

Katedra ekonometrie
Fakulty informatiky a statistiky VŠE v Praze



a

Katedra operačného výskumu a ekonometrie
Fakulty hospodárskej informatiky EU v Bratislave



a

Slovenská spoločnosť pre operačný výskum

MEZINÁRODNÍ VĚDECKÝ SEMINÁŘ NOVÉ TRENDY V EKONOMETRII A OPERAČNÍM VÝZKUMU

Katedry ekonometrie FIS VŠE v Praze

a

Katedry operačného výskumu a ekonometrie FHI EU v Bratislave

a

Slovenskej spoločnosti pre operačný výskum

ZBORNÍK

30. november – 2. december 2022

30. listopad – 2. prosinec 2022

Praha

Programový výbor:

prof. Ing. Ivan Brezina, CSc., EU Bratislava
prof. Ing. Mgr. Martin Dlouhý, Dr., MSc., VŠE Praha
prof. RNDr. Ing. Petr Fiala, MBA, CSc., VŠE Praha
prof. Ing. Josef Jablonský, CSc., VŠE Praha
prof. Ing. Martin Lukáčik, PhD., EU Bratislava
prof. Mgr. Juraj Pekár, PhD., EU Bratislava
Mgr. Jana Sekničková, Ph.D., VŠE Praha

Recenzenti:

Ing. Adam Borovička, Ph.D., VŠE Praha
Dr. h. c. prof. Ing. Michal Fendek, PhD., EU Bratislava
doc. Ing. Marián Goga, PhD., EU Bratislava
doc. Ing. Jaroslav Kultán, PhD., EU Bratislava
Ing. Tomáš Majer, PhD., UNIZA v Žiline
doc. Ing. Jan Zouhar, Ph.D., VŠE Praha

Organizačný výbor:

doc. Ing. Brian König, PhD.

AH 9,22

Kontakt: brian.konig@euba.sk

Editor: doc. Ing. Brian König, PhD.

Web editor: prof. Ing. Martin Lukáčik, PhD.

Za odbornú stránku a jazykovú úpravu zodpovedajú autori.

Fakulta informatiky a statistiky VŠE v Prahe,
nám. W. Churchilla 4, 130 67 Praha 3
a Fakulta hospodárskej informatiky EU v Bratislave,
Dolnozemska cesta 1, 852 35 Bratislava

© Ekonomická univerzita v Bratislave,
Vydavateľstvo EKONÓM, Bratislava 2022

ISBN: 978-80-225-5006-2

Obsah

<i>Zuzana Čičková, Patrícia Holzerová, Pavel Gežík:</i> <u>VIACKRITERIÁLNE BIMATICOVÉ HRY A PODMIENKY OPTIMÁLNOTI KARUSHA-KUHNA-TUCKERA</u>	5
<i>Ádám Csápai:</i> <u>UKÁŽKA VEKTOROVÉHO MODELU S KOREKČNÝM ČLENOM V SOFTVÉRI EVIEWS</u>	12
<i>Marek Daniel:</i> <u>ŘEŠENÍ ROZSÁHLÝCH DOPRAVNÍCH PROBLÉMŮ</u>	19
<i>Andrea Furková:</i> <u>MOTIVAČNÉ ASPEKTY K ŠTÚDIU NA VYSOKEJ ŠKOLE: KOMPARÁCIA EKONOMICKÝ VS TECHNICKÝ ŠTUDIJNÝ SMER</u>	27
<i>Dalibor Gonda, Tomáš Majer:</i> <u>MINIMALIZÁCII RIZIKA MEŠKANIA SPOJOV POMOCOU FUZZY ČÍSIEL</u>	35
<i>Jakub Hanousek:</i> <u>KONSTRUKCE KOMPOZITNÍHO INDIKÁTORU NA ZÁKLADĚ ANALÝZY OBALU DAT</u>	42
<i>Michaela Chocholatá:</i> <u>PRIESTOROVÁ ANALÝZA VÝSLEDKOV TESTOVANIA 9 Z MATEMATIKY V OKRESOCH SR PRED A PO COVID-19</u>	46
<i>Jaroslav Janáček, Marek Kvet:</i> <u>SOFISTIKOVANÉ HEURISTIKY PRO APROXIMACI PARETOVY FRONTY NÁVRHŮ VEŘEJNÉHO OBSLUŽNÉHO SYSTÉMU</u>	53
<i>Miroslava Jánošová:</i> <u>ZLOŽKY MEDZINÁRODNEJ INVESTIČNEJ POZÍCIE A ICH VÝVOJ VO VZŤAHU K HDP V OBDOBÍ ZMIEN V EKONOMIKE</u>	60
<i>Dominik Kavřík:</i> <u>POUŽITÍ ZOBECNĚNÝCH ADITIVNÍCH MODELŮ PRO PREDIKCI CEN OJETÝCH VOZŮ</u>	66
<i>Peter Knížat:</i> <u>FUNKCIONÁLNA ANALÝZA ROZPTYLU: PRÍPADOVÁ ŠTÚDIA Z POTRAVINÁRSKEHO PRIEMYSLU</u>	71
<i>Brian König:</i> <u>ODHAD DYNAMICKÉHO EFEKTU ČLENSTVA V EU POMOCOU GRAVITAČNÉHO MODELU</u>	78

<i>Petr Krautwurm:</i> <u>OČEKÁVANÉ INDIVIDUÁLNÍ NÁKLADY PROMISKUITY</u>	84
<i>Jindřich Lacko:</i> <u>SOLVING TRAVELING SALESMAN PROBLEM USING R IN CONTEXT OF {SF} & HERE API</u>	89
<i>Radoslav Lacko:</i> <u>VOLATILITA INDEXU S&P500 V DOBE ZVYŠUJÚCICH SA ÚROKOVÝCH SADZIEB</u>	95
<i>Martin Lukáčik, Adriana Lukáčiková, Karol Szomolányi:</i> <u>REPLIKOVATEĽNOSŤ VÝSKUMU A JEHO PUBLIKOVATEĽNOSŤ</u>	99
<i>Veronika Miťková:</i> <u>MAKROEKONOMICKÉ DÔSLEDKY EURÓPSKEHO EKOLOGICKÉHO DOHOVORU PRE SLOVENSKO</u>	105
<i>Vladimír Mlynarovič:</i> <u>VÝBER PORTFÓLIA A MIERY RIZIKA</u>	112
<i>Jakub Neugebauer:</i> <u>ANALÝZA VÝNOSŮ AKCIÍ NA ČESKÉM TRHU ZA POSLEDNÍCH 10 LET</u>	120
<i>Juraj Pekár, Ivan Brezina, Marian Reiff:</i> <u>MODEL VÝBERU PORTFÓLIA S VYUŽITÍM MIER RIZIKA CVAR A PRIEMERNÝ DRAW DOWN</u>	127
<i>Allan Jose Sequeira Lopez:</i> <u>OPAKOVANÉ HRY V TEÓRII HIER</u>	132
<i>Vojtěch Vávra:</i> <u>3D PROBLÉM OBCHODNÍHO CESTUJÍCÍHO</u>	137
<i>Lukáš Veverka:</i> <u>VYUŽITÍ LOGISTICKÉ REGRESE V CRM</u>	143

VIACKRITERIÁLNE BIMATICOVÉ HRY A PODMIENKY OPTIMÁLNOTI KARUSHA-KUHNA-TUCKERA

MULTICRITERIA BIMATRIX GAMES AND KARUSH-KUHN- TUCKER CONDITIONS

Zuzana Čičková, Patrícia Holzerová, Pavel Gežík

Abstrakt

V teórii hier sú bimaticové hry základnou hrou dvoch hráčov, ktorou modelujeme neantagonistický konflikt. Stanovenie equilibria v daných hrách je však pomerne komplikovaným problémom. Na rozdiel od štandardných bimaticových hier, viackriteriálne bimaticové hry umožňujú do rozhodovania zapojiť väčšie množstvo rozhodovacích kritérií pre oboch hráčov, čo síce približuje tieto hry požiadavkám praxe, súčasne však ešte viac komplikuje možnosti ich riešenia. V oboch prípadoch je stanovenie rovnovážnych stratégií hráčov vo všeobecnosti pomerne komplikované. Pre oba typy úloh však možno zapísať podmienky optimálnosti Karusha-Kuhna-Tuckera. Príspevok je zameraný na formuláciu týchto podmienok, čo vedie k úlohe nelineárneho programovania, ktorú možno použiť na stanovenie rovnovážnych stratégií hráčov v bimaticových hrách.

KLúčové slová: viackriteriálne hry, podmienky optimálnosti Karusha-Kuhna-Tuckera

Abstract

In game theory, bimatrix games are a basic two-player game by which we model non-antagonistic conflict. However, determining the equilibrium in the given games is a rather complicated. In contrast to standard bimatrix games, multi-criteria bimatrix games allow us to increase the number of criteria for both players in the decision-making process, which although brings these problems closer to practical requirements, but at the same time complicates their solving possibilities even more. In both cases, the determination of equilibrium strategies of the players are generally quite complicated. However, for both types of problems, the Karush-Kuhn-Tucker conditions can be written down. The paper is aimed at formulating these conditions. The result is a nonlinear programming problem, which can be used to determine the equilibrium strategies of players in bimatrix games.

Keywords: multicriteria games, Karush-Kuhn-Tucker conditions

1 ÚVOD

V teórii hier sú viackriteriálne hry (hry s vektorovými funkciami platieb) považované za rozšírenie štandardnej jednokriteriálnej hry. Rozšírenie spočíva v množstve kritérií, ktorými sa hráči ohodnocujú výsledky svojich rozhodnutí. Zatiaľ čo pri klasických hrách hráči zvažujú len jedno rozhodovacie kritérium (jednu funkciu platieb), pri viackriteriálnych hrách môžu mať hráči viac kritérií, ktoré ovplyvňujú ich rozhodovanie (viac funkcií platieb). Riešenie viackriteriálnych konfliktných situácií vychádza z princípov jednokriteriálnych hier, či už ide o hry v normálnom tvare, alebo tvare charakteristickej funkcie (Chobot, 1986).

Prvýkrát termín vektorové funkcie platieb použil vo svojej práci Blackwell (1956), ktorý bližšie popísal problém viackriteriálnych hier. Na riešenie takýchto hier bolo navrhnutých niekoľko možností. Shapley a Rigby (1959) vo svojej práci „Equilibrium Points in Games with Vector Payoffs“ predstavili základný koncept hľadania rovnovážnych bodov. Jedným zo známych spôsobov riešenia viackriteriálnych hier je proces parametrizácie, spočívajúci v transformácii viackriteriálnej hry na jednokriteriálnu (Chobot, 1986). V roku 1994 uviedli autori Kruš a Bronisz (1994) koncept skalarizácie (škálovania) hier, ktorý spočíva v priradovaní váh jednotlivým kritériám hráčov. Nakoľko však analytikom nie sú tieto váhy vopred známe a ich správne priradenie jednotlivým kritériám môže byť diskutabilné a zložité, autori Voorneveld, Grahn a Dufwenberg (2000) prišli s návrhom ideálnych rovnováh (stratégií). Tie predstavujú strategické profily, ktoré sú odolné voči jednostranným odchýlkam hráčov bez ohľadu na význam pripísaný kritériám. Jednoduchosť ich návrhu riešenia spočíva v tom, že horná hranica počtu skalarizácií je zároveň maximálnym počtom kritérií hráčov.

Vo všeobecnosti viackriteriálne rozhodovanie patrí k najzložitejším rozhodovacím procesom. Nie je preň charakteristický len počet kritérií ale aj vzťah medzi nimi a tieto kritériá sa zvyčajne nedajú jednoducho kvantitatívne sčítať (Zúna 2010).

Podmienky Karusha-Kuhna-Tuckera (KKT) (tiež známe ako podmienky Kuhna-Tuckera) sú pomienkami optimálnosti používané v oblasti riešenie nelineárnej extremalizácie s obmedzeniami (Fendek a Fendeková, 2008), (Akhbari a kol., 2015). Podmienky KKT boli pôvodne pomenované po Haroldovi W. Kuhnovi a Albertovi W. Tuckerovi, ktorý uverejnili podmienky v roku 1951 (Kuhn a Tucker, 1951). Podmienky nezávisle uviedol William Karush vo svojej diplomovej práci v roku 1939.

2 PODMIENKY OPTIMÁLNOTI KARUSHA-KUHNA-TUCKERA PRE BIMATICOVÉ HRY

Bimaticová hra je hra, kde vystupujú dvaja účastníci (hráči). Každý z hráčov volí nezávisle (bez informácie o voľbe protihráča) jednu z konečného počtu variantov správania (stratégií). Predpokladá sa, že záujmy hráčov nie sú diametrálne protikladné, t.j. zisk jedného z hráčov nemusí byť sprevádzaný stratou druhého a hráči sú inteligentní. Otázkou je akú stratégiu má hráč zvoliť tak, že voľbou inej stratégie nemôže zvýšiť svoj zisk (Goga, 2013, Chobot a kol.,1991).

Bimaticovú hru možno formalizovať takto: Nech $P=\{1,2\}$ je množina hráčov, z ktorých každý má konečnú množinu stratégií (X – hráč 1, Y – hráč 2}, t.j. hráč 1 volí $\mathbf{x} \in X$, hráč 2 volí $\mathbf{y} \in Y$. Množinu všetkých výsledkov hry možno označiť ako $(\mathbf{x}, \mathbf{y}) \in X \times Y$. Jednotlivé prvky množín X a Y možno usporiadať pomocou konečného počtu prirodzených čísel (prvky množiny X : $i = 1, 2, \dots, m$ a prvky množiny Y : $j = 1, 2, \dots, n$). Hodnoty hry pre hráča 1 možno zapísať do matice $\mathbf{A}_{m \times n} = \{a_{ij}\}$, kde a_{ij} udáva platbu hráča pri výsledku (i, j) . Výsledky hry pre hráča 2 zapisujeme do matice $\mathbf{B}_{n \times m} = \{b_{ji}\}$, kde b_{ji} udáva platbu hráča 2 pri výsledku (i, j) . Ak neexistuje riešenie hry v čistých stratégiách, hľadáme riešenie hry v stratégiách zmiešaných

Postup pri riešení bimaticových hier:

1. Eliminujeme dominované stratégie
2. Nájďme stĺpcové maximá v matici \mathbf{A} a riadkové maximá v matici \mathbf{B}^T , a identifikujeme rovnovážny bod (rovnovážne body), ak:
 - a) rovnovážny bod je práve jeden, potom je to rovnovážne riešenie hry
 - b) existuje viac rovnovážnych bodov a jeden z nich dominuje ostatné, potom je tento bod rovnovážnym riešením

- c) existuje viac rovnovážnych bodov a ani jeden nie je dominantný, hľadáme riešenie hry v zmiešaných stratégiách
- d) rovnovážny bod neexistuje, hľadáme riešenie hry v zmiešaných stratégiách

Ak neexistuje riešenie hry v čistých stratégiách, hľadáme riešenie hry v stratégiách zmiešaných, tj. určíme pravdepodobnostné rozdelenie výberu stratégií oboch hráčov.

Zmiešaná stratégia hráča 1: m -rozmerný vektor \mathbf{x} , $\sum_{i=1}^m x_i = 1, x_i \geq 0, i = 1, 2, \dots, m$

Zmiešaná stratégia hráča 2: n -rozmerný vektor \mathbf{y} , $\sum_{j=1}^n y_j = 1, y_j \geq 0, j = 1, 2, \dots, n$

Množina zmiešaných stratégií hráča 1: $S_m = \left\{ \mathbf{x}, \sum_{i=1}^m x_i = 1, x_i \geq 0, i = 1, 2, \dots, m \right\}$

Množina zmiešaných stratégií hráča 2: $S_n = \left\{ \mathbf{y}, \sum_{j=1}^n y_j = 1, y_j \geq 0, j = 1, 2, \dots, n \right\}$

Ak voľby stratégií \mathbf{x} a \mathbf{y} sú nezávislé náhodné javy, potom *stredná hodnota platby* (definovaná na $S_m \times S_n$) hráča 1:

$$E_1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j = \mathbf{x}^T \mathbf{A} \mathbf{y}$$

a hráča 2:

$$E_2(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^m \sum_{j=1}^n b_{ij} x_i y_j = \mathbf{x}^T \mathbf{B} \mathbf{y}$$

Rovnovážny bod bimaticovej hry je taká dvojica $\mathbf{x}^{(0)} \in S_m$ a $\mathbf{y}^{(0)} \in S_n$ pre ktorú platí:

$$E_1(\mathbf{x}^{(0)}, \mathbf{y}^{(0)}) \geq E_1(\mathbf{x}, \mathbf{y}^{(0)}) \text{ a } E_2(\mathbf{x}^{(0)}, \mathbf{y}^{(0)}) \geq E_2(\mathbf{x}^{(0)}, \mathbf{y})$$

Rovnovážne stredné hodnoty platieb:

$$v_1 = E_1(\mathbf{x}^{(0)}, \mathbf{y}^{(0)}) \text{ a } v_2 = E_2(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$$

dvojica $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$ môže byť rovnovážnou stratégiou hráčov vtedy a len vtedy, ak $\mathbf{x}^{(0)}$ a $\mathbf{y}^{(0)}$ sú riešeniami úloh lineárneho programovania (Čičková a Zagiba, 2018):

Hráč 1

$$\max \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j^{(0)}$$

$$\sum_{i=1}^m x_i = 1 \quad (1)$$

$$x_i \geq 0, i = 1, 2, \dots, m$$

Hráč 2

$$\max \sum_{i=1}^m \sum_{j=1}^n b_{ij} x_i^{(0)} y_j$$

$$\sum_{j=1}^n y_j = 1 \quad (2)$$

$$y_j \geq 0, j = 1, 2, \dots, n$$

Lagrangeove funkcie pre dané úlohy majú takýto tvar:

$$L(\mathbf{x}, u_1) = - \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j^{(0)} + u_1 \left(\sum_{i=1}^m x_i - 1 \right) \quad (3)$$

$$L(\mathbf{x}, u_2) = - \sum_{i=1}^m \sum_{j=1}^n b_{ji} x_i^{(0)} y_j + u_2 \left(\sum_{j=1}^n y_j - 1 \right) \quad (4)$$

Podmienky optimálnosti Karusha-Kuhna-Tuckera možno potom zapísať takto:

$$\sum_{j=1}^n a_{ij} y_j^{(0)} - u_1 \leq 0, \quad i = 1, 2, \dots, m \quad (5)$$

$$x_i^{(0)} \left(\sum_{j=1}^n a_{ij} y_j^{(0)} - u_1 \right) = 0, \quad i = 1, 2, \dots, m \quad (6)$$

$$\sum_{i=1}^m x_i = 1 \quad (7)$$

$$x_i \geq 0, \quad i = 1, 2, \dots, m \quad (8)$$

$$\sum_{i=1}^m b_{ji} x_i^{(0)} - u_2 \leq 0, \quad j = 1, 2, \dots, n \quad (9)$$

$$y_j^{(0)} \left(\sum_{i=1}^m b_{ji} x_i^{(0)} - u_2 \right) = 0, \quad j = 1, 2, \dots, n \quad (10)$$

$$\sum_{j=1}^n y_j = 1 \quad (11)$$

$$y_j \geq 0, \quad j = 1, 2, \dots, n \quad (12)$$

Po úpravách (uvedené v Čičková a Zagiba, 2017) možno formulovať úlohu kvadratického programovania:

$$\max \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j + \sum_{i=1}^m \sum_{j=1}^n b_{ji} x_i y_j - u_1 - u_2 \quad (13)$$

Za podmienok

$$\sum_{j=1}^n a_{ij} y_j - u_1 \leq 0, \quad i = 1, 2, \dots, m \quad (14)$$

$$\sum_{i=1}^m x_i = 1 \quad (15)$$

$$x_i \geq 0, \quad i = 1, 2, \dots, m \quad (16)$$

$$\sum_{i=1}^m b_{ji} x_i - u_2 \leq 0, \quad j = 1, 2, \dots, n \quad (17)$$

$$\sum_{j=1}^n y_j = 1 \quad (18)$$

$$y_j \geq 0, \quad j = 1, 2, \dots, n \quad (19)$$

Riešením tejto úlohy získavame rovnovážne stratégie $\mathbf{x}^{(0)}$ a $\mathbf{y}^{(0)}$. Hodnoty hry hráčov predstavujú u_1 a u_2 . Optimálnou hodnotou účelovej funkcie je 0, pričom však daný model nezohľadňuje dominované riešenia.

3 PODMIENKY OPTIMÁLNOSTI KARUSHA-KUHNA-TUCKERA PRE VIACKRITERIÁLNE BIMATICOVÉ HRY

Aj napriek tomu, že sa viackriteriálne hry od bimaticových líšia, ich formulácia prebieha obdobne.

Predpokladajme, že hráč 1 zvažuje pri svojom rozhodovaní l navzájom protichodných kritérií. Hodnoty hry pre hráča 1 v k -tom kritériu možno zapísať do matice $\mathbf{A}^{(k)} = a_{ij}^{(k)}$, kde $a_{ij}^{(k)}$ udáva platbu hráča pri stratégiách (i, j) , $i = 1, 2, \dots, m, j = 1, 2, \dots, n, k = 1, 2, \dots, l$. Nech hráč 2 zvažuje pri rozhodovaní r navzájom protichodných kritérií. Výsledky hry pre hráča 2 zapíšeme do matice $\mathbf{B}^{(s)} = b_{ij}^{(s)}$, kde $b_{ij}^{(s)}$ udáva platbu hráča 2 pri stratégiách (i, j) $i = 1, 2, \dots, m, j = 1, 2, \dots, n, s = 1, 2, \dots, r$.

Podľa (Isermann, 1974) dvojica $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$ môže byť rovnovážnou stratégiou hráčov vtedy a len vtedy, ak existujú parametre $\alpha_k, \beta_s > 0, k = 1, 2, \dots, l, s = 1, 2, \dots, r$, také, že $\mathbf{x}^{(0)}$ je riešením úlohy lineárneho programovania:

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^n \sum_{s=1}^r \beta_s b_{ij}^{(s)} y_j x_i^{(0)} &\rightarrow \max \\ \sum_{j=1}^n y_j &= 1 \\ y_j &\geq 0, j = 1, 2, \dots, m \end{aligned} \quad (20)$$

a $\mathbf{y}^{(0)}$ je riešením úlohy lineárneho programovania:

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^l \alpha_k a_{ij}^{(k)} x_i y_j^{(0)} &\rightarrow \max \\ \sum_{i=1}^m x_i &= 1 \\ x_i &\geq 0, i = 1, 2, \dots, m \end{aligned} \quad (21)$$

Úlohy sú úlohami lineárneho programovania, problémom však je (tak ako v predchádzajúcom prípade), že nepoznáme $\mathbf{y}^{(0)}$ a $\mathbf{x}^{(0)}$. Obdobným postupom možno formulovať Lagrangeove funkcie pre dané úlohy a podmienky optimálnosti Kuhna-Tuckera, a zapísať modifikované podmienky:

$$\sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^l \alpha_k x_i^{(0)} a_{ij}^{(k)} y_j^{(0)} = u_1 \quad (22)$$

$$\sum_{i=1}^m \sum_{j=1}^n \sum_{s=1}^r \beta_s x_i^{(0)} b_{ij}^{(s)} y_j^{(0)} = u_2 \quad (23)$$

$$\sum_{j=1}^n \sum_{k=1}^l \alpha_k a_{ij}^{(k)} y_j^{(0)} - u_1 \leq 0, i = 1, 2, \dots, m \quad (24)$$

$$\sum_{i=1}^m \sum_{s=1}^r \beta_s b_{ij}^{(s)} x_i^{(0)} - u_2 \leq 0, j = 1, 2, \dots, n \quad (25)$$

$$\sum_{i=1}^m x_i = 1 \quad (26)$$

$$\sum_{j=1}^n y_j = 1 \quad (27)$$

$$x_i \geq 0, i = 1, 2, \dots, m \quad (28)$$

$$y_j \geq 0, j = 1, 2, \dots, n \quad (29)$$

$$\alpha_k > 0, k = 1, 2, \dots, l \quad (30)$$

$$\beta_s > 0, s = 1, 2, \dots, r \quad (31)$$

na základe ktorých možno formulovať úlohu kvadratického programovania:

$$\sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^l \alpha_k x_i^{(0)} a_{ij}^{(k)} y_j^{(0)} + \sum_{i=1}^m \sum_{j=1}^n \sum_{s=1}^r \beta_s x_i^{(0)} b_{ij}^{(s)} y_j^{(0)} - u_2 - u_1 \quad (32)$$

s cieľovou hodnotou účelovej funkcie 0.

4 ZÁVER

Podmienky optimálnosti Karusha-Kuhna-Tuckera sú vo všeobecnosti používané na identifikáciu lokálnych extrémov v úlohách nelineárneho programovania. Aj keď ich priame analytické vyčíslenie môže byť niekedy problematické, často sú v rôznych formách súčasťou iteračných extremalizačných metód. V príspevku sme prezentovali formuláciu týchto podmienok pre nájdenie rovnovážnych stratégií v bimaticovej hre. Daný prístup sme rozšírili pre prípad viacerých kritérií hráčov.

Príspevok bol spracovaný v rámci riešenia grantovej úlohy VEGA 1/0427/20 Viackriteriálne modely teórie hier v ekonómii a politológii a v rámci riešenia grantovej úlohy VEGA 1/0339/20 Využitie skrytého Markovovho modelu vo finančnom modelovaní.

Použitá literatúra

1. AKHBARI, M., SADI-NEZHAD, S. 2015. Equilibrium Solution of Non-cooperative Bimatrix Game of Z-Numbers. In Bulletin of the Georgian National Academy of Sciences, roč. 9, 2015, č. 1.
2. BLACKWELL, D. 1956. An analog of the minimax theorem for vector payoffs. Pacific Journal of Mathematics, vol.6, 1956, č. 1, s. 1–8.
3. ČIČKOVÁ, Z., ZAGIBA, M. 2017. Podmienky optimálnosti Karusha-Kuhna-Tuckera a bimaticové hry. Aplikácia vybraných modelov teórie hier pri riešení niektorých ekonomických problémov Slovenska: 1. vedecký workshop k riešeniu projektu Vega č. 1/0351/17: zborník z workshopu, Bratislava. EKONÓM, s. 1-7.
4. FENDEK, M., FENDEKOVÁ, E. 2008. *Mikroekonomická analýza*. Bratislava: Iura Edition, spol. s r. o.
5. GOGA, M. *Teória hier*. 2013. Bratislava. Iura Edition. ISBN 978-80-8078-613-7.
6. CHOBOT, M. 1986. *Teória hier a rozhodovania*. Bratislava: Vysoká škola ekonomická.
7. CHOBOT, M., TURNOVEC, F., ULAŠIN, V. 1991. *Teória hier a rozhodovania*. Alfa, 1991.
8. ISERMANN, H. 1974. Proper efficiency and the linear vector maximum problem. Operations Research, roč. 22, 1974, č. 1, s. 189-191.
9. KRUŠ, L., BRONISZ, P. 1994. Onn-person noncooperative multicriteria games described in strategic form. Annals of Operations Research, roč. 51, 1994. č. 2, s. 83-97.
10. KUHN, H.W , TUCKER, A.W. 1951. Nonlinear programming. Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press. p. 481–492.
11. SHAPLEY, L. S., RIGBY, F. D. 1959. Equilibrium points in games with vector payoffs. Naval Research Logistics Quarterly, roč. 6, 1959, č. 1, s. 57-61.
12. VOORNEVELD, M., GRAHN, S., DUFWENBERG, M. 2000. Ideal equilibria in noncooperative multicriteria games. Mathematical Methods of operations research, roč. 52, 2000, č. 1, s. 65–77.
13. ZŮNA, P. 2010. Vojenský rozhodovací proces a pokročilé metody manažerského rozhodování. <https://vojenskerozhledy.cz/kategorie-clanku/teorie-a-doktriny/vojensky-rozhodovaci-proces-a-pokrocile-metody-manazerskeho-rozhodovani>

Kontaktné údaje

doc. Ing. Zuzana Čičková, PhD.

Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky

Dolnozemska 1/b, 852 35 Bratislava

Tel: (421 2) 67 295 830, email: zuzana.cickova@euba.sk

Ing. Patrícia Holzerová, PhD.

Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky

Dolnozemska 1/b, 852 35 Bratislava

Tel: (421 2) 67 295 826, email: patricia.holzerova@euba.sk

Ing. Pavel Gežík, PhD.

Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky

Dolnozemska 1/b, 852 35 Bratislava

Tel: (421 2) 67 295 820

email: pavel.gezik@euba.sk

UKÁŽKA VEKTOROVÉHO MODELU S KOREKČNÝM ČLENOM V SOFTVÉRI EViews

SHOWCASE OF A VECTOR ERROR CORRECTION MODEL IN EViews

Ádám Csápai

Abstrakt

V tomto článku ukážeme možnosti použitia VEC modelov pomocou EViews. Ako príklad analyzujeme vplyv exportu, priamych zahraničných investícií a tvorby hrubého kapitálu na ekonomický rast v Maďarsku. Testujeme stacionaritu časových radov pomocou ADF testu, preskúmame konitegračné vzťahy, testujeme prítomnosť autokorelácie a stabilitu modelu, ako aj slabú exogénnosť parametrov matice prispôsobenia. Na konci článku odhadneme dlhodobú rovnicu a interpretujeme naše zistenia.

Keľúčové slová: VEC, HDP, EViews

Abstract

Abstract in English. In this paper we show the possibilities of using VEC models in EViews. As an example, we analyze the effect of export, foreign direct investment and gross capital formation on economic growth in Hungary. We test the stationarity of the time series using ADF test, examine cointegrating relationships, test the existence of autocorrelation and model stability, as well as weak exogeneity of the parameters of the adjustment matrix. At the end of the paper, we estimate the long run equation and interpret our results.

Keywords: VEC, GDP, EViews

1 ÚVOD

V tomto článku analyzujeme vplyv exportu, priamych zahraničných investícií a tvorby hrubého kapitálu na ekonomický rast v podmienkach Maďarska. Tento vplyv analyzujeme pomocou vektorového modelu s korekčným členom (ďalej VEC). Popíšeme spôsob odhadu takéhoto modelu v softvéri EViews a ukážeme výstupy z neho. Článok má 2 kapitoly. V prvej kapitole uvedieme cieľ našej práce, metodiku skúmania a popisujeme použité dáta a spôsob ich úpravy. V tretej kapitole najprv analyzujeme použité časové rady, testujeme stacionaritu a vyberieme si optimálny počet oneskorení pre VEC model a testovanie kointegrácie. Ďalej, testujeme autokoreláciu náhodných zložiek a stabilitu modelu. Následne testujeme slabú exogénnosť jednotlivých koeficientov matice prispôsobenia. Na konci tejto kapitoly interpretujeme naše zistenia a odhadneme dlhodobú rovnicu. Záver článku obsahuje zhrnutie našich výsledkov.

2 METODOLÓGIA PRÁCE A METÓDY SKÚMANIA

Cieľom tohto článku je odhadnúť vplyv exportu, priamych zahraničných investícií a tvorby hrubého kapitálu na rast HDP v podmienkach Maďarska.

Na analýzu vplyvu týchto premenných na HDP najprv v programe EViews odhadneme neohraničený VEC model so štyrmi endogénnymi premennými a s jedným oneskorením. Na konci práce testujeme slabú exogénnosť parametrov a odhadneme dlhodobú rovnicu s korekčným členom. Náš model sme zostavili podľa práce Abdalla a Abdelbaki (2014).

Export (export), priame zahraničné investície (fdi) a tvorbu hrubého kapitálu (gcf) uvádzame vo forme ich percentuálneho podielu na HDP. Premenná HDP (gdp) vyjadruje rast hrubého domáceho produktu. Premenné nie sú zlogaritmované. V modeli okrem endogénnych premenných je zaradená aj exogénna umelá premenná (dummy). Pre odhad modelu použijeme štvrťročné dáta pre Maďarsko v období 2000:Q01 – 2018:Q4.

3 VÝSLEDKY

Našu analýzu začneme testovaním stacionarity. Stacionaritu časových radov testujeme v EViews pomocou rozšíreného Dickeyho-Fullerov testu (ADF-test), pomocou ktorého testujeme všetky endogénne premenné v modeli. Výsledok testu časového radu „export“ je zobrazený na obrázku 1.

Obr. 1: ADF test bez diferencie - export

Null Hypothesis: EXPORT has a unit root Exogenous: None Lag Length: 5 (Automatic - based on SIC, maxlag=11)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.446490	0.8081
Test critical values:		
1% level	-2.598416	
5% level	-1.945525	
10% level	-1.613760	

Zdroj: vlastné spracovanie

Ostatné premenné sme testovali tým istým spôsobom a zistili sme, že časové rady sú integrované 1. rádu. Ako príklad opäť uvedieme výsledky ADF testu rady export, ktoré sme diferencovali. Výstup zo softvéru EViews je na Obrázku 2.

Obr. 2: ADF test s 1 diferenciou - export

Null Hypothesis: D(EXPORT) has a unit root Exogenous: None Lag Length: 4 (Automatic - based on SIC, maxlag=11)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.483111	0.0136
Test critical values:		
1% level	-2.598416	
5% level	-1.945525	
10% level	-1.613760	

Zdroj: vlastné spracovanie

Ďalším predpokladom použitia VEC modelu je to, že premenné musia mať aspoň jeden spoločný kointegračný vzťah. Existenciu tohto vzťahu testujeme pomocou Johansenovho testu kointegrácie. Podľa informačných kritérií si zvolíme 2 oneskorenia. Výsledky sú zobrazené na Obrázku 3.

Podľa Pantulovho princípu si podľa lambda-trace štatistiky vyberieme najnižší variant kointegrácií a schém, pričom neberieme do úvahy prvú a piatu schému, keďže neobsahujú ani trend, ani úrovňovú konštantu v kointegračnom vzťahu. Po zohľadnení týchto skutočností si vyberieme druhú schému s počtom kointegračných vzťahov 1.

Obr. 3: Johansenovho test kointegrácie

Date: 06/08/20 Time: 07:06					
Sample: 2000Q1 2018Q4					
Included observations: 73					
Series: GDP EXPORT FDI GCF					
Lags interval: 1 to 2					
Selected (0.05 level*) Number of Cointegrating Relations by Model					
Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
Trace	1	1	2	1	2
Max-Eig	0	1	1	1	1
*Critical values based on MacKinnon-Haug-Michelis (1999)					

Zdroj: vlastné spracovanie

Obr. 4: Kointegračná rovnica

1 Cointegrating Equation(s):	Log likelihood	-555.9771
Normalized cointegrating coefficients (standard error in parentheses)		
GDP	EXPORT	FDI
1.000000	0.931101 (0.13697)	-0.273313 (0.05925)
	GCF	C
	0.894158 (0.61520)	-97.39523 (23.1187)
Adjustment coefficients (standard error in parentheses)		
D(GDP)	-0.226330 (0.06954)	
D(EXPORT)	-0.037629 (0.01747)	
D(FDI)	0.130836 (0.16496)	
D(GCF)	-0.002465 (0.01026)	

Zdroj: vlastné spracovanie

Johansenovho testom kointegrácie otestujeme aj druhú schému a dostaneme kointegračnú rovnicu, ktorú vidíme nižšie na Obrázku 4. Z výsledkov je vidno, že podľa podielu parametrov a ich rozptylov parameter “gcf” nie je štatisticky významný ani na 10%-nej hladine významnosti.

Obr. 5: LM test autokorelácie

VEC Residual Serial Correlation LM Tests						
Date: 06/08/20 Time: 11:24						
Sample: 2000Q1 2018Q4						
Included observations: 74						
Null hypothesis: No serial correlation at lag h						
Lag	LRE* stat	df	Prob.	Rao F-stat	df	Prob.
1	16.89550	16	0.3924	1.063505	(16, 187.0)	0.3930
2	9.797004	16	0.8770	0.605376	(16, 187.0)	0.8772

Zdroj: vlastné spracovanie

Obr. 6: Stabilita modelu

Roots of Characteristic Polynomial	
Endogenous variables: GDP EXPORT FDI GCF	
Exogenous variables:	
Lag specification: 1 1	
Date: 06/08/20 Time: 07:09	
Root	Modulus
1.000000	1.000000
1.000000	1.000000
1.000000	1.000000
0.851501	0.851501
0.641047 - 0.286750i	0.702258
0.641047 + 0.286750i	0.702258
0.609985	0.609985
0.440692	0.440692
VEC specification imposes 3 unit root(s).	

Zdroj: vlastné spracovanie

Našu analýzu pokračujeme odhadom VEC modelu, ktorý odhadneme s 1 oneskorením. Tak isto ako v prípade VAR modelov aj tu je dôležité zistiť, či je v modeli prítomná autokorelácia. Výsledky LM-testu prezentujeme na Obrázku 5.

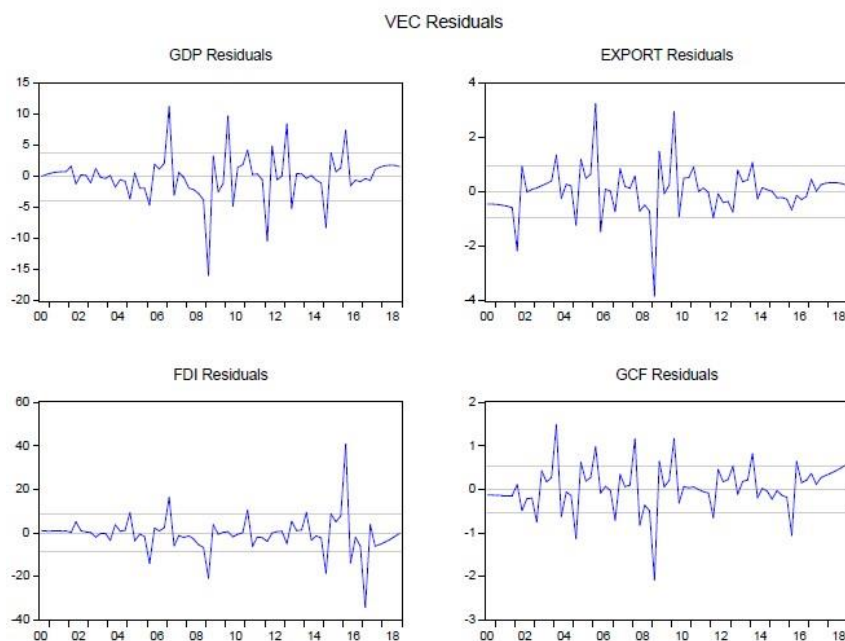
Vidíme, že v modeli nie je prítomná autokorelácia. Prejdeme teda na testovanie stability modelu. Je dôležité poznamenať, že v našom VEC modeli potrebujeme mať 3 jednotkové korene, nakoľko máme 4 endogénne premenné a 1 kointegračný vzťah, pričom ostatné

hodnoty nemôžu ležať mimo jednotkového koreňa. Táto podmienka je splnená, ako to môžeme vidieť nižšie na Obrázku 6.

Ešte predtým, než testujeme slabú exogénnosť parametrov, skúmame reziduály VEC modelu, ktoré sú graficky zobrazené na obrázku 8.

Podľa týchto grafov môžeme identifikovať významný negatívny výkyv v 2009:Q1 v reziduáloch HDP, exportu a tvorby hrubého kapitálu. Predpokladáme, že tento výkyv bol spôsobený finančnou krízou. Ďalšie náhle pozitívne a negatívne výkyvy identifikujeme v priamych zahraničných investíciách v obdobiach 2016:Q1 a 2017:Q1. Aby sme zmiernili tieto vplyvy, zavedieme exogénnu umelú premennú (dummy) do modelu, ktorá nadobudne hodnotu jedna v spomenutých kvartáloch a nulu v každom ostatnom období. Následne model testujeme doteraz použitými testmi. Výsledkom zavedenia umelej premennej je výrazne zlepšenie hodnôt koeficientu determinácie modelu a po testovaní dospejeme k záveru, že model je lepší po zaradení umelej premennej. Už môžeme testovať koeficienty matice prispôsobenia.

Obr. 7: Reziduály VEC modelu



Zdroj: vlastné spracovanie

Po odhade modelu testujeme slabú exogénnosť koeficientov matice alfa, teda matice koeficientov prispôsobenia. Testujeme koeficienty zvlášť aj spoločne pomocou LR testu. Keď parametre testujeme zvlášť tak iba prvá rovnica má štatisticky významné koeficienty.

Avšak keď jednotlivé koeficienty testujeme spolu, tak koeficienty sú štatisticky významné aj na 0.01 %-nej hladine významnosti. Dostaneme teda výsledný tvar VEC po testovaní slabej exogénnosti, čo prezentujeme na Obrázku 8.

Predtým, než výsledky interpretujeme, treba si pripomenúť, že vo VEC modeli dlhodobú rovnicu treba interpretovať opačnými znamienkami než majú parametre v softvéri EViews. Diferencované premenné spolu s umelou premennou nám udávajú krátkodobý vzťah medzi rastom HDP, exportom, priamymi zahraničnými investíciami a tvorbou hrubého kapitálu. Vidíme, že koeficient korigujúci chybu v modeli je buď nula alebo má záporné znamienko, čo svedčí o stabilite modelu. Odchýlka exportu od dlhobohovej rovnováhy v predchádzajúcom roku sa koriguje rýchlosťou 2,66%. Z výsledkov dlhobohovej rovnice vyplýva, že z dlhodobého

hľadiska rast podielu exportu a tvorby hrubého kapitálu má negatívny, pričom rast podielu priamych zahraničných investícií má pozitívny vplyv na hospodársky rast v podmienkach Maďarska v období 2000:Q1 až 2018:Q4. Treba si však poznamenať, že parametre nie sú štatisticky významné na 5 %-nej hladine významnosti.

Obr. 8: VEC model

Cointegration Restrictions:				
B(1,1)=1, A(1,1)=0, A(3,1)=0, A(4,1)=0				
Convergence achieved after 18 iterations.				
Restrictions identify all cointegrating vectors				
LR test for binding restrictions (rank = 1):				
Chi-square(3)	18.81568			
Probability	0.000771			
Cointegrating Eq:		CointEq1		
GDP(-1)	1.000000			
EXPORT(-1)	1.582953 (0.34904) [4.53513]			
FDI(-1)	-0.276924 (0.14446) [-1.91695]			
GCF(-1)	2.839854 (1.51617) [1.87305]			
C	-193.6273 (57.2112) [-3.38443]			
Error Correction:	D(GDP)	D(EXPORT)	D(FDI)	D(GCF)
CointEq1	0.000000 (0.00000) [NA]	-0.026646 (0.00693) [-3.84308]	0.000000 (0.00000) [NA]	0.000000 (0.00000) [NA]
D(GDP(-1))	0.620501 (0.10931) [5.67657]	-0.023476 (0.02498) [-0.93963]	-0.044021 (0.24534) [-0.17943]	0.007875 (0.01445) [0.54492]
D(EXPORT(-1))	-0.239334 (0.41954) [-0.57046]	0.652781 (0.09590) [6.80724]	1.263705 (0.94163) [1.34204]	-0.083925 (0.05547) [-1.51308]
D(FDI(-1))	-0.046357 (0.05399) [-0.85860]	0.005447 (0.01234) [0.44139]	0.459394 (0.12118) [3.79102]	-0.004155 (0.00714) [-0.58207]
D(GCF(-1))	0.386404 (1.12862) [0.34237]	0.092982 (0.25797) [0.36044]	-1.966843 (2.53311) [-0.77645]	0.646856 (0.14921) [4.33516]
DUMMY	-2.368240 (2.27075) [-1.04293]	-1.246872 (0.51903) [-2.40233]	-5.428963 (5.09653) [-1.06523]	-0.952504 (0.30021) [-3.17282]
R-squared	0.373457	0.563027	0.271477	0.475848

Zdroj: vlastné spracovanie

Pomocou tohto modelu môžeme dlhodobý vzťah odhadnúť ako jednorovnicový model. Príkazom v softvéri EViews si získame odhad náhodných zložiek z VEC modelu, pomocou ktorých následne odhadneme jednorovnicový model s oneskorenými náhodnými zložkami z dlhodobého vzťahu. Výsledný model vidíme na Obrázku 9.

Obr. 9: Dlhodobý model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	193.6273	4.86E-12	3.99E+13	0.0000
EXPORT	-1.582953	3.09E-14	-5.12E+13	0.0000
FDI	0.276924	8.93E-15	3.10E+13	0.0000
GCF	-2.839854	1.17E-13	-2.42E+13	0.0000
ECT	1.000000	1.86E-14	5.37E+13	0.0000
R-squared	1.000000	Mean dependent var	7.017996	
Adjusted R-squared	1.000000	S.D. dependent var	12.25817	
S.E. of regression	1.63E-12	Akaike info criterion	-51.38846	
Sum squared resid	1.88E-22	Schwarz criterion	-51.23512	
Log likelihood	1957.761	Hannan-Quinn criter.	-51.32718	
F-statistic	1.07E+27	Durbin-Watson stat	0.125697	
Prob(F-statistic)	0.000000			

Zdroj: vlastné spracovanie

4 ZÁVER

Naším cieľom je zistiť vplyv exportu, priamych zahraničných investícií a tvorby hrubého kapitálu na hospodársky rast v Maďarsku pomocou vektorového modelu s korekčným členom v období rokov 2000:Q1 až 2018:Q4. Dospejeme k záveru, že z dlhodobého hľadiska rast podielu exportu a tvorby hrubého kapitálu má negatívny, pričom rast podielu priamych zahraničných investícií má pozitívny vplyv na hospodársky rast. Práca je tiež ukázkou možnosti využitia softvéru EViews pre tvorbu a testovanie VEC modelov.

Použitá literatúra

1. ABDALLA, M. A. – ABDELBAKI, H. H. 2014. Determinants of Economic Growth in GCC Economies. In: Asian Journal of Research in Business Economics and Management, roč. 4, 2014, č. 6, s. 46-62.

ŘEŠENÍ ROZSÁHLÝCH DOPRAVNÍCH PROBLÉMŮ

SOLVING LARGE TRANSPORTATION PROBLEMS

Ing. Marek Daniel

Abstrakt

Cílem tohoto zkoumání je zjistit, jak rozsáhlé dopravní problémy jde řešit v řádech minut, popřípadě hodin, a jaký software na toto řešení použít. Jsou zde popsány a otestovány programy MPL, Matlab a Python. Testování probíhalo do rozměrů 1500×1500 pro úlohy se stejně dodavateli a odběrateli, pro úlohy s méně dodavateli do rozměrů 600×3000. Na všechny testy byl použit stejný počítač a pro každou velikost úlohy byl test proveden na pěti různých setech dat. Jako nerychlejší vychází Matlab (úlohu 1000×1000 řeší průměrně 141,5s), ale kvůli jeho velkým nárokům na paměť počítače se jako nejvhodnější program ukazuje MPL.

***Klíčová slova:** rozsáhlý dopravní problém, MPL, Matlab, Python*

Abstract

The aim of this research is to find how large transportation problems can be solved in the manner of minutes or hours and which software is the most suitable for doing so. The programmes and introduced and tested – MPL, Matlab and Python. The testing was done up to the size 1500×1500 for problems with the same number of suppliers and customers, for problems with less suppliers up to 600×3000. All testing was done on the same computer and for each size of the problem 5 different data sets were tested. Matlab appears to be the fastest option (solving 1000×1000 on average 141,5s), however MPL seems to be more suitable due to the large computer memory requirement by Matlab.

***Keywords:** large transportation problem, MPL, Matlab, Python*

1 ÚVOD

Cílem tohoto výzkumu je zjistit nejefektivnější způsob řešení rozsáhlých dopravních problémů. Počátky zkoumání této úlohy lineárního programování sahají do 30. a 40. let 20. století. Od té doby až do sedmdesátých let byl dopravní problém a metody na jeho řešení intenzivně zkoumány, poté však nastal útlum kvůli rozvoji výpočetní techniky, protože bylo jednodušší ho řešit obecně jako jiné úlohy lineárního programování. Jaké jsou tedy dnes možnosti řešení tohoto problému?

Pro zpřesnění výsledků je každá velikost dopravního problému testována na pěti různých setech dat. Rychlost řešení dopravního problému bude porovnávána ve třech programech. Jedná se o MPL, Matlab a Python. První dva zmíněné jsou programy s profesionálními řešiteli úloh lineárního programování, přesto jsou oba velmi rozdílné. O všech třech budou uvedeny základní informace, jejich výhody, nevýhody a popsán zdrojový kód, který je k měření rychlosti použit. Porovnány budou ve třech kategoriích – malé čtvercové úlohy do rozměrů 100×100 , velké čtvercové úlohy a úlohy, kde je méně dodavatelů než odběratelů.

