \frac{x-a}{b-a}, & x \in [a, b], \\ 1, & x \in [b, c], \\ \frac{x-d}{c-d}, & x \in]c, d]. \end{cases} \quad (3)$$

The symbol \mathbb{F}_{Tr} denotes the space of all TrFNs. Any TrFN $Tr(a, b, b, d)$ is called triangular FN. Let us note that any FN is TrFN if and only if its reference functions are linear.

Let symbol $*$ denote any arithmetic operation defined in \mathbb{R} . By \odot we denote an extension of arithmetic operation $*$. In line with the Zadeh Extension Principle [6] we can extend basic arithmetic operators to the case of \mathbb{F}_{Tr} , in such a way that for any pair $(Tr(a, b, c, d), Tr(e, f, g, h)) \in \mathbb{F}_{Tr}^2$ and $\beta \in \mathbb{R}$, arithmetic operations of extended sum \oplus and dot product \odot are determined as follows:

$$Tr(a, b, c, d) \oplus Tr(e, f, g, h) = Tr(a + e, b + f, c + g, d + h), \quad (4)$$

$$\beta \odot Tr(a, b, c, d) = \begin{cases} Tr(\beta \cdot a, \beta \cdot b, \beta \cdot c, \beta \cdot d) & \beta \geq 0, \\ Tr(\beta \cdot d, \beta \cdot c, \beta \cdot b, \beta \cdot a) & \beta < 0. \end{cases} \quad (5)$$

Any monotonic unary operator $G: \mathbb{R} \supset \mathbb{A} \rightarrow \mathbb{R}$ may be extended to TrFN case. The Zadeh Extension Principle implies that an extended unary operator $G: \mathbb{F}_{Tr} \supset \mathbb{H} \rightarrow \mathbb{F}$ is given by:

- for increasing operator

$$G(Tr(a, b, c, d)) = \mathcal{L}(G(a), G(b), G(c), G(d), L_L, R_L), \quad (6)$$

- for decreasing operator

$$G(Tr(a, b, c, d)) = \mathcal{L}(G(d), G(c), G(b), G(a), R_L, L_L), \quad (7)$$

where the reference functions are given by formulas

$$\forall y \in [G(a), G(b)[\quad L_L(y) = \frac{G^{-1}(y) - a}{b - a}, \quad (8)$$

$$\forall y \in]G(c), G(d)] \quad R_L(y) = \frac{G^{-1}(y) - d}{c - d}. \quad (9)$$

In this article we use the notation with $[a, b[$ interval, but it may be equivalently replaced with the interval $]a, b]$.

3. Present value

The starting point for our considerations is a notion of present value (PV), defined as a current equivalent of a future cash flow. PV understood this way may be imprecise [e.g. 3]. The natural consequence of this conclusion is estimating PV with FNs. In [4], the evolution of the fuzzy PV model is described in detail. In general, fuzzy PV is characterized by a non-decreasing sequence $(V_s, V_f, \check{P}, V_l, V_e)$, where:

- \check{P} is the quoted price,
- $[V_s, V_e] \subset \mathbb{R}^+$ is an interval of all possible values of PV,
- $[V_f, V_l] \subset [V_s, V_e]$ is an interval of all prices which do not noticeably differ from a quoted price \check{P} .

Due to the high complexity of arithmetic operations of FN, we estimate PV by TrFN

$$\bar{P}\check{V} = Tr(V_s, V_f, V_l, V_e). \quad (10)$$

4. Expected discount factor

Let us assume that the time horizon $t > 0$ of an investment is fixed. Then, the considered asset is determined by two values:

- anticipated future value (FV) V_t ,
- assessed PV V_0 .

The basic characteristic of benefits that come from owning this asset is a simple return rate r_t given by the identity:

$$r_t = \ln \frac{V_t}{V_0}. \quad (11)$$

In [4], it is justified that FV is a random variable $\tilde{V}_t: \Omega \rightarrow \mathbb{R}^+$. The set Ω is a set of elementary states ω , of the financial market. In a classical approach to a return rate estimation, PV is identified with the observed quoted price \check{P} . Thus, the return rate is a random variable determined by identity:

$$r_t(\omega) = \ln \frac{\tilde{V}_t(\omega)}{\check{P}}. \quad (12)$$

Uncertainty risk is a result of a lack of knowledge about the future state of affairs. In the practice of financial markets analysis, the uncertainty risk is usually described by the probability distribution of return rate (12) which may be given by a cumulative distribution function $F_r(\cdot | \bar{r}): \mathbb{R} \rightarrow [0,1]$. We assume that the expected value, \bar{r} , of this distribution exists. Then also the expected discount factor (EDF) \bar{v} exists and is determined by the dependency

$$\bar{v} = e^{-\bar{r}}. \quad (13)$$

If we take together (11) and (12), then we obtain the following formula describing the return rate

$$r_t = r_t(V_0, \omega) = \ln \frac{\check{P} \cdot e^{r_t(\omega)}}{V_0}. \quad (14)$$

It implies that the expected return rate may be expressed in the following way

$$\mathcal{R}(V_0) = \int_{-\infty}^{+\infty} \ln \frac{\check{P} \cdot e^y}{V_0} d F_r(y | \bar{r}) = \ln \frac{\check{P} \cdot e^{\bar{r}}}{V_0}. \quad (15)$$

This way we express return rate as a decreasing unary operator $\mathcal{R}: \mathbb{R}^+ \rightarrow \mathbb{R}$ that transforms the PV. If PV is imprecisely estimated by a TrFN (10), then in line with (7), (8), and (9), the imprecise logarithmic return rate is given as FN

$$\mathcal{R} = \mathcal{R}(\check{P}\bar{V}) = \mathcal{R}\left(\text{Tr}(V_s, V_f, V_l, V_e)\right) = \mathcal{L}\left(\ln \frac{\check{P} \cdot e^{\bar{r}}}{V_e}, \ln \frac{\check{P} \cdot e^{\bar{r}}}{V_l}, \ln \frac{\check{P} \cdot e^{\bar{r}}}{V_f}, \ln \frac{\check{P} \cdot e^{\bar{r}}}{V_s}, R_L, L_L\right). \quad (16)$$

where the reference functions are given by formulas

$$\forall y \in \left[\ln \frac{\check{P} \cdot e^{\bar{r}}}{V_s}, \ln \frac{\check{P} \cdot e^{\bar{r}}}{V_f} \right] \quad L_L(r) = \frac{\check{P} \cdot e^{\bar{r}-r} - V_s}{V_f - V_s}, \quad (17)$$

$$\forall y \in \left[\ln \frac{\check{P} \cdot e^{\bar{r}}}{V_l}, \ln \frac{\check{P} \cdot e^{\bar{r}}}{V_e} \right] \quad R_L(r) = \frac{\check{P} \cdot e^{\bar{r}-r} - V_e}{V_l - V_e}. \quad (18)$$

We see that the above reference functions are not linear. It implies that FN (16) is not a TrFN. Therefore, using logarithmic return rates does not allow for bypassing the complexity of arithmetic operations on FN.

In the next step, we consider the discount factor v_t determined by a logarithmic return rate r_t . From (11), we get

$$v_t = e^{-r_t}. \quad (19)$$

By combining together (19) and (12) we obtain the following random variable describing the discount factor

$$v_t = r_t(V_0, \omega) = \frac{V_0 \cdot e^{-r_t(\omega)}}{\check{P}}. \quad (20)$$

It implies that the EDF may be expressed by

$$\mathcal{V}(V_0) = \int_{-\infty}^{+\infty} \frac{V_0 \cdot e^{-y}}{\check{P}} d F_r(y | \bar{r}) = \frac{V_0 \cdot e^{-\bar{r}}}{\check{P}} = \frac{V_0 \cdot \bar{v}}{\check{P}}. \quad (21)$$

Thus, we can determine the imprecise EDF $\mathcal{V}: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ as an increasing operator transforming PV. If PV is imprecisely estimated by a TrFN (10) then in line with (6), (8), and (9), the imprecise EDF is also given as FN

$$\mathcal{V} = \mathcal{V}(\bar{P}\bar{V}) = \mathcal{V}\left(\text{Tr}(V_s, V_f, V_l, V_e)\right) = \mathcal{L}\left(\frac{V_s \cdot \bar{v}}{\bar{p}}, \frac{V_f \cdot \bar{v}}{\bar{p}}, \frac{V_l \cdot \bar{v}}{\bar{p}}, \frac{V_e \cdot \bar{v}}{\bar{p}}, L_L, R_L\right). \quad (22)$$

where the reference functions are given by formulas

$$\forall y \in \left[\frac{V_s \cdot \bar{v}}{\bar{p}}, \frac{V_f \cdot \bar{v}}{\bar{p}}\right] \quad L_L(v) = \frac{\frac{v \cdot \bar{p}}{\bar{v}} - V_s}{V_f - V_s} = \frac{\bar{p} \cdot v - V_s \cdot \bar{v}}{\bar{v} \cdot (V_f - V_s)}, \quad (23)$$

$$\forall y \in \left[\frac{V_l \cdot \bar{v}}{\bar{p}}, \frac{V_e \cdot \bar{v}}{\bar{p}}\right] \quad R_L(v) = \frac{\frac{v \cdot \bar{p}}{\bar{v}} - V_e}{V_l - V_e} = \frac{\bar{p} \cdot v - V_e \cdot \bar{v}}{\bar{v} \cdot (V_l - V_e)}. \quad (24)$$

We can see that both reference functions stated above are linear. This implies that if imprecise PV is given by a TrFN (10) then imprecise EDF \mathcal{V} is also a TrFN

$$\mathcal{V} = \mathcal{V}(\bar{P}\bar{V}) = \text{Tr}\left(\frac{V_s \cdot \bar{v}}{\bar{p}}, \frac{V_f \cdot \bar{v}}{\bar{p}}, \frac{V_l \cdot \bar{v}}{\bar{p}}, \frac{V_e \cdot \bar{v}}{\bar{p}}\right). \quad (25)$$

The presented explanation suggests that in case of financial analysis, imprecise EDF is more useful than an imprecise expected return rate. When using EDF, the criterion of maximizing expected return rate is replaced by one of minimizing the EDF.

5. Expected discount factor for a portfolio

By a financial portfolio we will understand an arbitrary, finite set of assets. Any asset can be determined as fixed security in the long position. On the other hand, a portfolio is also a security. Let us consider the case of a multi-asset portfolio π^* , built of securities Y_i . We can describe this portfolio as the set $\pi^* = \{Y_i: i = 1, 2, \dots, n\}$. Any security Y_i is characterized by its price $\bar{P}_i \in \mathbb{R}^+$, and by its imprecise PV evaluated with TrFN

$$\bar{P}\bar{V}_i = \text{Tr}(V_s^{(i)}, V_f^{(i)}, V_l^{(i)}, V_e^{(i)}), \quad (26)$$

and by its EDF \bar{v}_i determined by (13). Taking into account all the above, we evaluate any security Y_i by its imprecise EDF

$$\mathcal{V}_i = \text{Tr}\left(V_s^{(i)} \cdot \frac{\bar{v}_i}{\bar{p}_i}, V_f^{(i)} \cdot \frac{\bar{v}_i}{\bar{p}_i}, V_l^{(i)} \cdot \frac{\bar{v}_i}{\bar{p}_i}, V_e^{(i)} \cdot \frac{\bar{v}_i}{\bar{p}_i}\right). \quad (27)$$

Let imprecise PV of the portfolio π^* be denoted by the symbol $\bar{P}\bar{V}^*$. A portfolio PV is always equal to the sum of its components' PVs. Therefore, using (4) we calculate the PV of the portfolio π^* in the following way

$$\bar{P}\bar{V}^* = \text{Tr}(V_s^{(*)}, V_f^{(*)}, V_l^{(*)}, V_e^{(*)}) = \text{Tr}\left(\sum_{i=1}^n V_s^{(i)}, \sum_{i=1}^n V_f^{(i)}, \sum_{i=1}^n V_l^{(i)}, \sum_{i=1}^n V_e^{(i)}\right). \quad (28)$$

Next we calculate the EDFs of considered portfolios. The value of a portfolio π^* is calculated as follows

$$M^* = \sum_{i=1}^n \bar{P}_i. \quad (29)$$

The shares p_i of the asset Y_i in the portfolio π^* are given by the formula

$$p_i = \frac{\bar{P}_i}{M^*}. \quad (30)$$

The EDF \bar{v}^* of a portfolio π^* is given by

$$\bar{v}^* = \left(\sum_{i=1}^n \frac{p_i}{\bar{v}_i}\right)^{-1}. \quad (31)$$

Finally, using (25), (28), (29), (30), and (31) we calculate the imprecise EDF \mathcal{V}^* of a portfolio π^* . We get

$$\mathcal{V}^* = \text{Tr}\left(V_s^{(*)} \cdot \frac{\bar{v}^*}{M^*}, V_f^{(*)} \cdot \frac{\bar{v}^*}{M^*}, V_l^{(*)} \cdot \frac{\bar{v}^*}{M^*}, V_e^{(*)} \cdot \frac{\bar{v}^*}{M^*}\right). \quad (32)$$

On the other hand, we have the following important theorem

Theorem 2: If the discount factor is determined by a logarithmic return rate, then for any portfolio π^* we have

$$\mathcal{V}^* = \oplus_{i=1}^n \left(\left(\left(\sum_{i=1}^n \frac{p_i}{\bar{v}_i} \right)^{-1} \cdot \frac{p_i}{\bar{v}_i} \right) \odot \mathcal{V}_i \right). \quad (33)$$

Proof: Sequentially from (28), (30), (5), (4), (5), (27), and (31) we get

$$\begin{aligned} \mathcal{V}^* &= Tr \left(V_s^{(*)} \cdot \frac{\bar{v}^*}{M^*}, V_f^{(*)} \cdot \frac{\bar{v}^*}{M^*}, V_l^{(*)} \cdot \frac{\bar{v}^*}{M^*}, V_e^{(*)} \cdot \frac{\bar{v}^*}{M^*} \right) \\ &= Tr \left(\sum_{i=1}^n V_s^{(i)} \cdot \frac{\bar{v}^*}{M^*}, \sum_{i=1}^n V_f^{(i)} \cdot \frac{\bar{v}^*}{M^*}, \sum_{i=1}^n V_l^{(i)} \cdot \frac{\bar{v}^*}{M^*}, \sum_{i=1}^n V_e^{(i)} \cdot \frac{\bar{v}^*}{M^*} \right) \\ &= \bar{v}^* \odot Tr \left(\sum_{i=1}^n p_i \cdot \frac{V_s^{(i)}}{\check{P}_i}, \sum_{i=1}^n p_i \cdot \frac{V_f^{(i)}}{\check{P}_i}, \sum_{i=1}^n p_i \cdot \frac{V_l^{(i)}}{\check{P}_i}, \sum_{i=1}^n p_i \cdot \frac{V_e^{(i)}}{\check{P}_i} \right) \\ &= \bar{v}^* \odot \left(\oplus_{i=1}^n Tr \left(p_i \cdot \frac{V_s^{(i)}}{\check{P}_i}, p_i \cdot \frac{V_f^{(i)}}{\check{P}_i}, p_i \cdot \frac{V_l^{(i)}}{\check{P}_i}, p_i \cdot \frac{V_e^{(i)}}{\check{P}_i} \right) \right) \\ &= \bar{v}^* \odot \left(\oplus_{i=1}^n \left(\frac{p_i}{\bar{v}_i} \odot Tr \left(\bar{v}_i \cdot \frac{V_s^{(i)}}{\check{P}_i}, \bar{v}_i \cdot \frac{V_f^{(i)}}{\check{P}_i}, \bar{v}_i \cdot \frac{V_l^{(i)}}{\check{P}_i}, \bar{v}_i \cdot \frac{V_e^{(i)}}{\check{P}_i} \right) \right) \right) \\ &= \oplus_{i=1}^n \left(\left(\bar{v}^* \cdot \frac{p_i}{\bar{v}_i} \right) \odot \mathcal{V}_i \right) = \oplus_{i=1}^n \left(\left(\left(\sum_{i=1}^n \frac{p_i}{\bar{v}_i} \right)^{-1} \cdot \frac{p_i}{\bar{v}_i} \right) \odot \mathcal{V}_i \right). \end{aligned} \quad (34)$$

■

Above we have proved that the imprecise EDF of a portfolio is a linear combination of EDFs of their components. By showing this, we have justified the suitability of using EDF in portfolio analysis.

5. Final remarks

Obtained results may provide theoretical foundations for portfolio analysis of securities described with TrFNs. If applied, the criterion of maximization expected return rate is replaced by the criterion of minimizing the expected discount factor.

The obtained results may as well be useful for future research on the impact of the PV imprecision on portfolio analysis. It may also be beneficial to generalize the above stated results to the case of an arbitrary return rate, understood as an increasing function of the future value, decreasing function of the PV and is independent of the FV and PV measurement scale.

References

- [1] Dubois D., Prade H., (1978), *Operations on fuzzy numbers*, International Journal of System Science, vol. 9, pp. 613-629.
- [2] Delgado M., Vila M. A., Voxman W., (1998), *On a canonical representation of fuzzy numbers*, Fuzzy Sets and Systems, vol. 93, iss. 1, pp. 125-135.
- [3] Piasecki K., Siwek J., (2018), Multi-asset portfolio with trapezoidal fuzzy present values, *Przegląd Statystyczny*, vol. LXXV, iss. 1.
- [4] Piasecki K., Siwek J., (2018), Two-Asset Portfolio with Triangular Fuzzy Present Values—An Alternative Approach, in *Contemporary Trends in Accounting, Finance and Financial Institutions : Proceedings from the International Conference on Accounting, Finance and Financial Institutions (ICAFFI), Poznan 2016*, pp. 11-26.
- [5] Shyi-Ming C., (1994), Fuzzy system reliability analysis using fuzzy number arithmetic operations, *Fuzzy Sets and Systems*, vol. 64, pp. 31-38.
- [6] Zadeh L.A., (1965), *Fuzzy sets*, Information and control, vol. 8.

CONSUMPTION EXPENDITURES AND DEMANDS OF AGEING POPULATION

Jaroslav Sixta¹, Jakub Fischer²

Abstract. The paper deals with the consumption expenditures of Czech households in the situation of population ageing. The paper describes the structure of consumption expenditures and a broader definition of consumption (actual consumption) where selected expenditures of government and non-profit institutions are added. We present data for 2019 and estimate the development to 2060 in line with official demographic projections. We use data from national accounts and combine them with demographic indicators. The estimates of the size of the impacts on the economy are computed via Leontief input-output model (input-output analysis). We present significant increase of demands for health and social services due population ageing, nearly doubled. We estimate that health and social services will require much more costs and number of necessary workers will rise from 330 thousand to 460 thousand between 2020 and 2060. Input-output analysis also shows the impact on all industries in the economy despite the primary demand shock is aimed at health and social services.

Keywords: Households Consumption, Population Ageing, Input-Output Analysis

JEL Classification: J11, C67

AMS Classification: 62P20

1 Introduction

The quality of life has different levels, and nowadays, nearly everybody knows that this issue contains more components than just production or consumption. On the other side, the consumption component can be hardly omitted. It will be pleasant to live in a friendly and clean environment, but only if we are not starving. The future perspectives of our lives should be carefully studied since quantitative aspects are not sufficiently emphasized. This perspective is not connected with a retirement pension and enough material goods but also health and social services. The demographic ageing that is the case soon and will be observable in few years should lead to careful consideration of our policymakers.

The economic activity of older workers is one of the most important issues and we believe that this will be the driving factor of future development. Our estimates on the availability of health and social services and their impact take into account economic activity that is significantly influenced by state-fixed statutory retirement age of by 65 years in the Czech Republic. For demographers and economists, it is obvious that this capping is unsustainable. The sources for the study of economic activity can be found in Hiesinger and Tophoven, [2]. It is difficult to state the effect but we can also find studies that increases of the productivity in line with age of workers, see Posner [4]. The issue of productivity and its link to the age structure is deeply discussed in Šimková and Marek [7] or Tang and Macleod [8]. Similarly, Sewdas et al. [5] discussed the health status of workers and the risk of health-related job loss.

The crucial problem will arise in the availability of health and social services that form a significant part of our consumption in a broader sense. For national accountants', the actual consumption comprises households' consumption and non-market services provided mainly by the government. Part of this issue was discussed in Sixta and Fischer [6], and another part of the issue is discussed within this paper. A similar approach was used in Langamrová et al. [3] We provide the outlook of the development of ageing driven costs to health and social service by 2040 and 2060. We focus on the middle variant of the demographic projection, and we combine it with macro-economic data. The outcomes comprise the estimates of the health and social expenditures for services and a brief input-output analysis of the extrapolation. The input-output analysis provides a rough estimate of the impacts on specific industries.

¹ Prague University of Economics and Business, sixta@vse.cz.

² Prague University of Economics and Business, fischej@vse.cz

2 Data and Methodology

We use population projection conducted by the Czech Statistical Office and national accounts data from the same source. We select the middle variant of the population projection and three age groups (0-14, 15-64, 65 and over). We also use the structure of household expenditure that comes from the household budget survey from the Czech Statistical Office.

The analysis is conducted by using input-output analysis, similarly to Sixta and Fischer [6], where the impulse of the change is contained in \mathbf{y} , representing the change of the final use. Our analysis, $\Delta\mathbf{y}$ stands for the vector of changes of actual consumption of both households and government. The model is expressed by equation (1):

$$\Delta\mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1}\Delta\mathbf{y} \quad (1)$$

where

\mathbf{x} .. vector of output,

\mathbf{A} .. Leontief matrix of technical coefficients,

\mathbf{y} .. final demand – actual consumption in our case.

The matrix \mathbf{A} from equation (1) comes from the latest symmetric input-output tables available (2015), and estimates of the change of final use are for 2020, 2040 and 2060.

The impulse (\mathbf{y}) is estimated by the projection of actual consumption of specific goods and services describing health and social services. We count medicaments (CPA code 21), health services (CPA code 86) and social services (CPA code 87 and 88). Our analysis does not make sense to split these categories since this is a relatively simple economic outlook, but a deeper view could be provided. Therefore, we denote \mathbf{y} as health and social services costs associated with ageing. The estimates are based on data on the costs at comparable prices (2015), extrapolated by linear regression. The dependency of macroeconomic health and social services costs on the old population, people over 65 years, is very high. More profound studies aimed at the cost of Czech health insurance companies per person over 65 shows a significant increase. But this is not the aim of our analysis, and we aimed at total economic costs of health and social services. On the aggregated level, even simple linear regression provide valuable results.

The input-output analysis provides the estimates of the changes of output, value added, and employment and these indicators form the critical outcome of our analysis. In our opinion, our results are very optimistic since we do not consider fast running progress in medicine and public opinion of obtaining the most recent and successful care for everybody.

3 Population Outlook and Economic Consequences of Ageing by 2060

The Czech population is ageing, and its proportion does not provide any reason for the optimism. The change will come smoothly, and it is connected with shrinking the most productive generation of people between the ages of 15 and 64. Despite ongoing and never-ending discussion about the statutory retirement age between the policy-makers, it seems economically clear that it will be rising over 65 years inevitably. However, the rise of the statutory retirement age need not provide enough resources for maintaining the same living standard and quality of life of the pensioners. The Statistical Office published its demographic projection in 2018, see Czech Statistical Office [1]. Its middle variant shows a relatively stable number of people by 15 years but sharply decreasing share of people between 15 and 65, see figure 1.

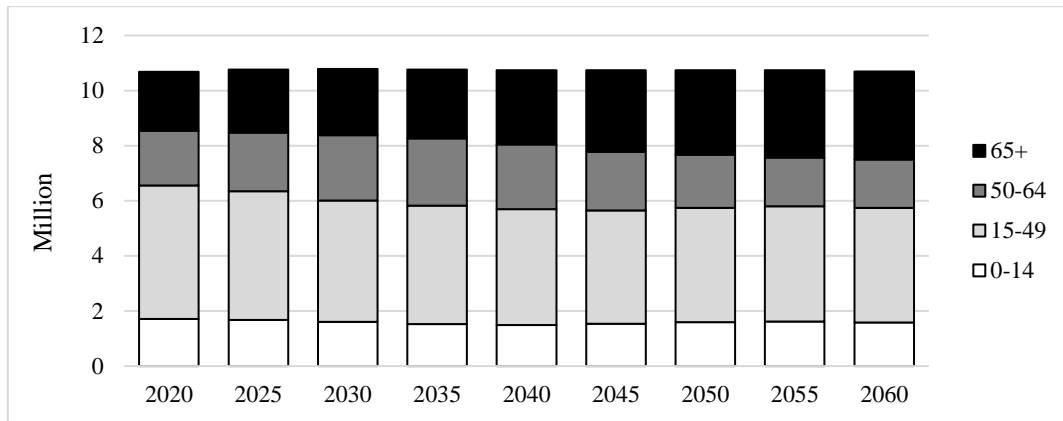


Figure 1 Population outlook by 2060 in the Czech Republic

The share of people between 0 and 14 years counted 16% in 2020, and it is projected to down by two percentage points in 2040, with a slight increase to 15% in 2060. The share of the people over 65 years will rise from 20% to 25% and 30% in 2040, 2060 respectively. It means that the most productive part of the population (15 to 64 years) is shrinking from 64% to 55%. It will be inevitably observable in the economy and financial balance of the state budget.

We should be aware that consumption habits are changing during individuals' life and expenditures are conditioned by age. Following figure 2 describes the structure of consumption by type products³. It is observable, the significantly different share of the household budget takes dwellings and associated costs (code L) and health and social services (Q). Of course, changes can be observed in manufacturing products, transport services and similar. When focusing on a relatively minor share of expenditures on health and social services of people between 15 and 49 years (0.7%), the same expenditures doubled for the group 65+. We are currently in a situation when people are not more or less softly motivated to spend money on health services. When measuring total expenditures for health and social service, counting about 436 CZK billion, the household share was less than 18%. The public insurance system finances the vast majority.

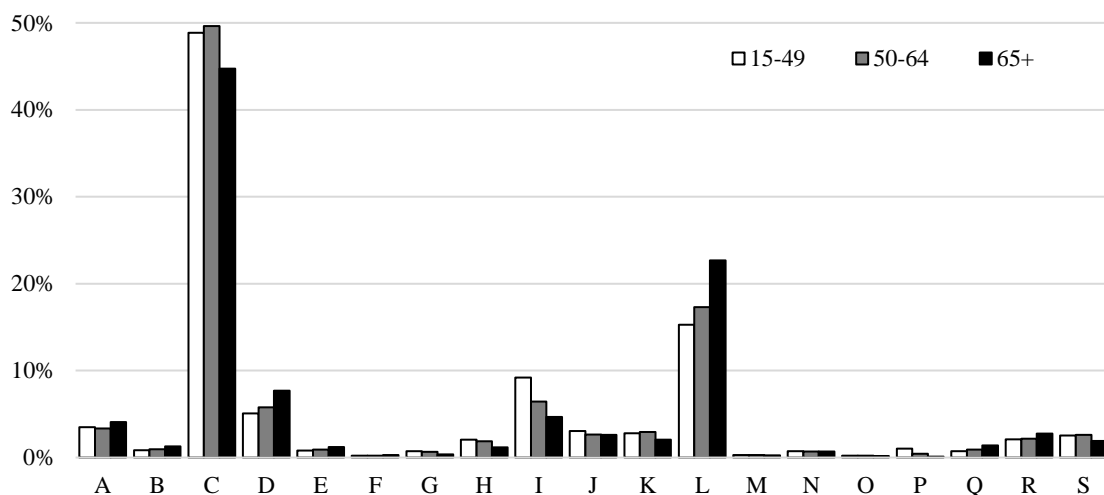


Figure 2 Structure of household consumption by age groups

³ Codes of the classification of industries and corresponding products are following: : A - Agriculture, forestry and fishing, B - Mining and quarrying, C - Manufacturing, D - Electricity, gas, steam and air conditioning supply, E - Water supply; sewerage, waste management and remediation activities, F - Construction, G - Wholesale and retail trade; repair of motor vehicles and motorcycles, H - Transportation and storage, Other: I - Accommodation and food service activities, J - Information and communication, K - Financial and insurance activities, L - Real estate activities, M - Professional, scientific and technical activities, N - Administrative and support service activities, O - Public administration and defense; compulsory social security, P - Education, Q - Human health and social work activities, R - Arts, entertainment and recreation, S - Other service activities, T - Activities of households as employers and producers for own use.

The age group significantly influences the structure of consumption, and this is reflected in estimated health expenditures. When analyzing data, we counted the final consumption of medicines, health and social services for households, government, and non-profit organizations serving households. Data comes from national accounts' supply and use tables, and they are valued at 2015 prices⁴.

Constant price valuations allow us further extrapolation by regression analysis and input-output analysis. The relationship between health and social services costs estimates is relatively straightforward. The increase in the quality of health services available in the Czech Republic in 1990s was evident, and the increase of costs accompanied it (see Figure 3. Future health and social costs will be under firm pressure given by the ageing population and ongoing progress in medicine. The second factor is deeply rooted in our modern society's behavior and is reflected in the highest quality and the availability of top outcomes of medicine. It means that our estimated future expenditures based on linear regression, with an adjusted R-squared of 0.73⁵, correspond with just a very conservative estimate. We assume that the reality will be much more costly since per person expenditures of health insurance companies are non-linearly rising with age. Anyway, the reason for our simple model is to provide a conservative estimate for input-output analysis.

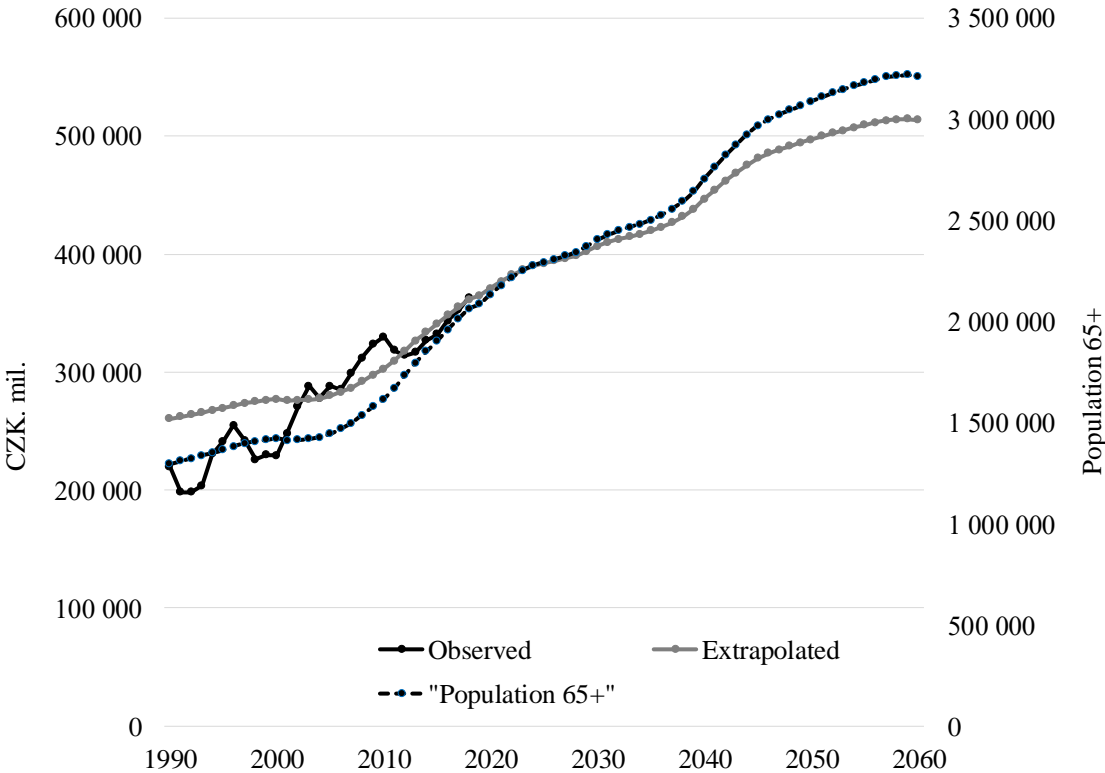


Figure 3 Estimates of health and social services costs, prices of 2015

4 Quantification of Health and Social Services Costs by Input-Output analysis

The link between the necessity of increased volume of health and social service and the number of older people is evident. We used the estimated amount of health and social costs for 2020, 2040 and 2060 to illustrate the effects of increased costs by input-output analysis. The following figure 4 illustrates the impact of future estimate health and social costs on the economy. According to formula 1, the induced effects lead to increased output of medicines, health and social services by 20.4% by 2040 and 38.4% by 2060. As stated before, this estimate is relatively conservative, and the results roughly correspond to the research published by Sixta and Fischer [6], where different

⁴ Constant price valuation reduces the risk of autocorrelation for further estimates.
⁵ All standard measures and assumptions for classic linear regression model are kept.

approach based on consumers' consumption of the middle variant of demographic projection leads to 26.1% and 67.7% increase of output by 2040, 2060 respectively.

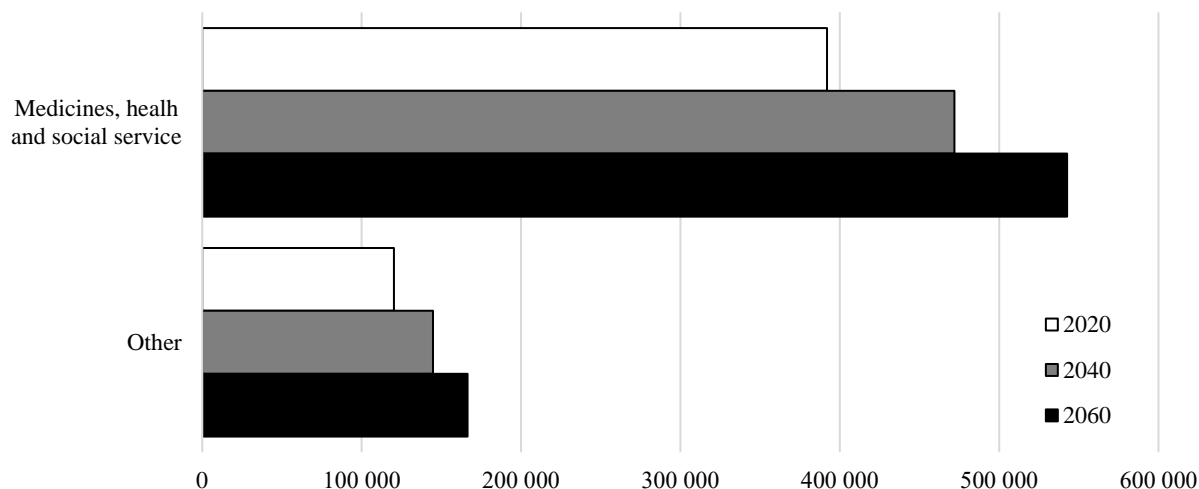


Figure 4 Impact on output by multiplication effects of health and social costs on industries, 2015 prices

We estimated the impact of increased demand for medicines, health and social services within two different papers to compare alternative estimates. Table 1 shows the impact of the multiplication of costs (second column) on output, gross value added and employment (in persons). The outcomes are apparent. The ageing population will require the allocation of resources to the health and social industries of the economy. The estimates will differ from the reality in 2040 and 2060, but no one can expect any relief in this issue. We expect increased necessity for resources from 20% to 26% by 2040 and from 38% to 68% by 2060. As well as there is substantial uncertainty in the future development, our estimates have high variation.

Year	Costs (CZK mil.)	Generated output (CZK mil.)	Generated value added (CZK mil.)	Employment (persons)
2020	370 876	512 167	279 962	332 592
2040	446 554	616 798	337 090	400 537
2060	513 438	709 179	387 578	460 528

Table 1 Impacts of the health and social costs, the Czech Republic

If we compare our estimates within the Czech Economy, the increase of necessary workers from about 333 thousand to 401 thousand in 2040 means a relatively moderate increase. Currently, the employment within health and social industries work about 6.5% of our labor force in national account terms. Despite the expected increase of a statutory retirement age, available labor force in optimistic scenario will not significantly shrink. If statutory retirement age did not exceed 65 years, the situation would be serious and economically untenable. We expect that the share can rise to 8% and 9% in 2040, 2060 respectively.

5 Conclusion

Consumption expenditures and demand of ageing population can be expressed by many means. We consider this issue as very relevant and therefore, we put the emphasis on such estimate. We are currently able to provide estimates for the research based on input-output analysis of future potential consumption of expenditures presented on the MME 2019 conference and alternative approach presented within this paper based on input-output analysis of extrapolated health and social costs. We focus on the expenditures of both private conducted by the households and non-profit institution serving households and public conducted by the government. All the estimates are based on actual individual consumption estimated at constant prices of 2015.

Obtained results reflected with previous approach and it allows us comparing potential effect very conservatively. We do not take into account progress in medicine and increased costs of better cure. The estimated effects of

increased demand for health and social services in 2040 and 2060 counts 70 and 130 thousand necessary workers more for health and social services needed in comparison with 2020. This means 38% increase between 2060 and 2020. Despite the continuous increase of the productivity, it will be very difficult to find such skilled employees since the age structure of doctors and nurses is worsening every year in the Czech Republic.

Acknowledgements

Supported by the Institutional Support for Long Period and Conceptual Development of Research and Science at Faculty of Informatics and Statistics, University of Economics, Prague and the project “Economy of Successful Ageing”, no. 19-03984S.

References

- [1] Czech Statistical Office (2018). *Population projection of the Czech Republic - 2018 - 2100 (Czech only)* Praha. <https://www.czso.cz/csu/czso/projekce-obyvatelstva-ceske-republiky-2018-2100>
- [2] Hiesinger, K.; Tophoven, S. (2017). Workloads as mediator for the association between job requirement level and health of older workers. *European Journal of Public Health*. Vol. 27, No. 3, 102–103.
- [3] Langhamrová, J., Šimková, M. & Sixta, J. (2018). Makroekonomické dopady rozšiřování sociálních služeb pro stárnoucí populaci České republiky. *Politická ekonomie*. Vol. 66, No. 2, 240–259.
- [4] Posner, R. (1995). *Aging and Old Age*. University of Chicago Press, Chicago.
- [5] Sewdas, R., Van Der Beek, A.J., Boot, C., D'Angelo, S., Syddal, H.E., Palmer, K.T., Walker-Bone, K. (2019). Poor health, physical workload and occupational social class as determinants of health-related job loss: results from a prospective cohort study in the UK. *BMJ OPEN*, Vol. 9, No. 7.
- [6] Sixta, J., Fischer, J. Multiplication Effects of Social and Health Services (2019). *37th International Conference on Mathematical Methods in Economics*, 246–250
- [7] Šimková, M., Marek, L. (2017). Age Structure of Labour Force and Its Impact on Wages and Product. *Applications of Mathematics and Statistics in Economics*. DOI: 10.15611/amse.2017.20.34., 419–430.
- [8] Tang, J., Macleod, C. (2006). Labour Force Ageing and Productivity Performance in Canada. *The Canadian Journal of Economics*, Vol. 39, No. 2, 582–603.

Central Moments and Risk-Sensitive Optimality in Markov Reward Processes

Karel Sladký¹

Abstract. In this note we consider discrete- and continuous-time Markov decision processes with finite state spaces. There is no doubt that usual optimality criteria examined in the literature on optimization of Markov reward processes, e.g. total discounted or mean reward, may be quite insufficient to select more sophisticated criteria that reflect also the variability-risk features of the problem. In this note we focus on models where the stream of rewards generated by the Markov processes is evaluated by an exponential utility function with a given risk sensitivity coefficient (so-called risk-sensitive models). For the risk-sensitive case, i.e. if the considered risk-sensitivity coefficient is non-zero, we establish explicit formulas for growth rate of expectation of the exponential utility function. Recall that in this case along with the total reward also its higher moments are taken into account. Using Taylor expansion of the utility function we present explicit formulae for calculating variance and higher central moments of the total reward generated by the Markov reward process along with its asymptotic behavior.

Keywords: Discrete- and continuous-time Markov reward chains, exponential utility, moment generating functions, formulae for central moments.

JEL classification: C44, C61

AMS classification: 90C40, 60J10

1 Introduction

The usual optimization criteria examined in the literature on stochastic dynamic programming, such as a total discounted or mean (average) reward structures, may be quite insufficient to characterize robustness of the problem from the point of a decision maker. To this end it may be preferable if not necessary to select more sophisticated criteria that also reflect stability and variability-risk features of the problem. Perhaps the best known approaches stem from the classical work of Markowitz (1952) on mean variance selection rules, i.e. we optimize the weighted sum of average or total reward and its variance, and from the seminal paper titled "Risk-sensitive Markov decision processes" of Howard and Matheson (1972), based on evaluating generated reward by exponential utility functions. Higher moments and variance of cumulative rewards in Markov reward chains have been originally studied for discrete time models. Research in this direction has been initiated in early papers Mandl (1971), Jaquette (1973) and Sobel (1982). For connections with risk sensitive models see e.g. Cavazos-Cadena and Fernandez-Gaucherand (1999), Cavazos-Cadena and Hernández-Hernández (2005) and Sladký (2005),(2008),(2018) and (2020). For basic facts on controlled Markov processes, see e.g. Puterman (1994) or Ross (1983).

The present paper is structured as follows. Section 2 contains notations and summary of basic facts on discrete- and continuous-time Markov reward processes. Discrete-time Markov models with exponential utility function (called risk-sensitive Markov chains) are studied in section 3 along the corresponding moment generating functions and explicit formulas of higher moments and higher central moments. Similar results for continuous-time Markov reward chains are sketched in Section 4.

2 Notations and Preliminaries

In this note we consider Markov decision processes with finite state space $\mathcal{I} = \{1, 2, \dots, N\}$ evolving in discrete- and continuous-time.

In the discrete-time case, we consider Markov decision chain $X^d = \{X_n, n = 0, 1, \dots\}$ with finite state space $\mathcal{I} = \{1, 2, \dots, N\}$, and finite set $\mathcal{A}_i = \{1, 2, \dots, K_i\}$ of possible decisions (actions) in state $i \in \mathcal{I}$. Supposing that in state $i \in \mathcal{I}$ action $a \in \mathcal{A}_i$ is selected, then state j is reached in the next transition with a given probability $p_{ij}(a)$ and one-stage transition reward r_{ij} will be accrued to such transition. Obviously, $r_i(a) = \sum_{j \in \mathcal{I}} p_{ij}(a)r_{ij}$ is the expected one-stage reward.

¹The Czech Academy of Sciences, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, sladky@utia.cas.cz

In the continuous-time setting, the development of the considered Markov decision process $X^c = \{X(t), t \geq 0\}$ (with finite state space \mathcal{I}) over time is governed by the transition rates $q(j|i, a)$, for $i, j \in \mathcal{I}$, depending on the selected action $a \in \mathcal{A}_i$. For $j \neq i$ $q(j|i, a)$ is the transition rate from state i into state j , $q(i|i, a) = -\sum_{j \in \mathcal{I}, j \neq i} q(j|i, a)$. Recall that on entering state i the process stays in state i for a random time that is exponentially distributed with parameter $q(i, a) = -q(i|i, a)$ and the next jump to state j occurs with probability $p_{ij}(a) = q(j|i, a)/q(i, a)$. As concerns reward rates, $r(i)$ denotes the rate earned in state $i \in \mathcal{I}$, and $r(i, j)$ is the transition rate accrued to a transition from state i to state j .

A (Markovian) policy controlling the decision process is given either by a sequence of decision at every time point (discrete-time case) or as a piecewise constant right continuous function of time (continuous-time case). Policy which takes at all times the same decision rule, i.e. $\pi \sim (f)$, is called stationary; in discrete-time models $P(f)$ is the transition probability matrix with elements $p_{ij}(f_i)$. Obviously, $r_i(f_i) = \sum_{j=1}^N p_{ij}(f_i) r_{ij}$ is the expected one-stage reward obtained in state $i \in \mathcal{I}$ and $r(f)$ denotes the corresponding N -dimensional column vector of one-stage rewards. $v(f, n) := \sum_{i=0}^{n-1} [P(f)]^n \cdot r(f)$ is the (column) vector of total rewards accrued after $n+1$ transitions, its i th entry $v_i(f, n)$ denotes expectation of the total reward if the process X^d starts in state i . Then $g(f) = P^*(f) \cdot r(f)$ (where $P^*(f) = \lim_{m \rightarrow \infty} m^{-1} \sum_{n=0}^{m-1} P^n(f)$) is the (column) vector of average rewards, its i th entry $g_i(f)$ denotes the average reward if the process starts in state i . If the Markov chain $P(f)$ is unichain, i.e. it has a single class of recurrent states, then $g(f)$ is constant vector of average rewards with elements $\bar{g}(f)$.

Similarly, for the continuous-time case policy controlling the chain, $\pi = f(t)$, is a piecewise constant, right continuous vector function where $f(t) \in \mathcal{F} \equiv \mathcal{A}_1 \times \dots \times \mathcal{A}_N$, and $f_i(t) \in \mathcal{A}_i$ is the decision (or action) taken at time t if the process $X(t)$ is in state i . Since π is piecewise constant, for each π we can identify the time points $0 < t_1 < t_2 < \dots < t_i < \dots$ at which the policy switches; we denote by $f^i \in \mathcal{F}$ the decision rule taken in the time interval $(t_{i-1}, t_i]$. Policy which takes at all times the same decision rule, i.e. $\pi \sim (f)$, is called stationary. $Q(f) = [q_{ij}(f_i)]$ for $f \in \mathcal{F}$ is an $N \times N$ matrix whose ij th element $q_{ij}(f_i) = q(j|i, f_i)$ for $i \neq j$ and for the ii th element we set $q_{ii}(f_i) = -q(i|i, f_i)$ (recall that the row sums in a transition rate matrix $Q(f)$ are equal to null). Expected value of the reward obtained in state $i \in \mathcal{I}$ equals $r_i(f_i) = [q(i|i, f_i)]^{-1} r(i) + \sum_{j \in \mathcal{I}, j \neq i} q(j|i, f_i) r(i, j)$ and $r(f) = [r_i(f)]$ is the (column) vector of reward rates at time t .

In this note we shall suppose that the obtained random reward, say ξ , is evaluated by an exponential utility function, say $u^\gamma(\cdot)$, i.e. utility functions with constant risk sensitivity depending on the value of the risk sensitivity coefficient γ .

In case that $\gamma > 0$ (the risk seeking case) the utility assigned to the (random) reward ξ is given by $u^\gamma(\xi) := \exp(\gamma\xi)$, if $\gamma < 0$ (the risk averse case) the utility assigned to the (random) reward ξ is given by $u^\gamma(\xi) := -\exp(\gamma\xi)$, for $\gamma = 0$ it holds $u^\gamma(\xi) = \xi$ (risk neutral case). Hence we can write

$$u^\gamma(\xi) = \text{sign}(\gamma) \exp(\gamma\xi) \quad (1)$$

and for the expected utility we have (\mathbf{E} is reserved for expectation)

$$\bar{U}^{(\gamma)}(\xi) := \mathbf{E}u^\gamma(\xi) = \text{sign}(\gamma)\mathbf{E}[\exp(\gamma\xi)] \quad \text{where} \quad U^{(\gamma)}(\xi) := \mathbf{E}[\exp(\gamma\xi)] = \sum_{k=0}^{\infty} \mathbf{E} \frac{1}{k!} (\gamma\xi)^k. \quad (2)$$

Then for the corresponding certainty equivalent $Z^\gamma(\xi)$ we have

$$u^\gamma(Z^\gamma(\xi)) = \text{sign}(\gamma)\mathbf{E}[\exp(\gamma\xi)] \iff Z^\gamma(\xi) = \gamma^{-1} \ln\{\mathbf{E}[\exp(\gamma\xi)]\}. \quad (3)$$

From (2),(3) we immediately conclude

$$Z^\gamma(\xi) \approx \mathbf{E}\xi + \frac{\gamma}{2} \text{Var} \xi \quad \text{where} \quad \text{Var} \xi = \mathbf{E}[\xi - \mathbf{E}\xi]^2. \quad (4)$$

In particular, considering discrete-time models, let ξ_n be the random reward obtained in the first n transitions of the considered Markov process, i.e.

$$\xi_n = r_{X_0, X_1} + r_{X_1, X_2} + \dots + r_{X_{n-1}, X_n}. \quad (5)$$

Supposing that stationary policy $\pi \sim (f)$ is followed and the process starts in state $i = X_0$ then $\mathbf{E}_i^\pi \xi_n = v_i(f, n)$ where $v(f, n) = [P(f)]^n r(f)$.

Similarly, for the continuous-time models, the (random) reward obtained in the first n jumps of the process depends not only on transition rewards (say r_{X_1, X_2}) connected with jumps of the process, but also on reward rates (say $r(\cdot)$) and random times (say τ) spent in the states, and is equal to

$$\xi_n = r_{X_0} \cdot \tau_{X_0} + r_{X_0, X_1} + r_{X_1} \cdot \tau_{X_1} + r_{X_1, X_2} + \dots + r_{X_{n-1}} \cdot \tau_{X_{n-1}} + r_{X_{n-1}, X_n}.$$

3 Discrete-Time Models: Exponential Utility and Higher Moments

It is well-known (see e.g. Puterman (1994), Ross (1983)) that if $P(f)$ is unichain then there exist vector $w(f)$ (unique up to constant vector) and constant vector $g(f)$ (with elements $\bar{g}(f)$) such that $w(f) + g(f) = r(f) + P(f)w(f)$. Then we can conclude that the vector of total expected rewards

$$v(f, n) = w(f) + n \cdot g(f) - [P(f)]^n \cdot w(f), \text{ in particular, } v_i(f, n) = w_i(f) + n \cdot \bar{g}(f) - [[P(f)]^n \cdot w(f)]_i, \quad (6)$$

i.e. the growth rate of $v_i(f, n)$ is linear over n .

Moreover, if the process starts in state i and stationary policy $\pi \sim (f)$ is followed from (2), (5) we can conclude that the growth of expected utility for risk-sensitive models in the n steps

$$U_i^{(\pi, \gamma)}(n) := E_i^{(\pi)} e^{\gamma(r_{x_0, x_1} + r_{x_1, x_2} + \dots + r_{x_{n-1}, x_n})} \quad (7)$$

well corresponds to the growth of $v_i(f, n)$. Observe that for γ near to zero $U_i^{(\pi, \gamma)}(\xi_n)$ risk-sensitive models can be well approximated by $v_i(f, n)$ of the classical models.

In what follows we show that if transition probability matrix $P(f) = [p_{ij}(f)]$ is unichain then:

There exist real $g^\gamma(f)$, $w_i^\gamma(f)$'s ($i \in \mathcal{I}$) such that for $\tilde{\varphi}_{ij}(w, g, f) := r_{ij} - g^\gamma(f) + w_j^\gamma(f) - w_i^\gamma(f)$

$$\sum_{j \in \mathcal{I}} p_{ij}(f_i) e^{\gamma[r_{ij} - w_i^\gamma(f) + w_j^\gamma(f) - g^\gamma(f)]} = 1 \iff \sum_{j \in \mathcal{I}} p_{ij}(f_i) e^{\gamma[r_{ij} + w_j^\gamma(f)]} = e^{\gamma[g^\gamma(f) + w_i^\gamma(f)]} \quad \text{for all } i \in \mathcal{I} \quad (8)$$

To verify (9), let $m_{ij}(f) := p_{ij}(f) \cdot e^{\gamma r_{ij}}$, $\rho(f) := e^{\gamma g^\gamma(f)}$, $x_j(f) := e^{\gamma w_j^\gamma(f)}$.

Then by the Perron–Frobenius theorem for nonnegative matrices (see e.g. Gantmacher (1959))

$$\sum_{i \in \mathcal{I}} m_{ij}(f) \cdot x_j(f) = \rho(f) \cdot x_i(f), \quad i \in \mathcal{I} \quad (9)$$

where $M(f) = [m_{ij}(f)]_{i,j}$ is an irreducible nonnegative matrix (or reducible nonnegative matrix with strictly positive right Perron eigenvector), $\rho(f)$ is the eigenvalue of $M(f)$ (equal to the spectral radius of $M(f)$) and $[x_i(f)]_i$ is the corresponding right eigenvector of $M(f)$.

