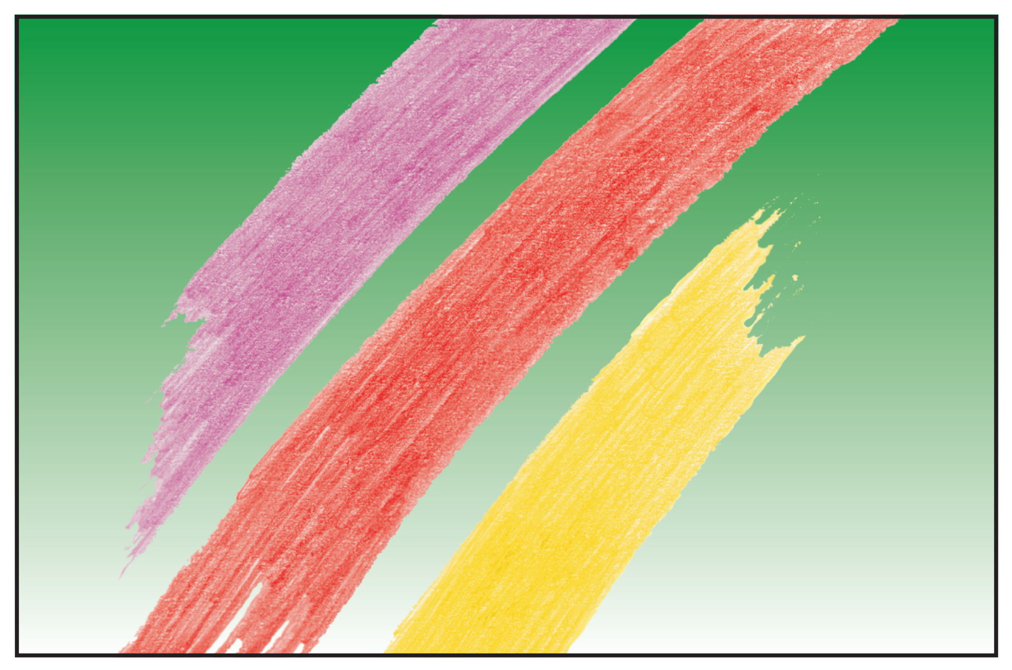
# QUANTITATIVE METHODS IN ECONOMICS

## Multiple Criteria Decision Making XXI



Proceedings of the International Scientific Conference

25th May - 27th May 2022

Púchov, Slovakia

Letra Edu

**The Slovak Society for Operations Research**
**Department of Operations Research and Econometrics**
**Faculty of Economic Informatics, University of Economics in Bratislava**

SSOR

SLOVAK SOCIETY FOR OPERATIONS RESEARCH

Proceedings of the International Scientific Conference

# QUANTITATIVE METHODS IN ECONOMICS
## Multiple Criteria Decision Making XXI

**25th May - 27th May 2022**
**Púchov, Slovakia**

**25th May - 27th May 2022**
**Púchov, Slovakia**

Technical Editor: Pavel Gežík, University of Economics, Bratislava
Web Editor: Martin Lukáčik, University of Economics, Bratislava
Web: https://fhi.euba.sk/katedry/katedra-operacneho-vyskumu-a-ekonometrie/konferencia-qme

Publisher: Letra Edu, s. r. o.
ISBN: 978-80-89962-93-8 (print)
ISBN: 978-80-89962-94-5 (online)

## CONTENT

# A METAFRONTIER APPROACH AND FARM PERFORMANCE BY THEIR MEMBERSHIP IN PRODUCER'S ORGANIZATIONS

***Lubica Bartova, Peter Fandel***
***Slovak University of Agriculture in Nitra***

**Abstract**
We demonstrate an application of a meta-efficiency approach based on the two-stage DEA (Charnes, Cooper and Rhodes, 1981) to account for farm technology heterogeneity due to their membership in Producers Organisations (PO) and production specialisation. We distinguished farms to non-members, permanent members, and members of newly established PO, which were supported under the Rural Development Program. Our sample consisted of 645 large farms. The farm performance in year 2014 was studied in three prevailing farm production specialisations. We found Livestock and Crop farms – members of the newly established POs, as the most technically efficient. Majority of these farms behaved opportunistically by cancelling their membership in the PO after the termination of the RDP support.

***Keywords:*** *Two-stage DEA metafrontier; Efficiency; Producers' organisations; Farms*

***JEL Classification:*** C61; D24; C12
***AMS Classification****:* 90C05; 90C90

## 1 INTRODUCTION

Underdeveloped vertical and horizontal cooperation have been recognised in the agro-food supply chains in Slovakia and addressed by the 2004-2006 Sectoral Operational Programme SAPARD and later by the 2007-2013 Rural Development Programme (RDP). Some PO have been active in Slovakia already before 2007. Support for newly established PO was provided under the 2007-2013 RDP. The aim of the new PO support was to assure common sale of the members' production for which the PO was recognised; to assist them in joint placing of goods on the market or to processing entities; to assure the quality of products; to centralise of sales and supply to improve competitiveness of the PO members. Eligible selected producer groups were allowed to draw funds for the period of 5 years. The funds depended on the value of production supplied to the market. The supported PO activities were oriented on enhancement of their market position by marketing and selling products of their members. Nevertheless, in the period of 2011-2012 there was relatively low market share for producers' organisations in Slovakia (Bijman et al, 2012).

There is growing number of studies in the EU member states observing effects of financial support provided to establishment of the new PO on either the PO performance or performance of their members. The objective of our study was to demonstrate an application of a meta-efficiency approach based on the two-stage DEA (Charnes, Cooper and Rhodes, 1981) in assessment of farm PO membership effect on their performance, by accounting for their technology heterogeneity due to membership in POs and production specialisation. We expected that farm membership in the newly established PO contributed significantly to improvement of their members' performance and to persistence of this membership.

The rest of the paper is organized as follows. Section 2 presents Literature review. Section 3 discusses used data and methods. In Section 4 we analyse obtained results. Finally, Section 5 presents discussion and conclusions.

## 2 LITERATURE REVIEW

Weak position of agricultural producers in agro-food supply chain, relations with intermediaries, processing industry and retail sector analysed McCorriston (2002), Reardon et al (2009), Fałkowski and Ciaian (2016). Farm membership in the POs could help to reduce price instability, transaction costs, total costs, and farms can benefit from higher selling prices (Cronin et al, 2018; Bijiman et al, 2012). In most of the new EU Member States (NMS) the PO have negative image and low degree of market penetration - below 25% compared to 40% or above in the old EU Member States (Copa-Cogeca, 2015).

Counterfactual approach using propensity score matching (PSM) and difference-in-differences (DID) or endogenous switching regression (ESR) are used to derive treatment effects of membership in producers' cooperation (Hoken and Su, 2015; Adjin et al, 2020). The last one estimated the effects of the RDP SR support of newly established PO on target indicators of their members and found positive effect on employment and in a less extent on gross value added (GVA) and labour productivity.

There is limited number of studies devoted to assessment of the PO or their members' efficiency. Non-parametric frontier production model was used in the study of Ferrier and Porter (1991). Guzmán et al (2009) examined the performance of PO using data envelopment analysis (DEA). DEA model and a bootstrap truncated regression was applied by Skevas and Grashuis (2019) to investigate spatial spillovers in the technical efficiency of farmer neighbouring cooperatives. Efficiency of POs in the EU selected agricultural sectors was assessed by Van Herck (2014).

Traditional methods of efficiency assessment assume the same production technologies for all production units. In case of different technologies, unobserved differences might be evaluated as inefficiencies. To address this issue, several approaches are suggested. Hayami (1969) and Hayami and Ruttan (1970) introduced the concept of meta-frontier approach in assessing differences in the agricultural productivity of selected countries. Battese and Rao (2002) and Battese, Rao and O'Donnell (2004) applied this concept using a stochastic meta-frontier framework and O'Donnell et al (2008) extended it using the DEA. According to this study, meta-frontier approach will allow the assessment of technical efficiency of decision-making units considering heterogeneity of production technology. Studies applying the metafrontier approach were conducted in various context. Chen, Ni and Liu (2022) applied a two-stage network slacks-based measure (NSBM) approach into a three-hierarchy meta-frontier framework to evaluate the value chain's innovation efficiency and the technology heterogeneity under the industrial chain structure. Alem et al (2019) estimated technical efficiency and technological gap ratios of Norwegian dairy farms using stochastic metafrontier approach developed by Huang, Huang and Liu (2014), to account for regional heterogeneity and true random-effect model to account for farm effects.

## 3 RESEARCH METHODOLOGY

In the study we analyse the effect of the membership in the PO on technical efficiency (TE) of their members. We assume positive effects of the membership in the newly established PO on their members' performance and persistence of this membership, although with differences by production specialisation. To assess the membership effect on farm TE by specialisation, we employ a DEA based two stage meta-frontier approach.

We used data of 645 farms for year 2014 from the database of the Slovak Ministry of Agriculture (IS MoA SR, 2018). The average farm utilised agricultural area (UAA) was 1342 ha. We distinguished three groups of agricultural producers by their PO membership over the 2007-2013 RDP, namely: the group 0 - the non-members of the PO; the group 1 consisted of the PO long-term members; group 2 comprised the members of newly established PO receiving the RDP support (new members). Furthermore, we divided the farms in the sample according to their production specialisation to those specialised to livestock or crop production, if more than 60% of their revenue comes from the livestock or crop production, respectively. Farms unspecialised are those with diversified production. Only one 2007-2013 RDP call was launched in Slovakia for the new PO project submission in 2008. The successful PO projects obtained financial transfer over next 5 years, therefore we assess farm performance in 2014.

We assume that within each PO membership farm group within production specialisation, inefficient farms have a potential to improve their performance by improvement of their low managerial and scale efficiency and farm performance is affected by their membership in the PO. In the 1st stage we estimate the overall TE of farms by specialisation and their PO membership, not considering differences in their managerial and scale inefficiency. The farm intra membership group-specific efficiency we assess by the input oriented CCR DEA model (Charnes, Cooper and Rhodes, 1978), assuming constant returns to scale (CRS).

$$TE^k(x,y)_{\theta,\lambda} \;=\; min\{\theta|\; \mathbf{Y_k}\,\boldsymbol{\lambda} \geq \mathbf{y_o},\; \mathbf{X_k}\boldsymbol{\lambda} \;\leq \theta\mathbf{x_o}, \boldsymbol{\lambda} \geq \mathbf{0}\} \qquad (1)$$

Where: $\boldsymbol{y_o}$ is the $M$×1 vector of output quantities for the farm under observation o; $o = 1, \dots, L_k$ $\boldsymbol{x_o}$ is the N×1 vector of input quantities for the farm under observation; $\mathbf{Y_k}$ is the $M \times L_k$ matrix of output quantities for all $L_k$ farms; $\mathbf{X_k}$ is the $N \times L_k$ matrix of input quantities for all $L_k$ farms; $\boldsymbol{\lambda}$ is a $L_k \times 1$ vector of intensity variables; $\theta$ is a scalar of TE measure; k =3 groups of farms with alternative PO membership.

Optimal solution to the DEA model (1) is $\hat{\theta}$ and $\hat{\lambda}$. If $\hat{\theta} = 1$, then the farm under observation is efficient. If $\hat{\theta} < 1$, the farm is inefficient and can become efficient if all inputs are proportionally reduced by $\hat{\theta}$. Model (1) we employ separately for three farm groups with alternative PO membership for each specialisation. Illustration is presented in the Figure 1A. Technical efficiency scores estimated for each farm within each farm group, by their specialisation and the PO membership, are used to calculate adjusted (target) group-specific values of inputs **x***.

$$\mathbf{x}^* = \hat{\theta}\mathbf{x_o} \qquad (2)$$

If constant returns to scale are assumed, calculation of adjusted values represents elimination of two inefficiencies: inefficiency due to poor management and inefficiency due to non-optimal farm scale (Charnes, Cooper and Rhodes, 1981). In the stage two we assess the meta (intergroup) efficiency. The metafrontier is constructed using a DEA model based on the pooled adjusted data of all farms, disregard of their PO membership, for each production specialisation separately.

$$TE^*(x^*,y) = \min\{\theta^*|\mathbf{Y}\boldsymbol{\lambda}^* \geq \mathbf{y_o}, \mathbf{X}^*\boldsymbol{\lambda}^* \leq \theta^*\mathbf{x}^*_\mathbf{o}, \boldsymbol{\lambda}^* \geq \mathbf{0}\} \qquad (3)$$

Where: $\mathbf{X}^*$ is the N×L matrix of adjusted input quantities for all L farms; $\mathbf{Y}$ is the M×L matrix of output quantities for all L farms; $\boldsymbol{\lambda}^*$ is a L ×1 vector of intensity variables; $\theta^*$ is a scalar of TE measure; $L = \sum_k L_k$ is number of farms.

Optimal solution to the DEA model (3) is $\theta^*$and $\boldsymbol{\lambda}^*$. If $\theta^* = 1$ then the farm under observation is efficient. If $\theta^* < 1$, the farm is inefficient. Applying model (3) on merged metadata of all three membership groups within each specialisation, we estimate technical efficiency of farms relative to TE of the best practice farms on metafrontier (Figure 1B). Deviation of inefficient farms from metafrontier now can be attributed to differences in their PO membership. To assess PO membership effects, we compare the distributions of the metafrontier technical efficiency scores between the three membership groups.

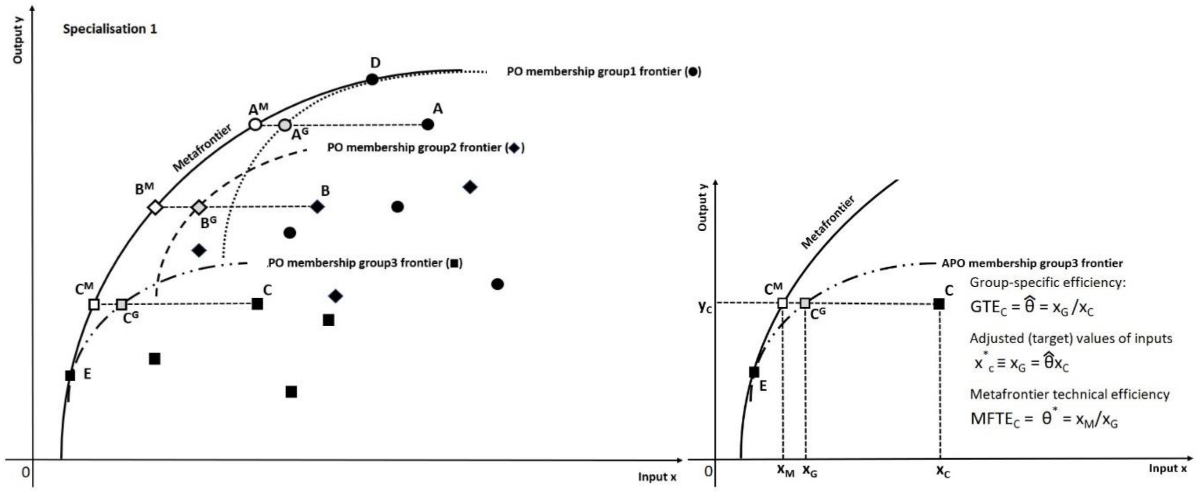



Figure 1 A) Group-specific frontiers and a meta-frontier. B) Estimation of group-specific efficiency and a metafrontier technical efficiency

Table 1. Frequency and average area of farms by specialisation and PO membership duration

| Production Specialisation | Number (%) of Farms by PO Membership | | | | Average UAA LPIS in ha by PO Membership duration | | | |
|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | Total | 0 | 1 | 2 | Average |
| Unspecialised | 30 | 19 | 59 | 108(17) | 1132 | 1710 | 2290 | 1867 |
| Crop | 130 | 44 | 152 | 326(50) | 947 | 1088 | 1542 | 1245 |
| Livestock | 85 | 48 | 78 | 211(33) | 1025 | 1408 | 1324 | 1222 |
| Total/Average | 245(38) | 111(17) | 289(45) | 645 | 997 | 1335 | 1636 | 1338 |

Note: 0 – non-members, 1- long-term members; 2- newly established PO members; UAA – Utilised Agricultural Area, LPIS – Land Parcel Identification System

Metafrontier technical efficiency score distributions and differences by PO membership groups were examined using the Shapiro-Wilk test, Levene's test, Welch's t-test, Kruskal-Wallis rank test and post hoc Dunn's test (Brockett and Golany,1996; Sueyoshi and Aoki, 2001). We used following variables: utilized agricultural area (in hectares); DEA output variable - total sales; DEA input variables - material and energy costs, land rent paid, wages paid, and total fixed assets (all in thousands of euros). Table 1 presents frequency and average size of farms by their PO membership and production specialisation.

## 4 RESULTS AND DISCUSSION

Traditional methods of farm performance assessment assume homogeneity of production technologies. Farm membership in producers' organisations as an alternative heterogeneous technology can contribute to farm performance differently in farms with the same production specialisation. The meta-frontier approach can solve the incomparability and allow to assess the alternative membership effect on farm technical efficiency. The crop farms dominated in our sample and in the subgroup of new members and non-members. Sample consisted of large commercial farms with 1338 ha of UAA on average. The average size of members of the supported POs ere even higher (Table 1). Since the results for specialized farms are analogous, we pay attention to the interpretation of the efficiency estimates for crop farms.

The average group-specific technical efficiency (GTE) was calculated relative to frontiers by farm specialisation, disregarding differences in technologies following from farm membership in the POs. They are not comparable. GTE represent managerial efficiencies (Charnes, Cooper and Rhodes, 1981), and indicate potential feasible farm productivity improvement within group by membership in PO. The most managerially efficient were farms with long-term membership in the PO, disregards their production specialisation. Their GTE score showed also low variability, indicating high competition within all specialised groups. Specialised farms, members of supported POs, exhibited in 2014 the lowest average managerial efficiency.

In the second stage of calculation, after the elimination of essential sources of inefficiencies (managerial and scale ones), the residual represents farm technical efficiency that can be attributed to the effect of farm membership in the PO. Metafrontier technical efficiency (MFTE) were calculated within each production specialization relative to the best practicing farms on the common metafrontier. Thus, MFTE expresses how efficient farms with different PO membership would perform under the current best technology. The average MFTE is comparable within production specialisation.

Farms specialized in crops exhibited on average the highest MFTE (0.869), followed by unspecialized farms (0.808) and livestock farms (0.6365) (Table 2, Figure 2). We found significant MFTE scores differences (p-value less than 0.01) between farm groups according to their membership in the PO (Welch's t-test, Kruskal-Wallis test) within all three production specializations. Further pairwise comparison (Dunn's test, Welch's test) revealed significant differences in MFTE score among members of all three farm groups by their PO membership for crop as well as unspecialised farms.

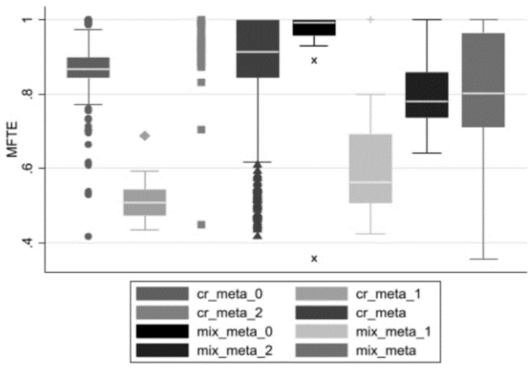



Figure 2. Distribution of crop and unspecialised farm technical efficiency score by membership in POs

Note: Cr – crop farms, Mix – unspecialised farms; 0 – non-members, 1- long-term members; 2- members of newly established PO. Source: own calculations

Members of supported POs revealed the significantly higher MFTE score in both crop (MFTE = 0.980), and livestock farms (MFTE = 0.999). The technical efficiency of these farms, however, was affected by managerial and scale inefficiencies (GTE). Long-term membership did not contribute significantly to farm average technical efficiency in 2014 compared to non-members.

More than half of the members of newly established POs were specialised on crop production. Out of 289 members of newly established POs, 66% quit their membership in 2014 when the RDP support to the POs terminated (Table 3). Most of these farms were crop specialised. Their behaviour demonstrates a weaker form opportunism by violation of relational norms (Luo, 2006; Williamson, 1993).

**Table 2:** Metafrontier technical efficiency by PO membership for each production specialization in 2014

| Spec PO Mem | Number of farms | Metafrontier technical efficiency MFTE Avg. | Std. Dev | Kruskal-Wallis $\chi^2$ | Dunn's post-hoc PO Membership 0 | 1 |
|---|---|---|---|---|---|---|
| Crop | | | | | | |
| 0 | 130 | 0.8606 | 0.0926 | 227.8*** | | |
| 1 | 44 | 0.5118 | 0.0493 | | 5.8727 *** | |
| 2 | 152 | 0.9802 | 0.0593 | | -11.2 *** | -13.8 *** |
| All | 326 | 0.8693 | 0.1686 | | | |
| Unspecialized | | | | | | |
| 0 | 30 | 0.9570 | 0.1166 | 56.9*** | | |
| 1 | 19 | 0.6067 | 0.1404 | | 7.3297 *** | |
| 2 | 59 | 0.7967 | 0.0905 | | 5.2963 *** | -3.6447 *** |
| All | 108 | 0.8078 | 0.1579 | | | |

Note: * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01; $\chi^2$ - Kruskal-Wallis H test, 2 d.f.; Spec PO Mem - farm PO membership: 0 – non-members, 1- long-term members; 2- members of newly established PO. Farm specialization (crop, unspecialized); All – all farms of particular specialization in the sample.
Source: own calculations based on IS MoA SR (2018) data

**Table 3:** Share and frequency of farms by specialization and membership in PO in 2014 (frequency)

| Specialisation | Membership in PO 0 | 1 | 2 | Total | Members of newly established PO in 2014 Members | Non-members |
|---|---|---|---|---|---|---|
| Total | 38.0 (245) | 17.2 (111) | 44.8 (289) | 100 (645) | 31.8 (99) | 68.2 (190) |
| Crop | 39.9 (130) | 13.5 (44) | 46.6 (152) | 100 (326) | 29.1 (53) | 70.9 (99) |
| Livestock | 40.3 (85) | 22.7 (48) | 37.0 (78) | 100 (211) | 31.8 (21) | 68.2 (57) |
| Unspecialised | 27.8 (30) | 17.6 (19) | 54.6 (59) | 100 (108) | 39.8 (25) | 60.2 (34) |

Notes: Farm PO membership: 0 – non-members, 1- long-term members; 2- newly established PO members.

## 5 CONCLUSIONS

Assessment of production units' performance using traditional methods assume homogenous production technologies. The farm option for membership in the POs within production specialisation can be assumed as different production technologies affecting their performance. Producers' organisations assist farms mainly in joint placing of goods on the market, to assure the quality of products, to centralise of sales and supply to improve competitiveness of the PO members. To assess the effect of this membership in the PO on farm technical efficiency we used a DEA based two stage metafrontier approach. Technical efficiency gained in the first stage allows to assess unobserved, mainly managerial and scale inefficiencies. By elimination of this inefficiency in the second stage, deviation of technical efficiency from metafrontier, can be attributed to the effect of farm membership in the PO.

We found that membership in newly established POs, supported from the RDP, contributed significantly to crop and livestock specialised farm technical efficiency in 2014. Their performance however was reduced by managerial and scale inefficiency. Crop farms demonstrated opportunistic behaviour when the support to POs terminated. To avoid inefficient use of public funds, financial support of producer cooperation should be further adjusted and better targeted to the maintenance of the viability of producer organisations.

**Acknowledgments:** Authors acknowledge the financial support of the Slovak Scientific Grant Agency VEGA 1/0845/17.

**References**

[1] Adjin, K. Ch., Goundan, A., Henning, Ch. H.C.A. and Sarr, S. (2020). Estimating the impact of agricultural cooperatives in Senegal: Propensity score matching and endogenous switching regression analysis. *Working Papers of Agricultural Policy, No. WP2020-10. Kiel University* [online]. Available at: http://hdl.handle.net/10419/235900 (Accessed: 14 November 2021)

[2] Alem, H., Lien, G., Hardaker, J. B. and Guttormsen, A. (2018). Regional differences in technical efficiency and technological gap of Norwegian dairy farms: a stochastic meta-frontier model. *Journal Applied Economics*, 51*(4*), pp. 409-421. https://doi.org/10.1080/00036846.2018.1502867

[3] Battese, G. E. and Rao, D. S. P. (2002). Technology Gap, Efficiency and a Stochastic Metafrontier Function. *International Journal of Business and Economics*, 1, pp.87-93.

[4] Battese, G. E., Rao, D. S. P. and O'Donnell, C. J. (2004). A Metafrontier Production Function for Estimation of Technical Efficiencies and Technology Gaps for Firms Operating under Different Technologies, *Journal of Productivity Analysis*, 21, pp. 91-103. https://doi.org/10.1023/B:PROD.0000012454.06094.29

[5] Bijman, J., Iliopoulos, C., Poppe, K., Gijselinckx, C., Hagedorn, K., Hanisch, M., Hendrikse, G.W.J., Kühl, R., Ollila, P., Pyykkönen, P. and Van Der Sangen, G. (2012). *Support for Farmers' Cooperatives.* Report prepared for the European Commission. Wageningen UR [online]. Available at: https://library.wur.nl/WebQuery/wurpubs/fulltext/245008 (Accessed: 6 March 2018)

[6] Brockett, P. L. and Golany, B. (1996) Using rank statistics for determining programmatic efficiency differences in data envelopment analysis. *Management Science, 42*(3), pp. 466–472. doi: 10.1287/mnsc.42.3.466

[7] Hayami, Y. (1969). Sources of agricultural productivity gap among selected countries. *American Journal of Agricultural Economics*, 51(*3*), pp. 564-575.

[8] Hayami, Y. and Ruttan, V.W. (1970). Agricultural productivity differences among countries. *The American economic review*, 60(*5*), pp. 895-911.

[9] Charnes, A., Cooper, W.W. and Rhodes, E. (1978). Measuring the Efficiency of Decision- Making Units. *European Journal of Operational Research*, 2, pp. 429–444. doi: 10.1016/0377-2217(78)90138-8

[10] Charnes, A., Cooper, W.W. and Rhodes, E. (1981). Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to program follow through. *Management Science*, 27(*6*), pp. 668-697. doi: 10.1287/mnsc.27.6.668

[11] Copa-Cogeca (2015). Development of agricultural cooperatives in the EU 2014. Brussels [online]. Available at: http://www.copa-cogeca.eu/ (Accessed: 20 May 2018)

[12] Cronin, E., Selten, M., Galen, M.A., Bijman, J., Viaggi, D., Arevalo, I., Smith, A.B., Ruijs, M.N.A., Meulen, B.M.J. and Vollaro, M. (2018). *Study on Producer Organisations and their activities in the olive oil, beef and veal and arable crops sectors.* Brussel [online]. Available at: https://edepot.wur.nl/456160 (Accessed: 14 November 2021)

[13] Fałkowski, J. and Ciaian, P. (2016). *Factors Supporting the Development of Producer Organizations and their Impacts in the Light of Ongoing Changes in Food Supply Chains. JRC Technical Report*. doi: 10.2791/21346

[14] Ferrier, G. D. and Porter, P. K. (1991). The productive efficiency of US milk processing cooperatives, *J. Agric. Econ.* 42(*2*), pp.161–174. doi: 10.1111/j.1477-9552.1991.tb00344.x

[15] Guzmán,I., Arcas, N., Ghelfi, R. and Rivaroli, S. (2009). Technical efficiency in the fresh fruit and vegetable sector: a comparison study of Italian and Spanish firms. *Fruits,* 64(*4*), pp. 243-252. doi: 10.1111/j.1467-8292.2007.00354.x
[16] Hoken, H. and Su, Q. (2015). Measuring the effect of agricultural cooperatives on household income using PSM-DID: A case study of a rice-producing cooperative in China. Institute of Developing Economies, *Japan External Trade Organization (JETRO*) [online]. Available at: https://ideas.repec.org/p/jet/dpaper/dpaper539.html (Accessed: 14 November 2021)
[17] Huang, C. J., Huang, T.H. and Liu, N.H. (2014). A New Approach to Estimating the Metafrontier Production Function Based on A Stochastic Frontier Framework. *Journal of Productivity Analysis,* 42(*3*), pp. 241–254. https://doi.org/10.1007/s11123-014-0402-2
[18] IS MoA SR. (2018). *Database Information Sheets* of the Ministry of Agriculture and Rural Development of the Slovak Republic.
[19] Luo, Y. (2006) Opportunism in Inter-Firm Exchanges in Emerging Markets, Management and Organization Review, 2, pp. 121–47. https://doi.org/10.1111/j.1740-8784.2006.00032.x
[20] McCorriston, S. (2002). Why Should Imperfect Competition Matter to Agricultural Economists? *European Review of Agricultural Economics,* 29, pp. 349-372. https://doi.org/10.1093/eurrag/29.3.349
[21] O'Donnell, C.J., Rao, D.S.P. and Battese, G.E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics*, 34(*2*), pp. 231–255.
[22] Reardon, T., Barrett, C.B., Berdegué,J.A., and Swinnen, J.F.M. (2009). Agrifood industry transformation and small farmers in developing countries. *World development*, 37(*11*), pp. 1717-1727. https://doi.org/10.1016/j.worlddev.2008.08.023
[23] Skevas, T. and Grashuis, J. (2019). Technical efficiency and spatial spillovers: Evidence from grain marketing cooperatives in the US Midwest. *Agribusiness*. 2019; pp. 1–16. doi: 10.1002/agr.21617
[24] Sueyoshi, T. and Aoki, S. (2001). A use of a nonparametric statistic for DEA frontier shift: the Kruskal and Wallis rank test. *Omega*. 29(*1*), pp. 1–18. doi: 10.1016/S0305-0483(00)00024-4
[25] Van Herck, K. (2014). *Assessing efficiencies generated by agricultural Producer Organisations*. doi: 10.2763/76733
[26] Williamson, O. E. (1993). Opportunism and its Critics. *Managerial and Decision Economics*, 14(*2*), pp. 97–107. https://doi.org/10.1002/mde.4090140203

**Authors' address**
prof. Ing. Lubica Bartova, CSc.
doc. Ing. Peter Fandel, CSc.
Slovak University of Agriculture in Nitra, Institute of Statistics, Operations Research and Mathematics, Faculty of Economics and Management
tr. A. Hlinku 2, 949 76 Nitra
Slovak Republic
email: lubica.bartova@uniag.sk, peter.fandel@uniag.sk

# SET-COVERING BASED FORMULATION OF THE DYNAMIC CONTAINER DRAYAGE PROBLEM WITH MODULAR CONCEPT AND LONGER COMBINATION VEHICLES

***Nenad Bjelić, Milorad Vidović, Branislava Ratković***
***University of Belgrade – Faculty of Transport and Traffic Engineering***

***Abstract*** Consideration of only a static container drayage problem leads to good start of controlling such systems. However, since drayage is realized in a stochastic surrounding, chances that initial conditions could be applied until the end of a planning period are minimal. Accordingly, some sort of dynamic controlling mechanism has to be applied. In this paper we propose a new mathematical formulation for solving the dynamic container drayage problem with longer combination vehicles and modular concept. Numerical tests showed model's potential for the implementation in a rolling horizon approach, especially for the case when one chassis vehicle is used.

***Keywords:*** *Container drayage, Mixed integer linear programming, Dynamism, Rolling horizon*

***JEL Classification:*** C61
***AMS Classification:*** 90B06, 90C11, 90C90

## 1. INTRODUCTION

Distribution and collection of containers from and to a container hub by the utilization of trucks is known as the drayage. Although the drayage distances are considerably shorter compared to the distances between intermodal hubs, due to the specificity of the processes, their ratio in the overall intermodal transportation costs is disproportionally larger. According to Macharis and Bontekoning (2004) 25 to 40% of overall intermodal costs are related to the drayage operations. In that sense, it is obvious that any improvement in the drayage process can significantly influence the attractiveness of the intermodal transport. Moreover, Morlok et al (1995) state that drayage activities have the largest potential for reducing overall cost of the intermodal transport. Of course, not only financial benefits of the drayage improvements are noticeable since ecological aspect of the transportation of heavy load by trucks cannot be neglected.

Usually, the distribution of containers implies that a container (mostly of 20 or 40 feet in the length) is loaded onto a chassis, transferred to the destination and unloaded at the client's location. In the case when there are no possibilities for unloading the container, loaded chassis (with or without a truck) is left at the location until the container content is (un)loaded. Collection of containers is executed in the reverse order. Unloaded container is collected and returned to the hub location for realization of maintenance activities (e.g., cleaning, disinfection, small repairs etc.), or, if it satisfies all requirements of the next goods, it is transferred to the location of next client. During the drayage operations containers on chassis can be either loaded with the content, or empty. It should be noted that, until recently, the legislation of the most world countries allowed transportation of only one container on a chassis. Nevertheless, the drayage improvement potential, as well as the development in chassis construction, which provided better spread of container weight on the road surface, forced governments to allow modular concept of transportation, i.e., transportation of two 20-ft container on a single chassis.

Moreover, some of them, like the USA, Sweden, Finland and Australia (Vidović et al, 2017) now allow employment of longer combination vehicles transportation, which implies that more than one chassis can be towed by a truck, where all chassis apply modular concept. Modular and longer combination vehicles concepts (Vidović et al, 2017; Funke and Kopfer, 2016), the possibility of transferring emptied containers directly to the next client without returning to the depot(Breakers, Caris and Janssens, 2013; Zhang, Yun and Kopfer, 2015), the concept of unloading and loading containers at clients premises instead of waiting for their (un)loading(Breakers, Caris and Janssens, 2013; Zhang, Yun and Kopfer, 2015; Vidović et al, 2017), use of platooning concept of truck driving (You et al, 2020), and consideration of tasks for multiple successive periods (Clemente et al, 2017; Fanti et al, 2019; Ghezelsoflu et al, 2021) are just some of concepts proposed in the literature regarding the improvement of the drayage process.

Regardless of the used concept the common feature of the vast majority of researches is that they consider the static version of the problem, i.e., they all consider the case in which all relevant data are known at the beginning of the planning horizon and that during the realization of the process there are no disturbances that could lead to the changes that may affect the provided operational plan. However, since drayage takes place on the road network, that waiting at the clients' premises might differ from time to time, that breakdowns may occur, that new tasks could arise, or some of the existing could change during the planning horizon, it is obvious that there is a need for the dynamic approach in managing the drayage activities.

So far there are several papers that treat the dynamic nature of the drayage operations. In the basic paper that treats dynamism in transportation tasks Pillac et al (2013) define the class of dynamic problems treated in this research as problems in which all or a part of task information is dynamically reviled during the realization of on-going operational plan. They classify such problems as the dynamic and deterministic problems, but they are also referred to in the literature as on-line and real-time problems. According to Máhr et al (2010) the usual sources of dynamism in vehicle routing problems include job arrival (or cancelation), changes in task duration, and resource failures. However, sources of dynamism in the drayage process that are considered so far in the literature are more specific and include orders' locations (Zhang, Smilowitz and Erera, 2011; Zhan, Lu and Wang, 2014), container packing and unpacking time (Shiri, Ng and Huynh, 2019), transit time (Escudero et al, 2013) and new job arrivals and container service time (Máhr et al, 2010). Usually, only one or two sources of dynamism are treated in researches. To the best of our knowledge only Bjelić et al (2022) presented an approach capable of dealing more than two sources of dynamism. Therefore, this research is the extension of the research conducted in Bjelić et al (2022) in the sense that it presents additional mixed integer programming model based on the set covering concept.

In that sense, the rest of the paper is organized in the following order. In Section 2 we give a detailed problem description while in Section 3 we present the set covering based mathematical model. In Section 4 we test the model on a set of test instances and comment the most interesting results. Eventually, we give the concluding remark with the directions for future work.

## 2. PROBLEM DESCRIPTION

We considered a problem from a perspective of a decision maker in a drayage company in a moment when an actual operational plan is no longer applicable because of changes that occurred in the controlled system. In that situation a reassignment of a set of drayage tasks to a set of vehicles, and their sequencing, must be undertaken, with respect to all relevant characteristics and constraints of the problem. Eventually, a new operational plan defines a sequence of tasks that each vehicle realizes in the following period in such a way that objective function is optimized.
Each task for reassignment is defined by a task type, location, expected service duration (for the realization of all administrative and (un)loading activities) and a time window during which it is possible to start its service. However, it should be mentioned that we considered only the earliest time of a time window interval as the hard constraint, while the right-hand side of a window is treated as a soft one, meaning that it can be broken but with the cost of paying additional fees to the client. There are four types of tasks defined by the size of a container included in a task (20-ft or 40-ft), and an operation that has to be undertaken (loading or unloading a container). From the perspective of a vehicle, where by a vehicle we consider a truck trailing one or two chassis for modular drayage operations, loading of a container reduces available capacity of chassis by the length of the container, while unloading of a container increases the available capacity. Accordingly, set of considered task types is {-20, +20, -40, +40}. Set of tasks to be assigned to vehicles consists of all tasks that occurred after the last assignment, and from all tasks assigned in the previous planning period whose service has not started in the task's location until the moment of re-optimization. However, it should be mentioned that unload task types from previous planning period whose containers are already on a vehicle are not subjects of the assignment. Nevertheless, they are subjects of the optimization process in the sequencing part of the task.

The rest of presumptions on which the modeling is based are:

- All emptied picked-up containers are returned to the depot for the maintenance tasks
- Containers are (un)loaded at tasks' locations, i.e., vehicles do not wait for emptying or fulfilling containers, nor there is a possibility of separating chassis from towing vehicles.
- Set of vehicles consists of vehicles already executing drayage tasks (we refer them as non-depot vehicles – NDVs) and vehicles in the depot (referred as depot vehicles – DVs)
- Number of DVs is unbounded, i.e., it is supposed that there are sufficiently large number of them in the depot. All of them start from the depot with the same number of chassis with modular structure.
- Maximally two chassis with modular structure is considered, i.e., maximum capacity of a vehicle is 80-ft
- A location, an active task and its status (travelling to the task's location, waiting at the task's premise for the start of service, or at the service) are known for all NDVs.
- NDV's capacity depends on the realization of an active sequence of tasks, i.e., by lengths of picked-up containers that must be transferred to the depot, and by the length of containers to be unloaded at locations of unfinished tasks of active task sequence.
- A sequence of task types that can be assigned to a vehicle is defined by using the load state (LoSt) concept defined in Bjelić et al (2022).

# 3. SET-COVERING FORMULATION OF THE PROBLEM

This formulation is based on the formulation of the multisize container drayage problem given in Vidović et al (2017) in which the main idea is to find such matching of variables that minimizes objective value and satisfies all constraints. Each variable represents a feasible route *k* of a vehicle *v*, respecting its capacity, feasible order of pickup and delivery tasks, and possible existence of delivery tasks on NDVs, remained from an ongoing operational plan. All features the defined model are taken into consideration in the part of defining a set of available routes, i.e., a set of model variables. For the model formulation we used the following notation:

*Sets*
$V$ – set of vehicles
$T$ – set of tasks
$K_v$ – set of feasible routes for vehicle *v*

*Parameters*
$c_k$ –cost of route *k*
$s_t$ – service time of task *t*
$d_{k_{j,j+1}}$– travel time from task on a position *j* in route *k*, to task on a position *j*+1 in route *k*

*Variables*

$$Y_{v,k} = \begin{cases} 1 - \text{if vehicle } v \text{ traverses according to the route } k \\ 0 - \text{otherwise} \end{cases}$$

$B_{k_j}$– service start time for task $k_j$ , where *j* is the position of task node in route *k*
$D_t$- lateness of service start (from the right-hand side time window boundary) for task node *t*

According to the definition of variables $Y_{v,k}$ it is obvious that their generation might be challenging and that it depends from the initial state of a vehicle, in terms of starting location, capacity and previously assigned tasks, as well as from a set of unassigned tasks. In defining sets of vehicles' feasible routes, we used a load state concept, given in more detail in Section 2.2 in Bjelić et al (2022). With the given notation a set covering based formulation can be defined as:

$$min \rightarrow \sum_{v\in V}\sum_{k\in K_v} c_k \cdot Y_{v,k} + W \cdot \sum_{t\in T} D_t \tag{1}$$

Subject to

$$D_t \geq B_t - l_t \qquad \forall t \in T \tag{2}$$

$$\sum_{v\in V}\sum_{k\in K_v \wedge t\in k} Y_{v,k} = 1 \quad \forall t \in T \tag{3}$$

$$\sum_{v\in V}\sum_{k\in K_v} Y_{v,k} \leq 1 \quad \forall v \in V \tag{4}$$

$$B_{k_{j+1}} \geq B_{k_j} + s_{k_j} + d_{k_{j,j+1}} - M \cdot (1 - Y_{v,k}) \quad \forall k \in K_v, \forall v \in V, j \in [2, |k| - 1] \tag{5}$$

$$e_t \leq B_t \quad \forall t \in T \tag{6}$$

$$D_t \geq 0 \quad \forall t \in T \tag{7}$$

$$Y_{v,k} \in \{0,1\} \quad \forall v \in V, \forall k \in K \tag{8}$$

Objective function minimizes the sum of the time required to finish all tasks and the weighted sum of disobeying time windows. $W$ is a sufficiently large number to force the model to prefer satisfaction of time windows over the minimization of distances (or some other cost type parameter). Constraints (2) and (7) determine delays of tasks. Constraints (3) provide that every task can be found in only one route, while constraints (4) provide that a feasible assignment of tasks could be assigned to a vehicle but it does not have to. Constraints (5) are tracking values of tasks' service start times on routes. Constraints (6) forbid the beginning of a service prior to task's time window left boundary. Constraints (8) define the binary nature of the variable $Y_{v,k}$.

## 4. NUMERICAL EXPERIMENT

Since this research is the extension of the previous research, we used the test instances set from that research, and expanded it with new instances with the goal of performing more exhaustive comparing of linear formulations of the considered problem. Of course, due to the problem's complexity we considered only the small size problem instances, i.e., problems of only 10 and 15 tasks. Also, it should be emphasized that all additional tests are made in accordance to the generation of the test set described in the mentioned paper. The set of test instances, as well as the technical version of the paper are available upon an interested reader's request to the corresponding author.

Each instance is solved once for the case of the network flow formulation (NFF), given in Bjelić et al (2022), and once for the case of the given set covering formulation (SCF) of the problem. Mathematical formulations are solved by Gurobi 9.1 MILP solver with the implementation of appropriate Python based API. Testing is done as a single thread on an Intel Core i5-347 CPU 3.2GHz configuration with 8GB of RAM on a 64 bits Windows 10 OS.

Results regarding the time aspect of utilizing NFF and SCF are given in Table 1. First column contains information regarding the size of instances; the second denotes the number of considered depots; while the third one contains the number of considered instances. Columns four and five contains an average CPU time required for obtaining optimal solutions in the case of NFF and SCF, respectively, for the chassis capacities of 40 feet. Similarly, columns six and seven contain data regarding 80 feet capacity. It should be noted that due to the increased problem sizes of the SCF formulation in the cases of 15 tasks, for four instances GUROBI solver reported the Out of Memory status. Data for those instances are excluded from the analysis. Classes of such instances are denoted by a superscript number representing the number of instances with the Out of Memory status.

As it can be seen from Table 1 for both NFF and SCF the number of depots (number of depots includes the depot and all NDVs as virtual depots with one available vehicle) in a problem does not play an important role in a time required for obtaining an optimal solution. On the other hand, the number of tasks and the chassis' capacities do, in the sense that with the increase in the number of tasks, or in the chassis capacity, CPU time increases as well. The reasons for that are quite straightforward since there are more assignments of tasks to vehicles to be considered. Nevertheless, it should be noted that for the cases of 40 feet chassis capacity, SCF performs significantly time efficiently.

Table 1. Comparison of average CPU time required for obtaining optimal solutions

| | | | Average of CPU time [s] | | | |
|---|---|---|---|---|---|---|
| | | | Chassis' capacities | | | |
| | | | 40-ft | | 80-ft | |
| Task number | Depots number | Instance number | NFF | SCF | NFF | SCF |
| 10 | 1 | 4 | 35.10 | 0.31 | 15.65 | 16.60 |
| | 2 | 1 | 2.08 | 0.38 | 4.96 | 13.79 |
| | 3 | 4 | 1.95 | 0.60 | 0.81 | 19.61 |
| | 4 | 10 | 3.05 | 0.66 | 0.69 | 14.67 |
| | 5 | 1 | 2.69 | 0.47 | 1.14 | 4.16 |
| | 6 | 2 | 2.17 | 0.54 | 1.13 | 4.00 |
| | 7 | 1 | 0.78 | 0.48 | 0.80 | 30.44 |
| | 8 | 1 | 51.23 | 1.32 | 1.53 | 36.34 |
| | 9 | 1 | 1.37 | 0.45 | 0.64 | 1.29 |
| 15 | 1 | 3[1] | 20.67 | 2.88 | 2871.66 | 923.74 |
| | 2 | 2 | 63.58 | 2.28 | 749.32 | 339.49 |
| | 3 | 2 | 18.37 | 3.29 | 583.42 | 852.91 |
| | 4 | 12[3] | 30.22 | 3.50 | 240.01 | 261.97 |
| | 5 | 1 | 32.58 | 3.90 | 304.37 | 765.86 |
| | 6 | 2 | 29.63 | 2.62 | 278.08 | 507.80 |
| | 7 | 1 | 107.78 | 2.42 | 9.38 | 145.11 |
| | 8 | 1 | 11.37 | 6.02 | 21.93 | 1077.97 |
| | 9 | 1 | 3.14 | 1.50 | 14.00 | 60.45 |

Table 2 Comparative overview of model sizes for the cases of NFF and SCF

| | Average number of constraints | | Average number of variables | | | |
|---|---|---|---|---|---|---|
| | | | Binaries | | All | |
| Task number | 10 | 15 | 10 | 15 | 10 | 15 |
| | Chassis' capacity of 40-ft | | | | | |
| NFF | 4287.48 | 11955.4 | 1243.2 | 3634.28 | 1764.08 | 4560.52 |
| SCF | 2697.92 | 16889.36 | 1002.16 | 5488.64 | 1030.08 | 5526.8 |
| | Chassis' capacity of 80-ft | | | | | |
| NFF | 1954.04 | 5150.32 | 563.88 | 1561.68 | 803.6 | 1963.64 |
| SCF | 76891.92 | 2618146 | 15363.8 | 450217.9 | 15391.72 | 450256 |

Table 2 contains data regarding the model sizes of considered formulations. As it can be seen, SCF is quite more sensitive to the increase of task number and the chassis capacity from both perspectives, i.e., number of constraints, as well as number of variables increases much more than number of tasks or chassis capacity. It must be emphasized that although for the NFF data show decrease in numbers of constraints and variables with the increase of the chassis capacities, that is just an impression. Namely, since number of available vehicles largely determines the number of variables and constraints in the NFF, due to vehicles' doubled capacity in 80 feet instances, i.e., larger number of tasks that can be handled in a route, their number is reduced in all 80 feet test instances. Eventually, the number of variables and constraints is also reduced.

## 5. CONCLUSION

In this research we considered the dynamic version of the container drayage problem with the implementation of the longer combination vehicles and modular concept of loading two 20-ft containers at the 40-ft chassis. For that problem we presented a new set covering based mathematical formulation of the problem and executed numerical experiments that compared it with the existing network flow based mathematical formulation. Testing results are executed on a set of small problem instances and from the results it can be seen that proposed formulation is time efficient in solving problems for the cases when overall capacity of vehicles is 40-ft. However, increase of the chassis capacity leads to significantly more possible assignments of tasks to routes which eventually lead to extremely large models. For example, in the case of 80-ft chassis capacity and 15 tasks four instances required more memory than it was available on a computer.

The proposed formulation may have significant role in the rolling horizon controlling approach of the drayage operations when re-optimizations are done in moments which provide that relatively small number of tasks has to be reassigned. In that sense, implementation of the rolling horizon approach in controlling drayage operations with the proposed formulation is one direction for future researches. The other direction is related to the general role of set covering formulations in developing column-based solution methods. More precisely, in order to avoid consideration of all possible assignments of routes to vehicles, as in the case of given formulation, only small, but wisely chosen, number of possible assignments should be considered. Application of such approach is provided by the shape of the given formulation in which there are considerably more variables (routes to vehicles assignments) then constraints.

**Acknowledgements:** This work was supported by the Ministry of education, science and technological development of the Government of the Republic of Serbia through the institutional funding of University of Belgrade - The Faculty of Traffic and Transport Engineering

## References

[1] Bjelić, N., Vidović, M., Popović, D. and Ratković, B. (2022). Rolling-horizon approach in solving Dynamic Multisize Multi-trailer Container Drayage Problem. *Expert Systems with Application*, 201, p. 117170.

[2] Braekers, K., Caris, A. and Janssens, G.K. (2013). Integrated planning of loaded and empty container movements. *OR Spectrum*, 35(2), pp. 457–478.

[3] Clemente, M., Fanti, M.P., Nolich, M., Stecco, G. and Ukovich, W. (2017). Modelling and solving the multi-day container drayage problem. In *2017 13th IEEE Conference on Automation Science and Engineering (CASE)*. 2017 13th IEEE Conference on Automation Science and Engineering (CASE), pp. 231–236.

[4] Escudero, A., Muñuzuri, J., Guadix, J. and Arango, C. (2013). Dynamic approach to solve the daily drayage problem with transit time uncertainty. *Computers in Industry*, 64(2), pp. 165–175.

[5] Fanti, M.P., Locatelli, A., Stecco, G. and Ukovich, W. (2019). A New ILP Formulation for the Multi-Day Container Drayage Problem. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), pp. 655–660.

[6] Funke, J. and Kopfer, H. (2016). A model for a multi-size inland container transportation problem. *Transportation Research Part E: Logistics and Transportation Review*, 89, pp. 70–85.

[7] Ghezelsoflu, A., Di Francesco, M., Frangioni, A. and Zuddas, P. (2021). A multiperiod drayage problem with customer-dependent service periods. *Computers & Operations Research*, 129, p. 105189.
[8] Macharis, C. and Bontekoning, Y.M. (2004). Opportunities for OR in intermodal freight transport research: A review. *European Journal of Operational Research*, 153(2), pp. 400–416.
[9] Máhr, T., Srour, J., de Weerdt, M. and Zuidwijk, R. (2010). Can agents measure up? A comparative study of an agent-based and on-line optimization approach for a drayage problem with uncertainty. *Transportation Research Part C: Emerging Technologies*, 18(1), pp. 99–119.
[10] Morlok, E.K., Sammon, J.P., Spasovic, L.N. and Nozick, L.K. (1995). Improving Productivity in Intermodal Rail-Truck Transportation. In P. T. Harker, ed. *The Service Productivity and Quality Challenge*. International Studies in the Service Economy. Dordrecht: Springer Netherlands, pp. 407–434.
[11] Pillac, V., Gendreau, M., Guéret, C. and Medaglia, A.L. (2013). A review of dynamic vehicle routing problems. *European Journal of Operational Research*, 225(1), 1.–11.
[12] Shiri, S., Ng, M. and Huynh, N. (2019). Integrated drayage scheduling problem with stochastic container packing and unpacking times. *Journal of the Operational Research Society*, 70(5), pp. 793–806.
[13] Vidović, M., Popović, D., Ratković, B. and Radivojević, G. (2017). Generalized mixed integer and VNS heuristic approach to solving the multisize containers drayage problem. *International Transactions in Operational Research*, 24(3), pp. 583–614.
[14] You, J., Miao, L., Zhang, C. and Xue, Z. (2020). A generic model for the local container drayage problem using the emerging truck platooning operation mode. *Transportation Research Part B: Methodological*, 133, pp. 181–209.
[15] Zhang, G., Smilowitz, K. and Erera, A. (2011). Dynamic planning for urban drayage operations. *Transportation Research Part E: Logistics and Transportation Review*, 47(5), pp. 764–777.
[16] Zhang, R., Lu, J.-C. and Wang, D. (2014). Container drayage problem with flexible orders and its near real-time solution strategies. *Transportation Research Part E: Logistics and Transportation Review*, 61, pp. 235–251.
[17] Zhang, R., Yun, W.Y. and Kopfer, H. (2015). Multi-size container transportation by truck: modeling and optimization. *Flexible Services and Manufacturing Journal*, 27(2), pp. 403–430.

**Author's address**
Nenad Bjelić, PhD
University of Belgrade, Faculty of Transport and Traffic Engineering, Department of Logistics
Vojvode Stepe 305, 11000 Belgrade
Republic of Serbia
email: n.bjelic@sf.bg.ac.rs

# MATHEMATICAL MODEL FOR OPTIMIZING STAFFING OF AIRCRAFT GROUND HANDLING

***Jakub Cilecek, Dusan Teichmann***
***VSB-Technical university of Ostrava***

**Abstract**
Presented article deals with optimizing staffing of ground handling process for aircraft at international regional airport´s apron. Solution of optimization problem was conducted in two phases. In the first phase of solution, time reserves were found for each activity of ground handling process. Said reserves provide space for creation of delay of any activity which can lead to decrease of number of workers which are needed for realization of whole process. Critical path method was used for the calculation of time reserves. In the second phase, minimal number of workers which are needed for process of ground handling was calculated. Mathematical programming was used for calculation of solution. In the presented article we used mixed linear mathematical model. Optimizing criterium in the mentioned model was total number of workers needed for realization of all activities of ground handling process. Value of optimizing criterium was minimalized. Calculation experiment was conducted in the conditions of international regional airport Ostrava.

***Keywords:*** *Aviation, Critical Path Method, Ground Handling, Mathematical Programming, Operational Research*

***JEL Classification:*** C69, C44
***AMS Classification:*** *90C10*

## 1 INTRODUCTION

Technical handling is process which deals with ground handling of aircrafts after their arrival and before theirs departure. Capacity of technical handling is mostly limited by restricted personal resources. This problem appears mainly at regional airports in summer season, when higher capacities for technical handling are needed. Better organization of work allows us to increase the use of staff and trouble-free course of technical handling process even with lower number of staff. This allows us to lower prices at regional airports and increase the competitiveness of said regional airports. This problem which is about increasing of usage of staff can be understand as optimisation task which can be solve with the methods of operation analysis.

## 2 CURRENT STATE ANALYSIS

The aircraft ground handling process at the airport, which also includes technical ground handling, is a complex and time-consuming process during which both the aircraft and passengers are provided with the required services (IATA 2018) and (Cervinka 2017). The total duration of the aircraft ground handling process depends on the airport technical equipment, the type of aircraft operated and the range of services required for the ground handling by the air transport operator.

The complexity of the entire ground handling process shows a need to remove time reserves, reduce the overall time required for ground handling, or simply, to make the process more efficient. Numerous approaches can be taken to achieve this. The key to the entire streamlining process involves, in many cases, the methods of managerial decision-making, specifically the methods of network analysis, and namely, CPM.

CPM (Critical Path method) is a basic deterministic method of a network analysis (i.e. project management). Every project (i.e., also the process of technical ground handling of aircraft) can be represented by an oriented chart, in which the nodes represent the beginnings and ends of individual activities, the edges represent the activities themselves and the evaluation of edges represents the duration of individual activities. One of the goals of CPM is to identify the so-called critical path, which is the longest sequence of consecutive activities starting at the starting node (node representing the beginning of the project) and ending at the end node (node representing the completion of the project), and its length represents the shortest possible duration of the solved project.

The activities that make up a critical path are called critical activities. Any extension of the duration of any of the critical activities will result in an extension of the total project implementation time. CPM allows you to quantify the time reserve for each activity. The critical activities are then all activities with zero time reserve. Detailed information on CPM can be found, for example, in (Fiala 2004), (Jablonsky 2007), (Kluson 1973) and (Sajdlerova and Konecny 2008).

The authors of the article (Andreatta et al. 2014) address the issue of increasing the efficiency of the aircraft handling process by assigning workers to handling equipment. The authors created a heuristic approach, GRAP (Ground - service Resource Allocation Problem). The method assigns workers to the equipment in the shortest possible time intervals. This eliminates staff downtime and possible flight delays. They obtained input data for their experiments at Berlin Tegel Airport, where the INFORM software is used to solve the problem, and the mathematical model created by the authors is only an extension of it. During testing, the authors achieved a reduction in times by an average of 7.5%.

The author of the paper (Kwasiborska 2010) analyzes the overall process of handling by ATR and Embraer type aircrafts. The handling process and the time required by individual activities in her work are represented using a Gantt chart. The article does not reject the hypothesis of the Poisson probability distribution, which models the input flow of aircraft into the system. Further, the author monitored the times of individual phases of handling – boarding and disboarding of passengers, loading and unloading of luggage, goods and mail, refueling, replenishment of catering and cabin cleaning. Based on the analyzed data, the author simulated a ground services model.

The authors of the work (Al-Bazi et al. 2016) investigated the possibilities of improving the aircraft handling process in the conditions of a specific low-cost air carrier Turkish Airline using CPM. The authors focus mainly on the passenger disembarkation and embarkation phases, which they consider critical and the most time-consuming. In the article, the authors define different types of flights and divide them into national, international and their combination according to their nature. Input data were obtained for 3 days by observation at the airport serving the carrier as a hub.

# 3 FORMULATION OF TASK AND PROPOSED MATHEMATICAL MODEL

## 3.1 Formulation of optimization task

In the presented paper, we follow up on the article (Teichmann and Dorda 2016), in which a linear mathematical model was proposed to solve the general problem of allocating staff to activities. However, the mentioned linear mathematical model was significantly modified for the needs of the solved problem.

The creation of a mathematical model is preceded by the creation of a network graph of the project and the acquisition of basic time characteristics of individual activities using CPM.

The objective function remains the same as in the previous model; the weights of the decision variables do not change either. In addition to the previous model, a constant $C_j$ is introduced for each activity $j \in N$, modeling the number of employees to be assigned to perform the given activities. Furthermore, a set of qualifications $P$ is introduced into the model, deviating from the previous model, and a constant $c_p$, representing the number of employees of the given qualification, is defined for each qualification $p \in P$. The competence of a qualification $p \in P$ to perform an activity $j \in N$ is modeled by means of an incidence matrix, $\boldsymbol{A}$, while if the activity $j \in N$ can be performed by an employee with a $p \in P$ qualification, then $a_{jp} = 1$, and in the opposite case, $a_{jp} = 0$.

To solve the problem, two groups of variables will be introduced into the model:

$x_{ijpl}$…bivalent variable representing the transfer of the worker $l$ with qualification $p$ from activity $i$ to activity $j$ (this, of course, occurs when both activities can be performed by a qualified worker $p$). When $x_{ijpl}$ = 1, the transfer of the given worker between activities will take place, when $x_{ijpl}$= 0, the transfer of the given worker between activities will not take place.

$z_i$…a non-negative variable representing the possible shift in the start time of activity $i \in N$.

## 3.2 Mathematical model

Mathematical model has the following form:

$$min\, f(x) = \sum_{j \in N} \sum_{p \in P} \sum_{l \in L_p} x_{0jpl} \tag{1}$$

Under the conditions:

$$\sum_{i \in N} \sum_{p \in P_j} \sum_{l \in L_p} a_{jp} x_{ijpl} = C_j \qquad for\ j \in N \tag{2}$$

$$\sum_{j \in N} \sum_{l \in L_p} x_{0jpl} \le c_p \qquad for\ p \in P \tag{3}$$

$$x_{ijpl} \le a_{jp} \qquad for\ i \in N \cup \{0\}; j \in N;\ p \in P;\ l \in L_p \tag{4}$$

$$b_{ij} \ge x_{ijpl} \qquad for\ i \in N \cup \{0\}; j \in N;\ p \in P;\ l \in L_p \tag{5}$$

$$\sum_{j \in N} x_{ijpl} \leq 1 \qquad for\ i \in N;\ p \in P; l \in L_p \qquad (6)$$

$$\sum_{i \in N \cup \{0\}} x_{ijpl} = \sum_{i \in N} x_{jipl} \qquad for\ j \in N;\ p \in P; l \in L_p \qquad (7)$$

$$\overline{t_i} + z_i + T_i \leq \overline{t_j} + z_j + M \cdot \left(1 - x_{ijpl}\right) \qquad for\ i \in N; j \in N;\ p \in P; l \in L_p \qquad (8)$$

$$z_i \leq \overline{\overline{t_i}} - \overline{t_i} \qquad for\ i \in N \qquad (9)$$

$$x_{ijpl} \in \{0,1\} \qquad for\ i \in N; j \in N;\ p \in P; l \in L_p \qquad (10)$$

$$z_i \geq 0 \qquad for\ i \in N \qquad (11)$$

Function (1) represents the number of employees deployed. The group of restrictive conditions (2) will ensure that the required number of staff is allocated to each activity $j \in N$. The group of restrictive conditions (3) will ensure that no more staff are used in each qualification group $p \in P$ to solve the task being addressed than is currently available. The group of restrictive conditions (4) ensures that if the activity $j \in N$ cannot be performed by a $p \in P$ qualified worker, then the worker of the given profession will not be deployed to perform the given activity. The group of restrictive conditions (5) will ensure that no $p \in P$ qualified worker $l \in L$ can return to an activity $i \in N$ after performing the activity $j \in N$; this condition is ensured by means of an incidence matrix $\boldsymbol{B}$. The group of restrictive conditions according to relation (6) will ensure that each worker $l \in L_p$ who can perform activities $i \in N$ and $j \in N$ will, upon completion of the activity $i \in N$, switch, at most, to one activity $i \in N$. The group of restrictive conditions according to relation (7) ensures the continuity of the worker $l \in L_p$, with qualification $p \in P_i \cap P_j$. This means that when a worker $l \in L_p$ with qualification $p \in P_i \cap P_j$ is engaged in an activity, he or she must also be removed from this activity after its completion. The group of restrictive conditions according to relation (8) prevents temporally inadmissible time transfers of workers. This means that if a qualification $p \in P_i \cap P_j$ worker $l \in L_p$ assigned to the activity $i \in N$ does not make start time of the activity $j \in N$, then the worker's transition between activities $i \in N$ and $j \in N$ will not occur (the symbol $M$ represents a prohibitive constant). The group of restrictive conditions according to relation (9) ensures that the time shift of the activity $i \in N$ can take place only in the interval delimited by the earliest possible time of the activity and no later than the permissible start of the activity. The groups of constraints (10) and (11) determine the domain of variables used in the model.

# 4 COMPUTIONAL EXPERIMENTS

Computational experiments with the proposed model were carried out during handling of charter flights operated with Boeing 737 aircraft at Leoš Janáček International Airport in Ostrava, Czech Republic (Ostrava Airport). The charter flights under consideration were handled without a stopover, so that no transit passengers remained on board during ground handling during their stay at Ostrava Airport. As the calculation relates to technical ground handling, it does not matter whether the flight took place inside or outside the Schengen area.

In the conditions of Ostrava Airport, one handling agent is assigned to the aircraft. Two exit steps were used to disembark passengers, two belt conveyors were used for unloading luggage, and only luggage, no cargo or mail, was loaded in the cargo area of the aircraft, and luggage was not loaded in containers. The luggage was loaded in the front and rear cargo area. Refueling of the aircraft was performed from a vehicle after departing of all passengers, so there was no need for assistance from firefighters during handling.

The following technological specifics for performing technical handling are defined in the conditions of Ostrava Airport:

1. Departing and boarding passengers cross the route from the plane to the gate and back on foot under the supervision of a handling agent,
2. Boarding of passengers begins immediately after the completion of refueling of the aircraft,
3. Aircrafts are not pushed back from the apron by push-back vehicle.

Table 1 shows the activities that can be performed by a given group of occupations. This table serves as the base for creation of matrix $\boldsymbol{A}$. Table 1 also shows how many members each workgroup has.

Table 1 Authorization to perform component activities

| | Activity/Qualification | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 1 | Guiding to the apron | | 1 | | | | | |
| 2 | Securing the landing gear | | 1 | | | | | |
| 3 | GPU connection | | 1 | 1 | | | | |
| 4 | Safety zone | | 1 | 1 | 1 | | | |
| 5 | Visual check of aircraft | | 1 | | | | | |
| 6 | Delivery of stairs | | 1 | 1 | 1 | | | |
| 7 | Opening of cargo space | | 1 | 1 | 1 | | | |
| 8 | Delivery of belt conveyors | | 1 | 1 | 1 | | | |
| 9 | Baggage unloading | | | | 1 | | | |
| 10 | Passenger exiting | 1 | | | | | | |
| 11 | Refueling | | | | | | | 1 |
| 12 | Refilling catering | | | | | | 1 | |
| 13 | Refilling drinking water | | 1 | 1 | | | | |
| 14 | Draining of toilets | | 1 | 1 | | | | |
| 15 | Cabin cleaning | | | | | 1 | | |
| 16 | Baggage loading | | | | 1 | | | |
| 17 | Removal of conveyors | | 1 | 1 | 1 | | | |
| 18 | Closing of cargo space | | 1 | 1 | 1 | | | |
| 19 | Passenger boarding | 1 | | | | | | |
| 20 | Delivery of documents | 1 | | | | | | |
| 21 | Removal of stairs | | 1 | 1 | 1 | | | |
| 22 | GPU disconnection | | 1 | 1 | | | | |
| 23 | Cancellation of the safety zone | | 1 | 1 | 1 | | | |
| 24 | Visual check of aircraft | | 1 | | | | | |
| 25 | Removal of chocks | | 1 | | | | | |
| 26 | Marshalling | | 1 | | | | | |
| | Number of workers | 1 | 1 | 1 | 6 | 5 | 1 | 1 |

Group job titles from Table 1:

1 – Handling agent, 2 – Marshall, 3 – Specialist, 4 – Loader, 5 – Cleaner, 6 – Catering worker, 7 – Aircraft refueling worker.
The solution process started with the creation of a network chart. The network chart was created in compliance with all the above input conditions. Table 2 lists all the activities that were performed during the technical ground handling, their time intensity (duration) and the edge that represents the activity in the network chart. This table lists the values of the earliest possible starts of activities ($\bar{\bar{t}}$), the latest allowable starts of activities ($t$), the time requirements of each activity ($t$) and the calculated time values of reserves. The activities are arranged in a logical sequence.

After completing the optimization calculation it was found, that in process of technical handling, one qualification group can be saved. This qualification group is Specialist. Table 2 contains all the results from optimization calculations. Activities which were completed by Specialist are now completed by Marshall and with activity 23 qualification group – Loader was assigned. You can see which qualification group was assigned to which activity from process of technical handling as well as the worker from each group.

Table 2 Results of computational experiment

| Phase 1 | | | |
|---|---|---|---|
| Activity | Assumed Qualification | Deployed Qualification | Employee |
| 1 | 2 | 2 | 1 |
| 2 | 2 | 2 | 1 |
| 3 | 3 | 2 | 1 |
| 4 | 2 | 2 | 1 |
| 5 | 2 | 2 | 1 |
| 6 | 4 | 4 | 1,3 |
| 7 | 4 | 4 | 6 |
| 8 | 4 | 4 | 1,4 |
| 9 | 4 | 4 | 1,2,3,4,5,6 |
| 10 | 1 | 1 | 1 |
| 11 | 7 | 7 | 1 |
| 12 | 6 | 6 | 1 |
| 13 | 3 | 2 | 1 |
| 14 | 3 | 2 | 1 |
| 15 | 5 | 5 | 1,2,3,4,5 |
| 16 | 4 | 4 | 1,2,3,4,5,6 |
| 17 | 4 | 4 | 2,4 |
| 18 | 4 | 4 | 3,5 |
| 19 | 1 | 1 | 1 |
| 20 | 1 | 1 | 1 |
| 21 | 4 | 4 | 2,3 |
| 22 | 3 | 2 | 1 |
| 23 | 2 | 4 | 6 |
| 24 | 2 | 2 | 1 |
| 25 | 2 | 2 | 1 |
| 26 | 2 | 2 | 1 |

## 5 CONCLUSIONS

Traffic at regional airport Ostrava is mainly seasonal, in the winter months is minimal. Summer months are distinguished by increase of traffic as well as increase of requirements for staffing. The base for mathematical modeling of process of technical handling of aircraft at apron is time plan created on the base of CPM. Every particular activity was identified and process of technical handling was designed to suit real conditions. Results of CPM (mainly time reserves) serve as input information for follow-up calculations.

Result of mathematical model, created for solution of proposed problem and described in article, is to obtain information about minimal number of staff needed for process of technical handling of aircraft in defined time span. With the minimization of workers by one qualification group the flight schedule must not be disrupted and no delay cannot be created.

**References**

[1] Al-Bazi, A., Gok, Y.S., Ozturk, C. and Guimarans, D. (2016). Developing a mathematical model for scheduling of turnaround operations (low cost airline as a case study). In *The 3rd International Aviation Management Conference 2016* (p. 16). Emirates Aviation University.
[2] Andreatta, G., De Giovanni, L. and Monaci, M. (2014). A fast heuristic for airport ground-service equipment–and–staff allocation. *Procedia-Social and Behavioral Sciences*, *108*, pp.26-36.
[3] IATA.(2018). *IATA Ground Operations Manual*, Montreal. Montreal, Geneva: International Air Transport Association.
[4] Červinka, M. (2017). *Airport Operation*. VSB-Technical university of Ostrava. ISBN 978–80-248-4115-1
[5] Fiala, P. (2004). *Project management: models, methods, analysis*. Professional Publishing.
[6] Jablonský, J. (2007). *Operational Research :quantitative models for economical decisioning*. Praha: Professional Publishing.
[7] Klusoň, V. (1973). *Critical path and PERT in operational practice*. SNTL.
[8] Kwasiborska, A. (2010). Modelling of ground handling operations at airport. *Journal of KONES*, *17*, pp. 253-261.
[9] Teichmann, D. and Dorda, M. (2016). About an Unconventional Application of Assignment Problem. *Proceedings of the international conference: Quantitative Methods in Economics: Multiple Criteria Decision Daking XVIII*, pp.388–392.
[10] Šajdlerová I. and Konečný M. (2008). *Project management*. Ostrava: VSB - Technical univerzity of Ostrava.