2 TEORIE DOPRAVNÍHO PROBLÉMU

Cílem klasického dopravního problému je naplánování přepravy nějakého zboží nebo materiálu od dodavatelů (zdroje) k odběratelům (požadavky) při minimalizaci celkových nákladů vynaložených na tuto přepravu. Je zde použita stejná definice a označení, jako uvádí Lagová a Jablonský [1]. V úloze je definováno m dodavatelů $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_m$ s kapacitami jejich zdrojů a_1, a_2, \dots, a_m a n odběratelů $\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_n$ s požadavky b_1, b_2, \dots, b_n . Přeprava mezi každou dvojicí dodavatele a odběratele je kladně oceněna a tato cena představuje vypočítané náklady na přepravu jednoho kusu zboží mezi touto dvojicí. V modelu bude označena jako $c_{ij}, i = 1, 2, \dots, m, j = 1, 2, \dots, n$. Množství zboží přepravovaného mezi dodavatelem a odběratelem je definován jako $x_{ij}, i = 1, 2, \dots, m, j = 1, 2, \dots, n$. Formulace matematického modelu obecného dopravního problému vypadá následovně.

Minimalizovat:

$$z = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij},$$

za podmíněk:

$$\sum_{i=1}^m x_{ij} = a_i, \quad i = 1, 2, \dots, m,$$

$$\sum_{j=1}^n x_{ij} = b_j, \quad j = 1, 2, \dots, n,$$

$$x_{ij} \geq 0, \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n.$$

3 DOBA HLEDÁNÍ OPTIMÁLNÍHO ŘEŠENÍ

3.1 Historie rychlostí hledání optimálního řešení

V roce 1951 představil G. B Dantzig [2] upravenou verzi simplexové metody na řešení dopravního problému, které se také říká PSTM (primal simplex transportation method) nebo MODI (modified distribution method) – modifikovaná distribuční metoda. O tři roky později přišli A. Charnes a W. W. Copper s takzvanou stepping-stone method [3]. Oproti MODI byla tato metoda lehčí na uchopení a efektivnější na počítání úloh malých rozsahů. Pro větší úlohy byla ovšem stále lepší MODI.

S rozvojem výpočetní techniky, programovacích jazyků a programování obecně se v šedesátých a sedmdesátých letech vědci snažili přijít s co nejefektivnější implementací dříve vyvinutých algoritmů. Na konci šedesátých let byla obecně považována za nejlepší implementace od Forda a Fulkersona [4], která řešila dopravní problém o rozměrech 30×220 v průměru za 4 minuty a 58 sekund na počítači IBM 704. Tato implementace dokázala vyřešit problémy o velikosti $m+n \leq 8000$.

Přelom v tomto zkoumání přinesli Glover a kolektiv [5], kteří v roce 1974 přišli s implementací MODI, která byla devětkrát rychlejší než doposud známé kódy – OPHELIE/LP a Out-of-kilter. Během porovnávání byli velmi důslední, neboť nejenže porovnávali rychlosti hledání optima na různých velikostech úloh, ale i na různých počítačích. Hlavní testy byly prováděny na prvogeneračním superpočítači CDC 6600, ale porovnání bylo provedeno i na běžnějších počítačích IBM 360/65 a UNIVAC 1108, na kterých byly prezentovány dřívější výsledky pro OPHELIE/LP a Out-of-kilter.

3.2 Parametry počítače pro výpočty

Aby byly výsledky porovnatelné s výsledky zjištěnými na jiných počítačích, uvedu zde několik základních parametrů o počítači, na kterém testy probíhaly.

Operační systém: Microsoft Windows 10 Enterprise, x64

Processor: Intel Core i5-3570 CPU @ 3.40GHz, 4 fyzická jádra, 4 logická jádra

RAM: 8 GB

Virtuální paměť: 26,8 GB

3.3 Malé čtvercové úlohy

Jako první sérii měření jsem zvolil malé čtvercové úlohy, tedy takové, kde $m=n$ a $m,n < 100$. V Tabulce 1 jsou uvedeny celkové časy potřebné pro řešení úlohy. Sloupec MVB představuje celkový čas potřebný pro vyřešení úlohy metodou vnitřních bodů v Matlabu. Sloupec DSM referuje na duálně simplexovou metodu ve stejném programu. Všechny uvedené časy jsou průměr z pěti měření. Nikdy se však měření neodchýlila více než o 10 % od uvedeného průměru, a to platí pro všechny testované rozměry ve všech použitých programech.

rozměry	Matlab		MPL	Python
	MVB	DSM		
5×5	0,010	0,024	0,047	0,005
10×10	0,048	0,056	0,100	0,065
20×20	0,019	0,058	0,343	0,476
40×40	0,065	0,094	1,253	4,884
60×60	0,040	0,070	2,540	19,808
80×80	0,057	0,086	4,678	46,147

Tabulka 1: Čas řešení v sekundách pro malé čtvercové úlohy

Jak je z Tabulky vidět, do úloh rozměrů 20×20 řeší všechny programy úlohu pod 1 vteřinu. Poté můžeme vidět, že čas potřebný pro řešení v Pythonu prudce stoupá. To, že bude Python horší než Matlab a MPL, bylo očekáváno, protože se zde data nedrží v žádném speciálním formátu a ani nevyužívá žádné speciální balíčky pro řešení úloh lineárního programování. MPL je trochu pomalejší než Matlab, ale rozdíl není tak signifikantní jako rozdíl u Pythonu.

3.4 Velké čtvercové úlohy

Obě metody pro Matlab vykazovaly pro malé úlohy téměř shodné časy a MPL o trochu vyšší, ale ne o tolik. V Tabulce 2 jsou prezentovány výsledky měření pro mnohem větší úlohy. Uvedené hodnoty jsou opět průměr z pěti měření kromě hodnot označených hvězdičkou. Z důvodu jejich extrémní délky byly otestovány pouze na jednom setu dat.

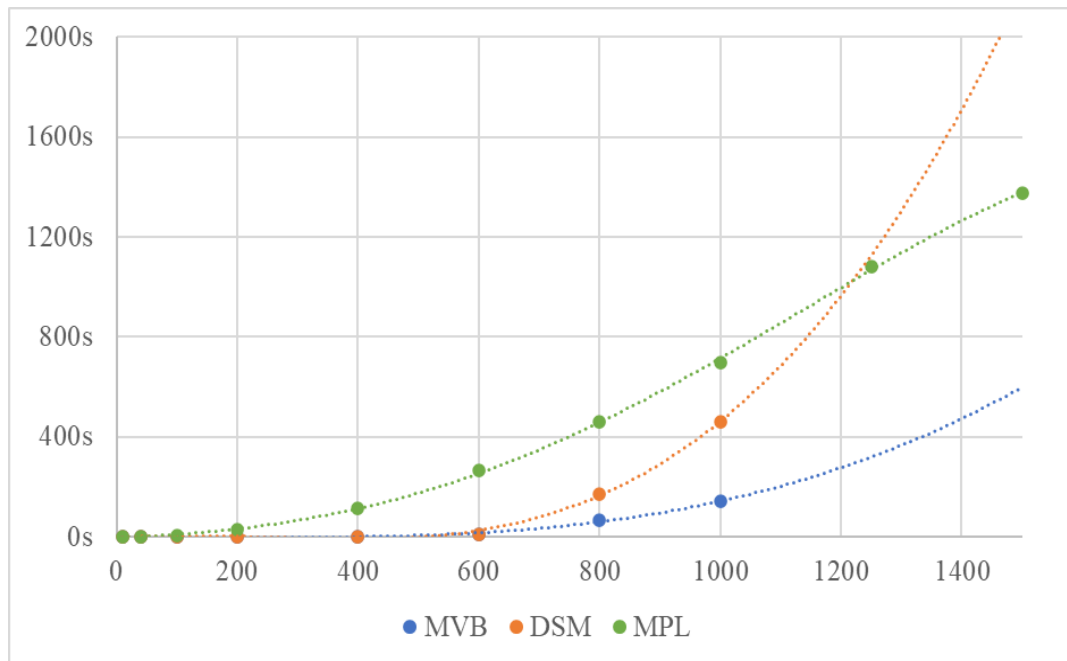
rozměry	Matlab		MPL	Python
	MVB	DSM		
100×100	0,127	0,13	7	20,9
200×200	0,463	0,427	27,8	236
400×400	2,9	3,1	112,5	2229
600×600	10,4	12,5	265	7890
800×800	66,1	173	460	20367*
1000×1000	141,5	461,8	698	40937*
1250×1250	chyba	chyba	1080	82354*
1500×1500	chyba	chyba	1375	147521*

Tabulka 2: Čas řešení v sekundách pro velké čtvercové úlohy

Jako nejlepší způsob řešení se zdá být metoda vnitřních bodů implementovaná v Matlabu, která je více než čtyřikrát rychlejší než MPL pro úlohy 1000×1000. Pro větší úlohy však řešení v Matlabu skončí chybou, jelikož nestačí velikost RAM a virtuální paměti k vytvoření matice strukturních koeficientů. Tento problém se dá řešit koupením lepší paměti do počítače, ale velikost matice strukturních koeficientů neroste úměrně velikosti úlohy. Například pro úlohu 100×100 má matice strukturních koeficientů dva miliony prvků. Pro „dvojnásobnou“ úlohu 200×200 je to šestnáct milionů prvků, tedy pokud se rozměry zvětší dvakrát, počet prvků matice se zvětší osmkrát.

Pokud odhlédneme od problému s pamětí, jako druhý nejlepší způsob vypadá duálně simplexová metoda v Matlabu, ale při bližším zkoumání můžeme vidět, že nárůst času pro řešení touto metodou roste rychleji než pro MPL. Toto pozorování je lépe vidět v Grafu 1, ve

kterém nejsou zobrazeny hodnoty pro Python, jelikož by pak z grafu nebylo možné vyčíst nic o Matlabu a MPL.



Graf 1: Závislost délky řešení na rozměrech čtvercové úlohy

grafu je na vodorovné ose rozměr úlohy, na svislé ose čas v sekundách. Pro jednotlivé metody řešení jsou vyneseny body z Tabulky 1 a Tabulky 2 a jsou propojeny přerušovanou čarou, která představuje odhad chování mimo zkoumané body. Pro úlohy 600×600 a větší to vypadá, že časová náročnost řešení MPL roste téměř lineárně k rozměrům úlohy. Oproti tomu doba řešení duálně simplexovou metodou v Matlabu roste na tomto úseku mnohem strměji. Zřejmě zde dochází k tomu, že Matlabu už k výpočtu nestačí RAM paměť a používá virtuální paměť, a proto dochází ke zpomalení výpočtu. Dá se tedy očekávat, že pro úlohy kolem rozměrů 1300×1300 a více by bylo MPL lepší než duálně simplexová metoda z Matlabu.

3.5 Úlohy s méně dodavateli

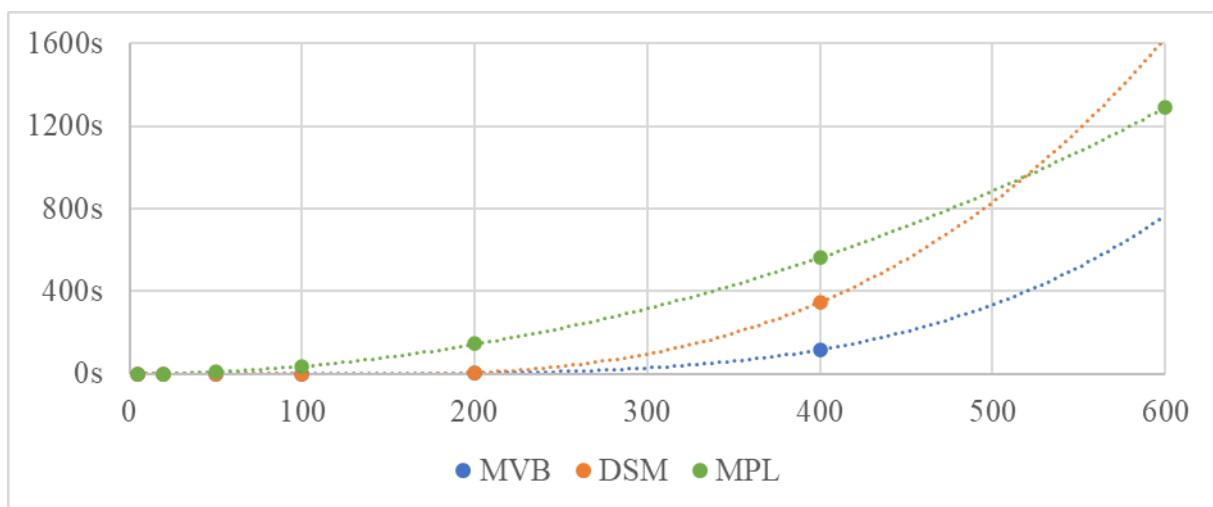
Jako poslední set dat k měření jsem vybral „obdélníkové“ úlohy, kde je méně dodavatelů než odběratelů. To je pro dopravní problém v praxi typičtější podoba, například automobilové společnosti zřejmě budou mít méně závodů na výrobu než prodejen automobilů nebo obchodní řetězce budou mít omezené počty skladů, odkud bude pokračovat distribuce do jednotlivých supermarketů. Jako poměr dodavatelů k odběratelům jsem zvolil 1:5, tedy $n = 5m$. Výsledky jsou ukázány v Tabulce 3.

rozměry	Matlab		MPL	Python
	MVB	DSM		
5×25	0,054	0,085	0,123	0,013
10×50	0,034	0,07	0,390	0,170
20×100	0,340	0,071	1,4	1,8
50×250	0,121	0,145	9,1	32,9
100×500	0,531	0,741	35,1	322
200×1000	3,8	6,4	145,1	2985
400×2000	118,2	347,7	562,4	28120*
600×3000	chyba	chyba	1288,7	98765*

Tabulka 3: Čas řešení v sekundách pro obdélníkové úlohy

Stejně jako pro čtvercové úlohy i zde je nejrychlejší metoda vnitřních bodů z Matlabu. Avšak pro některé úlohy již nenajde řešení opět kvůli matici strukturních koeficientů. To by nemělo být překvapivé, neboť tato matice pro úlohu 1250×1250 má zhruba 3,9 miliardy prvků a paměť na to nestačila, pro úlohu 600×3000 to je 6,48 miliardy prvků, a tedy paměť nemůže stačit ani zde.

Data z Tabulky 3 budou opět přehledněji vidět v Grafu 2, kde je na svislé ose čas v sekundách a na vodorovné ose počet dodavatelů (rozměr m). Body v grafu znázorňují měření z Tabulky a přerušované čáry jsou odhady mimo naše měření. Odhady pro MVB a DSM pro úlohy větších rozměrů než 400×2000 mohou být dost nepřesné kvůli malému množství dat.



Graf 2: Závislost délky řešení na rozměrech obdélníkové úlohy

3.6 Zjištění

Jako nejrychlejší metoda řešení se ukazuje použití funkce *linprog* v Matlabu s řešením metodou vnitřních bodů. Má však zásadní nedostatek – ukládání matice strukturních koeficientů, což vede k omezení rozměrů řešitelných úloh. Největší matice strukturních koeficientů, se kterou Matlab na testovacím počítači dokáže pracovat, má přibližně 3,57 miliardy prvků, což odpovídá čtvercové úloze 1214×1214 nebo obdélníkové 492×2460.

Co se týče duálně simplexové metody v Matlabu, je pro všechna měření pomalejší než metoda vnitřních bodů a má stejnou nevýhodu, jelikož je také omezena maximální velikostí matice strukturních koeficientů. Není proto důvod tuto metodu při libovolných rozměrech úlohy preferovat.

Zajímavým zjištěním o Matlabu je fakt, že čas na načtení dat a tvorbu matice strukturních koeficientů (v Tabulce 4 jako *čas na přípravu*) je několikanásobně kratší než čas na vlastní řešení funkcí *linprog* a to pro obě použité metody.

rozměry	čas na přípravu	čas řešení MVB	čas řešení DSM
100×100	0,03	0,1	0,1
200×200	0,07	0,4	0,4
400×400	0,41	2,5	2,7
600×600	1,28	9,2	11,3
800×800	6,33	59,7	166,5
1000×1000	18,44	123,1	443,3

Tabulka 4: Čas na přípravu a řešení v Matlabu

Pro úlohy, který již Matlab nezvládne, je nejlepší možností MPL. U čtvercových i obdélníkových úloh to vypadá, že jeho výpočetní čas roste lineárně s rozměry úlohy, takže můžeme dopředu docela dobře odhadnout čas, jaký bude na výpočet potřeba. Oproti Matlabu drtivou většinu času, který MPL potřebuje k získání řešení, trvá načítání dat z Excelu. Samotné řešení pak trvá jen zlomek času, jak lze vidět v Tabulce 5, kde jsou jednotlivé časy uvedeny v sekundách.

rozměry	čas na přípravu	čas na řešení
100×100	6,97	0,03
200×200	27,6	0,14
400×400	111,7	0,87
600×600	262,7	2,40
800×800	454,3	5,43
1000×1000	689,3	8,70
1250×1250	1065,7	14,87
1500×1500	1352,7	22,55

Tabulka 5: Čas na přípravu a řešení v MPL

Je možné, že načítání dat ze sofistikovanějších databázových programů, jako například SQL, by mohlo být rychlejší než načítání z Excelu. Další výhodou oproti Matlabu je, že kód potřebný k řešení je výrazně kratší a intuitivnější. Není také potřeba psát si nové funkce pro načtení dat a pro výrobu matice strukturních koeficientů.

Řešení pomocí zde použitého kódu pro Python je pro větší úlohy řádově časově náročnější, ale na rozdíl od Matlabu nepotřebuje tolik paměti a dokáže spolehlivě řešit i úlohy, na které již Matlab nestačí. Navíc je oproti ostatním dvěma programům zdarma.

Použitá literatura

1. LAGOVÁ, M., JABLONSKÝ, J. *Lineární modely*. Praha: Oeconomica, 2009. ISBN 978-80-245-1511-3.
2. DANTZIG, G.B. Application of simplex method to a transportation problem. In: *Activity analysis of production and allocation*, New York: Wiley, 1951.
3. CHARNES, A., COOPER, W.W. The stepping-stone method for explaining linear programming calculations in transportation problems. *Management Science*, 1954, roč. 1, č. 1-4, s. 46-69.
4. FORD, L.R., FULKERSON, D.R. Solving the transportation problem. *Management Science*, 1956, roč. 3, č. 1, s. 24-32.
5. GAVISH, B., SCHWEITZER, P., SHLIFER, E. The zero pivot phenomenon in transportation and assignment problems and its computational implications. *Mathematical Programming*, 1977, roč. 12, č.1, s. 226-240.

Poděkování

Tento příspěvek byl vytvořen za podpory projektu IGA F4/42/2021 Interní grantové agentury, Fakulty informatiky a statistiky, Vysoké školy ekonomické v Praze.

Kontaktní údaje

Ing. Marek Daniel

Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky, Katedra ekonometrie
nám. Winstona Churchilla 1938/4, 130 67 Praha

Tel: (420) 721826879

email: danm02@vse.cz

MOTIVAČNÉ ASPEKTY K ŠTÚDIU NA VYSOKEJ ŠKOLE: KOMPARÁCIA EKONOMICKÝ VS TECHNICKÝ ŠTUDIJNÝ SMER

MOTIVATION ASPECTS TO STUDY AT THE UNIVERSITY: COMPARISON ECONOMIC VS. TECHNICAL STUDY PROGRAM

Andrea Furková¹

Abstrakt

Príspevok skúma charakteristiky motivácie k štúdiu na vysokej škole u študentov dvoch vysokých škôl a to Fakulty informatiky a informačných technológií Slovenskej technickej univerzity v Bratislave a Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave. Hlavným cieľom bolo overiť hlavnú hypotézu rozdielných motivačných aspektov študentov s ekonomickým a technickým zameraním štúdia, resp. zistiť, či je možné nájsť štatisticky významné rozdiely medzi týmito dvomi skupinami študentov vo faktoroch vonkajšej ako aj vnútornej motivácie.

Kľúčové slová: vnútorná motivácia, vonkajšia motivácia, motivácia k štúdiu

Abstract

The paper examines the characteristics of motivation to study at university among students of two universities, namely the Faculty of Informatics and Information Technologies of the Slovak University of Technology in Bratislava and the Faculty of Economic Informatics of the University of Economics in Bratislava. The main goal was to verify the main hypothesis of different motivational aspects of students with an economic and technical focus of study, that is, to find out whether it is possible to find statistically significant differences between these two groups of students in the factors of external as well as internal motivation.

Keywords: internal motivation, external motivation, motivation to study

1. ÚVOD

Konanie každého človeka je motivované mnohými faktormi, ktoré dodávajú energiu ľudskému konaniu, zameriavajú ho určitým smerom, radia jeho priebeh, ovplyvňujú prežívanie človeka a jeho vzťahy k ostatným ľuďom. Dve základné kategórie tohto v súčasnosti veľmi frekventovaného psychologického pojmu sú vnútorná a vonkajšia motivácia (Družka, 205; László, 2004; Žiaková a Balogová, 2012). Pojem vnútorná motivácia môžeme chápať ako stav kedy jednotlivec niečo robí alebo sa učí pre vlastnú potrebu, z vlastného záujmu, pre vlastné potešenie, atď. Zdrojom vnútornej motivácie sú hlavne poznávacie potreby, osobné hodnoty, túžby alebo stanovené ciele. Medzi motívy vonkajšej motivácie zaradujeme napríklad finančné ohodnotenie, iné materiálne výhody, pozitívnu sociálnu klímu v práci alebo škole, atď.

Cieľom tohto príspevku bolo preskúmať charakteristiky motivácie k štúdiu na vysokej škole u študentov dvoch vysokých škôl a to Fakulty informatiky a informačných technológií (FIIT) Slovenskej technickej univerzity (STU) v Bratislave a Fakulty hospodárskej

¹ Príspevok bol spracovaný v rámci riešenia grantovej úlohy VEGA 1/0193/20 a VEGA 1/0211/21.

informatiky (FHI) Ekonomickej univerzity v Bratislave. Hlavným cieľom bolo overiť hlavnú hypotézu rozdielnych motivačných aspektov študentov ekonomických a technických vied, resp. zistiť, či je možné nájsť štatisticky významné rozdiely medzi týmito dvoma skupinami študentov vo faktoroch vonkajšej ako aj vnútornej motivácie. Vychádzali sme z predpokladu, že hlavnými faktormi vonkajšej motivácie študenta pre štúdium na vysokej škole sú rodičia, priatelia, potreba získania dobre platenej práce po ukončení štúdia, potreba dobrého spoločenského postavenia po absolvovaní štúdia a perspektíva lepšej budúcnosti. Ako faktory vnútornej motivácie sme zvolili záľubu v predmete štúdia, túžbu študovať daný odbor, túžbu zlepšiť si poznatky v danej oblasti, získanie nových skúseností, priateľov a túžbu osamostatniť sa. Cieľom bolo taktiež kvantifikovať významnosť rozdielov medzi odpoveďami dvoch skupín študentov prostredníctvom analýzy rozptylu (ANOVA). Ako základný zdroj informácií pre daný výskum sme zvolili dotazníkovú formu získania informácií.

Zaujímavý prieskum o voľbe vysokej školy zrealizovala pre STU v Bratislave agentúra GfK (Growth from Knowledge, 2017). Na rozdiel od našej analýzy, prieskum bol cielený už na stredoškolských študentov plánujúcich vysokoškolské štúdium. Z výsledkov vyplynulo, že pri výbere vysokej školy zohrávajú najvýznamnejšiu rolu kritériá: potenciálne uplatnenie, osobná skúsenosť alebo referencie pri výbere konkrétnej univerzity, prepojenie štúdia s praxou, potenciálne uplatnenie sa s primeraným príjmom. Silný vplyv na vnímanie a výber vysokej školy má aj osobná skúsenosť ako sú návšteva školy, referencie alebo návšteva akcie „dňa otvorených dverí“.

Štruktúra tohto príspevku je nasledujúca: empirickú časť príspevku spolu so získanými výsledkami obsahuje druhá kapitola a príspevok je ukončený záverom a zoznamom literatúry.

2. MOTIVAČNÉ ASPEKTY K ŠTÚDIU NA VYSOKEJ ŠKOLE – EMPIRICKÉ VÝSLEDKY

Výskumnú vzorku výskumu tvorilo celkovo 113 študentov denného štúdia – respondentov. Z toho 64 respondentov tvorili študenti Fakulty informatiky a informačných technológií (FIIT) Slovenskej technickej univerzity v Bratislave, obory bakalárskeho štúdia Informatika a Počítačové a komunikačné systémy a siete. 49 respondentov tvorili študenti z Fakulty hospodárskej informatiky (FHI) Ekonomickej univerzity v Bratislave, odbory bakalárskeho štúdia Manažérske rozhodovanie a informačné technológie a Hospodárska informatika. Študenti boli oslovení formou dotazníka² a všetci študenti boli študentami tretieho ročníka daného bakalárskeho štúdia. Pre výskum boli zámerné zvolené tieto dve skupiny študentov, t.j. študenti s ekonomickým zameraním (FHI) a študenti s technickým zameraním (FIIT) aby bolo možné overiť hlavnú hypotézu rozdielnych motivačných aspektov študentov tzv. humanitných a technických vied. Ako potenciálne motivačné faktory pre štúdium na vysokej škole sme zvolili faktory vnútornej ako aj vonkajšej motivácie. Vychádzali sme z predpokladu, že hlavnými faktormi vonkajšej motivácie sú rodičia, priatelia, potreba získania dobre platenej práce po ukončení štúdia, potreba dobrého spoločenského postavenia po absolvovaní štúdia, perspektíva lepšej budúcnosti. Ako faktory vnútornej motivácie sme zvolili záľubu v predmete štúdia, túžbu študovať daný odbor, túžbu zlepšiť si poznatky v danej oblasti, získanie nových skúseností, priateľov a túžbu osamostatniť sa. Okrem týchto základných faktorov vnútornej a vonkajšej motivácie k štúdiu na vysokej škole dotazník obsahoval aj otázky týkajúce sa hlavne dôvodov pre výber konkrétnej univerzity/fakulty ako aj otázky súvisiace so spokojnosťou, resp. nespokojnosťou študenta so štúdiom a výberom školy.

² Prieskum bol vykonaný v roku 2016 (Furková, 2016).

Z priestorových dôvodov neuvádzame úplné znenie dotazníka (t.j. znenie otázky a možné odpovede) a venujeme pozornosť vybraným získaným výsledkom³.

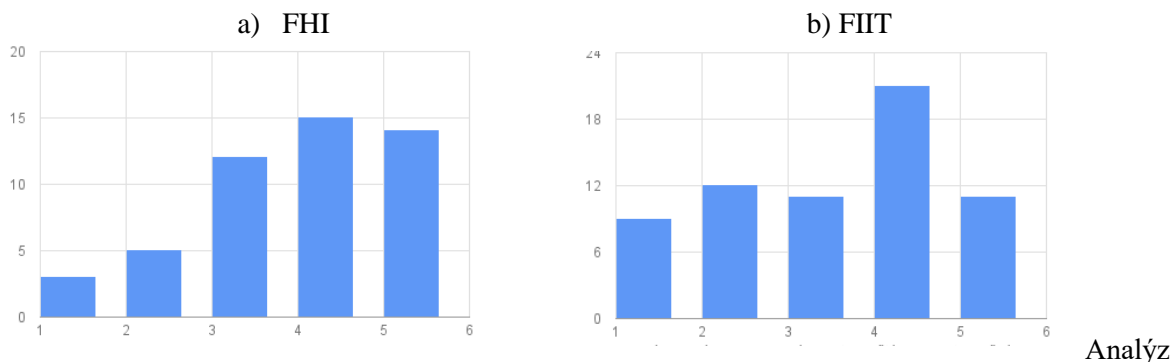
Za najvýznamnejšie faktory vonkajšej motivácie boli u oboch skupín identifikované rovnaké faktory a to perspektíva lepšej budúcnosti a získanie dobre platenej práce. Avšak významné rozdiely nachádzame vo hlavných faktoroch vnútornej motivácie. U študentov FHI sa pri otázke „Aký bol Váš hlavný dôvod pre štúdium na vysokej škole?“ faktory vnútornej motivácie takmer vôbec nevyskytovali, týchto študentov motivujú prevažne potenciálne finančné a materiálne výhody po ukončení školy. Na druhej strane, študenti FIIT okrem potenciálnych finančných a materiálnych výhod sú výrazne motivovaní aj svojimi záľubami a teda predmet štúdia ako dlhodobá záľuba študentov (faktor vnútornej motivácie) má pre túto skupinu rovnako významné postavenie ako najvýznamnejšie faktory vonkajšej motivácie.

Ďalej nás zaujímalo ako významné boli jednotlivé faktory k štúdiu vonkajšej a vnútornej motivácie. Študenti vyjadrili vplyv jednotlivých faktorov na škále od 1 po 5 (1 - žiadny vplyv, 5 - veľmi významný vplyv) a výsledky prezentujeme prostredníctvom histogramov (graf 1 až graf 10). Navyše s cieľom kvantifikovať významnosť rozdielov medzi odpoveďami dvoch skupín študentov sme sa rozhodli využiť metódy kvantitatívnej štatistiky – analýzu rozptylu – ANOVA. V nasledujúcej časti sa bližšie pozrieme na výsledky vzťahujúce sa na jednotlivé motivačné faktory. Vzhľadom na limitujúci rozsah príspevku, grafickú prezentáciu výsledkov (histogramy) a výsledky analýzy rozptylu uvedieme iba pre motivačné faktory, pre ktoré boli zistené štatisticky významné rozdiely medzi skupinami študentov.

Najskôr sa zameriame na výsledky faktorov vonkajšej motivácie. Z výsledkov pre motivačný faktor „*bol som motivovaný rodičmi (okolím, strednou školou, atď.)*“ dedukujeme, že rodičia ako motivačný faktor bol rovnako významný pre obe skupiny študentov. Tento záver potvrdzujú aj výsledky analýzy rozptylu pre tento faktor. Medzi študentami FHI a FIIT nie sú štatisticky signifikantné rozdiely v odpovediach na danú otázku. Obdobné výsledky ako v predchádzajúcom prípade sme získali pre faktor „*štúdium odporúčali priatelia (študenti, resp. absolventi daného štúdia)*“. Výsledky analýzy pre daný faktor potvrdzujú, že obe skupiny študentov považujú tento motivačný faktor za rovnako významný, t. j. nenachádzame štatisticky signifikantné rozdiely medzi odpoveďami týchto dvoch skupín študentov. Vyššie sme už uviedli, že „*získanie dobre platenej práce po ukončení štúdia*“ bolo hlavným motivačným faktorom pre obe skupiny študentov a zistili sme, že tento aspekt je rovnako dôležitý motivačný faktor pre štúdium na vysokej škole pre študentov ekonomického ako aj technického zamerania. Ďalším skúmaným faktorom vonkajšej motivácie bolo „*získanie dobrého spoločenského postavenia po absolvovaní štúdia*“. Ako výsledky naznačujú (pozri graf 1), v rámci tohto skúmaného faktora zisťujeme, že rozdiely medzi skupinami študentov sú štatisticky významné na 10 % nej hladine významnosti. Posledným skúmaným faktorom vonkajšej motivácie bol faktor „*perspektíva lepšej budúcnosti*“. U študentov FIIT ako aj študentov FHI sa tento faktor medzi odpoveďami na otázku „Aký bol Váš hlavný dôvod pre štúdium na vysokej škole?“ vyskytoval najčastejšie. Výsledky pre tento faktor potvrdzujú skutočnosť, že medzi odpoveďami dvoch skupín študentov nie je štatisticky významný rozdiel.

³ Úplné znenie dotazníka ako aj všetky výsledky sú k dispozícii na požiadanie autora.

Graf 1 Histogram a ANOVA „získanie dobrého spoločenského postavenia po absolvovaní štúdia“

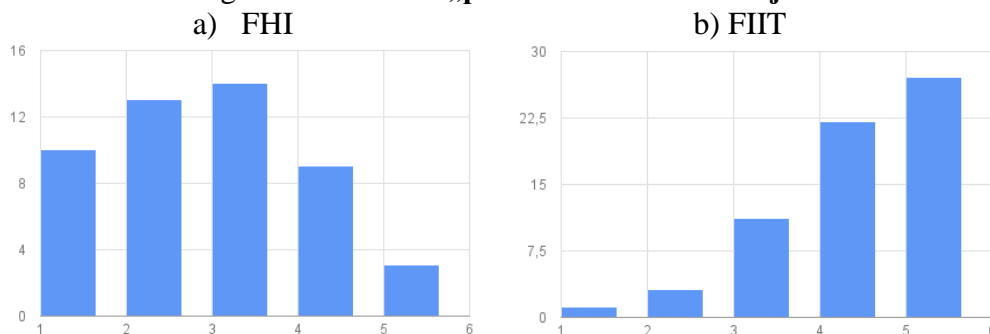


a rozptylu (ANOVA)

Zdroj variability	SS	df	MS	F	P-value	F crit
Medzi skupinami	5,61823	1	5,61823	3,514136	0,063476	3,926607
V rámci skupín	177,4614	111	1,598751			
Spolu	183,0796	112				

Ako faktory vnútornej motivácie sme skúmali: „predmet štúdia bol moja záľuba“, „vždy som túžil študovať daný odbor“, „štúdiom som si chcel zlepšiť poznatky v danej oblasti“, „získanie nových skúseností, priateľov“ a „túžba osamostatniť sa“. Na základe výsledkov (pozri grafy 2 až 6) môžeme konštatovať, že pre všetky tieto motivačné faktory nachádzame štatisticky významné rozdiely medzi študentami FHI ako zástupcami ekonomického zamerania a študentami FIIT ako zástupcami technického smeru. Veľmi významné rozdiely sú pre „predmet štúdia bol moja záľuba“, „vždy som túžil študovať daný odbor“ a „štúdiom som si chcel zlepšiť poznatky v danej oblasti“. Tu je zjavné, že študenti FIIT sú významne motivovaný aj týmito faktormi, o predmete štúdia, t. j. o informatike a informačných technológiách majú už pred nástupom do školy istú predstavu, je to ich záľuba, chcú sa v danej oblasti zdokonaľovať a venovať sa tomuto štúdiu. Študenti FHI naopak faktorom „predmet štúdia bol moja záľuba“, a „vždy som túžil študovať daný odbor“, pripisujú nízku významnosť a skôr sú motivovaný ako sme už uviedli faktormi finančného a materiálneho charakteru.

Graf 2 Histogram a ANOVA „predmet štúdia bol moja záľuba“

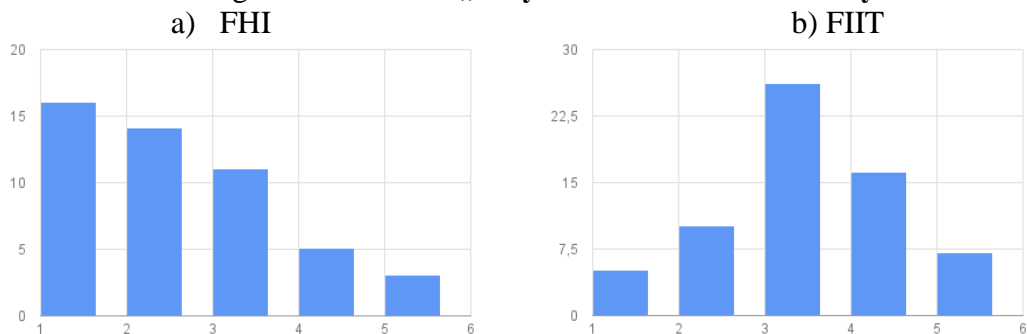


Analýza rozptylu (ANOVA)

Zdroj variability	SS	df	MS	F	P-value	F crit
-------------------	----	----	----	---	---------	--------

Medzi skupinami	60,51946	1	60,51946	53,47514	4,33E-11	3,926607
V rámci skupín	125,6221	111	1,131731			
Spolu	186,1416	112				

Graf 3 Histogram a ANOVA „vždy som túžil študovať daný odbor“

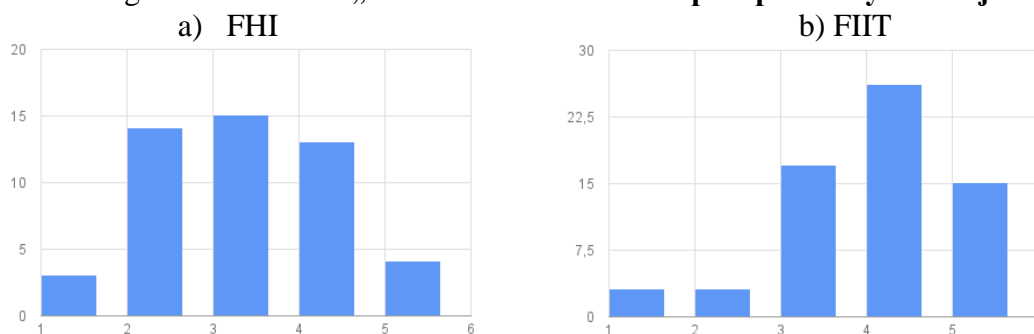


Analý

za rozptylu (ANOVA)

Zdroj variability	SS	df	MS	F	P-value	F crit
Medzi skupinami	21,03153	1	21,03153	16,38964	9,57E-05	3,926607
V rámci skupín	142,4375	111	1,283221			
Spolu	163,469	112				

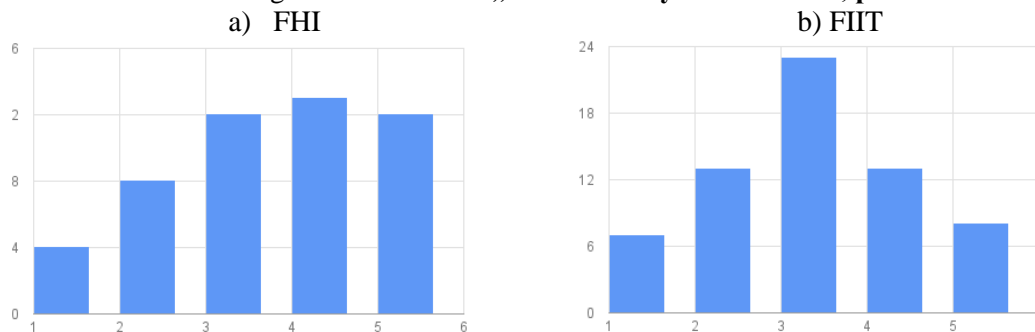
Graf 4 Histogram a ANOVA „štúdiom som si chcel zlepšiť poznatky v danej oblasti“



Analýza rozptylu (ANOVA)

Zdroj variability	SS	df	MS	F	P-value	F crit
Medzi skupinami	14,14665	1	14,14665	12,92794	0,000484	3,926607
V rámci skupín	121,464	111	1,09427			
Spolu	135,6106	112				

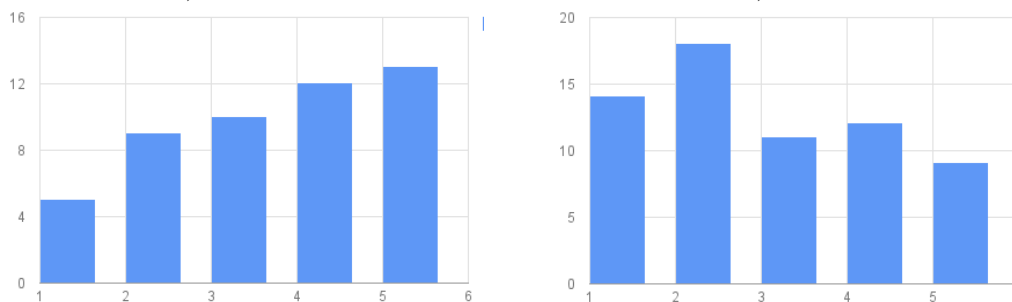
Graf 5 Histogram a ANOVA „získanie nových skúseností, priateľov“



Analýza rozptylu (ANOVA)

Zdroj variability	SS	df	MS	F	P-value	F crit
Medzi skupinami	4,381084	1	4,381084	3,003012	0,085887	3,926607
V rámci skupín	161,9375	111	1,458896			
Spolu	166,3186	112				

Graf 6 Histogram a ANOVA „túžba osamostatniť sa“
a) FHI b) FIIT



Analýza rozptylu (ANOVA)

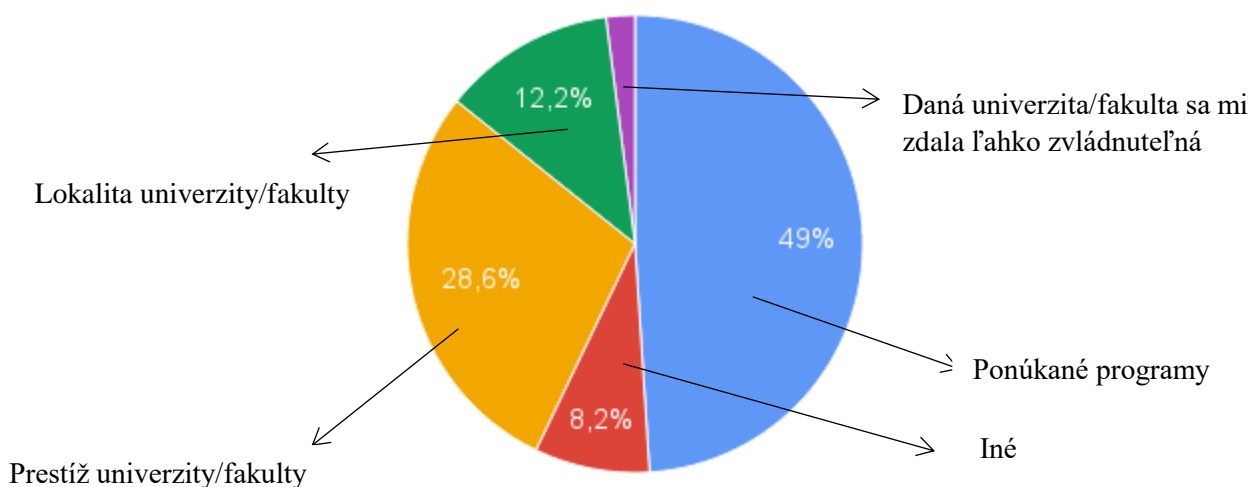
Zdroj variability	SS	df	MS	F	P-value	F crit
Medzi skupinami	11,2877	1	11,2877	6,152917	0,01462	3,926607
V rámci skupín	203,6327	111	1,834528			
Spolu	214,9204	112				

Ďalšia skupina dotazníkových otázok bola zameraná na motivačné aspekty v súvislosti s výberom univerzity/fakulty. Medzi potenciálne dôvody sme zaradili nasledujúce faktory: *prestíž univerzity/fakulty, lokalita univerzity/fakulty, ponúkané programy, vyučujúci učitelia, zvládnuteľnosť danej školy*. Až 49 % študentov FHI uviedlo ako hlavný dôvod pre výber univerzity ponúkané programy (pozri graf 7a) a študentov FIIT najviac oslovuje prestíž univerzity/fakulty (pozri graf 7b), takto sa vyjadrilo až 59,4 % respondentov. Na základe grafu 7 môžeme skonštatovať, že motivačné faktory súvisiace z výberom školy sa medzi sledovanými skupinami študentov odlišujú.

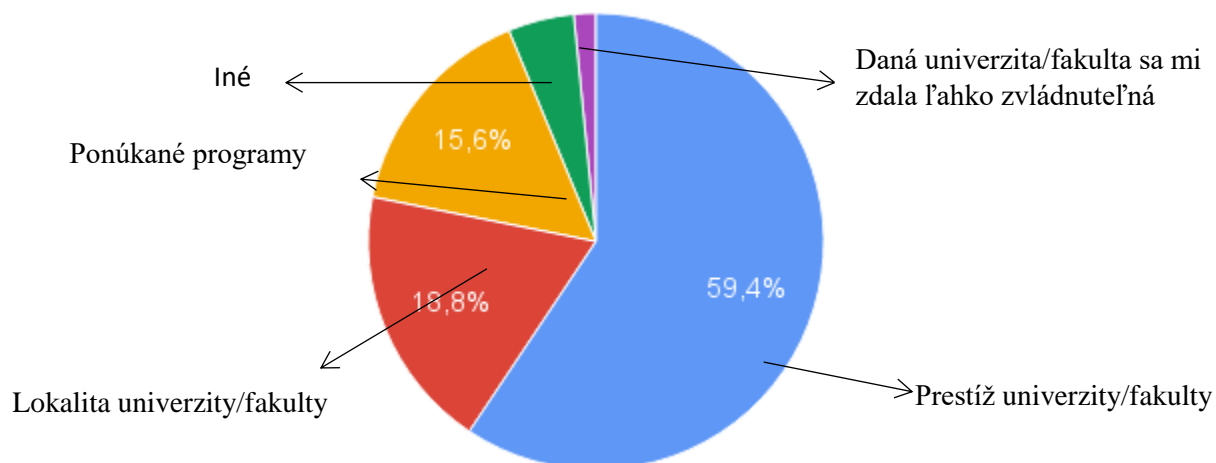
Posledná skupina dotazníkových otázok súvisela so spokojnosťou študentov so štúdiom. Na otázku „Na základe Vašich skúseností vybrali by ste si inú školu pre štúdium Vášho odboru?“ odpovedalo „áno“ 26,5 % študentov FHI a 34,4 % študentov FIIT. Čiže za spokojnejších študentov v súvislosti s výberom školy môžeme považovať študentov FHI. Čo sa týka spokojnosti študentov s výberom študijného odboru, môžeme na základe odpovedí na otázku „Na základe Vašich skúseností vybrali by ste si úplne iný odbor pre Vaše štúdium?“ konštatovať väčšiu spokojnosť u študentov FIIT kde až 85,9 % sa vyjadrilo, že by si nevybrali úplne iný odbor pre štúdium. U študentov FHI to bolo výrazne menej a to 46,9 %.

Graf 7 Výsledky prieskumu na otázku „Aký bol Váš hlavný dôvod pre výber univerzity/fakulty?“

a) FHI



b) FIIT



3. ZÁVER

Analýza bola zameraná na aspekty vnútornej a vonkajšej motivácia k štúdiu na vysokej škole. Ako základný zdroj informácií sme zvolili dotazníkovú formu získania informácií, do ktorého boli zahrnutých 113 študentov dvoch fakúlt a to FIIT Slovenskej technickej univerzity v Bratislave FHI Ekonomickej univerzity v Bratislave. Hlavným cieľom bolo overiť hypotézu o rozdielnych motivačných aspektov študentov tzv. humanitných a technických vied, resp.

zistiť, či je možné nájsť štatisticky významné rozdiely medzi týmito dvoma vysokými školami vo faktoroch vonkajšej ako aj vnútornej motivácie. Zistili sme, že ako najvýznamnejšie faktory vonkajšej motivácie boli u oboch skupín identifikované rovnaké faktory a to perspektíva lepšej budúcnosti a získanie dobre platenej práce. Avšak významné rozdiely sme identifikovali v hlavných faktoroch vnútornej motivácie. U študentov FHI sa pri otázke „Aký bol Váš hlavný dôvod pre štúdium na vysokej škole?“ faktory vnútornej motivácie takmer vôbec nevyskytovali, týchto študentov motivujú prevažne potenciálne finančné a materiálne výhody po ukončení školy. Na druhej strane, študenti FIIT okrem potenciálnych finančných a materiálnych výhod sú výrazne motivovaní aj svojimi záľubami a teda predmet štúdia ako dlhodobá záľuba študentov (faktor vnútornej motivácie) má pre túto skupinu rovnako významné postavenie ako najvýznamnejšie faktory vonkajšej motivácie. V rámci empirickej časti sme preskúmali aj motivačné aspekty v súvislosti s výberom univerzity/fakulty a vyhodnotili sme spokojnosť študentov so štúdiom.

Použitá literatúra

1. DŽUKA, J. 2005. Motivácia a emócie človeka. Prešovská univerzita, 167 s.
2. FURKOVÁ, A. 2016. Motivačné aspekty k štúdiu na vysokej škole. Pedagogické vzdelávanie učiteľov EU v Bratislave.
3. GFK. GROWTH FROM KNOWLEDGE 2017. Motinácia študentov pre štúdium na STU. Správa z kvalitatívneho prieskumu. Dostupné: https://www.stuba.sk/buxus/docs//stu/informacie_o/stu/organy_akademicke_samospravne/zasadnutia_vedenie/11_Analyza_PR_na_STU_priloha_1_2019-06-05.pdf [20/10/2022].
4. LÁSZLÓ, K. 2004. Motivácia v edukačnom prostredí. Banská Bystrica: UMB. 90 s.
5. ŽIAKOVÁ, E. a BALOGOVÁ, B. 2012. Vonkajšia a vnútorná motivácia k štúdiu sociálnej práce. In Výzvy a trendy vo vzdelávaní v sociálnej práci. Zborník príspevkov z medzinárodnej vedeckej konferencie, s. 158 – 167.

Kontaktné údaje

doc. Ing. Andrea Furková, PhD.

Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky

Dolnozemska 1/b, 852 35 Bratislava

Tel: (421 2) 67 295 832

Email: andrea.furkova@euba.sk

MINIMALIZÁCII RIZIKA MEŠKANIA SPOJOV POMOCOU FUZZY ČÍSIEL

MINIMIZING THE RISK OF CONNECTION DELAYS USING FUZZY NUMBERS

Dalibor Gonda, Tomáš Majer

Abstrakt

Problémom plánovania vo verejnej autobusovej doprave je zoradiť daný súbor spojov na minimálny počet turnusov, ktoré spĺňajú viaceré dodatočné obmedzenia. Optimalizácia počtu turnusov bola prevedená vo viacerých optimalizačných cykloch. Najskôr bol minimalizovaný počet autobusov, následne počet prázdnych prejazdov a v záverečnej fáze je potrebné minimalizovať riziko meškania spojov. Plánovanie príchodu spoja v čase zvyšujúcej sa hustoty dopravy je ovplyvnené mnohými faktormi, ktoré možno zaradiť do kategórie náhodných udalostí. Náhodné javy so sebou prinášajú určitý stupeň neistoty, ktorý je možné kvantifikovať pomocou fuzzy čísel.