Similarly, from (7) we conclude that for $\pi \sim (f)$

$$U_i^{(\pi, \gamma)}(n) = E_i^\pi e^{\gamma \sum_{k=0}^{n-1} r_{x_k, x_{k+1}}} = e^{\gamma[n g^\gamma(f) + w_i^\gamma(f)]} \times E_i^\pi e^{\gamma[\sum_{k=0}^{n-1} \tilde{\varphi}_{x_k, x_{k+1}}(w^\gamma(f), \gamma(f)) - w_{x_n}^\gamma(f)]}. \quad (10)$$

Observe that the first term on the RHS of (10) is non-random and hence (for $|c| \leq w_i(f)$)

$$E_i^\pi e^{\gamma[\sum_{k=0}^{n-1} \tilde{\varphi}_{x_k, x_{k+1}}(w^\gamma(f), \gamma(f)) - c]} \leq \frac{U_i^{(\pi, \gamma)}(n)}{e^{\gamma[n g^\gamma(f) + w_i^\gamma(f)]}} \leq E_i^\pi e^{\gamma[\sum_{k=0}^{n-1} \tilde{\varphi}_{x_k, x_{k+1}}(w^\gamma(f), \gamma(f)) + c]} \quad (11)$$

If (stationary) policy $\pi \sim (f)$ is followed then by (2) $U_i^\pi(\gamma, n) = E_i^\pi [\exp(\gamma \xi^{(n)})]$ is also the moment generating function of $\xi^{(n)}$. Hence (cf. (2)) for some $h > 0$ and any $|\gamma| < h$

$$\frac{d}{d\gamma} E_i^\pi [\exp(\gamma \xi^{(n)})] = E_i^\pi \xi^{(n)} [\exp(\gamma \xi^{(n)})], \quad \text{hence for } k = 0, 1, 2, \dots, n = 0, 1, 2, \dots$$

$$M_i^{(\pi, k)}(n) := E_i^\pi (\xi^{(n)})^k = \frac{d^k}{d\gamma^k} E_i^\pi [\exp(\gamma \xi^{(n)})]_{\gamma=0} \text{ is the } k\text{th moment of } \xi^{(n)} \quad (12)$$

and (cf. (2)) the Taylor expansion around $\gamma = 0$ reads for $|\gamma| < h$

$$U_i^\pi(\gamma, n) = 1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} M_i^{(\pi, k)}(n) \quad \text{and} \quad e^{\gamma r_{ij}} = 1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} [r_{ij}]^k. \quad (13)$$

From (12),(13) we immediately get

$$1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} M_i^{(\pi, k)}(n+1) = \sum_{j \in \mathcal{I}} p_{ij}(f) \left(1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} [r_{ij}]^k \right) \left(1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} M_j^{(\pi, k)}(n) \right). \quad (14)$$

Similarly on introducing the moment generating function for the central moments of $\xi^{(n)}$ by

$$\tilde{U}_i^\pi(\gamma, n) := \mathbb{E}_i^\pi [\exp(\gamma(\xi^{(n)} - \mathbb{E}_i^\pi \xi^{(n)}))] \quad \text{for all } i \in \mathcal{I} \quad (15)$$

for the k th central moment of $\xi^{(n)}$ we have

$$\widetilde{M}_i^{(k, \pi)}(n) := \mathbb{E}_i^\pi [\xi^{(n)} - \mathbb{E}_i^\pi \xi^{(n)}]^k = \frac{d^k}{d\gamma^k} \mathbb{E}_i^\pi [\exp(\gamma(\xi^{(n)} - \mathbb{E}_i^\pi \xi^{(n)}))] |_{\gamma=0} \quad (16)$$

and the Taylor expansion around $\gamma = 0$ for $|\gamma| < h$ reads

$$\tilde{U}_i^\pi(\gamma, n) = 1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} \cdot \widetilde{M}_i^{(\pi, k)}(n). \quad (17)$$

For the central moments similarly to (15), (16) we can conclude from (10), (11) that

$$\tilde{U}_i^\pi(\gamma, n) := \mathbb{E}_i^\pi e^{\gamma[\xi^{(n)} - (ng - w_i + w_{X_n})]} = \sum_{j \in \mathcal{I}} p_{ij}(f) \cdot e^{\gamma(r_{ij} - g + w_i - w_j)} \tilde{U}_j^\pi(\gamma, n-1) \quad (18)$$

where $\tilde{U}_j^\pi(\gamma, n-1) = \mathbb{E}_j^\pi e^{\gamma[\xi^{(1, n)} - (n-1)g + w_j - w_{X_n}]}$.

In analogy to (14) we get

$$1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} \widetilde{M}_i^{(\pi)}(k, n+1) = \sum_{j \in \mathcal{I}} p_{ij}(f) \left(1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} [r_{ij} - (g + w_i - w_j)]^k\right) \times \left(1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} \widetilde{M}_j^{(\pi)}(k, n)\right). \quad (19)$$

Similarly as in the previous section our analysis based on (15), (16), (17) and (18) enables to generate recursively all central moments of $\xi^{(n)}$.

By comparing in (18) the terms γ^k ($k = 1, 2, \dots$) we obtain the following recursive formulas for the central moments (obviously, the first central moment $\widetilde{M}_i^{(\pi, 1)}(n) \equiv 0$ for all n). In particular,

$$\text{For } k = 2 : \widetilde{M}_i^{(\pi, 2)}(n+1) = \sum_{j \in \mathcal{I}} p_{ij}(f) [(r_{ij} + w_j) - (g + w_i)]^2 + \sum_{j \in \mathcal{I}} p_{ij}(f) \widetilde{M}_j^{(\pi, 2)}(n). \quad (20)$$

$$\begin{aligned} \text{For } k = 3 : \widetilde{M}_i^{(\pi, 3)}(n+1) &= \sum_{j \in \mathcal{I}} p_{ij}(f) [(r_{ij} + w_j) - (g + w_i)]^3 \\ &+ 3 \sum_{j \in \mathcal{I}} p_{ij}(f) [(r_{ij} + w_j) - (g + w_i)] \widetilde{M}_j^{(\pi, 2)}(n) + \sum_{j \in \mathcal{I}} p_{ij}(f) \widetilde{M}_j^{(\pi, 3)}(n). \end{aligned} \quad (21)$$

In general:

$$\begin{aligned} \widetilde{M}_i^{(\pi, s)}(n+1) &= \sum_{j \in \mathcal{I}} p_{ij}(f) \{[(r_{ij} + w_j) - (g + w_i)]^s\} \\ &+ \sum_{j \in \mathcal{I}} p_{ij}(f) \left\{ \sum_{k=1}^{s-1} \binom{s}{k} [(r_{ij} + w_j) - (g + w_i)]^k \widetilde{M}_j^{(\pi, s-k)}(n) \right\} + \sum_{j \in \mathcal{I}} p_{ij}(f) \widetilde{M}_j^{(\pi, s)}(n) \end{aligned} \quad (22)$$

that can be also written as

$$\widetilde{M}_i^{(\pi, s)}(n+1) = \sum_{k=0}^s \binom{s}{k} \sum_{j \in \mathcal{I}} p_{ij}(f) \{[(r_{ij} + w_j) - (g + w_i)]^k \widetilde{M}_j^{(\pi, s-k)}(n)\} \quad (23)$$

From these equations we immediately conclude that if stationary policy $\pi \sim (f)$ is followed the variance (i.e. the central second moment) of the total reward grows linearly over time and the growth rate $g^{(2)}$ of $\widetilde{M}_i^{(\pi, 2)}(n)$ in (21) can be found as a solution of

$$\widetilde{M}_i^{(\pi, 2)}(n) = ng^{(2)} + w_i^{(2)} \quad \text{where} \quad (24)$$

$$g^{(2)} + w_i^{(2)} = s_i^{(2)}(f) + \sum_{j \in \mathcal{I}} p_{ij}(f) w_j^{(2)}, \quad s_i^{(2)}(f) = \sum_{j \in \mathcal{I}} p_{ij}(f) [(r_{ij} + w_j) - (g + w_i)]^2. \quad (25)$$

To establish the growth rate of $\widetilde{M}_i^{(\pi,3)}(n)$, it suffices to insert into (21) from (20). Since $\sum_{j \in \mathcal{I}} p_{ij}(f) [(r_{ij} + w_j) - (g + w_i)] g^{(2)} = 0$ we can conclude

$$\sum_{j \in \mathcal{I}} p_{ij}(f) [(r_{ij} + w_j) - (g + w_i)] (ng^{(2)} + w_j^{(2)}) = \sum_{j \in \mathcal{I}} p_{ij}(f) [(r_{ij} + w_j) - (g + w_i)] w_j^{(2)}.$$

Hence using the same arguments as for the second central moment we can conclude that

$$\widetilde{M}_i^{(\pi,3)}(n) = ng^{(3)} + w_i^{(3)} \quad \text{where} \quad g^{(3)} + w_i^{(3)} = s_i^{(3)}(f) + \sum_{j \in \mathcal{I}} p_{ij} w_j^{(3)} \quad (26)$$

$$s_i^{(3)}(f) = \sum_{j \in \mathcal{I}} p_{ij}(f) \left\{ [(r_{ij} + w_j) - (g + w_i)]^3 + 3 [(r_{ij} + w_j) - (g + w_i)] w_j^{(2)} \right\}. \quad (27)$$

4 Continuous-time Models: Exponential Utility and Higher Moments

Supposing that the obtained random reward up to time t , say $\xi(t)$, is evaluated by an exponential utility function, say $u^\gamma(\cdot)$, with the risk sensitivity coefficient γ , let for $\pi \sim (f)$, $U_i^{(\gamma)}(t, f) := \mathbb{E}_i^\pi[\exp(\gamma \xi(t))]$ considered as the moment generating function of $\xi(t)$, we can conclude that for $k = 0, 1, 2, \dots, n = 0, 1, 2, \dots$

$$M_i^{(k,\pi)}(t) := \mathbb{E}_i^\pi(\exp(\xi(t)^k)) = \frac{d^k}{d\gamma^k} \mathbb{E}_i^\pi[\exp(\gamma \xi(t))] |_{\gamma=0} \quad \text{is the } k\text{th moment of } \xi(t) \quad (28)$$

and the Taylor expansion around $\gamma = 0$ reads

$$U_i^{(\gamma)}(t, f) = 1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} M_i^{(k,\pi)}(t). \quad (29)$$

Similarly on introducing the moment generating function for the central moments of $\xi(t)$ by

$$\widetilde{U}_i^{(\gamma)}(t, f) := \mathbb{E}_i^\pi[\exp(\gamma(\xi(t) - \mathbb{E}_i^\pi \xi(t)))]^k \quad \text{for all } i \in \mathcal{I} \quad (30)$$

for the k th central moments of $\xi(t)$ we have

$$\widetilde{M}_i^{(k,\pi)}(t) := \mathbb{E}_i^\pi[\xi(t) - \mathbb{E}_i^\pi \xi(t)]^k = \frac{d^k}{d\gamma^k} \mathbb{E}_i^\pi[\exp(\gamma(\xi(t) - \mathbb{E}_i^\pi \xi(t)))] |_{\gamma=0} \quad (31)$$

and the Taylor expansion around $\gamma = 0$ for sufficiently small γ reads

$$\widetilde{U}_i^{(\gamma)}(t, f) = 1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} \widetilde{M}_i^{(k,\pi)}(t). \quad (32)$$

Let $M^{(k,\pi)}(t)$, $\widetilde{M}^{(k,\pi)}(t)$ be the (column) vector of the k moments, central k moments respectively, with elements $M_i^{(k,\pi)}(t)$, $\widetilde{M}_i^{(k,\pi)}(t)$ respectively.

In particular, if the system starts in state i , the expected total reward earned in state i up to the first exit of state i within time interval $[t, t + \delta)$ is equal to

$$M_i^{(1,\pi)}(t + \delta) = M_i^{(1,\pi)}(t) \cdot [1 - q_i(f_i) \cdot \delta] + \delta \cdot r(i) \cdot [1 - q_i(f_i) \cdot \delta] + o(\delta^2) \quad (33)$$

and for the s -th power of this reward it holds

$$\begin{aligned} M_i^{(s,\pi)}(t + \delta) &= \{M_i^{(1,\pi)}(t)[1 - q_i(\cdot) \cdot \delta] + [r(i) \cdot \delta] \cdot [1 - q_i(f_i) \cdot \delta]\}^s \\ &= M_i^{(s,\pi)}(t) + s \cdot M_i^{(s-1,\pi)}(t) \cdot \delta + o(\delta^2) \end{aligned} \quad (34)$$

(Observe that $M_i^{(s,\pi)}(t) = [M_i^{(1,\pi)}(t)]^s$, $[M_i^{(1,\pi)}(t) + r(i) \cdot \delta]^s = M_i^{(s,\pi)}(t) + s \cdot r(i) \cdot M_i^{(s-1,\pi)}(t) + o(\delta^2)$.)

On inserting from (28),(30),(33) after some algebra and for δ tending to null we arrive at differential equations similar to that of discrete time models. In particular,

$$\text{For } k = 1 : \quad \frac{d}{dt} M_i^{(1,\pi)}(t) = r(i) + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) r_{ij} + \sum_{j \in \mathcal{I}} q_{ij}(f_i) M_j^{(1,\pi)}(t). \quad (35)$$

$$\begin{aligned} \text{For } k = 2 : \quad \frac{d}{dt} M_i^{(2,\pi)}(t) &= 2 \cdot M_i^{(1,\pi)}(t) \cdot r(i) + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \left\{ [r_{ij}]^2 + 2 \cdot r_{ij} \cdot M_j^{(1,\pi)}(t) \right\} \\ &+ \sum_{j \in \mathcal{I}} q_{ij}(f_i) \cdot M_j^{(2,\pi)}(t). \end{aligned}$$

In general:

$$\begin{aligned} \frac{d}{dt} M_i^{(s,\pi)}(t) &= s \cdot M_i^{(s-1,\pi)}(t) \cdot r(i) + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \left\{ \sum_{k=1}^s \binom{s}{k} \cdot M_i^{(s,\pi)}(t) \cdot [r_{ij}]^k M_j^{(s-k,\pi)}(t) \right\} \\ &+ \sum_{j \in \mathcal{I}} q_{ij}(f_i) \cdot M_j^{(s,\pi)}(t). \end{aligned} \quad (36)$$

Supposing that higher moments are known, the corresponding central moments can be easily computed. To this end, on recalling definition central moments, we can easily conclude that the if the system starts in state i and policy π is followed then the n th central moment at time t

$$\widetilde{M}_i^{(n,\pi)}(t) := \mathbb{E}_i^\pi [\xi(t) - M_i^{(1,\pi)}(t)]^n \quad (37)$$

Since $M_i^{(j,\pi)}(t) := \mathbb{E}_i^\pi [\xi(t)]^j$, after little algebra we arrive at

$$\begin{aligned} \widetilde{M}_i^{(n,\pi)}(t) &:= \sum_{j=0}^n \binom{n}{j} \cdot (-1)^{n-j} \cdot M_i^{(j,\pi)}(t) \cdot [M_i^{(1,\pi)}(t)]^{n-j} \\ &= \sum_{j=0}^{n-2} \binom{n}{j} \cdot (-1)^{n-j} \cdot M_i^{(j,\pi)}(t) \cdot [M_i^{(1,\pi)}(t)]^{n-j} + (-1)^{n-1} \cdot (n-1) \cdot [M_i^{(1,\pi)}(t)]^n. \end{aligned} \quad (38)$$

More details can be found in Sladký (2020).

5 Conclusions

The paper presents explicit formulas for higher moments and higher central moments of cumulative rewards in Markov reward chains that can be applied for more sophisticated optimality criteria in many dynamic problems.

Acknowledgements: This work was supported by the Czech Science Foundation under Grant 18-02739S.

References

- [1] Cavazos-Cadena, R. and Fernandez-Gaucherand, F. (1999). Controlled Markov ..Chains with Risk-Sensitive Criteria: Average Cost, Optimality Equations and Optimal Solutions. *Mathematical Methods of Operations Research*, 43, pp. 121–139.
- [2] Cavazos-Cadena, R. and Hernández-Hernández, D. (2005). A Characterization of the Optimal Risk-Sensitive Average Cost Infinite Controlled Markov Chains. *Annals of Applied Probability*, 15, pp. 175–212.
- [3] Gantmakher, F. R. (1959). *The Theory of Matrices*. Chelsea, London.
- [4] Howard, R. A. and Matheson, J. (1972). Risk-Sensitive Markov Decision Processes. *Management Science*, 18(7), pp. 356–369.
- [5] Jaquette, S. C. (1973). Markov Decision Processes with a New Optimality Criterion: Discrete Time. *Annals of Statistics*, 1, pp. 496–505.
- [6] Mandl, P. (1971). On the Variance in Controlled Markov Chains. *Kybernetika*, 7(1), pp. 1–12.
- [7] Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, 7, 77–92.
- [8] Prieto–Rumeau, T. and Hernández-Lerma, O. (2009). Variance Minimization and the Overtaking Optimality Approach in Continuous-Time Controlled Markov Chains. *Mathematical Methods of Operations Research*, 70, pp. 527–540.
- [9] Puterman, M. L. (1994). *Markov Decision Processes – Discrete Stochastic Dynamic Programming*. Wiley: New York.
- [10] Ross, S. M. (1983). *Introduction to Stochastic Dynamic Programming*. Academic Press: New York.
- [11] Sladký, K. (2005). On Mean Reward Variance in Semi-Markov Processes. *Mathematical Methods of Operations Research*, 62(3), pp. 387–397
- [12] Sladký, K. (2008). Growth Rates and Average Optimality in Risk-Sensitive Markov Decision Chains. *Kybernetika*, 44(2), pp. 205–216.
- [13] Sladký, K. (2018). *Central Moments and Risk-Sensitive Optimality in Markov Reward Chains*. In: Quantitative Methods in Economics – Multiple Criteria Decision Making XIX (M. Reiff, P. Gežík, Eds). University od Economics, Bratislava 2018, pp. 325–331.
- [14] Sladký, K. (2020). *Central Moments and Risk-Sensitive Optimality in Continuous-Time Markov Reward Processes*. In: Quantitative Methods in Economics – Multiple Criteria Decision Making XX (M. Reiff, P. Gežík, Eds). University od Economics, Bratislava 2020, pp. 305–311.
- [15] Sobel, M. (1982). The Variance of Discounted Markov Decision Processes. *Journal of Applied Probability*, 19, pp. 794–802.
- [16] van Dijk, N.M. and Sladký, K. (2006). On the Total Reward Variance for Continuous-Time Markov Reward Chains. *Journal of Applied Probability*, 43(4), pp. 1044–1052.

Construction and optimization of HFT systems with the use of binary-temporal state model

Michał Dominik Stasiak¹

Abstract. The state model for a binary-temporal representation (SMBTR) is one of the main models dedicated to the binary-temporal data format. This representation is characterized by a higher accuracy than the commonly used candlestick one. Using the candlestick data format can lead to a significant decrease in the informative value of data, especially regarding the course trajectory changes. SMBTR model allows for analysis of course changes defined as transitions between defined market states, and, as a result, for approximation of probability distribution, calculated for the future direction of the next change in rate trajectory. In the article, we present a schema for constructing an algorithmic trade system that uses the mentioned probability distribution. Also, optimization of model parameters was performed, to obtain the highest financial efficiency of the proposed system. The research was carried out for the gold market, using historical data from 6 years (2014-2020). Calculations were performed using dedicated MQL4 and C++ software.

Keywords: high frequency econometric, technical analysis, investment decision support, algorithmic trading.

JEL Classification: C53, G11, C49

AMS Classification: 91G70, 62P20

1 Introduction

In recent years we observe the increasing popularity of an algorithmic trade, especially of the HFT systems, which make hundreds or thousands of transactions, where the profit comes from the statistical advantage of the profitable transactions over the lossy ones [1]. New platforms are constantly being developed, which offer so-called ‘social trading’, where investors can invest according to a given strategy by paying a set subscription fee for the system author (e.g. ZuluTrade, Etoro, etc.). Construction of profitable algorithmic trade systems that make thousands of fast transactions requires using a new approach and prediction methods different from the commonly used methods of technical analysis (e.g. visual analysis) [3,6,13]. Even the very format of analysed data is problematic – for example, the vastly used candlestick representation leads to a loss of an undefined informative value, which could result in the falsification of further data analysis [15]. Therefore, the application of proper formatting of the exchange rate data and choosing advanced modelling methods is crucial for the construction of an effective algorithmic trade system.

In the following article, we propose using state modelling of the exchange rate quotations given in a binary-temporal representation to construct an algorithmic trade system. The modelling process uses the state model in binary-temporal representation (SMBTR), which allows for an approximation of the probability distribution calculated for the future direction of a course change [14]. This distribution stands as a basis for the proposed algorithmic trade system. The financial efficiency of the obtained system is dependent on the assumed parameters of the used representation and the model. The paper presents a schema for appointing optimal parameters and means for assessing the obtained financial efficiency. Also, empirical research was performed, based on an exemplary course of one of the most popular financial instrument, that is the exchange rate between gold and dollar (XAU/USD).

The article is organized as follows. After a brief introduction, in the second section, we present the assumptions of the binary-temporal representation. In the third section, we introduce the assumptions of the binary-temporal state model. The fourth section includes a detailed description of the construction of an algorithmic trade system and the means of assessing its financial efficiency. In the fifth section, a schema for appointing system parameters assuring the optimal efficiency is given. Additionally, research results of an exemplary analysis of historical data are provided, for the gold exchange rate from the period of 2014-2020.

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 Binary-temporal representation

Investors, as well as analysts and researchers, use mainly the historical quotations in the so-called candlestick representation to analyze the exchange rate trajectories. In this representation, the rate is described by four parameters (minimum, maximum, opening and closing price), calculated for a given timeframe. This kind of representation is commonly used in all broker platforms. One should take notice that the representation is also a basis for almost all visual methods of technical analysis (e.g. formation analysis, etc.) and appointment of indicators (e.g. RSI, MACD, etc.) [3,5,6]. Yet, the candlestick representation causes a loss of information about price changes "inside" the candle. The consequence of this may be erroneous results of the further analysis - detailed description of dangers of using the candlestick representation are described in [15]. An intuitive solution would be to use raw tick data, but the data is burdened with a lot of *noise* - many small changes (of 1-2 pips) of a random nature [7,8,9]. Therefore, in order to allow a possible application in predictive modelling algorithmic trade systems, the used representation should filter the mentioned noise.

The binary-temporal representation was inspired by the visual point-symbolic method [4]. The basic assumption of this representation is to assign two values to each i -th change in the exchange rate by a given, fixed value (the so-called discretization unit (δ)). The first value describes the direction of the change ($\varepsilon_i = 1$ for an increase and $\varepsilon_i = 0$ for a decrease), and the second one is the duration of the change given in the time units (Δt_i) [14]. The course is considered as a process of changes, so we assume that the next change and its duration is registered in relation to the end of the previous change. The course trajectory of a given financial instrument can be described as follows:

$$\mathcal{E} = \{(\varepsilon_i, \Delta t_i)\}_{i=1}^N, \quad (1)$$

where N is the number of registered changes.

Using the binary-temporal representation allows for noise filtration – the changes smaller than the discretization unit are not registered. At the same time, the representation allows for retaining all information about the changes higher than the discretization unit – all of them will be registered, while in the candlestick representation some of them could have been omitted. Due to those characteristics, the modelling results can be efficiently applied to the construction of algorithmic trade systems, which will be introduced in Section 5.

3 Model in binary-temporal representation

In line with the basic assumption of the technical analysis, investors make decisions according to some specific, statistically more frequent patterns [2,13]. This fact has been analyzed many times with the use of statistical and behavioural analysis [10], etc., and it does not raise much controversy today (the results obtained in this article also confirm the fact). State modelling considers the exchange rate as a process of transitions between some precisely defined states of the market. The statistically more frequent transitions between states represent characteristic patterns of the investors' behaviour. The state modelling of exchange rates in a binary-temporal representation consists of defining market states as specific binary sequences that represent historical price changes and some additional parameters describing e.g. the duration of change, etc.

The state model in binary-temporal representation (SMBTR) is one of the basic models dedicated to binary-temporal representation. The state in this model consists of two sets, the first one defined as the direction of m preceding course trajectory changes, while the second one consists of the value parameter p for n previous changes. Thus, the state can be written as:

$$T = \{\varepsilon_1 \dots \varepsilon_m; p_1 \dots p_n\}, \quad (2)$$

where the parameter p is calculated based on the change duration analysis (if the duration of i -th change is higher than the assumed threshold (thr), then the parameter takes the value 1, if not, the value 0):

$$p_i = \begin{cases} 1, & \Delta t_i \geq thr, \\ 0, & \Delta t_i < thr. \end{cases} \quad (3)$$

This way of appointing p allows for change duration analysis while at the same time maintaining a finite number of states. Based on the values of m and n parameters, we can calculate the number of states in the model:

$$k = 2^{m+n} \quad (4)$$

Therefore, the SMBTR model takes the following parameters: m , n and thr . The model will be further referred to as SMBTR (m,n,thr). Let us now consider the model in binary-temporal representation (SMBTR) (2,1,300). We will assume that the two last changes were decreases and the last change took 200 seconds – thus the market is now in state $\{\{0,0\},0\}$. If in the following 280 seconds an increase in the course occurs, the model will make a transition into a $\{\{0,1\},0\}$ state. On the other hand, if the change will take 340 seconds, the model will find itself in the state $\{\{0,1\},1\}$, and so on. It is important to note that the process presented in the article, i.e. the process constructed based on the binary-temporal representation, is not a Markov process [12]. This comes, among others, from the very definition of states, where each state (i.e. binary series of m changes), includes $m-1$ changes from the previous state.

Based on the historical data for the given financial instrument we can create a graph presenting the process of transitions between states, corresponding to the investors behaviour. The vertices of the graph depict states and the weights of edges present frequencies of transitions. Figure 1 presents an exemplary transition graph for SMBTR(2,1,300) model for XAU/USD exchange rate, from the period of 2014-2020.

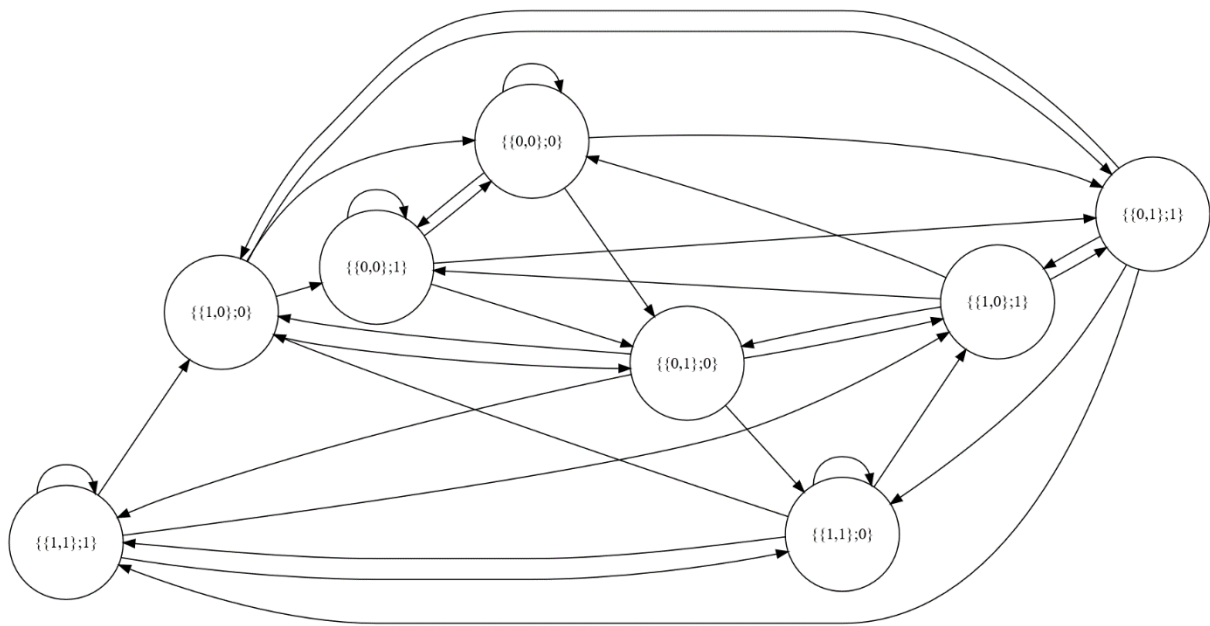


Figure 1 Graph of the transition process in SMBTR(2,1,300) model.

The frequencies are interpreted as estimators of the probability of transitions between states. Because the transition into each state is connected with a particular direction of the next change, based on the frequency analysis we can approximate the distribution of the future direction of a change (a detailed description of this process can be found in [14]). An exemplary distribution obtained for the SMBTR(2,1,300) model for XAU/USD exchange rate in binary-temporal distribution ($\delta = 350$) from the 01.01.2014-31.12.2017 period can be found in Table 1.

Table 1 Probability distribution for the direction of future course change in SMBTR(2,1,300) model, for XAU/USD.

State	Probability of an increase	Probability of a decrease	Number of occurrences
$\{\{0,0\},0\}$	0,53192	0,46809	1410
$\{\{0,0\},1\}$	0,49115	0,50885	452
$\{\{0,1\},0\}$	0,50946	0,49054	1427
$\{\{0,1\},1\}$	0,48073	0,51927	441
$\{\{1,0\},0\}$	0,47088	0,52912	1425
$\{\{1,0\},1\}$	0,50901	0,49091	444
$\{\{1,1\},0\}$	0,48585	0,51415	1449
$\{\{1,1\},1\}$	0,52439	0,47561	410

This distribution stands as a basis for the construction of an algorithmic trade system which will be described in the following section.

4 Construction of an HFT system based on a SMBTR model

Let us now consider algorithmic trade systems dedicated to binary-temporal representation. The main idea of their construction consists in assigning each course change of a given discretization unit length a notion of a separate transaction [12]. Each transaction is opened at the beginning of the change and closed at its end (that is the beginning of another one). The direction of the transaction is defined based on the probability distribution of the direction of a future change. If the occurrence of 1 in the binary representation is more probable, a BUY order is opened, and in case of 0 being more probable, a SELL order is made. Since each transaction ends with the occurrence of the next change, all transactions generate identical profit (in case of a right decision), or identical loss (in case of a wrong decision). If the i -th transaction ends with a profit, it makes the balance of the investor increase by:

$$S_i = S_{i-1} + (\delta - r) * v, \quad (5)$$

where S_i is the investor's account balance after an i -th transaction, r is the spread understood as the broker's provision, and v is the value of change equal to 1 pip. In the case of XAU/USD it equals:

$$v = 100 * l, \quad (6)$$

where l is the transaction size expressed in Lots. Because of the spread, the value of a lossy transaction is always higher and results in a balance change after i -th position:

$$S_i = S_{i-1} - (\delta - r) * v, \quad (7)$$

As an effect of the system performance, we obtain a series of transactions. The investor's profit results from the statistical dominance of the profitable transactions over the lossy ones. In this types of systems, because of the leverage of CFD contracts, one should take into account the risk of an investor's bankruptcy. In such kind of systems, the risk is calculated by the so-called maximal capital drawdown (*mdd*) [1,11]:

$$mdd(T) = \max_{t \in (0, T)} \left\{ \max_{s \in (0, t)} CR(s) - CR(t) \right\}, \quad (8)$$

where CR is the cumulative return rate. The measure that would allow for assessing which system is characterized by the best profit to risk ratio is the financial efficiency. One of the most popular efficiency indicators used by investors is Calmar's indicator. Its value can be calculated from [16]:

$$Calmar = \frac{E}{mdd}, \quad (9)$$

where E is the cumulative annual return rate.

5 Optimization of parameters for an algorithmic trade system

The financial efficiency of an algorithmic trade system depends on the approximation of the probability distribution calculated for the direction of a future course change, obtained based on the modelling. The direction distribution depends, on the other hand, on the applied discretization unit (δ) and the model parameters – in the case of considered SMBTR models – on m , n and thr parameters. For the purpose of this article, dedicated C++ software was created, which allows determining optimal parameters by testing all their possible combinations.

Historical tick data for XAU/USD instrument from Ducascopy broker were analyzed. The data was divided into two 3-year periods – first (01.01.2014-01.01.2017), where we appointed the distribution of a future change direction, and second (01.01.2017-01.01.2020), where we conducted the backtest and calculated financial efficiency. In the backtest, the investor's balance of 10000\$ was assumed, and the position size of 10 Lot (these values scale the return rate and maximal capital drop, and therefore have no influence on the obtained financial efficiency). We also assumed the fixed spread of 20 pips.

Due to the high number of parameters' combinations, and because of the equipment limitations (historical data weight about 6 GB and have to be computed in each iteration), the research was divided into two stages. In the first stage, we appointed a discretization unit and the value of m , for which the obtained system is characterized by the highest efficiency. In this research, the time analysis was not included ($n=0$). In the second stage, the time

analysis was introduced for the last change, and all possible threshold values (*thr*) were analysed, from the shortest to the longest registered change duration, by increasing the value by 5 sec in each iteration.

In the first part of the research, the highest financial efficiency equal to 0.22 was obtained for the discretization unit of $\delta=487$ and the BCMS model analysing the last three changes, i.e. $m=3$. In the second stage, after the time introduction, the most effective model was BCMS(3,1,12240), which was characterized by an efficiency of 0.68. Figure 2 presents the backtest results for the mentioned system.

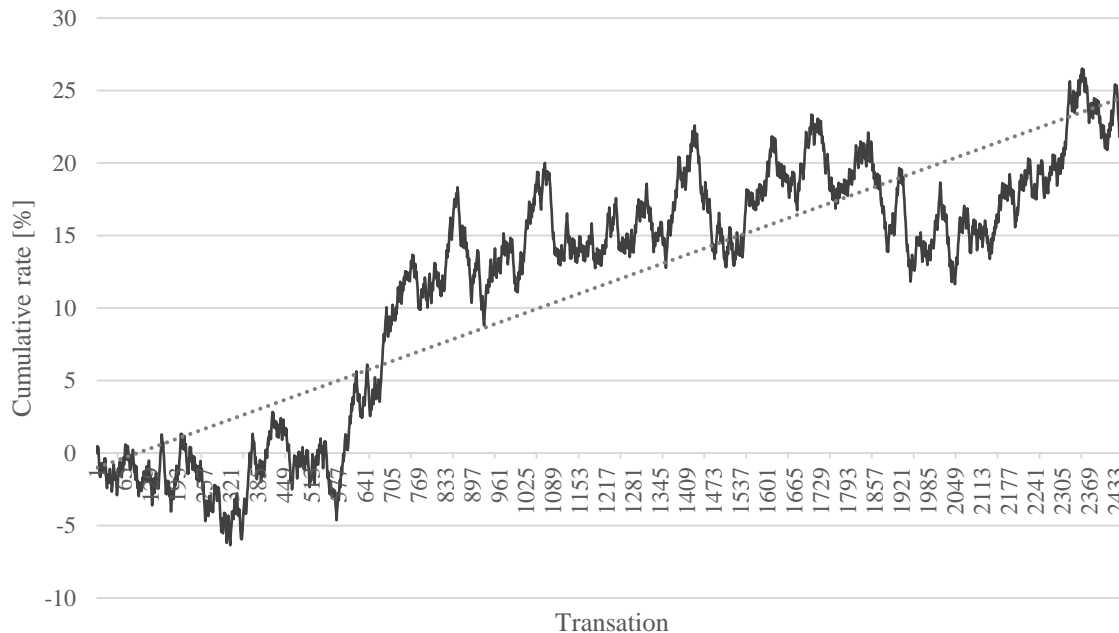


Figure 2 Backtest of the HFT system constructed based on SMBTR (3,1,583) model, for the binary-temporal representation ($\delta = 487$ pips) and XAU/USD

6 Summary

The article presents a schema for constructing an HFT system based on an analysis of a binary-temporal representation of data, with the use of state modelling. The state model in binary-temporal representation (SMBTR) was used for approximation of the probability distributions calculated for the directions of future changes of the exchange rate. Since the obtained financial efficiency is dependent on the assumed discretization unit and model parameters, the paper presents a schema for appointing optimal values of these parameters. To do so, dedicated software was created, allowing for an analysis of historical data from a period of six years, performed for one of the most popular financial instruments, that is XAU/USD.

Results presented in the article justify the validity of using the state modelling of an exchange rate in a binary-temporal representation in order to construct algorithmic trade systems that would be characterized by high financial efficiency. The research performed in two phases additionally justifies the use of analysis of course change duration in the modelling process. The analysis leads to a significant increase in the financial efficiency of the obtained algorithmic trade system.

The paper also presents modelling results and construction of the algorithmic trade system for the XAU/USD exchange rate. Yet, it is important to note here that the presented method for modelling, constructing a system and appointing parameters assuring the highest financial efficiency, is in fact universal and can be applied for an arbitrary financial instrument.

7 References

- [1] Aldridge I. (2009) *High-frequency trading: a practical guide to algorithmic strategies and trading systems*, John Wiley and Sons.
- [2] Alonso-Monsalve, S., Suárez-Cetrulo, A. L., Cervantes, A., & Quintana, D. (2020). Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. *Expert Systems with Applications*, 149, 113250.
- [3] Burgess, G. (2010). *Trading and Investing in the Forex Markets Using Chart Techniques* (Vol. 543), John Wiley & Sons.

- [4] De Villiers, V.: *The Point and Figure Method of Anticipating Stock Price Movements: Complete Theory and Practice*, Stock Market Publications, 1933.
- [5] 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.
- [6] Kirkpatrick, C. D., and Dahlquist, J. A., (2010). *Technical analysis: the complete resource for financial market technicians*. FT press.
- [7] Lo A. W., Mamaysky H., & Wang J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The Journal of Finance*, 55(4), 1705-1770.
- [8] Logue, D. E., and Sweeney, R. J.: *White noise in imperfect markets: the case of the franc/dollar exchange rates*. *The Journal of Finance*, 32(3), 761-768, 1977.
- [9] Neely, C. J., & Weller, P. A. (2011). *Technical analysis in the foreign exchange market*. Federal Reserve Bank of St. Louis Working Paper No.
- [10] Oberlechner, T. (2005). *The psychology of the foreign exchange market*, John Wiley & Sons.
- [11] Pardo R. (2011). *The evaluation and optimization of trading strategies*, John Wiley & Sons.
- [12] Piasecki K. & Stasiak M. D. (2019). *The Forex Trading System for Speculation with Constant Magnitude of Unit Return* , *Mathematics*, 7(7), 1-23.
- [13] Schlossberg, B. (2006) *Technical Analysis of the Currency Market*. John Wiley & Sons.
- [14] Stasiak, M. D. (2016). *Modelling of currency exchange rates using a binary-temporal representation*, Proceedings of: International Conference on Accounting Finance and Financial Institution: Theory meets Practice ICAFFI 2016, Springer.
- [15] Stasiak, M. D. (2020). Candlestick—The Main Mistake of Economy Research in High Frequency Markets. *International Journal of Financial Studies*, 8(4), 59.
- [16] Young, T. W. (1991). Calmar ratio: A smoother tool. *Futures*, 20(1), 40.

Possibilistic median of a fuzzy number

Jan Stoklasa¹, Pasi Luukka²

Abstract. Fuzzy sets, fuzzy numbers and other entities of fuzzy set theory have become a frequently used part of many economic models including predictive, decision-support and control ones. In many applications the need for the representation of a fuzzy number by a single characteristic may arise. Many analogies of the moments of random variables have been introduced as a direct analogy to the formulas used in probability theory, such as the center of gravity, variance of a fuzzy number, etc. For most of these characteristics the fully possibilistic counterpart has also been proposed. But not for all. This paper introduces the formulation of a fully possibilistic median. We apply the LSC transformation introduced by Luukka, Stoklasa and Collan in 2019 to the fuzzy number median formulation proposed by Bodjanova in 2005 and derive the fully possibilistic counterpart of the median of a fuzzy number, present the necessary formulas and discuss its possible use in economic and business applications.

Keywords: possibilistic mean, fuzzy number, LSC transformation, decision support, defuzzification

JEL Classification: C44, D81, G11

AMS Classification: 91B06, 03E72

1 Introduction

Fuzzy numbers and fuzzy sets have become an everyday part of economic, social sciences and human sciences models [16, 15, 19]. The tools of fuzzy set theory [6, 24] are being used in a variety of economic applications ranging from fuzzy control, through timeseries prediction [9], multiple-criteria decision-making [8], strategic management [12] to real options valuations [17], scientific outputs evaluations [11] and even to new methods for managerial [22] and questionnaire data summarization and analysis [21, 18, 20].

Particularly in expert systems and in applications where the fuzzy sets model uncertainty and imprecision, i.e. when the fuzzy sets can be considered disjunctive or epistemic according to Dubois and Prade [7], it makes sense to understand the membership functions of these fuzzy sets as possibility distributions over the given universe of discourse. The question of summarization of this information using the moments of the distribution therefore appears often in the practical applications and also in fuzzy control, where defuzzification is a necessary step of the control loop. The possibilistic moments have, however, not been investigated much in the literature, with several exceptions such as the proposal of possibilistic mean [5, 2] and variance [2], possibilistic correlation [4]. Alternatively the direct analogies of “probabilistic” moments and characteristics of distributions are being used. These usually simply standardize the membership function of the fuzzy set so that the area underneath it is equal to one and then apply the formulas used in standard statistics. The center of gravity of a fuzzy number is a very good example of such an approach. Even though technically there is nothing incorrect with the analogies of standard formulas for probability distributions in fuzzy set theory, it seems that the use of fully possibilistic measures would make much better sense. Even more so when the membership functions of fuzzy sets represent expert estimates and as such can really be treated as possibility distributions. It is this line of research that we want to contribute to in this paper.

Recently Luukka, Stoklasa and Collan have introduced the LSC transformation [14] that allows for the conversion of formulas from the “probabilistic” framework to the possibilistic one and vice versa. They have clearly shown in [14] the connection between the center of gravity of a fuzzy number and the possibilistic mean [2] as well as between the centre-of-gravity-based variance of a fuzzy number and the possibilistic variance of a fuzzy number proposed in [2]. In this paper we apply the LSC transformation to derive the formula for a possibilistic median using the median value of a fuzzy number proposed by Bodjanová in [1] as a departing point. As such this paper introduces the notion of a possibilistic median of a fuzzy number. We also discuss the reasonability of the obtained formula in the context of possibility theory. We thus contribute to the fuzzy MCDM and data analysis methods by

¹ LUT University, School of Business and Management, Yliopistonkatu 34, 53850 Lappeenranta, Finland, & Palacký University Olomouc, Faculty of Arts, Department of Economic and Managerial Studies, Křížkovského 12, 77900 Olomouc, Czech Republic, jan.stoklasa@lut.fi; jan.stoklasa@upol.cz

² LUT University, School of Business and Management, Yliopistonkatu 34, 53850 Lappeenranta, Finland, pasi.luukka@lut.fi

providing another fully possibilistic measure of location for fuzzy numbers and also confirm the usefulness of the LSC transformation in the process.

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 $A : U \rightarrow [0, 1]$. 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 $A(\cdot)$ is called a *membership function* of the fuzzy set A . Let A be fuzzy set on the same universe U . The set $\text{Ker}(A) = \{x \in U | A(x) = 1\}$ denotes the *kernel* of A , ${}^\alpha A = \{x \in U | A(x) \geq \alpha\}$ denotes an α -*cut* of A for any $\alpha \in [0, 1]$, $\text{Supp}(A) = \{x \in U | A(x) > 0\}$ denotes a *support* of A . $\text{Hgt}(A) = \sup\{A(x) | x \in U\}$ denotes a height of fuzzy set. The *core* of a fuzzy set is the set $\text{Core}(A) = \{x \in U | A(x) = \text{hgt}(A)\}$. The set of all fuzzy sets on $[r, s]$ will be denoted $\mathcal{F}([r, s])$; for any $A \in \mathcal{F}([r, s])$ we assume $A(x) = 0$ for $x \notin [r, s]$.

A *fuzzy number* is a fuzzy set A defined on the set of real numbers which satisfies the following conditions: $\text{Ker}(A) \neq \emptyset$ (A is *normal*); ${}^\alpha A$ are closed intervals for all $\alpha \in (0, 1]$; and $\text{Supp}(A)$ is bounded. A fuzzy number A is said to be defined on $[r, s] \subset \mathbb{R}$, if $\text{Supp}(A)$ is a subset of the interval $[r, s]$. The set of all fuzzy numbers on an interval $[r, s]$ will be denoted $\mathcal{F}_N([r, s])$. Real numbers $a_1 \leq a_2 \leq a_3 \leq a_4$ are called *significant values* of the fuzzy number A if $[a_1, a_4] = \text{Cl}(\text{Supp}(A))$ and $[a_2, a_3] = \text{Ker}(A)$, where $\text{Cl}(\text{Supp}(A))$ denotes a closure of $\text{Supp}(A)$.

To fully characterize a fuzzy number $A \in \mathcal{F}_N([r, s])$, $A \sim (a_1, a_2, a_3, a_4)$, we need to specify the shape of the left part of the membership function ($A_L(x)$) between a_1 and a_2 and the right part of the membership function ($A_R(x)$) between a_3 and a_4 . We can use the pseudo-inverse functions to $A_L(x) : [a_1, a_2] \rightarrow [0, 1]$ and $A_R(x) : [a_3, a_4] \rightarrow [0, 1]$ (that is, the functions $a_l(\alpha) : [0, 1] \rightarrow [a_1, a_2]$ and $a_u(\alpha) : [0, 1] \rightarrow [a_3, a_4]$, $\alpha \in [0, 1]$ respectively) to represent the fuzzy number A in the following way: $A = \{[a_l(\alpha), a_u(\alpha)]\}_{\alpha=0}^1$. In this representation ${}^\alpha A = [a_l(\alpha), a_u(\alpha)]$ for all $\alpha \in (0, 1]$ and ${}^0 A = [a_l(0), a_u(0)] = [a_1, a_4]$.

The fuzzy number A is called *trapezoidal* if its membership function is linear on $[a_1, a_2]$ and $[a_3, a_4]$ and $a_1 \neq a_4$. A 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 $A_G(x) = w_A \cdot A(x)$, $x \in U$. If we want to relax the assumption of linearity of the membership function and the normality of the fuzzy sets, we can define triangular-type and trapezoidal-type fuzzy sets in the following way. A *triangular-type fuzzy set* is a convex fuzzy set $A \in \mathcal{F}([r, s])$ such that $a_2 = a_3$, in other words its core contains a single value $x \in [r, s]$ such that $A(x) = \text{hgt}(A)$. A *trapezoidal-type fuzzy set* is a convex fuzzy set $A \in \mathcal{F}([r, s])$ such that $\text{core}(A) = [a_2, a_3]$, $a_2 \neq a_3$ and $A(x) = \text{hgt}(A)$ for all $x \in [a_2, a_3]$. For more information about fuzzy sets see e.g. [10].

Definition 1 (Median value of a fuzzy number [1]). The median value of a fuzzy number $A \in \mathcal{F}_N([r, s])$ is the real number $m_A \in \text{Supp}(A)$ such that

$$\int_r^{m_A} A(x)dx = \int_{m_A}^s A(x)dx. \quad (1)$$

The equation (1) can be also formulated in the following way:

$$\int_r^{m_A} A(x)dx = \frac{1}{2} \int_r^s A(x)dx = \frac{1}{2} \text{Card}(A). \quad (2)$$

In line with the usual statistical use the median of a fuzzy number is defined as the 50th percentile; in other words as the value from the universe of discourse that divides the fuzzy number into two fuzzy sets with equal cardinalities. A “probabilistic” analogy would be the value that splits the universe of discourse into two subsets with equal probabilities. Even though this formulation seems intuitive, it might not be so in the context of understanding the membership function of A as a possibility distribution. Note, that $A(m_A)$ does not need to be equal to 1, or in other words it is not guaranteed that $m_A \in \text{Ker}(A)$. From a possibilistic point of view one would expect that the median of a possibility distribution would split the distribution into two with identical possibilities. Note, that this understanding would imply that a possibilistic median should lie in the kernel of the fuzzy number.

Luukka et al. [14] introduced a two-way transformation between the analogies of formulas for probability distribution (random variables) and their possibilistic counterparts on the case of the center of gravity and the possibilistic mean [2] for triangular and trapezoidal fuzzy numbers. Note that this LSC transformation is not an integral transform, neither does it represent integration by substitution. The LSC transform is introduced as means of converting integral-based formulas for the moments of fuzzy numbers from the domain of probabilistic analogies to the possibilistic domain and vice versa.

Definition 2 (The Luukka-Stoklasa-Collan (LSC) transformation [13, 14]). The LSC transformation for the “ x to α ” direction can be defined for convex fuzzy sets in the following way (more details on the transformation and the related proofs are available directly in the source paper [14]).

Let us consider a convex fuzzy set $A \in \mathcal{F}([r, s])$, where $[r, s] = \text{Cl}(\text{Supp}(A))$, $A \sim (a_1, a_2, a_3, a_4)$. Then

$$T_{x \rightarrow \alpha} : \begin{cases} dx & \rightarrow d\alpha \\ x & \rightarrow a(\alpha) \\ A(x) = a^{-1}(x) & \rightarrow \alpha \\ \text{integration limits } [r, s] & \rightarrow [c, d], \end{cases} \quad (3)$$

where, $[c, d] = [\min\{a^{-1}(x)|x \in U_k\}, \max\{a^{-1}(x)|x \in U_k\}]$, U_k are the respective subintervals of $[r, s]$, i.e. $U_1 = [a_1, a_2]$, $U_2 = [a_2, a_3]$ and $U_3 = [a_3, a_4]$. In this formulation $a(\alpha) = a_l(\alpha)$ on the $[a_1, a_2]$ interval, $a(\alpha) = a_m(\alpha) = \text{core}(A) = [a_2, a_3]$ on the $[a_2, a_3]$ interval, $a(\alpha) = a_u(\alpha)$ on the $[a_3, a_4]$ interval, and ${}^\alpha A = [a_l(\alpha), a_u(\alpha)]$ for all $\alpha \in [0, \text{hgt}(A)]$. In other words the $a_l(\alpha)$ and $a_u(\alpha)$ functions define the left and right endpoints of the α -cuts of A .

Luukka et al. define the transformation also in the direction “ α to x ”, that is for the transformation of possibilistic formulas to their “probabilistic” counterparts. This direction is not necessary to recall in this paper, we therefore refer the interested readers to [14] for more details on the opposite direction of the transformation.

3 Deriving the possibilistic median of a fuzzy number

In this section we will apply the LSC transformation (3) of the definition of the median of a fuzzy number (1). The possibilistic median of a fuzzy number $A \in \mathcal{F}_N([r, s])$ will be denoted \bar{m}_A . We start with with the formulation for a triangular-type fuzzy number, i.e. we assume $A \sim (a_1, a_2, a_3, a_4)$ is a convex, normal fuzzy set on $[r, s]$ such that $a_2 = a_3$.

Theorem 1 (Possibilistic median of a triangular-type fuzzy number). *Let $A \in \mathcal{F}_N([r, s])$, $A \sim (a_1, a_2, a_4)$, be a triangular-type fuzzy number. Then $\bar{m}(A) = a_2$.*

Proof. As mentioned before, we are going to apply the LSC transformation to the formulation of the median of the fuzzy number proposed by Bodjanová and represented by (1) reformulated as (2):

$$\int_r^{m_A} A(x)dx = \frac{1}{2} \int_r^s A(x)dx.$$

Applying the LSC transformation $T_{x \rightarrow \alpha}$ to the RHS of the equation we get:

$$\frac{1}{2} \int_r^s A(x)dx \xrightarrow{(T_{x \rightarrow \alpha})} \frac{1}{2} \left(\int_0^1 \alpha d\alpha + \int_0^1 \alpha d\alpha \right) = \frac{1}{2}$$

To be able to apply the LSC transformation to the LHS of the equation, we need to consider two possible cases: $m_A \in [a_1, a_2]$ and $m_A \in (a_2, a_4]$.