**Authors address**

Ing. Jakub Cíleček
VSB-Technical university of Ostrava, Faculty of Mechanical Engineering, Institute of Transport
17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic
email: jakub.cilecek@vsb.cz
doc. Ing. Dušan Teichmann, Ph.D.
VSB-Technical university of Ostrava, Faculty of Mechanical Engineering, Institute of Transport
17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic
email: dusan.teichmann@vsb.cz

# RANDOM WALK AND RESERVES MODELING IN STUDYING PENSIONS FUNDS SUSTAINABILITY

***Manuel Alberto M. Ferreira***
***Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, BRU-IUL***
***Portugal***

**Abstract**
Random walk is a stochastic process classic example, used to study a set of phenomena and, particularly, as in this article, models of reserves evolution. Random walks also allow the construction of significant complex systems and are also used as an instrument of analysis, being used in the sense of giving a theoretical characteristic to other types of systems. Our goal is primarily to study reserves to see how to ensure that pension funds are sustainable. This classic approach to the pension funds study makes it possible to draw interesting conclusions about the problem of reserves.

***Keywords:*** *Reserves, Ruin, Random walks, Pensions funds*

***JEL Classification:*** C18
***AMS Classification:*** 60G99

## 1 INTRODUCTION

Gambler's ruin suggests that growth events occur randomly, depending the survival on the stock of stored resources. In Gambler's ruin problem, reserves behave according to a simple random walk. This problem has been often presented in many works considering stochastic processes theory in contexts under the frameworks of Markov Chains, Random Walks, Martingales or even others. Billingsley (1986) or Feller (1968) solve problems in this topic using the classic first step analysis to obtain a difference equation, which approach we use in our study. In another basis, (Grimmett and Stirzaker, 1992) and (Karlin and Taylor, 1975) get resolutions through the Martingales Theory and apply it as an applications' example of the Martingales Stopping Time Theorem.
The paper is an updated approach of the work presented in SPMCS 2012, see (Filipe, Ferreira, and Andrade, 2012), and is organized as follows. In section 2. gambler's ruin problem is presented. In section 3. we approach random walks, modeling the general problem. Section 4. provides a ruin probability's particularization. In section 5. we approach the definition of pensions fund management policies to guarantee its sustainability. Section 6. concludes.
In Ferreira (2017, 2018), (Ferreira, Andrade, and Filipe, 2012) and (Andrade, Ferreira, and Filipe, 2012) more applications of stochastic processes can be seen in problems of reserves and pensions funds.

## 2 GAMBLER'S RUIN

We consider a gambler with an initial capital of $x$ monetary units, intending to play a sequence of games till the gambler's wealth reaches $k$ monetary units. We suppose that $x$ and $k$ are integer numbers satisfying the conditions $x > 0$ and $k > x$. In each game, there are two possible situations: winning 1 monetary unit with probability $p$ or losing 1 monetary unit with probability $q = 1- p$. A question is posed in terms of knowing which will be the probability the gambler

gets in ruin before attaining his/her objective. This means that it is intended to know which the probability of losing the $x$ monetary units is before adding gains of $k - x$ monetary units.
Be $X_n, n = 1, 2, \ldots$ the outcome of the $n^{\text{th}}$ game. Obviously, the variables $X_1, X_2, \ldots$ are *i.i.d.* random variables, which common probability function is: $P(X_n = 1) = p, P(X_n = -1) = q = 1 - p$.
Consequently, the gambler's wealth, his/her reserves after the $n^{\text{th}}$ game represent the simple random walk as follows: $S_0 = x, S_n = S_{n-1} + X_n, n = 1, 2, \ldots$ .
Aiming to get the gambler's ruin probability, let's consider this probability as $\rho_k(x)$. It relates to the probability that $S_n = 0$ and $0 < S_i < k, i = 0, 1, \ldots, n-1$ for $n = 1$ or $n = 2$ or … . If $\rho_k(x)$ is conditioned to the result of the first game, and considering the Total Probability Law, we get the following:

$$\rho_k(x) = p\rho_k(x+1) + q\rho_k(x-1),\ 0 < x < k. \qquad (1)$$

If $0 \leq x \leq k$, the difference equation presented is easily solved considering the border conditions

$$\rho_k(0) = 1,\ \rho_k(k) = 0 \qquad (2)$$

getting:

$$\rho_k(x-1) = \begin{cases} \dfrac{1-(p/q)^{k-x}}{1-(p/q)^{k}}, p \neq \frac{1}{2} \\ \\ \dfrac{k-x}{k},\ p = \frac{1}{2} \end{cases} \qquad (3)$$

Be $N_a$ the random walk $S_n$first passage time by $a$: $N_a = \min\{n \geq 0: S_n = a\}$. Now we can write $\rho_k(x) = P(N_0 < N_k | S_0 = x)$. It is appropriate to consider in (3) the limit as $k$ converges to $\infty$ to evaluate $\rho(x)$, the ruin probability of a gambler infinitely ambitious. In this context, considering the simple random walk $S_n, \rho(x) = P(N_0 < \infty | S_0 = x)$, after
(3)

$$\rho(x) = \lim_{k\to\infty} \rho_k(x) = \begin{cases} (q/p)^x, if\ p > \frac{1}{2} \\ \\ 1, \qquad \text{if}\, p \leq \frac{1}{2} \end{cases} \qquad (4)$$

Being $\mu = E[X_n] = 2p - 1,$ after (4), the ruin probability is 1 for the simple random walk at which the mean of the step is $\mu \leq 0 \Leftrightarrow p \leq \frac{1}{2}$.

# 3 THE GENERAL RANDOM WALK

In this section, we begin considering a fund in which contributions/pensions received/paid, per time unit, are described as a sequence of random variables $\xi_1, \xi_2, \ldots$ $(\eta_1, \eta_2, \ldots)$. Let's consider that $\xi_n(\eta_n)$ is the amount of the contributions/pensions that are received/paid by the fund during the $n^{\text{th}}$ time unit and consequently $X_n = \xi_n - \eta_n$ is the reserves variation in the fund at the $n^{\text{th}}$ time unit. If $X_1, X_2, \ldots$ is a sequence of non-degenerated *i.i.d* random variables, the stochastic process, representing the evolution of the fund reserves, since the value $x$ till the amount $\tilde{S}_n$ after $n$ time units, will be defined recursively as $\tilde{S}_0 = x, \tilde{S}_n = \tilde{S}_{n-1} + X_n,$ with $n = 1, 2, \ldots$ . Such a process is a general random walk.
The aim is to study the fund ruin probability, i.e., the game reserves exhaustion probability.

We consider $x$ and $k$ real numbers fulfilling $x > 0$ and $k > x$. First, the evaluation of $\rho_k(x)$, the probability that the fund reserves decrease from an initial value $x$ to a value in $(-\infty,0]$ before reaching a value in $[k,+\infty)$, is considered. Then, after the calculus of the limit, as seen in the previous section, the evaluation of $\rho(x)$, the eventual fund ruin probability is considered, by admitting in such a situation that the random walk - that represents its reserves - evolves with no restrictions at the right of 0.
This method that was presented here is known in the literature of stochastic processes as Wald's Approximation. The explanations by (Grimmett and Stirzaker,1992) and (Cox and Miller, 1965), relating this issue are diligently considered in our work. Here, we consider the $S_n = \tilde{S}_n - x$ process, i.e., the random walk $S_0 = 0, S_n = S_{n-1} + X_n, n = 1, 2, \ldots$, instead of $\tilde{S}_n$ process.
Accordingly, when we evaluate $\rho_k(x)$, it is the probability the process $S_n$ is visiting the set $(-\infty,-x]$ before visiting the set $[k-x,+\infty)$ that is considered. And when we evaluate $\rho(x)$ it is only the probability that the process $S_n$ goes down from the initial value 0 till a level lesser or equal than $-x$ that is considered.
We begin considering now the non-null value $\theta$ for which the $X_1$ moments generator function assumes the value 1. It is assumed that such a $\theta$ exists, that is, $\theta$ satisfies

$$E\left[e^{\theta X_1}\right] = 1, \theta \neq 0 \tag{5}$$

Defining the process as $M_n = e^{\theta S_n}, n = 0, 1, 2, \ldots$, it is now evident that $E[|M_n|] < \infty$ and also, after (5), $E[M_{n+1}|X_1, X_2, \ldots, X_n] = E\left[e^{\theta(S_n + X_{n+1})}|X_1, X_2, \ldots, X_n\right] = e^{\theta S_n} E\left[e^{\theta X_{n+1}}|X_1, X_2, \ldots, X_n\right] = M_n$. Consequently, the process $M_n$ is a Martingale in what concerns the sequence of random variables $X_1, X_2, \ldots$. Consider now $N$, the $S_n$ first passage time to outside the interval $(-x, k-x)$, $N = \min\{n \geq 0: S_n \leq -x \text{ or } S_n \geq k - x\}$.

The random variable $N$ is a stopping time – or a Markov time – for which the following conditions are satisfied:

- $E[N] < \infty$,
- $E[|M_{n+1} - M_n||X_1, X_2, \ldots, X_n] \leq 2e^{|\theta|a}$, for $n < N$,
- $a = -x \lor a = k - x$.

We suggest giving a look at the work by (Grimmett and Stirzaker, 1992) on this issue. Under these conditions, we can resort to the Martingales Stopping Time Theorem and so:

$$E[M_n] = E[M_0] = 1 \tag{6}$$

Besides,

$$E[M_n] = E\left[e^{\theta S_N}|S_N \leq -x\right]P(S_N \leq -x) + E\left[e^{\theta S_N}|S_N \geq k - x\right]P(S_N \geq kx) \tag{7}$$

Realizing the approximations $E\left[e^{\theta S_N}|S_N \leq -x\right] \cong e^{-\theta x}$ and $E\left[e^{\theta S_N}|S_N \geq k - x\right] \cong e^{\theta(k-x)}$, and considering $P(S_N \leq -x) = \rho_k(x) = 1 - P(S_N \geq k - x)$, after (6) and (7), we obtain

$$\rho_k(x) \cong \frac{1 - e^{\theta(k-x)}}{e^{-\theta x} - e^{\theta(k-x)}}, \text{when } E[X_1] \neq 0 \tag{8}$$

This result is the Classic Approximation for the Ruin Probability in the conditions stated in (5). To admit a non-null solution $\theta$ for the equation $E\left[e^{\theta X_1}\right] = 1$ implies in fact to assume that $E[X_1] \neq 0$.
Going farer, we may consider a particular case, beyond the studied above, looking at the situation for which the equation $E\left[e^{\theta X_1}\right] = 1$ only solution is precisely $\theta = 0$; it means, the situation at which $E[X_1] = 0$. This case may be considered through the following passage to the limit:

$$\rho_k(x) \cong \lim_{\theta \to 0} \frac{1-e^{\theta(k-x)}}{e^{-\theta x}-e^{\theta(k-x)}} = \frac{k-x}{k}, \text{ when } E[X_1] = 0 \qquad (9)$$

As for $\rho(x)$, the probability that the process $S_n$ decreases eventually from the initial value 0 to a level lesser or equal than – $x$, is also got from (8), now for a different passage to the limit:

$$\rho(x) \cong \lim_{k \to \infty} \frac{1-e^{\theta(k-x)}}{e^{-\theta x}-e^{\theta(k-x)}} = e^{\theta x}, \text{ if } \theta < 0 \Leftrightarrow E[X_1] > 0 \qquad (10)$$

Considering the previous section results on the simple random walk, it is effective to accept $\rho(x) = 1$ when $\theta \geq 0 \Leftrightarrow \mathrm{E}[\mathrm{X}_1] \leq 0$.

## 4 A RUIN PROBABILITY'S PARTICULARIZATION

Consider $X_1, X_2, \ldots$ is a sequence of independent random variables with normal distribution with mean $\mu$ and standard deviation $\sigma$. So, we can suppose $X_n$, is the fund reserves variation at the $n^{\text{th}}$ time unit, normally distributed with those parameters. Now, the moments' generator function is: $E\left[e^{\theta X_1}\right] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} e^{\theta x - \frac{(x-\mu)^2}{2\sigma^2}} dx = e^{\theta\mu + \frac{\theta^2\sigma^2}{2}}$. And equation (5) solution is $\theta = \frac{-2\mu}{\sigma^2}, \mu \neq 0$.

- So $\rho_k(x)$, the ruin probability, is acquired when substituting this result in (8), as follows:

$$\rho_k(x) \cong \frac{1-e^{-\frac{2\mu(k-x)}{\sigma^2}}}{e^{\frac{2\mu x}{\sigma^2}}-e^{-\frac{2\mu(k-x)}{\sigma^2}}}, \text{when } \mu \neq 0 \qquad (11)$$

- It is obvious that this particularization does not influence the approximation to $\rho_k(x)$ when $\mu = 0$. As it was seen before, it is given by (9) and after (10), we have

$$\rho(x) \cong e^{-\frac{2\mu x}{\sigma^2}}, \text{ when } \mu > 0 \qquad (12)$$

## 5 ASSETS AND LIABILITY MANAGEMENT POLITICS

It evidently follows from the matters presented so far that it is imperative to define pension fund management policies to guarantee its sustainability.
To do so, see for instance Ferreira (2017,2018), we assume that the assets value process of a pensions fund, may be represented by a geometric Brownian motion process $A(t) = be^{a+(\rho+\mu)t+\sigma B(t)}, \mu < 0, ab\rho + \mu\sigma > 0$, where $B(t)$ is a standard Brownian motion process. Suppose also that the fund liabilities value process performs such as the deterministic process $L(t) = be^{\rho t}$. Be also the stochastic process $Y(t) = ln\frac{A(t)}{L(t)} = a + \mu t + \sigma B(t)$: it is a generalized Brownian motion process, starting at $a$, with drift $\mu$ and diffusion coefficient

$\sigma^2$.Note also that the first passage time of the assets process $A(t)$ by the mobile barrier $T_n$, the liabilities process, is the first passage time of $Y(t)$ by 0, with finite expected time.
Then, for instance, be a pensions fund management scheme that raises the assets value by a positive constant $\theta_n = \theta,$ when the assets value falls equal to the liabilities process in the $n^{th}$ time. This corresponds to consider a modification in the process A(t), we call it $\bar{A}(t)$ , that now restarts at instants $T_n$, when hits the barrier $L(t)$, by the $n^{th}$ time at level $L(T_n) + \theta_n$.
A convenient choice for management policy is, for example, one in which:

$$\theta_n = L(T_n)\left(e^{\theta} - 1\right), \text{for some } \theta > 0 \tag{13}$$

Then, the corresponding modification of $Y(t)$, $\bar{Y}(t)$, behaves as a generalized Brownian motion process that restarts at the level $\theta = \ln\frac{L(T_n)+\theta_n}{L(T_n)}$ when it its 0, at times $T_n$.
The **present value of the cost of perpetual maintenance of the pension fund,** see (Ferreira and Filipe, 2021, 2021a), is considered, due to the proposed asset-liability management scheme, given by the random variable:$V(r,a,\theta) = \sum_{n=1}^{\infty} \theta_n e^{-rT_n} = \sum_{n=1}^{\infty} b\left(e^{\theta} - 1\right) e^{-(r-\rho)\,T_n}, r > \rho,$ where $r$ represents the due discount rate. This random variable expected value is

$$v_r(a,\theta) = \frac{b(e^{\theta}-1)e^{-K_{r-\rho}a}}{1-e^{-K_{r-\rho}\,\theta}},\ K_i = \frac{\mu+\sqrt{\mu^2+2i\sigma^2}}{\sigma^2} \tag{14}$$

Note that $\lim_{\theta\to 0} v_r(a,\theta) = \frac{be^{-K_{r-\rho}a}}{K_{r-\rho}}$.

## 6 CONCLUSIONS

These results were obtained considering the simple and general random walk that are classic and widely studied stochastic processes. Since its general ideas are easily grasped by everyone, which quickly connect them with real systems, the random walk is used to model situations more disparate realities, far beyond the reserve evolution models considered in this work. They are also, and very often, used to build other complex systems, sometimes much more complex, for other types of systems.
We highlight in our approach a set of different methodologies that were applied to the study of this type of processes such as the cases of Difference Equations and Martingales Theory.
In our approach, reserves systems are treated as physical systems. In our study, we recognize that this may be a limitation to be considered, since it is not obvious that the direct application of these principles to financial reserve funds can be legitimate when their own dynamics of appreciation and devaluation over time are ignored.
The models themselves, and the consequent valuation of stability systems based on the assessment of the probability of depletion or ruin of reserves, seem to be valid only in constant price contexts.
But, in section 5. we consider the integration of factors that are associated with the process of temporal depreciation of the value of money while considering the modeling of financial reserves, when dealing with possible politics aiming to guarantee the fund sustainability.

**Acknowledgement:** This work is partially financed by national funds through FCT - Fundação para a Ciência e Tecnologia, I.P., under the project FCT UIDB/04466/2020. Furthermore, the author thanks the ISCTE-IUL and ISTAR-IUL, for their support.

## References

[1] Andrade, M., Ferreira, M. A.M. and Filipe, J. A. (2012). Representation of reserves through a Brownian motion model, In *Journal of Physics: Conference Series* (Vol. 394, No.1, P. 012036), IOP Publishing. DOI:10.1088/1742-6596/394/1/012036.
[2] Billingsley, P. (1986). *Probability and Measure*. 2nd Ed. New York: John Wiley & Sons.
[3] Cox, D. and Miller, H. (1965), *The Theory of Stochastic Processes*. London: Chapman & Hall.
[4] Feller, W. (1968). *An Introduction to Probability Theory and its Applications*. Vol. 1,3rd Ed. New York: John Wiley & Sons.
[5] Ferreira, M. A. M. (2017). Searching for answers to the maintenance problem of insufficiently financed, financially dependent pension funds through stochastic diffusion processes. In *Pensions: Global Issues, Perspectives and Challenges. Hauppauge*, New York: Nova Science Publishers, pp. 113–126.
[6] Ferreira, M. A.M. (2018). Financially dependent pensions funds maintenance approach through Brownian motion processes. In *17th Conference on Applied Mathematics, APLIMAT 2018*. Slovak University of Technology in Bratislava, pp. 348–355.
[7] Ferreira, M A.M, Andrade, M. and Filipe, J. A. (2012). Studying pensions funds through an infinite servers nodes network: A theoretical problem, In *Journal of Physics: Conference Series* (Vol. 394, No.1, P. 012035), IOP Publishing. DOI: 10.1088/1742-6596/394/1/012035.
[8] Ferreira, M.A.M. and Filipe, J. A. (2021). Diffusion and Brownian Motion Processes in Modeling the Costs of Supporting Non-autonomous Pension Funds. In Nermend K., Łatuszyńska M., Thalassinos E. (eds) Decision-Making in Management. CMEE 2019. Contributions to Management Science. Springer, Cham. https://doi.org/10.1007/978-3-030- 67020-7_5
[9] Ferreira, M.A.M. and Filipe, J. A. (2021a). Addressing Reserves and Pension Funds through Gambler's Ruin and Generalized Brownian Motion Process. *Recent Advances in Mathematical Research and Computer Science* Vol. 4, pp. 15–24. https://doi.org/10.9734/bpi/ramrcs/v4/14551D.
[10] Filipe, J.A., Ferreira, M. A. M. and Andrade, M. (2012). Reserves represented by random walks, In *Journal of Physics: Conference Series* (Vol. 394, No. 1, P. 012034), IOP Publishing. DOI:10.1088/1742-6596/394/1/012034.
[11] Grimmett, G. and Stirzaker, D. (1992). *Probability and Random Processes*. 2nd Ed. Oxford: Clarendon Press.
[12] Karlin, S. and Taylor, H. (1975). *A First Course in Stochastic Processes*. 2nd Ed. New York: Academic Press.

## Author's address

Manuel Alberto M. Ferreira, Emeritus Full Professor
Instituto Universitário de Lisboa (ISCTE-IUL)
ISTAR-IUL, BRU-IUL
Av. das Forças Armadas, 1649-026 Lisboa, Portugal.
Email: manuel.ferreira@iscte.pt

# MODELS OF CO-OPETITION

***Petr Fiala, Prague University of Economics and Business, Prague***
***Renata Majovska, University of Finance and Administration, Prague***

**Abstract**
Co-opetition is a business strategy that combines the advantages of competition and cooperation into new dynamic, which can be used to not only generate more profits but also to change nature of the business environment in benefit of users. The coopetition business model has PARTS of a business strategy - five dimensions a company can use to identify strategies that change the game: Players, Added value, Rules, Tactics, and Scope. The players are the firm, customers, suppliers, competitors and complementors, competitors whose products add value. The paper presents game theory models that represent the co-opetition concept. Traditional game theory is divided into non-cooperative and cooperative models. Biform games combine non-cooperative and cooperative models. An important part of the biform game is to learn which conditions will influence the players to either compete or cooperate.

***Keywords:*** *Co-opetition, Business model, Game theory, Biform games*

***JEL Classification:*** L22, C70
***AMS Classification:*** 93A30, 91A40

## 1 INTRODUCTION

The traditional concept of business as a “winner takes all” contest is replaced by a new concept. Co-opetition is a business strategy that goes beyond the rules of competition and cooperation to combine the advantages of both. Co-opetition combines the advantages of both competition and cooperation into new dynamic, which can be used to not only generate more profits but also to change nature of the business environment in benefit of users (see Brandenburger and Nalebuff, 2011).

The co-opetition is based on game theory. The work of John von Neumann and Oskar Morgenstern (1944) is the classic work upon which modern game theory is based. Since then, an extensive literature on game theory was published. For example, Myerson's book (1997) provides a clear and thorough examination of the models, solution concepts, results, and methodological principles of non-cooperative and cooperative game theory. Game theory models analyse situations where players make decisions to maximize their own utility, while considering that other players are doing the same, and that decisions, made by players, impact others utilities. Traditional game theory is divided into non-cooperative and cooperative models. The non-cooperative theory of games is strategy oriented; it studies what one may expect the players to do. The non-cooperative theory is a “micro” approach in that it focuses on precise descriptions of what happens. The co-opetition concept combines non-cooperative and cooperative game theory models.

Dynamics is a very important part of co-opetition models. Current business conditions are changing rapidly. New products are evolving faster. Searching for relationships with complementors brings ever new opportunities that bring added values. This is one of the basics of coopetition. This dynamic must be included in the new cooperative models.

Searching for new opportunities with complementors and negotiations with competitors are carried out in the face of multiple criteria. There are multiple criteria in co-opetition models, such as economic, social, environmental, technological, and others. Multicriteria evaluation approaches will be tested for inclusion in the model of co-opetition.

The rest of the paper is organized as follows. Section 2 presents co-opetition business model. Section 3 summarizes the basics of the game theory concepts. Biform games as basic models for modelling of co-opetition problems are analysed in Section 4. Extensions of biform games are presented in Section 5. Section 6 presents conclusions.

## 2 CO-OPETITION BUSINESS MODEL

The co-opetition concept (Brandenburger and Nabuleff, 2011) goes beyond the use of separate approaches to competition and cooperation goes beyond the use of separate approaches to competition and cooperation by combining their advantages together. The co-opetition business model has PARTS of a business strategy - five dimensions a company can use to identify strategies that change the game:

- Players,
- Added value,
- Rules,
- Tactics,
- Scope.

**Players** are divided by types into producers, customers, suppliers, competitors and complementors. Complementors are the competitors whose products increase the value of the producer's product. The situation is analysed as a game between the relevant players. Expanding the number of player types from a wider environment provides a deeper analysis of the situation (relationships between suppliers and producers affect costs, relationships between producers and customers affect demand and prices, relationships between producers and competitors affect producer behaviour, relationships between producers and competitors bring mutual added value). The model should analyse the impulses that affect players to be competitors or to cooperate.

The Value Net is a scheme that identifies relevant players and the relationships between them (see Fig. 1):

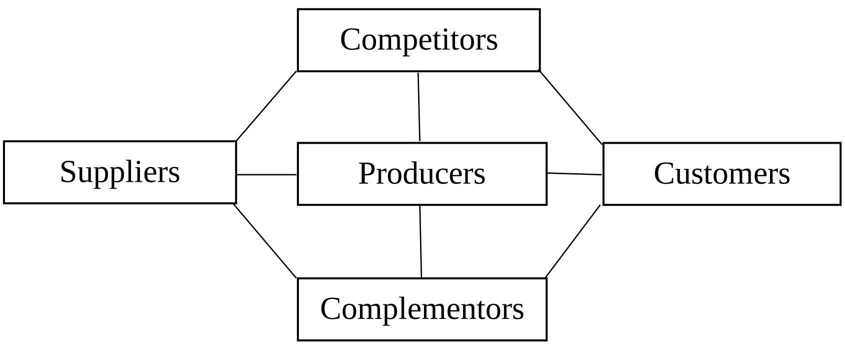



**Figure 1** The Value Net.

**Added values** is provided by complementors if their products bring an extension of possibilities for producers' own products. The producers can recognize these added values and take activities that increase their profitability.

**Rules** create a structure of negotiation between players. Some rules are hard and cannot be changed during negotiations. Other rules are softer and can be changed in the negotiation of contracts.

**Tactics** are sequences of activities that form the monitoring of the negotiation process by other players. Players can use these certain activities to intentionally influence the behaviour of other players. It is useful to monitor these activities and respond to them accordingly.

**Scope** is determined by the interconnections between the PART elements of the game model and other possible games in which players from this model participate. Extending the scope with more games can increase profitability. Leaving games separate may prove advantageous if the interconnection would limit some businesses. Joining and separating games is determined by changes in conditions over time.

Doing business should mean more than just competing for market share. The cooperation model forms the basis for capturing and explaining the effects of the wider environment and their use to gain an advantage and increase profitability.

The concept of co-opetition has been refined and applied in many cases. The paper (Zineldin, 2004) predicts that strategic planners in organizations of the future need to consider the potential benefits of collaborating, co-operating and co-ordinating with others serving the same markets, rather than pursuing conventional "competition". The paper (Gnyawali, He, and Madhavan, 2006) examines empirically the impact of co-opetition on firm competitive behaviour. The co-opetition concept with technological innovation is designed for small and medium-sized enterprises (see Gnyawali and Park, 2009) and for large firms (see Gnyawali and Park, 2011). Specific issues of co-opetition were analysed, such as dynamics (see Bengtsson, Eriksson and Wincent, 2010a), tension between cooperation and competition (see Bengtsson, Eriksson and Wincent, 2010b), and information sharing (see Renna and Argoneto, 2012). The co-opetition is based on game theory (Okura and Carfì, 2014). The players are connected to the supply chain. The paper (Okura and Carfì, 2014) describes coordination game model of co-opetition relationship on cluster supply chains. The paper (Gurnani, Erkoc and Luo, 2007) analyses impact of product pricing and timing of investment decisions on supply chain co-opetition. There are some critical reviews of the co-opetition concept (see Armstrong, 1997). Deeper use of game theory, the theory of supply chain and other disciplines is required for new more precise models of co-opetition.

## 3 BASIC GAME THEORY CONCEPTS

We will start from the basic non-cooperative and cooperative concepts of the game theory that are applied in the proposed approach for co-opetition models (see Myerson, 1997).

### 3.1 Non-cooperative concepts

Most non-cooperative allocation strategies in distributed systems consist of following steps:

- The formulation of utility (pay-off) functions for the system participants.
- The formulation of best response strategies.
- The existence of Nash equilibrium is proved in the system of multiple agents.
- Efficiency is measured compared to achievable welfare.

An *n*-player non-cooperative game in the normal form is a collection

$$\{N = \{1,2,\dots,n\};\ X_1, X_2, \dots, X_n;\ \pi_1(x_1, x_2, \dots, x_n), \pi_2(x_1, x_2, \dots, x_n), \dots, \pi_n(x_1, x_2, \dots, x_n)\},\tag{1}$$

where $N$ is a set of $n$ players; $X_i$, $i$ = 1, 2, …, $n$, is a set of strategies for player $i$; $\pi_i(x_1, x_2, \dots, x_n)$, $i$ = 1, 2, …, $n$, is a pay-off function for player $i$, defined on a Cartesian product of $n$ sets $X_i$, $i$ = 1, 2, …, $n$.

Decisions of other players than player $i$ are summarized by a vector

$$\mathbf{x}_{-i} = (x_1, \dots, x_{i-1}, x_{i+1} \dots, x_n).\tag{2}$$

A vector of decisions $(x_1^0, x_2^0, \dots, x_n^0)$ is Nash equilibrium of the game if

$$x_i^0(\mathbf{x}_{-i}^0) = \underset{x_i}{\operatorname{argmax}}\, \pi_i(x_i, \mathbf{x}_{-i}) \forall i = 1,2, \dots, n.\tag{3}$$

Nash equilibrium is a set of decisions from which no player can improve the value of his pay-off function by unilaterally deviating from it.

### 3.2 Cooperative concepts

Cooperative game theory looks at the set of possible outcomes, studies what the players can achieve, what coalitions will form, how the coalitions that do form divide the outcome, and whether the outcomes are stable and robust.

The maximal combined output is achieved by solving the following task

$$\mathbf{x}^0 = \underset{\mathbf{x}}{\operatorname{argmax}} \sum_{i=1}^{n} \pi_i(x_i).\tag{4}$$

Allocation mechanisms are based on different approaches such as negotiations, auctions, Shapley values, etc. When modelling cooperative games is advantageous to switch from the game in normal form to the game in the characteristic function form. The characteristic function of the game with a set $N$ of $n$ players is such function $v(S)$ that is defined for all subsets $S \subseteq N$ (i.e. for all coalition) and assigns a value $v(S)$ with following characteristics:

$$v(\emptyset) = 0,\quad v(S_1 \cup S_2) \geq v(S_1) + v(S_2),\tag{5}$$

where $S_1$, $S_2$ are disjoint subsets of the set $N$. The pair ($N$, $v$) is called a cooperative game of $n$ players in the characteristic function form.

A particular allocation policy, introduced by Shapley (see Shapley, 1953) has been shown to possess the best properties in terms of balance and fairness. So-called Shapley vector is defined as

$$\mathbf{h} = (h_1, h_2, \dots, h_n),\tag{6}$$

where the individual components (Shapley values) indicate the mean marginal contribution of $i$-th player to all coalitions, which may be a member. Player contribution to the coalition $S$ is calculated by the formula

$$v(S) - v(S - \{i\}).\tag{7}$$

Shapley value for the $i$-th player is calculated as a weighted sum of marginal contributions according to the formula:

$$h_i = \sum_S \left\{ \frac{(|S|-1)!(n-|S|)!}{n!} . \left[ v(S) - v(S - \{i\}) \right] \right\},\tag{8}$$

where the number of coalition members is marked by symbol $|S|$ and the summation runs over all coalition $i \in S$.

# 4 BIFORM GAMES

We will use biform games as basic models for modelling of co-opetition problems. A biform game is a combination of non-cooperative and cooperative games (see Brandenburger and Stuart, 2007). The goal of co-opetition is to move the players from a zero-sum game to a plus-sum game, a scenario in which the end result is more profitable when the competitors work together. We propose to divide biform games into sequential and simultaneous (Fiala and Majovská, 2020).

## 4.1 Sequential biform games

The sequential biform game is a two-stage game: in the first stage, players choose their strategies in a non-cooperative way, thus forming the second stage of the game, in which the players cooperate. First, suppliers make initial proposals and take decisions. This stage is analysed using a non-cooperative game theory approach. The players search for Nash equilibrium by solving the problem (3).

Then, players negotiate among themselves. In this stage, a cooperative game theory is applied to characterize the outcome of negotiation among the players over how to distribute the total surplus. Each player's share of the total surplus is the product of its added value and its relative negotiation power. We propose to distribute the total surplus to players by Shapley values (8).

The sequential biform game is used f. e. in supply chains (see Fiala, 2015, 2016). In the non-cooperative part, a coordination mechanism based on a specific buy-back contract is applied between producers and customers with price-dependent stochastic demand. The contract has desirable features: full coordination of the supply chain, flexibility to allow any division of the supply chain's profit, and easy to use. The cooperative part is merely focused on two concepts, coalition formations by resource capacity constraints and profit sharing. Profit sharing is carried out on the recognized concept of Shapley value.

## 4.2 Simultaneous biform games

The simultaneous biform game is a one-stage model where combinations of concepts for cooperative and non-cooperative games are applied. The combinations will be changed according situations in co-opetition problems. The first problem is a classification of situations. The situations are affected by:

- which players can cooperate,
- to what scope they can cooperate.

If all players can cooperate fully, then a standard cooperative model can be used (4) with subsequent distribution of the result according to the Shapley values (8). If no one can cooperate even in a partial content, a standard non-cooperative model is used (3). These are two extreme cases.

The players in the co-opetition model are the firm, customers, suppliers, competitors and complementors (competitors whose products add value). The relationship between the firm and the direct competitors is non-cooperative. The relationship between the firm and the complementors in a search for common added values is cooperative. The relationship between the firm and the suppliers can be partly cooperative; for some criteria cooperative (f. e. timing of deliveries), for others non-cooperative (f. e. price).

The scope of cooperation is determined by the various constraints that result from the fact that players are under internal and external pressures. The effects of pressures will be reflected in restrictive conditions. The scope of cooperation is dynamic and given by multiple criteria.

# 5 EXTENSIONS OF BIFORM GAMES

The biform game models can be more adapted to respect the co-opetition concept with basic elements such as players, added values, rules, tactics, and scope. Players are represented by negotiators who use their tactics to look for added value. Players negotiate under the pressure of the rules with varying scope.

The decision space $X$ is formulated by hard rules and defines hard restrictions on negotiations (Fiala, 1999). Elements of the decision space are vector strategies $\mathbf{x} \in X$. A consensus $\mathbf{x}^*$ should be chosen from the decision space $X$. Each negotiator $i = 1, 2, \dots, n$, evaluates the decisions $\mathbf{x}$ by the pay-off function $\pi_i(\mathbf{x})$. The process takes place in discrete time units $t = 1, 2, ..., T$ that represent multi-round negotiations.
Each negotiator, $i = 1, 2, ..., n$, has a set $X_i(t)$ of admissible decisions at discrete time points $t = 1, 2, ..., T$

$$X_i(t) = \{\mathbf{x};\ \mathbf{x} \in X,\ \pi_i(\mathbf{x}) \geq b_i(t)\},\ i = 1, 2, ..., n. \tag{9}$$

where the aspiration levels $b_i(t)$, $i = 1, 2, ..., n$, $t = 1, 2, ..., T$, of pay-off functions represent soft rules that can change in time and look for a consensus with added values.
The negotiation space is an intersection of sets of admissible decisions of all negotiators

$$X(t) = \bigcap_{i=1}^{n} X_i(t). \tag{10}$$

If the negotiation space $X(t)$ is a single element set, then this element is the consensus $\mathbf{x}^*$. If the negotiation space is not one-element, then negotiations on the aspiration levels of the negotiators begin. If the negotiation space is empty, then negotiators have to reduce some or all of the aspiration levels $b_i(t)$. If the negotiation space contains more than one element, then negotiators have to increase some or all of the aspiration levels $b_i(t)$.

Each negotiator negotiates under different internal and external pressures. The consequences of the pressures will cause changes in the constraints in admissible decision sets and thus a change in the negotiation space and, as a result, a consensus can be found.

# 6 CONCLUSION

New models of co-opetition are developed. Relevance of the project can be seen at two levels: theoretical and practical. The theoretical contribution can be seen in the creation of new models for biform games that will capture the sequential or the simultaneous use of non-cooperative and cooperative models. The models will include dynamic changes of parameters in the games, unlike traditional models of game theory. Multiple criteria evaluation of strategies can be also included in the models.

The practical contribution consists in the use of models for applications. Co-opetition business strategy is considered to be very promising. Changes in the new models will better reflect the real conditions. The proposed models can be used not only for co-opetition strategy, but also for many other decision problems which combine the advantages of competition and cooperation.

The authors of this paper have tried to apply the principles of co-opetition and biform game models to some specific problems. The basic areas of application were supply chains (Fiala, 2015), their design (Fiala and Majovská 2020) and profit allocation among chain members (Fiala, 2016). The paper (Fiala, 2022) extends the area of supply chains to network industries and focuses on modelling and analysis of co-opetition in network industries by biform games. Another area of application is environmental negotiations between polluters and authority.

**Acknowledgements:** This work was supported by the grant No. IGA F4/42/2021, Faculty of Informatics and Statistics, Prague University of Economics and Business, Prague.

**References**

[1] Armstrong J. S. (1997). Review of Barry J. Nalebuff and Adam N. Brandenburger, Co-opetition 1. Revolutionary Mindset that Redefines Competition and Co-operation 2. The Game Theory Strategy that's Changing the Game of Business. *Journal of Marketing*, 61(2), pp. 92-95.

[2] Bengtsson, M., Eriksson, J. and Wincent, J. (2010a). Co-opetition dynamics–an outline for further inquiry. *Competitiveness review: An international business journal*, 20(2), pp. 194-214.

[3] Bengtsson, M., Eriksson, J. and Wincent, J. (2010b) Tension in Co-opetition. In *Creating and Delivering Value in Marketing*. Springer International Publishing, Berlin, pp. 38-42.

[4] Brandenburger, A. M. and Nalebuff, B. J. (2011) *Co-opetition*. New York: Crown Business.

[5] Brandenburger, A. and Stuart, H. (2007). Biform games. *Management science*, 53(4), pp. 537-549.

[6] Fiala, P. (1999). Modeling of negotiation process with multiple evaluation criteria (in Czech). *Politická ekonomie*, 47(2), pp. 253-268.

[7] Fiala, P. (2015). Biform games in supply chains. In: *Mathematical Methods in Economics 2015*. University of West Bohemia, Pilsen, pp. 177–183.

[8] Fiala, P. (2016). Profit allocation games in supply chains. *Central European Journal of Operations Research*, 24(2), pp. 267-281.

[9] Fiala, P. (2022). Modelling and analysis of co-opetition in network industries by biform games. *Central European Journal of Operations Research*, 30(2), pp. 647-665.

[10] Fiala, P. and Majovská, R. (2020). Modeling the Design Phase of Sustainable Supply Chains. In: Khan, S. A. R. (ed) *Global Perspectives on Green Business Administration and Sustainable Supply Chain Management*. Hershey PA: IGI Global.

[11] Gnyawali, D. R., He, J. and Madhavan, R. (2006). Impact of co-opetition on firm competitive behavior: An empirical examination. *Journal of Management*, 32(4), pp. 507-530.

[12] Gnyawali, D. R. and Park, B. J. R. (2009). Co-opetition and technological innovation in small and medium-sized enterprises: A multilevel conceptual model. *Journal of small business management*, 47(3), pp. 308-330.

[13] Gnyawali, D. R. and Park, B. J. R. (2011). Co-opetition between giants: Collaboration with competitors for technological innovation. *Research Policy*, 40(50), pp. 650-663.

[14] Gurnani, H., Erkoc, M. and Luo, Y. (2007). Impact of product pricing and timing of investment decisions on supply chain co-opetition. *European Journal of Operational Research*, 180(1), pp. 228-248.

[15] Myerson, R. B. (1997). *Game Theory: Analysis of Conflict*. Cambridge: Harvard University Press.

[16] Okura, M. and Carfì, D. (2014). Coopetition and Game Theory. *Journal of Applied Economic Sciences*, 9(3), pp. 1-29.

[17] Renna, P. and Argoneto, P. (2012). Capacity investment decision in co-opetitive network by information sharing. *Computers & Industrial Engineering*, 62(1), pp. 359-367.

[18] Šauer, P., Dvořák, A. and Fiala, P. (1998). Negotiation between authority and polluters-model for support of decision making in environmental policy. *Politická ekonomie*, 46(6), pp. 772-787.
[19] Shapley, L. S. (1953). A value for n-person games. In: *Contributions to the Theory of Games* II. (Tucker, A. W., Luce, R. D., eds.). Princeton: Princeton University Press, pp. 307-317.
[20] von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.
[21] Zineldin, M. (2004). Co-opetition: the organisation of the future. *Marketing Intelligence & Planning*, 22(7), pp. 780-790.

**Author's address**

Prof. RNDr. Ing. Petr Fiala, CSc., MBA
Prague University of Economics and Business, Prague, Faculty of Informatics and Statistics, Department of Econometrics
W. Churchill Sq. 4, 130 67 Prague 3
Czech Republic
email: pfiala@vse.cz

PaedDr. Renata Majovska, PhD.
University of Finance and Administration, Prague, Faculty of Economic Studies, Department of Computer Science and Mathematics
Estonska 500, 101 00 Praha 10
Czech Republic
email: renata.majovska@mail.vsfs.cz

# THE ROLE OF PROXIMITY MEASUREMENT IN THE EVALUATION OF REGIONAL INNOVATION PROCESSES

***Andrea Furková***
***University of Economics in Bratislava***

**Abstract**
The paper deals with two different degrees of proximity between regions in relation to the spatial econometric modelling of the innovation problem. In addition to geographical proximity defined as the spatial distance between the actors – European regions, technological similarity based on the indicator - Small and Medium-Sized Enterprises (SMEs) introducing product or process innovations was also considered. Based on the dataset consisting of 220 European regions, two spatial autoregressive models were estimated. Patent applications were considered as a substitute for innovative output. Spatial innovation spilloves were indicated on the basis of both estimates, but spatial regional interactions were significantly weaker when technological similarity is considered as the degree of proximity of regions.

***Keywords:*** *European innovation, Geographical proximity, Technological similarity, Spatial econometrics*

***JEL Classification:*** O31, R12
***AMS Classification:*** 91B72

## 1 INTRODUCTION

For a long time, in the theoretical as well as in the empirical literature, the issue of innovation has played an important role in connection with many economic phenomena. We also see in the long run that a lot of research focuses on the spatial component of innovation and many studies have used proximity to examine spatial and relational mechanisms that lead to knowledge spillovers and sustainable technology development (Omobhude and Chen, 2019). In connection with the innovation process, geographical proximity of actors is emphasized, as clusters of innovative companies, research institutions or universities create suitable conditions for the production of innovation, knowledge sharing and their rapid dissemination. Such an approach is found in most of studies that deal with econometric modelling of regional innovation processes. In the context of examining European regional innovation processes, the findings of such studies point to an uneven distribution of innovation processes across European regions (e.g., Moreno et al., 2005a; Moreno eta al., 2005b; Kumar, 2008); Khan, 2012; Charlot et al., 2014 or Furková, 2019). In terms of proximity, study of Boschma (2005) brought critical remarks towards the role of geographical proximity in learning and innovation processes and Boschma assumes that other forms of proximity may substitute for geographical proximity. According to Boschma (2005) other forms of proximity's dimensions might be suitable such as cognitive, organisational, social and institutional proximity. In this paper, we consider two different degrees of proximity between regions in relation to the spatial econometric modelling of the innovation problem. In addition to geographical proximity, defined as the spatial distance between the actors – European regions, we consider technological similarity based on the indicator - Small and Medium-Sized Enterprises (SMEs) introducing product or process innovations. In the latter case, the neighbourhood of the regions is defined based on the difference of the number of SMEs. Higher shares of SMEs - technological innovators should reflect a higher level of innovation activities. In general, SMEs are vital to economic development in both developed and less developed economies.

The paper is structured as follows: section 2 includes the data description, spatial weighting matrices specifications and the main theoretical backgrounds of the study. In section 3, empirical results are presented and interpreted. The main concluding remarks can be found in section 4. The paper closes with references.

# 2 METHODOLOGY

In the first part of this section, the data and spatial weighting matrix specifications will be presented. The most important methodological issues related to empirical analysis will be part of the second part of this section.

## 2.1 Data and spatial weighting matrix specifications

This paper uses a set of data from the Regional Innovation Scoreboard (RIS) 2019 (Hollanders et al. 2019) and the Eurostat regional statistical database (Eurostat, 2021) to model the European innovation problem. As European Union (EU) level policies focus mainly on the regional level, our database contains 220[1] European regions at NUTS 1 or NUTS 2 level (NUTS - Nomenclature of territorial units for statistics). Most of regions are classified at NUTS 2 level, but countries that have a NUTS 1 and NUTS 2 level identical to the country's territory, are included at NUTS 1 levels (country levels). Figure 1 provides an overview of the study area. This figure shows a real spatial distribution of Patent Cooperation Agreement (PCT) applications that are used as a proxy of innovation output. PCT applications will be considered as a modelled variable of the following econometric analysis.

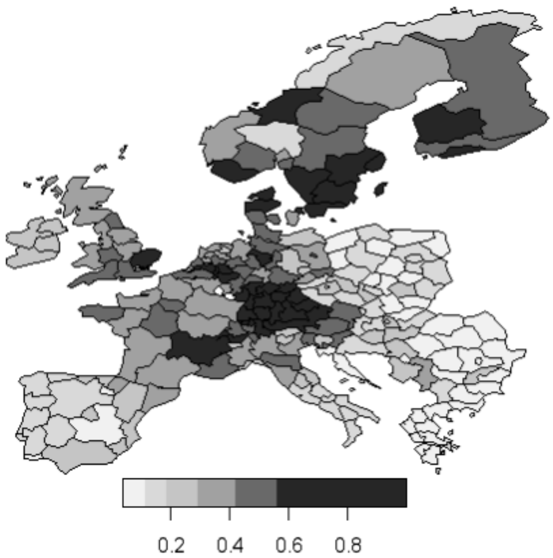



*Figure 1 Spatial distribution of Patent applications*
*Source: author's elaboration based on the RIS 2019 data (Hollanders et al., 2019) in RStudio*

Figure 1 shows that most regions in Eastern Europe, Greece, Spain and southern Italy have relatively low levels of patents. On the other side, the regions of Germany, Finland, Sweden, Switzerland, the Netherlands and Denmark belong to the top actors in this field. Already these geographical differences in the distribution of innovation (proxied by patents) suggest that the geographical location of the region plays an important role in generating regional innovation and the innovation process might be spatially autocorrelated. The problem of spatial autocorrelation and potential spatial innovation spillovers encouraged us to consider spatial

[1] The regions of 23 EU Member States, Norway, Serbia and Switzerland.

econometric approach to model the innovation problem. It has already been mentioned that the geographical proximity of the regions is likely to have an impact on innovation output. It will also be our intention to examine whether technological similarity will also be a suitable measure of the proximity of regions. As modelled variable – innovation output, PCT applications at the European Patent Office - EPO (PCT applications at the EPO per billion regional GDP) is used. As innovation inputs following variables are considered: top 10% most-cited publications (as percentage of total scientific publications), R&D expenditure in the business sector (as percentage of regional GDP), human resources in science and technology (as percentage of the active population) and population density (persons per square kilometre). All data are normalised using the min-max procedure (the data from RIS 2019 are already normalised). In addition, dummy variables for lagging regions[2] are considered to control additional regional characteristics ($DUMMY_{LAGG} = 1$, if region belongs to the category of lagging regions, others=0).

The estimation of the spatial econometric model (presented in the next section) is based on the two specifications of spatial weight matrix $\mathbf{W} = \left(w_{ij} : i, j = 1,..., N\right)$, which summarizes the spatial relationships between $N$ regions:

- Spatial queen contiguity weights (**$\mathbf{W}_{\mathbf{QUEEN}}$**) – these weights indicate whether regions share a boundary or not. Corresponding spatial weights are taking non-zero values when the regions are neighbours and zero values otherwise.
- Technological similarity/proximity weights (**$\mathbf{W}_{\mathbf{TECH.PROX}}$**) based on the SMEs introducing product or process innovations (defined as percentage of SMEs) – these weights are a negative power distance function of SMEs distance[3] of the form:

$$w_{ij} = d_{ij}^{(-\alpha)} \tag{1}$$

where $\alpha = 1$.

In Figure 2, corresponding connectivity histograms are shown. From the left side of Figure 2, we see that the number of neighbours based on the matrix **$\mathbf{W}_{\mathbf{QUEEN}}$** ranges from 1 to 11. However, significantly higher variability in the number of neighbours occurs when we consider the matrix **$\mathbf{W}_{\mathbf{TECH.PROX}}$**. In this case, the range of the number of the neighbours varies from 1 to 34.

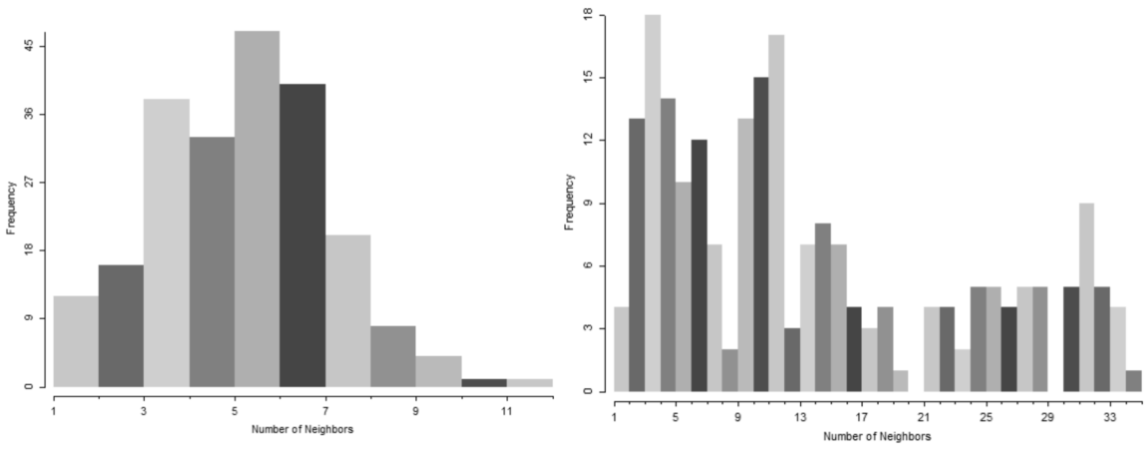



*Figure 2 Connectivity histograms – queen contiguity weights (left side) and inverse distance technological proximity weights (right side)*

[2] The regions whose GDP per capita is less than 75% of the average GDP of the EU-27.
[3] Euclidean distance metric was applied.

To illustrate the weights of the neighbourhood, we present Figure 3 which depicts the German region Oberplatz (DE23) and its neighbouring regions on the basis of both considered weight specifications. This figure clearly illustrates the differences in neighbouring regions.

*Figure 3 German region Oberplatz - DE23 (dark marked region) and its neighbouring regions - queen contiguity weights (left side) and inverse distance technological proximity weights (right side)*

## 2.2 Econometric models

Empirical part of this paper consists of estimation of three econometric models:

- Non-spatial Ordinary Least Squares (OLS) regression:
  $$\mathbf{y} = \mathbf{l}_N \alpha + \mathbf{X\beta} + \mathbf{u} \tag{2}$$
- Spatial Autoregressive model (SAR) - $\mathbf{W}_{\mathbf{QUEEN}}$ spatial weights:
  $$\mathbf{y} = \rho \mathbf{W}_{\mathbf{QUEEN}} \mathbf{y} + \mathbf{l}_N \alpha + \mathbf{X\beta} + \mathbf{u} \tag{3}$$
- Spatial Autoregressive model (SAR) - $\mathbf{W}_{\mathbf{TECH.PROX}}$ spatial weights:
  $$\mathbf{y} = \rho \mathbf{W}_{\mathbf{TECH.PROX}} \mathbf{y} + \mathbf{l}_N \alpha + \mathbf{X\beta} + \mathbf{u} \tag{4}$$

where $\mathbf{y}$ is the $N$-vector of a dependent variable; $\mathbf{X}$ is a matrix of $k$ explanatory variables ($k$ is a number of explanatory variables); $\mathbf{l}_N$ represents $N \times 1$ vector of ones associated with the intercept $\alpha$; $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters, $\mathbf{u} \sim i.i.d.\left(\mathbf{0}, \sigma_u^2\right)$ is $N \times 1$vector of random errors, $\sigma_u^2$ is random error variance and $\rho$ is spatial autoregressive parameter.

The spatial regression models defined in (3) and (4) contain a spatial lag vector of dependant variable and, the models state that innovation output in each region is related to the innovation activities performed in neighbouring regions. This means that the expected value of innovation output in the *i*th region is no longer influenced only by exogenous regional characteristics, but also by the exogenous characteristics of all other regions through a spatial multiplier $\left(\mathbf{I}_N - \rho \mathbf{W}\right)^{-1}$. Thus, corresponding marginal effects can be decomposed into direct and indirect impacts. LeSage and Pace (2009) suggested a summary measures of these impacts (for more details and formulas see LeSage and Pace 2009). Simultaneous feedback effects are major feature of the spatial regression models. Since both SAR models contain spatially-lagged dependent variable which introduces a source of endogeneity, we chose spatial two-stage least square (S2SLS) procedure to estimate these models.

# 3 EMPIRICAL RESULTS

The main results of the estimates are shown in Table 1. Given the results of the spatial dependency tests (see the lower part of Table 1) and the resulting problem of spatial autocorrelation, the OLS estimate will not be further addressed.

Table 1 Estimation results generated by OSL and SAR models

| Variable | OLS estimation | | Spatial 2SLS estimation | |
|---|---|---|---|---|
| | | | $W_{QUEEN}$ | $W_{TECH.PROX}$ |
| Intercept | -0.1691*** | | -0.1693*** | -0.1831*** |
| *Most-cited publications* | | | | |
| Parameter estimate | 0.2440*** | | 0.1341** | 0.1948*** |
| Direct impact | – | | 0.1417** | 0.1959*** |
| Indirect impact | – | | 0.0984** | 0.0294* |
| Total impact | – | | 0.2402** | 0.2253*** |
| *R&D expenditure in the business sector* | | | | |
| Parameter estimate | 0.5899*** | | 0.4914*** | 0.5806*** |
| Direct impact | – | | 0.5193*** | 0.5836*** |
| Indirect impact | – | | 0.3606*** | 0.0877** |
| Total impact | – | | 0.8798*** | 0.6713*** |
| *Human resources in science and technology* | | | | |
| Parameter estimate | 0.2898*** | | 0.1536*** | 0.2777*** |
| Direct impact | – | | 0.1623*** | 0.2792*** |
| Indirect impact | – | | 0.1127** | 0.0420* |
| Total impact | – | | 0.2751*** | 0.3211*** |
| *Population Density* | | | | |
| Parameter estimate | -0.3498*** | | -0.2594*** | -0.3386*** |
| Direct impact | – | | -0.2741*** | -0.3404*** |
| Indirect impact | – | | -0.1903*** | -0.0512* |
| Total impact | – | | -0.4644*** | -0.3916*** |
| *DUMMY(Lagging regions)* | | | | |
| Parameter estimate | 0.1265*** | | 0.1237*** | 0.1196*** |
| Direct impact | – | | 0.1307*** | 0.1202*** |
| Indirect impact | – | | 0.0907** | 0.0181 |
| Total impact | – | | 0.2214*** | 0.1383*** |
| *DUMMY(Lagging regions)* Human resources in science and technology* | | | | |
| Parameter estimate | -0.5002*** | | -0.3911*** | -0.4690*** |
| Direct impact | – | | -0.4133*** | -0.4715*** |
| Indirect impact | – | | -0.2870*** | -0.0708* |
| Total impact | – | | -0.7002*** | -0.5423*** |
| *Spatial Lag* | – | | 0.4415*** | 0.1352** |
| **GOODNESS OF FIT STATISTICS** | | | | |
| $R^2$ | 0.7472 | | 0.8113 | 0.7582 |
| Spatial pseudo $R^2$ | – | | 0.7883 | 0.7525 |
| *AIC* | -314.465 | | – | – |
| **DIAGNOSTICS FOR SPATIAL DEPENDENCE** | | | | |
| | $W_{QUEEN}$ | $W_{TECH.PROX}$ | | |
| Moran's *I* (*Patent applications*) | 0.6731*** | 0.3107*** | – | – |
| Moran's *I* (OLS residuals) | 4.7391*** | 3.3453*** | – | – |
| Lagrange Multiplier (SAR) | 50.3119*** | 8.9227*** | – | – |
| Robust Lagrange Multiplier (SAR) | 31.9960*** | 4.4134** | – | – |
| Lagrange Multiplier (SEM) | 18.4882*** | 4.8214** | – | – |
| Robust Lagrange Multiplier (SEM) | 0.1724 | 0.3121 | – | – |
| Anselin–Kelejian test | – | – | 0.830 | 0.321 |

*Note*: Symbols ***, **, * indicate statistical significance at 1%, 5 % and 10 % level of significance, respectively.
Source: own calculations in RStudio and GeoDaSpace

Two realized estimates of the SAR spatial model point to the importance of all considered innovative determinants. However, the population density as innovation input did not have the expected positive sign of the parameter. R&D expenditure in the business sector appears to be the most significant innovation input, both in terms of direct impact and indirect impact. The models also included dummy variables to capture the diversity of the innovation process in lagging regions. The statistical significance of both dummy variables (in absolute and interaction forms) indicated the expected differences. The statistical significance of almost all indirect impacts confirms the important role of neighbouring regions in generating innovation in the region. Spatial innovation spillovers were indicated on the basis of both estimates, and thus not only the geographical proximity of the regions is likely to have an impact on innovation output but also technological similarity as a measure of the proximity of the regions is not negligible. However, in this case, the spatial regional interactions are considerably weaker, which is evident from the estimation of the spatial autoregressive parameter (see spatial lag parameter in Table 1). Subsequently, we can notice that all indirect impacts resulting from the estimate with the matrix $\mathbf{W}_{\mathbf{TECH.PROX}}$ are significantly weaker than these impacts obtained from the estimate with the matrix $\mathbf{W}_{\mathbf{QUEEN}}$. The geographical proximity of the regions and the resulting benefits for driving innovation seem to be a more important aspect than the similarity of the regions in terms of SMEs. Clusters of innovative companies, the proximity of research centres or universities allows their cooperation, sharing and faster dissemination of new knowledge and technologies.

# 4 CONCLUSIONS

The paper deals with spatial spillovers in terms of the European innovation problem. Based on 220 European regions classified at NUTS 1 or NUTS 2 level and the RIS 2019 dataset and the Eurostat regional statistical database, two SAR models were estimated. Patent applications were considered as a proxy of innovation output. In addition to the geographical proximity of the regions as a traditional measure of spatial interconnection, technological similarity as the proximity of the regions was also examined. Spatial innovation spilloves were indicated on the basis of both estimates, but we found that spatial regional interactions are significantly weaker when technological similarity is taken into account as the degree of proximity of regions.

**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"

**References**

[1] Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, 39, pp. 61–74.
[2] Charlot, S., Crescenzi, R. and Musolesi, A. (2014). Econometric modelling of the regional knowledge production function in Europe. *Journal of Economic Geography*, 15(6), pp.1227-1259.
[3] Eurostat (2021). Eurostat regional statistical database [online] Available at: http://ec.europa.eu/eurostat/ [Accessed 1 Jun 2021].
[4] Furková, A. (2019). Spatial spillovers and European Union regional innovation activities. *Central European Journal of Operations Research*, 27(3), pp.815-834.
[5] Hollanders, H., Es-Sadki, N. and Mekelbach I. (2019). Regional Innovation Scoreboard 2019. [online] Available at: https://ec.europa.eu/growth/sites/growth/files/ris2019.pdf [Accessed 1 March 2020].

[6] Khan, B. Z. (2012). Of Time and Space: A Spatial Analysis of Technological Spillovers among Patents and Unpatented Innovations in the Nineteenth Century. NBER Working Paper No. 20732. [online] Available at: http://www.nber.org/papers/w20732 [Accessed 11 Feb. 2015].
[7] Kumar, I. (2008). Innovation Clusters: A Study of Patents and Citations. In: *ESRI International User Conference*. [online] Available at: https://www.pcrd.purdue.edu/files/media/Innovation-Clusters-A-Study-of-Patents-and-Citations.pdf [Accessed 11 Jun 2015].
[8] LeSage, J. and Pace, K. (2009). *Introduction to Spatial Econometrics*. Boca Raton: Chapman and Hall/CRC.
[9] Moreno, R., Paci, R. and Usai, S., (2005a). Spatial Spillovers and Innovation Activity in European Regions. *Environment and Planning A: Economy and Space*, 37(10), pp.1793-1812.
[10] Moreno, R., Paci, R. and Usai, S., (2005b). Innovation clusters in the European regions. [online] Available at: http://crenosunicait/crenos/sites/default/files/wp/05-12pdf [Accessed 2 March 2017].
[11] Omobhude, C. and Chen, S. (2019). The Roles and Measurements of Proximity in Sustained Technology Development: A Literature Review. *Sustainability*, 11(1), pp. 224.

**Author's address**
doc. Ing. Andrea Furková, PhD.
University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics
Dolnozemská 1, 852 35 Bratislava
Slovak Republic
email: andrea.furkova@euba.sk

# MODELING OF UNCERTAINTY OF CONNECTION ARRIVAL USING FUZZY NUMBERS

***Dalibor Gonda, Tomáš Majer***
***University of Žilina, Faculty of Managements Science and informatics***

**Abstract**
The problem of planning in public bus transport is to arrange a given set of connections for a minimum number of rounds, which meet several additional restrictions. The planning of the arrival of the connection at a time of increasing traffic density is influenced by many factors, which can be classified into the category of random events. Random phenomena bring with them a degree of uncertainty that can be quantified using fuzzy numbers. In this paper, we present the theoretical basis of the application of fuzzy numbers in estimating the arrival of a connection, considering random variables that may affect the time required to implement the connection. We assume that considering the random factor when planning bus connections will contribute to optimizing the load on drivers and vehicles in the form of proportional load distribution.

***Keywords:*** *Fuzzy number, Connection, Optimization*

***JEL Classification:*** C610
***AMS Classification:*** 0C850, 65K05, 90C05

## 1 INTRODUCTION

In recent years, we have seen an ever-increasing number of cars on the road, especially near larger cities. The increase in individual population mobility in Slovakia was confirmed by Pucher and Buehler (2005). The reason for this negative phenomenon is, among other things, the expansion of the construction of residential houses in the narrower or wider surroundings of cities. One way to reduce road congestion is to increase the attractiveness of public transport through the gradual introduction of integrated transport (Mervart and Novák, 2020). Integrated transport means the harmonization of the operation of urban public transport, suburban transport, regional transport. This requires the optimization and coordination of public transport through unified timetables and a single transfer tariff. One of the effects of using integrated transport is to reduce energy consumption and emissions (Pettersson and Frisk, 2016). Integrated transport is relatively widespread, especially in the western part of the European Union, where the energy intensity of transport fell by 7.4% between 1995 and 2013, while it rose by 21.4% in the eastern part of the European Union (Andrés and Padilla, 2018). As we have already mentioned, an important element in the implementation of integrated transport is the planning of individual connections and the optimization of tours for individual vehicles. Within the framework of transport integration, it is necessary to minimize connection delays, especially at the interconnection points of various types of public transport. The establishment of a transport plan depends on various quantitative and qualitative criteria affecting the transport process, which in many cases cannot be precisely determined. This is due to the uneven flow of passengers, which affects the choice of transport services. Multicriteria analysis is an effective decision-making tool in choosing a suitable alternative depending on various quantitative and qualitative criteria (Stoilova and Munier, 2021). One way to handle uncertainty during the decision-making process is to apply fuzzy numbers. According to (Ďuriš, Bartková and Tirpáková, 2021), fuzzy numbers make it possible to reduce the number of variables causing uncertainty in the arrival of a connection while maintaining the maximum amount of information that the input data carries. Therefore, Fuzzy's theory became very popular and widely used, mainly due to its ability to work with inaccurate information relatively easily (Koláček and Matulová, 2018).

# 2 Methodology

## 2.1 Mathematical model

The ordered quadruple is called the connection $(cod_k, cpr_k, mod_k, mpr_k)$, where $cod_k$ is the departure time from the first stop $mod_k$ (departure point) and $cpr_k$ is the arrival time to the last stop $mpr_k$ (arrival point) of the connection $k$.

The individual elements of the matrix $M = \{m\ (u, v)\}$ represent the time required for the void transit of the bus from stop $u$ to stop $v$.

The connection $j$ may be operated by the same bus after the connection $i$ if:

$$cod_j \geq cpr_i + m(mpr_i, mod_j). \quad (1)$$

Thus, the connection $j$ can be operated by a bus which has stopped operating the connection $i$ at the stop $mpr_i$, if it manages to move to the stop $mod_j$ fast enough to catch the start of the operation of the connection $j$. We will call this that connection $j$ can follow connection $i$ and denote $i \prec j$.

The course of the vehicle is the sequence of connections $i_1, i_2, \dots, i_k,$ which can be operated by one bus. For the sequence of connections $i_1, i_2, \dots, i_k$ must apply

$$i_1 \prec i_2 \prec \cdots \prec i_k$$

The course is a one-day work schedule for one bus.

The relation of the connections' sequence $\prec$ can be represented by the digraph $\vec{G} = (S, E)$, where

$$E = \{(i, j) | i \in S, j \in S, i \prec j\}.$$

Let $S = \{1,2,3\dots, n\}$ be the set of connections. Then the binary variable $x_{ij}$ will express whether the connection $j$ will be $(x_{ij} = 1)$ or not be $(x_{ij} = 0)$ operated immediately after the connection $i$ by one bus in one course.

## 2.2 Minimizing the number of needed buses

The goal of optimization is to divide the set of connections S into as few courses as possible, so that each connection belongs to just one of them. This will achieve the smallest number of buses. A set of connections contains $n$ connections. If no connection could follow another, we would need $n$ buses, because each course could contain only one connection. This would be the case if all values of the variable $x_{ij}$ were equal to zero. It is obvious that the number of required buses will decrease with the number of cases when $x_{ij} = 1$, with each value $x_{ij} = 1$ we save one bus. Based on the above, the number of buses can be expressed in relation

$$p = n - \sum_{(i,j)\in E} x_{ij}. \quad (2)$$

Then the mathematical model for solving a problem with a minimum number of courses would have the form:

$$\text{Maximize} \quad \sum_{ij,(i,j)\in E} x_{ij} \quad (3)$$

Under conditions $$\sum_{i,(i,j)\in E} x_{ij} \leq 1\ for\ j = 1..n \tag{4}$$

$$\sum_{j,(i,j)\in E} x_{ij} \leq 1 \text{ for } i = 1..n \tag{5}$$

$$x_{ij} \in \{0,1\}\ for\ all\ \ i,j = 1..n \tag{6}$$

The mathematical model (3) to (6) is a model of the assignment problem that will give an integer solution even without condition (6), which can be replaced by the condition:

$$x_{ij} \geq 0\ for\ all\ \ i,j = 1..n. \tag{7}$$

The task of designing a course plan with a minimum number of buses without additional requirements is therefore solvable in polynomial time.

**2.3 Minimizing the length of void transits**

By solving the model (3) - (6) we found the required number of buses, let's denote it by the symbol T. If we force the maintenance of this number by adding a structural condition, we can change the purpose function and minimize the sum of empty passes. The model should then have the form:

Minimize $$\sum_{ij,(i,j)\in E} x_{ij}\, m\left(mpr_i, mod_j\right) \tag{8}$$

Under conditions $$\sum_{i,(i,j)\in E} x_{ij} \leq 1\ for\ j = 1..n \tag{9}$$

$$\sum_{j,(i,j)\in E} x_{ij} \leq 1\ for\ i = 1..n \tag{10}$$

$$\sum_{ij,(i,j)\in E} x_{ij} = |S| - T \tag{11}$$

$$x_{ij} \in \{0,1\}\ for\ all\ \ i,j = 1..n \tag{12}$$

**2.4 The concept of a fuzzy set**

Fuzzy sets can be interpreted as a set that assigns uncertainty to elements in the form of so-called degree of affiliation (Ďuračiová et al., 2011). Fuzzy set *A* is a set of elements $x \in U$ (*U* is a so-called universe), where each of them is assigned a degree of affiliation $\mu_A(x)$, whose values are in the interval 〈0; 1〉 . Fuzzy sets can identify with a function

$$\mu_A(x): U \to \langle 0; 1\rangle,$$

which we call the fuzzy set A membership function.

**Definition**: Let there be a set *U*, which we call a universe. Then a fuzzy set A∈U means a set that is defined by a representation

$$\mu_A: U \to \langle 0; 1\rangle.$$

For each x∈U we call the value $\mu_A(x)$µ the degree of belonging of the element *x* to the fuzzy set *A*.

Despite the possibility of identifying the fuzzy set *A* with the membership function $\mu_A$, we will distinguish between the fuzzy set *A* and its membership function. We will perceive the fuzzy set as an ordered pair, which we will formally write in the form

$$A = \{(x, \mu_A(x)); x \in U\}.$$

The shape and parameters of the membership function can be determined in individual cases based on practical experience or known properties of the analyzed phenomenon.
We will define other important terms that are needed to define the term fuzzy number.
**Definition**: Let *A* be a fuzzy set of elements x∈U. Then

- The sharp set S$upp\ A = \{x \in U; \mu_A(x) > 0\}$ is called the fuzzy set carrier;
- The sharp set $Ker\ A = \{x \in U; \mu_A(x) = 1\}$ we call the core of the fuzzy set,
- The sharp set $A_\alpha = \{x \in U; \mu_A(x) > \alpha\}$ we call the $\alpha$-section of a fuzzy set.

**2.5 Fuzzy numbers**
Most of us imagine a term number as some positive integer. A fuzzy number does not represent one specific value, but it is a notation for a fuzzy set if that set satisfies several conditions, such as that for at least one element *x* of the fuzzy set *A*

$$\mu_A(x) = 1.$$

**Definition**: Fuzzy set C defined on a set of real numbers *R* is called a fuzzy number if it meets the following properties:

(i) Core of the fuzzy set $C$ is nonempty, i.e., $Ker\ A \neq 0$;
(ii) $\alpha$-sections $C_\alpha$are closed intervals for all $\alpha \in (0, 1\rangle$;
(iii) $Supp\ C\ carrier$ is limited.