KLúčové slová: fuzzy číslo, rozvrhovanie vozidiel, optimalizácia

Abstract

The planning problem in public bus transport is to sort a given set of connections into the minimum number of shifts that satisfy several additional constraints. The optimization of the number of shifts was carried out in several optimization cycles. First, the number of buses was minimized, then the number of empty trips, and in the final phase it is necessary to minimize the risk of delays in connections. Planning the arrival of a connection at a time of increasing traffic density is influenced by many factors that can be classified as random events. Random phenomena bring with them a certain degree of uncertainty, which can be quantified using fuzzy numbers.

Keywords: fuzzy number, vehicle scheduling, optimization

1 ÚVOD

V posledných rokoch sledujeme stále narastajúci počet áut na cestách, predovšetkým v blízkosti väčších miest. Nárast individuálnej mobility obyvateľstva na Slovensku potvrdili Pucher a Buehler (2005). Dôvodom tohto negatívneho javu je okrem iného aj rozširovanie výstavby obytných domov do užšieho či širšieho okolia miest. Jednou z možností ako znížiť preplnenosť ciest je zvýšenie atraktivity hromadnej dopravy prostredníctvom postupného zavádzania integrovanej dopravy (Mervart a Novák, 2020). Pod integrovanou dopravou sa rozumie harmonizácia prevádzky mestskej hromadnej dopravy, prímestskej dopravy, regionálnej dopravy. K tomu je potrebná optimalizácia a koordinácia verejnej hromadnej dopravy pomocou zjednotených cestovných poriadkov a jednotnej prestupnej tarify. Jedným z efektov využívania integrovanej dopravy je aj zníženie spotreby energie a tvorby emisií (Pettersson & Frisk, 2016). Integrovaná doprava je pomerne rozšírená predovšetkým v krajinách západnej časti Európskej únie, kde medzi rokmi 1995 a 2013 klesla energetická náročnosť dopravy o 7,4%, kým v rovnakom období vo východnej časti Európskej únie stúpla o 21,4% (Andrés & Padilla, 2018). Ako sme už uviedli, dôležitým prvkom zavádzania

integrovanej dopravy je plánovanie jednotlivých spojov a optimalizácia turnusov pre jednotlivé vozidlá. V rámci integrácie dopravy je potrebné minimalizovať meškanie spojov, predovšetkým v bodoch prepojenia rôznych typov hromadnej prepravy obyvateľov. Stanovenie dopravného plánu závisí od rôznych kvantitatívnych a kvalitatívnych kritérií ovplyvňujúcich proces prepravy, ktoré sa v mnohých prípadoch nedajú presne určiť. Je to spôsobené nerovnomernosťou tokov cestujúcich, ktorá ovplyvňuje výber dopravných služieb. Multikritériálna analýza je účinným nástrojom rozhodovania pri výbere vhodnej alternatívy v závislosti od rôznych kvantitatívnych a kvalitatívnych kritérií (Stoilova & Munier, 2021). Jednou z možností ako zvládnuť neistotu počas rozhodovacieho procesu je aplikácia fuzzy čísel. Podľa (Ďuriš et al., 2021) fuzzy čísla umožňujú zníženie počtu premenných spôsobujúcich neistotu príchodu spoja pri zachovaní maxima informácií, ktoré vstupné údaje nesú. Preto sa Fuzzy teória stala veľmi populárnou a široko používanou, predovšetkým kvôli jej schopnosti relatívne jednoducho pracovať s nepresnými informáciami (Koláček & Matulová, 2018).

2 MATEMATICKÝ MODEL

Spojom nazveme usporiadanú štvoricu $(cod_k, cpr_k, mod_k, mpr_k)$, kde cod_k je čas odchodu z prvej zastávky mod_k (miesto odchodu) a cpr_k je čas príchodu do poslednej zastávky mpr_k (miesto príchodu) spoja k .

Jednotlivé elementy matice $M = \{m(u, v)\}$ predstavujú čas potrebný na prázdny prejazd autobusu zo zastávky u do zastávky v .

Spoj j môže byť obsluhovaný tým istým autobusom po spoji i , ak platí:

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

Teda spoj j môže byť obsluhovaný autobusom, ktorý ukončil obsluhu spoja i na zastávke mpr_i , ak sa stihne presunúť na zastávku mod_j dostatočne rýchlo nato, aby stihol začiatok obsluhy spoja j . Budeme tomu hovoriť, že spoj j môže nasledovať za spojom i a značiť $i < j$.

Turnusom vozidla nazveme takú postupnosť spojov i_1, i_2, \dots, i_k , ktoré môžu byť obslužené jedným autobusom. Pre postupnosť spojov i_1, i_2, \dots, i_k musí platiť

$$i_1 < i_2 < \dots < i_k$$

Turnus predstavuje jednodňový rozvrh práce pre jeden autobus.

Reláciu následnosti spojov $<$ môžeme reprezentovať pomocou digrafu $\vec{G} = (S, E)$, kde S je množina spojov a E je množina možných spojení spojov do turnusu

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

Nech $S = \{1, 2, 3, \dots, n\}$ je množina spojov. Potom binárna premenná x_{ij} bude vyjadrovať, či bude ($x_{ij} = 1$) alebo nebude ($x_{ij} = 0$) spoj j obsluhovaný bezprostredne za spojom i jedným autobusom v jednom turnuse.

2.1 Minimalizácia počtu potrebných autobusov

Cieľom optimalizácie je rozdeliť množinu spojov S do čo najmenšieho počtu turnusov tak, aby každý spoj patril práve do jedného z nich. Tým dosiahneme najmenší počet autobusov. Množina spojov obsahuje n spojov. Ak by žiaden spoj nemohol nasledovať po niektorom inom, tak by sme potrebovali n autobusov, pretože každý turnus by mohol obsahovať iba jeden spoj. Tento prípad by nastal v prípade, ak by všetky hodnoty premennej x_{ij} boli rovné

nule. Je zrejmé, že počet potrebných autobusov bude klesať s počtom prípadov, keď $x_{ij} = 1$, pričom s každou hodnotou $x_{ij} = 1$ ušetríme jeden autobus. Na základe uvedeného počet autobusov p možno vyjadriť vzťahom

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

Potom matematický model na riešenie problému s minimálnym počtom turnusov by teda mal tvar:

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

$$\text{za podmienok} \quad \sum_{i:(i,j) \in E} x_{ij} \leq 1 \text{ pre } j = 1..n \quad (4)$$

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

$$x_{ij} \in \{0,1\} \text{ pre všetky } i,j = 1..n \quad (6)$$

Matematický model (3) až (6) predstavuje model priradovacieho problému, ktorý bude dávať celočíselné riešenie aj bez podmienky (6), ktorá môže byť nahradená podmienkou:

$$x_{ij} \geq 0 \text{ pre všetky } i,j = 1..n \quad (7)$$

Úloha navrhnúť turnusový plán s minimálnym počtom autobusov bez ďalších požiadaviek je teda riešiteľná v polynomiálnom čase.

2.2 Minimalizácia počtu potrebných autobusov

Vyriešením modelu (3) až (6) sme zistili potrebný počet autobusov, označme ho symbolom T . Ak zachovanie tohto počtu vynútime pridaním štrukturálnej podmienky, môžeme účelovú funkciu zmeniť a minimalizovať súčet prázdnych prejazdov. Model by mal potom tvar:

$$\text{Minimalizuj} \quad \sum_{i,j:(i,j) \in E} x_{ij} m(mpr_i, mod_j) \quad (8)$$

$$\text{za podmienok} \quad \sum_{i:(i,j) \in E} x_{ij} \leq 1 \text{ pre } j = 1..n \quad (9)$$

$$\sum_{j:(i,j) \in E} x_{ij} \leq 1 \text{ pre } i = 1..n \quad (10)$$

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

$$x_{ij} \in \{0,1\} \text{ pre všetky } i,j = 1..n \quad (12)$$

2.3 Koncept fuzzy množiny

Fuzzy množinu možno interpretovať ako množinu, ktorá prvkom priraduje neurčitosť v podobe tzv. miery príslušnosti (Ďuriačová et al., 2011). Fuzzy množina A je množina prvkov $x \in U$ (U je tzv. univerzum), kde každému z nich je priradený stupeň príslušnosti $\mu_A(x)$, ktorej hodnoty sa pohybujú v intervale $\langle 0; 1 \rangle$. Fuzzy množiny sa môžu stotožňovať s funkciou

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

ktorú nazývame funkcia príslušnosti fuzzy množiny A .

Definícia: Nech je daná množina U , ktorú nazývame univerzum. Potom fuzzy množinou $A \in U$ rozumieme takú množinu, ktorá je definovaná zobrazením

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

Pre každé $x \in U$ nazveme hodnotu $\mu_A(x)$ stupeň príslušnosti prvku x k fuzzy množine A .

Napriek možnosti stotožniť fuzzy množinu A s funkciou príslušnosti μ_A budeme rozlišovať medzi fuzzy množinou A a jej funkciou príslušnosti. Fuzzy množinu budeme vnímať ako usporiadanú dvojicu, ktorú formálne zapíšeme v tvare

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

Tvar a parametre funkcie príslušnosti môžu byť v jednotlivých prípadoch stanovené na základe praktických skúseností alebo známych vlastností analyzovaného javu.

Zadefinujeme ešte ďalšie dôležité pojmy, ktoré sú potrebné pre vymedzenie pojmu fuzzy číslo.

Definícia: Nech A je fuzzy množina prvkov $x \in U$. Potom

- ostrú množinu $Supp A = \{x \in U; \mu_A(x) > 0\}$ nazývame nosičom fuzzy množiny;
- ostrú množinu $Ker A = \{x \in U; \mu_A(x) = 1\}$ nazývame jadrom fuzzy množiny,
- ostrú množinu $A_\alpha = \{x \in U; \mu_A(x) > \alpha\}$ nazývame α -rezom fuzzy množiny.

3 OPERÁCIE S FUZZY ČÍSLAMI

Definícia: Fuzzy množinu A definovanú na množine reálnych čísel \mathbb{R} nazývame fuzzy číslo, ak spĺňa nasledujúce vlastnosti:

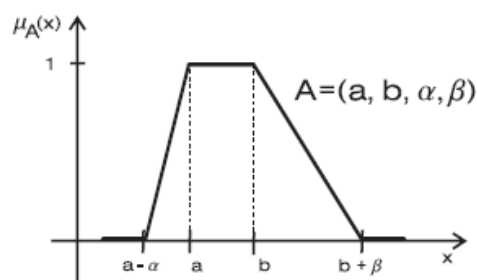
- A je normálna, tj. existuje $x \in \mathbb{R}$ také, že $\mu_A(x) = 1$
- α -rezy A_α sú pre všetky $\alpha \in (0, 1)$ uzatvorené intervaly;
- nosič $Supp A$ je ohraničená podmnožina množiny \mathbb{R} .

Definícia: Fuzzy číslo A sa nazýva LR fuzzy číslo, ak jeho funkcia príslušnosti je daná predpisom

$$\mu_A(x) = \begin{cases} 1 & x \in \langle a, b \rangle \\ L\left(\frac{a-x}{\alpha}\right) & x \in \langle a-\alpha, b \rangle \\ R\left(\frac{x-b}{\beta}\right) & x \in \langle a, b+\beta \rangle \\ 0 & inde \end{cases}$$

kde $a, b \in \mathbb{R}, \alpha, \beta > 0$.

Funkcie L a R sa nazývajú tvarové funkcie a čísla α, β sú ľavá, resp. pravá neurčitosť. LR fuzzy čísla budeme označovať $A = (a, b, \alpha, \beta)_{LR}$. Význačným typom LR fuzzy čísel sú lineárne fuzzy čísla, ktoré dostaneme pre lineárne tvarové funkcie $L(x) = R(x) = 1 - x$. Lineárne fuzzy čísla sú buď trojuholníkové ($a = b$), alebo lichobežníkové ($a \neq b$).



Obrázok 1 Lichobežníkové fuzzy číslo $A = (a, b, \alpha, \beta)$

Veta 1: Štandardným súčtom LR fuzzy čísel

$$A_1 = (a_1, b_1, \alpha_1, \beta_1)_{LR} \quad a \quad A_2 = (a_2, b_2, \alpha_2, \beta_2)_{LR}$$

je LR číslo

$$A_1 \oplus A_2 = (a_1 + a_2, b_1 + b_2, \alpha_1 + \alpha_2, \beta_1 + \beta_2)_{LR}$$

Poznámka: Štandardným súčtom dvoch lineárnych fuzzy čísel je lineárne fuzzy číslo.

Ľavá (pravá) neurčitosť súčtu dvoch lineárnych fuzzy čísel je súčtom ľavých (pravých) neurčítostí sčítovaných fuzzy čísel, a teda pri postupnom sčítovaní viacerých čísel dochádza k zväčšovaniu neurčitostí – výsledok je dosť „rozmazaný“. Tomuto javu pri sčítovaní sa dá predísť pomocou t -noriem. Dá sa dokázať, že pre každú t -normu T , pre ktorú platí $T \leq T_L$ (Lukasiewiczova t -norma) a pre každú spojitú archimedovskú t -normu s konkávnym aditívnym generátorom platí

$$A_1 \oplus_T A_2 = (a_1 + a_2, b_1 + b_2, \max(\alpha_1, \alpha_2), \max(\beta_1, \beta_2)).$$

Podrobnejšie o t -normách v Klement et al (2000).

3.1 Usporiadanie fuzzy čísel

Pre usporiadanie fuzzy čísel existuje viacero metód, z ktorých uvedieme:

- usporiadanie podľa α -rezov;
- usporiadanie podľa číselných charakteristík.

Usporiadanie fuzzy čísel pomocou α -rezov vychádza z relácie usporiadania definovanej na množine uzatvorených intervalov (α -rezy sú uzatvorené intervaly).

Definícia: Hovoríme, že fuzzy číslo A je menšie alebo rovné podľa α -rezov ako fuzzy číslo B , označujeme $A \leq_\alpha B$, ak platí

$$\forall \alpha \in (0, 1): A_\alpha \leq B_\alpha.$$

Ak navyše platí $A \leq_\alpha B$ a $A \neq B$, potom fuzzy číslo A je menšie ako fuzzy číslo B podľa α -rezov.

Usporiadanie fuzzy čísel podľa číselnej charakteristiky spočíva v tom, fuzzy čísla nahradíme reálnymi číslami, ktoré ich z určitého hľadiska najlepšie vystihujú.

Definícia: Nech $A = (a, b, \alpha, \beta)$ je lineárne fuzzy číslo. Pre medián m_A fuzzy čísla A platí

$$m_A = \frac{2a + 2b - \alpha + \beta}{4}.$$

Definícia: Hovoríme, že fuzzy číslo A je menšie alebo rovné fuzzy číslu B vzhľadom k ich mediánom, označujeme $A \leq_m B$, ak platí $m_A \leq m_B$.

Definícia: Nech $A = (a, b, \alpha, \beta)$ je lineárne fuzzy číslo. Pre ťažisko t_A fuzzy čísla A platí

$$t_A = \frac{3b^2 - 3a^2 + \beta^2 - \alpha^2 + 3b\beta - 3a\alpha}{3(2b - 2a + \alpha + \beta)}.$$

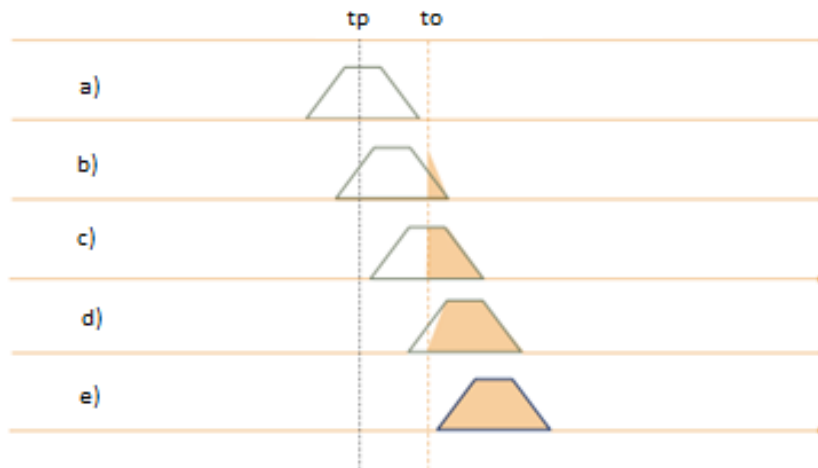
Definícia: Hovoríme, že fuzzy číslo A je menšie alebo rovné fuzzy číslu B vzhľadom k ich ťažiskám, označujeme $A \leq_t B$, ak platí $t_A \leq t_B$.

4 PRÍCHOD A ODKHOD SPOJA VYJADRENÝ POMOCOU FUZZY ČÍSIEL

Cieľom ďalšieho kola optimalizácie je rovnomerne rozložiť riziko meškania medzi jednotlivé turnusy. Príchod spoja je zaťažený väčšou či menšou mierou neistoty. Túto neistotu vyjadríme tak, že čas príchodu spoja vyjadríme pomocou obdĺžnikového fuzzy čísla

$$C = (c_1, c_2, \alpha, \beta)$$

kde c_1 bude predstavovať najskorší očakávaný príchod spoja, pre ktorý platí $\mu_C(t) = 1$ a c_2 bude predstavovať najneskorší očakávaný príchod spoja, pre ktorý platí $\mu_C(t) = 1$. Hodnota $c_1 - \alpha$ je najskorší možný príchod spoja a hodnota $c_2 + \beta$ je najneskorší možný príchod spoja. Na obrázku 2 sú znázornené základné možnosti, ktoré môžu nastať vzhľadom na tarifný príchod predchádzajúceho spoja (t_p) a tarifný odchod nasledujúceho spoja.



Obrázok 2 Príchod a odchod spoja

Z obrázku 2 vyplýva, že v prípadoch b) až e) je potrebné pomocou fuzzy čísla vyjadriť aj odchod nasledujúceho spoja. Možnosti a) a b) považujeme za prípady s malou mierou rizika oneskoreného odchodu nasledujúceho spoja. Ostatné možnosti považujeme za prípady s veľkou mierou rizika meškania odchodu nasledujúceho spoja. Uvedeným spôsobom budú ohodnotené všetky spoje v danom turnuse. Tým budú rozdelené turnusy na dve skupiny: turnusy s malou mierou rizika meškania; turnusy s veľkou mierou rizika meškania. Následne bude realizované kríženie turnusov tak, aby novo navrhnuté turnusy mali čo najmenšiu mieru rizika meškania spojov.

5 ZÁVER

Cieľom je vytvoriť turnusy tak, aby sme minimalizovali celkový počet zmeškaných spojov (pri zadanej fuzzy hladine). Samozrejme bez navýšenia potrebného počtu autobusov a vodičov a len s miernym navýšením veľkosti prázdnych prejazdov.

Ak za spojom, pri ktorom je veľké riziko meškania, nie je možné urobiť dlhšiu pauzu a meškanie sa preniesie aj na v poradí druhý spoj, je vhodné zabezpečiť dlhšiu pauzu aspoň za druhým spojom tak, aby sa znížilo riziko prenesenia meškania aj na tretí spoj. Práve toto sa budeme snažiť dosiahnuť krížením turnusov. Takýmto prekrižením môže vzniknúť alebo narásť veľkosť prázdneho prejazdu medzi spojmi. Prejazd ale zároveň skracuje pauzu medzi spojmi a tak kritérium minimalizácie prázdnych prejazdov nie je v protiklade s maximalizáciou pauzy medzi spojmi.

Pod'akovanie

Tento príspevok vznikol vďaka podpore grantu VEGA 1/0776/20 " Obehové rozvrhy vozidiel v podmienkach neurčitosti".

Použitá literatúra

1. Andrés, L., & Padilla, E. (2018). Driving factors of GHG emissions in the EU transport activity. *Transport Policy*, 61, 60-74.
2. Ďuračiová, R., Lieskovský, T., Kročková, K., & Sabo, M. (2011). Multicriteria Decision Making Using Fuzzy Sets in GIS Environment and Its Application in the Archaeological Prediction.
3. Ďuriš, V.; Bartková, R.; Tirpáková, A. Principal Component Analysis and Factor Analysis for an Atanassov IF Data Set. *Mathematics* 2021, 9, 2067. <https://doi.org/10.3390/math9172067>
4. Koláček, M., & 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., & Novák, R. (2020). PRAŽSKÁ A STŘEDOČESKÁ INTEGROVANÁ DOPRAVA. *Perner's Contacts*, 15(2).
6. Pettersson, F., & 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), 64-82.
7. Pucher, J., & Buehler, R. (2005). Transport policies in central and Eastern Europe. *Transport strategy, policy, and institutions*, 860.
8. Stoilova, S., & Munier, N. (2021). A novel fuzzy SIMUS multicriteria decision-making method. An application in railway passenger transport planning. *Symmetry*, 13(3), 483.

Kontaktné údaje

doc. PaedDr. Dalibor Gonda, PhD.

Žilinská univerzita v Žiline, Fakulta riadenie a informatiky,

Univerzitná 8215/1, 01026 Žilina

Tel: (421 41) 513 4251

email: dalibor.gonda@fri.uniza.sk

Ing. Tomáš Majer, PhD.

Žilinská univerzita v Žiline, Fakulta riadenie a informatiky,

Univerzitná 8215/1, 01026 Žilina

Tel: (421 41) 513 4251

email: tomas.majer@fri.uniza.sk

KONSTRUKCE KOMPOZITNÍHO INDIKÁTORU NA ZÁKLADĚ ANALÝZY OBALU DAT

COMPOSITE INDICATOR CONSTRUCTION BASED ON DATA ENVELOPMENT ANALYSIS

Jakub Hanousek

Abstrakt

Kompozitní idikátory jsou populární nástroj pro hodnocení komplexních systémů, kde jednotlivá kritéria mohou být velice různorodá. Samotný výpočet kompozitního indikátoru může být do jisté míry zkreslen rozhodovatelem. Tento příspěvek ukazuje, jak vypočítat kompozitní indikátor nezávislý na rozhodovatel. Výpočet kompozitního indikátoru je realizován skrze model analýzy obalu dat. Takto získaný kompozitní indikátor je dále podroben testu robustnosti.

***Klíčové slová:** Kompozitní idikátory, analýza obalu dat, robustnost*

Abstract

Composite indicators are a popular tool for evaluating complex systems where individual criteria can be very diverse. The calculation of the composite indicator itself may be biased by the decision-maker to some extent. This post shows how to calculate a decision maker independent composite indicator. The calculation of the composite indicator is implemented through the data envelopment analysis model. The composite indicator thus obtained is further subjected to a robustness test.

***Keywords:** Composite indicators, data envelopment analisys, robustness*

1 ÚVOD

Poměřovat, vytvářet žebříčky, pořadí, být schopen vyhodnotit jednotlivé varianty. Tato potřeba je lidstvu známa několik tisíc let. Za tímto účelem vznikla celá vědní disciplína teorie rozhodování.

Do teorie rozhodování bezesporu náleží i kompozitní indikátory. Kompozitní indikátor je bezrozměrný index, který vznikne agregací jednotlivých kritérií. Samotná konstrukce kompozitního indikátoru však může být do značné míry závislá na rozhodovateli. Takto vytvořený kompozitní indikátor může být zkreslující. Tento příspěvek ukazuje postup výpočtu kompozitního indikátoru, který je nezávislý na rozhodovateli, a tudíž maximálně objektivní. Kompozitní indikátor bude konstruován prostřednictvím modelů analýzy obalu dat.

2 VÝPOČET KOMPOZITNÍCH INDIKÁTORŮ

Mějme množinu kritérií K . Tuto množinu rozdělme na maximalizační kritéria X a minimalizační kritéria Y . Vytvoříme skupiny podobných kritérií podle jejich charakteristiky. Skupiny kritérií dle charakteristiky budeme značit indexem b . V rámci zachování homogenity hodnocených jednotek, jednotky rozdělíme do skupin dle určitého klastrovacího

kritéria. Tyto skupiny budeme značit indexem p . Tímto způsobem dostaneme matice X^{pb} a Y^{pb} .

$$\mathbf{X}^{pb} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \quad (1)$$

$$\mathbf{Y}^{pb} = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1r} \\ y_{21} & y_{22} & \dots & y_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ y_{m1} & y_{m2} & \dots & y_{mr} \end{bmatrix} \quad (2)$$

Dílčí kompozitní indikátory získáme na základě výsledků z modelu (3).

Minimalizovat: θ_q^{pb} ,

Za podmíněk: $\sum_{i=1, i \neq q}^m x_{ij}^{pb} \lambda_i^{pb} + s_j^{pb-} = \theta_q^{pb} x_{qj}^{pb}$, $j = 1, 2, \dots, n$,

$$\sum_{i=1, i \neq q}^m y_{ik}^{pb} \lambda_i^{pb} - s_k^{pb+} = y_{qk}^{pb}$$
, $k = 1, 2, \dots, r$, (3)

$$\lambda_q^{pb} = 0,$$

$$\lambda_i^{pb} \geq 0, s_k^{pb+} \geq 0, s_j^{pb-} \geq 0, \quad j = 1, 2, \dots, n, \quad k = 1, 2, \dots, r, \quad i = 1, 2, \dots, m.$$

Výsledný kompozitní indikátor získáme agregací výsledků dílčích kompozitních indikátorů pomocí modelu (3), kde matice X a Y jsou definovány následujícím vztahem.

$$\mathbf{X}^{p0} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad \mathbf{Y}^{p0} = \begin{bmatrix} I_1^{p1} & \dots & I_1^{pb} \\ \vdots & \ddots & \vdots \\ I_m^{p1} & \dots & I_m^{pb} \end{bmatrix} \quad (4)$$

3 TESTOVÁNÍ ROBUSTNOSTI VÝSLEDKŮ

Jedním z potencionálních problémů modelů analýzy obalu dat je robustnost výsledků. Hodnoty jednoho z kritérií mohou zásadně ovlivnit výsledky celého modelu.

Tato práce navrhuje, jak změřit sílu dílčích kompozitních indikátorů při výpočtu výsledného kompozitního indikátoru pomocí korelace. Korelace je měřena Pearsonovým korelačním koeficientem. Algoritmus výpočtu je představen níže.

Nechť I_i^{p0} je celkový kompozitní indikátor i -té jednotky ve skupině p .

Pro $k = 1, \dots, r$:

Vypuť dílčí kompozitní indikátor k z datasetu.

Pro $i = 1, \dots, m$ spočti celkové kompozitní indikátory $I_i^{p,-k}$ pro jednotlivé jednotky ve skupině p .

Pro $k_1 = 0, 1, \dots, r-1$ a $k_2 = k_1 + 1, k_1 + 2, \dots, r$, spočti korelaci r_{k_1, k_2}^p mezi vektory indexu $I^{p,-k_1}$ a $I^{p,-k_2}$.

Jsou-li korelace nízké, znamená to, že výsledný kompozitní indikátor dosažený výpočtem modelem analýzy obalu dat je robustní. Žádný dílčí indikátor neovlivňuje zásadně dosažené výsledky.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

4 ZÁVĚR

Cílem příspěvku bylo definovat výpočet kompozitních indikátorů nezávislých na rozhodovateli. Pro výpočet kompozitního indikátoru byl zvolen model analýzy obalu dat. V příspěvku bylo ukázáno, jak je možné vypočítat dílčí kompozitní indikátory. Dále byl navržen postup, jak z dílčích kompozitních indikátorů získat výsledný kompozitní indikátor.

Pro výsledný kompozitní indikátor byl navrhnout algoritmus pro test robustnosti. Výpočet testu robustnosti je realizován pomocí Pearsonova korelačního koeficientu. Jsou – li získané korelace nízké. Znamená to, že příslušná kritéria, či dílčí kompozitní indikátory jsou mezi sebou odlišné. V takovém případě může jedno kritérium, či dílčí kompozitní indikátor zásadně ovlivnit výsledky. Pokud jsou korelace vysoké, výsledky jsou robustní a žádné kritérium či dílčí kompozitní indikátor výsledky zásadně neovlivňuje.

Další předmět výzkumu bude tyto navržené algoritmy otestovat na reálných datech.

Poděkování:

Článek vznikl za podpory projektu IGA F4/42/2021 Interní grantové agentury Vysoké školy ekonomické v Praze.

Použitá literatura

1. Allen, R., Athanassopoulos, A., Dyson, R.G., Thanassoulis, E., 1997. Weights restrictions and value judgements in Data Envelopment Analysis: Evolution, development and future directions. *Ann. Oper. Res.* 73, 13–34.
2. Belton, V., Stewart, T.J., 1999. DEA and MCDA: Competing or Complementary Approaches?, in: Meskens, N., Roubens, M. (Eds.), *Advances in Decision Analysis, Mathematical Modelling: Theory and Applications*. Springer Netherlands, Dordrecht, pp. 87–104.
3. Hanousek, J. Evaluation of the health condition and medical resources in the municipalities in the Czech Republic. In 40th International Conference on Mathematical Methods in Economics 2022. Jihlava: College of Polytechnics Jihlava, 2022, s. 86 - 90.
4. Hwang, C. L., Yoon, K.: *Multiple Attribute Decision Making. Methods and Application*. Berlin: Springer-Verlag, 1981.
5. CHARNES, A., COOPER, W.W., LEWIN, A., SEIFORD, L.: *Data envelopment analysis: Theory, methodology and applications*. Boston: Kluwer Academic Publisher, 1994, 422 s.
6. Cherchye, L., Moesen, W., Rogge, N., Puyenbroeck, T., 2007. An Introduction to ‘Benefit of the Doubt’ Composite Indicators. *Soc. Indic. Res.* 82, 111–145.
7. JABLONSKÝ, J., DLOUHÝ, M.: *Modely hodnocení efektivnosti produkčních jednotek*. Praha: Professional Publishing, 2004, 183 s.

Kontaktné údaje

Ing. Jakub Hanousek

Vysoká škola ekonomická v Praze, fakulta informatiky a statistiky

Náměstí Winstona Churchilla 1938/4, 130 67 Praha 3

Email: xhanj52@vse.cz

PRIESTOROVÁ ANALÝZA VÝSLEDKOV TESTOVANIA 9 Z MATEMATIKY V OKRESOCH SR PRED A PO COVID-19¹

SPATIAL ANALYSIS OF THE TESTING 9 RESULTS FROM MATHS IN DISTRICTS OF SR BEFORE AND AFTER COVID-19

Michaela Chocholatá

Abstrakt

Predmetom príspevku je priestorová analýza priemernej výkonnosti žiakov v okresoch SR na báze výsledkov Testovania 9 z matematiky v období pred a po pandémie Covid-19, t.j. v rokoch 2019 a 2022. Príspevok aplikuje exploratóčnú analýzu priestorových dát (ESDA) ako sú box-mapy, Moranov rozptylový diagram, lokálna Moranova I štatistika a lokálna Gearyho c štatistika. Výsledky potvrdzujú pokles priemernej výkonnosti žiakov v sledovanom období a umožňujú odhaliť priestorové prepojenia medzi jednotlivými okresmi.

Príučové slová: Testovanie 9, exploratóčná analýza priestorových dát, priestorová autokorelácia

Abstract

This paper deals with the spatial analysis of the average performance of pupils in the districts of SR based on the results of Testing 9 in Mathematics in the period before and after the Covid-19 pandemic, i.e., in 2019 and 2022, respectively. The paper applies exploratory spatial data analysis (ESDA) such as box-maps, Moran's scatter plot, local Moran's I statistic and local Geary's c statistic. The results confirm the drop in the average performance of pupils in the observed period and allow revealing the spatial linkages between individual districts.

Keywords: Testing 9, exploratory spatial data analysis, spatial autocorrelation

1 ÚVOD

Po dvoch rokoch pandémie Covid-19 sa v školskom roku 2021/2022 opäť uskutočnilo celoslovenské testovanie žiakov 9. ročníka základných škôl a 4. ročníka gymnázií a stredných športových škôl s osemročným vzdelávacím programom známe pod názvom Testovanie 9. Riadny termín Testovania 9 2022 bol 6. apríla 2022 z predmetov matematika, slovenský jazyk a literatúra, maďarský jazyk a literatúra a 7. apríla 2022 z predmetu slovenský jazyk a slovenská literatúra na školách s vyučovacím jazykom iným ako slovenským (Národný ústav certifikovaných meraní vzdelávania, 2022a). Cieľom Testovania 9 je jednak porovnávať výkony žiakov v testoch, ako aj získať obraz o výkonoch žiakov na výstupe zo ZŠ.

Predmetom príspevku je porovnanie výkonnosti žiakov pred a po pandémie Covid-19 (t.j. v roku 2019 a v roku 2022) v jednotlivých okresoch Slovenska na základe priemerných výsledkov Testovania 9 z matematiky s využitím nástrojov priestorovej analýzy dát.

Štruktúra príspevku je nasledujúca: po úvode nasleduje časť venovaná metodológii, v ktorej predstavíme vybrané nástroje exploratóčnej analýzy priestorových dát (ESDA), empirická časť a príspevok uzatvára záver.

¹ Príspevok bol spracovaný v rámci riešenia grantových úloh VEGA 1/0193/20 a VEGA 1/0211/21.

2 VYBRANÉ NÁSTROJE ESDA

Na preskúmanie priestorovej štruktúry analyzovaných regiónov možno využiť prístupy exploratórnej analýzy priestorových dát (ESDA), kde patrí napr. znázornenie pozorovaných hodnôt na mape. Hoci niektoré mapy (napr. box-mapy) umožňujú identifikovať určité priestorové vzory, zhluky hodnôt, extrémne hodnoty či odľahlé hodnoty, neposkytujú však informáciu o štatistickej významnosti zhlukovania. Ďalším krokom ESDA je preto spravidla analýza priestorovej autokorelácie na globálnej i lokálnej úrovni. Pojem priestorová autokorelácia označuje koreláciu premennej s jej priestorovo posunutou hodnotou, čo indikuje, že geograficky blízke regióny sú vzájomne prepojené. Globálne štatistiky poskytujú informáciu o priestorovej autokorelácii z globálneho hľadiska, t.j. ide o jednu hodnotu pre celý analyzovaný súbor dát. Lokálne štatistiky poskytujú informáciu o priestorovej autokorelácii pre jednotlivé analyzované regióny (okresy). Medzi najčastejšie používané štatistiky patria Moranovo I , Gearyho c a Getisova-Ordova G štatistika. Veľmi často sa možno stretnúť s grafickým znázornením pomocou Moranovho rozptylového diagramu, ktorého vodorovná os obsahuje analyzovanú premennú a zvislá os jej priestorovo posunuté hodnoty. Tento diagram obsahuje okrem hodnoty globálnej štatistiky aj hodnoty lokálnych štatistik a je rozdelený do štyroch kvadrantov, z ktorých dva zodpovedajú pozitívnej autokorelácii (pravý horný a ľavý dolný) a dva zodpovedajú negatívnej autokorelácii (pravý dolný a ľavý horný). S ohľadom na empirickú časť príspevku predstavíme lokálnu Moranovu I štatistiku a lokálnu Gearyho c štatistiku.

Vzťah pre výpočet lokálnej Moranovej I štatistiky pre jednotlivé regióny (okresy) je nasledovný (Anselin, 1995):

$$I_i = \frac{(x_i - \bar{x}) \sum_{j=1}^N w_{ij} (x_j - \bar{x})}{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad i \neq j \quad (1)$$

kde symbol x_i označuje hodnotu analyzovanej premennej v regióne i , \bar{x} označuje výberový priemer, N je počet analyzovaných regiónov a w_{ij} sú prvky priestorovej matice váh \mathbf{W} rozmeru $N \times N$ zodpovedajúce dvojici regiónov i, j .

Lokálna Gearyho c štatistika vychádza zo štvorcových diferencií, resp. nepodobnosti (rozdielnosti). Nízke hodnoty tejto štatistiky označujú pozitívnu priestorovú autokoreláciu (podobnosť), kým jej vysoké hodnoty svedčia v prospech negatívnej autokorelácie (rozdielnosti). Lokálnu Gearyho štatistiku možno zapísať v tvare (Anselin, 2019; Fischer a Wang, 2011; Furková, 2019):

$$c_i = \sum_j w_{ij} (x_i - x_j)^2 \quad (2)$$

kde x_i označuje hodnotu analyzovanej premennej v regióne i ($i = 1, 2, \dots, N$) a symbol w_{ij} označuje prvky priestorovej matice váh \mathbf{W} .

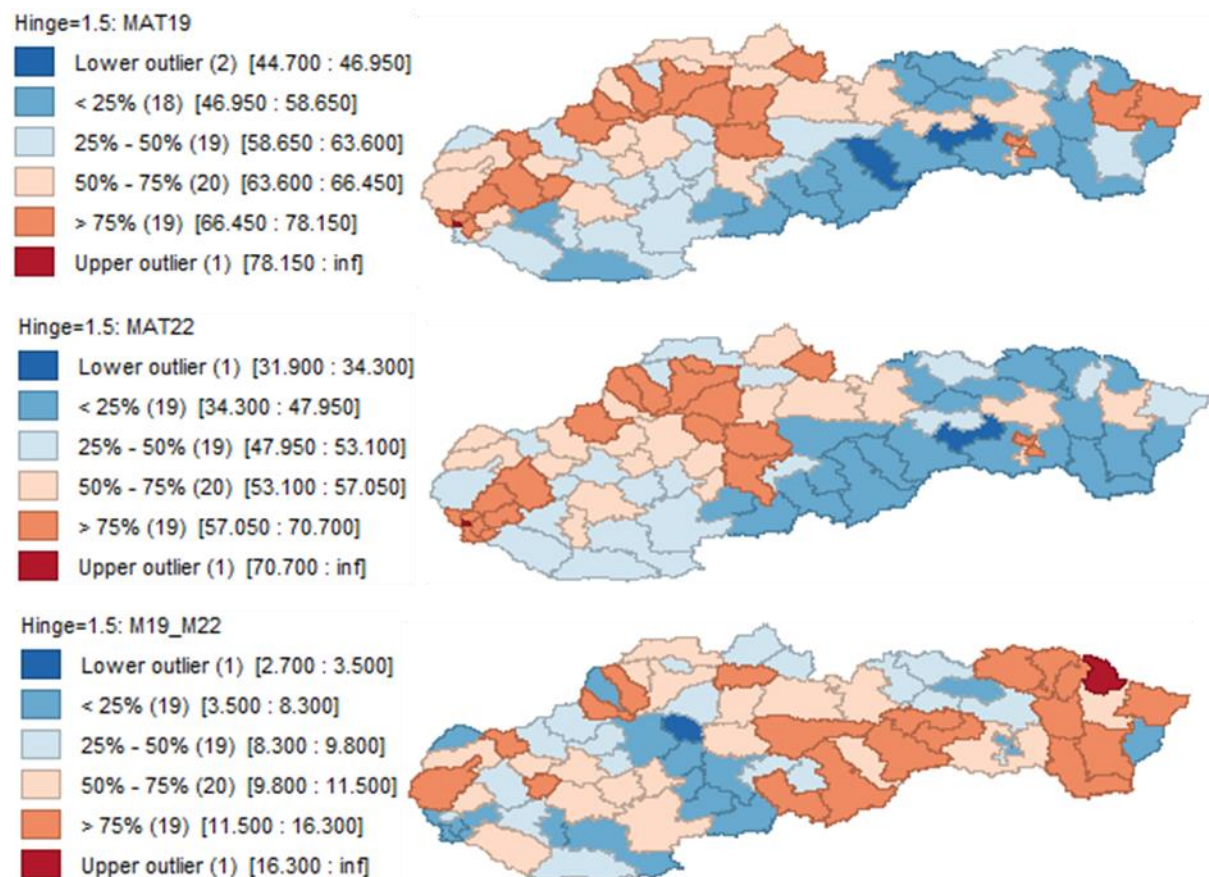
3 EMPIRICKÉ VÝSLEDKY ANALÝZY

Predmetom príspevku je priestorová analýza výsledkov Testovania 9 z matematiky za jednotlivé okresy Slovenskej republiky (SR) pred a po pandémie Covid-19, t.j. analýza výsledkov z roku 2019 a 2022. Priemerné výsledky dosiahnuté v 79 okresoch SR boli získané z Národného ústavu certifikovaných meraní vzdelávania (2019, 2022b). Analýza bola

zrealizovaná vo voľne dostupnom softvéri GeoDa (Geoda, 2022). Mapa v podobe .shp súboru pre okresy Slovenskej republiky bola stiahnutá z webovej stránky Freemap Slovakia (2015).

Obr. 1 obsahuje box-mapy ilustrujúce výkonnosť žiakov v Testovaní 9 pred a po pandémii Covid-19 (MAT19 a MAT22) ako aj rozdiel medzi výsledkami v roku 2019 a 2022 (M19_M22) v jednotlivých okresoch Slovenska. Box-mapa umožňuje rýchlu a efektívnu identifikáciu odľahlých hodnôt a priestorových vzorcov – ide v podstate o rozšírenú verziu kvartilovej mapy na šesť kategórií, čo umožňuje samostatnú identifikáciu dolných a horných odľahlých hodnôt. Je zrejmé, že priemerná výkonnosť žiakov v jednotlivých okresoch v roku 2022 bola v porovnaní s rokom 2019 nižšia, najväčší pokles priemernej výkonnosti možno pozorovať v okrese Medzilaborce, najmenší v okrese Turčianske Teplice. Z hľadiska výkonnosti v oboch analyzovaných obdobiach sa najlepšie umiestnili žiaci z okresu Bratislava I, najhoršiu priemernú výkonnosť sme zaznamenali v oboch obdobiach v okrese Gelnica.

Obr.1: Box-mapy výkonnosti žiakov v Testovaní 9 z matematiky v roku 2019 (MAT19), 2022 (MAT22) a rozdielu vo výkonnosti pred a po pandémii Covid-19(M19_M22)

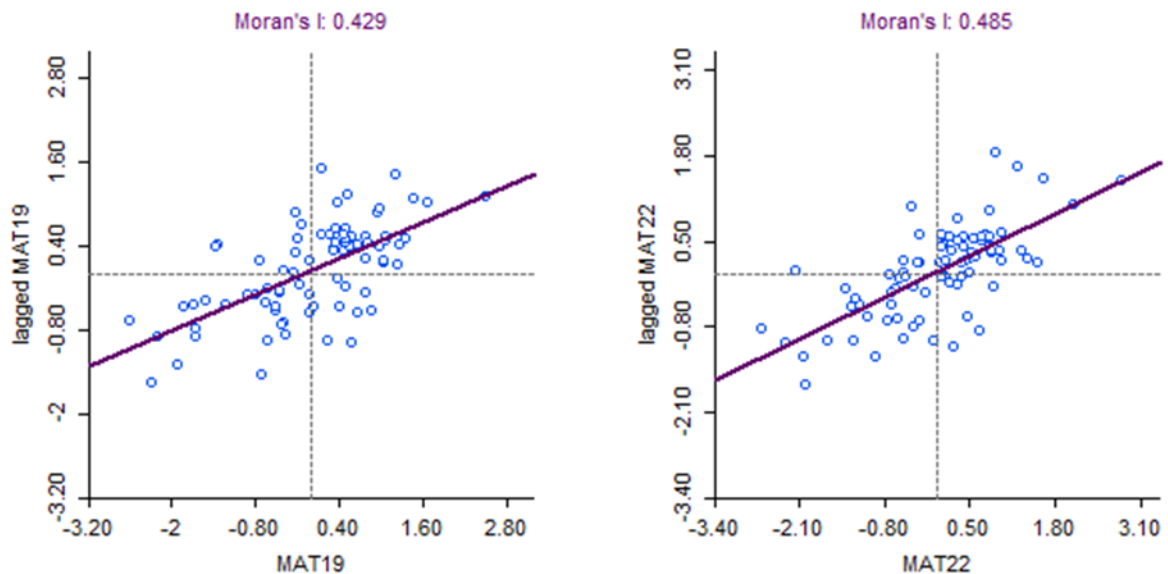


Zdroj: vlastné spracovanie v GeoDa

Na vizuálne posúdenie lokálnych priestorových asociácií možno využiť Moranov rozptylový diagram znázornený na Obr. 2 pre výsledky testovania v oboch uvažovaných obdobiach. Pri výpočte bola použitá matica váh typu kráľovná (viac o maticiach váh pozri napr. Furková, 2013). Hodnoty globálnej Moranovej I štatistiky 0,429, resp. 0,485 naznačujú existenciu globálnej pozitívnej priestorovej autokorelácie. Na báze hodnôt lokálnych Moranových I štatistík možno identifikovať okresy s pozitívnou priestorovou autokoreláciou typu vysoká-vysoká (pravý horný kvadrant), resp. nízka-nízka (ľavý dolný kvadrant), pričom zvyšné dva

kvadranty obsahujú okresy, ktoré majú záporný priestorový vzťah so svojimi susednými okresmi (ide teda o okresy s negatívnou priestorovou autokoreláciou).

Obr. 2: Moranove diagramy pre výsledky Testovania 9 z matematiky v roku 2019 (MAT19) a 2022 (MAT22)



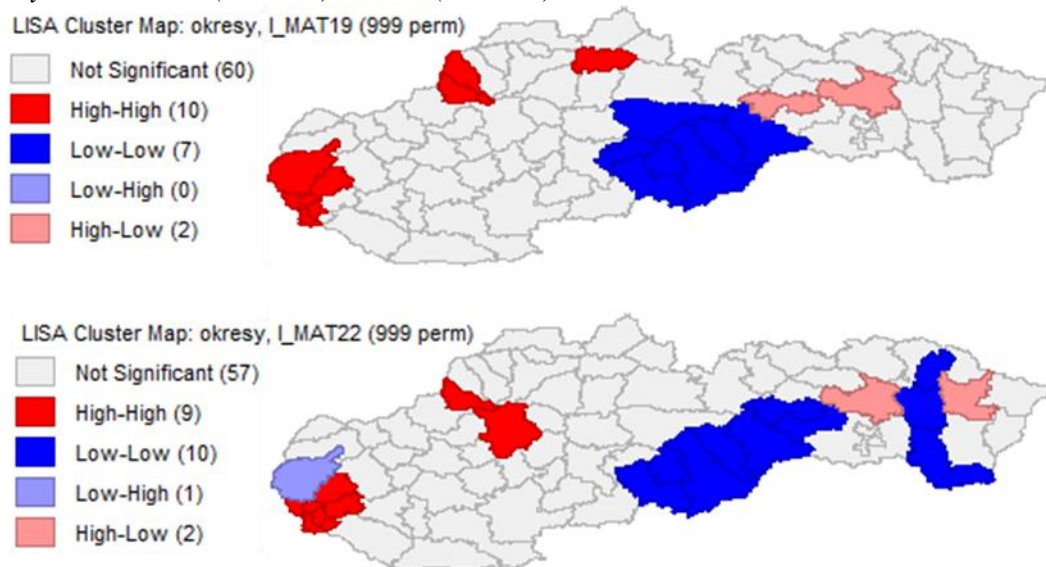
Zdroj: vlastné spracovanie v GeoDa

Obr.3 obsahuje zhlukové mapy skonštruované na základe lokálnej Moranovej I štatistiky pre výsledky Testovania 9 v jednotlivých rokoch, ktoré umožňujú identifikovať okresy, resp. ich zhluky so štatisticky významnou priestorovou autokoreláciou (hladina významnosti 0,05). Je zrejmé, že počet okresov so štatisticky významnou pozitívnou priestorovou autokoreláciou (vysoká-vysoká a nízka-nízka) bol značne vyšší ako počet okresov so štatisticky významnou negatívnou priestorovou autokoreláciou. Za zaujímavý možno považovať napr. vývoj výkonnosti v okresoch Malacky a Spišská Nová Ves. Kým okres Malacky bol v roku 2019 okresom s pozitívnou priestorovou autokoreláciou typu vysoká-vysoká, v roku 2022 bol identifikovaný ako okres so štatisticky významnou negatívnou priestorovou autokoreláciou typu nízka-vysoká. Znamená to, že kým v roku 2019 bola priemerná výkonnosť žiakov v tomto okrese porovnateľne vysoká ako v okolitých okresoch (uvažujeme maticu váh typu kráľovná), v roku 2022 bola už štatisticky významne nižšia ako v susedných okresoch. V prípade okresu Spišská Nová Ves bola situácia odlišná - v roku 2019 išlo o okres so štatisticky významnou negatívnou priestorovou autokoreláciou typu vysoká-nízka, čo naznačuje odlišnú úroveň v porovnaní s okolitými okresmi, v roku 2022 však možno hovoriť už o štatisticky významnej pozitívnej priestorovej autokorelácií typu nízka-nízka.