- Let us start with the case when $m_A \in [a_1, a_2]$. Then

$$\int_r^{m_A} A(x)dx \xrightarrow{(T_{x \rightarrow \alpha})} \int_0^{A(m_A)} \alpha d\alpha = \left[\frac{\alpha^2}{2} \right]_0^{A(m_A)} = \frac{A^2(m_A)}{2}.$$

Putting the LHS=RHS we get

$$\frac{A^2(m_A)}{2} = \frac{1}{2} \Rightarrow A^2(m_A) = 1 \Rightarrow A(m_A) = 1.$$

- Next, let us consider $m_A \in (a_2, a_4]$. Then

$$\int_r^{m_A} A(x)dx \xrightarrow{(T_{x \rightarrow \alpha})} \int_0^1 \alpha d\alpha + \int_{A(m_A)}^1 \alpha d\alpha = \frac{1}{2} + \frac{1}{2} - \frac{A^2(m_A)}{2} = 1 - \frac{A^2(m_A)}{2}.$$

Putting the LHS=RHS we get

$$1 - \frac{A^2(m_A)}{2} = \frac{1}{2} \Rightarrow A^2(m_A) = 1 \Rightarrow A(m_A) = 1.$$

Regardless of the actual position of m_A , after applying the LSC transformation we get $A(m_A) = 1$. As applying the LSC transformation to the fuzzy number median formula gets us the possibilistic counterpart thereof, we can rewrite the condition as $A(\bar{m}_A) = 1$. Therefore \bar{m}_A has to lie in the kernel of A , in other words for a triangular-type fuzzy set we have $\bar{m}_A = a_2$. \square

Let us now consider a more general type of a fuzzy number, frequently denoted as fuzzy interval, in other words a trapezoidal-type fuzzy number. Along with triangular-type fuzzy numbers these mathematical entities are frequently used in economic models to represent expert estimates including uncertainty or to represent future values of projects, options, investments etc.

Theorem 2 (Possibilistic median of a trapezoidal-type fuzzy number). *Let $A \in \mathcal{F}_N([r, s])$, $A \sim (a_1, a_2, a_3, a_4)$, be a trapezoidal-type fuzzy number. Then $\bar{m}(A) \in [a_2, a_3]$.*

Proof. We will again be applying the LSC transformation to the formulation of fuzzy-number median represented by (2), that is to

$$\int_r^{m_A} A(x)dx = \frac{1}{2} \int_r^s A(x)dx.$$

Applying the LSC transformation $T_{x \rightarrow \alpha}$ to the RHS of the equation we get:

$$\frac{1}{2} \int_r^s A(x)dx \xrightarrow{(T_{x \rightarrow \alpha})} \frac{1}{2} \left(\int_0^1 \alpha d\alpha + \int_1^1 [a_2, a_3] d\alpha + \int_0^1 \alpha d\alpha \right) = \frac{1}{2}$$

To be able to apply the LSC transformation to the LHS of the equation, we need to consider three cases: $m_A \in [a_1, a_2]$, $m_A \in (a_2, a_3)$ and $m_A \in [a_3, a_4]$.

- For $m_A \in [a_1, a_2]$ the proof is identical to the respective part of the proof of Theorem 1 and we get $A(m_A) = 1$.
- For $m_A \in (a_2, a_3]$ we get for the LHS of the equation

$$\int_r^{m_A} A(x)dx \xrightarrow{(T_{x \rightarrow \alpha})} \int_0^1 \alpha d\alpha + \int_{A(a_2)}^{A(m_A)} \alpha d\alpha = \frac{1}{2} + \frac{A^2(m_A)}{2} - \frac{A^2(a_2)}{2} = \frac{1}{2} + \frac{A^2(m_A)}{2} - \frac{1}{2}.$$

Again putting LHS=RHS we get that $A(m_A) = 1$.

- Lastly for $m_A \in (a_3, a_4]$ we get for the LHS of the equation

$$\int_r^{m_A} A(x)dx \xrightarrow{(T_{x \rightarrow \alpha})} \int_0^1 \alpha d\alpha + \int_1^1 \alpha d\alpha + \int_{A(m_A)}^1 \alpha d\alpha = \frac{1}{2} + 0 + \frac{1}{2} - \frac{A^2(m_A)}{2} = 1 - \frac{A^2(m_A)}{2}.$$

Putting LHS=RHS we get

$$1 - \frac{A^2(m_A)}{2} = \frac{1}{2} \Rightarrow \frac{A^2(m_A)}{2} = \frac{1}{2} \Rightarrow A(m_A) = 1.$$

In all three instances we get the condition that $A(m_A) = 1$, and as we are in the possibilistic setting after the application of the LSC transformation, this means that $A(\bar{m}_A) = 1$ and thus $\bar{m}_A \in \text{Ker}(A) = [a_2, a_3]$. \square

Theorem 2 proposes that any value from the kernel of the trapezoidal-type fuzzy number can be considered to be its possibilistic median. Note, that if a possibilistic median is expected to be a possibilistic analogy to the fuzzy-number median (1) then \bar{m}_A needs to divide the support of A into two equally possible subsets.

As we can see, Theorems 1 and 2 specify the possibilistic median \bar{m}_A of a general convex fuzzy number $A \in \mathcal{F}_N([r, s])$ as any value from the Kernel of the respective fuzzy number. Given the fact that the possibility of a set is defined as the maximum of the possibilities of its elements, then defining the possibilistic median as a value that lies in the Kernel of the fuzzy number guarantees, that the intervals $[r, \bar{m}_A]$ and $[\bar{m}_A, s]$ both have the same possibility equal to 1. This result is intuitive and well in line with our expectations.

4 Discussion and conclusions

This paper introduces the fully possibilistic median of a fuzzy number. It starts from the definition of a median of a fuzzy number proposed by Bodjanová in [1] as a direct analogy to the statistical understanding of the median. We apply the LSC transformation proposed by Luukka et al. [14] to transform the formula for the calculation of the median to the possibilistic context. The possibilistic median of a fuzzy number A is then proposed to be any number $\bar{m}_A \in \text{Ker}(A)$. This intuitive result also provides another piece of evidence of the strength of the LSC transformation and its ability to link the characteristics of fuzzy numbers proposed in direct analogy to the characteristics of random variables to their fully possibilistic counterparts.

This way the paper introduces the possibilistic median of a fuzzy number as another moment of possibility distributions to be applied when the fuzzy number can be considered to be epistemic, that is when it represents for example an uncertain expert estimate of quantity or value or when it represents an estimate of a future value of a variable. As such it can provide an alternative to the possibilistic mean [2] in possibilistic portfolio optimization [25, 3, 23], creation on managerial summaries that are based on fuzzy monetary values [22] but also in the defuzzification in the context of fuzzy control and fuzzy expert systems.

Acknowledgements

This research would like to acknowledge the partial support by the Specific University Research Grant IGA_FF_2021_001, as provided by the Ministry of Education, Youth and Sports of the Czech Republic in the year 2021, by the Finnish Strategic Research Council, grant number 313396 / MFG40 - Manufacturing 4.0, and by LUT research platform ABMI- Analytics-based management for business and manufacturing industry.

References

- [1] Bodjanova, S.: Median value and median interval of a fuzzy number. *Information Sciences* **172** (2005), 73–89.
- [2] Carlsson, C., and Fuller, R.: On possibilistic mean value and variance of fuzzy numbers. *Fuzzy Sets and Systems* **122** (2001), 315–326.
- [3] Carlsson, C., Fullér, R., and Majlender, P.: A possibilistic approach to selecting portfolios with highest utility score. *Fuzzy Sets and Systems* **131** (2002), 13–21.
- [4] Carlsson, C., Fullér, R., and Majlender, P.: On possibilistic correlation. *Fuzzy Sets and Systems* **155** (2005), 425–445.
- [5] Dubois, D., and Prade, H.: The mean value of a fuzzy number. *Fuzzy sets and systems* **24** (1987), 279–300.
- [6] Dubois, D., and Prade, H., eds.: *Fundamentals of Fuzzy Sets*. Kluwer Academic Publishers, Massachusetts, 2000.
- [7] Dubois, D., and Prade, H.: Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets. *Fuzzy Sets and Systems* **192** (2012), 3–24. URL <http://dx.doi.org/10.1016/j.fss.2010.11.007>.
- [8] 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.
- [9] Jónás, T., Eszter Tóth, Z., and Dombi, J.: A Fuzzy Time Series Model with Customized Membership Functions. In: *Advances in Time Series Analysis and Forecasting* (Rojas, I., Pomares, H., and Valenzuela, O., eds.). Springer International Publishing AG, Cham, 2017, 285–298.
- [10] Klir, G. J., and Yuan, B.: *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, New Jersey, 1995.
- [11] Krejčí, J., and Stoklasa, J.: Fuzzified AHP in the evaluation of scientific monographs. *Central European Journal of Operations Research* **24** (2016), 353–370.
- [12] Kumbure, M. M., Tarkiainen, A., Luukka, P., Stoklasa, J., and Jantunen, A.: Relation between managerial cognition and industrial performance : An assessment with strategic cognitive maps using fuzzy-set qualitative comparative analysis. *Journal of Business Research* **114** (2020), 160–172.
- [13] Luukka, P., Stoklasa, J., and Collan, M.: Transformation of Variance to Possibilistic Variance and Vice Versa. In: *Advances in Fuzzy Logic and Technology 2017. IWIFSGN 2017, EUSFLAT 2017*. (Kacprzyk, J., Szmidt, E., Zadrozny, S., Atanassov, K. T., and Krawczak, M., eds.). Springer International Publishing AG, Cham, 2018, 456–467.

- [14] Luukka, P., Stoklasa, J., and Collan, M.: Transformations between the center of gravity and the possibilistic mean for triangular and trapezoidal fuzzy numbers. *Soft Computing* **23** (2019), 3229–3235.
- [15] Smithson, M.: Fuzzy Sets and Fuzzy Logic in the Human Sciences. In: *Fuzzy Logic in Its 50th Year* (Kahraman, C., Kaymak, U., and Yazici, A., eds.), volume 341. Springer international publishing, 2016, 175–186.
- [16] Smithson, M., and Verkuilen, J.: *Fuzzy set theory: applications in the social sciences*. Sage Publications, Thousand Oaks, London, New Delhi, 2006.
- [17] Stoklasa, J., Luukka, P., and Collan, M.: Possibilistic fuzzy pay-off method for real option valuation with application to research and development investment analysis. *Fuzzy Sets and Systems* **409** (2020), 153–169.
- [18] Stoklasa, J., Talášek, T., and Luukka, P.: Fuzzified Likert scales in group multiple-criteria evaluation. In: *Soft Computing Applications for Group Decision-making and Consensus Modeling* (Collan, M., and Kacprzyk, J., eds.), volume 357. Springer International Publishing AG, 2018, 165–185.
- [19] Stoklasa, J., Talášek, T., and Musilová, J.: Fuzzy approach - a new chapter in the methodology of psychology? *Human Affairs* **24** (2014), 189–203.
- [20] Stoklasa, J., Talášek, T., and Stoklasová, J.: Reflecting emotional aspects and uncertainty in multi-expert evaluation: one step closer to a soft design-alternative evaluation methodology. In: *Advances in Systematic Creativity: Creating and Managing Innovations* (Chechurin, L., and Collan, M., eds.). Palgrave Macmillan, 2019, 299–322.
- [21] Stoklasa, J., Talášek, T., and Stoklasová, J.: Semantic differential for the twenty-first century: scale relevance and uncertainty entering the semantic space. *Quality & Quantity* **53** (2019), 435–448.
- [22] 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).
- [23] Vercher, E., and Bermudez, J. D.: A possibilistic mean-downside risk-skewness model for efficient portfolio selection. *IEEE Transactions on Fuzzy Systems* **21** (2013), 585–595.
- [24] Zadeh, L. A.: Fuzzy sets. *Information and control* **8** (1965), 338–353.
- [25] Zhang, W. G., Wang, Y. L., Chen, Z. P., and Nie, Z. K.: Possibilistic mean-variance models and efficient frontiers for portfolio selection problem. *Information Sciences* **177** (2007), 2787–2801.

Asymmetric Transmission of Crude Oil Prices to Retail Gasoline and Diesel Prices in U.S. Market

Karol Szomolányi¹, Martin Lukáčik², Adriana Lukáčiková³

Abstract. Numerous studies dealing with the transmission of crude oil prices to retail gasoline prices indicate that retail gasoline prices respond more quickly when crude oil price rises rather than when it decreases. An explanation of the effect can be addressed to the oligopolistic coordination theory, the production and inventory cost of adjustment, and the search theory. The paper is studying the asymmetric transmission of crude oil prices to retail gasoline and diesel prices in U.S. market. An approach, based on the adjustment cost function in linear-exponential form is provided. An advantage of the approach is its linkage with the theory. Corresponding econometric specification of a gasoline and diesel price reaction function are derived. Estimating the specification, we do not reject the asymmetric effect in U.S. gasoline and diesel market. The weekly average U.S. retail price bias is about 0.15 cents per litre of gasoline and about 0.1 cents per litre of diesel.

Keywords: asymmetric transmission, retail fuel prices, linex adjustment cost function

JEL Classification: C26, C51, Q41

AMS Classification: 62P20

1 Introduction

Numerous studies dealing with the transmission of crude oil prices to retail gasoline prices indicate that retail gasoline prices respond more quickly when crude oil price rises rather than when it decreases, e.g. [10], [5], [6], [8], [7], [11] and [12]. Bacon [1] called this asymmetric retail gasoline price adjustment as "rockets and feathers" effect. His study was followed by paper of Borenstein et al. [2] who provide strong evidence of asymmetry in the US market between 1986 and 1992 in different stages of the production and distribution of gasoline.

Several theoretical explanations of the asymmetry have been proposed and tested. A short review is provided by Brown and Yücel [4] and Radchenko [10]. Borenstein et al. [2] suggest three possible explanations for the asymmetric response of gasoline prices:

- the oligopolistic coordination theory,
- the production and inventory cost of adjustment,
- the search theory.

The commonly used methods in given empirical studies are error correction models (ECM) and vector auto-regressive models (VAR); e.g. [10], [5], [6], [7] and [16]. In the paper, we provide an alternative empirical approach based on linex adjustment costs formulation [17]. The nonlinear reaction function of fuel prices is derived from the linex adjustment costs function form. Estimating the coefficients of the reaction function, one can verify the asymmetric reactions of fuel prices on price changes in crude oil. The approach allows to estimate an average weekly gasoline or diesel price bias per litre of gasoline or fuel. By the traditional approach, Szomolányi et al. [17] could not confirm the asymmetric reactions of fuel prices in Slovak market. However, by the linex approach, the asymmetries were confirmed. We prefer this approach, because in our opinion the character of fuel price making is rather discretionary, and it better corresponds to the process of minimizing price adjustment costs.

Our ideas are inspired by empirical studies analysing the U.S. and EMU monetary policy asymmetries provided by Surico [13], [14] and [15]. He responded on questions, whether central bankers weight differently positive and negative deviations of inflation, output, and the interest rate from their reference values. Such behaviour corresponds to a discretionary monetary policy making.

The aim of this paper is to verify the asymmetric transmission of crude oil prices to retail gasoline and diesel prices in the U.S. market using the linex approach. We will find that the weekly average U.S. retail price bias is about 0.15 cents per litre of gasoline and about 0.1 cents per litre of diesel.

¹ University of Economics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, karol.szomolanyi@euba.sk.

² University of Economics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, martin.lukacik@euba.sk.

³ University of Economics, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia, adriana.lukacikova@euba.sk.

The paper is divided into the sections as follows. The short literature review is in this section. From the problem of firm minimizing its adjustment costs in the linex form, the reaction function of retail gasoline and diesel prices is derived in the next section. Following section describes data and methodology used to estimate the coefficients of reaction function. The results of estimates are in the same section. The final section concludes.

2 Model

Consider that gasoline and diesel seller react asymmetrically on changes in crude oil price. His adjustment costs will be lower after the crude oil price rises and they will be higher, after the crude oil price lowers. Therefore, after the fashion of Surico [13], [14] and [15] we consider the adjustment costs function F is in the linex form:

$$F[p_t, E_{t-1}(c_t)] = \frac{-\gamma[p_t - kE_{t-1}(c_t)] + e^{\gamma[p_t - kE_{t-1}(c_t)]} - 1}{\gamma^2} \quad (1)$$

where p_t is retail gasoline or diesel price, c_t is crude oil price, k is technology coefficient and γ is an asymmetry coefficient. A negative value of the coefficient γ implies that a negative value of the difference $p_t - kE_{t-1}(c_t)$ causes higher costs to the price-maker than it would if γ were positive. The linex specification nests the quadratic form as a special case, so that applying l'Hôpital's rule twice when γ tends to zero results in a reducing in the loss function (1) to the following symmetric parameterization:

$$\lim_{\gamma \rightarrow 0} \{F[p_t, E_{t-1}(c_t)]\} = \frac{1}{2}[p_t - kE_{t-1}(c_t)]^2 \quad (2)$$

The fuel price-maker chooses p_t to minimize the cost function (1). The first-order condition with respect to p_t is in the form:

$$\frac{-1 + e^{\gamma[p_t - kE_{t-1}(c_t)]}}{\gamma} = 0 \quad (3)$$

Condition (3) is a general description of the reaction function of the fuel price-maker. Performing the second-order Taylor expansion of the exponential terms in (3), we gain:

$$p_t - kE_{t-1}(c_t) + \frac{\gamma}{2}[p_t - kE_{t-1}(c_t)]^2 + v_t = 0 \quad (4)$$

The remainder of the approximation is v_t and it contains terms of the third or higher orders of the expansion.

We solve equation (4) for p_t and, prior to generalised method of moment's estimation (GMM) of the short-run relation, we replace expected values with actual values and we take the first differences of the relation. In practise we estimate the following nonlinear specification:

$$\Delta p_t = k\Delta c_t - \frac{1}{2}\gamma\Delta[(p_t - kc_t)^2] + u_t \quad (5)$$

From (3), we can also express the average weekly gasoline or diesel price bias caused by the rockets and feathers effect, $\gamma < 0$. Assuming that oil shocks Δc_t is a normally distributed process with zero mean and variance σ^2 , taking the first differences, expected values and logarithms of (3) and after rearranging terms we gain the price bias in the form:

$$E(\Delta p_t) = -\frac{k^2\gamma}{2}\sigma^2 \quad (6)$$

3 Data, Methodology and Results

3.1 Data

Our analysis focuses on the U.S. fuel markets and all data are gathered from the U.S. Energy Information Administration website. We consider the relations between the spot price of West Texas Intermediate (WTI) light crude oil and retail fuel prices. The retail fuel prices are Weekly U.S. Regular All Formulations Retail Gasoline Prices and Weekly U.S. No 2 Diesel Retail Prices.

Crude oil and gasoline spot prices are collected at daily sampling frequency, while retail gasoline and diesel prices are available only at weekly frequency. The spot and retail prices of petroleum products are denominated in dollars per gallon, while the spot price of oil is expressed in dollars per barrel. According to the U.S. Energy Information Administration website, the retail fuel prices are collected every Monday, therefore the daily spot prices are aggregated to weekly to match their Monday spot values.

Our weekly dataset consists of 1503 observations of retail gasoline prices from the period 08/20/1990 – 07/15/2019, 1322 observations of retail diesel prices from the period 03/21/1994 – 07/15/2019 and 1509 observations of crude oil prices from the period 08/20/1990 – 07/15/2019.

3.2 Methodology

The orthogonality condition implied by the rational expectation hypothesis makes the general method of moments (GMM) a natural candidate to estimate equation (5) for both retail gasoline prices and diesel prices. A Newey–West estimator is used to provide the estimates of standard errors. The most important feature of the Newey–West estimator [9] is its consistency in the presence of both heteroskedasticity and the autocorrelation of unknown forms.

The one-period lags of the first differences of retail gasoline prices and crude oil prices were used as instruments in gasoline equation (5), i.e. $\Delta p_{t-1}^g, \Delta c_{t-1}$; the one-period lags of the first differences of retail diesel prices, retail gasoline prices and crude oil prices were used as instruments in diesel equation (5), i.e. $\Delta p_{t-1}^d, \Delta p_{t-1}^g, \Delta c_{t-1}$. By the orthogonality test of instruments, two model versions were estimated for all cases: one with an instrument and second without him. The difference in the corresponding J statistics is subject to χ^2 statistics. The corresponding results of the orthogonality test are in the Table 1. We do not reject the validity of instruments in all cases.

Due to the robustness, we estimated (5) with the different methods (ordinary least square OLS, two-stage least square 2SLS). Besides, the Breusch and Pagan [3] test confirmed that gasoline and diesel equations are correlated in all cases. The corresponding χ^2 distributed LM statistics are in the Table 2. Therefore, we also estimated both relations as system. Seemingly unrelated model (SUR), GMM, 2SLS and three-stage least square (3SLS) methods were used to estimate the system. By GMM, 2SLS and 3SLS estimates, the one-period lags of the first differences of retail gasoline prices and crude oil prices were used as instruments, i.e., $\Delta p_{t-1}^g, \Delta c_{t-1}$.

Method	Eq./Inst.	Δc_{t-1}	Δp_{t-1}^g	Δp_{t-1}^d
GMM	Gasoline	0.011737	0.011737	–
	Diesel	0.193302	0.03739	0.121524
2SLS	Gasoline	0.007508	0.007508	–
	Diesel	0.228668	0.070399	0.159209

Table 1 Orthogonality Test of Instruments

3.3 Results

The results of estimates are in the Table 2. The corresponding standard errors, computed with the Newey and West estimator [9], are in parentheses. Three asterisks denote the statistical significance at 1% level: two at 5% level and one at 10% level. All the estimates are statistically significant at 1% level, but the k coefficient at diesel 2SLS equation is significant at 10% level, and the γ coefficient at diesel 2SLS equation is significant at 5% level. Average weekly one-gallon gasoline and diesel price biases from the rockets and feathers effect are in cents.

From the Table 2, we state that OLS and SUR methods underestimate the asymmetry coefficient and price bias. This result is in the line with the result of the Slovak retail fuel market [17]. Using the GMM, 2SLS and 3SLS estimates, we estimate the average weekly gasoline price bias caused by the rockets and feather effect to be between 0.13 and 0.15 and the average weekly gasoline price bias caused by the rockets and feather effect to be between 0.07 and 0.08.

Estimating the system, we could compare the values of the estimated coefficients for gasoline and diesel equations. While the k coefficients are approximately equal, the estimated asymmetry coefficients γ are in absolute value higher in the gasoline equations. The estimated gasoline price biases are essentially higher than diesel biases as well. However, testing the linear coefficient restrictions, we could confirm the statistically significant higher absolute value of gasoline asymmetry coefficient γ at 10% level only by the 3SLS estimate.

Method	Equation	K	gamma	Instrument Set	R ² /J stat	LM	Bias
OLS	Gasoline	0.004145*** (0.000932)	-0.386711*** (0.016127)	-	0.949420	533.8481	0.002350
	Diesel	0.004955*** (0.000822)	-0.348729*** (0.011290)	-	0.943457		0.003029
GMM	Gasoline	0.022190*** (0.004323)	-0.862580*** (0.216680)	$\Delta p_{t-1}^g \Delta c_{t-1}$	0.011737	511.9801	0.148082
	Diesel	0.020424*** (0.007079)	-0.640156*** (0.218444)	$\Delta p_{t-1}^d \Delta c_{t-1} \Delta p_{t-1}^g$	0.335267		0.103942
2SLS	Gasoline	0.022126*** (0.004451)	-0.859281*** (0.221396)	$\Delta p_{t-1}^g \Delta c_{t-1}$	0.887053	497.0945	0.146659
	Diesel	0.018876* (0.010879)	-0.595338** (0.289097)	$\Delta p_{t-1}^d \Delta c_{t-1} \Delta p_{t-1}^g$	0.896620		0.082563
SUR	Gasoline	0.004205*** (0.000336)	-0.383911*** (0.004187)	-	0.949349	-	0.002402
	Diesel	0.004852*** (0.000349)	-0.346048*** (0.003945)	-	0.943433		0.002882
system GMM	Gasoline	0.021838*** (0.005128)	-0.849000*** (0.246194)	$\Delta c_{t-1} \Delta p_{t-1}^g$	0.000160	-	0.143236
	Diesel	0.020527*** (0.007734)	-0.644814*** (0.239499)				0.096117
system 2SLS	Gasoline	0.020992*** (0.002533)	-0.810544*** (0.114822)	$\Delta c_{t-1} \Delta p_{t-1}^g$	0.884954	-	0.126363
	Diesel	0.018728*** (0.005860)	-0.593231*** (0.158860)		0.896684		0.073606
system 3SLS	Gasoline	0.021146*** (0.002436)	-0.817718*** (0.112316)	$\Delta c_{t-1} \Delta p_{t-1}^g$	0.883670	-	0.129359
	Diesel	0.018794*** (0.005735)	-0.595045*** (0.156409)		0.896252		0.074355

Table 2 Estimates of the U.S. Gasoline and Diesel Reaction Function Coefficients

Note: Specification is (5). The standard errors are computed with the Newey and West estimator. The sets of instruments consist of combinations of one-period lags of the first differences of retail gasoline and diesel prices and crude oil prices, i.e., Δp_{t-1}^g , Δp_{t-1}^d , Δc_{t-1} . Three asterisks denote the statistical significance at 1% level: two at 5% level and one at 10% level. In the last but one column, LM statistics of the Breusch and Pagan test are computed. The average weekly one-gallon gasoline and diesel price biases from the rockets and feathers effect are in cents.

4 Conclusion

In the line with the earlier studies [2] we confirm the asymmetric transmission of crude oil prices to retail gasoline and diesel prices in the U.S. retail fuel market. Our paper differs from earlier studies using different methodological approach based on the linear adjustment cost function. The weekly average U.S. retail price bias is about 0.15 cents per litre of gasoline and about 0.1 cents per litre of diesel. This cognition may be useful for the theoretical discussion studying sources of asymmetries in retail fuel price-making process. It may also be useful for explanation of price rigidities, currently vehemently used in the New Keynesian models of business cycles.

The surprising result is that estimated diesel price bias is at first sight higher than gasoline. Such a result contradicts the New Zealand [7] and Slovak [17] retail fuel market results. According to the explanation of Liu et al. [7], we should expect the revert result: “The diesel price is not as competitive as that of petrol. As diesel is mainly used by the business sector, the results suggest that commercial customers are not as price sensitive as individual motorists. As a result, oil companies have been able to take advantage of the relatively inelastic demand for diesel to increase their profits.”

However, on the other hand, testing the linear coefficient restrictions, we cannot reject that gasoline and diesel asymmetry coefficients differ at 5% statistically significant level, supposing that the transmission of crude oil

prices to gasoline prices is as asymmetric as the transmission of crude oil prices to diesel prices. A possible explanation, why these asymmetries are roughly equal in the U.S. retail fuel market is that U.S. individual motorists have higher demand for diesel cars than motorists in Slovakia or New Zealand.

Acknowledgements

The Grant Agency of Slovak Republic - VEGA, supports this paper by grant no. 1/0211/21, "Econometric Analysis of Macroeconomic Impacts of Pandemics in the World with Emphasis on the Development of EU Economies and Especially the Slovak Economy" and by grant no. 1/0193/20, "Impact of spatial spillover effects on innovation activities and development of EU regions".

References

- [1] Bacon, R. (1991). Rockets and Feathers: The Asymmetric Speed of Adjustment of UK Retail Gasoline Prices to Cost Changes. *Energy Economics*, 13, 211–218.
- [2] Borenstein, S., Cameron, A. & Gilbert, R. (1997). Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes? *Quarterly Journal of Economics*, 112, 305–339.
- [3] Breusch, T. S. & Pagan, A. R. (1980). The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics. *The Review of Economic Studies*, 47, 239–253.
- [4] Brown, S. P. A. & Yücel, M. K. (2000). Gasoline and crude oil prices: Why the asymmetry? *Economic and Financial Policy Review*, 2000, Q3, 23–29.
- [5] Grasso, M. & Mannera, M. (2007). Asymmetric Error Correction Models for the Oil–Gasoline Price Relationship. *Energy Policy*, 35, 156–177.
- [6] Honarvar, A. (2009). Asymmetry in Retail Gasoline and Crude Oil Price Movements in the United States: An Application of Hidden Cointegration Technique. *Energy Economics*, 31, 395–402.
- [7] Liu, M., Margaritis, D. & Tourani-Rad, A. (2010). Is there an Asymmetry in the Response of Diesel and Petrol Prices to Crude Oil Price Changes? Evidence from New Zealand. *Energy Economics*, 32, 926–932.
- [8] Meyler, A. (2009). The Pass-Through of Oil Prices into Euro Area Consumer Liquid Fuel Prices in an Environment of High and Volatile Oil Prices. *Energy Economics*, 31, 867–881.
- [9] Newey, W. & West, K. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55, 703–708.
- [10] Radchenko, S. (2005). Oil Price Volatility and the Asymmetric Response of Gasoline Prices to Oil Price Increases and Decreases. *Energy Economics*, 27, 708–730.
- [11] Rahman, S. (2016). Another Perspective on Gasoline Price Responses to Crude Oil Price Changes. *Energy Economics*, 55, 10–18.
- [12] Sun, Y., Zhang, X., Hong, Y. & Wang, S. (2018). Asymmetric pass-through of oil prices to gasoline prices with interval time series modelling. *Energy Economics*, 78, 165–173.
- [13] Surico, P. (2007a). The Fed's Monetary Policy Rule and US inflation: The Case of Asymmetric Preferences. *Journal of Economics Dynamics and Control*, 31, 305–324.
- [14] Surico, P. (2007b). The Monetary Policy of the European Central Bank. *Scandinavian Journal of Economics*, 109, 115–135.
- [15] Surico, P. (2008): Measuring the Time Inconsistency of US Monetary Policy. *Economica*, 75, 22–38.
- [16] Szomolányi, K., Lukáčik, M. & Lukáčiková, A. (2020a). Are Slovak Retail Gasoline and Diesel Price Reactions on Crude Oil Changes Asymmetric? *Statistika: Statistics and Economy Journal*, 100, 138–153.
- [17] Szomolányi, K., Lukáčik, M. & Lukáčiková, A. (2020b). Asymmetric Retail Gasoline and Diesel Price Reactions in Slovak Market. *Ekonomický časopis*, 68, 115–133.

DEA Window Analysis of Engineering Industry Performance in the Czech Republic

Eva Štichhauerová¹, Miroslav Žížka²

Abstract. The article analyses the influence of the existence of a cluster organization in the engineering industry on the degree of efficiency of companies. For research purposes, the companies were divided into three groups. The first group of companies included members of the Czech Machinery Cluster which represent an organized cluster. The second group consisted of companies that do business in the same region as the organized cluster, but are not its members. The third group contained companies in the engineering industry that operate in other regions of the Czech Republic. The technical efficiency of companies in all these groups was examined using DEA window analysis for the period of 2009–2019. A total of 256 companies were examined. Differences between groups of companies were analysed using the Games-Howell multiple range test. The results of the analysis showed that the member companies of the cluster organization had the highest average level of efficiency throughout the monitored period. At the significance level of 5%, it was better than in the group of non-member companies in the given region for almost the whole period. The members of the cluster organization also achieved a significantly higher rate of efficiency growth. Since 2014, their average efficiency has been significantly higher compared to companies in other regions.

Keywords: Data envelopment analysis, window analysis, technical efficiency, cluster organization, economic value added, engineering industry.

JEL Classification: C61, L25, L64.

AMS Classification: 90B90, 90C90

1 Introduction

Clusters are a voluntary form of cooperation between companies, universities, research institutes and other professional organizations in a certain industry which is geographically concentrated in a certain area. The founder of cluster theory is considered to be Michael Porter, who published a book in the 1990s [9]. Clusters most often arise quite naturally on the basis of long-term formal and informal ties between these entities. In some countries, including the Czech Republic, efforts to institutionalize such a group are supported. The result of such efforts is the creation of a cluster organization that oversees and manages the joint activities of cluster members. A cluster organization has the form of a legal entity, such as an association. The members of a cluster organization are usually entities located in the territory of a natural cluster, but also other institutions from other regions. The cluster organization and the natural cluster exist in symbiosis. However, not all entities in the region where the natural cluster operates are members of a cluster organization.

There are many studies in the literature that look at the impact of clusters on economic performance and productivity. For example, one recent survey [5] examined the impact of agglomerations in the industry on economic performance, namely on wage levels in 28 European countries. The results showed a clear relationship between wage levels and the existence of strong clusters, and this link grew with the degree of cluster specialization. However, on the other hand, for example, the research [3] has found that the positive effect of a cluster on performance can occur only with a significant time lag. In the case of the cork industry the performance increased after more than 20 years of the cluster's existence. Kukalis [6] even concluded in research on clustered and non-clustered companies in the semiconductor and pharmaceutical industries that clusters have no positive effect on performance.

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.

Therefore, the aim of the research is to find out whether the performance of Czech engineering companies operating in natural and organized clusters differs from those that are not part of a cluster. That is, whether the existence of a cluster brings companies either direct economic benefits (in the case of cluster organizations) or positive externalities (in the case of natural clusters).

2 Theoretical background

Data envelopment analysis (DEA) is a non-parametric method used to evaluate the technical efficiency of homogeneous decision making units (DMUs). Such units can be, for example, companies in a certain industry, as it is in our research. The efficiency of each unit can be examined based on a larger number of inputs and outputs. DEA uses linear programming to find an empirical efficient frontier. The basic model with the assumption of constant returns to scale (CRS) was developed in 1978 by the author trio Charnes, Cooper and Rhodes, according to which it is referred to as the CCR model. In 1984, it was extended by the assumption of variable returns to scale (VRS). According to Banker, Charnes and Cooper, this model is called BCC, see for example [13]. Basic DEA models are static. It evaluates the effectiveness of units in a given period, or in several consecutive periods, but in each period separately.

In the case of evaluating the efficiency or performance of companies in a particular industry, it is desirable to monitor the development trend. For example, in our research, we attempt to find out how the performance of companies develops over time depending on whether they are members of a cluster or not. Dynamic DEA models, such as Malmquist index (MI) or window analysis, are suitable for such situations.

In the classic DEA model, the efficient frontier is different each year. The efficiency score is thus difficult to compare in time series. Therefore, most studies use the MI to analyse panel data, but even this technique may not fully account for technological developments. The traditional MI assumes technological progress. For example, it does not take into account industries where technological decline can be observed. Window analysis can therefore be considered a more suitable method for the analysis of panel data [8], [11]. The problem also occurs when the MI is calculated based on scores from the DEA window analysis. Then there is the problem of defining the efficient frontier of a given period [1].

DEA window analysis works on the principle of moving averages. Each unit in different periods is considered separate. This means that the efficiency of a unit in one period is measured with the efficiency of the same unit in another period, and moreover with the efficiency of other units. This increases the amount of data in the analysis, which is especially beneficial for small sets of units. This is precisely the situation in our research, where there are a small number of units forming the core of a cluster organization, see Chapter 3. This improves the discriminatory power of analysis [12]. Changing the width of the window (i.e. the number of periods included in the analysis) then allows you to perform the analysis both in terms of one period and in terms of several periods. This is a special case of sequential analysis, where the width of the window varies between one and all periods. On the other hand, it is assumed that there are no major technological changes within the windows. This is a general problem of this analysis, especially when combined with the MI approach, as it estimates technological change. To reduce this problem, a narrower window width can be used [1].

We assume N units ($n = 1, \dots, N$) that use r inputs to produce s outputs in time series T , where $t = 1, 2, \dots, T$. The sample has $N \times T$ observations. The DMU_n^t unit represents the observation of entity n in period t with r -dimensional input vector (1) and with s -dimensional output vector (2).

$$\mathbf{x}_t^n = (x_{1t}^n, x_{2t}^n, \dots, x_{rt}^n)' \quad (1)$$

$$\mathbf{y}_t^n = (y_{1t}^n, y_{2t}^n, \dots, y_{st}^n)' \quad (2)$$

Window analysis begins at time k ($1 \leq k \leq T$) with width w ($1 \leq w \leq T - k$), is denoted by k_w and has $N \times w$ observations. The matrices of inputs X_{kw} and outputs Y_{kw} are given by relations (3) and (4), see [1].

$$X_{kw} = (x_k^1, x_k^2, \dots, x_k^N, x_{k+1}^1, x_{k+1}^2, \dots, x_{k+1}^N, \dots, x_{k+w}^1, x_{k+w}^2, \dots, x_{k+w}^N) \quad (3)$$

$$Y_{kw} = (y_k^1, y_k^2, \dots, y_k^N, y_{k+1}^1, y_{k+1}^2, \dots, y_{k+1}^N, \dots, y_{k+w}^1, y_{k+w}^2, \dots, y_{k+w}^N) \quad (4)$$

By substituting these matrices of inputs and outputs into CCR (5) or BBC (6) models, in this case input-oriented, see for example [1], we obtain the results of DEA window analysis. The BCC model differs only in the summation

condition for the lambda vector elements, see (6). For each unit we obtain $w(T - w + 1)$ efficiency results. To interpret the results, it is then appropriate to calculate the arithmetic mean of all the values found [4].

$$\begin{aligned} \min_{\theta, \lambda} \theta &= \theta'_{kwt} \\ \theta x'_t - X_{kw} \lambda &\geq 0 \\ Y_{kw} \lambda - y_t &\geq 0 \\ \lambda_n &\geq 0 \quad (n = 1, 2, \dots, N \times w) \end{aligned} \quad (5)$$

$$\sum_{n=1}^N \lambda_n = 1 \quad (6)$$

3 Data and methodology

The engineering industry was chosen for the research because one of the first Czech Machinery Cluster organizations was founded in 2003. The research process can be divided into the following five steps.

Step 1: Creation of a database of companies in the industry – entities that operate in the NACE 251 (Manufacture of structural metal products) and NACE 28 (Manufacture of machinery and equipment) industries were included in the research. These two industries represent the core of the engineering cluster. Lists of companies were filtered from the MagnusWeb commercial database [2]. The entities were divided into three groups:

- members of the engineering cluster (group designated as CLU);
- companies operating in the same region and industry as the Czech Machinery Cluster (Moravian-Silesian, Olomouc and South Moravian regions), but which are not its members (group designated as REG);
- other companies in the same industries, but doing business in other regions of the Czech Republic (designation CZE).

The first group included 10 companies, the second 1,011 companies and the third 1,836 companies.

Step 2: Definition of inputs and outputs for DEA Analysis – due to continuity with previous research [10], capital resources (equity and liabilities) were chosen as inputs and revenues from own products and services and economic value added as outputs. Economic value added (furthermore EVA) was calculated according to the procedure of the Ministry of Industry and Trade [7].

Step 3: Data collection for individual groups of companies – inputs and outputs were determined for all three groups of companies listed in Step 1. The data source was balance sheets and profit and loss accounts obtained from the MagnusWeb database for the period 2009 to 2019. Unfortunately, data availability in a complete time series was limited. Small businesses are not required to disclose financial statements, however, even many larger companies do not comply with the legal obligation at all or publish data with a long time lag. Complete accounting information was obtained for 4 companies in the core of the cluster (CLU), 118 companies from the region where the cluster operates (REG) and for 134 companies doing business in other regions (CZE). Each company is represented in only one group.

Step 4: Calculation of the DEA window analysis score - for all companies listed in Step 3, efficiency scores were determined using the DEA window analysis method. Radial models, input-oriented with assumptions of constant (CCR) and variable returns to scale (BCC), were used. MaxDEA 7 Ultra software was used to calculate efficiency scores. The length of the window was chosen to be 2 years, taking into account the fact that engineering cannot be considered a rigid industry with a constant or slow change in technology. For this reason, a shorter window length was preferred. For each company, the efficiency score in a given year was determined as the average of two values of that year from all windows in which it was represented (i.e. from two adjacent windows). Furthermore, the average efficiency scores in the two-year rolling periods (2009–2010, 2010–2011, etc.) were determined for each company. The individual efficiency scores were then aggregated according to the affiliation of the companies to the CLU, REG or CZE groups. The result is the average efficiency scores of individual groups of companies, according to both CCR and BCC models.

Step 5: Comparison of efficiency scores among groups of engineering companies – the Games-Howell multiple range test was used to identify differences between the average efficiency scores in individual groups of companies. This test was chosen because the data in the individual groups do not have a normal distribution. In some

years, heteroskedasticity was also detected using the Levene's variance check test, and the sizes of the individual groups vary considerably. Games-Howell is a nonparametric test used to compare multiple observation groups. It is suitable for cases where the sizes of individual groups of data differ significantly. All tests were performed at significance level alpha 5%.

4 Research results

Table 1 shows the average efficiency scores for each group of companies in each year using the CCR model. The scores of individual companies arose as an average of efficiency measures from individual two-year time windows. It appears from Table 1 that the average efficiency of the members of the cluster organization was higher throughout the period than for other groups of companies. This means that it was higher both in comparison with companies operating in the region where the cluster organization is based and in comparison with companies operating in the same industry in other regions.

In terms of development over time, a gradual increase in average efficiency can be observed in the group of members of the engineering cluster until 2014, when it reached its peak. This was followed by a gradual decline until 2017. In the last two years, the efficiency of cluster companies has been growing again. In contrast, the development in the other two groups of companies was different. In the group of companies in the region of operation of the cluster organization, the efficiency rate decreased slightly during the period under review. In the group of other companies, it was almost constant throughout the period. The higher degree of fluctuations in the efficiency of clustered companies can be explained by the different success rates in obtaining contracts, taking into account the size of the sample. Using the Games-Howell multiple range test, it was found that, except in 2009, the average efficiency of the members of the engineering cluster was significantly higher throughout the period (at a significance level of 5%) compared to companies from the same region. However, the difference in average efficiency compared to other companies was not statistically significant. Likewise, no significant differences in efficiency rates were identified between non-clustered companies in a given region and in the rest of the Czech Republic.

Group	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mean
CLU	0.56	0.56	0.69	0.71	0.74	0.82	0.67	0.64	0.63	0.73	0.78	0.68
REG	0.48	0.46	0.48	0.46	0.43	0.41	0.42	0.40	0.40	0.40	0.41	0.43
CZE	0.44	0.45	0.45	0.45	0.46	0.47	0.46	0.46	0.45	0.45	0.45	0.45
Total	0.46	0.46	0.47	0.46	0.45	0.45	0.45	0.44	0.43	0.43	0.44	0.45

Table 1 Average efficiency score by groups of companies in individual years (CCR Model)

Table 2 shows the average efficiency scores over two-year time windows. Compared to the previous table, this jumps fluctuations in efficiency. However, this analysis also shows that in the entire period under review, the members of the cluster organization achieved the highest efficiency, both in comparison with other companies in the region in which the cluster organization operates and in the rest of the Czech Republic. However, based on the Games-Howell test, only the differences between the scores of the companies in the core of the cluster and in the given region of operation proved to be statistically significant, starting from the 2010–2011 window until the end of the monitored period. No statistical differences were identified between companies in the given region and in the rest of the Czech Republic in terms of average efficiency.

Group	2009– 2010	2010– 2011	2011– 2012	2012– 2013	2013– 2014	2014– 2015	2015– 2016	2016– 2017	2017– 2018	2018– 2019	Mean
CLU	0.53	0.66	0.70	0.69	0.82	0.73	0.67	0.62	0.66	0.78	0.69
REG	0.46	0.49	0.46	0.45	0.41	0.42	0.42	0.40	0.41	0.40	0.43
CZE	0.44	0.39	0.40	0.44	0.53	0.52	0.47	0.46	0.45	0.43	0.45
Total	0.46	0.44	0.43	0.45	0.48	0.48	0.45	0.43	0.44	0.42	0.45

Table 2 Average efficiency score by groups of companies and windows (CCR Model)

In terms of development over time, an increase in average performance can be observed for companies in the core of the cluster until the period 2013–2014. Then the average efficiency decreased. In the last two time windows, however, the level of efficiency grew again and began to approach the peak of the period 2013–2014.

The next two Tables 3 and 4 show the results for the BCC models. Looking at the average efficiency scores of groups (see Table 3), it is clear that the average efficiency of companies in the core of the cluster organization in

all periods exceeds the efficiency of other companies both in the region in which the cluster organization operates and in the rest of the Czech Republic. However, these differences were not always statistically significant. While companies from the core of the cluster achieved a statistically significantly higher score throughout the observed period in comparison with other companies in the given region, this was true only in comparison with companies from the rest of the Czech Republic since 2014. Among companies in the region and in the rest of the Czech Republic, no statistical differences were identified in terms of efficiency.

Group	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mean
CLU	0.71	0.69	0.71	0.75	0.84	0.95	0.89	0.86	0.89	0.91	0.94	0.83
REG	0.48	0.49	0.54	0.56	0.53	0.52	0.55	0.55	0.55	0.55	0.57	0.53
CZE	0.48	0.49	0.48	0.49	0.50	0.51	0.52	0.52	0.52	0.50	0.48	0.50
Total	0.48	0.49	0.51	0.53	0.52	0.52	0.54	0.54	0.54	0.53	0.53	0.52

Table 3 Average efficiency score by groups of companies in individual years (BCC Model)

Looking at the average efficiency scores of groups in the time windows (see Table 4), it is clear that the average efficiency of companies in the core of the cluster organization in all periods grew faster than in other companies in the region in which the cluster organization operates and in other regions of the Czech Republic. These differences were not always significant, though. The average efficiency of clustered companies was significantly higher compared to other companies in the region throughout the period under review. However, in comparison with companies from other regions, it was higher starting from the 2013–2014 window. In the case of companies operating in the region of cluster organizations and in other regions, it can be stated that the growth of their average efficiency was almost the same.

Group	2009– 2010	2010– 2011	2011– 2012	2012– 2013	2013– 2014	2014– 2015	2015– 2016	2016– 2017	2017– 2018	2018– 2019	Mean
CLU	0.69	0.72	0.73	0.75	0.96	0.91	0.87	0.88	0.90	0.93	0.83
REG	0.49	0.52	0.53	0.56	0.52	0.53	0.56	0.54	0.56	0.56	0.54
CZE	0.50	0.42	0.44	0.47	0.51	0.54	0.54	0.54	0.53	0.49	0.50
Total	0.50	0.47	0.49	0.52	0.52	0.54	0.55	0.54	0.55	0.53	0.52

Table 4 Average efficiency score by groups of companies and windows (BCC Model)

The results show that members of a cluster organization would have to reduce their inputs by an average of 17%, at a given level of outputs, in order to achieve full efficiency. For non-member companies operating in the cluster region, the reduction in inputs would have to reach 46%. Companies operating in the engineering industry in other regions would then have to halve their inputs in order to become efficient.

5 Conclusions

The research results showed that the member companies of the cluster organization had a significantly higher average efficiency rate for the whole period under review (in the case of the CCR model only the initial year 2009 was an exception) than other companies operating in the same region but that are not members of the cluster organization. From this it can be concluded that from the very beginning the cluster organization rather associates more successful companies in the industry. In contrast, the average efficiency of companies in a region with a cluster organization did not differ from that of a group of companies in other regions. The member companies of the cluster organization also show the fastest growth in efficiency. The average efficiency of cluster companies improved by about 3% per year (in the case of both models). In contrast, for non-members of the cluster organization in the region, it almost stagnated. In the case of the CCR model, an average decrease of 1.6% per year was even found. In the group of other companies from other regions, only a slight growth of the efficiency rate was found (by about 0.1% to 0.2% per year). Due to the faster growth rate of efficiency in the group of cluster members, the average efficiency rate of this group is higher even in comparison with the group of companies operating in other regions since 2014. This is true for the BCC model. In a situation of CRS, the difference between these groups of companies is not significant. However, the assumption of VRS is closer to the real situation.

In a previous study [10], using the MI, it was found that in the period of 2009–2016, companies in a cluster organization showed the strongest growth. However, the differences compared to other groups of companies were not significant. At the same time, it was found that the desired effect could occur with a certain time delay. After extending the time series for another three years and applying window analysis to increase discriminatory power,

especially for a small sample of cluster organizations, it was confirmed that the efficiency rate of cluster members is improving faster than other groups of companies. From 2014, it is better compared to both groups of companies.

It can therefore be stated that the existence of an engineering cluster brings companies a direct economic benefit – i.e. increased efficiency. However, it has not been confirmed that it would bring positive externality to other companies outside the cluster organization. Given that the tendency of the development of efficiency in the time series of 2009–2019 is stable for other groups of companies (REG, CZE), it is not very realistic to expect that the effect of the cluster's existence will manifest itself with a further time lag.

It is also necessary to draw attention to the limits of the research. They are mainly affected by the availability of financial statements, which is worse from year to year. Thus, only relatively small samples of companies can be examined. In terms of further research, it will be interesting to monitor the development of average efficiency in other industries in which there are institutionalised or natural clusters and compare them with the results of the engineering industry.

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] Asmild, M., Paradi, J. C., Aggarwall, V. & Schaffnit, C. (2004). Combining DEA Window Analysis with the Malmquist Index Approach in a Study of the Canadian Banking Industry. *Journal of Productivity Analysis*, 21(1), 67–89.
- [2] Bisnode: *Magnusweb: Komplexní informace o firmách v ČR a SR* [online]. Bisnode ČR, Prague, 2021 [cit. 2021-03-08]. Available at: <https://magnusweb.bisnode.cz>
- [3] Branco, A. & Lopes, J. C. (2018). Cluster and business performance: Historical evidence from the Portuguese cork industry. *Investigaciones de Historia Económica*, 14(1), 43–53.
- [4] Dlouhý, M., Jablonský, J. & Zýková, P. *Analýza obalu dat*. Prague: Professional Publishing.
- [5] Ketels, C. & Protsiv, S. (2021). Cluster presence and economic performance: A new look based on European data. *Regional Studies*, 55(2), 208–220.
- [6] Kukalis, S. (2010). Agglomeration Economies and Firm Performance: The Case of Industry Clusters. *Journal of Management*, 36(2), 453–481.
- [7] MPO: *Finanční analýza podnikové sféry za rok 2018*. Ministry of Industry and Trade, Prague, 2019.
- [8] Oh, D. & Heshmati, A. (2010). A sequential Malmquist–Luenberger productivity index: Environmentally sensitive productivity growth considering the progressive nature of technology. *Energy Economics*, 32(6), 1345–1355.
- [9] Porter, M. E. (1990). *The competitive advantage of nations*. New York: Free Press.
- [10] Štichhauerová, E. & Žižka, M. (2020). Impact of Cluster Organizations on Financial Performance in Selected Industries: Malmquist Index Approach. In S. Kapounek & H. Vránová (Eds.), *38th International Conference on Mathematical Methods in Economics 2020. Conference Proceedings* (pp. 572–578). Brno: Mendel University in Brno.
- [11] Wu, D., Wang, Y. & Qian, W. (2020). Efficiency evaluation and dynamic evolution of China's regional green economy: A method based on the Super-PEBM model and DEA window analysis. *Journal of Cleaner Production*, 264, 121630.
- [12] Yang, H.-H. & Chang, C.-Y. (2009). Using DEA window analysis to measure efficiencies of Taiwan's integrated telecommunication firms. *Telecommunications Policy*, 33(1–2), 98–108.
- [13] Zhu, J. (2014). *Quantitative Models for Performance Evaluation and Benchmarking* (Vol. 213). Springer International Publishing.

Work Contour Models in Projects

Tomáš Šubrt¹, Jan Bartoška², Daniel Chamrada³

Abstract. The paper describes work contour models in project management and offers new mathematical concepts of this phenomenon. The authors focus on the course of resource's work effort in the project activities, the exact mathematical formulation of which is not often stated in the literature. In a brief introduction of the paper, the authors deal on common work contours in project management or another work contour of the human resource in working teams. The core of the paper will be analysis and mathematical formulations of specific work contours in projects managed on the basis of the Critical Chain method and using agile approaches. The results of the paper contain brief case study for hypothetical use in praxis. The paper aims at the proposal new concept of the mathematical model for the work contours, based on variable scenarios of resource behavior.

Key words: Project management, Work Contour, Resource Allocation, Critical Chain, Agile approaches, Mathematical model.

JEL Classification: C44

AMS Classification: 90C15

1 Introduction

Project management is currently probably the most effective tool for managing changes within organizations. At least partial knowledge and ability to apply project management is required of managers at all levels, as this skill represents a strong competitive advantage and the ability to work efficiently with resources. As Abuhantash [1] states, absolutely key part of the resources in projects are human resources, characterized by their necessity (due to the management and execution of actions) and at the same time limitedness (in terms of disposition and expertise). In addition, human resources may cause problems based on people's behaviour that may negatively affect systems development project productivity, and project outcomes [8]. Human resources are thus an important part of all projects and, due to their nature, need to be approached individually during the project planning process.