Theorem: Fuzzy set C is a fuzzy number just when there exist real numbers $x_1 \leq x_2 \leq x_3 \leq x_4$ such that for the membership function $\mu_C$ holds

$$\mu_C(x) = \begin{cases} L(x) & \text{for} \quad x \in (-\infty, x_2) \\ 1 & \text{for} \quad x \in \langle x_2, x_3 \rangle \\ P(x) & for \quad x \in (x_3, \infty) \end{cases}$$

where function $L: (-\infty, x_2) \to \langle 0; 1 \rangle$ is non-decreasing, continuous from the right and applies to it $L(x) = 0$ for $x \in (-\infty, x_1)$; function $P: (x_3, \infty) \to \langle 0; 1 \rangle$ j is non-increasing, continuous from left and applies to it $P(x) = 0$ for $x \in (x_4, \infty)$.
The real numbers $x_1, x_2, x_3, x_4$ are called significant values of the fuzzy number C and the following applies:

- the interval $\langle x_1, x_4 \rangle$ is the closure of the fuzzy carrier of the set C;
- the interval $\langle x_2, x_3 \rangle$ is a core of fuzzy set C

In solving our problem, we can assume that the indefinite input values will be from some closed interval 〈a, b〉. We will assume that $a \leq x_1 \leq x_2 \leq x_3 \leq x_4 \leq b$ holds, thus a narrower interval $\langle x_2, x_3 \rangle$ will be defined, the given value will occur and a wider interval $(x_1, x_4)$, outside which the occurrence of the given value is quite negligible. The simplest class, this way fuzzy numbers entered, are linear fuzzy numbers.
**Definition:** A linear fuzzy number on the interval $\langle a, b \rangle$,, which is determined by four points

$$[x_1, 0], [x_2, 1], [x_3, 1], [x_4, 0]$$

where $a \leq x_1 \leq x_2 \leq x_3 \leq x_4 \leq b$, we understand fuzzy number C, whose function of belonging has the form

$$\forall x \in \langle a,b \rangle : \mu_{\tilde{C}}(x, x_1, x_2, x_3, x_4) = \begin{cases} 0 & \text{pro } x < x_1, \\ \frac{x-x_1}{x_2-x_1} & \text{pro } x_1 \le x < x_2, \\ 1 & \text{pro } x_2 \le x \le x_3, \\ \frac{x_4-x}{x_4-x_3} & \text{pro } x_3 < x \le x_4, \\ 0 & \text{pro } x_4 < x. \end{cases}$$

If $x_2 = x_3$ we get the so-called triangular fuzzy number C - type number Λ. The triangular fuzzy number C can be written using a trio of numbers such as

$$C = (c_l, c_m, c_u),$$

where $c_l$ is the lower significant value or is also called the infimum, $c_m$ is the middle significant value or mode and $c_U$_is the upper significant value or supremum.

**3. THIRD ROUND OF OPTIMIZATION**

In the first round of optimization, courses were created to achieve the minimum number of buses needed. Subsequently, the model was improved to minimize the length of void bus crossings. The problem is that due to unexpected circumstances, some connections may be delayed. For each connection, the random variable of the size of the delay has a different distribution function i.e., this distribution function is location and time dependent (traffic jams only occur in some places and at different times). If the delayed connection is followed by another connection in the course without a sufficient break, the next connection will be missed from the very beginning. In this way, the delay could be passed on to other connections.

It used to be that there was a mandatory break between every two connections in the shift, e.g., 5 minutes, so-called connection handling. During this break, the driver had to check the bus, replace the information boards (when they were not yet electronic) and thus create a reserve to reduce delays. Condition (1) after taking the manipulation M into account should take the form

$$cod_j \ge cpr_i + m(mpr_i, mod_j) + M.$$

Excessive values for flat-rate manipulations reduce labor productivity and increase the required number of buses and drivers. Therefore, even the transport companies themselves do not comply with this condition in exceptional cases (if, for example, this saves the course). This is especially true during peak hours, when most connections run and at the same time the need for a break to cover possible delays is highest.

The goal of the next round of optimization is to evenly spread the risk of connection delays caused by delays of previous connections in the course. The arrival of the connection is burdened with a greater or lesser degree of uncertainty. We express this uncertainty by expressing the arrival time of the connection using a triangular fuzzy number (

$$C_{pr} = (c_o, c_t, c_n)$$

where $c_1$ will represent the tariff arrival of the connection, $c_0$ the expected possible arrival of the connection and $c_n$ the latest possible arrival of the connection. The value of $c_t$ is given by the timetable, the other two values are obtained by considering various factors that can cause delays. These factors can be:

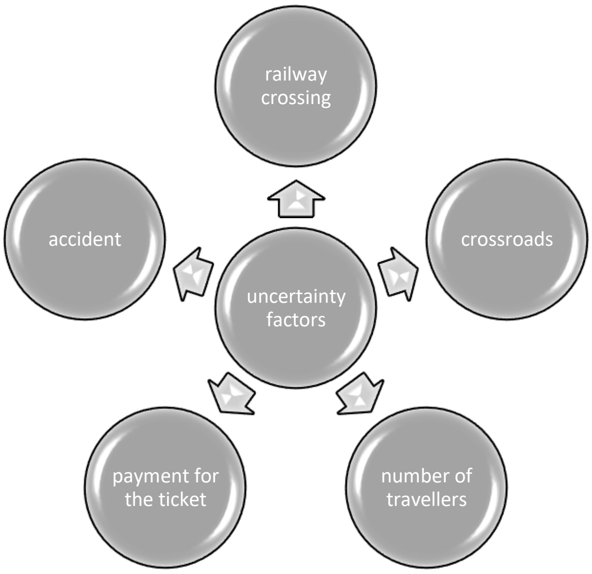



Figure 1 Factors causing uncertainty of the connection arrival

To obtain the data needed to quantify the uncertainty caused by these factors, it is necessary to use fuzzy numbers to analyze both connections geographically and demographically. The connection will be assigned by a parameter in the form of a fuzzy number

$$C_i = (c_t, c_o, c_n),$$

which will represent the delay of the i connection in the course expressed in minutes. Subsequently, for connections included in one course, the break between connections will be considered and for the selected fuzzy level of membership, courses with many delayed connections and courses with a small number of delayed connections at a certain level of risk - selected level of membership will be identified. Such courses will be heuristically crossed to minimize the risk of delays in high-risk courses at the expense of increasing the risk of delays in low-risk courses. For example, we have courses:

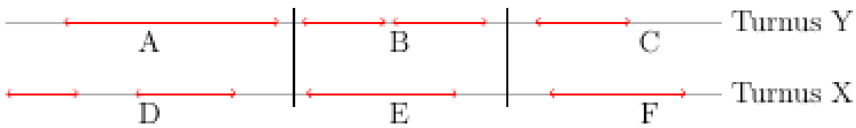



Figure 2 Crossing of the connections

Using fuzzy numbers, the course X was marked as a course with a low risk of delay, and course Y as a course with a high risk of delay. These courses are divided into sections that can be interchanged to create proposals for new courses. To solve this problem, we will use the Grouping Genetic Algorithm (GGA), which is described in detail in Mutingi and Mbohwa (2017). From the GGA we will use the method of crossing, in which two courses are first selected (based on the above-mentioned criteria), which are called parents. The selected solutions are combined, creating new solutions, also called descendants. From the descendants, the solution is selected which, based on the recalculation of delayed connections as a fuzzy criterion function, shows the lowest degree of risk of delay.

**4 CONCLUSIONS**

In our paper, we have introduced a model that uses fuzzy numbers to assess connection delays, while the delay of a particular connection will be affected by the delay of the previous connection in the course and the pause between connections. The goal will be to create courses so that we minimize the total number of delayed connections at a certain risk - the level of membership of the fuzzy number. Of course, without increasing the required number of buses and drivers and only with a slight increase in the size of void crossings.

If it is not possible to take a longer break after a connection where there is a high risk of delay and the delay is also transmitted to the second connection, it is appropriate to provide a longer pause at least after the second connection to reduce the risk of transmission of delay to the third connection. This is exactly what we will try to achieve by crossing course. Such a crossing can create or increase the size of the empty passage between the connections. However, the crossing also shortens the pause between connections and so the criterion of minimizing void passes is not in conflict with maximizing the pause between connections.

**Acknowledgements:** This work was supported by the research grants VEGA 1/0776/20 „Vehicle routing and scheduling in uncertain conditions“.

**References**

[1] Andrés, L. and Padilla, E. (2018). Driving factors of GHG emissions in the EU transport activity. *Transport Policy*, *61*, pp.60-74.
[2] Ďuračiová, R., Lieskovský, T., Kročková, K. and Sabo, M. (2011). Multicriteria Decision Making Using Fuzzy Sets in GIS Environment and Its Application in the Archaeological Prediction.
[3] Ďuriš, V., Bartková, R. and Tirpáková, A. (2021). Principal Component Analysis and Factor Analysis for an Atanassov IF Data Set. *Mathematics*, *9*(17), pp. 2067.
[4] Koláček, M. and Matulová, M. (2018). An application of hybrid fuzzy multi-criteria decision-making method on a location-selection problem. Letra Edu, s. r. o.
[5] Mervart, M. and Novák, R. (2020). Pražská a Středočeská integrovaná doprava. *Perner's Contacts*, *15*(2).
[6] Mutingi, M. and Mbohwa, C. (2017). Optimizing order batching in order picking systems: Hybrid grouping genetic algorithm. In *Grouping Genetic Algorithms* Springer, Cham, pp. 121-140.
[7] Pettersson, F. and Frisk, H. (2016). Soft space regional planning as an approach for integrated transport and land use planning in Sweden–challenges and ways forward. *Urban, Planning and Transport Research*, *4*(1), pp. 64-82.
[8] Pucher, J. and Buehler, R. (2005). Transport policies in central and Eastern Europe. *Transport strategy, policy, and institutions*, pp. 860.
[9] Stoilova, S. and Munier, N. (2021). A novel fuzzy SIMUS multicriteria decision-making method. An application in railway passenger transport planning. *Symmetry*, *13*(3), pp. 483.

**Author's address**

Ing. Tomáš Majer, PhD.
University of Žilina, Faculty of Managements Science and informatics,
Univerzitná 8215/1, 01026 Žilina
Slovak republik
Tel: +421-41-513 4251
email: tomas.majer@fri.uniza.sk

doc. PaedDr. Dalibor Gonda, PhD.
University of Žilina, Faculty of Managements Science and informatics,
Univerzitná 8215/1, 01026 Žilina
Slovak republik
Tel: +421-41-513 4251
email: dalibor.gonda@fri.uniza.sk

# DEA CLUSTERING AND MONTE CARLO APPROACHES

*Jakub Hanousek, Prague University of Economics and Business*

**Abstract**
Data envelopment analysis has some relatively strong assumptions for proper application. We discuss violating two assumptions. The first assumption is homogeneity between (across) production units. The second assumption is a need for a sufficient amount of production units. These problems are represented on data of the resource´s capacities against Covid – 19 in the Czech Republic. The production units in this article are the regions in the Czech Republic. The regions in the Czech Republic are very different and maybe isn´t fair to compare all regions in one model as there is a potential heterogeneity problem. Our idea is to cluster the regions in similar clusters. Data envelopment analysis models are then applied separately for each cluster. The second problem is the number of production units in the clusters. The clusters with a low number of production units gave skewed results. We try to solve this problem by the Monte Carlo simulation in the article.

***Keywords:*** *DEA, Monte Carlo, Clustering*

***JEL Classification:*** C44
***AMS Classification:*** 90–11

## 1 INTRODUCTION

Data envelopment analysis is a method based on linear programming used to measure the efficiency of production units. The method of data envelopment analysis has some assumptions, which should be complied with. This article deals with breaches two of this assumption. The assumption of homogeneity of production units and the assumption of sufficient amount of production units for calculating undistorted results. These problems are illustrated in the data of hospital capacities for patients with Covid – 19 in the Czech Republic. The production units are the regions in the Czech Republic. The input is the number of inhabitants in each region. The outputs are the number of beds with oxygen, the number of beds with HFNO and the number of beds with ALV.

## 2 DATA ENVELOPMENT ANALYSIS

DEA was developed by Charnes, Cooper and Rhodes in 1978 (Charnes et all., 1978) 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 defined as the ratio of its total weighted output to its total weighted input or, vice versa, as the ratio of its total weighted input to its total weighted 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. The production frontier represents the maximum amounts of output that is produced by given amounts 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 (Dlouhý and Hanousek, 2019), (Dlouhý and Jablonský, 2004).

### 2.1 Output oriented model with constant return of scale

$$\text{Maximize: } \boldsymbol{\varphi}_\mathbf{q} + \boldsymbol{\varepsilon}\left(\mathbf{e}^\mathbf{T}\mathbf{s}^+ + \mathbf{e}^\mathbf{T}\mathbf{s}^-\right),$$

$$\text{Subject to: } \begin{aligned} &\mathbf{X\lambda} + \mathbf{s}^- = \mathbf{x}_\mathbf{q}, \\ &\mathbf{Y\lambda} - \mathbf{s}^+ = \boldsymbol{\varphi}_\mathbf{q}\mathbf{y}_\mathbf{q}, \\ &\boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^- \geq \mathbf{0}. \end{aligned} \qquad (1)$$

$\boldsymbol{\varphi}_\mathbf{q}$ is a variable which represents efficiency rate of a production unit. $\mathbf{x}_\mathbf{q}$ is a vector of inputs of an optimized production unit. $\mathbf{y}_\mathbf{q}$ is a vector of outputs of an optimized production unit. $\boldsymbol{\varepsilon}$ is an infinitesimal constant. The infinitesimal constant $\boldsymbol{\varepsilon} = 10^{-8}$. $\mathbf{e}^\mathbf{T} = (1,1,1...,1)$. $\mathbf{s}^+$ and $\mathbf{s}^-$ are vectors of additional variables. $\mathbf{X}$ is a matrix of inputs. $\mathbf{Y}$ is a matrix of outputs. $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_n)$ is a vector of weights which are assign to productions units. Weights are the variables in a model (Dlouhý and Jablonský, 2004).

### 2.2 Output oriented model with variable return of scale

$$\text{Maximize: } \boldsymbol{\varphi}_\mathbf{q} + \boldsymbol{\varepsilon}\left(\mathbf{e}^\mathbf{T}\mathbf{s}^+ + \mathbf{e}^\mathbf{T}\mathbf{s}^-\right),$$

$$\text{Subject to: } \begin{aligned} &\mathbf{X\lambda} + \mathbf{s}^- = \mathbf{x}_\mathbf{q}, \\ &\mathbf{Y\lambda} - \mathbf{s}^+ = \boldsymbol{\varphi}_\mathbf{q}\mathbf{y}_\mathbf{q}, \\ &\mathbf{e}^\mathbf{T}\boldsymbol{\lambda} = \mathbf{1}, \\ &\boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^- \geq \mathbf{0}. \end{aligned} \qquad (2)$$

Model with variable return of scale has one extra condition compared with model constant return of scale. The extra condition is condition of convexity $\mathbf{e}^\mathbf{T}\boldsymbol{\lambda} = 1$ (Dlouhý and Jablonský, 2004).

## 3 CLUSTERING APPROACH

Homogenous production units are one of the premises in the application of DEA models. We know lots of examples in a real-life that the homogenous of production units isn´t fully complied. For example, small and big companies have other production opportunities. Socioeconomics differences in the regions. Regions with big cities are more equipped than countryside regions in many ways. One of the possible solutions is clustering production units into more similar clusters.

We will transform our case into a network case. Each off production is one node. The distance between nodes is a difference between the output or input of the production unit. Now we can use an optimization algorithm for network optimization or a heuristic algorithm for network optimization. The partial paths or the partial cycles from this optimization are your clusters.

**Modified nearest neighbor method**

1. **Step 1:** Define the input or output according to which we cluster.
2. **Step 2:** Calculate the differences between all production units according to the input or output which we cluster.
3. **Step 3:** Define the maximum value of the partial path.
4. **Step 4:** Start the path in the node which represent the smallest value of input or output according to which we cluster (not yet selected).
5. **Step 5:** Find the nearest node (not yet selected) to the node on the path and add to the end of the path. If the path is higher than the maximum path. Go to step 4. If such a node doesn't exist (all nodes have already been selected). Go to step 6.
6. **Step 6:** Finish.

## 4 MONTE CARLO APPROACH

The number of the production units should be at least three times more than a number of inputs and outputs. If we have least the production units, the results are not so relevant. Many of the production units come out efficient.

This pitfall is solved by the Monte Carlo simulation in this article. We create fictive production units. We use a pseudorandom number generator to generate pseudorandom numbers in interval $\langle 0,1 \rangle$ from the standardized normal distribution $N(0,1)$. Then we subtraction the value $-0.5$ from each pseudorandom number. This gives us the pseudorandom numbers from interval $\langle -0.5, 0.5 \rangle$. We multiply the pseudorandom number by the value of 10% from each input and output of each production units. (Dlouhý and Fábry, 2007), (Langville and Meyer, 2012).

## 5 APPLICATION

The data comes from the Institute of Health Information and Statistics of the Czech Republic. There is a potential problem with homogeneity in the regions. The regions with a lot of inhabitants include large cities with big hospitals. The big hospitals have bigger capacities than smaller countryside hospitals.

Table 1: Data

| Region | Population | Hospital bed with oxygen | Hospital bed with HFNO | Hospital bed with ALV |
|---|---|---|---|---|
| **Hlavní město Praha** | 1 309 000 | 3 154 | 312 | 473 |
| **Středočeský** | 1 369 000 | 1886 | 122 | 104 |
| **Jihočeský** | 642 000 | 1634 | 79 | 64 |
| **Plzeňský** | 585 000 | 1 609 | 104 | 146 |
| **Ústecký** | 821 000 | 1505 | 146 | 121 |
| **Královéhradecký** | 551 000 | 1402 | 116 | 114 |
| **Jihomoravský** | 1 189 000 | 2524 | 197 | 235 |
| **Moravskoslezský** | 1 203 000 | 2824 | 200 | 309 |
| **Karlovarský** | 295 000 | 418 | 40 | 44 |
| **Liberecký** | 442 000 | 813 | 20 | 81 |
| **Pardubický** | 520 000 | 889 | 69 | 62 |
| **Vysočina** | 509 000 | 1739 | 37 | 62 |
| **Zlínský** | 583 000 | 902 | 91 | 92 |
| **Olomoucký** | 633 000 | 1458 | 52 | 139 |

(Český statistický úřad, 10.12.2021), (Ministerstvo zdravotnictví České republiky, 10.12.2021)

## 5.1 Clustering approach

This approach clusters the regions in the Czech Republic in the more homogenous group according to the population. We will work in this section with the abbreviation of regions because region abbreviation is more useful and more practical. Region abbreviations are in Table 2.

Table 2: Region abbreviation

| Region | Region abbreviation |
|---|---|
| **Hlavní město Praha** | A |
| **Plzeňský** | P |
| **Královéhradecký** | H |
| **Karlovarský** | K |
| **Vysočina** | J |
| **Moravskoslezský** | T |
| **Jihomoravský** | B |
| **Jihočeský** | C |
| **Olomoucký** | M |
| **Liberecký** | L |
| **Ústecký** | U |
| **Zlínský** | Z |
| **Pardubický** | E |
| **Středočeský** | S |

The first differences in the absolute value between regions are in Table 3. We can say that the regions are the nodes and the first difference between regions are the length of an edge. Now, we can solve the clustering problems as an optimization network problem to find the shortest partial path between nodes.

Table 3: Distance between nodes in thousand

| **Region** | A | S | C | P | U | H | B | T | K | L | E | J | Z | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0 | 60 | 667 | 724 | 488 | 758 | 120 | 106 | 1014 | 867 | 789 | 800 | 726 | 676 |
| S | 60 | 0 | 727 | 784 | 548 | 818 | 180 | 166 | 1074 | 927 | 849 | 860 | 786 | 736 |
| C | 667 | 727 | 0 | 57 | 179 | 91 | 547 | 561 | 347 | 200 | 122 | 133 | 59 | 9 |
| P | 724 | 784 | 57 | 0 | 236 | 34 | 604 | 618 | 290 | 143 | 65 | 76 | 2 | 48 |
| U | 488 | 548 | 179 | 236 | 0 | 270 | 368 | 382 | 526 | 379 | 301 | 312 | 238 | 188 |
| H | 758 | 818 | 91 | 34 | 270 | 0 | 638 | 652 | 256 | 109 | 31 | 42 | 32 | 82 |
| B | 120 | 180 | 547 | 604 | 368 | 638 | 0 | 14 | 894 | 747 | 669 | 680 | 606 | 556 |
| T | 106 | 166 | 561 | 618 | 382 | 652 | 14 | 0 | 908 | 761 | 683 | 694 | 620 | 570 |
| K | 1014 | 1074 | 347 | 290 | 526 | 256 | 894 | 908 | 0 | 147 | 225 | 214 | 288 | 338 |
| L | 867 | 927 | 200 | 143 | 379 | 109 | 747 | 761 | 147 | 0 | 78 | 67 | 141 | 191 |
| E | 789 | 849 | 122 | 65 | 301 | 31 | 669 | 683 | 225 | 78 | 0 | 11 | 63 | 113 |
| J | 800 | 860 | 133 | 76 | 312 | 42 | 680 | 694 | 214 | 67 | 11 | 0 | 74 | 124 |
| Z | 726 | 786 | 59 | 2 | 238 | 32 | 606 | 620 | 288 | 141 | 63 | 74 | 0 | 50 |
| M | 676 | 736 | 9 | 48 | 188 | 82 | 556 | 570 | 338 | 191 | 113 | 124 | 50 | 0 |

**Modified nearest neighbor method**

The classical nearest neighbor method doesn´t define in which node you must start the optimization. In our case, the distance between nodes is the difference between regions. We define the first node as a region with the smallest population. From this node, we find the nearest path between nodes. We can define the maximum length of the partial path. We define a maximum of the partial path to 300 000. The border of 300 000 was chosen due to a number of inhabitants in the Czech Republic. The bigger regional city in the Czech Republic has about 300 000 inhabitants.

**Results of the modified nearest neighbor method**

We applied the modified method of nearest neighbor to find the partial nearest path between the number of inhabitants in the region of the Czech Republic. The partial nearest path must be less than 300 000. The results are in equation 3.

$$\begin{gathered}
\overset{0}{K} \xrightarrow{147} \overset{147}{L} \xrightarrow{67} \overset{214}{J} \xrightarrow{11} \overset{225}{E} \xrightarrow{31} \overset{256}{H} \xrightarrow{32} \overset{288}{Z} \xrightarrow{2} \overset{290}{P} \\
\overset{0}{M} \xrightarrow{9} \overset{9}{C} \xrightarrow{179} \overset{188}{U} \\
\overset{0}{B} \xrightarrow{14} \overset{14}{T} \xrightarrow{106} \overset{120}{A} \xrightarrow{60} \overset{180}{S}
\end{gathered} \tag{3}$$

The algorithm of the modified nearest neighbor method gave us results for 3 partial paths. These 3 paths represent 3 clusters.

We applied DEA optimization for the first cluster. The results are in Table 4. We can see that we have a low number of production units. As a consequence a lot of production units become efficient.

Table 4: Results of first Cluster CCR and BCC

| Region | Score Eff. CCR in cluster | Score Eff. CCR between all regions | Region | Score Eff. BCC in cluster | Score Eff. BCC between all regions |
|---|---|---|---|---|---|
| **J** | 1.00 | 1.00 | **K** | 1.00 | 1.00 |
| **P** | 1.00 | 1.01 | **J** | 1.00 | 1.00 |
| **H** | 1.00 | 1.01 | **P** | 1.00 | 1.00 |
| **Z** | 1.33 | 1.53 | **H** | 1.00 | 1.00 |
| **L** | 1.36 | 1.51 | **L** | 1.18 | 1.21 |
| **K** | 1.48 | 1.72 | **Z** | 1.26 | 1.37 |
| **E** | 1.52 | 1.53 | **E** | 1.48 | 1.48 |

### 5.2 Monte Carlo approach

The number of production units, which we gained by clustering is low for DEA optimization. We simulate new fictive production units by Monte Carlo simulation.

We will analyse the first cluster. We simulate each production unit ten times. Every input and out are simulated by the original pseudorandom number. We get 70 new fictive production units by this approach. Then we applied DEA optimization. We repeat this operation 100 times. We get results from one hundred data sets. We average the results to one result.

Results of the BCC output-oriented model with fictive production units are in Table 4. We can see, original production units are inefficient.

Table 5: Comparison of the results

| Region | Score Eff. CCR | Score Eff fictive CRR | Region | Score Eff. BCC | Score Eff. Fictive BCC |
|---|---|---|---|---|---|
| **J** | 1.00 | 1.52 | **K** | 1.00 | 1.12 |
| **P** | 1.00 | 1.33 | **J** | 1.00 | 1.05 |
| **H** | 1.00 | 1.45 | **P** | 1.00 | 1.19 |
| **Z** | 1.33 | 1.84 | **H** | 1.00 | 1.52 |
| **L** | 1.36 | 1.98 | **L** | 1.18 | 1.88 |
| **K** | 1.48 | 2.12 | **Z** | 1.26 | 2.05 |
| **E** | 1.52 | 1.55 | **E** | 1.48 | 1.18 |

Results of original CCR and BCC oriented models and CCR and BCC output-oriented models with fictive production units are compared in table 5. The majority of origin production scolded your score in models with fictive production units.

## 6 CONCLUSIONS

This article compared hospital resources against disease Covid – 19 in the regions of the Czech Republic. The hospital resources are the number of beds with oxygen, the number of beds with HFNO and the number of beds with ALV. Article deals with the non-homogeneous environment. Non-homogeneous environment means in this article the number of inhabitants in the region. Because there is a premise that the high equipped hospitals are in the big cities. The big cities are in the regions with a high number of inhabitants.

We clustered the region to a more similar cluster according to the number of inhabitants by the modified nearest neighbor method. Then we applied DEA optimization. There is a second pitfall, that there is a broken premise of the sufficient amount of the production units in the cluster. This article suggests this pitfall solve by Monte Carlo simulation. We simulated fictive production units by Monte Carlo simulation. Then we applied DEA optimization with sufficient fictive production units.

Not one of the original production units is efficient in optimization with fictive production units. Most of the production units have lower score. Some of the production units improve up the score, but a minimum of them. For example region Pardubice improve up the score from 1.48 to 1.18 in the BCC output oriented-model. We can discuss why.

## References


[1] Dlouhy M.and Fábry J. (2007). *Simulace podnikových procesů.* Praha: Computer Press.

[2] Dlouhý, M. and Hanousek, J. (2019). An Assesment of Regional Variations: An Application 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

[3] Charnes A., Cooper W. W. and Rhodes E. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research,* 2, pp.429– 444.

[4] Jablonský J. and Dlouhý, M. (2004). *Modely hodnocení efektivnosti produkčních jednotek*. Praha: Professional Publishing.

[5] Langville N.A. and Meyer D.C. (2012). *Who´s # 1? The Science of Rating and Ranking.* USA(New Jersey): Princeton University Press.

[6] Český statistický úřad, [online] Available at: https://www.czso.cz/csu/czso/obyvatelstvo_lide [Accessed 10.12.2021]

[7] Ministerstvo zdravotnictví České republiky [online] Available at: https://onemocneni-aktualne.mzcr.cz/kapacity-luzkove-pece [Accessed 10.12.2021]


## Author's address


Ing. Jakub Hanousek
Prague University of Economics and Business, Faculty of Informatics and Statistics, Department Econometrics
Nám Winstona Churchilla 1938/8, 130 67 Praha-3 Žižkov
Czech Republic
email: xhanj52@vse.cz

# SPATIAL MODEL WITH THE INFLUENCE OF THE REGULATOR ON THE DECISIONS OF DUOPOLISTS

*Patrícia Holzerová, Zuzana Čičková, University of Economics in Bratislava*

**Abstract**
The consideration of space in economic analysis and economic theory is continuously gaining more attention. This is also important from the individual point of view of the company on its operation in the market, in which they try to operate as efficiently as possible. Therefore, they have to make strategic decisions, whether in the location of their operation or financing of their activities. With the need of financial support, companies often turn to the state. However, state aid is important not only for companies, but also for the provider itself as a tool to pursue its own interest. In this paper we focus on the situation in which companies deciding on their location of their branches are influenced by the regulator as authority trying to achieve its own goals, which is the preference of a particular location among the possible ones. He chooses effective tools for this goal. The subject of our paper is to present model of spatial competition using game theory, effective due to the competitive character of the presented situation. The result is information on the final decisions of companies, amount of financial support as the first tool used and the costs of the media campaign as the second tool of the regulator.

***Keywords:*** *Spatial competition, Regulator, Subsidy, Multicriteria game*

***JEL Classification:*** C62, C72
***AMS Classification:*** 90C29, 91A05, 91A10

## 1 INTRODUCTION

The company's operation in the market and its success is determined by a number of factors accompanied by its strategic decisions. These include the choice of location, whether of branches or production facilities. The issue of locating entities in the market is addressed by location models, which are closely related to the concept of spatial competition. It transfers the competitive environment to space, that is gaining more and more attention in economic theory. There are several approaches to it. Combining spatial competition with game theory, which is a tool for analyzing strategic behavior, is a specific and very promising topic for the future.

One of the first to deal with the issue of spatial competition was the mathematician and economist Harold Hotelling (1929), who presented a model based on the presence of two companies looking for the most advantageous position in the linear market. The model is the basis of a number of theories of product differentiation and location, as evidenced by publications such as (C. D'Aspremont et al., 1979), (Graitson, 1982), (Anderson, 1988) or more recent work by (Goryunov et al., 2017). Despite the fact that the beginnings of the issue of spatial models are associated with Hotelling and his work is followed by the largest number of authors, in fact, the first known attempt to analyze economic activity in space is associated with the German J.H. von Thunen (1826). His theory explained the location of production activities in an isolated city-state with land and homogeneous resources (Gehling, 1965). Another was Weber (1909), who developed the theory of the location of industry (Fearon, 2002).

When it comes to the most important factors influencing the success of companies in the market, method of financing is as important as the location. Sufficient funds enable the implementation of business plans and the achievement of the goals of new entities without major complications.

The financial situation of companies is also affected by a number of determinants, the unfavorable development of which may negatively affect their further operation. In such cases, companies can also turn to the state. However, in addition to the aid, state aid may distort competition and disadvantage of certain entities, as this is one of the ways in which the state can influence and control the market. Who the state provides aid within the set criteria is therefore affected by its preferences and intentions.

# 2 MODEL WITH THE INFLUENCE OF THE REGULATOR

In our paper we will present an original mathematical model in which two players (companies) decide on the locations of their branches at known prices $p_1$ and $p_2$ of products they offer. The products are homogeneous and in unlimited quantities. The basic idea of spatial game is based on (Lopez and Čičková 2018). Let $V = \{1,2, \dots n\}, n \in Z^+$ be the set of customers and let there be graph $G = (V, H)$ where $V$ represents nodes of the graph and $H \subset VxV$ represents set of the edges $h_{ij} = (v_i, v_j)$ from node $v_i$ to node $v_j$, while for each oriented edge $h_{ij}$ there is assigned real number $d_{ij}$ referred to as a valuation or value $h_{ij}$, representing the minimum distance (the shortest path length) between the nodes $v_i$ and $v_j$. The matrix $\mathbf{D}_{nxn} = \{d_{ij}\}$ is then the matrix of the shortest distances between the nodes $v_i$ and $v_j$. Spatial game was formulated in so-called full-valued graph $\bar{G} = (V, \bar{H})$ with the same set of nodes as graph $G$, where $\bar{H}$ is set of the edges between each pair of nodes $v_i$ and $v_j$, while their valuation is equal to the minimum price between nodes $v_i$ and $v_j$ of the original graph, $i, j \in V$. We also consider a constant (unit) demand at each node, while each customer makes a purchase from any company (service is always carried out). When choosing a company, customers consider the total cost of purchasing the product, which consists of the price of the product and the cost of transportation to the selected company. Transport costs are expressed as $t$ per unit distance. If player 1 places his store in the $i$ −th node ( $i \in V$) and player 2 places his store in the $j$ −th node ( $j \in V$), player 1 gets the customer from the $k$ −th node ($k \in V$) only if $t * d_{ki} + p^{(1)} < t * d_{kj} + p^{(2)}$, otherwise the customer from the $i$ −th node is served by player 2. If $t * d_{ki} + p^{(1)} = t * d_{kj} + p^{(2)}$, players share the demand equally.
The situation is extended by the assumption that regulator enters the decision-making process of the players as an authority pursuing its own intentions. In this case, its intention is to motivate the companies to build branches in its preferred areas. The preferred and unpreferred areas are represented by sets of nodes $V^{(p)} \subseteq V$ and $V^{(u)} \subseteq V$, $V = V^{(p)} \cup V^{(u)}$. The regulator uses effective tools for the motivation of the companies. In our case we assume financial subsidy and also influencing consumers, for example by advertising, which influences the solution of the game. As these tools represent costs for the regulator, its goal is to minimize them. The solution of such a situation will take place in a total of three consecutive stages. Each of these stages is given by a criterion in the form of a goal that the regulator is trying to achieve. In the first phase, it is important for the regulator to ensure the establishment of branches of companies in its preferred areas using the mentioned tools. Subsequently, in the second phase, it tries to minimize the number of affected (conscious) customers following its recommendations, which are associated with the cost of advertising it uses. In the last phase, it tries to minimize the amount of provided subsidies. We will present the phases formulated in this way in the following model, in which the first goal is infinitely more important than the second and this in turn is infinitely more important than the third one. The presented situation is formulated as a goal programming problem.

The application of the regulator's tools and the amount of costs associated with them depends on the final choice of companies, i.e. whether they decide to build their branches in preferred areas. From this point of view, a total of 3 cases can occur:

1. both companies decide to build their branches in preferred nodes, and thus the regulator will provide a price subsidy to both companies,
2. only one of the companies decides to build its branch in the preferred nodes, and it will be provided with a subsidy,
3. none of the companies decides to build its branch in the preferred nodes, and thus regulator will not provide subsidy to any of the companies.

That means that the final price at which consumers will buy the offered products depends on the final choice of the location. The above situation, where player 1 chooses node $i$, player 2 chooses node *j* and under the conditions of distinguishing whether these are preferred locations, it is possible to write:

$$t * d_{kj} + p_2 - c - (t * d_{ki} + p_1 - c) = l_{kij}; k, i \in V^{(p)}, j \in V^{(p)} \tag{1}$$

$$t * d_{kj} + p_2 - (t * d_{ki} + p_1 - c) = l_{kij}; k, i \in V^{(p)}, j \in V^{(u)} \tag{2}$$

$$t * d_{kj} + p_2 - c - (t * d_{ki} + p_1) = l_{kij}; k, i \in V^{(u)}, j \in V^{(p)} \tag{3}$$

$$t * d_{kj} + p_2 - (t * d_{ki} + p_1) = l_{kij}; k, i \in V^{(u)}, j \in V^{(u)} \tag{4}$$

where $V^{(p)} \subseteq V$ is the set of preferred nodes, $V^{(u)} \subseteq V$, is the set of non-preferred nodes, $V = V^{(p)} \cup V^{(u)}$, $c$ is the financial subsidy provided by regulator and $l_{kij}$ represents the difference in the costs borne by the consumer from the *k*th node in choosing the player from whom to buy (player 1 chooses node $i$, player 2 chooses node *j*).

(1) occurs when both players decide to build their branches in preferred nodes,
(2) occurs when only player 1 decides to build his branch in the preferred node,
(3) occurs when only player 2 decides to build his branch in the preferred node,
(4) occurs when none of the players decides to build their branch in the preferred node.

The assignment of the customer to the player can be realized using the Signum function, the problem of which is solved by using the binary variables $b_{kij}^{(1)} \in \{0,1\}; k, i, j \in V$, $b_{kij}^{(2)} \in \{0,1\}; k, i, j \in V$, continuous variable $b_{kij} \in \langle -1,1 \rangle;\ k, i, j \in \mathrm{V}$ and constraints

$$l_{kij} \leq M * b_{kij}^{(1)}; k, i, j \in V \tag{5}$$

$$l_{kij} \geq -M * b_{kij}^{(2)}; k, i, j \in V \tag{6}$$

Constraint (5) ensures that if the difference in cost $l_{kij} > 0$, $b_{kij}^{(1)} = 1$. Constraint (6) on the other hand ensures that if $l_{kij} < 0$, $b_{kij}^{(2)} = 1$. The situation in which the following two cases could occur at the same time is treated by:

$$b_{kij}^{(1)} + b_{kij}^{(2)} \leq 1; k, i, j \in V \tag{7}$$

Using of variables $b_{kij} \in \langle -1,1 \rangle;\ k, i, j \in \mathrm{V}$ serves for calculation of elements $a_{ij}$:

$$b_{kij} = b_{kij}^{(1)} - b_{kij}^{(2)}; k, i, j \in V \tag{8}$$

$$b_{kij}^{(1)} * l_{kij} \geq \varepsilon * b_{kij}^{(1)}; k, i, j \in V \tag{9}$$

$$b_{kij}^{(2)} * l_{kij} \leq -\varepsilon * b_{kij}^{(2)}; k, i, j \in V \tag{10}$$

$$s_{kij} = \frac{b_{kij} + 1}{2}; k, i \in V^{(p)}, j \in V^{(p)} \tag{11}$$

$$s_{kij} = \frac{b_{kij} + 1}{2} * (1 - \lambda) + \lambda; k, i \in V^{(p)}, j \in V^{(u)} \tag{12}$$

$$s_{kij} = \frac{b_{kij} + 1}{2} * (1 - \lambda); k, i \in V^{(u)}, j \in V^{(p)} \tag{13}$$

$$s_{kij} = \frac{b_{kij} + 1}{2}; k, i \in V^{(n)}, j \in V^{(n)} \tag{14}$$

$$a_{ij} = \sum_{k \in V} s_{kij}, i, j \in V \tag{15}$$

Constraints (9) and (10) handle the situation when $l_{kij} = 0$. At the same time, elements $a_{ij}$ depend on the number of customers affected by the regulator, which consider the constraints (11)-(15).

Let us now recapitulate the model solving the problem in which the controller pursues its goals and which is given by parameters and sets:

- $n \in Z^+$ – number of nodes,
- $V = \{1,2, \dots n\}$ – set of all nodes,
- $V^{(p)} \subseteq V$ – set of preferred nodes,
- $V^{(u)} \subseteq V$, $V = V^{(p)} \cup V^{(u)}$- set of non-preferred nodes,
- $d_{ij} \geq 0, i, j \in V$ – the shortest distance between nodes $i$ and $j$,
- $t > 0$ – unit transport costs,
- $p_1 > 0$ – price of the product offered by the player 1,
- $p_2 > 0$ – price of the product offered by the oponent (player 2),
- $\lambda^H$ – maximum percentage of customers that the regulator is able to influence,
- $c^H$ – upper bound of financial subsidy provided by regulator,
- $M$ – big positive number,
- $\varepsilon$ − small positive number,

and variables:

- $w_1 \in \langle 0, n \rangle$ – value of the game of player 1,
- $w_2 \in \langle 0, n \rangle$ – value of the game of player 2,
- $x_i \in \langle 0,1 \rangle$, $i \in V$ – $i$th mixed strategy of player 1,
- $y_j \in \langle 0,1 \rangle$, $j \in V$ – $j$th mixed strategy of player 2,
- $a_{ij} \in \langle 0, n \rangle$ , $i, j \in V$ – payment matrix of player 1,
- $b_{kij}^{(1)} \in \{0,1\}; k, i, j \in V$ – binary variable,
- $b_{kij}^{(2)} \in \{0,1\}; k, i, j \in V$ – binary variable,
- $b_{kij} \in \langle -1,1 \rangle;\ k, i, j \in V$ – continuous variable,
- $z_{ij} \in \{0,1\};\ i, j \in V$ – binary variable,
- $\lambda \in \langle 0, \lambda^H \rangle$ – percentage of conscious consumers influenced by regulator,
- $c \in \langle 0; c^H \rangle$ – regulator subsidy,
- $d_1^-$ – first level deviation,
- $l_{kij}$ – difference in costs compared to the player 2,
- $s_{kij}$ – auxiliary variable for determining elements $a_{ij}$.

The final model is:

$$d_1^- \to \min \tag{16}$$

$$\sum_{i \in V^{(p)}} x_i + \sum_{j \in V^{(p)}} y_j + d_1^- = 2 \tag{17}$$

$$t * d_{kj} + p_2 - c - (t * d_{ki} + p_1 - c) = l_{kij}; k, i \in V^{(p)}, j \in V^{(p)} \tag{18}$$

$$t * d_{kj} + p_2 - (t * d_{ki} + p_1 - c) = l_{kij}; k, i \in V^{(p)}, j \in V^{(u)} \tag{19}$$

$$t * d_{kj} + p_2 - c - (t * d_{ki} + p_1) = l_{kij}; k, i \in V^{(u)}, j \in V^{(p)} \tag{20}$$

$$t * d_{kj} + p_2 - (t * d_{ki} + p_1) = l_{kij}; k, i \in V^{(u)}, j \in V^{(u)} \tag{21}$$

$$l_{kij} \le M * b_{kij}^{(1)}; k, i, j \in V \tag{22}$$

$$l_{kij} \ge -M * b_{kij}^{(2)}; k, i, j \in V \tag{23}$$

$$b_{kij}^{(1)} + b_{kij}^{(2)} \le 1; k, i, j \in V \tag{24}$$

$$b_{kij} = b_{kij}^{(1)} - b_{kij}^{(2)}; k, i, j \in V \tag{25}$$

$$b_{kij}^{(1)} * l_{kij} \ge \varepsilon * b_{kij}^{(1)}; k, i, j \in V \tag{26}$$

$$b_{kij}^{(2)} * l_{kij} \le -\varepsilon * b_{kij}^{(2)}; k, i, j \in V \tag{27}$$

$$s_{kij} = \frac{b_{kij} + 1}{2}; k, i \in V^{(p)}, j \in V^{(p)} \tag{28}$$

$$s_{kij} = \frac{b_{kij} + 1}{2} * (1 - \lambda) + \lambda; k, i \in V^{(p)}, j \in V^{(u)} \tag{29}$$

$$s_{kij} = \frac{b_{kij} + 1}{2} * (1 - \lambda); k, i \in V^{(u)}, j \in V^{(p)} \tag{30}$$

$$s_{kij} = \frac{b_{kij} + 1}{2}; k, i \in V^{(n)}, j \in V^{(n)} \tag{31}$$

$$a_{ij} = \sum_{k \in V} s_{kij} \, ; i, j \in V \tag{32}$$

$$w_1 \le \sum_{i \in V} a_{ij} * x_i \, ; j \in V \tag{33}$$

$$w_2 \ge \sum_{j \in V} a_{ij} * y_j \, ; i \in V \tag{34}$$

$$w_1 \le w_2 \tag{35}$$

$$\sum_{i \in V} x_i = 1 \tag{36}$$

$$\sum_{j \in V} y_j = 1 \tag{37}$$

The value of the objective function (16) represents deviation from the goal of building player branches in preferred areas (17). Conditions (18) to (32) are used to calculate the payment matrix of player 1. Both players decide on their locations and, based on preferences and final decisions on selected consumption points, cases (1) - (4) may occur. Conditions (28) - (31) determine the elements of matrix **A** depending on the percentage of conscious customers according to the regulator's preference. Conditions (33) - (37) make it possible to determine the strategies of player 1 and player 2, where (33)-(37) are based on the dual linear programming problems and ensure finding equilibrium strategies.

The result of the first stage of the problem should therefore be to ensure the construction of a point of consumption at designated nodes. In the second stage, we introduce the second criterion and objective function that minimizes the deviation from it into the model. We will therefore replace the objective function (16) and the goal (17).

The new objective function minimizes deviation $d_2$:

$$d_2^+ \rightarrow \min \tag{37}$$

from the second level goal

$$\lambda - d_2^+ = 0 \tag{38}$$

while adding equation

$$\sum_{i \in V^{(p)}} x_i + \sum_{j \in V^{(p)}} y_j + d_1^{-*} = 2 \tag{39}$$

where $d_1^{-*}$ is the optimal solution of the problem (16) - (36) while $\lambda \in \langle \lambda^D, \lambda^H \rangle$, where the upper and lower bounds are set by the regulator based on how much it is willing to invest in the tool, which subsequently influences the players' choice of location.

This goal represents the regulator's interest in determining the percentage of convinced consumers. The smallest possible values are desirable, as the motivation and acquisition of these consumers entails costs that the regulator has to bear (such as advertising costs).

The regulator's tools also include financial subsidy for players (companies) who decide to build their branches in the preferred areas. The regulator has chosen a price subsidy as financial assistance, which influences the decision-making process of customers choosing one of the players. Considering the information on conscious customers obtained from stage 2, in the third stage of the model, the objective function and the goal will be replaced by function

$$d_3^+ \rightarrow \min \tag{40}$$

minimizing deviation from the goal

$$c - d_3^+ = 0 \tag{41}$$

adding equation

$$\lambda - d_2^{-*} = 0 \tag{42}$$

where $d_2^{-*}$ is optimal solution of (18) - (39) and $c \in \langle c^D; c^H \rangle$, where the upper and lower bounds are also set by the regulator and similarly influence the players' choice of location, i.e. whether they decide for the preferred or non-preferred.

# 3 CONCLUSIONS

Spatial competition models are generally discussed topic. The financial situation of companies, their strategic decisions regarding the choice of prices of offered products and services, the form of presentation and advertising or the choice of location of their operation are important topics that they must address at the moment of entering the market. However, the financial situation is also related to external factors, such as economic cycles or the state of the world. These are the reasons when they have to use forms of assistance, including state subsidies. However, the established criteria must be met when the authority enters the decision-making process. In such a case, the goals of both parties, regulator and corporate, can be met. The subject of our paper is the presentation of original model in which companies entering the market decide on the place of their operation. The regulator enters their decision-making process with intention to motivate them to choose its preferred locations. It uses financial assistance in the form of a subsidy. At the same time, it seeks to influence consumers through targeted advertising, whose behavior is an important factor in the process of finding optimal business strategies. In the next phase, the regulator tries to minimize these tools so that its costs are as low as possible. Such a situation is formulated as a goal programming problem solved in three stages represented by individual goals.

**Acknowledgements:** This work was supported by the project VEGA 1/0427/20 Multi-criteria models of game theory in economics and politology.

**References**

[1] Anderson, S. P. (1988). Equilibrium Existence in the Linear Model of Spatial Competition. *Economica*, pp. 479-491.
[2] Čičková, Z. and Lopez, A. J. S. (2018). Price Policy in Games in Spatial Competition. Proceedings of the International Scientific Conference in Trenčianske Teplice, Slovakia. In *Quantitative Methods in Economics*: *Multiple Criteria Decision Making. XIX.* Bratislava: Letra Edu, pp. 313-319. ISBN 978-80-89962-07-5.
[3] d'Aspremont, C., Gabszewicz, J. J. and Thisse, J. F. (1979). On Hotelling's" Stability in competition". *Econometrica: Journal of the Econometric Society*, pp. 1145-1150.
[4] Fearon, D. (2006). Alfred weber: theory of the location of industries, 1909. *Ŕ Center for Spatially Integrated Social Science.*
[5] Gehling, G. J. (1968). An evaluation of spatial equilibrium models for policy decisions in the dairy industry.
[6] Goryunov, M., Kokovin, S. and Tabuchi, T. (2017). *Continuous spatial monopolistic competition: matching goods with consumers.* National Research University Higher School of Economics.
[7] Graitson, D. 1982. Spatial Competition a la Hotelling: A Selective Survey. *The Journal of Industrial Economics*, pp.11-25.
[8] Hotelling, H. 1929. Stability in competition. *Economic Journal*, 39(153), pp. 41-57.

**Author’s address**
doc. Ing. Zuzana Čičková, PhD
University of Economics in Bratislava,
Faculty of business informatics, Department of Operations research and Econometrics
Dolnozemská 1/b, 852 35 Bratislava
Slovak Republic
email: zuzana.cickova@euba.sk

**Author’s address**
Ing. Patrícia Holzerová
University of Economics in Bratislava,
Faculty of business informatics, Department of Operations research and Econometrics
Dolnozemská 1/b, 852 35 Bratislava
Slovak Republic
email: patricia.holzerova@euba.sk

# OPERATION CONTROL PROBLEMS OF THE AIRLINE CARRIERS IN THE EVENT OF IRREGULARITIES DURING FLIGHT OPERATIONS

***Jiri Horinka, Lenka Kontrikova, Dusan Teichmann***
***VSB - Technical University of Ostrava***

**Abstract**
Airline transport operation is a complicated process affected by a number of factors. Despite perfect planning, it is common for real operations to be burdened by irregularities of varying range. These irregularities can be caused by meteorological phenomena, aircraft technical failures, crew shortage, awaiting for a delayed connecting flight, etc. If there is not enough aircraft available in case of irregularity during operation, it is necessary to postpone the realization of some flights to the next day or even cancel it. If such a decision is required, the staff of the operational control center (OCC) must identify which flights will be affected by the delay or cancellation. The task of identifying flights that will be affected by the delay or cancellation can be drawn up as an optimization problem. The optimization approach used for its solution will be mixed integer linear programming. The paper will contain a mathematical model in which the total number of passengers whose flights will not be delayed or canceled will be maximized. Computational experiments will be performed on model data in the optimization software Xpress-IVE.

***Keywords:*** *Mathematical model, Optimization, Airline transport operation, Carried passengers*

***JEL Classification:*** C44
***AMS Classification:*** A90C08

## 1. INTRODUCTION - MOTIVATION FOR THE PROBLEM SOLVING

The organization of the operation of an air carrier operating a network of scheduled routes is a complex problem which can withstand very few comparisons with other business activities or models provided by either transport or other logistics entities. One of its characteristic features is its extreme financial complexity. The average airline currently has an annual turnover of billions of euros, but due to pressure on air fares, airlines often only achieve profitability at the percentage of annual turnover.

The air carrier's traffic management process requires very precise planning of the entire process and subsequent continuous management and continuous control of compliance with the accuracy of the planned processes. Despite the better planning of all activities essential for the implementation of flights according to the flight schedule, it is a common reality that the implementation of the flight schedule does not proceed exactly according to plan.

Typical questions that managers must look for in the event of irregularity are:
• Is an aircraft available?
• Is a crew available?
• What will be the impact on passengers?
• What will be the immediate direct costs of addressing irregularity?
• What will be the future indirect costs of addressing irregularities?

• What will be the impact of irregularity on operations and how long will it take to eliminate irregularity?
• What will be the effects of irregularity on revenues?
• Are there any special circumstances that must be considered when dealing with irregularities?

While parameters for addressing aircraft and crew availability have historically been addressed as primary in the original operational control decision-making processes, revenue and especially cost parameters have become important in modern decision-making processes, the impact on passengers has long been neglected. According to the latest trends, however, it is one of the key parameters, as it can affect both the immediate costs of solving irregularities (costs of alternative transport, compensation, accommodation, etc.) and in the long run significantly affects revenues (loss of existing customers and subsequently domino effect and reducing a certain number of new potential customers). Therefore, the passenger impact parameter will be considered as a key in the presented article.

At present, the human – subjective view of the solved situation is dominant in solving irregularities, although it is often based on the use of partial software tools. An objective decision support tool or system is absent.

## 2. ANALYSIS OF THE CURRENT STATE OF AFFAIRS IN THE FIELD OF OPERATION CONTROL DURING IRREGULARITY OPERATIONS

Several works devoted to increasing the efficiency of air carriers in the establishment of IRROP (Irregularity operations) have been found in the literature. A global view of the IRROP solution is provided by the publication (Ball et al., 2007) Collaborative decision making (CDM) is considered in the publication to be the main tool for solving IRROP.

Publication (Evler et al., 2021) provides partial insights into the issue of minimization of impacts. The work deals with the issue of minimizing the total tactical costs caused by departures delays at the airport. The input data are sets of input parameters such as the set of affected aircraft, set of aircraft turnover activities, passenger exit and boarding, refuelling, aircraft cleaning and related activities, unloading and loading of luggage and cargo, etc. turnover, costs of cancellation of passenger transport (costs of flight re-routing, in case of longer delay costs of accommodation and compensation resulting from currently valid legislation), number of buses for passenger transport, number of employees involved in aircraft turnover, aircraft turnover time and others. The model in the published article works with 31 decision variables and constraints. Computational experiments were performed at Frankfurt Airport for 20 departures at the peak of the morning wave departures 7:30 - 11:00. The solution is inspiring, unfortunately it only deals with the issue of minimizing the delay in the turn of the group of aircraft in the transit hub.

Also, other work (Wei Yu, and Song 1997) deals with the creation of an optimization model, but again only for the subsystem of operational management. The work deals with the issue of crew movement control at IRROP in the operation of air carriers. The authors consider the movement of crews to be a bottleneck in the whole process of restoring the functionality of the system. The authors cite the complexity of crew schedules resulting from the flight duty limitations and the size and scope of hub-and-spoke networks operated by large air carriers as the reasons for this phenomenon. The input data are sets of parameters such as the number of aircraft, the number of crews, the number of flights, the limiting flight load standards. The model in the published article makes it possible to decide on the number of exchanged crews in the event of an irregularity and the total number of crews required to cover the scope of operation. Extensive testing has been performed to solve various IRROPs with real data, and the algorithm has proven to be sufficiently efficient for practical applications. The main benefits of the article can be seen in the provision of a comprehensive framework for managing the movement of crews during the establishment of IRROP in the operation of the air carrier.

From the same author, there is a paper (Graf et al., 2020) where the author demonstrates the use of the optimization method in the conditions of air transport in situations where operational rescheduling of crews is necessary due to the occurrence of an irregularity operation. The paper represents one of the possible optimization approaches, in which the objective function includes the cost of non-productive transfer of pilots to the departure point and the cost of compensating passengers in the event of a delay.

# 3. PROBLEM FORMULATION AND MATHEMATICAL MODEL

## 3.1 Problem formulation

A set of flights $I$ is given. For each flight $i \in I$, the earliest scheduled pre-flight commencement time for flight operations is defined – $t_i$, flight duration including times required to perform all pre-flight and flight operations $T_i$, maximum pre-flight commencement time delay for the flight (if pre-flight operations are commenced no later than at the end of this period, there will be no flight delays) – $a_i$ and an indication of the number of passengers who have purchased a ticket $c_i$ for the flight $i \in I$. For each pair of flights $i \in I$ and $j \in I$, the "repositioning (empty) flight time" is further defined (including cases where the destination is also the departure airport of the next flight) after the end of the flight operation $i \in I$ until the time of the start of the flight $j \in I - \tau_{ij}$. In the general formulation of the task of optimizing the daily utilization of the aircraft fleet, let us assume the conditions of a homogeneous aircraft fleet, i.e., the possibility that all aircraft in the fleet are interchangeable. The optimization problem arises in a situation where irregularities occur to such an extent that the flight plan is adversely affected in the near future with a direct impact on the number of flights served. We are therefore in a situation where irregularity has arisen in operation. The task also includes information on the total number of aircraft (excluding the aircraft whose flight has become irregular) that can be used to operate flights.

The task is to optimize the daily use of the aircraft fleet, i.e., decide on the sequence of flight operations by individual aircraft so that the total number of aircraft used is minimal.

## 3.2 Mathematical model

In order to model the decision in the optimization problem, we introduce a group of variables $x_{ij}$, for $i \in I \cup \{0\}$, $j \in I \cup \{0\}$ and $i \neq j$ with domains containing values 0 and 1 (these will be bivalent variables). The bivalent variables $x_{ij}$ will model the decision that the same aircraft will be deployed to operate flight $j \in I$ after flight control $i \in I$. The value $x_{ij} = 1$ after the end of the optimization calculation will represent the decision that the flight $i \in I$ and the flight $j \in I$ will be included in the daily operation plan of the aircraft in the presence of an immediate precedence i≺≺j, i.e. that among the flight operators $i \in I$ a $j \in I$ it will not be operated in any other flight. The value $x_{ij} = 0$ after the end of the optimization calculation will represent the decision that there will be no immediate precedence between flights $i \in I$ and $j \in I$ in the daily operation plan of the aircraft.

The variables $x_{0j}$ for $j \in I$ and $x_{i0}$ for $i \in I$ will have a specific position between the variables $x_{ij}$, for $i \in I \cup \{0\}$ and $j \in I \cup \{0\}$. If $x_{0j} = 1$ applies to the bivalent variable $x_{0j}$ after the optimization calculation, then the flight operator $j \in I$ will not be preceded by the operation of another flight. If $x_{0j} = 0$ applies to the bivalent variable $x_{0j}$ after the end of the optimization calculation, then in the plan of daily operation of the aircraft, the flight operator $j \in I$ will be preceded by the operation of another flight. The variables x_i0 will have analogous meanings. If $x_{i0} = 1$ applies to the bivalent variable $x_{i0}$ after the end of the optimization calculation, then in the plan of daily operation of the aircraft, the operation of another flight will not follow the service $i \in I$. If $x_{i0} = 0$ applies to the bivalent variable $x_{i0}$ after the end of the optimization calculation, then the flight $i \in I$ will be followed by the operation of another flight.

The mathematical model has the form:

$$\max f(x,z) = \sum_{j \in I} c_j \sum_{i \in I \cup \{0\}} x_{ij} \tag{1}$$

subject to:

$$\sum_{i \in I \cup \{0\}} x_{ij} \leq 1 \qquad \text{for } j \in I \tag{2}$$

$$\sum_{i \in I} x_{ji} \leq \sum_{i \in I \cup \{0\}} x_{ij} \qquad \text{for } j \in I \tag{3}$$

$$\sum_{j \in I} x_{0j} \leq N \tag{4}$$

$$t_i + T_i + z_i + \tau_{ij} \leq t_j + z_j + M \cdot (1 - x_{ij}) \qquad \text{for } i \in I \text{ and } j \in I \tag{5}$$

$$z_i \leq a_i \qquad \text{for } i \in I \tag{6}$$

$$x_{ij} \in \{0; 1\} \qquad \text{for } i \in I \cup \{0\} \text{ and } j \in I \tag{7}$$

$$z_i \in R_0^+ \qquad \text{for } i \in I \tag{8}$$

Function (1) represents the optimization criterion - the total number of passengers whose flights the aircraft will be deployed. The group of restrictive conditions (2) shall ensure that a maximum of one aircraft is deployed to operate each flight. The group of restrictive conditions (3) ensures that if the flight control $j \in I$ does not take place (situation when $\sum_{i \in I \cup \{0\}} x_{ij} = 0$), then the flight control $i \in I$ cannot occur after it either. If flight $j \in I$ occurs (situation where $\sum_{i \in I \cup \{0\}} x_{ij} = 1$), then the group of conditions ensures that it will be followed by a maximum of one flight $i \in I$. Restriction (4) will ensure that no more aircraft are used to operate the flights than the carrier has available. The group of restrictive conditions (5) will ensure that if the flight operator $j \in I$ after the flight $i \in I$ is not time-permissible, then it will not be scheduled. The group of restrictive conditions (6) shall ensure that the delay in the time of commencement of pre-flight operations at the aerodrome of departure of the flight does not exceed the time that would cause the flight to be delayed. The groups of constraints (7) and (8) define the domains of the variables used in the proposed model.

## 4. COMPUTING EXPERIMENTS

We verify the functionality of the proposed mathematical model (1) - (8) on the following optimization problem. At the time of the IRROP occurrence, let the air carrier to have 10 flights to serve until the end of the planning period (e.g., until the end of the given day) and 4 operational aircraft remain for this service. Each aircraft is technically and capacitively able to serve any flight. It is true that not all flights must be served by four aircrafts in a given planning period. This means that some flights have to be delayed (e.g., for the next day). Our goal is to propose a solution that maximizes the number of passengers whose flights will not be postponed. To verify the functionality of the model (1) - (8), we need to know the time information about these flights and the number of passengers with the purchased ticket for the given flights, see Table 1.

Table 1: Time information about the flights and numbers of passengers with a purchased ticket for the given flights

| let $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $t_i$ | 100 | 230 | 240 | 290 | 340 | 400 | 450 | 480 | 500 | 630 |
| $T_i$ | 100 | 160 | 140 | 130 | 90 | 135 | 145 | 180 | 160 | 200 |
| $a_i$ | 0 | 30 | 100 | 0 | 40 | 60 | 35 | 25 | 40 | 80 |
| $c_i$ | 135 | 120 | 140 | 130 | 100 | 95 | 140 | 140 | 130 | 95 |

Flight time information are expressed in minutes. Time information of type $t_i$, where $i = 1, \dots, 10$, (expressing starting times of pre-flight activities for every flight) are the number of minutes that elapse from a certain preselected time point to the start of pre-flight activities according to the original schedule. For example, if the pre-flight start time for flight 2 is scheduled for 15:50 and the selected time point is 12:00 (i.e., noon of the day), then the number of minutes between 12:00 and 15:50 is $t_2 = 230$.
We also need to know the time required for unproductive repositioning flights between scheduled flights, see Table 2.

Table 2: Time required for non-productive repositioning flights between scheduled flights

| let $i$ \ let $j$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | x | 25 | 80 | 100 | 120 | 100 | 80 | 90 | 130 | 140 |
| 2 | 65 | x | 45 | 55 | 70 | 130 | 140 | 100 | 90 | 65 |
| 3 | 70 | 120 | x | 130 | 140 | 95 | 160 | 140 | 100 | 105 |
| 4 | 120 | 120 | 120 | X | 135 | 60 | 95 | 140 | 130 | 160 |
| 5 | 145 | 130 | 145 | 120 | x | 90 | 105 | 60 | 160 | 200 |
| 6 | 130 | 125 | 100 | 130 | 160 | X | 140 | 100 | 90 | 100 |
| 7 | 100 | 75 | 60 | 120 | 130 | 100 | x | 50 | 100 | 60 |
| 8 | 110 | 130 | 120 | 100 | 105 | 60 | 195 | x | 130 | 160 |
| 9 | 100 | 140 | 120 | 105 | 45 | 160 | 125 | 130 | X | 60 |
| 10 | 110 | 65 | 130 | 115 | 175 | 140 | 130 | 125 | 100 | x |

Computational experiments with the proposed model were performed in the optimization software Xpress-IVE. The result is the following flight sequences, see Table 3. Deviating values of pre-flight start time shifts are given in the Table 4.

Table 3: Results of the optimization calculation

| Aircraft | Sequence of flights served |
|---|---|
| 1 | $1 \longrightarrow 2 \longrightarrow 8$ |
| 2 | $3 \longrightarrow 9$ |
| 3 | $4 \longrightarrow 10$ |
| 4 | 7 |

Table 4: Results of the optimization calculation

| Flight | Time shift value [min] |
|---|---|
| 3 | 40 |
| 8 | 25 |
| 9 | 20 |
| 10 | 130 |

Operators of flights 5 and 6 will be postponed to the following period. The value of the objective function (total number of passengers whose flight will not be delayed due to IRROP) is 1,030.

## 5. CONCLUSIONS

The concept of the proposed decision support tool based on fuzzy linear programming seems to be applicable in the conditions of medium-sized air carriers operating both scheduled commercial air transport and charter transport. It should be applicable to both network and low-cost airlines, as well as to airlines with a homogeneous and inhomogeneous aircraft fleet in terms of capacity.

The total number of passengers who will be allowed to depart in conditions with a limited number of aircraft was chosen as the optimization criterion. A variant optimization criterion could also be chosen, such as the total number of passengers who will not be allowed to depart.

In the case of choosing a variant criterion (the value of which would be minimized), a solution would be sought in the fuzzy version of the model, where the value of the objective function at the maximum level of confidence will reach sufficiently small values.

**Acknowledgements:** The paper was supported by internal project from the Faculty of Mechanical Engineering Faculty of VSB – Technical University of Ostrava, SP 2022/62 Development and research in transport and logistics.

### References

[1] Ball, M., Barnhart, C., Nemhauser, G. and Odoni, A. (2007). Air Transportation: Irregular Operations and Control. *Handbooks in operations research and management science*, 14. pp. 1-67. https://doi.org/10.1016/S0927-0507(06)14001-3

[2] Evler, J., Asadi, E., Preis, H. and Fricke, H. (2021). Airline ground operations: Schedule recovery optimization approach with constrained resources. *Transportation Research Part C: Emerging Technologies, 128*. p. 103129. https://doi.org/10.1016/j.trc.2021.103129

[3] Graf, V., Teichmann, D., Horinka, J. and Dorda, M. (2020). Dynamic Model for Scheduling Crew Shifts. *Mathematical Problems in Engineering 2020*. https://doi.org/10.1155/2020/5372567

[4] Wei, G., Yu, G. and Song, M. (1997). Optimization Model and Algorithm for Crew Management During Airline Irregular Operations, *Journal of Combinatorial Optimization*. 1(3), pp. 305-321

**Author's address**

Hořínka Jiří
Faculty of Mechanical Engineering, Institute of Transport, VSB – Technical University of Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic
e-mail: horinka@centrum.cz

Kontriková Lenka, Ph.D.
Faculty of Mechanical Engineering, Institute of Transport, VSB – Technical University of Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic
e-mail: lenka.kontrikova@vsb.cz

Teichmann Dušan, Assoc. Prof., Ph.D.
Faculty of Mechanical Engineering, Institute of Transport, VSB – Technical University of Ostrava
17. listopadu 15/2172, 708 33 Ostrava-Poruba, Czech Republic
e-mail: dusan.teichmann@vsb.cz

# MATHEMATICAL MODEL FOR LOCALIZATION OF THE FINAL SHIPMENT DISTRIBUTION WAREHOUSE AT A DISTRIBUTOR OPERATING IN VIRTUAL REALITY

***Andrea Hrníčková, Matúš Kováč, Dušan Teichmann, Denisa Mocková***
***Czech Technical University in Prague***

**Abstract**

Distribution of goods to the final customer is the last link in the logistics chain. This is a topical and important issue, as the efficiency of the distribution processes can have a significant positive effect on the economy of the whole chain. The processes taking place in the parts of the logistics chains dealing with the distribution of goods can be of a different nature. The reasons for the different nature of processes may be specific requirements for the process of completion of goods before final shipment, specific requirements for the final shipment process depending on the type of goods removed, end customer service technology resulting from their specific requirements for delivery dates, number of units, periodicity of service, etc. The presented article deals with the task of selecting the distribution warehouse of the final shipment in the conditions of a distributor with whom orders are placed electronically (in virtual reality). The problem is conceived as an optimization problem, which is solved by methods of mathematical programming. The optimization criterion is the cost of completing customer orders. When creating a mathematical model, knowledge from the field of transport and location tasks is used. Computational experiments will be performed with the proposed model. Computational experiments will be performed for different numbers of warehouses, while only one of them will be designated for the final shipment.

***Keywords:*** *Mathematical Model, Optimization, Distribution, Warehouse, Assembly, Localization*

***JEL Classification:*** C610
***AMS Classification:*** 90B06, 90C08

## 1 INTRODUCTION

The issue of completing shipments is based primarily on customer requirements, even when shopping in a virtual environment on the Internet. Buying goods online today is an increasingly accessible service for the general public. The diversity of goods in individual stores gives the consumer the opportunity to buy everything at one place. It is therefore common that the consumer buys several types of goods in one purchase and expects that all the ordered consignments will be delivered together. At the moment there is a problem for the distributor, who has several warehouses with the following characteristics (geosize): SPO - designation for smaller products (consider dimensions 20x30x40 cm), BPO - designation for medium-sized products (consider dimensions 40x60x60 cm) and XPO - designation for large products (consider dimensions 80x90x100 cm).

The distributor-trader is forced to process such orders and, of course, complete them before sending them to the customer himself. Distribution costs include shipping costs between warehouses and the cost of completing shipments. Completion of shipments can be solved in the following ways:

- assembly in one warehouse - includes the transport of individual parts of the consignment to one warehouse, where sorting and assembly will take place. Such a warehouse must be adapted to receive a larger volume of consignments,
- assembly in a larger geosize warehouse - this method involves the transport of smaller products always to warehouses with larger products (i.e. from SPO to BPO / XPO and from BPO to XPO),

- assembly at the carrier - the method does not involve transport between warehouses but the dispatch of products in each warehouse to an external carrier, which will then complete the consignment at its capacity.

# 2 MOTIVATION TO SOLVE THE PROBLEM AND ANALYSIS OF THE CURRENT STATE OF KNOWLEDGE

Optimization in the distribution parts of logistics chains takes place both within the processes taking place within the individual chain links and between the chain links themselves.
For example, work (Pelikán, 2021) containing two models of mathematical programming deals with the optimization of internal processes in individual links in the distribution parts of logistics chains. In the first model, the total distance travelled by the employee designated in the warehouse for picking the items of individual orders is at first minimized. The second model addresses the minimization of the number of occupied seats in the warehouse by stored items.

The issue of process optimization between individual links in the distribution parts of logistics chains is dealt with mainly by well-known optimization tasks - e.g., transport tasks, location tasks, vehicle routing problem tasks, etc. Approaches in these tasks can be based on linear programming, integer linear programming, mixed linear programming, heuristic methods or metaheuristic methods (e.g., genetic algorithms). It is possible to distinguish approaches that work in deterministic conditions or approaches that take into account uncertainty. When working with uncertainty, for example, approaches based on fuzzy modeling, robust programming or simulation methods can be used.

The use of integer linear programming methods can be found, for example, in the works (Wang and Sheu 2019) and (Čičková, Brezina, Pekár, 2017). A relatively new problem is being addressed (Wang and Sheu 2019), which focuses on route planning for drones (VRPD). The work (Čičková, Brezina, Pekár, 2017) focuses on the problem of minimizing fuel consumption depending on the distance travelled and the vehicle load.

Other approaches are work based on robust programming, which allows inconporation of uncertainty into optimization approaches. It is possible to work with uncertainty with the use of robust optimization, simulation methods, etc. The issue of robust optimization is solved, for example, in works (Shi et al. 2020) or (Borčinová and Peško, 2020). In the work (Shi et al. 2020), the authors focus on the issue of VRP with synchronized visits and uncertain scenarios with respect to greenhouse gas emissions. The authors use hybrid taboo search and simulated annealing to solve this problem. The thesis (Borčinová and Peško, 2020) deals with the issue of CVRP with uncertain customer demand, the probability distribution of which is unknown. The solution to robust programming in this case is a route plan that optimizes the worst-case scenario that can occur.

The issue of simulation of distribution processes is solved, for example, in the work (Pekarčíková et al. 2021). The work deals with the issue of bottlenecks. Using a software tool, bottlenecks and deficiencies in the company's production, logistics and transportation systems are detected. By combining simulation methods and lean logistics tools, they eliminate losses in business processes and further improve flow throughput, reduce inter-operational inventories and increase the efficiency of the production system as a whole.