Podobnosť, resp. nepodobnosť jednotlivých okresov z pohľadu priemernej výkonnosti žiakov v Testovaní 9 z matematiky možno znázorniť na báze využitia hodnôt lokálnej Gearyho štatistiky – zhlukové mapy pre analyzované obdobia sú na Obr. 4. V prípade lokálnej Gearyho štatistiky teda rozlišujeme tri prípady pozitívnej lokálnej priestorovej autokorelácie, a to okrem typov vysoká-vysoká a nízka-nízka tiež tzv. typ iná v prípade, že hodnoty v analyzovanej dvojici okresov sú síce z hľadiska ich veľkosti podobné, jedna z nich je však nižšia ako stredná hodnota, druhá naopak vyššia ako stredná hodnota. V prípade negatívnej lokálnej priestorovej autokorelácie však vzhľadom na spôsob výpočtu tejto štatistiky (2) nie je možné špecifikovať, či ide o negatívnu autokoreláciu typu vysoká-nízka alebo nízka-vysoká.

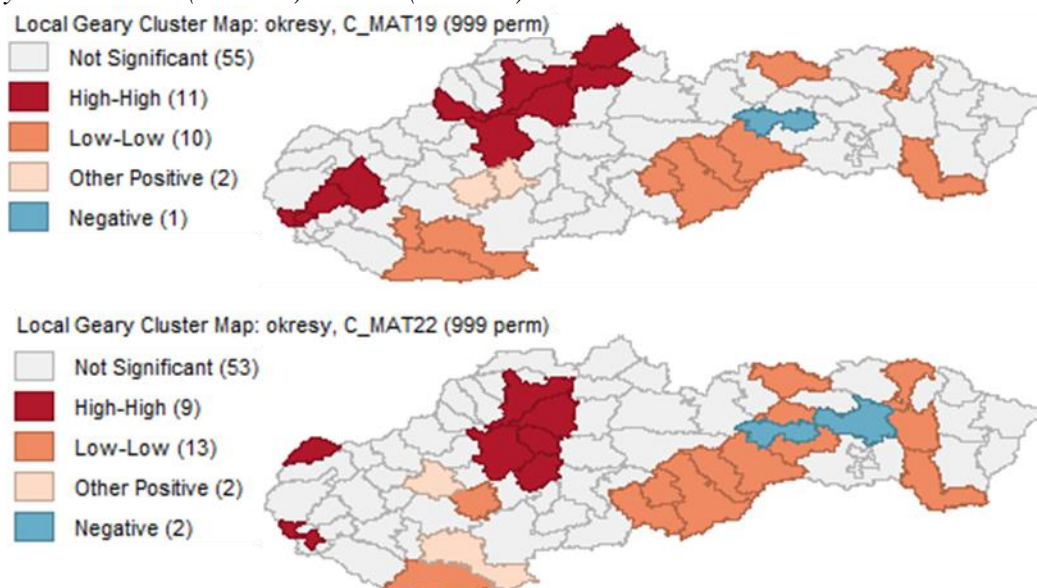
Na základe znázornených výsledkov je zrejmé, že kým ako okresy so štatisticky významnou pozitívnou priestorovou autokoreláciou typu vysoká-vysoká sú identifikované väčšinou „iné“ okresy ako v prípade lokálnej Moranovej I štatistiky, pri okresoch so štatisticky významnou pozitívnou priestorovou autokoreláciou typu nízka-nízka možno pozorovať väčšiu „zhodu“ s výsledkami dosiahnutými na báze Moranovej I štatistiky. Podobnosť výsledkov možno pozorovať tiež v prípade identifikovania okresov so štatisticky významnou negatívnou priestorovou autokoreláciou.

Obr.3: Zhlukové mapy na základe lokálnej Moranovej štatistiky pre výsledky Testovania 9 z matematiky v roku 2019 (MAT19) a 2022 (MAT22)



Zdroj: vlastné spracovanie v GeoDa

Obr.4: Zhlukové mapy na základe lokálnej Gearyho štatistiky pre výsledky Testovania 9 z matematiky v roku 2019 (MAT19) a 2022 (MAT22)



Zdroj: vlastné spracovanie v GeoDa

4 ZÁVER

Cieľom príspevku bola priestorová analýza priemernej výkonnosti žiakov v okresoch SR na báze výsledkov Testovania 9 z matematiky v období pred a po pandémie Covid-19, t.j. v rokoch 2019 a 2022. Okrem posúdenia priestorového rozloženia okresov z pohľadu výkonnosti žiakov na báze box-máp boli využité tiež zhlukové mapy skonštruované na báze lokálnej Moranovej I štatistiky a Gearyho c štatistiky. Prítomnosť štatisticky významnej pozitívnej globálnej priestorovej autokorelácie indikuje dôležitosť skúmať polohu analyzovaného okresu v geografickom priestore. Výsledky preukázali nižšiu úroveň výkonnosti žiakov v postpandemickom období v porovnaní s obdobím pred pandemiou a umožnili špecifikovať podobnosť, resp. rozdielnosť jednotlivých okresov z pohľadu analyzovanej výkonnosti žiakov. Je zrejmé, že na výkonnosť žiakov malo vplyv okrem pandémie Covid-19 tiež viacero iných faktorov, ako napr. sociálna situácia (veľkosť príjmu, resp. nezamestnanosť), ale tiež vzťah rodičov k vzdelaniu. Predmetom ďalšieho výskumu by preto mohla byť konštrukcia priestorových ekonometrických modelov so zahrnutím relevantných vysvetľujúcich premenných.

Použitá literatúra

1. ANSELIN, L. (1995). Local Indicators of Spatial Association – LISA. *Geographical Analysis*, 27(2), 93-115.
2. ANSELIN, L. (2019). A Local Indicator of Multivariate Spatial Association: Extending Geary's c . *Geographical Analysis*, 51(2), 133-150.
3. FISCHER, M.M. - WANG, J. (2011). *Spatial Data Analysis. Models, Methods and Techniques*. Heidelberg: Springer.
4. FREEMAP SLOVAKIA. (2015). *Vektorové dáta administratívnych území Slovenska*. [online] Dostupné na: <http://wiki.freemap.sk/HraniceAdministrativnychUzemi> [cit. 11-02-2015].
5. FURKOVÁ, A. (2013). Spatial dependence econometrics models. *Ekonomika a informatika*, 11(2), 41-52.
6. FURKOVÁ, A. (2019). Spatial Cluster Analysis of the Regional Poverty in the European Union. In *Proceedings of the International Conference on Contemporary Issues in Economics, Management and Business*. Hanoi: National Economics University Publishing House, 873-883.
7. GEODA. (2022). *Download GeoDa Software* [online] Dostupné na: <https://geodacenter.github.io/download.html> [cit. 07-06-2022].
8. NÁRODNÝ ÚSTAV CERTIFIKOVANÝCH MERANÍ VZDELÁVANIA. (2019). *Príloha 4.1: Výsledky žiakov v Testovaní 9 2019 podľa okresov* [online]. Dostupné na: https://www.nucem.sk/dl/4422/S_T9_2019_Priloaha_4.1.pdf [cit. 07-07-2019].
9. NÁRODNÝ ÚSTAV CERTIFIKOVANÝCH MERANÍ VZDELÁVANIA. (2022a). *Testovanie 9 2022* [online]. Dostupné na: https://www.nucem.sk/dl/5303/TS_T9_2022-NUCEM.pdf [cit. 29-06-2022].
10. NÁRODNÝ ÚSTAV CERTIFIKOVANÝCH MERANÍ VZDELÁVANIA. (2022b). *Príloha 1.1: Výsledky žiakov v Testovaní 9 2022 podľa okresov v teste z matematiky* [online]. Dostupné na: https://www.nucem.sk/dl/5301/Pr%C3%ADloaha_1.1_07062022.pdf [cit. 29-06-2022].

Kontaktné údaje

doc. Ing. Michaela Chocholatá, PhD.

Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky

Dolnozemska 1/b, 852 35 Bratislava
Tel: (421 2) 67 295 832
email: michaela.chocholata@euba.sk

SOFISTIKOVANÉ HEURISTIKY PRO APROXIMACI PARETOVY FRONTY NÁVRHŮ VEŘEJNÉHO OBSLUŽNÉHO SYSTÉMU

SOPHISTICATED HEURISTICS FOR APPROXIMATION OF PARETO FRONT OF PUBLIC SERVICE SYSTEM DESIGNS

Jaroslav Janáček, Marek Kvet

Abstrakt

Príspevek referuje o výsledcích výzkumu v oblasti použití metaheuristik a hyperheuristik pro získání dobré aproximace Paretoovy fronty návrhů veřejného obslužného systému s dvěma konfliktními kritérii.

Klíčové slová: Paretova fronta, Metaheuristika, Hyperheuristika, veřejné obslužné systémy

Abstract

The contribution reports on results of a research in the field of usage of metaheuristics and hyperheuristics for obtaining a good approximation of Pareto front of public service system designs with two conflicting criteria.

Keywords: Pareto front, Metaheuristic, Hyperheuristic, public service systems

1 ÚVOD

V tomto příspěvku se zabýváme veřejnými obslužnými systémy, v kterých je služba jednotlivým uživatelům poskytována z několika vhodně rozmístěných středisek [2, 3, 4, 5, 6, 9, 13, 19]. Systémové kritérium běžně používané pro hodnocení daného rozmístění obslužných středisek bývá průměrná doba dostupnosti poskytované služby, tj. doba, za kterou z nejbližšího střediska s volným obslužným vozidlem toto přesune na místo, kde je zapotřebí poskytnout službu. Výhradní naplňování systémového kritéria vytváří nefér situaci pro minority uživatelů hodně vzdálených od větších aglomerací. Proto je při navrhování veřejného obslužného systému často vznášen požadavek, aby počet uživatelů, resp. jejich požadavků, za jistou časovou hranicí T byl co nejmenší. Toto férové kritérium je v konfliktu s výše uvedeným systémovým kritériem, tj. čím je menší hodnota jednoho kritéria, tím větší bude hodnota druhého. Vzhledem k vzájemné neporovnatelnosti obou kritérií, není možno vytvořit jednu účelovou funkci jako kombinaci obou kritérií a potom řešit úlohu návrhu veřejného obslužného systému jako běžnou umísťovací úlohu. Protože výsledný návrh veřejného obslužného systému se systémovým a férovým kritériem je věcí politického rozhodnutí hledajícího politicky únosný kompromis mezi benefitem průměrného uživatele systému a disutilitou uživatelů patřících k postižené minoritě, může být výsledkem exaktního zpracování dané úlohy jen reprezentativní série návrhů systému, kde je lepší hodnota jednoho kritéria vyvažovaná horší hodnotou druhého kritéria. Ideálním výsledkem exaktního zpracování úlohy by byla takzvaná Paretova fronta. Paretova fronta je taková množina vzájemně se nedominujících řešení, že každé jiné přípustné řešení úlohy je dominované některým řešením obsaženým v Paretově frontě. Slovy, že jedno řešení dominuje druhé, rozumíme stav, kdy dominující řešení nemá ani jednu z hodnot obou kritérií horší než dominované řešení a navíc je alespoň v jednom z kritérií lepší [7, 8, 10].

I samotný exaktní výpočet optimálního návrhu veřejného obslužného systému s jediným kritériem je pro běžný rozsah obsluhovaného regionu početně značně náročný. Exaktní výpočet Paretovy fronty je za předpokladu konečnosti množiny všech přípustných řešení možný, ale je zaplacen mnohonásobným opakováním výše uvedeného exaktního výpočtu, kdy jedno z kritérií je optimalizováno za předpokladu, že hodnota druhého nepřekročí zadanou mez [8]. Časová náročnost získání Paretovy fronty návrhů veřejného obslužného systému iniciovala výzkumný projekty APVV-19-0441 a VEGA 1/0216/21 zaměřené na získání dostatečně přesné aproximace Paretovy fronty s vynaložením akceptovatelného výpočetního času. Doposud dosažené výsledky jsou shrnuté v následující části tohoto příspěvku.

2 FORMULACE ÚLOHY NÁVRHU VEŘEJNÉHO OBSLUŽNÉHO SYSTÉMU S NÁHODNÝMI POŽADAVKY NA OBSLUHU

Předpokládáme, že řešením úlohy je vybrání p umístění obslužných středisek z m možných umístění v obsluhovaném regionu, kde náhodné požadavky na obsluhu mohou vznikat v kterémkoliv z n umístění uživatelů systému. Uživatel j generuje požadavek na obsluhu nákladů s frekvencí b_j . Dobu potřebnou k poskytnutí služby uživateli j z možného umístění střediska i označíme t_{ij} . Přípustné řešení úlohy může být modelované m -rozměrným vektorem \mathbf{y} nula-jednotkových rozhodovacích proměnných y_i , $i = 1, \dots, n$, kde proměnná y_i nabyde hodnoty 1 má-li být v možném umístění i vybudováno středisko a hodnoty 0 v opačném případě. Přípustnost řešení \mathbf{y} je dána podmínkou, že právě p rozhodovacích proměnných nabyde hodnoty 1. V případě \mathbf{z} může být řešení modelováno seznamem $P = \{i: i \in \{1, \dots, n\}, y_i = 1\}$ Množina všech přípustných řešení \mathbf{Y} je specifikovaná definicí (1).

$$\mathbf{Y} = \left\{ \mathbf{y} : \mathbf{y} \in \{0,1\}^m, \sum_{i=1}^m y_i = p \right\} \quad (1)$$

Při formulaci systémové účelové funkce vyjadřující průměrnou dobu, za kterou k uživateli dorazí služba z nejbližšího volného střediska, musí být zohledněno, že se jedná o distribuovaný systém hromadné obsluhy s p linkami obsluhy, které v okamžiku vzniku požadavku u uživatele j nemusí být volné. Tuto charakteristiku systému hromadné obsluhy budeme modelovat pomocí pravděpodobností q_1, \dots, q_r , kde q_k je hodnota pravděpodobnosti, že k -té časově nejbližší středisko obsluhy bude to nejbližší volné. S využitím předpokladu, že pravděpodobnosti q_1, \dots, q_r jsou v celém regionu přibližně stejné, můžeme systémovou účelovou funkci definovat výrazem (2), kde symbol $\min_k \{a_1, \dots, a_m\}$ označuje k -tou nejmenší hodnotu z hodnot a_1, \dots, a_m .

$$f_1(\mathbf{y}) = \sum_{j=1}^n b_j \sum_{k=1}^r q_k \cdot \min_k \{t_{ij} : i=1, \dots, m, y_i=1\} \quad (2)$$

Férová účelová funkce byla v naší studii formulovaná jako požadavek, aby počet požadavků na službu za hranicí T od nejbližšího umístěného střediska byl co nejmenší. Zde bude férová účelová funkce definovaná výrazem (3).

$$f_2(\mathbf{y}) = \sum_{j=1}^n b_j \cdot \max\{0, \text{sign}(\min\{t_{ij} : i=1, \dots, m, y_i=1\} - T)\} \quad (3)$$

Paretova fronta PF pro minimalizační úlohu s dvěma konfliktními kritérii f_1 a f_2 je množina řešení specifikovaná zápisem (4), kde \mathbf{Y} je množina všech unikátních řešení, tj. řešení pro která platí buď $f_1(\mathbf{x}) \neq f_1(\mathbf{y})$ a nebo $f_2(\mathbf{x}) \neq f_2(\mathbf{y})$.

$$PF = \{\mathbf{y} \in \mathbf{Y} : (\forall \mathbf{x} \in \mathbf{Y})(f_1(\mathbf{y}) < f_1(\mathbf{x}) \text{ nebo } f_2(\mathbf{y}) < f_2(\mathbf{x}))\} \quad (4)$$

Protože Y i \underline{Y} jsou konečné množiny, uvažovaná Paretova fronta bude též konečná množina vzájemně nedominovaných řešení, a proto každé řešení z \underline{Y} bude dominované alespoň jedním prvkem z PF .

3 APROXIMACE PARETOVY FRONTY

Paretovu frontu budeme aproximovat množinou vzájemně nedominovaných řešení y^1, \dots, y^{noNDSS} uspořádaných vzestupně podle hodnot f_2 . Řešení y^1 budeme nazývat levým hraničním řešením a y^{noNDSS} pravým hraničním řešením v dané aproximaci. Pro aproximující množinu budeme používat označení $NDSS$ a pro počet prvků této množiny $noNDSS$. Otázka měření kvality dané aproximace byla řešena v [12, 15], kde jako míra kvality byla navržena plocha pod schodovitou klesající funkcí, jejíž hodnoty jsou postupně $f_1(y^1), \dots, f_1(y^{noNDSS-1})$, a kde plocha označená jako $NDSS_Area$ může být vyčíslena podle vztahu (5).

$$NDSS_Area = \sum_{k=1}^{noNDSS-1} (f_1(y^k) - f_1(y^{noNDSS})) (f_2(y^{k+1}) - f_2(y^k)) \quad (5)$$

Aby hodnota $NDSS_Area$ byla konzistentní míra kvality aproximace Paretovy fronty, je třeba, aby hodnota $NDSS_Area$ byla větší anebo rovna hodnotě vypočítané podle (5) pro Paretovu frontu. V [12, 15] bylo ukázáno, že postačující podmínkou je aby $f_2(y^1)$ bylo řešením úlohy (6) a $f_1(y^{noNDSS})$ byla hodnotou optimálního řešení úlohy (7).

$$\min\{f_2(y) : y \in Y\} \quad (6)$$

$$\min\{f_1(y) : y \in Y\} \quad (7)$$

Na základě výsledků doporučujeme u iterativních algoritmů inicializovat počáteční $NDSS$ exaktními hraničními řešeními z Paretovy fronty místo heuristicky získanými řešeními úlohy (6) a (7).

Pokud je $NDSS$ udržovaná ve výše uvedeném tvaru a inicializovaná hraničními prvky Paretovy fronty, tak následující jednorůchodový algoritmus je pro libovolné přípustné řešení $x \in Y$ schopen rozhodnout, zda řešení je dominováno některým z prvků současného $NDSS$ a nebo jestli je to řešení, které zlepší (zmenší) $NDSS_Area$. Algoritmus aktualizace $NDSS$ pracuje následujícím způsobem:

Nejdříve je určeno maximální $k < noNDSS$ pro které platí $f_2(x) \geq f_2(y^k)$. Pokud $f_2(x) \geq f_2(y^{noNDSS})$, potom je x dominováno řešením y^{noNDSS} a je vyloučeno z aktualizace.

V opačném případě pokračuje algoritmus zjištěním zda $f_1(x) \geq f_1(y^k)$. V tomto případě je x dominováno řešením y^k a je rovněž vyloučeno z dalšího zpracování. Pokud $f_1(x) < f_1(y^k)$, tak je řešení x vsunuto jako $k+1$ vý prvek do $NDSS$. Nakonec jsou probírána řešení $y^{k+1}, \dots, y^{noNDSS}$ a pokud $f_1(x) \geq f_1(y^i)$, tak je řešení y^i vyloučeno z $NDSS$ a počet $noNDSS$ je aktualizován.

Uvedený algoritmus umožňuje využít jakoukoliv posloupnost přípustných řešení úlohy k postupnému zlepšování Paretovy fronty.

4 ZÁKLADNÍ OPERACE NA MNOŽINĚ PŘÍPUSTNÝCH ŘEŠENÍ Y

Všechny dále zkoumané přístupy k získání dobré aproximace Paretovy fronty návrhů veřejných obslužných systémů jsou založeny na prozkoumávání okolí daného řešení y . Okolí současného řešení y na množině Y jako topologický pojem je určeno množinou povolených operací, kterými je možno dané řešení měnit a také podmínkami, za kterých je možno povolenou operaci aplikovat. V prezentovaném výzkumu využíváme jako jedinou povolenou

operaci tzv. 1-1 výměnu umístění i střediska, které je součástí současného řešení za umístění j střediska, které součástí řešení není.

Pro výpočet hodnot účelových funkcí f_1 a f_2 změněného řešení lze použít rychlý algoritmus o složitosti $O(rn)$, čímž se prozkoumání okolí daného řešení výrazně zkrátí. Podmínky, redukující velikost zkoumaného okolí daného řešení mohou být stanoveny jednak na základě omezení hodnoty kombinované účelové funkce v tvaru $a_1.f_1(\mathbf{x}) + a_2.f_2(\mathbf{x})$ a nevyužitých umístění středisek na základě redukce množiny možných nepoužitých umístění středisek $C = \{i = 1, \dots, m: y_i = 0\}$, které je možné použít při výměně.

Uvedená operace výměny spolu s případnou redukcí příslušného okolí byla použita pro návrh dvou základních metod prohledávání. První z nich je perturbační algoritmus startující z výchozího řešení, kterým je inicializované tzv. současné řešení. Okolí současného řešení je prohledáváno spolu s vyhodnocováním kombinované účelové funkce $w_1.f_1(\mathbf{x}) + w_2.f_2(\mathbf{x})$ a podle aplikované strategie, je-li nalezeno řešení s lepší hodnotou kombinované funkce než hodnota současného řešení, je proveden přechod k tomuto řešení, které se stane novým současným řešením. Tento postup je opakován, dokud prozkoumávané okolí poskytuje možnost přechodu k řešení s lepší hodnotou kombinované funkce. Druhá metoda nazývaná „path relinking“ [1, 16] postupuje po nejkratší cestě mezi dvěma přípustnými řešeními po povrchu m - rozměrné jednotkové krychle. Jeden posun v daném směru ze současného řešení je proveden vybráním výměny minimalizující hodnotu kombinované funkce.

Řešení vyhodnocovaná v jedné i druhé metodě představují posloupnost kandidátů pro aktualizaci *NDSS*.

5 SCHÉMATA ZLEPŠOVÁNÍ *NDSS*

V rámci prezentovaného výzkumu bylo vyvinuto a testováno několik schémat aplikací metod z předchozí sekce pro zlepšování *NDSS*.

První schéma [14] spočívalo v použití prostého směrového prohledávání, kde na náhodně vygenerované řešení byl použit výměnný algoritmus minimalizující lokální účelovou funkci $\lambda.f_1(\mathbf{x}) + (1-\lambda).f_2(\mathbf{x})$, kde s malým krokem byla měněna hodnota parametru λ .

Dalším schématem [17] bylo použití adaptivního algoritmu, který na základě informace o předchozích úspěších a neúspěších prohledávání s jistou pravděpodobností měnil parametr λ a i výchozí řešení směrového prohledávání.

Nejúspěšnějším schématem se stalo schéma postupného zjemňování [11, 16, 18]. Postup podle tohoto schématu začíná s malou množinou *NDSS* obsahující pouze dva prvky a to hraniční řešení z Paretovy fronty. Dále jsou pro $k = 1, \dots, noNDSS-1$ probírány a zpracovávány dvojice $\mathbf{y}^k, \mathbf{y}^{k+1}$ za neustálé aktualizace *NDSS*. Pokud po zpracování dvojice $\mathbf{y}^k, \mathbf{y}^{k+1}$ zůstane \mathbf{y}^k nezměněné, posune se zpracovávaná dvojice o jedno umístění v současné *NDSS* k vyššímu indexu. Základní cyklus tohoto schématu skončí zpracováním dvojice $\mathbf{y}^{noNDSS-1}, \mathbf{y}^{noNDSS}$. Základní cyklus je opakován pro získanou *NDSS* od dvojice $\mathbf{y}^1, \mathbf{y}^2$ dokud není vyčerpán výpočetní čas určený pro běh algoritmu. Vlastní zpracování dvojice $\mathbf{y}^k, \mathbf{y}^{k+1}$ může být realizováno více způsoby. Prozatím byly testované postupy směrového prohledávání v několika směrech a možným omezením přechodu k některým sousedním řešením. Kromě toho byly testované postupy, které obsahovaly prvky metody simulated annealing [18]. Také bylo testované použití učící se hyperheuristiky, která volila vhodnou metodu směrového prohledávání na základě informace získané ze zpracování předchozích dvojic $\mathbf{y}^k, \mathbf{y}^{k+1}$.

6 VÝSLEDKY VÝPOČETNÍCH EXPERIMENTŮ

Oba dva zmíněné postupy byly implementovány v jazyku java a testovány na osobních počítačích s 11th Gen Intel® Core™ i7 11700KF procesor s parametry: 3,6 GHz a 16 GB RAM. Jako testovací úlohy byly použity benchmarky odvozené z existujících zdravotních záchranných systémů jednotlivých samosprávných krajů SR. Tyto benchmarky jsou označeny dále uvedenými jmény sídel samosprávy a zkratky uvedené v závorkách jsou použity k označení výsledků vztahujících se k jednotlivým krajům: Bratislava (BA), Banská Bystrica (BB), Košice (KE), Nitra (NR), Prešov (PO), Trenčín (TN), Trnava (TT) and Žilina (ZA). Exaktní Paretovy fronty pro jednotlivé kraje byly převzaty z [7, 8] a příslušná plocha PF_Area a počet členů $noPF$ Paretovy fronty je spolu s parametry p a m jednotlivých úloh uvedena v tabulce 1.

Tabulka 1. Popis benchmarků

Region	m	p	$noPF$	PF_Area
BA	87	14	34	569039
BB	515	36	229	1002681
KE	460	32	262	1295594
NR	350	27	106	736846
PO	664	32	271	956103
TN	276	21	98	829155
TT	249	18	64	814351
ZA	315	29	97	407293

Výsledky numerických experimentů pro jednotlivé přístupy (Semi_SA a Hyperheuristika) jsou uvedeny v tabulkách 2 a 3 a pro větší přehlednost místo absolutní velikosti plochy výsledné aproximace jsou uvedeny pouze relativní odchylky (gap) od plochy exaktní Paretovy fronty.

$$gap = 100 * \frac{|NDSS_Area - PF_Area|}{PF_Area} \quad (8)$$

Tabulka 2. Výsledky metody Semi_SA pro různé parametry zahřívání podle schématu $T = \alpha T$

Base α	BA	BB	KE	NR	PO	TN	TT	ZA
1.0	0.2	0.4	1.2	0.4	1.4	0.7	0.1	0.0
1.5	2.2	0.4	1.2	0.4	0.6	0.7	0.6	0.1
1.7	2.2	0.5	1.1	0.7	0.8	0.7	0.6	0.1
1.8	2.1	0.5	1.4	0.7	1.8	0.7	0.7	0.2
1.9	2.7	0.5	1.2	0.8	1.0	0.7	0.6	0.1
2.0	3.1	0.4	1.1	1.1	0.3	0.7	0.5	0.1

Tabulka 3. Výsledky metody Hyperheuristiky pro různé počáteční hodnoty Score

Score	BA	BB	KE	NR	PO	TN	TT	ZA
1	1.47	0.95	3.04	0.99	1.20	4.37	0.06	1.25
10	1.47	1.14	3.14	0.99	0.85	4.37	0.00	1.25
100	1.47	1.08	3.23	0.99	1.08	4.37	0.00	1.25

7 ZÁVĚR

Výpočetní experimenty prokázaly velkou úspěšnost schématu postupného zjemňování kombinovaného jak s hyperheuristikou nastavující jednotlivé algoritmy směrového prohledávání, tak i s metaheuristikou typu simulated annealing. Při vhodném nastavení parametrů uvedených metod je možné dosáhnout takové aproximace Paretovy fronty, že se v kritériu kvality až na výjimky liší nejvýše o 2% od exaktní Paretovy fronty.

Další výzkum v této oblasti bude zaměřen na metody ladění parametrů hyperheuristik a metaheuristik a případně na vytváření jiných strategií prohledávání.

Poděkování

Tento příspěvek byl podporován výzkumnými granty VEGA 1/0216/21 “Design of emergency systems with conflicting criteria using artificial intelligence tools”, VEGA 1/0077/22 „Innovative prediction methods for optimization of public service systems“ a 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“. Rovněž jsme získali podporu agentury „Slovak Research and Development Agency“ v rámci projektu č. APVV-19-0441.

Použitá literatura

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. Avella, P., Sassano, A. and Vasil'ev, I. (2007). Computational study of large scale p-median problems. *Mathematical Programming*, 109, pp. 89-114.
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. (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.
10. Janáček, J. and Kvet, M. (2021b). Emergency Medical System under Conflicting Criteria. In: SOR 2021 Proceedings, pp. 629-635.
11. Janáček, J. and Kvet, M. (2021b). Repeated Refinement Approach to Bi-objective p-Location Problems. In: INES 2022: 26th IEEE International Conference on Intelligent Engineering Systems 2022, pp. 41-45.

12. Janáček, J. and Kvet, M. (2022). Quality Measure of Pareto Front Approximation. In: Proceedings of the International Scientific Conference QUANTITATIVE METHODS IN ECONOMICS: Multiple Criteria Decision Making XXI, Bratislava: Letra Edu, 2022, pp. 95-103.
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. Kvet, M. and Janáček, J. (2022a). Evaluation of dynamics of heuristics used for pareto front approximation. In: Proceedings of the International Scientific Conference QUANTITATIVE METHODS IN ECONOMICS: Multiple Criteria Decision Making XXI, Bratislava: Letra Edu, 2022, pp. 132-140.
16. Kvet, M. and Janáček, J. (2022b). Directed Search for Pareto Front Approximation with Path-relinking Method. In: Mathematical Methods in Economics 2022, Jihlava. pp. 212-217.
17. Kvet, M. and Janáček, J. (2022c). Adapted Path-relinking Based Search for Nondominated Set of Emergency Medical System Designs. In: International Carpathian Control Conference 2022, IEEE, ISBN 978-1-6654-6636-3, pp. 266-270.
18. Kvet, M. and Janáček, J. (2022d). Non-dominated Set of Public Service System Designs using Simulated Annealing Approach. In: INES 2022: 26th IEEE International Conference on Intelligent Engineering Systems 2022, pp. 59-63.
19. 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.

Kontaktné údaje

Prof. RNDr. Jaroslav Janáček, CSc.

Žilinská univerzita v Žiline, Fakulta riadenia a informatiky

Univerzitná 1, 010 26 Žilina

Tel: (421 41) 513 4204

email: jaroslav.janacek@fri.uniza.sk

ZLOŽKY MEDZINÁRODNEJ INVESTIČNEJ POZÍCIE A ICH VÝVOJ VO VZŤAHU K HDP V OBDOBÍ ZMIEN V EKONOMIKE

COMPONENTS OF THE INTERNATIONAL INVESTMENT POSITION AND THEIR DEVELOPMENT IN RELATION TO GDP DURING A PERIOD OF CHANGES IN THE ECONOMY

Miroslava Jánošová

Abstrakt

Príspevok skúma vzťah hrubého domáceho produktu (HDP) ako ukazovateľa vyjadrujúceho ekonomický rast krajiny a medzinárodnej investičnej pozície (MIP) a jej jednotlivých zložiek. Dosiahnuté výsledky potvrdzujú existenciu prepojenia medzi medzinárodnou investičnou pozíciou a hrubým domácim produktom, čo sa potvrdilo aj pri samotných zložkách medzinárodnej investičnej pozície. Tieto výsledky sú porovnateľné aj s ekonomickou teóriou a rovnako aj s inými medzinárodnými štúdiami, ktoré uvádzame v príspevku.

***Kľúčové slová:** medzinárodná investičná pozícia, priame zahraničné investície, hrubý domáci produkt*

Abstract

The paper examines the relationship between gross domestic product (GDP), as an indicator of a country's economic growth, and the international investment position and its individual components. The results obtained confirm the existence of a relationship between the international investment position and GDP, which was also confirmed for its components. These results are comparable with both economic theory and other international studies.

***Keywords:** International Investment Position, Foreign Direct Investment, Gross Domestic Product*

ÚVOD

V tomto príspevku sa venujeme ukazovateľu medzinárodnej investičnej pozície, ktorý je súvahou zahraničných aktív a pasív krajiny a v mnohých prípadoch sa používa na meranie stupňa finančnej otvorenosti krajiny. Rovnako je aj užitočným nástrojom na monitorovanie finančnej stability. V úvodnej kapitole ponúkame prehľad prác autorov, ktorí sa snažia zhodnotiť zmeny v tokoch vybraných foriem kapitálu, a prehľadu prác, v ktorých je potvrdená vzájomná prepojenosť medzi tokmi finančného kapitálu a ekonomického rastu. V druhej kapitole poskytneme komparatívnu analýzu vývoja jednotlivých zložiek ukazovateľa MIP a ich vzťahu k hrubému domácemu produktu na Slovensku.

1 PREHĽAD VEDECKÝCH PRÁC

Nasledujúca časť príspevku je prehľadom vedeckých prác, ktoré sa aktuálne venujú problematike medzinárodného pohybu kapitálu, ku ktorému dochádza v dôsledku

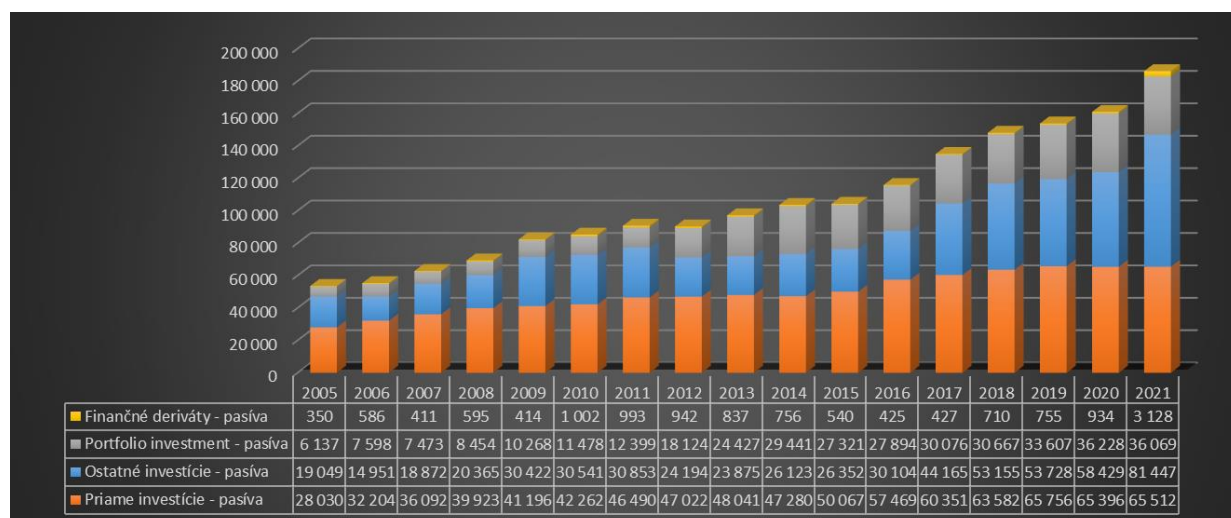
dynamických zmien v globálnej ekonomike. Príspevok autorky Petrikova (2022), analyzuje súčasnú situáciu vo svetovej ekonomike v roku 2020, ktorá je spôsobená energetickou krízou v dôsledku zlyhania dohôd OPEC, politických konfliktov, ale aj v dôsledku COVID-19. Autorka sa zamerala na štatistickú analýzu dopadov pandémie COVID-19 na platobnú bilanciu a MIP Ruska, ktorá v súhrnnej podobe odzrkadľuje výsledky zahraničnej ekonomickej aktivity krajiny a porovnáva ich s globálnou krízou v roku 2008 a zavedením medzinárodných finančných sankcií po roku 2014. Kolosok a kol. (2018) svoj výskum zamerali na vplyv verejnej politiky na vzťah medzi HDP a MIP a jej vybranými zložkami a to konkrétne priamymi portfóliovými investíciami a inými investíciami Ukrajiny a to v kontexte snahy Ukrajiny o európsku integráciu. Výsledky, ktoré dosiahli v danom výskume naznačujú, že šokové zmeny v štruktúre MIP boli spôsobené predovšetkým politickými a ekonomickými faktormi v krajine a rovnako boli aj výsledkom nesystematickej verejnej politiky Ukrajiny. Korelačná analýza, ktorú použili na vyhodnocovanie poukázala na významný vzťah medzi HDP a investíciami a to najmä portfóliovými a priamymi. Autori Andries, Chiper, Ongena a Sprincean (2022) poukazujú na to, že vonkajšia nerovnováha zohrávala kľúčovú úlohu v období pred globálnou finančnou krízou a bola príčinou následných otrasov. Zdôrazňujú, že nerovnováha bežného účtu sa po kríze zmenšila avšak nerovnováha medzinárodnej investície pretrváva. V príspevku rovnako skúmajú aj dôsledky čistých zahraničných pozícií na systémové riziko. Tento výskum realizujú na vzorke 450 bánk, ktoré sa nachádzajú v 46 vyspelých, pričom uvádzajú, že tieto banky môžu znížiť svoje vystavenie sa systémovému riziku tým, že si udržia svoje veriteľské pozície voči zvyšku sveta. Autori Slepecký a kol. (2022) skúmajú tradičných a ESG akciových indexov na čistú medzinárodnú investičnú pozíciu krajiny. Výsledky ich deskriptívnej štatistiky ukazujú, že indexy ESG sú volatilnejšie ako tradičné a využitím analýzy ANOVA štatisticky nevýznamné. Ich výsledky zároveň ukazujú, že tradičné a ESG akciové indexy nemajú významný vplyv na čistú medzinárodnú investičnú pozíciu a poskytujú odporúčania na zlepšenie rámca zodpovedného investovania. Zaujímavým príspevkom je aj práca autorky Knap (2019), ktorá dáva do vzťahu vplyv starnutia populácie a medzinárodnú investičnú pozíciu Poľska. Výsledky jej štúdie ukazujú, že očakávaný intenzívny postup starnutia populácie prispeje do roku 2065 k zvýšenému dopytu po zahraničnom kapitáli, k čoraz zápornejšej medzinárodnej investičnej pozícii, ale aj k nepriaznivým zmenám v samotnej štruktúre zahraničných pasív. Z jej výsledkov je zaujímavé aj poznanie, že prílev zahraničného kapitálu, ktorý bude spôsobený starnutím populácie sa nemusí premietnuť do očakávaného zvýšenia tempa hospodárskeho rastu, ale práve naopak zvyšujúca negatívna medzinárodná investičná pozícia môže byť prekážkou rozvoja ekonomiky. Autorka Knap (2020) použila aj deskriptívnu štatistickú analýzu na posúdenie špecifického ukazovateľa medzinárodnej investičnej pozície na identifikovanie globálnych platobných nerovnováh z pohľadu tokov a zásob v rokoch 2000-2017. Z jej analýzy vyplýva, že po vypuknutí globálnej finančnej krízy sa rovnováha globálnych tokov znížila, zatiaľ čo nerovnováha globálnych zásob sa prehĺbila. Maciejczyk-Bujnowicz (2016) vo svojej práci poskytuje zhodnotenie zmien v tokoch vo vybraných formách kapitálu naprieč členskými krajinami Európskej únie. Tieto krajiny rozdelila do dvoch skupín a to členské krajiny v eurozóne a členské krajiny mimo eurozóny. Výsledky jej analýzy poukazujú na výrazne rozdiely medzi vyspelými a rozvojovými krajinami. Okrem tejto analýzy zhodnotila v príspevku aj medzinárodnú investičnú pozíciu Poľska na pozadí vybraných krajín strednej a východnej Európy a členov eurozóny. Okrem vyššie spomínaných príspevkov je aktuálne publikovaných viacero prác, ktoré sa zameriavajú na zhodnotenie medzinárodnej investičnej pozície v kontexte aktuálnych udalostí pre viaceré ekonomiky sveta ako napríklad práca autorov Yosifov (2020), Zhang (2019) alebo Durkalić a Čurčić (2019). Pri štúdiu vedeckých prác sme sa zamerali aj na práce, ktoré sa zameriavajú na potvrdenie vzájomného vzťahu ekonomického rastu vyjadreného prostredníctvom makroekonomického ukazovateľa HDP

a MIP prípadne vybraných zložiek tohto ukazovateľa a to najmä PZI. Prepojenie medzi PZI a ekonomickým rastom bolo popísané a určené už predstaviteľom neoklasickej teórie Solowom (1957). Pozitívny vzťah medzi PZI a ekonomickým rastom predpokladal v práci aj Romer (1986). Autor Borensztein (1998) uvádza, že interakcia medzi PZI a ekonomickým rastom je závislá aj od charakteristík hostiteľskej krajiny a jej úrovne technologického pokroku a ľudského kapitálu. Okrem týchto autorov aj iné empirické štúdie ako napr. Chan (2000), Pelinescu a Radulescu (2009), Pece a kol. (2015).

2 KOMPARATÍVNA ANALÝZA VÝVOJA ZLOŽIEK MEDZINÁRODNEJ INVESTIČNEJ POZÍCIE

Ukazovateľ MIP je charakterizovaný ako súhrn zahraničných aktív a pasív. Tento ukazovateľ je využívaný aj na posúdenie finančnej otvorenosti krajiny, ale aj na sledovanie finančnej stability krajiny. Vykazovanie MIP a jej zložiek podlieha 6 vydaniu Manuálu platobnej bilancie a investičnej pozície voči zahraničiu, ktoré je aktualizáciou 5. vydania. Zavedenie tohto nového manuálu v krajinách Európskej únie bolo v roku 2014, a je koordinované Európskou centrálnou bankou a Eurostatom. MIP môžeme rozdeliť na aktíva a pasíva, ktoré sú tvorené priamymi investíciami v zahraničí (majetkový kapitál a reinvestovaný zisk, ostatný kapitál), portfóliovými investíciami (majetkové cenné papiere, dlžobné cenné papiere), ostatnými investíciami (obchodné úvery, pôžičky, hotovosť a vklady, ostatné aktíva) a rezervnými aktívami. Národná banka Slovenska vykazuje ukazovateľ MIP na mesačnej, štvrtročnej a ročnej báze. V nasledujúcej časti príspevku si popíšeme vývoj pasív MIP zobrazené na Obrázku 1 a jej jednotlivých zložiek v priebehu sledovaného obdobia od roku 2005 do roku 2021. Pričom vidíme, že počas celého sledovaného obdobia MIP rástla, avšak niektoré jej zložky v posledných uplynulých rokoch zaznamenali pokles. Zo zobrazených dát môžeme vidieť, že išlo o prílev priamych zahraničných investícií a portfóliových investícií, ktoré tvoria najväčšiu časť medzinárodnej investičnej pozície. K tomuto poklesu došlo presne v rokoch 2019 – 2021 čo môžeme pripísať globálnym ekonomickým zmenám vo vzťahu k pandémie COVID-19 aj vo vzťahu k vojne na Ukrajine. Zaujímavým pozorovaním je, že veľmi výrazný nárast prílevu je zaznamenaný pri Finančných derivátoch, ktoré predstavujú finančné nástroje viazané na konkrétny finančný nástroj, ukazovateľ alebo komoditu, prostredníctvom ktorých sa dá na finančných trhoch obchodovať s konkrétnymi finančnými rizikami.

Obrázok 1 Vývoj pasív MIP a jej zložiek v rokoch 2005 – 2021 v mil. EUR



Zdroj: Spracované na základe dát NBS

Tieto obchody s finančnými sa považujú za samostatné transakcie a nie za súčasť hodnoty podkladových obchodov, na ktoré sa môžu viazať. Ide o transakcie a pozície v opciách, futures, swapoch, termínovaných devízových obchodoch, úverových derivátoch a zabudovaných derivátoch. Pod finančné deriváty by mali byť zaradené aj platby dodatočných záloh typu futures. Ako môžeme vidieť prílev týchto finančných derivátov na Slovensko bol v roku 2020 na úrovni 934 mil. Eur a v roku 2021 bol tento prílev na úrovni 3 128 mil. Eur čo predstavuje 234% nárast.

3 ZÁVISLOŤ MEDZINÁRODNEJ INVESTIČNEJ POZÍCIE A HOSPODÁRSKEHO RASTU

Tento príspevok využíva makroekonomické údaje Slovenska za obdobie rokov 2005 – 2021 pričom ide o obdobie, za ktoré je vykazovaná štatistika ukazovateľa Medzinárodnej investičnej pozície podľa metodiky BPM6. Údaje sú získané z databázy Národnej banky Slovenska. Hlavnými premennými je HDP a MIP-pasíva, pričom MIP-pasíva sme si rozložili aj na jej jednotlivé zložky a teda Priame zahraničné investície-pasíva, Portfólio investment – pasíva, Finančné deriváty – pasíva a Ostatné investície – pasíva.

Ako nám naznačuje ekonomická teória Medzinárodnej investičnej pozície a teda jej jednotlivých zložiek majú tendenciu sa vyvíjať rovnakým smerom ako hrubý domáci produkt. To znamená, že ak PZI rastú tak HDP má tendenciu rásť tiež.

Z výsledkov korelácie uvedených v Tabuľke 1 vidíme, že sa nám potvrdil náš predpoklad o vývoji jednotlivých ukazovateľov rovnakým smerom, pričom môžeme vidieť, že existujú silné korelácie medzi hrubým domácim produktom a pasívami medzinárodnej investičnej pozície, pričom tieto silné korelácie sú hlavne medzi hrubým domácim produktom a priamymi zahraničnými investíciami, portfólio investment a ostatnými investíciami. Korelácia medzi HDP a finančnými derivátmi na základe korelačnej analýzy vykazuje stredne silnú závislosť.

Tabuľka 1 Korelácia medzi HDP a pasívami MIP a jej jednotlivými zložkami

	HDP - hrubý domáci produkt	MIP - pasíva	Priame investície pasíva	Portfolio investment - pasíva	Finančné deriváty - pasíva	Ostatné investície pasíva
HDP - hrubý domáci produkt	1.000000	0.941820	0.984056	0.927010	0.421411	0.808295
MIP - pasíva	0.941820	1.000000	0.966332	0.923355	0.582624	0.944751
Priame investície - pasíva	0.984056	0.966332	1.000000	0.929288	0.420768	0.849987
Portfolio investment - pasíva	0.927010	0.923355	0.929288	1.000000	0.399101	0.759833
Finančné deriváty - pasíva	0.421411	0.582624	0.420768	0.399101	1.000000	0.710194
Ostatné investície - pasíva	0.808295	0.944751	0.849987	0.759833	0.710194	1.000000

Zdroj: Spracované za základe dát z NBS

4 ZÁVER

Súčasný výskum prispieva teoretickými konceptami, ktoré sú vhodné ako podnety k ďalším úvahám. Zdôrazňuje úlohu zahraničného finančného kapitálu či už v podobe priamych zahraničných investícií alebo iných zložiek medzinárodnej investičnej pozície a to vo vzťahu ku globálnym ekonomickým zmenám a k hrubému domácemu produktu. Tak ako viaceré

štúdie tento vzťah medzi HDP a MIP prípadne PZI potvrdzujú Pelinescu a Radulescu (2009); Borensztein (1998), Pece a kol. (2015) tak aj na analýze dát pre Slovensko sa nám ten vzťah potvrdil korelačnou analýzou.

Tento príspevok bol vypracovaný v rámci riešenia projektu: VEGA 1/0628/20 „Medzinárodná investičná pozícia a jej vzťah k ekonomickému rastu - možný ukazovateľ vonkajších nerovnováh krajiny“ a VEGA 1/0668/22 „Migrácia a strieborná ekonomika: Ekonomické, politické a právne dôsledky riešenia problému starnutia a migrácie populácie v krajinách EÚ“

Použitá literatúra

1. ANDRIES, A. M., CHIPER, A. M., ONGENA, S., & SPRINCEAN, N. (2022). External Wealth of Nations and Systemic Risk. Swiss Finance Institute Research Paper, (22-74).
2. BORENSZTEIN, E., D. G. JOSE (1998). How Does Foreign Direct Investment Affect Economic Growth? *Journal of International Economics* 45, p.115-135
3. CEZAR, R., & SILVESTRINI, M. (2021). Impact of the ECB quantitative easing on the international investment position. *International Economics*, 165, 241-263.
4. DURKALIC, D., a CURCIC, M. (2019). Mutual influence of the international investment position and the net government position with the banking sector of the Republic of Serbia. *Ekonomika, Journal for Economic Theory and Practice and Social Issues*, 65(1350-2019-2799), 95-106
5. KNAP, R. (2019). Global imbalances from the flow and stock perspective. *Journal of Management and Financial Sciences*, (38), 19-31.
6. KNAP, R. (2020). Population Ageing and Poland's International Investment Position. *Argumenta Oeconomica Cracoviensia*, (2 (21)), 129-147.
7. KOLOSOK, S., DEMENTOV, V., KOROL, S., a PANCHENKO, O. (2018). Public policy and international investment position in European integration of Ukraine. *Journal of Applied Economic Sciences*, 13(8), 2375-2384.
8. MACIEJCZYK-BUJNOWICZ, I. (2016). Capital Flows in European Union on the Basis of International Investment Position–Selected Aspects. *Copernican Journal of Finance & Accounting*, 5(1), 152-172.
9. PECE, A. M., SIMONA, O. E., AND SALISTEANU, F. (2015), "Innovation and economic growth: An empirical analysis for CEE countries", *Procedia Economics and Finance*, Vol. 26, pp. 461-467
10. PELINESCU, E., AND RADULESCU, M. (2009), "The impact of foreign direct investment on the economic growth and countries' export potential", *Romanian Journal of Economic Forecasting*, Vol. 4, No. 1, pp. 153-169.
11. PETRIKOVA, E. (2022). Assessment of the Impact of the Covid-19 Pandemic on the Balance of Payments and the International Investment Position of Russia. *Voprosy statistiki*. 29. 33-42. 10.34023/2313-6383-2022-29-2-33-42.
12. ROMER, PAUL M. (1986). Increasing Returns and Long Run Growth. *Journal of Political Economy*, 94, 1002–37
13. SOLOW, R. (1957). Technical change and the aggregate production function. *Review of economics and statistics*, vol 39, No 3, pp 312-320.
14. SLEPECKÝ, J., VORONTSOVA, A. S., PLASTUN, O. L., MAKARENKO, I. O., a ZHYHLEI, I. (2022). Net investment position and the stock market: The case of traditional and ESG indices.