When performing a simple time analysis of the project, it is advisable to choose adequate tools that prevent the above-mentioned problem of human behaviour. Generally known Critical Path Method cannot always properly implement delays in projects, however, the more sophisticated Critical Chain Method is able to include to the project and also respond to them [3]. This possibility of using time reserves is mentioned in the Theory of Constraints [5]. Delays in projects, which require the creation of these time reserves, are then usually caused by three main factors: 1) No activity ends before it was originally planned; 2) The planned reserve for the fulfilment of the task is exhausted before its beginning and; 3) Something will always fail, but everything will never succeed. These formulations are generally used in the Theory of Constraints used to denote 1) Parkinson's design law; 2) Student's syndrome and; 3) Murphy's law.

In this article, we focus mainly on the second of these concepts, which is the student's syndrome, generally associated with the term procrastination. The research of procrastination and its elimination in university or higher education courses is dealt with by Tadepalli et al [9] and Korstange [7]. The Student Syndrome can occur during any human activity, especially when fulfilling tasks during the work on project. However, as Smith [8] states, Student Syndrome not only affects students, but can also occur in various fields of work (such as IT, banking, ...) An analysis of the ability to work with resources in the context of the impact of Student Syndrome on the inclusion of work schedules can provide insights for more effective human resource management in projects. Belout [2] states that human resources are today almost the only reason that causes the project to fail. On the other hand, it should be noted that the proportion of successfully completed projects is growing.

The importance of human resource management in project management is pointed out by Dwivedula [4], who focuses on research on topics linking project management and human resource management. The author also draws attention to the individual behaviour of people in projects and mentions the need to monitor their work efforts.

¹ Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, subrt@pef.czu.cz.

² Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, bartoska@pef.czu.cz.

³ Czech University of Life Sciences Prague, Department of Systems Engineering, Kamýcká 129, Prague, chamrada@pef.czu.cz.

Zwikael [10] also deals with the possibility of the monitoring of key employees on the project, which is a necessary part of its control. He states, that control is used as the monitoring mechanism to ensure that each of planning and execution is properly implemented, corrective actions being introduced where there are undesired discrepancies between the project's plan and its execution. The presented approach suggests mathematical models for Work Contour (WC) and Student Syndrome phenomenon and their use for modifying EVM (especially for computation of BCWS and ACWP), with use Cone of Uncertainty.

2 Mathematical expression of standard work contours

The progress of work on project activities is almost never uniform (flat). During the work on the task work effort changes. It can increase or decrease linearly or nonlinearly, both planned or unplanned. However, it should always be met, that the amount of work performed on the task (work, expressed for example in man-days), is retained. The graphical expression of it is a curve of the work effort in time - $p_0(t)$, usually with functional values from the interval $\langle 0,1 \rangle$, because the resource can never be used from more than 100 %. In the majority of work contour consideration, we suppose at least one moment (or interval), where work effort is maximal. A possible expression on the y-axis is also the immediate consumption of work (work effort) in hours or units of resource used. Due to the requirement of maintaining the total amount of work on a task (expressed by the area under this curve), it should always be true that the integral of the curve:

$$\int_0^1 p_0(t)dt = 1 \quad (1)$$

Contemporary project management uses seven basic uneven work contours:

1. **Front loaded**, where the work effort increases linearly, usually from a non-zero initial value.
2. **Back loaded**, where the work effort decreases linearly, usually to a non-zero final value.
3. **Early Peak**, with a linear increase up of work effort to about 1/3 of the duration, then decreasing to the initial value
4. **Late Peak**, with a linear increase up of work effort to about 2/3 of the duration, then decreasing to the initial value
5. **Turtle**, where the work effort concavely increases to about 1/2 duration and similarly decreases to the initial value
6. **Bell**, where the work effort increases in the shape of the S-curve about 1/2 of duration and again decreases similarly to the initial value
7. **Double Peak**, where the work effort increases in the shape of the S-curve about 1/4 duration and again decreases similarly to the initial value in 1/2, then repeatedly in the same shape increase to 3/4 and again decreases similarly to the initial value

Sometimes the work contour Double Peak is still modelled, with a linear increase to about 1/3 of the duration, then with a decrease to the initial value in about 1/2 duration, then with a repeated increase to about 2/3 of the duration, then with a decrease to the initial value. The initial value is usually 10% of the maximum work effort. By linear, trigonometric and polynomial approximation, under the conditions of relative duration $t \in \langle 0; 1 \rangle$ and work effort $p(t) \in \langle 0; 1 \rangle$ these work contours can be approximated by the following functions:

$$1 \quad \text{Flat} \quad p_0 = 1 \quad (2)$$

$$2 \quad \text{Back Loaded} \quad p_0 = 0,8t + 0,2 \quad (3)$$

$$3 \quad \text{Front Loaded} \quad p_0 = -0,8t + 1 \quad (4)$$

$$4 \quad \text{Late Peak} \quad p_0 = 1,2t + 0,2 \text{ for } t \in \langle 0; \frac{2}{3} \rangle, p_0 = -2,4t + 2,6 \text{ for } t \in \langle \frac{2}{3}; 1 \rangle \quad (5)$$

$$5 \quad \text{Early Peak} \quad p_0 = 2,4t + 0,2 \text{ for } t \in \langle 0; \frac{1}{3} \rangle, p_0 = -1,2t + 1,4 \text{ for } t \in \langle \frac{1}{3}; 1 \rangle \quad (6)$$

$$6 \quad \text{Turtle} \quad p_0 = -(1,948t - 0,974)^4 + 1 \quad (7)$$

$$7 \quad \text{Bell} \quad p_0 = 0,9\sin^4\pi t + 0,1 \quad (8)$$

$$8 \quad \text{Double Peak} \quad p_0 = 0,9\sin^42\pi t + 0,1 \quad (9)$$

Kinser [6] gives top 10 List paradigm (laws) influencing good practices in project management, three of which have a significant impact on the real work contour or work effort itself: Parkinson's Law: "Work expands to fill the time available." originally stated as "Work expands so as to fill the time available for its completion," Saint Exupéry's Law: "Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away." It means that misguided attempts to exceed customers' expectations can cause problems ranging from padding estimates through tying up resources needed elsewhere, to sacrificing other projects of more value to the organization. Kinser's Law: "About the time you finish doing something, you know enough to start." You cannot consider a project (task) finished unless you learn from it. It can cause an instantly repairing task immediately before it ends.

3 Analysis of linear work contours

As mentioned above, linear work contours include "Back loaded", "Front Loaded", Early Peak "and "Late Peak". The first two are expressible by a linear function, the second two by a linear polynomial function.

- 1) Work Contour Back Loaded – It is generally assumed that the initial value of work effort is 0.2 (20 %), and with a slope of the curve $38,66^\circ$ ($k = \tan \alpha = 0,8$). It reaches full effort (100%) upon completion of the activity, i.e. in the value of $t = 1$. The area representing the total workload of the task defined by this function can be determined by simple trigonometric relations. Compared to the value 1 for Flat, it is reduced to the value $W = 0,6$, i.e. to 60 % of the overall workload. In order to achieve the original workload, it is (again using simple trigonometric relations) to extend the duration of the task to 1.67 and reduce the slope of the curve by about 13° , i.e. to the value $k = 0,48$, i.e. to adjust the function to the shape:

$$p_1 = 0,48t + 0,2 \quad (10)$$

- 2) Work Contour Front Loaded – Also here, it is generally assumed that the initial value of work effort is 1 (100 %), and with a negative slope of the curve $-38,66^\circ$ ($k = \tan \alpha = -0,8$) it reaches the final value of 0.2 (20 %), at completion of the activity, i.e. in the value $t = 1$. The area representing the total workload of the task defined by this function can be determined by simple trigonometric relations. Compared to the value 1 for Flat, it is reduced to the value $W = 0,6$, i.e. to 60 % of original workload. In order to achieve the original workload, it is necessary to extend the duration of the task to 1.67 and change the slope of the curve by about 13° , ie to the value $k = -0,48$, i.e. to adjust the function to the shape

$$p_1 = -0,48t + 1 \quad (11)$$

- 3) Work Contour Late Peak – This type of work contour is expressed by a fractional linear function. It is generally assumed that the start and end values of work effort are 0.2 (20 %), peak deployment 1 is reached in 2/3 of the task. The slope of the first part of the fractional linear function is about $50,2^\circ$ ($k = \tan \alpha = 1,2$), the second part of the fractional linear function is then about $-67,38^\circ$ ($k = \tan \alpha = -2,4$). Similarly, the area representing the total workload of the task defined by this function can be determined by simple trigonometric relations, with the proviso that, compared to the value 1 at Flat, it is reduced to the value $W = 0,6$, i.e. to 60 %. In order to achieve the original workload, it is therefore necessary to extend the duration of the task again to 1.67 and change the shape of the contour linear polynomial function to

$$p_1 = 0,72t + 0,2 \text{ for } t \in \langle 0; 1,111 \rangle, p_1 = -1,44t + 2,6 \text{ for } t \in \langle 1,111; 1,667 \rangle \quad (12)$$

- 4) Work Contour Early Peak – This type of work contour is again expressed by a fractional linear function. It is generally assumed that the start and end value of Work Effort is 0.2 (20 %), peak deployment 1 is reached in 1/3 of the task. The slope of the first part of the polynomial is about $67,39^\circ$ ($k = \tan \alpha = 2,4$), the second part of the polynomial is then about $20,2^\circ$ ($k = \tan \alpha = -1,2$). Similarly, the area representing the total workload of the task defined by this function can be determined by simple trigonometric relations, with the proviso that, compared to the value 1 at Flat, it is reduced to the value $W = 0,6$, i.e. to 60 %. In order to achieve the original workload, it is therefore necessary to extend the duration of the task again to 1.67 and change the shape of the contour linear polynomial function to

$$p_1 = 1,44t + 0,2 \text{ for } t \in \langle 0; 0,555 \rangle, p_1 = -0,72t + 1,4 \text{ for } t \in \langle 0,555; 1 \rangle \quad (13)$$

For all standard linear work contours, the task duration is extended by approx. 2/3 when the unit workload is required. This extension is achieved by changing the slope of the contour curves.

4 Analysis of non-linear work contours

As mentioned above, non-linear work contours include "Turtle", "Bell", and "Double Peak". The authors decided to slightly modify this type of work contours due to greater accuracy and more accurate analysis compared to their linear concept using fractional linear functions. The first WC is expressed by a polynomial function of the 4 degree, the other two are formalized by trigonometric functions. Originally our approach was based on using spline functions, but later on these formulas have been rearranged using simulations and experimenting while strictly preserving all the properties of generally defined work contours. In contrast to linear WCs, the coefficient q is used to transform the function into the contour function of the unit area, changing the shape of the function while maintaining its course on the newly created interval.

- 1) Work Contour Turtle – This WC assumes a gradual concave increase in work effort from an initial value of 0.1 (10 %) to full deployment at about 1/3 of the duration and a subsequent decrease from about 2/3 of the duration to a target value of 10 % at the end. The area representing the total workload of a task defined by this function can be determined using a defined integral

$$\int_0^1 p_0(t)dt = \int_0^1 (-(1,948t - 0,974)^4 + 1)^4 dt = 0,102669(1,948t - 0,974)^5 = 0,102669(1,948x - 0,974)^5 \quad (14)$$

Compared to the value 1 for Flat, it will be reduced to a value of approx. $W = 0.91$, ie to 91 %. In order to achieve the original workload 1, it is necessary to introduce the coefficient q into the equation WC, ie

$$p_0 = -(1,948qt - 0,974)^4 + 1 \quad (15)$$

and solve the equation

$$\int_0^{Ts} p_1(t)dt = \int_0^{Ts} (-(1,948qt - 0,974)^4 + 1)^4 dt = 0,102669(1,948qt - 0,974)^5 = 1 \quad (16)$$

where for the value $q = 0.91$ the duration of the task will be extended by approx 8 % - $Ts \cong 1,08$.

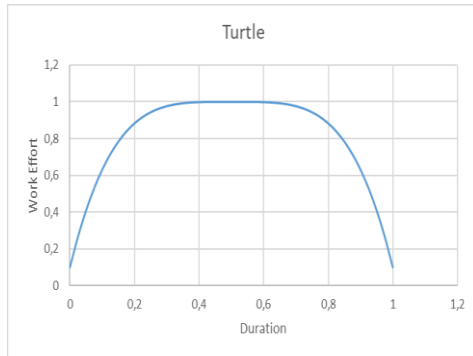


Figure 1 Original WC Turtle

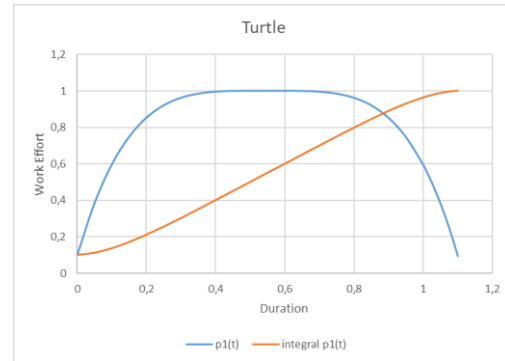


Figure 2 WC Turtle with unit amount of work

- 2) Work Contour Bell – This WC assumes a gradual increase in work effort from an initial value of 0.1 (10%) to full deployment in 1/2 duration and a subsequent immediate rapid decline to a target value of 10% at the end. The area representing the total workload of a task defined by this function can be determined using a defined integral

$$\int_0^1 p_0(t)dt = \int_0^1 (0,9\sin^4\pi t + 0,1)dt = 0,4375t + 0,9\left(-\frac{1}{4\pi}\sin 2\pi t + \frac{1}{32\pi}\sin 4\pi t\right) = 0,102669(1,948t - 0,974)^5 = 0,102669(1,948x - 0,974)^5 \quad (17)$$

Compared to the value 1 for Flat, it will be reduced to a value of approx. $W = 0.4375$, ie to 43.75 %. In order to achieve the original workload 1, it is necessary to introduce the coefficient q into the equation WC, ie

$$p_0 = 0,9\sin^4\pi qt + 0,1 \quad (18)$$

and solve the equation

$$\int_0^{Ts} p_1(t)dt = \int_0^{Ts} (0,9\sin^4\pi qt + 0,1)dt =$$

$$= 0,4375t + 0,9\left(-\frac{1}{4\pi q} \sin 2\pi qt + \frac{1}{32\pi q} \sin 4\pi qt\right) = 1 \quad (19)$$

where for the value $q = 0.4375$ the duration of the task will be extended by approx. 130 % - $Ts \cong 2,3$.

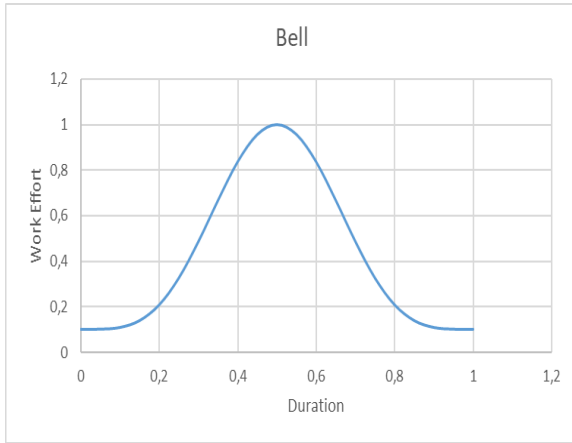


Figure 3 Original WC Bell

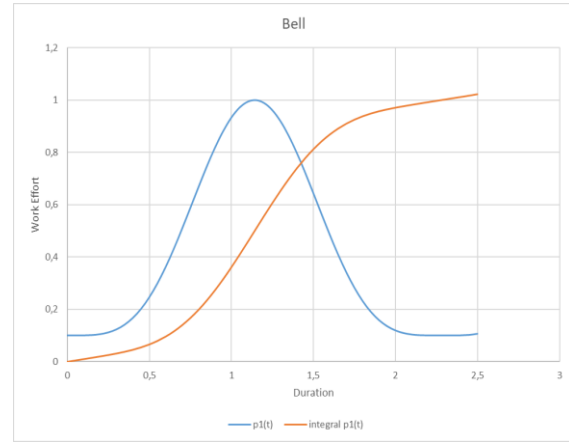


Figure 4 WC Bell with unit amount of work

- 3) Work Contour Double Peak – This WC assumes a gradual increase in work effort from an initial value of 0.1 (10%) to full deployment in 1/4 of the duration and a subsequent immediate decrease to 10% in the middle of the duration, then again an increase to full deployment in 3/4 of the time duration and a decrease again to 10% at the end. The area representing the total workload of a task defined by this function can be determined using a defined integral

$$\int_0^1 p_0(t)dt = \int_0^1 (0,9\sin^4 2\pi t + 0,1)dt =$$

$$= 0,4375t + 0,9\left(-\frac{1}{8\pi} \sin 4\pi t + \frac{1}{64\pi} \sin 8\pi t\right) \quad (20)$$

Compared to the value 1 for Flat, it will be reduced again to a value of approx. $W = 0.4375$, ie to 43.75%. In order to achieve the original workload 1, it is necessary to introduce the coefficient q into the equation WC, ie

$$p_0 = 0,9\sin^4 2\pi qt + 0,1 \quad (21)$$

and solve the equation

$$\int_0^{Ts} p_1(t)dt = \int_0^{Ts} (0,9\sin^4 2\pi qt + 0,1)dt =$$

$$= 0,4375t + 0,9\left(-\frac{1}{8\pi} \sin 4\pi qt + \frac{1}{64\pi q} \sin 8\pi qt\right) = 1 \quad (22)$$

where for the value of $q = 0.4375$ the duration of the task will be extended again by approx. 128.5 % - $Ts \cong 2,285$.

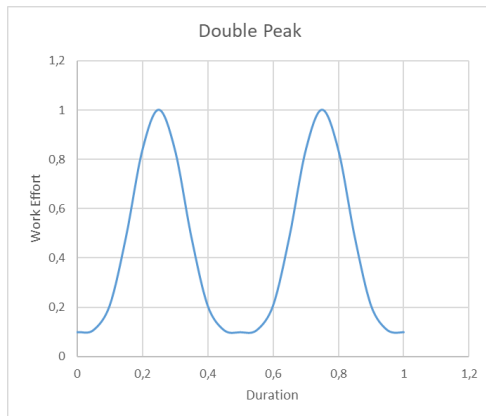


Figure 5 Original WC Double Peak

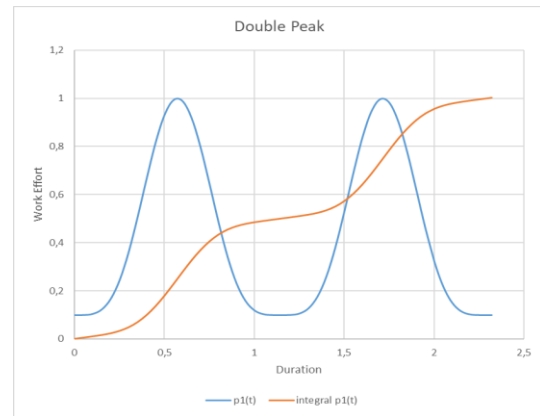


Figure 6 WC Double Peak with unit amount of work

5 Conclusion and future research

This paper deals with the possibility of mathematical formalization of work contours in projects using linear and nonlinear contour functions. The aim was to formulate such a relationship between the shape of the contour function and its cut-out surface so that it is possible to put them into context and thus prepare the basis for subsequent multi-criteria optimization models of work contours in projects with respect to quality and cost of work. A possible reduction in the planned workload in favor of a smaller extension of the task presupposes a deterioration in the quality of the work performed. It would thus be possible to formulate a multi-criteria optimization problem, enabling to find a compromise solution between extending the duration of the task and the quality of the work performed. Another possibility of using the proposed formulations of work contours is resource leveling models, e.g. in resolving resource conflicts, or in optimizing the use of resources over time.

References

- [1] Abuhantash, A., Dokras, U. V., Lewin, J. (2019). Human Resources in Project Management A Critical Analysis of existing literature and evolving factors. *Management* 7(3), 12-21.
- [2] Belout, A., Gauvreau, C. (2003). Factors influencing project success: the impact of human resource management. *International Journal of Project Management* 22(2004), 1-11
- [3] Deacon, H., Lingen, E. (2015). The Critical Path and Critical Chain Methods usage in the South African Construction Industry. *Oorsigartikels* 22(2015), 73-95,
- [4] Dwivedula, R. (2019). Human Resource Management in Project Management: Ideas at the Cusp. *European Project Management Journal* 9(1), 34-41.
- [5] Kalender, Z., Günay, N., Vayvay, Ö. (2014). Theory of Constraints: A Literature Review. *Social and Behavioral Sciences* 150(10), 930-936.
- [6] Kinser, J. (2008). *The top 10 laws of project management*. PMI® Global Congress 2008—North America, Denver, CO. Newtown Square, PA: Project Management Institute Inc.
- [7] Korstange, R., Craig, M., Duncan, M. D. (2019). Understanding and Addressing Student Procrastination in College. *The Learning Assistance Review* 1(24), 57-70
- [8] Smith, D. (2010). The Effects of Student Syndrome, Stress, and Slack on Information Systems Development Projects. *Issues in Informing Science and Information Technology* 7(2010), 489-494.
- [9] Tadepalli, S. R., Pryor, M. W., Booth, C. (2009). *Evaluating academic procrastination in a personalized system of instruction based curriculum*. American Society for Engineering Education.
- [10] Zwikael, O., Globerson, S., Raz, T. (2000). Evaluation of models for forecasting the final cost of a project. *Project Management Journal* 31(1), 53-57.

Visualising HR data concerning the performance of university staff - career development assessment and assistance perspective

Tomáš Talášek¹, Jana Stoklasová², Jan Stoklasa³, Jana Talašová⁴

Abstract. The employees of universities are, without a doubt, their most valuable resource. They require appropriate care, guidance, support and proper management. All of this is conditioned by the existence of proper information and decision support that is available to the superiors of the staff members and to the management of these tertiary education institutions in general. Recent changes in the tertiary education legislation and in the requirements for the accreditations of study programmes facilitated the development and improvement of quality assurance procedures and systems in universities. Nowadays the information systems of universities store large amounts of data that can be accessed by the management, but might be in a difficult to understand form. This paper aims on the support of development and management of academic staff members at universities through novel visualization of HR data concerning the activities and performance of these staff members. It assumes the existence of a staff evaluation system of the IS HAP type and suggests how career paths of the academics can be tracked in time and how the need for training/development can be identified from the already available data.

Keywords: Human resource management, university, evaluation models, visualization, performance, trajectory, efficiency.

JEL Classification: O15,C44

AMS Classification: 91B84, 91B06

1 Introduction

With the new legislation that is governing the tertiary education sector in the Czech Republic [1] and with the requirements introduced by the university funding methodologies [7], quality assurance has become a crucial activity at universities. Even though the importance of this area was unquestionable before the new legislation, the assessment of the existence and quality of quality assurance systems and processes at universities being one of the inputs of the funding process of tertiary education institutions has motivated universities to gather data in a higher quality and in a more structured form than before [3, 6]. The requirement of a working quality assurance information system has also facilitated the adoption of systems like IS HAP [15] to store and process data on the activities and performance of academic staff members of the universities. The wide adoption of the information systems for managerial decision support also in the tertiary education sector has motivated a lot of research in the field of linguistic decision support [8], human resources data processing [6, 14, 9, 10], and generally in the area of providing summaries of available data to the decision-makers or people in charge [11] in an understandable and well designed form [2].

In this paper, we are going to assume the availability of data on the activities of academic staff members in one place (system). We are not directly assuming that IS HAP has to be used for decision support in academic staff evaluation and management for the methods of data visualisation proposed in this paper to be applicable. Although the below listed assumption fit the evaluation methodology and data processing techniques used in IS HAP, they are considered as general requirements for all our suggestions and comments to be relevant. We simply assume (without a significant loss of generality) that:

- There are two areas of activities that are being evaluated - pedagogical activities (PA) and research, development and creative activities (RDC).

¹ Palacký University Olomouc, Faculty of Education, Department of Mathematics, Žižkovo nám. 5, 77140 Olomouc, Czech Republic, tomas.talasek@upol.cz.

² LUT University, School of Business and Management, Yliopistonkatu 34, 53850 Lappeenranta, Finland

³ LUT University, School of Business and Management, Yliopistonkatu 34, 53850 Lappeenranta, Finland, & Palacký University Olomouc, Faculty of Arts, Department of Economic and Managerial Studies, Křížkovského 12, 77900 Olomouc, Czech Republic

⁴ Palacký University Olomouc, Rector's Office, Strategy and Quality Control Office, Křížkovského 8, 77900 Olomouc, Czech Republic

- Each of the areas can be further divided into subareas that cover specific families of activities the grouping of which makes sense in the context of the evaluation or decision support being provided. For example, if we simplify the descriptions of subcategories of activities that IS HAP uses, we get
 - PA1 - direct teaching and assessment of students
 - PA2 - supervision of students
 - PA3 - activities supporting the teaching - development of the fields of study, teaching-related projects, study materials development etc.
 - RDC1 - higher-level research and development outcomes (publications, patents etc.) reflected in the funding of the universities or otherwise considered important for the institution
 - RDC2 - other research and development outcomes (publications, other types of R&D outputs) that are either not reflected in the funding of the university, are generally considered lower-level, or are associated with not easily verifiable sources of data
 - RDC3 - activities supporting research and development - obtaining research funding, involvement in the scientific community, editorial activities, reviewing activities, conference and seminar organization etc.
 - RDC4 - outputs of creative activities registered in the Registry of artistic performances (RUV, [12, 15])
- the activities (their value for the institution) are quantified in some way. It is not necessary to use the same units for the quantification of the activities in different areas.
- different requirements can be set for different academic positions - this allows us to reflect different academic expectations/responsibilities of the positions and also possibly different salary levels of these positions. IS HAP uses the so called standardized evaluations, that is evaluations expressed in terms of multiples of the standard score set specifically for each academic position and for each area of interest.
- an aggregation of the evaluations in separate areas of activities is available. This aggregation should reflect the potentially different nature of quantification in different areas of activities, it should allow for compensation of performance among the activities, if desired, it should reflect the needs of the decision support (i.e. of the users of the decision support, of the management of the university etc.).
- the data is being gathered and stored in a time invariant manner to allow for intertemporal comparison, for the construction of “time series” of data for the purpose of the analysis of development of performance of staff members, the identification of irregularities, problems or need for support, and for the purpose of predictions thereof, if needed.

Note that it is the availability of an overall evaluation, that can reflect the needs of the institution as well as the level to which one area can compensate for the other, that makes it possible to properly reflect the compensability of the evaluated areas in the graphical representations that we will discuss in this paper. Also the availability of “raw evaluations” as well as “standardized evaluations” makes it possible to reflect the different requirements on different academic positions when needed (then standardized evaluations are used), but also to compare the simple amount of activities performed by the academic staff members across all the academic positions.

In this paper we will assume the non-repressive point of view and we will be investigating how the data on the performance of academic staff members can be visualized and analyzed in order to understand the development of the staff in time, to identify well performing staff members (potentially performing at a level sufficient for a higher academic position), to identify the top-performing staff members in each academic position that can be used as role models or to get an idea whether resources (e.g. in terms of wages) are being efficiently distributed. We will point out some of the shortcomings of the usual time series visualization and analysis for the above-mentioned purposes and suggest methods for getting the necessary insights. We will start from the individual performance development level and then move to the efficiency and relative comparison of performance among the staff members of a given unit.

In the practical examples of visualization, we will be using artificial data representing an artificial university unit. The data does not describe any real-life unit, but is constructed in such a way that it represents a viable piece of data that could be obtained for academic staff members on the chosen positions. The data consists of 5 years of observations (denoted 2016-2020) of the performance of 13 academic staff members denoted “staff member 1” to “staff member 13”. We assume that the raw scores are available for all of them for PA, RDC and also for PA1, PA2, PA3 and RDC1, RDC2, RDC3 and RDC4. We also assume that standardized evaluations for PA and RDC and overall evaluations are available. We will be presenting graphical summaries of the dataset. Its full presentation is not necessary for the purpose of this paper given the space restrictions for the proceedings.

2 Career paths of academic staff members - tracking staff performance in the context of academic freedom

The tertiary education sector has its specific features when it comes to HR management. Academic staff members have the so called academic freedom to choose what the focus of their research will be, to express their scientific opinion, to select the desired publication channels and in many cases also to choose the form and content of their lectures as well as the topics of the theses being supervised by them. Although the freedom is not limitless and has to be understood in the context of the needs of the unit/institution where the staff member works, it creates potential problems in the evaluation of the staff and in defining the precise content of their work. For the units/institutions to function best and to be competitive, different academic staff members that have the same academic position can be required/encouraged to focus on the teaching and research areas in a different way - either equally, or to focus more on the area in which the given academic staff member performs better. The “optimal focus” can vary from year to year and might also depend on the development of the skills of the staff. Evaluation systems designed for the decision support of the managers of academic staff need to reflect this aspect and need to allow for compensability of performance in one area of interest by the performance in the other and, at the same time, need to allow for the shift of focus from one area to the other depending on the needs of the unit/institution. Although this is well possible to achieve [4, 9], it creates problems in data visualization. Time series describing the performance of staff members in isolated areas might no longer be informative enough. Visualizing more time series for the same staff member (e.g. one for PA, one for RDC) is possible but if the compensability of the performance in these areas cannot be described by a simple linear relationship, the effect of simultaneous changes might not be simple to interpret. With the exception of the simultaneous increase in performance in both areas that should be interpreted as a positive development (until a certain point when the problem of overheating and burnout can occur) and the simultaneous decrease in performance in both areas, which can be interpreted as undesirable. Note that, for example, IS HAP uses linguistic fuzzy rule bases to define the compensability and to derive the aggregation function for PA and RDC [13].

Figure 1 shows the time series of the evaluations of staff members 10 and 12 in the two evaluated areas (PA and RDC) between the years 2016 and 2020. It also shows their overall evaluations obtained on a [0,2] benefit-type scale. It is clearly visible that the predictability of the overall evaluation is not very high based solely on the visual inspection of the time series for PA and RDC evaluations. The reason for this is that the overall evaluation might be obtained by complex evaluation function that reflects a nonlinear compensability relationship of the performance in the two evaluated areas. As such the individual time series are not a good summary of the performance of the staff members.



Figure 1 The usual time series presentation of the development of performance in two evaluated areas (PA and RDC) for two selected academic staff members (staff member 10 and 12) and the development of their overall performance score measured on a [0,2] benefit type scale.

If we consider the nature of the time series in detail, we can realize that presenting each time series individually

for the staff members may even be misleading and borderline incorrect. Given the academic freedom and the variable focus on one area or the other that can change from one year to the other, it does not make sense to view the time series individually. Instead we should understand the data as a two-dimensional time series for each staff member. The performance of the staff member needs to be seen as a 2-tuple of values, as only these two values provide a “complete” view of all the activities performed by the academic staff member during the period that is being evaluated. The changes in performance should therefore be understood as a movement of this 2-tuple, in other words represented as a movement of a point in a 2-dimensional Cartesian space. This movement can define a “*performance trajectory*” of the staff member in time, as shown in Figure 2 in the right subplot. The trajectory can be also enhanced with the overall evaluation to reflect the compensability of the two areas and other possible nonlinearities - see the left subfigure in Figure 2.

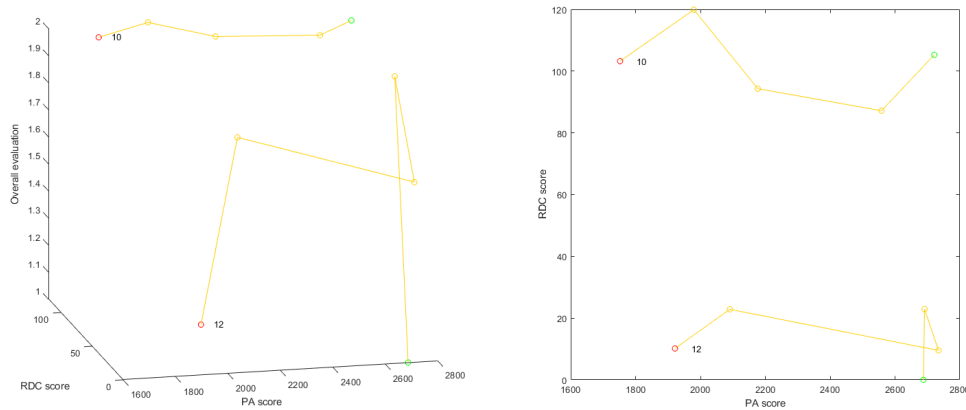


Figure 2 The visualization of the 2-dimensional performance time series (performance trajectories) for staff members 10 and 12. Each trajectory starts in the numbered point, the number is the index of the staff member, the year corresponding with the starting point is 2016. The subsequent points on the trajectory represent the subsequent years’ evaluations/scores. The right subplot represents just the PA-RDC visualization, the left subplot adds the overall evaluation dimension to see the effect of the change of the 2-dimensional performance value on the overall evaluation, thus reflecting the compensability of the evaluated areas and the performance therein.

Both staff members whose performance is visualized in Figures 1 and 2 are assistant professors. We can see that while their performance in PA is comparable throughout the years, staff member 10 clearly outperforms staff member 12 in RDC. For both of them the teaching load is increasing in time and the performance in RDC is more or less the same. Regardless of the fluctuations staff member 10 has a constant maximum overall evaluation in all the analysed years whereas the overall evaluation of staff member 12 is fluctuating a lot (see the left subfigure of Figure 2). The performance trajectories tell us that the teaching load is increasing for both staff members and yet they manage to keep their performance in RDC at a more or less constant level. On the other hand the last years performance of staff member 12 plummets to one, which might require an intervention by his/her superior. Apart from performance trajectories, we can construct also curves that would represent “*career paths*” - in this case the 2-tuples might not consist of PA and RDC scores, but instead might focus on those subareas that are relevant for professional advancement, for example RDC1 representing top-tier research and PA2 representing the supervision of students. In this case the connection with the overall evaluation is not straightforward any more, as the overall evaluation reflects all the activities, but it might still be needed to track whether a staff member aiming for a habilitation or full professorship retains the needed level of publication and student supervision without compromising other duties (and thus still having reasonably high overall evaluation). In this sense the performance trajectories and career paths provide much needed insights into the development of the staff in time. Such insights cannot be easily obtained by analysing the single time series. We therefore strongly suggest the use of career paths and performance trajectories for the visualization of performance data, particularly in the tertiary education sector. We intend to further investigate the possibilities of this representation of performance development visualization and to develop further analytical methods for this purpose.

3 Performance assessment of individuals in the context of the unit/institution

In many cases the assessment of individual performance is difficult, as there might be underlying effects that affect all the staff members in the given unit. In this perspective even a worsening of performance in both evaluated areas

might not be considered as a very undesirable thing as long as the overall performance is still very good in the given unit/institution. A unit/institution-wide comparison of performance might therefore bring interesting insights into the dynamics and performance relations among its staff. This relative comparison of staff members with all the others can also help discover discrepancies in remuneration, benefits, assistance and support the members of the staff are being provided with.

In the light of the existence of several evaluated areas with possibly different units of performance measurement, the aggregation to a single-number overall evaluation might result in a loss of information, or might simply reflect the specifically chosen aggregation function (or linguistic fuzzy rule base). However, a comparison of the performance of different members of the academic staff can be performed also without the aggregation. We can simply build on the ideas of the data envelopment analysis [5] and consider the performance areas to represent the outputs of the academic staff while the remuneration or the support provided to them would be the input.

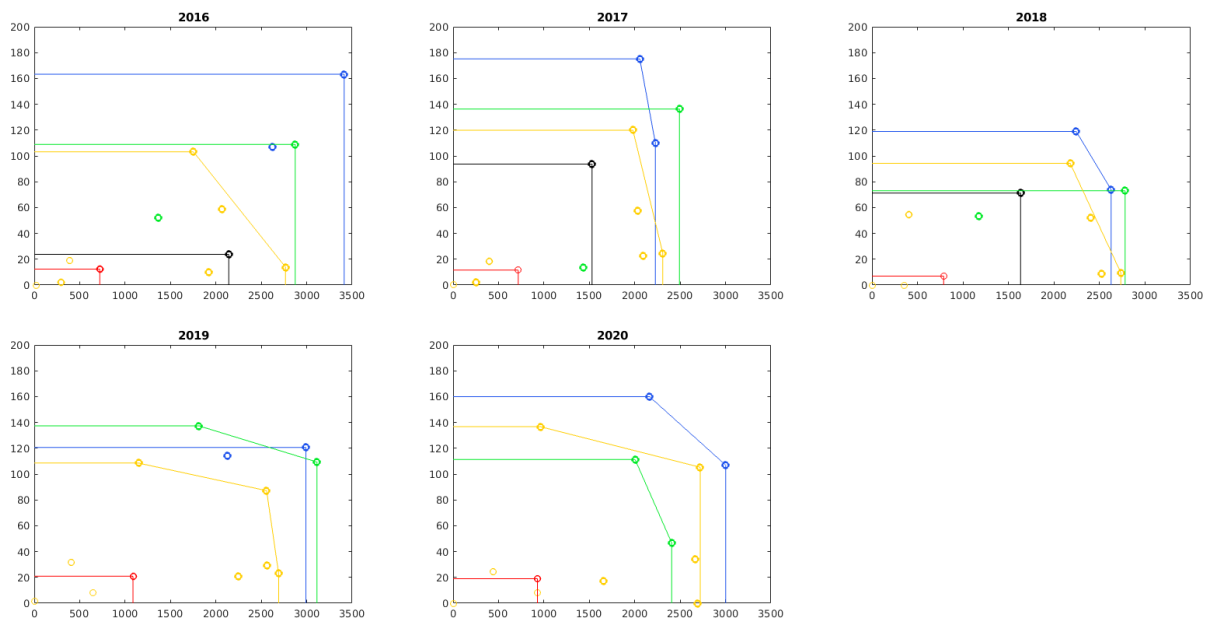


Figure 3 Efficient performance frontiers for different academic positions: assistants (red), assistant professors (yellow), associate professors (green), full professors (blue), researchers (black). The x-axes represent PA performance, y-axes represent RDC performance. All the values are in raw scores.

Using the raw scores, Figure 3 represents the comparison of performance of the academic staff members regardless of the requirements set on their positions. It provides a “fair comparison” of the amount of work being performed by the staff members. It also defines the efficient frontiers for each academic position in every year. This way if a point representing a specific staff member lies inside the area of its color, he/she is not performing in an efficient way and may require some kind of assistance to be able to reach the PA/RDC performance of the efficient representatives of that position. Note that this holds particularly if all the contracts of the staff members are of the same size and if they all receive the same salary. Still the plots tell us who does how much and might help us identify discrepancies in remuneration - for example if a staff member appears to be more efficient (performing not-worse in any area) than another staff member, the outperforming staff member should probably be receiving a larger salary. The connection with remuneration is and cannot be direct here, but e.g. an associate professor appearing more efficient than full professors in these plots might indicate a talented young academic who needs appropriate conditions and motivation to stay and provide optimal performance for the unit.

One can also use the standardized evaluations (the ones that reflect different requirements on the academic positions and thus also different salaries) to get rid of the influence of the academic position and to compare all the staff members in terms of how much they fulfill or exceed the requirements set for their position. In this sense the most efficient units, that is those defining the efficient frontier, would be those that given their academic position perform the best. This point of view could prove particularly useful when the resources to be provided to assist the academic staff members in their development are limited. In this case the ones that need support the most can be easily identified as being the least efficient. This way we are, however, losing easy comparability with the remuneration.

4 Conclusions

In this paper we have proposed several visualization techniques for the data on the performance of academic staff members. We have introduced the 2-dimensional time series perspective on single staff member performance visualization and proposed a performance trajectory and career path view of the data. This paper identifies appropriate and novel ways of the visualization of academic staff performance data. Having done so the next steps of our research will necessarily involve the design of analytical procedures for these 2-dimensional time series representations of performance and their development in time. We have also proposed a DEA-inspired visualization of the performance of the whole units for the purpose of identification of resource/support/remuneration discrepancies. All the data visualization methods proposed in this paper aim on maximal utilization of the available HR data and the minimization of data distortion or information loss.

Acknowledgements

This research would like to acknowledge the partial support by the grant *Mathematical education and its reflection in publication outputs according to Methodology 17+* of Faculty of Education of Palacký University Olomouc, and by LUT research platform ABMI- Analytics-based management for business and manufacturing industry.

References

- [1] Act no. 111/1998 sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), 1998.
- [2] Collan, M., Stoklasa, J., and Talašová, J.: Examples of Academic Faculty Evaluation Systems from the Czech Republic and Finland. In: *Peer Reviewed Full Papers of the 8th International Conference on Evaluation for Practice: "Evaluation as a Tool for Research, Learning and Making Things Better"*, June. University of Tampere, 111–120.
- [3] Collan, M., Stoklasa, J., and Talašová, J.: On Academic Faculty Evaluation Systems: More than just Simple Benchmarking. *International Journal of Process Management and Benchmarking* **4** (2014), 437–455.
- [4] Collan, M., Stoklasa, J., and Talasova, J.: Staff Evaluation Systems-Shaping Autonomy through Stakeholders. In: *(Re)Discovering University Autonomy: The Global Market Paradox of Stakeholder and Educational Values in Higher Education* (Turcan, R. V., Reilly, J. E., and Bugaian, L., eds.). New York, 2016, 97–106.
- [5] Cooper, W. W., Seiford, L. M., and Zhu, J., eds.: *Handbook on data envelopment analysis*. 2nd edition. Springer Science+Business Media, New York, Dordrecht, Heidelberg, London, 2011.
- [6] Holeček, P., Stoklasa, J., and Talašová, J.: Human resources management at universities - a fuzzy classification approach. *International Journal of Mathematics in Operational Research* **9** (2016), 502–519.
- [7] Office of the Government of the Czech Republic: *Methodology for Evaluating Research Organisations and Research, Development and Innovation Purpose-tied Aid Programmes*. Praha, 2018.
- [8] Stoklasa, J.: *Linguistic models for decision support*. Lappeenranta University of Technology, Lappeenranta, 2014.
- [9] Stoklasa, J., Holeček, P., and Talašová, J.: A holistic approach to academic staff performance evaluation - a way to the fuzzy logic based evaluation. In: *Peer Reviewed Full Papers of the 8th International Conference on Evaluation for Practice: "Evaluation as a Tool for Research, Learning and Making Things Better"*, June. University of Tampere, 121–131.
- [10] Stoklasa, J., Talášek, T., and Musilová, J.: Fuzzy approach - a new chapter in the methodology of psychology? *Human Affairs* **24** (2014), 189–203.
- [11] 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.
- [12] Stoklasa, J., Talášek, T., and Talašová, J.: AHP and weak consistency in the evaluation of works of art – a case study of a large problem. *International Journal of Business Innovation and Research* **11** (2016), 60–75.
- [13] Stoklasa, J., Talašová, J., and Holeček, P.: Academic Staff Performance Evaluation - Variants of Models. *Acta Polytechnica Hungarica* **8** (2011), 91–111.
- [14] Talašová, J., Stoklasa, J., Holeček, P., and Talášek, T.: Mathematical support for human resource management at universities. In: *Proceedings of the 35th International Conference on Mathematical Methods in Economics* (Pražák, P., ed.). University of Hradec Králové, Hradec Králové, 783–788.
- [15] Talašová, J., Stoklasa, J., Holeček, P., and Talášek, T.: Informační systém pro hodnocení akademických pracovníků – IS HAP. *AULA* **26** (2018), 32–43.

Combined Time Coordination of Connections in Public Transport

Dušan Teichmann¹, Michal Dorda², Denisa Mocková³, Pavel Edvard Vančura⁴, Vojtěch Graf⁵, Ivana Olivková⁶

Abstract. Time coordination of connections in transfer nodes is one of the basic pillars of attractive public transport. Coordination of the connections must be implemented in such a way as to ensure not only the continuity between the required connections, but also that the operating conditions of the public transport provider are not adversely affected. Coordination is also not usually limited to an isolated transport node, because there are more transfer nodes in the network of public transport lines. The presented article contains a mathematical model mixing node and section time coordination of connections in public transport system at the same time. Factors that affect the coordination process are discussed and some problems that occur when solving tasks of time coordination using mathematical models are pointed out.

Keywords: time coordination, network, mathematical programming, optimization, public transport

JEL Classification: C44

AMS Classification: 90C11

1 Introduction – Motivation for Solution

Coordination of connections is one of the fundamental qualitative parameters of public transport. The basic purpose of the coordination of public transport connections is to shorten the wait time between connections or to evenly distribute the offer of connections over a block of time. Coordination of connections can take place between connections of one mode of transport, but also between connections of different modes of transport. In addition, we can distinguish between node, sectional or combined coordination.

The coordination problem addressed is based on a real situation in which a transfer from two different directions to two different directions is provided. The coordination node is not an end node for connections of lines serving the coordination node. This means that the connections of all lines entering the coordination node also exit from this node continuing in the same direction. The coordination situation is complicated by the fact that, during coordination, one incoming direction and one outgoing direction are served by a pair of lines. The remaining entry direction and exit direction are served by only one line. The transfer coupling from the incoming direction served by the pair of lines must be provided only from the connection of one line. The transfer connection from the incoming direction served by a single line must be provided only to the connection of one of the pair of lines. However, the situation is also complicated by the fact that in the sections adjacent to the coordination node (incoming and outgoing) served by the pair of lines, it is necessary to perform so-called section coordination. Section coordination is coordination in which a certain minimum time interval must be ensured between the connections of lines serving the same section. The reason for ensuring a minimum time interval is to ensure a predefined time uniformity of section service.

Therefore, in the task under consideration, the coordination model must fulfill two main requirements:

1. The preference of transfer connections with maximum intensity of transferring passengers.
2. Provision of an acceptable time interval between the connections of lines serving the same section.

¹ CTU in Prague, Faculty of Transportation Sciences, Department of Logistics and Management of Transport, Prague 1, Konviktská 20, 110 00, teichdus@fd.cvut.cz

² VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, Ostrava – Poruba, 17. listopadu, 15/2172, 708 33, michal.dorda@vsb.cz

³ CTU in Prague, Faculty of Transportation Sciences, Department of Logistics and Management of Transport, Prague 1, Konviktská 20, 110 00, mockova@fd.cvut.cz

⁴ CTU in Prague, Faculty of Transportation Sciences, Department of Logistics and Management of Transport, Prague 1, Konviktská 20, 110 00, vancupav@fd.cvut.cz

⁵ VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, Ostrava – Poruba, 17. listopadu, 15/2172, 708 33, vojtech.graf@vsb.cz

⁶ VSB – Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, Ostrava – Poruba, 17. listopadu, 15/2172, 708 33, ivana.olivkova@vsb.cz

2 Analysis of the State of the Art

Numerous authors have dealt with the issue of time coordination of connections in the past. Paper [1] deals with the coordination of bus timetables. The proposed approach considers the current demand for transport and at the same time allows searching for alternative routes to destinations in the event that the transfer is not made. The paper contains an integer nonlinear model, the optimization criterion of which is the total cost of system operation, including operating costs and user costs. The solution of the problem is based on the heuristic method based on successive averages. The functionality of the proposed solution was verified on a model task.

In paper [2], the authors propose an integer linear model for coordination to increase the smoothness of passenger transfers between different directions. The optimization criterion is the passengers' total waiting time, and the goal of optimization is to minimize this time. The actual solution of the problem is performed by genetic algorithms. The proposed model is being tested on a real transport network in Beijing.

In paper [3], its authors deal with the problem of finding links between the offered capacity of buses and passenger waiting time. They use two methods to solve the problem – an approach based on timed Petri nets and max plus algebra. The proposed approach is tested on a model task.

Paper [4] is devoted to the coordination of connections between urban railway lines and other train transport. The authors focus primarily on the coordination of the last connections (night connections), to ensure that passengers can always reach their destination without the risk of not being able to continue on the next part of their journey at the interchange node. Synchronization is achieved using a linear mathematical model with two criteria – the total number of passengers who can successfully reach the destination of their journey and the time of termination of operation of urban railway lines. The functionality of the proposed approach has been verified on a real task with the coordination of connections between Beijing's urban railway lines and the high-speed railway lines between Beijing and Shanghai and Beijing and Tianjin.

In paper [5], its authors deal with the coordination of city tram lines and surrounding traffic. The authors of the article point out that trams spend a relatively large amount of unproductive time waiting at intersections. They propose an integer linear model with the aim of increasing the smoothness of tram traffic by coordinating with the traffic light plans at intersections. The objective function of the proposed model contains two criteria – the total running time of the tram between the final stops and the time of passing through the intersections during green light periods. In an experiment carried out on a part of the real tram network in Shanghai, the average total tram travel time was reduced by 11% and the number of stops at intersections was reduced by 80%. The proposed solution did not have an impact on the surrounding traffic, because it fully adapts to the already established control plans of individual traffic lights.

The authors of paper [6] address the same problem. Unlike the previous paper, its authors approach the problem in a completely different way. On the main tram lines, they propose coordinated green waves at intersections depending on the passage of trams. The individual tram connections are coordinated so as to ensure that multiple trams are not passing through a single intersection and avoiding long periods of blocking of the intersection for the surrounding traffic in the same time period. The proposed method was verified on a part of the tram network in the city of Huaian.

In our paper, we will partially follow the approach for node time coordination proposed by prof. Jaroslav Janáček from the University of Žilina, which has not yet been published in its original form, and partly follow the approach for section coordination proposed by prof. Jan Černý from the University of Economics in Prague and published, e.g., in paper [7]. Prof. Janáček's nodal time coordination model was created for an isolated transfer node with one or more incoming directions and one exiting direction. That is, at the transfer node, transfers are made from connections coming from several directions to connections departing in one direction. None of the lines serving the transfer node is a transit line. The optimization criterion was the total time loss of transferring passengers. The mathematical approaches of prof. Černý were designed to optimize the time positions of connections of different lines running in a cycle and serving the same section in the same direction. The optimization criterion was a minimum time interval between the connections of two different, consecutive lines and the goal was to maximize its value. There was also an alternative criterion representing the maximum time spacing between the connections of two consecutive different lines, and the goal of optimization was to minimize its value. The problem of section coordination of connections is also known in the literature as the problem of Žilina n – angles. In the past, for example, authors also dealt with tasks on section coordination in bus transport [8].

3 Preparatory considerations related to the construction of a mathematical model

Let us consider a transfer node into which the connections of three lines come together from two directions, entering from one direction the junction of one line (isolated direction) and from the other direction entering at the junction of two lines (common direction). From the transfer node, the connections of the three lines also depart in two directions, while the connections of the line of the isolated direction depart in one direction and the connections of the common direction depart in the other direction (the routes of the common direction lines also split later). Let us include a line serving an isolated direction into set I and a line serving a common direction into set I' . The situation is demonstrated by the following diagram – see Fig. 1. In Fig. 1, the isolated direction served by one line is indicated by a solid line and the common direction served by a pair of lines by a dashed line.

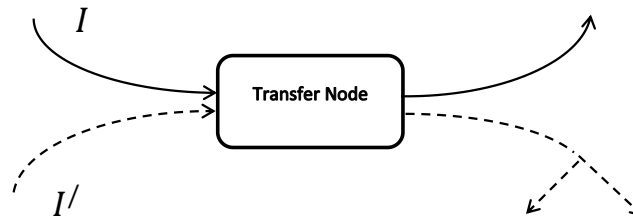


Figure 1 Diagram of the task under consideration

Periodic operation is implemented on all three lines, i.e. the connections run on them at regular intervals. This is characteristic, for example, for weekend operation. However, the implementation of periodic operation is important in one other respect from the point of view of modeling. With periodic operation, not all connections for the entire coordination period need to be included in the optimization model. Only two connections from each coordinated line need to be included in a coordination task. Coordination will always be performed from the first connection of the line entering the transfer node to one of the departing connections of the coordinated line. This radically reduces the mathematical model. With a variable intensity of transferring passengers during the coordination period, it is possible to express the intensity of transferring passengers from a given connection as a cumulative intensity value for all connections operated in the coordination period. We also assume the situation common in tram transport that the service time of the transfer node is given by one element of time data (the times of arrival of connections to the transfer node and the times of their departures are in the same minute).