# 3 PROBLEM FORMULATION AND ITS MATHEMATICAL MODEL

In this chapter, the solved problem is introduced, and a mathematical model is introduced, which is used to perform the experiments described in the following chapter.

### 3.1. Formulation of optimization problem

The set of warehouses $I$ of the distributor is defined in the distribution network. For each warehouse $i \in I$, the requirement $a_i$ is defined for stored products (the basic product type is a product of the type "SPO", products of types "BPO" and "XPO" are standardized for products of the type "SPO"). The transport of products between warehouses is ensured by vehicles with the same capacity $c$. A distance matrix $D$ is known, where its elements $d_{ij}$ represent the distances between individual warehouses. Furthermore, the costs per 1 kilometer driven by the vehicle $l$, the costs of handling the product of the "SPO" type in the warehouse $i \in I$ marked $e_i$ are known. Handling costs are only included for products transported from other warehouses.

The task is to decide on the location of the dispatch center with assembly in one of the existing warehouses so that the total cost of dispatch of the completed warehouse products is minimal.

### 3.2. Design of optimization model

In order to model decisions in a mathematical model, we introduce two groups of variables into the optimization problem. The first group of variables consists of bivalent variables $x_i$. If $x_i = 1$ after the optimization calculation, it means that the dispatch center will be located in the warehouse $i \in I$. If, after the optimization calculation, $x_i = 0$, -it means that the dispatch center will not be in warehouse $i \in I$. The second group of variables consists of non-negative variables $y_{ij}$, modeling the number of trips made between warehouse $j \in I$ without a distribution center and warehouse $i \in I$ with a dispatch center. The symbol M represents a prohibitive constant, the value of which can be determined in two ways.

It is either possible to use a number several orders of magnitude larger than all the others or its minimum value can generally be determined as $M_i \geq \ min \{ai, c\} \, for \ \ i = 1, \dots, n$. If the value of the prohibitive constant was lower than the minimum possible volume of transport in the given session, the binding conditions would malfunction and thus the optimal solution would be negatively affected.

The mathematical model can be written in this form

$$\min f(x,y) = \sum_{i \in I} \sum_{j \in I} d_{ij} \cdot y_{ij} \cdot l + \sum_{i \in I} \sum_{j \in I, j \neq i} e_i \cdot a_j \cdot x_i \qquad (1)$$

Subjet to

$$a_i \cdot (1 - x_i) \leq c \cdot \sum_{j \in J} y_{ij} \qquad \text{for } i \in I \qquad (2)$$

$$y_{ij} \leq M \cdot (1 - x_i) \qquad \text{for } i \in I \text{ a } j \in I \qquad (3)$$

$$y_{ij} \leq x_j \cdot M \qquad \text{for } i \in I \text{ a } j \in I \qquad (4)$$

$$\sum_{i \in I} x_i = 1 \qquad (5)$$

$$x_i \in \{0,1\} \qquad \text{for } i \in I \qquad (6)$$

$$y_{ij} \in Z_0^+ \qquad \text{for } i \in I \text{ a } j \in I \qquad (7)$$

Function (1) represents the optimization criterion - the total cost of shipping the completed stock products. The first part of the criterion represents the total transport costs, the second part of the criterion is the handling costs related to the completion of shipments in the dispatch center. The group of restrictive conditions (2) will ensure the conversion of requests transported from the given warehouses to the dispatch center into the number of journeys. The group of restrictive conditions (3) will ensure that if the warehouse is also a dispatch center, the products will not be transported. The group of conditions (4) provides links between the values of the variables $x_i$ and $y_{ij}$. Condition (5) ensures that the dispatch center will be located in just one warehouse. The groups of conditions (6) and (7) represent the domains of variables used in the mathematical model.

# 4 CALCULATION EXPERIMENTS WITH THE MATHEMATICAL MODEL

We will demonstrate the functionality of the model in the following example. We solve the task of assessing the possibility of completing multi-product orders in one of the warehouses located in the distributor's network. The following model data were used for computational experiments. The distributor has a total of 4 warehouses at his disposal in Kladno (BPO, SPO, XPO), Kutná Hora (BPO), Mělník (SPO) and Mladá Boleslav (BPO and SPO). It is assumed that all warehouses have sufficient capacity for assembly and subsequent dispatch.

*Table 1 Distances, weekly customer requirements for individual warehouses and handling costs for individual warehouses*

| | **Kladno** | **Kutná Hora** | **Mělník** | **Mladá Boleslav** | $a_i$ [$pc\ SPO$] | $e_i$[$K\check{c} \cdot pc\ SPO^{-1}$] |
|---|---|---|---|---|---|---|
| **Kladno** | 0 | 86 | 35 | 65 | 5 367 | 1,89 |
| **Kutná Hora** | 86 | 0 | 72 | 57 | 2 784 | 2,05 |
| **Mělník** | 35 | 72 | 0 | 32 | 3 579 | 1,92 |
| **Mladá Boleslav** | 65 | 57 | 32 | 0 | 1 367 | 1,73 |

Handling costs in individual warehouses vary depending on the technical equipment, technologies used and the location of the warehouse. There is a homogeneous vehicle fleet. The capacity of the vehicle is 40 m$^3$. For the purposes of the experiment, the capacity of the vehicle corresponds to 90% of the maximum capacity of the vehicle, which is 1,500 "SPO". Based on the data from Chapter 1, the number of "BPO" and "XPO" products in the vehicle was determined and subsequently their number was converted to "SPO" products, as shown in Table 2.

*Table 2 Conversion of products per vehicle*

| **Product** | **Number of products/vehicle [pc]** |
|---|---|
| SPO | 1 500 |
| BPO | 250 „BPO“ ≅ 1 500 „SPO“ |
| XPO | 50 „XPO“ ≅ 1 500 „SPO“ |

Transport between warehouses is provided by the distributor's internal carrier and the cost per 1 km is $l = 29\ \text{Kč} \cdot km^{-1}$.

The Xpress-IVE optimization software was used to obtain the optimal solution. The experiments were performed on a personal computer equipped with Intel Celeron with parameters 1.8 GHz and 4.0 GB RAM.

In addition to the basic experiment with 4 warehouses, other experiments with higher numbers of warehouses were performed. The maximum number of warehouses was 8. Warehouse 5 was located in Beroun, 6 in Kolín, 7 in Benešov and 8 in Nymburk. Customer requirements for transport from individual warehouses were $a_5 = 1\ 489\ pcs$, $a_6 = 2\ 680\ pcs$, $a_7 = 1\ 754\ pcs$ and $a_8 = 7\ 425\ pcs$. All requirement values are expressed in "SPO" type units. Handling costs are $e_5 = 1{,}84\ \text{Kč} \cdot \text{pcs SPO}^{-1}$, $e_6 = 1{,}94\ \cdot \text{pcs SPO}^{-1}$, $e_7 = 2\ \text{Kč} \cdot \text{pcs SPO}^{-1}$ and $e_8 = 1{,}8\ \text{Kč} \cdot \text{pcs SPO}^{-1}$. Distances were determined according to (www.mapy.cz). The results of the optimization calculations are summarized in Table 3.

*Table 3 Overview of results of optimization calculations for 4 - 8 warehouses in the distribution network*

| Number of warehouses | Model outputs | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | OV value [Kč] | Total number of rides | | | | | | | |
| | | | | | | | | | | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ |
| **4** | 0 | 0 | 1 | 0 | | | | | 34 310 | 7 | | | | | | | |
| | | | | | | | | | | $y_{13} = 4$ | $y_{23} = 2$ | 0 | $y_{43} = 1$ | | | | |
| ***5*** | *1* | *0* | *0* | *0* | *0* | | | | *38 530* | 7 | | | | | | | |
| | | | | | | | | | | 0 | $y_{21} = 2$ | $y_{31} = 3$ | $y_{41} = 1$ | $y_{51} = 1$ | | | |
| ***6*** | *0* | *0* | *1* | *0* | *0* | *0* | | | *44 716* | 10 | | | | | | | |
| | | | | | | | | | | $y_{13} = 4$ | $y_{23} = 2$ | 0 | $y_{43} = 1$ | $y_{53} = 1$ | $y_{63} = 2$ | | |
| ***7*** | *0* | *0* | *1* | *0* | *0* | *0* | *0* | | *53 477* | 12 | | | | | | | |
| | | | | | | | | | | $y_{13} = 4$ | $y_{23} = 2$ | 0 | $y_{43} = 1$ | $y_{53} = 1$ | $y_{63} = 2$ | $y_{73} = 2$ | |
| ***8*** | *0* | *0* | *0* | *0* | *0* | *0* | *0* | *1* | *73 902* | 15 | | | | | | | |
| | | | | | | | | | | $y_{18} = 4$ | $y_{28} = 2$ | $y_{38} = 3$ | $y_{48} = 1$ | $y_{58} = 1$ | $y_{68} = 2$ | $y_{78} = 2$ | 0 |

In the case of a task with 4 warehouses in the distribution network, the dispatch center with the completion of consignments was placed in warehouse 3 (Mělník). This location was chosen mainly due to the relatively high customer demand and relatively short distances from other warehouses. The value of the purpose function after rounding to whole crowns was CZK 34,310. The total number of trips between warehouses without a dispatch center and a warehouse with a dispatch center is 7, specifically 4 trips will be dispatched from warehouse 1, 2 trips from warehouse 2 and 1 trip from warehouse 4.

Table 3 shows that for most of the tested cases, the dispatch center was placed in warehouse 3. In the case of five warehouses, the location of the dispatch center was changed, as warehouse 1 (Kladno) was significantly closer than the other warehouses to the newly included warehouse in the network. In the case of eight warehouses in the network, the result of the optimization was strongly influenced by the high customer requirements for the newly located warehouse in the network. The location of the dispatch center is therefore significantly affected by the distance between warehouses and the scale of customer requirements.

# 5 CONCLUSION

The presented article deals with the issue of locating a dispatch center within the existing distribution warehouse network. The aim was to present a mathematical model enabling the minimization of the costs of internal distribution between the warehouses and the dispatch center, including the costs of handling and assembly. The functionality of the model was tested on tasks with 4 - 8 warehouses located in the distribution network.

The obtained results clearly show that the distance of the warehouse from other warehouses located in the network and the scale of the customer's request have a significant importance for the location of the dispatch center. If the distance of one warehouse from the others is significantly lower, the location of the center can be expected in this warehouse, the same applies to a warehouse that has significantly higher customer requirements than the others. Another factor influencing the location of the dispatch center is the amount of handling costs, but this is not significantly different from the others in any of the cases we describe, thus its influence on the solved tasks is not so important.

In future research, we could deal with the influence of an inhomogeneous vehicle fleet on the results of optimization, the introduction of uncertainty into the model or the existence of an asymmetric distance matrix.

**Acknowledgements:** This work was supported by SGS22/125/OHK2/2T/16 Research in the field of computational methods for process optimization in specific postproduction segments of the logistics chains.

## References

[1] Borčinová, Z. and Peško, S. (2020). The Maximum Feasible Scenario Approach for the Capacitated Vehicle Routing Problem with Uncertain Demands. In: *9th International Conference on Operations Research and Enterprise Systems, 2020*. Valetta, pp. 159-164.

[2] Čičková, Z., Brezina, I. and Pekár, J. (2017). Capacitated Vehicle Routing Problem Depending on Vehicle Load. In: *35th International Conference Mathematical Methods in Economics 2017*. Hradec Králové: University of Hradec Králové, pp. 108-112.

[3] Pekarčíková, M., Trebuna, P., Kliment, M. and Dic, M., (2021). Solution of Bottlenecks in the Logistics Flow by Applying the Kanban Module in the Tecnomatix Plant Simulation Software. *Sustainability*, *13*(14), p. 1-21. https://doi.org/10.3390/su13147989.

[4] Pelikán, J. (2021). Optimal routing order-pickers in a warehouse. In: *39th International Conference on Mathematical Methods in Economics* [online]. Praha: Czech University of Life Sciences, pp. 370-374.

[5] Shi, Y., Zhou, Y., Ye, W. and Zhao, Q. Q., (2020). A relative robust optimization for a vehicle routing problem with time-window and synchronized visits considering greenhouse gas emissions. *Journal of Cleaner Production*, *275*, p. 1-24. https://doi.org/10.1016/j.jclepro.2020.124112.

[6] Wang, Z. and Sheu, J. B. (2019). Vehicle routing problem with drones. *Transportation research part B: methodological*, *122*, pp. 350-364. https://doi.org/10.1016/j.trb.2019.03.005.

[7] Mapy (2022) Available at: https://www.mapy.cz (Accessed: 1.2.2022)

**Author's address**

Ing. Andrea Hrníčková, Ing. Bc. Matúš Kováč, doc. Ing. Dušan Teichmann, Ph.D., doc. Ing. Denisa Mocková, Ph.D.
Czech Technical University in Prague, Faculty of Transport Sciences, Department of Logistics and Management of Transport
Horská 3, 128 03 Praha 2
Czech Republic
email: hrnicand@fd.cvut.cz, kovacm10@fd.cvut.cz, teichdus@fd.cvut.cz, mockova@fd.cvut.cz

# ANALYSIS OF STOCK RETURNS WITH RESPECT TO COVID-19 PANDEMIC AND WAR IN UKRAINE

*Michaela Chocholatá, University of Economics in Bratislava*

**Abstract**

This paper deals with the analysis of S&P 500, DAX and FTSE 100 daily stock returns based on the Markov switching model (MSM) and cross-sectional standard deviation approach. During the analysed period January 3, 2018 – March 30, 2022 both the calm and turbulent periods (Covid-19 pandemic and war in Ukraine) are covered. The two-regime MSM is estimated in order to capture the specifics of the probability distribution of returns as well as to dynamically describe the different behaviour of stock returns and their volatility during calm and turbulent periods. Furthermore, the cross-sectional standard deviation is used to specify the convergence and divergence periods of analysed stock returns data. The results proved that the probability that the analysed stock market is in a calm regime was significantly higher compared to the probability that the market is in the turbulent regime. With regard to the cross-sectional standard deviation, the Covid-19 pandemic outbreak period and the period indicating the outbreak of the war in Ukraine were clearly identified as divergence periods.

***Keywords:*** *S&P 500, DAX, FTSE 100, Stock returns, Markov switching model, Cross-sectional standard deviation*

***JEL Classification:*** G15, C22, C58
***AMS Classification:*** 62M05, 91B84

## 1 INTRODUCTION

Monitoring of the development of stock indices allows us to describe the behaviour of a particular market over time (Reiff, Brezina and Pekár, 2018). Facing the process of globalization, it seems to be more and more important to analyse not only the development of individual markets, but to consider the growing interconnectedness among them. It is commonly known, that the developments of the financial markets are extremely sensitive to information of various types like e.g., economic and political shocks, health crises (including the Covid-19 pandemic) and the war conflicts (like the ongoing war in Ukraine). The closer relationship among the markets may result in huge vulnerability of economies during such turbulent periods.

In order to capture the specifics of the probability distribution of returns as well as to dynamically describe the different behaviour of stock returns and their volatility during calm and turbulent periods the Markov switching models (MSM) of Hamilton (1989) have become very popular. With regard to the Covid-19 pandemic, several studies have been published to capture its impacts on the development of stock returns across different countries and areas, e.g., Baiardi et al. (2020), Just and Echaust (2020), Chocholatá (2021) and Reiff (2021). In the context of the ongoing war in Ukraine, it is possible to find several comments on the impact of this war on the development of world financial markets, however, more comprehensive analyses are currently being processed.

The risks connected with the strong mutual relationships of the stock markets led analysts to carefully assess the stock market developments during various periods and to apply different methodologies to give the relevant recommendations and future implications, e.g., for investors and policy makers. For a survey of some methodological approaches see e.g., Chocholatá

(2016). Regarding the empirical part of this paper, the concept of the *sigma*-convergence indicator based on the cross-sectional dispersion of stock returns (Adam et al., 2002; Chaloupka, 2012; Seďa, 2020) can be mentioned. This approach enables to reveal whether the returns tend to become more similar during the analysed period, i.e. converge or reversely have a divergent tendency.

The aim of this paper is two-fold. Firstly, to describe the dynamic behaviour of daily stock returns (January 3, 2018 – March 30, 2022) of the three stock indices (S&P 500, DAX and FTSE 100) based on use of the two-regime MSM methodology. This aim enables to assess the switching between the calm and turbulent regimes as well as to calculate the transition probabilities between the regimes. Secondly, this paper tries to capture and identify the time slots of stock markets convergence and divergence, respectively. Considering the analysed period, both the impacts of Covid-19 pandemic and of the war in Ukraine are investigated.

The rest of the paper is organized as follows: section 2 deals with the methodology, section 3 presents the data and empirical results and section 4 concludes.

# 2 METHODOLOGY

In general, the behaviour of financial markets is very unstable and can change quite dramatically during some periods of time. In recent decades various time series models have been presented to capture the occurrence of different regimes generated by a stochastic process.

The MSM, popularized by Hamilton (1989), enables to capture processes driven by heterogeneous states of the world supposing that the occurrence of the particular regime is determined by an unobservable stochastic process usually denoted as $s_t$. The most commonly used, the two-regime MSM, distinguishes only two states of the world – calm periods with normal behaviour of stock returns and turbulent periods with dramatically changing behaviour of stock returns characterised as "positive mean-low volatility" period and "negative mean-high volatility" period, respectively (see e.g., Kośko, 2006 Almonares, 2019 and Chocholatá, 2021). The regime indicator variable $s_t$ takes on two values – 1 and 2, i.e., $s_t = 1$ indicates that the process is in regime 1 at time *t* and $s_t = 2$ indicates that the process is in regime 2 at time *t*.

We will assume that the behaviour of the stock returns[1] $r_t$ is described as follows (Hamilton, 2010; Chocholatá, 2021)[2]:

$$r_t = c_{s_t} + \emptyset r_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_{s_t}^2) \tag{1}$$

where $s_t$ is a regime indicator variable governed by a first-order Markov process.

Transition probabilities for the two-regime Markov chain i.e. probabilities that the regime *i* (at time $t-1$) will be followed by the regime *j* (at time *t*) are specified as follows (Franses and van Dijk, 2000):

$$P\{s_t = j | s_{t-1} = i\} = p_{ij} \tag{2}$$

Formula (2) specifies the conditional probability which is non-negative and it should hold that:

$$p_{i1} + p_{i2} = 1, \qquad i = 1, 2. \tag{3}$$

Furthermore, the concept of *sigma* – convergence (Adam et al., 2002; Chaloupka, 2012; Seďa, 2020), i.e. calculation of the cross-sectional standard deviations for each business day enable to assess whether the differences between the log returns of individual markets are increasing or

[1] Stock returns $r_t$ , i.e. continuously compounded returns, were calculated using the formula $r_t = \ln(P_t) - \ln(P_{t-1})$, where symbol $P_t$ denotes the closing price of the corresponding stock index in time *t*.
[2] Symbol $\emptyset$ denotes the autoregressive parameter.

decreasing. The formula for the calculation of the cross-sectional standard deviation is as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(r_i - \bar{r})^2}{n}} \tag{3}$$

where $n$ denotes the quantity of analysed log returns. While the declining values of the cross-sectional standard deviation (3) speak for the convergence of the analysed log returns, the increasing values indicate the divergence.

## 3 DATA AND EMPIRICAL RESULTS

The analysed data set consists of daily data of the three major stock indices – the US stock index S&P 500, the German stock index DAX and the British stock index FTSE 100 spanning from January 3, 2018 to March 30, 2022 thus including 976 observations[3]. The data were retrieved through the R package "quantmod" from the web-page of Yahoo! (Ryan et al., 2020). The whole analysis was carried out with the use of softwares EViews and R.

The analysis is focused on the log returns data, i.e. continuously compounded returns $r_t$, which are graphically depicted in Figure 1 together with the basic descriptive statistics. The daily percentage returns were slightly positive with volatilities spanning between 1.17 % and 1.38 %. The behaviour of all three return series clearly reflect the huge volatility attributable to the Covid-19 pandemic. Since the S&P 500 stock returns does not indicate increased volatility implied by the war in Ukraine, the developments of the two European return series, DAX and FTSE 100, markedly mirror this event (especially the outbreak of the war).

To capture the switching behaviour of the S&P 500, DAX and FTSE 100 log returns during the analysed period, parameters of the two-regime Markov switching models with regime invariant AR(1) process were estimated. The estimation results – regime specific mean, regime specific variance[4] and regime invariant AR(1) parameter are outlined in Table 1. The presented results indicate that not all estimated parameters were statistically significant. While regime 1 is characterized by positive mean returns and low volatilities (0.62 %, 0.92 % and 0.73 %, respectively), regime 2 shows statistically insignificant negative mean returns and higher volatilities (2.16 %, 2.97 % and 2.49 %, respectively).

The transition probabilities and expected durations are gathered in Table 2. The calculated values of constant transition probabilities indicate that the regime 1 is much more persistent than the regime 2. Furthermore, the transition probabilities from the low volatility regime 1 to the high volatility regime 2 are lower in comparison to transition probabilities from regime 2 to regime 1. The corresponding constant expected durations in a regime are considerably higher for the low volatility regime 1 in comparison to the high volatility regime 2.

[3] Due to omitting of the missing data.
[4] Calculated as exp( log($\sigma_1$) ) and exp( log($\sigma_2$) ), respectively.

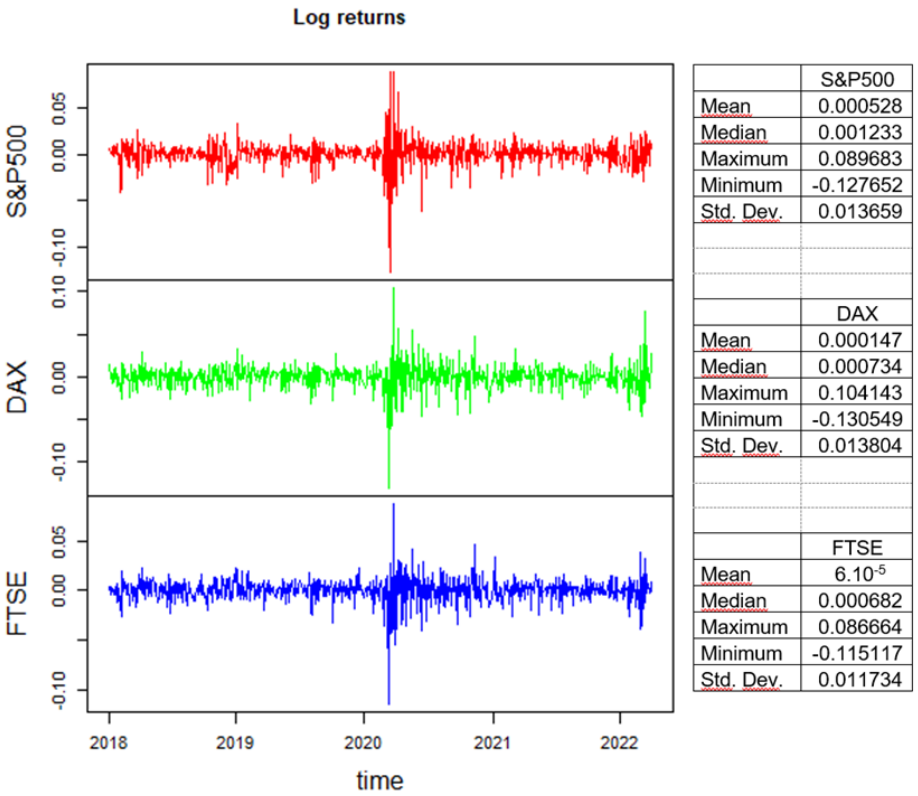



**Figure 1** Log returns of S&P 500, DAX and FTSE 100 and basic descriptive statistics

Source: author's own calculations in R

**Table 1** Two-regime Markov switching model: estimation of parameters.

| | Parameter | S&P 500 | DAX | FTSE 100 |
|---|---|---|---|---|
| **Regime 1** | $c_1$ | 0.0016*** | 0.0005 | 0.0005* |
| | $\log(\sigma_1)$ | -5.0867*** | -4.6826*** | -4.9101*** |
| **Regime 2** | $c_2$ | -0.0016 | -0.0022 | -0.0022 |
| | $\log(\sigma_2)$ | -3.8337*** | -3.5154*** | -3.6940*** |
| **Common** | $\emptyset$ | -0.1114*** | -0.0635** | -0.0758** |

Note: Symbols ***, ** and * denote rejection of the null hypothesis at the 1%, 5 % and 10% significance level, respectively.

Source: author's own calculations in EViews

**Table 2** Transition probabilities and expected durations

| | | S&P 500 | | DAX | | FTSE 100 | |
|---|---|---|---|---|---|---|---|
| **Constant transition probabilities** | **Reg.** | **1** | **2** | **1** | **2** | **1** | **2** |
| | **1** | 0.9682 | 0.0318 | 0.9902 | 0.0098 | 0.9828 | 0.0172 |
| | **2** | 0.0680 | 0.9320 | 0.0662 | 0.9338 | 0.1030 | 0.8970 |
| **Constant expected durations** | **Reg.** | **1** | **2** | **1** | **2** | **1** | **2** |
| | | 31.4631 | 14.7018 | 101.8429 | 15.1083 | 58.1298 | 9.7129 |

Source: author's own calculations in EViews

Figure 2 presents Markov switching smoothed regime probabilities of being in regime 1 at time *t*. The above presented transition probabilities, expected durations as well as smoothed regime probabilities show that the highest probability and longest expected duration of being in the low volatility regime was detected for the DAX log returns followed by FTSE 100 and S&P 500 log returns, respectively. Thus, least frequently, there was a switch between regimes in the case of DAX log returns (corresponding to Covid-19 issues and to the outbreak of the war in Ukraine, respectively). More frequently switching between the two regimes was detected for FTSE 100 log returns where besides the turbulences during the Covid-19 pandemic and war in Ukraine, some switches during the first quarter of 2018 and the third quarter of 2019 were recorded. Clearly lower persistence of regime 1 was recorded in case of S&P 500 log returns in comparison to the remaining two log returns. The depicted S&P 500 smoothed probabilities indicate that besides the generally turbulent periods of Covid-19 pandemic and outbreak of the war in Ukraine, various switches to the high volatility regimes were marked for the 2018 and 2019 attributable to the global economic slowdown, uncertainty about the Brexit issues, the U.S. presidential elections, etc.

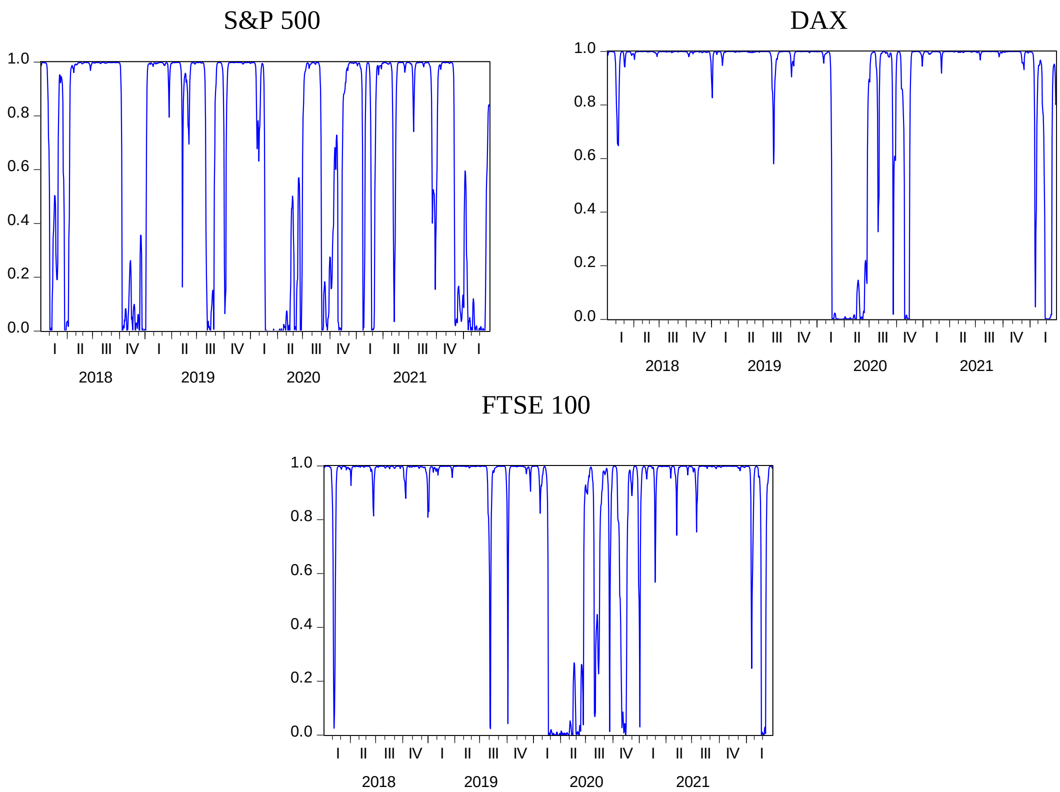



**Figure 2** MSM smoothed regime probabilities for the regime 1 – P(S(t)=1)

Source: author's own calculations in EViews

In order to assess the convergent or divergent tendency of the three log returns during the analysed time period, the cross-sectional standard deviation was calculated and further smoothed by the Hodrick-Prescott (HP) filter. The graphical illustration in Figure 3 identifies the divergent tendency of log returns corresponding to the outbreak of the Covid-19 pandemic in March 2020 and the Covid-19 wave of autumn 2021 with continuing divergence till the end of analysed period probably attributable to the growing tensions between Russia and Ukraine as well as to the ongoing war in Ukraine.

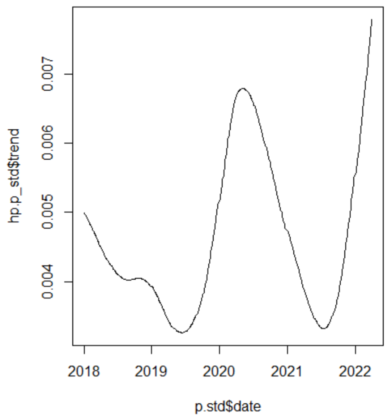



**Figure 3** Cross-sectional standard deviation among S&P500, DAX and FTSE log returns (smoothed by HP filter)

Source: author's own calculations in R

# 4 CONCLUSION

The paper analysed the daily log returns of the three major stock indices – the US stock index S&P 500, the German stock index DAX and the British stock index FTSE 100 spanning from January 3, 2018 to March 30, 2022 with respect to the Covid-19 pandemic and war in Ukraine. Two approaches were adopted – the two-regime MSM and the cross-sectional standard deviation. The MSM results proved positive mean returns during the calm periods (regime 1) while in turbulent periods (regime 2) the mean returns were negative. Furthermore, the results revealed that the frequency of switches between regimes was in case of individual log returns different, but all there analysed log returns indicated the switch to the high volatility regime during the Covid-19 pandemic and the war in Ukraine. These two turbulent period were, based on the cross-sectional standard deviation, clearly identified as divergent periods of log returns.

**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".

**References**

[1] Adam, K., Jappelli, T., Menichini, A., Padula, M. and Pagano, M. (2002). *Analyse, Compare, and Apply Alternative Indicators and Monitoring Methodologies to Measure the Evolution of Capital Market Integration in the European Union.* [Online]. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.126.6040&rep=rep1&type=pdf [cited 2019-04-12].

[2] Almonares, R.A.L. (2019). Markov Switching Model of Philippine Stock Market Volatility. *DLSU Business & Economics Review*, 29 (1), pp.24-30.

[3] Baiardi, L., Costabile, M., Giovanni, D. D., Lamantia, F., Leccadito, A., Massabò, I., Menzietti, M., Pirra, M., Russo, E. and Staino, A. (2020). The dynamics of the S&P 500 under a crisis context: Insights from a three-regime switching model. *Risks*, *8*(3), 71, pp.1-15.

[4] Franses, P.H. and van Dijk, D. (2000). *Non-linear time series models in empirical finance.* Cambridge: Cambridge University Press.

[5] Hamilton, J.D. (1989). A New approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), pp.357-384.

[6] Hamilton, J. D. (2010). Regime switching models. In: Durlauf, S.N. and Blume, L.E. (eds). *Macroeconometrics and Time Series Analysis*. The New Palgrave Economics Collection. London: Palgrave Macmillan, pp.202-209.
[7] Chaloupka, J. (2012). Ekonometrická analýza vzájemných vazeb akciových trhů nových členských zemí EU před vypuknutím a po vypuknutí finanční krize. *Ekonomická revue – Central European Review of Economic Issues*, 15(3), pp.157-175.
[8] Chocholatá, M. (2016). Different approaches to stock market linkages: evidence from CEE-3 countries. In *Advances in applied business research: the L.A.B.S. initiative*. New York: Nova Science Publishers, pp.49-70.
[9] Chocholatá, M. (2021). Czech Stock Market Volatility Before and During the Covid-19 Crisis. In *Strategic Management and Its Support by Information Systems 2021*: Proceedings of the 14th International Conference, Ostrava, pp.112-119
[10] Just, M. and 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, 101775.
[11] Kośko, M. (2006). An Application of Markov-switching Model to stock Returns Analysis. *Dynamic Econometric Models,* 7, pp.259-268.
[12] Reiff, M., Brezina, I. and Pekár, J. (2018). Skrytý Markovov model finančného indexu S&P Europe 350. In Proceedings from the international conference *Trendy v podnikání 2018*. Plzeň, pp.1-8.
[13] Reiff, M. (2021). Vplyv pandémie COVID-19 na akciový index SDAX. In Proceedings from the conference *Využitie kvantitatívnych metód vo vedeckovýskumnej činnosti a v praxi XIV*. Bratislava: EKONÓM, pp.96-100.
[14] Ryan, J. A., Ulrich, J. M., Thielen, W., Teetor, P. and Bronder, S. (2020). Package ‘quantmod’. [Online]. Available at: https://cran.r-project.org/web/packages/quantmod/quantmod.pdf [cited 2022-04-01].
[15] Seďa, P. (2020). *Teoretické aspekty hypotézy efektivního trhu a empirie na akciových trzích v zemí střední Evropy*. Ostrava: VSB-TU Ostrava.

**Author’s address**

doc. Ing. Michaela Chocholatá, PhD.
University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics
Dolnozemská cesta 1/b, 852 35 Bratislava
Slovak Republic
email: michaela.chocholata@euba.sk

# PROGRESS IN OPTIMIZATION SOFTWARE IN THE LAST DECADE

***Josef Jablonský***
***Prague University of Economics and Business***

**Abstract**
Linear programming (LP) and mixed integer linear programming (MILP) problems are frequently used modelling tools in practice. The aim of this study is to discuss the progress in optimization software for LP and MILP problems. The paper compares computational results of GUROBI solver releases 6.0.1 to 9.5.1 on the set of 30 instances from the MIPLIB 2017 library of test instances. GUROBI solver belongs to the top optimization tools at all. The results show a significant progress and reduction of the solution time in the latest versions. In addition, the comparison of another top optimization solver CPLEX 20.10 with Gurobi solvers is presented and discussed.

***Keywords:*** *Optimization, Mixed Integer Linear Programming, Solver, Gurobi*

***JEL Classification:*** C44
***AMS Classification:*** 90-04, 90-08

## 1 INTRODUCTION

Linear and mixed integer linear optimization problems (LP and MILP) belong to frequently applied modelling tools. Their solving requires (especially in case of MILP problems) availability of high-quality optimization solvers. IBM ILOG CPLEX, GUROBI and FICO XPRESS Optimizer are the triple of solvers that belong currently to the top software optimization tools at all. The aim of this paper is to analyze the progress in optimization software in the last 8 years using the set of 30 test instances from the MIPLIB 2017 library (Gleixner et al., 2017). The analysis is performed with the GUROBI solver starting by version 6.0.1 released in December 2014 until the current version 9.5.1 released in February 2022. Except these two versions, the releases 7.0.1 (November 2016) and 8:1.0 (October 2018) are tested. All versions of the solver successfully reached the optimum solution of the problems. The study measures the progress in solution time. For comparison purposes, the latest version of CPLEX is considered in the analysis.

GUROBI is a product of Gurobi, Inc. that was established in 2008 by Z. Gu, E. Rothberg and R.E. Bixby. This system contains top LP, MILP, and quadratic programming (QP) solvers with top results in benchmarking tests – (Mittelmann, 2022). GUROBI is free for academic purposes and the full professional license can be activated online very easily within few seconds. GUROBI offers links to important modelling languages (GAMS, MPL for Windows, AIMMS, and others), Python and R system. First version of GUROBI solver was released in 2008. It is possible to find comparison of this version with the newer ones in (Gurobi 9.5 Performance Benchmarks, 2022).

IBM ILOG CPLEX was developed in the 80s of the last century by CPLEX Optimization that was founded by R.E. Bixby. As in the previous case, this solver contains LP, MILP and QP solvers. CPLEX became the most powerful LP and MILP solver on the software market until 2008 when the company was sold to IBM, and the first version of GUROBI solver was released. The newest version of CPLEX is 20.10 (December 2021). This solver together with powerful modelling environment IBM ILOG CPLEX Optimization Studio is also available to researchers and students under the IBM Academic Initiative. More information about this system can be found in (IBM ILOG CPLEX Optimizer, 2022). Nice benchmarking results of the first versions of this optimizer are available in (Bixby, 2002).

The next top MILP optimization solver is FICO XPRESS Optimizer distributed by British firm FICO, Inc. It is a powerful system with similar properties as the previous two ones. The Optimizer is a part of the FICO XPRESS Optimization Suite that offers very nice tools for building models using own Mosel modelling language, and presentation the results. Recently the Optimizer version 39.01.04 was released. FICO XPRESS Optimization Suite is available for the members of FICO Academic Partner Program. More about this system in (FICO Xpress Optimizer, 2022). Unfortunately, FICO does not support benchmarking and not all instances from the MIPLIB library can be solved using this software. That is why this solver was not included in the study finally.

The next section of the paper contains a detailed information about the datasets included in the MIPLIB libraries and about the history of these libraries. Section 3 presents benchmarking results for four versions of GUROBI solvers, and the latest version of CPLEX solver. The final section discusses the results and concludes the paper.

## 2 MIPLIB 2017 LIBRARY

The first edition of the Mixed Integer Programming Library (MIPLIB) was released in 1992. In 1996, this library was updated to the edition 2.0 and later 3.0. The set of instances included in MIPLIB 3.0 is still downloadable – in total, it is 65 instances in three main categories as in the later editions:

- *Easy* - instances that are solved using professional commercial solvers quite easily (within one hour).
- *Hard* - the set of instances that are already solved using specialized software tools and their optimum solution is known.
- *Open* (not solved) – the instances where the optimal solution is not known and are not solved up to now.

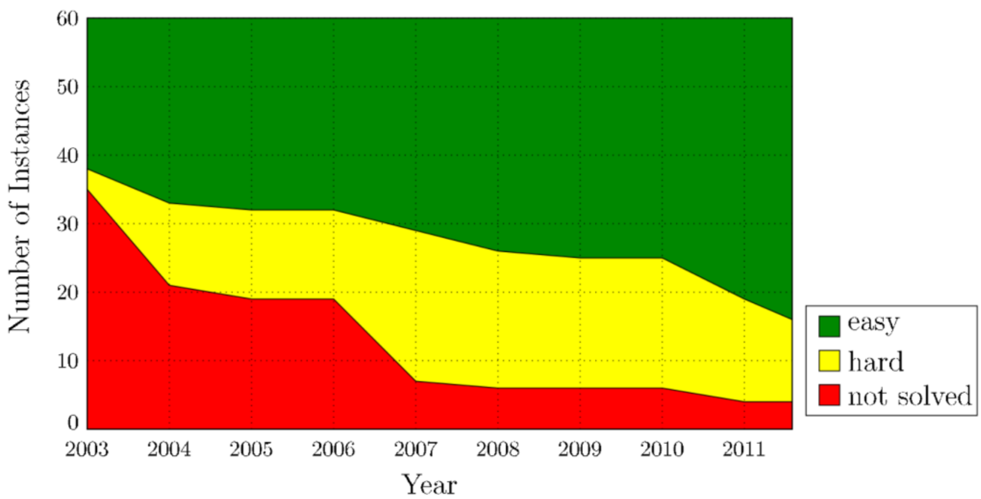



Figure 1: Progress in MILP solvers – MIPLIB 2003 instances

MIPLIB 3.0 was slightly updated in 2003. Some of the instances were removed, and several new ones added. The total number of instances in MIPLIB 2003 was 60. A detailed description of this library is included in (Achtenberg et al., 2006). Figure 1 from (Koch et al.

2011) shows progress in solving the instances within 8 years until 2011. In 2003, more than half instances were unsolvable. In 2011 just 4 instances belong to the category “not solved”. It is interesting that these four instances are still currently open.

The fifth edition of the MIPLIB library was released in 2010 and the number of instances was substantially extended – from 60 in MIPLIB 2003 to 361 in MIPLIB 2010. The subset of selected 87 instances from the whole set (the whole set is called the collection set) is the benchmark set – all the instances of this set are in the category “easy”. So far, the last revision of the MIPBIL library was made in 2017. The collection set of this 6$^{th}$ edition contains 1065 instances. The benchmark set has 240 of them belonging to the “easy” category. A detailed description of the MIPLIB 2017 library is available in (Gleixner et al., 2021).

Table 1: Benchmark data set

| Instance | Constr | Var | Integer | Binary | Cont | Obj |
|---|---|---|---|---|---|---|
| 30n20b8 | 0 | 18380 | 7344 | 11036 | 0 | 302 |
| aflow40b | 1442 | 2728 | 0 | 1364 | 1364 | 1168 |
| air04 | 823 | 8904 | 0 | 8904 | 0 | 56137 |
| bab5 | 4964 | 21600 | 0 | 21600 | 0 | -106411.8401 |
| beasleyC3 | 1750 | 2500 | 0 | 1250 | 1250 | 754[1] |
| cov1075 | 637 | 120 | 0 | 120 | 0 | 20 |
| csched010 | 351 | 1758 | 0 | 1457 | 301 | 408[2] |
| danoint | 664 | 521 | 0 | 56 | 465 | 65.666666 |
| eil33-2 | 32 | 4516 | 0 | 4516 | 0 | 934.007915999 |
| gmu-35-40 | 424 | 1205 | 0 | 1200 | 5 | -2406733.3688 |
| map20 | 328818 | 164547 | 0 | 146 | 164401 | -922 |
| mcsched | 2107 | 1747 | 14 | 1731 | 2 | 211913 |
| mzzv11 | 9499 | 10240 | 251 | 9989 | 0 | -21718 |
| n3seq24 | 6044 | 119856 | 0 | 119856 | 0 | 52200 |
| n4-3 | 1236 | 3596 | 174 | 0 | 3422 | 8993 |
| neos13 | 20852 | 1827 | 0 | 1815 | 12 | -95.4748 |
| neos18 | 11402 | 3312 | 0 | 3312 | 0 | 16 |
| net12 | 14021 | 14115 | 0 | 1603 | 12512 | 214 |
| newdano | 576 | 505 | 0 | 56 | 449 | 65.6666667 |
| noswot | 182 | 128 | 25 | 75 | 28 | -41 |
| ns1758913 | 624166 | 17956 | 0 | 17822 | 134 | -1454.67 |
| pg5_34 | 225 | 2600 | 0 | 100 | 2500 | -14339.4 |
| qiu | 1192 | 840 | 0 | 48 | 792 | -132.87313695 |
| ran16x16 | 288 | 512 | 0 | 256 | 256 | 3823 |
| reblock67 | 2523 | 670 | 0 | 670 | 0 | -3.46306e+07 |
| rmine6 | 7078 | 1096 | 0 | 1096 | 0 | -457.186 |
| roll3000 | 2295 | 1166 | 492 | 246 | 428 | 12889.999992 |
| sp98ir | 1531 | 1680 | 809 | 871 | 0 | 219676790.4 |
| tanglegram2 | 8980 | 4714 | 0 | 4714 | 0 | 443 |
| vpphard | 47280 | 51471 | 0 | 51471 | 0 | 5 |

The current MIPLIB 2017 benchmark set contains 240 instances. In our study, we have reduced the original benchmark set to 30 instances. The same set of instances was solved and

[1] Exact optimum objective for beasleyC3 is 753.9999999999128
[2] Exact optimum objective for csched010 is 407.999999999994

analyzed in detail in (Jablonský, 2015), and the conclusions of this study can be compared with the results coming from the current experiments. The entire dataset is presented in Table 1. The first column of this table contains identification of the instance as it is denoted in the MIPLIB library. The next five columns inform about the number of constraints (Constr), the total number of variables (Var), the number of integer, binary and continuous variables of the problem, and the optimum objective function value (Obj).

Table 2: Solution time [sec]

| Solver | GUROBI | | | | CPLEX |
|---|---|---|---|---|---|
| Instance | 6.0.1 | 7.0.1 | 8.1.0 | 9.5.1 | 20.10 |
| 30n20b8 | 2.29 | 2.71 | 1.79 | **1.43** | 2.56 |
| aflow40b | 390.42 | 95.64 | 59.35 | 39.86 | **28.02** |
| air04 | 10.11 | 4.60 | **3.26** | 6.30 | 4.80 |
| bab5 | 448.26 | 44.91 | **14.86** | 31.80 | 111.31 |
| beasleyC3 | 12.21 | 5.41 | **2.10** | 2.57 | 10.76 |
| cov1075 | 4.06 | 1.35 | 1.24 | 3.92 | **1.05** |
| csched010 | 1266.42 | 318.11 | 247.01 | **157.00** | 216.09 |
| danoint | 2876.83 | 584.66 | 454.85 | 244.55 | **151.09** |
| eil33-2 | 43.75 | 1.86 | **1.08** | 2.15 | 9.17 |
| gmu-35-40 | 958.03 | 72.52 | 27.78 | 39.35 | **7.77** |
| map20 | 74.11 | 33.11 | **19.69** | 20.85 | 78.42 |
| mcsched | 44.34 | 10.89 | **9.17** | 9.61 | 18.58 |
| mzzv11 | 15.31 | 6.15 | 5.00 | **4.48** | 6.64 |
| n3seq24 | 228.15 | 29.62 | **18.26** | 43.18 | 83.24 |
| n4-3 | 1010.50 | 33.36 | 61.57 | **24.03** | 8.11 |
| neos13 | 20.09 | **10.71** | 31.10 | 41.07 | 112.39 |
| neos18 | 6.31 | 1.37 | 1.34 | **1.24** | 2.73 |
| net12 | 77.21 | 23.21 | **19.72** | 21.03 | 30.97 |
| newdano | 365.08 | **93.68** | 187.67 | 108.95 | 137.36 |
| noswot | 18.43 | 20.30 | 8.66 | 21.39 | **3.36** |
| ns1758913 | 11.92 | **9.68** | 20.84 | 11.51 | 53.39 |
| pg5_34 | 287.97 | **3.98** | 7.92 | 7.60 | 4.66 |
| qiu | 18.98 | 6.64 | **2.73** | 2.57 | 2.97 |
| ran16x16 | 26.56 | **6.25** | 6.93 | 6.50 | 7.64 |
| reblock67 | 275.07 | 113.43 | 12.80 | **11.12** | 24.00 |
| rmine6 | 524.49 | 85.09 | 76.61 | **52.09** | 137.05 |
| roll3000 | 27.98 | 7.54 | 8.25 | **7.02** | 7.70 |
| sp98ir | 23.28 | 18.46 | 11.92 | **5.89** | 6.75 |
| tanglegram2 | 0.51 | 0.46 | 0.69 | 0.66 | **0.13** |
| vpphard | 403.45 | 128.67 | **100.79** | 114.43 | 505.22 |
| **Sum abs** | **9472.12** | **1774.37** | **1424.98** | **1044.15** | **1773.93** |
| **Sum rel** | **xxx** | **19.77** | **16.25** | **15.85** | **21.07** |

## 3 RESULTS

The instances listed in Table 1 were tested using GUROBI solver versions 7.0.1, 8.1.0 and 9.5.1 on desktop machine with processor i7-8700 3.20 GHz and 8GB RAM. In all cases, the optimum solution was found, and the maximum solution time was less than 10 minutes (danoint instance was the hardest for all solvers). The results (solution times) are presented in Table 2. In addition, Table 2 also presents the results for older version of this software 6.0.1.

Unfortunately, this older version uses different license policy, and cannot be used with current license files. That is why the first column of Table 2 contains solution times taken from (Jablonský, 2015). Of course, they are derived using an older and slower computer. The best results for each instance are in bold.

The results show that the GUROBI solver in its version 7.0.1 is the fastest in 5 instances, the next version in our study 8.1.0 was the best in 10 cases, and the newest release 9.5.1 in 9 instances. The latest version of CPLEX solver was the best in 6 test problems. Maybe it is surprising that the latest GUROBI version is not the fastest in more cases but the differences in solution times are often very tiny, especially for the instances that are easily solved, and their solution time is short.

The solvers can be evaluated using various criteria. One of them may be the number of instances solved fastest in comparison to other solvers. According to this criterion, GUROBI 8.1.0 as mentioned above is the best. This criterion is probably not the most appropriate. Another possibility is to compare the absolute values of solution times. The pre-last row of Table 2 (Sum abs) contains the total time needed for solving of all instances. The result for GUROBI solver 6.0.1 (even another and slower computer than in case of other solvers was used) was almost 9500 seconds. The fastest solver was the newest version of GUROBI with slightly more than 1000 seconds. This clearly illustrates the progress in optimization software in the last years. GUROBI 7.0.1 was slower by 70%, and version 8.1.0 slower by 36% than the latest version of this solver. This progress was reached in 4 and 2 years respectively. According to the sum of solution times the latest version of CPLEX solver has similar results as GUROBI 7.0.1 solver.

It can be argued that the evaluation by absolute solution times is not correct because the differences in solving harder instances have much higher impact in the final sum. That is why the solution times are normalized – for all instances, the times are divided by the worse (highest) value. Then, the worse solver has relative value 1, the other ones have fractions of these maximum values. This normalized sum can be used as another criterion for evaluation of the solvers. It is presented in the last row of Table 2. Due to the incomparable values of the old GUROBI solver, just the last four columns of Table 4 have been considered in the analysis. The best solver according to this criterion is again GUROBI 9.5.1 but the differences are not so high as in the previous case.

# 4 CONCLUSIONS

Discrete optimization problems find their applications in many real cases. Their solving is usually a complex task, and the development of MILP solvers is of a high importance. This study dealt with testing and comparison of GUROBI MILP solvers in its four versions where the oldest was released in 2014. The test problems considered for testing purposes are defined as the subset of the benchmark set of MIPLIB 2017 library. Comparison of the latest version of GUROBI solver shows that it is better than the competing IBM ILOG CPLEX solver. A significant improvement of the solving time of the latest GUROBI version comparing to the previous ones is clearly seen especially for the instances with the higher solution time. The important feature of GUROBI solver consists in its possible connection with other modelling systems and programming languages. In this way the uses can create their own applications for solving special classes of optimization problems. For academic purposes it is useful that GUROBI may be used for academic purposes in its full version for free, and the license can be activated and extended very easily. Due to this friendly policy, this solver is an ideal optimization tool for teaching and research.

**Acknowledgements:** The research is supported by the Faculty of Informatics and Statistics, Prague University of Economics and Business.

## References

[1] Achtenberg, T., Koch, T. and Martin, A. (2006). MIPLIB 2003. *Operations Research Letters*, 34(4), pp. 361–372. https://doi.org/10.1016/j.orl.2005.07.009.
[2] Bixby, R.E. (2002). Solving real-world linear programs: a decade and more of progress. *Operations Research*, 50(1), pp. 3–15. https://doi.org/10.1287/opre.50.1.3.17780.
[3] FICO Xpress Optimization. (2022) [Online]. Available at: https://www.fico.com/en/products/fico-xpress- optimization [Accessed: 6 March 2022].
[4] Gleixner, A., Hendel, G., Gamrath, G. *et al.* (2021). MIPLIB 2017: data-driven compilation of the 6th mixed-integer programming library. *Mathematical Programming Computation,* 13**,** pp. 443–490. https://doi.org/10.1007/s12532-020-00194-3.
[5] GUROBI 9.5 Performance Benchmarks (2022). [Online]. Available at: https://www.gurobi.com/wp-content/uploads/2022/02/Performance-Gurobi-9_5.pdf. Accessed: 6 March 2022].
[6] IBM ILOG CPLEX Optimizer (2022). [Online]. Available at: https://www.ibm.com/analytics/cplex-optimizer. [Accessed: 6 March 2022].
[7] Jablonský, J. (2015). Benchmarks for current linear and mixed integer optimization solvers. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 63(6), pp. 1923-1928. https://doi.org/10.11118/actaun201563061923.
[8] Koch, T., Achtenberg, T., Andersen, E., *et al.* (2011). MIPLIB 2010 – Mixed Integer Programming Library. *Mathematical Programming Computation*, 3(2), pp. 103-163. https://doi.org/10.1007/s12532-011-0025-9.
[9] Mittelmann, H. (2022). Benchmarks for optimization software. [Online]. Available at: http://plato.asu.edu/bench.html. [Accessed: 6 March 2022].

## Author's address

prof. Ing. Josef Jablonský, CSc.
Prague University of Economics and Business
Department of Econometrics
W. Churchill Sq. 4, 13067 Praha 3
Czech Republic
e-mail: jablon@vse.cz

# QUALITY MEASURE OF PARETO FRONT APPROXIMATION

***Jaroslav Janáček, Marek Kvet***
***University of Žilina, Faculty of Management Science and Informatics***

**Abstract**
Pareto front of public service system designs is an important tool for negotiation of public representatives and the system founder, especially, when a system criterion and a fair criterion are in a conflict. As the optimal design of the system for only one criterion combining the conflicting criteria using fixed weights is considerably computational time demanding, the complete Pareto front determination cannot be usually performed in acceptable time. That is why various heuristics are employed to obtain a Pareto front approximation by a finite set of non-dominated designs. To be able to compare outputs of different heuristic approaches to the Pareto front approximation, we suggested a way to measure quality of a resulting set of non-dominated designs and studied weak sites of the measure. Furthermore, we deal with algorithms for determination of bordering members of the set of non-dominated designs, which considerably influence the accuracy of the approximation.

***Keywords:*** *Public service system designing, Conflicting criteria, Pareto front, Heuristic approaches, Quality measurement*

***JEL Classification: C61***
***AMS Classification: 90C27***

## 1 INTRODUCTION

Local politicians responsible for satisfaction of specific demands of population have to choose an admissible compromise between the fairness and system objectives (Bertsimas, Farias, Trichakis, 2011, Buzna, Koháni, Janáček, 2013, Current, Daskin, Schilling, 2002, Doerner et al, 2005, Jánošíková, 2007, Marianov, Serra, 2002). The Pareto front of public service system designs would be an excellent tool for negotiation between public representatives and the system founder, because the Pareto front elements represent the complete set of relevant designs for negotiation. Each other design outside the Pareto front is dominated by some design of the Pareto front, i.e. both objective values of the dominating are better than or at least equal to the objective values of the dominated design (Arroyo et al, 2010).

Unfortunately, determination of the complete Pareto front is a very complex task, which demands huge computational time (Grygar, Fabricius, 2019, Janáček, Fabricius, 2021). This obstacle can be overcome by replacing the exact Pareto front with its good approximation, which can be obtained by some of quick heuristic method (Janáček, Kvet, 2021a, Janáček, Kvet, 2021b). To verify this approach, it is necessary to evaluate quality of produced approximations. This contribution is focused on suggestion and analysis of a quality measure together with improving procedures. The associated computational study is performed for the public service system design problem with two conflicting criteria.

## 2 PARETO FRONT APPROXIMATION BY NDSS

Let $\boldsymbol{Y}$ denote a finite set of all feasible solutions of the public service system design problem, where each solution $\boldsymbol{y} \in \boldsymbol{Y}$ can be evaluated according to two conflicting criteria $f_1$ and $f_2$. The conflict of the criteria means that minimizing process of one criterion causes undesirable increase of the other one. In the two-dimensional space of the criteria, a relation of dominancy can be established by following definition. We say that a solution $\boldsymbol{z}$ is dominated by a solution $\boldsymbol{y}$, if the clauses $f_1(\boldsymbol{y}) \leq f_1(\boldsymbol{z})$ and $f_2(\boldsymbol{y}) \leq f_2(\boldsymbol{z})$ hold. For practical reasons, the solution $\boldsymbol{z}$ corresponding to the same point in the two-dimensional space as the solution $\boldsymbol{y}$, i.e. $f_1(\boldsymbol{y}) = f_1(\boldsymbol{z})$ and $f_2(\boldsymbol{y}) = f_2(\boldsymbol{z})$, is considered to be equivalent to $\boldsymbol{y}$. As the common objective is to minimize both of the criteria, dominated solution of $\boldsymbol{Y}$ does not need to be taken into account. Thus the final solution for implementation can be chosen from such set of solutions, which are not dominated by any solution of $\boldsymbol{Y}$. Such a set of all non-dominated solutions is called Pareto front (Grygar, Fabricius, 2019, Janáček, Fabricius, 2021, Janáček, Kvet, 2021a, Janáček, Kvet, 2021b). Even if Pareto front for considered set of $\boldsymbol{Y}$ is finite, its determination by an exact approach is computational time-demanding task. That is why, we focus on ways of its approximation, where incrementing heuristic will be employed.

As the incrementing heuristics are based on changing a feasible solution by operations, which keep feasibility of newly obtained solutions, then performance of any of the heuristics yields a sequence of feasible solutions, which can be evaluated and used to update a set of non-dominated solutions (*NDSS*) with respect to processed solutions.

The further algorithm *UpdateNDSS* keeps the current set *NDSS* of non-dominated solutions $\boldsymbol{y}^1$, $\boldsymbol{y}^2, \ldots, \boldsymbol{y}^n$ in the form of a list of solutions ordered according to the values of $f_2$ in the ascending order, i.e. $f_2(\boldsymbol{y}^1) < f_2(\boldsymbol{y}^2) < \ldots. < f_2(\boldsymbol{y}^n)$. It implies that also the following inequalities must hold $f_1(\boldsymbol{y}^1) > f_1(\boldsymbol{y}^2) > \ldots. > f_1(\boldsymbol{y}^n)$.

The algorithm processes a newly obtained solution $\boldsymbol{z}$ according to the following steps.

*UpdateNDSS*(*NDSS*, $n$, $\boldsymbol{z}$)

1. Evaluate $f_2(\boldsymbol{z})$ and distinguish, which of the cases is valid.
   a) $f_2(\boldsymbol{z}) < f_2(\boldsymbol{y}^1)$
   b) It holds $f_2(\boldsymbol{y}^n) \leq f_2(\boldsymbol{z})$ or there exists $k$ so that $f_2(\boldsymbol{y}^k) \leq f_2(\boldsymbol{z}) < f_2(\boldsymbol{y}^{k+1})$.
2. According to the valid case perform
   a) Insert $\boldsymbol{z}$ into *NDSS* as the first solution together with deleting each current solution, which is dominated by $\boldsymbol{z}$.
   b) If $f_1(\boldsymbol{z}) \geq f_1(\boldsymbol{y}^k)$, then abandon $\boldsymbol{z}$, because it is dominated by $\boldsymbol{y}^k$. Otherwise insert $\boldsymbol{z}$ into *NDSS* as the $k$ + 1 st solution together with deleting each current solution, which is dominated by $\boldsymbol{z}$.

## 3 QUALITY OF NDSS

To be able to measure proximity of a current set of non-dominated solutions to the Pareto front (*PF*), a piece-wise constant curves determined by both current *NDSS* and Pareto front solutions can be used. The curve $c_{sY}$ for a sequence $s\boldsymbol{Y}$ of non-dominated solutions $\{\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots, \boldsymbol{y}^n\}$ ordered according to the values of $f_2$ in the ascending order is defined on the interval $[f_2(\boldsymbol{y}^1), f_2(\boldsymbol{y}^n)]$ by the prescription: If $x \in [f_2(\boldsymbol{y}^k), f_2(\boldsymbol{y}^{k+1}))$ for $k \in \{1, \ldots, n-1\}$, then $c_{sY}(x) = f_1(\boldsymbol{y}^k)$. The area $A_{sY}$ under the curve can be computed by the expression (1).

$$A_{sY} = \sum_{k=1}^{n-1} f_1\left(\boldsymbol{y}^k\right) * \left(f_2\left(\boldsymbol{y}^{k+1}\right) - f_2\left(\boldsymbol{y}^k\right)\right) \tag{1}$$

The basic idea of measuring quality of a Pareto front approximation *NDSS* consists in computing positive difference $A_{NDSS} - A_{PF}$.

Nevertheless, this approach may fail, if bordering solutions $\boldsymbol{y}^1$ and $\boldsymbol{y}^n$ of the approximation *NDSS* are too distant from bordering solutions of the Pareto front in the two-dimensional space $f_1 \times f_2$ (see Fig. 1). Due to inaccuracy of the *NDSS* bordering points, the difference can be negative.

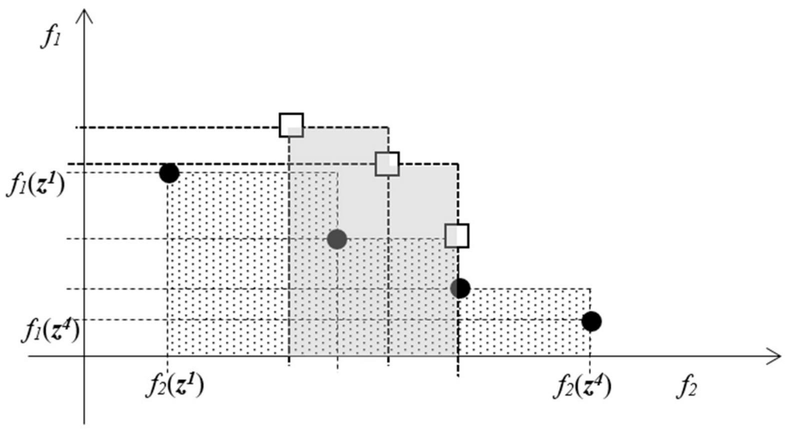



**Figure 1** The black circles denote points corresponding to Pareto front solutions $z^k$, whereas the squares denote points corresponding to solutions of *NDSS*. The dotted area corresponds to $A_{PF}$ and the grey area corresponds to $A_{NDSS}$.

If the Pareto front for the problem is known, or if there are known at least its bordering solutions $\boldsymbol{z}^1$ and $\boldsymbol{z}^{n(PF)}$, then the drawback can be removed by so called calibration of *NDSS* by adding dummy bordering solutions $\boldsymbol{y}^0$ and $\boldsymbol{y}^{n(NDSS)+1}$. The points for the dummy solutions can be defined by (2).

$$\begin{gathered}
\left[f_2\left(\boldsymbol{y}^0\right), f_1\left(\boldsymbol{y}^0\right)\right] = \left[f_2\left(\boldsymbol{z}^1\right), \max\left\{f_1\left(\boldsymbol{z}^1\right), f_1\left(\boldsymbol{y}^1\right)\right\}\right] \\
\left[f_2\left(\boldsymbol{y}^{n(NDSS)+1}\right), f_1\left(\boldsymbol{y}^{n(NDSS)+1}\right)\right] = \left[\max\left\{f_2\left(\boldsymbol{z}^{n(PF)}\right), f_2\left(\boldsymbol{y}^{n(NDSS)+1}\right)\right\}, f_1\left(\boldsymbol{z}^{n(PF)}\right)\right]
\end{gathered} \tag{2}$$

Regardless of the possible dominancy, the $A_{NDSS}$ can be computed according to (3).

$$A_{NDSS} = \sum_{k=0}^{n(NDSS)} f_1\left(\boldsymbol{y}^k\right) * \left(f_2\left(\boldsymbol{y}^{k+1}\right) - f_2\left(\boldsymbol{y}^k\right)\right) \tag{3}$$

The situation after addition of the dummy solutions is depicted in Fig. 2.

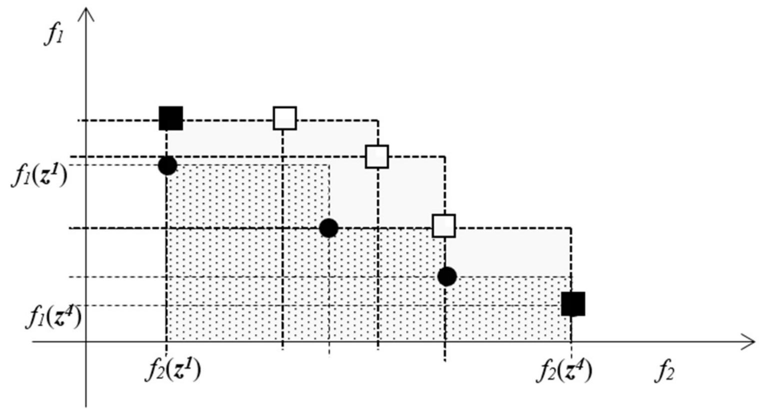



**Figure 2** The black circles denote points corresponding to Pareto front solutions $z^k$, whereas the squares denote points corresponding to solutions of enlarged *NDSS*, where the black squares denotes the added dummy points. The dotted area corresponds to $A_{PF}$ and the grey area corresponds to $A_{NDSS}$.

In the case, when the bordering solutions of the Pareto front are not known, the above correction cannot be used and thus the situation depicted in Fig. 1 may happen. In such situation the updated *NDSS* corresponds with smaller area than the PF area, which contradicts to purpose of the suggested measure.

This analysis shows that it is necessary to start each process employing the measure with proper determination of good approximations of the Pareto front bordering points.

# 4 INCREMENTING ALGORITHM FOR BI-CRITERIA LOCATION PROBLEM

The studied bi-criteria $p$-location problem formulation covers a broad family of the public service system designs, where $p$ facilities should be deployed in the set of $m$ possible service system center locations to service $n$ users in a serviced region. Each user $j$ calls for $b_j$ demand satisfactions. The set $\boldsymbol{Y}$ of all relevant feasible solutions is given by (4).

$$\boldsymbol{Y} = \left\{ \boldsymbol{y} \in \{0,1\}^m : \sum_{i=1}^{m} y_i = p \right\} \tag{4}$$

Each problem solution $\boldsymbol{y}$ can be evaluated by two criteria, which are in conflict. The first one is so called system criterion described by objective function (5), which is proportional to average response time of users' demand satisfaction.

$$f_1(\boldsymbol{y}) = \sum_{j=1}^{n} b_j \sum_{k=1}^{r} q_k \min_k \left\{ t_{ij} : i = 1,...,m;\ y_i = 1 \right\} \tag{5}$$

In the criterion description, $q_k$ stands for probability of the case that $k$-th nearest facility is the closest one, which is accessible. Constant $t_{ij}$ gives traversing time from the service center location $i$ to the user's location $j$. Operation $\min_k\{\}$ returns the $k$-th minimal value of the specified list of values.

The second (fair) criterion (6) expresses the number of users' demands, which are distant more than time $T$ from the nearest facility.

$$f_2(\boldsymbol{y}) = \sum_{j=1}^{n} b_j \max\left\{0,\ sign\left(\min\left\{t_{ij}\ : i=1,\dots,m;\ y_i=1\right\} - T\right)\right\} \tag{6}$$

To obtain a good solution of the problem for weighted criterion $a_1f_1(\boldsymbol{y}) + a_2f_2(\boldsymbol{y})$, an incrementing heuristic based on swapping two facility locations was suggested. The heuristic starts with an input solution $\boldsymbol{y}$ and for given vector $\boldsymbol{a} = [a_1, a_2]$ performs according the following steps. The heuristic uses the procedure $Exch(\boldsymbol{y}, i, j)$, which is applicable to any solution $\boldsymbol{y} \in \boldsymbol{Y}$, where $y_i = 1$ and $y_j = 0$. The procedure returns vector $\boldsymbol{y}' \in \boldsymbol{Y}$, where $y'_k = y_k$ for $k = 1, \dots, m$ ; $k \neq i$, $k \neq j$, and $y'_i = 0$ and $y'_j = 1$.

*Heuristic*($\boldsymbol{y}$, $\boldsymbol{a}$)
0. Set $F = a_1f_1(\boldsymbol{y}) + a_2f_2(\boldsymbol{y})$ and initialize $F^* = F$.
1. For each pair $(i, j) \in \{1, \dots, m\}^2$, where $y_i = 1$ and $y_j = 0$, do the following decisions.
   Compute $E = a_1f_1(Exch\ (\boldsymbol{y}, i, j)) + a_2f_2(Exch\ (\boldsymbol{y}, i, j))$.
   If $F > E$, then set $F = E$, $i^* = i$, $j^* = j$.
2. If $F^* = F$, then update $\boldsymbol{y} = Exch\ (\boldsymbol{y}, i^*, j^*)$ and continue with step 1, otherwise terminate with returning current $\boldsymbol{y}$ and $F^*$.

# 5 NUMERICAL EXPERIMENTS

The following computational study is focused on studying various approaches to determination of good approximations of the Pareto front bordering solutions. The used methods are based on usage of the presented swap algorithm *Heuristic*($\boldsymbol{y}$, $\boldsymbol{a}$). Especially, two ways of the *Heuristic*($\boldsymbol{y}$, $\boldsymbol{a}$) usage are studied. The first of them consists in repeating the algorithm run with different starting solutions $\boldsymbol{y}$ chosen from a set of uniformly deployed solutions (Janáček, Kvet, 2019a, Janáček, Kvet, 2019b). Influence of the number of repetitions on quality of the obtained *NDSS* bordering points is mapped.

The second way deals with perturbation of the input vector $\boldsymbol{a} = [a_1, a_2]$, which originally equals to for the most right or left bordering solution search. This vector can be changed to [1, eps] or [eps, 1] and effective size of perturbation coefficient eps is searched for.

The methods have been implemented in in the programming language Java within the NetBeans IDE 8.2 environment. The experiments were run on a common PC equipped with the Intel® Core™ i7 4790 CPU@ 3.60 GHz processor and 8 GB of RAM.