15. ZHANG, S. (2019). Human Rights and International Investment Agreements: How to Bridge the Gap? *The Chinese Journal of Comparative Law*, 7(3), 457-483.
16. YOSIFOV, T. (2020). The International Investment Position of Bulgarian Economy in the Period from 2010 to 2019.

Author's address

Ing. Miroslava Jánošová, PhD.
Fakulta sociálnych a ekonomických vied UK
Mlynské luhy 4
821 05 Bratislava 7
email: miroslava.janosova@fsev.uniba.sk

POUŽITÍ ZOBECNĚNÝCH ADITIVNÍCH MODELŮ PRO PREDIKCI CEN OJETÝCH VOZŮ

USING GENERALIZED ADDITIVE MODELS TO PREDICT USED CAR PRICES

Dominik Kavřík¹

Abstrakt

Trh ojetých automobilů je typem trhu s výraznou informační asymetrií mezi kupujícím a prodávajícím. Potenciální prodávající může chtít využít dostupné informace z nabídek automobilů, stejně jako kupující, aby zjistil, jak ocenit svůj vůz v případě koupě nebo prodeje. K modelování a předpovědi ceny na základě uvedených atributů vozidla jsou v této studii použity zobecněné aditivní modely. Data odhalila nelineární vliv stáří vozidel a jejich najetých kilometrů na výslednou cenu.

***Klíčová slova:** aditivní modely, trh ojetých aut, Škoda*

Abstract

The used-car market is a type of market with significant information asymmetry between the buyer and seller. The potential seller might want to use the available information from the car listings as well as the buyer to find out how to evaluate their car in case of the purchase or selling. Generalized additive models are applied to model and predict the price based on the vehicle's listed attributes. The data revealed non-linear propagation of the effect of vehicles age and its mileage on the resulting price.

***Keywords:** additive models, used-car market, Skoda*

1 INTRODUCTION

This study aims to model and predict the price of used Skoda vehicles on the market in the Czech Republic. The used-car market is specific for the information asymmetry between the buyer and seller. The quality uncertainty on the market affects the resulting price that, in the end, affects the average quality and condition of the offered used cars put on the market. There is an assumed non-linear relationship between the price of a vehicle and its features. Generalized Additive Models (GAM) are used in the analysis to capture the relationships. The data are retrieved from the used car advertisement server Sauto using web scraping techniques in R. The study is structured as follows. The methodology section describes the model used to predict used Skoda price. The data section describes the procedures necessary to obtain the original data from the primary source. The empirical section describes the model equation and the estimation result, which are presented graphically in partial dependence plots for all continuous and factor variables. The conclusion section contains the results summary.

¹ This research project was supported by Grant No. IGA F4/27/2020 of the Internal Grant Agency, Faculty of Informatics and Statistics, Prague University of Economics and Business

2 METHODOLOGY

Generalized additive model (GAM) of authors Hastie and Tibshirani (1987) is used to predict Skoda vehicle price. In general, GAM model has the following form:

$$y_i = \beta_0 + f(x_{i,1}) + f(x_{i,2}) + \dots + f(x_{i,k}) + u_i \quad (1)$$

In this analysis, the function $f(\cdot)$ used to model the vehicle's price will be a natural cubic spline which from all splines are the smoothest interpolators, as noted by Wood (2006). See e. g. Green and Silverman (1993) for a proof of this statement.

GAM model will help to uncover possible non-linearities in the models feature. To do exactly this, the natural cubic splines will be used to fit Generalized additive model is used to capture any non-linearity in the model's feature. To model the non-linearity in the continuous variables, I decided to use natural cubic splines, as did e. g. Cao, Valois, and Goldberg (2006).

The predictive power of GAM model will be tested against the OLS benchmark, with the following population regression function

$$y_i = \beta_0 + x_{i,1} + x_{i,2} + \dots + f(x_{i,k}) + u_i, \quad (2)$$

by comparing the predictions against the values from the test sample, which is later described in the next section. The metric used to evaluate the prediction will be RMSE:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_{\text{train}} - y_{\text{test}})^2}{n}}$$

To be able to test the hypothesis, that one prediction is better then another, the test-train split is repeated for multiple random seeds, in order to obtain a distribution of the above-mentioned metric.

3 DATA

The price and vehicle features are retrieved from the used-car advertisement server Sauto. The data are obtained algorithmically using iterative API GET requests to the database from which the web retrieves the values. Due to the limitation of the total number of listings the API can provide at once, the iterative requests are sent for every offered car model respectively. Those are then merged into the final data set. A few outliers were present and consequently removed. The cleaned dataset contains information from a total of 18417 listings. The data are then randomly split into the testing and training sample in the 75:25 ratio. The following variables were retrieved from the listings:

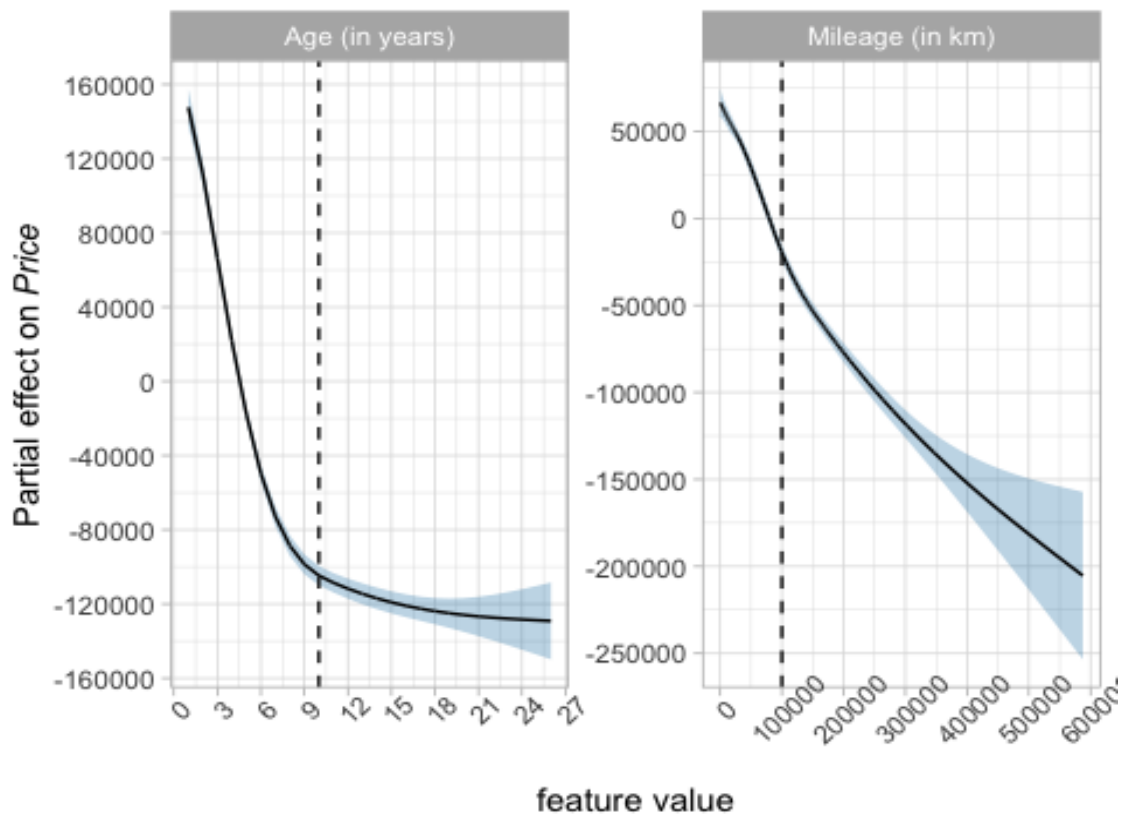
- continuous variables
 - *Price*
 - *Mileage*
 - *Vehicle age*
- categorical variables
 - *Diesel*
 - *SUV*
 - *First owner*
 - *Automatic transmission*
 - *Superb (model name)*
 - *Air condition*
 - *Service history*

4 EMPIRICAL PART

The generalized additive model (GAM) is estimated in the following functional form.

$$\text{price}_i = \beta_0 + f(\text{mileage}) + f(\text{age}) + \text{diesel} + \text{SUV} + \text{owner} + \text{automatic} + \text{AC} + \text{history} + u \quad (3)$$

where “u” stands for the random disturbance term. The natural cubic splines were chosen as the smoother for *age* and *mileage*, each with 5 degrees of freedom. To make the interpretation of this model easier, the partial dependence plots are used further in the empirical part. The partial dependence plots in the following figure represent the effect of each explanatory variable as a function of the feature’s value.

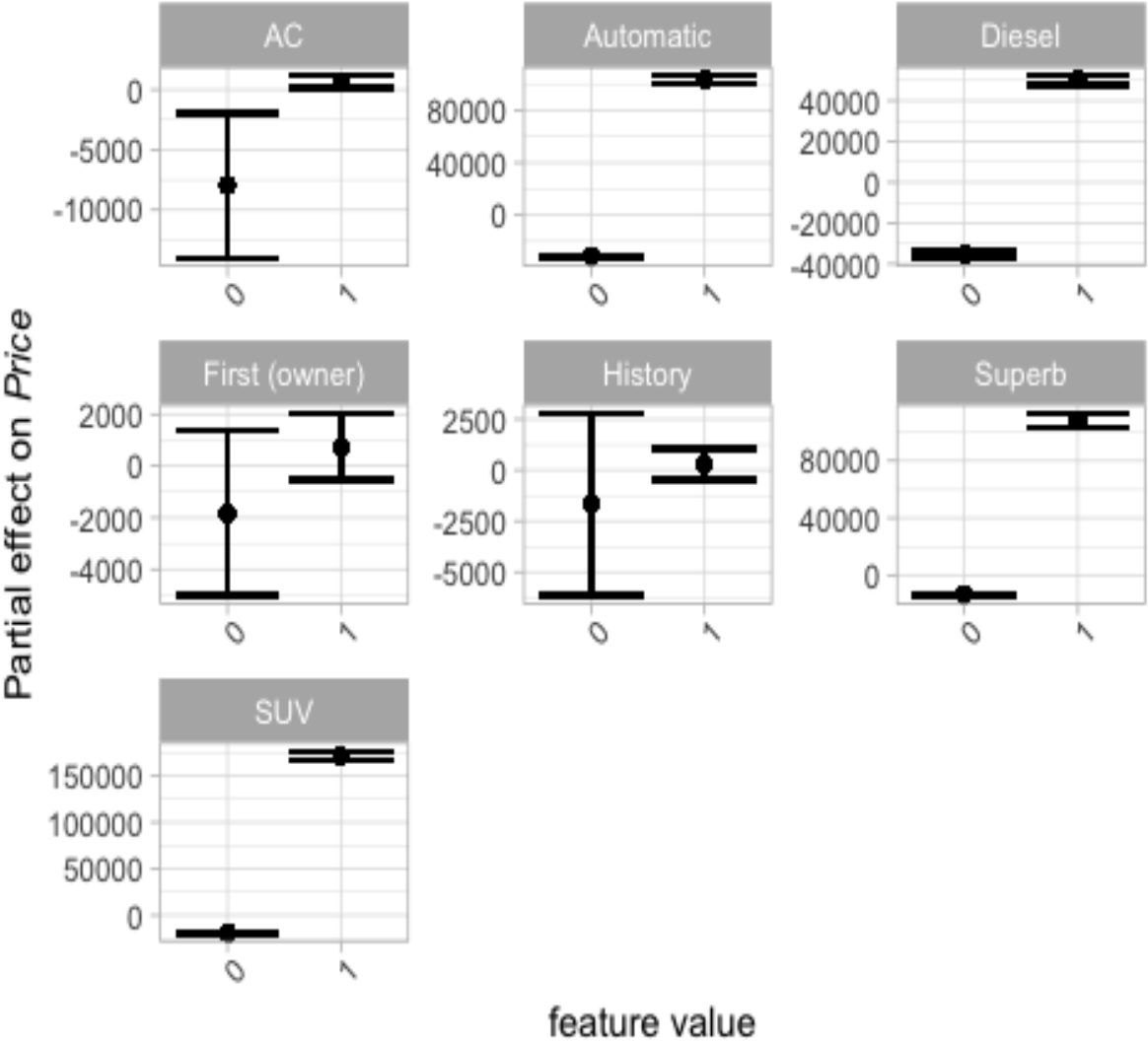


In the figure on the left, there is an apparent switch in the magnitude of the effect of the vehicle's *age* on its price. It appears to be stronger if the vehicle is not older than 10 years. The marginal effect of the additional year then diminishes once the vehicle is older than 10 years up until the age of 20.

Based on the point estimates, the data suggests that the effect becomes positive, which might suggest an effect of rarity, but since there are fewer data points for *age* there is also high uncertainty which is represented by the wider confidence interval for $age > 20$. Due to this fact, it is not possible to evaluate whether the rarity effect exists. The effect might very well be decreasing for vehicles older than 20 years.

The figure on the right displays the negative effect of mileage on the vehicle's price. This finding is not by any means surprising. On the other hand, the effect's non-linear propagation across the mileage values is. The line decreases somewhat rapidly up to 100,000 km compared to the rest of the line beyond this threshold. The price is relatively sensitive to an additional kilometer, but the effect of the marginal increase in mileage above the threshold is lower.

The effect of the categorical variables is summarized in the figure below. the service history and first owner are statistically insignificant.



The predictive performance of GAM model is tested against the prediction obtained using OLS. The RMSE is calculated from the predictions which are compared to the actual value of *price* from the test sample. the GAM model's $RMSE_{GAM} = 96218.83$, which is lower than that of OLS: $RMSE_{OLS} = 104335.4$. Simulating RMSE's from a thousand different test-train splits allowed to test the hypothesis of equality of means between GAM RMSE and OLS RMSE. Based on the p-value from the standard ANOVA test, the hypothesis of equality of the means between each group was rejected. The RMSE of the GAM is significantly lower, hence the predictions from GAM are better compared to the benchmark.

5 CONCLUSION

In this study, the price of used Skoda vehicles was modelled and predicted using the original data from the used car advertisement server. A generalized additive model (GAM) with natural cubic splines estimated a non-linear relationship between the vehicle's age and mileage. There appears to be a threshold of 100.000 km, after which the effect of an additional kilometer diminishes. There is a threshold of 10 years in the variable age. From the point estimates, the slope became positive for vehicles older than 20 years. However, there were fewer data points in the age > 20 region, which increased the estimated standard error. Hence, there is no conclusion about the effect of rarity. The predictive power of GAM was superior to those obtained by the OLS. By simulating multiple test-train splits of the data and re-estimating the models, the average RMSE of the GAM model was significantly different and lower than the one of OLS.

References

1. Cao, Jiguo, Marie-France Valois, and Mark S Goldberg. 2006. "An s-Plus Function to Calculate Relative Risks and Adjusted Means for Regression Models Using Natural Splines." *Computer Methods and Programs in Biomedicine* 84 (1): 58–62.
2. Green, Peter J, and Bernard W Silverman. 1993. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Crc Press.
3. Hastie, Trevor, and Robert Tibshirani. 1987. "Generalized Additive Models: Some Applications." *Journal of the American Statistical Association* 82 (398): 371–86.
4. Wood, Simon N. 2006. *Generalized Additive Models: An Introduction with r*. chapman; hall/CRC.

Contact

Ing. Dominik Kavříkⁱ
Prague University of Economics and Business
Department of Econometrics
W. Churchil Sq. 4
130 67 Prague 3
email: kavd00@vse.cz

ⁱ This research project was supported by Grant No. IGA F4/27/2020 of the Internal Grant Agency, Faculty of Informatics and Statistics, Prague University of Economics and Business.

FUNKCIONÁLNA ANALÝZA ROZPTYLU: PRÍPADOVÁ ŠTÚDIA Z POTRAVINÁRSKEHO PRIEMYSLU

FUNCTIONAL ANALYSIS OF VARIANCE: CASE STUDY FROM FOOD INDUSTRY

Peter Knížat

Abstrakt

Technologický pokrok v informačných systémoch poskytol možnosť zaznamenávať veľmi veľké súbory dát, ktorý tiež umožňuje meranie fyzikálnych alebo sociálnych fenoménov presnejším spôsobom. Dáta v takýchto viacrozmerných súboroch sa dajú transponovať do matematických, tzv. funkcionálnych, objektov pre pozorované vzorky. Štatistická terminológia určená na spracovanie tohto druhu viacrozmerných matematických objektov sa nazýva analýza funkcionálnych dát. V tomto článku predstavíme teoretický rámec, ktorý transponuje reálne dáta na funkcionálne a definuje tieto matematické objekty v štatistickej metóde analýzy rozptylu. V prípadovej štúdií z potravinárskeho priemyslu ukážeme aplikáciu tohto teoretického rámca na pozorované dáta, ktorej cieľom bude overenie, či je možné určiť obsah tuku vo vzorke mäsa na základe spektrálnych absorbancií meraných infračerveným analyzátorom.

KLúčové slová: analýza funkcionálnych dát, splajnová expanzia, funkcionálna analýza rozptylu, spektrometrické dáta

Abstract

The technological progress in information technology provided the ability of recording very large datasets that offers the opportunity to observe phenomena in a more accurate way. The information in such high-dimensional datasets allows such data to be transposed to a collection of mathematical, also called functional, objects for observed samples. The statistical terminology designated to deal with this kind of high-dimensional mathematical objects is called functional data analysis. In this article, we outline a theoretical framework that transposes a real data into functional and defines these mathematical objects in the statistical methodology of analysis of variance. In the case study from the food industry, we demonstrate the application of this theoretical framework on observed data, of which the aim is to evaluate the fat content in a meat sample based on its absorbance spectrum measured by infrared analyzer.

Keywords: functional data analysis, spline expansion, functional analysis of variance, spectrometric data

1 ÚVOD

Téma analýzy funkcionálnych dát je v štatistike veľmi nedávny fenomén, ktorý sa v posledných desaťročiach intenzívne skúmal. Hlavná charakteristika funkcionality v pozorovaných dátach nastáva za predpokladu ak je možné opísať pozorovanú štruktúru vzoriek dát matematickou funkciou tak aby neprišlo k nežiadúcej strate variability v pozorovaných dátach.

Jednou z prelomových štúdií aplikácie analýzy funkcionálnych dát v štatistickej lineárnej regresii je (Cardot a kol., 1999). V tejto štúdií je predstavený teoretický rámec, ktorý definuje matematické objekty v matici nezávislých premenných s korešpondujúcimi funkcionálnymi regresnými parametrami, ktoré sú vypočítané odhadom použitím metódy minimalizácie najmenších štvorcov, a skalárnou závislou premennou.

Podobná štruktúra pre funkcionálne dáta je navrhnutá vo vedeckej štúdií (Ferraty a kol., 2002), s tým, že štatistická regresia je predstavená v neparametrickej forme, tzv. formou jadrovej (kernel) funkcie, a aplikovaná na spektrometrické dáta.

(Ramsay a kol., 2005) sa rozsiahle venuje problematike transponovania reálnych dát do matematických objektov pomocou metód vyrovnávania (smoothing) dát a ich aplikácie v štatistických lineárnych regresných modeloch. V ďalšej knihe (Ramsay a kol., 2009) sú tieto metódy aplikované v softvéroch R a MATLAB.

Zámerom tohto článku bude predstaviť teoretický rámec, ktorý transponuje reálne dáta na funkcionálne a definuje tieto matematické objekty v štatistickej metóde analýzy rozptylu. V prípadovej štúdií z potravinárskeho priemyslu našim cieľom bude aplikovať tento teoretický rámec na pozorované dáta a overiť, či je možné určiť obsah tuku vo vzorke mäsa na základe spektrálnych absorbancií meraných infračerveným analyzátorom.

2 THEORETICKÉ VÝCHODISKÁ

V tejto kapitole predstavíme teoretický rámec pre transponovanie reálnych dát na funkcionálne a funkcionálnu analýzu rozptylu (ANOVA).

2.1 Funkcionálne dáta

Štandardne, zo štatistického pozorovania získame maticu reálnych čísel vo forme X_{ij} , kde $j = 1, \dots, p$ reprezentuje počet pozorovaných nezávislých náhodných premenných, a korešpondujúci vektor reálnych čísel y_i , ktorý reprezentuje závislú premennú, obidve pozorované pre vzorky $i = 1, \dots, n$.

Matematické funkcie sú interpolované do matice reálnych čísel X_{ij} pre každú pozorovanú vzorku i , takže p -rozmerný priestor reálnych čísel sa mapuje do nekonečného, alebo funkcionálneho, priestoru. Zápis pre takúto maticu funkcionálnych premenných je $\chi_i(t)$, kde $i = 1, \dots, n$ reprezentuje počet funkcií vzoriek a t je kontinuum, na ktorom môžu byť dané vzorky pozorované.

Pre transponovanie reálnych dát do funkcionálnych je možné použiť rôzne typy matematických funkcií. V tejto štúdií použijeme metódu expanzie bázických splajnov, tzv. B-splajnov, ktorá je lineárna kombinácia splajnových kriviek B_k . Takže $\chi_i(t)$ je možné vyjadriť vo forme bázických splajnových kriviek nasledovne (Ramsay a kol., 2005)

$$\chi_i(t) = \sum_{k=1}^K c_k B_k(t, \tau_l), \quad l = 1, \dots, L - 1 \quad (1)$$

kde B_k sú čiastkové polynómy radu m , ktoré sú spojené v bode uzla τ_l , kde $L - 1$ je počet uzlov a K je počet kriviek v expanzii. Počtom kriviek v expanzii kontrolujeme stupeň vyhladenia dát. Vektor koeficientov c_k môže byť vypočítaný odhadom použitím štandardnej metódy minimalizácie najmenších štvorcov. Podrobný popis teoretického rámca splajnových expanzií je poskytnutý v (De Boor, 2001).

Aplikácia štatistických metód regresie a analýzy rozptylu, ktoré skúmajú spôsob akým možno variabilitu pozorovaných reálnych dát vysvetliť pozorovanými funkciami, je použiteľná na analýzu funkcionálnych dát. Takúto štruktúru môžeme formalizovať do nasledovného funkcionálneho modelu

$$y_i = f(\chi_i(t)) + \varepsilon_i, \quad i = 1, \dots, n \quad (2)$$

kde ε_i je vektor náhodných chýb, ktorý má normálne rozdelenie s priemerom 0 a konštantnou štandardnou odchýlkou σ .

2.2 Funkcionálna analýza rozptylu

Konvenčná analýza rozptylu môže byť rozšírená do funkcionálnej formy definovaním splajnových expanzií v kritériu minimalizácie najmenších štvorcov. V analýze rozptylu závislá premenná je funkcionálna a nezávislá premenná je skalárna. Ďalší rozdiel je, že regresné parametre majú funkcionálnu formu, takže $\beta(t)$ sú pozorovateľné na kontinuálnej doméne t .

Terminológiu môžeme sformalizovať nasledovne. Takže máme počet pozorovaných funkcií N_c , v každej kategórii $c = 1, \dots, C$ a model pre i -tú funkciu v c -tej kategórii, označený ako $y_{ic}(t)$, (Ramsay a kol., 2005), je

$$y_{ic}(t) = \mu(t) + \alpha_c(t) + \varepsilon_{ic}(t) \quad (3)$$

kde funkcia $\mu(t)$ je celkový priemer pre všetky N funkcie, $N = \sum_{c=1}^C N_c$. $\alpha_c(t)$ sú špecifické efektové funkcie, ktoré reprezentujú odchýlky od celkového priemeru pre každú kategóriu c . Funkcia náhodných chýb $\varepsilon_{ic}(t)$ je nevysvetlená variácia špecifická pre i -tú funkciu pre každú kategóriu c .

Aby bolo možné vypočítať odhadom špecifické efektové funkcie pre jednotlivé kategórie, obmedzenie $\sum_{c=1}^C \alpha_c(t) = 0$ pre všetky t musí byť splnené.

Rovnicu (3) možno predefinovať nasledovne (Ramsay a kol., 2005). Matica $X_{(ic)j}$ s rozmermi $N \times (C + 1)$ pre i -tú funkciu v c -tej kategórii má hodnoty 1 ak funkcia spadá do danej kategórie a 0 v opačnom prípade, kde stĺpec j pre celkový priemer funkcií má hodnoty 1. Funkcionálne regresné koeficienty $\beta_j(t)$ sú definované ako $\beta_1(t) = \mu(t), \beta_2(t) = \alpha_1(t) \dots \beta_{C+1}(t) = \alpha_C(t)$, takže $\beta_j(t) = (\mu(t), \alpha_1(t), \dots, \alpha_C(t))^T$. Model v rovnici (3) je ekvivalentný s nasledovným modelom

$$y_{ic}(t) = \sum_{j=1}^{C+1} X_{(ic)j} \beta_j(t) + \varepsilon_{ic}(t) \quad (4)$$

Každá funkcia závislej premennej $y_{ic}(t)$ a funkcionálne regresné koeficienty $\beta_j(t)$ môžu byť vyjadrené v splajnovej expanzii nasledovne

$$y_{ic}(t) = \sum_{k=1}^{K_y} C_{ik} \theta_k(t), \quad \beta_j(t) = \sum_{k=1}^{K_\beta} B_{jk} \psi_k(t) \quad (5)$$

kde $\theta_k(t)$ je K_y -rozmerný vektor, ktorý obsahuje splajnové krivky, a matica C_{ik} s rozmermi $N \times K_y$ obsahuje koeficienty expanzie splajnových kriviek pre $y_{ic}(t)$. Podobne $\psi_k(t)$ je K_β -rozmerný vektor, ktorý obsahuje splajnové krivky, a matica B_{jk} s rozmermi $(C + 1) \times K_\beta$ obsahuje koeficienty expanzie splajnových kriviek pre $\beta_j(t)$.

Kritérium najmenších štvorcov (KNŠ) pre výpočet odhadom $\beta_j(t)$, rozšírené do funkcionálnej formy (Ramsay a kol., 2005), je

$$KN\check{S}(\beta_j(t)) = \sum_{c=1}^C \sum_{i=1}^{N_c} \int \left(y_{ic}(t) - \sum_{j=1}^{C+1} X_{(ic)j} \beta_j(t) \right)^2 dt \quad (6)$$

Aby bolo možné vypočítať odhadom $\beta_j(t)$, minimalizáciou kritéria najmenších štvorcov, obmedzenie $\sum_{j=2}^{C+1} \beta_j(t) = 0$, ekvivalentné s $\sum_{c=1}^C \alpha_c(t) = 0$, pre všetky t musí byť splnené. Detailný popis výpočtu je uvedený v (Ramsay a kol., 2005).

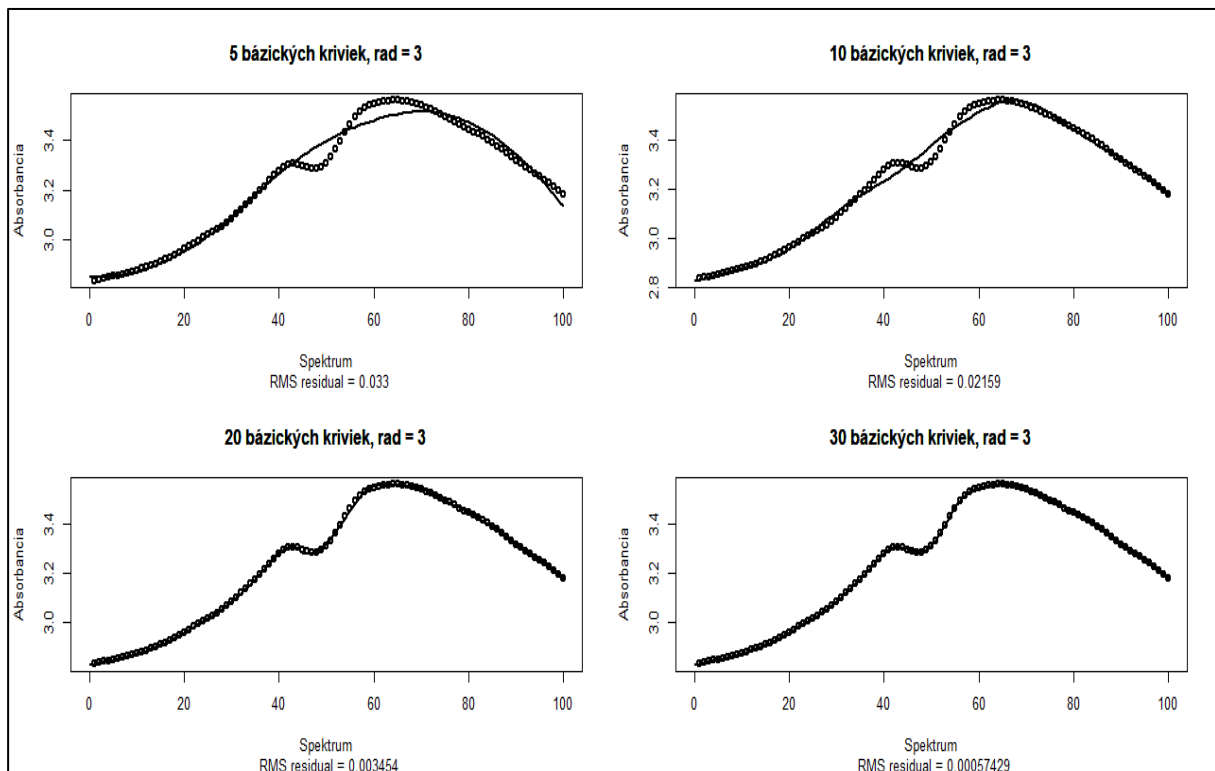
3 VÝSLEDKY

V tejto časti aplikujeme teoretický rámec uvedený v predchádzajúcej časti na pozorované dáta. Použijeme dáta, ktoré pochádzajú zo skúmania kontroly kvality v potravinárskom priemysle, tzv. spektrometrické dáta. Detailný popis dát, tzv. Tecator data set, je na webovej stránke <http://lib.stat.cmu.edu/datasets/tecator>.

Každá vzorka v súbore dát reprezentuje kúsok mäsa s rôznym obsahom tuku, ktorý je meraný v percentách, kde máme celkom 215 vzoriek. Korešpondujúce merania spektrálnych absorbcíí pre každú vzorku mäsa boli zaznamenané pomocou analyzátora, ktorý pracuje na princípe infračerveného prenosu. Pre každú vzorku bolo zaznamenaných 100 spektrálnych absorbcíí.

V prvom kroku transponujeme pozorované reálne dáta pre 100 premenných, spektrálnych absorbcíí, na funkcionálne dáta pre každú vzorku mäsa. Použijeme expanziu bázických splajnov definovanú v rovnici (1) s čiastkovými polynómami 3-teho radu s rôznym počtom bázických kriviek (5, 10, 20 a 30). Graf 1 zobrazuje vyhladenie pre jednu vzorku mäsa.

Graf 1: Jedna vzorka mäsa



Zdroj: Vlastné vypracovanie

Graf 1 ukazuje, že pri vyhladení pozorovaných dát dochádza len k minimálnej strate variability s porovnaním s funkcionálnymi dátami ak použijeme 20, alebo viac, bázických kriviek v splajnovej expanzii. V ďalšej analýze použijeme splajnovú expanziu s 30-timi bázickými krivkami. Takže máme $y_i(t), i = 1, \dots, 215$ funkcií vzoriek mäsa, ktoré sú pozorovateľné na kontinuálnej doméne t .

Vzhľadom na stanovený cieľ tohto príspevku, pokračujeme s využitím metodologického aparátu uvedeného v kapitole 2. Hlavnou otázkou je preskúmať, či obsah tuku v mäse je možné vysvetliť na základe meraní spektrálnych absorbancií.

Funkcie vzoriek mäsa rozdelíme do 5-tich kategórií podľa obsahu tuku v mäse. Rozdelenie do kategórií je stanovené autorom a nie je založené na žiadnom štatistickom modeli alebo expertíze v danej problematike. Tiež vytvoríme maticu $X_{(ic)j}$ podľa definície v rovnici (4).

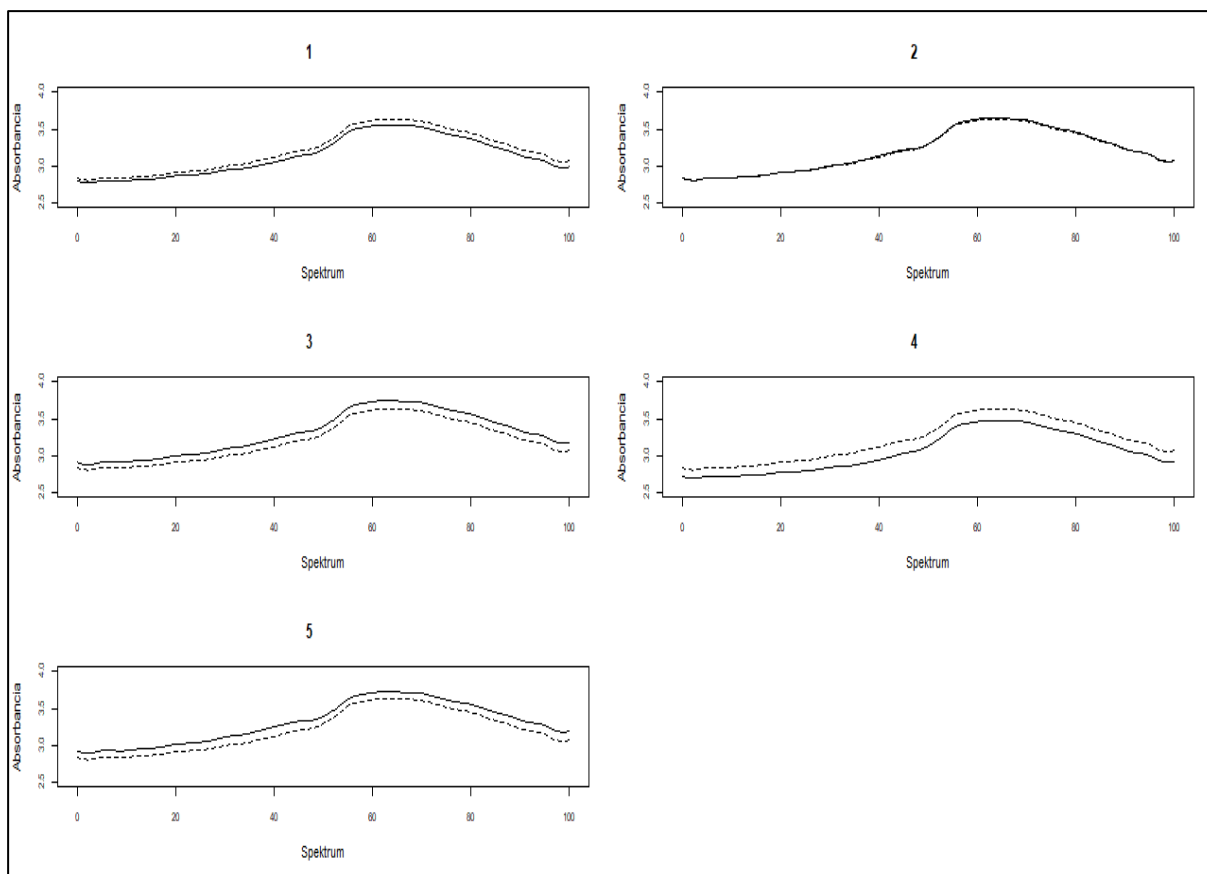
Tabuľka 1: Kategórie – počet funkcií vzoriek mäsa podľa obsahu tuku

Kategória	Obsah tuku v mäse v % (OTM)	Počet funkcií v kategórii
1	$0 \leq \text{OTM} < 10$	77
2	$10 \leq \text{OTM} < 20$	61
3	$20 \leq \text{OTM} < 30$	36
4	$30 \leq \text{OTM} < 40$	23
5	$40 \leq \text{OTM} < 50$	18

Zdroj: Vlastné vypracovanie

Po aplikovaní modelu v rovnici (3), výpočtom odhadované parametre sú zobrazené graficky v nasledujúcom grafe, kde plné čiary znázorňujú odhadované profily $\hat{\mu}(t) + \hat{\alpha}_c(t)$ pre 5 kategórií a prerušovaná čiara je odhadovaný celkový priemer $\hat{\mu}(t)$ pre všetky N funkcie vzoriek mäsa.

Graf 2: Odhadované funkcionálne parametre



Zdroj: Vlastné vypracovanie

Graf 2 ukazuje, že odhadovaný profil pre kategóriu 2 je takmer zhodný s odhadovaným celkovým priemerom funkcií, ale odhadované profily pre kategórie 1, 3, 4 a 5 sa odchyľujú od odhadovaného celkového priemeru funkcií. Na základe tejto grafickej vizualizácie predpokladáme, že spektrálne absorbancie môžu byť determinant pre obsah tuku v mäse.

4 ZÁVER

Cieľom tejto štúdie bolo predstavenie teoretického rámca analýzy funkcionálnych dát a jeho následné aplikovanie v praxi. V prípadovej štúdií zo skúmania kontroly kvality v potravinárskom priemysle sme použili merania spektrálnych absorbancií, ktoré boli zaznamenané pomocou infračerveného analyzátoru, na aplikáciu teoretického rámca na pozorované dáta. V prvom kroku pozorované reálne dáta jednotlivých vzoriek skúmaní boli transponované do funkcionálnych, pomocou metódy expanzie bázických spaljnov. Následne sme predstavili teoretický rámec pre funkcionálnu analýzu rozptylu a demonštrovali aplikáciu v prípadovej štúdií, ktorej cieľom bolo objasniť, že či obsah tuku vo vzorke mäsa je možné vysvetliť na základe meraní spektrálnych absorbancií. Funkcionálna analýza rozptylu a grafické vyobrazenie odhadovaných parametrov modelu ukázali, že obsah tuku môže byť vysvetlený spektrálnymi absorbanciami.

Avšak, ďalšia analýza je potrebná na vyhodnotenie štatistickej významnosti odhadovaných funkcionálnych špecifických efektov pre jednotlivé kategórie a ich odlišnosti od odhadovaného celkového priemeru pre všetky funkcie vzoriek mäsa.

Použitá literatúra

1. Anderson T.W. (2003). An Introduction to Multivariate Statistical Analysis. Third edition. New York: Wiley.
2. Cardot H., Ferraty F., Sarda P. (1999). Functional linear model. Statistics and Probability Letters 45, 11-22.
3. De Boor C. (2001). A Practical Guide to Splines, Revised Edition. New York: Springer.
4. Ferraty F. and Vieu P. (2002). The Functiona Nonparametric Model and Application to Spectrometric Data. Computational Statistics 17, 545-564.
5. Ferraty F. and Romain Y. (2011). The Oxford Handbook of Functional Analysis. Oxford University Press.
6. Ramsay J. O. and Silverman B. W. (2005). Functional Data Analysis. Second Edition. Springer Series in Statistics, Springer.
7. Ramsay J. O., Hooker G., Graves S. (2009). Functional Data Analysis with R and MATLAB. Use R!, Springer.
8. Tecator data set: <http://lib.stat.cmu.edu/datasets/tecator>

Kontaktné údaje

Peter Knížat

Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky

Dolnozemská 1/b, 852 35 Bratislava

Tel: (421) 910412189

email: peter.knizat@euba.sk

ODHAD DYNAMICKÉHO EFEKTU ČLENSTVA V EU POMOCOU GRAVITAČNÉHO MODELU

ESTIMATION OF THE DYNAMIC EFFECT OF EU MEMBERSHIP USING A GRAVITY MODEL

Brian König

Abstrakt

Článok sa zaoberá odhadom efektu členstva v EU na bilaterálny obchodný tok pomocou gravitačných modeloch. Dôraz je kladený na komparáciu odhadu statického a dynamického efektu členstva. Z výsledkov dosiahnutých v aplikačnej časti článku možno predpokladať, že efekt členstva v európskej únii má štatisticky významný dopad na bilaterálny obchodný tok obchodujúcich krajín. V prípade kontroly dynamického efektu členstva je celkový efekt o takmer 10 percentuálnych bodov vyšší, ako v modeli so statickým efektom.

Kľúčové slová: gravitačný model, dynamický, členstvo v európskej únii

Abstract

The article deals with estimating the effect of EU membership on bilateral trade flow using gravity models. Emphasis is put on the comparison of static and dynamic membership effect estimates. From the results obtained in the application part of the article, it can be stated that the effect of membership in the European Union has a statistically significant impact on the bilateral trade flow of trading countries. While controlling for the dynamic effect of membership, the total effect is almost 10 percentage points higher as compared to the model with a static effect.

Keywords: gravity model, dynamic effect, EU membership

1 ÚVOD

Vstup do EU je významným krokom v dejinách jednotlivých európskych krajín. Každá krajina pred vstupom do EU zvažovala výhody a záväzky, ktoré jej zo vstupu do EU budú vyplývať. Jednou z ekonomických výhod členstva v EU je voľný obchod medzi krajinami EU a zvyčajne výhodnejšie obchodné dohody so zvyškom sveta, akoby krajina bola schopná vyrokovať sama o sebe bez členstva v EU. Uvedená skutočnosť má zvyčajne pozitívny dopad na zahraničný obchod členskej krajiny EU. Poznanie vplyvu členstva v EU na zahraničný obchod je dôležité ako pre potenciálne prístupové krajiny do EU, tak aj pre krajiny, ktoré zvažujú odchod z EU. Vstup členstva do EU sa nemusí hneď úplne premietnuť do nárastu bilaterálneho obchodu, ale môže skôr dochádzať k postupnému prispôbovaniu obchodu členstvu v EU. Túto hypotézu možno overiť pomocou kontrolovania dynamického efektu členstva v EU. Cieľom príspevku je odhadnúť a porovnať statický a dynamický efekt členstva v EU na bilaterálny obchod.

2 METODOLÓGIA

Gravitačný model zahraničného obchodu je historicky spájaný s holandským ekonómom Jánom Tinbergenom, ktorý tento koncept predstavil v roku 1962 (Tinbergen, 1962). V tejto

časti budeme uvažovať s gravitačným modelom odvodeným z dopytovej strany ekonomiky podľa Andersona a van Wincoopa (2003) založeného na Anderson (1979).

Ak takto upravený vzťah (1.14) následne dosadíme do vzťahov (1.8) a (1.9) získame štruktúrnu podobu gravitačného modelu Andersona a Wincoopa (2003) v tvare:

$$X_{ij} = \frac{E_j Y_i}{Y_w} \left[\frac{t_{ij}}{\Pi_i P_j} \right]^{1-\sigma}, \quad (2.1)$$

$$P_j^{1-\sigma} = \sum_i \left(\frac{t_{ij}}{\Pi_i} \right)^{1-\sigma} \frac{Y_i}{Y_w}, \quad (2.2)$$

$$\Pi_i^{1-\sigma} = \sum_j \left[\frac{t_{ij}}{P_j} \right]^{1-\sigma} \frac{E_j}{Y_w}. \quad (2.3)$$

kde X_{ij} sú bilaterálne obchodné toky z krajiny i do krajiny j ; Y_w je príjem sveta (resp. produkcia); Y_i je domáca produkcia krajiny i ; E_j sú celkové výdavky krajiny j ; t_{ij} reprezentuje bilaterálne obchodné náklady medzi krajinami i a j ; σ je elasticita substitúcie medzi tovarmi z jednotlivých krajín;

Rovnicu (2.1) možno rozložiť na dve časti. Časť mimo zátvorky, hovoriacu o hypotetickej úrovni obchodu, ak by neexistovali obchodné náklady, bariéry a za predpokladu rovnakých cien. Druhá časť pozostáva z obchodných bariér, resp. nákladov a multilaterálnych rezistencií. Obchodné náklady t_{ij} sú v literatúre často aproximované geografickými premennými a premennými reprezentujúcimi politiku zahraničného obchodu ako napr. vzdialenosť, clá, spoločný jazyk, dohody o voľnom obchode, existencia spoločnej hranice atď. Vzťah (2.2) vyjadruje vnútornú multilaterálnu rezistenciu importéra j , ktorá hovorí o náročnosti vstupu na trh importéra j . Rovnica (2.3) reprezentuje vonkajšiu multilaterálnu rezistenciu a tá vyjadruje náročnosť vstupu exportéra i na trh. Multilaterálne rezistencie budú aproximované pomocou fixných efektov premenlivých v rozmeroch importér - čas, exportér - čas a importér-exportér (Harrigan, 1996; Baier a Bergstrand, 2007). Na odhad parametrov bude použitý pseudo-Poissonov estimátor maximálnej vierohodnosti PPML (Silva a Tenreyro, 2006)

2.1 Odhad multilaterálnych rezistencií pomocou fixných efektov

V prvej časti preskúmame všeobecne vplyv členstva v EU na bilaterálny tok obchodu na celom časovom intervale. Keďže budú v modeli použité tri typy fixných efektov (importér - čas, exportér - čas, importér - exportér) všetky premenné, ktoré sú premenlivé iba v týchto troch rozmeroch budú absorbované fixnými efektmi, a teda nebude možné odhadnúť ich vplyv priamo (všetky premenné, ktoré sú špecifické pre jednu z obchodujúcich krajín a sú premenlivé v čase ako napr. HDP, veľkosť populácie atď.; rovnako aj všetky premenné premenlivé iba v rozmere importér - exportér napr. vzdialenosť, koloniálne väzby, spoločný oficiálny jazyk atď.). Multiplikatívny gravitačný model možno zapísať v tvare:

$$x_{ij,t} = \exp(\beta_1 eu_{oaij,t} + \phi_{ij} + \gamma_{i,t} + \delta_{j,t}) \varepsilon_{ij,t}, \quad (2.4)$$

kde t je index času; $x_{ij,t}$ teraz reprezentuje bilaterálny obchodný tok z krajiny j do krajiny i (už nie normalizovaný pomocou HDP krajín ako v predchádzajúcom prípade); ϕ_{ij} sú fixné efekty v rozmere importér - exportér. Keďže prispôsobenie obchodného toku medzi krajinami na zmenu obchodnej politiky alebo obchodných nákladov môže trvať istú dobu, viacerí autori (napr. Baier a Bergstrand, 2007; Bergstrand, Larch a Yotov, 2015) odporúčajú použiť na odhad gravitačného modelu namiesto po sebe idúcich rokov časové intervaly (intervaly troch, štyroch, resp. piatich rokov).

3 DÁTA

Ako závislá premenná bude použitý bilaterálny import tovarov databázy DOTS (Direction of Trade Statistics) medzinárodného menového fondu. Databáza pokrýva údaje o bilaterálnom obchode za 184 krajín, pričom ročné údaje sú k dispozícii od roku 1948 až po súčasnosť zhruba s dvojročným oneskorením. Údaje sú publikované v amerických dolároch. Import bol zvolený za závislú premennú z toho titulu, ako uvádzajú viacerí autori (napr. Yotov a kol., 2016), že údaje o importe sú zvyčajne spoľahlivejšie ako údaje o exporte, pretože importy do krajiny sú o mnoho viac monitorované v porovnaní s exportom (najmä kvôli clu). Okrem toho v DOTS databáze je viacej pozorovaní importu ako exportu, a preto budeme využívať ako závislú premennú import vyjadrený v amerických dolároch (USD). Za účelom našej analýzy boli použité údaje za roky 1950 - 2015 pre 138 najväčších obchodných partnerov EU v roku 2015, pričom medzi nimi sa nachádza všetkých 28 štátov EU. Údaje o zmluvách o voľnom obchode (fta) boli čerpané z databázy Egger, Peter H. and Mario Larch (2008), ktorá okrem iného obsahuje aj informácie o colných úniách a dohodách o ekonomickej integrácii. Rok 2015 bol zvolený kvôli tomu, že väčšina premenných je použitých z databáz francúzskeho centra pre výskum a expertízy o svetovej ekonomike CEPPII (databázy Gravity a GeoDist), ktoré poskytuje rôzne databázy (okrem iného aj databázy slúžiace k tvorbe gravitačných modelov). Databáza „Gravity“ obsahuje informácie za obdobie rokov 1948 – 2015, pričom poskytuje zhruba 70 premenných za 224 krajín. Informácia o členstve v EU je použitá práve z tejto databázy.

4 ODHAD DYNAMICKÉHO EFEKTU ČLENSTVA V EU POMOCOU GRAVITAČNÉHO MODELU

V tejto časti porovnáme statický a dynamický odhad efektu členstva v EU na bilaterálny obchod získané z gravitačných modelov. V jednotlivých modeloch budeme popri členstve v EU kontrolovať aj vplyv premennej reprezentujúcej *fta*. V prvom kroku budeme skúmať, či efekt členstva v EU je v čase stabilný alebo sa mení v závislosti od skúmaného obdobia.