4 Mathematical formulation of the problem and design of a model for the solution

Let us assume that coordinated lines enter the transfer node from two directions I and I' and exit the transfer node in the same two directions. For each line $i \in I \cup I'$, the value of the earliest possible service of the transfer node, period (regular time interval between two connections of the same line serving the route in the same direction) T_i , the value of the maximum allowed time shift of connections a_i and the cumulative intensity of passengers f_i changing in the selected transfer node are defined. For the pair of lines I' serving the same section, the minimum value of the time interval of the connections of these lines in the same direction r is defined. The task is to plan the time positions of the connections of individual lines so as to minimize the total time loss for transferring passengers while maintaining the minimum time intervals between the connections of the lines serving the common section. The symbol M represents a large enough constant.

For the purpose of modeling the decision, we will define four groups of variables:

- x_i – the non-negative variable modeling the time shift of line connections $i \in I \cup I'$.
- h_i – the non-negative variable modeling the accumulated time loss for the passengers transferring from the connections $i \in I \cup I'$ in the transfer node, between which the transfer connection is created.
- z_{ijk} – the auxiliary binary variable modeling the creation of the transfer connection between the first connection of line $i \in I$ and the connection $k \in K$ of line $j \in I'$ or between the first connection of line $i \in I'$ and the connection $k \in K$ of line $j \in I$ – when $z_{ijk} = 1$, then a transfer link arises, while in the opposite case, no transfer link arises.
- u – the auxiliary binary variable modeling the sequence of connections of the lines servicing the common section, when $u = 1$, then in the solved period, the first connection of one of the lines will precede

the first connection of the second of the lines, when $u = 0$, the sequence of the first connections will change on both lines.

The mathematical model for the problem solved will have the form:

$$\min f(x, h, z, u) = \sum_{i \in I \cup I'} f_i h_i \quad (1)$$

subject to:

$$\left[t_{j1} + (k-1)T_j + x_j \right] - (t_{i1} + x_i) \geq M(z_{ijk} - 1) \text{ for } i \in I, j \in I', k \in K \quad (2)$$

$$\left[t_{j1} + (k-1)T_j + x_j \right] - (t_{i1} + x_i) \geq M(z_{ijk} - 1) \text{ for } i \in I', j \in I, k \in K \quad (3)$$

$$\left[t_{j1} + (k-1)T_j + x_j \right] - (t_{i1} + x_i) \leq h_i + M(z_{ijk} - 1) \text{ for } i \in I, j \in I', k \in K \quad (4)$$

$$\left[t_{j1} + (k-1)T_j + x_j \right] - (t_{i1} + x_i) \leq h_i + M(z_{ijk} - 1) \text{ for } i \in I', j \in I, k \in K \quad (5)$$

$$\sum_{j \in I'} \sum_{k \in K} z_{ijk} = 1 \text{ for } i \in I \quad (6)$$

$$\sum_{j \in I} \sum_{k \in K} z_{ijk} = 1 \text{ for } i \in I' \quad (7)$$

$$x_i \leq a_i \text{ for } i \in I \cup I' \quad (8)$$

$$(t_{jk} + x_j) + r - M(1-u) \leq (t_{ik} + x_i) \text{ for } i \in I', j \in I', i \neq j, k = 1 \quad (9)$$

$$(t_{ik} + x_i) + r \leq (t_{jk+1} + x_j) \text{ for } i \in I', j \in I', i \neq j, k = 1 \quad (10)$$

$$(t_{ik} + x_i) + r - M u \leq (t_{jk} + x_j) \text{ for } i \in I', j \in I', i \neq j, k = 1 \quad (11)$$

$$(t_{jk} + x_j) + r \leq (t_{ik+1} + x_i) \text{ for } i \in I', j \in I', i \neq j, k = 1 \quad (12)$$

$$x_i \in R_0^+ \text{ for } i \in I \cup I' \quad (13)$$

$$h_i \in R_0^+ \text{ for } i \in I \cup I' \quad (14)$$

$$z_{ijk} \in \{0;1\} \text{ for } i \in I, j \in I', k \in K \quad (15)$$

$$z_{ijk} \in \{0;1\} \text{ for } i \in I', j \in I, k \in K \quad (16)$$

$$u \in \{0;1\} \quad (17)$$

Function (1) represents the optimization criterion. The groups of constraints (2) and (3) ensure that in the case of temporal non-viability of connection positions, no transfer links are formed. Constraints (4) and (5) quantify the time losses generated by the transfer links. Constraints (6) and (7) ensure the formation of transfer links. Constraints (8) ensures that their maximum values will not be exceeded during time shifts. Constraints (9) – (12) ensures compliance with the prescribed minimum time interval between connections of lines serving a common section (section coordination). Constraints (13) – (17) define the domains of variables used in the model.

5 Computational experiment with the proposed model

We will test the functionality of the proposed model on a real coordination task. On the tram network of the city of Ostrava in the Czech Republic, we will select one of the transfer nodes – specifically, the “Svinov, mosty” node – see Fig. 2. From Fig. 2 it is clear that, at this node, it is possible for the connections of two lines to stop at the same time at the same platform to allow the transfer of passengers. The transfer node therefore has sufficient capacity conditions to allow transfers. The coordination will take place for operation on weekend days, when the connections of individual tram lines are run in a 20-minute interval. The earliest possible service times of the “Svinov, mosty” node were set to time 0. The maximum time shifts of the connections of individual lines at an interval time of 20 minutes take on values of 19 min. A minimum time interval of 6 minutes is required between the connections of the various lines serving the common section.

Line no.	Line Route
4	Martinov – Svinov, mosty – Mariánské Hory – Centrum
7	Poruba – Svinov, mosty – Výškovice
17	Poruba – Svinov, mosty – Dubina

Table 1 Routes of the coordinated lines

The connections of three lines must be coordinated in the transfer node. Their routes are listed in Table 1.

In the transfer node, it is necessary to create transfer connections either between the connections of lines 4 and 7 or between the connections of lines 4 and 17. A fragment of the tram line network with the designation of the coordination node is shown in Fig. 3. The values of the binary variable u were chosen as follows: If, after the optimization calculation, it applies that $u = 1$, then in the coordination period, the first connection of line 17 will precede in time the first connection of line 7 in the common section. If, after the optimization calculation is completed, it applies that $u = 0$, the first connection of line 7 will then precede the first connection of line 17 within the common section in the coordination period.



Figure 2 Transfer node Svinov, mosty [8]

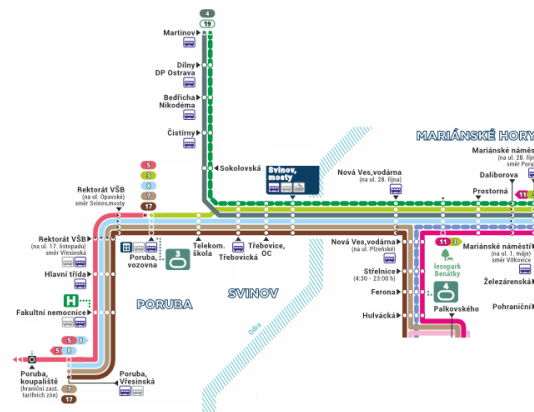


Figure 3 Diagram of tram network with indication of transfer node [9]

This coordination will ensure smooth transport from stops in the Poruba – Svinov, mosty section, to the Centrum stop and from stops in the Martinov – Svinov, mosty section, in the direction of the Výškovice and Dubina housing estates located in the southern part of Ostrava. We consider lines 7 and 17 on the exit direction from the Svinov, mosty node, as interchangeable, because after leaving the node, they service 6 common stops. At the same time, it is necessary to ensure that a certain minimum distance is ensured between the connections of lines 7 and 17 serving the same section Poruba – Svinov, mosty, so that the service of this section is time equalized (at the same moment in time then, only the connections of lines 4 and 7 or the connections of lines 4 and 17 can come together). The diagram of coordinated directions and lines is shown in Fig. 4.

Computational experiments with the proposed model were performed in Xpress-IVE optimization software installed on a personal computer with the Windows 10 operating system, an Intel i5-5200U processor with a frequency of 2.2 GHz and an operating memory of 8 GB. After the completion of optimization calculation, the optimal solution given in Table 2 was obtained. The value of the variable u after the end of the optimization calculation was $u = 1$, i.e., it applies that in the coordination period on the common section the first connection of the line precedes the first connection of line 7. Furthermore, after the end of the optimization calculation it was true that $z_{131} = 1$ a $z_{311} = 1$. Thus, transfer connections were created in the transfer node between the first connections of lines no. 4 and 17. This also follows from the results given in Table 2, where both lines have the same service time of the transfer node.

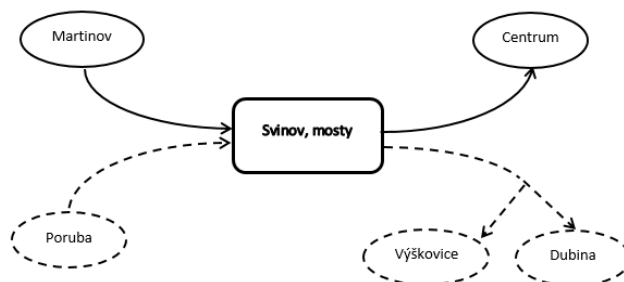


Figure 4 Diagram under the conditions of the computational experiment

Line number i	Time shift of line connection i (x_i)	Time of service of transfer node 1. by the connection of line i after its shift to $t_{i1} + x_i$
4	0	0
7	6	6
17	0	0

Table 2 Results of the optimization experiment

6 Conclusions

In the paper, we dealt with the optimization of transfer connections in the conditions of the transfer node. The subject of the solution was a specific situation occurring in the tram network in Ostrava. The optimization model designed to coordinate the connections of different lines contains elements of node and section coordination of connections. Node coordination focuses on the creation of transfer connections between the connections of tram lines entering the transfer node from two directions and also exiting the transfer node in two directions. The model presented in the paper will be the basis of the so-called network coordination model, which will be prepared in the future and presented at one of the future years of this conference.

Acknowledgements

The creation of this paper was supported by the project CK01000043 System for Support of Network Time Coordination of Connections at Transfer Nodes.

References

- [1] Wu, W., Liu, R., Jin, W. and Ma, C. (2019). Stochastic bus schedule coordination considering demand assignment and rerouting of passengers. *Transportation Research Part B: Methodological*, 121, pp. 275-303.
- [2] Cao, N., Tang, T. and Gao, C. (2020). Multiperiod Transfer Synchronization for Cross-Platform Transfer in an Urban Rail Transit System. *Symmetry*, 12(10), p. 1665.
- [3] Idel Mahjoub, Y., Chakir El-Alaoui, E. and Nait-Sidi-Moh, A. (2020). Modeling and developing a conflict-aware scheduling in urban transportation networks. *Future Generation Computer Systems*, 107, pp. 1026-1036.
- [4] Long, S., Meng, L., Miao, J., Hong, X. and Corman, F. (2020). Synchronizing Last Trains of Urban Rail Transit System to Better Serve Passengers from Late Night Trains of High-Speed Railway Lines. *Networks and Spatial Economics*, 20(2), pp. 599-633.
- [5] Ji, Y., Tang, Y., Du, Y. and Zhang, X. (2019). Coordinated optimization of tram trajectories with arterial signal timing resynchronization. *Transportation Research Part C: Emerging Technologies*, 99, pp. 53-66.
- [6] Xu, Z., Lv, S., Li, D. and Quan, M. (2020). Coordinated Control Method for Trams on Urban Arterial. *IOP Conference Series: Earth and Environmental Science*, 587, p. 012092.
- [7] Černý, J., Kluvánek, P. (1991). *Základy matematickej teórie dopravy*. Bratislava: VEGA, 1991. ISBN: 80-224-0099-8.
- [8] Kozel, P. and Michalcová, Š. (2017). Time coordination of bus arrivals on a set of bus stops. In *Proceedings of the 12th International Conference on Strategic Management and its Support by Information Systems* (pp. 267-273). Ostrava: VSB-Technical University of Ostrava.
- [9] <https://www.nejkacka.eu/projekt/prednadrizi-svinov/>.
- [10] <https://www.dpo.cz/soubory/jr/schema-tram-dopravy-2021-01-02.pdf>.

Measuring Efficiency of Football Clubs: DEA approach

Michal Tomíček¹, Natalie Pelloneová²

Abstract. Professional football clubs have a special characteristic and that is a combination of sports and financial performances. Currently, performance analysis of football clubs is a useful tool to help managers with evaluation and feedback with respect to the real context of the market. The aim of the presented paper is to use data envelopment analysis (DEA) to calculate efficiency and then evaluate and compare the efficiency of selected European football competitions. For comparison, two European competitions with the potential for similarity were selected from the football environment. They are the Czech Fortuna:Liga and the Danish 3F Superliga. Both competitions are the highest professional football competitions in their countries and were chosen on the basis of the same game model and similar placement in the UEFA league coefficients. The researched period are the seasons 2015/2016 to 2019/2020. The data used were obtained from the official databases of both examined sports competitions and subsequently supplemented with private databases of companies from the football environment. In terms of results, some implications for the management of football clubs are discussed and suggestions for increasing efficiency in inefficient clubs are made.

Keywords: sport management, football, DEA, efficiency, performance factors

JEL Classification: C10, L83, C67, C44

AMS Classification: 90B90, 90C90

1 Introduction

At present, football is one of the most important sports, and at the same time a business of high economic importance. Every sport organization strives to evaluate its performance: its weaknesses and strengths. Nowadays, success in the professional football league goes hand in hand with successful coaching and leading the entire team. In the real world, team leadership always requires not only evaluating one's own performance, but also monitoring the activities of competitors. So it is often a crucial condition to learn from your own mistakes and from your competitors' innovations. However, evaluating performance is not an easy task at all.

Football is a global phenomenon that prevails mainly in Europe. The football industry has changed significantly over the last two decades. Economic survival has become increasingly important in recent years as the more restrictive future forces football clubs to rethink the amounts they pay for players and their wages. Based on this outlook, the technical efficiency analysis appears to be an important tool for evaluating the activities of clubs and their sports performance. The aim of this paper is to contribute to previous research and to evaluate and compare the efficiency of selected European football competitions using data envelopment analysis. The presented research evaluates the performance of professional football clubs that participated in five seasons (2015/16 to 2019/20) of the highest football competitions in the Czech Republic and Denmark.

2 Literature review

Sports statistics and performance appraisal are currently more and more frequently used words. There are several peer-reviewed journals looking for innovative methodologies for data analysis in sports, and several major conferences are dedicated each year to presenting statistics in sports research. As in many other fields, there is currently an influx of large amounts of available sports data, creating an opportunity for advanced statistical analysis. The following part of the paper briefly summarizes selected authors who apply various statistical and mathematical methods in their research to evaluate the performance of various sports disciplines.

Tunaru et al. [18] use contingent claims methodology and standard techniques in stochastic calculus to develop a framework for determining the financial value of professional football players to prospective-acquiring clubs. In

¹ Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, michal.tomicek@tul.cz.

² Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, natalie.pelloneova@tul.cz.

their research, they use the Opta Index to model the uncertainty of players' performance, including possible injuries. They also consider uncertainty about the club's income, player income, club image, fan loyalty and so on. Kirschstein and Liebscher [11] use robust analysis techniques to examine the extent to which a player's market value depends on his football skills. In their research, they analyze a data set comprising 28 performance metrics and the market value of 493 players from the 1st and 2nd Bundesliga. Kim et al. [10] use several techniques (for example, regression analysis or cluster analysis) to propose an approach for classifying and evaluating footballers based on their performance and transfer fee. Lee and Harris [13] examine factors that could affect Major League Soccer players' salaries. By applying four different methods of regression analysis, they revealed these three key determinants affecting players' salaries: the number of goals, the number of assists and the minutes in the game. Simmons [16] examines the relationship between a player's salary and his performance in two major sports: American football and European football.

The following is a brief summary of selected authors who apply the DEA method in their research. This method has been applied successfully to measure efficiency in different cases as well as in sport. In his research, [12] applies the non-parametric DEA method to the evaluation of hockey teams in the NHL. Using this method, each team receives an efficiency score and is then compared to other NHL teams. The salaries of goalkeepers, defenders and attackers were chosen as inputs to the model. League points in the base season model and winnings in the play-off model were chosen as outputs. Halkos and Tzeremes [7] evaluate the career performance of 229 professional tennis players using the DEA methodology. Gutierrez and Ruiz [5] evaluate the game performance of 24 teams participating in the 2011 Men's Handball World Championships in Sweden using the DEA and the cross-efficiency evaluation. Espitia-Escue and García-Cebrián [3] measure the efficiency of professional football teams playing in the Spanish First Division. The timeline of the study covers three seasons from 1998 to 2001. To this end, they apply the DEA methodology, taking into account players used, attacking moves, the minutes of possession of the ball, and the shots and headers as input variables. They consider the achieved number of points as an output. Palafox-Alcantar and Vargas-Hernández [14] measure the wage efficiency of 32 football teams in the National Football League in the 2014 season using data analysis. Halkos and Tzeremes [8] use the DEA method to compare the current level of value of football clubs and their performance. The research concluded that the current level of value of football clubs has a negative effect on their performance, which suggests that the high value of football clubs does not guarantee their higher performance. Zambom-Ferraresi et al. [19] have employed DEA to estimate the technical efficiency of English football teams playing in the Premier League. The researched period are the seasons 2012/13 and 2014/15. They consider sports results measured by the total number of points achieved in the league, revenues, attendance, and fans impact on social media as the outputs. The input variables include the amount of the team's salary, the market value and the plays performed during the match. In their research, [6] apply an input-oriented efficiency model to Spanish football clubs. They use staff costs and other expenses such as inputs and turnover and points won as outputs. Their model of analysis mixes economic and sport performance, and so a non-economic additional output, such as points won in competitions, was also considered to calculate efficiency scores.

From the above, it can be stated that for football clubs, there is no agreement on the input and output variables that should be used in the DEA model, each study uses different combinations of inputs and outputs. Some works consider only sports variables (shots on goal, minutes of control over the ball, points obtained), while others also include economic factors (rotation, personnel expenses) in the research. Some papers combine both types of variables. The literature also discusses what type of inputs better explains the performance of sports clubs. On the one hand, several contributions analyzed the performance of sports clubs using so-called ex-post inputs (e.g. wages). On the other hand, some authors criticized the use of these ex-post inputs and recommend the use of ex-ante inputs such as the market value of players. In the presented paper, an approach was chosen that uses both economic and non-economic variables. In terms of input variables, an ex-ante approach was used.

2.1 Data Envelopment Analysis

To calculate efficiency, two methodological approaches are possible [15]: parametric models and non-parametric models, which use the conditions that must be met by the set of production possibilities. The advantage of the non-parametric approach is its flexibility to adapt to multi-product and lack-of-price environments, although it has the disadvantage of being deterministic in character, so that any deviation from the efficiency frontier is considered to be inefficient behavior of the evaluated unit. In this paper, a non-parametric method of data envelopment analysis will be used. It is very often utilized to evaluate the technical efficiency of production units. The DEA technique is a linear and parameter-free programming method. DEA is especially suitable for determining the technical efficiency of units that are comparable to each other. The method analyzes and calculates the relative efficiency of the DMUs from multiple inputs and output factors. The units are compared with each other and

it is determined which of them are efficient and which are inefficient. In the case of inefficient units, depending on the type of model, it can be determined how the inefficient unit should reduce/increase its inputs/outputs to become efficient. There are many DEA models that can be used in different situations.

One of the frequently used models is the CCR model which works under the conditions of constant returns to scale and which enables the calculation of overall technical efficiency score (ET_{CCR}). The second alternative is the BCC model, which is based on variable returns to scale and calculates the so-called pure technical efficiency score (TE_{BCC}). The efficiency frontier defines the maximum combinations of outputs that can be produced for a given set of inputs. Assuming a set of n scenarios as DMUs, each DMU j ($j = 1, \dots, n$) uses m inputs x_{ij} ($i = 1, 2, \dots, m$) to produce s outputs y_{rj} ($r = 1, 2, \dots, s$). The input-oriented envelopment models with constant returns of scale, can be formulated in equation (1) to minimize the inputs while keep the outputs at their current levels.

$$\begin{aligned} & \min \theta - \varepsilon \left(\sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \right) \\ \text{s. t. } & \sum_{j=1}^n \lambda_j x_{ij} + s_i^- = \theta x_{ip} \quad i = 1, 2, \dots, m \\ & \sum_{j=1}^n \lambda_j y_{rj} - s_r^+ = y_{rp} \quad r = 1, 2, \dots, s \\ & \lambda_j, s_i^-, s_r^+ \geq 0, \quad j = 1, 2, \dots, n. \theta \text{ unrestricted in sign.} \end{aligned} \quad (1)$$

The technical efficiency of the unit is measured in relation to the other analyzed units using the efficiency score. The overall technical efficiency rate (ET_{CCR}) takes values in the range from 0 to 1. Technically efficient units achieve an efficiency rate of 1, for inefficient units the efficiency rate is less than 1.

3 Data and methodology

The aim of the research is to use DEA to calculate the degree of technical efficiency and then evaluate and compare the efficiency of selected European football competitions. Two European competitions with the potential for similarity were selected for comparison from the football environment. They are the Czech Fortuna:Liga and the Danish 3F Superliga. The evaluation of efficiency is based on the indicators such as number of players, total squad market value and total points in the season. Due to the availability of data in both countries, the efficiency was analyzed in the seasons 2015/16 to 2019/20. The research process can be divided into the following stages.

1. Selection of sports competitions – for the needs of the research, two sports football clubs were selected from the total number of 55 UEFA member associations – the Czech Fortuna:Liga and the Danish 3F Superliga. The similarity of these competitions is based on the same game model and similar placement in the UEFA league coefficients. A close ranking in the UEFA league coefficient table indicates relatively similar qualities of the best units of each competition. These and other similarities are a prerequisite for making comparisons. After the 2019/20 season, Denmark was two places ahead of the Czech Republic in the UEFA league coefficient. The Czech top football competition called Fortuna:Liga is attended by a total of 16 teams from Bohemia, Moravia and Silesia. A total of 12 teams take part in Denmark's highest football competition called the 3F Superliga.

2. Collection of the data – the data used for the purposes of the research come from the official databases of both examined sports competitions and are supplemented by private databases of companies from the football environment. The core of the research are data from InStat [9], which analyzes the performance of athletes and sports teams. They are supplemented by the database of the Transfermarkt.com server [17] and the databases of the Czech Fortuna:Liga [4] and the Danish 3F Superliga [1]. The research period was limited by InStat's data for Denmark's highest competition when the oldest possible sports data are for the 2015/16 season.

3. Definition of inputs and outputs – another important step is to choose the appropriate number of inputs and outputs. Bowlin [2] set the rule that the number of DMUs should be at least three times as high as the number of the inputs and outputs reasoned. A large amount of data was available to evaluate the performance, the selection of the most suitable variables was then performed using correlation analysis. Number of players and total squad market value were used as inputs; total points in the season were outputs.

4. Construction of DEA model and determination of the ET_{CCR} score – an input-oriented CCR model was used to measure the performance of football clubs. The CCR model measures overall technical efficiency (ET_{CCR}), by aggregating pure technical efficiency and scale efficiency into a single value. All calculations were made using the OSDEA-GUI software.

5. Comparison of the performance of football clubs in both countries – a Kolmogorov-Smirnov test was used to compare the performance of selected sports competitions. A non-parametric test was chosen because using the Shapiro-Wilk test, it was proved that the values of the individual variables do not have a normal distribution. Statistical testing was performed at a significance level of 5% using STATGRAPHICS Centurion XVIII.

4 Research results

Applying an input-oriented CCR efficiency model, the performance levels achieved by Czech football teams for seasons 2015/16 to 2019/20 are reported in Table 1. The analysis thus shows how clubs convert the inputs into points gained within competition tables. The values refer to the current members of the season.

The CCR-I model described 2 to 3 football clubs as efficient each year. This means that the size of these clubs is optimal, they have chosen the appropriate sports tactics and at the same time the clubs are able to efficiently transform the inputs into outputs. The most efficient club in the researched period of five seasons in the Czech highest competition is Viktoria Plzeň. A total of three times out of five seasons, Viktoria Plzeň has become efficient from the point of view of ET_{CCR} . Prague's Slavia and Ostrava's Baník achieved an efficient score twice. However, in the 2015/16 season, when Baník descended from the highest competition, it also achieved the lowest score from the totals of the given season. The fact that a total of 39 different players in the Baník Ostrava jersey intervened in the matches in the 2015/16 season has a significant influence on this fact. This is the highest number of the whole observations. The FC Viktoria Plzeň club achieved an average ET_{CCR} score of 0.9560 in the monitored period and became the most efficient club in terms of long-term observation. The SK Slavia Praha club, which achieved an average ET_{CCR} score of 0.9130, took the second place. Far beyond expectations in terms of efficiency remains the historically most successful football club of the Czech highest competition, the AC Sparta Praha team. In the monitored period, the club playing its home matches at Prague's Letná achieved an average ET_{CCR} score of 0.6294. This result corresponds to the fact that, according to the Transfermarkt.com server, the value of the Sparta Praha team, together with Slavia Praha, is one of the utmost in the entire Czech highest competition every year. Unlike its biggest rival from the other side of the Vltava River, however, Sparta failed to transfer the high value of the team and the number of players into an adequate points gain. Specific levels of the ET_{CCR} value found according to the CCR-I method lower than 1 show that inefficient clubs should have had a lower level of inputs to achieve the given point gains, represented by a lower number of players and a lower team value. If their technical inefficiency were caused only by scale inefficiency, it could be stated that the size of these clubs is not optimal and inappropriate sports tactics was chosen.

Club	Season					Avg.
	2015/16	2016/17	2017/18	2018/19	2019/20	
1. FC Slovácko	0.8481	0.8940	0.8748	0.7188	0.8462	0.8364
1. FK Příbram	0.5160	0.7058	-	0.7421	0.4954	-
AC Sparta Praha	0.6867	0.6265	0.5521	0.6594	0.6223	0.6294
Bohemians Praha 1905	0.6354	0.7596	0.8537	0.8246	0.7605	0.7668
FC Baník Ostrava	0.2268	-	0.7077	1.0000	1.0000	-
FC Fastav Zlín	0.6912	0.9941	0.5762	0.8484	0.5360	0.7292
FC Hradec Králové	-	0.9572	-	-	-	-
FC Slovan Liberec	0.8515	0.5972	0.6251	0.7078	0.7151	0.6993
FC Viktoria Plzeň	1.0000	0.7807	1.0000	0.9994	1.0000	0.9560
FC Vysočina Jihlava	0.6527	0.8934	0.7008	-	-	-
FC Zbrojovka Brno	1.0000	0.8682	0.5479	-	-	-
FK Dukla Praha	0.6560	0.8333	0.7866	0.4763	-	-
FK Jablonec	0.7008	0.8503	0.8098	0.8286	0.8557	0.8090
FK Mladá Boleslav	0.8657	0.8357	0.4613	0.5371	0.6093	0.6618
FK Teplice	0.5125	1.0000	0.5952	0.6306	0.6652	0.6807
MFK Karviná	-	0.8909	0.7956	0.5604	0.4432	-
SFC Opava	-	-	-	1.0000	0.6483	-
SK Dynamo Č. Budějovice	-	-	-	-	1.0000	-
SK Sigma Olomouc	0.5327	-	1.0000	0.6849	0.7297	-
SK Slavia Praha	0.9663	1.0000	0.7024	1.0000	0.8961	0.9130
Avg.	0.6559	0.8429	0.7243	0.7637	0.7389	x

Table 1 ET_{CCR} score of Czech clubs playing Fortuna:Liga

Applying an input-oriented CCR efficiency model, the performance levels achieved by Danish football teams for seasons 2015/16 to 2019/20 are reported in Table 2. The analysis thus shows how clubs convert the inputs into a point gain within competition tables. The values refer to the current members of the season.

According to the CCR-I model, in the first to the fourth monitored seasons, two teams were always efficient. In the last fifth season, the number of efficient teams doubled. We can observe that the clubs that achieved the lowest ET_{CCR} score in the season either decreased or reached the efficient ET_{CCR} value in the following season. This also applies to Silkeborg IF, which after the 2019/20 season descended to the second highest Danish football competition. On average, the most efficient team of the Danish league in the monitored period was FC Copenhagen, which due to the less efficient examined third season reached the average value of ET_{CCR} "only" 0.8862. Other clubs in terms of ET_{CCR} value are Aalborg BL (0.8511), SønderjyskE Fodbold (0.8500) and FC Midtjylland (0.8075). SønderjyskE Fodbold reached the efficient ET_{CCR} value in the first monitored season, FC Midtjylland in the last season. Despite the fact that it was not efficient in any of the monitored seasons according to the ET_{CCR} value, the Aalborg BL club steadily maintained a score of around 0.8.

Club	Season					Avg.
	2015/16	2016/17	2017/18	2018/19	2019/20	
Aalborg BK	0.8312	0.8966	0.8728	0.8694	0.7853	0.8511
Aarhus GF	0.6158	0.5834	0.6716	0.5666	1.0000	0.6875
AC Horsens	-	0.8648	0.8773	0.9062	1.0000	0.9121
Brøndby IF	0.6436	0.9814	1.0000	0.8029	0.7432	0.8342
Esbjerg fB	0.4140	0.5324	-	0.9161	0.5022	-
FC Copenhagen	1.0000	1.0000	0.5057	1.0000	0.9254	0.8862
FC Helsingør	-	-	0.7112	-	-	-
FC Midtjylland	0.6677	0.6798	0.6991	0.9911	1.0000	0.8075
FC Nordsjælland	0.4382	0.6155	0.7787	0.5890	0.6142	0.6071
Hobro IK	0.3887	-	1.0000	0.6626	0.6796	-
Lyngby Boldklub	-	1.0000	0.6875	-	1.0000	-
Odense BK	0.5352	0.8518	0.7451	0.9720	0.6670	0.7542
Randers FC	0.7211	0.7618	0.4531	1.0000	0.8795	0.7631
Silkeborg IF	-	0.8815	0.7644	-	0.4640	-
SønderjyskE Fodbold	1.0000	0.9590	0.8027	0.7371	0.7510	0.8500
Vejle Boldklub	-	-	-	0.5993	-	-
Vendsyssel FF	-	-	-	0.6315	-	-
Viborg FF	0.7062	0.6846	-	-	-	-
Avg.	0.6635	0.8066	0.7549	0.8031	0.7865	x

Table 2 ET_{CCR} score of Danish clubs playing in the 3F Super League

In the last lines of Tables 1 and 2, the average ET_{CCR} scores of the clubs that participated in a given season are calculated. In both professional competitions, the highest average ET_{CCR} score was achieved in the 2016/17 season, which followed the season when, on the contrary, the lowest average ET_{CCR} score was achieved. The development of the values of the average ET_{CCR} score was very similar in both competitions. The performed non-parametric Kolmogorov-Smirnov test also confirmed that there are no significant differences between the two competitions at the level of significance of 5%.

5 Conclusion and discussion

In this article, we have measured the efficiency of the professional football clubs playing in the Czech Fortuna:Liga and Danish 3F Superliga. To that end, we have taken the time span of the five seasons from 2015/16 to 2019/20. Efficiency of football clubs was measured using non-parametric DEA method. Total squad market value and number of players were chosen as club inputs. Output was measured by total points in the season. This particular specification proved to be suitable for this application, but it can also be applied to efficiency analysis of team sports other than football.

Given the current economic and financial situation of football clubs, there is an increasing need to know how efficiently a club uses its resources. In addition, this analysis is also important for evaluating the club's sports performance. Based on the analyzed period, the first conclusion of the research can be reached: the winner of Fortuna:Liga was, with the exception of the 2019/20 season, always marked as efficient and the club with the lowest ET_{CCR} value mostly descended. Similar conclusions apply to Danish clubs: the winner of the 3F Superliga

has always been described as efficient and the club with the lowest ET_{CCR} value has either dropped or been considered efficient the following season. With the help of the research, it was possible to state the second conclusion: in the five seasons analyzed, no Czech or Danish club was able to maintain efficiency throughout the observed period. It is important to note that the clubs and resources used are constantly changing from season to season, just as the opposing teams are changing. So when an efficient club uses the same resources in the same way in previous seasons, it is not enough to be efficient in the coming seasons. Furthermore, it can be stated that there are no significant differences between Czech and Danish clubs. It is also possible to discuss the sources of clubs inefficiency. The first source of clubs inefficiency is related to the waste of resources. To achieve the same output, clubs should only need a lower value of inputs (i.e. a lower total squad market value or a lower number of players). The second source of inefficiency could be determined by calculating the scale efficiency. The problem is not only how they use their sports resources, but what sports tactics they use. These clubs should seek to develop a medium and long-term strategy to develop new and different tactics. The purchase of players should also take place in the context of the development of these new sports tactics. There are also clubs that suffer from both sources of inefficiency (the BCC model should be applied in further research to assess the sources of inefficiency in detail). In this case, clubs should reduce their resources and, in terms of size, find out how efficient clubs are developing sports tactics.

References

- [1] 3F Superliga. (2021). *Superliga.dk: Performance Center* [online]. Kodaň, 2021 [cit. 2021-02-15]. Available at: <https://www.superliga.dk/performance-center-superliga-2020-2021>.
- [2] Bowlin, W. F. (1998). Measuring Performance: An Introduction to Data Envelopment Analysis (DEA). *Journal of Cost Analysis*, 7, 3–27.
- [3] Espitia-Escue, M. & García-Cebrián, L. I. (2004). Measuring the Efficiency of Spanish First-Division Soccer Teams. *Journal of Sports Economics*, 5(4), 329–346.
- [4] Fortuna:Liga. (2021). *FortunaLiga.cz: Statistiky* [online]. Praha, 2021 [cit. 2021-02-15]. Available at: <https://www.fortunaliga.cz/statistiky?unit=1¶meter=1>.
- [5] Gutiérrez, O. & Ruiz, J. L. (2013). Data envelopment analysis and cross-efficiency evaluation in the management of sports teams: The assessment of game performance of players in the Spanish handball league. *Journal of Sport Management*, 27(2013), 217–229.
- [6] Guzmán-Raja, I. & Guzmán-Raja, M. (2021). Measuring the Efficiency of Football Clubs Using Data Envelopment Analysis: Empirical Evidence From Spanish Professional Football. *SAGE Open*, 11(1), 1–13.
- [7] Halkos, G. & Tzeremes, N. (2012). *Evaluating professional tennis players' career performance: A Data Envelopment Analysis approach*. MPRA Paper 41516. Munich: University Library of Munich.
- [8] Halkos, G. & Tzeremes, N. (2013). A two-stage double bootstrap DEA: the case of the top 25 European football clubs' efficiency levels. *Managerial and Decision Economics*, 34(2), 108–115.
- [9] InStat. (2020). *InStatSport.com: Football* [online]. Dublin, 2021 [cit. 2021-01-02]. Available at: <https://instatsport.com/football>.
- [10] Kim, Y., Nam Bui, K. & Jung, J. J. (2019). Data-driven exploratory approach on player valuation in football transfer market. *Concurrency and Computation Practice and Experience*, 33(3).
- [11] Kirschstein, T., & Liebscher, S. (2018). Assessing the market values of soccer players - a robust analysis of data from German 1. and 2. Bundesliga. *Journal of Applied Statistics*, 46(7), 1336–1349.
- [12] Kuosmanen, T. (1998). *Efficiency of Hockey Teams in NHL*. Fifth European Workshop on Efficiency and Productivity Analysis, 9 - 11 October 1997, Denmark: Copenhagen.
- [13] Lee, S. & Harris, J. (2012). Managing excellence in USA Major League Soccer: an analysis of the relationship between player performance and salary. *Managing Leisure*, 17(2/3), 106–123.
- [14] Palafox-Alcantar, P. G. & Vargas-Hernández, J. G. (2015). Measuring the efficiency of the 32 franchises in the NFL during the 2014 season. *Journal of Sports Economics & Management*, 5(1), 37–53.
- [15] Parkan, C. (2002). Measuring the operational performance of a public transit company. *International Journal of Operations & Production Management*, 22(6), 693–720.
- [16] Simmons, R. (2007). Overpaid Athletes? Comparing American and European Football. *Working USA*, 10(4), 457–471.
- [17] Transfermarkt. (2021). *Transfermarkt.com: Intern* [online]. Berlin, 2021 [cit. 2021-01-02]. Available at: <https://www.transfermarkt.com/intern/stellenangebote>.
- [18] Tunaru, R., Clark, E. & Viney, H. (2005). An option pricing framework for valuation of football players. *Review of Financial Economics*, 14(3-4), 281–295.
- [19] Zambom-Ferraresi, F. Iráizoz, L. I. & Lera-López, F. (2019). Performance Evaluation in the UEFA Champions League. *Journal of Sports Economics*, 18(5), 448–470.

Analysis of profitability of cryptocurrencies trading strategy based on technical analysis

Quang Van Tran,²1, Jaromír Kukal²

Abstract. Cryptocurrencies are a new phenomenon of the last decade which has attracted very much attention of theoreticians and practitioners. The volumes of their trades have immensely increased and the price development of some of them is so dramatic that they have become unignorable assets for potential investment target. A substantial amount of researches has been devoted to prediction of their price dynamics. To contribute to this stream, we propose a trading strategy that is based on technical analysis. We use ten various technical indicators and combine them together to form our trading algorithm for two most noticeable cryptocurrencies: Bitcoin and Ethereum. Then we evaluate its profitability and compare it to the one obtained from the traditional buy and hold strategy. The preliminary results show that if properly proceeded, the profitability of our strategy exceeds the one of buy and hold strategy. The results hence indicate the existence of inefficiencies in the cryptocurrencies trading market and the well-known Fama's efficiency market hypothesis may not fully hold in the case of these two cryptocurrencies.

Keywords: Technical analysis, Cryptocurrencies, Trading strategy, Profitability

JEL Classification: C13, C46

1 Introduction

Cryptocurrencies have attracted so much attention of theoreticians as well as practitioners for quite some time. Recently, they have reached new highs almost every day and certainly become interesting assets for investment. As such, reliable prediction of their price dynamics is an indispensable goal for researchers. Many forecasting techniques have been applied for this purpose [6, 9, 10]. Further, they seem to exhibit non-random price patterns which can be converted into profitable trading opportunities [7]. In this regard, technical analysis has also been used [2, 4, 5]. This also signals that the technical analysis may not be totally dead and it can be revived to detect these patterns. Though there is a number of studies devoted to the use of technical indicators on cryptocurrencies, we have not seen their combined effect in the literature. To verify this hypothesis, we propose a novel trading strategy, unlike in those previously cited works, based on ten technical analysis indicators. These ten most often used technical indicators are combined together in order to identify the potentially profit making opportunities. Based on the conclusion of technical analysis, we identify the trading opportunities as a trading strategy and then we evaluate its profitability when trading according to these signals and compare it to the traditional buy and hold strategy. In the framework of this research, two most attractive cryptocurrencies are chosen and they are Bitcoin and Ethereum. Due to very limited space for a contribution to the proceedings of the conference, the rest of our contribution is structured as follows: we will give a very brief description on the indicators we use in the next section and then go straight to the presentation of the results of our research and wrap up this report with some concluding remarks.

2 Technical Analysis and Its Indicators

Technical analysis is one of standard tools for analyzing and forecasting price movements in the financial markets using historical prices. From historical price series different charts and indicators (sometimes they are also euphemistically called market statistics) can be constructed. The principal idea of technical analysis is that prices in the financial markets based on the idea that price movements of financial assets tend to repeat themselves due to the collective and patterned behavior of investors therefore it is sufficient for an investor to identify the current price movement patterns and make his or her investment decisions based on this observed trends.

¹ University of Economics and Business, Prague, Department of Monetary Theory and Policy, W. Churchill sq. 4, Prague, Czech Republic
e-mail: tran@vse.cz., tranvqua@fjfi.cvut.cz

² Czech Technical University/Faculty of Nuclear Sciences and Physical Engineering, Trojanova 13, Praha 2, Czech Republic,
jaromir.kukal@fjfi.cvut.cz

Previously, technical analysts used to construct various charts from historical data and then they searched for such chart patterns like head-and-shoulders, lines of support, resistance, channels, double top/bottom reversal patterns, and other archetypes as flags, pennants, balance days and cup and handle. Later, these chart patterns have been often replaced by the so called technical indicators. Due to new developments in computational technology and techniques, these indicators can be easily computed a patterns can be identified through the values of these indicators.

So many technical indicators have been invented and an extensive overview on technical analysis can be found in works of Achelis or Murphy [1, 8]. Generally, technical indicators can be divided into several types: trend, momentum, volatility, volume related indicators. From a different perspective, they can also be divided into two groups: leading and lagging indicators. A lagging technical indicator is one that falls behind the price movement and it can be used to generate trading signals or to confirm the current trend of the movement. A leading indicator is a technical indicator that uses past price data to forecast future price movements in the market. A leading indicator is the one that can be used to forecast future price movements and therefore, hence it can give a signal to start a trade. For this research, ten indicators are selected. They are listed in Table 1 below.

Indicator	Full Name	Sell	Buy
ADI	Accumulation Distribution Indicator	Decreasing ADI	Increasing ADI
		Uptrend in Prices	Downtrend in Prices ADI
ADX	Average Directional Index	ADX > 20	AU < AD
		+DI > -DI	+DI < -DI
Aroon	Aroon Oscillator	AU < AD	AU > AD
		AU < 30, AD > 70	AU > 70, AD < 30
BB	Bollinger Bands	≤ Lower Band	≥ Upper Band
CCI	Commodity Channel Index	CCI ≤ -100	CCI ≥ 100
IC	Ichimoku Cloud	above green cloud	under red cloud
MACD	Moving Average Convergence Divergence	Under Signal Line	Above Signal Line
OBC	On Balance Volume	Decreasing trend	Increasing trend
RSI	Relative Strength Index	RSI ≥ 70	RSI ≤ 30
SOS	Stochastic Oscillator	SOS > 80	SOS < 20
		lower than its 3-day MA	higher than its 3-day MA

Table 1 The List of Used Indicators of Technical Analysis

Since there is not enough room for a more detailed description of these selected indicators, those who are interested in getting more information on them can turn to the already referred sources above or in the original works of these indicators, for example [12]. Other valuable sources of these indicators can be these two websites: **investopedia** and **school.stockcharts** [13]. From a typological point of view, the selection is made in such a way so that the indicators are from all types. Namely, Aroon, CCI, RSI and SOS are momentum indicators, ADX, MACD, IC are trend indicators (IC is also a momentum and support-resistance indicator), ADI, OBV are volume indicators, and BB is a volatility indicator. From another perspective ADI, CCI, OBV, RSI and SOS are leading indicators and ADX, BB, IC, MACD are lagging indicators. Ichimoku cloud due to its nature is both a leading and at the same time a lagging indicator.

3 Analysis of Profitability of a Trading Strategy based on TA Indicators

3.1 Data for Analysis

For this research, two most traded cryptocurrencies at this moment are chosen: Bitcoin and Ethereum. We use daily data in the the typical fashion OHLCV (Open, High, Low, Close and Volume data). Data are publicly available from website Bitstamp: **Data: cryptodatadownload**. Data for Bitcoin are from November of 2014 to May of 2021 [13]. Data for Ethereum are from November of 2017 to May of 2021. The price development of these two cryptocurrencies in the described periods are shown in Figure 1.

One can observe in Figure 1 a dramatic increase in prices of these two currencies since the end of the last year. From the data set some descriptive statistics are computed and they are displayed in Table 2. In this table what is striking is that with these two currencies one can make a profit of almost 30 % or also lose around 40 % over a night.

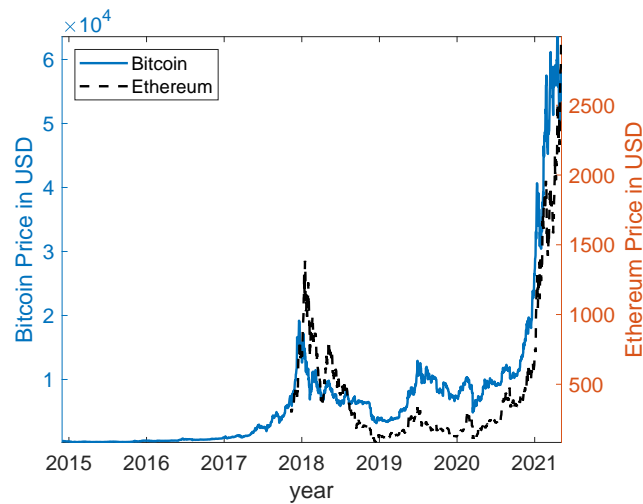


Figure 1 The development of price of cryptocurrencies of interest

We also compute another statistic which is the hypothetical return of each cryptocurrency from the beginning of each year upto now. The results are shown in Table 3. From this table, one can see that if someone bought Bitcoin at the beginning of this year, he can make profit of 95 % in May and over 300 % with Ethereum. And if someone bought Ethereum at the beginning of the last year, he could make profit of almost 22 times of his initial investment. The most profitable case is the one if someone bought Bitcoin at the beginning of 2015, he could make a profit of almost 180 times now in May of 2021. These numbers make the strategy buy and hold really hard to beat.

Characteristic	Bitcoin	Returns [%]	Ethereum	Returns [%]
Mean	7334.9	0.2944	470.1	0.3180
Median	4605.8	0.1798	260.1	0.1321
Mode	276.8	0	276.7	-43.76
Minimum	162	-38.98	82.9	-43.76
Maximum	63564.5	28.89	2991.8	26.27
Std deviation	10979.7	3.983	496.4	5.283
Skewness	3.165	-0.1252	2.346	-0.2749
Kurtosis	13.771	12.297	8.553	8.716
Obs.	2349	2348	1272	1271

Table 2 Descriptive Statistics of two cryptocurrencies and their returns

Year	Bitcoin in [%]	Ethereum in [%]
2021	94.64	309.17
2020	697.23	2196.60
2019	1396.84	2040.81
2018	325.72	297.43
2017	5629.95	831.09
2016	13057.09	-
2015	17993.24	-
2014	15109.61	-

Table 3 The cummulative returns of cryptocurencies from a certain year up to 2021

3.2 Analysis of the Profitability of the Trading Strategy

In this experiment we focus on identifying trading opportunities in the last six months. The reason of this choice are both the dynamics of the two cryptocurrencies in this period and the extremely high profitability of the buy and

hold strategy in the longer time frame. First, we identify the ideal trading opportunities. For this purpose, we set a threshold (in this case the value of this threshold is 3 % with respect to the transaction costs of trades). We define trading opportunities (or trading rules like in [2, 11]) as a series of peaks and troughs where the distance from a peak to a trough must exceed the preset threshold value ignoring the changes under the threshold value. The results are shown in Figure 2 and in the bottom of Table 4. The results can be used as a benchmark for technical analysis.

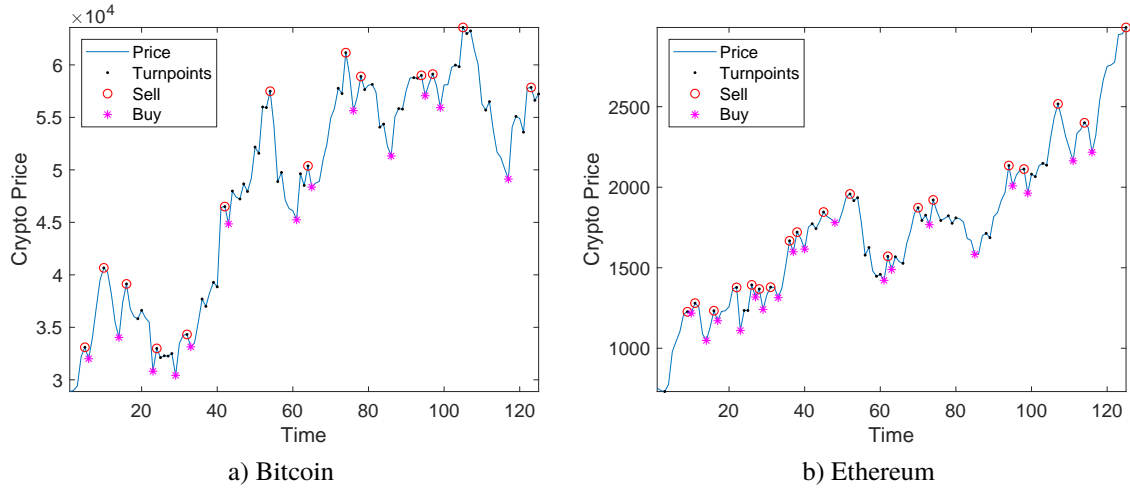


Figure 2 The Identified Hypothetical Trading Opportunities

Indicator	Bitcoin		Ethereum	
	Sell	Buy	Sell	Buy
ADI	0	5	0	2
ADX	1	6	0	12
Aroon	6	6	9	1
BB	12	0	24	2
CCI	15	3	9	41
IC	16	35	5	50
MACD	1	40	6	6
OBC	5	5	6	6
RSI	20	0	21	2
SOS	3	44	21	2
Hypothetical	14	13	19	18
Combination	6	22	3	30

Table 4 The Number of Trading Opportunities Generated by Technical Analysis

The main focus of this experiment is to find the trading windows using technical indicators. First we compute values of these indicators using data from the chosen period. Then these values are converted into the trading signal in three forms: buy, hold and sell. The results are displayed in Table 4. We can observe that the first two indicator generate very low number of trading signals in terms of buy and sell. Some indicators generate more buy signals than sell signals (BB, CCI, RSI) and vice versa (SOS). The rest produces a mix result. The asymmetry and inconsistency of the results makes us do the following correction. As indicators can generate a sequence of sell or buy signals and after selling cryptocurrencies we have money (in this case we get US dollars) and we need to buy cryptocurrencies back, we have to change the generated signals as follows. If we have a sequence of buy signals, then we leave only the first one unchanged and the rest in the sequence is switched to the hold signals. We do the same with a sequence of sell signals. By doing so, we obtain a series of alternating buy and sell signal. However, the correction technique also substantially reduces the number of trading opportunities. The results after reduction are shown in Table 5.

After that, we perform another experiment. We combine the trading signals generated by individual indicators

Indicator	Bitcoin		Ethereum	
	Sell	Buy	Sell	Buy
ADI	0	1	0	1
ADX	1	2	0	1
Aroon	5	6	2	1
BB	1	0	3	2
CCI	4	3	9	10
IC	12	13	5	6
MACD	1	2	5	6
OBC	5	5	5	6
RSI	1	0	3	2
SOS	3	4	3	2
Combination	5	6	3	4

Table 5 The Number of Trading Opportunities after Correction

together to get potentially more trading opportunities. The rule for combination is following. For each trading day, as the signals given by different indicators may be contradictory: ones tell us to sell and others tell us to buy, we ignore the hold signals and if that day the number of buy signals is higher than the number of sell signals, the resulting signal will be to buy and vice versa. The result for the combination is shown at the bottom of Table 4. We also apply the reduction technique described above to the result of our combination experiment and the result is at the bottom of Table 5. It is clear that this combination attempt does not lead to the increase of the number of trading opportunities.

As for the evaluation of profitability of our trading strategy, instead of the more common logarithmic returns, we compute return r_t as follows to get more accurate results:

$$r_t = \left(\frac{P_t}{P_{t-k}} - 1 \right) * 100, \quad (1)$$

where P_t is the price of a cryptocurrency at trading day t , P_{t-k} is its price k days earlier. The profitability of our strategies is shown in Table 6.

Strategy	Bitcoin in [%]	Ethereum in [%]
Hypothetical	313.63	473.44
Buy and Hold	98.16	297.47
Technical Analysis	60.38	242.50
Modified With Stop Loss	142.30	337.81

Table 6 The profitability of Various Trading Possibilities

It is clear that the hypothetical strategy can earn more money than strategy buy and hold as it can profit from rises and falls of the price of two cryptocurrencies. Unfortunately, trading based on technical indicators cannot bring profit as high as the one of buy and hold strategy. If we look at the price development of Ethereum in the period of our interest, we can clearly observe a growing trend with sharp drops at times which quickly return back to the trend. It seems that technical analysis can capture the general trend but it is unable to take into account these stiff falls. That could be the reason while technical analysis cannot generate a sufficient number of trading opportunities. As the dynamics of price of Bitcoin is more complex, it provides more trading signals detectable by technical analysis. However, it still cannot generate profit as good as the buy and hold strategy. In order to increase the profitability of our trading strategy, we modify our original strategy "buy all and sell all we have" to the one that allows to buy more times in a sequence and also to build the stop loss order into the strategy. This modification actually helps to increase the profitability of the original strategy and the resulting profit exceeds the profitability of the buy and hold strategy, see Table 6. This results show one can invent a strategy based on technical indicators that can identify nonrandom patterns in the price dynamics of cryptocurrencies and this strategy can beat the traditional buy and hold strategy. Though the increase is not extremely high, it may be considered to be an

evidence of how the well-known Fama's [3] efficiency market hypothesis may not totally hold in the case of these two cryptocurrencies.