To evaluate the results, the benchmarks derived from the emergency medical systems currently implemented in eight Slovak self-governing regions were used. Members of the list of benchmarks denoted by name of capitals of the regions follows: Žilina (ZA), Trnava (TT), Trenčín (TN), Prešov (PO), Nitra (NR), Košice (KE), Banská Bystrica (BB) and Bratislava (BA). The abbreviations in brackets will be used to denote the benchmarks in the tables.

The objective function (5) minimizing the average response time of the system to clients is based on the assumption, that the nearest located facility may be temporarily unavailable and in such a case, the demand for service is assigned to the nearest available ambulance. Therefore, we take into account $r$ nearest service centers, which may provide the patient with the necessary care. In our experiments, the parameter $r$ was set to 3. This value is the smallest that sufficiently describes common performance of an Emergency Medical Service system (Kvet, 2014, Kvet, Janáček, 2021). The associated coefficients $q_k$ for $k$=1, …, $r$ were set in percentage in the following way: $q_1 = 77.063$, $q_2 = 16.476$ and $q_3 = 100 - q_1 - q_2$. These values were obtained from a simulation model of existing Emergency Medical Service system in Slovakia (Jankovič, 2016). Parameter $T$ used in the objective function (6) was set to the value of 10 minutes in accordance with our previous research published in (Grygar, Fabricius, 2019, Janáček, Kvet, 2021a, Janáček, Kvet, 2021b). The Pareto front bordering solutions of the very Pareto fronts were obtained from (Janáček, Fabricius, 2021). Characteristics of the bordering solutions together with other characteristic of the benchmarks are presented in Table 1. The symbols $DemF_1$ and $DemF_2$ stand for domains of the functions $f_1$ and $f_2$ of the Pareto front members.

**Table 1** Basic characteristics of used benchmarks and the bordering solutions

| Region: | ZA | TT | TN | PO | NR | KE | BB | BA |
|---|---|---|---|---|---|---|---|---|
| $m$ | 315 | 249 | 276 | 664 | 350 | 460 | 515 | 87 |
| $p$ | 29 | 18 | 21 | 32 | 27 | 32 | 36 | 14 |
| $RightF_1$ | 4210 | 41338 | 35273 | 56703 | 48939 | 45587 | 44751 | 26649 |
| $RightF_2$ | 728 | 921 | 567 | 1282 | 1001 | 816 | 935 | 28 |
| $LeftF_1$ | 48025 | 48964 | 45863 | 65944 | 59415 | 61241 | 53445 | 42912 |
| $LeftF_2$ | 254 | 450 | 223 | 711 | 557 | 276 | 453 | 0 |
| $DomF_1$ | 5915 | 7626 | 10590 | 9241 | 10475 | 15654 | 8694 | 16263 |
| $DomF_2$ | 474 | 471 | 344 | 571 | 444 | 540 | 482 | 280 |

The two approaches – repeating and perturbation were tested for several variants. The variants of repeating are denoted as *Rep*1*, Rep*2*, Rep*4*, Rep*8 *and* Rep16, where the number attached to the label *Rep* denotes the number of repeated runs. Each of the run is performed with different input solution withdrawn from a set of uniformly deployed solutions (Janáček, Kvet, 2019a, Janáček, Kvet, 2019b). The vector $\boldsymbol{a}$ from the repeating approach was set at either [1, 0] or [0, 1] for searching the most right and most left solutions respectively.

The variants of the perturbation approach are denoted by *Per*1, *Per*2 *and Per*4. These variants include only one run of the *Heuristic*($\boldsymbol{y}$, $\boldsymbol{a}$) for the vectors taking the forms [1, 0.01**rat*], [1, 0.02**rat*], and [1, 0.04**rat*] respectively for the most right bordering solution search. The coefficient *rat* is computed from the input solution $\boldsymbol{y}$ according to $rat = f_1(\boldsymbol{y})/ f_2(\boldsymbol{y})$ to balance influence of different sizes objectives domains.

The most left bordering solution search uses the following settings of the vector $\boldsymbol{a}$ [0.01, *rat*], [0.02, *rat*], and [0.04, *rat*] for then variants *Per*1, *Per*2 *and Per*4.

The results of all tested variants for the most right bordering solution are plotted in Table 2. In the rows denoted by $F_1$ and $F_2$, there are given the best found objective function values for the given benchmark. The rows denoted by $Dif_1$ and $Dif_2$, contains the differences of the obtained objective function values from the values of the right bordering solution of the Pareto front. These differences are given in percentage, where the associated base is the size of the associated objective function domain ($DomF_1$ or $DomF_2$). In the remaining rows of the table, we marked by asterisk that the given variant (row) for the given benchmark (column) reached the best result plotted in the first and second rows.

**Table 2** The results of all tested variants for the most right bordering solutions

| Region: | ZA | TT | TN | PO | NR | KE | BB | BA |
|---|---|---|---|---|---|---|---|---|
| $F_1$ | 42100 | 41338 | 35274 | 56703 | 48940 | 45587 | 44751 | 266649 |
| $F_2$ | 728 | 921 | 567 | 1282 | 996 | 816 | 935 | 280 |
| $Dif_1$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $Dif_2$ | 0 | 0 | 0 | 0 | 1.1 | 0 | 0 | 0 |
| *Rep*1 | | * | * | | * | | * | * |
| *Per*1 | | * | * | | | | * | * |
| *Per*2 | | * | * | | | | * | * |
| *Per*4 | | | * | | | | | * |
| *Rep*2 | * | * | * | | * | * | * | * |
| *Rep*4 | * | * | * | * | * | * | * | * |
| *Rep*8 | * | * | * | * | * | * | * | * |
| *Rep*16 | * | * | * | * | * | * | * | * |

The results of the most left bordering solution search are plotted in Table 3.

**Table 3** The results of the most left bordering solution search

| Region: | ZA | TT | TN | PO | NR | KE | BB | BA |
|---|---|---|---|---|---|---|---|---|
| $F_1$ | 48358 | 49418 | 47429 | 6603 | 59841 | 63263 | 53843 | 43095 |
| $F_2$ | 255 | 457 | 223 | 728 | 563 | 290 | 464 | 0 |
| $Dif_1$ | 8.7 | 6.0 | 14.8 | 0.1 | 4.1 | 12.9 | 4.6 | 1.1 |
| $Dif_2$ | 0.2 | 1.5 | 0 | 2.3 | 1.4 | 2.6 | 2.3 | 0 |
| *Rep*1 | | | | | | | | |
| *Per*1 | * | | | | | | | * |
| *Per*2 | * | | | | | | | * |
| *Per*4 | * | | | | | | | * |
| *Rep*2 | | | | | | | | |
| *Rep*4 | | | * | | | | | |
| *Rep*8 | | * | * | | * | | | |
| *Rep*16 | | * | * | * | * | * | * | |

The swap heuristic proves to be very efficient algorithm to approximate the most right member of the Pareto front of the solved problem. In all cases, almost exact solution has been found by the most of tested variants. The variants *Rep*4, *Repi*8 and *Rep*16 achieved the exact solution in all benchmarks. The perturbation approach proved not to have any effect. It achieved good solution only in cases, when each variant with exception of *Per*4 produced exact solution. Obviously the variant *Per*4 uses exaggerative size of the perturbation coefficient.

The most left Pareto front member approximation has proved to be much harder task than the most right member approximation. The results plotted in Table 3 show that the optimal value of the minimized fair objective $f_2$ has been achieved with accuracy of 2.6% of the fair objective domain and, furthermore, the inaccuracy in the system objective varied up to 15% of the system objective domain. In addition, the performed competition of the tested variants has no unambiguous winner. The variant *Rep*16 won in 6 benchmarks of 8, but the perturbation variants won in benchmarks ZA and BA.

# 6 CONCLUSIONS

The main goal of this paper was to design a good measure of Pareto front approximation quality for usage in heuristic processes, which are developed to determine a good approximation of the Pareto front of the bi-criteria public service system designs. We suggested the quality measure and yielded analysis of the possible weak sides of the measure. Furthermore, we performed a research of suitability of swap methods for determination of bordering points of the Pareto front to mitigate the weak sides of the suggested measure. The preliminary research confirmed the hypothesis that the success of the complexity of the successful method for the bordering solution determination depends on form of the decisive objective, where min-max objective demands much bigger computational time to determine the good approximation.

**Acknowledgements:** This work was supported by the research grants VEGA 1/0216/21 "Design of emergency systems with conflicting criteria using artificial intelligence tools" and VEGA 1/0654/22 "Cost-effective design of combined charging infrastructure and efficient operation of electric vehicles in public transport in sustainable cities and regions". This work was supported also by the Slovak Research and Development Agency under the Contract no. APVV-19-0441.

**References**

[1] Arroyo, J. E. C., dos Santos, P. M., Soares, M. S. and Santos, A. G. (2010). A Multi-Objective Genetic Algorithm with Path Relinking for the p-Median Problem. In: Proceedings of the 12th Ibero-American Conference on Advances in Artificial Intelligence, pp. 70-79.

[2] Bertsimas, D., Farias, V. F. and Trichakis, N. (2011). The Price of Fairness. In: *Operations Research*, 59, pp. 17-31.

[3] Buzna, Ľ., Kohány, M. and Janáček, J. (2013). Proportionally Fairer Public Service Systems Design. *Communications - Scientific Letters of the University of Žilina*, 15(1), pp. 14-18.

[4] Current, J., Daskin, M. and Schilling, D. (2002). Discrete network location models, Drezner Z. et al. (ed) *Facility location: Applications and theory*, Springer, pp. 81-118.

[5] Doerner, K. F., Gutjahr, W. J., Hartl, R. F., Karall, M. and 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.

[6] Grygar, D. and Fabricius, R. (2019). An efficient adjustment of genetic algorithm for Pareto front determination. In: TRANSCOM 2019: conference proceedings, Amsterdam: Elsevier Science, pp. 1335-1342.

[7] Janáček, J. and Fabricius, R. (2021). Public service system design with conflicting criteria. *IEEE Access: practical innovations, open solutions*, 9, pp. 130665-130679.

[8] Janáček, J. and Kvet, M. (2019a). Uniform Deployment of the p-Location Problem Solutions. In: Operations Research Proceedings 2019, Springer, pp. 315-321.

[9] Janáček, J. and Kvet, M. (2019b). Usage of Uniformly Deployed Set for p-Location Min-Sum Problem with Generalized Disutility. In: SOR 2019 proceedings, pp. 494-499

[10] Janáček, J. and Kvet, M. (2021a). Swap Heuristics for Emergency System Design with Multiple Facility Location. In: Proceedings of the 39th International Conference on Mathematical Methods in Economics, 2021, pp. 226-231.

[11] Janáček, J. and Kvet, M. (2021b). Emergency Medical System under Conflicting Criteria. In: SOR 2021 Proceedings, pp. 629-635.

[12] Jankovič, P. (2016). Calculating Reduction Coefficients for Optimization of Emergency Service System Using Microscopic Simulation Model. In: 17th International Symposium on Computational Intelligence and Informatics, pp. 163-167.

[13] Jánošíková, Ľ. (2007). Emergency Medical Service Planning. *Communications - Scientific Letters of the University of Žilina*, 9(2), pp. 64-68.

[14] Kvet, M. (2014). Computational Study of Radial Approach to Public Service System Design with Generalized Utility. In: Proceedings of International Conference DT 2014, Žilina, Slovakia, pp. 198-208.

[15] Kvet, M. and Janáček, J. (2021). Incrementing Heuristic for Non-Dominated Designs of Emergency Medical System. In: SOR 2021 Proceedings, pp. 429-474.

[16] Marianov, V. and Serra, D. (2002). Location problems in the public sector, Drezner Z. et al. (ed) *Facility location - Applications and theory*, Berlin, Springer, pp. 119-150.

**Author's address**
Prof. RNDr. Jaroslav Janáček, CSc., doc. Ing. Marek Kvet, PhD.
University of Žilina - Faculty of Management Science and Informatics
Univerzitná 8215/1, 010 26 Žilina, Slovakia
Email: jaroslav.janacek@fri.uniza.sk, marek.kvet@fri.uniza.sk

# THE HIERARCHICAL *PQ*-MEDIAN MODEL FOR AMBULANCE LOCATION IN A CITY

***Ľudmila Jánošíková, Peter Jankovič, Marek Kvet***
***University of Žilina, Faculty of Management Science and Informatics***

**Abstract**
In this paper, we optimize the locations of the ambulance base stations in a city using the hierarchical *pq*-median model. The methodology is verified in real-world setting. The stations in all eight regional capitals in Slovakia are optimally re-distributed. The computer simulation model is used to evaluate the current and proposed sets of locations. The simulation study proves that the hierarchical *pq*-median model results in a considerable reduction of the response times, mainly to the most serious patients. The greatest reduction is achieved in Banská Bystrica (by almost 5 minutes on average).

***Keywords:*** *Emergency medical service, Response time, Hierarchical pq-median problem, Computer simulation*

***JEL Classification:*** JEL C61, JEL C63
***AMS Classification:*** AMS 65C05, AMS 90B80, AMS 90C10

## 1 INTRODUCTION

Emergency medical service (EMS) plays an important role in pre-hospital health care. Most European countries apply the Franco-German model of EMS delivery where the ambulance crew is qualified to treat patients in their homes or at the scene of an emergency incident. If necessary, afterwards they transport the patient to an appropriate hospital. In Slovakia, we have the two-tiered EMS system where two types of ambulances are in operation. They differ in the equipment and the qualification of the rescue team. Most ambulances provide basic life support (BLS; Slovak abbreviation RZP) and have only a paramedic and a rescue driver on board. About one third of ambulances are well-equipped advanced life support units (ALS; Slovak abbreviation RLP). An ALS crew consists of a physician, a paramedic and a driver. The staff is capable of performing additional life-saving procedures, e.g. inserting breathing tubes. The closest available ambulance to the emergency site is always dispatched regardless of its type. If it is a BLS ambulance and the incident is life-threatening, then the closest available ALS ambulance is dispatched concurrently. The rationale is that any medical treatment is better than waiting without a professional intervention for the arrival of a doctor. The consequence of such organization is that ALS ambulances transport few patients to a hospital than BLS ambulances. The data on rescue trips performed in January to September 2020 provided us by the National Dispatch Center reveals that the proportion of transported patients was 53 % and 71 % for ALS and BLS ambulances, respectively.

Patients' survivability and morbidity are affected by the efficiency of the interventions that, in turn, depend on the ambulance base location, the ambulance allocation to the stations, and ambulance dispatching decisions. This study deals with the location of the base stations. Currently (September 2021), there are 274 stations distributed all over the country. The term "station" has an organizational, rather than a physical, meaning. The government determines the number of stations and their distribution across the country, assuming that every station houses one ambulance. The government regulations specify which municipality (a town, city district, or village) has one or multiple stations, but it does not specify the particular addresses of the stations. After a provider of urgent healthcare gets a license to operate a station, they choose a suitable building with a garage for the ambulance and a room where the crew waits in-between rescue trips. If one company operates in multiple stations in a town, it may decide to locate multiple stations under the same address, so the stations share the same building. For example, in the city of Banská Bystrica, Záchranná zdravotná služba Bratislava operates in four stations located at the same address. We believe that using operations research methods a better distribution of the stations can be proposed with the view to increase the availability of the service, especially for patients in critical conditions.

In this research, we concentrate on locating EMS stations in an urban environment. We respect the current number of stations but look to locate them in a better way. The costs of relocating the stations from their current positions are not considered because it is beyond the scope of our research to look for potential locations where the stations can be sited and thus to estimate investment costs related to the candidate locations. Instead of considering the costs associated with every closed or open station, one can limit the number of stations that can be relocated under a given budget constraint. Currently, we suppose that no stations can be added, but there are no limitations regarding the number of relocated stations.

The number of articles related to EMS planning has grown rapidly during the last two decades. This growth may be attributed to increases in the availability of data and in the power of decision-supporting software tools. Several survey articles have been published during this period. A recent survey of location models applied to healthcare facilities was presented by (Ahmadi-Javid, Seyedi and Syam, 2017). The study (Van Den Berg and Van Essen, 2019) assessed several ambulance location models in the view of the performance indicators frequently used in practice. Simulation experiments with real-world case studies showed that the maximum expected covering location problem (MEXCLP) and the expected response time model (ERTM) performed the best. However, the ERTM model is rather complicated, and it takes a long computing time. That is why the average response time model is recommended as a good alternative. The average response time model is an application-related name for the model of the well-known *p*-median problem (*p*MP) that seeks the location of *p* facilities so that the average travel time between customers and their closest facility is as short as possible.

The EMS system with two types of rescue units can be viewed as a hierarchical facility system. Using the classification by (Şahin and Süral, 2007), we face a multi-flow, nested, and non-coherent system. If the objective is to minimize the total distance (or travel time, respectively) from demand zones to the closest ALS and BLS stations, then the hierarchical *pq*-median problem is to be solved. In our previous study (Jánošíková, Jankovič, Kvet and Zajacová, 2021) we proposed a modification of the *pq*-median model by (Serra and ReVelle, 1993). The paper (Serra and ReVelle, 1993) focused on coherent systems where all demand areas assigned to a lower level facility must be assigned to one and the same upper level facility. However, this condition does not hold in the EMS system. Thus, we amended their model for a non-coherent system and applied it for the location of the EMS stations across the country. In this paper, we apply the aforementioned hierarchical *pq*-median model in an urban area.

## 2 THE HIERARCHICAL *PQ*-MEDIAN MODEL

To model demand zones in a city, a common approach is to cover the city by a rectangular grid where grid elements correspond to demand zones (Aringhieri, Carello and Morale, 2013; Schmid and Doerner, 2010; Zaffar et al., 2016). The demand for EMS is usually derived from records on historical interventions archived by an information system. The demand points are identified with the centres of the cells, or with the nodes on the road network that are the nearest ones to the centres of the cells. We apply the latter approach since it allows for the calculation of the distance matrix by using the road network and the estimated speed of ambulances (Jánošíková, Kvet, Jankovič and Gábrišová, 2019). Further, we suppose that the demand points are also candidate locations for the stations.

The model is formulated using the following sets, indices, parameters and decision variables.

*Sets and indices*

$I$ set of candidate locations
$J$ set of demand points
$i \in I$ candidate location
$j \in J$ demand point

*Parameters*

$p$ number of stations (regardless of their type) to be sited in the city
$q$ number of ALS stations to be sited
$b_j$ the number of EMS patients in demand zone $j$
$t_{ij}$ shortest travel time from candidate location $i$ to demand point $j$

*Decision variables*

$$y_i = \begin{cases} 1, & \text{if a station is located at site } i \\ 0, & \text{otherwise} \end{cases}$$

$$x_{ij} = \begin{cases} 1, & \text{if demand point } j \text{ is served by the station located at site } i \\ 0, & \text{otherwise} \end{cases}$$

$$u_i = \begin{cases} 1, & \text{if an ALS station is located at site } i \\ 0, & \text{otherwise} \end{cases}$$

$$v_{ij} = \begin{cases} 1, & \text{if demand point } j \text{ is served by the ALS station located at site } i \\ 0, & \text{otherwise} \end{cases}$$

At the lower level of the hierarchical problem we locate all stations regardless of their type and create their catchment areas, i.e. the following $p$-median model is solved:

$$\textit{minimize} \quad \sum_{i\in I}\sum_{j\in J} t_{ij}\, b_j x_{ij} \tag{1}$$

$$\textit{subject to} \quad \sum_{i\in I} x_{ij} = 1 \quad \textit{for } j \in J \tag{2}$$

$$x_{ij} \le y_i \quad \textit{for } i \in I,\ j \in J \tag{3}$$

$$\sum_{i\in I} y_i = p \tag{4}$$

$$y_i, x_{ij} \in \{0,1\} \quad \textit{for } i \in I,\ j \in J \tag{5}$$

The objective function (1) minimizes the total travel time needed by the ambulances to reach the patients. The average travel time is equal to the total travel time divided by the number of all patients. The average travel time is a lower bound of the real response time since the model assumes that there is always an ambulance available to answer to the call and therefore an ambulance from the closest station is always dispatched to the patient. Constraints (2) assign every demand point $j$ to exactly one station $i$. Constraints (3) ensure that if a demand point $j$ is assigned to a node $i$, then a station will be opened at the node $i$. Constraint (4) limits the number of stations. The obligatory constraints (5) specify the definition domains of the variables.

The upper level of the model decides which stations opened at the lower level will house ALS ambulances:

$$\text{minimize} \quad \sum_{i\in I}\sum_{j\in J} t_{ij}\, b_j v_{ij} \tag{6}$$
$$\text{subject to} \quad \sum_{i\in I} v_{ij} = 1 \quad \text{for } j \in J \tag{7}$$
$$v_{ij} \le u_i \quad \text{for } i \in I,\ j \in J \tag{8}$$
$$u_i \le y_i \quad \text{for } i \in I \tag{9}$$
$$\sum_{i\in I} u_i = q \tag{10}$$
$$u_i, v_{ij} \in \{0,1\} \quad \text{for } i \in I,\ j \in J \tag{11}$$

The objective function (6) minimizes the total travel time needed by the ALS ambulances to reach all patients. Constraints (7) assign every demand point $j$ to the catchment area of exactly one ALS station $i$. Constraints (8) say that a demand point $j$ can be assigned only to an open ALS station. Constraints (9) allow an ALS ambulance to be allocated only to a station located at node $i$ at the lower level of hierarchy (optimal values of variables $y_i$ are regarded as coefficients in this model). Constraint (10) limits the number of located ALS stations to their current amount $q$. The remaining constraints (11) specify binary variables.

## 3 RESULTS AND DISCUSSION

The National Dispatch Center provided us grids for all Slovak regional capitals. The size of a grid cell is 250 × 250 m. Demand consists of high- and middle-priority patients served from August 2018 to December 2021 by all ambulances located in the town. A call-taker assigns high priority (denoted as K) to patients who are supposed to be in a life-threatening condition because the failure of at least one life function was detected during the emergency call. It is supposed that without providing a professional rescue the person would die. Middle priority (denoted as N) is assigned to patients who have difficulties related to at least one basic life function. Without dispatching a rescue team the person's health would be damaged. Most high-priority patients have severe diagnoses that are denoted as the First Hour Quintet (FHQ), and they include: chest pain, severe trauma, stroke, severe respiratory difficulties, and cardiac arrest. For these patients the response time is critical because every minute of delay reduces their chance of survival.

We implemented the hierarchical *pq*-median model for the regional capitals with the parameters that are summarized in Table 1.

The proposed and current sets of locations were evaluated using a detailed computer simulation model described in (Jánošíková, Kvet, Jankovič and Gábrišová, 2019). Computer simulation enables us to estimate the impact of the proposed changes in the EMS system infrastructure. We focus on the following performance indicators:

1. Average response time to all patients, since it has been monitored by the Operation Centre of the EMS of the Slovak Republic;
2. Percentage of all calls responded to within 15 min, because a 15-minute response time is a target for 95% calls in Slovakia;
3. Average response time to FHQ patients;
4. Percentage of FHQ patients responded to within 8 min.

**Table 1.** Characteristics of the instances

| Town | Number of BLS stations | Number of ALS stations | Number of demand zones | Total demand (K + N patients) |
|---|---|---|---|---|
| Banská Bystrica | 2 | 2 | 395 | 8093 |
| Bratislava | 15 | 3 | 1823 | 33424 |
| Košice | 6 | 3 | 865 | 23886 |
| Nitra | 2 | 2 | 533 | 7686 |
| Prešov | 3 | 2 | 445 | 9611 |
| Trenčín | 2 | 2 | 355 | 7801 |
| Trnava | 1 | 1 | 286 | 6426 |
| Žilina | 2 | 2 | 448 | 9886 |

The computational experiments were performed on a personal computer equipped with the Intel Core i7 processor with 1.60 GHz and 8 GB RAM. The solver Xpress Optimizer 8.3 (64-bit, release 2017) was used to solve the location models. The agent-based simulation model was implemented in AnyLogic simulation software.

The current and optimal distribution of the stations were simulated. The simulation experiment for one set of station locations consisted of 10 replications. One replication took 94 days. The mean values from 10 replications are reported.

The results of the simulation study are in Table 2. The values of the indicators for the optimal locations of the stations are in the columns denoted as "Opt". The difference in relation to the current distribution of the stations are in the "Diff" columns. The table contains the indicators for the regional capitals, and also for the whole country. The capital Bratislava and the second largest Slovak town Košice are assessed at the district levels. The best improvements of the indicators are displayed in bold.

The computer simulation of the current system revealed that the system is short of the target to reach 95% of patients within 15 min. The real accessibility within this time limit is barely 71%. The Slovak system also exhibits poor performance regarding the 8-minute response-time standard for critical patients. Only 34% of FHQ patients are responded to within 8 min. The EU average is 67% (Kraft et al., 2006).

Further, there are big differences in EMS availability among towns. The best situation is in Košice I district and Prešov, where the average response time to all as well as to FHQ patients is approximately 6 min. Approximately 80% of FHQ patients gets professional first aid within 8 min. On the other side, response times in Trnava are about 13 min and only 16% FHQ patients is responded to within 8 min. Optimization reduces the response times in all regional capitals and districts with the only exception of Bratislava IV district, where the average response time to all patients is increased by about half minute. The most significant improvement is observed in Banská Bystrica, where the average response time to all patients is reduced by almost 4 min and to the most serious patients by almost 5 min. The accessibility of FHQ patients is improved from 16% to 72%.

**Table 2.** Performance indicators for the optimal locations

| Area | Average response time to all patients (min) | | % of calls responded to within 15 min | | Average response times to FHQ patients (min) | | % of FHQ patients responded to within 8 min | |
|---|---|---|---|---|---|---|---|---|
| | Opt | Diff | Opt | Diff | Opt | Diff | Opt | Diff |
| Slovakia | 11.97 | -0.35 | 71.66 | 1.05 | 11.40 | -0.45 | 37.43 | 3.23 |
| Banská Bystrica | 7.17 | **-3.93** | 95.62 | 8.95 | 6.52 | **-4.79** | 72.01 | **55.86** |
| Nitra | 7.64 | -1.09 | 89.46 | 2.23 | 7.00 | -1.20 | 73.00 | 11.67 |
| Prešov | 5.52 | -0.84 | 91.99 | -0.38 | 5.15 | -0.87 | 88.12 | 5.77 |
| Trenčín | 9.68 | -0.71 | 81.96 | 1.69 | 8.88 | -0.73 | 53.80 | 7.56 |
| Trnava | 11.39 | -1.76 | 77.38 | 6.26 | 10.28 | -2.77 | 41.77 | 25.95 |
| Žilina | 10.64 | -0.05 | 81.27 | -1.67 | 9.54 | -1.03 | 46.87 | 14.34 |
| Bratislava I | 6.48 | -1.71 | 97.18 | 1.55 | 5.93 | -1.59 | 81.86 | 16.88 |
| Bratislava II | 8.67 | -1.33 | 90.30 | 3.73 | 7.78 | -1.26 | 65.66 | 10.26 |
| Bratislava III | 10.75 | -2.85 | 80.47 | **9.84** | 9.54 | -2.91 | 52.77 | 20.16 |
| Bratislava IV | 11.65 | 0.45 | 79.16 | 0.58 | 10.14 | -0.13 | 45.45 | -3.83 |
| Bratislava V | 6.36 | -2.37 | 95.02 | 4.36 | 5.79 | -2.26 | 80.88 | 15.83 |
| Košice I | 4.76 | -1.37 | 98.76 | 0.56 | 4.69 | -1.05 | 89.53 | 8.94 |
| Košice II | 8.41 | -1.10 | 82.68 | 0.49 | 8.45 | -0.78 | 63.33 | 3.85 |
| Košice III | 5.10 | -3.05 | 98.14 | 4.56 | 4.98 | -3.08 | 87.47 | 30.10 |
| Košice IV | 5.08 | -1.59 | 98.77 | 1.59 | 4.91 | -1.19 | 87.34 | 9.22 |

Re-distribution of the stations in regional capitals has the impact on the EMS provision in their surrounding areas as well because the ambulances located in the towns also serve patients in neighbouring villages. Therefore, better distribution of the stations in all regional capitals would result in statistically significant reduction of the response times even at the countrywide level.

## 4 CONCLUSIONS

The research applies the hierarchical *pq*-median model to optimal location of the EMS stations in an urban environment. While currently multiple stations are located at the same place, the stations in the optimal solution are distributed across the whole city territory. It enables the responsiveness of EMS to be increased not only in the city but also in the surrounding area.

The weak point of our research is that there were no limits on the potential locations of the stations. We suppose that the station can be located at any central node of the grid. It was beyond our capabilities to investigate potential locations with all the aspects that have to be taken into account in practical station relocation (i.e., the existence of a proper building, the costs of its tenancy, etc.). Thus, our results should be interpreted in terms of the wards surrounding the calculated optimal station locations rather than stations' precise addresses. At present, such an interpretation is satisfactory for the decision-makers, as follows from the discussions with the EMS authorities.

We note that in practice it may not be acceptable to relocate all stations operating in the city, mainly for economic reasons that are not incorporated into our study. However, the proposed model can easily cope with such a limitation by the appropriate setting of the $p$ and $q$ parameters.

Probably the most important input parameter that affects not only the optimal locations but also the simulated results is the travel time along the road network. The travel time depends on the average speed of the ambulances. It in turn depends on the traffic intensity in the course of a day and on the road category. In this research, we adopted rather old estimates from the study (Jánošíková, Kvet, Jankovič and Gábrišová, 2019). However, we hope that in the future we will be able to obtain up-to-date GPS traces of ambulances' trips and derive more precise estimates such as in (Buzna and Czimmermann, 2021).

**Acknowledgements:** This work was supported by the research grant VEGA 1/0216/21 "Designing of emergency systems with conflicting criteria using tools of artificial intelligence" and by 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] Ahmadi-Javid, A., Seyedi, P. and Syam, S. (2017). A survey of healthcare facility location. *Computers & Operations Research*, 79, pp.223-263.
[2] Aringhieri, R., Carello, G. and Morale, D. (2016). Supporting decision making to improve the performance of an Italian Emergency Medical Service. *Annals of Operations Research*, 236(1), pp.131-148.
[3] Buzna, Ľ. and Czimmermann, P. (2021). On the Modelling of Emergency Ambulance Trips: The Case of the Žilina Region in Slovakia. *Mathematics*, 9(17), p.2165.
[4] Jánošíková, Ľ., Jankovič, P., Kvet, M. and Zajacová, F. (2021). Coverage versus response time objectives in ambulance location. *International Journal of Health Geographics*, 20(1).
[5] Jánošíková, Ľ., Kvet, M., Jankovič, P. and Gábrišová, L. (2019). An optimization and simulation approach to emergency stations relocation. *Central European Journal of Operations Research*, 27(3), pp.737-758.
[6] Krafft, T., García-Castrillo Riesgo, L., Fischer, M., Lippert, F., Overton, J. and Robertson-Steel, I. (2006). *Health Monitoring & Benchmarking of European EMS Systems: Components, Indicators, Recommendations*. Köln, Germany: Arbeitsgruppe GEOMED, Geographisches Institut der Universität zu Köln.
[7] Şahin, G. and Süral, H. (2007). A review of hierarchical facility location models. *Computers & Operations Research*, 34(8), pp.2310-2331.
[8] Schmid, V. and Doerner, K. (2010). Ambulance location and relocation problems with time-dependent travel times. *European Journal of Operational Research*, 207(3), pp.1293-1303.
[9] Serra, D. and ReVelle, C. (1993). The pq-median problem: Location and districting of hierarchical facilities. *Location Science*, 1(4), pp.299-312.
[10] Van Den Berg, P. and Van Essen, J. (2019). Comparison of static ambulance location models. *International Journal of Logistics Systems and Management*, 32(3/4), pp.292-321.
[11] Zaffar, M., Rajagopalan, H., Saydam, C., Mayorga, M. and Sharer, E. (2016). Coverage, survivability or response time: A comparative study of performance statistics

used in ambulance location models via simulation–optimization. *Operations Research for Health Care*, 11, pp.1-12.

**Author's address**
prof. Ing. Ľudmila Jánošíková, PhD.
University of Žilina, Faculty of Management Science and Informatics, Department of Mathematical Methods and Operations Research
Univerzitná 1, 010 26 Žilina
Slovak Republic
e-mail: Ludmila.Janosikova@fri.uniza.sk

Ing. Peter Jankovič, PhD.
University of Žilina, Faculty of Management Science and Informatics, Department of Mathematical Methods and Operations Research
Univerzitná 1, 010 26 Žilina
Slovak Republic
e-mail: Peter.Jankovic@fri.uniza.sk

doc. Ing. Marek Kvet, PhD.
University of Žilina, Faculty of Management Science and Informatics, Department of Informatics
Univerzitná 1, 010 26 Žilina
Slovak Republic
e-mail: Marek.Kvet@fri.uniza.sk

# GAME THEORY AS AN APPROACH TO THE ANALYSIS OF THE INDIVIDUAL DECISIONS AGAINST POPULATION

*Miroslava Jánošová, Veronika Mitkova*
*Faculty of Social and Economic Sciences, Comenius University Bratislava*

**Abstract**
Game theory is a scientific discipline applicable in various areas of everyday life. In the current pandemic situation and the associated need to ensure vaccination are the principles of the game theory representative example for analysis of the individuals‘ decisions in relation to the population as a whole. The paper aims in the first part to introduce construction methodology a non-zero-sum bi-matrix game with two players represented by the individual and whole population as describe by (Georgiou, 2020). Subsequently on the base of this methodology we identify equilibrium through strict dominance strategy and Nash equilibrium in the pure strategies. In the second part of the paper, through the groupings of individual strategies, we formulate a game of 2by2 which will be analyzed in terms of finding Nash equilibrium in pure strategies.

***Keywords:*** *Vaccination Game, Static Game, Game Theory, Nash Equilibrium in Pure Strategies*

***JEL Classification:*** C720
***AMS Classification:*** 91A35, 91A80

## 1 INTRODUCTION

Individuals' decisions in relation to health issues are a frequently debated issue nowadays because of the COVID 19 pandemic and are subject to the use of different methods of analysis. It is important to have in mind that the decisions in relation to health realized by individuals do not necessarily benefit the whole population and vice versa. The possible conflicting interests can be obstacles to making crucial decisions by authorities, as it is mandatory vaccination. An individual should consider not only the possible side effects of the vaccines but also the possibility of being infected by the new coronavirus despite all available vaccines being declared safe and their use being proven by the European Medicines Agency. Since the COVID 19 may be a fatal disease in some cases, the individual's decision may seem unambiguous, and it may be considered an acceptance of the side effects. However, in case of high vaccination coverage rate, this individual may consider being protected by collective immunity hence being safe against the virus, and the best decision for him is to not vaccinate and eliminate its side effects. This decision may be made by many individuals leading to the impairment of collective immunity, increasing the likelihood of infection, and faster spread of infection. The game theory proves that, in this case, the best decision for the individual is the best for the whole population, but it fulfills the Nash equilibrium, which is not the Pareto effective equilibrium. The result is given by the interaction among engaged individuals, and how the risk is perceived by the individuals and the population. What is the best strategy in this situation? We use Nash equilibrium, assuming that the change of our strategy does not improve our situation if the strategy of other players remains the same.

There may be found many benefits of highlighting the potential value of game theory in the development and evaluation of public health policy concerning vaccination. Group interest versus self-interest was examined using game theory and epidemic modeling on smallpox vaccination by Bauch, Galvani, and Earn (2003). Their results illustrate how this conflict typically causes large differences between vaccine coverage levels, that are best for the group, and uptake levels that might be achieved under a voluntary vaccination program. They prove that self-interested decision-making could lead to seriously suboptimal vaccine uptake levels. Bauch and Earn (2004) use a formal game-theoretical analysis to show that if a sufficient proportion of the population gains immunity (either by vaccination or naturally), then even a small risk linked to vaccination outweighs the risk from infection. Individual self-interest then may lead to vaccine-preventable disease eradication. Next, using game and epidemic models Reluga, Bauch, and Galvani (2006), demonstrate that the optimal individual choice between vaccination and non-vaccination depends on the self-interest which often leads to stable dynamics but also to oscillations in vaccinations over time. The more homogenous the population is in the perception of vaccine and infection risk, the more unstable the choice is.

Chapman et al. (2012) examine the conditions under which young individuals, who contribute most to disease spread of influenza, would get vaccinated to protect elderly individuals who face the highest mortality risk. They proved that the payout structure of this game influences whether individuals got vaccinated for group benefit or self-interest. The Nash equilibrium, predicting self-interested behavior, is consistent with a situation in which players are paid according to individual point totals and more elderly got vaccinated. However, more young individuals got vaccinated when players were paid according to group point totals, finding consistent with the utilitarian equilibrium predicting group. Bai (2016) proves that there is a unique Nash equilibrium for each vaccination game in a homogenous mixing population. The critical condition for the uniqueness of the Nash equilibrium set is that the attack ratio monotonically decreases as the coverage level increases.

Piraveenan et al. (2021) use game theory tools to model either the policymakers (vaccine givers) or individuals' (vaccine takers) decision-making process in SARS-CoV-2 vaccination uptake. The cooperative game theory is used to model the vaccination process as a public good and the non-cooperative game theory to model the selfish decisions of the individuals. They applied game theory also in the allocation and prioritization of resources needed for the implementation of a vaccination program against SARS-CoV-2. Choi and Shim (2020) developed an epidemiological game-theoretic model for vaccination and social distancing during the ongoing COVID-19 pandemic with individual maximization of payoffs. They identified optimal strategy based on the Nash strategy under two different circumstances: both available instruments and only one available instrument. One of the first and most cost-effective instruments during the COVID-19 pandemic was wearing a face mask which may be considered a public good game structure in which an individual does not want to wear a mask but requires others to wear it. Kabir et al. (2021) developed a new intervention game model combining the mathematical models of epidemiology with evolutionary game theory to quantify how people use mask-wearing.

## 2 METHODOLOGY

We will formulate the game as a mathematical model in which we assume that all players in the game are rational players who implement their decision at the same time. The first player is represented by the individual, and the second player is represented by the population. This is the reason why the game is formulated as a non-zero-sum game we used methodology as describe by (Georgiou, 2020) in his technical report and we enlargement with another strategy.

## 2.1 Game in normal form with pure strategies

Based on a methodology of a static game with complete information, as described by (Tadelis 2013). The formal framework of such methodology is defined by capturing the strategic nature of such a game. The normal form of the game is defined by the set of players, the set of players' strategies and the set of payoff functions for each player - the following sets create a payoff matrix capturing the value of the payoff by combining all the options from the set of their strategies.

**Definition 1:** „A normal-form game includes three components as follows:

1. *A finite set of players,* $N = \{1, 2, \dots n\}$.
2. *A collection of sets of pure strategies,* $\{S_1, S_2, \dots, S_n\}$
3. *A set of payoff functions,* $\{v_1, v_2, \dots, v_n\}$*, each assigning a payoff value to each combination of chosen strategies, that is, a set of function* $v_i: S_1 \times S_2 \times \dots \times S_n \to R$ *for each* $i \in N$*.“*[1]

All the options from which players can choose their action can be formalized through a set of pure strategies.

**Definition 2:** *„A pure strategy for player* $i$ *is a deterministic plan of action. The set of all pure strategies for player* $i$ *is denoted* $S_i$ *A profile of pure stategies* $s = (s_1, s_2, \dots, s_n)$*,* $s_i \in S_i$ *of all* $i = 1, 2, \dots, n$*, describes a particular combination of pure strategies chosen by all* $n$ *players in the game. “*[2]

In relation to strategies, it is also appropriate to define the concept of strict dominance based on which we can clearly say that the strictly dominated strategy will not be played by any of the rational players, because in that case it would not minimize its loss or maximize its profit.

**Definition 3:** „$s_i \in S_i$ *is a strictly dominant strategy for* $i$ *if every other strategy of* $i$ *is strictly dominanated by it, that is,* $v_i(s_i, s_{-i}) > v_i(s_i', s_{-i})$ *for all* $s_i' \in S_i$*,* $s_i' \neq s_i$*, and all* $s_{-i} \in S_{-i}$*.“*[3]

Except for the pure strategies, the player can also express the selection of a strategy by means of a probability distribution between all strategies from a set of pure strategies, which is then referred to as the mixed strategy of the player.

## 2.2 Dominant strategy Equilibrium

The strict dominant strategy defined above can be used as one of the concepts of finding an equilibrium solution, precisely because rational player will never play a strictly dominated strategy.

[1] Tadelis, S., 2013. Game theory: an introduction. Princeton university press, p. 47
[2] Tadelis, S., 2013. Game theory: an introduction. Princeton university press, p. 46
[3] Tadelis, S., 2013. Game theory: an introduction. Princeton university press, p. 61

**Definition 4:** „*The strategy profile $s^D \in S$ is a strčit dominant strategy equilibrium if $s_i^D \in S_i$ is a strčit dominant strategy for all $i \in N$.*"[4]

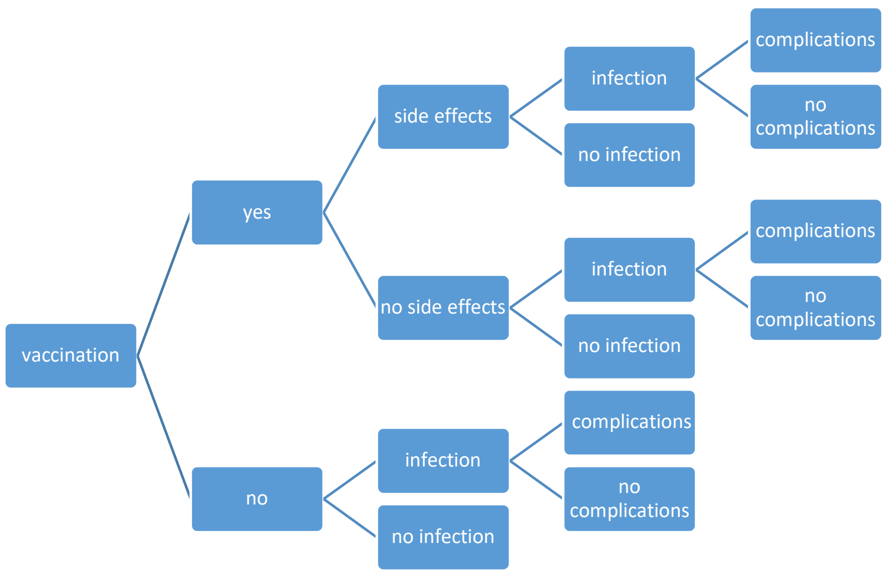
### 2.3 Nash equilibrium in pure strategies

Nash equilibrium is characterized by the interest of the individual player in choosing the optimal strategy for him in best response to the choice of all others' strategies.

**Definition 6:** „*The pure-strategy profile $s^* = (s_1^*, s_2^*, \dots, s_n^*) \in S$ is a Nash equilbrium if $s_i^*$ is the best response to $s_{-i}^*$ for all $i \in N$, that is, $v_i(s_i^*, s_{-i}^*) \geq v_i(s_i', s_{-i}^*)$ for all $s_i' \in S_i$ and all $i \in N$*"[5]

### 2.4 Vaccination as a static game with non-zero sum

The assumption of our analysis through the game was to construct a static mathematical model of the non-zero-sum game. The main reason for the construction of this type of game was the assumption of two rational players, the first being defined as the individual and the second as the whole population. Given this fact, we compare the gains and losses of an individual against the whole population and not against another individual. The gains and losses appear to be different. Other assumption that had to be introduced before the actual construction of the game model is that vaccination resp. non vaccination is decided only by the individual, which we consider to be the only factor decided by the individual himself. The second assumption is that vaccination does not provide complete protection against the disease. The probability of infection in case of vaccination is significantly different from the case of non-vaccination. The Graph 1 shows the game as a sequence of steps made as an initial decision by individual.




Graph 1 Possible situations based on vaccination, Source: Own processing

We can also write down individual's decisions as the Table 1. It captures the possibilities of decision-making and consequently possible processes that may occur with a certain probability, to which we have also assigned a valuation based on our subjective evaluation. The rating range is 1 – 9, where 1 is the worst rating and 9 is the best rating. At the same time, we consider the scenario to get a vaccination, and subsequently to have the side effects of vaccination or non-

[4] Tadelis, S., 2013. Game theory: an introduction. Princeton university press, p. 61

[5] Tadelis, S., 2013. Game theory: an introduction. Princeton university press, p. 80

vaccination and become infected with a disease that will have a complicated course, the second one is the worst-case scenario. On the contrary, we consider the best possible situation to decide not to be vaccinated and to avoid the side effects of the vaccine, not to become infected with the disease against which vaccination is to be avoided, and to avoid the complicated course of the disease.

For each individual, we consider the following factors based on the vaccination decision made:
1. Vaccination → Yes (A) → Side effects (U) → Infection (C) → Complications (L)
2. Vaccination → Yes (A) → Side effects (U) → Infection (C) → No complications (S)
3. Vaccination → Yes (A) → Side effects (U) → No infection (Z)
4. Vaccination → Yes (A) → No side effects (B) → Infection (C) → Complications(L)
5. Vaccination → Yes (A) → No side effects (B) → Infection (C) → No complications (S)
6. Vaccination → Yes (A) → No side effects (B) → No infection (Z)
7. Vaccination → No (N) → Infection (C) → Complications(L)
8. Vaccination → No (N) → Infection (C) → No complications (S)
9. Vaccination → No (N) → No infection (Z)

Table 1 Preferentially arranged situations and values

| Vaccination (A/N) | Side effects | Infection | Complications | Marks | Value |
|---|---|---|---|---|---|
| YES | YES | YES | YES | AUCL | 1 |
| | | | NO | AUCS | 4 |
| | | NO | --------- | AUZ | 5 |
| | NO | YES | YES | ABCL | 3 |
| | | | NO | ABCS | 8 |
| | | NO | --------- | ABY | 6 |
| NO | NO | YES | YES | NCL | 2 |
| | | | NO | NCS | 7 |
| | | NO | --------- | NZ | 9 |

Source: Own processing

Given the records of the game and the given values for individual's options, which can be marked as the strategy of player 1 (Table 1), we can write down the payoff matrix of this player (Table 2).

Table 2 Payoffs based on the awards used for each of his game options

| | $x_{AUCL}$ | $x_{AUCS}$ | $x_{AUZ}$ | $x_{ABCL}$ | $x_{ABCS}$ | $x_{ABZ}$ | $x_{NCL}$ | $x_{NCS}$ | $x_{NZ}$ * |
|---|---|---|---|---|---|---|---|---|---|
| $x_{AUCL}$ | (1, 1) | (1, 4) | (1, 5) | (1, 3) | (1, 8) | (1, 6) | (1, 2) | (1, 7) | (1, **9**) |
| $x_{AUCS}$ | (4, 1) | (4, 4) | (4, 5) | (4, 3) | (4, 8) | (4, 6) | (4, 2) | (4, 7) | (4, **9**) |
| $x_{AUZ}$ | (5, 1) | (5, 4) | (5, 5) | (5, 3) | (5, 8) | (5, 6) | (5, 2) | (5, 7) | (5, **9**) |
| $x_{ABCL}$ | (3, 1) | (3, 4) | (3, 5) | (3, 3) | (3, 8) | (3, 6) | (3, 2) | (3, 7) | (3, **9**) |
| $x_{ABCS}$ | (8, 1) | (8, 4) | (8, 5) | (8, 3) | (8, 8) | (8, 6) | (8, 2) | (8, 7) | (8, **9**) |
| $x_{ABZ}$ | (6, 1) | (6, 4) | (6, 5) | (6, 3) | (6, 8) | (6, 6) | (6, 2) | (6, 7) | (6, **9**) |
| $x_{NCL}$ | (2, 1) | (2, 4) | (2, 5) | (2, 3) | (2, 8) | (2, 6) | (2, 2) | (2, 7) | (2, **9**) |
| $x_{NCS}$ | (7, 1) | (7, 4) | (7, 5) | (7, 3) | (7, 8) | (7, 6) | (7, 2) | (7, 7) | (7, **9**) |
| $x_{NZ}$ * | (**9**, 1) | (**9**, 4) | (**9**, 5) | (**9**, 3) | (**9**, 8) | (**9**, 6) | (**9**, 2) | (**9**, 7) | (**9**, **9**)* |

Source: Own processing

Table 2 can be characterized as the notation of the analytical game by the non-zero-sum game with payoffs determined based on the constructed preferential order shown in Table 1. Based on Definition 4 and Definition 5 we see that the equilibrium strategy of the game is, as expected, strategy and choice of the option not to be vaccinated and without the disease.

### 2.5 Nash equilibrium matrix by 2x2

We modified the static game with a non-zero-sum, which we formulated in the previous part of the article, by grouping all the options for individual decisions made by individuals into a game in the normal form of 2x2 dimensions. We have implemented this grouping in a way that in case the player chooses the strategy to be vaccinated, the minimum values from all payoffs listed in Table 2 is min {1,4,5,3,8,6} = 1 and in case of opposite strategy the min {2,7, 9} = 2, what will provide the player with the highest win.

Table 3 Formulation of game dimensions 2x2

| | $Vaccination$ | $No - vaccination$ |
|---|---|---|
| $Vaccination$ | (1, 1) | (1,2) |
| $Non - vaccination$ | (2, 1) | **(2, 2)** |

Source: Own processing

Based on Definition 5 for Nash equilibrium in pure strategies, we can see that there is the possibility (non-vaccination, non-vaccination) with payoffs (2.2). This Nash Equilibrium is not Pareto effective, because both players may be better off if they choose a strategy (vaccination, vaccination) with payoffs (1, 1). However, such a result can be considered as an interaction between these players since the individual approach of the individual prevents the achievement of the Pareto effective solution. In view of the rationality of this individual, i.e., the first player, and the individual decision made by this player, the possibility of non-vaccination will always appear to be a more advantageous decision, regardless of what strategy the second player chooses. This type of game is a typical example of game theory, namely the Prisoner Dilemma.

## 3 DISCUSSION

The aim of this paper is to explore the individual's behavior against the population regarding a health issue, namely the vaccination decision, from the game theory's point of view. A model, in which the only individual's decision is whether to vaccinate or not, was formulated and the payment matrix with a subjective scale was constructed with the follow-up situations occurring with a given probability. The game was analyzed using the dominance strategy and Nash equilibrium, and the optimal strategy for the individual is: to not vaccinate, to not vaccinate. In the next part of the paper, the individual strategies were clustered, and the game was restricted to the sole option decided by the individual, which is a 2by2 game. The Nash equilibrium gives the same strategy (to not-vaccinate, to not-vaccinate). The game formulated in this way is a typical example of the Prisoner's dilemma wherein the equilibrium is not the optimal decision, since in the case of the individual's and population's strategy (vaccinate, vaccinate) both players would gain. In this case, the cooperative strategy between individuals and the population seems to be very unstable. Similar analyses and their extensions could be helpful in guiding many government actions on compulsory vaccination.

# 4 CONCLUSION

A possible extension of the thus formulated game, which we want to address subsequently is the enumeration of given payments based on costs associated with vaccination and non-vaccination of an individual as well as vaccination and non-vaccination of a population. Depending on those costs, in case that individual decides to vaccinate while the population opts to not vaccinate, and the costs of side effects were lower than the costs associated with non-vaccination and infection, the game would not represent the prisoner's dilemma. In this case, there would be two non-dominated equilibrium solutions and the best individual ‘s decision would be a mixed strategy. A prerequisite for finding a mixed strategy would be to define appropriate probabilities for the individual strategies as the best response to the opposite player's strategy.

**Acknowledgements:** This work was supported by a grant APVV-17-0551 “Balancing out the imbalances: redefining the view on macroeconomic imbalances under the european governance framework” and VEGA 1/0628/20 “The relationship between international investment position and economic growth – potential indicator of country ‘s external imbalances”

**References**

[1] Bai, F. (2016). Uniqueness of Nash equilibrium in vaccination games. *Journal of biological dynamics*, *10*(1), pp. 395-415.
[2] Bauch, C. T. and Earn, D. J. D. (2004). Vaccination and the theory of games. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), pp. 13391-13394.
[3] Bauch, C. T., Galvani, A. P. and Earn, D. J. D. (2003). Group interest versus self-interest in smallpox vaccination policy. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(18), pp. 10564-10567.
[4] Chapman, G. B., Li, M., Vietri, J., Ibuka, Y., Thomas, D., Yoon, H. and Galvani, A. P., (2012). Using game theory to examine incentives in influenza vaccination behavior. *Psychological science*, *23*(9), pp. 1008-1015.
[5] Choi, W. and Shim, E. (2020). Optimal strategies for vaccination and social distancing in a game-theoretic epidemiologic model. *Journal of theoretical biology, 505*, p.110422.
[6] Georgiou, H. V. (2020). Social responsibility is the crucial factor in adopting early vaccination plans. medRxiv, pp. 4
[7] Kabir, K.M., Risa, T. and Tanimoto, J. (2021). Prosocial behavior of wearing a mask during an epidemic: an evolutionary explanation. *Scientific reports, 11*(1), pp.1-14.
[8] Piraveenan, M., Sawleshwarkar, S., Walsh, M., Zablotska, I., Bhattacharyya, S., Farooqui, H.H., Bhatnagar, T., Karan, A., Murhekar, M., Zodpey, S. and Rao, K.M. (2021). Optimal governance and implementation of vaccination programmes to contain the COVID-19 pandemic. *Royal society open science*, *8*(6), pp.210429.
[9] Reluga, T. C., Bauch, C. T. and Galvani, A. P. (2006). Evolving public perceptions and stability in vaccine uptake. *Mathematical biosciences*, *204*(2), pp. 185-198.
[10] Tadelis, S. (2013). Game theory: an introduction. *Princeton university press,* pp.396

**Author's address**

Ing. Miroslava Jánošová, Phd., Ing. Veronika Mitkova, PhD.
Comenius University in Bratislava, Faculty of Social and Economic Sciences, Institute of Economics
Mlynské luhy 4, 821 05 Bratislava
Slovak Republic
email: miroslava.janosova@fses.uniba.sk, veronika.mitkova@fses.uniba.sk

# COMPARISON OF TWO SELECTED APPROACHES TO THE TRANSPORT INFRASTRUCTURE INVESTMENT PROJECTS PORTFOLIO SELECTION FOR IMPLEMENTATION

*Karel Ječmen, Daniel Pilát, Dušan Teichmann, Olga Mertlová*
*Czech Technical University in Prague*

**Abstract**
Financial needs of the prepared transport infrastructure investment projects are generally very high and exceed the available budgets. Therefore, it is necessary to set the selection process to find a project portfolio to be proposed for realization. Due to the significant socio-economic impact of transport infrastructure investments, the selection of projects for realization is a crucial decision-making step at the level of the central state authorities.
The article compares two approaches to mathematical programming, which can contribute to the optimization of the portfolio of transport infrastructure investment projects to be financed. Optimization of the selection aims at the maximization of the benefits of the projects and the efficient use of allocated funds. The paper focuses on the differences in application of the use of the linear and goal programming models and compares the outputs of both optimization approaches.

***Keywords:*** *Linear programming, Goal programming, Transport infrastructure financing*

***JEL Classification:*** C610
***AMS Classification:*** 90B06, 90B50, 90C29

## 1 INTRODUCTION – MOTIVATION TO SOLVE THE PROBLEM AND ANALYSIS OF THE CURRENT STATE OF KNOWLEDGE

Transport infrastructure investments present a significant part of the society growth. The scope and quality of transport infrastructure has a major impact on the development of society and the economy (Mačiulis, Vasiliauskas, Jakubauskas, 2009) and transport infrastructure investments also have an impact on the course of GDP indicators in the long run (Kalantzis, Arnoldus, Brons, 2015).

The value of investment projects under preparation, which can be included in the implementation, is usually higher in terms of investment costs than the volume of available investment funds. The decision-making bodies are therefore forced to select only some projects from the set of projects being prepared for implementation. In the case of investment programs financed from EU resources (eg. the Operational Program Transport), indicators are considered that monitor the degree of contribution of individual projects to meet the program objectives and thus achieve the required societal benefits.

The above-mentioned decision-making task solves the maximization of the fulfilment of indicators within the limited amount of available resources, which can be solved as an optimization task of linear programming (LP) (Ječmen, 2021). However, the solution of the LP problem is limited by the fact that the minimum values of all indicators must be saturated. If the minimum values of all indicators cannot be met, no feasible solution can be found.
In real life application, however, there are possible situations where no combination of selected projects can meet all the indicators to the required minimum level, and yet we require to identify

a solution with maximum possible degree of indicator values fulfilment. In such a situation, when finding a solution using LP, it is possible to achieve an acceptable solution only by adjusting the input minimum values of indicators. However, this can affect finding a global optimal solution or also make the decision-making non-transparent, which is not convenient when managing public funds.

An alternative to LP to solve the mentioned problem seems to be the use of goal programming (GP). The purpose of GP (Chang, 2007) is to minimize the deviation from the required target values of the monitored parameters. The article (Teichmann, Dorda, 2015) compares the approach to solving an analogous problem using LP and GP, showing that both approaches offer comparable solutions. GP allows to solve the task even in case of impossibility to reach the minimum values of indicators.

The aim of this paper is to compare the solution of the task of selecting the optimal investment projects portfolio in a situation where it is not possible to meet the required saturation of indicators. The paper compares approaches based on GP and LP with artificial adjustment of minimum indicator values.

## 2 PROBLEM FORMULATION AND ITS MATHEMATICAL MODEL

This chapter formulates the problem and theoretically introduces both approaches to solving the optimization problem, the results of which will be compared in Chapter 3.

### 2.1 Problem formulation

Consider the set $I$ of isolated investment projects, intended for implementation in a predefined period and the set of indicators for a given period $K$. The indicators can be interpreted as a measurable form of societal benefit. Each of the projects $i \in I$ has defined investment costs $n_i$, which need to be spent on its implementation, and its contribution to the fulfilment of the indicator $k \in K$ called $a_{ik}$. Each indicator $k \in K$ has defined weight $w_k$, which represents its importance or penalty rate for each unit of deviation from target value. In addition, the minimum values of individual $b_k$ indicators that need to be met are determined for a given investment period. The task is to create an optimal portfolio of projects for implementation from the set of projects $I$. Optimization criteria will be different in each type of approach.

Sets common to both models

| | |
|---|---|
| $I$ | set of projects to be implemented |
| $K$ | set of indicators |

Input variables common to both models

| | |
|---|---|
| $n_i$ | total investment costs of the project $i \in I$ |
| $N$ | financial resources available |
| $b_k$ | requirements for minimum values of indicators fulfilment |
| $a_{ik}$ | project contribution $i \in I$ to fulfil indicator $k \in K$ |
| $w_k$ | weight of indicator $k \in K$ |

### 2.2 Design of optimisation model 1 – LP-based approach

The optimization criterion in this approach is the cumulative value of all indicators, and the goal of optimization is to maximize it.

Variables used in the model

$x_i$ binary variable representing the financing of the $i \in I$ project; if $x_i = 1$ applies after the optimization calculation, the project $i \in I$ will be financed, if $x_i = 0$ applies after the optimization calculation, the project $i \in I$ will not be financed.

The mathematical model can be written in this form

$$f(x) = \sum_{i\in I}\sum_{k\in K} a_{ik}x_i w_k \to max \tag{1}$$

Subject to

$$\sum_{i\in I} x_i n_i \le N \tag{2}$$

$$\sum_{i\in I} a_{ik}x_i \ge b_k \qquad \text{for } k\in K \tag{3}$$

$$x_i \in \{0,1\} \qquad \text{for } i\in I \tag{4}$$

The objective function (1) represents the optimization criterion – the weighted cumulative fulfilment value of all indicators. The constraint (2) will ensure that the available budget is not exceeded. The group of constraints (3) will ensure that the minimum required values of indicators are met. The group of constraints (4) defines the domains of variables used in the model.

### 2.3 Design of optimisation model 2 – GP-based approach

The optimization calculation will take place in two phases.

Variables used in the model

- $x_i$ binary variable representing the financing of the $i\in I$ project; if $x_i = 1$ applies after the optimization calculation, the project $i\in I$ will be financed, if $x_i = 0$ applies after the optimization calculation, the project $i\in I$ will not be financed,
- $d_k^+$ variable presenting a positive deviation – the value of exceeding the minimum value of the indicator $k\in K$,
- $d_k^-$ variable presenting the negative deviation – the value of non-compliance with the minimum value of the $k\in K$ indicator.

In the first phase of the calculation, the cumulative negative deviation of the achieved values of indicators from the minimum target values of indicators will be minimized. The mathematical model is designed as follows.

$$f(x, d^+, d^-) = \sum_{k\in K} d_k^- w_k \to min \tag{5}$$

Subject to

$$\sum_{i\in I} x_i n_i \le N \tag{6}$$

$$\sum_{i\in I} a_{ik}x_i + d_k^- - d_k^+ = b_k \qquad \text{for } k\in K \tag{7}$$

$$x_i \in \{0,1\} \qquad \text{for } i\in I \tag{8}$$

$$d_k^- \ge 0 \qquad \text{for } k\in K \tag{9}$$

$$d_k^+ \ge 0 \qquad \text{for } k\in K \tag{10}$$

In the first phase of the calculation, the objective function (5) represents the optimization criterion – the weighted cumulative negative deviation. The constraint (6) has the same meaning as the constraint (2). The group of constraints (7) creates binding relations between the variables and the optimization criterion. The group of constraints (8), (9) and (10) defines the domains for variables used in the model.

In the second phase of the calculation, the goal will be to maximize the cumulative positive deviation from the minimum target values of the indicators. The second phase occurs only if $N - \sum_{i \in I} x_i n_i \geq \min\{n_i;\ x_i = 0\}$. For the purposes of the second phase, a constant $Y$ will be introduced, which corresponds to the resulting value of the objective function (5) after the first phase of the optimization calculation.

The mathematical model for the second stage of the calculation has the form

$$f(x, d^+, d^-) = \sum_{k \in K} d_k^+ \, w_k \rightarrow max \tag{11}$$

Subject to (6)-(10) and (12)

$$\sum_{k \in K} d_k^- \leq Y \tag{12}$$

The meanings of the group of constraints (6)-(10) were explained in the model for the first phase of the optimization calculation, the constraint (12) ensures compliance with the result from the first phase of the optimization calculation.

## 3 COMPUTING EXPERIMENTS

In this chapter, the results of both approaches are presented and compared. The experiments were performed on a personal computer equipped with Intel Core i7-8750H CPU Duo E6850 with parameters 2.20 GHz and 16 GB RAM.

Data according to (Ječmen, 2021) were used for the computational experiment. These are 20 isolated investment projects that are intended to be implemented within a certain investment program period. The identification numbers of these projects (1-20 in Table 1) and the parameters $a_{ik}$, $b_k$, $n_i$ and $N$ serve as input data for the calculation. The values in the rows of the Table 1 marked as GP and LP restriction express in the case of columns representing indicators $a_{ik}$ values of $b_k$, which are the requirements of the minimum values of indicators to fulfil, and in the case of the column $n_i$ total budget $N$. The reduced values of $b_k$ for the LP-based approach were found by gradually decreasing the values of the indicators until the moment of the first feasible solution, the values of $a_{ik}$ are identical for both approaches. An overview of input data, including restrictions, is given in Table 1. For simplification purposes is the parameter $w_k = 1$ for all $k \in K$.

Table 1 Overview of the input data

| Project ID | $a_{ik}$ | | | | | | | $n_i$ |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | (mil. CZK) |
| 1 | 3,1 | 0 | 0 | 0 | 0 | 3,1 | 0 | 1 275 |
| 2 | 0 | 6,13 | 0 | 6,13 | 0 | 0 | 0 | 880 |
| 3 | 10,45 | 0 | 0 | 0 | 0 | 10,45 | 0 | 3 745 |
| 4 | 8,577 | 0 | 0 | 0 | 0 | 0 | 0 | 2 507,5 |
| 5 | 5,95 | 0 | 0 | 0 | 0 | 5,95 | 0 | 2 500 |
| 6 | 0 | 0 | 5,096 | 0 | 3 | 0 | 0 | 2 500 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 7 | 0 | 0 | 0 | 0,42 | 0 | 0,42 | 0 | 160 |
| 8 | 16,346 | 0 | 0 | 0 | 6 | 0 | 0 | 5 000 |
| 9 | 0 | 0 | 7,1 | 0 | 0 | 7,1 | 0 | 840 |
| 10 | 0 | 0 | 0 | 2,879 | 0 | 2,879 | 0 | 1 030 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 200 |
| 12 | 0 | 0 | 0 | 0 | 10 | 10 | 0 | 650 |
| 13 | 0 | 0 | 9,51 | 0 | 0 | 9,51 | 0 | 1 500 |
| 14 | 0 | 0 | 1 | 7,527 | 0 | 0 | 0 | 1 400 |
| 15 | 0 | 2,766 | 0 | 0 | 2,766 | 0 | 0 | 1 200 |
| 16 | 0 | 0 | 10,3 | 0 | 0 | 10,3 | 0 | 2 000 |
| 17 | 7,537 | 0 | 0 | 0 | 0 | 7,537 | 0 | 2 000 |
| 18 | 0 | 0 | 0 | 0 | 5 | 5 | 0 | 350 |
| 19 | 0 | 0 | 6,02 | 2 | 0 | 0 | 0 | 3 150 |
| 20 | 17,3 | 0 | 0 | 0 | 3 | 0 | 0 | 3 900 |
| Total | 69,26 | 8,896 | 39,026 | 18,956 | 29,766 | 72,246 | 5 | 36 787,5 |
| GP restriction | 30 | 16 | 20 | 24 | 16 | 20 | 10 | 12 500 |
| LP restriction | 17 | 6 | 20 | 16 | 16 | 20 | 5 | 12 500 |

As can be seen in Table 2, the LP-based approach proposes to fund 11 projects. The GP-based approach proposes to fund 10 projects. The selection of the project for financing is marked with the symbol X. The resulting values of invested funds are also included.

Table 2 - Overview of implemented projects after LP and GP calculation and invested funds (stated in millions of CZK))

| Projects realization | | | | | | | | | | | | | | | | | | | | | Spent funds | Unspent funds |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Project | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | | |
| LP | | X | | X | | | X | | X | | X | X | X | X | | X | X | X | | | 12 487,5 | 12,5 |
| GP | | X | | X | | | X | | X | | X | X | X | X | | | | X | | X | 12 387,5 | 112,5 |

The resulting indicator values and penalty values indicating the occurrence of negative deviations from the minimum indicator values are shown in Table 3. (The fulfilment of indicators is always compared to the required target value, despite the reduction of the LP limit.)

Table 3 - Final fulfilment of indicators (output values were rounded)

| Indicator fulfilment | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Indicator | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
| Target value | 30 | 16 | 20 | 24 | 16 | 20 | 10 | 136 |
| Value achieved by LP | 16 | 6,1 | 28 | 14 | 15 | 50 | 5 | 134,1 |
| Value achieved by GP | 26 | 6,1 | 18 | 14 | 18 | 32 | 5 | 119,1 |
| LP penalty | 14 | 9,9 | 0 | 10 | 1 | 0 | 5 | 39,9 |
| GP penalty | 4 | 9,9 | 2 | 10 | 0 | 0 | 5 | 30,9 |

Table 3 represents target values (row 2), achieved values of benefits (row 3 for LP, row 4 for GP) and their deviation from target value (row 5 for LP, row 6 for GP) for each indicator and cumulatively.

The results of both approaches offer significantly different solutions. Although the LP approach (1)-(4) offers a higher cumulative fulfilment of the monitored indicators, the GP approach (5)-(12) brings a significantly lower negative deviation from the required values of the indicators as well as a lower volume of invested funds. The LP approach will therefore theoretically bring a relatively higher social benefit, however, it does not address the need for highest possible fulfilment of all indicators.

## 4 CONCLUSION

The presented paper deals with the use of two approaches to mathematical programming to optimize the investment portfolio of transport infrastructure projects. The aim of the article was to compare approaches based on LP and GP and to assess the differences in the results achieved when applying them to the same task. The achieved values of indicators, which represent the degree of projects contribution to the investment program objectives fulfilment and are assessed by their control authorities, were monitored.

The fundamental difference in the application of both approaches is evident from the course of the experimental phase. The approach based on LP (1)-(4) will not find a feasible solution if all the minimum target values of indicators cannot be met. Therefore, in order to use the LP-based approach, it was necessary to reduce the minimum target values of the indicators so that they were achievable and to allow a feasible solution to be found. The GP (5)-(10) approach is not burdened by this problem. From the point of view of practical use, this is a fundamental reason for choosing GP if target values of one or more indicators cannot be met considering all available resources.

The results achieved by both approaches are also different. It is clear from the calculated values of the indicators that the results of the approach based on LP (1)-(4) reached a better overall cumulative value. The results of the GP-based approach (5)-(10) better met the requirements of the investment program and its main benefit is in minimizing the unfulfilled required values of indicators. In future research, we will look at which option also considers other socio-economic needs and is more economically advantageous and develop methodology for determining the weight of indicators.

**Acknowledgements:** This work was supported by project SGS22/126/OHK2/2T/16 Design of Computational Methods for Optimizing Investment Projects Portfolio in Transport Infrastructure.

**References**

[1] Chang, Ching-Ter. (2007). Multi-choice goal programming, *Omega*, 35(4), pp. 389-396, ISSN 0305-0483. doi: 10.1016/j.omega.2005.07.009.
[2] Ječmen, K. (2021). *Optimalizační metody při investičním rozhodování na dopravní infrastruktuře*. Master’s Thesis. Czech Technical University in Prague.
[3] Kalantzis, F., Arnoldus, P. and Brons, M., European Commission, Directorate-General for Economic and Financial Affairs (2015). *Infrastructure in the EU: developments and impact on growth*. Publications Office. doi:10.2765/85301.
[4] Mačiulis, A., Vasiliauskas, A. V. and Jakubauskas, G. (2009). The impact of transport on the competitiveness of national economy, *Transport*, 24(2), pp. 93-99. doi: 10.3846/1648-4142.2009.24.93-99.
[5] Teichmann, D. and Dorda, M. (2015). Comparison of two selected methods in evaluating of investments in transport infrastructure, *Finance and Performance of Firms in Science, Education and Practice 2015*, pp. 1524-1536.