Na základe výsledkov v tabuľke (Tabuľka 1) možno konštatovať, že efekt členstva v EU na bilaterálny obchod nebol v čase stabilný, ale súvisel so špecifikami jednotlivých období. Do roku 1990 možno pozorovať nárast vplyvu členstva v EU, pričom medzi prvým a druhým obdobím sa efekt takmer zdvojnásobil. Uvedený jav môže byť dôsledkom toho, že druhé obdobie zachytáva vstup Spojeného kráľovstva do EU, ktoré predstavuje ekonomicky veľkú a vyspelú krajinu.

Tabuľka 1 – Odhad gravitačného modelu s časovými intervalmi

	(1) <1970	(2) <1980	(3) <1990	(4) 1950-2015	(5) >2000
import					
fta	0.2500*** (0.0525)	0.0941*** (0.0243)	0.1434*** (0.0329)	-0.0119 (0.0492)	0.0153 (0.0339)
eu_od	0.3307*** (0.0704)	0.6543*** (0.0835)	0.6766*** (0.0805)	0.4259*** (0.0559)	0.2382*** (0.0522)
<i>N</i>	21122	46392	68652	194352	93442
<i>R</i> ²	0.992	0.992	0.992	0.991	0.994

Standard errors in parentheses

* $p < 0.10$, ** $p < .05$, *** $p < .01$

Model (1) reprezentuje odhad na období <1970. Model (2) reprezentuje odhad na období <1980. Model (3) reprezentuje odhad na období <1990. Model (4) reprezentuje odhad na období 1950-2015. Model (5) reprezentuje odhad na období >2000.

Ak zohľadníme celé obdobie, možno pozorovať, že efekt členstva v EU začal v záverečnej časti klesať, čo môže byť spôsobené najmä nižším efektom po roku 2000. Uvedené môže byť v dôsledku nárastu vplyvu ekonomickej integrácie a globalizácie, ale aj finančnej globalizácie. Rovnako zaujímavý je aj vývoj efektu dohôd o voľnom obchode, ktorý taktiež nie je stabilný v čase. Do roku 1990 bol efekt *fta* štatisticky významne odlišný od nuly, pričom najsilnejší bol v prvom období, kedy ešte ekonomická spolupráca a integrácia medzi jednotlivými krajinami nebola tak dobre rozvinutá ako tomu bolo neskôr. V poslednom období po roku 2000 je efekt *fta* štatisticky nevýznamný a relatívne blízky nule. Existencia tohto javu môže byť spôsobená z rovnakej príčiny, z akej dochádza k poklesu vplyvu členstva v EU.

Bilaterálny obchod sa nemusí okamžite prispôbiť zmene obchodnej politiky a prispôbenie zahraničného obchodu na vstup do EU môže trvať niekoľko období. Aby sme zachytili možnosť postupného prispôbenia sa zahraničného obchodu vstupu do EU, v ďalšej časti sa pokúsime explicitne zahrnúť pozvoľné prispôbovanie sa bilaterálneho obchodu členstvu v EU zavedením oneskorených premenných.¹ Konkrétne budeme skúmať ako je schopný sa bilaterálny obchod prispôbiť vstupu do EU približne za jednu dekádu. Budeme uvažovať s modelom v tvare:

$$x_{ij,t} = \exp(\beta_1 fta_{ij,t} + \beta_2 fta_{ij,t-3} + \beta_3 fta_{ij,t-6} + \beta_4 fta_{ij,t-9} + \beta_5 eu_{od_{ij,t}} + \beta_6 eu_{od_{ij,t-3}} + \beta_7 eu_{od_{ij,t-6}} + \beta_8 eu_{od_{ij,t-9}} + \phi_{ij} + \gamma_{i,t} + \delta_{j,t}) \varepsilon_{ij,t}. \quad (4.1)$$

Tabuľka 4.2 – Odhad dynamického gravitačného modelu

	(1) basic	(2) lag
import		
eu_od	0.4259*** (0.0559)	0.3400*** (0.0419)
eu_od_lag3		0.0575** (0.0246)
eu_od_lag6		0.0525** (0.0241)
eu_od_lag9		0.0378 (0.0282)
fta	-0.0119 (0.0492)	-0.0234 (0.0330)
fta_lag3		0.0255 (0.0228)
fta_lag6		0.0093

¹ Obdobný postup použili Baier a Bergstrand (2007) na skúmanie efektu dohôd o voľnom obchode (*fta*).

6. HARRIGAN, J. (1996). Openness to Trade in Manufactures in the OECD. *Journal of international economics* 40 (1-2), 23–39.
7. SANTOS SILVA, J. – TENREYRO, S. (2006). The log of gravity. *The Review of Economics and Statistics* 88 (4), 641–658.
8. TINBERGEN, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. 1st ed. New York: The Twentieth Century Fund.
9. YOTOV, Y. V. – PIERMARTINI, R. – MONTEIRO, J. A. – LARCH, M. (2016). An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model (Geneva: WTO, 2016).

Kontaktné údaje

Ing. Brian König, PhD.,

a, Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky,
Dolnozemska 1/b, 852 35 Bratislava.

b, Ekonomický ústav SAV,
Šancová 56, 811 05 Bratislava.

e-mail: brian.konig@euba.sk

OČEKÁVANÉ INDIVIDUÁLNÍ NÁKLADY PROMISKUITY

EXPECTED INDIVIDUAL COSTS OF PROMISCUITY

Petr Krautwurm

Abstrakt

Práce se zabývá očekávanými individuálními náklady promiskuity, přičemž je vypočítává pro osobu ve věku 20 let, která se účastní nechráněné soulože s neznámou stejně starou osobou. Práce udává, že se tyto náklady na jednu nechráněnou soulož pohybují okolo 19 238 Kč. Hlavní položku v nákladech tvoří možnost početí dítěte, zatímco možnost nakazit se pohlavní nemocí je položkou spíše marginální.

***KLíčové slová:** Promiskuita, Očekávané náklady, QALY*

Abstract

The paper investigates the expected individual costs of promiscuity, calculating them for a 20-year-old person who engages in unprotected intercourse with an unknown person of the same age. The paper indicates that these costs for one unprotected intercourse are around CZK 19,238. The main item in the costs is the possibility of conceiving a child, while the possibility of contracting a venereal disease is a rather marginal item.

***Keywords:** Promiscuity, Expected costs, QALY*

ÚVOD

Náklady jsou esenciálním faktorem v rozhodování, neboť nákladnost určitých aktivit determinuje vůli rozhodovatele je provádět. V případech nejistoty však není možné náklady vyčíslit s jednoznačnou přesností a tedy je nutné počítat pouze s očekávanými náklady. Očekávané náklady v takovém případě odpovídají střední hodnotě konkrétního prospektu, kterému rozhodovatel čelí. Toto je zvláště relevantní v případě, že jednotlivá rozhodnutí tvoří předzávazky.

V případě promiskuity rozhodovatel čelí prospektu, s nímž se pojí náklady jak zdravotní, tak sociální, a tedy by bylo smysluplné určit jeho střední hodnotu, aby bylo možné dále analyzovat rozhodovatelovo jednání v rámci promiskuitního chování. Předchozí práce na toto téma sice ilustrují, že analýza nákladů a výnosů v rámci daných aktivit predikuje smysluplné výsledky, ale tyto práce jsou spíše obecného charakteru a již se nezabývají přesnými hodnotami jednotlivých faktorů (Houdek & Koblavský, 2017).

Tato práce vyčísluje celkovou individuální očekávanou hodnotu nákladů jedné jednotky promiskuity, tedy právě očekávaných nákladů spojených s jedním aktem nechráněné soulože s neznámou osobou.

1 PŘEDPOKLADY

Práce předpokládá, že náklady promiskuity lze zobecnit do jedné jednotky v podobě jedné nechráněné soulože s neznámou osobou, která je zakončena ejakulací. Daný předpoklad je

zvolen z důvodu jednoznačnosti, neboť mnohá data týkající se pravděpodobností početí jsou získána na párech, které se o početí snažily, a tedy docházelo k plnohodnotnému dokončení soulože.

Zároveň, náklady jsou počítány pro rozhodovatele ve věku 20 let, který se účastní pohlavního styku se stejně starou osobou. Tento předpoklad je zvolen z důvodu, že se jednotlivé náklady mohou napříč věkovými skupinami výrazně lišit, především kvůli plodnosti.

2 ZÁKLADNÍ MODEL

Model předpokládá, že základní náklady spojené s promiskuitou (konkrétně s jedním nechráněným pohlavním stykem s neznámou osobou) lze rozložit především na náklady týkající se neplánovaného početí dítěte a náklady související s nákazou pohlavní nemocí. Ostatní potenciální náklady, jako například mechanické poranění či skandál, mohou být zanedbány buď z důvodu jejich nepříliš velkého významu nebo z důvodu jejich zřejmé závislosti na jiných charakteristikách rozhodovatelů. V případě mechanického poranění lze argumentovat, že pravděpodobnost jeho výskytu je malá (Abasiattai et al., 2005), přičemž i náklady s ním spojené bývají zanedbatelné (Dao et al., 1995). V případě skandálu sice mohou být náklady mnohonásobně vyšší, ale rovněž bývají závislé na společenském postavení zúčastněných (Borelli-Kjaer, 2021) a tedy budou mít vysokou variabilitu, z důvodu čehož by bylo vhodné je zkoumat pro každou ze společenských vrstev zvlášť. Na druhou stranu, náklady jsou nezáporná veličina, tudíž započítání dodatečných položek by nikdy nemohlo vést k poklesu celkových nákladů. Z toho tedy plyne, že vyčíslené náklady, které neberou v potaz ty ostatní, jsou pouze spodní hranicí celkových očekávaných individuálních nákladů promiskuity.

Model předpokládá aditivní vztah mezi jednotlivými náklady. Jedná se o zjednodušující předpoklad, který v reálném světě nemusí být splněn. Na druhou stranu, vzhledem k relativně malé prevalenci pohlavních chorob v populaci a jejich nízké korelaci s neplánovanými početími lze všechny nelineární vztahy mezi jednotlivými položkami v nákladech považovat za nevýznamné. Základní model pro výpočet nákladů promiskuity lze tedy zapsat následovně:

$$TC_{PROM} = TC_{TEH} + TC_{NEM} + TC_{IIN} \quad (1)$$

Kde TC_{PROM} jsou celkové očekávané individuální náklady promiskuity, TC_{TEH} odpovídá nákladům souvisejícím s neplánovaným početím, TC_{NEM} označuje náklady spojené s pohlavními nemocemi a TC_{IIN} znázorňuje všechny ostatní náklady.

Je nutné zde poznamenat, že odhad celkových nákladů promiskuity se bude napříč jednotlivými rozhodovateli lišit. Dané náklady mohou být různé napříč věkovými skupinami, pohlavími, či sociálními vrstvami. Při jejich odhadu tedy záleží na informaci, kterou jsou podmíněné, přičemž v této práci jsou podmíněné nízkým věkem zúčastněných.

Co se týče všech složek celkových nákladů, je nutné vzít v potaz, že takovéto náklady budou vždy záviset na pravděpodobnosti vzniku určitého jevu a následných dílčích nákladech s ním spojených. Jedná se tedy o střední hodnotu diskrétního rozdělení mnohdy kategoriální náhodné veličiny.

Ohledně celkových nákladů spojených s neplánovaným početím platí, že primárně závisí na pravděpodobnosti, že dojde k samotnému početí a následně na pravděpodobnostech, souvisejících s jeho následky, jakož i na vyčíslení daných následků. Tedy, náklady spojené s neplánovaným početím lze zapsat následovně:

$$TC_{TEH} = P_{TEH}(P_{potr}C_{potr} + P_{dite}C_{dite}) \quad (2)$$

Kde P_{TEH} odpovídá pravděpodobnosti, že dojde k těhotenství, která je rovna průniku pravděpodobností, že dojde ke styku dvou plodných osob P_{plod} a obecné pravděpodobnosti, že důsledkem nechráněného pohlavního styku bude početí P_{poc} , tedy platí: $P_{TEH} = P_{plod} * P_{poc}$. Dále, P_{potr} je pravděpodobnost, že dojde k potratu a C_{potr} jsou očekávané náklady s ním spojené. P_{dite} je poté pravděpodobnost, že dojde k porodu, která je rovna $P_{dite} = (1 - P_{potr})$ a C_{dite} analogicky označuje očekávané náklady spojené s dítětem.

Co se týče dílčích očekávaných nákladů, platí, že C_{potr} závisí na pravděpodobnosti, s jakou dojde ke kterému typu potratu (tedy přirozený či umělý) a na konkrétních nákladech spojených s jednotlivými typy potratů. Náklady spojené se vznikem dítěte C_{dite} závisí na mnohem více faktorech, a to především na pravděpodobnostech, s nimiž se dítě dožije konkrétního věku, a následně na nákladech spojených s vyživováním dítěte. Danou veličinu by bylo možné modelovat markovským řetězcem s jediným absorbčním stavem a s možnými přechody buď o stav výše nebo do absorbčního stavu. Očekávané náklady na dítě jsou tedy následující:

$$C_{dite} = \sum_{t=1}^T P_{ziv,t} C_{ziv,t} \quad (3)$$

Kde $P_{ziv,t}$ je pravděpodobnost, že se dítě dožije konkrétního věku t a $C_{ziv,t}$ jsou očekávané náklady spojené s jeho vyživováním. V tomto kontextu je nutné uvést, že pravděpodobnost úmrtí se výrazně liší mezi novorozenci a všemi dětmi, které prožily svůj první rok života, přičemž u druhé skupiny je tato pravděpodobnost téměř zanedbatelná (pod 1 promile) a to až do věku okolo 20 let (SSA, 2022). Z toho tedy plyne, že lze pro jednoduchost uvažovat dva scénáře: dítě buď přežije stádium novorozence či nikoliv. Pokud toto stádium přežije, lze předpokládat, že se dožije až dospělosti. Rovněž je v tomto ohledu nezbytné zohlednit pravděpodobnost, že se narodí vícero dětí najednou, kterou udává Hellinovo pravidlo (Fellman & Erikson, 2009).

Co se týče vyživovacích nákladů na dítě, při jejich analýze je nutné uvažovat jak inflaci, tak úrokovou míru, využitelnou pro zjištění současné hodnoty. Současnou očekávanou hodnotu nákladů na dítě lze interpretovat jako nutný objem peněžních prostředků, jimiž subjekt musí disponovat nyní, aby zvládl dítě do budoucna zabezpečit. Tedy zjednodušující mechanismus je následující:

$$C_{dite}^{**} \cong \sum_{t=0}^T \frac{VÝŽ_t}{(1+r)^t} \cong \sum_{t=0}^T \frac{(1+\pi)^t}{(1+r)^t} VÝŽ_0 \quad (4)$$

Kde C_{dite}^{**} je současná očekávaná hodnota nákladů na dítě, $VÝŽ_t$ je výživné v čase t , které lze chápat jako funkci inflace π , přičemž je jeho hodnota redukována o mocniny výnosové míry. Intuice tohoto kroku je taková, že výživné v čase sice neustále roste, ale zároveň na něj lze spořit. Spornou veličinou dané rovnice je délka výživného, neboť je variabilní v závislosti na

tom, zdali se dítě rozhodne studovat či nikoliv a tedy se do analýzy musí započítat i tato pravděpodobnost.

Co se týče dalších nákladů spojených s nechráněným pohlavním stykem, je nutné uvažovat pravděpodobnost nákazy pohlavními nemocemi a s nimi spojené nákladné komplikace. Jedny z hlavních faktorů s nimi související jsou buď ztráta života nebo úbytek kvality života. Vhodná metrika pro měření daných efektů je parametr QALY, který je charakterizován jako hodnota jednoho roku života při výborném zdravotním stavu (Lee et al., 2009; Neumann et al., 2014). Z toho tedy plyne, že je nutné vyčíslit pravděpodobnosti, že dojde k nakažení se pohlavně přenosnou nemocí, jakož i jaký úbytek QALY vznikne v jejím důsledku. Kromě toho, s problematou nákazy pohlavními nemocemi se pojí často i ztráta zaměstnání, tedy je nutné uvažovat i náklady v důsledku menší mzdy. Celkové náklady související s pohlavně přenosnými chorobami lze vyčíslit následujícím mechanismem:

$$TC_{NEM} = \sum_{j=1}^J P_{nem,j} C_{nem,j} \quad (5)$$

Kde J označuje celkový počet pohlavně přenosných chorob, jimiž se subjekt může nakazit, $P_{nem,j}$ označuje pravděpodobnost nakažení se konkrétní nemocí a $C_{nem,j}$ jsou náklady s ní spojené. Je nutné dodat, že se lze při jednom nechráněném pohlavním styku nakazit více nemocemi naráz a tedy nákaza jednou nemocí nevyklučuje možnost nakazit se nemocí druhou. Analýza však rovněž předpokládá, že pravděpodobnost nakažení se konkrétní nemocí je nezávislá na pravděpodobnostech nakažení se ostatními nemocemi, což je zjednodušující předpoklad. Náklady pohlavní choroby se řídí následujícím mechanismem:

$$C_{nem,i} = (1 - \gamma_i)QALY + (1 - \theta_i)W \quad (6)$$

Kde γ_j označuje procentuální redukci v kvalitě života vzhledem ke konkrétní nemoci, θ_j znázorňuje procentuální redukci ve mzdě W .

3 VÝSLEDKY

V pilotní studii (Krautwurm, 2019) bylo zjištěno, že očekávané individuální náklady promiskuity, tedy jednoho nechráněného pohlavního styku, se pohybují okolo 19 238 Kč. Hlavní položkou v těchto nákladech je možnost početí dítěte, jehož nákladnost se pohybuje okolo dvou milionů (ČSÚ, 2004). Ačkoliv cena života je dle QALY poměrně vysoká (cf. Neumann et al., 2014), nízká pravděpodobnost nakažení se pohlavními nemocemi a následně relativně dobrá lékařská péče, pomocí níž je možné i v takovém případě velkou část kvality života zachovat, způsobuje, že tento aspekt nechráněného pohlavního styku není příliš významný.

V dalším výzkumu je nutné aplikovat současný postup na nová data a případně komparovat jednotlivé výsledky napříč různými skupinami a vzhledem k jiným podmiňovacím informacím.

ZÁVĚR

Práce se zabývá očekávanými individuálními náklady promiskuity, přičemž tu zobecňuje do jedné nechráněné soulože s neznámou osobou. Práce dokládá mechanismus, s nímž lze dané náklady získat a vyčísluje je na 19 238 Kč. Hlavní nákladná položka je spojena s možností početí dítěte. Z toho tedy plyne, že ochrana dramaticky snižuje nákladnost nechráněného pohlavního styku a bylo by možné tedy predikovat, že v zemích s levnou antikoncepcí bude nechráněného styku více (cf. Houdek & Kobloušský 2017).

Použitá literatura

1. ABASIATTAI, AM, ETUK, SJ, BASSEY, EA, ASUQUO, EE. Vaginal injuries during coitus in Calabar: a 10-year review. *The Nigerian Postgraduate Medical Journal*. 2005 Jun;12(2):140-144. PMID: 15997266.
2. BORELLI-KJAER, M., MOEHL S. L., & NIELSSON, U. MeToo: Sexual harassment and company value. *Journal of Corporate Finance*, 2021 67, 101875. <https://doi.org/10.1016/j.jcorpfin.2020.101875>
3. ČESKÝ STATISTICKÝ ÚŘAD. Náklady na výchovu a výživu dětí – 2003. Prosinec, 2004. [online] Dostupné z: <https://www.czso.cz/csu/czso/naklady-na-vychovu-a-vyzivu-deti-2003-n-a6vxfdndb0> [cit. 17. 3. 2019]
4. DAO, B, DIOUF, A, BAMBARA, M, BAH MD, DIADHIOU, F. Vaginal injuries during coitus: 98 cases. *Contraception, Fertilité, Sexualité* (1992). 1995 Jun;23(6):420-422. PMID: 7627286
5. FELLMAN, J., ERIKSSON, A. W. Statistical Analyses of Hellin's Law. *Twin Research and Human Genetics*. April, 2009. 12(02). 191–200. doi:10.1375/twin.12.2.191
6. HOUDEK, Petr, KOBLOVSKÝ, P, Economics of Sex: Cost-Benefit Analysis. *Society*. 54(1). 2017. 18-22. DOI: <https://doi.org/10.1007/s12115-016-0089-2>
7. KRAUTWURM, Petr, Analýza individuálních nákladů dvacetiletých na nechráněný pohlavní styk za rok 2019. Seminární práce v předmětu Bakalářský seminář. Vysoká škola Ekonomická. Březen, 2019.
8. LEE, Chris P., CHERTOW, Glenn M., ZENIOS, Stefanos A. An Empiric Estimate of the Value of Life: Updating the Renal Dialysis Cost-Effectiveness Standard. *Value in Health*. 2009. 12(1), 80–87. doi:10.1111/j.1524-4733.2008.00401.x
9. NEUMANN, Peter J., COHEN, Joshua T., WEINSTEIN, Milton C. Updating Cost-Effectiveness — The Curious Resilience of the \$50,000-per-QALY Threshold. *New England Journal of Medicine*. 2014. 371(9), 796–797. doi:10.1056/nejmp1405158
10. THE UNITED STATES SOCIAL SECURITY ADMINISTRATION. Actuarial life table. 2022. [online] Dostupné z: <https://www.ssa.gov/oact/STATS/table4c6.html> [cit. 13. 11. 2022]

Kontaktné údaje

Ing. Petr Krautwurm

Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky
nám. Winstona Churchilla, 1938/4

email: petr.krautwurm@vse.cz

Článek vznikl s podporou projektu VŠE IGS F4/38/2022 Interní grantové agentury Vysoké školy ekonomické v Praze

ŘEŠENÍ PROBLÉMU OBCHODNÍHO CESTUJÍCÍHO V KONTEXTU PROGRAMOVACÍHO JAZYKA R, BALÍČKU {SF} & HERE API

SOLVING TRAVELING SALESMAN PROBLEM USING R IN CONTEXT OF {SF} & HERE API

Jindřich Lacko

Abstrakt

Problém obchodního cestujícího je jedním z nejstarších a teoreticky nejlépe zpracovaných témat v oblasti optimalizačních úloh, s vysokou mírou standardizace řešení. Příspěvek ilustruje možnosti praktického využití jednoho z nástrojů - rozšiřujícího balíčku {TSP} - v kontextu statistického programovacího jazyka R. Dále demonstruje využití tří alternativních metrik pro optimalizaci: vzdálenosti vzdušnou čarou, dojezdovou dráhu autem a čas cesty autem.

Klíčové slova: *optimalizace, TSP, R, HERE API*

Abstract

The traveling salesman problem is one of the oldest and theoretically best understood problems in optimization techniques. From theoretical point of view the approaches to its solution are highly standardized. This article demonstrates several empirically advantageous approaches to solving Traveling Salesman Problem within context of the statistical programming language R. It further demonstrates three alternative metrics for optimization: Euclidean distance, travel distance by car and travel time by car.

Keywords: *optimization, TSP, R, HERE API*

1. PROBLEM STATEMENT

The Traveling Salesman Problem (TSP) is an optimization classic, with a number of well understood and highly standardized solutions available in the context of statistical programming language R.

The problem can be mathematically formulated, following (1), as: Given a weighted complete digraph (K_n, c) , find a Hamiltonian cycle in K_n of minimum cost. The cost function c can be determined by the cost matrix $[c_{ij}]$. The cost – or distance – matrix thus plays a key role in defining, and solving, the TSP.

In this article I would like to share a practical example of solving the TSP using Open Street Map (2) data of bars via {osmdata} (3) and routing engine provided by HERE Global B.V. Kennedyplein 222-226, 5611 ZT Eindhoven, Netherlands as an Application Programming Interface (API) service and accessed from R via {hereR} (4). The actual solution will be found by utilizing the {TSP} (5) package.

The Prague School of Economics, is located in Žižkov. A formerly working class neighborhood, now rather gentrified, it has to this day retained some traces of its former rougher edges. One of these is an active night life.

A visit to the bars of Žižkov is therefore a familiar activity for many VŠE students, and can serve as an introduction to serious optimization techniques.

2. DATA ACQUISITION

The first step in our exercise is acquiring data of Žižkov bars. A search is performed over the area of *core Žižkov*, defined as a polygon, using the OSM Overpass API. As there seems not to be a clear consensus over what constitutes a bar, restaurant or a pub in Prague I am including all three of the possible amenities.

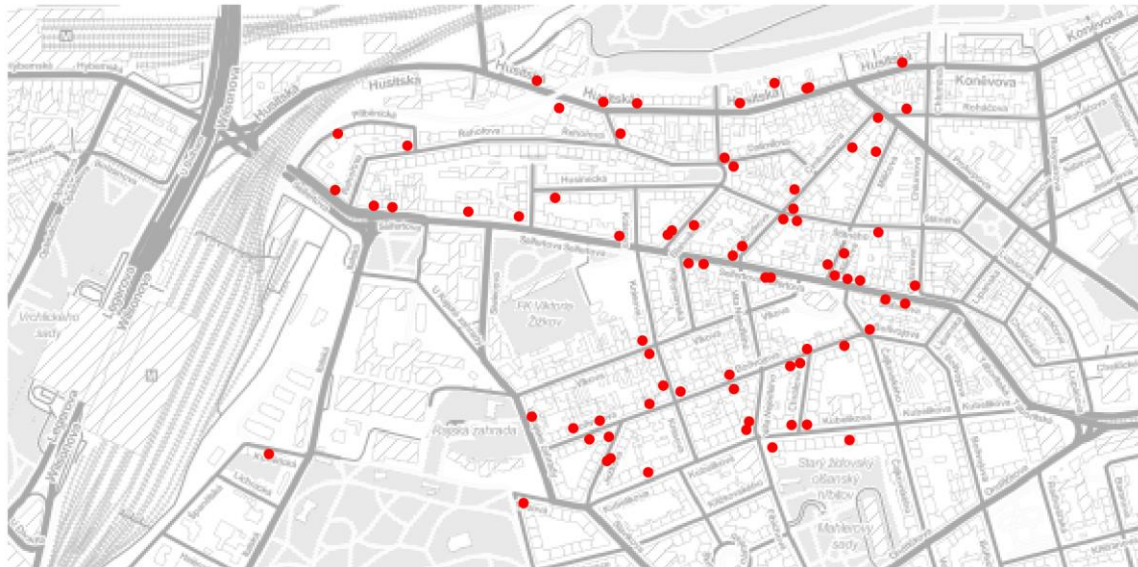


Figure 1: Map of Žižkov with full set ($n = 75$) of bars.

Source: OpenStreetMap + authors own work

We have located 75 bars, implying a cost matrix of 5625 elements. Not a huge one by today's standards – but big enough to think twice about trying to solve using a pen and a piece of paper.

3. EUCLIDEAN DISTANCE TSP

The easiest cost matrix to calculate is plain “as the crow flies” distance.

The resulting matrix (Table 1) will be based on pure distance, with some differences in interpretation depending on coordinate reference system of underlying data (Euclidean (6) in projected CRS and spherical in unprojected CRS).

Calculating the cost matrix using plain distance is easy, and the cost matrix is symmetrical [c_{ij}] equals [c_{ji}]. The problem can be formulated such as the K_n being considered is an undirected graph.

Solving the TSP for such a matrix is straightforward, as optimization toolset required is readily implemented in the {TSP} package.

The optimal route cost is 1697 and the metric minimized is meters of length.

From the visual overview (Figure 2) we can see an obvious shortcoming of the “as the crow flies” approach: it completely ignores other constraints except for distance – such as the layout of a road network.

Thus while the route shown is “optimal” in the sense that it forms the shortest path joining the five bars selected, it is not one that we could actually follow (unless we were a flying crow).

This shortcoming can be resolved by using an alternative cost matrix as input, while retaining the techniques of {TSP} for the actual route selection. A possible source of more applicable data are routing engines, available to R users via API interfacing packages.

The routing results give us several pieces of data:

- the routes as linestring objects in WGS84 (7) (for visualization later on)
- distance of the route (in meters)
- travel time (in seconds) both raw and adjusted for traffic
- fuel consumption (both combustion type and EV)

Having a variety of metrics will be helpful in construction of alternative distance matrices that can be applied under different empirical conditions.

4. TRAVEL DISTANCE TSP

The next cost matrix will be based on route distance; notice that the matrix (Table 1) is not symmetrical. This is not surprising, as routing is not commutative – optimal route from A to B need not be the same as from B to A, given constraints such as one way roads. Žižkov is a veritable warren of one way streets. In this context the K_n being considered is a directed graph, unlike what was the case for the Euclidean distance.

It will need to be declared as asymmetrical to {TSP} solver; other than that the actual process of solving the matrix will be analogical to the “as the crow flies” matrix.

The optimal route cost is 3569 and the metric minimized is meters of length.

Since the HERE API is generous in terms of results provided it is not difficult to construct an alternative cost matrix, using a different metric. This could be either trip duration or petrol consumption.

In our specific situation both of these can be expected to be highly correlated with the plain distance results. All the streets in Žižkov are of very similar type, and both the average speed and petrol consumption are unlikely to vary greatly between the routes.

The most significant difference between the distance and time based TSP will be driven by current traffic, which is a factor HERE routing engine considers.

5. TRAVEL TIME TSP

Finally one can focus on travel time optimization – a task relevant for emergency services / first responders as well as commercial fast delivery services.

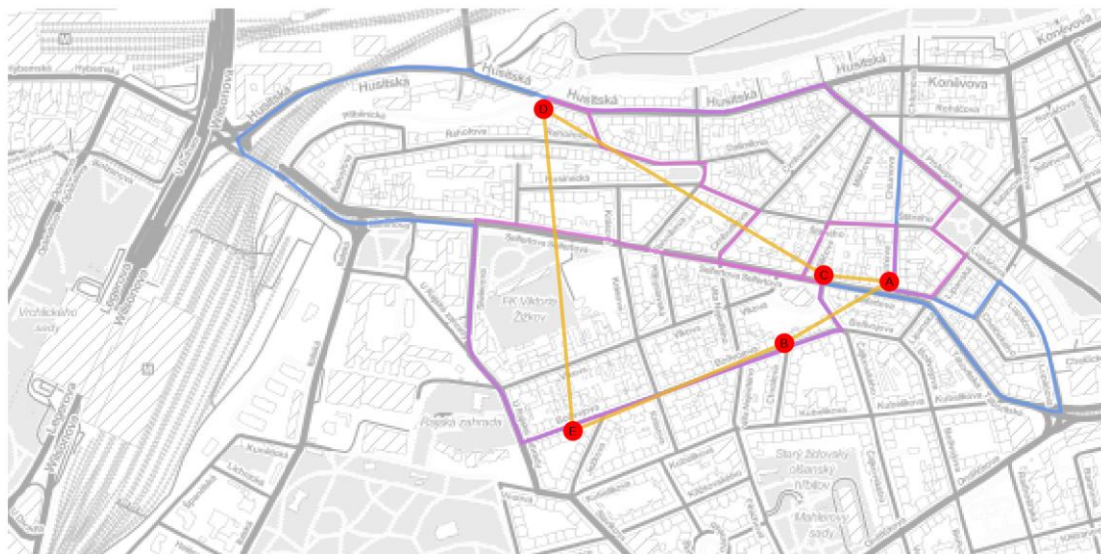
The optimal route cost is 576 and the metric minimized is seconds of time.

The actual route is in our use case very similar to the distance based one. This will not necessarily be the case in other contexts, especially ones with greater variation of road types (city streets vs. highways).

6. CONCLUSION

This article demonstrates relevant empirical advantages to be gained from using:

- a standardized solution (the {TSP} package) to a well known and well understood problem (the Traveling Salesman Problem) within the context of R ecosystem
- construction of distance matrices from HERE API routing results, with option to optimize for multiple metrics (such as minimizing the travel distance, travel time and fuel consumption)



TSP optimized for: — Euclidean distance — Route length — Route time

Figure 2: Map of Žižkov with sample ($n = 5$) of bars and connecting routes.

Source: OpenStreetMap + authors own work

Table 1: Cost Matrices
Euclidean Distance Cost Matrix

	A	B	C	D	E
A	0	194	105	616	558
B	194	0	125	535	365
C	105	125	0	519	471
D	616	535	519	0	514
E	558	365	471	514	0

<i>Route Length Cost Matrix</i>						<i>Route Time Cost Matrix</i>					
	A	B	C	D	E		A	B	C	D	E
A	0	829	329	821	1286	A	0	160	84	137	206
B	1096	0	183	893	951	B	189	0	37	196	135
C	913	500	0	710	1023	C	144	74	0	139	144
D	1027	1352	852	0	1412	D	150	212	136	0	212
E	1299	363	546	1655	0	E	220	46	83	248	0

References

1. GUTIN, G. and PUNNEN, A. P. *The Traveling Salesman Problem and Its Variations*. Springer Science & Business Media, 2006. ISBN 978-0-306-48213-7.
2. OPENSTREETMAP CONTRIBUTORS. *Planet dump retrieved from <https://planet.osm.org>*. Online. 2017. Available from: <https://www.openstreetmap.org/>
3. PADGHAM, Mark, LOVELACE, Robin, SALMON, Maëlle and RUDIS, Bob. *osmdata. Journal of Open Source Software*. Online. 26 June 2017. Vol. 2, no. 14, p. 305. [Accessed 7 October 2022]. DOI [10.21105/joss.00305](https://doi.org/10.21105/joss.00305).
4. UNTERFINGER, Merlin and POSSENRIEDE, Daniel. *hereR: “sf”- Based Interface to the “HERE” REST APIs*. Online. 0.9.0. R. 5 August 2022. [Accessed 7 October 2022]. Available from: <https://CRAN.R-project.org/package=hereR>
5. HAHSLER, Michael and HORNIK, Kurt. TSP—Infrastructure for the Traveling Salesperson Problem. *Journal of Statistical Software*. Online. 2008. Vol. 23, p. 1–21. [Accessed 7 October 2022]. DOI [10.18637/jss.v023.i02](https://doi.org/10.18637/jss.v023.i02).
6. HARTSHORNE, Robin. *Geometry: Euclid and Beyond*. Springer Science & Business Media, 2013. ISBN 978-0-387-22676-7.
7. DMA TR 8350.2: *Department of Defense World Geodetic System 1984, Its Definition and Relationships With Local Geodetic Systems*. Defense Mapping Agency Systems Center 8613 Lee Highway, Fairfax VA 22031-2138, 1991.

Contact

University of Economics in Prague, Faculty of Informatics and Statistics, Department of
Econometrics

W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic

E-mail: jindra.lacko@vse.cz

VOLATILITA INDEXU S&P500 V DOBE ZVYŠUJÍCICH SA ÚROKOVÝCH SADZIEB

VOLATILITY OF S&P500 IN THE PERIOD OF INCREASING INTEREST RATES

Radoslav Lacko

Abstrakt

Všetky neočakávané situácie na burzových trhoch majú svojich víťazov a porazených. V posledných troch rokoch bolo niekoľko nepredpovedateľných situácií, napr. vojna, obdobie vysokej inflácie a pandémie. Príspevok rozoberá a odhaduje volatilitu počas nestálych časov na finančných trhoch.

Kľúčové slová: volatilita, Garch, S&P500

Abstract

All unexpected situations on the stock markets has winners and losers. There were unpredictable situation in the last three year, such as war, high inflation and pandemic. The article examine and estimate the volatility fluctuations during the unstable period on financial markets in that time.

Keywords: volatility, Garch, S&P500

Acknowledgements The work on this paper was supported by the grant No. F4/27/2020 of the Internal Grant Agency of University of Economics, Prague

1 INTRODUCTION

It is more than decade since the last economic crisis. A lot of economists are predicting that it may happen soon and we are not far from recession. The last three years were so turbulent e.g. with a lot of financial injections by central banks. The aim of the article is to estimate volatility and evaluate its changes as a response to actions of central banks, rising inflation and war on the Ukraine.

2 DATA

As the topic says, the American stock index S&P 500 daily data are employed for the volatility investigation. The examined period of one year starts on 2017/11/13 and ends on 2022/11/10. The daily data are obtained from Yahoo Finance website. The Figure 1 shows how index moved during the selected period.

The figure 2 shows daily returns of index prices as the result of the first difference of logarithm. This process made data stationary which is confirmed by ACF and PACT test visualized in the Figure 3.

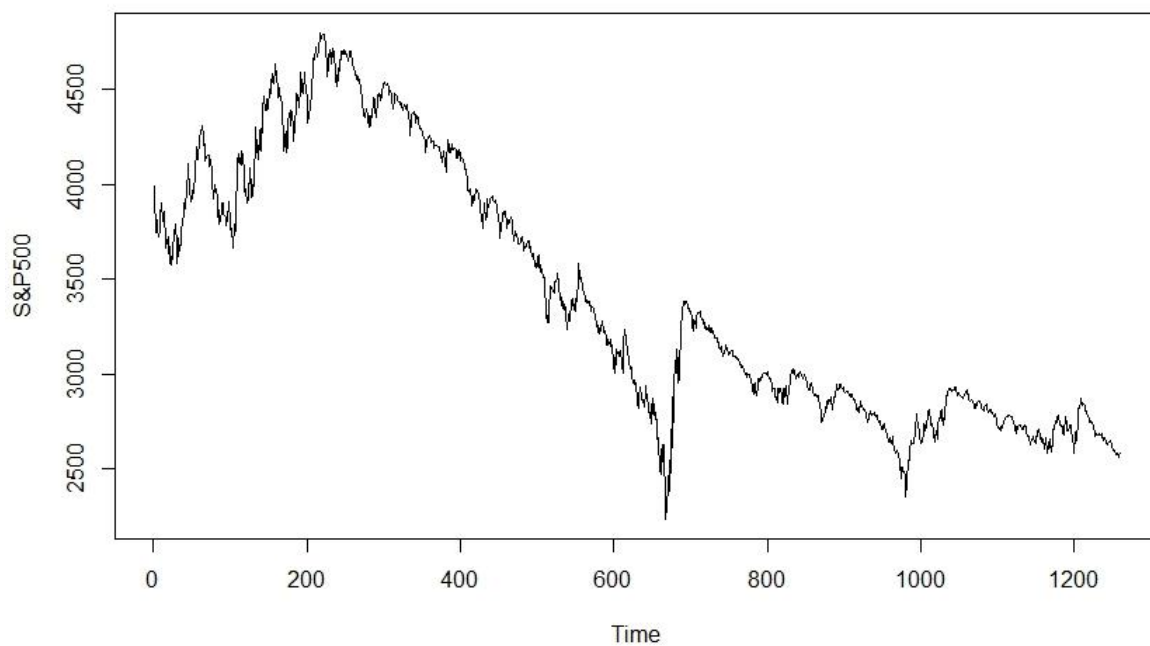


Figure 1 S&P 500 stock index during the examined period

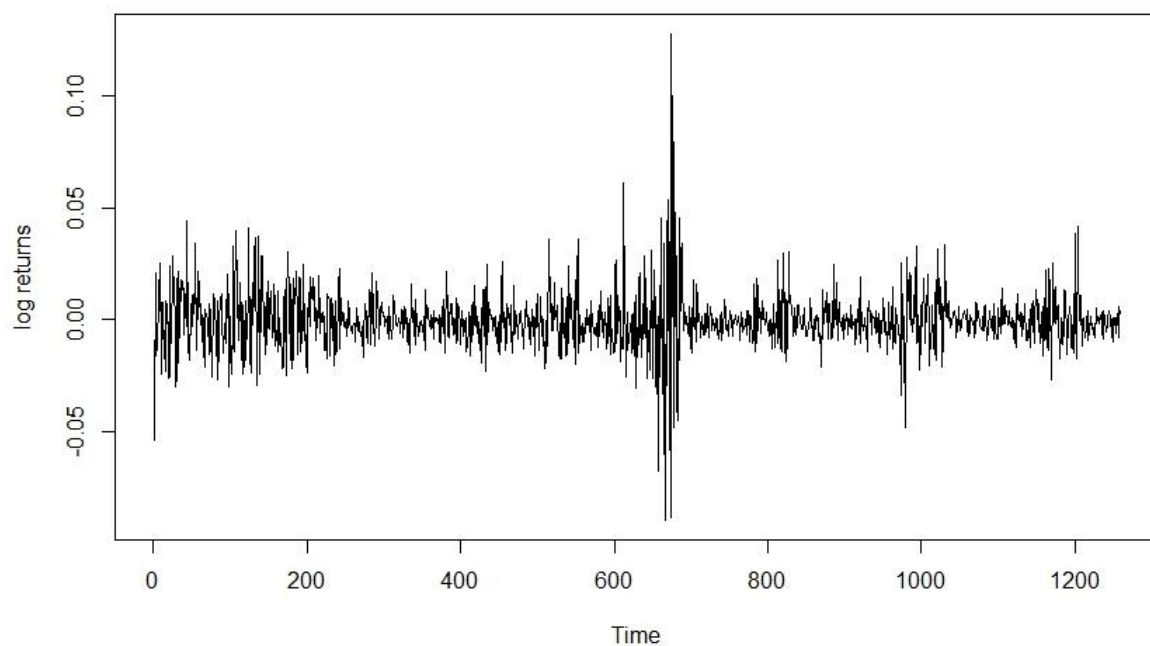


Figure 2 S&P 500 stock index daily return during the examined period

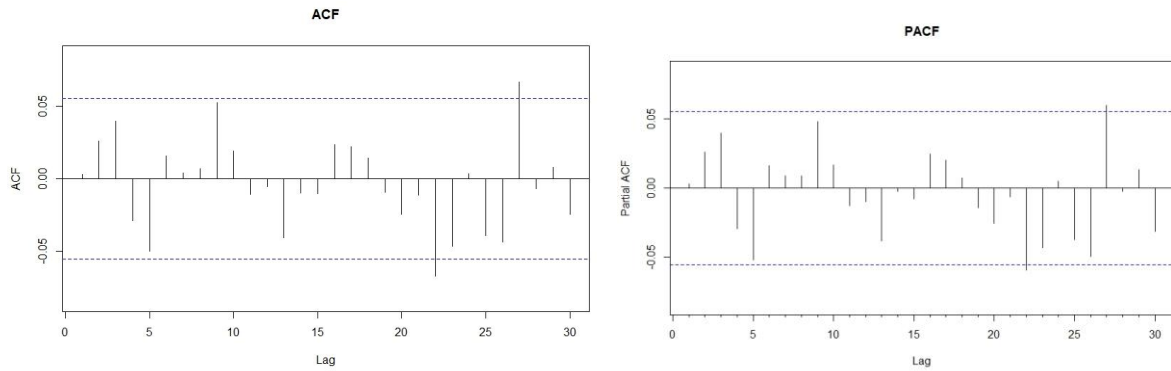


Figure 3 ACF and PACF tests

3 METHODOLOGY

The main source of information presented in the methodology part is the book *Introductory Econometrics for Finance* by Chris Brook (2019)[3].

First of all there is necessary to test for ARCH effects and then build volatility (GARCH type) models. The test employed by Brook can be formulated in several steps.

1. Run any postulated linear regression and save the residuals u_t .

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$$
2. Square the residual, and regress them on q own lags to test for ARCH of order q and obtain the R^2 .

$$\hat{u}_t^2 = \gamma_0 + \gamma_1 \hat{u}_{t-1}^2 + \dots + \gamma_q \hat{u}_{t-q}^2 + v_t$$
3. The test statistic is defined as TR_2 , which is the number of observations multiplied by the coefficient of multiple correlation, from the last regression.
4. The null and alternative hypotheses are
 $H_0 : \gamma_1 = 0 \text{ and } \gamma_2 = 0 \text{ and } \gamma_3 = 0 \text{ and } \dots \text{ and } \gamma_q = 0$
 $H_1 : \gamma_1 \neq 0 \text{ and } \gamma_2 \neq 0 \text{ and } \gamma_3 \neq 0$
 $\text{and } \dots \text{ and } \gamma_q \neq 0$

Based on the hypothesis results we may build the model. The most common ARCH family model for volatility modeling is Generalized ARCH (GARCH) model that allows the conditional variance to be dependent upon previous own lags, with the simplest version GARCH(1,1):

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (3)$$

Where $\alpha_1 u_{t-1}^2$ is information about volatility during the previous period and $\beta \sigma_{t-1}^2$ is the fitted variance from the model during the previous period.

4 RESULTS

First of all, it is necessary to mention that the ARCH-LM test rejected the null hypothesis. Hence, the conditional heteroskedasticity is presented in residuals. Moreover, we confirmed our findings by running the Ljung-Box test. GARCH was the most suitable model and according to the optimal information criteria we have chosen GARCH (1,1). Both variables (β and α) were significant. The figure 4 shows estimated volatility for S&P 500. The times of high inflation and falling markets are not the period of the highest estimated volatility, but the

corona crisis is. The peak during the pandemic was estimated at 6.2 % but in the times of high inflation and war it has not exceeded more than 3.0 %.

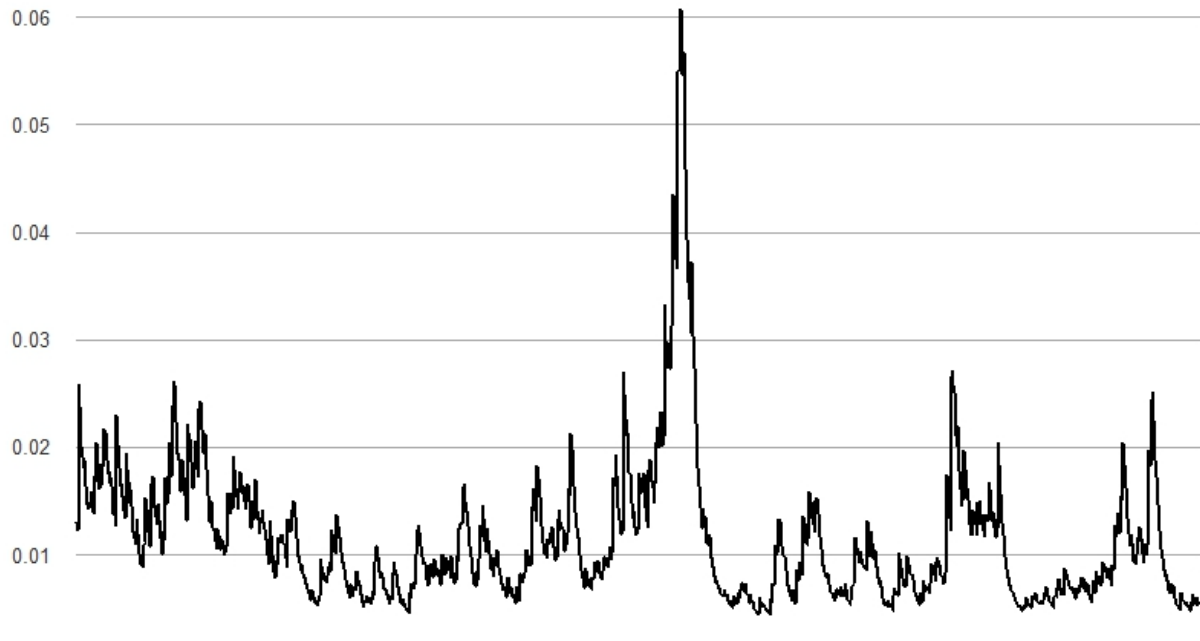


Figure 4 Estimated volatility of S&P 500 in the examined period.

Literature

1. Adrian T. & Rosenberg J. (2006). Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk. SSRN Electronic Journal.
2. Baillie R. & DeGennaro R. (1996). Stock Returns and Volatility. *The Journal of Financial and Quantitative Analysis* 25.2.
3. Brooks Ch. (2019). *Introductory Econometrics for Finance*. Cambridge University Press.
4. French K., Schwert G. & Robert F. Stambaugh. (1987). Expected stock returns and volatility. *Journal of Financial Economics* 19.1.
5. Karolyi A. (1995). A Multivariate GARCH Model of International Transmissions of Stock Returns and Volatility: The Case of the United States and Canada. *Journal of Business & Economic Statistics* 13.1

Kontakné údaje

Radoslav Lacko

Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky
nám. Winstona Churchilla 1938/4, 130 67 Praha 3-Žižkov

Tel: +421 948 321 340

Email: rl.lacko@gmail.com

REPLIKOVATEĽNOSŤ VÝSKUMU A JEHO PUBLIKOVATEĽNOSŤ¹

REPLICABILITY OF RESEARCH AND ITS PUBLICABILITY

Martin Lukáčik, Adriana Lukáčiková, Karol Szomolányi

Abstrakt

Replikovateľnosť výskumu je jedným z jeho kľúčových aspektov. Vďaka nej sa môžu poznatky rýchlejšie šíriť medzi ďalšie a ďalšie skupiny záujemcov, ktoré môžu pokračovať v začatom výskume a rozširovať poznanie o nové aspekty. Takisto sa vďaka replikovateľnosti výskumu ľahšie odhaľujú chyby a nedostatky pri používaní zložitých vedeckých metód. V neposlednom rade umožňuje učiteľom v danom odbore prezentovať svojim študentom aktuálne problémy a postupy používané pri ich riešení. Táto doména kvalitných vedeckých časopisov, ktoré vyžadujú poskytnutie kompletnej údajovej základne a programové riešenie vedúce k všetkým publikovaným výsledkom, sa dá v súčasnosti jednoducho uchopiť pomocou nástrojov, ktoré vznikli ako podpora existujúcich programových jazykov. V tomto príspevku sa zameriavame na R Markdown ako nástroj podpory analýz a ich publikovateľnosti v jazyku R.