4 Conclusion

Cryptocurrencies have become very attractive financial assets and a huge volume of researches has been devoted to identification of their nonrandom price patterns to earn some extra profits. One out of many possible tools to enter this space is technical analysis. We propose a trading strategy based on ten various technical indicators combined to generate trading signal for two most noticeable cryptocurrencies: Bitcoin and Ethereum. Our results show that technical indicators, alone or as a group, tend to underidentify the number of trading signals which in the end leads to insufficient profitability level compared to the one obtained from the traditional buy and hold strategy. This result is similar to the results reported in [11]. However, the profitability of our strategy exceeds the one of buy and hold strategy if it is properly modified in combination with the use of stop loss order like in [2]. The results hence indicate the existence of inefficiencies in the cryptocurrencies trading markets. In our opinion, the results may be substantially improved if technical analysis is used in combination with techniques that are capable of absorbing irregular changes round the main trend of the price dynamics which should be the topic of our future research.

Acknowledgements

This paper is financially supported by grant SGS20/190/OHK4/3T/14 and by the University of Economics and Business, Prague as a part of Institutional Research Support no. IP 100040.

References

- [1] Achelis, S. B. (1995). *Technical Analysis from A to Z*, Second Edition. New York: McGraw-Hill.
- [2] Corbet, S., Eraslan, V., Lucey, B., & Sensoy, A. (2019). The effectiveness of technical trading rules in cryptocurrency markets. *Finance Research Letters*, 31, 32-37.
- [3] Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25, 383–417.
- [4] Grobys, K., Ahmed, S., & Sapkota, N. (2020). Technical trading rules in the cryptocurrency market. *Finance Research Letters*, 32, 101396.
- [5] Hudson, R., & Urquhart, A. (2019). Technical trading and cryptocurrencies. *Annals of Operations Research*, 297, 191-220.
- [6] Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert systems with Applications*, 38(5), 5311-5319.
- [7] Kristoufek, L. (2018). On Bitcoin markets (in) efficiency and its evolution. *Physica A: Statistical Mechanics and its Applications*, 503, 257-262.
- [8] Murphy, J. J. (1991). *Intermarket technical analysis: trading strategies for the global stock, bond, commodity, and currency markets*. Hoboken, NJ: John Wiley & Sons.
- [9] Nakano, M., Takahashi, A., & Takahashi, S. (2018). Bitcoin technical trading with artificial neural network. *Physica A: Statistical Mechanics and its Applications*, 510, 587-609.
- [10] Pokorný, M. (2021). Neural Networks as a Tool of Technical Analysis for Cryptocurrencies' Price Prediction. *Bakalářská práce*, FJFI, Czech Technical University in Prague.
- [11] Ready, M. J. (2002). Profits from technical trading rules. *Financial Management* 31(3), 43-61.
- [12] Wilder, W. J. Jr (1978). *New Concepts in Technical Trading Systems*, 1st Edition. Winston-Salem, NC: Hunter Publishing Company.
- [13] Other Sources (2021). Technical Analysis and Data. <https://www.investopedia.com>, <https://school.stockcharts.com> and <https://www.cryptodatadownload.com/data/bitstamp/>.

GLMM Based Segmentation of Czech Households Using the EU-SILC Database

Jan Vávra¹

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 or general categorical. In our previous contribution to MME 2020 we presented a clustering method for such a type of data. The used thresholding approach of latent numeric counterparts of binary and ordinal outcomes suffered from slow convergence and unclear interpretation of resulting estimates. Hence we propose an alternative approach which again exploits a Bayesian variant of the model based clustering (MBC). Nevertheless, the underlying models are all of a generalized linear mixed model (GLMM) nature: (proportional odds) logit model for (ordinal) or binary indicators, multinomial logit model for general categorical outcomes and a standard linear mixed model for numeric outcome. Czech households interviewed within the EU-SILC project between 2005 and 2018 are then divided into several groups of similar evolution of income, housing costs, self-evaluations and other indicators.

Keywords: Multivariate panel data, Mixed type outcome, GLMM, Model based clustering, Classification

JEL Classification: C33, C38

AMS Classification: 62H30

1 Introduction

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), *ordinal* (e.g., level of ability to make ends meet) or *categorical* (e.g., ownership of a car/computer). It is our primary aim to use such longitudinally gathered outcomes towards segmentation of households according to typical patterns of their temporal evolution.

In our previous contribution [6] to MME 2020 we proposed a statistical model capable of joint modelling of longitudinal outcomes of diverse nature (*numeric, binary, ordinal*). We improve it by replacing thresholding of latent numeric outcomes with generalized linear mixed models (GLMM, [4]) appropriate to the type of the modelled outcomes. This change also allows us to extend the model for completely general categorical outcomes using multinomial logistic regression with random effects. This is a topic of Section 2. Consequently, we apply model based clustering (MBC, [3]) 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. We also dedicate one additional Section 4 to explain in more details, how do we use Markov chain Monte Carlo (MCMC, [2]) methods to estimate the model in Bayesian setting. The final Section 5 describes the use of this methodology on the Czech subset of the EU-SILC dataset. The paper is finalized by conclusions in Section 6.

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, level of a financial burden of housing, ...). Let $\mathbf{Y}_i = (\mathbf{Y}_{i,1}^\top, \dots, \mathbf{Y}_{i,R}^\top)^\top$ stand for a vector consisting of all the values $\mathbf{Y}_{i,r} = (Y_{i,r,1}, \dots, Y_{i,r,n_i})^\top$ of the r th outcome ($r = 1, \dots, R$) of the i th unit ($i = 1, \dots, n$) obtained at n_i occasions. Let C_i stand for available covariates

¹ Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic, Department of Probability and Mathematical Statistics, vavraj@karlin.mff.cuni.cz

(the times of measurements, possibly other explanatory variables) of i -th unit. Finally, let $h(\mathbf{y}_i; C_i, \zeta)$ represent the assumed distribution of the outcome vector \mathbf{Y}_i which possibly depends on the covariates C_i and also on a vector ζ of unknown parameters.

This setting corresponds to the one considered in [6], however the assumed distribution h will be different. Previously, we assumed linear mixed model (LMM, [5]) for numeric outcomes and for latent numeric counterparts of binary or ordinal outcomes. The observed categorical outcomes were then considered to be result of segmentation into intervals by unknown set of thresholds. However, the estimation of these parameters exhibited very slow convergence to posterior distribution (see Section 4). Moreover, estimated coefficients were tied to the dimensionless latent outcome and not directly to the observed ordered categories. For these reasons we decided to leave thresholding approach in favour of generalized linear mixed models (GLMM, [4]), which still allows us to join models for different outcomes through general combined distribution of random effects.

For each outcome r we consider a predictor $\eta_{i,r,j} = \eta_{i,r,j}^F + \eta_{i,r,j}^R$ for j th observation of i th unit that consists of

- fixed part $\eta_{i,r,j}^F = \mathbf{X}_{i,r,j}^\top \boldsymbol{\beta}_r$, where $\boldsymbol{\beta}_r$ are unknown fixed effects of regressors $\mathbf{X}_{i,r,j}$ created from covariates $C_{i,j}$,
- random part $\eta_{i,r,j}^R = \mathbf{Z}_{i,r,j}^\top \mathbf{B}_{i,r}$, where $\mathbf{B}_{i,r}$ are unit-specific unknown random effects of regressors $\mathbf{Z}_{i,r,j}$ created from covariates $C_{i,j}$.

The distribution of observed outcomes $Y_{i,r,j}$ is then supposed to depend on this predictor $\eta_{i,r,j}$ depending on what nature the outcome is. Let us for the moment drop the indices i, r, j .

- *Numeric* outcome Y is assumed to follow normal distribution with mean value η and precision parameter τ , which is a reciprocal value to standard variance parameter. This distribution can be summarized via logarithm of probability density function (log-pdf) $\ell^N(Y|\eta, \tau) = -\frac{1}{2} \log(2\pi) + \frac{1}{2} \log \tau - \frac{1}{2} \tau (Y - \eta)^2$.
- *Binary* outcome $Y \in \{0, 1\}$ is assumed to follow logistic regression model, which parametrizes probability of $Y = 1$ by inverse of *logit* function: $\text{P}[Y = 1|\eta] = \text{logit}^{-1}(\eta) = e^\eta / (1 + e^\eta)$. The corresponding log-pdf is then of form $\ell^B(Y|\eta) = Y \cdot \eta - \log(1 + e^\eta)$.
- *Ordinal* outcome $Y \in \{1, \dots, K\}$ of K ordered levels is assumed to follow ordinal logistic regression (ORL), where the probability of attaining level greater than $k \in \{1, \dots, K\}$ is parametrized by $p_k := \text{P}[Y > k|\eta, \mathbf{c}] = \text{logit}^{-1}(\eta - c_k)$, where $\mathbf{c} = (c_1, \dots, c_{K-1})$ is a set of ordered thresholds $-\infty = c_0 < c_1 < \dots < c_{K-1} < c_K = \infty$ that play a role of intercept. For identifiability purposes, the predictor η must not contain a fixed intercept term (or fix corresponding β parameter to zero). We extend our notation to $p_0 = 1$ and $p_K = 0$. The probability q_k of attaining level k can then be expressed as $q_k = p_{k-1} - p_k$, which finally yields the corresponding log-pdf $\ell^O(Y|\eta, \mathbf{c}) = \log q_Y = \log(p_{Y-1} - p_Y) = \text{logit}^{-1}(\eta - c_{Y-1}) - \text{logit}^{-1}(\eta - c_Y)$.
- *General categorical* outcome $Y \in \{1, \dots, K\}$ of K unordered levels is the newest addition to the model, for which we decided to use multinomial logistic regression (MLR) model that parametrizes probability of attaining level $k \in \{1, \dots, K\}$ by K different predictors η_k that share the same structure, however, they use different set of fixed effects $\boldsymbol{\beta}_{r,k}$. For simplicity of the model we use only one set of random effects $B_{i,r}$ creating only one η^R for all predictors η_k . In order to identify these effects, it is necessary to fix them to zero for one k , let us use $\eta_K = 0$. Then the probability that Y falls into category $k \in \{1, \dots, K\}$ is k th element of so called softmax function $\text{P}[Y = k|\eta_1, \dots, \eta_{K-1}] = \text{softmax}_k(\boldsymbol{\eta}) = e^{\eta_k} / \left(1 + \sum_{k'=1}^{K-1} e^{\eta_{k'}}\right)$, where $\boldsymbol{\eta} = (\eta_1, \dots, \eta_K)$ is vector of all predictors corresponding to categorical outcome Y . The resulting log-pdf is then of the form $\ell^C(Y|\boldsymbol{\eta}) = \eta_Y - \log\left(1 + \sum_{k=1}^{K-1} e^{\eta_k}\right)$.

Note that under $K = 2$ both ordinal and categorical setting reduce to the logistic regression assumed for binary outcomes. Such an ordinal outcome would require lone threshold c_1 that would correspond to negative intercept term in logistic regression since $q_1 = 1 - p_1$ and $q_2 = p_1$. On the other hand, a categorical outcome would then have one actual predictor $\eta = \eta_1$ and $\eta_2 = 0$ which trivially means that the first element of softmax coincides with logit^{-1} in logistic regression. To avoid such special cases we limit ourselves to $K \geq 3$ when using ORL or MLR, and hence any categorical outcome of $K = 2$ levels will be treated by logistic regression directly.

Now, when supposed distributions for observed outcomes are set, we discuss the other object of randomness - random effects. Let us for i th unit denote \mathbf{B}_i a vector of all random effects $\mathbf{B}_{i,r}$. For clarity, we will always consider outcomes to be ordered by their type - first numeric, then binary, ordinal and categorical as last and in this way we also order elements of vector of random effects \mathbf{B}_i . By the assumption that $\mathbf{B}_i \stackrel{\text{iid}}{\sim} \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} > 0$ is unknown completely general covariance matrix, we incorporate possible dependencies among outcomes into our model. Because even when the conditional distributions of different outcomes given random effects are independent, the unconditional distributions become dependent. Such distribution is possible to derive when only numeric outcomes are considered, however, once we add any categorical outcome modelled by GLMM it becomes

much more challenging task. In fact, the overall pdf h for \mathbf{Y}_i in general takes the form of

$$h(\mathbf{y}_i; C_i, \zeta) = \int \prod_{r=1}^R \prod_{j=1}^{n_i} \exp \left\{ \ell^{\text{type}(r)}(Y_{i,r,j} | C_{i,j}, \mathbf{B}_{i,r}, \zeta_r) \right\} \cdot (2\pi)^{-\frac{\dim \mathbf{B}_i}{2}} |\Sigma|^{-1} \exp \left\{ -\frac{1}{2} \mathbf{B}_i^\top \Sigma^{-1} \mathbf{B}_i \right\} d\mathbf{B}_i, \quad (1)$$

where $\text{type}(r) \in \{N, B, O, C\}$ denotes the type of r th outcome and ζ consists of all ζ_r that cover all unknown parameters tied with r th outcome. We propose to solve this integration by multivariate version of adaptive Gauss-Hermite quadrature using Laplace approximation (Breslow and Clayton, [1]). Our MCMC based estimation completely avoids the necessity of performing this integration. However, it is still needed for a detailed exploration of posterior distribution of several useful characteristics.

3 Model based clustering

In this section we discuss the way we can distinguish different patterns in joint model for mixed type panel data. We are given with the number of groups G into which we intend to classify the units in advance and $G \geq 2$. The classification proceeds by using the model outlined in Section 2 within the Bayesian model based clustering procedure (MBC, [3]). It is assumed that the overall model, f , is given as a finite mixture of certain group-specific models f_g , $g = 1, \dots, G$. That is, $f(\mathbf{y}_i; C_i, \theta) = \sum_{g=1}^G w_g f_g(\mathbf{y}_i; C_i, \psi, \psi^g)$, where $\mathbf{w} = (w_1, \dots, w_G)^\top$ are the mixture weights (proportions of the groups in the population), ψ is a vector of unknown parameters common to all groups and ψ^g , $g = 1, \dots, G$, are vectors of group-specific unknown parameters. Hence the vector θ of all unknown parameters is $\theta \equiv \{\mathbf{w}, \psi, \psi^1, \dots, \psi^G\}$.

Using the notation from previous section we set the group-specific density f_g to be the density h , however, depending on parameter ζ^g elements of which $(\beta_r^g, \tau_r^g, \mathbf{c}_r^g, \Sigma^g)$ 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 $\psi = \{\tau, \mathbf{c}, \Sigma\}$ and $\psi^g = \beta^g$.

Further, let $U_i \in \{1, \dots, G\}$ denote the unobserved allocation of the i th unit into one of the G groups. As it is usual with the mixture models, the group-specific distribution $f_g(\mathbf{y}_i; C_i, \psi, \psi^g)$, $g = 1, \dots, G$, can be viewed as a conditional distribution of the outcome \mathbf{Y}_i given $U_i = g$ while the mixture weights \mathbf{w} determine the marginal distribution of the allocations, i.e., $P(U_i = g) = w_g$, $g = 1, \dots, G$. Classification of the i th unit is then based on suitable estimates of the conditional individual allocation probabilities $p_{i,g}(\theta)$, $g = 1, \dots, G$, calculated by the Bayes rule:

$$p_{i,g}(\theta) = P(U_i = g | \mathbf{Y}_i = \mathbf{y}_i; C_i, \theta) = \frac{w_g h(\mathbf{y}_i; C_i, \psi, \psi^g)}{f(\mathbf{y}_i; C_i, \theta)}, \quad (2)$$

where $\theta = \{\psi, w_g, \psi^g, g = 1, \dots, G\}$ is the set of all unknown parameters. Unfortunately, calculation of such probabilities would require immense amount of calculation of the integral (1). However, we can still bypass this issue by taking advantages of Bayesian approach and MCMC estimation, see the end of Section 4.

4 Estimation by MCMC

A Bayesian approach that treats unknown parameters to be random by assigning an uninformative prior is a natural choice for our model. It allows us to consider all random effects \mathbf{B}_i and group allocation indicators U_i as additional unknown parameters of the model. Markov chain with states consisting of all unknown parameters θ , latent \mathbf{B}_i and U_i needs to be constructed in such a way that its limiting distribution corresponds to their posterior distribution, i.e. the distribution of $\theta, \mathbf{B}_i, U_i, i = 1, \dots, n$ given all available data. Once the chain reaches its stationary distribution (equal to the limiting one) we follow Monte Carlo principles and continue in sampling to obtain a sample of state values on which an estimation of posterior distribution is based.

The question is, how do we construct such a Markov chain. The most straightforward way (taken in [6]) using Gibbs sampling cannot be applied directly, since needed full-conditional distributions of fixed effects β_r^g , random effects \mathbf{B}_i and ordered intercepts \mathbf{c}_r^g do not fall into known distributional families. Nevertheless, we still keep the main idea of sampling from full-conditional distributions but in cases where the full-conditional distribution is unclear we rather replace it with a Metropolis step that accepts a new sampled value from proposal distribution with a corresponding acceptance probability. This so-called *Metropolis-within-Gibbs* approach satisfies all conditions necessary for MCMC to work [2].

However, there still remains the problem of selecting the proposal distributions in Metropolis steps. A vague choice could lead to poor acceptance probability and inefficient Markov chain. In this regard we adapt the

proposal distribution to reflect the true full-conditional distribution behind. By Newton-Raphson method we find the parameter value maximizing full-conditional distribution and evaluate the second order derivatives at this parameter value to obtain a variance matrix for proposal distribution. To be more specific, we transition from current value to the new one where the direction is sampled from centered multivariate normal distribution with such variance matrix possibly modified by a constant to reach better acceptance rate. To reduce autocorrelation (and speed up convergence) we can perform a walk consisting of more than one step. The frequent update of variance matrices of proposal distributions is crucial at the burn-in period of the Markov chain, however, once the stationary distribution is reached it could be updated at lower rate to speed up the sampling.

Finding proposal distributions for fixed effects β_r^g is analogous to finding maximum likelihood estimates in GLM models on subset of units in g th cluster. However, under the Bayesian setting we do not maximize only log-pdf, but we also add a logarithm of prior distribution. During Newton-Raphson only fixed part of predictor η^F changes while the random part η^R is kept the same since we condition by the random effects in this step. The process very analogous for random effects \mathbf{B}_i , however, this time η^F remains the same while the random part η^R changes. Unlike with the fixed effects β_r^g , where each was done separately for all $r = 1, \dots, R$, random effects \mathbf{B}_i have to be sampled jointly across all outcomes due to our common general prior $\mathbf{B}_i|U_i = g \sim \mathbf{N}(\mathbf{0}, \Sigma^g)$. Note that each unit $i = 1, \dots, n$ has to be treated separately while using parameters specific to cluster to which unit i currently belongs.

A unique treatment is required for c_r^g parameter of ordered intercepts for ordinal outcomes. Derivation of Newton-Raphson method leads to an algorithm that does not necessarily guarantee the ordinality of elements of c_r^g . Hence, we propose the following transformation

$$\begin{aligned} c_1 &= a_1 & a_1 &= c_1, \\ c_2 &= a_1 + e^{a_2} & a_2 &= \log(c_2 - c_1), \\ c_3 &= a_1 + e^{a_2} + e^{a_3} & a_3 &= \log(c_3 - c_2), \\ &\vdots & &\vdots \\ c_{K-1} &= a_1 + \sum_{k=2}^{K-1} e^{a_k} & a_{K-1} &= \log(c_{K-1} - c_{K-2}). \end{aligned}$$

There is one to one correspondence between the ordered intercepts c_r^g and newly defined parameter a_r^g , elements of which are no longer restricted. Therefore, we set multivariate normal prior over parameter \mathbf{a}_r^g and apply Metropolis step to it with the use of proposal distribution found by Newton-Raphson method (the transformation preserves smoothness), which now does not face any limitations. After sampling \mathbf{a}_r^g we transform it back to obtain ordered intercepts c_r^g .

In order to fully approximate posterior distribution of parametric functions of allocation probabilities $p_{i,g}(\theta)$ we would need to efficiently compute integral (1). Nevertheless, if we settle for posterior mean only, we can utilize full-conditional (also given \mathbf{B}_i) classification probabilities that are already calculated for each sampled state in order to obtain new U_i indicators. By the fact that

$$\int \mathbb{P}[U_i = g | \mathbf{Y}_i; C_i, \theta] \cdot p(\theta | \mathbf{Y}_i; C_i) d\theta = \int \int \mathbb{P}[U_i = g | \mathbf{Y}_i; C_i, \theta, \mathbf{B}_i] \cdot p(\theta, \mathbf{B}_i | \mathbf{Y}_i; C_i) d\mathbf{B}_i d\theta,$$

where $p(\cdot | \circ)$ denotes conditional pdf of \cdot given \circ , we can approximate the posterior mean of $p_{i,g}(\theta)$ by arithmetic mean of full-conditional probabilities. This approximation can then be used for classifying units into cluster with highest posterior mean probability. However, without full exploration of posterior distribution we can hardly create more sophisticated classification rules that would adequately account for possible indecisiveness.

5 Application to EU-SILC

The Czech subset of EU-SILC dataset gathered between 2005 and 2018 consists of $n = 23\,360$ households that were followed for exactly $n_i = 4$ consecutive years which is induced by rotational design. Each year a quarter of the followed households is dropped to be replaced by a set of new households. For the analysis we primarily use data gathered on household level, however, on special occasions we use gathered personal data to create a new indicator that summarizes the whole household. Outcomes of interest are listed below with respect to their type:

- Numeric outcomes (used on log-scales)
 - HX090 - Equivalised total disposable income [€/year]
 - HS130 - Lowest monthly income to make ends meet (to pay for its usual necessary expenses) [€/month]

- Binary outcomes (Yes / No)
 - HS040 - Capacity to afford paying for one week annual holiday away from home
 - HS060 - Capacity to face unexpected financial expenses
- Ordinal outcomes (the higher, the less problematic)
 - HS120 - Ability to make ends meet (self-evaluation by respondent)
 1. with great difficulty
 2. with difficulty
 3. with some difficulty
 4. fairly easily
 5. easily
 6. very easily
 - HS140 - Financial burden of the total housing cost (self-evaluation by respondent)
 1. a heavy burden
 2. a slight burden
 3. not a burden at all
- Categorical outcomes (Yes / No - cannot afford / No - other reason)
 - HS090 - Do you have a computer?
 - HS110 - Do you have a car?

The dataset also offers plenty of potential regressors to be used in the fixed part of the model.

- *Time* is the most important one due to the economical crisis that has struck during the follow-up time. This had a great impact on the households in several different ways. Hence, we decided for quadratic spline parametrization with one inner knot to allow for the change in evolution.
- *Equivalised household size* expresses how large the household is while taking age into consideration. The head of the household (respondent) has unit weight, while other members have either 0.5 or 0.3 depending on their age - older or younger than 14, respectively. We keep it in a linear parametrization as one of the fixed effects.
- *Level of urbanisation* was originally divided by the population density and minimum population into three categories: densely populated area, intermediate area and thinly-populated area. However, the capital city Prague behaves in many aspects very distinctly. For that reason we created fourth category dedicated to households in Prague exclusively.
- *The highest education level achieved* within the whole household rarely attains the lowest possible option of primary education. The most common one - secondary education was divided into lower, upper and post secondary categories. The highest group - tertiary education - are households where at least one member has some kind of university degree.
- *Presence of student* or *baby* are binary indicators whether some household member currently attends any educational institution or is younger than 3 years, respectively.

Usage all of the above suggested regressors creates the fixed part of the predictor dependent on 13-dimensional vector of β coefficients plus an intercept term for non-ordinal outcome. Each of the outcomes shares the same structure of the fixed effects but is supposed to have his own group-specific β_r^g . On the other hand, we try to keep random structure as simple as possible to not to increase the dimension of Σ (common to all clusters). Therefore, a simple random intercept term for each considered outcome suffices for our purposes of joint modelling. Then the dimension of \mathbf{B}_r is the lowest possible and corresponds to the number of considered outcomes, which means 8 when all outcomes suggested above are used.

During the analysis we treat the number of groups G to be fixed and known in advance. However, nothing stops us from trying several potential values of $G \in \{2, 3, 4, 5, 6\}$ and choose the one that fits best according to cluster interpretations. In [6] we also used posterior distribution of deviance $D(\theta; \mathbf{Y}_1, \dots, \mathbf{Y}_n) = -2 \sum_{i=1}^n \log f(\mathbf{Y}_i; C_i, \theta)$.

However, this approach is as computationally demanding as full exploration of posterior distribution of all classification probabilities $p_{i,g}(\theta)$, which for Markov chain of length M includes $M \cdot n \cdot G$ approximations of integral (1).

We classify households into the cluster with the highest posterior mean of classification probabilities (2) only when this estimate dominates other probabilities by a certain margin, e.g. 0.1. Apart from cluster characteristics given by specific fixed effects β_r^g we can relate the clustering with other available data. A very interesting comparisons are with the type of the household (its family composition) and with the poverty indicator (whether a household has equivalised total disposable income below 60 % of the median). It shows that our clustering method provides an alternative and more elaborated approach for identification of households endangered by poverty.

6 Conclusion

Although, replacement of latent variable approach by GLMM brings several difficulties with respect to the sampling process for estimation, we were able to circumvent them by the use of Metropolis proposal steps combined with Newton-Raphson method. For this effort we were rewarded by significantly improved convergence properties. However, some of the difficulties still remain in the integration (1). For example, when we want to explore posterior distribution of classification probabilities or deviance in more depth. Although, suggested solution by multivariate version of adaptive Gauss-Hermite quadrature using Laplace approximation (Breslow and Clayton, [1]) can handle the integration quite satisfactorily, the enormous amount of its applications significantly prolongs the computational time.

The next step in the improvement could be to replace generality of covariance matrix Σ by some block structure to make the dependence clearer and to allow for much higher dimensions of random effects and possibly much larger set of outcomes of interest. The GLMM framework and the use of Newton-Raphson method also opens the door for other types of models for numeric outcomes, choice of different link functions than logit or even for addition of count-type outcomes.

In terms of real data application, our methodology can identify households of both regular and extraordinary behaviour. Classical poverty indicators tend to be one-dimensional and do not take all household aspects into consideration. Our method can absorb diverse kinds of information and then provide a more complex insight. One of the advantages of being still a fully parametric model is that we can describe exactly how the clusters differ among themselves.

Acknowledgements

This research was supported by the Czech Science Foundation (GAČR) grant 19-00015S.

References

- [1] Breslow, N. E. and Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, **88**, 9–25.
- [2] Brooks, S., Gelman, A., Jones, G. and Meng, X. (2011). *Handbook for Markov chain Monte Carlo*. Taylor & Francis.
- [3] Fraley, C. and Raftery, A. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, **97**, 611–631.
- [4] Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer-Verlag New York.
- [5] Laird, N. M. and Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, **38**, 963–974.
- [6] Vávra, J. and Komárek, A. (2020). Identification of Temporal Patterns in Income and Living Conditions of Czech Households: Clustering Based on Mixed Type Panel Data from the EU-SILC Database. *Conference Proceedings of 38th International Conference on Mathematical Methods in Economic (MME2020)*, 612–617.

State-space modeling of claims reserves in non-life insurance

Petr Vejmelka¹

Abstract. One of the fundamental activities of insurance companies is the calculation of technical reserves. The available data are usually considered in the form of run-off triangles. Prediction of unknown values, which are the basis for the calculation of reserves, can be constructed not only using simple approaches such as chain-ladder, but also using advanced methods such as state-space modeling. This paper focuses on the construction of Kalman projections of the values in dependent run-off triangles, which can be considered in the form of a time series with missing observations. Since the quantiles, currently the preferred risk measure, or even the whole distribution of the reserves are highly important, their estimation using smoothed bootstrap is also the content of this work. The proposed methods are applied to real data consisting of two dependent run-off triangles in order to verify their applicability in practice. The obtained results are compared to the ones obtained by other procedures.

Keywords: run-off triangle, reserve, IBNR, state-space model, time series, bootstrap

JEL Classification: C32, C53, G22

1 Introduction

In this paper we focus on the estimation of reserves in dependent run-off triangles of similar lines of business in non-life insurance. An insurance company observes reported claims, but also needs to estimate the volume of incurred but not reported claims (IBNR) to create IBNR reserves. One of the possible approaches how to estimate these reserves is state-space modeling. The log-normal model presented in Hendrych and Cipra (2020) is used in this paper, but a different principle of quantile estimation of the reserves is introduced, namely a method based on the smoothed bootstrap. The advantage of this multivariate approach lies in the use of all available information for reserving. Hence, one can expect more reliable results in this case compared to those obtained separately. Moreover, in practice, the state-space model approach achieves better results than the usually used simple approaches.

The paper is organized as follows. Section 2 presents multivariate reserve estimation using the log-normal model in the context of state-space modeling. In Section 3, particular steps of the smoothed bootstrap for estimating quantiles of the reserves are introduced. Section 4 is devoted to the practical implementation of the proposed estimation methods and the comparison of different approaches. Finally, Section 5 summarizes conclusions achieved in this paper.

2 Reserve estimation using state-space modeling

Since the method of reserve estimation is based on state-space modeling, we will introduce the linear state-space model, which can be seen as the fundamental model. The model presented in Section 2.2 follows that structure.

2.1 Linear state-space model

Linear state-space model, also called dynamic linear model, is given by the following system of equations:

$$y_t = Z_t \alpha_t + \varepsilon_t, \quad (1)$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \quad (2)$$

where y_t is a p -dimensional observation vector at time t , α_t is a m -dimensional state vector at time t and further Z_t , T_t and R_t are matrices of parameters of types $(p \times m)$, $(m \times m)$ and $(m \times k)$. It is assumed that $\varepsilon_t \sim N(0, H_t)$, $\eta_t \sim N(0, Q_t)$ and $\alpha_1 \sim N(a_1, P_1)$ are independent, where H_t and Q_t are covariance matrices of types $(p \times p)$ and $(k \times k)$ and both a_1 and P_1 are some initial estimates.

Two main approaches how to estimate α_t and y_t in state-space models are the Kalman filtration and the Kalman smoothing. They differ by the information which is taken into account when α_t is estimated. Since there are

¹ MFF UK, KPMS, Sokolovská 83, 186 75 Praha 8, vejmelp@karlin.mff.cuni.cz

missing values in our data that correspond to the values in the estimated triangles, the Kalman smoothing will be used for their estimation. This procedure provides us with the smoothed values $\alpha_{t|n}$, where $t < n$ and in the assumed context, n is the number of observations (the number of cells in each run-off triangle). Thus,

$$\alpha_{t|n} = E(\alpha_t | y_1, \dots, y_n), \quad (3)$$

$$P_{t|n} = \text{Var}(\alpha_t | y_1, \dots, y_n). \quad (4)$$

2.2 Multivariate log-normal state-space model

When using state-space models for the reserve estimation, one can apply various models. We will present here only one of them which was proposed in Hendrych and Cipra (2020) and which will be used in the numerical analysis. This so-called log-normal model assumes that the incremental claims X_{ij} have log-normal distribution, i.e., $X_{ij} \sim LN(\mu_{ij}, \sigma_\varepsilon^2)$. Therefore, for $Y_{ij} = \log X_{ij}$ we assume

$$Y_{ij} = \mu_{ij} + \varepsilon_{ij}, \quad \varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2), \quad (5)$$

where

$$\mu_{ij} = c + a_i + b_j. \quad (6)$$

In (6) the parameters a_i represent the row effects and b_j the column effects.

This double-index form must be transformed into the time format. One of the possible options is to assume that the values are stacked row-wise. In order to set some initial levels in the observation equation, the values in the first column will be subtracted from the values in each row and will not appear in the vector y_t , see (7). If there are $s + 1$ columns in the run-off triangles (where $i = 0, \dots, s$ and $j = 0, \dots, s$), the corresponding time index is $t = i \cdot s + j$. Generally, the number of rows does not need to be the same as the number of columns, but for simplicity let us assume that it holds. However, the notation used so far supposed only one run-off triangle. For this reason, the multivariate version of the state-space model will be introduced. Let us suppose that there are N run-off triangles available with values $X_{ij}(n)$, where n denotes the n th run-off triangle for $n = 1, \dots, N$, which can be logarithmized to $Y_{ij}(n)$. Then the log-normal model can be written as follows:

$$y_t(n) - y_t^0(n) = \alpha_t(n) + \varepsilon_t(n) \quad (7)$$

$$\alpha_{t+1} = \alpha_{t-s+1}(n) + \eta_t(n), \quad (8)$$

where $y_t(n)$ and $y_t^0(n)$ represent the corresponding terms $Y_{ij}(n)$ and $Y_{i0}(n)$ according to the time index formula (see above), $\varepsilon_t(n) \stackrel{\text{ind.}}{\sim} N(0, \sigma_\varepsilon(n, n))$ and $\eta_t(n) \stackrel{\text{ind.}}{\sim} N(0, \sigma_\eta(n, n))$. According to the used notation, $\sigma_\varepsilon(m, n) = \text{Cov}(\varepsilon_t(m), \varepsilon_t(n))$ and similarly for $\sigma_\eta(m, n)$, where $m, n = 1, \dots, N$.

It is necessary to adjust these equations into the form of dynamic linear model which can be further estimated by means of a suitable software. The final form of the model is

$$y_t - y_t^0 = \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix} \alpha_t + \varepsilon_t. \quad (9)$$

$$\alpha_{t+1} = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ & & \ddots & & \\ & & & 0 & 0 & \dots & 0 & 1 \\ & & & 1 & 0 & \dots & 0 & 0 \\ & & & \vdots & \vdots & & \vdots & \vdots \\ & & & 0 & 0 & \dots & 1 & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ & & \ddots & & \\ & & & 1 & 0 & \dots & 0 & 0 \\ & & & 0 & 0 & \dots & 0 & 0 \\ & & & \vdots & \vdots & & \vdots & \vdots \\ & & & 0 & 0 & \dots & 0 & 0 \end{pmatrix} \eta_t, \quad (10)$$

where $y_t = (y_t(1), \dots, y_t(N))'$, $y_t^0 = (y_t^0(1), \dots, y_t^0(1))'$, $\alpha_t = (\alpha_t(1), \dots, \alpha_{t-s+1}(1), \dots, \alpha_t(N), \dots, \alpha_{t-s+1}(N))'$, $\varepsilon_t = (\varepsilon_t(1), \dots, \varepsilon_t(N))'$ and $\eta_t = (\eta_t(1), 0, \dots, 0, \dots, \eta_t(N), 0, \dots, 0)'$. The covariance matrices of the residual vectors ε_t and η_t are then $\text{Var}(\varepsilon_t) = H_t = (\sigma_\varepsilon(m, n))_{m,n=1,\dots,N}$ and $\text{Var}(\eta_t) = Q_t = (\sigma_\eta(m, n))_{m,n=1,\dots,N}$.

Our main interest lies in reserve calculation. Therefore, it is necessary to estimate the missing values from the run-off triangles. Their sum corresponds to the reserve. In practice, one can use various software systems, where the estimation procedures are implemented. One of them is the KFAS package in R, which is also used in the numerical part, see Helske (2017). As the first step, the model is fitted and the unknown parameters are estimated via maximum likelihood, and further the smoothed values are calculated. Since we assumed the logarithmized values Y_{ij} in the state space model, it is necessary to transform the smoothed values back. In Atherino et al. (2010) a method using the property of the log-normal distribution is proposed. The desired values \hat{X}_{ij} are calculated as $\hat{X}_{ij} = \exp \left\{ \hat{Y}_{ij} + \sigma_{Y_{ij}}^2 / 2 \right\}$.

Although in most situations the log-normal model achieves good results, its use is limited if there are negative values in the development triangles. In that case it is not possible to calculate logarithms of negative values and these values must be either substituted by some positive values (e.g., means of neighbouring values) or taken as further missing values. However, these adjustments deteriorate the accuracy of the estimates.

3 Quantile estimation

In the previous section, we have shown how to estimate the reserves. In practice, the reserve itself is often insufficient. In many situations, one also need to know whether reserve would be high enough to cover the claims with a given probability, e.g., 95%. Bootstrap is a popular method how to estimate the distribution of a sample statistics. In literature, many modifications of the basic principle have been proposed. In order to estimate the distribution of the reserves, we will use the so-called smoothed bootstrap. The main difference from the non-parametric bootstrap is that in the smoothed bootstrap we resample the residuals from their estimated distribution, not only from the residuals themselves. Thus, before resampling, it is necessary to calculate the residuals and estimate their density. This is done for each development triangle separately and (n) will not be written in the following text. As the basis for the residuals serve differences between the values in the observed triangle and the smoothed values. The first column is not included since these values were not smoothed. The actual residuals will be constructed in a similar way as Pearson residuals in GLM. For distributions from the exponential family, Pearson residuals would be calculated as the quotients of $X_{ij} - \hat{X}_{ij}$ and $\sqrt{V(\hat{X}_{ij})}$, where $V(x)$ denotes the variance function of that distribution and . Log-normal distribution is not a member of the exponential family, but since the variance is related to the mean squared, we will assume $V(x) = x^2$. Hence, the residuals, whose density is to be estimated, are constructed in the following way:

$$r_{ij}^p = \frac{X_{ij} - \hat{X}_{ij}}{\hat{X}_{ij}}. \quad (11)$$

Since our aim is to estimate the density of the calculated residuals, we will use the kernel estimation of the form:

$$\hat{f}_w(x) = \frac{1}{wh_w} \sum_{i=1}^w K \left(\frac{x - r_i^p}{h_w} \right), \quad (12)$$

where w is the number of r_{ij}^p that are available and h_w is the bandwidth. There exist several kernels and rules how to set the bandwidth, but we will not present them here.

The next step is to resample B samples of the residuals for each triangle from the estimated densities, construct the resampled run-off triangles and with the use of the log-normal model calculate the IBNR reserves for each of them. The usual choice of B is 1,000. Finally, one can calculate the sample quantiles or estimate the density of the reserves.

For the sake of clarity, an overview of steps leading to the reserves calculation and quantiles estimation follows:

1. construction of vectors $y_t(n)$ and $y_t^0(n)$ from the observed values $X_{ij}(n)$,
2. estimation of the smoothed values $\hat{y}_t(n)$ and their back-transformation to $\hat{X}_{ij}(n)$,
3. calculation of the residuals and kernel density estimation in each run-off triangle,
4. resampling B samples of residuals and their use for construction of B sets of run-off triangles,
5. reserves calculation for each resampled set of run-off triangles,
6. quantile and density estimation of the reserves.

4 Numerical analysis

In this numerical study, real data from non-life insurance will be used to demonstrate the use of the multivariate log-normal model and the estimation of the distribution and quantiles of IBNR reserves. We will work with data from Shi, Basu and Meyers (2012). These consist of two run-off triangles corresponding to an insurance portfolio of personal auto and commercial auto lines of business from a major property-casualty insurer in the United States. These triangles are presented in Tables 1 and 2. Due to the nature of the data, these triangles can be expected to be highly correlated. This expectation is in accordance with the calculated sample correlation coefficient, which is equal to 0.73.

Acc. year	Development year									
	0	1	2	3	4	5	6	7	8	9
0	1,376,384	1,211,168	535,883	313,790	168,142	79,972	39,235	15,030	10,865	4,086
1	1,576,278	1,437,150	652,445	342,694	188,799	76,856	35,042	17,089	12,507	
2	1,763,277	1,540,231	678,959	364,199	177,108	78,169	47,391	25,288		
3	1,779,698	1,498,531	661,401	321,434	162,578	84,581	53,449			
4	1,843,224	1,573,604	613,095	299,473	176,842	106,296				
5	1,962,385	1,520,298	581,932	347,434	238,375					
6	2,033,371	1,430,451	633,500	432,257						
7	2,072,061	1,458,541	727,098							
8	2,210,754	1,517,501								
9	2,206,886									

Table 1 Run-off triangle - data 1

Acc. year	Development year									
	0	1	2	3	4	5	6	7	8	9
0	33,810	45,318	46,549	35,206	23,360	12,502	6,602	3,373	2,373	778
1	37,663	51,771	40,998	29,496	12,669	11,204	5,785	4,220	1,910	
2	40,630	56,318	56,182	32,473	15,828	8,409	7,120	1,125		
3	40,475	49,967	39,313	24,044	13,156	12,595	2,908			
4	37,127	50,938	34,154	25,455	19,421	6,728				
5	41,125	53,302	40,289	39,912	6,650					
6	57,515	67,881	86,734	18,109						
7	61,553	132,208	20,923							
8	112,103	33,250								
9	37,554									

Table 2 Run-off triangle - data 2

Various methods are offered for reserve calculation. The basic approach is based on Mack's model, where unknown values are estimated using development factors. A nonparametric bootstrap can then be used to calculate the quantiles. Since this multivariate case can be viewed as a SUR system and contemporaneous correlations should be taken into account, a method for multidimensional estimation was proposed in Zhang (2010). We focus on the state-space modeling as presented in Section 2 with the use of the log-normal model. The difference from Hendrych and Cipra (2020) is that the quantiles and the density of the reserves are estimated via the procedure described in Section 3. For the estimation of the missing values and the subsequent determination of reserves, we will use the KFAS package in R. Figure 1 shows both the actual claims (full black line) and the projections (dashed red line).

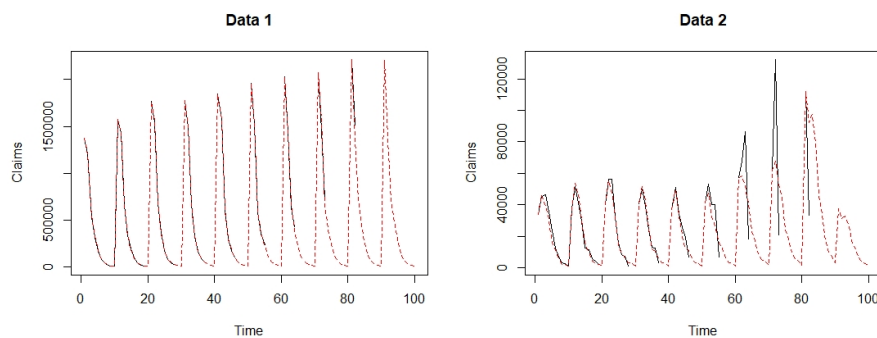


Figure 1 Original and projected claims

In the case of the personal auto line of business (Data 1), the estimated curve fits the original claims very well,

while in the case of the commercial auto line of business (Data 2), there are some noticeable differences between the actual and the estimated values, especially in later years. This is caused by a significant increase in claims in the sixth, seventh and eighth year. Some of these values could be considered as outliers, and closer attention should be paid to them during the estimation. Since the Kalman smoothing is used for the estimation, this issue is already partially taken into account.

We use the estimated values corresponding to the claims from the observed triangles to calculate the residuals, for which kernel estimates of densities are then constructed. These estimates are considered as one-dimensional for simplicity. When estimating the densities, we use the Gaussian kernel and the bandwidth h_n calculated using Silverman's rule of thumb

$$h_n = 0.9n^{-1/5} \min\{S_n, IQR_n\}, \tag{13}$$

where n is the sample size (in our case $n = 45$), S_n stands for the standard deviation and IQR_n corresponds to the interquartile range of the residuals. Estimated densities of residuals are plotted in Figure 2, where also the calculated bandwidths are stated.

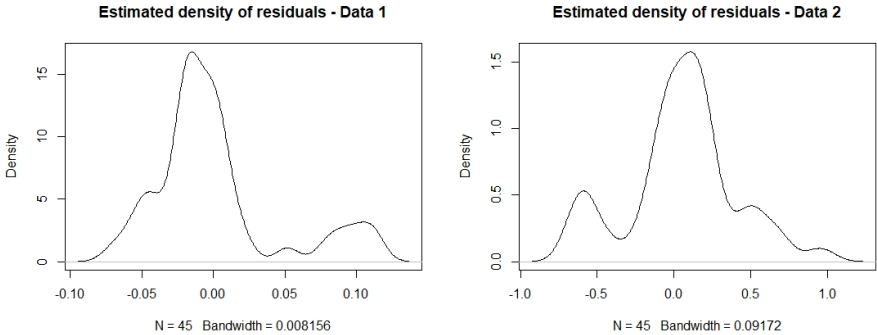


Figure 2 Kernel density estimates of residuals

These densities serve as the bases for the bootstrap. We will generate 1,000 pairs of vectors of bootstrapped residuals in total, which are transformed back in order to create 1,000 pairs of development triangles. For each pair there are again calculated the reserves. Finally, these values are used to obtain quantiles. Furthermore, kernel density estimates of IBNR reserves can be also constructed. Figure 3 shows the estimated densities of the reserves for the considered portfolio.

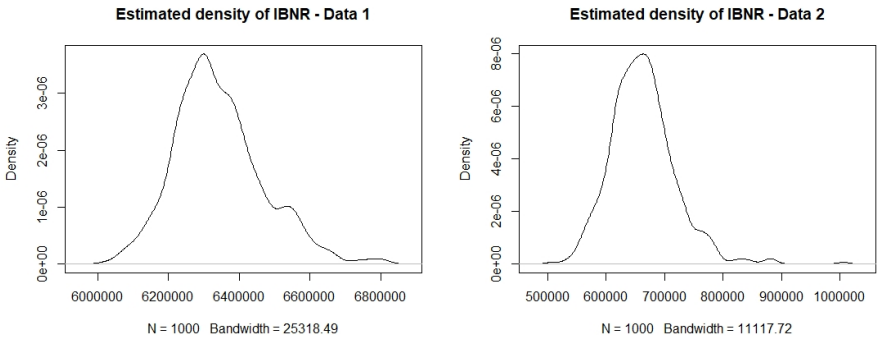


Figure 3 Kernel density estimates of IBNR

Due to the fact that for the second triangle there are larger differences between the original and estimated values, the quantiles of the reserve are also significantly wider if one takes into account the amount of the reserve. The obtained results are summarized in Table 3, where for comparison also corresponding results for other possible methods mentioned above are presented. Unfortunately, there is no suggestion on how to calculate quantiles in Zhang (2010) and thus, they are not part of Table 3.

Method	Data	IBNR	CV (%)	75% quantile	90% quantile	95% quantile	99% quantile
Log-normal - kernel	Data 1	6,327,917	1.99	6,407,219	6,526,457	6,566,421	6,678,027
	Data 2	643,776	8.46	691,490	729,950	761,566	839,624
Log-normal - smoother	Data 1	6,579,354	4.00	6,827,001	7,058,335	7,201,213	7,456,239
	Hendrych&Cipra (2020)	Data 2	639,516	17.00	702,078	784,325	856,898
MackChain	Data 1	6,439,764	4.39	6,639,997	6,820,014	6,933,749	7,168,620
Mack (1993)	Data 2	487,401	20.03	548,763	618,248	654,568	739,130
MultiChain	Data 1	6,439,974	5.01	-	-	-	-
Zhang (2010)	Data 2	488,348	18.50	-	-	-	-

Table 3 Comparison of different methods

The estimate of data 1 reserve is similar for all considered methods. However, in the case of data 2, there is a significant difference between the reserve determined by the state-space model and the reserve calculated by the chain-ladder, which is significantly lower. Even though the multivariate chain-ladder takes into account the correlation structure, the resulting reserves are almost identical to those calculated individually. The quantiles calculated by kernel estimation are considerably lower than those obtained by simulation smoother technique, and even the coefficients of variation are approximately half.

5 Conclusion

This paper deals with IBNR reserve estimation for correlated run-off triangles. Firstly, the multivariate log-normal model was introduced. Secondly, the bootstrap method and procedure how to estimate quantiles of the reserves was presented. This method is an alternative to that used in Hendrych and Cipra (2020) based on the simulation smoother proposed in Durbin and Koopman (2002). In the numerical part the real data were examined and the results were compared with the ones by other methods. It was shown that the reserve obtained by Mack's chain ladder method for the second run-off triangle was underestimated compared to the log-normal model. On the other hand, with higher number of negative values in the data, one should consider usage of a different model than the log-normal. The quantile estimation method assumed in this paper showed significantly lower coefficient of variation than the other methods which is a benefit of this approach. Another positive aspect of this method is its universal use.

Acknowledgements

This paper was supported by the grant 19-28231X provided by the Czech Science Foundation.

References

- [1] Atherino, R., Pizzinga, A. & Fernandes, C. (2010): A row-wise stacking of the runoff triangle: State space alternatives for IBNR reserve prediction. *ASTIN Bulletin*, **40**(2), 917–946.
- [2] Brockwell, P. J. & Davis, R. A. (1991): *Time Series: Theory and Methods* (Second Edition). Springer-Verlag, New York, ISBN 978-0-387-97429-3.
- [3] Durbin, J. & Koopman, S.J. (2002): A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, **89**(3), 603–615.
- [4] Helske, J. (2017): KFAS: Exponential family state space models in R. *Journal of Statistical Software*, **78**(10), 1–38.
- [5] Hendrych, R. & Cipra, T. (2020): Applying state space models to stochastic claims reserving. *ASTIN Bulletin*, **51**(1), 267–301.
- [6] Mack, T. (1993): Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, **23**(2), 213–225.
- [7] Shi, P., Basu, S. & Meyers G.G. (2012): A Bayesian log-normal model for multivariate loss reserving. *North American Actuarial Journal*, **16**(1), 29–51.
- [8] Silverman, B.W. (1986): *Density estimation for statistics and data analysis*. Chapman & Hall/CRC, London, ISBN 978-0-412-24620-3.
- [9] Zhang, Y. (2010): A general multivariate chain ladder model. *Insurance: Mathematics and Economics*, **46**, 588–599.

Forecasting third party insurance: A comparison between Random forest and Generalized Additive Model

Lukáš Veverka¹

Abstract. As the computing power and the amount of data grow, machine learning approaches tend to be more popular. It makes sense to compare them with the econometric approach and describe both advantages and disadvantages. This work will be focused on third party insurance which depends on lots of factors and non-linear dependencies are expected. Therefore we will use the Generalized Additive Model utilizing smoothing splines and compare it with Random forest. Since the computational demands to build and optimize a Random forest are high, parallelization is described and performed.

Keywords: GAM, Smoothing splines, Parallelization, Random Forest, Interpretable Machine Learning

JEL Classification: C530

AMS Classification: 62P20

1 Introduction

When estimating complex models, it is convenient to use more flexible approaches which can reflect the real relationships among variables. Two popular examples of these are Generalized Additive Models and Random Forests – both have their advantages and disadvantages. In this paper, they will be briefly described and since Random Forest is very computationally demanding, also a little part of computer science will be introduced with its possible application to Random Forest tuning (or any other model tuning). All these notes will be then applied to a case study of forecasting third-party insurance based on a quite complex dataset.

2 Literature review

Machine learning methods apply to the insurance industry are nowadays common and performs great on big data. In the automotive segment, the growing trend of insurance claims forces companies to focus on correct insurance height estimation. It was found out that Random Forest outperformed other models in its accuracy [1]. There are great advantages of non-parametric approaches compared to common methods for risk evaluation which are discussed in [2]. In addition, another study pointed out the non-stationary, non-linearity and non-normality of data in the automobile industry which also testify for machine learning models [3]. However, data mining and machine learning methods have some disadvantages – e.g. computational demands. Other challenges are mentioned in [4] where also the possibility of distributed computing is discussed. Possible improvement of Random forest estimation was suggested in [5] which describes the optimization of random forest based on task-parallel optimization. The paper summarizes all the knowledge and compares the accuracy and computation time of parallelized Random Forest and Generalized Additive Model.

3 Non-linearity in econometrics

3.1 GAM – Generalized Additive Model

Generalized Additive Model combines the additive structure of a linear model and non-linear functions of each of the variables which results in a flexible and interpretable model. Mathematically it is done by replacing $\beta_j x_{ij}$ in a multiple linear regression model with a non-linear function $f_j(x_{ij})$ which approximates the relationship between each feature and its response. The model can be written as

$$y_i = \beta_0 + \sum_{j=1}^p f_j(x_{ij}) + u_i. \quad (1)$$

¹ University of Economics, Prague, Department of Econometrics, Winston Churchill Square 4, CZ13067 Prague, Czech Republic, vevl00@vse.cz

The greatest advantage of this model is the ability to look at non-linear functions. Individual coefficients for the non-linear functions (typically splines) are not that interesting since there is almost no interpretation of them.