**Author’s address**

Ing. Karel Ječmen, Ing. Daniel Pilát, doc. Ing. Dušan Teichmann, Ph.D., Ing. Olga Mertlová, Ph.D.
Czech Technical University in Prague, Faculty of Transport Sciences, Department of Logistics and Management of Transport
Horská 3, 128 03 Praha 2
Czech Republic
email: jecmekar@fd.cvut.cz, pilatdan@fd.cvut.cz, teichdus@fd.cvut.cz, pokorol1@fd.cvut.cz

# DOES THE EFFECT OF ACCESSION TO THE EUROPEAN UNION VARY ACROSS COUNTRIES? CROSS COUNTRY ANALYSIS OF SLOVAKIA AND THE CZECH REPUBLIC BASED ON THE STRUCTURAL GRAVITY MODEL

***Brian König,***
***University of Economics in Bratislava***
***Economic Institute of Slovak Academy of sciences***

**Abstract**
The main objective of presented research was to use the examples of Slovakia and the Czech Republic to enquire whether the average effect of accession to the EU on the bilateral trade between the EU countries differs across various member states. Based upon the findings, we may assume that the effect of the EU membership on the bilateral trade flows of Slovakia and the Czech Republic is rather different with a much stronger impact on the latter.

***Keywords:*** *Structural gravity model, EU membership, Slovakia, Czech Republic*

***JEL Classification:*** C13, F14
***AMS Classification:*** 91B60

## 1 INTRODUCTION

It has been over 60 years since European Economic Community (EEC) was created by six founding countries in 1957. Since then, 22 countries joined EEC later formed as today known European Union (1993). Among crucial benefits of this integration belong free movement of goods, services and people respectively. The question is to what extent individual country is able to transform these advantages into its own benefit. In other words, is the effect of these benefits similar among accession countries or for some reasons it differentiates and how big these changes are. Such information might be interesting especially for potential candidates to accession (e.g. Albania, Republic of North Macedonia, Montenegro, Serbia and Turkey). In the following research we want to address the effect of EU accession on bilateral trade. The effect of the EU membership can be analyzed in many different ways, but one of the most common is a gravity model which is regarded as a key tool designed for the international trade analysis. The gravity model of international trade is historically connected with the personality of the Dutch economist Jan Tinbergen who introduced this concept in 1962. The original gravity equation (Tinbergen, 1962) assumed the bilateral trade between two countries depends on their economic sizes and distance. Later, gravity models were heavily criticized for being mere economic tools without a sufficient theoretical background. The subsequent research showed, however, that gravity models can be supported by a number of international trade theories (e.g. Anderson, 1979; Bergstrand, 1985; Trefler, 1995; McCallum, 1995; Deardorff, 1998; Eaton and Kortum, 2002; Anderson and van Wincoop, 2003). The article by Anderson and van Wincoop (2003) was regarded as the breakthrough for gravity models as they proposed a structural gravity model based on the pure microeconomic theory. Their model was considered as a base line for modelling international trade. Anderson and van Wincoop pointed to a key fact related to a gravity model construction. They introduced trade costs relative to the costs of other countries (i.e. multilateral resistances - MR), which are necessary for a well-specified structural model. The elements representing multilateral resistances are theoretical variables which cannot be directly observed. In their original

article, Anderson and van Wincoop (2003), in order to take account of the multilateral resistances, programmed their own iterative non-linear least squares method. The elements of multilateral resistances are often approximated by fixed-effects variables within the dimensions - importer - time and exporter - time which were introduced by Baier and Bergstrand (2007). Gravity models have long been criticized for the fact that several trade cost variables are not considered as exogenous variables. The authors Baier and Bergstrand also suggested solving the endogeneity problem of some bilateral trade cost variables related to the different kinds of economic integration (e.g. free trade agreements, regional free trade agreements, custom union, etc.) by introducing the fixed-effect variables within the dimension exporter - importer. Another issue often addressed during the estimation process of gravity models is the presence of the zero values of bilateral trade flows. Unless the distribution of zeros is random and zero values are excluded from the sample, a sample selection bias occurs due to non-random selection of observations. Various authors in their research dealt with the problems of the existence of zeros and inconsistency of gravity models estimates in case the least squares method (OLS) is used. These attempts have led to new specific methods for estimation of gravity models parameters being proposed (e.g. Silva and Tenreyro, 2006; Helpman, Melitz and Rubinstein, 2008; Egger, Larch, Staub and Winkelmann, 2011; Egger and Nigai, 2015; Larch, Wanner, Yotov and Zylkin, 2019; Zylkin, 2018; Oberhofer and Pfaffermayr, 2021; Correia, Guimarães and Zylkin, 2019).
Most of the studies that have been published have dealt with the effect of EU membership in general. The estimation of the effect of the EU membership on bilateral trade can be observed in the papers published by: Hufbauer and Schott (2009), Baier, Bergstrand, Egger and McLaughlin (2008), Eicher, Henn and Papageorgiou (2012), Fournier et al. (2015), Grančay, Grančay and Drutarovská (2015), Gudgin, Coutts, Gibson and Buchanan (2017).
The main objective of our research is to use the example of Slovakia (SK) and the Czech Republic (CZ) to point out that the average effect of entry into the EU on bilateral trade across the EU Member States can be different. Our intention was to verify this hypothesis by applying the gravity model and to study whether the effect of entry on the SK / CZ bilateral trade with the EU countries differs from the European average. The countries were selected because of their similar economic development and close economic and cultural identity as they used to be part of the same federation. The article is structured as follows: the Chapter II deals with the applied methodology and empiric models that are used in the article. The third chapter describes the data that have been applied. The last part presents and discusses the results of the estimated models.

# 2 METHODOLOGY

For the purpose of analyzing the effect of EU membership on international trade, the gravity model is used. The former gravity equation (Tinbergen, 1962) supposed that the size of bilateral trade of the two countries is proportionate to their respective economic size and distance. The gravity models were further developed and the subsequent research showed, that gravity models can be supported by a number of theories of international trade e.g. Anderson(1979), Bergstrand (1985), Eaton and Kortum (2002), Anderson and van Wincoop (2003) Chaney (2008) and others. Anderson and van Wincoop (2003) proposed a structural gravity model based on the clearly theoretical baseline which is considered as a starting point when modelling the international trade. The final form of the structural gravity model as designed by Anderson and van Wincoop can be described as follows:

$$x_{ij} = \frac{y_i y_j}{y^w}\left(\frac{t_{ij}}{\Pi_i P_j}\right)^{1-\sigma}, \tag{1}$$

$$\Pi_i^{1-\sigma} = \sum_j \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} \frac{y_j}{y^w}, \tag{2}$$

$$\mathrm{P}_j^{1-\sigma} = \sum_i \left(\frac{t_{ij}}{\Pi_i}\right)^{1-\sigma} \frac{y_i}{y^w}, \tag{3}$$

where $x_{ij}$ are bilateral trade flows from the country i to the country j; $y^w$ refers to the world income (or production); $y_i$ is the domestic production of the country i; $y_j$ captures total expenditures of the country j; $t_{ij}$ represents bilateral trade costs between the countries i and j often approximated in the literature by variables such as: distance of countries, free trade agreements on the bilateral or regional levels (FTA, RTA), customs, single currency, etc; $\sigma$ captures the elasticity of substitution of goods between particular countries; $P_j$ refers to the inward multilateral resistance, which points out that the volume of import from country i to country j is a function of trade costs across all possible importers; $\Pi_i$ represents outward multilateral resistance stating that export between the countries i and j depends on the trade costs across all possible exporters. Since the original gravity model did not entail these MR terms, the estimation of parameters was rather biased as they excluded relevant variables correlating with variables representing trade costs. The patterns representing multilateral resistance are theoretical variables which cannot be directly observed. One of the ways how to measure their influence in the model is to use so-called remoteness indexes catching the function of countries' distance and weighted GDP (Wei, 1996). Nevertheless, this approach was often criticised because the remoteness indexes contained only a small part of the theory including original theoretical multilateral resistances (Head and Mayer, 2014). The theoretically consistent multilateral resistance terms can be obtained upon applying fixed effects within the dimensions importer - time and exporter - time. One of the first authors to have applied these pairs of fixed effects in the gravity models was (Harrigan, 1996). In order to take into account the issue of endogeneity of variables within international integration (free trade agreements, EU membership), the model will also have to entail fixed effects varying in dimensions of importer - exporter (Baier and Bergstrand, 2007). Baseline fixed effects model has a form as follows:

$$x_{ij,t} = \exp(\beta_1 fta_{ij,t} + \beta_2 eu_{od_{ij,t}} + \phi_{ij} + \gamma_{i,t} + \delta_{j,t})\varepsilon_{ij,t}. \tag{4}$$

where $t$ is the time index; $x_{ij,t}$ refers to the nominal value of import from the country *j* to the country *i;* $eu_{od_{ij,t}}$ represents a dummy variable equal to one if importer and exporter are EU members at the same time and zero otherwise; $fta_{ij,t}$ is a dummy variable expressing the existence of trade agreement between the importer and the exporter, which is equal to one if there exist *fta*, between trading countries and zero otherwise; $\phi_{ij}$ refers to fixed effects for the pairs importer - exporter; $\gamma_{i,t}$ and $\delta_{j,t}$ are fixed effects varying within the dimensions importer - time and exporter - time; $\varepsilon_{ij,t}$ is an error term. The specified model with three types of fixed effects is estimated in the multiplicative form through the Poisson pseudo-maximum-likelihood (PPML estimator). The model will be regarded as a benchmark estimator of the EU membership effect on international trade.

# 3 DATA

The bilateral import of goods from the DOTS database (Direction of Trade Statistics) of the International Monetary Fund will be used as a dependent variable. The database covers data about bilateral trade for 184 countries worldwide from 1948 until now with a rough two-years' delay. The data are published in the US dollars. Import was selected as a dependent variable due to the fact that various authors (e.g. Yotov et al., 2016) pointed out that import data are much more reliable than export data because the country import balance tends to be more carefully monitored in contrast to export (mainly due to customs). Besides that, the DOTS database contains more data about import than export and therefore, import in US dollars (USD) will be used as a dependent variable. For the purpose of the analysis, we used the data from 1950 - 2015 for 138 largest business partners of the EU in 2015 including all 28 EU Member States. The data about free trade agreements and customs unions were taken from the Egger and Larch (2008) database. The upper interval limit (2015) was selected because the information about the EU membership is used from the databases of the French Centre for Research and Expertise on the World Economy CEPII (the databases Gravity), providing various types of databased (among others, the databases designed for gravity models). The database "Gravity" contains information from 1948 to 2015 providing approx. 70 variables for 224 countries. As the adaptation of trade flows among countries to the change in the trade policy or trade costs can take a little time, various authors (e.g. Baier and Bergstrand, 2007; Bergstrand, Larch and Yotov, 2015) recommend the use of time intervals (the intervals of three, four or five years) instead of the consecutive years to estimate the gravity model. In order to use the highest number of observations when estimating the gravity model, we have decided to apply three-years' long intervals.

# 4 RESULTS

In order to verify our hypotheses, the estimation of the EU membership on international trade of the selected countries (SK, CZ) was carried out in two steps. In the first step, we estimated average effect of the EU membership on bilateral trade obtained from gravity model.

Estimate of the Model (1) with approximation of MR trough the fixed effects points out on the fact that the effect of the EU membership on bilateral trade to be measured by import is statistically significant. It value ranges around 0.426 (53.1%).[1] The parameter of the variable eu_od describes trade flow percentage change of the countries entering the EU with other member states. As this effect is average for all EU countries, we may expect that the effects across particular countries are likely to vary. As we are interested in the possible effect of the EU membership on international trade of Slovakia and the Czech Republic, the second step will use alternative formulations complemented by the control of effects on these two countries. As described by Campos and Timini (2019), the alternative formulation will use - instead of eu_od - the variable that will only replace the observations for Slovakia or the Czech Republic for zeros $eu_{NSK}, eu_{NCZ}$.[2] Simultaneously, the dummy variable will be added with the value 1 only if the exporter or importer is Slovakia (or the Czech Republic) and at the same time, both trading countries are EU members ($eu_{SK}, eu_{CZ}$). In this way, the individual effect of the entry of Slovakia and the Czech Republic on their international trade with other EU countries can be controlled. The alternative specification will therefore satisfy:

[1] Average effect is calculated as follows: exp(0.426)-1=53.1%

[2] These variables are equal to one if importer and exporter are EU members and none of them is Slovakia or Czech Republic.

$$x_{ij,t} = \exp(\beta_1 fta_{ij,t} + \beta_2 eu_{nsk_{ij,t}} + \beta_3 eu_{sk_{ij,t}} + \phi_{ij} + \gamma_{i,t} + \delta_{j,t})\varepsilon_{ij,t}, \quad (5)$$

where the index *sk* can be replaced by *cz*. It is worth saying that by this approach, we gain the effect of the EU membership on international trade of Slovakia or the Czech Republic as the average value of trade with other EU countries. The results of particular estimations (5) are in the Table below.

Table 1 – estimation of the average effect of the EU membership for Slovakia and the Czech Republic.

| | (1)<br>cons | (2)<br>sk | (3)<br>cz |
|---|---|---|---|
| eu_od | 0.4259***<br>(0.0243) | | |
| eu_nsk | | 0.4301***<br>(0.0244) | |
| eu_sk | | 0.2431**<br>(0.1139) | |
| eu_ncz | | | 0.4339***<br>(0.0246) |
| eu_cz | | | 0.2938***<br>(0.0606) |
| *N* | 193736 | 193736 | 193736 |

*Standard errors in parentheses * p < 0.10, ** p < .05, *** p < .01*
*Model (1) represents the theoretically consistent model (4). The model (2)-(3) represent the estimation of an average effect of the EU membership on international trade of SK and CZ according to the relation (5). All models stated in the table are estimated by means of high dimensional-fix effects ppml (hdfeppml) method. In all cases, fixed effects varying in dimension importer - time, exporter - time and importer – exporter are used. Simultaneously, all models apply the explanatory variable fta which is statistically insignificant and therefore its effect is not shown in the Table 1.*

The results in the Table 1 show that the effect of the entry of Slovakia and the Czech Republic into EU on international trade with other EU Member States is considerably lower than the overall average effect of all EU Member States. In light of the aforementioned results, it can be observed that the average effect of the EU membership on Slovakia is lower than on the Czech Republic and therefore, international trade of the Czech Republic was a bit more profitable in comparison with the Slovak one. The results suggest that the effect of EU membership on bilateral trade is country-specific. Such information might be interesting especially for potential candidates to EU accession. Further research could analyzed

# 5 CONCLUSION

The main objective of this study was to verify the hypothesis whether the effect of EU membership on trade flow across the EU countries is similar or not. The hypothesis was tested in Slovakia and the Czech Republic which are proportionate in their economic development and possess similar cultural identity. Based upon the findings, we may observe the difference in the effect of the EU membership on the foreign flow measured by import between Slovakia and the Czech Republic, with the latter showing a much stronger impact. Future research could focus on analysing the factors affecting the different impact of EU membership on bilateral trade.

**Acknowledgement**
This contribution was supported by VEGA 1/0193/20 and APVV-17-0551.

**References**

[1] Anderson, J. E. (1979). A Theoretical Foundation for the Gravity Equation. *American Economic Review*, 69(1), pp. 106-116.
[2] Anderson, J. E. and van Wincoop, E. (2003). Gravity With Gravitas: a Solution to the Border Puzzle. *American Economic Review*, 93(1), pp. 170-192.
[3] Baier, S. L. and Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade, *Journal of International Economics*, 71(1), pp. 72-95.
[4] Baier, S. L., Bergstrand, J. H., Egger, P. and McLaughlin, P. A. (2008). Do economic integration agreements actually work? Issues in understanding the causes and consequences of the growth of regionalism. *World Economy*, 31(4), pp. 461-497.
[5] Bergstrand, J. H. (1985). The gravity equation in international trade: some microeconomic foundations and empirical evidence. *The Review of Economics and Statistics,* 67(3), pp. 474–81.
[6] Bergstrand, J. H., Larch, M. and Yotov, Y. V. (2015). Economic integration agreements, border effects, and distance elasticities in the gravity equation. *European Economic Review*, 78, pp. 307-327.
[7] Correia, S., Guimarães, P. and Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal*, 20(1), pp. 95-115.
[8] Deardorff, A. (1998). Determinants of bilateral trade: does gravity work in a frictionless Model? *The Regionalisation of the World Economy*, edited by J. Frankel, 7-28.
[9] Chaney, T. (2008). Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review*, 98(4), pp. 1707–21.
[10] Eaton, J. and Kortum, S. (2002). Technology, geography and trade. Econometrica, 70(5), pp. 1741–79.
[11] Egger, P. and Larch, M. (2008). Interdependent Preferential Trade Agreement Membership: an Empirical Analysis, *Journal of International Economics*, 76(2), pp. 384-399.
[12] Egger, P., Larch, M., Staub, K. E. and Winkelmann, R. (2011). The trade effects of endogenous preferential trade agreements. *American Economic Journal: Economic Policy*, 3(3), pp. 113-43.
[13] Egger, P. and Nigai, S. (2015). Structural Gravity With Dummies Only: Constrained ANOVA-Type Estimation of Gravity Models. *Journal of International Economics,* 97(1), pp. 86-99.
[14] Eicher, T. S., Henn, C. and Papageorgiou, C. (2012). Trade creation and diversion revisited: Accounting for model uncertainty and natural trading partner effects. *Journal of Applied Econometrics*, 27(2), pp. 296-321.
[15] Fournier, J. M., Domps, A., Gorin, Y., Guillet, X. and Morchoisne, D. (2015). *Implicit Regulatory Barriers in the EU Single Market: New Empirical Evidence from Gravity Models* (No. 1181). OECD Publishing.
[16] Grančay, M., Grančay, N., Drutarovská, J. and Mura, L. (2015). Gravitačný model zahraničného obchodu českej a slovenskej republiky 1995–2012: ako sa zmenili determinanty obchodu. *Politická ekonomie*, 63(6), pp. 759-777.
[17] Gudgin, G., Coutts, K., Gibson, N. and Buchanan, J. (2017). The Role of Gravity Models in Estimating the Economic Impact of Brexit. Centre for Business Research, University of Cambridge, working paper no. 490.
[18] Harrigan, J. (1996). Openness to Trade in Manufactures in the OECD. *Journal of international economics*, 40(1-2), pp. 23–39.

[19] Head, K. and Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics* (Vol. 4, pp. 131-195). Elsevier.
[20] Helpman, E., Melitz, M. and Rubinstein, Y. (2008). Trading Partners and Trading Volumes. *Quarterly Journal of Economics*, 123(2), pp. 441-487.
[21] Hufbauer, G. C. and Schott, J. J. (2009). Fitting Asia-Pacific agreements into the WTO system. *Multilateralizing regionalism: Challenges for the global trading system*, pp. 554-635.
[22] Larch, M., Wanner, J., Yotov, Y. V. and Zylkin, T. (2019). Currency unions and trade: A PPML re-assessment with high-dimensional fixed effects. *Oxford Bulletin of Economics and Statistics*, 81(3), pp. 487-510.
[23] Oberhofer, H. and Pfaffermayr, M. (2021). Estimating the trade and welfare effects of Brexit: A panel data structural gravity model. *Canadian Journal of Economics*, 54(1), pp. 338-375.
[24] McCallum J (1995). National borders matter: Canada–US regional trade patterns. *American Economic Review*, 85(3), pp. 615–623.
[25] Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics,* 88(4), pp. 641–658.
[26] Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. 1st edn. New York: The Twentieth Century Fund.
[27] Trefler, D. (1995). The case of the missing trade and other mysteries. *The American Economic Review*, 85(5), pp. 1029–1046.
[28] Wei S J (1996). Intra-National versus International Trade: How Stubborn are Nations in Global Integration? Working Paper 5531, NBER.
[29] Yotov, Y. V., Piermartini, R. and Larch, M. (2016). *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*. WTO iLibrary.
[30] Zylkin, T. (2018). PPML_PANEL_SG: Stata module to estimate" structural gravity" models via Poisson PML.

**Author address**
doc. Ing. Brian König, PhD.
University of Economics in Bratislava
Faculty of business informatics
Department of Operations research and Econometrics,
Dolnozemská cesta 1/b, 852 35 Bratislava
Slovak Republic
email: brian.konig@euba.sk

Institute of Economic Research SAS
Šancová 56, 811 05 Bratislava
Slovak Republic

# EVALUATION OF DYNAMICS OF HEURISTICS USED FOR PARETO FRONT APPROXIMATION

***Marek Kvet, Jaroslav Janáček***
***University of Žilina, Faculty of Management Science and Informatics***

**Abstract**
Heuristic algorithms used for such a complex structure determination like a large set of non-dominated solutions of the emergency medical system design problem with conflicting criteria yield solutions, quality of which depends on a limit of computational time permitted for the algorithm run. When studying efficiency of different heuristics for non-dominated solution set determination, quality of the result achieved in a given computational time limit cannot be taken for an exhausted assessment. It holds especially in the case, when the studied heuristics are used as building blocks of more complex hyperheuristics, where the time limit will be adaptively adjusted. That is why, we suggested and studied a special measure of dynamic characteristic of heuristic algorithms developed for obtaining a good Pareto front approximation. The suggested measure is used in the computational study, when the studied heuristics are used for obtaining a set of non-dominated designs of the emergency medical system.

***Keywords:*** *Discrete location problems, Public service systems, Set of non-dominated solutions, Dynamics of heuristics performance, Pareto optimization*

***JEL Classification:*** C61, C63
***AMS Classification:*** 90C27

## 1 INTRODUCTION

There are host of hard mathematical programming problems, solving of which consists of repeated application of computational process of NP complexity. One of the tough problems is determination of the Pareto front of bi-criteria public service system designs. Since the exact approaches (Grygar and Fabricius, 2019, Janáček and Fabricius, 2021) demand for almost unacceptable computational time, the attention of researches has returned to more complex heuristics such as hybridized evolutionary algorithms or hyperheuristics (Janáček and Kvet, 2021a, Janáček and Kvet, 2021b, Kvet and Janáček, 2021). These metaheuristics control their computational processes by dynamic setting of control parameters or by choosing a suitable solving method of a lower rank, to perform solving process of a partial problem (Gendreau and Potvin, 2010). To enable rational choice of a method, it is necessary to know its characteristic, especially its contribution to an increment of the Pareto front approximation quality.

Quality of a bi-criteria Pareto front approximation can be measured by an area bounded by a curve induced by points representing solutions of a set of non-dominated solutions, which is used as the approximation of the unknown Pareto front (Janáček and Kvet, 2022). Nevertheless, this approach evaluates found set of non-dominated solutions, but it is insufficient evaluation of the heuristic process, which reached the associated result. The core of the insufficiency consists in the fact that most of experiments for heuristic comparison are organized so that a time threshold is determined and the heuristics are compared according to the results achieved in the given time. Usage of the heuristics in hyperheuristics or hybridized metaheuristics considers not only choice of a heuristic of a lower rank, but also determining the time interval, for which the heuristic should be run. To overcome the shortcoming resulting from the use of the area given by the non-dominated set of solutions, we propose an evaluation of heuristic performance using the mean value of area determined by the subsequently updated set of non-dominated solutions.

The rest of the paper is arranged as follows. In the next section, the process of non-dominated solution set updating is explained and the measure of heuristic quality is defined. In section 3, the public service system design problem with conflicting criteria is formulated. The tested heuristics are described in section 4 and the numerical experiments with application of the measure to heuristic comparison are described in section 5. The closing section summarizes the obtained results and gives the perspectives of their further usage.

## 2 DYNAMIC CHARACTERISTIC OF PROCESSES FOR PARETO FRONT APPROXIMATION

We consider a discrete optimization problem with a finite set $\boldsymbol{Y}$ of feasible solutions, where each solution can be evaluated according to two conflicting objectives formulated by objective functions $f_1$ and $f_2$. It holds that the less objective function value is the better solution from the associated point of view. The contradiction of the objectives means that improvement of one of the objectives is paid by degradation of the other one. Since both objective functions are to be minimized on the set $\boldsymbol{Y}$, the solution $\boldsymbol{z}$, for which such solution $\boldsymbol{y}$ is found that $f_1(\boldsymbol{y}) \leq f_1(\boldsymbol{z})$ and $f_2(\boldsymbol{y}) \leq f_2(\boldsymbol{z})$, does not need to be taken into account. We say that the solution $\boldsymbol{y}$ dominates the solution $\boldsymbol{z}$. The goal of bi-criteria problem solving is to determine so-called Pareto front (*PF*), what is a subset of $\boldsymbol{Y}$, for which each solution $\boldsymbol{z} \in \boldsymbol{Y}$ is dominated by at least one member of *PF* (Grygar and Fabricius, 2019, Janáček and Fabricius, 2021, Janáček and Kvet, 2022, Janáček and Kvet, 2021b). We deal here with the incrementing heuristic approaches, which construct a Pareto front approximation by step-by-step updating a current set of non-dominated solutions (*NDSS*), where the non-dominancy condition holds only for such solutions, which have been met during the incrementing process (Janáček and Kvet, 2021a, Kvet and Janáček, 2021). Arbitrary incrementing heuristic is based on exploration of feasible solutions, which can be obtained from a current solution by performing some of permitted operations. This way, an incrementing heuristic yields a flow of feasible solutions, which can be evaluated by $f_1$ and $f_2$ and compared to the members of current *NDSS*. Result of the comparison is finding that either some of *NDSS* members dominates the evaluated solution or any member of *NDSS* dominates it. In the first case, the candidate solution is refused and in the second case, the candidate is inserted into *NDSS*, what can be accompanied by excluding some former members of *NDSS* due to being dominated by the candidate. We assume that *NDSS* is a set of non-dominated solutions $\{\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots, \boldsymbol{y}^n\}$ ordered increasingly according to the values of $f_2$, i.e. $f_2(\boldsymbol{y}^1) < f_2(\boldsymbol{y}^2) < \ldots < f_2(\boldsymbol{y}^n)$. The solutions $\boldsymbol{y}^1$ and $\boldsymbol{y}^n$ will be denoted as the most left and right bordering solutions of *NDSS* respectively and the elements $[f_1(\boldsymbol{y}^1), f_2(\boldsymbol{y}^1)]$ and $[f_1(\boldsymbol{y}^n), f_2(\boldsymbol{y}^n)]$ of $E^2$ will be called the bordering points. The *NDSS* induces piece-wise constant decreasing function defined on the range $[f_2(\boldsymbol{y}^1), f_2(\boldsymbol{y}^1)]$ (see Fig. 1).

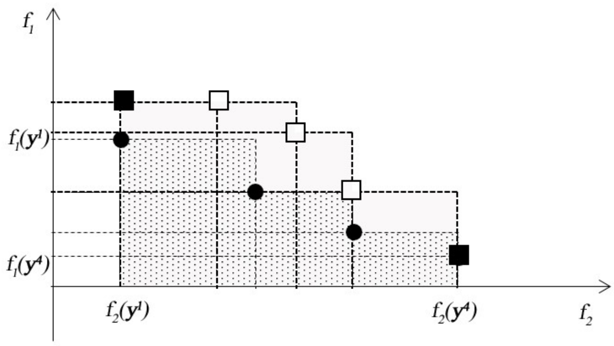



**Fig. 1** The black circles denote the points corresponding to Pareto front solutions $\boldsymbol{y}^k$, whereas the white squares denote the points corresponding to solutions of *NDSS*. To compare the areas correctly, it may be necessary to add dummy points, which are depicted by the black squares. The dotted area corresponds to $A_{PF}$. Similarly, it is possible to compute also the area $A_{NDSS}$ corresponding to members of *NDSS*.

The area $A_{NDSS}$ computed according to (1) converges to $A_{PF}$ under the assumption that the bordering points of *NDSS* closely approximate the bordering solutions of the Pareto front.

$$A_{NDSS} = \sum_{k=1}^{n-1} f_1\left(\boldsymbol{y}^k\right)\left(f_2\left(\boldsymbol{y}^{k+1}\right) - f_2\left(\boldsymbol{y}^k\right)\right) \tag{1}$$

The area $A_{NDSS}$ or the difference $A_{NDSS} - A_{PF}$ may express quality of the *NDSS* as approximation of the Pareto front. Based on the $A_{NDSS}$, we suggest a dynamic characteristic of an incrementing heuristic. We exploit the fact that any incrementing heuristic inspects the set $\boldsymbol{Y}$ solution by solution and only occurrence of successful candidate corresponds with the instant, when *NDSS* is changed and, simultaneously, the area $A_{NDSS}$ decreases.

Let us consider $T$ as a time horizon of testing an incrementing heuristic used as a generator of solutions-candidates for *NDSS* update. Let the sequence $\{t_0, t_1, \ldots, t_s\}$ be a set of instants, when the heuristic produced a successful candidate. Furthermore, let $NDSS(t_r)$ denote the *NDSS* after the $r$-th candidate insertion and let $NDSS(t_0)$ correspond to a starting *NDSS*. Then, the dynamic characteristic of the heuristic can be described by a volume $V_{Heur}$ computed according to (2), or it can be described by the mean value of area $\underline{A}$ computed as $\underline{A} = V_{Heur}/T$.

$$V_{Heur} = \sum_{r=0}^{s-1} A_{NDSS(t_r)}\left(t_{r+1} - t_r\right) + A_{NDSS(t_s)}\left(T - t_s\right) \tag{2}$$

## 3 DESIGN OF SERVICE SYSTEM WITH CONFLICTING CRITERIA

Simple or bi-criteria optimization finds its application in many fields like medical sphere (Doerner et al, 2005, Jankovič, 2016, Jánošíková, 2007) or in different areas of public sector (Avella, Sassano and Vasil'ev, 2007, Current, Daskin and Schilling, 2002, Marianov and Serra, 2002). Within this paper, we restrict ourselves on public service system designing with two contradictory objectives.

To formulate the problem, let $m$ denote the number of candidate locations for a service center. Our task is to choose $p$ location, in which a service center should be located in order to minimize given objectives. The decision on a service center locating will be modelled by a binary variable $z_i$, which takes the value of one, if a center should be located at the location $i = 1, \ldots, m$. Then, the set $\boldsymbol{Y}$ of all feasible solutions $\boldsymbol{z}$ takes the following form of (3).

$$\boldsymbol{Y} = \left\{ \boldsymbol{z} \in \{0, 1\}^m : \sum_{i=1}^{m} z_i = p \right\} \tag{3}$$

As far as the objective functions are concerned, let $n$ denote the number of system users' locations. From the bi-criteria point of view, we distinguish so-called *system* and *fair* criteria.

The system criterion expresses average response time of the system to randomly emerging demand for service of a system user. It is often used, when only one objective is considered (Avella, Sassano and Vasil'ev, 2007, Jankovič, 2016, Jánošíková, 2007). Let $b_j$ denote the number of demands for service at the location $j = 1, \ldots, n$. Under the assumption that the service is provided from the nearest available center, which does not have to be the nearest located one, we need the number of centers $r$, which share system providing to the concrete user. The time distance from a user location to the $k$-th nearest center is weighted by a positive coefficient $q_k$, $k = 1, \ldots, r$. Finally, let $t_{ij}$ denote the time distance from the location $i = 1, .., m$ to the location $j = 1, \ldots, n$. Based on these preliminaries, the system criterion $f_1(\boldsymbol{z})$ for any solution $\boldsymbol{z}$ can be formulated by the expression (4).

$$f_1(\boldsymbol{z}) = \sum_{j=1}^{n} b_j \sum_{k=1}^{r} q_k \min{}_k \left\{ t_{ij} : i = 1, \ldots, m;\ z_i = 1 \right\} \tag{4}$$

The fair criterion takes into account the worst situated users and it can be expressed as a number of users' demands outside a given radius $T$ from the nearest located center (Bertsimas, Farias and Trichakis, 2011, Buzna, Koháni and Janáček, 2013). The formula (5) can mathematically describe it.

$$f_2(\boldsymbol{z}) = \sum_{j=1}^{n} b_j \max\left\{ 0,\ sign\left( \min\left\{ t_{ij} : i = 1, \ldots, m; z_i = 1 \right\} - T \right) \right\} \tag{5}$$

These two objectives may conflict (Grygar and Fabricius, 2019, Janáček and Fabricius, 2021, Janáček and Kvet, 2021b). It means that a decrease of one objective function value is paid by an increase of the other, when minimizing one of the objective functions.

# 4 INCREMENTING HEURISTICS FOR PARETO FRONT APPROXIMATION

The incrementing heuristics for Pareto front approximation is based on a swap algorithm, which proved to be a very useful tool, which could be used in combination with updating the set of non-dominated solutions (Janáček and Kvet, 2021a, Janáček and Kvet, 2021b, Kvet and Janáček, 2021). Thus, it represents the base for a solving approach to two objective public service system design problem.

The basic operation of the swap algorithm is applied on a feasible solution $\boldsymbol{y}$ and it consists of moving a center from its current location $i$ to an unoccupied location $j$. The resulting solution can be denoted as $Exchange(\boldsymbol{y}, i, j) = [y_1, y_2, \ldots, y_{i-1}, 0, y_{i+1}, \ldots, y_{j-1}, 1, y_{j+1}, \ldots, y_m]$.

The algorithm described by the following pseudocode starts with an initial solution $\boldsymbol{y}$, initial set *NDSS* of non-dominated solutions and with the vector $\boldsymbol{a}$ of coefficients $a_1$ and $a_2$.

*Search*($\boldsymbol{y}$, $\boldsymbol{a}$, *NDSS*)
0. Initialize $F = a_1 f_1(\boldsymbol{y}) + a_2 f_2(\boldsymbol{y})$, and $BestF = F$.
1. For each $i = 1, \ldots, m, y_i = 1$ and $j = 1, \ldots, m, y_j = 0$ repeat step 2 and then go to step 3.
2. Perform *UpdateNDSS*(*NDSS*, *Exchange*($\boldsymbol{y}$, $i, j$)).
   If $F > a_1 f_1(Exchange(\boldsymbol{y}, i, j)) + a_2 f_2(Exchange(\boldsymbol{y}, i, j))$, then
   put $F = a_1 f_1(Exchange(\boldsymbol{y}, i, j)) + a_2 f_2(Exchange(\boldsymbol{y}, i, j))$ and $i^* = i$, $j^* = j$.
3. If $F < BestF$ then update $\boldsymbol{y} = Exchange(\boldsymbol{y}, i^*, j^*)$ and go to step 1. Otherwise, return $\boldsymbol{y}$ and *NDSS* and finish the process.

The above algorithm *Search*($\boldsymbol{y}$, $\boldsymbol{a}$, *NDSS*) has been embedded in repetitive search strategy that processes solutions from the current $NDSS = \{\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots, \boldsymbol{y}^n\}$ in order from solution $\boldsymbol{y}^1$ to solution $\boldsymbol{y}^{n-1}$. The solution $\boldsymbol{y}^k$ is used as a starting solution and the algorithm minimizes the surrogate objective function $a_1 f_1(\boldsymbol{y}) + a_2 f_2(\boldsymbol{y})$, i.e. it proceeds in the direction $\boldsymbol{d} = [-a_1, -a_2]$ in the two-dimensional space $f_1 \times f_2$. During this minimization process, the current *NDSS* is updated and, after the process has finished, the strategy continues depending on the case that occurred. If the solution $\boldsymbol{y}^k$ has not been modified, then strategy continues with $\boldsymbol{y}^{k+1}$, otherwise it repeats the search with modified solution $\boldsymbol{y}^k$. After the complete current *NDSS* is processed, the strategy repeats *NDSS* processing until the consumed computational time exceeds a given time limit. As concerns the direction $\boldsymbol{d} = [-a_1, -a_2]$, it is chosen from the cone bounded by so-called tangential and radial directions. The tangential direction $\boldsymbol{d}^T$ is determined by solutions $\boldsymbol{y}^k$ and $\boldsymbol{y}^{k+1}$ as $\boldsymbol{d}^T = [f_1(\boldsymbol{y}^{k+1}) - f_1(\boldsymbol{y}^k), f_2(\boldsymbol{y}^{k+1}) - f_2(\boldsymbol{y}^k)]$. The radial direction $\boldsymbol{d}^R$ cam be determined as a perpendicular vector to the $\boldsymbol{d}^T$, i.e. $\boldsymbol{d}^R = [f_2(\boldsymbol{y}^k) - f_2(\boldsymbol{y}^{k+1}), -f_1(\boldsymbol{y}^k) + f_1(\boldsymbol{y}^{k+1})]$.

The direction $\boldsymbol{d}$ is determined by the expression $\boldsymbol{d} = \alpha \boldsymbol{d}^R + (1-\alpha)\boldsymbol{d}^T$ for $\alpha \in [0, 1]$, where coefficient $\alpha$ is a parameter of the method. If $\alpha = 1$, the search process in the radial direction is performed and if $\alpha = 0$, then the tangential direction is used. Impact of a used value of $\alpha$ on the method efficiency is studied in the next section.

# 5 NUMERICAL EXPERIMENTS

This section summarizes the results of performed numerical experiments and brings the analysis of the impact of parameter settings on the result accuracy.

As far as used hardware and software tools are concerned, all algorithms were implemented in the Java language within the NetBeans IDE 8.2 environment. The experiments were run on a PC equipped with the Intel® Core™ i7 8550U CPU@1,8 GHz processor and 32 GB RAM.

The dataset of benchmarks is the same as in our previous research reported in (Janáček and Fabricius, 2021, Janáček and Kvet, 2022, Janáček and Kvet, 2021a, Janáček and Kvet, 2021b, Kvet and Janáček, 2021). As a source if time distances we used the road network of eight self-governing regions Slovakia, in which the Emergency Medical Service System is operated. Those benchmarks are denoted by their abbreviations. The list of problem instances contains Bratislava (BA), Banská Bystrica (BB), Košice (KE), Nitra (NR), Prešov (PO), Trenčín (TN), Trnava (TT) and Žilina (ZA). In the problem instances, all inhabited network nodes represent both the set of candidates for a service center and the set of system users' locations. From the viewpoint of studied objectives (4) and (5), the parameter $r$ was set to 3. The associated coefficients $q_k$ for $k$=1, …, $r$ were set in percentage in the following way: $q_1 = 77.063$, $q_2 = 16.476$ and $q_3 = 100 - q_1 - q_2$. These values were obtained from a simulation model of existing Emergency Medical Service system in Slovakia (Jankovič, 2016). Parameter $T$ used in the fair objective function (5) was set to the value of 10 minutes (Janáček and Kvet, 2021b). Basic characteristics of the benchmarks are summarized in Table 1, the structure of which takes the following form. For each benchmark corresponding to one row of the table, the column denoted by $m$ contains the number of possible service center locations. The parameter $p$ expresses the number of facilities to be located. The middle part of the table contains the value of *NoS* that means the number of solutions forming the Pareto front. In the column denoted by *Area* we provide the readers with the size of $A_{PF}$ (see Fig. 1) computed according to the formula (1). The right part of the table contains additional information about the bordering points of the Pareto front. Let the symbol *MLM* denote the most left member of the Pareto front and let *MRM* denote the most right member. Then, the objective function value (4) for *MLM* will be denoted as $f_1(MLM)$ and the criterion (5) for *MLM* will be denoted by $f_2(MLM)$. We will use an analogical denotation also for *MRM*.

**Table 1** Benchmarks characteristics and the exact Pareto fronts features

| Region | $m$ | $p$ | *NoS* | *Area* | $f_1(MLM)$ | $f_2(MLM)$ | $f_1(MRM)$ | $f_2(MRM)$ |
|---|---|---|---|---|---|---|---|---|
| BA | 87 | 14 | 34 | 569039 | 42912 | 0 | 26649 | 280 |
| BB | 515 | 36 | 229 | 1002681 | 53445 | 453 | 44751 | 935 |
| KE | 460 | 32 | 262 | 1295594 | 61241 | 276 | 45587 | 816 |
| NR | 350 | 27 | 106 | 736846 | 59415 | 557 | 48940 | 996 |
| PO | 664 | 32 | 271 | 956103 | 65944 | 711 | 56703 | 1282 |
| TN | 276 | 21 | 98 | 829155 | 45865 | 223 | 35274 | 567 |
| TT | 249 | 18 | 64 | 814351 | 48964 | 450 | 41338 | 921 |
| ZA | 315 | 29 | 97 | 407293 | 48025 | 254 | 42110 | 728 |

An individual experiment was organized so that the suggested solving procedure was run with different settings of parameter $\alpha$. The studied values were 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0. The values 0.0 and 1.0 mean that only one of radial or tangential strategy will be applied. Other values represent a convex combination of both approaches. For completeness, an individual run of the whole solving process was limited by a time threshold, which was set to 5 minutes. The obtained results are reported in the following tables. Each table summarizes the results for a specific setting of $\alpha$. All the following tables have the same structure. Each column of the table represents one solved instance. For each benchmark, several output data are reported. The row *CT* contains the computational time in seconds. The objective functions $f_1$ and $f_2$ of the most left and the most right member of *NDSS* are not reported due to the fact that they are almost the same as in the exact Pareto front. The number of *NDSS* members is reported in the row denoted by *noNDSS*. The size of area formed by the *NDSS* members is given in the row denoted by *FinArea*. The row denoted by *MeanArea* contains the mean value of area. The denotation *noOS* expresses the number of outer cycles performed during the solving procedure.

**Table 2** Results of numerical experiments for $\alpha = 0.0$

| Region | BA | BB | KE | NR | PO | TN | TT | ZA |
|---|---|---|---|---|---|---|---|---|
| *CT* | 300.1 | 614.1 | 433.8 | 348.0 | 842.0 | 308.5 | 307.3 | 323.3 |
| *noNDSS* | 30 | 155 | 159 | 102 | 160 | 85 | 62 | 89 |
| *FinArea* | 600880 | 984317 | 1319957 | 729825 | 858890 | 865033 | 824330 | 411994 |
| *MeanArea* | 601158.3 | 1059427.4 | 1533179.2 | 751472.8 | 930406.6 | 870884.7 | 829757.6 | 424487.5 |
| *noOS* | 705 | 1 | 1 | 4 | 1 | 12 | 28 | 7 |

**Table 3** Results of numerical experiments for $\alpha = 0.2$

| Region | BA | BB | KE | NR | PO | TN | TT | ZA |
|---|---|---|---|---|---|---|---|---|
| *CT* | 300.4 | 674.4 | 432.1 | 318.1 | 958.9 | 300.9 | 304.9 | 305.0 |
| *noNDSS* | 30 | 147 | 157 | 103 | 203 | 83 | 61 | 86 |
| *FinArea* | 598294 | 989677 | 1340588 | 757646 | 827702 | 865586 | 824352 | 412233 |
| *MeanArea* | 598461.8 | 1056333.4 | 1528712.0 | 789088.4 | 891076.9 | 871115.3 | 829459.9 | 423072.6 |
| *noOS* | 631 | 1 | 1 | 3 | 1 | 13 | 27 | 7 |

**Table 4** Results of numerical experiments for $\alpha = 0.4$

| Region | BA | BB | KE | NR | PO | TN | TT | ZA |
|---|---|---|---|---|---|---|---|---|
| *CT* | 300.2 | 731.9 | 455.6 | 385.0 | 782.6 | 313.1 | 300.9 | 305.9 |
| *noNDSS* | 29.0 | 184.0 | 176.0 | 103.0 | 175.0 | 82.0 | 64.0 | 89.0 |
| *FinArea* | 582609.0 | 979169.0 | 1223679.0 | 758714.0 | 847284.0 | 865602.0 | 818920.0 | 411992.0 |
| *MeanArea* | 582877.6 | 1039337.9 | 1422294.5 | 779816.6 | 921942.1 | 870478.3 | 823219.3 | 424980.6 |
| *noOS* | 729 | 1 | 1 | 4 | 1 | 13 | 27 | 7 |

**Table 5** Results of numerical experiments for $\alpha = 0.6$

| Region | BA | BB | KE | NR | PO | TN | TT | ZA |
|---|---|---|---|---|---|---|---|---|
| *CT* | 300.2 | 619.5 | 425.5 | 338.4 | 643.8 | 317.8 | 306.9 | 300.4 |
| *noNDSS* | 30 | 157 | 171 | 102 | 160 | 83 | 63 | 89 |
| *FinArea* | 600880 | 986021 | 1310798 | 740930 | 864419 | 866075 | 819601 | 412087 |
| *MeanArea* | 601022.3 | 1045677.0 | 1513893.7 | 760550.6 | 937825.9 | 870192.9 | 827157.7 | 419651.3 |
| *noOS* | 816 | 1 | 1 | 5 | 1 | 13 | 28 | 8 |

**Table 6** Results of numerical experiments for $\alpha = 0.8$

| Region | BA | BB | KE | NR | PO | TN | TT | ZA |
|---|---|---|---|---|---|---|---|---|
| *CT* | 300.1 | 608.1 | 444.5 | 324.7 | 700.9 | 316.8 | 302.0 | 322.2 |
| *noNDSS* | 32 | 168 | 187 | 104 | 199 | 83 | 64 | 89 |
| *FinArea* | 581872 | 977746 | 1251284 | 779284 | 865086 | 865507 | 818920 | 412066 |
| *MeanArea* | 582067.6 | 1021252.8 | 1351548.1 | 783541.1 | 906049.8 | 868641.1 | 822805.4 | 419795.3 |
| *noOS* | 880 | 1 | 1 | 5 | 1 | 14 | 33 | 9 |

**Table 7** Results of numerical experiments for $\alpha = 1.0$

| Region | BA | BB | KE | NR | PO | TN | TT | ZA |
|---|---|---|---|---|---|---|---|---|
| *CT* | 300.2 | 337.2 | 489.2 | 311.2 | 434.6 | 301.3 | 304.3 | 306.6 |
| *noNDSS* | 33 | 203 | 230 | 105 | 249 | 95 | 62 | 90 |
| *FinArea* | 592690 | 939307 | 1244591 | 775288 | 817426 | 834807 | 815095 | 407905 |
| *MeanArea* | 592825.7 | 969564.1 | 1275918.5 | 784559.2 | 867147.1 | 838091.0 | 817742.5 | 414238.0 |
| *noOS* | 1435 | 1 | 2 | 8 | 1 | 20 | 56 | 14 |

To make the comparison of the obtained results simpler, the following Table 8 summarizes the results for the self-governing region of Žilina (ZA) for different settings of parameter $\alpha$ to see the impact of its value on the result accuracy.

**Table 8** Results of numerical experiments for the self-governing region of Žilina and different values of $\alpha$

| $\alpha$ | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |
|---|---|---|---|---|---|---|
| *CT* | 323.3 | 305.0 | 305.9 | 300.4 | 322.2 | 306.6 |
| *noNDSS* | 89 | 86 | 89 | 89 | 89 | 90 |
| *FinArea* | 411994 | 412233 | 411992 | 412087 | 412066 | 407905 |
| *MeanArea* | 424487.5 | 423072.6 | 424980.6 | 419651.3 | 419795.3 | 414238.0 |
| *noOS* | 7 | 7 | 7 | 8 | 9 | 14 |

# 6 CONCLUSIONS

This paper was aimed at such public service system design problem, in which two conflicting criteria need to be taken into account. When bi-criteria optimization is to be performed, a Pareto front of non-dominated system designs or its approximation seem to be a sufficient output for the decision makers. The reported research was focused on a special measure of dynamic characteristic of heuristic algorithms developed for obtaining a good Pareto front approximation. Suggested approaches were experimentally verifies to study their performance characteristic and result accuracy. The achieved results have proved suitability of the heuristic method for practical usage, because it enables us to obtain good results within acceptably short time. If we compare the Pareto front approximation to the exact set of non-dominated solution, we can see that the cardinality of the approximate set is very near to the exact one. As far as the approximation quality is concerned, the suggested approach seems to be sensitive to the setting of the input parameter. The best results were obtained in such a case, when the search process in the radial direction was performed. Thus, we can conclude that we have significantly contributed to the state-of the-art methods and analyses connected with bi-criteria optimization and Pareto front construction.

Future research in this field could be aimed at other possible ways of parameter settings and different heuristic methods, which could lead to better results. Another scientific goal could consists in generalization for multi-criteria optimization problems.

**Acknowledgements:** This work was supported by the research grants VEGA 1/0216/21 “Design of emergency systems with conflicting criteria using artificial intelligence tools” and VEGA 1/0654/22 “Cost-effective design of combined charging infrastructure and efficient operation of electric vehicles in public transport in sustainable cities and regions”. This work was supported also by the Slovak Research and Development Agency under the Contract no. APVV-19-0441.

**References**

[1] Avella, P., Sassano, A. and Vasil'ev, I. (2007). Computational study of large scale p-median problems. *Mathematical Programming*, 109, pp. 89-114.

[2] Bertsimas, D., Farias, V. F. and Trichakis, N. (2011). The Price of Fairness. In: *Operations Research*, 59, pp. 17-31.

[3] Buzna, Ľ., Koháni, M. and Janáček, J. (2013). Proportionally Fairer Public Service Systems Design. *Communications - Scientific Letters of the University of Žilina*, 15(1), pp. 14-18.

[4] Current, J., Daskin, M. and Schilling, D. (2002). Discrete network location models, Drezner Z. et al. (ed) *Facility location: Applications and theory*, Springer, pp. 81-118.

[5] Doerner, K. F., Gutjahr, W. J., Hartl, R. F., Karall, M. and 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.

[6] Gendreau, M. and Potvin, J. (2010). *Handbook of Metaheuristics*, Springer Science & Business Media.

[7] Grygar, D. and Fabricius, R. (2019). An efficient adjustment of genetic algorithm for Pareto front determination. In: TRANSCOM 2019: conference proceedings, Amsterdam: Elsevier Science, pp. 1335-1342.

[8] Janáček, J. and Fabricius, R. (2021). Public service system design with conflicting criteria. *IEEE Access: practical innovations, open solutions*, 9, pp. 130665-130679.

[9] Janáček, J. and Kvet, M. (2022). Quality Measure of Pareto Front Approximation. In: Proceedings of the international conference on Quantitative Methods in Economics 2022, in print.

[10] Janáček, J. and Kvet, M. (2021a). Swap Heuristics for Emergency System Design with Multiple Facility Location. In: Proceedings of the 39th International Conference on Mathematical Methods in Economics, 2021, pp. 226-231.

[11] Janáček, J. and Kvet, M. (2021b). Emergency Medical System under Conflicting Criteria. In: SOR 2021 Proceedings, pp. 629-635.

[12] Jankovič, P. (2016). Calculating Reduction Coefficients for Optimization of Emergency Service System Using Microscopic Simulation Model. In: 17th International Symposium on Computational Intelligence and Informatics, pp. 163-167.
[13] Jánošíková, Ľ. (2007). Emergency Medical Service Planning. *Communications - Scientific Letters of the University of Žilina*, 9(2), pp. 64-68.
[14] Kvet, M. and Janáček, J. (2021). Incrementing Heuristic for Non-Dominated Designs of Emergency Medical System. In: SOR 2021 Proceedings, pp. 429-474.
[15] Marianov, V. and Serra, D. (2002). Location problems in the public sector, Drezner Z. et al. (ed) *Facility location - Applications and theory*, Berlin, Springer, pp. 119-150.

**Author's address**
doc. Ing. Marek Kvet, PhD., prof. RNDr. Jaroslav Janáček, CSc.
University of Žilina - Faculty of Management Science and Informatics
Univerzitná 8215/1, 010 26 Žilina, Slovakia
Email: marek.kvet@fri.uniza.sk, jaroslav.janacek@fri.uniza.sk

# APPLICATION OF PREDICTIVE ANALYSIS METHOD TO OPTIMIZE WAREHOUSE WORKFORCE IN A 3PL COMPANY USING WEKA SOFTWARE TOOLS

***Milica Mitrović, Dražen Popović, Gordana Radivojević***
***University of Belgrade, Faculty of Transport and Traffic Engineering, Serbia***

**Abstract**
Logistics companies continuously collect large datasets using modern logistics information systems. Modern technologies enable managers to make decisions based on data-driven methods. Based on the collected historical data and by applying of predictive analysis in data mining software tools, companies can predict the demand for goods or services. In this paper, we have applied Weka software tool for prediction of daily number of orderpicking items and the number of orderpicker operations. The goal is to plan for a workforce load in the warehouse based on the forecast. The data were obtained from a logistics company that operates in Serbia and provides warehousing and transportation services. Conducted predictive analysis can help managers make decisions when managing the workforce in the warehouse. The best results in the forecast of the workload, in terms of orderpicking items and number of operations, was obtained by the M5P, SMOreg and Linear Regression algorithms.

***Keywords:*** *logistics, machine learning, forecasting, workforce management*

***JEL Classification:*** C, L
***AMS Classification:*** 62-07

## 1 INTRODUCTION

Warehouses are important places to achieve potential cost savings in the supply chain. Warehouse costs include costs: space, facilities, workforce, transhipment and handling equipment, information and communication systems, etc. By applying and implementing certain measures and solutions, potential cost savings can be achieved. One of the possible potentials for optimizing warehouse operation and reducing operating costs is workforce. The challenge for warehouse managers is to determine the workforce load for warehousing operations at different time horizons. Managers tend to determine the workforce they need, so that there is no shortage of workforce, and no overload of work tasks. Growing market demand is forcing companies to be flexible and able to predict future demand for a more profitable business. Machine learning methods and data mining tools are increasingly being used to improve supply chain efficiency. Data mining tools have been shown to provide better organizational efficiency compared to traditional methods of predictive analysis (Franz, 2020).

The aim of this paper is to obtain a daily forecast of the number of orderpicking items (total lines in picking orders) and the number of orderpicking operations (picking item quantities, e.g. pieces, inner packages, transportation packages etc.) to determine the workforce load using the predictive analysis method in the Weka software tool. All phases of the machine learning process have been conducted in the Weka software tool and different models have been created to forecast target variables. Determining the required number of workers in a warehouse is often challenging because of the variable load in time. The cause of variable load is variable demand in the market. Predicting the workload and capacity required is essential in planning the workforce in the warehouse. Variability in the supply chain can be caused by seasonality of products, serial production, product consolidation, product refinement, etc. (Krishna and Prabhu, 2016).

The paper is organized into six chapters, in addition to the introduction and conclusion. The second chapter provides a review of the literature on workforce management. The third chapter describes the problem of workforce management in warehouses. The fourth chapter describes the demand forecast. The fifth chapter describes a case study. In the sixth chapter, the results of the case study are presented and analyzed. The last chapter contains the conclusion about the conducted study and the directions of future research.

## 2 REVIEW OF LITERATURE

Competition pressure in the market is forcing distribution centers to increase the number of order processing in a shorter period of time. Vanheusden et al. (2021) state that the order piker's load forecast can reduce uneven load redistribution and thus improve process efficiency. The authors used different statistical approaches to determine the workload of workers (e.g. Rawlsian's approach, range, mean-based) in the orderpicking process. The results of the research showed that managerial activities significantly affect the workload of orderpicker to be to be more even than the warehouse layout.

Improving the workforce management process in orderpicking activities can contribute to a more efficient and effective process of picking goods (Van Gils et al., 2017). The authors applied time series models for daily order forecast items. Such a forecast can be used to determine the required number of orderpicker, as well as to deploy orderpicker by storage zones. Fluctuation in the workforce and a large number of customer demands create certain challenges for distribution centers. Popović et al. (2021) dealt with the problem of workforce deployment in the distribution center using integer linear programming. Variable customer demands are the cause of the problem of variable need for workers during the week. Also, variable demand is the cause of labor shortages or work overload in certain periods of the work day. Warehouse managers deploy workers to different activities in full-time or part-time shifts, but regardless of the defined working hours, there may again be overtime or inadequate use of working hours in certain periods.

Kim et al. (2017) presented a warehouse work planning methodology that is based on an analytical relationship between the demand forecast model and the labor productivity model. Sainathuni et al. (2014) presented a model of nonlinear integer programming to determine the optimal distribution plan from supplier to customer to reduce distribution costs. In the process of developing the model, the authors took into account the workload of warehouse workers that you affect their productivity. The results of the model compared to the sequential approach showed less variance of the load on warehouse workers and lower total distribution costs. Mundschenk and Drexl (2007) presented an integer programming model for determining the number of employees in the long term. The developed model reduces workforce costs by 26-39%. The authors note that effective workforce management is a key factor in achieving the company's long-term market success.

# 3 WORKFORCE MENAGMENT IN WAREHOUSES

Workforce management is a set of techniques used to increase the productivity and efficiency of employees in the company. Techniques for workforce management include human resource management, human capital management, performance management, data collection, budgeting, forecasting, etc. The goal of effective workforce management is to hire, develop, and deploy competent workers to the right jobs at the right time to meet organizational and individual goals (McDonald et al., 2017). Effective workforce planning and management is a challenge for companies in many industries, especially for service companies (Lee et al., 2008).

A series of activities and processes are realized in the warehouse from receipt to shipment of goods to the recipient. Some of the limiting factors in warehousing processes are the space and workforce capacity. Workforce scheduling and planning is an important task for managers in distribution centers. The complexity of this task arises from the variability of the workload of workers during different periods, i.e. variable demand affects the variable workload of workers in different time horizons (Popović et al., 2021). When it comes to 3PL companies that provide last mile distribution service and from their own warehouses, the workload depends on the volume of order picking, i.e. the number of items and the intensity of operations on items of orderpicking orders.

Two problems in workforce planning are fluctuations and uncertainty of demand, often solved in practice by concluding short-term employment contracts. However, the long recruitment process and high cost of training employees have shown that short-term contracts are not the best way to match the number of employees with the real needs of the workforce (Zhu and Sherali, 2009). Many warehouse managers hire temporary workforce, in addition to full-time employment, to better respond to variable demand for goods and to achieve an even distribution of tasks to workers. The problem with hiring temporary workforce can be the qualification of potential workers to perform the required jobs. Therefore, there are warehouses that deliberately work with excess full-time workers without wanting to risk whether skilled workers will be available at the right time (Kim et al., 2017).

# 4 DEMAND FORECAST

The demand forecast is a technique for estimating the possible demand for a good or service in the future based on historical data. Based on forecasts of future demand, distribution centers can define the quantity and time horizon of demand for goods. Precisely, demand forecast is a key planning process in the supply chain. Using different periods to forecast demand requires different analytical methods (Souza, 2014). In literature, there are several divisions of forecasting methods, and the general division consists of two categories: qualitative forecasting and quantitative forecasting. Quantitative forecasting is based on historical data, and qualitative based on the estimates and opinions of experts (Petković, 2013).

Quantitative forecasting is used when there are enough numerical values to observe historical data to identify patterns. It can be assumed that the identified patterns from the past are likely to be repeated in the future. Time horizon components that are useful to observe are: trend components, seasonal components and error components. A trend is the direction of movement in a time horizon. The seasonal component is a pattern that always occurs in the same period. The error component represents the value of the random error in the period without the existence of a trend and seasonal component (Fischetti, 2018).

Modern technologies enable the collection and processing of large amounts of data in the supply chain. Namely, demand forecasting is based on the data-driven method that managers use to decision making. Demand forecasting is an indispensable process in the inventory planning and operations system in warehouses. Demand forecasting shows companies what to expect in the coming days, weeks or months. Forecasts are characterized by a certain uncertainty that are sampled by fluctuations and other external factors in the market. Therefore, different prediction techniques are used to minimize errors between projected and actual values (Franz, 2020). Machine learning methods have been proposed as an alternative to statistical methods for forecasting demand (Evangelos et al., 2020).

# 5 CASE STUDY

In this paper, a case study has been presented to determine the workforce load in the warehouse. To determine the workforce load, daily forecasts have been made for the target variable orderpicking items and the target variable number of operations. Orderpicking items represent the total number of lines on all orderpicking orders (one line represents an item). The number of operations represents the sum of the orderpicking units of goods, i.e. the sum of the orderpicking units of packaging and pieces of goods. Based on these two measures, the orderpicking time can be calculated and therefore determined the workload, e.g. in the case of observed logistics company from Serbia one worker in average can pick 350 orderpicking items or can do 1750 orderpicking operations. The conducted forecasts are based on real data from the orderpicking process, obtained from leading logistics provider in providing warehousing and transportation services in Serbia.

The initial dataset contains real purchase order data of one client and 379 items, with total of 1.007.156 instances (rows that represents lines or items in picking orders) and contains the following data: date, order number, item code, orderpicking transport packages and orderpicking pieces. The dataset used to train machine learning models refers to the period in 2020, and the dataset used in the first quarter of 2021 was used to testing the model. In the case study, all available algorithms for model training were applied in the Weka software tool, and the paper presents models for algorithms that gave the best results of the observed performance. The observed performances of predictive models are: correlation coefficient (r), mean absolute error (MAE), square root of mean square error (RMSE), relative absolute error (RAE), and root relative square error (RRSE).

# 6 RESULTS

The machine learning process was realized in the data mining software tool Weka. Weka is open source software, developed at the University of Waikato in New Zealand. This software contains a set of algorithms used to detect patterns in data. Weka's software tool enables the following data mining tasks: data preparation, classification, regression analysis, clustering, association rule learning, selection of relevant attributes and data visualization. In the process of machine learning, the technique used to validate and evaluate the model is 10-fold cross-validation. Cross-validation is a statistical method for obtaining a reliable estimate of the performance of created models using only a set of training data (Witten and Frank, 2005). The basic dataset is divided into two sets, one set is used to train the model and the other to test the accuracy of that model.

The training dataset consists of 254 instances and the test dataset of 62 instances (one instance represents one day). For the forecast of orderpicking items (target variable), the independent attributes that affect the target variable are: date, month, day of the week and number of operations. Table 1 shows the performance of five models for the daily forecast of the number of orderpicking items measured on the training dataset. Based on the presented results, it can be observed that the Multilayer Perceptron algorithm has worse performance results than other algorithms. The correlation coefficient of models based on the M5P, SMOreg, Linear Regression and Gaussian Processes algorithms is greater than 0.9. Therefore, the process of machine learning continues with models based on the mentioned algorithms.

**Table 1**. Performance of Prediction Models obtained using Training Dataset for a target variable orderpicking item

| Algorithm | r | MAE | RMSE | RAE % | RRSE % |
|---|---|---|---|---|---|
| M5P | 0.9352 | 246.4734 | 335.2989 | 34.8157 | 35.3547 |
| SMOreg | 0.9348 | 242.1274 | 336.2703 | 34.2018 | 35.4572 |
| LinearRegression | 0.9334 | 247.8759 | 339.8637 | 35.0138 | 35.8361 |
| GaussianProcesses | 0.9078 | 302.0875 | 428.1838 | 42.6714 | 45.1488 |
| MultilayerPerceptron | 0.8974 | 338.7385 | 426.6221 | 47.8486 | 44.9841 |

Table 2 shows the performance of the four models measured on a dataset for testing. A comparative analysis of the performance of the models measured on the test dataset with the performance of the models measured on the training set based on the SMOreg, M5P and Gaussian Processes algorithms shows that the performance of the training model is better. However, the model based on the Gaussian Processes algorithm shows the presence of the problem of underfitting of actual and projected values. The correlation coefficient of the model of the mentioned algorithm is less than 0.9, which indicates that it cannot be used to predict the target variable. Also, it has been observed that the performance results are slightly better on testing than on training models based on the Linear Regression algorithm. Precisely, this result indicates the tendency of the overfitting problem. Models based on M5P and SMOreg algorithms have the best results of the observed performance and can be used to forecast the number of orderpicking items.

**Table 2**. Performance of Prediction Models obtained using Test Dataset for a target variable orderpicking item

| Algorithm | r | MAE | RMSE | RAE % | RRSE % |
|---|---|---|---|---|---|
| M5P | 0.9331 | 285.3429 | 338.6614 | 44.9134 | 43.3057 |
| SMOreg | 0.9320 | 263.6544 | 321.3267 | 41.4996 | 41.0891 |
| Linear Regression | 0.9337 | 285.8687 | 338.8290 | 44.9962 | 43.3272 |
| GaussianProcesses | 0.8831 | 301.5625 | 391.8845 | 47.4664 | 50.1115 |

The graph in Figure 1 shows the ratio of actual and projected values of the target variable for the period from January to March 2021. The forecast of the target variable was made using two models with the best performance. The graph shows that the forecast of the target variable conducted on the test dataset closely follows the actual values of the orderpicking items.

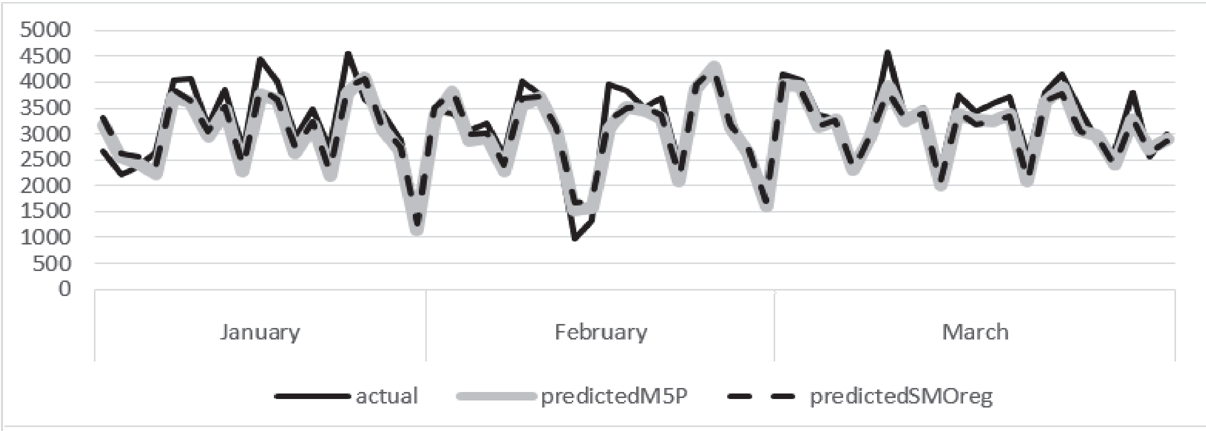



**Figure 1**. Actual and projected values orderpicking items on a daily basis

To forecast the number of operations (the second target variable being observed), the independent attributes that affect the target variable are: date, month, day of the week, and orderpicking items. Table 3 shows the performance of the four best models for predicting the number of operations measured on a training dataset. The further process of machine learning continues with models based on the algorithms Linear Regression, SMOreg, M5P and Gaussian Processes.

**Table 3**. Performance of Prediction Models obtained using Training Dataset for the target variable of the number of operations

| Algorithm | r | MAE | RMSE | RAE % | RRSE % |
|---|---|---|---|---|---|
| Linear Regression | 0.9398 | 1385.4202 | 1947.1289 | 32.1314 | 34.0514 |
| SMOreg | 0.9395 | 1357.0873 | 1962.0951 | 31.4743 | 34.3131 |
| M5P | 0.9366 | 1435.1901 | 1997.9102 | 33.2857 | 34.9394 |
| GaussianProcesses | 0.9065 | 1814.0755 | 2573.0998 | 42.0730 | 44.9983 |

Table 4 shows the performance of the four models measured on the test dataset. It has been observed that models based on Linear Regression, SMOreg and M5P algorithms have slightly better performance on the training dataset than on the test dataset. Therefore, these models can be used for daily forecasting of the target variable of the number of operations. However, with a model based on the Gaussian Processes algorithm, it is observed that the performance results are better on the test dataset than on the training dataset. This result indicates the problem of overfitting and this model cannot be used to predict.

**Table 4**. Performance of Prediction Models obtained using Test Dataset for the target variable of the number of operations

| Algorithm | r | MAE | RMSE | RAE % | RRSE % |
|---|---|---|---|---|---|
| LinearRegression | 0.9347 | 1548.149 | 1820.5441 | 49.8918 | 47.3206 |
| SMOreg | 0.9339 | 1851.84 | 2080.6675 | 59.6788 | 54.0819 |
| M5P | 0.9356 | 1499.91 | 1778.3824 | 48.3372 | 46.2247 |
| GaussianProcesses | 0.9221 | 1941.853 | 2242.5606 | 62.5796 | 58.2899 |

The graph in Figure 2 shows the ratio of actual and projected values of the target variable of the number of operations for the period from January to March 2021. The forecast of the target variable was made using three models with the best performance. Also, this graph shows that the forecast of the number of operations conducted on the test dataset closely follows the actual values of the number of operations.

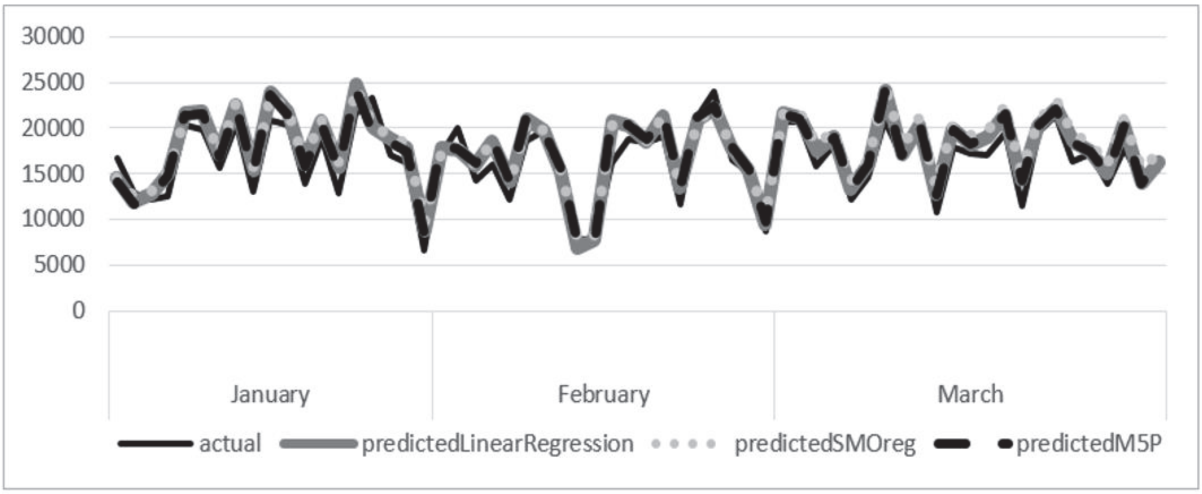



**Figure 2**. Actual and projected values of the number of operations on a daily basis

The graph shown in Figure 3 shows the percentage of absolute average errors in the forecast of orderpicking items and the number of operations by working days for the first quarter of 2021. Percentage absolute average error forecasts of orderpicking items are shown for the SMOreg algorithm because it has the smallest absolute error. It does not notice the trend of the largest error for the forecast of orderpicking items in the observed period, while the smallest error was on Wednesday or Thursday. Percentage average errors in the forecast of the number of operations are shown for the M5P algorithm because it has the lowest absolute error and the highest correlation coefficient. The largest percentage error in the forecast of the number of operations was mostly on Fridays, and the smallest on Wednesdays or Thursdays.