KLúčové slová: *replikovateľnosť, podpora analýz, publikovateľnosť, R Markdown*

Abstract

The replicability of research is one of its key aspects. Thanks to it, knowledge can be spread more quickly among more and more groups of interests, which can continue the research that has started and expand knowledge with new aspects. Also, the replicability of research makes it easier to detect errors and flaws in the use of complex scientific methods. Last but not least, it allows teachers in the given field to present their students with current problems and procedures used in solving them. This domain of quality scientific journals, which require the giving of a complete database and a programming solution leading to all published results, can now be easily taken over using tools that were created to support existing programming languages. In this paper, we focus on R Markdown as a tool to support analytics in R language.

Keywords: *replicability, support analytics, publicability, R Markdown*

1 ÚVOD

Ako na svojich stránkach konštatujú editori známeho vedeckého časopisu Journal of Applied Econometrics (The JAE Replication Section, 2022): „Replikovateľnosť je dôležitou súčasťou každej vedy, hoci to, čo sa rozumie pod pojmom replikácia, sa v rôznych odboroch značne líši. V našom časopise sme vždy požadovali od autorov, aby poskytli svoje dátové súbory a akékoľvek špecializované počítačové programy ako podmienku zverejnenia, s cieľom uľahčiť replikáciu výsledkov, ktoré publikujeme.“ Dôvod pre takéto rozhodnutie renomovaného časopisu je zrejmy. Overenie postupu analýzy širokým auditóriom čitateľov a zároveň prezentovanie výskumu, ktoré bude oveľa ľahšie uchopiteľné a pravdepodobne takisto citované je v ich eminentnom záujme. Rovnako by mali uvažovať aj editori časopisov doposiaľ nezverejňujúcich údaje a programy vedúce k celej publikovanej analýze.

¹ Príspevok vznikol s podporou projektov VEGA č. 1/0211/21, Ekonomická analýza makroekonomických dopadov vplyvu pandémie vo svete s dôrazom na vývoj ekonomík EÚ a zvlášť ekonomiky Slovenska a VEGA 1/0193/20, Vplyv priestorových spillover efektov na inovačné aktivity a rozvoj regiónov EÚ.

Replikovateľnosť výskumu umožňuje učiteľom v každom odbore, a teda aj v ekonometrii, ktorou sa zaoberajú autori tohto príspevku, prezentovať svojim študentom aktuálne problémy a postupy používané pri ich riešení. Tým je skvele zabezpečená pedagogika metodológie. Aj keď zdanlivo na túto úlohu stačí napísať skript v ľubovoľnom softvéri, pre lepšie uchopenie problematiky je vhodné, keď je doplnený o vysvetlenia, motivácie a rozhodnutia pre dané kroky, ktoré v kontexte čistého programového riešenia môžu neskúsenému čitateľovi zanikať.

V súčasnosti existuje viacero nástrojov podporujúcich tieto činnosti. Pre používateľov programovacieho jazyka Python je to napríklad Jupyter Notebook. Jupyter je webové interaktívne vývojové prostredie, ktoré používateľom umožňuje vytvárať výsledný súbor tzv. *notebook* na dokumentovanie svojej práce spolu s kódmi a ich výstupmi. Toto prostredie si zvolili napríklad tvorcovia ekonometrického programu EViews, keď v jeho aktuálnej verzii 13 sa tento dá použiť ako jadro pre Jupyter. To znamená, že môžete použiť Jupyter Notebook na spustenie a organizáciu programu EViews a zobrazenie jeho výsledkov v Jupyter Notebooku (EViews 13 – Jupyter Notebook Support. 2022).

Obdobným nástrojom pre používateľov programovacieho jazyka R sú súbory R Markdown podporované softvérovým rozhraním RStudio. R Markdown je jednoduchý a ľahko použiteľný nástroj pracujúci s obvyčajným textom, ktorý sa používa na spojenie vášho R kódu, výsledkov analýzy údajov a písaného komentára do jedného naformátovaného a reprodukovateľného dokumentu alebo webovej stránky. R Markdown predstavuje variant jazyka Markdown a oba patria do skupiny značkovacích jazykov. Napríklad štatistický softvér STATA ponúka okrem využitia Jupyter Notebooku aj možnosti Markdownu prostredníctvom príkazu *markstat* (Rodríguez, 2017).

2 NÁSTROJ R MARKDOWN

Programovacím jazykom R ktorý je voľne dostupný, sme sa zaoberali vo viacerých príspevkoch (Lukáčik, 2016 a 2019; Lukáčik a Lukáčiková, 2018; Lukáčik, Lukáčiková a Szomolányi, 2016 a 2019). RStudio (rovnako ako aj R bez neho) využíva dva základné spôsoby zadávania príkazov, a to buď pomocou konzoly (*Console*), v ktorej sa po zadaní *Enter* na konci príkazu, tento ihneď vykoná alebo pomocou tzv. *scriptu* teda sekvencie príkazov, ktorá sa ukladá do súborov s príponou (.R). To znamená, že reprodukovateľnosť je zabezpečená aj obvyčajným spustením *scriptu*, ktorý obsahuje celý postup analýzy. Ale využitím jazyka Markdown (po nainštalovaní zodpovedajúceho balíčka *rmarkdown*) sa celý výstup môže rozšíriť o komentáre a usporiadať do elegantného výstupu, v ktorom sa dokáže zorientovať každý záujemca o danú problematiku.

Formát dokumentu R Markdown sa prvýkrát objavil v balíčku *knitr* (Xie, 2015) v roku 2012. Myšlienkou bolo vložiť kúsky kódu (R aj iných jazykov) do dokumentov Markdown. Balíček *knitr* od začiatku podporoval okrem Markdownu aj iné jazyky. Medzi všetkými podporovanými formátmi dokumentov vynikol pre svoju jednoduchosť Markdown. Pôvodná verzia Markdown, sa však považovala za príliš jednoduchú a nevhodnú na písanie technických dokumentov. Ale neskôr bol vytvorený balíček *pandoc* (Pandoc, 2022) na konverziu dokumentov Markdown (a mnohých iných typov dokumentov) do širokej škály výstupných formátov. R Markdown stojí na balíčkoch *knitr* a *pandoc*.



Obrázok 1: Proces v pozadí tvorby výstupných formátov z R Markdown súboru

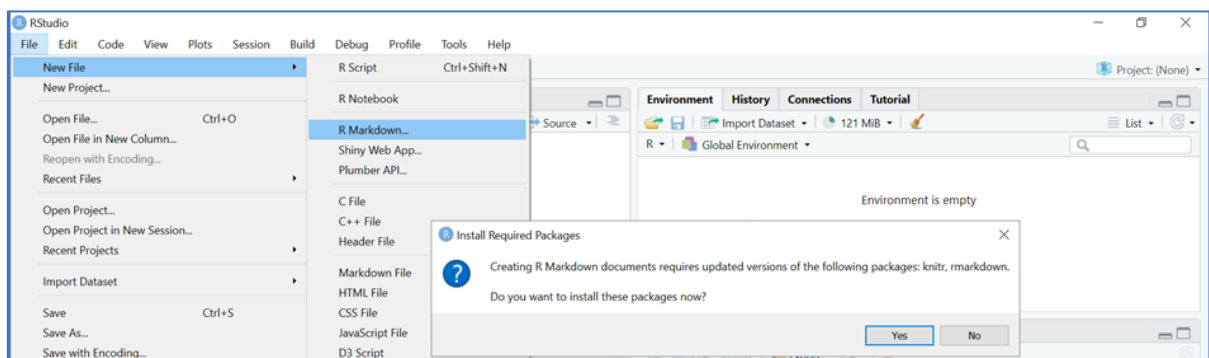
Balíček *knitr* spustí počítačový kód vložený do Markdownu a skonvertuje R Markdown (Rmd) na súbor typu Markdown (md). Balíček *pandoc* vykreslí Markdown do požadovaného výstupného formátu (napríklad pdf, html, docx a rôzne iné).

Balíček *rmarkdown* (Allaire, Xie, McPherson, et al. 2022) bol prvýkrát vytvorený začiatkom roka 2014. V súčasnosti sa vyvinul do kompletného systému na tvorbu dokumentov, umožňujúceho riešiť množstvo úloh. Po vytvorení jedného dokumentu R Markdown sa dá vytvoriť výsledná zostava v rôznych formátoch, ako je *pdf*, *html* alebo *docx*. Dajú sa vytvoriť poznámkové bloky, v ktorých sa interaktívne spúšťajú časti kódu, dajú sa tvoriť snímky pre prezentácie v rôznych formátoch, ako je *html5*, *LaTeX Beamer* alebo *PowerPoint*. Môžeme vytvárať informačné panely s flexibilným a interaktívnym výstupom. Dajú sa tvoriť interaktívne aplikácie založené na *Shiny* (Shiny, 2022), priamo písať články do časopisov alebo knihy s viacerými kapitolami (Bookdown, 2022) a jednoducho tvoriť webové stránky a blogy.

Tvorcovia R Markdownu nevytvorili nový editor so všetkými známymi nastaveniami textového alebo grafického editora. Pri výbere funkcionality ponúkli iba základné nástroje: odseky, hlavičky (sekcii), blokové citácie, bloky kódu, číslované a nečíslované zoznamy, tabuľky, vložené formátovanie ako zvýraznenie, prečiarknutie, horné a dolné indexy, doslovný text a text s veľkými písmenami, matematické výrazy cez LaTeX, rovnice, odkazy, obrázky, poznámky pod čiarou, citácie a linky, ktoré by mali postačovať na tvorbu zaujímavých a dostatočne personifikovaných výstupov analýz vlastného výskumu.

3 TVORBA R MARKDOWN SÚBORU

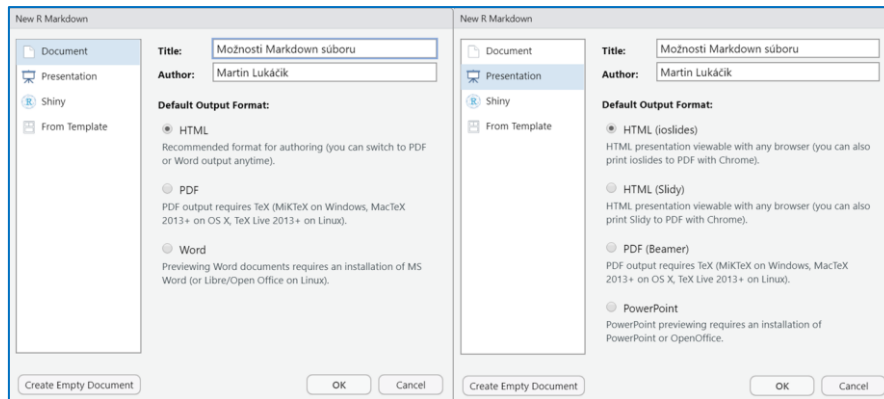
Balíček R Markdown môžeme nainštalovať pomocou štandardného postupu inštalácie balíčkov v R, teda `install.packages("rmarkdown")`. Dokonca ho nie je potrebné ani inštalovať, ale po začatí prvého vytvárania tohto typu dokumentu v prostredí RStudio sa ponúkne inštalácia tohto balíčka (obvykle spolu s balíčkom *knitr*).



Obrázok 2: Spustenie tvorby R Markdown súboru v programe R Studio

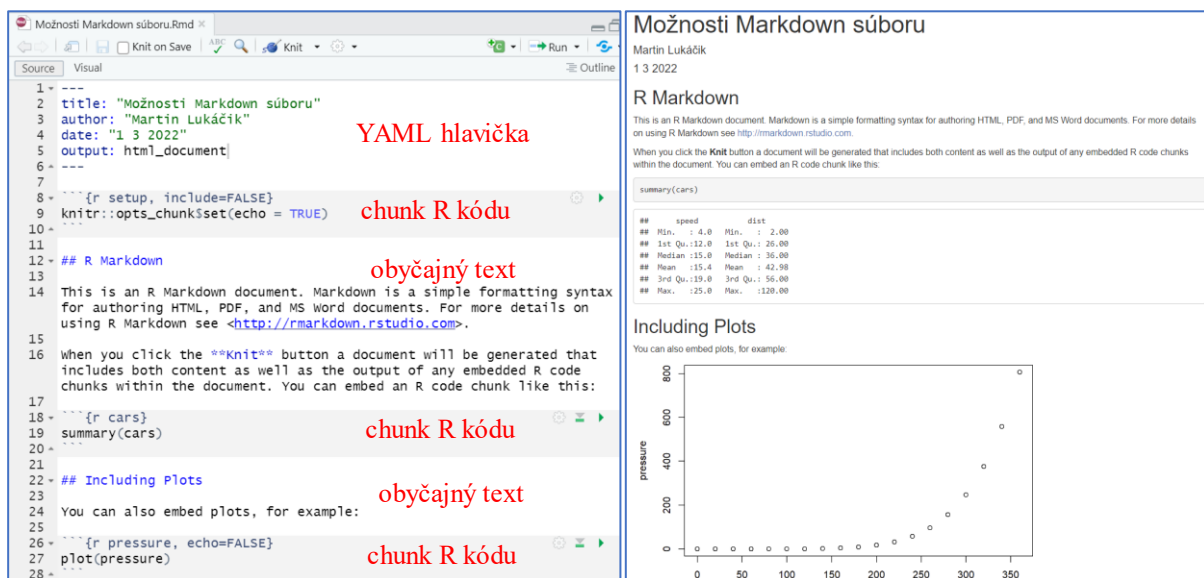
Okno vytvorenia nového R Markdown súboru umožňuje výber typu výsledného: dokumentu (html, pdf, docx, odt, rtf); knihy; prezentácie (html, pdf, powerpoint); interaktívneho *shiny* dokumentu alebo prezentácie pomocou existujúcich šablón (napr. *prettydoc*, ktoré získame v rovnomennom balíčku) pomocou jednoduchej preddefinovanej ukážky, alebo umožňuje vytvoriť prázdny dokument, ktorý vytvárame od začiatku samostatne.

Ak zvolíme prvú možnosť vytvorenie dokumentu typu html (a zadáme nadpis dokumentu a meno autora), RStudio vytvorí jednoduchý ukázkový R Markdown súbor. V hornej časti tvorí začiatok dokumentu tzv. YAML hlavička vymedzená 3 pomlčkami na začiatku a na konci. Obsahuje základné informácie ako nadpis dokumentu a výstup (output), prípadne ďalšie informácie ako autor, dátum, obsah a podobne. Zadané informácie o výstupe určujú, aký typ dokumentu bude vytvorený a s akými vlastnosťami po potvrdení výberu ikonou *Knit*.



Obrázok 3: Výber výsledného výstupu pri začatí tvorby R Markdown v programe R Studio

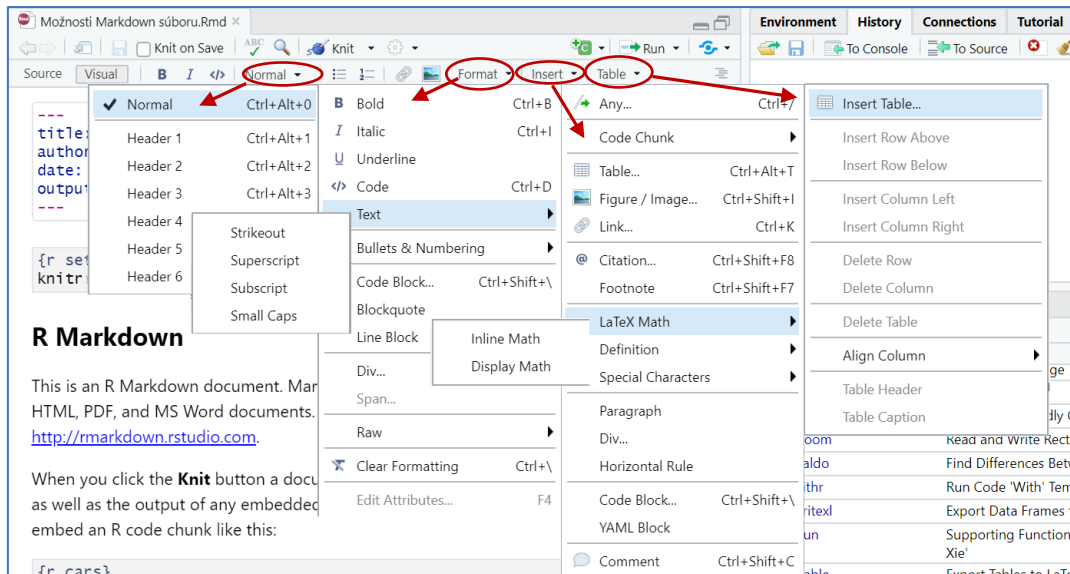
Na ľavej strane obrázku 4 vidíme v zdroji šedým podkladom zvýraznené kúsky kódu R jazyku tzv. *chunks*. Sú vymedzené tromi spätnými apostrofmami. Za nimi nasledujú zložené zátvorky, kde je názov jazyka, ktorého kód nasleduje. V našom prípade je to `{r}`, prípadne sa dajú vkladať kúsky kódu jazyka Python, SQL, Stan alebo iných. Voľbou nastavovacích parametrov (*include*, *echo*, atď.) sa dá ovplyvniť, či sú kúsky kódu viditeľné vo výstupe (ako v prípade príkazu `summary(cars)`) alebo ich tam nevidíme (ako v prípade zvyšných chunkov v ukážke). Vo výstupnom súbore sa pod kúskami kódu zobrazí výstup vyvolaný príkazmi kódu, pričom sa dá nastavovacími parametrami upraviť, či sa zobrazí (ako v prípade `summary` a grafu) alebo nie.



Obrázok 4: Vytvorenie jednoduchého ukážkového R Markdown súboru, zdroj a výstup

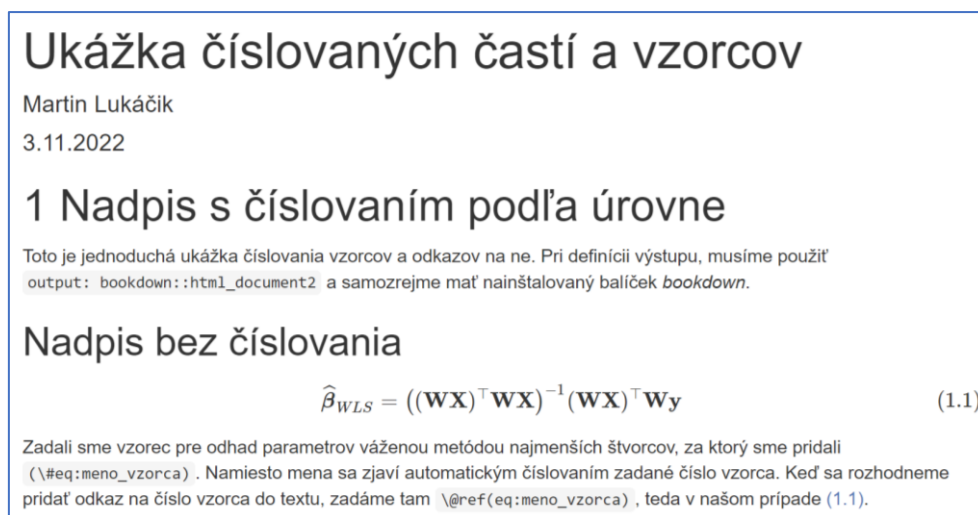
Obyčajný text sa upravuje riadiacimi znakmi na začiatku, kde napríklad počet hashtagov znamená úroveň nadpisov alebo sa táto uvedie medzi tagmi `<h2>nadpis</h2>` ako pri HTML. Pri spustení tvorby R Markdown súboru sa obvykle zobrazuje zdrojový súbor, teda vidíme editor s číslovanými riadkami a nad ním je zvýraznený prepínač *Source*. Môžeme v ňom ľahko editovať text podľa pravidiel Markdownu. Napríklad text, ktorý chceme zvýrazniť **boldom**, vložíme medzi dve hviezdičky pred a za (`**text boldom**`). Alebo text, ktorý chceme zvýrazniť kurzívou, vložíme medzi jednu hviezdičku pred a za (`*text kurzívou*`).

Program R Studio ponúka aj jednoduchší spôsob úpravy obyčajného textu, ktorý si nevyžaduje pamätať tieto pravidlá, stačí prepnúť zobrazenie súboru na *Visual*. Možnosti editácie v tomto zobrazení sa podobajú na klasické editory textu a ich možnosti sú na obrázku 5. Vzájomné prepínanie režimov *Source* a *Visual* dokáže výrazne zjednodušiť úpravu výsledného vzhľadu.



Obrázok 5: Možnosti editácie textu R Markdown súboru cez zobrazenie Visual

Matematický výraz v texte získame vložením syntaxe LaTeXu napísanej medzi 1 znak dolára. Napríklad $e = \frac{dy}{dx}$. Ak chceme zadať matematický výraz na samostatný riadok a prípadne ho aj číslovať, zadáme ho medzi 2 znaky dolára a zadáme na jeho konci ($\#eq:memo_vzorca$). Napríklad, estimátor váženej MNS
$$\widehat{\boldsymbol{\beta}}_{WLS} = \left((\mathbf{WX})^{\top} \mathbf{WX} \right)^{-1} (\mathbf{WX})^{\top} \mathbf{Wy} \quad (\#eq:WLS)$$
 sa zobrazí s aktuálnym číslom poradia vzorca v číslovanej kapitole pri vhodnom nastavení výstupu a za podpory balíčka *bookdown*.



Obrázok 6: Automatické číslovanie R Markdown súboru cez zobrazenie Visual

Automatické číslovanie kapitol a vzorcov si vyžaduje inštalovanie balíčka *bookdown* a nahradenie typu výstupu zadaného ako `output: html_document` v YAML hlavičke za `output: bookdown::html_document2`. Číslovanie kapitol sa automaticky pridáva za nadpisy s hashtagmi podľa úrovne nadpisov. Ak chceme nečíslovaný nadpis, ten uvedieme medzi klasické tagy.

Rovnako jednoducho sa dajú číslovať a robiť odkazy aj na chunky kódu, obrázky alebo tabuľky. Prehľad a postupy v jednom príklade uvádza Xie et al (2022) v časti 4.7 Cross-referencing within documents. Na uvádzaných odkazoch v zozname použitej literatúry sú vysvetľované mnohé ďalšie zaujímavé úpravy a nastavenia R Markdownu, či sprievodných balíčkov.

4 ZÁVER

Replikovateľnosť výskumu a jeho publikovateľnosť je dôležitou nielen pre vedecké časopisy, ale je podstatnou súčasťou práce každého veda a univerzitného učiteľa. V súčasnosti existuje mnoho nástrojov umožňujúcich plniť túto úlohu na veľmi dobrej úrovni. V tomto príspevku sme sa pokúsili prezentovať možnosti jedného z nich a to nástroja R Markdown. Predstavuje veľmi užitočný inštrument v podpore analýz realizovaných v jazyku R v prostredí R Studio.

Použitá literatúra

1. ALLAIRE, J.J, XIE, Y., MCPHERSON, J., LURASCHI, J., USHEY, K., ATKINS, A., WICKHAM, H., CHENG, J., CHANG, W., IANNONE, R. 2022. *Rmarkdown: Dynamic Documents for R*.
2. *Bookdown Write HTML, PDF, ePub, and Kindle books with R Markdown*. 2022. Dostupné na: <https://bookdown.org/> [03.11.2022]
3. EViews 13 – Jupyter Notebook Support. 2022. *EViews homepage*. Dostupné na: https://eviews.com/EViews13/ev13general_n.html#jup [03.11.2022]
4. LUKÁČIK, M. 2016. *Úvod do ekonometrickej analýzy so systémom R*. In Zborník medzinárodného vedeckého seminára Nové trendy v ekonometrii a operačným výzkumu, Bratislava: Ekonóm, 2016, s. 77-87.
5. LUKÁČIK, M. 2019. *Odhad vektorovo autoregresných modelov v R*. In Využitie kvantitatívnych metód vo vedeckovýskumnej činnosti a v praxi XIII, Bratislava: Ekonóm, 2019, s. 118-126.
6. LUKÁČIK, M., LUKÁČIKOVÁ, A. 2018. *Úvod do analýzy panelových dát v systéme R*. In Zborník medzinárodného vedeckého seminára Nové trendy v ekonometrii a operačným výzkumu. Bratislava: Ekonóm, 2018, s. 118-127.
7. LUKÁČIK, M., LUKÁČIKOVÁ, A., SZOMOLÁNYI, K. 2016. *Úvod do práce so systémom R*. In: Zborník medzinárodného vedeckého seminára Nové trendy v ekonometrii a operačným výzkumu, Bratislava: Ekonóm, 2016, s. 88-98.
8. LUKÁČIK, M., LUKÁČIKOVÁ, A., SZOMOLÁNYI, K. 2019. *Bayesovský odhad vektorovo autoregresných modelov v R*. In Využitie kvantitatívnych metód vo vedeckovýskumnej činnosti a v praxi XIII, Bratislava: Ekonóm, 2019, s. 127-136.
9. *Pandoc a universal document converter*. 2022. Dostupné na: <http://pandoc.org> [03.11.2022]
10. *R Markdown*. 2022. Dostupné na: <https://rmarkdown.rstudio.com/index.html> [03.11.2022]
11. RODRÍGUEZ, G. 2017. Literate data analysis with Stata and Markdown. *The Stata Journal*. roč. 17, 2017, č. 3, s. 600-618.
12. *Shiny*. 2022. Dostupné na: <https://shiny.rstudio.com/> [03.11.2022]
13. The JAE Replication Section. 2022. *Journal of Applied Econometrics*. Dostupné na: <https://onlinelibrary.wiley.com/page/journal/10991255/homepage/news.html#replication> [03.11.2022]
14. XIE, Y. 2015. *Dynamic Documents with R and Knitr. 2nd ed*. Boca Raton: Chapman; Hall/CRC.
15. XIE, Y., DERVIEUX, C., RIEDERER, E. 2022. *R Markdown Cookbook*. Dostupné na: <https://bookdown.org/yihui/rmarkdown-cookbook/> [03.11.2022]

Kontaktné údaje

prof. Ing. Martin Lukáčik, PhD., Ing. Adriana Lukáčiková, PhD., doc. Ing. Karol Szomolányi, PhD.
Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky
Dolnozemska 1/b, 852 35 Bratislava

Tel: (421 2) 67 295 822

email: martin.lukacik@euba.sk, adriana.lukacikova@euba.sk, karol.szomolanyi@euba.sk

MAKROEKONOMICKÉ DÔSLEDKY EURÓPSKEHO EKOLOGICKÉHO DOHOVORU PRE SLOVENSKO

THE MACROECONOMIC IMPLICATIONS OF THE EUROPEAN GREEN DEAL ON SLOVAKIA

Veronika Mitkova

Abstrakt

V príspevku sa venujeme modelovaniu dôsledkov Európskeho klimatického dohovoru na makroekonomické agregáty na Slovensku v porovnaní s krajinami Európskej únie. Využívame mierne modifikovaný model všeobecnej ekonomickej rovnováhy vytvorený na Purdue University GTAP-E a údajovú základňu z roku 2014. Prostredníctvom externého šoku zníženia emisných kvót oxidu uhličitého o 55% sme ukázali vplyv na celkový output komodít, ich domáci predaj a dovoz, domácu trhovú cenu a nakoniec aj na hrubý domáci produkt v oboch modelovaných regiónoch.

***Kľúčové slová:** CGE model, environmentálne ukazovatele, Európsky ekologický/klimatický dohovor*

Abstract

In this paper we model the impact of the European Climate Convention on macroeconomic aggregates in Slovakia compared to European Union countries. We use a slightly modified general economic equilibrium model developed at Purdue University GTAP-E and the 2014 database. Through an exogenous shock of a 55% reduction in carbon dioxide emission quota, we show the impact on total commodity production, domestic sales and imports, domestic market price, and finally gross domestic product in both modeled regions.

***Keywords:** CGE model, environmental indicators, European Green Deal*

1 EURÓPSKY EKOLOGICKÝ DOHOVOR

Zhoršujúce sa merateľné ale aj nemerateľné indikátory kvality životného prostredia priamo súvisia s ľudskou činnosťou a rastúcou svetovou populáciou. Je preto na nás ľuďoch tento trend zvrátiť, zastaviť alebo aspoň spomaliť. Európsky ekologický, resp. klimatický dohovor stanovuje Európu za prvý klimaticky neutrálny kontinent. Dohovor tvorí súbor takých zmien v klimatických, energetických, dopravných a daňových oblastiach, ktoré povedú k zníženiu emisií skleníkových plynov do roku 2030 najmenej o 55% v porovnaní s rokom 1990. Za cieľ si tiež kladie využívať aspoň 32% energie z obnoviteľných zdrojov a zvýšiť energetickú efektívnosť o 32,5%.

V predložennom článku budeme modelovať Európsky ekologický dohovor prostredníctvom modelu všeobecnej ekonomickej rovnováhy a zvlášť sa zameriame na jeho makroekonomické implikácie pre Slovensko. Výstupy prezentované v tomto článku sú súčasťou pripravovanej publikácie *Environomics* (Grisáková a kol., 2022).

2 GLOBAL TRADE ANALYSIS PROJECT - E

Modely všeobecnej ekonomickej rovnováhy (CGE- Computable General Equilibrium model) sú triedou makroekonomických modelov založených na mikroekonomických princípoch. Využívajú údaje z jedného časového obdobia a nevyžadujú na odhad parametrov dlhé časové rady, čím sú mimoriadne vhodné na modelovanie rozvíjajúcich sa ekonomík, a takých sociálno-ekonomických oblastí, v ktorých sa stretávame s nedostatkom pozorovaní. Tieto modely vďaka spôsobu stanovenia hodnôt parametrov umožňujú bohatú ekonomickú štruktúru. Vstupné dáta sú usporiadané v matici spoločenského účtovníctva, ktorá je postavená na dvoch základných pilieroch: matica je štvorcová, využívajúc princípy input-output analýzy a sumy príslušných riadkov a stĺpcov sa rovnajú.

Na modelovanie efektov zníženia emisií skleníkových plynov využijeme model Global Trade Analysis Project - E, ktorý vznikol modifikáciou GTAP modelu. GTAP-E model prezentovali Burniaux a Truong (2002) a neskôr McDougall a Golub (2009) a v tomto príspevku ho upravíme pre naše potreby.

2.1 Databáza údajov

Údaje, s ktorými v modeli pracujeme, sú prevzaté z Global Trade Analysis Project a pokrývajú 65 komodít a prislúchajúcich odvetví v dezagregácii na krajiny Európskej únie, Slovenskej republiky a zvyšných krajín sveta v roku 2014. Komodity a odvetvia sú následne zoskupené do nasledovných množín:

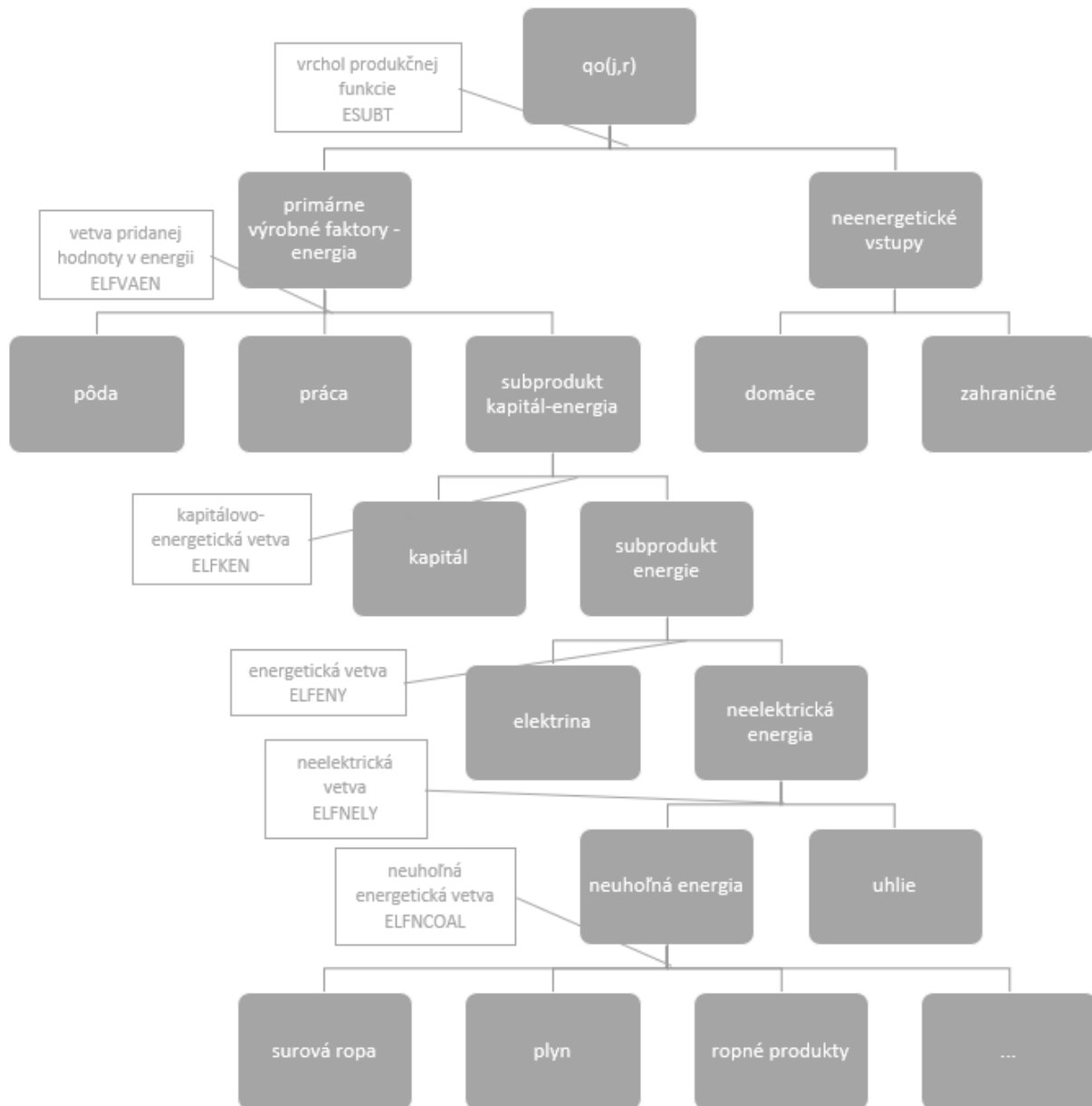
1. Poľnohospodárstvo (Agr): nelúpaná ryža, pšenica, obilniny i.n.¹, zelenina, ovocie, orechy, olejnaté semená, cukrová trstina, cukrová repa, vláknina rastlinného pôvodu, plodiny i.n., hovädzí dobytok, ovce, kozy, živočíšne produkty i.n., surové mlieko, vlna, kokóny priadky morušovej, lesníctvo a rybolov.
2. Ťažba uhlia (Coal): uhlie.
3. Surová ropa (Oil): ropa.
4. Ťažba zemného plynu (Gas): výroba a distribúcia plynu.
5. Rafinované ropné produkty (Oil_pcts): ropné a uhoľné produkty.
6. Elektrická energia (Electricity): elektrina.
7. Energeticky náročné odvetvia (EN_int): nerasté suroviny i.n., chemické výrobky, základné farmaceutické výrobky, výrobky z gummy a plastov, minerálne výrobky i.n., železné kovy, kovy i.n..
8. Ostatné priemyselné odvetvia a služby (Oth_ind_ser): výrobky z hovädzieho mäsa, mäsové výrobky i.n., rastlinné oleje a tuky, mliečne výrobky, spracovaná ryža, cukor, potravinárske výrobky i.n., nápoje a tabakové výrobky, textil, odevy, výrobky z kože a vlny, výrobky z papiera, vydavateľské práce, kovové výrobky, počítačové, elektronické, optické, elektrické zariadenia, stroje a zariadenia i.n., motorové vozidlá a ich časti, dopravné zariadenia i.n., výroba i.n., voda, stavebníctvo, obchod, ubytovanie, stravovanie a služby, doprava i.n., vodná, letecká doprava, skladovanie a pomocné činnosti, komunikácie, finančné služby i.n., poisťovníctvo, služby v oblasti nehnuteľností, obchodné služby i.n., rekreačné a iné služby, verejná správa a obrana, vzdelávanie, zdravotníctvo a sociálna pomoc, bývanie.

2.2 Produkcia v GTAP-E

Produkcia je v modeli modelovaná pomocou takzvaných vnorených produkčných funkcií s konštantnou elasticitou substitúcie, ako to znázorňuje Obrázok 1. Pri výrobe každej komodity j v každom odvetví r sa využívajú dva druhy vstupov, tzv. energetické vstupy

¹ inde nezaradené

primárných výrobných faktorov a neenergetické vstupy, ktoré ďalej delíme na domáce a zahraničné.

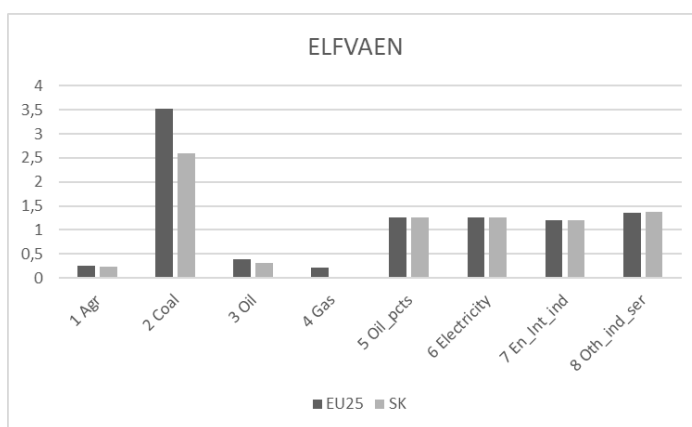


Obrázok 1: Štruktúra produkcie (vlastné spracovanie podľa Burniaux a Truong, 2002)

Elasticita substitúcie, označená ESUBT je pre všetky agregované odvetvia nulová, čo znamená, že energetické a neenergetické vstupy nie sú vzájomne substituovateľné. Vetva pridanej hodnoty s elasticitou substitúcie ELFVAEN využíva primárne výrobné faktory pôdu a prácu a sub-produkt kapitál-energia. Hodnoty elasticity zobrazuje Tabuľka 1, kde je výrazná možnosť substitúcie práve v odvetví ťažby uhlia s hodnotou presahujúcou 2,5 a takmer nulovou substituovateľnosťou v plynovom odvetví 0,02. Pre porovnanie uvádzame aj hodnoty elasticít pre zvyšok krajín Európskej únie. Odvetvia ropných produktov, elektriny, energeticky náročné odvetvia a ostatné odvetvia substituujú primárne výrobné faktory v podobnej miere, v rozmedzí 1,2 až 1,37, podobne v EÚ ako aj na Slovensku. Ďalšie vetvy sú kapitálovo-energetická vetva s elasticitou ELFKEN, kde jej hodnoty indikujú ľahkosť, s akou sa výroba energie prispôbuje úrovni investícií. Nasleduje energetická vetva s elasticitou EFENY, neelektrická vetva s ELFNELY a nakoniec neuhofná vetva s ELFCOAL.

Tabuľka 1: Hodnoty elasticít substitúcie vo vetve pridanej hodnoty a energie

Elasticita substitúcie vo vetve pridanej hodnoty a energie		
ELFVAEN	EU25	SK
1 Poľnohosp.	0,25	0,24
2 Uhlie	3,52	2,59
3 Ropa	0,39	0,32
4 Plyn	0,22	0,02
5 Ropné prod.	1,26	1,26
6 Elektryna	1,26	1,26
7 En. nár. odv.	1,21	1,20
8 Ostatné odv.	1,36	1,37

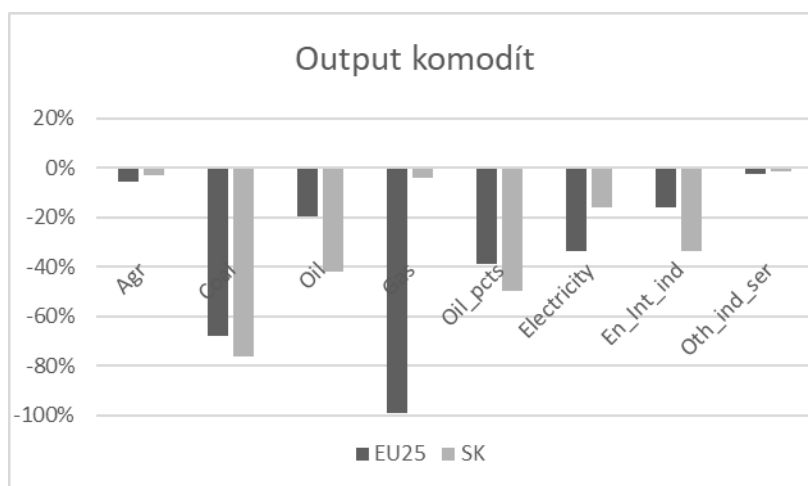


3 MODELOVANIE ZNÍŽENIA EMISNÝCH KVÓT

Východiskom bol model opísaný v McDougall a Golub (2009), ktorý sme modifikovali v niekoľkých aspektoch. Najprv sme upravili uzávery modelu tak, že exogénna premenná reálna sadzba dane z oxidu uhličitého v regióne r bola zamenená za endogénnu nominálnu sadzbu dane z oxidu uhličitého v regióne r . Ďalej boli nákupy emisií, pôvodne exogénna premenná, vymenená za nominálnu sadzbu dane z oxidu uhličitého v príslušných regiónoch. A nakoniec premenná emisií oxidu uhličitého bola zamenená za nákupy emisií v regiónoch r . Následne bol sformulovaný šok zníženia emisných kvót oxidu uhličitého o 55%.

3.1 Makroekonomické dôsledky

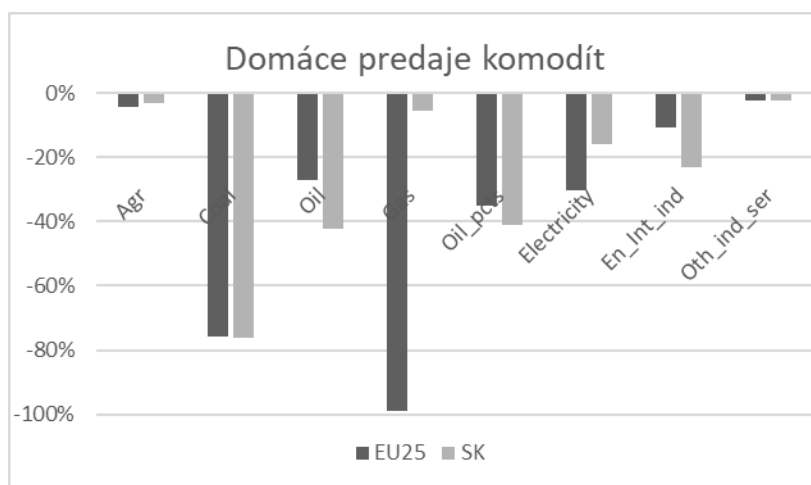
Externý šok vo forme zníženia emisných kvót mal nasledovné dôsledky na makroekonomické agregáty output komodít, domáce predaje, celkový dovoz, trhovú cenu komodít a hrubý domáci produkt:



Obrázok 2: Output komodít

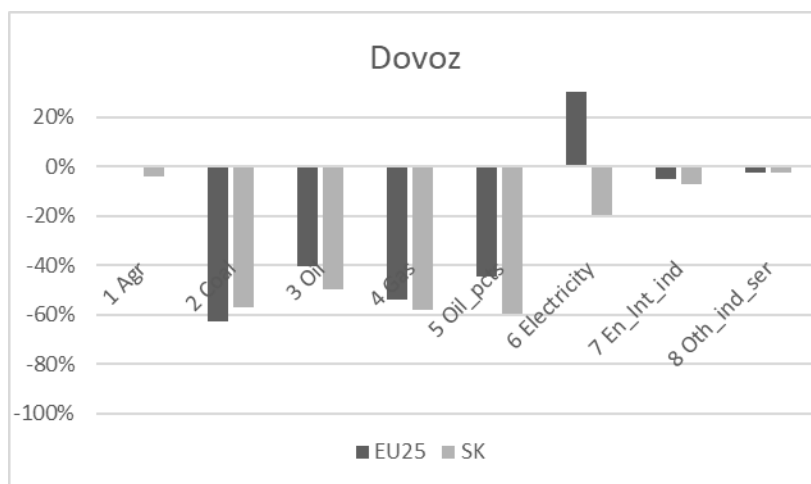
Priamy efekt mal modelovaný šok na output komodít, Obrázok 2, ktorý sa v každom z pozorovaných odvetví znížil, dramaticky v ťažbe uhlia, až o 76% na Slovensku a 68% v EU. Najcitlivejší bol dopad v plynárenskom odvetví v EU, a to pokles až o -99%, kým na

Slovensku len o 4%. Najmenej bude zasiahnutá výroba v poľnohospodárstve a ostatných odvetviach, pokles o 2% až 6%.



Obrázok 3: Domáce predaje komodít

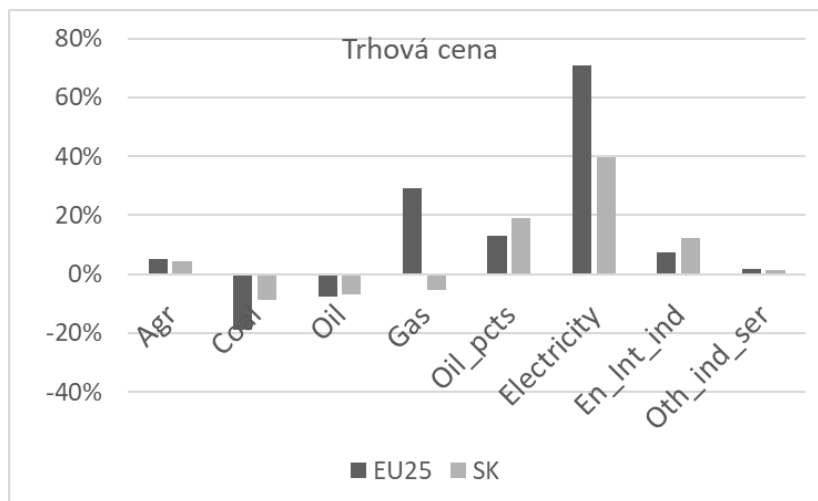
Zmeny v domácich predajoch komodít kopírujú dopad simulovaného šoku na output komodít. Najvyšší prepád v predajoch bude na Slovensku v sektore ťažby uhlia, spracovania ropných produktov a surovej ropy, kým najmenej zasiahnutými budú opäť poľnohospodárstvo, plyn a ostatné sektory. V EÚ bude najvyšší prepád znova v sektore plynu - vid'. Obrázok 3.



Obrázok 4: Celkový dovoz

Celkový dovoz do regiónu klesne na Slovensku vo všetkých odvetviach, pri uhlí, rope, plyne a ropných produktoch v rozmedzí -50% až -60%, v elektrine o -20% a v poľnohospodárstve, energeticky náročných a ostatných odvetviach je pokles do -7%. Obdobné sú dopady aj na sektory EÚ s výnimkou odvetvia elektriny, kde naopak dovoz vzrastie až o 51%, ako to zobrazuje Obrázok 4.