The main limitation of GAMs is that the interactions can be missed. It is possible to add them to the model by including another predictor $x_j \times x_k$ as it is possible in the linear model. It is also possible to add low-dimensional interaction functions $f_{jk}(x_j, x_k)$ – fitted by local regression or two-dimensional splines (two-dimensional smoothers).

3.2 Splines & Smoothing splines

Splines in general are an extension of polynomial regression and piecewise constant regression. They take both advantages from both attitudes – they are local (as in piecewise constant regression) and smooth (as in polynomial regression). This results in the ability to reflect unusual local behaviour without too many degrees of polynomial which means also a more robust solution.

It is important to determine the number of knots and where to place them. Usually, the knots are placed on suspicious breakpoints provided by experts in the examined area – this might be complicated as the knowledge is not always available. The regression function itself is defined as a piecewise polynomial function with the condition of continuity in every knot. Mathematically it can be written as

$$y_i = \beta_0 + \beta_1 b_1(x_i) + \beta_2 b_2(x_i) + \dots + \beta_{K+3} b_{K+3}(x_i) + u_i, \quad (2)$$

where K is the number of knots and b_k are basis functions. In the case of splines basis functions are polynomial and for example, a cubic spline (i.e. third degree of a polynomial) can be represented as

$$\begin{aligned} b_1(x_i) &= x_i \\ b_2(x_i) &= x_i^2 \\ b_3(x_i) &= x_i^3 \\ b_{k+3}(x_i) &= (x_i - \xi_k)_+^3, \quad \forall k. \end{aligned} \quad (3)$$

This can be extended to as many degrees of polynomial as the data allows. Typically just cubic splines are good enough and widely used. The way to represent any spline is to start with a basis for the polynomial then add one truncated power basis function per knot. In the sample equation (3) the truncated power basis function is $(x_i - \xi_k)_+^3$ where ξ_k are the values of the knots and the $()_+$ means a positive part (which makes it truncated) i.e.

$$(x_i - \xi_k)_+^3 = \begin{cases} (x_i - \xi_k)^3 & \text{if } x_i > \xi_k \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

This ensures the condition of continuity in knots. Moreover, the amount of continuity is one less than the degree of the polynomial (i.e. cubic polynomial (3rd degree) is continuous at each of the knots in first and second derivatives).

Since it is necessary to determine the number and position of the knots which requires some deeper knowledge of the examined area and it might not be easy to obtain, the smoothing splines were invented. The most significant difference is that they do not need knots to be determined. On the contrary, the knot is placed at each training point x_i .

The general problem of smoothing splines is defined as

$$\sum_{i=1}^n (y_i - g(x_i))^2 + \lambda \int g''(t)^2 dt \rightarrow \text{MIN}, \quad (5)$$

where the minimized function consists of two parts – RSS and penalty for complexity. The penalty is based on adding up (integral) all the variability in the function g . The square there is to get rid of the sign and the second derivative reflects how wiggly the function is (the amount by which the slope is changing). Lastly, λ serves as a roughness penalty - a tuning parameter. In the case of $\lambda = 0$, the function g would be allowed to interpolate, on the contrary, $\lambda = \infty$ would result in a linear function (linear regression) since the second derivative of a linear function is zero and the penalty will be zero (therefore irrelevant).

The solution to this problem is a natural spline with a knot at every unique value of x_i - that would consume all degrees of freedom. However, the effective degrees of freedom are much less because of the roughness penalty. The estimate of the effective degrees of freedom is:

$$df_\lambda = \sum_{i=1}^n \{S_\lambda\}_{ii}, \quad (6)$$

where S_λ is known as a smoother matrix and there is a formula in [6] to obtain it. The principle is that it is possible to rewrite the fitted values of smoothing spline as $\hat{g}_\lambda = S_\lambda \mathbf{y}$ where \hat{g}_λ are the fitted values of g which is the solution to (5) based on the training points x_1, \dots, x_n .

Once we can calculate effective degrees of freedom based on the roughness parameter λ , it is also possible to reverse it and calculate λ based on the required degrees of freedom. Another option is the well-known cross-validation and especially *leave-one-out* cross-validation error can be calculated very efficiently.

4 Random forest tuning

4.1 Tuning parameters

Before the parameters of Random Forest will be introduced and described, it is necessary to mention that the first important step in improving Random Forest results is feature engineering and feature reduction. Once this is done it makes sense to tune the parameters.

- Number of variables (features) randomly sampled to consider at each split. This parameter is one of the most important tuning parameters since it affects the correlation of trees inside the random forest and therefore the robustness of the model and its ability to predict properly. It is hard to choose a correct value and the best way to obtain it is using grid search and cross-validation. The typical default value is \sqrt{n} (n = number of variables) for classification and $\frac{n}{3}$ for regression.
- Number of trees to build. More trees inside the random forest result in a more robust aggregate model with less variance. However, the training time increases alongside the number of trees. On the other hand, the number of trees should not be small because of the possibility of not including an important variable with high predictive power due to the randomness of selecting variables.
- Minimum size of terminal nodes. When this number is set large, it leads to smaller trees (and thus less time taken) since the nodes stop to split earlier because they fulfil the condition of terminal node size.
- The maximum depth of each tree affects the possible number of variable interactions. The deeper the tree is, the more splits it has and therefore it takes more possible combinations into account. Very deep trees on the other hand can cause overfitting.
- The size of the bootstrapped dataset to train each tree with. Since it is bootstrapped with replacement, even though the size of the bootstrapped dataset may be the same as the whole training set, both datasets will differ. Common practice is to leave this parameter at the size of the training sample.

There are even more parameters to tune however, these are the most common. The optimal combination of these parameters, so that the Random Forest performs at best in the case of forecasting, can be chosen by cross-validation. It is known as a grid search (of parameters) which is a method that, instead of sampling randomly parameters, evaluates every combination of settings we specified [7]. This however requires a huge computational power. The creation of a random forest itself is already demanding because of the mechanics behind (computing the information gain at each node for each split point) and when it is multiplied by all the combinations of parameters it starts to consume an unbearable amount of time. Therefore it is convenient to utilize parallelization.

4.2 Parallelization

Computations in either R or Python are designed to process data serially. Some computations can be made faster when they are run in parallel. Generally, parallelization is a concept of independently executing smaller pieces of a larger computation simultaneously by discrete computing resources. If the computation is optimized for parallelization it can achieve the results much faster than programs executing processes in serial. Since processors are approaching their fastest possible speeds, parallelization is a way how to improve computing performance. There are two different ways of parallelization:

- **Multi-core computing** began as a link of computers to create a large cluster of CPUs. Nowadays, multiple cores on a single machine are common and that is also a little cluster of computing power. However, the most popular data science programming languages are designed for serial computations – therefore parallelization

can help and make the computations faster. In order to utilize more cores (or whole CPUs for example in the case of parallel computing via the network-of-workstations), it is necessary to break down a demanding task into a set of functions, with one function going to each core (or CPU). A typical example is a for loop – each run can be distributed to different cores. It is necessary to utilize sockets if the processes need to communicate among each other (typically passing back and forth the data and results between parent and child processes). For example, in Python, the well-known package Scikit-learn offers parallelization of a Random Forest without any other coding required.

- **Multi-threading** is an ability of a CPU to run more threads of execution simultaneously. By way of contrast to multi-core computing, in multi-threading, the threads share the same CPU and cache. It utilizes the time of reading or writing operations (i.e. I/O operations) when the CPU has to wait – instead of waiting, the CPU switches to another thread and executes it. A lot of packages in both R and Python are developed to utilize multi-threading (e.g. data.table in R which includes lots of I/O operations).

Parallelization is, however, not always suitable for reducing the computation time. There is an overhead that reduces the efficiency – e.g. the data need to be copied to each CPU, threads need to be created by an operating system. Therefore the performance improvement is less than theoretical. It depends on the ratio of the computation and preparation time. This may be approximated by Amdahl’s Law [8], which describes the dependency of the speedup of the computation on both the number of cores and the proportion of the computation that can be parallelized. Tasks very suitable for parallelization are repetitive tasks since they can be broken down into completely independent runs and executed in parallel, each on its own core.

In section 4.1, it was suggested to optimize the tuning parameters with the grid search cross-validation – the repetitive task of estimating the computationally demanding random forest with a certain set of parameters. This is a great option to parallelize on multiple cores and reduce the computation time significantly.

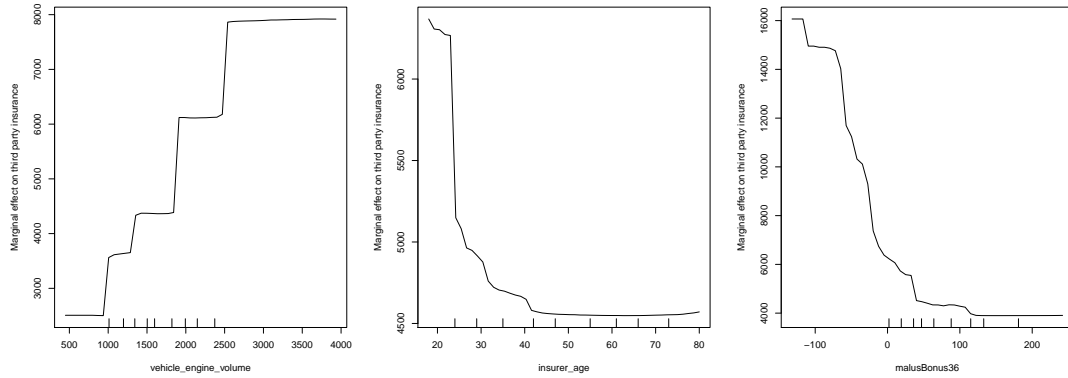
Finally, in the case of really big datasets which cannot be processed on a single computer, the cloud computing options come in handy. For example, Google Cloud or Czech product Keboola offers a cloud platform for data processing where it is possible to buy the required computer performance.

5 Case study

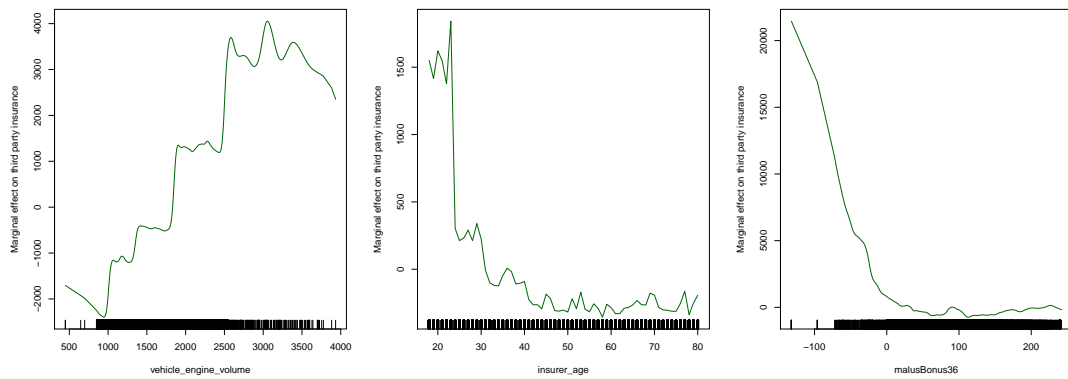
The main intention of this work is to figure out which factors and how they affect the height of third party insurance for 7 different companies, then to compare the accuracy and computational demands of each model. Even though the dataset is not huge (approx. 25 000 rows and 27 columns), it already takes a significant amount of time. The data are split into training, testing and validation samples, of which the training sample has 14 000 observations. This information is necessary to mention to have a benchmark for the resulting time necessary to search through the grid. The grid in this research contains 32 combinations of tuning parameters. It takes more than 2.5 hours to go through the grid and create the random forest with set parameters. When the grid search is parallelized on 3 cores, it results in a significant reduction of 2.6x of the computation time which means approx. 1 hour per optimized Random Forest. This optimization is run for each company since the mechanics of determining the insurance might differ for every company. The note on the inefficiency of adding cores/CPU’s which was described in 4.2, applies also here – theoretical time reduction on 3 cores is 3x less time, however, the reality was only 2.6x due to the overhead of importing data and packages to each socket.

Table 1: Comparison of pseudo R-squared of Random Forest and GAM on the training and testing samples

pseudo R^2	Random Forest		GAM	
	Train	Test	Train	Test
1	0.985	0.979	0.887	0.866
2	0.926	0.924	0.806	0.795
3	0.881	0.875	0.808	0.777
4	0.808	0.825	0.777	0.765
5	0.812	0.814	0.834	0.811
6	0.956	0.952	0.907	0.882
7	0.888	0.890	0.800	0.776



(a) Random Forest model – Partial Dependence Plot (PDP)



(b) GAM model – visualization of smoothing splines

Figure 1: Comparison of the non-linear dependencies between the target variable and top 3 most significant predictors (based on variable importance in Random Forest).

source: author

The importance of parallelization of specific tasks is indisputable as the time difference is significant. However, machine learning methods, in general, have one great disadvantage – difficult (or impossible) interpretation. There is an option of partial dependence plots which show the relationship between a dependent and an independent variable. On contrary, GAMs using smoothing splines estimates a non-linear mathematical function which is easy to understand, visualize and explain. The comparison of the non-linearity estimation by a tuned Random Forest and GAM is shown in Figure 1. It is possible to see that the estimated functions are quite similar which proves the ability of GAM to approximate a model in a completely unknown area very precisely. Even the results on training and testing samples in Table 1 suggest that the robustness and accuracy – measured in pseudo R^2 which is calculated as $cor(y_i, \hat{y}_i)^2$ – of both models is satisfactory. However, there is a significant difference in the computational demands of both approaches.

Table 2: Comparison of average accuracy and time necessary to compute the model of Random Forest and GAM

	Random Forest	GAM
Time (minutes)	4.91	0.13
Accuracy (pseudo R^2 on the testing sample)	0.894	0.810

The random forest has a slightly better accuracy at a cost of almost 38 times longer time to estimate the model. When is it worth it? For a brief dive in the area and a quick look at the relations in the model, GAM is good enough, however, Random Forest with its ability to estimate even interactions performs better, therefore once we are sure that the model is correctly set, it is sensible to estimate and tune the Random Forest.

6 Conclusion

In the case of models which has to be quickly recalculated (e.g. part of a trading algorithm in case of some dramatic changes in a market) GAM serves better because of a quite similar accuracy as Random Forests while preserving quick estimation. In case of accuracy is the most important (e.g. credit risk modelling) and estimation time does not matter that much, Random Forests are better.

As for the estimation of the third party insurance, it was shown that a tuned Random Forest performs better at the cost of longer computation time. Variables with the highest impact on the insurance height are volume of the engine, insurer age and bonus-malus (system of penalization or discount based on previous insurance claims). It might be a good idea to enrich the model with a spatial approach as it was done in this research about non-life insurance [9].

Acknowledgements

The work on this paper was supported by the Internal Grant Agency of the Prague University of Economics and Business under project F4/27/2020.

References

- [1] Hanafy, M., & Ming, R. (2021). Machine Learning Approaches for Auto Insurance Big Data. *Risks*, 9(2), 42. <https://doi.org/10.3390/risks9020042>
- [2] Kaščelan, V., Kaščelan, L., & Burić, M. N. (2016). A nonparametric data mining approach for risk prediction in car insurance: A case study from the Montenegrin market. *Economic Research-Ekonomska Istraživanja*, 29(1), 545–558. <https://doi.org/10.1080/1331677X.2016.1175729>
- [3] Shalini, V., Krishnamurthy, S., & Narasimhan, S. (2017). Predictive Analytics in Automobile Industry: A Comparison between Artificial Intelligence and Econometrics. *WCX 17: SAE World Congress Experience*. <https://doi.org/10.4271/2017-01-0238>
- [4] Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, 237, 350–361. <https://doi.org/10.1016/j.neucom.2017.01.026>
- [5] Chen, J., Li, K., Tang, Z., Bilal, K., Yu, S., Weng, C., & Li, K. (2017). A Parallel Random Forest Algorithm for Big Data in a Spark Cloud Computing Environment. *IEEE Transactions on Parallel and Distributed Systems*, 28(4), 919–933. <https://doi.org/10.1109/TPDS.2016.2603511>
- [6] Hastie, T., Tibshirani, R., & Friedman, J. (2009). Basis Expansions and Regularization. In T. Hastie, R. Tibshirani, & J. Friedman (Eds.), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (pp. 139–189). Springer.
- [7] Campos, R., Canuto, S., Salles, T., de Sá, C. C. A., & Gonçalves, M. A. (2017). Stacking Bagged and Boosted Forests for Effective Automated Classification. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 105–114. <https://doi.org/10.1145/3077136.3080815>
- [8] Almasi, G. S., & A. Gottlieb. Highly parallel computing. *Benjamin-Cummings Publishing Co., Inc.*, 1989.
- [9] Gschlößl, S., & Czado, C. (2007). Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal*, 2007(3), 202–225. <https://doi.org/10.1080/03461230701414764>

Efficiency Verifications of Classical Portfolio Optimization Models

Anlan Wang

Abstract. Starting from the mean-variance model proposed by Markowitz in 1952, the approaches of determining the optimal portfolio compositions have been well developed. In the portfolio investments, most investors believe in that the portfolios selected by applying optimization models would have better performance. However, in the research of DeMiguel et al. in 2009, evaluations are made on 14 portfolio optimization models and according to the empirical results, it shows that none of the models is consistently better than the 1/N naive portfolio in terms of applied performance measures. So, in this paper, we make the efficiency verifications on the historical risk and performance of portfolios obtained by applying the classical portfolio optimization models. From the empirical results of our research, by applying the historical trading data of Dow Jones Industrial Average, we find that the models aiming at risk-minimizing are efficient, while the models aiming at performance-maximizing are not.

Keywords: hypothesis testing, performance ratios, portfolio optimization, random-weights, risk measures

JEL Classification: G11, G17

AMS Classification: 46N10, 91G10

1 Introduction

Starting from the introduction of the *mean-variance* (MV) model by Harry M. Markowitz in 1952, the approaches which aim at determining the optimal portfolio compositions have been well developed. In Markowitz [8], the expected return and risk of the portfolio is measured as the mean and variance of the historical returns, thus, there are two objectives of the portfolio selection problem, which are the maximization of mean return and the minimization of variance, see Martínez-Nieto et al. [10]. Moreover, in the following study of Markowitz, few valuable topics such as the efficient frontier and the expected utility are added, see Markowitz [9]. The MV model set the foundation of portfolio optimization by providing a straightforward way to understand the risk and return, while there are some limitations of its applications in later studies by practitioners, for example, sensitivity to input errors. In the past few decades, various models have been proposed to address the effects of estimation errors, such as the Bayes-Stein shrinkage portfolio, which integrated with the idea of shrinkage estimation pioneered by Stein [13] and James and Stein [13]. However, in DeMiguel et al. [2], evaluations are made on the out-of-sample performance of MV model and its extensions designed to reduce estimation errors, and by applying 14 models across 7 empirical datasets in their study, the empirical results indicate that none of the applied models is consistently better than the 1/N naive portfolio in terms of applied performance measures. Thus, we raise a question in this paper, that is whether the portfolios generated from the optimization models provide an advantage compared to the ones selected by people who lack of professional investment knowledge.

To address this question, the goal of this paper is to make the efficiency verifications on the historical risk and performance of portfolios obtained by applying the classical portfolio optimization models. More specifically, in our research, we apply the models which are objective to risk-minimizing and performance-maximizing under the MV framework. In the meanwhile, to compare the performance of classical portfolio optimization models with that of the portfolios selected without professional managements, the 1/N naive portfolio and the random-weights portfolios are applied as the benchmarks.

The structure of this paper is constructed as follows. In section 1, the background of the research and the goal of this paper are introduced. In section 2, we describe the risk measures and performance ratios which are applied in finding the optimal portfolios in our research. Next, we describe the frameworks of the applied classical portfolio optimization models as well as the methods of efficiency verifications in section 3. The empirical analysis is made in section 4. In the last section, we make conclusions of the paper.

2 Risk Measures and Performance Ratios

The objectives in the portfolio selection problems are different across different categories of investors, thus, there is no unique definition of risk measure or performance measure. The risk measures and performance measures are investors-specific, for example, one of the most classical pairs of risk and performance is the variance and mean of portfolio returns, which indicate the volatility and average of the future benefits, respectively. While more than the mean and variance, in this section several types of the risk and performance measures are described.

2.1 Risk Measures

According to Artzner et al. [1], “A risk measure satisfying the four axioms of translation invariance, subadditivity, positive homogeneity, and monotonicity is called coherent.” Moreover, based on the properties of risk measures, in Rachev et al. [11], a broad categories of risk measures are introduced. In this paper, we apply 5 widely used risk measures in our research.

Standard Deviation

The standard deviation is used to measure the amount by which returns deviate from the mean of portfolio return (or from the expected portfolio return), and to calculate the standard deviation of portfolio, we first estimate the returns of individual assets. The most common method to estimate the individual assets returns is the historical sample estimation, based on this method, the extensions such as the Bayes-Stein shrinkage estimation (henceforth BS estimation) and fuzzy estimation can also be applied.

- Historical Sample Estimation

In the historical sample estimation method, the standard deviation of returns of asset i can be calculated as follows,

$$\sigma_i = \sqrt{\frac{1}{M-1} \sum_{m=1}^M (R_{i,m} - E(R_i))^2} \quad (1)$$

where R_i is the historical returns of asset i and M is the number of historical observations.

- BS Estimation

According to Mansour et al. [7], most of the historical data in the financial market are not accurate enough to predict the future condition of the market due to the high volatility of market environments. So, there might be estimation errors in the probability distribution of the estimates which are only based on the historical observations. To reduce the estimation errors, the BS estimation takes the subjective (a priori) assumption of the shape of the asset’s returns distribution into account, and the resulting (a posteriori) assumption of the shape of the probability distribution is then a combination of the priori assumption and the probability distribution of the observed sample. In this paper, we apply the shrinkage estimation method suggested by Jorion [4], in which the shrinkage factor ξ is introduced as follows,

$$\xi = \frac{N+2}{(N+2)+M \cdot (\overline{E(R_i)} - \overline{E(R_i)})^T \cdot \hat{Q}^{-1} \cdot (\overline{E(R_i)} - \overline{E(R_i)})} \quad (2)$$

where N is the number of assets. From equation (2), we can see that ξ depends on the dispersion of historical estimated expected returns from a priori assumption (the greater the dispersion, the greater the estimation error).

- Fuzzy Estimation

Tanaka et al. [15] proposed a fuzzy probability model which helps to handle the uncertainty of the individual assets returns distribution by involving the fuzzy theory. In the proposed fuzzy probability model, the possibility grade is introduced to reflect the similarity degree between the future state of the asset return and the same asset’s m -th historical return, for the specific fuzzy estimation methods, see Kresta and Wang [6].

After we get the estimates of individual assets returns, then we can calculate the portfolio standard deviation according to equation (3),

$$\sigma(R_p) = \sqrt{E(R_p - E(R_p))^2} \quad (3)$$

where R_p is the portfolio returns and $E(R_p)$ is the expected return of portfolio.

Mean Absolute Moment (henceforth MAM) & Mean Absolute Deviation (henceforth MAD)

According to Rachev and Mittnik [12], asymptotic mean dispersion is a special variant of the classical mean absolute deviation, it assumes that the returns are jointly α -stable sub-Gaussian distributed with $\alpha > 1$. Moreover, the MAM is the logical generalization of the asymptotic mean dispersion, it can be calculated as follows,

$$MAM(R_p, q) = (E(|R_p - E(R_p)|)^q)^{1/q} \quad (4)$$

where q is the order of the absolute moment and it satisfies $q \geq 1$. When $q = 2$, the MAM is just equal to the standard deviation of portfolio return, and when $q = 1$, the MAM is just equal to the MAD.

Conditional Value at Risk (henceforth CVaR)

Value at Risk (henceforth VaR) is defined as the worst-case loss associated with a probability and a time horizon, while CVaR indicates the expected loss under the condition of exceeding VaR, which can be calculated as follows,

$$CVaR_\alpha(R_p) = E(-R_p | -R_p \geq VaR_\alpha(R_p)) \quad (5)$$

where α is the probability level which is usually set to 1% or 5%.

Lower Partial Moment (henceforth LPM)

The LPM is a downside risk measure that depends on a target rate of return (e.g. the minimum acceptable return) and the power index which is used to indicate the investor's degree of risk aversion, the LPM of portfolio return can be calculated as follows,

$$LPM(R_p, u) = \left\{ E \left[(MAR - R_p)_+^u \right] \right\}^{1/u} \quad (6)$$

where $u \geq 1$ is the order of the lower partial moment (power index) and the function $(v)_+^u = (\max(v, 0))^u$. MAR is a minimum acceptable return, which is set by investors and according to Kouaissah and Hocine [5], it may be represented by a risk-free rate.

2.2 Performance Ratios

According to Rachev et al. [13], in the ex-post analysis, the performance measures are applied to evaluate the performance of a portfolio over historical time series, and in the ex-ante analysis, the performance measures are applied to express the investors' objectives in the portfolio optimization procedures. In this paper, we apply 5 performance measures in our research.

Expected Return of Portfolio

Similarly to the calculation of portfolio standard deviation, the method of calculating the portfolio expected return also varies from the estimation methods of individual assets returns. By applying the three methods introduced in section 2.1, the expected return of an individual asset can be calculated as in equation (7), (8) and (9) differently,

$$E(R_i) = \frac{1}{M} \sum_{m=1}^M R_{i,m} \quad (7)$$

$$E(R_i^{BS}) = (1 - \xi) \cdot \overline{E(R_i)} + \xi \cdot \overline{E(R_i)} \quad (8)$$

$$E(R_i^F) = \frac{\sum_{m=1}^M h_m \cdot R_{i,m}}{\sum_{m=1}^M h_m} \quad (9)$$

in equation (9), h_m is the possibility grade introduced in the fuzzy probability model.

Then, the expected return of a portfolio can be calculated as the weighted average of the components assets expected returns, which can be shown as follows,

$$E(R_p) = w^T \cdot E(R) = \sum_{i=1}^N w_i \cdot E(R_i) \quad (10)$$

where w is the vector of weights of component assets of the portfolio.

Jensen's Alpha (henceforth JA)

JA measures the excess return of a portfolio over the theoretical expected return predicted by Capital Asset Pricing Model. The calculation of JA can be shown as follows,

$$\alpha_j = R_p - [R_f + \beta_p \cdot (R_m - R_f)] \quad (11)$$

where $E(R_m)$ is the expected return of the market, $E(R_m) - R_f$ is the market risk premium and β_p is the systematic risk. Parameter β_p is a portfolio's beta that measures the tendency of a portfolio's return to change in response to changes in the return of the overall market.

Sharpe Ratio (henceforth SR) & Treynor Ratio (henceforth TR)

Taking into account the variability of portfolio returns, the SR and TR are both reward-to-variability ratios used to measure the adjusted return of a portfolio, the only difference between SR and TR is the risk measure.

$$Sharpe(R_p) = \frac{E(R_p - R_f)}{\sigma(R_p)} \quad (12)$$

$$Treynor(R_p) = \frac{E(R_p - R_f)}{\beta_p} \quad (13)$$

Sortino-Satchell Ratio (henceforth SSR)

The SSR can be defined as in equation (14), in which u is the order of the lower partial moment and we set $u = 2$ in our empirical analysis.

$$SSR(R_p, u) = \frac{E(R_p - MAR)}{\{E[(MAR - R_p)_+^u]\}^{1/u}} \quad (14)$$

3 Portfolio Optimization Models & Efficiency Verification Methods

Based on the risk measures and performance measures introduced in section 2, we find optimal portfolio strategies with the objectives of risk-minimization or performance-maximization. For finding the minimum-risk strategies, the constraints of the optimization model can be set as follows,

$$\begin{aligned} & \min RM_S \\ & \text{s.t } \sum_{i=1}^N w_i = 1 \\ & w_i \geq 0, i = 1, \dots, N. \end{aligned} \quad (15)$$

where RM_S is the risk measurement of strategy portfolio and w_i is the weight of component asset i . While for finding the maximum-performance strategies, we apply an indirect optimization procedure in the model, which can be shown as follows,

$$\begin{aligned} & \max (k - 1) \cdot PM_S - k \cdot RM_S \\ & \text{s.t } \sum_{i=1}^N w_i = 1 \\ & w_i \geq 0, i = 1, \dots, N. \end{aligned} \quad (16)$$

where PM_S is the performance measurement of strategy portfolio and the parameter $k \in [0,1]$ represents that the investor is risk-averse. When $k = 1$ we have an absolute risk-averse investor seeking a minimum-risk portfolio, when $k = 0$ we have a risk-neutral investor seeking the maximum-performance portfolio.

After we obtain the optimal weights of strategy portfolios by applying the optimization models, then we apply the obtained weights to test the out-of-sample performance. The efficiency verifications on the historical risk and performance of the obtained portfolios are made through the statistical testing by applying the benchmarks. The benchmarks applied in our research is the naive portfolio and random-weights portfolios. The naive portfolio is the portfolio has equal weight $1/N$ to each asset in the portfolio. As for the random-weights portfolio, it is the portfolio in which the component assets are invested with random weights. For generating the random-weights, see the method introduced in Wang and Kresta [16].

In the efficiency verifications, we apply the hypothesis testing. Since the strategy portfolios are more likely to be demonstrated as having lower risk than the benchmark portfolios, so, we make the null hypothesis and alternative hypothesis as follows,

$$\begin{aligned} \text{null hypothesis—}H_0: RM_s &= E(RM_r), \\ \text{alternative hypothesis—}H_A: RM_s &< E(RM_r) \end{aligned} \quad (17)$$

where $E(RM_r)$ is the expected value of the risk measurements of random-weights portfolio. The significance level of the tests is set to 5% in our empirical analysis and the p-value is the proportion of the random-weights portfolios which meets the condition $E(RM_r) < RM_s$. For the hypothesis testing on the basis of performance measurements, the null hypothesis and alternative hypothesis are stated as follows,

$$\begin{aligned} \text{null hypothesis—}H_0: PM_s &= E(PM_r), \\ \text{alternative hypothesis—}H_A: PM_s &> E(PM_r) \end{aligned} \quad (18)$$

in this case the p-value is the proportion of the random-weights portfolios which meets the condition $E(PM_r) > PM_s$.

4 Empirical Analysis

In this section we make the empirical analysis. The data input in the analysis is the historical daily adjusted-closing prices of 27 components of Dow Jones Industrial Average (henceforth DJI) starting from April 29th, 2011 to April 30th, 2021. The missing three components are the stocks of *General Motors Corporation*, *Hewlett-Packard Company* and *United Technologies Corporation* due to the incomplete data in the selected period. To make the efficiency verifications on the historical risk and performance of the strategy portfolios, we divide the whole dataset into in-sample and out-of-sample. In the in-sample period (4.29.2011 to 4.29.2016) we find the optimal weights of strategy portfolios and in the out-of-sample period (5.2.2016 to 4.30.2021) we measure the strategies' risk and performance.

In the in-sample period, we find 9 strategy portfolios, which are aiming at minimizing risk or maximizing performance. The naive portfolio is generated as the benchmark. In Figure 1, there shows the proportions of component stocks in each strategy portfolio, we can see the minimum-risk portfolios are more diversified than the maximum-performance portfolios, especially in the two maximum-E(Rp) portfolios (estimated by the historical sample method and the fuzzy method), there is merely one component stock invested in the portfolio. In the out-of-sample period, we test the efficiency of the strategy portfolios, the values of risk and performance measurements as well as the corresponding testing results of p-values are shown in Table 1.

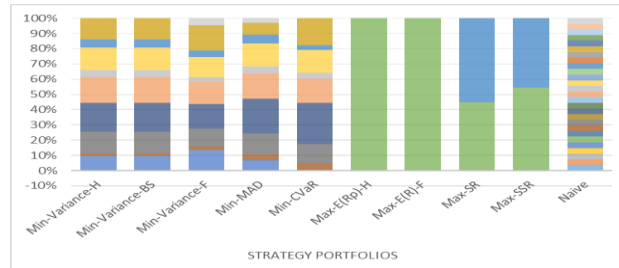


Figure 1 In-sample proportions of component stocks of strategy portfolios

	Min-Variance-H	Min-Variance-BS	Min-Variance-F	Min-MAD	Min-CVaR	Max-E(Rp)-H	Max-E(Rp)-F	Max-SR	Max-SSR	Naive
σ_p	0.99% ****	0.99% ****	0.99% ****	0.99% ****	1.01% ****	1.60% ns	1.60% ns	1.32% ns	1.32% ns	1.28% ns
MAD	0.59% ****	0.59% ****	0.59% ****	0.59% ****	0.60% ****	1.00% ns	1.00% ns	0.84% ns	0.83% ns	0.74% ns
MAM	2.68% ****	2.68% ****	2.69% ****	2.72% ****	2.83% ****	5.04% ****	5.04% ****	3.38% ****	3.60% ****	3.60% ****
cVaR	0.06% ****	0.06% ****	0.07% ****	0.06% ****	0.07% ****	0.09% ns	0.09% ns	0.10% ns	0.09% ns	0.09% ns
LPM	1.03% ns	1.03% ns	1.04% ns	1.03% ns	1.05% ns	1.71% ns	1.71% ns	1.44% ns	1.45% ns	1.36% ns
E(Rp)	0.05% ns	0.05% ns	0.05% ns	0.05% ns	0.05% ns	0.09% ***	0.09% ***	0.05% ns	0.06% ns	0.06% ns
JA	0.05% ns	0.05% ns	0.05% ns	0.05% ns	0.05% ns	0.09% ***	0.09% ***	0.05% ns	0.06% ns	0.06% ns
SR	5.04% ns	5.04% ns	4.79% ns	5.32% ns	5.10% ns	5.69% *	5.69% *	3.65% ns	4.20% ns	4.55% ns
TR	0.08% *	0.08% *	0.07% *	0.08% **	0.08% *	0.09% ****	0.09% ****	0.06% ns	0.07% ns	0.06% ns
SSR	4.84% ns	4.84% ns	4.57% ns	5.09% *	4.90% ns	5.34% *	5.34% *	3.34% ns	3.83% ns	4.29% ns

Table 1 Out-of-sample risk and performance measurements of strategy portfolios

Notes: ns (p-value>0.05), * (p-value<0.05), ** (p-value<0.01), *** (p-value<0.001), **** (p-value<0.0001).

5 Conclusion

In this paper, based on the pioneered research on performance testing of portfolio optimization models, we raise a question that whether the portfolios generated from the classical optimization models provide an advantage compared to the ones selected by people who lack of professional investment knowledge. To address this question, the goal of this paper is to make the efficiency verifications on the historical risk and performance of portfolios obtained from the portfolio optimization models which aim at risk-minimizing and performance-maximizing.

In our research, to compare the performance of strategies with that of the portfolios selected without professional managements, the 1/N naive portfolio and the random-weights portfolios are applied as the benchmarks. From the out-of-sample testing results, we make the following conclusions. Firstly, contrary to the finding in DeMiguel [2], the naive portfolio does not perform better than the strategies generated from the applied optimization models. Secondly, considering the random-weights portfolios as the benchmarks, we find that the values of risk measurements (except for LPM) in the obtained minimum-risk portfolios are significantly low, which means that the optimization models aiming at risk-minimizing are efficient. However, for the models aiming at performance-maximizing, the testing results indicate that only the maximum-E(Rp) portfolios (estimated by the historical sample method and the fuzzy method) are efficient considering the applied performance measures, while the other maximum-performance portfolios are not.

Acknowledgements

The author was supported through the Czech Science Foundation (GACR) under paper 18-13951S and moreover by SP2021/15, an SGS research paper of VSB-TU Ostrava. The support is greatly acknowledged.

References

- [1] Artzner, P., Delbaen, F., Eber, J. M. & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203–228.
- [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(5), 1915–1953.
- [3] James, W. & Stein, C. (1961). Estimation with Quadratic Loss. *Proceedings of the 4th Berkeley Symposium on Probability and Statistics I*. Berkeley, CA: University of California Press.
- [4] Jorion, P. (1986). Bayes-Stein Estimation for Portfolio Analysis. *Journal of Financial and Quantitative Analysis*, 21(3), 279–292.
- [5] Kouaissah, N. & Hocine, A. (2020). Forecasting systemic risk in portfolio selection: The role of technical trading rules. *Journal of Forecasting*, 10.1002/for. 2741.
- [6] Kresta A. & Wang A. (2020). Portfolio Optimization Efficiency Test Considering Data Snooping Bias. *Business Systems Research*, 11(2), 73–85.
- [7] Mansour, N., Cherif, M. S. & Abdelfattah, W. (2019). Multi-objective imprecise programming for financial portfolio selection with fuzzy returns. *Expert Systems with Applications*, 138.
- [8] Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91.
- [9] Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press.
- [10] Martínez-Nieto, L., Fernández-Navarro, F., Carbonero-Ruz, M. & Montero-Romero, T. (2021). An experimental study on diversification in portfolio optimization. *Expert Systems With Applications*, 181(1).
- [11] Rachev, S., Menn, C. & Fabozzi, F. (2005). *Fat-tailed and skewed asset return distributions*. New York: Wiley.
- [12] Rachev, S. & Mitnik, S. (2000). *Stable Paretian Model in Finance*. Chichester: Wiley.
- [13] Rachev, S., Stoyanov, S. & Fabozzi, F. (2008). *Advanced stochastic models, risk assessment, and portfolio optimization: The ideal risk, uncertainty, and performance measures*. Hoboken, N.J.: Wiley.
- [14] Stein, C. (1955). Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. *In 3rd Berkeley Symposium on Probability and Statistics I* (pp.197–206). Berkeley, CA: University of California Press.
- [15] Tanaka, H., Guo, P. & Türksen, B. (2000). Portfolio selection based on fuzzy probabilities and possibility distributions. *Fuzzy Sets and Systems*, 111(3), 387–397.
- [16] Wang A. & Kresta A. (2020). Statistical Testing of Portfolio Performance Ratios. In M. Čulík (Eds), *Proceedings of the 10th International Scientific Conference on Managing and Modelling of Financial Risks* (pp. 224–231). Ostrava: VSB-Technical University of Ostrava.

Scoring Applications in Early Collection

Jiří Witzany¹, Anastasiia Kozina²

Abstract. The goal of this contribution is to propose, empirically test and compare different scoring techniques in order to optimize the debt collection process. This process uses various actions, such as phone calls, mails, visits, or legal steps to recover past due loans. We focus on the soft collection part, where the question is whether and when to call a past-due debtor with regards to the expected financial return of such an action. We propose to use the survival analysis technique, in which the phone call can be compared to a medical treatment, and repayment to the recovery of a patient. We show on a real banking dataset that, unlike ordinary logistic regression, this model provides the expected results and can be efficiently used to optimize the soft collection process.

Keywords: credit risk modeling; survival analysis; scoring; debt recovery

JEL Classification: G21, G28, C14

AMS Classification: 90C15

1 Introduction

The aim of this study is to apply and compare logistic and survival regressions developing a system that streamlines the process of debt recovery and creates more efficient soft collection strategies through telephone communication with the clients.

There is relatively limited research on the subject of the optimization of the recovery process. One of the first papers by De Almeida Filho et al. (2010) proposes building a dynamic programming model to optimize collections. The dynamical programming approach has been followed by several other papers (e.g. van de Geer et al., 2018, or So et al., 2019 using Bayesian dynamic programming). Chehrazi et al. (2015, 2019) model repayments as a self-exciting stochastic process and propose using stochastic optimization approaches. He et al. (2015) and Liu et al. (2019) model state transitions of loan accounts using Markov transitions matrices and determine the optimal action conditional on the state and time. Surprisingly, there are not many (academic) applications of classical regression methods such as logistic regression or survival analysis in the recovery process optimization. Murgia and Sbrilli (2012) test logistic regression against other methods in a collection decision support system. Thomas et al. (2016) develop a Markov chain model and a hazard rate model in order to study the payment and nonpayment state transitions and impacts of different write-off strategies. Besides loan lending, Thomas et al. (2017) mention other fields where scorecard techniques can be used, such as prescreening, preapproval, fraud prevention, and also debt recovery.

Generally, there are two phases of the debt collection process: early collection and late collection. Different methods of recovery can be used in each of these two phases, such as phone communication, sending SMS and email messages, and even handing over the debt portfolio management to an external firm. Due to the still relatively high probability of repayments and the lower accrued penalty interest, the early collection phase is the more appropriate phase to apply statistical methods. The key problem in the early collection modeling we are going to focus on can be formulated as a classical binary classification problem: Is the marginal effect of calling a past-due debtor (or another soft collection action) on the probability of repayment positive or negative? Is the administrative and personal cost of the call offset by the increased recovery return? We propose to apply the survival analysis approach to solve the question of timing, and, at the same time, to handle the data when the calls historically took place at different times and the outcome of the recovery is often unknown (censored).

Survival analysis is a common statistical method from the medical and healthcare sectors (see Marubini & Valsecchi, 1995 or Collet, 2003), but has also found widespread use in credit risk (Witzany et al., 2012 or Witzany, 2017). One of the first uses of survival analysis in banking can be attributed to Narain (1992). Recently, a significant amount of research has been conducted using survival analysis in the area of credit risk modeling in banking

¹ Corresponding author: Prague University of Business and Economics, Faculty of Finance and Accounting, W. Churchill Sq. 4, 130 67, Prague, Czech Republic, +420 224 095 174, e-mail: jiri.witzany@vse.cz.

² Prague University of Business and Economics, Faculty of Finance and Accounting.

(Stepanova & Thomas, 2002, Cao et al., 2009, Hosmer et al., 2008, Bellotti & Crook, 2010, 2013, Thomas et al., 2016, or Djeundje & Crook, 2018, 2019).

The main contribution of our paper lies in the survival analysis set up allowing to estimate effects of time varying collection actions such as making a call, sending a letter, etc. The outlined models will be estimated on a real banking dataset.

The remainder of the paper is divided into the following parts: Section 2 describes the dataset, its preprocessing, and also discusses the logistic regression results; Section 3 examines the survival analysis methodology; Section 4 reports and discusses the survival models' results; and finally, Section 5 provides conclusions.

2 Data Description and Preprocessing

This analysis is based a dataset provided by a Czech bank for the recovery process period 10/2017 to 11/2019 (Kozina, 2020). As a real banking dataset, it is confidential and it has been provided only for the purpose of this research. The input data set contains 42,382 observations with more than 30 variables. The data represent historical records of past due retail exposures and debtors, including personal information, the characteristics of the products with which the debtors have become past due and entered the recovery process during the observation window, and also whether and when telephone communication took place. The dataset also contains the information whether and when the recovery process was successful, i.e. whether full recovery (repayment of all past due amounts) took place or not (exit to late collection). However, for some exposures, the outcome is not known due to the limited observation window, i.e. some of the observations are censored.

There are 13,003 observations for which the call took place. The decisions to call were based on a relatively simple set of rules and their timing often depended on call center capacity. The calls were realized mostly in a few days after the early collection process entry, in majority of cases within 30 days after the entry (9,573 observations, i.e. 74% of the total number of calls), but in more than 15% cases the number of days was larger than 60.

The standard logistic regression (Witzany, 2017) can be set up with the binary target variable “repaid” and with a selection of the explanatory variables including the binary variable *tel* indicating whether the collection call was made or not. Note that in this case we are losing the information on timing of the phone call and of the repayment (if it takes place). The phone call does not always take place immediately after the debtor enters the soft collection process and the repayment indicator (*repaid*) is based on the information whether the full repayment of past due amounts took place during the soft collection phase which may take up to 762 days (in the dataset). We will show that it is more appropriate to use the survival analysis modeling the (full) repayment time with the time dependent explanatory variable *tel*, while all the other variables remain constant. We believe that this approach utilizing the available timing information is also more robust, flexible, and parsimonious than an approach analyzing the partial repayment patterns such as in Tomas et al. (2016). The estimated survival function allows us to optimize in a relatively straightforward way the timing of soft collection actions such as making a call which is the main goal of the modeling exercise.

In order to apply any of the techniques, it is important to preprocess the data, which includes, among other things, selection of only the statistically significant variables and elimination of those with low informative value. This can be achieved following the standard logistic scoring function development process (see e.g. Witzany, 2017) on the basis of a univariate Gini coefficient values, Weight of Evidence and variable Information Values (for categorical and binned numerical variables).

3 Survival Analysis Methodology

The goal of survival analysis is to estimate the probability distribution of the time of exit of an object conditional on a set of explanatory variables (see Hosmer et al., 2008). The distribution can be specified by the cumulative distribution function $F(t) = F(t|\mathbf{x})$, or by the density function $f(t)$, or the survival function $S(t) = 1 - F(t)$, or by the hazard function $h(t) = f(t)/S(t)$. The classical examples of exit are the death of a patient or the breakage of a machine part, the default of an exposure, etc. In our case, it will be the event of repayment, with the time measured (in days) from the exposure entry in the soft collection process.

A dataset to estimate the model contains observations with explanatory variables, and not just with binary outcomes, but also with the time of exit outcomes. Another advantage of this class of models is that we can also use observations where the time of exit is censored. Here, we only know that the object has survived until a time limited by our observation window, which is the case of the soft collection repayment data. In addition, we can

also work with left censored observations, which can be useful for dealing with explanatory variables that change their value during the life of an object. A survival dataset can be used to calculate Kaplan-Meier empirical hazard and survival functions simply by counting the number of exits over a period (e.g. on daily basis) out of all cases alive.

There are two broad classes of survival analysis models: parametric models, where the hazard function or survival function has a parametric form with coefficients to be estimated, and semi-parametric models, where the shape of the hazard function is not specified (e.g. the Cox model). The simplest parametric model is the exponential model where the hazard $h(t) = \lambda$ is constant conditional on the explanatory variables, typically in the form $\lambda = \exp(\mathbf{x}'\boldsymbol{\beta})$. We will also test Accelerated Failure Time (AFT) models, which are characterized by a specific distribution of the log-time of exit, e.g. normal for the lognormal model, or logistic for the loglogistic model, etc.

The vector of coefficients $\boldsymbol{\theta}$ of a parametric model is estimated by maximizing the total log-likelihood function

$$LL(\boldsymbol{\theta}) = \sum_{i=1}^n \ln L(T_i, c_i, \mathbf{x}_i, \boldsymbol{\theta}), \quad (1)$$

where $L(T_i, c_i, \mathbf{x}_i, \boldsymbol{\theta})$ is the likelihood of the observation $i = 1, \dots, n$ with the time T_i of exit or censoring indicated by $c_i \in \{0,1\}$ and with covariates \mathbf{x}_i . In the case of exit, the likelihood is $L = f(T_i; \mathbf{x}_i, \boldsymbol{\theta}) = h(T_i; \mathbf{x}_i, \boldsymbol{\theta})S(T_i; \mathbf{x}_i, \boldsymbol{\theta})$, while in the case of censoring, it is just the survival probability $L = S(T_i; \mathbf{x}_i, \boldsymbol{\theta})$. If an observation is left censored from the time $T_{i,0}$ (which needs to be indicated by a left censoring variable), then the likelihood, defined as above, is simply divided by $S(T_{i,0}; \mathbf{x}_i, \boldsymbol{\theta})$.

We will start by estimating the Cox semi-parametric model where the hazard function shape is given by a nonparametric baseline function $h_0(t)$ and $h(t) = h_0(t)\exp(\mathbf{x}'_i\boldsymbol{\beta})$. This model belongs to the class of proportional hazard models, where the coefficients $\boldsymbol{\beta}$ can be estimated by maximizing the partial log-likelihood function (assuming no ties)

$$PLL(\boldsymbol{\theta}) = \sum_{i=1}^n \ln \left(\frac{\exp(\mathbf{x}'_i\boldsymbol{\beta})}{\sum A_{ij} \exp(\mathbf{x}'_j\boldsymbol{\beta})} \right), \quad (2)$$

where A_{ij} ($j = 1, \dots, n$) is an indicator which takes the value 1 if the object j is still at risk (alive) at the beginning of period t_i and 0 otherwise. The baseline hazard function can then be directly calculated by the Breslow-Crowley estimator. The survival models can generally be estimated with time varying covariates. In this case the parametric or Cox model hazard function at time t depends on the covariates specific for this time. The coefficients of the time varying covariates remain constant and the maximum-likelihood estimation method can be still applied.

Notice that the partial log-likelihood given by (2) is not the same as the one given by (1), even if the hazard functions are the same. Therefore, if we want to compare consistently the Cox model with a parametric one using the log-likelihood or Akaike criterion (AIC), we have to calculate the log-likelihood of the Cox model according to (1).

4 Survival Analysis Empirical Results

We are going to estimate the Cox model and then several alternative parametric models on the full dataset (for all products) and separately for the specific product classes, as in the case of logistic regression. In order to select the best model, we will calculate and report the AIC for all models based on the log-likelihood (1), which differs from the partial log-likelihood given by (2). The coefficients of the Cox model, expressing the linear effects of covariates on the log-hazard independent of time, are easy to interpret, but the proportional hazard assumptions also need to be tested in order to decide between the Cox and the parametric models.

Figure 1 shows, as a preliminary inspection, the Kaplan-Meier empirical estimates of the survival, cumulative hazard, and hazard functions (where the time is measured in days) for the portfolio of all products conditional only on calling or not calling at the beginning of the collection process. It is obvious that the effect of calling on the survival probability, i.e. on the probability of repayment over time, is substantial. The empirical cumulative hazard function shows that the effect of calling is larger in the days immediately following the call and diminishes over time.

Table 1 shows the Cox model output for the entire input data set and individual products, and the impact of the explanatory variables on the hazard level. Notice that the coefficient of the time varying variable *TEL* is now significantly positive for all models and does not vary much. The overall multiplicative effect of calling, i.e. $TEL =$

1, on the hazard function (for the all-product portfolio) is $\exp(2.223) \cong 9.2$. Practically, this means that the daily probability of repayment increases more than 9-times due to the call.

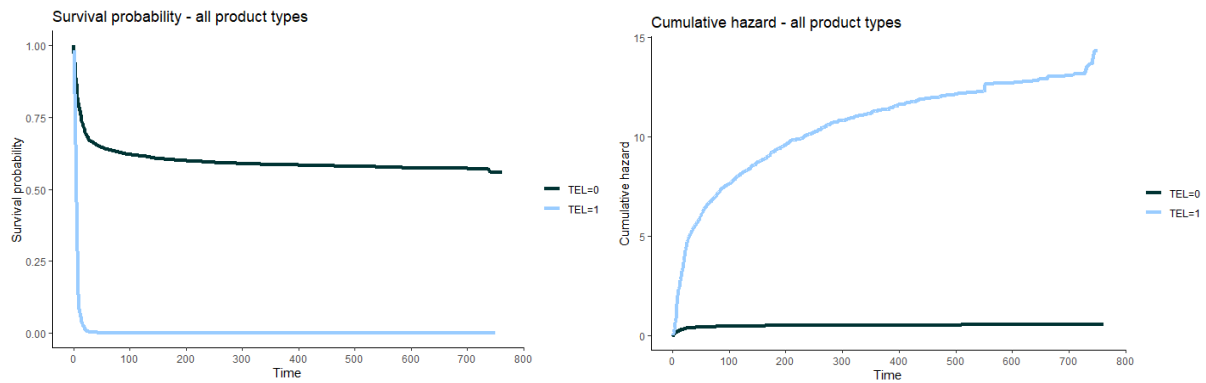


Figure 1 Kaplan-Meier estimates of the survival, cumulative hazard, and hazard functions

Variables / Est. Coef.	All products	Current account unauthorized debits	Overdrafts	Mortgages	Credit cards	Consumer and Investment loans	Current account small debits
TEL	2.223*** (0.016)	4.209*** (0.036)	3.865*** (0.059)	1.758*** (0.050)	2.694*** (0.028)	1.752*** (0.031)	1.273*** (0.056)
prod_type_1_woe	0.516*** (0.007)	-	-	-	-	-	-
exist_time_woe	0.119*** (0.008)	0.205*** (0.058)	0.620*** (0.037)	0.182*** (0.028)		-0.176*** (0.020)	0.080*** (0.018)
ovd_amount_woe	-0.088*** (0.008)	-0.278* (0.148)	0.132** (0.042)	-0.086*** (0.025)	-0.080*** (0.013)	-0.215*** (0.016)	-
resident_flag_woe	0.150*** (0.013)	0.714*** (0.056)	0.218*** (0.043)	-	-	-	0.182*** (0.019)
age_woe	- not.sign.	-	0.355*** (0.067)	0.200* (0.079)		-0.118*** (0.027)	-
cnb_class_woe	0.167*** (0.004)	-	-	0.434*** (0.020)	0.394*** (0.012)	0.514*** (0.011)	-
limit_woe	-0.230*** (0.007)	-	-	-	0.046*** (0.010)	-	-
segment_1_woe	0.221*** (0.010)	0.737*** (0.060)	-	0.290*** (0.026)	-	0.168*** (0.023)	0.116*** (0.019)
int_rate_woe	-0.377*** (0.007)	-	-	-	-	-	-
Likelihood ratio test	60, 656	1,222	5,663	2,043	13,184	7,077	586
Partial-log-likelihood	-237,557	-4,486	-9,673	-21,435	-62,218	-63,566	-33,085
Log-likelihood	-94,949	-4,071	-6,383	-9,094	-28,398	-27,012	-17,090
AIC (LL)	189,907	8,148	12,771	18,193	56,799	54,030	34,185

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 1 Cox model results

The survival data on which the Cox model is applied should also be tested for the proportional hazard assumptions. An exact option for testing the proportional hazards assumption is to perform the Xue and Schifano, (2017) statistical test based on the Schoenfeld residuals. Since the Cox model does not satisfy the proportional hazard tests very well, we shall estimate and compare several standard parametric survival models, namely the lognormal, Weibull, and loglogistic models. In order to select the best fitting model, we have calculated the Akaike criterion (AIC) shown in Table 2. The best AIC values are given by the lognormal model for all the product classes.