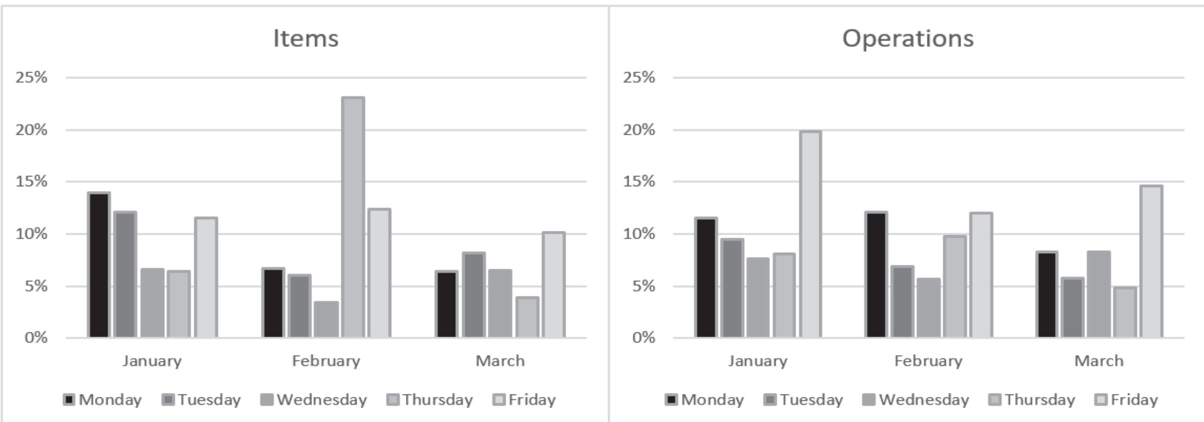



**Figure 3**. Display percentage errors forecasting orderpicking items and number of operations by working days

## 7 CONCLUSIONS

Forecasting demand and labor workload provides the possibility of more efficient implementation of warehousing operations. Planning based on accurate forecasts allows a high level of service for clients/customer and lower labor costs. The results of a case study show that the machine learning process can be of significant benefit to logistics companies in forecasting demand and planning the workforce. In this case study, the machine learning process for forecasting the workload, in terms of orderpicking items and number of operations, was conducted in the Weka software tool. On the training dataset and on the test dataset, models based on the SMOreg and M5P algorithms showed the best performance for forecasting orderpicking items. Further research direction should be focused on the determining orderpickers working hours based on the prediction of orderpicking items and operations, and specific daily norms for these measures.

## References


[1] Evangelos, S., Makridakis, S., Semenoglou, A. A. and Assimakopoulos, V. (2020). Comparison of statistical and machine learning methods for daily SKU demand forecasting. *Operational Research*. pp. 1-25

[2] Fischetti, T. (2018). *Data Analysis with R. Second Edition.* Packt Publishing.

[3] Franz, S. (2020). A Comparison of Machine Learning and Traditional Demand Forecasting Methods. Clemson University. All Theses. 3367.

[4] Kim, T. Y., Dekker, R. and Heij, C. (2017). Improving warehouse labour efficiency by intentional forecast bias. *International Journal of Physical Distribution & Logistics Management*, 4(1), pp. 93-110.

[5] Krishna, A. G. and Prabhu, V. V. (2016). Workforce Planning Models for Distribution Center Operations. In *IFIP International Conference on Advances in Production Management Systems*, Springer, Cham. pp. 206-213.

[6] Lee, Y. M., An, L. and Connors, D. (2008). Controlling Workforce Resource Actions for Demand Disturbances in Services Supply Chains. In *2008 IEEE International Conference on Service Operations and Logistics, and Informatics*, IEEE, pp. 1493-1499.

[7] McDonald, K., Fisher, S. and Connelly, C.E. (2017). E-HRM Systems in Support of "Smart" Workforce Management: An Exploratory Case Study of System Success. In *Electronic HRM in the smart era.* Emerald Publishing Limited, pp. 87-108.

[8] Mundschenk, M. and Drexl, A. (2007). Work force planning in the printing industry. *International Journal of Production Research*, 45(20), pp. 1-21.

[9] Petković, J.G. (2013). Razvoj modela tehnološkog predviđanja u preduzeću. Doktorski rad. Univerzitet u Beogradu – Fakultet organizacionih nauka.

[10] Popović, V., Kilibarda, M., Andrejić, M., Jereb, B. and Dragan, D. (2021). A New Sustainable Warehouse Management Approach for Workforce and Activities Scheduling. *Sustainability*, 13(4), pp. 1-19.

[11] Sainathuni, B., Parikh, P.-J., Zhang, X. and Kong, N. (2014). The warehouse-inventory-transportation problem for supply chains. *European Journal of Operational Research*, 237(2), pp. 690-700.

[12] Souza, G.C. (2014). Supply chain analytics. *Business Horizons*, 57(5), pp. 595-605.

[13] Van Gils, T., Ramaekers, K., Caris, A. and Cools, M. (2017). The use of time series forecasting in zone orderpicking systems to predict order pickers' workload. *International Journal of Production Research*, 55(21), pp. 6380-6393.
[14] Vanheusden, S., Van Gils, T., Braekers, K., Ramaekers, K. and Caris, A. (2021). Analysing the effectiveness of workload balancing measures in orderpicking operations. *International Journal of Production Research*, pp. 1-25.
[15] Witten, I.H. and Frank, E. (2005). D*ata Mining Practical Machine Learning Tools and Techniques,* Second Edition. Elsevier Inc.
[16] Zhu, X., Sherali, H.D. (2009). Two-stage workforce planning under demand fluctuations and uncertainty. *Journal of the Operational Research Society*, 60(1), pp. 94-103.

**Author's address**
Milica Mitrović, Teaching and Research Assistant
University of Belgrade, Faculty of Transport and Traffic Engineering, Department of Logistics
Vojvode Stepe 305, Belgrade
Republic of Serbia
email: milica.mitrovic@sf.bg.ac.rs

Dražen Popović, Assistant Professor
University of Belgrade, Faculty of Transport and Traffic Engineering, Department of Logistics
Vojvode Stepe 305, Belgrade
Republic of Serbia
email: d.popovic@sf.bg.ac.rs

Gordana Radivojević, Professor
University of Belgrade, Faculty of Transport and Traffic Engineering, Department of Logistics
Vojvode Stepe 305, Belgrade
Republic of Serbia
email: g.radivojevic@sf.bg.ac.rs

# PORTFOLIO SELECTION MODEL USING CVaR AND MDD RISK MEASURES

***Juraj Pekár, Ivan Brezina, Marian Reiff, University of Economics in Bratislava***

**Abstract**
Various decision-making models described in the literature can be used to determine the optimal investor strategy. This group of models may also include portfolio selection models that diversify assets to reduce the overall risk of the investment under consideration. In general, risk minimization and return maximization can be identified as the main objectives of the selection models. In minimizing risk, the investor may use several risk measures that complement and provide more excellent coverage of the risk investment to adequately reflect the corresponding asset's short-term and long-term capital shortfalls. The paper focuses on the description of the construction of a portfolio selection model that takes into account both aspects of investment risk. The risk measures conditional value at risk (CVaR), and maximum drawdown (MDD) have been chosen. The result of the model solution is a set of effective solutions for the determined parameters of the model, which reflect different combinations of the determined values of expected returns and risks.

***Keywords:*** *Portfolio Optimization, Risk Measure, CVaR, Drawdown*

***JEL Classification:*** G11, C60
***AMS Classification:*** 91B30, 90C90

## 1 INTRODUCTION

Portfolio optimization is crucial for investing available funds, as investing is one of the most effective ways to capitalize on available funds. From the nature of investing, it is clear that every investor seeks to maximize their return while minimizing risk. However, it is necessary to assume that higher profits also bring higher risk.
The risk represents the possibility that the actual return differs from the expected return, thus representing uncertainty about future income. Thus, investment risk is represented because the investment may not achieve the expected return, or it may even mean a loss. Although this risk cannot be avoided entirely, it can be monitored, optimized, and significantly reduced with appropriate tools.
This risk can be minimized by diversification, i.e., the risk of individual assets is spread over the portfolio and thus reduce this risk (Alexander 2009). Depending on the financial instruments themselves, the investment risk is influenced by various factors. The most common are: interest rate risk (in the case of bond and money investments), market risk (risk of changing the financial market situation), currency risk (in the case of foreign currency investments, it is the risk of loss), the purchasing power of a given currency during the conversion), liquidity risk (the investment will become very difficult or utterly non-tradable on the financial market at a given time), respectively credit risk (potential inability of a borrower to repay its obligations).
Portfolio selection models that diversify assets to minimize risk and maximize return generally present a two-criteria challenge. The article is focused on a problem when it is necessary to minimize losses for one period and minimize the maximum capital drop for the entire analyzed period. It is necessary to formulate a mathematical programming problem with three criteria. First, express the objectives of maximizing expected return. Second, minimize short-term losses using the selected risk measure, conditional value at risk (CVaR). Third, minimize possible

long-term capital outflows using the selected risk measure maximum drawdown (MDD). The result will be a set of effective solutions for the specified parameters of the model, which reflect various combinations of the set values of expected returns and risks (Pekár 2015).
In the first part of the article, we define the method of calculating the input parameters of the model and the rates of expected returns and risks measures used in the construction of the model. Subsequently, we define a model used to generate efficient portfolios taking into account the three criteria.

## 2 RISK MEASURE AND EXPECTED RETURN

In this section, we present the method of calculating the model input data as yield vector, cumulative return vector, and the method of obtaining the expected return value of the asset. Also, we define the selected risk measures maximum drawdown (MDD) and conditional value at risk (CVaR), which are used in our formulation of the mathematical programming model.
Let us consider having $n$ assets and let $P_{jt}$ represent the price of a $j$-th asset, $j$ = 1, 2,...$n$, in time $t$ =1, 2,...$T$, where $T$ is the length of the observed time period.

The rate of the return of a $j$-th asset, $j$ = 1, 2,...$n$, between time $t$ and $t$ –1 can be expressed as a relative rate of return $r_{jt}$ and is expressed by equation:

$$r_{jt} = \frac{P_{jt} - P_{jt-1}}{P_{jt-1}} \tag{1}$$

Subsequently, the cumulative return rate of $j$-th asset, $j$ = 1, 2,...$n$, in time $t$ can be expressed by equation:

$$y_{jt} = \frac{P_{jt} - P_{j1}}{P_{j1}} \tag{2}$$

Expected return of $j$-th asset, $j$ = 1, 2,...$n$, based on the value of the cumulative return is

$$E_j = \left(1 + y_{jT}\right)^{\frac{1}{T}} - 1 \tag{3}$$

Drawdown value $DD_j(t)$ of $j$-th portfolio asset, $j$ = 1, 2,...$n$, in time $t$, $t$ =1, 2,...$T$, can be calculated as the difference of the maximum value in $v$-th time $v$ = 1,2 ,...$t$ and the value in time $t$. According to (Grossman and Zhou 1993), (Cheklov, Uryasev and Zabarankin 2003), (Brezina, Pekár and Brezina jr. 2018), (Pekár, Brezina and Brezina jr. 2018) it can be calculated as follows

$$DD_j\left(t\right) = \max_{v=1,2,\ldots t} y_{jv} - y_{jt} \tag{4}$$

In addition to the considered indicator $DD_j(t)$, we will also deal with the derived risk measure [2], the maximum drawdown (MDD). Because, naturally, the investor is primarily interested in the probability expressing that the investment in the asset will not reach the expected return, it is the maximum risk level that drawdown allows for observing the most significant drop in the value of the corresponding asset. This indicator is based on historical asset price values and describes the most significant decrease in the value of an asset during the reference period ($T$). Maximum drawdown is the most significant cumulative percentage decline in the net asset value of a portfolio from the highest or peak value to the lowest or trough value after the peak. Maximum drawdown $MDD_j\left(T\right)$ for $j$-th asset, $j$ = 1, 2,...$n$, can be calculated as follows

$$MDD_j\left(T\right) = \max_{t=1,2,\ldots T} DD_j\left(t\right) \tag{5}$$

As mentioned in the introduction, the model, in addition to a long-term capital collapse (MDD), also uses the indicator of short-term capital decline, represented by the currently most used measure of risk conditional value at risk. Its calculation is based on the level of VaR risk, which is generally defined as the maximum possible loss over a given period of time over a certain confidence interval α (Yamai and Yoshiba 2002), (Alexander 2009). Then, based on the VaR value, the CVaR value can be defined as the average of losses higher than the VaR value. Formulation of the equation for the calculation of CVaR, which during the period under review ($T$) for $j$-th asset, $j = 1, 2, \ldots n$, can be defined (Rockafellar and Uryasev 2000), (Pekár 2015) in the following manner

$$CVaR_\alpha\left(\mathbf{r}_j\right) = VaR_\alpha - \frac{1}{\alpha} E\left[\left|\mathbf{r}_j + VaR_\alpha\right|_-\right] \tag{6}$$

where $\mathbf{r}_j$ represents the vector of returns of $j$-th asset and α represents the level of significance.

# 3 PORTFOLIO OPTIMIZATION

This section presents the construction of the mathematical programming model based on the expected return and CVaR and MDD risk measures. In the first step, we define the model variables and the parameters used in its construction. The solution results to this model will be efficient portfolios that take into account two levels of risk reflecting short-term and long-term capital outflows in contrast to traditional portfolio models (Pekár, Čičková and Brezina 2016), (Pekár, Brezina and Brezina jr. 2018).

When constructing the model, we will use the variables $z_t \geq 0$ ($t = 1, 2, \ldots T$), which acquire the value of the difference between VaR and the portfolio return in state $t$, if the return is lower than VaR, or will be equal to zero based on the significance level $\alpha$. Let us further assume that the vector $\mathbf{u} = (u_0, u_1, \ldots u_T)$ of free variables represents the maximum value of the accumulated return $u_t$ in the time 0, 1, ... $T$. Let $VaR_\alpha$ be a variable that represents the threshold for portfolio return based on the significance level $\alpha$. In addition to the mentioned free variables, the variables $w_1, w_2, \ldots w_n$, also appear in the model, which acquires optimal values of the weights of individual assets ($j = 1, 2, \ldots n$). Except to model variables definition, the input parameters given by the investor need to be also defined. In addition to the returns, accumulated returns, and expected returns of individual assets, the values of the $MDD_p$ parameters must be determined. Which represents the maximum allowable value of the portfolio capital loss, i.e. the value determined by the investor below which the loss must not fall. Subscript $p$ indicates that this is an input parameter. Parameter $Ep$ indicates the minimum expected return of the portfolio, i.e. the value determined by the investor that the calculated portfolio must reach, subscript $p$ indicates that this is an input parameter. By combining these two parameters, a set of effective portfolios can be obtained that provide the investor with proposals for investment strategies with different values of required returns and risk expressed by MDD.

The three-criteria task (all three formulated goals are respected) can be transformed into a mathematical programming problem minimizing the CVaR risk objective function with regard to the constraints that will model the maximum loss of capital allowed and the minimum required value of the expected return:

$$\min CVaR\left(w_1, w_2, ...w_n, z_1, z_2, ...z_T, u_0, u_1, ...u_T, VaR_\alpha\right) = VaR_\alpha + \frac{1}{\alpha T}\sum_{t=1}^{T} z_t$$

$$z_t + \sum_{j=1}^{n} r_{jt} w_j + VaR_\alpha \geq 0, t = 1,\ 2, ...T$$

$$\sum_{j=1}^{n} E_j w_j \geq E_P$$

$$u_t - \sum_{j=1}^{n} y_{jt} w_j \leq MaxDD_p, t = 1, 2, ...T$$

$$u_t \geq \sum_{j=1}^{n} y_{jt} w_j, t = 1, 2, ...T \tag{7}$$

$$u_t \geq u_{t-1}, t = 1, 2, ...T$$

$$u_0 = 0$$

$$\sum_{j=1}^{n} w_j = 1$$

$$w_1, w_2, ...w_n \geq 0, z_1, z_2, ...z_T \geq 0$$

# 4 CONCLUSION

The paper deals with modifying the portfolio selection model based on the criteria of maximizing expected return, risk minimization based on CVaR rates and maximum drawdown, in which, unlike models known from the literature, we consider three criteria. In contrast to known approaches, this model considers two levels of risk, which will ensure more significant reduction in the overall risk of the investment. Testing of this model was performed in the Python programming language, and an exciting part, which we do not discuss in the paper, is to determine the values of the maximum drawdown model and expected yield, which consists of solving a system of problems ensuring the subsequent feasibility of the problem. The result of the whole process is to obtain effective solutions for the model in the Python programming environment, MIP module, which documents the possibility of calculating the solution, which serves as the base for the investor's decision.

**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] Alexander, C. (2009). *Market risk analysis, value at risk models*. John Wiley & Sons.

[2] Brezina, I., Pekár, J. and Brezina, I. jr. (2018). Manažment portfólia na základe miery výkonnosti DrawDown. In *Competition: Proceedings of 10th Annual International Scientific Conference*, 17th–18th May, 2018, Jihlava: College of Polytechnics Jihlava.

[3] Cheklov A, Uryasev S and Zabarankin M. (2003). Portfolio optimization with drawdown constraint. In: Scherer B (ed) *Asset and liability management tools. Risk Books*, London, pp. 263–278.

[4] Grossman, S. J. and Zhou, Z. (1993). Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, 3(3), pp. 241–276.

[5] Pekár, J. (2015). *Modely matematického programovania na výber portfólia*. Bratislava: Vydavateľstvo EKONÓM.

[6] Pekár, J., Brezina, I. and Brezina, I. jr. (2018). Portfolio Selection Model Based on Drawdown Performance Measure. In *International Conference. Mathematical Methods in Economics - MME 2018: Proceedings from the 36th International Conference*, Jindřichuv Hradec, Czech republic, September 12-14, 2018. Praha: Matfyzpress.
[7] Pekár, J., Čičková, Z. and Brezina, I. (2016). Portfolio performance measurement using differential evolution. *Central European Journal of Operations Research*, 24(2), pp. 421-433.
[8] Rockafellar, T. R. and Uryasev, S. (2000). Optimization of Conditional Value-at-Risk. *The Journal of Risk*, 2(3), pp. 21-42.
[9] Yamai, Y. and Yoshiba, T. (2002). On the validity of value-at-risk: comparative analyses with expected shortfall. *Monetary and economic studies*, 20(1), pp. 57-85.

**Author's address**

prof. Mgr. Juraj Pekár, PhD.
University of Economics in Bratislava,
Faculty of Economic Informatics
Department of Operations Research and Econometrics
Dolnozemská 1/b, 852 35 Bratislava
Slovakia
Email: juraj.pekar@euba.sk

prof. Ing. Ivan Brezina, CSc.
University of Economics in Bratislava,
Faculty of Economic Informatics
Department of Operations Research and Econometrics
Dolnozemská 1/b, 852 35 Bratislava
Slovakia
Email: ivan.brezina@euba.sk

doc. Ing. Marian Reiff, PhD.
University of Economics in Bratislava,
Faculty of Economic Informatics
Department of Operations Research and Econometrics
Dolnozemská 1/b, 852 35 Bratislava
Slovakia
Email: marian.reiff@euba.sk

# VEHICLE ROUTING PROBLEM WITH OPTIONAL NODES

***Jan Pelikán, Prague University of Economics and Business***

**Abstract**
In the standard vehicle routing problem (VRP) a list of nodes with their transport requirements is given and optimal routes are found to meet all nodes requirements. The paper addresses VRP in which some or all nodes are optional (except for a depot). In a lot of applications of this problem, the requirements for the transport of items from depot come continually in time and they allow to delay the delivery. Therefore, there is no need to deliver this packet immediately at the time the requirement arrives. In this case we need to decide which nodes will be in these routes and which not and to search for the optimal routes throughout the selected nodes. Two objective functions are proposed for VRP with optional nodes. The first objective function represents the total gain due the difference in revenues from realized requirements of transport and costs for vehicle routes. The second objective function is a revenue and a costs ratio. Both functions are maximized and interpreted on a numerical example.

***Keywords:*** *Vehicle routing problem, Integer programming, Heuristic methods*

***JEL Classification:*** JEL C44
***AMS Classification:*** 90C15

## 1 INTRODUCTION

VRP consists in searching for optimal routes for a vehicle with given capacity. The aim is to minimize the total length of routes so that the requirements of the nodes for transfer of given amount of items from depot are fulfilled (Bräysy and Gendreau , 2005; Laporte, 1992).

In (Pelikán, 2019) the VRP was solved with a condition that the requirements for transfer into the nodes were divided into the urgent requirements, which have to be realized immediately, and others, which might be delayed into other days.

In this article we will not be working with urgent requirements. Therefore, all the requirements are optional that are or are not necessary to be included into the routes of vehicle. Nevertheless, there can be an added condition into the suggested model such that some nodes are urgent. A list of nodes is given with their requirements and distances between the nodes. We will work in this article with costs and revenue of solutions, where costs are made up by total length of routes multiplied by the unit costs of distance (e.g. km), revenue is derived from total volume of transport multiplied by revenue from one unit of transferred items.

This problem is solved in (Pelikán and Jablonský, 2020) with objective function, which is a ratio of revenue and costs and therefore consisting the index of revenue on unit of cost (alternatively, we can work with the ratio of transported items and total length of routes, therefore the average volume of transported items on a unit of length of routes). This function is non-linear with linear-fractional objective function. The model with this non-linear objective function was transformed by using the Charles-Cooper method into the linear model with binary variables.

The next objective function is a total profit made by the difference of total revenue and total costs. This function is linear and maximized.

We will work with both non-linear and with linear objective functions and therefore with two VRP models with optional nodes. Both models described in this article are illustrated on a numerical example. It is shown that the optimal solution may differ between the first and the second objective functions.

# 2 MATHEMATICAL MODELS

## 2.1 The first mathematical model

The objective function in this model will be the difference between revenue and costs. All nodes except for depot are optional. Costs will be made by the total length of routes multiplied by costs on a unit of length of the route (e.g. km), which come out of the type of a vehicle. Revenue of the route is the volume of transported items into selected nodes multiplied by revenue on a unit of items.

*Parameters of the model:*
$n$ number of nodes, node 1 is depot,
$d_{ij}$ distance between node $i$ and node $j$,
$q_i$ demand of node $i$,
$c$ unit costs (costs per unit of distance),
$r$ unit revenue (revenue of unit of delivered goods)
$W$ capacity of vehicle.

*Variables of the model are:*
$x_{ij}$ binary, equals 1 if a vehicle travels from node $i$ to node $j$,
$u_j$ variables in anti-cyclic constraints.

Linear objective function $z(x)$ is the difference between total revenue and total costs (1).
Equation (2) means condition: if a vehicle enters a node, it has to leave it. Anti-cyclic conditions are in (3). Inequality (4) assures that the capacity of vehicles is not exceeded.

$$z(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} r\, q_i\, x_{ij} - \sum_{i=1}^{n} \sum_{j=1}^{n} c\, d_{ij} x_{ij} \qquad \rightarrow\ max \tag{1}$$

$$\sum_{i=1}^{n} x_{ij} = \sum_{i=1}^{n} x_{ji} \quad j = 1,2, \dots, n \tag{2}$$

$$u_i + q_j - W\left(1 - x_{ij}\right) \le u_j \qquad i = 1,2, \dots, n, j = 2,3, \dots, n, i \ne j \tag{3}$$

$$u_j \le W \quad j = 2,3, \dots, n \tag{4}$$

$$x_{ij} \quad i, j = 1,2, \dots, n, \qquad i \ne j \ \ binary \tag{5}$$

## 2.2 The second mathematical model.

In this model the constraints remain the same as in the first model, only the objective function is a ratio of revenue and costs. The second model is (1’) and (2), (3), (4), (5).

$$I(x) = \frac{\sum_{j=1}^{n} r\, q_i\, x_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} c\, d_{ij} x_{ij}} \qquad \rightarrow max \tag{1'}$$

The objective function *I(x)* (1') is a ratio with numerator of total amount of loads of all routes multiplied by unit revenue $r$ (i.e. total revenue of delivered goods) and denominator of total length of all routes multiplied by unit costs $c$ (i.e. total costs of all routes).
The second model is not linear in its objective function, but can be transformed into a linear program rather easily using Charnes-Cooper transformation. Non-linear model is solved by using the Charles-Cooper method (Barros, 1998; Martos, 1975).

## 3 COMPARISON OF OPTIMAL SOLUTIONS OF BOTH MODELS

In this paragraph we will show the difference in optimal solutions of both models. First, we need to state some notation.

**Notation:**

$\boldsymbol{X}$ is the set of feasible routes through a nonempty subset of nodes from the set $\{2,3,\ldots,n\}$,
$f(X)>0$ the revenue of route $X \in \boldsymbol{X}$,
$g(X)>0$ the costs of route $X \in \boldsymbol{X}$,
profit $z(X)=f(X)-g(X)$,
$I(X)= f(X)/g(X)$,
$X_0$ maximizes $I(X)$ on $\boldsymbol{X}$,
$X'$ maximizes $z(X)$ on $\boldsymbol{X}$.

**Proposition.**

a) If $z(X') > 0$ then $g(X') \geq g(X_0)$ and $f(X') \geq f(X_0)$,
b) If $z(X') < 0$ then $g(X') \leq g(X_0)$ and $f(X') \leq f(X_0)$.

Providing $\frac{f(X_0)}{g(X_0)} > \frac{f(X')}{g(X')}$, it holds

c) If $z(X') > 0$ $then$ $g(X') > g(X_0)$ $and$ $f(X') > f(X_0)$ (costs and revenue of X' are higher than costs and revenue of $X_0$ ),
d) if $z(X') < 0$ $then$ $g(X') < g(X_0)$ and $f(X') < f(X_0)$ (costs and revenue of X' are lower than costs and revenue of $X_0$ ),
e) if $z(X')=0$ then $I(X_0)=I(X')=1$ and $X_0$ and X' are optimal for both objective functions $z(X)$ and $I(X)$.

**Proof.**

Because $X_0$ maximizes the function $I(X)$ on $X$, it holds

$$\frac{f(X_0)}{g(X_0)} \geq \frac{f(X')}{g(X')} \tag{6}$$

We easily see that from (6) follows next inequalities:

$$\frac{f(X_0)}{g(X_0)} - 1 \geq \frac{f(X')}{g(X')} - 1$$

$$\frac{f(X_0) - g(X_0)}{g(X_0)} \geq \frac{f(X') - g(X')}{g(X')}$$

$$\frac{f(X_0)-g(X_0)}{g(X_0)} \geq \frac{f(X')-g(X')}{g(X')}$$

$$f(X')-g(X') \geq f(X_0)-g(X_0) \geq \frac{g(X_0)}{g(X')}\big(f(X')-g(X')\big). \quad (7)$$

From (7) it follows two cases: a) and b).

**Case a):**

If $z(X') = f(X')-g(X') > 0$ then it follows from (7) by dividing it by $z(X') = f(X')-g(X')$ the inequality $\frac{g(X_0)}{g(X')} \leq 1$ and $g(X_0) \leq g(X')$ and

$$f(X')-g(X') \geq f(X_0)-g(X_0) \geq f(X_0)-g(X')$$

and finally $f(X') \geq f(X_0)$.

Costs at the solution $X_0$ are not higher than at *X*' (they can be lower see c)) and the same for revenue.

**Case b):**

If $z(x') = f(X')-g(X') < 0$ then from (7) it follows by dividing it by $z(X') = f(X')-g(X')$ that $\frac{g(X_0)}{g(X')} \geq 1$ and $g(X_0) \geq g(X')$. Then

$$\frac{f(X_0)}{g(X_0)} \geq \frac{f(X')}{g(X')} \geq \frac{f(X')}{g(X_0)}$$

hence $f(X') \leq f(X_0)$.

Costs at the solution $X_0$ are not lower than at *X*' and the same for revenue.

Cases c) and d) follow from a) and b) such that in (6) we assume a strong inequality.

Case e) is trivial. QED.

Note. If two routes *X* and *Y* contain the same subset of nodes, then *f(X)* = *f(Y)*, then they can differ only in the lengths of routes and therefore in the values of *g(X)* and *g(Y)*.

## 4 NUMERICAL EXAMPLE

Both mathematical models were verified on an illustrative example. Consider 11 nodes where node 1 is a depot. Capacity of each vehicle is *W*=100. The requirements of the nodes are *q* = ( 0, 5, 20, 10, 20, 85, 65, 30, 20, 70, 30 ). The distance matrix *D* is as below:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 13 | 6 | 55 | 93 | 164 | 166 | 168 | 169 | 241 | 212 |
| 13 | 0 | 11 | 66 | 261 | 175 | 177 | 179 | 180 | 239 | 208 |
| 6 | 11 | 0 | 60 | 97 | 168 | 171 | 173 | 174 | 239 | 209 |
| 55 | 66 | 60 | 0 | 82 | 113 | 115 | 117 | 117 | 295 | 265 |
| 93 | 261 | 97 | 82 | 0 | 113 | 115 | 117 | 118 | 333 | 302 |
| 164 | 175 | 168 | 113 | 113 | 0 | 6 | 7 | 2 | 403 | 374 |
| 166 | 177 | 171 | 115 | 115 | 6 | 0 | 8 | 7 | 406 | 376 |
| 168 | 179 | 173 | 117 | 117 | 7 | 8 | 0 | 3 | 408 | 378 |
| 169 | 180 | 174 | 117 | 118 | 2 | 7 | 3 | 0 | 409 | 379 |
| 241 | 239 | 239 | 295 | 333 | 403 | 406 | 408 | 409 | 0 | 46 |
| 212 | 208 | 209 | 265 | 302 | 374 | 376 | 378 | 379 | 46 | 0 |

Price of a unit of delivered items is stated as *r*=11 and costs for a unit of a route as *c*=3.

In Tab. 1 there are given costs and revenue at 10 chosen routes. Columns are numbered by numbers 1 – 10. Each column represents a specific subset of nodes and routes over these nodes and hence corresponding profits *f(X)* and costs *g(X)*. Based on profits and costs there is calculated a value of function *z(X)* and *I(X)*. It is obvious that if *I(X)* > 1, then the profit *z(X)* > 0 and on the contrary for *I(X)* < 1, then the profit *z(X)* < 0, for zero profit *z(X)* = 0, the costs equal to revenue and value *I(X)* is equal to 1.

Column 5 represents a route with maximal value of profit of 253, column 10 represents the solution with the highest value of function *I(X)*= 6.11. At this solution, the profits are 6,11 times higher than the costs.

On the Fig. 1 there are points marked as values of profits and costs in a form of a point [*g(X), f(X)*] at the solutions presented in a Tab. 1. Contour lines of the function *z(X)* are parallel lines with axis of the first quadrant. Contour lines of the function *I(X)* are straight lines passing through the origin.

**Tab 1.** A subset of feasible routes.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| f(x) | 3905 | 3685 | 3465 | 2585 | 2365 | 1320 | 1320 | 1265 | 275 | 275 |
| g(x) | 4713 | 4161 | 3609 | 2664 | 2112 | 1125 | 1086 | 1032 | 90 | 36 |
| I(x) | 0,82 | 0,88 | 0,96 | 0,97 | 1,11 | 1,17 | 1,21 | 1,22 | 3,05 | **6,11** |
| z(x) | -808 | -476 | -144 | -79 | **253** | 195 | 234 | 233 | 185 | 184 |

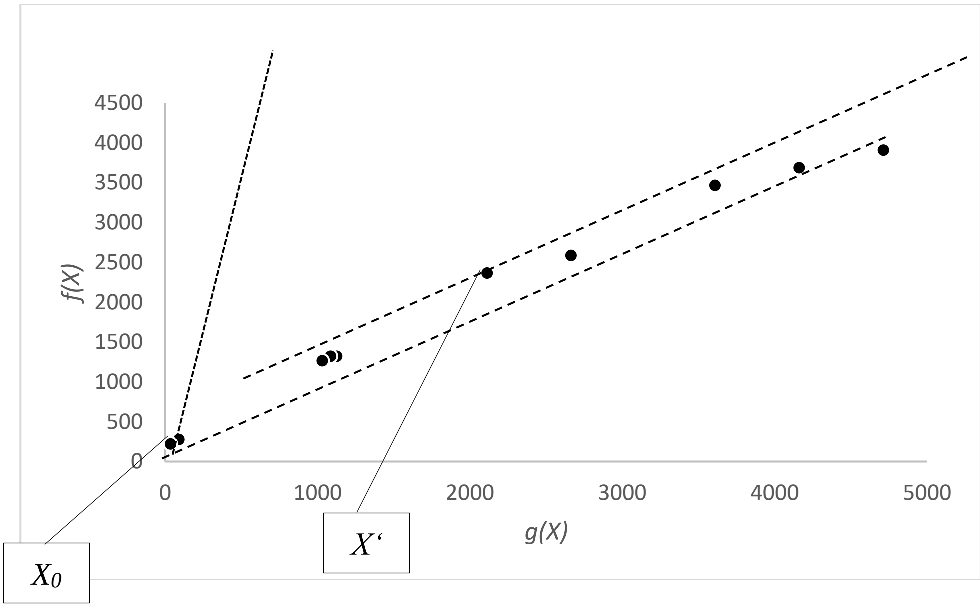



**Fig. 1.** Revenue-costs graph

# 5 CONCLUSION

In the paper a VRP with optional nodes is solved. There are suggested two types of objective function, one linear and the other one non-linear. The optimal solutions are analysed according to the first and the second objective function and everything is illustrated on a numerical example.

## References

1. Barros, A. I. (1998). *Discrete and fractional programming techniques for location models.* Combinatorial Optimization. 3. Dordrecht: Kluwer Academic Publishers.
2. Bräysy O. and Gendreau M. (2005). Vehicle routing problem with time windows, Part I: route construction and local search algorithms. *Transportation Science,* 39(1), pp. 104-118.
3. Laporte G. (1992). The vehicle routing problem: an overview of exact and approximate algorithms. *European Journal for Operations Research*, 59 (3), pp. 345-358.
4. Martos, B. (1975). Nonlinear programming: *Theory and methods.* Amsterdam-Oxford: North-Holland Publishing Co.
5. Pelikán, J. (2019). VRP with loading time window. In: *37th International Conference on Mathematical Methods in Economics 2019.* České Budějovice : Ekonomická fakulta, Jihočeská univerzita v Českých Budějovicích, pp. 20–24.
6. Pelikán, J. and Jablonský, J. (2020). Nonlinear vehicle routing problem. In: *HED 2020 – Hradec Economic Days.* Hradec Králové, University of Hradec Králové, pp. 593–597.

## Author's address

Prof. RNDr. Jan Pelikán, CSc..
Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky
Nám. W. Churchilla 4, 13067 Praha 3
Czech Republic
email: pelikan@vse.cz

# MEASURING THE EFFICIENCY OF THE SLOVAK FIRST AND SECOND LEAGUE FOOTBALL CLUBS

*Natalie Pelloneová, Technical University of Liberec*

**Abstract**
Data envelopment analysis (DEA) is widely used for efficiency assessment in many sports, specifically football. This study applies the DEA method to measure the efficiency of the Slovak first and second league football clubs in the season 2020/21. In this context, each football club is a decision-making unit (DMU), where two inputs are selected: the number of players and the total squad market value. The total points obtained during the season 2020/21 is the output. The input-oriented CCR and BCC model evaluate the football clubs' efficiency. Overall, the research results show that Slovak clubs are inefficient and use more resources than necessary or choose a bad strategy to achieve their goals. The analysis results also show that the 1st and 2nd leagues winners were always labelled as efficient DMUs during the period studied. The applied approach can help football club managers identify relevant weaknesses and focus on efficiency strategies in the future.

***Keywords:*** *Football, Efficiency, Sport, Data envelopment analysis, Football club efficiency*

***JEL Classification:*** L83, C67, C44
***AMS Classification:*** 90C08, 90C15

## 1 INTRODUCTION

Football is one of the most popular team sports, played by more than 250 million registered players in more than 200 countries. Hundreds of millions of people play football without being registered, simply as a leisure activity. Football is the most television viewed sport worldwide (SportSen, 2022). By playing football, society strengthens its national pride, the feeling of togetherness, and last but not least, it brings joy to people. Football has been popular in Slovakia for many decades. Generally speaking, it is a top-rated sport in Slovakia, which, in 2010, was played (professionally or not) by almost 11 % of the inhabitants (i.e. 622 thousand people). That fact placed Slovakia among the first ten countries out of 130 in football popularity. There is a relatively high number of football clubs and football players compared to the number of inhabitants (Machlica and Šiškovič, 2010). In comparison with other countries, Slovakia has been achieving relatively above-average results. The Slovak national football team was ranked 17-29th in the UEFA international ranking in 2020 (Melek and Dedík, 2020).

Some football clubs make significant financial profits, while others do not; in fact, they do the opposite. In this case, it is essential to ask how a particular club manages its resources while achieving its outcomes. Currently, it seems appropriate, especially in the context of the worldwide COVID-19 pandemic, which has significantly impacted all football clubs, to focus on the professional football clubs' efficiency. In particular, attention should be paid to determining whether a club utilises its resources efficiently to achieve its best results. One can imagine that a club whose players are relatively cheap can get "interesting" results. A low-budget club can be viewed as efficient if the players' market prices are in line with the club's performance. This point, in particular, will be addressed in this paper. The paper aims to analyse the efficiency of the Slovak first and second league football clubs, whose financial resources are considerably different. The study intends to discuss whether the clubs' resources are allocated efficiently regarding their achieved sports results. The structure of the paper is as follows. Firstly, the literature review focuses on authors who measure and describe efficiency in the football sphere. Secondly, research methodology and data are described. Thirdly, the obtained results and their analysis are presented along with the discussion.

## 2 LITERATURE REVIEW

One of the most valuable tools to help football managers make better decisions and become more efficient in using their current resources is the awareness of efficiency and effectiveness. Data envelopment analysis (DEA) is a popular non-parametric method used to measure the relative efficiency of similar decision-making units (DMUs). DEA is frequently applied to evaluate efficiency in various team sports, especially in football. Most of the authors using the DEA method in football focus primarily on the teams' performance during given football seasons. Professional foreign research focuses on two main areas.

First, a group of authors evaluate football clubs' efficiency using economic indicators (e.g. Barros and Garcia-del Barrio, 2011). They assess Spanish football clubs' efficiency by applying a two-stage data envelopment analysis procedure proposed by Simar and Wilson. The inputs to the model are: operating cost, total assets, and team's payroll. The output variables are: the team's audience attendance and other receipts. Second, a group of authors, e.g. Petrović Djordjevic, Vujosevics and Martic (2015) assesses the football clubs' efficiency using sports variables. They analyse the technical efficiency of the Croatian national football team during the World Cup qualifications in 2010 using the non-parametric output orientated DEA model. The output variables are: effect of successful ones out of total crosses, the effect of successful ones out of the total crosses in play, the effect of successful ones out of the total passes in play, the effect of successful ones out of the total number of air duels, the effect of successful ones out of the total number of ground duels, the number of shots, number of shots on target, the total number of fouls made and number of successful dribbling. The output variable is the total number of recognised scored goals. Espitia-Escuer and García-Cebrián (2010) evaluate the football teams in the Champion League in 2003-2007 using DEA. They include the number of players, the number of attacking moves, the number of minutes the teams had possession of the ball and the number of shots and headers into their input model. The output variable is the number of scored goals.

The presented research applies an approach that uses both economic and sports variables. This approach is pursued by the majority of the authors. Barros and Leach (2006) conduct similar research using DEA. They evaluate the performance of the Premier League football clubs in the seasons 1998/99-2002/03, whilst combining the sports and financial variables. Barros, Assaf and Sa-Earp (2010) analyse the technical efficiency of the Brazilian first league football clubs. Their variables are sport-related and financial. They conclude that the Brazilian clubs are very inefficient. Soleimani-Damaneh, Hamidi and Sajadi (2011) take advantage of a hybrid approach combining DEA and AHP to analyse the Iranian football premier league teams' performance in the 2009/10 season. The input variables are each team's fixed assets, the players, the coach, and the staff's wages. The output variables include the points received by the team, the number of spectators watching the team's matches and the team's income at the end of the season. The research results have indicated that seven teams (i.e. 39 %) were efficient in the 2009/10 season. Ribeiro and Lima (2012) use the DEA method to measure the efficiency of Portuguese first league football clubs during the 2002/03 to 2008/09 seasons. The research supports the fact that some clubs spend more money on the players' wages than they should, their higher expenses do not correlate with the efficient utilisations of their resources. Miningou and Vierstraete (2012) use the DEA model to examine the effectiveness of French football clubs in Ligue 1 and Ligue 2 in the 2002/03 and 2007/08 seasons. The input variables are compensation of employees, other expenses, Europe categorical variable, league category variable. The output variables are the attendance in %, number of points. Miningou and Vierstraete (2012) conclude that the French teams are inefficient and use more sources than is necessary to achieve their goals. The Ligue1 teams are less efficient than the Ligue 2 clubs; the efficiency decreases with time. Chaiwuttisak (2018) evaluates the sport and financial performance of Thai Premier League football clubs during the 2014/15 season with the help of the DEA model. He uses the CCR model and the super-efficiency model. The input variables are the stadium capacity, the capital investment, and the administrative costs. He uses nine variables, the average number of spectators, the total number of scores in 2015, the total sales, net assets, the number of trophies, the qualification for AFC for the last and the following season, the qualification for the Thai Premier League for the previous and the next season. Keskin and Öndes (2020) analyse the sport and financial performance of 50 big football clubs in 10 European countries. The efficiency of clubs is examined in two phases. The first phase explores the clubs' technical performance scores in the 2007/08 to 2017/18 seasons. The second phase identifies the causes of the efficiency differences with the help of a random effect panel Tobit model.

# 3 METHODOLOGY AND DATA

The paper aims to use the data envelopment analysis to calculate the degree of technical efficiency. Consequently, it evaluates and compares the Slovak 1$^{st}$ and 2$^{nd}$ leagues football clubs' efficiency during the 2020/21 season. The efficiency was assessed only based on the primary data from the regular season of the two investigated football leagues.

## 3.1 Data set

The research in the first part of the paper focuses on evaluating the efficiency of the top Slovak football league called Fortuna Liga. Twelve football clubs participated in the Slovak highest football competition in the 2020/21 season. The second part of the paper evaluates the efficiency of the Slovak football clubs playing in the 2$^{nd}$ Slovak league in the 2020/21 season, where15 clubs participated.

As mentioned above, the research focused only on the regular season of the league season. It did not include any playoffs/payouts. The Fortuna Liga and the 2$^{nd}$ Slovak League operate on the classical basis when all teams play all other teams both home and away. In the regular season of the Fortuna Liga, 22 games were played in the 2020/21 season. In the regular season of the 2$^{nd}$ Slovak league, 28 matches were played in the same season. Teams in both leagues play each other twice, once at home and once away, earning three points for a victory and one point for a draw.

The data used for the purposes of the research were obtained from the official database of the Fortuna Liga (Unia ligovych klubov and eSports, 2022) and are supplemented by the information from the databases of companies that operate in the football environment, mainly from the companies EuroFotbal (2022) and Transfermarkt (2022).

## 3.2 Data Envelopment Analysis

The input-oriented CCR-I and BCC-I model was used to calculate the technical efficiency of each club in the 1$^{st}$ and 2$^{nd}$ Slovak football leagues in the 2020/21 season. The DEA methodology was first introduced by Charnes et al. (1978). DEA is a non-parametric technique that helps estimate a decision-making unit's (DMU) technical efficiency. The relative efficiency of each DMU depends on its ability to improve its performance (outputs) or reduce its resource consumption (inputs) under certain constraints that reflect the performance of other units. This approach is frequently used due to its flexibility. In addition, the DEA method can be used in a situation of multiple inputs and outputs. Among other things, the DEA method also provides sample unit values for improving the inefficient DMUs. The DEA method constructed a weighted average output-to-input ratio for each decision unit. The weights are chosen by the method itself when solving a linear programming. The weight vectors are set to be the most advantageous for the evaluated DMU (Charnes et al., 1985). In the DEA method, the efficiency score is limited by the interval 0 to 1. Score 1 is achieved by an efficient unit (in this case, a football club).

The input-oriented models help determine how much inputs must be reduced for an evaluated unit to become efficient. The output-oriented models quantify how much the outputs need to be increased for the evaluated unit to become efficient. Another decision that should be made before measuring efficiency is the selection of so-called returns to scale according to the technology used in the sector. According to the returns to scale, two primary models can be distinguished – the CCR and the BCC models. The CCR model works on the assumption of constant returns to scale, which calculates the overall technical efficiency (OTE) score regardless of the returns to scale. The input-oriented CCR model evaluates the efficiency by solving the following linear programming problem. Where $\lambda_j$, $j$ = 1, 2, ..., $n$ are weights of all DMUs, $s_i^-$, $i$ = 1, 2, ..., $m$ and $s_k^+$, $k$ = 1, 2, ..., $r$ are slack/surplus variables, $\theta_q$ is the efficiency score of the DMU$_q$.

$$\textit{Minimize} \qquad \theta_q \tag{1}$$

$$\begin{aligned} S.t. \qquad & \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_q x_{iq}, \qquad i = 1,2,\dots,m, \\ & \sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq}, \qquad k = 1,2,\dots,r, \\ & \lambda_j \geq 0, \;\; j = 1,2,\dots,n, \end{aligned} \tag{2}$$

The BCC model works on the assumption of variable returns to scale. The convexity condition supplements the previous model $\sum_{j=1}^{n} \lambda_j = 1$. The BCC model provides pure technical efficiency (PTE) scores. When both models are compared by considering the production function of a given DMU situated on the efficiency frontier of the BCC and CCR models, it is possible calculate the scale efficiency (SE), see (3). Generally, if SE = 1 means that the DMU operate at optimal scale. Otherwise, if SE < 1, the DMU doesn't work in the most productive scale size (Guzmán-Raja and Guzmán-Raja, 2021).

$$SE = \frac{OTE}{PTE} \tag{3}$$

However, the DEA method also has disadvantages. The main disadvantage of this method is the fact that the efficiency score is relatively sensitive to the number of inputs and outputs involved. The efficiency score can be overestimated when the number of inputs and outputs is too high concerning the number of DMUs (Othman et al., 2010). According to Dyson et al. (2001), the number of inputs and outputs should be at least three times smaller than the number of DMUs. Two input and one output variables were included in the DEA model due to the number of DMUs (the lower limit is 12 clubs). Three variables were included in the efficiency model for 12 football clubs (DMUs). Such a model has sufficient discriminating power to calculate the subsequent efficiency scores. Individual variables were selected using correlation analysis. The number of players is the first input variable. The second input variable is the total squad market value. The output variable is the number of points achieved in the season.

# 4 RESULTS

This part of the paper is devoted to empirical research, where the non-parametric DEA methodology was utilised. In this research, the 1st and 2nd Slovak football league clubs are evaluated as the DMUs. An input-oriented DEA models were applied to measure the clubs' efficiency. The input-oriented models were chosen due to the difficulty of controlling the football clubs' outputs (a club cannot gain more than 3 points for winning one match). Unlike the outputs, most inputs can be controlled: a football club can reduce the number of players or buy cheaper players. Both the CCR model and the BCC model were used in the research. The efficient clubs characterised by OTE = 1 and PTE = 1 are highlighted in grey in the tables below. Each table also shows inefficient clubs (OTE < 1), which are further broken down. The clubs characterised by PTE = 1 and SE <1 did not waste their resources however made bad tactical choices. The inefficient clubs, which wasted their resources and, at the same time, made bad tactical choices (PTE < 1 and SE < 1), are also described in this section.

Table 1 displays the Fortuna Liga football clubs' efficiency results of the season 2020/21. In this season, Slovan Bratislava and Zlaté Moravce – Vráble were described as efficient clubs in terms of overall technical efficiency. Slovan Bratislava was also the winner of the regular season mentioned above. Zlaté Moravce - Vráble finished fifth. It can be stated that both clubs used their resources optimally (PTE = 1), and at the same time, they made good tactical choices (SE = 1). In terms of pure technical efficiency (OTE < 1 and PTE = 1), no club was identified as efficient in the 2020/21season. All other Fortuna Liga football clubs' inefficiency was due to both wasting resources and bad tactical choices (OTE < 1, PTE < 1 and SE < 1).

Table 1: Efficiency scores of Slovak Fortuna Liga clubs

| Football club | Overall technical efficiency | Pure technical efficiency | Scale efficiency | Position in the table |
|---|---|---|---|---|
| Slovan Bratislava | **1.000** | **1.000** | **1.000** | 1. |
| Z. Moravce-Vráble | **1.000** | **1.000** | **1.000** | 5. |
| Dun. Streda | 0.837 | 0.885 | 0.945 | 2. |
| Žilina | 0.762 | 0.766 | 0.995 | 3. |
| Trnava | 0.700 | 0.759 | 0.921 | 4. |
| Ružomberok | 0.616 | 0.900 | 0.685 | 7. |
| Pohronie | 0.600 | 0.991 | 0.606 | 12. |
| Michalovce | 0.599 | 0.899 | 0.667 | 9. |
| Senica | 0.561 | 0.881 | 0.636 | 11. |
| Trenčín | 0.541 | 0.659 | 0.821 | 6. |
| Nitra | 0.538 | 0.808 | 0.667 | 8. |
| Sereď | 0.481 | 0.722 | 0.667 | 10. |

The efficiency results of the football clubs playing in the 2nd league in the 2020/21 season are shown in Table 2. In the 2020/21 season, only one club was described as efficient in terms of overall technical efficiency - Liptovský Mikuláš, the regular season winner, and got promoted to Fortuna Liga for the season 2021/22. Liptovský Mikuláš used its resources optimally (PTE = 1), and at the same time, it chose good game tactics (SE = 1). In terms of pure technical efficiency (OTE < 1 and PTE = 1), three clubs (Skalica, Trebišov and Púchov) were identified as efficient in the 2020/21 season. It can be concluded that these clubs used their resources optimally. Their inefficiency is given only by their bad game tactical choices. The inefficiency of all other 2nd league clubs in the 2020/21 season was due to wasting their resources and wrong choice of game tactics (OTE < 1, PTE < 1 and SE < 1).

Table 2: Efficiency scores of Slovak clubs playing 2nd league

| Football club | Overall technical efficiency | Pure technical efficiency | Scale efficiency | Position in the table |
|---|---|---|---|---|
| Lip. Mikuláš | **1.000** | **1.000** | **1.000** | 1. |
| Skalica | 0.985 | **1.000** | 0.985 | 3. |
| B. Bystrica | 0.967 | 0.979 | 0.988 | 2. |
| Trebišov | 0.837 | **1.000** | 0.837 | 10. |
| Podbrezová | 0.816 | 0.842 | 0.969 | 4. |
| Púchov | 0.712 | **1.000** | 0.712 | 7. |
| Dubnica | 0.643 | 0.892 | 0.720 | 12. |
| FC Košice | 0.607 | 0.683 | 0.889 | 5. |
| Komárno | 0.573 | 0.780 | 0.735 | 8. |
| Bardejov | 0.539 | 0.926 | 0.582 | 13. |
| Šamorín | 0.431 | 0.569 | 0.757 | 6. |
| Petržalka | 0.416 | 0.605 | 0.688 | 9. |
| Žilina B | 0.301 | 0.491 | 0.613 | 11. |
| Slovan Bratislava B | 0.181 | 0.464 | 0.390 | 14. |
| Poprad | 0.085 | 0.884 | 0.096 | 15. |

# 5 CONCLUSIONS

Considering the current economic and financial situation of football clubs affected by the COVID-19 pandemic, the need of know about how efficiently a club use its resources increase. In this paper, the DEA method was used to analyze the efficiency of Slovak football clubs playing in the 1st and 2nd leagues in the 2020/21 season. There were 12 football clubs that played the Fortuna Liga and 15 football clubs that played the 2nd league.

From OTE's point of view, both the Fortuna Liga winner and the 2nd league winner have always been described as efficient. It was a surprise to find that Z. Moravce-Vráble, which placed 5th in the regular season of the Fortuna Liga, was also an efficient club from OTE's point of view. From the point of view of the lowest OTE values, Sereď took the last place, which finished in 10th place in the regular season of the Fortuna Liga. Pohronie took the last place in the league table. By applying the DEA method, it was found that the problems of this club could be caused by poor game tactics. In the case of 2nd league, the worst club was Poprad, which placed in the last part of the table in the regular season. Poprad was also the worst club from OTE's point of view. The surprise was the interesting position of Trebišov, which is an efficient club from the point of view of PTE. Behind his placement in the 10th place of the league table is probably a poorly chosen game tactics.

The research identified various sources of inefficiency in both football leagues. The first source of inefficiency is observed by PTE and is related to waste of inputs. To achieve the same number of points in the league table, a lower value of input variables (i.e. a lower total squad market value or a lower number of players) should be enough for these clubs. The second source of inefficiency can be observed by calculating scale efficiency and is associated with the selection of inappropriate sports tactics. In this case, of course, the head coach of the team is the most involved. The problem is not just how football clubs use their resources. These clubs should try to develop new and different sports tactics. The selection and subsequent purchase of players should also take place in the context of the development of new sports tactics. Most clubs in the 1st and 2nd leagues suffer from both sources of inefficiency. In this case, clubs should reduce resources and find out how efficient clubs are developing sports tactics. Further research will focus on evaluating the performance of players in both football competitions.

## References


[1] Barros, C. P., and Garcia-del-Barrio, P. (2011). Productivity drivers and market dynamics in the Spanish first division football league. *Journal of Productivity Analysis*, 35(1), pp. 5-13.

[2] Barros, C. P. and Leach, S. (2006). Performance evaluation of the English Premier Football League with data envelopment analysis. *Applied Economics*, 38(12), pp. 1449-1458.

[3] Barros, C. P., Assaf, A. and Sa-Earp, F. (2010). Brazilian Football League Technical Efficiency: A Simar and Wilson Approach. *Journal of Sports Economics*, 11(6), pp. 641-651.

[4] Chaiwuttisak, P. (2018). Measuring Efficiency of Thailand's Football Premier Leagues using Data Envelopment Analysis. In: *5th International Conference on Business and Industrial Research (ICBIR)*. Bangkok: IEEE.

[5] Charnes A., Cooper, W.W. and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), pp. 429-444.

[6] Charnes A., Cooper W., Golany B., Seiford L. and Stutz J. (1985). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, 30(1/2), pp. 91-107.

[7] Dyson, R., Allen, R., Camanho, A., Podinovski, V., Sarrico, C. and Shale, E. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2), pp. 245-259.

[8] Espitia-Escuer, M. and García-Cebrián, L. I. (2010). Measurement of the Efficiency of Football Teams in the Champions League. *Managerial and Decision Economics*, 31(6), pp. 373-386.

[9] EuroFotbal, (2022). Kluby [online] Available at: https://www.eurofotbal.cz/ [Accessed 9 Feb. 2022].

[10] Guzmán-Raja, I. and Guzmán-Raja, M. (2021). Measuring the Efficiency of Football Clubs Using Data Envelopment Analysis: Empirical Evidence from Spanish Professional Football. *SAGE Open*, 11(1), pp 1-13.

[11] Keskin, H. İ. and Öndes, H. (2020). Measuring the Efficiency of Selected European Football Clubs: DEA and Panel Tobit Model. *Sosyoekonomi Journal*, 28(43), pp. 153-174.

[12] Machlica, G. and Šiškovič, M. (2010). Ak v Afrike nasneží, budeme majstri sveta: Majstrovstvá sveta vo futbale v roku 2010 z pohľadu ekonómov. Bratislava: Inštitút finančnej politiky Ministerstvo financií SR.

[13] Melek, P. and Dedík, P. (2020). Spoločenská návratnosť investícií do futbalu na Slovensku. [online] Available at: https://futbalsfz.sk/spolocenska-navratnost-investicii-do-futbalu-na-slovensku/ [Accessed 22 Feb. 2022].

[14] Miningou, E. W. and Vierstraete, V. (2012). Efficience des Clubs Français de Football des Ligues 1 et 2. *Revue d'économie politique*, 122(1), pp. 37-66.

[15] Othman M., Foo L., Karim M. and Andaziz Y. (2010). Total Factor Productivity Efficiency Changes in a Malaysian Hotel Chain. *International Journal of Revenue Management*, 4(3/4), pp. 327-343.

[16] Petrović Djordjevic, D., Vujosevics, M. and Martic, M. (2015). Measuring efficiency of football teams by multi-stage DEA model. *Tehnički vjesnik*, 22(3), pp. 763-770.

[17] Ribeiro, A. S. and Lima, F. (2012). Portuguese football league efficiency and players' wages. *Applied Economics Letters*, 19(6), pp. 599-602.

[18] Soleimani-Damaneh, J., Hamidi, M. and Sajadi, N. (2011). Evaluating the Performance of Iranian Football Teams Utilizing Linear Programming. *American Journal of Operations Research*, 1(02), pp. 65-72.
[19] SportSen, (2022). Historie fotbalu [online] Available at: https://www.sportsen.cz/blog/historie-fotbalu [Accessed 25 Feb. 2022].
[20] Transfermarkt, (2022). Transfermarkt.com: Competitions [online]. Available at: https://www.transfermarkt.com/ [Accessed 10 Feb. 2022].
[21] Unia ligovych klubov and eSports, (2022). Tabuľka [online]. Available at: https://www.fortunaliga.sk/tabulka/2020 [Accessed 5 Feb. 2022].

**Author's address**
Ing. Natalie Pelloneová, Ph.D.
Technical University of Liberec, Faculty of Economics, Department of Business Administration and Management
Studentská 1402/2, 461 17 Liberec
Czech Republic
email: natalie.pelloneova@tul.cz

# QUANTIFYING THE ATTRACTIVENESS OF A LOCATION: THE CASE OF A BRICK-AND-MORTAR STORE

*Ondřej Sokol, Prague University of Economics and Business,*
*Tomáš Formánek, Prague University of Economics and Business,*
*Michal Černý, Prague University of Economics and Business*

**Abstract**
The sales of brick-and-mortar stores are to a large extent influenced by the attractiveness of their location. But the challenge arises in quantifying such attractiveness. The ideal would be to measure the number of passes by potential customers, but this data is almost impossible to access -- even if mobile operators have it, the cost is so high that it is unaffordable even for retail chains. However, with the development of geographic information systems, it is possible to use alternative geographic data, for example, the number of restaurants in a neighborhood can be a good indicator of pedestrian density. In this paper, we discuss possible approaches to quantify the attractiveness of locations with respect to retail objectives.

***Keywords:*** *Retail store, Location attractiveness, Geographic information systems*

***JEL Classification:*** C51
***AMS Classification:*** 62H12

## 1 INTRODUCTION

When deciding on the location of a brick-and-mortar store, knowledge of the surrounding area is essential, especially information regarding the pedestrian traffic density on the street in front of the location under consideration. The problem is that this information is not readily available if we require quantification of such density. In the case of selecting a location for a single store, it might even be possible to estimate or at least compare a few locations in a predetermined area, but if the objective of a retail chain is to determine the appropriate location for huge expansion, this is not realistically possible and mathematical or econometric models must be used to approximate the effect of location.

Of course, there are some other methods of estimating pedestrian density at a particular location. One possibility is to use cameras aimed at the vicinity of the shop. A significant disadvantage in this case is the need to monitor the location for a longer period, which greatly limits this method in the case of finding a suitable location for a new store from many options, as it is not possible to monitor a large number of locations simultaneously. In all aspects, a better alternative would be to use data from mobile operators, who also sell it to interested parties, but the price tends to be so high that it is unaffordable even for multinational retail chains.

In this short paper, we discuss the main approaches to quantifying of location attractiveness, with focus on modeling based on geographic data and their advantages and disadvantages. We focus on pedestrian traffic modelling using geographic data of the immediate surroundings of the point of interest.

## 2 STATE OF ART

Two main streams of research dealing with quantitative estimation of location attractiveness in relation to retail business are emerging in the literature. The first focuses on the calculation of indicators of the attractiveness of a location within a city using commonly available data, while the second is complex modelling using based on spatial interaction models (SIM), which, however, require a lot of difficult-to-access data.

## 2.1 Spatial interactions models

An interesting and complex approach is presented by spatial interaction models, which became popular in recent years, see for example Beckers et al (2021), Hood et al (2020) and Hood et al (2021). They aim to predict sales depending on geographic data, using modelling at the level of the customer segments.

The retail SIM typically takes the following form (Beckers et al, 2021):

$$S_{ij}^{k} = A_{i}^{k} O_{i}^{k} W_{j}^{\alpha^{k}} \, exp^{-\beta^{k} c_{ij}}, \tag{1}$$

where $S_{ij}^{k}$ represents the total expenditure flow between demand zone $i$ and store $j$, by consumer type $k$, $A_{i}^{k}$ is a balancing factor, $O_{i}^{k}$ is a measure of the available demand for groceries in zone $i$, and is disaggregated by person type. $W_{j}^{\alpha^{k}}$ represents the attractiveness of store $j$, measured here by store size; $c_{ij}$ is the travel cost between demand zone and store, measured as the Euclidean distance; $\alpha^{k}$ and $\beta^{k}$ disaggregate the model by including brand attractiveness.

Calibration is necessary for the correct use of these models. And for that reason, it is necessary to have information about the customer's location and his segmentation based on shopping behavior. Then his behavior as he moves around the city is estimated and the result is extrapolated.

The main problem is getting this information about the customers. For example, loyalty program data can be used, but it may not always be in place and selection bias can be introduced as the comparison of customers who have joined a loyalty club with those who refuse to do so is always problematic. The other option is to conduct a survey, but this is very expensive if it is to be truly representative.

Therefore, for the correct use of spatial interaction models, there are relatively high requirements for data on customer shopping behavior, which cannot be obtained entirely from the commonly available transaction data of a retail chain, even if it has an extensive loyalty program.

## 2.2 Centrality indicators

Even before the introduction of spatial interaction models, centrality indicators were used. These were basically very simple indicators based on the idea that each city can be represented as a graph (Hillier et al, 1993). Streets are then individual edges and intersections, or other points of interest are vertices. The basic centrality indicator, called closeness, is then calculated as follows

$$Closeness_{\ell} = (N-1) \sum_{\substack{s=1 \\ s \neq \ell}}^{N} \frac{1}{d_{\ell s}}, \tag{2}$$

where $s = 1,2, \dots, N$ is the $s$th node, $d_{\ell s}$ is the shortest transportation (street-based) distance between the two distinct nodes, and $N$ is the total number of nodes. For a given graph, $Closeness_{\ell}$ is the inverse average distance from $\ell$ to all other nodes. Graph elements with higher scores have shorter distances to other nodes (i.e. increased closeness). Closeness can be regarded as a measure of how long it will take to spread information from given node to all other nodes sequentially. An example of closeness indicator computed on the map of Prague is shown in Figure 1, which we have plotted in R using *sf* and *ggplot* packages. Sometimes, inverted value of closeness is used, called farness or peripherality.

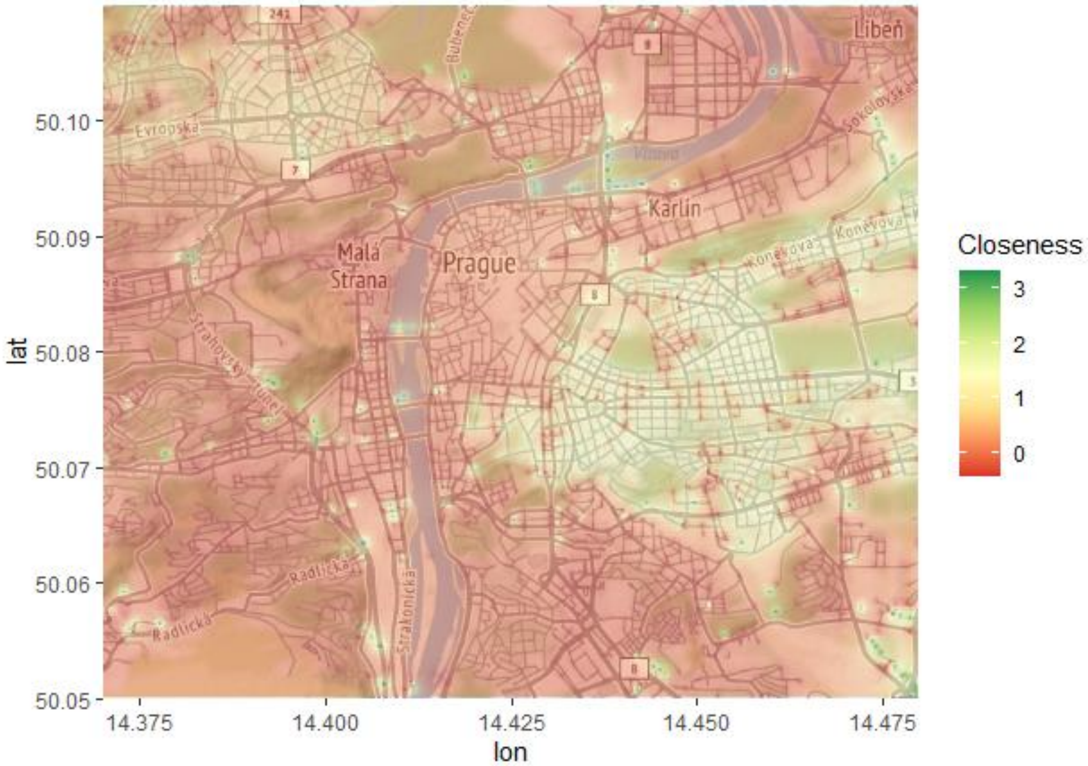



***Figure 1:*** *Map of Prague with closeness indicator computed.*

There are several variants of centrality indicators, e.g. betweenness. Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

$$Betweenness_{\ell} = \sum_{s,v \in V} \frac{\sigma_{sv}(\ell)}{\sigma_{sv}}, \quad (3)$$

where $s, v$ are the nodes, $V$ is the set of all nodes, $\sigma_{sv}$ is the number of shortest paths (street-based) between the two distinct nodes and $\sigma_{sv}(\ell)$ is the number of shortest paths (street-based) between the two distinct nodes going through node $\ell$. An example of betweenness indicator computed on the city of Prague is shown in Figure 2.

The advantage is their computational effortlessness and, from a certain point of view, their versatility. The only prerequisite for their correct use is the correct definition of edges and vertices. For vertices, there may be a problem of deciding whether a given node is a sufficient attractor and should be included – here it really depends on the expert. The alternative is to use only intersections.

Although these indicators enjoyed great popularity at the end of the last century, they have recently come under criticism for their inadequacy in failing to estimate true densities in empirical studies, see Kang (2016) and Kickert et al (2020).

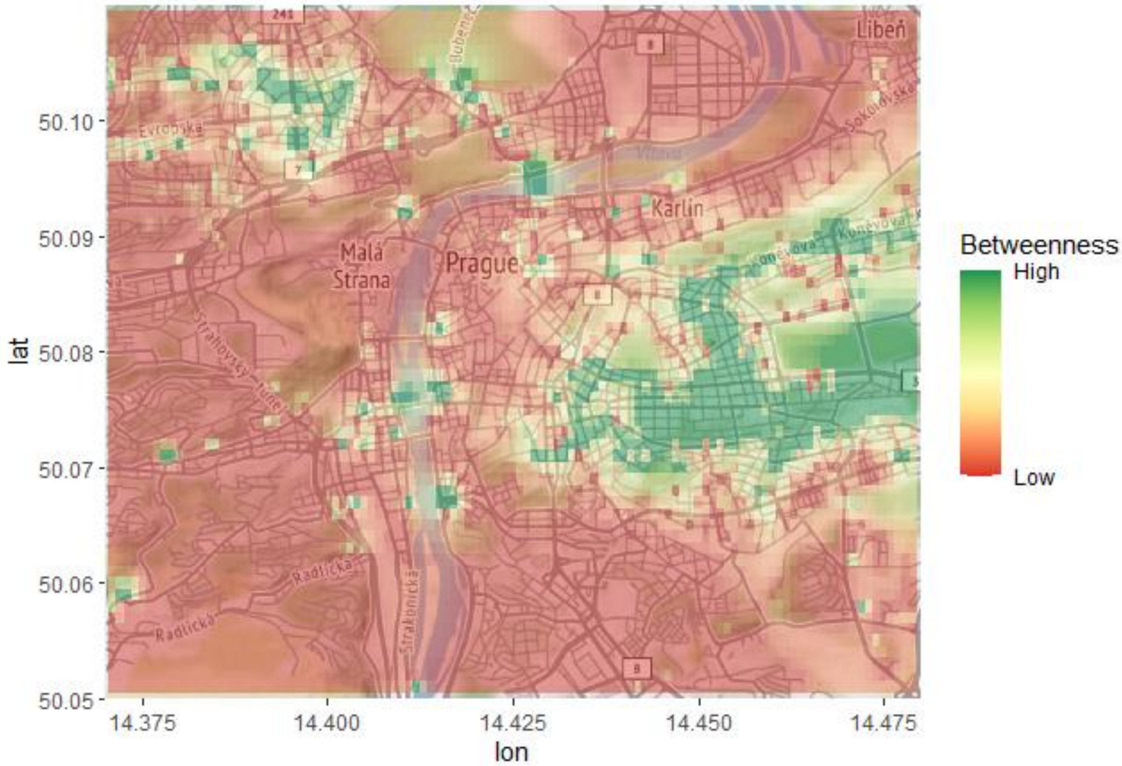



***Figure 2:*** *Map of Prague with betweenness indicator computed.*

### 2.3 Attractors in the surroundings

In the case of a retail chain, it is straightforward to use sales data in combination with geographic data to estimate the attractiveness of a location, especially at the micro level, i.e. the immediate surroundings of the store, for example, within walking distance of 5 minutes. Such attractors can be public transport stops, railway stations, post offices, banks, shopping centres, but also restaurants. Such data can be obtained from publicly available sources such as OpenStreetMaps or from commercially available sources such as Google Maps. And even the processing of this data is relatively simple and does not require advanced knowledge.