Distribution / AIC value	All products	Current account unauthorized debits	Overdrafts	Mortgages	Credit cards	Consumer and Investment loans	Current account small debits
lognormal	217,755	8,770	14,527	20,775	64,416	60,638	39,256

Weibull	223,386	8,825	15,148	-	-	-	40,526
loglogistic	222,995	-	15,278	21,240	67,480	61,383	39,325

Table 2 AIC values for the selected parametric models and product classes

The lognormal model coefficients estimated for the overall portfolio and individual product classes are reported in Table 3. In this case, the interpretation of the parameters is not as straightforward as in the case of the Cox model. In the lognormal model the log-time of exit is normally distributed with the mean $\mu(\mathbf{x}_i) = \beta_0 + \mathbf{x}'_i\boldsymbol{\beta}$, where meanlog given in Table 3 is the intercept β_0 , and with the standard deviation σ given by sdlog. Therefore, the negative coefficients of the time varying variable *TEL* in all models indeed significantly reduce the expected time of repayment as expected.

Variable / Est. Coef.	All products	Current account unauthorized debits	Overdrafts	Mortgages	Credit cards	Consumer and Investment loans	Current account small debits
meanlog	5.464*** (0.019)	14.10*** (0.452)	7.870*** (0.074)	4.704 (0.255)	6.793*** (0.092)	5.564*** (0.053)	2.982*** (0.025)
sdlog	1.881*** (0.011)	4.810*** (0.196)	2.470*** (0.024)	1.198*** (0.020)	1.598*** (0.017)	1.319*** (0.014)	1.657*** (0.019)
TEL	-8.585*** (0.128)	-24.2*** (2.23)	-15.6*** (0.071)	-3.586*** (0.172)	-8.108*** (0.186)	-3.251*** (0.098)	-3.233*** (0.223)
prod_type_1_woe	-0.871*** (0.009)	-	-	-	-	-	-
exist_time_woe	-0.354*** (0.016)	-0.538*** (0.150)	-2.37*** (0.011)	-0.293*** (0.034)	-	0.208*** (0.026)	-0.147 (0.863)
ovd_amount_woe	0.203*** (0.016)	1.46*** (0.453)	-	0.044*** (0.030)	-	0.333*** (0.022)	-
resident_flag_woe	-0.332*** (0.020)	-1.67*** (0.131)	-0.891*** (0.029)	-	-	-	-0.286*** (0.029)
age_woe	0.083*** (0.025)	-	-1.99*** (0.029)	0.460*** (0.093)	-	0.330*** (0.036)	-
cnb_class_woe	-0.329*** (0.009)	-	-	-0.710*** (0.026)	-0.811*** (0.019)	-0.880*** (0.015)	-
limit_woe	0.384*** (0.014)	-	-	0.112*** (0.092)	-0.133*** (0.024)	-	-
segment_1_woe	-0.414*** (0.018)	-1.66*** (0.145)	-	-0.446*** (0.032)	-	-0.119*** (0.026)	-0.191*** (0.030)
int_rate_woe	0.872*** (0.014)	-	-	-	-	-	-
AIC	217,754.8	8,769.97	14,527.19	20,774.53	64,416.36	60,637.66	39,255.65
Log-likelihood	-108,865.4	-4,377.985	-7,257.594	-10,378.26	-32,203.18	-30,310.83	-19,621.83

Significance. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 3 Lognormal model results

5 Conclusions

We have tested the logistic regression and several survival models in order to estimate the effect of calling a debtor in terms of the probability of the repayment of an exposure in the early collection process. The models were estimated on a relatively large recovery process dataset of retail defaulted loans from the period 2017-2019. We have performed a univariate analysis and preselected the most important variables. The categorical values were replaced by the WoE evidence values, with the exception of the *tel* (call) indicator. The analysis has shown different repayment patterns for individual products and a need to estimate the models separately for individual products rather than for the overall portfolio. The results of the standard logistic regression demonstrated that this type of model is not appropriate, due to the issue of call timing. The calls were, in this case, made based on call center capacity and simple prioritization rules, and so smaller and quickly repaid exposures were not usually called. This interdependence provides an explanation for the unexpected signs of the *tel* indicator for some of the products, and an argument for applying a survival analysis, in which the call can be compared to a medical treatment. The estimated Cox and selected parametric models have shown the consistent directional effects of the main time varying variable *tel* as well as of the other explanatory variables. Regarding the choice of model, the conclusions are mixed. Based on the log-likelihood Akaike Information Criterion (AIC) we would recommend the Cox model, but its

violation of the proportional hazard assumptions for most of the products led us to prefer the lognormal model, which has the best AIC values among the various parametric models considered.

Acknowledgements

This research has been supported by the Czech Science Foundation Grant 18-05244S "Innovation Approaches to Credit Risk Management" and by the VSE institutional grant IP 100040.

References

- [1] Bellotti A. & Crook J. (2013). *Retail Credit Stress Testing Using A Discrete Hazard Model With Macroeconomic Factors*. Journal of the Operational Research Society, vol 65, 2014, 340-350.
- [2] Bellotti A. & Crook J. (2010). *Time varying and dynamic models for default risk in consumer loans*. Journal of the Royal Statistical Society Series A., 173(2), 283-305.
- [3] Cao R., Vilar J.M., & Devia A. (2009). *Modelling consumer credit risk via survival analysis*. SORT 33(1): 3–30.
- [4] Chehrazi, N. & Weber, T. A. (2015). *Dynamic valuation of delinquent credit-card accounts*. Management Science, 61(12), 3077-3096.
- [5] Chehrazi, N., Glynn, P. W., & Weber, T. A. (2019). *Dynamic credit-collections optimization*. Management Science, 65(6), 2737-2769.
- [6] Collet D. (2003). *Modelling Survival Data in Medical Research*, Chapman & Hall/CRC, London.
- [7] De Almeida Filho, A. T., Mues, C., & Thomas, L. C. (2010). *Optimizing the collections process in consumer credit*. Production and Operations Management, 19(6), 698-708.
- [8] Djeundje V.B. & Crook J. (2019). *Identifying hidden patterns in credit risk survival data using generalised Additive Models*. European Journal of Operational Research, v277(1), 2019, 366-276.
- [9] Djeundje V.B. & Crook J. (2018). *Dynamic survival models with time varying coefficients for credit risks*. European Journal of Operational Research v275(1) 2019, 319-333.
- [10] He, P., Hua, Z., & Liu, Z. (2015). *A quantification method for the collection effect on consumer term loans*. Journal of Banking & Finance, 57, 17-26.
- [11] Hosmer D.W., Lemeshow S., & May S. (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. 2nd Edition, John Wiley & Sons, Chichester, United Kingdom
- [12] Kozina A. (2020). *Využití skóringu v procesu vymáhání pohledávek*. Diploma Thesis, University of Economics, Prague
- [13] Liu, Z., He, P., & Chen, B. (2019). *A Markov decision model for consumer term-loan collections*. Review of Quantitative Finance and Accounting, 52(4), 1043-1064.
- [14] Marubini E., & Valsecchi M.G. (1995). *Analysing Survival Data from Clinical Trials and Observational Studies*. John Wiley & Sons, Chichester, United Kingdom.
- [15] Murgia, G., & Sbrilli, S. (2012). *A decision support system for scoring distressed debts and planning their collection*. In Methods for Decision Making in an Uncertain Environment (pp. 69-89).
- [16] Narain, B. (1992). *Survival Analysis and the Credit Granting Decision*. In: Thomas, L.C., Crook, J.N. and Edelman, D.B., Eds., Credit Scoring and Credit Control, OUP, Oxford, 109-121.
- [17] So, M. C., Mues, C., de Almeida Filho, A. T., & Thomas, L. C. (2019). *Debtor level collection operations using Bayesian dynamic programming*. Journal of the Operational Research Society, 70(8), 1332-1348.
- [18] Stepanova, M., & Thomas, L. (2002). *Survival analysis methods for personal loan data*. Operations Research, 50(2), 277-289.
- [19] Thomas, L., Crook, J., & Edelman, D. (2017). *Credit scoring and its applications*. SIAM.
- [20] Thomas, L. C., Matuszyk, A., So, M. C., Mues, C., & Moore, A. (2016). *Modelling repayment patterns in the collections process for unsecured consumer debt: A case study*. European Journal of Operational Research, 249(2), 476-486.
- [21] Van de Geer, R., Wang, Q., & Bhulai, S. (2018). *Data-driven consumer debt collection via machine learning and approximate dynamic programming*. Available at SSRN 3250755.
- [22] Witzany, J. (2017). *Credit Risk Management*. Springer Books.
- [23] Witzany, J., Rychnovský, M., & Charamza, P. (2012). *Survival Analysis in LGD Modeling*. European Financial and Accounting Journal, 2012(1), 6-27.
- [24] Xue, Y., & Schifano, E. D. (2017). *Diagnostics for the Cox model*. Communications for statistical Applications and Methods, 24(6), 583-604.

On Modelling Dependencies between Criteria in PROMETHEE

František Zapletal¹

Abstract. PROMETHEE ranking method is a popular multi-criteria decision-making (MCDM) method which is based on pair-wise comparison using a special comparing (preference) function. As well as many other MCDM algorithms, it does not allow for modelling the dependencies between criteria. Namely, the evaluation in terms of a given criterion is completely independent from performances according to other criteria. However, this assumption is really simplifying because it is quite natural to assume *if-then* conditions like *"It is important to keep the luggage as light as possible unless one travels by car."* (i.e., when a car is used, it does not matter so much/at all how heavy the luggage is), or *"If a student did not prove sufficient knowledge of theoretical background, it does not matter what result he/she achieved in the practical part"* etc. The aim of this paper is to adjust the original PROMETHEE algorithm to model such dependencies. The proposed algorithm is demonstrated using a simple numerical example.

Keywords: PROMETHEE, conditional evaluation, multi-criteria decision-making, dependence.

JEL classification: C44

AMS classification: 90C15

1 Introduction

It is more than 30 years, since the family of PROMETHEE methods has been introduced by [3]. During this long history, many different extensions and modifications have been presented to solve various problems of multi-criteria decision making under various conditions – the PROMETHEE ranking can be found, e.g., if the input data are fuzzy [4], or random [5]; if there is a single decision-maker, or if a group of decision-makers is considered. However, there is still one factor, which has not been discussed so far. Namely, the evaluations in terms of particular criteria are always regarded as absolutely independent. To the best knowledge of the author, the only trace of dependence can be found in PROMETHEE V algorithm [7] where the optimization algorithm is used to find the best portfolio of alternatives, but in this case, the dependence can be captured using the mathematical programming models (like *The best portfolio of cars must contain a van, or at least two large family cars.*). The multiple-choice selection problem (portfolio selection) is a quite special problem and, moreover, it requires advanced skills in optimization.

The fact that the dependence has not been included in PROMETHEE does not mean that this feature is not relevant for decision-makers. It is quite usual that a good performance in terms of one criterion brings the added value only if some condition is met, i.e., if the alternative performs well in terms also another criterion. When travelling by plane, the 20kg limit for the baggage weight is relevant only for the economy class; if a passenger travels with a checked baggage, he or she needs more time to get to the airport shuttle bus, etc. Such *if-then* dependencies occur more often than we admit in our doing. Most of popular methods of multi-criteria decision-making (MCDM) force us to omit these relationships with some exceptions like the popular Saaty's method Analytical Network Process (ANP) [6], or the Choquet integral decision model [1]

The aim of this paper is to introduce a new modification of the PROMETHEE ranking algorithm to capture the dependencies between criteria. In particular, the situations where if some condition is not fully met, another criterion is not applicable for the pairwise evaluation. This proposed extension should not suppress the main advantages of the PROMETHEE approach, i.e., the ease of use and good traceability of the algorithm.

The rest of this paper is organised as follows. Sec. 2 is devoted to the brief description of the PROMETHEE algorithm. In Sec. 3, possible ways how the dependencies between criteria could be modelled are outlined and, mainly, the proposed extension of PROMETHEE with dependencies is introduced. Sec. 4 compares the original and new algorithms using a numerical example.

¹VŠB - Technical university of Ostrava, Sokolská 33, Ostrava, Czech Republic, frantisek.zapletal@vsb.cz

2 PROMETHEE ranking without dependencies

In this section, the original PROMETHEE ranking algorithm for complete ranking, i.e., PROMETHEE II, is briefly introduced. For more details, see [3].

Let us have a decision-making problem with m alternatives and k criteria. All PROMETHEE algorithms, including the PROMETHEE II ranking, share the same idea. A decision-maker makes pairwise comparisons of pairs of alternatives in terms of each criterion. This comparison is made using the preference function P_i , which is usually set individually for each criterion i . The preference function assigns the preference degree (in $[0, 1]$) according to the difference in performance between two options and can be viewed as to what extent one alternative is preferred to another with respect to the given criterion. The choice of the preference function's shape is usually essential for further results. The authors of [3] have developed 6 basic shapes of the preference functions, see Fig. 1. The suitability of each type depends on, e.g., the subjective opinion of a decision-maker, and the data type of a criterion (V-shape, linear and Gaussian preference functions are usually used to handle quantitative criteria, and U-shape, usual and level types are more common for qualitative data where substantially smaller variability in values can be expected).

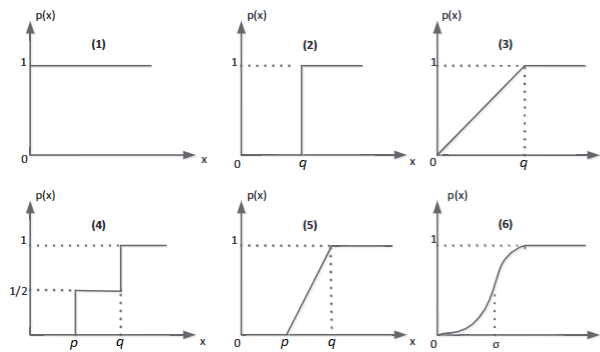


Figure 1 Types of preference functions [3]:(1) Usual, (2) U-shape, (3) V-shape, (4) Level, (5) Linear, (6) Gaussian

The PROMETHEE II algorithm brings the complete order of the alternatives with respect to their net flows ϕ , see (1).

$$\phi(A_t) = \frac{\sum_{j=1, j \neq t}^s \sum_{i=1}^k w_i \cdot P_i(v_{i,t} - v_{i,j})}{s-1} - \frac{\sum_{j=1, j \neq t}^s \sum_{i=1}^k w_i \cdot P_i(v_{i,j} - v_{i,t})}{s-1}, \quad \forall t, \quad (1)$$

where

- $v_{i,t}$ is the performance value of t -th alternative in terms of i -th criterion;
- k stands for the number of criteria;
- s stands for the number of alternatives;
- w_i represents the weight of i -th criterion;
- $\phi(A_t)$ is a net flow of the t -th alternative.

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 **indifference**

$$A_x =^{II} A_y \Leftrightarrow \phi(A_x) = \phi(A_y).$$

3 New proposed algorithm

In the previous section, the role of a preference function for the final ranking was explained. In the original algorithm, degrees of preference in terms of each criterion when comparing a pair of alternatives are mutually independent. Let us consider some conditions, which can influence the view of preference between alternatives in terms of criterion c_j . In this case, the dependence cannot be modelled using different weights because the weight is usually set for a criterion regardless the particular performance values of alternatives. A better approach is to adjust the preference degree (thus also the preference function), depending on whether the condition is met or not. Namely, if the condition is satisfied, the preference degree between the given pair of alternatives is different than if the condition is not satisfied (despite the performance values of both alternatives remain the same). Recalling the example from introduction, the 20kg weight limit is usually applied when travelling in Economy class. The

condition would be *The baggage does not exceed the weight capacity*. If it is satisfied, a passenger does not care about its weight (at least in terms of fees). For instance, let A and B be two alternatives (weighting 18kg and 20kg, respectively). If both satisfy the limit, none of them is preferred to another (the preference degree is zero). However, if their values increase by 2kg (to 20 and 22 kg, respectively), it would be reasonable to prefer the lighter baggage with some positive preference degree (because the penalty has to be paid). In this case, the condition and the impacted evaluation belong to the same criterion.

If we add to the presented "travelling example" the criterion of *travelling class*, the situation would be more complicated. If the passenger travels in the business class, the 20kg limit is not applied and even when comparing 20- and 22-kg options, the preference degree should be equal to zero. The condition related to the criterion *class* impacts the perception of preference for the criterion *weight*. Let us call this type of dependence *prohibitive* – if the condition is met, the dependent criterion can be ignored at all for the corresponding pairwise evaluations and neither positive nor negative flow of the compared alternatives are influenced by this criterion. Each alternative is described by a 2-tuple – namely, the information, if the condition is met (Y), or not (N), and the performance value in terms of the dependent criterion. Then, when comparing two alternatives, the following cases can occur:

- a) $[Y, \bullet] \times [Y, \bullet]$,
- b) $[N, \bullet] \times [Y, \bullet]$,
- c) $[N, \bullet] \times [N, \bullet]$,

where \bullet stands for the performance value of the alternative in terms of the dependent criterion. If the condition is satisfied by both criteria (a), i.e., if both alternatives are business class, regardless the performance values in weight, the preference degree must be zero. The other way around, if none of the pair satisfies the condition (c), the difference in weight matters, and the decision-maker should tend to prefer the lighter baggage. The question is how to deal with the comparison of the pair where one alternative meets the condition and the second one does not (b). Here, it depends on whether the condition is taken as another criterion, or if it is used only to describe the dependence. It is more transparent not to keep the condition aside the decision matrix and work with the corresponding parameter like with a regular criterion (dichotomous criterion *class* which can be either economy or business for the sake of simplicity). In this case, the preference of *business* over *economy* is treated with the corresponding criterion and it is not necessary to double/duplicate this difference also by *weight* criterion. Therefore, the author proposes to handle the comparison between a condition-satisfying alternative and a condition-unsatisfying alternative (b) in the same manner as the comparison between two condition-satisfying options (a). I.e., the preference degrees of the comparisons in both orders (the first vs. the second, and the second vs. the first) should be always zero. In other words, for the alternatives for which the condition holds cannot either increase, or decrease their net flow through the dependent criterion. The other way around, the net flow of an alternative not satisfying the condition can be increased by *weight* criterion only if it weighs less than at least one another not satisfying alternative (and vice versa for the decrease of the net flow). The rules are comprehensibly summarized using the example in Tab. 1 and the resulting preference degrees in Tab. 2. It can be seen that the passenger appreciates only the potential saving of 3kg by which economic V_1 is better than economic V_2 .

Alternative	Class	Weight [kg]
V_1	Economy	20
V_2	Economy	23
V_3	Business	21
V_4	Business	24
Preference function	Usual	V-shape*

Table 1 Didactical example with 4 alternatives. *the preference threshold was set to 6, i.e., the difference 6kg and greater brings the absolute preference

	V_1	V_2	V_3	V_4
V_1	–	0.5	0	0
V_2	0	–	0	0
V_3	0	0	–	0
V_4	0	0	0	–

Table 2 Resulting preference degrees for the example provided in Tab. 1

At the end of this section, there is one potential drawback of the proposed algorithm which must be taken into account. In extreme cases, it could happen that a dominated alternative can be ranked better than the dominating one (i.e., despite one alternative performs not worse than another in terms of all criteria, it is ranked worse). This

phenomenon contradicts the natural logic and is not welcome. However, at the end of this section, it is shown that the problem relates mainly with the weights and the demonstration how to avoid this problem is provided.

If another alternative V_5 is added to the problem in Tab. 1, and assuming that it is a business class and weight 18kg, V_1 will be dominated by V_5 (V_5 is lighter with better class). However, if the weight of *Weight* criterion is 0.9 and the weight of *Class* is 0.1, and if the usual shape of the preference function (see Fig. 1) is applied for both criteria, the resulting net flows will be $\phi(V_1) = \frac{7}{30}$ and $\phi(V_5) = \frac{2}{30}$, respectively. It is apparent that the illogical choice of the weights is to blame for the troubles. The weight of the baggage can be regarded as a subordinate criterion of the class criterion (only if the class is Economy, the baggage weight is applied), and therefore it is not reasonable to assign a greater weight to the dependent criterion. Below, the proof that if the weight of the dependent criterion is greater than the weight of the conditioning criterion, the inconsistency cannot occur.

Let us take only a problem with two criteria (one conditional with the weight w_c and the second one depending with the weight w_d) and let n and m be the number of alternatives satisfying and not satisfying the condition, respectively. The positive flow ϕ^+ of an alternative Y satisfying the condition is:

$$\phi^+(Y) = \frac{w_d \cdot 0 + w_c \cdot m}{m + n - 1},$$

i.e. the depending criterion is not relevant for the evaluation but the alternative is absolutely preferred over all m alternatives not satisfying the condition. The negative flow of Y must be zero because none of the other alternatives can be preferred over this one. Then, the resulting net flow $\phi(Y)$ is given as:

$$\phi(Y) = \frac{w_c \cdot m}{m + n - 1}.$$

The greatest possible positive flow of an alternative N not satisfying the condition will be:

$$\phi^+(N) = \frac{(m - 1) \cdot w_d + w_c \cdot 0}{m + n - 1},$$

i.e., if this alternative is absolutely preferred over all remaining $m - 1$ not satisfying alternatives. On the other hand, the negative flow of N must be positive because it is not preferred over all n alternatives satisfying the condition:

$$\phi^-(N) = \frac{0 \cdot w_d + w_c \cdot n}{m + n - 1},$$

and then, the resulting net flow $\phi(N)$ cannot be greater than:

$$\phi(N) = \frac{(m - 1) \cdot w_d - w_c \cdot n}{m + n - 1}.$$

The only case when the apparent inconsistency can occur is the one when the net flow of an alternative not satisfying the condition is greater than the net flow of an alternative satisfying the condition. This means that no problem can occur if $\phi(Y) > \phi(N)$, i.e., when

$$\frac{w_c \cdot m}{m + n - 1} > \frac{(m - 1) \cdot w_d - w_c \cdot n}{m + n - 1},$$

thus when

$$\frac{w_c}{w_d} > \frac{m - 1}{m + n}.$$

Because $\frac{m-1}{m+n}$ is always non-negative and less than 1, the inconsistency problem cannot occur for sure if $w_c > w_d$. If the conditional criterion is not considered binary, it can be easily proved that the results are similar with only one difference: $\phi^-(Y)$ can be positive, but always less than or equal to $\frac{(n-1) \cdot w_c}{m+n-1}$ (i.e., in the case that this alternative is absolutely worse than other satisfying alternatives in terms of the conditional criterion). Then, the ratio between w_c and w_d must be greater, but still less than 1 (namely, $\frac{m-1}{m+1}$).

4 Numerical example

In this section, an apartment selection example is provided to demonstrate the proposed algorithm. Six available apartments (V_1 to V_6) are to be evaluated using 4 criteria (C_1 price, C_2 mortgage conditions, C_3 size, C_4 cellar). Let us assume that the mortgage conditions (including the interest rate) differ for each option (real estate brokers can have different contracts with banks). This criterion is important only if the mortgage is necessary that happens

	Price [K CZK]	Mortgage conditions	Size [m ²]	Cellar
V_1	850	1	65	0
V_2	950	3	75	1
V_3	1000	2	85	1
V_4	1100	2	90	0
V_5	1150	1	70	0
V_6	1200	3	100	1
Preference function	V-shape	Level	V-shape	Usual
Parameters	$q = 300$	$p = 1, q = 2$	$q = 40$	-

Table 3 Input data for the numerical example

if the price exceeds 1M CZK (this is the amount saved by the decision-maker). Thus, if the price exceeds 1M, the mortgage conditions matter. Then, a cellar is highly appreciated if the size is low. Namely, if the size is greater than 80m², the cellar is unnecessary. The decision matrix is provided in Tab. 3.

The mortgage condition (C_2) is assessed using the special scale: 1 = without benefits, 2 = premium interest rate, 3 = premium interest rate + free insurance. The cellar is dichotomous - an apartment either has the cellar (1) or not (0). The preference functions and their parameters can be found in Tab. 3 and are set so that they have a good distinguishing power for the problem. All criteria are taken as equally important ($w_1 = w_2 = w_3 = w_4 = 0.25$).

The matrices 2 show the preference degrees of the alternatives for which the dependent criteria (mortgage conditions and cellar) are relevant. An interested reader can compare those values with the values obtained using the original algorithm, see Tab. 4.

$$\begin{array}{cccc|cccc}
 C_2 & V_4 & V_5 & V_6 & C_4 & V_1 & V_2 & V_5 \\
 V_4 & - & 1/2 & 0 & V_1 & - & 0 & 0 \\
 V_5 & 0 & - & 0 & V_2 & 1 & - & 1 \\
 V_6 & 1/2 & 1 & - & V_5 & 0 & 0 & -
 \end{array} \quad (2)$$

C_2	V_1	V_2	V_3	V_4	V_5	V_6	C_4	V_1	V_2	V_3	V_4	V_5	V_6
V_1	-	0	0	0	0	0	V_1	-	0	0	0	0	0
V_2	0	-	0	1/2	1	0	V_2	1	-	0	1	1	0
V_3	0	0	-	0	1/2	0	V_3	1	0	-	1	1	0
V_4	1/2	0	0	-	1/2	0	V_4	0	0	0	-	0	0
V_5	0	0	0	0	-	0	V_5	0	0	0	0	-	0
V_6	1	0	1/2	1/2	1	-	V_6	1	0	0	1	1	-

Table 4 Preference degrees for c_1 and c_2 according to the original algorithm

The final results of the example solved by the proposed algorithm and their comparison with the original PROMETHEE II ranking are provided in Tab. 5 and Tab. 6, respectively.

New algorithm	V_1	V_2	V_3	V_4	V_5	V_6
ϕ^+	0.183	0.227	0.131	0.131	0.191	0.219
ϕ^-	0.167	0.099	0.073	0.171	0.329	0.2
ϕ	0.015	0.128	0.058	-0.04	-0.138	0.019
Ranking	4	1	2	5	6	3

Table 5 PROMETHEE evaluation of the alternatives using the new algorithm

The ranking of the alternatives is similar for the algorithms, only two pairs switched their positions: V_3 with V_6 and V_1 with V_4 . On the other hand, one can see that the dependencies between criteria can influence the final evaluation. Above that, the original algorithm apparently overestimates the flows of the alternatives performing well in terms of the depending criteria which are not relevant for them. V_6 has excellent mortgage conditions and cellar, however, none of these benefits is relevant here in fact.

Original algorithm	V_1	V_2	V_3	V_4	V_5	V_6
ϕ^+	0.183	0.427	0.331	0.156	0.191	0.444
ϕ^-	0.419	0.099	0.135	0.346	0.504	0.2
ϕ	-0.235	0.328	0.196	-0.19	-0.313	0.244
Ranking	5	1	3	4	6	2

Table 6 PROMETHEE evaluation of the alternatives using the original algorithm

5 Conclusions

This paper introduces an alternative algorithm of the PROMETHEE method, which takes into account the dependencies between criteria. Good performance in one criterion can be worthless if some condition is/is not satisfied and vice versa. Several examples, where these dependencies can play an important role, were presented in this paper. The proposed algorithm was demonstrated using the numerical example. This demonstration proved that the dependencies can influence the final ranking and that the new algorithm allows to model the selected decision making problems better. On the other hand, special attention has to be paid to the weights of the mutually dependent criteria. The future research in this field will be focused on generalized dependencies between the criteria of any data type and dependencies among more than two criteria.

Acknowledgements

This work was supported by the SGS project No. SP2021/86. This support is gratefully acknowledged.

References

- [1] Angilella, S., Corrente, S., and Greco, S. SMAA-Choquet: Stochastic multicriteria acceptability analysis for the Choquet integral. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Springer, Berlin, Heidelberg (2012), 248–257.
- [2] Behzadian, M., Kazemzadeh, R. B., Albadvi, A., and Aghdasi, M.: PROMETHEE: A comprehensive literature review on methodologies and applications, *European journal of Operational research* **200(1)** (2010), 198–215.
- [3] Brans, J. P., Mareschal, B., and Vincke, P.: A New Family of Outranking Methods in Multicriteria Analysis, *Operational Research* (1984).
- [4] Geldermann, J., Spengler, T., and Rentz, O. Fuzzy outranking for environmental assessment. Case study: iron and steel making industry. *Fuzzy sets and systems* **115(1)** (2000), 45-65.
- [5] Mareschal, B. Stochastic multicriteria decision making and uncertainty. *European Journal of Operational Research* **26(1)** (1986), 58-64.
- [6] Saaty, T. L., and Vargas, L. G. *Decision making with the analytic network process*, Springer Science+ Business Media **282** (2006).
- [7] Vetschera, R., and De Almeida, A. T.: A PROMETHEE-based approach to portfolio selection problems. *Computers & Operations Research* **39(5)** (2012), 1010–1020.

Consumer Dividend Aristocrats: Dynamic DEA Approach

Petra Zýková¹

Abstract. The paper deals with the efficiency analysis of Consumer Dividend Aristocrats. Dividend aristocrats are companies that are members of the S&P 500 and have increased dividend payment for at least the past 25 consecutive years. The study analyses 18 companies. These companies could be analysed from many points of view. One of the possibilities is using Data Envelopment Analysis (DEA). Usually, DEA models analyse DMUs in one period. This paper proposes a dynamic DEA model with a convex vector of time weights, which compute the overall efficiency score for every DMU. This paper also proposes a super-efficiency modification of this dynamic DEA model. This paper aims to find the most efficient company or the set of the most efficient companies. These companies could be suitable for potential dividend investors.

Keywords: data envelopment analysis, dynamic efficiency, dividend aristocrats

JEL Classification: C44

AMS Classification: 90C05, 90C90

1 Introduction

Investment is recently a hot topic. The uncertain situation and increasing inflation lead more people to think about their hard money. There are many possibilities where invest the money: bonds, shares, real estate, option, cryptocurrencies, forex, precious metals, commodities and others. This article focuses on the shares and dividends approach. Concretely on US market and companies called Dividend Aristocrats which pay dividend more than 25 years consecutive. There exist many common attitudes on how to find a suitable investment for the concrete investor. Investors usually use technical analysis or fundamental analysis [12]. PE (Price to Earnings ratio), PB (Price to Book ratio), or PS (Price to Sales) are often used ratios of fundamental analysis.

This article offers an analysis based on dynamic Data Envelopment Analysis. DEA models have been first developed by [3] based on the concept introduced by [6]. Using DEA models, the efficiency scores of decision-making units are computed. These models classified decision-making units into two subsets – efficient and inefficient. Inefficient units can be easily ranked according to efficiency scores, but efficient ones cannot be ranked due to their identical maximum efficiency scores. Therefore the super-efficiency model has been proposed by [1]. These models do not cover the influence of time, which means these models are static. If we want to cover the impact of time in counting efficiency scores, we have to use dynamic DEA models. Firstly were dynamic data envelopment analysis showed by [7] and then [5]. This article proposed an original dynamic DEA model and its super-efficiency modification. Those models are used for our analysis of Dividend Aristocrats. DEA models for different investment analysis have been used by [8] and [2].

The paper is organised as follows. The following section presents the definition of DEA models generally. Section three proposed new dynamic DEA models based on the convex vector of time weights. Section 4 contains the analysis of Dividend Aristocrats in the Consumer sector of the market. The proposed models from section 3 are used in section four. 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 applying DEA models. Let us denote $\mathbf{Y} = (y_{rj}, r = 1, \dots, t, j = 1, \dots, n)$ a non-negative matrix of outputs and $\mathbf{X} = (x_{kj}, k = 1, \dots, w, j = 1, \dots, n)$ a non-negative matrix of inputs. The efficiency score of the unit under evaluation DMU j_0 is derived as follows:

¹ Prague University of Economics and Business, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic, petra.zykova@vse.cz.

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

$$\begin{aligned} \frac{\sum_{r=1}^t u_r y_{rj}}{\sum_{k=1}^w v_k x_{kj}} &\leq 1, \quad j = 1, \dots, n, \\ u_r &\geq \varepsilon, \quad r = 1, \dots, t, \\ v_k &\geq \varepsilon, \quad k = 1, \dots, w, \end{aligned} \quad (1)$$

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

$$U_{j_0} = \sum_{r=1}^t u_r y_{rj_0}$$

subject to

$$\begin{aligned} \sum_{k=1}^w v_k x_{kj_0} &= 1, \\ \sum_{r=1}^t u_r y_{rj} - \sum_{k=1}^w v_k x_{kj} &\leq 0, \quad j = 1, \dots, n, \\ u_r &\geq \varepsilon, \quad r = 1, \dots, t, \\ v_k &\geq \varepsilon, \quad k = 1, \dots, w. \end{aligned} \quad (2)$$

Above mentioned DEA models analyse the efficiency score of decision-making units only in one period.

3 Dynamic DEA models

Dynamic data envelopment models divide into two groups. In the first group, the efficiency score is computed for every year or every group of years (windows). Window analysis is one example of these dynamic models [4]. The disadvantage of these models is in the aggregation of single efficiency scores into one efficiency score for the period. In the second group, the efficiency score is computed for every decision-making unit in one stage. The efficiency score is called the overall efficiency score. Let us denote $\mathbf{Y} = (y_{rj}^\tau, \tau = 1, \dots, T, r = 1, \dots, t, j = 1, \dots, n)$ a non-negative three-dimensional matrix of outputs and $\mathbf{X} = (x_{kj}^\tau, \tau = 1, \dots, T, k = 1, \dots, w, j = 1, \dots, n)$ a non-negative three-dimensional matrix of inputs.

We introduce the new dynamic DEA models, which computed the overall efficiency score. These dynamic DEA models' efficiency score summarises the influence of all inputs and outputs during the investigated time series. Every single input and output of every single unit has an impact on the final efficiency score. That means that all inputs and outputs overall years have some effect.

The model with the overall efficiency score of the unit under evaluation DMU_{j_0} is the following:

Maximise

$$\begin{aligned}
 AR_{j_0} &= \sum_{\tau=1}^T \phi_{j_0}^{\tau} U_{j_0}^{\tau} \\
 \phi_j^{\tau} &\geq \delta, \quad \tau = 1, \dots, T, \quad j = 1, \dots, n, \\
 \sum_{\tau=1}^T \phi_j^{\tau} &\leq 1, \quad j = 1, \dots, n, \\
 \phi_j^{\tau} - 2\phi_j^{\tau-1} + \phi_j^{\tau-2} &\geq \delta, \quad \tau = 1, \dots, T-2, \quad j = 1, \dots, n, \\
 \text{subject to} \quad \sum_{r=1}^t u_r^{\tau} y_{rj_0}^{\tau} &= U_{j_0}^{\tau}, \quad \tau = 1, \dots, T, \\
 \sum_{k=1}^w v_k^{\tau} x_{kj_0}^{\tau} &= 1, \quad \tau = 1, \dots, T, \\
 \sum_{r=1}^t u_r^{\tau} y_{rj}^{\tau} - \sum_{k=1}^w v_k^{\tau} x_{kj}^{\tau} &\leq 0, \quad \tau = 1, \dots, T, \quad j = 1, \dots, n, \\
 u_r^{\tau} &\geq \varepsilon, \quad \tau = 1, \dots, T, \quad r = 1, \dots, t, \\
 v_k^{\tau} &\geq \varepsilon, \quad \tau = 1, \dots, T, \quad k = 1, \dots, w,
 \end{aligned} \tag{3}$$

where AR_{j_0} is the overall efficiency score for j_0 th unit under evaluation, ϕ_j^{τ} is time weight of τ th year and j th unit, U_j^{τ} is a year efficiency in τ th year and for j th unit, δ is a small number, u_r^{τ} is a positive weight of the r th output in a year τ , v_k^{τ} is a positive weight of the k th input in the year τ and T is the number of years.

The first set of constraints $\phi_j^{\tau} \geq \delta$ assures that every year has an impact on the overall efficiency score AR_{j_0} .

The second set of constraints $\sum_{\tau=1}^T \phi_j^{\tau} \leq 1$ says that the sum of time weights for j th unit is smaller or equal to one. It guarantees that the overall efficiency score AR_j is smaller or equal to one, and also, it means that the j_0 th decision-making unit is efficient if AR_{j_0} is equal to one. The set of constraints

$$\phi_j^{\tau} - 2\phi_j^{\tau-1} + \phi_j^{\tau-2} \geq \delta, \quad \tau = 1, \dots, T-2, \quad j = 1, \dots, n, \tag{4}$$

says that the time weights ϕ_j^{τ} are convex with time. The rest constraints are usual as CCR input-oriented DEA model. The (4) says that the last time's weight must be higher than the weight of the last but one time. The unit under evaluation j_0 is efficient if AR_{j_0} is equal to one. Generally, it could be more than one efficient unit.

Therefore the original super-efficiency form of the mentioned model is proposed. The first super-efficiency model introduces [1]. The super-efficiency form of the model (3) is the following:

Maximise

$$\begin{aligned}
 AR_{j_0}^* &= \sum_{\tau=1}^T \phi_{j_0}^{\tau} U_{j_0}^{\tau} \\
 \phi_j^{\tau} &\geq \delta, \quad \tau = 1, \dots, T, \quad j = 1, \dots, n, \\
 \sum_{\tau=1}^T \phi_j^{\tau} &\leq 1, \quad j = 1, \dots, n, \\
 \phi_j^{\tau} - 2\phi_j^{\tau-1} + \phi_j^{\tau-2} &\geq \delta, \quad \tau = 1, \dots, T-2, \quad j = 1, \dots, n, \\
 \text{subject to} \quad \sum_{r=1}^t u_r^{\tau} y_{rj_0}^{\tau} &= U_{j_0}^{\tau}, \quad \tau = 1, \dots, T, \\
 \sum_{k=1}^w v_k^{\tau} x_{kj_0}^{\tau} &= 1, \quad \tau = 1, \dots, T, \\
 \sum_{r=1}^t u_r^{\tau} y_{rj}^{\tau} - \sum_{k=1}^w v_k^{\tau} x_{kj}^{\tau} &\leq 0, \quad \tau = 1, \dots, T, \quad j = 1, \dots, n, \quad j \neq j_0 \\
 u_r^{\tau} &\geq \varepsilon, \quad \tau = 1, \dots, T, \quad r = 1, \dots, t, \\
 v_k^{\tau} &\geq \varepsilon, \quad \tau = 1, \dots, T, \quad k = 1, \dots, w,
 \end{aligned} \tag{5}$$

where $AR_{j_0}^*$ is the overall super-efficiency score for j_0 th unit under evaluation in the model (5).

All proposed models contain three indexes. The first one is for years (or generally index of time series). The second one is for the decision-making unit, and the third index is for input or output. It means that the model counted with a three-dimensional matrix. This occasion is problematic by solving the models. It is necessary to transfer the data set in MS Excel, which is only two-dimensional to the three-dimensional **X** and **Y** matrixes. The transfer produces by three-cycle in the LINGO program. The model's code is not simple, and the creation of the code spent quite enough time.

4 Dividend Aristocrats

There are two major stocks exchange in the United States: NYSE (New York Stock Exchange) and NASDAQ (National Association of Securities Dealers Automated Quotations). NYSE is the biggest stock exchange in the world by market capitalisation of shares traded and is situated in New York City. NASDAQ is the second biggest stock exchange in the world by market capitalisation.

The S&P 500 (Standard and Poor's 500) is the stock market index of 500 large companies listed on the stock exchanges in the United States. Dividend aristocratic companies have to fulfil two qualifications. Firstly must be a member of the S&P 500, and secondly, must have increasing dividend payment for at least the past 25 consecutive years. Up to date, there are 65 Dividend Aristocrats. The companies are divided into sectors according to the business. The sectors are Basic Materials, Communication Services, Consumer Cyclical, Consumer Defensive, Energy, Financial Services, Healthcare, Industrials, Real Estate, Technology, Utilities. The distribution of Dividend Aristocrats by sectors is shown in Figure 1. The study analyses sectors Consumer Cyclical ad Consumer Defensive, e.g. 29 % the Dividend Aristocrats market.

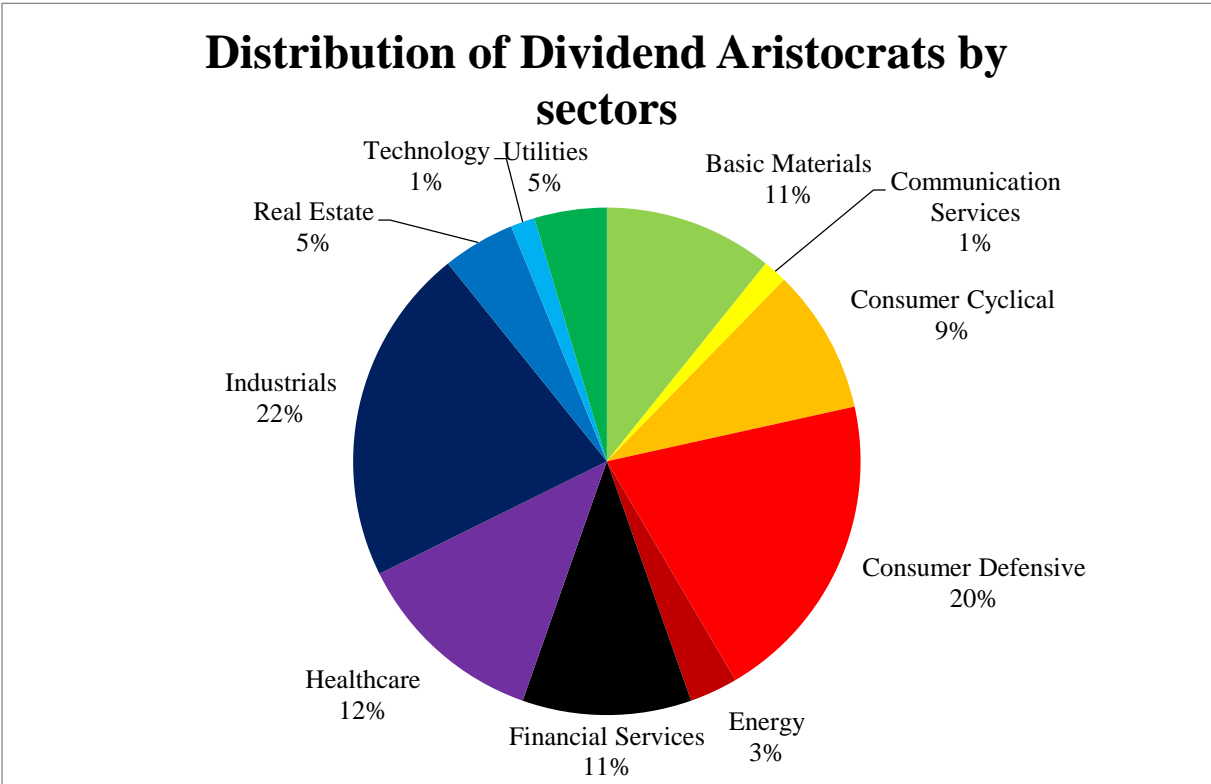


Figure 1 Distribution of Dividend Aristocrats by sectors [14].

There would be investigated Dividend Aristocrats form sectors: Consumer Defensive and Consumer Cyclical. Consumer Defensive are companies: Archer Daniels Midland Co., Brown-Forman Corp., Clorox Co., Coca-Cola Co., Colgate-Palmolive Co., Hormel Foods Corp., Kimberly-Clark Corp., McCormick & Co. Inc., PepciCo Inc., Procter & Gamble Co., Sysco Corp., Target Corp., Walmart Inc. Consumer Cyclical are companies Genuine Parts Co., Leggett & Platt, Inc., Lowe's Cos., Inc., McDonald's Corp., VF Corp.

The analysis investigated 18 companies. The study should help dividend investors advise in which company he/she should invest. The data is up to the end of 2020. The investors decide at the end of the year 2020. The data was obtained from [10],[11] and [13]. This analysis investigates the overall efficiency for 16 consecutive years from 2005 to 2020. We have used two inputs: Market Capitalisation in billions of US dollars and Share-

holder's equity in billions of US dollars. And two outputs: Earning Per Share in US dollars and Dividend Per Share in US dollars. Shareholder's equity and Earning Per Share could reach negative values. This problem can be easily solved by way of the translation invariance property [9]. We have used translation in this study. We computed with $\varepsilon = \delta = 10^{-8}$.

Ticker	Model (3)		Model (5)		Dividend Yield	
	1	1 - 3	2.867435	1	3.91%	1
LEG	1	1 - 3	2.867435	1	3.91%	1
CLX	1	1 - 3	1.345288	2	2.03%	12
MCD	1	1 - 3	1.290993	3	2.30%	8
GPC	0.999907	4	1.118764	4	3.21%	2
KMB	0.965258	5	0.985804	5	3.07%	4
LOW	0.854390	6	0.854390	6	1.46%	15 - 16
TGT	0.822454	7	0.822462	7	1.63%	14
MKC	0.731965	8	0.731990	8	1.34%	17
CL	0.731822	9	0.731171	9	2.19%	9 - 10
PEP	0.724231	10	0.724230	10	2.75%	7
VFC	0.669592	11	0.669600	11	2.17%	11
BF.B	0.656638	12	0.658093	12	0.85%	18
HRL	0.647341	13	0.649393	13	1.84%	13
SYN	0.641684	14	0.641684	14	2.84%	6
ADM	0.636143	15	0.636143	15	2.89%	5
PG	0.536445	16	0.536403	16	2.19%	9 - 10
KO	0.513546	17	0.513546	17	3.14%	3
WMT	0.432440	18	0.432440	18	1.46%	15 - 16

Table 1 The overall efficiency and super-efficiency score for 18 analysed companies, dividend yield of companies in 2020

The results of the analysis are shown in Table 1. According to the model (3), there are three efficient companies: Leggett & Platt, Inc., Clorox Co., McDonald's Corp., with an overall efficiency score equal to one. The model (5) determined the following results. Four companies were identified as efficient. The most efficient company is Leggett & Platt, Inc. with the overall super-efficiency score equal to 2.867435. The second most efficient company is Clorox Co. with the overall super-efficiency score equal to 1.345288. The third most efficient company is McDonald's Corp. with the overall super-efficiency score equal to 1.290993. And the fourth efficient company is Genuine Parts Co. with the overall super-efficiency score equal to 1.118764.

The dividend investors should invest in one of the efficiency companies according to the models (3) and (5). There is also a dividend yield in 2020 [13] of investigated companies in Table 1. A dividend yield is a ratio of dividend and price. According to the dividend yield in 2020, the best company for investing is Leggett & Platt, Inc. with a dividend yield equal to 3.91%. The second-highest dividend yield in 2020 had Genuine Parts Co. (3.21%). The third highest dividend yield in 2020 had Coca-Cola Co. (3.14%). The fourth highest dividend yield in 2020 had Kimberly-Clark Corp. (3.07%). The proposed models and dividend yield agreed with investing in Leggett & Platt, Inc. and Genuine Parts Co. The proposed models involve more information about the company than only price and dividend as a dividend yield. The proposed models could help the dividend investors with decision-making about where they will invest their money. They are also other types of investors, less conservative than dividend investors, deciding according to different ratios. Other possible and often used are PE (Price to Earnings ratio), PB (Price to Book ratio), or PS (Price to Sales).

5 Conclusions

The paper dealt with dynamic efficiency analysis based on Data Envelopment Analysis. The article proposed two new models. The models compute the overall efficiency (super-efficiency) score for the decision-making unit with covering the impact of all inputs and outputs during the investigated time series. Thus, every single input and output of every single unit affects the final efficiency score.

The analysis investigated Dividend Aristocrats in sectors Consumer Cyclical and Consumer Defensive. Dividend aristocrats companies that are members of the S&P 500 and have increased dividend payment for at least the past 25 consecutive years. The study has analysed 18 companies.

According to the proposed models, dividend investors at the end of 2020 should invest in one of these companies: Leggett & Platt, Inc., Clorox Co., McDonald's Corp. or Genuine Parts Co.

The numerical experiments were realised using original procedures written in the LINGO modelling language. Dynamic DEA models are an interesting subject for further research.

Acknowledgements

This work was supported by the Internal Grant Agency of the Faculty of Informatics and Statistics, Prague University of Economics and Business, project F4/29/2020 (*Dynamic data envelopment analysis models in economic decision making*).

References

- [1] Andersen, P. & Petersen, N.C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, 39(10), pp.1261–1264.
- [2] Chang, T.S., Tone, K. & Wu, Ch.-H. (2021). Nested dynamic network data envelopment analysis models with infinitely many decision-making units for portfolio evaluation, *European Journal of Operational Research*, 291 (2), pp. 766-781.
- [3] Charnes, A., Cooper, W. & Rhodes, E. (1978). Measuring the efficiency of decision-making units. *European Journal of Operational Research*, 2(6), pp.429-444.
- [4] Dlouhý, M., Jablonský, J. & Zýková P. (2018). *Analyza obalu dat*. Praha: Professional Publishing.
- [5] Färe, R. & Grosskopf, S. (1996). *Intertemporal Production Frontiers: With Dynamic DEA*, Kluwer, Dordrecht.
- [6] Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), pp. 253-290.
- [7] Sengupta, J.K. (1995). Dynamics of Data Envelopment Analysis, *Theory of Systems Efficiency*, Kluwer, Dordrecht.
- [8] Zamani, L., Beegam, R. & Borzoian, S. (2014). Portfolio Selection using Data Envelopment Analysis (DEA): A Case of Select Indian Investment Companies. *International Journal of Current Research and Academic Review*, 2(4), pp. 50-55.
- [9] Zhu, J. (2014). *Quantitative Models for Performance Evaluation and Benchmarking*. Springer International Publishing.
- [10] Companies Market Capitalization. (2021). [online] Available at: <https://companiesmarketcap.com/>, [Accessed 31 March 2021].
- [11] Dividend. (2021). [online] Available at: <https://www.dividend.com/>, [Accessed 31 March 2021].
- [12] Investopedia. (2021). [online] Available at: <https://www.investopedia.com/>, [Accessed 31 March 2021].
- [13] Macrotrends. (2021). [online] Available at: <https://www.macrotrends.net/>, [Accessed 31 March 2021].
- [14] Suredividend. (2021). [online] Available at: <https://www.suredividend.com/>, [Accessed 31 March 2021].

**39th International Conference on
Mathematical Methods in Economics (MME 2021)
Conference Proceedings**

Editor: Robert Hlavatý
Cover: Jiří Fejfar
Technical editors: Jiří Fejfar, Michal Hruška
Publisher: Czech University of Life Sciences Prague
Kamýcká 129, Prague 6, Czech Republic

546 pages

First edition

Prague 2021

ISBN 978-80-213-3126-6

MME2021

Mathematical Methods in Economics



Czech University Of Life Sciences Prague

Faculty of Economics
and Management

<https://mme2021.czu.cz/en>