To estimate the effect of individual attractors on pedestrian density in front of a store, methods based on regression analysis can be used, where instead of pedestrian traffic density – which we consider unobservable – the store revenue is often used. Recall that we are assuming the retail chain's situation and the data it has available by default. The question then, of course, is what attractors are worth examining, but standard statistical procedures can be used for this as well. We discuss the problems of using attractors in regression, as well as the measures of centrality, in the next section.

## 3 PEDESTRIAN DENSITY ANALYSIS

Expenditure flow modelling is difficult, and data is usually not available, see Section 2.1. However, the retailer's objective is mainly to model the traffic at the location, the origin of the traffic is secondary. This can be aided by information on attractors, which can possibly be supplemented by centrality indicators. A natural approach is to use regression models.

In contrast to an approach based on centrality indicators alone, linear regression model with attractors as regressors offers a measure of the type of pedestrian traffic itself, see Formánek and Sokol (2022). Information on pedestrian density is not sufficient because there may be multiple very different segment of pedestrians – it is fundamentally different for a retailer to have a large amount of young people in front of a ski shop versus traffic consisting mostly of retirees. With regression and fusion of sales and geographic data, these differences can be modeled. We stem from linear regression models with following basic form

$$y = \beta_0 + \alpha C + \gamma A + \epsilon, \tag{4}$$

where $y$ is a vector of dependent variable, $C$ is a vector (or matrix) of centrality indicators, $A$ is a matrix of selected attractors and $\epsilon$ is the random error. Of course, a non-linear approach may be used, but in our experience linear (and log linear) form is often sufficient.

However, it is necessary to obtain information on how each attractor works on different types of pedestrian densities. But if sales data is available, this information can be estimated by estimating various dependent variables, for example sales of different product categories or number of specific transactions. A retailer usually knows how its customers behave and what customer segments are present and how they differ.

If we base the regression on the idea that customers from segment $A$ purchases certain product categories above average compared to other segments, or that their baskets are above average, this knowledge can be used to estimate the effect of various attractors directly on said customers from segment $A$. Here we assumed that customer segments can be estimated purely from sales data, using identification of customers who joined loyalty program, but customer segments and their shopping behavior can be alternatively estimated by surveys.

The advantage of such an approach is the ease of interpretation and the relatively low requirements for assumptions and available data. A possible disadvantage is the relatively high level of aggregation. Spatial interactions models also provide more information that a retailer could use.

# 4 CONCLUSIONS

In this paper, we discussed main approaches to quantify the attractiveness of locations with respect to retail objectives and their advantages and disadvantages.

The ideal method would accurately measure the number of passes by potential customers, but the needed data is basically impossible to obtain. Great proxy would be mobile operator's data about pedestrian density, but they are extremely costly. However, with the development of geographic information systems, it is possible to use geographic data as a viable alternative. Especially the discussed model in Section 3, with both attractors and centrality indicators in linear regression model, seems to be an accurate indicator of pedestrian density.

**Acknowledgements:** The work of Michal Černý was supported by the Czech Science Foundation (project 20-17529S) and the work of Ondřej Sokol was supported by the Internal Grant Agency of Prague University of Economics and Business (project F4/27/2020).

## References

[1] Beckers, J., Birkin, M., Clarke, G., Hood, N., Newing, A. and Urquhart, R. (2021). Incorporating E-commerce into Retail Location Models. *Geographical Analysis*.

[2] Formánek, T., & Sokol, O. (2022). Location effects: Geo-spatial and socio-demographic determinants of sales dynamics in brick-and-mortar retail stores. *Journal of Retailing and Consumer Services*, 66, p. 102902.

[3] Hillier, B., Penn, A., Hanson, J., Grajewski, T. and Xu, J. (1993). Natural movement: or, configuration and attraction in urban pedestrian movement. *Environment and Planning B: planning and design*, 20(1), pp. 29-66.

[4] Hood, N., Urquhart, R., Newing, A. and Heppenstall, A. (2020). Sociodemographic and spatial disaggregation of e-commerce channel use in the grocery market in Great Britain. *Journal of Retailing and Consumer Services*, 55, p. 102076.

[5] Hood, N., Clarke, G., Newing, A. and Rains, T. (2021). Applications of store loyalty card big data in the location planning process. In *Big Data Applications in Geography and Planning*. Edward Elgar Publishing.

[6] Kang, C. D. (2016). Spatial access to pedestrians and retail sales in Seoul, Korea. *Habitat international*, 57, pp. 110-120.

[7] Kickert, C., Vom Hofe, R., Haas, T., Zhang, W. and Mahato, B. (2020). Spatial dynamics of long-term urban retail decline in three transatlantic cities. *Cities*, 107, p. 102918.


**Author's address**


Ing. Ondřej Sokol, PhD.
Prague University of Economics and Business, Department of Econometrics
Náměstí Winstona Churchilla 1938/4, 130 67 Praha
Czech Republic
email: ondrej.sokol@vse.cz

# FINDING THE OPTIMAL FOOTBALL TEAM BASED ON GPS DATA AND MULTI-CRITERIA DECISION MAKING

***Kateřina Svitková, Prague University of Economics and Business,***
***Martina Kuncová, Prague University of Economics and Business,***
***Jana Sekničková, Prague University of Economics and Business***

**Abstract**
Nowadays, the use of technology has already reached all areas of everyday life. Sport is no exception. GPS chips (so-called GPS locators) belong among the technologies commonly used today. They monitor the location of athletes and are used by endurance runners, cross-country skiers and biathletes. Surprisingly, they can also be found in footballers, in the so-called GPS vest, where they monitor not only the movement of footballers but also their biometrics. On the basis of the monitored data, teams and coaches then carry out extensive analyses and plan the most appropriate training for their athletes. This technology is also used by West Ham United, one of the England's top women's football teams. In this article, we analyze the players of this team based on real data and try to describe a possible way of selecting the ideal team for the upcoming match. Four methods of multi-criteria decision-making, namely the conjunctive method combined with the PRIAM method and the SAW and WSA methods, are used to make the selection. The obtained results are then compared. SAW and WSA methods cover the feedback of coaches responsible for the team selection. They quantified the weights of each criterion and thus enabled the construction of their own player rankings.

***Keywords:*** *Multi-criteria decision-making, Football player selection, GPS locators data*

***JEL Classification:*** C60
***AMS Classification:*** 90B50, 90C29, 91B06

## 1 INTRODUCTION

Personnel selection plays a significant role in human resource management. A typical goal in this process is to choose the best candidate for a given workplace or a vacancy in a company (Dursun and Karsak 2010). The organizations are seeking more powerful ways of ranking employees or personnel and multi-criteria decision-making (MCDM) methods seem to be one of the appropriate tools: Seol and Sarkis (2005) used AHP method to select the best internal auditor; Saremi et al. (2009) evaluated TQM consultants using TOPSIS method; project management team members were evaluated by COPRAS method by Zavadskas et al. (2008) or by AHP method by Bazsova (2017); Karabasevic et al. (2017) applied the combination of SWARA and ARAS methods to find out the best sales manager; Widianta et al. (2018) compared the results of AHP, SAW, TOPSIS and PROMETHEE methods in the employee placement process; Afshari et al. (2010) used SAW method to select a senior IT officer.

Various evaluation systems or methods can also be found in the field of sports and football (called soccer in the USA), which is also the focus of this article. In addition to clubs or players comparison using statistical methods, a comparison of different groups of football fans or female visitors can also be found (Scholz 2020). But MCDM usage is typical to rank the football teams: Kiani Mavi et al. (2012) used AHP and TOPSIS to rank the football teams in Germany; Gökgöz and Yalçın (2019, 2021) evaluated the performance of World Cup 2018 teams first using ARAS and SAW methods and second with TOPSIS and WASPAS methods. TOPSIS method was also used by Qader et al. (2017) for the football player selection problem. Sianturi

(2019) used WSM for the football athletes' comparison. Nasiri et al. (2018) combined ANP with PROMETHEE II and DEA methods to compute the score attained by the different players in each potential position and to identify the most efficient Pareto solution determining the status of each player. From the above analyzes it is clear that various MCDM methods and especially its combinations are used in research. The same approach was chosen in this article. Our comparison concerns women footballers from the West Ham United club. The conjunctive method combined with the PRIAM method and the SAW and WSA methods are used to select the best team.

## 2 DATA AND METHODOLOGY

The data in this paper is real data from the top women's football competition in England, specifically West Ham United. For privacy reasons, no names will be mentioned in this paper. Players (as alternatives in this paper) are identified by our numbers (which is different from their squad number). They are evaluated according to the 8 criteria described below. It is evident from the meaning that all the criteria are of the maximizing type. All individual data comes from the GPS chip on the vest that the players wear on their backs. The relevant data for this paper shows Table 1. It contains data from this season's first five games, and the values are averaged over one game.

Table 1 – GPS data

| Player | Total Distance (m) | Run (m/min) | HSR (m) | HSR% (%) | Sprints | Decels | Accels | Top Speed (m/s) |
|---|---|---|---|---|---|---|---|---|
| **1** | 12503.06 | 112.00 | 718.03 | 6 % | 10.27 | 228.29 | 210.93 | 7.66 |
| **2** | 11310.54 | 112.22 | 341.48 | 3 % | 5.32 | 162.00 | 157.10 | 7.22 |
| **3** | 9725.11 | 98.62 | 421.15 | 4 % | 6.34 | 108.79 | 108.51 | 7.91 |
| **4** | 10823.47 | 101.96 | 791.94 | 7 % | 16.73 | 179.26 | 191.91 | 8.49 |
| **5** | 10281.44 | 106.37 | 501.76 | 5 % | 9.64 | 195.14 | 136.76 | 7.92 |
| **6** | 11043.31 | 156.90 | 734.30 | 7 % | 17.91 | 197.91 | 162.51 | 7.44 |
| **7** | 11676.95 | 107.33 | 607.45 | 5 % | 10.61 | 179.83 | 173.89 | 7.75 |
| **8** | 8547.78 | 74.53 | 859.45 | 10 % | 12.29 | 175.28 | 162.66 | 7.99 |
| **9** | 10947.73 | 112.80 | 592.41 | 5 % | 13.03 | 184.27 | 180.85 | 7.74 |
| **10** | 11348.26 | 106.04 | 544.61 | 5 % | 12.30 | 187.01 | 206.81 | 7.68 |
| **11** | 10258.27 | 117.12 | 598.28 | 6 % | 10.42 | 165.42 | 198.89 | 7.98 |
| **12** | 9271.29 | 78.96 | 680.23 | 7 % | 13.01 | 121.88 | 143.15 | 7.90 |
| **13** | 11831.59 | 128.50 | 831.54 | 7 % | 12.10 | 247.98 | 281.21 | 7.45 |
| **14** | 12341.40 | 138.88 | 696.57 | 6 % | 8.97 | 180.27 | 229.37 | 7.73 |
| **15** | 10207.52 | 95.75 | 561.45 | 6 % | 13.87 | 143.66 | 187.42 | 8.43 |
| **16** | 9923.74 | 181.21 | 336.68 | 3 % | 3.22 | 233.42 | 166.20 | 6.98 |
| **17** | 13176.17 | 123.63 | 685.70 | 5 % | 8.63 | 100.82 | 112.52 | 6.73 |
| **18** | 10686.39 | 118.40 | 719.14 | 7 % | 13.09 | 206.41 | 187.60 | 8.24 |

- *Total Distance (m)* – distance measured in meters and indicates how much a player runs on average per game.
- *Run (m/min)* – interpreted as speed in meters per minute, shows how many meters a player runs per minute and the values are averaged per game.
- *HSR (m)* – high-speed running or running at high intensity. This distance starts to count once you are running over 26 km/h. Its value is in between conventional running and sprinting. It indicates how many meters on average a player runs in a match at speeds higher than 26 km/h.
- *HSR% (%)* – the relative proportion of running at high intensity, calculated as a proportion of total distance and running at high intensity.

- *Sprints* – average number of sprints per game. A sprint counts if a player runs more than 10 meters over 28 km/h.
- *Decels* – deceleration's means the number of times a player is in a hard brake on average per game.
- *Accels* – acceleration's means the number of times a player is in a burst on average per game.
- *Top Speed (m/s)* – is the highest measured speed in a match, averaged over one match.

For a better understanding of the data, let's use player number 1 as an example (see the first row in Table 1). On average, she runs a total of 12,503.06 meters per game, with 112 meters per minute. Of this distance, she runs 718.03 meters at speeds over 26 km/h which is a total of 6% of her total distance in a match. On average, she will make 10 sprints per game. She stops a total of 228 times abruptly and starts abruptly 210 times on average per game. Her top speed in a match is 7.66 m/s on average.

For the best 10 players' selection, aspiration levels or criteria weights can be used. Aspiration levels represent the acceptable limiting values for each criterion. For this article, the aspiration levels were set in agreement with the coaches. Since the criteria are maximizing, the goal is to exceed all the pre-set values.

The weights of the criteria were set according to the professional evaluation of the coaches based on their personal importance. It is based on their style of play based on attacking and quick counterattacks. HSR is, therefore an important criterion. The second set of weights was created by one of the players, according to her importance of the criteria for football. Both weight vectors, as well as the specified aspirational levels, are presented in Table 2.

Table 2 – Weights and aspiration levels

| | **Total Distance (m)** | **Run (m/min)** | **HSR (m)** | **HSR% (%)** | **Sprints** | **Decels** | **Accels** | **Top Speed (m/s)** |
|---|---|---|---|---|---|---|---|---|
| weight (coach) | 0.20 | 0.05 | 0.25 | 0.10 | 0.05 | 0.10 | 0.10 | 0.15 |
| weight (player) | 0.25 | 0.07 | 0.10 | 0.05 | 0.15 | 0.09 | 0.09 | 0.20 |
| aspiration levels | 11000 | 100 | 700 | 0.07 | 11 | 170 | 150 | 7.40 |
| ideal values | 13176.17 | 181.21 | 859.45 | 0.10 | 17.91 | 247.98 | 281.21 | 8.49 |

The selection of the team of 10 players is made via multi-criteria evaluation of alternatives' methods. To solve this kind of model it is necessary to know the decision-maker's preferences. These preferences can be described by aspiration levels (or requirements), criteria order or by the weight of the criteria. The model of multi-criteria evaluation of alternatives contains a list of alternatives $A = \{a_1, a_2, \dots, a_p\}$, a list of criteria $F = \{f_1, f_2, \dots, f_k\}$ and $y_{ij}$ as an evaluation of the alternatives by each criterion, $i = 1, 2, \dots, p, j = 1, 2, \dots, k$ (Evans 1984). For the analysis several methods can be used. As the aim is the football team (10 players) selection and we have numerical data and information from the coach and footballer, we decided to use the conjunctive method and its combination with the PRIAM method, and then WSA and SAW methods to see whether there is a difference in the final team selection.

Conjunctive method and PRIAM (Programme utilisant l´Intelligence Artificiele en Multicritere) method belong to the group of methods that need information about aspiration levels for each criterion that should be matched (Figueira et al. 2005). The conjunctive method separates alternatives into acceptable and unacceptable. Acceptable alternatives meet all aspiration levels. The aim of PRIAM method is usually to find out one alternative that meets the requirements (aspiration levels). If there is no such alternative, the aspiration levels should be changed. Another possibility is to calculate the distance of each alternative from the vector of aspiration levels – for all alternatives acceptable by the conjunctive method – or from the vector of ideal criteria values. This distance is calculated for each alternative. The distance from aspiration level is based upon the formula:

$$d_i = \sum_{j=1}^{n} \frac{|a_j - y_{ij}|}{b_j} \tag{1}$$

where $d_i$ is the relative index for the distance of alternative $i$, $a_j$ is the aspiration level for the criterion $j$, $b_j$ is the ideal level for the criterion $j$ and $y_{ij}$ is the real value for the alternative $i$ and criterion $j$ $(i = 1, 2, \ldots, m, j = 1, 2, \ldots, n)$. The distance from the ideal values is calculated according to the formula (1) but instead of $a_j$ the $b_j$ values are used.

SAW (Simple Additive Weighting) and WSA (Weighted Sum Approach) sort the alternatives based on the values of their utility functions which in this case are assumed to be linear. Both methods require information about the weights of the criteria. The higher value of utility means the better alternative. For both methods, the formula (2) 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).

$$u(a_i) = \sum_{j=1}^{k} w_j r_{ij},\ \forall\, i = 1, \ldots, p. \tag{2}$$

SAW and WSA use formulas (3) for the maximization type of criterion (left one) or for minimization type (right one), where in SAW $\min(y_{ij}) = 0$ in both formulas while WSA uses $\min(y_{ij})$ taken from the real data.

$$r_{ij} = \frac{y_{ij} - \min\limits_i y_{ij}}{\max\limits_i y_{ij} - \min\limits_i y_{ij}}\ ,\ r_{ij} = \frac{\max\limits_i y_{ij} - y_{ij}}{\max\limits_i y_{ij} - \min\limits_i y_{ij}}. \tag{3}$$

## 3 RESULTS

The first method used was the conjunctive method. Only players who reach all the defined aspiration levels are acceptable. The three aspiration levels were easy to reach (run, accels and top speed), while for HSR the aspiration level was set quite high. Table 3 shows that only two players with numbers 6 and 13 are acceptable. As we need 10 players to have a team, changing the aspiration levels to have more acceptable alternatives would be necessary.

Table 3 – Acceptable alternatives for conjuntive method

| Player | Total Distance (m) | Run (m/min) | HSR (m) | HSR% (%) | Sprints | Decels | Accels | Top Speed (m/s) |
|---|---|---|---|---|---|---|---|---|
| **6** | 11043.31 | 156.90 | 734.30 | 0.07 | 17.91 | 197.91 | 162.51 | 7.44 |
| **13** | 11831.59 | 128.50 | 831.54 | 0.07 | 12.10 | 247.98 | 281.21 | 7.45 |
| *aspiration level* | 11000.00 | 100.00 | 700.00 | 0.07 | 11.00 | 170.00 | 150.00 | 7.40 |

As it was mentioned above, one possibility is not to change the aspiration levels but to use the principle of the PRIAM method and measure the distance from the ideal values for each player. Each player is the greatest asset in a team if she achieves values close to the ideal according to all criteria. We, therefore, set the maximum values of each criterion as aspiration values. These distances from the ideal are calculated in Table 4. We can see that the player with the number 13 is the winner, the second place was again taken by the player with the number 6. The players who would be selected for the TOP 10 team are marked in Table 4 in green color. In conjunctive and PRIAM methods there is not possible to use criteria weights to express the importance of each criterion. That is why we decided to use WSA and SAW methods. Both calculate the utility for each player. Two weight vectors, one taken from the coach and one from the player, were used. Table 5 shows the final utilities and ranking of all players.

Table 4 – PRIAM method results – distances from the ideal values

| ideal values / player | Total Distance (m) | Run (m/min) | HSR (m) | HSR% (%) | Sprints | Decels | Accels | Top Speed (m/s) | SUM | Rank |
|---|---|---|---|---|---|---|---|---|---|---|
| | 13176.17 | 181.21 | 859.45 | 0.10 | 17.91 | 247.98 | 281.21 | 8.49 | | |
| 1 | 0.051 | 0.382 | 0.165 | 0.429 | 0.427 | 0.079 | 0.250 | 0.098 | 1.880 | 5 |
| 2 | 0.142 | 0.381 | 0.603 | 0.700 | 0.703 | 0.347 | 0.441 | 0.150 | 3.465 | 17 |
| 3 | 0.262 | 0.456 | 0.510 | 0.569 | 0.646 | 0.561 | 0.614 | 0.068 | 3.687 | 18 |
| 4 | 0.179 | 0.437 | 0.079 | 0.272 | 0.066 | 0.277 | 0.318 | 0.000 | 1.627 | 3 |
| 5 | 0.220 | 0.413 | 0.416 | 0.515 | 0.462 | 0.213 | 0.514 | 0.067 | 2.819 | 14 |
| 6 | 0.162 | 0.134 | 0.146 | 0.339 | 0.000 | 0.202 | 0.422 | 0.124 | 1.528 | 2 |
| 7 | 0.114 | 0.408 | 0.293 | 0.483 | 0.408 | 0.275 | 0.382 | 0.087 | 2.449 | 11 |
| 8 | 0.351 | 0.589 | 0.000 | 0.000 | 0.314 | 0.293 | 0.422 | 0.059 | 2.028 | 7 |
| 9 | 0.169 | 0.378 | 0.311 | 0.462 | 0.272 | 0.257 | 0.357 | 0.088 | 2.294 | 8 |
| 10 | 0.139 | 0.415 | 0.366 | 0.523 | 0.313 | 0.246 | 0.265 | 0.095 | 2.362 | 9 |
| 11 | 0.221 | 0.354 | 0.304 | 0.420 | 0.418 | 0.333 | 0.293 | 0.060 | 2.403 | 10 |
| 12 | 0.296 | 0.564 | 0.209 | 0.270 | 0.273 | 0.508 | 0.491 | 0.069 | 2.682 | 13 |
| 13 | 0.102 | 0.291 | 0.032 | 0.301 | 0.325 | 0.000 | 0.000 | 0.122 | 1.173 | 1 |
| 14 | 0.063 | 0.234 | 0.190 | 0.439 | 0.499 | 0.273 | 0.184 | 0.090 | 1.971 | 6 |
| 15 | 0.225 | 0.472 | 0.347 | 0.453 | 0.226 | 0.421 | 0.334 | 0.007 | 2.483 | 12 |
| 16 | 0.247 | 0.000 | 0.608 | 0.663 | 0.820 | 0.059 | 0.409 | 0.178 | 2.983 | 16 |
| 17 | 0.000 | 0.318 | 0.202 | 0.482 | 0.518 | 0.593 | 0.600 | 0.207 | 2.921 | 15 |
| 18 | 0.189 | 0.347 | 0.163 | 0.331 | 0.269 | 0.168 | 0.333 | 0.029 | 1.829 | 4 |

Table 5 – WSA and SAW results

| | weights given by coach | | | | weights given by player | | | |
|---|---|---|---|---|---|---|---|---|
| Player | WSA utility | Rank WSA | SAW utility | Rank SAW | WSA utility | Rank WSA | SAW utility | Rank SAW |
| 1 | 0.6587 | 3 | 0.8177 | 3 | 0.6395 | 3 | 0.8094 | 4 |
| 2 | 0.2580 | 17 | 0.5956 | 17 | 0.3148 | 16 | 0.6364 | 17 |
| 3 | 0.2378 | 18 | 0.5803 | 18 | 0.2756 | 18 | 0.6068 | 18 |
| 4 | 0.6876 | 2 | 0.8328 | 2 | 0.6880 | 2 | 0.8399 | 2 |
| 5 | 0.3989 | 15 | 0.6741 | 15 | 0.4325 | 13 | 0.7007 | 15 |
| 6 | 0.5959 | 6 | 0.8097 | 4 | 0.6089 | 6 | 0.8378 | 3 |
| 7 | 0.5147 | 8 | 0.7362 | 8 | 0.5316 | 8 | 0.7519 | 9 |
| 8 | 0.5702 | 7 | 0.8043 | 5 | 0.4595 | 12 | 0.7478 | 11 |
| 9 | 0.4960 | 9 | 0.7352 | 9 | 0.5244 | 10 | 0.7634 | 7 |
| 10 | 0.4879 | 10 | 0.7266 | 11 | 0.5290 | 9 | 0.7615 | 8 |
| 11 | 0.4862 | 11 | 0.7276 | 10 | 0.4925 | 11 | 0.7374 | 12 |
| 12 | 0.4264 | 14 | 0.7093 | 13 | 0.4023 | 15 | 0.7072 | 13 |
| 13 | 0.7524 | 1 | 0.8922 | 1 | 0.6884 | 1 | 0.8626 | 1 |
| 14 | 0.6323 | 4 | 0.8003 | 7 | 0.6185 | 5 | 0.7930 | 6 |
| 15 | 0.4804 | 12 | 0.7116 | 12 | 0.5334 | 7 | 0.7502 | 10 |
| 16 | 0.2596 | 16 | 0.6179 | 16 | 0.2865 | 17 | 0.6436 | 16 |
| 17 | 0.4417 | 13 | 0.7090 | 14 | 0.4219 | 14 | 0.7069 | 14 |
| 18 | 0.6284 | 5 | 0.8031 | 6 | 0.6220 | 4 | 0.8043 | 5 |

From the results in Tables 4 and 5 we can see that the players can be divided into several groups based on the analysis. Player **number 13** is the winner according to all the methods used. The second group is made up of 6 players (**numbers 1, 4, 6, 9, 14 and 18**) who were selected for the top 10 team by all methods regardless of the weights used, so we could say that they took 2nd-7th place. On the other hand, player number 3 finished in the last 18th place using all methods. The 16th and 17th places are taken by players number 2 and 16 regardless of the method and players number 5, 12 and 17 will not be included in the selection team as they are ranked 13th-15th in all methods.

The remaining 5 players (numbers 7, 8, 10, 11, 15) form a sort of middle group and deserve their own analysis because three of them should be selected for the top 10 team. Player number 7 is ranked 9th by all methods, except PRIAM, which puts her as high as 11th and therefore will not make the team at the given aspiration levels. Players number 8 and 15 will fight for a place in the elite team. Player number 8 will even finish 8th by all methods respecting the coaches' preferences, compared to player number 15 who will finish 12th. The latter, in turn, wins over player number 8 in the methods that take into account the player's preferences. The last pair is represented by players with numbers 10 and 11. Player 10 is rated better than player 11 in all methods except SAW with coaches' weights. Therefore, if we were to select the last three players from this group for the elite team, they would be **number 7, 10** and, according to the coaches' preferences, player **number 8**.

In the process of personnel selection, it is typical that MCDM methods have different results, and that is why it is better to use more methods and compare the results. According to our analysis it is clear that the results are similar and the methods could be used to find out the best 10 players. The decision can also be influenced by the number of players needed for a special position, such as strikers or defenders (this fact was not included in our analysis).

# 4 CONCLUSIONS

The main aim of this paper was to evaluate the physical fitness of West Ham United women's team players based on data obtained from GPS chips. Based on the team coaches' data, we attempted to select the top 10 female players using methods of multi-criteria evaluation of alternatives. The results show that only two female players (numbers 6 and 13) satisfy the coaches' aspiration levels and they are therefore certain players for the final squad. The remaining eight players are selected for the team using the PRIAM, WSA and SAW methods. We have shown how chip data, coaches' knowledge and multi-criteria decision-making methods can be combined to analyze sports teams and make representative team selections. We have also demonstrated that simple methods of multicriteria evaluation of alternatives provide very similar results and can therefore be used for analysis without much difficulty.

**Acknowledgements:** This work was supported by the grant No. F4/42/2021 of the Internal Grant Agency and by an institutional fund IP400040 for long-term conceptual development of science and research, Faculty of Informatics and Statistics, Prague University of Economics and Business.

**References**

[1] Afshari, A., Mojahed, M. and Yusuff, R.M. (2010). Simple Additive Weighting approach to Personnel Selection problem. *International Journal of Innovation, Management and Technology*, 1(5), pp. 511-515.

[2] Bazsova, B. (2017). Evaluation of the Project Management Team Members by Using the MCDM. In: *Key Issues for Management of Innovative Projects – Chapter 9*

[online] IntechOpen, pp. 151-166. Available at: https://www.intechopen.com/chapters/56003 [Accessed 7 March 2022].
[3] Dursun, M. and Karsak, E. (2010). A fuzzy MCDM approach for personnel selection. *Expert Systems with Applications*, 37, pp. 4324–4330.
[4] Evans, G.W. (1984). An Overwiev of Techniques for Solving Multiobjective Mathematical Programs. *Management Science*. 30(11), pp. 1268-1282.
[5] Figueira, J., Greco, S. and Ehrgott M. (2005). *Multiple Criteria Decision Analysis – State of the Art Surveys*. New York: Springer Science + Business Media Inc.
[6] Gökgöz, F., & Yalçın, E. (2019). An Integrated Approach to the World Cup Teams Using Entropy based ARAS and SAW Methods. In *25th ISTANBUL Int'l Conference on Literature, Languages, Humanities & Social Sciences*, pp. 1-4.
[7] Gökgöz, F., & Yalçın, E. (2021). A comparative multi criteria decision analysis of football teams: evidence on FIFA world cup. *Team Performance Management: An International Journal*. 27(3/4), pp. 177-191.
[8] Karabasevic, D., Stanujkic, D., and Urosevic, S. (2017). The MCDM Model for Personnel Selection Based on SWARA and ARAS Methods. *Management:Journal Of Sustainable Business And Management Solutions In Emerging Economies*, 20(77), pp. 43-52.
[9] Kiani Mavi, R., Kiani Mavi, N., and Kiani, L. (2012). Ranking football teams with AHP and TOPSIS methods. *International Journal of Decision Sciences, Risk and Management*, 4(1-2), pp. 108-126.
[10] Nasiri, M. M., Ranjbar, M., Tavana, M., Santos Arteaga, F. J., and Yazdanparast, R. (2019). A novel hybrid method for selecting soccer players during the transfer season. *Expert Systems*, 36(1), e12342.
[11] Qader, M. A., Zaidan, B. B., Zaidan, A. A., Ali, S. K., Kamaluddin, M. A., and Radzi, W. B. (2017). A methodology for football players selection problem based on multi-measurements criteria analysis. *Measurement*, 111, pp. 38-50.
[12] Saremi, M., Mousavi, S. F. and Sanayei, A. (2009). TQM consultant selection in SMEs with TOPSIS under fuzzy environment, *Expert Systems with Applications*, 36(2), pp.2742–2749.
[13] Scholz, P. (2020). Female visitors and their behavior at football stadiums: Czechia case study. *Journal of Physical Education and Sport (JPES),* 20 (4), pp. 1803 – 1808.
[14] Seol, I. and Sarkis, J. (2005). A multi-attribute model for internal auditor selection, *Managerial Auditing Journal,* 20(8), pp. 879-892.
[15] Sianturi, L. T. (2019). Implementation of Weight Sum Model (WSM) in the Selection of Football Athletes. *The IJICS (International Journal of Informatics and Computer Science)*, 3(1), pp. 24-27.
[16] Widianta, M.M.D et al. (2018). Comparison of Multi-Criteria Decision Support Methods (AHP, TOPSIS, SAW & PROMETHEE) for Employee Placement. *Journal of Physics: Conference Series*, 953 [online] IOP Publishing. Available at: https://iopscience.iop.org/article/10.1088/1742-6596/953/1/012116/pdf [Accessed 7 March 2022].
[17] Zavadskas, E. K., Turskis, Z., Tamošaitiene, J. and Marina, V. (2008). Multicriteria selection of project managers by applying grey criteria. *Technological and Economic Development of Economy*, 14(4), pp. 462–477.

**Author's address**
Kateřina Svitková
Prague University of Economics and Business, Faculty of Informatics and Statistics, Dpt. of Econometrics
W. Churchill Sq. 4, 130 67 Prague 3
Czech Republic
email: svik02@vse.cz

Ing. Martina Kuncová, Ph.D.
Prague University of Economics and Business, Faculty of Informatics and Statistics, Dpt. of Econometrics
W. Churchill Sq. 4, 130 67 Prague 3
Czech Republic
email: kuncovam@vse.cz

Mgr. Jana Sekničková, Ph.D.
Prague University of Economics and Business, Faculty of Informatics and Statistics, Dpt. of Econometrics
W. Churchill Sq. 4, 130 67 Prague 3
Czech Republic
email: jana.seknickova@vse.cz

# LINEAR-EXPONENTIAL ADJUSTMENT COSTS IN THE FORM OF LINEAR SPECIFICATION

***Karol Szomolányi, Martin Lukáčik, Adriana Lukáčiková***
***University of Economics in Bratislava***

**Abstract**
A rockets and feathers hypothesis states that output price reactions are prompt when an input price rises, and output price reactions are small when an input price falls. Many method approaches study the hypothesis. This paper focuses on the method based on the linear-exponential form of the firm's adjustment costs. The first-order conditions of the firm expand to the linear econometric specification of input price reactions. The theory supposes coefficient restrictions. The paper verifies the supposition in Slovak and U.S. retail gasoline markets. The orthogonality conditions implied by the rational expectation hypothesis makes the generalized method of moments a natural candidate for estimating equation. The choice of instruments corresponds to the orthogonality test, the exogeneity test, and the weakness of instruments test. Estimating the equation and testing the restrictions, the results support the theory in both markets.

***Keywords:*** *Linear-exponential adjustment costs, Generalized method of moments, Restriction test*

***JEL Classification:*** C26, C51, Q41
***AMS Classification:*** 62P20

## 1 INTRODUCTION

A rockets and feathers hypothesis states that output price reactions are prompt when an input price rises, and output price reactions are small when an input price falls. A typical example of the rockets and feathers hypothesis realization is in a fuel market. Borenstein, Cameron and Gilbert (1997) referred that U.S. gasoline reactions are more extensive when crude oil price increases.

Many papers studying the hypotheses use various methods: quadratic adjustment function (Bacon, 1991), distributed lag model (Karrenbrock 1991), first differences model (Duffy-Deno 1996), cointegration techniques (Borenstein, Cameron and Gilbert 1997), Markov-switching regime model (Boroumand et al. 2016) and vector autoregressive model (Balke, Brown and Yucel 1998 and Kang, Gracia and Ratti 2019). Recently, authors of empirical studies have used threshold models with many regimes (Douglas 2010 and Bagnai and Ospina 2018).

Szomolányi, Lukáčik and Lukáčiková (2020, 2022) derived a specification of the reactions of fuel prices to changes in crude oil prices from the linear-exponential adjustment cost function. By the approach, this paper presents the reaction function in the form of linear econometric specification. A test of the coefficient restriction confirms the theory based on the linear-exponential adjustment cost function.

# 2 MODEL AND METHODS

A reaction function derived from the linear-exponential adjustment cost function is in the form (Szomolányi, Lukáčik and Lukáčiková 2020, 2022):

$$\Delta p_t = k\Delta c_t - \frac{1}{2}\gamma\Delta\left[\left(p_t - kc_t\right)^2\right] + u_t \tag{1}$$

where $p$ denotes an output price, $c$ denotes an input price, $u$ denotes a stochastic term, $k$ denotes the technology coefficient, and $\gamma$ is the asymmetry term. The first-difference operator is $\Delta$. The rockets and feathers hypothesis supposes a negative value of the asymmetry coefficient $\gamma$. For $\gamma = 0$, the specification (1) is linear, and output price reactions are symmetric.

The specification (1) expands to a linear form:

$$\Delta p_t = \beta_0 + \beta_1\Delta c_t + \beta_2\Delta\left(p_t^2\right) + \beta_3\Delta\left(p_t c_t\right) + \beta_4\Delta\left(c_t^2\right) + u_t \tag{2}$$

The specification (2) corresponds to (1) if the following coefficient restrictions hold:

$$\beta_0 = 0,\ \beta_3 = -2\beta_1\beta_2 \text{ and } \beta_1^2\beta_2 = \beta_4 \tag{3}$$

The orthogonality conditions implied by the rational expectation hypothesis make the generalized method of moments (GMM) a natural candidate for estimating equation (2). An advantage of (2) is that it can apply well-known linear instrument tests. These are the Hausman test of the endogeneity of regressors (Hausman 1978), the Cragg and Donald *F* statistic, and the Stock and Yogo bias critical values of instrument weakness (Cragg and Donald 1993 and Stock and Yogo 2005). The paper results report the Cragg and Donald statistic only for comparative purposes.

The paper estimates standard errors using the Newey and West procedure. The most important feature of the procedure explained by Newey and West (1987) is its consistency in the presence of both heteroskedasticity and the autocorrelation of unknown forms.

The nonlinear coefficient restriction test (3) uses test statistics subject to a $\chi 2$ asymptotic distribution with degrees of freedom equal to the number of restrictions (3 in this case).

# 3 DATA

This work applies the linear specification (2) to both Slovak and U.S. retail gasoline markets. Statistical Office of Slovak Republic provides weekly time series of Slovak retail gasoline prices in euros per liter.

The remaining energy price data series come from U.S. Energy Information Admini-stration – the agency responsible for collecting, analyzing, and disseminating energy information. The series includes weekly Europe Brent Spot Price FOB Dollars per Barrel, daily Cushing OK WTI Spot Price FOB in Dollars per Barrel, weekly U.S. Regular All Formulations Retail Gasoline Prices in Dollars per Gallon.

The conversion of Brent oil prices to euros per liter uses daily reference exchange rates gathered from the European Central Bank, and 1 barrel equals 158.99 liters. The corresponding weekly exchange rate is computed by averaging. The data corresponding to the Slovak market pertains to the period from the first week of 2009 till the second week of 2019; 524 observations are available.

The conversion of daily WTI oil prices to weekly prices in Dollars per gallon uses averaging, and 1 barrel equals 42 gallons. The U.S. market data period is from 8/20/1990 to 7/15/2019; 1503 observations are available.

# 4 RESULTS

The estimates of specification (2) in Slovak and U.S. markets follow. The Slovak gasoline reactions represents an equation in the form:

$$\begin{aligned}\Delta\hat{p}_t = &\ 2.49\times10^{-5} + 1.186\times\Delta c_t + 0.563\times\Delta\left(p_t^2\right) - 0.009\times\Delta\left(p_t c_t\right) + 0.900\times\Delta\left(c_t^2\right)\\ &\left(4.21\times10^{-5}\right)\left(0.117\right) \qquad \left(0.023\right) \qquad \left(-0.001\right) \qquad \left(0.138\right)\end{aligned}$$

Apart from the level coefficient, all the estimated coefficients are statistically significant at a level of 1%. The estimate uses the instrument set $\{\Delta p_{t-1}, \Delta c_{t-1}, \Delta(p_{t-1}c_{t-1}), \Delta(p_{t-1}^2), \Delta(c_{t-1}^2)\}$. The corresponding $J$ statistics equals 0.08, implying the orthogonality of instruments. The study employs the endogeneity test by estimating two versions of the specification (2). In the first one, explanatory variables were considered to be endogenous and so the original instrument set $\{\Delta p_{t-1}, \Delta c_{t-1}, \Delta(p_{t-1}c_{t-1}), \Delta(p_{t-1}^2), \Delta(c_{t-1}^2)\}$ was used. In the second one, the explanatory variables were exogenous, and the original instrument set was extended by the explanatory variables. The difference in the $J$ statistics with $\chi 2$ asymptotic distribution and four degrees of freedom (number of explanatory variables) equals 22.675, rejecting the exogeneity of explanatory variables. The result suggests that using instrumental methods such as GMM is proper to estimate the (2) specification.

The Cragg-Donald $F$ statistics of the instrument weakness test equals 6.361. Ensuring that instruments are not weak, the study applied first-stage regressions for endogenous right-hand-side variables of the linear specification on a constant and the instrument set. The corresponding $F$ statistics are higher than 10 in all regressions.

The coefficient signs support the rockets and feathers hypothesis. Applying the restriction test (3), testing statistics with $\chi 2$ asymptotic distribution and three degrees of freedom (number of restrictions) equals 1.091, not rejecting the hypotheses. The result strongly supports the theoretical model of adjustment costs in linear-exponential form.

An equation for U.S. retail gasoline market is in the form:

$$\Delta\hat{p}_t = -2.77\times10^{-6} + 1.373\times\Delta c_t + 0.649\times\Delta\left(p_t^2\right) - 1.498\times\Delta\left(p_t c_t\right) + 0.749\times\Delta\left(c_t^2\right)$$
$$\left(6.31\times10^{-4}\right)\left(0.293\right) \qquad \left(0.069\right) \qquad \left(0.231\right) \qquad \left(0.179\right)$$

All the coefficients, but the level one, are statistically significant at a 1% level. $J$ statistics equals 0.398, implying the orthogonality of the used instrument set $\{\Delta p_{t-1}, \Delta c_{t-1}, \Delta(p_{t-1}^2), \Delta(c_{t-1}^2), \Delta(cb_{t-1})\}$, where $cb$ denotes Brent oil prices. The Cragg and Donald $F$ statistic equals 4.3033; the $F$ statistics of first-stage regressions are higher than 10 for all endogenous right-hand-side variables. The difference in $J$ statistics with $\chi 2$ asymptotic distribution and 5 degrees of freedom when considering the regressors to be exogenous and endogenous is 47.28, rejecting the exogeneity.

The U.S. estimate confirms the theory as well. The coefficient signs correspond to the rockets and feathers hypothesis. The restriction test (3) statistics equals 0.67.

## 5 CONCLUSION

Solving a firm problem with linear-exponential adjustment cost formulations expands to the linear specification (2). The theory based on the rockets and feathers hypothesis supposes the exact coefficient restriction (3). Using Slovak and U.S. data, GMM, and proper instrument sets, the estimates of (2) correspond to hypothesis (3) and the rockets and feathers theory. The instrument set choice corresponds to the well-known tests. These are the orthogonality test, the exogeneity test, and the weakness of instruments test.

**Acknowledgements:** This work was supported by The Grant Agency of Slovak Republic - 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] Bacon, R. W. (1991). Rockets and feathers: the asymmetric speed of adjustment of UK retail gasoline prices to cost changes. *Energy Economics*, 13(3), pp. 211-218.

[2] Bagnai, A. and Ospina, C. A. M. (2018). Asymmetries, outliers and structural stability in the US gasoline market. *Energy Economics*, 69, pp. 250-260.

[3] Balke, N. S., Brown, S. P. and Yucel, M. K. (1998). Crude oil and gasoline prices: an asymmetric relationship?. *Economic and Financial Policy Review*, 1998(Q1), pp. 2-11.

[4] Borenstein, S., Cameron, A. C. and Gilbert, R. (1997). Do gasoline prices respond asymmetrically to crude oil price changes? *The Quarterly Journal of Economics*, 112(1), pp. 305-339.

[5] Boroumand, R. H., Goutte, S., Porcher, S. and Porcher, T. (2016). Asymmetric evidence of gasoline price responses in France: A Markov-switching approach. *Economic Modelling*, 52, pp. 467-476.

[6] Cragg, J. G. and Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, 9(2), pp. 222-240.

[7] Douglas, C. C. (2010). Do gasoline prices exhibit asymmetry? Not usually!. *Energy Economics*, 32(4), pp. 918-925.

[8] Duffy-Deno, K. T. (1996). Retail price asymmetries in local gasoline markets. *Energy Economics*, 18(1-2), pp. 81-92.
[9] Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, pp. 1251-1271.
[10] Karrenbrock, J. D. (1991). The behavior of retail gasoline prices: symmetric or not?. *Federal Reserve Bank of St. Louis Review*, 73(4), pp. 19-29.
[11] Kang, W., de Gracia, F. P. and Ratti, R. A. (2019). The asymmetric response of gasoline prices to oil price shocks and policy uncertainty. *Energy Economics*, 77, pp. 66-79.
[12] Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), pp. 703-708.
[13] Stock, J. H. and Yogo, M. (2005). 'Testing for weak instruments in linear IV regression' in Andrews, D. W. K., and Stock, J. H. (eds.) *Identification and inference for econometric models: essays in honor of Thomas J. Rothenberg*. Cambridge: Cambridge University Press, pp. 80-108.
[14] Szomolányi, K., Lukáčik, M. and Lukáčiková, A. (2020). Asymmetric retail gasoline and diesel price reactions in slovak market. *Ekonomický časopis*, 68(2), pp. 115-133.
[15] Szomolányi, K., Lukáčik, M. and Lukáčiková, A. (2022). Estimation of asymmetric responses of U.S. retail fuel prices to changes in input prices based on a linear exponential adjustment cost approach. *Central European Journal of Operations Research*, 30(2), pp. 757-779.

**Authors address**

doc. Ing. Karol Szomolányi, PhD.
University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics
Dolnozemská 1/b, 852 35 Bratislava
Slovak Republic
email: karol.szomolanyi@euba.sk

prof. Ing. Martin Lukáčik, PhD.
University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics
Dolnozemská 1/b, 852 35 Bratislava
Slovak Republic
email: martin.lukacik@euba.sk

Ing. Adriana Lukáčiková, PhD.
University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics
Dolnozemská 1/b, 852 35 Bratislava
Slovak Republic
email: adriana.lukacikova@euba.sk

# DESIGN OF OPTIMIZATION MODEL FOR NETWORK COORDINATION OF TRANSFERS BETWEEN CONNECTIONS - LINE NETWORK WITH APPLICATION OF HOMOGENEOUS HEADWAY

*Dusan Teichmann, CTU in Prague*
*Denisa Mockova, CTU in Prague*
*Michal Dorda, VSB - Technical University of Ostrava*
*Pavel Edvard Vancura, CTU in Prague*
*Ivana Olivkova, VSB - Technical University of Ostrava*
*Vojtech Graf, VSB - Technical University of Ostrava*

**Abstract**
The basic aspect of public transport in terms of its use must be its attractiveness to passengers. Attractiveness is characterized by many factors. One of them is, for example, the offer of a direct connection. However, it is not always possible to provide a direct connection. In this case, passenger transfers cannot be avoided. During transfers, it is important that passengers do not wait unnecessarily for consecutive connections. The presented article deals with the issue of network coordination of tram connections in the Ostrava public transport system. The optimization criterion is the total time loss of transferring passengers in all nodes in which there is a transfer between tram lines.

***Keywords:*** *Network coordination, Linear programming, Optimization, Public transport*

***JEL Classification:*** *C610*
***AMS Classification:*** *90C05*

## 1 INTRODUCTION – OUR MOTIVATION TO DEAL WITH THE PROBLEM AND STATE OF THE ART

Time coordination of connections is a hot topic in public transport. The issue finds its application in all modes of transport - in rail transport (Long et al., 2020), bus transport (Wu et al., 2019) or tram transport (Ji et al., 2019). The most common optimization methods used to solve time coordination problems are mathematical programming (Cao, Tang and Gao, 2020), (Gábrišová, 2010), (Janáčková and Szendreyová, 2010), (Ji et al., 2019), (Kozel and Michalcová, 2017), (Long et al., 2020), (Teichmann et al., 2021), (Wu et al., 2019), different heuristics (Černá and Černý, 2004), (Černá and Černý, 2014), (Černý and Kluvánek, 1991), genetic algorithms (Cao, Tang and Gao, 2020), colored Petri nets and Max plus algebra (Idel Mahjoub, Chakir El-Alaoui and Nait-Sidi-Moh, 2020).

In practice, two basic types of time coordination of public transport connections are known - node coordination and section coordination. In the case of node coordination, the time loss of transferring passengers is usually minimized. In the case of section coordination, the maximum time between consecutive connections is usually minimized or the minimum time between consecutive connections is maximized (Černá and Černý, 2004), (Černá and Černý, 2014) or (Černý and Kluvánek, 1991). When minimizing the maximum time or maximizing the minimum time between consecutive connections, it is also possible to use the cascade approach – see for example (Janáčková and Szendreyová, 2010). However, it is also possible to use the time loss of waiting passengers who come to a public transport stop without knowledge of the timetable as an optimization criterion. The goal of optimization is to minimize it. Because the time loss of a waiting passenger is a nonlinear function, it is possible to make its partial linearization (Gábrišová, 2010). The task of section coordination can be conceived as one-way or two-way (Kozel and Michalcová, 2017). A very promising area is the union of node and section coordination, to which, for example, the publication (Teichmann et al., 2021) is dedicated. Time coordination of connections can also bring other effects. One of them is, for example, the minimization of the number of shunting tracks at tram turning loops, as shown, for example, in (Teichmann et al., 2020).

# 2 PROBLEM FORMULATION AND ITS MATHEMATICAL MODEL

Let a set of transfer nodes $U$ be given. For each node $u \in U$, a set of lines $L_u$ is defined, whose connections should be coordinated in the node $u \in U$. Furthermore, for each line $i \in L_u$ a set of directions $S_i$ is defined, in which its connections are run. In the following text we assume that all the lines have two directions, i.e., the sets of possible directions are the same for all the lines. Therefore, it is possible to omit the index in the case of sets $S_i$ and we will use the simplified notation $S$ in the following text.

Each request on coordination in the solved network is defined by an ordered hexad $[u; i; l; j; s; f]$. The first symbol in the ordered hexad identifies the transfer node $u \in U$. The second symbol represents the line number $i \in L_u$ and the third symbol represents the direction $l \in S$ from which the transfer is requested. The fourth and fifth symbols indicate the line number $j \in L_u$ and the direction $s \in S$ to which the transfer is requested. It is logical that it must hold that $i \neq j$, because the coordination of transfers between connections of the same lines is not relevant in practice. However, in the case of directions, it may be true that $l = s$ – transfers between different lines to the same direction may be required. The last symbol in the hexad represents the volume of passengers who transfer in the node $u \in U$ from the connection of the line $i \in L_u$ going in the direction $l \in S$ to the connections of the line $j \in L_u$ going in the direction $s \in S$. For each line $i \in L_u$ whose connections serve the transfer node $u \in U$ in the direction $l \in S$, the headway (a regular time interval between two connections of the same line serving the route in the same direction) $T_i$ and the earliest possible service times of the transfer node by two connections $t_{uil1}$ and $t_{uil2}$ of a given line are defined. It must be satisfied that $t_{uil2} = t_{uil1} + T_i$.

For each pair of lines $i \in L_u$ and $j \in L_u$, where $u \in U$, the value of the minimal transfer time $tprest_{uij}$ is defined (it is assumed that the value of the minimal transfer time between connections of coordinated lines does not depend on which pair of connections it is) and the volume of passengers $f_{uiljs}$ transferring (in the chosen time period) from the connections of the line $i \in L_u$ going in the direction $l \in S$ to the connections of the line $j \in L_u$ going in the direction $s \in S$ is known.

Our task is to decide on time offsets of the lines in individual directions coordinated in the given transfer nodes so that the offsets of connections of individual lines running in the same direction are uniform (to preserve the values of pre-defined headways on the lines) and at the same time we want to minimize the total time loss (CCZ) of all transferring passengers.

To simplify the mathematical notation of the optimization criterion, we first introduce the incidence matrix $\boldsymbol{A}$. If it is required that there is coordination in the node $u \in U$ between the connections of the line $i \in L_u$ going in the direction $l \in S$ and the connections of the line $j \in L_u$ going in the direction $s \in S$, then $a_{uiljs} = 1$, otherwise $a_{uiljs} = 0$.

Due to applying of a headway and the impossibility of guaranteeing all transfers between the pairs of the first connections, a pair of connections is always introduced for each line whose connections are mutually coordinated. By using the second connection it is possible to ensure the continuity of the first connection, from which it is transferred, which arrived at the transfer node after the departure of the first connection of the line to which it is transferred.

To model all the decisions, we further introduce some variables with the following meaning and their domains of definition into the optimization problem - see Table 1.

Table 1: Summary of the variables, their meaning, and domains of definition

| Variable | Meaning of the variable | Domain of definition |
| --- | --- | --- |
| $x_{il}$ | A time offset of all connections of the line $i \in L_u$ in the direction $l \in S$ calculated from their earliest possible time positions. | $R_0^+$ |
| $h_{uiljs}$ | A time loss of each passenger who transfers in the node $u \in U$ from the connection of the line $i \in L_u$ going in the direction $l \in S$ to the nearest connection of the line $j \in L_u$ going in the direction $s \in S$. | $R_0^+$ |
| $z_{uiljsk}$ | An auxiliary variable modelling the creation of a transfer link between the first connection of the line $i \in L_u$ going in the direction $l \in S$ to the nearest connection $k \in \{1; 2\}$ of the line $j \in L_u$ going in the direction $s \in S$ via the node $u \in U$. | $\{0; 1\}$ |

Mathematical model of the problem can be written in the following form:

$$min\, f(x,h,z) = \sum_{u \in U} \sum_{i \in L_u} \sum_{l \in S} \sum_{\substack{j \in L_u \\ j \neq i}} \sum_{s \in S} a_{uiljs} \cdot f_{uiljs} \cdot h_{uiljs} \quad (1)$$

subject to:

$$\left[t_{ujs1} + (k-1) \cdot T_j + x_{js}\right] - (t_{uil1} + x_{is}) \geq M\left(z_{uiljsk} - 1\right) \quad \text{for } u \in U, i \in L_u, j \in L_u, l \in S, s \in S, k \in \{1; 2\}, a_{uiljs} = 1 \quad (2)$$

$$\left[t_{ujs1} + (k-1) \cdot T_j + x_{js}\right] - (t_{uil1} + x_{is}) \leq h_{uiljs} + M\left(1 - z_{uiljsk}\right) \quad \text{for } u \in U, i \in L_u, j \in L_u, l \in S, s \in S, k \in \{1; 2\}, a_{uiljs} = 1 \quad (3)$$

$$\sum_{k \in \{1;2\}} z_{uiljsk} = 1 \quad \text{for } u \in U, i \in L_u, j \in L_u, l \in S, s \in S, a_{uiljs} = 1 \quad (4)$$

$$x_{il} \leq a_{il} \quad \text{for } i \in L_u, l \in S \quad (5)$$

$$x_{il} \in R_0^+ \quad \text{for } i \in L_u, l \in S \quad (6)$$

$$h_{uiljs} \in R_0^+ \quad \text{for } u \in U, i \in L_u, j \in L_u, l \in S, s \in S, a_{uiljs} = 1 \quad (7)$$

$$z_{uiljsk} \in \{0; 1\} \quad \text{for } u \in U, i \in L_u, j \in L_u, l \in S, s \in S, a_{uiljs} = 1 \quad (8)$$

Formula (1) represents the optimization criterion - the total time loss of all transferring passengers. The group of constraints (2) ensures that in case of inadmissibility of the time positions of the connections of the coordinated lines going in the affected directions, there will be no transfer link. The group of constraints (3) quantifies the time loss of the transferring passenger generated by the creation of the transfer link. The group of constraints (4) ensures the creation of transfer links. The group of constraints (5) guaranties that any time shift of the connections, generated in order to reduce the total time loss, will not exceed its maximum value. Constraints (6) - (8) define the domains of definition of the variables used in the model.

## 3 CALCULATION EXPERIMENTS WITH THE MATHEMATICAL MODEL - CASE STUDY

The article presents a computational experiment focused on network coordination of connections in the Ostrava public transport tram network. The experiment is realised in conditions of weekend days (Saturdays and Sundays), when, due to the lower number of connections, the importance of time coordination increases.

Although the number of tram lines operated during the day (approximately from 4:00 to 23:00) is 14, only selected tram lines are subject to coordination, namely lines 1, 2, 3, 4, 7, 8, 11, 12 and 17. Since the tram lines are not circular and, therefore, the connections of these lines must be coordinated in both directions, it is necessary to distinguish the individual directions of the lines in the model – see Table 2.

Table 2: Directions of the coordinated lines

| The number of the line | The number of the direction | First stop | Last stop |
|---|---|---|---|
| 1 | 1 | Dubina | Hlavní nádraží |
| | 2 | Hlavní nádraží | Dubina |
| 2 | 1 | Výškovice | Hlavní nádraží |
| | 2 | Hlavní nádraží | Výškovice |
| 3 | 1 | Prostorná (Hulváky) | Dubina |
| | 2 | Dubina | Prostorná (Hulváky) |
| 4 | 1 | Martinov | Hranečník/Nová Huť |
| | 2 | Hranečník/Nová Huť | Martinov |
| 7 | 1 | Poruba, Vřesinská | Výškovice |
| | 2 | Výškovice | Poruba, Vřesinská |
| 8 | 1 | Poruba, Vřesinská | Hlavní nádraží/Přívoz, Hlučínská |
| | 2 | Hlavní nádraží/Přívoz, Hlučínská | Poruba, Vřesinská |
| 11 | 1 | Zábřeh | Prostorná |
| | 2 | Prostorná | Zábřeh |
| 12 | 1 | Dubina | Hranečník |
| | 2 | Hranečník | Dubina |
| 17 | 1 | Poruba, Vřesinská | Dubina |
| | 2 | Dubina | Poruba, Vřesinská |

We must assign an identification number to each transfer node due to its unique identification in the model. The list of transfer nodes in the Ostrava public transport tram network, the identification numbers assigned to them and the list of coordination requirements in the transfer nodes are given in Table 3.

In case of tram lines whose connections are coordinated simultaneously in several nodes of the tram network, travel times for individual route sections of the tram network, bounded by the transfer nodes, are also important for the optimization calculation – see Table 4. The same travel times in both directions are considered.

Because the connections of the coordinated tram lines are run with the line headway of 20 minutes in the coordination period, it is always possible to use only a pair of connections for optimization. Thus, a pair of connections will be established for each coordinated tram line, the transfer will be always enabled from the first connection of the given line, either to the first or to the second connection of another coordinated line. The restriction on the pair of connections of coordinated lines does not affect the quality of the obtained results of the optimization calculation, even in a situation where the volumes of transferring passengers change during the day. It is sufficient to replace the volume of transferring passengers between the two connections by the volume of transferring passengers for the whole coordination period.

Table 3: List of the transfer nodes in the Ostrava public transport tram network

| The transfer node | From the connections of the line… | From the direction | To the connections of the line… | To the direction |
|---|---|---|---|---|
| Svinov, mosty (1) | 7 | Výškovice | 4 | Martinov |
| | 4 | Martinov | 7 | Výškovice |
| Nová Ves, vodárna (2) | 17 | Dubina | 4/8 | Karolina |
| | 4/8 | Karolina | 17 | Dubina |
| Prostorná (3) | 11 | Zábřeh | 8 | Hlavní nádraží |
| | 8 | Hlavní nádraží | 11 | Zábřeh |
| Karolina (4) | 1 | Dubina | 4 | Hranečník |
| | 4 | Hranečník | 1 | Dubina |
| | 8 | Poruba | 12 | Hranečník |
| | 12 | Hranečník | 8 | Poruba |
| Hulvácká (5) | 17 | Poruba | 11 | Zábřeh |
| | 11 | Zábřeh | 17 | Poruba |
| Palkovského (6) | 7 | Výškovice | 12 | Hranečník |
| | 12 | Hranečník | 7 | Výškovice |
| | 7 | Poruba | 12 | Dubina |
| | 12 | Dubina | 7 | Poruba |
| Vítkovice, Mírové náměstí (7) | 3 | Dubina | 2 | Hlavní nádraží |
| | 2 | Hlavní nádraží | 3 | Dubina |
| Kino Luna (8) | 11 | Zábřeh | 2 | Hlavní nádraží |
| | 2 | Hlavní nádraží | 11 | Zábřeh |

Table 4: Travel times between the transfer nodes

| The number of the line | The route section | The travel time [min] |
|---|---|---|
| 2 | Kino Luna – Mírové náměstí | 12 |
| 4 | Svinov, mosty – Nová Ves, vodárna | 2 |
| | Nová Ves, vodárna – Karolina | 10 |
| 7 | Svinov, mosty – Palkovského | 11 |
| 8 | Nová Ves, vodárna – Prostorná | 3 |
| | Prostorná – Karolina | 7 |
| 11 | Kino Luna – Hulvácká | 15 |
| | Hulvácká – Prostorná | 7 |
| 12 | Palkovského – Karolina | 16 |
| 17 | Nová Ves, vodárna – Hulvácká | 3 |

In the optimization model, the initial departure time of connections in the earliest possible time positions is considered. When identifying the earliest possible time positions of the connections of the coordinated lines, it is necessary to consider the travel times of the connections of the coordinated tram lines between the individual transfer nodes.

When designing the mathematical model, it was first considered that the basic reference time of the coordinated line, from which other times will also be derived, will be the departure time of the first line from the first stop at time 00. However, this proved not entirely appropriate because in conditions of the Ostrava public transport tram network, it is common that especially in the peripheral parts of the city (outside the area in which the transfer nodes are located) the lines run along different routes with different lengths.

To coordinate the connections, the reference time points were therefore located to the first transfer node on the route of the given line in the individual directions. Thus, in the optimization model, the earliest possible time position of the coordinated connection in each direction in the first transfer node equals to 00. If there is only one transfer node on the tram line route, then time 00 is the reference time point of the line in both directions. If there are several transfer nodes on the tram line route, then the reference time point for both directions of the line is time 00 in the first transfer node, where the connections of the given tram line are coordinated.

In case of some coordinated tram lines, the sum of the travel times between some transfer nodes at which the connections of the given line are coordinated exceeds the line headway. A typical case of such tram line in Ostrava is tram line 11, for which it holds that the travel time from the first (Kino Luna) to the last transfer node (Prostorná) equals to 22 minutes. The earliest possible service time of the transfer node of the tram line 11 is 02.

When calculating the value of the optimization criterion, all the volumes of transferring passengers for the individual transfer nodes were assumed to be 1. This corresponds to the situation where all transfers are assigned the same importance (preference).

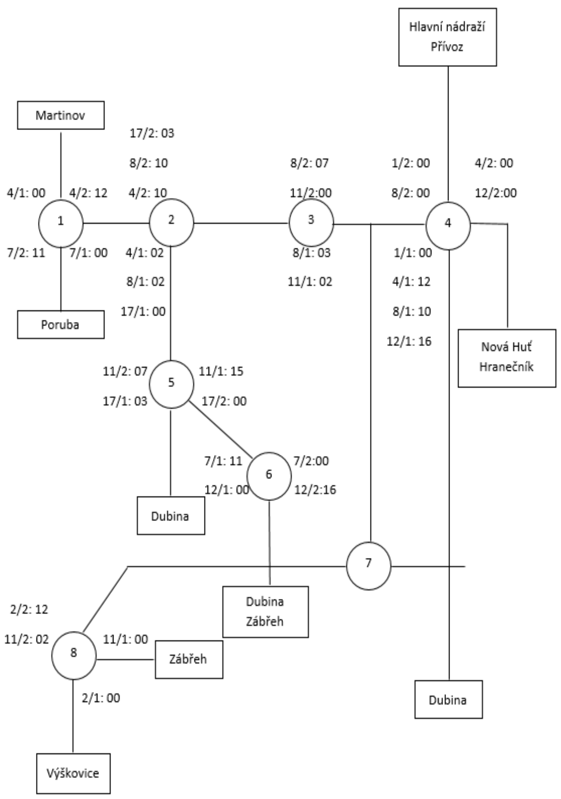



Figure 1: Test network with positions of the transfer nodes and the earliest possible service times of the transfer nodes

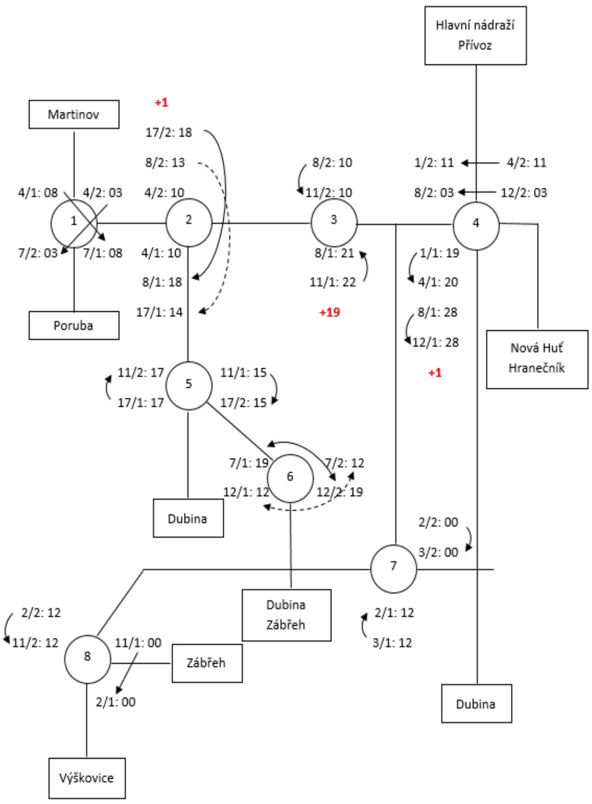



Figure 2: Results of the optimization experiment depicted in the test network

A fragment of the Ostrava public transport tram network - the part with the transfer nodes, for which the presented model was tested, is shown in Figure 1. Dimensions related to individual transfer nodes express the first possible time positions of the first connections in the individual transfer nodes, where the connections of the tram lines should be coordinated. The notation used in Figure 1 has the following meaning. For example, in case of transfer node 1 (Svinov, mosty) the symbol 4/1: 00 means the earliest possible arrival time of the first connection of line 4 in direction 1 (Martinov - Nová Huť) to the transfer node occurs at time 00 (the second connection arrives at the transfer node in case of the line headway of 20 minutes at time 00 + 20 = 20).

The achieved solution for the test network of the Ostrava public transport tram lines with positions of the transfer nodes and the service times of the transfer nodes by the tram lines with required transfers after coordination is shown in Figure 2. Red numbers in Figure 2 represent the time loss of each passenger. The time loss therefore occurs during the transfers at transfer node 2 (Nová Ves, vodárna) in the amount of 1 minute when changing from the connection of line 8 in direction 2 (from the Ostrava centre) to the connection of line 17 in direction 1 (to Dubina), in transfer node 3 (Prostorná) in the amount of 19 minutes when changing from the connection of line 11 in direction 1 (the route of the line ends at the stop Prostorná) to the connection of line 8 in direction 1 (Hlavní nádraží) and in transfer node 4 (Karolina) in the amount of 1 minute when changing from the connection of line 1 in direction 1 (Hlavní nádraží) to the connection of line 4 in direction 1 (Hranečník / Nová Huť). The total time loss is therefore 21 minutes, which corresponds to the value of the optimization criterion of 21 minutes.

# 4 CONCLUSIONS

The presented article deals with the network coordination of connections in public transport. Specifically, it deals with the mathematical model for the coordination of connections in the Ostrava tram network during weekend days. A specific feature of the tram traffic in Ostrava is the regular line headway 20 minutes which is applied to all lines operated during the weekend. This fact allowed us to reduce the optimization model. The reduction is reflected mainly in the number of the coordinated connections. A pair of connections was established for each coordinated line. The subject of coordination is one connection on arrival and two connections on departure from a given transfer node. When designing the initial time positions of individual connections coordinated in several transfer nodes, it is necessary to respect the travel times between the individual transfer nodes.

There are 9 transfer nodes in the Ostrava tram network. In the mathematical model, a separate group of constraints is created for each transfer node. They ensure the creation of the transfer links and the calculation of the time loss arising during transfers of passengers. The total time loss of passengers is 21 minutes after optimization. The computational experiments confirmed the functionality of the presented model.

The calculation experiment must be understood as preliminary only, because all the transfer links had the same weight in the optimization calculation (the volume of transferring passengers was always equal to 1). The same weight of the transfer links does not affect the demonstration of the functionality of the model.

In the future research, our attention will be focused on the coordination of lines on which heterogenous headway between connections is applied. An example of such case can be found in public bus transport in the city of Prague. Bus connections are run with alternating headways of 7 and 8 minutes. If the first connection of a line in a certain direction serves a transfer node at time 00, then the next connections of the same line serve the transfer node in the same direction at times 07, 15, 22, 30, 37 etc. or 08, 15, 23, 30, 38 etc.

**Acknowledgements:** This work was supported by project CK01000043 System for supporting network time coordination of connections at interchange nodes.

## References


[1] 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.
[2] Černá, A. and Černý, J. (2004). *Teorie řízení a rozhodování v dopravních systémech*. Pardubice: Institut Jana Pernera.
[3] Černá, A. and Černý, J. (2014). *Manažerské rozhodování o dopravních systémech*. Pardubice: Institut Jana Pernera.
[4] Černý, J. and Kluvánek, P. (1991). *Základy matematickej teorie dopravy*. Bratislava: VEGA.
[5] Gábrišová, L. (2010). Presnosť riešenia vybraných nelineárních úloh při použití po čiastkách lineárnej aproximácie. In: *Conference Úlohy diskrétní optimalizace v dopravní praxi*. Pardubice, pp. 43-51.
[6] 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.
[7] Janáčková, M. and Szendreyová, A. (2010). Kaskádový prístup k riešeniu úloh typu maxmin. In: *Conference Úlohy diskrétní optimalizace v dopravní praxi*. Pardubice, pp. 94-102.
[8] 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.
[9] Kozel, P. and Michalcová, Š. (2017). Time coordination of bus arrivals on a set of bus stops. In: *12th International Conference on Strategic Management and its Support by Information Systems (SMSIS 2017)*. Ostrava, pp. 267-273.
[10] 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.
[11] Teichmann, D., Dorda, M., Mocková, D., Vančura, P. E., Graf, V. and Olivková, I. (2020). Minimizing Number of Shunting Tracks at Tram Turning Loops. In: *Proceedings of 38th International Conference on Mathematical Methods in Economics (MME 2020)*. Brno, pp. 585-591.
[12] Teichmann, D., Dorda, M., Mocková, D., Vančura, P. E., Graf, V., Olivková, I. (2021). Combined Time Coordination of Connections in Public Transport. In: *Proceedings of 39th International Conference on Mathematical Methods in Economics (MME 2021)*. Praha, pp. 487-492.
[13] 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.


## Author's address


doc. Ing. Dusan Teichmann, Ph.D., doc. Ing. Denisa Mockova, Ph.D., Ing. Bc. Pavel Edvard Vancura, Ph.D.
CTU in Prague, Faculty of Transportation Sciences, Department of Logistics and Management of Transport, Horska 3, 128 03 Praha, Czech Republic
E-mail: teichdus@fd.cvut.cz, mockova@fd.cvut.cz, vancupav@fd.cvut.cz
doc. Ing. Michal Dorda, Ph.D., doc. Ing. Ivana Olivkova, Ph.D., Ing. Vojtech Graf, Ph.D.
VSB - Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15/2172, 708 00 Ostrava-Poruba, Czech Republic
E-mail: michal.dorda@vsb.cz, ivana.olivkova@vsb.cz, vojtech.graf@vsb.cz

ISBN: 978-80-89962-93-8 (print)
ISBN: 978-80-89962-94-5 (online)