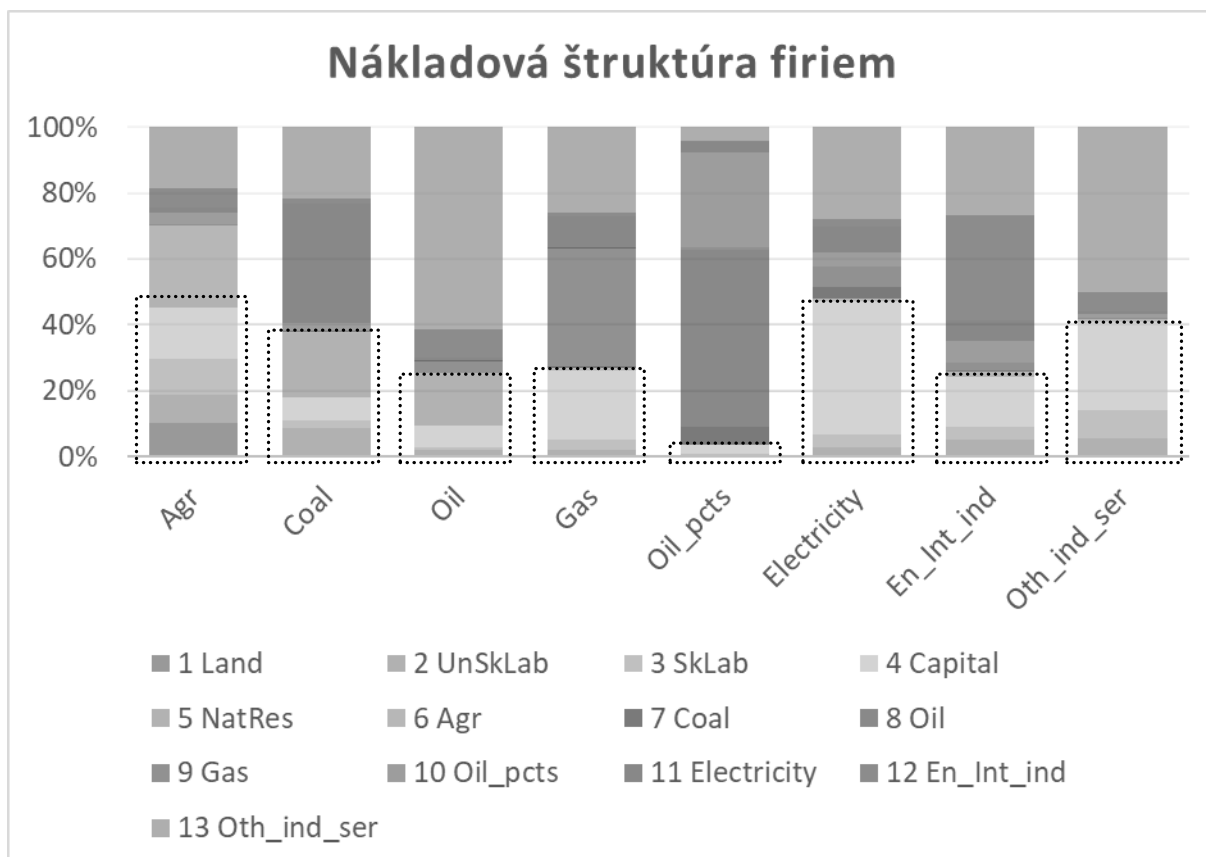
Dopad modelovanej zmeny na cenu komodít je rôzny. Najsilnejšie sa efekt zníženia emisných kvót prejaví v cenách elektriny, ktorá na Slovensku vzrastie o 40%, kým v EÚ o 71%. Na Slovensku sa tiež zvýši domáca trhovú cenu produktov vyrábaných z ropy (19%), energeticky náročných produktov (12%), poľnohospodárskych produktov (4%) a nepatrne aj ostatných produktov (1%), kým trhové ceny ostatných komodít klesnú, najvýraznejšie uhlia o 9% a ropy o 7%. V EÚ je opačný smer zmeny len pri plyne, kde cena vzrastie o 29%. Zmeny sú zachytené na Obrázku 5.



Obrázok 5: Trhová cena komodít

Posledným sledovaným agregátom je hrubý domáci produkt (HDP), ktorý v dôsledku zmien, ktoré v oboch regiónoch prebehnú, na Slovensku vzrastie o 1,76% a v ostatných krajinách EÚ o 0,86%.

Na záver uvádzame aj nákladovú štruktúru firiem v jednotlivých odvetviach, pričom zvlášť poukazujeme na podiel nákladov primárnych výrobných faktorov a medziproduktov. Najvyšší podiel zvýšenia nákladov na primárne výrobné faktory sú v odvetviach poľnohospodárstva a elektriny, naopak najnižšie v ropných produktoch.



Obrázok 6: Nákladová štruktúra firiem v Slovenskej republike

4 DISKUSIA A ZÁVER

Dôsledné uplatnenie Európskej klmatickej dohody by viedlo k všeobecnému poklesu výroby väčšiny komodít a domáceho predajov týchto komodít, pričom trhové ceny väčšiny komodít s výnimkou surovín (uhlie, ropa, plyn) by na Slovensku vzrástli. Celkové efekty merané zmenou hrubého domáceho produktu pre Slovensko a krajiny Európskej únie sa ukazujú ako pozitívne, s rastom HDP pre oba regióny.

Použitá literatúra

1. BURNIAUX, J. M., a TRUONG, T. P. 2002. *GTAP-E: An Energy-Environmental Version of the GTAP Model*. West Lafayette, In: Global Trade Analysis Project (GTAP). Získané z <https://www.gtap.agecon.purdue.edu/resources/download/1203.pdf>
2. GRISÁKOVÁ a kol. 2022. *Environomics: Economic Approach to Environmental Issues*. České Budějovice (Czech Republic): The College of European and Regional Studies.
3. MCDUGALL, R. a GOLUB, A. 2009. *GTAP-E: A Revised Energy-Environmental Version of the GTAP Model (GTAP Research Memorandum No. 15)*. West Lafayette, In: Global Trade Analysis Project (GTAP). Získané z https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=2959

Pod'akovanie

Výskum bol podporený grantovou schémou Ministerstva školstva, vedy a výskumu Slovenskej republiky grantom VEGA 1/0668/22 „Migrácia a striedaná ekonomika: Ekonomické, politické a právne dôsledky riešenia problému starnutia a migrácie populácie v krajinách EÚ“ a VEGA 1/0628/20 „Medzinárodná investičná pozícia a jej vzťah k ekonomickému rastu - možný ukazovateľ vonkajších nerovnováh krajiny”.

Kontaktné údaje

Ing. Veronika Mit'ková, PhD.

Univerzita Komenského v Bratislave, Fakulta sociálnych a ekonomických vied

Mlynské luhy 4, 821 05 Bratislava

email: veronika.mitkova@fses.uniba.sk

VÝBER PORTFÓLIA A MIERY RIZIKA¹

PORTFOLIO SELECTION AND RISK MEASURES

Vladimír Mlynarovič

Abstrakt

Článok prezentuje klasifikáciu a prehľad mier rizika aplikovaných v úlohách výberu portfólia. Postupne sa analyzuje variancia, stredná absolútna odchýlka, dolné parciálne momenty a centrálné semi odchýlky. Na báze kvantilov je zavedená ako miera rizika hodnota v riziku a podmienená hodnota v riziku.

Kľúčové slová: *variancia, stredná absolútna odchýlka, dolný parciálny moment, hodnota v riziku, podmienená hodnota v riziku.*

Abstract

The paper presents a classification and an overview of risk measures applied in portfolio selection models. Such measures as variance, mean absolute deviation, lower partial moment and central semi deviation are analyzed. On the base of quantile definitions, the value at risk and conditional value at risk are introduced.

Keywords: *variance, mean absolute deviation, lower partial moment VaR, CVaR*

1 ÚVOD

Výber portfólia je dobré známy ako kľúčový problém teórie a praxe financií, pretože pri neznámych budúcich výnosov aktív v čase investičného rozhodovania, ide o úlohu rozhodovania v podmienkach neistoty, resp. rizika. Modely výberu portfólia predstavujú *ex-ante* nástroje rozhodovanie: rozhodnutia prijaté dnes možno vyhodnotiť až v budúcnosti, keď sa odhalí neurčitost' týkajúca sa výnosov aktív.

Formálne možno problém sformulovať nasledovne: ako pri danej množine n aktív, do ktorých možno investovať, *teraz* rozdeliť objem peňazí medzi aktíva tak, aby sa *po špecifikovanom časovom období* T dosiahol čo možno najvyšší výnos z investície?

Budúce výnosy aktív sú náhodné premenné, označme ich r_1, r_2, \dots, r_n . Portfólio označíme ako $\mathbf{w} = (w_1, w_2, \dots, w_n)$, kde w_j je podiel kapitálu investovaný do aktíva $j, j = 1, 2, \dots, n$. Aby váhy w_1, w_2, \dots, w_n predstavovali portfólio, musia vyhovovať množine ohraničení, ktorá vytvára množinu W prípustných rozhodovacích vektorov. Najjednoduchší spôsob ako definovať prípustnú množinu je požiadavka, aby súčet nezáporných (čo znamená, že krátke predaje sú vylúčené) váh bol rovný 1. Množina W môže vo všeobecnosti byť množinou prípustných

¹ Výskum bol podporený grantovou schémou Ministerstva školstva, vedy a výskumu Slovenskej republiky grantom VEGA 1/0668/22 „Migrácia a strieborná ekonomika: Ekonomické, politické a právne dôsledky riešenia problému starnutia a migrácie populácie v krajinách EÚ“ a VEGA 1/0628/20 „Medzinárodná investičná pozícia a jej vzťah k ekonomickému rastu - možný ukazovateľ vonkajších nerovnováh krajiny“.

riešení úlohy lineárneho programovania v štandardnom tvare bez toho, aby to ovplyvnilo nasledujúcu analýzu.

Výnos portfólia \mathbf{w} je náhodná premenná $R_{\mathbf{w}}$ a jej závislosť od výnosov individuálnych aktív je opísaná vzťahom

$$R_{\mathbf{w}} = w_1 r_1 + w_2 r_2 + \dots + w_n r_n$$

Hlavné otázky, ktoré tu vznikajú, smerujú do troch hlavných oblastí výskumu, a to ko opísať rozdelenie náhodných výnosov, ako vybrať medzi substitučnými možnosťami rozdelení pravdepodobnosti a, ak sa prijalo investičné rozhodnutie, kedy a ako ho prehodnotiť?

Pri výbere portfólia je obvyklé vyjadriť náhodné výnosy prostredníctvom diskretných náhodných premenných pri ich známych realizáciách pre špecifikovaný počet S stavov sveta (scenárov) s nejakými priradenými pravdepodobnosťami p_1, p_2, \dots, p_S ich výskytu. Označíme S možných výsledkov náhodného výnosu $r_j, j = 1, 2, \dots, n$, ako $r_{1j}, r_{2j}, \dots, r_{Sj}$. Takto, výnos $R_{\mathbf{w}}$ portfólia \mathbf{w} je diskretnou náhodnou premennou s S možnými výsledkami $R_{1\mathbf{w}}, R_{2\mathbf{w}}, \dots, R_{S\mathbf{w}}$ vyskytujúcimi sa s pravdepodobnosťami p_1, p_2, \dots, p_S , kde

$$R_{i\mathbf{w}} = w_1 r_{i1} + w_2 r_{i2} + \dots + w_n r_{in}, \quad i = 1, 2, \dots, S$$

V tomto príspevku sa zameriame na druhú z vyššie uvedených otázok. Predpokladáme, že (diskrétno, združené) rozdelenie náhodných výnosov r_1, r_2, \dots, r_n je známe a ako r_{ij} označíme výnos aktíva j pri scenári $i, i = 1, 2, \dots, S, j = 1, 2, \dots, n$. Budeme uvažovať investičný problém len v jednom investičnom období, takže tretia z vyššie uvedených otázok je tiež mimo predmetu tejto analýzy.

Medzi všeobecne akceptované modely týkajúce sa výberu medzi substitúciami rozdelení patria: modely stredná hodnota – riziko a maximalizácia očakávanej užitočnosti/stochastická dominancia. Prvý definuje preferenčnú reláciu (kritérium výberu) medzi náhodnými premennými. Na základe tohto kritéria výberu sú investičné rozhodnutia prijímané prostredníctvom riešenia optimalizačných úloh. V ďalšej časti tejto analýzy budeme detailne skúmať formulácie takýchto rozhodovacích problémov.

Táto analýza poskytuje prehľad existujúcich model výberu a mier rizika používaných pri výbere portfólia. Je známe, že takéto modely, hoci sú teoreticky otáznе, sú široko využívané v praxi finančnej komunity. Iné modely, napriek popularite a odporúčaní vedeckej komunity neboli v praxi využívané v širšom meradle. Nakoniec, existujú modely, ktoré sú teoreticky zvučné, ale ťažko prakticky implementovateľné. Naším zámerom nie je obhajovať konkrétny prístup, ale poskytnúť objektívny súhrn toho, čo je dostupné pre výber portfólia.

2 MIERY RIZIKA

V modeloch typu stredná hodnota – riziko sú rozdelenia výnosov charakterizované a porovnávané prostredníctvom dvoch skalárnych charakteristík pre každé rozdelenie. Jednou skalárnou hodnotou je očakávaná hodnota (priemer), a žiaduce sú vysoké očakávané hodnoty. Druhou skalárnou charakteristikou je hodnota „miery rizika“. Ináč povedané, miera rizika je nejaká funkcia, ktorá spája s každým rozdelením číslo, ktoré opisuje jeho „rizikovosť“, pričom, samozrejme, žiaduce sú nízke hodnoty rizika. Preferencia medzi rozdeleniami je definovaná prostredníctvom substitúcie medzi strednou hodnotou a rizikom

Uvažujme dve prípustné portfólia $\mathbf{w} = (w_1, w_2, \dots, w_n)$ a $\mathbf{v} = (v_1, v_2, \dots, v_n)$ s výnosmi $R_{\mathbf{w}}$ a $R_{\mathbf{v}}$. V prístupe stredná hodnota – riziko, s mierou rizika označenou ako ρ , náhodná premenná $R_{\mathbf{w}}$ dominuje (je preferovaná pred) náhodnou premennou $R_{\mathbf{v}}$ vtedy a len vtedy, ak

$$E(R_w) \geq E(R_v) \text{ a } \rho(R_w) \leq \rho(R_v)$$

pričom aspoň jedna z týchto dvoch nerovníc je splnená ako ostrá.

Alternatívne možno povedať, že portfólio w je preferované pred portfóliom v . Zámerom je nájsť také portfólia, ktoré sú nedominované vzhľadom k vyššie definovanej preferenčnej relácii, čo znamená, že neexistuje žiadne iné prípustné portfólio, ktoré môže súčasne zlepšiť ako stredný výnos tak aj riziko. Formálne: prípustné portfólio $w = (w_1, w_2, \dots, w_n)$ je nedominované v prístupe stredná hodnota – riziko, s rizikom označeným ako ρ , ak pre akékoľvek iné prípustné portfólio $v = (v_1, v_2, \dots, v_n)$ nemôže súčasne platiť

$$E(R_w) \leq E(R_v) \text{ a } \rho(R_w) \geq \rho(R_v)$$

s výnimkou situácie, kde

$$E(R_w) = E(R_v) \text{ a } \rho(R_w) = \rho(R_v)$$

To znamená, že nedominované portfólio, nazývané tiež „efektívne“ portfólio, má najnižšiu úroveň rizika pre daný stredný výnos (a najvyššiu stredný výnos, pre danú úroveň rizika).

Efektívne portfólia sú získavané riešením optimalizačných úloh, ktoré možno sformulovať tromi alternatívnymi spôsobmi. Najobvyklejšia formulácia je založená na použití špecifikovanej minimálnej hodnoty pre očakávaný výnos portfólia a minimalizáciou rizika portfólia:

$$\min \rho(R_w)$$

za podmienok

(1)

$$E(R_w) \geq d, w \in W$$

kde d vyjadruje požadovanú úroveň očakávaného výnosu portfólia (špecifikovanú rozhodovateľom). Variovanie d a opakované riešenie zodpovedajúcej optimalizačnej úlohy identifikuje portfólia s minimálnym rizikom pre každé d . Alternatívnou formuláciou, ktorá v účelovej funkcii explicitne „obchoduje“ riziko proti výnosu, je:

$$\max E(R_w) - \lambda \rho(R_w) \quad (\lambda \geq 0)$$

za podmienok

(2)

$$w \in W$$

Prostredníctvom variovania substitučného koeficienta λ a opakovaným riešením zodpovedajúcich optimalizačných úloh skonštruujeme efektívnu hranicu. Poznamenajme, že alternatívne môžeme tiež maximalizovať očakávaný výnos portfólia pre špecifikovanú maximálne možnú úroveň rizika.

Ako merať riziko a vybrať vhodnú mieru rizika v úlohe výberu portfólia je predmetom stáleho výskumu a rozsiahlych diskusií. Je zrejmé, že použitá miera rizika hrá dôležitú úlohu v rozhodovacom procese a portfólia vybrané pri rôznych mierach rizika môžu byť veľmi odlišné. Miery rizika možno klasifikovať do dvoch kategórií. Miery rizika, ktoré uvažujú odchýlku od stanoveného cieľa, vytvárajú prvú kategóriu. Cieľ môže byť fixovaný (napríklad minimálny akceptovaný výnos), závislý od rozdelenia výnosov (napríklad očakávaná hodnota), alebo dokonca ako stochastický benchmark (napríklad nejaký index). Takéto miery rizika môžu nadobúdať len kladné hodnoty. Tieto miery rizika možno ďalej rozčleniť na symetrické (dvojstranné) miery rizika a nesymetrické (jednostranné, dolné strany, straty) miery rizika. Symetrické miery rizika kvantifikujú riziko prostredníctvom pravdepodobnostne váženej

disperzie výsledkov v okolí predom špecifikovaného cieľa, obvykle očakávanej hodnoty. Miery tejto kategórie penalizujú ako negatívne tak aj kladné odchýlky od stanoveného cieľa. Zvyčajne používanými symetrickými mierami rizika sú variancia a stredná absolútna odchýlka (MAD). Asymetrické miery rizika kvantifikujú riziko podľa výsledkov a pravdepodobnosti pod cieľovými hodnotami. Výsledky sú tu penalizované len v prípadoch, keď sú nižšie ako stanovaná cieľová hodnota. Argumentom je, že to je oveľa bližšie k intuitívnemu chápaniu rizika ako nežiaduceho výsledku, „nepriaznivej“ situácie, kde sa uvažuje, že väčšina investorov spája riziko so zlyhaním možnosti dosiahnuť špecifikovaný cieľ. Globálna myšlienka týkajúca sa dolného rizika je taká, že ľavá strana rozdelenia výnosov obsahuje riziko, zatiaľ čo jeho pravá strana obsahuje lepšie investičné príležitosti. Medzi asymetrickými mierami rizika patrí významné postavenie dolným parciálnym momentom a centrálnym semi – odchýlkam.

Miery rizika v druhej kategórii pracujú s celkovou nebezpečnosťou možných strát. Niekedy sú nazývané „chvostové miery rizika“, pretože uvažujú len ľavý chvost rozdelenia, korešpondujúci s najvyššími stratami. Do úvahy sa berie len určitý počet najhorších výsledkov (tento počet závisí od špecifikovanej hladiny významnosti α (napríklad pri $\alpha = 0.05$ sa berie do úvahy len najhorších 5% výsledkov). Často používanými mierami rizika tejto kategórie sú hodnota v riziku (VaR) podmienená hodnota v riziku (CVaR). Hodnoty týchto mier rizika môžu byť tak kladné ako aj záporné. Ďalej opíšeme niektoré najčastejšie miery rizika.

2.1 Variancia

Variancia je historicky prvou mierou rizika použitou v úlohe výberu portfólia. Markowitz (1952) zaviedol prístup stredná hodnota – variancia, ktorý bol neskôr zovšeobecnený na modely typu stredná hodnota – variancia. Okrem toho, vyvinul v rámci tohto prístupu výpočtový model na nájdenie efektívnych portfólií.

Variancia bola široko využívaná v štatistike ako indikátor rozpätia v okolí očakávanej hodnoty. Jej druhá odmocnina, štandardná odchýlka, je najznámejšou mierou štatistickej disperzie.

Variancia náhodnej premennej R_w , označená ako $\sigma^2(R_w)$, je definovaná ako jej druhý centrálny moment: očakávaná hodnota štvorca odchýlok R_w od jeho vlastného priemeru, t.j.

$$\sigma^2(R_w) = E \left[(R_w - E(R_w))^2 \right]$$

Medzi vlastnosťami variancie je zvlášť významná jedna, využívaná pri výpočte variancie pre lineárnu kombináciu náhodných premenných. V kontexte optimalizácie portfólia umožňuje vyjadriť varianciu (rozptyl) výnosu portfólia

$$R_w = w_1 r_1 + w_2 r_2 + \dots + w_n r_n$$

vyplývajúceho z výberu $\mathbf{w} = (w_1, w_2, \dots, w_n)$ ako

$$\sigma^2(R_w) = \sum_{j=1}^n \sum_{k=1}^n w_j w_k \sigma_{jk} \quad (3)$$

kde σ_{jk} označuje kovarianciu medzi R_j a R_k . Znamená to, že variancia portfólia je vyjadrená ako kvadratická funkcia rozhodnutia w_1, w_2, \dots, w_n

2.2 Stredná absolútna odchýlka

Stredná absolútna odchýlka (*MAD*) náhodnej premennej R_w je definovaná ako očakávaná hodnota absolútnych hodnôt odchýlok od strednej hodnoty, t.j.

$$MAD(R_w) = E[|(R_w - E(R_w))|]$$

MAD je tiež známou mierou štatistickej disperzie. Konceptuálne je veľmi podobná s varianciou. V praxi výberu portfólia bola zavedená hlavne preto, že model stredná hodnota - variancia bol v minulosti ťažko riešiteľný pre veľké množiny údajov, kvôli kvadratickej účelovej funkcii (5). Využitia *MAD*, namiesto variancie, tento problém prekonalo, pretože model typu stredná hodnota – *MAD* je úlohou lineárneho programovania (LP). V súčasnosti, pri moderných solveroch, možno rozsiahle KP problémy riešiť úplne rutinne. Využitie *MAD* ako miery rizika však stále zostáva alternatívou pre klasický model typu stredná hodnota – variancia.

2.3 Dolné parciálne momenty a centrálné semi - odchýlky

Asymetrické miery rizika, alebo miery rizika pod stanoveným cieľom, poskytujú lepšie opísanie rizika ako nežiadúceho výsledku. Dolné parciálne momenty predstavujú generický názov pre asymetrické miery rizika, ktoré pracujú s fixným cieľom.

Dolné parciálne momenty merajú očakávanú hodnotu odchýlok pod *fixnou* cieľovou hodnotou τ . Možno ich vyjadriť nasledujúcou formuláciou: Nech τ je vopred definovaná (investorom špecifikovaná) cieľová hodnota pre výnos portfólia R_w a nech $\alpha > 0$. *Dolný parciálny moment stupňa α okolo hodnoty τ* náhodnej premennej R_w s distribučnou funkciou F je definovaný ako

$$LPM_\alpha(\tau, R_w) = E[(\max(0, \tau - R_w))^\alpha] = \int_{-\infty}^{\tau} (\tau - r)^\alpha dF(r) \quad (4)$$

pri často využívanej „normalizovanej verzii“ v tvare

$$LPM_\alpha(\tau, R_w) = \{E[(\max(0, \tau - R_w))^\alpha]\}^{\frac{1}{\alpha}}$$

Úloha τ je jednoznačná. Každý rozhodovateľ stanoví jeho vlastný cieľ, pod ktorý nechce aby výnos klesol ($\tau = 0$, t. j. žiadne straty, čo je najobvyklejší prípad) Úloha parametra α už nie je taká priamočiara a jej analýzu možno nájsť v Fishburn (1977). Ak je hlavným zámerom vylúčiť zlyhanie v splnení cieľa, bez ohľadu na veľkosť zlyhania, potom je rozhodovateľ ochotný prijať riziko, aby minimalizoval možnosť, že výnos čo len slabšie klesne pod τ . V takomto prípade je vhodným výberom pre merania rizika malá hodnota α . Ak sú malé odchýlky pod cieľ relatívne neškodné v porovnaní s veľkými odchýlkami, potom rozhodovateľ preferuje padnúť slabšie pod τ o malé množstvo pred akceptovaním rizika, že výsledkom by mohla byť veľká strata. V tomto prípade sú indikované vyššie hodnoty α . To teda znamená, že α je parameter opisujúci investorovu averziu k riziku. Čím vyššia je α , tým vyššia je investorova averzia k riziku. Ako najznámejšie formulácie mier rizika typu „pod cieľom“ využijúc formuláciu (α, τ) uvedieme:

1. Safety First: ak $\alpha \rightarrow 0$

$$SF(R_w) = LPM_{\alpha \rightarrow 0}(\tau, R_w) = E[(\max(0, \tau - R_w))^{\alpha \rightarrow 0}]$$

SF je pravdepodobnosť schodku, ktorý meria možnosť, že výnos portfólia klesne pod nejakú predefinovanú katastrofálnu úroveň τ :

$$SF(R_w) = Prob(R_w < \tau)$$

2. Očakávané dolné riziko: ak $\alpha = 1$

$$LPM_1(\tau, R_w) = E[\max(0, \tau - R_w)]$$

3. Cieľová semi-variácia: ak $\alpha = 2$

$$LPM_2(\tau, R_w) = E\left[\left(\max(0, \tau - R_w)\right)^2\right]$$

Centrálne semi-odchýlky sú podobné dolným parciálnym momentom, avšak, merajú očakávanú hodnotu odchýlky pod strednou hodnotou (zatiaľ čo dolné parciálne momenty merajú očakávanú hodnotu rovnakej odchýlky pod nejakou fixovanou cieľovou hodnotou).

Centrálne semi-odchýlka stupňa α ($\alpha = 1, 2, \dots$) náhodnej premennej R_w je definovaná ako

$$CSD_\alpha(R_w) = E\left[\left(\max(0, E(R_w) - R_w)\right)^\alpha\right] \quad (5)$$

Najznámejšími mierami rizika tejto kategórie sú miery pre $\alpha = 1$ (absolútna semi-odchýlka) a pre $\alpha = 2$ (semi-variácia), ktorú ako prvý zaviedol Markowitz (1959), kde absolútna semi-odchýlka je definovaná ako

$$\delta^-(R_w) = CSD_1(R_w) = E[\max(0, E(R_w) - R_w)]$$

a odvodená štandardná semi-odchýlka ako

$$\sigma^-(R_w) = E\left\{\left[\left(\max(0, E(R_w) - R_w)\right)^2\right]\right\}^{\frac{1}{2}}$$

2.4 Hodnota v riziku

Nech R_w je náhodná premenná opisujúca výnos portfólia w v danom období jeho držby a $A\% = \alpha \in (0, 1)$ je percento, ktoré vyjadruje nejakú vzorku „najhorších prípadov“ pre výsledky R_w . Zaujímavé sú hodnoty α blízke k 0 (napr. $\alpha = 0.01 = 1\%$, alebo $\alpha = 0.05 = 5\%$).

Výpočet hodnoty v riziku (VaR – Value at Risk) náhodnej premennej R_w (alebo portfólia w) pre stupeň α indikuje že, s pravdepodobnosťou najmenej $(1 - \alpha)$ strata neprevýši VaR v priebehu daného obdobia držby. Súčasne platí, že, pravdepodobnosť straty prevyšujúcej VaR je ostro nižšia ako α . Poznajme, že strata je uvažovaná vzhľadom k 0: záporné výnosy sú považované za kladné straty. To znamená, že „strata“ spojená s náhodnou premennou R_w je opísaná náhodnou premennou $-R_w$.

Matematicky, VaR v stupni α náhodnej premennej R_w je definovaná prostredníctvom pojmu α - kvantilov. Nižšie uvádzame definície a vlastnosti týkajúce sa kvantilov, ktoré sú dôležité pre pochopenie definície VaR. V uvádzaných definíciách F vyjadruje kumulatívnu distribučnú funkciu náhodnej premennej R_w . Nejaký α -kvantil premennej R_w je také reálne číslo r , že

$$Prob(R_w < r) \leq \alpha \leq Prob(R_w \leq r)$$

Dolný α -kvantil premennej R_w , označený ako $q_\alpha(R_w)$ je definovaný ako

$$q_\alpha(R_w) = \inf\{r \in \mathcal{R}: F(r) = Prob(R_w \leq r) \geq \alpha\}$$

a horný α -kvantil premennej R_w , označený ako $q^\alpha(R_w)$ je definovaný ako

$$q^\alpha(R_w) = \inf\{r \in \mathcal{R}: F(r) = \text{Prob}(R_w \leq r) > \alpha\}$$

Hodnota v riziku (VaR) premennej R_w v úrovni α je definovaná ako záporný horný α -kvantil premennej R_w :

$$\text{VaR}_\alpha(R_w) = -q^\alpha(R_w)$$

Znamienko *minus* v definícii VaR je preto, lebo je pravdepodobné, že $q^\alpha(R_w)$ nadobudne zápornú hodnotu, ale ak hovoríme o „strate“, uvažujeme s absolútnou hodnotou. Vychádzajúc z (9) môžeme VaR definovať tiež ako

$$\text{VaR}_\alpha(R_w) = q_{1-\alpha}(-R_w)$$

a teda

$$\text{VaR}_\alpha(R_w) = -q^\alpha(R_w) = q_{1-\alpha}(-R_w) \quad (6)$$

Vo vzťahu (6) je VaR vyjadrená vo forme výnosu. Možno ju tiež vyjadriť v peňažných jednotkách. Ak je daná peňažná hodnota portfólia w : napríklad, ak $\text{VaR}_\alpha(R_w) = 0.1$, potom, vyjadrené v peňažných jednotkách to znamená, VaR je 10% východiskovej investície.

2.5 Podmienená hodnota v riziku

Nech, ako vyššie, R_w je náhodná premenná opisujúca výnos portfólia w v danom období jeho držby a $A\% = \alpha \in (0, 1)$ je percento, ktoré vyjadruje nejakú vzorku „najhorších prípadov“ pre výsledky R_w (obvykle $\alpha = 0.01 = 1\%$, alebo $\alpha = 0.05 = 5\%$). Definícia podmienenej hodnoty v riziku *CVaR* (*Conditional Value at Risk*) pre konkrétnu úroveň α je matematickým prepisom koncepcie „priemernej straty v najhorších $A\%$ prípadoch“, kde „strata“ spojená s R_w je opísaná náhodnou premennou $-R_w$.

CVaR je aproximačne rovná priemeru strát väčších alebo rovných ako *VaR* (pri tej istej úrovni spoľahlivosti α), pričom v niektorých prípadoch to platí presne. Ak existuje jediné také r , pre ktoré $\text{Prob}(R_w \leq r) = \alpha$, potom je situácia jednoduchá: *VaR* je záporom r a *CVaR* očakávaná strata zo strát väčších ako *VaR*. Platí to vždy, keď je R_w spojitá náhodná premenná. (Tiež v prípade takých násobných hodnôt r , pre ktoré $\text{Prob}(R_w \leq r) = \alpha$). *CVaR* je podmienené očakávanie strát, ktorý prevyšujú *VaR*. V prípade, že neexistuje taká hodnota r , pre ktorú $\text{Prob}(R \leq r) = \alpha$ (ako je to napríklad na obr. 2 a 4) použije sa α -kvantil $q^\alpha(R_w)$. Avšak $\text{Prob}(R_w \leq q^\alpha(R_w)) > \alpha$ a „dodatočná pravdepodobnosť“ rovná $\text{Prob}(R_w \leq q^\alpha(R_w)) - \alpha$, zodpovedajúca $q^\alpha(R_w)$, je vyňatá z podmienene očakávaných výsledkov pod α -kvantilom. Formálne budeme *CVaR* nasledovne:

Podmienená hodnota v riziku náhodnej premennej R_w pre stupeň α je definovaná ako

$$\text{CVaR}_\alpha = -\frac{1}{\alpha} \{E(R_w 1_{\{R_w \leq q^\alpha(R_w)\}}) - q^\alpha(R_w) [\text{Prob}(R_w \leq q^\alpha(R_w)) - \alpha]\} \quad (7)$$

kde

$$1_{\{relácia\}} = \begin{cases} 1, & \text{ak je výraz pravdivý} \\ 0, & \text{v opačnom prípade} \end{cases}$$

čo znamená, že $R_w 1_{\{R_w \leq q^\alpha(R_w)\}}$ je získané z R_w tak, že sa uvažujú len výsledky pod horným α -kvantilom $q^\alpha(R_w)$.

3 ZÁVER

Pri výbere portfólia je zásadnou otázkou model použitý na výber medzi rozdeleniami platieb (výnosov). Doteraz je najviac akceptovaným a často aplikovaným prístup poskytovaný modelmi typu stredná hodnota – riziko. V týchto modeloch je rozdelenie výnosov opísané dvomi skalárnymi hodnotami: očakávanou hodnotou („priemerom“) o veľkosťou rizika. Riziko spojené s rozdelením výnosov možno kvantifikovať variadou spôsobov. Obvyklý prístup (chronologický prvý) je využitie variancie ako miery rizika. Model typu priemer – variancia je doteraz v praxi najčastejšie aplikovaný.

Variancia ako miera rizika však bola kritizovaná ako nevhodná miera, kvôli jej symetrickej podstate a chýbajúcej konzistencii s reláciami stochastickej dominancie. Alternatívne miery boli najskôr navrhnuté s cieľom intuitívnejšie podchytiť riziko, ako nežiadúcu udalosť. Dolné parciálne momenty kvantifikujú riziko uvažujúc len výsledky pod definovaným cieľom. Centrálné odchýlky sú podobné, ale považujú za cieľ očakávanú hodnotu. Všetky tieto miery rizika kvantifikujú riziko vo forme odchýlok od cieľa (fixného alebo závislého od rozdelenia pravdepodobností). Odlišnou kategóriou mier rizika sú tzv. „chvostové“ miery rizika, ktoré uvažujú s malým percentom najhorších scenárov a oceňujú rozsah strát za týchto podmienok. Povšimnutia hodnými sú VaR a CVaR.

Z výpočtového hľadiska, minimalizácie väčšiny z vyššie uvedených mier rizika vedie k realizovateľným optimalizačným modelom. Pre miery rizika typu „pod cieľom“, je väčšinou využívaný stupeň 1 (očakávané dolné riziko) a stupeň 2 (semi-variancia). Minimalizácia variancie a (cieľovej) semi-variance sú modelmi kvadratického programovania. Minimalizácia MAD, očakávaného dolného rizika a CVaR predstavujú úlohy LP. VaR je jediná miera rizika, ktorej minimalizácia vedie k úlohe nekonvexného programovania, ťažko riešiteľnej. Teda, hoci sa široko využíva pre vyhodnocovanie rizika portfólia (je štandardom pri regulácii), táto miera nebola použitá ako parameter v modeloch konštrukcie portfólia.

Použitá literatúra

1. BERNOULLI, D. 1954. Exposition of a new theory on the measurement of risk, *Econometrica* 22 (1954), 23–36.
2. FISHBURN, P.C. 1977. Mean-risk analysis with risk associated with below-target returns. *American Economic Review* 67, no. 2, 116-126.
3. MARKOWITZ, H. 1952. Portfolio selection. *Journal of Finance* 7, 77–91.
4. MARKOWITZ, H. 1959. Portfolio selection, John Wiley and Sons, New York.
5. ROMAN, D. and G. MITRA. 2008. Portfolio Selection Models: A Review and New Directions. *WILMOTT Journal. Volume 1 Number 2*, 69 – 85.
6. ROMAN, D., DARBY-DOWNAN, D., and G. MITRA. 2007. Mean-risk models using two risk measures: A multi-objective approach, *Quantitative Finance* 7, 443–458.
7. ROY, A. D. 1952. Safety first and the holding of assets, *Econometrica* 20 (1952), 431–449.

Kontaktné údaje

doc. Ing. Vladimír Mlynarovič, CSc.

Univerzita Komenského v Bratislave, Fakulta sociálnych a ekonomických vied

Mlynské luhy 4 1/b, 821 05 Bratislava

Tel: +421 2 20 66 98 60

email: vladimir.mlynarovic@fses.uniba.sk

ANALÝZA VÝNOSŮ AKCIÍ NA ČESKÉM TRHU ZA POSLEDNÍCH 10 LET

ANALYSIS OF DISTRIBUTIONS OF STOCK RETURNS IN PAST 10 YEARS

Ing. Jakub Neugebauer

Abstrakt

Původní Markowitzův model předpokládá normální rozdělení výnosů akcií. Je však tento předpoklad dodržen? Ve světě financí se objevují důkazy, že tomu tak není. Mezi často navrhované alternativy normálního rozdělení patří například zešikmené studentovo rozdělení či normální-inverzní Gaussovské rozdělení. Tento článek sleduje, jaká rozdělení měly výnosy akcií českých firem v období posledních 10 let.

Klíčová slova: akcie, výnosy, rozdělení výnosů, hustota pravděpodobnosti, zešikmené studentovo rozdělení

Abstract

Basic Markowitz's model has an assumption of normally distributed stock returns. Does this assumption hold in real life? In the world of finance there are proofs, that it is not like that. Ones of many suggested alternatives commonly used are skew student distribution or normal-inverse Gaussian distribution. This article finds out, which distribution does returns of stocks traded on Czech market in the past 10 years have.

Keywords: stocks, returns, distribution of returns, probability density function, skew student distribution

1 VYBRANÁ ROZDĚLENÍ PRAVDĚPODOBNOSTI

Jak již název napovídá, tématem článku je analýza rozdělení výnosů akcií na českém trhu. Primárním účelem této analýzy bylo vytvoření základu pro vybudování modelu sloužícímu k optimalizaci portfolia v diplomové práci autora s názvem *Optimalizace portfolia za různých měř rizika a rozdělení výnosů s využitím genetického algoritmu* (1). Cílem práce bylo vytvoření dvoufázového modelu, kde v první fázi bylo odhadnuto vhodné rozdělení výnosů a v druhé fázi aplikován genetický algoritmus k řešení Markowitzova modelu.

Výchozím bodem pro celou analýzu byly zejména práce M. Erlinga (2) a C. J. Adcocka (3), kteří doporučují využívat pro popis rozdělení výnosů akcií zešikmené studentovo rozdělení. M. Erling pak testoval poměrně velkou škálu rozdělení na indexech a fondech různých trhů

a zjistil tak, že zešikmené studentovo rozdělení je vhodné pro využití v oblasti financí. Dalším takovým rozdělením pak bylo normální-inverzní Gaussovské rozdělení. V tomto článku bude částečně zreplicována jmenovaná analýza, ovšem se zaměřením na český trh. Vybrána k tomu byla následující rozdělení:

1.1 Normální rozdělení

Nejznámějším pravděpodobnostním rozdělením je bezesporu normální rozdělení. Toto rozdělení jistě není třeba podrobně představovat, jelikož se s ním každý setkal, i když o tom neví. Uved' me však formálně jeho hustotu pravděpodobnosti:

$$f_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

Toto rozdělení je určeno parametrem plohu μ a parametrem určujícím rozptyl σ^2 .

1.2 Zešikmené normální rozdělení

Poněkud zajímavější variantou normálního rozdělení je jeho zešikmená verze. Toto rozdělení by mohlo lépe popisovat rozdělení výnosů z důvodu zavedení parametru pro ovlivnění šikmosti rozdělení. Hustota tohoto rozdělení je následující:

$$f_{zN}(x) = \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \int_{-\infty}^{\alpha\left(\frac{x-\mu}{\sigma}\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (2)$$

K parametrům určujícím polohu a rozptyl zde přibyl parametr určující právě zešikmení α .

1.3 Zobecněné studentovo rozdělení

Dalším ve statistice hojně používaným rozdělením, které se také objevuje ve světě financí, je studentovo t. Jedním z možných zápisů jeho hustoty, v tomto případě za pomoci Beta funkce, je:

$$f_{t'}(x) = \frac{1}{\sigma\sqrt{v}B\left(\frac{v}{2}, \frac{1}{2}\right)} \left(1 + \frac{1}{v} \left(\frac{x-\mu}{\sigma}\right)^2\right)^{-(1+v)/2}, \quad (3)$$

Kde $B(x, y)$ je Beta funkce.

Parametry tohoto rozdělení jsou parametr polohy μ , parametr rozptylu σ^2 a počet stupňů volnosti v . Pokud se položí $\mu = 0$ a $\sigma^2 = 1$, dostaneme původní studentovo rozdělení.

1.4 Zešikmené studentovo rozdělení

Dalším a vyšším zobecněním studentova rozdělení je jeho zešikmená verze. Opět je zde přidán parametr pro zešikmení q . Jedná se o rozdělení, které má již 4 parametry a jeho hustota je následující:

$$f_{zt}(x) = \frac{\Gamma(\frac{1}{2} + q)}{v\sigma(\pi q)^{1/2}\Gamma(q) \left(\frac{|x-\mu+m|^2}{q(v\sigma)^2(\lambda \text{sign}(x-\mu+m)+1)^2} + 1 \right)^{\frac{1}{2}+q}}, \quad (4)$$

kde

$$m = \frac{2v\sigma\lambda q^{1/2}\Gamma(q - \frac{1}{2})}{\pi^{1/2}\Gamma(q + \frac{1}{2})}, \quad (5)$$

$$v = q^{-\frac{1}{2}} \left[(3\lambda^2 + 1) \left(\frac{1}{2q-2} \right) - \frac{4\lambda^2}{\pi} \left(\frac{\Gamma(q - \frac{1}{2})}{\Gamma(q)} \right)^2 \right]^{-\frac{1}{2}} \quad (6)$$

a $\Gamma(x)$ je Gamma funkce.

1.5 Zešikmené, zobecněné studentovo rozdělení

Ještě větším zobecněním studentova rozdělení je zešikmené, zobecněné studentovo rozdělení. Je zde přidán parametr p , který ještě více modifikuje „tvar“ rozdělení.

$$f_{zzt}(x) = \frac{p}{2v\sigma q^{1/p} B\left(\frac{1}{p}, q\right) \left(\frac{|x-\mu+m|^p}{q(v\sigma)^p(\lambda \text{sign}(x-\mu+m)+1)^p} + 1 \right)^{\frac{1}{p}+q}}, \quad (7)$$

kde

$$m = \frac{2v\sigma\lambda q^{1/p} B\left(\frac{2}{p}, q - \frac{1}{p}\right)}{B\left(\frac{1}{p}, q\right)}, \quad (8)$$

$$v = q^{-\frac{1}{p}} \left[(3\lambda^2 + 1) \left(\frac{B\left(\frac{3}{p}, q - \frac{2}{p}\right)}{B\left(\frac{1}{p}, q\right)} \right) - 4\lambda^2 \left(\frac{B\left(\frac{2}{p}, q - \frac{1}{p}\right)}{B\left(\frac{1}{p}, q\right)} \right)^2 \right]^{-\frac{1}{2}}. \quad (9)$$

Toto zobecnění studentova rozdělení je nejobecnější ze všech zmiňovaných, ze kterého vhodnou fixací některých parametrů dostaneme jakékoliv jiné. Tyto fixace jsou spolu s podrobnějším popisem rozdělení hezky zapsány v dokumentaci balíčku do R od Cartera Davise (4).

1.6 Normální-inverzní Gaussovské rozdělení

Posledním rozdělením, pomocí kterého se budeme snažit popsat rozdělení výnosů akcií je normální-inverzní Gaussovské rozdělení (NIG). Jeho hustota je:

$$f_{NIG}(x) = \frac{\alpha \delta K_1 \left(\alpha \sqrt{\delta^2 + (x - \mu)^2} \right)}{\pi \sqrt{\delta^2 + (x - \mu)^2}} e^{\delta \gamma + \beta (x - \mu)}, \quad (10)$$

kde K_1 je modifikovaná Besselova funkce třetího druhu a $\gamma = \sqrt{\alpha^2 - \beta^2}$.

Parametry určujícími chování tohoto rozdělení jsou parametr polohy μ , parametr rozptylu δ , parametr určující zešikmení β a nakonec parametr určující, jak má rozdělení těžké konce α .

2 ANALÝZA ROZDĚLENÍ VÝNOSŮ

Tato část článku se dostává k samotné analýze rozdělení výnosů akcií na českém trhu. Analyzované hodnoty výnosů patří 8 vybraným firmám obchodovaným na Pražské burze. Jedná se o ty největší firmy, které jsou také zařazeny v indexu PX. Konkrétně se jedná o firmy Erste, Avast, ČEZ, Kofola, KB, MONETA, Philip Morris ČR (PM) a Viena Insurance Group (VIG).

Data byla stažena z portálu [Patria](https://www.patria.cz/)¹ ve formě závěrečných, denních kurzů z období od 1. 1. 2012 do 17. 3. 2022, pokud byly tyto kurzy k dispozici. U těch akcií, kde nebyly k dispozici kurzy za celých 10 let, bylo použito maximální dostupné množství dat. 10leté období bylo vybráno, aby byl dostatek dat pro odhad rozdělení výnosů.

Tabulka 1: Hodnoty AIC dle rozdělení a akcie

	ERSTE	AVAST	ČEZ	KOFOLA	KB	MON.	PM	VIG
N	-12488	-4637	-14765	-9260	-14121	-8319	-1591	-14302
zN	-12486	-4643	-14763	-9318	-14115	-8317	-1591	-14378
t'	-12961	-5006	-15158	-9957	-14571	-9107	-1677	-14635
zt	-12956	-5001	-15153	-9951	-14565	-9102	-16766	-14630
zzt	-12954	-5000	-15159	-10187	-14582	-9100	-16821	-14638
NIG	-12949	-4998	-15157	-9951	-14575	-9094	-16766	-14634

Pozn.^a Tučně zobrazené hodnoty s nejnižším AIC pro jednotlivé akcie.

Zdroj: vlastní zpracování

První statistikou, pomocí které jsou srovnána jednotlivá rozdělení pro všechny firmy je AIC. Hodnoty pro jednotlivá rozdělení a firmy jsou zaznamenány v tabulce 1. Tučně jsou vyznačené hodnoty s nejnižší hodnotou AIC pro jednotlivé firmy. Na první pohled je vidět, že se hod-

¹<https://www.patria.cz/> [cit. 10. 11. 2022]

noty AIC pro odlišné firmy velmi výrazně liší. Je to způsobeno nejen tím, že statistika AIC je bezrozměrná, ale také tím, že pro jednotlivé akcie je k dispozici různý počet pozorování.

Nejllepšími rozděleními výnosů sledovaných akcií dle hodnot AIC se jeví zobecněné studentovo rozdělení a zešikmené, zobecněné studentovo rozdělení. První jmenované má nejlepší statistiku AIC pro 3 ze sledovaných firem, zatímco druhé je nejlepší pro zbylých 5. Nutno však podotknout, že ve většině případů se hodnoty AIC u rozdělení t' , zt , zzt a NIG liší jen nepatrně na rozdíl od rozdělení normálního a jeho zešikmené varianty, které dosahují stabilně významně nižších hodnot.

Podle hodnot AIC se zdá, že existují rozdělení lepší, nežli je normální. Pojd' me se však podívat také na Kolmogorovu-Smirnovovu statistiku KS , abychom měli porovnání i z jiného úhlu pohledu. Zaznamenané hodnoty KS jsou zapsané v tabulce 2. Tučně zvýrazněné jsou opět hodnoty pro nejlepší rozdělení (minimální KS). Dále jsou v tabulce kritické hodnoty testu o rozdělení KS^* závislé na počtu pozorování n .

Tabulka 2: Hodnoty KS dle rozdělení a akcie

	ERSTE	AVAST	ČEZ	KOFOLA	KB	MONETA	PM	VIG
n	2555	963	2555	1574	2555	1469	2555	2555
KS^*	322	525	322	411	322	425	322	322
N	57*	99*	25*	852	235*	258*	384	200*
zN	57*	60*	25*	807	235*	258*	384	67*
t'	29*	546	109*	846	203*	308*	454	103*
zt	46*	540	119*	848	206*	307*	455	94*
zzt	45*	530	152*	970	204*	302*	375	111*
NIG	42*	505*	124*	843	206*	290*	435	94*

Pozn.^a Hodnoty KS včetně KS^* jsou v řádu 10^{-4} .

^b KS^* je kritická hodnota Kolmogorova-Smirnovova testu pro $\alpha = 5\%$

^c Tučně zobrazené hodnoty s nejnižší hodnotou KS pro jednotlivé akcie

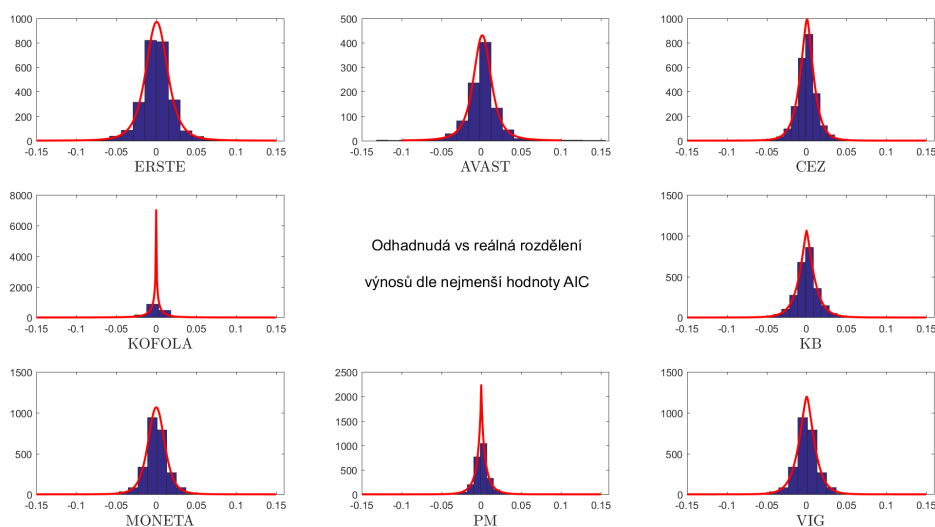
^d U hodnot s hvězdičkou nezamítáme hypotézu o tom, že výnosy pochází z daného rozdělení

Zdroj: vlastní zpracování

V tabulce 2 je vidět, že se hodnocení vhodnosti rozdělení pro daný výběr dle kritéria KS rozchází s AIC . Jedinými dvěma akciemi, kde se kritéria schází, jsou ERSTE a PM, u kterého je však na 5% hladině významnosti zamítnuta hypotéza, že výnosy pochází ze zešikmeného, zobecněného studentova rozdělení. V ostatních případech, kromě KB, kde je jako nejlepší rozdělení zvoleno t' , se jako nejlepší jeví zešikmené normální rozdělení. Toto rozdělení dosahuje nejnižších hodnot KS u 5 z 8 sledovaných firem, přičemž stejnou hodnotu má i normální rozdělení u firem ČEZ a MONETA.

Při pohledu na Kolmogorův-Smirnovův test o původu náhodného výběru z daného rozdělení je vidět, že kromě akcií AVAST, kde pro rozdělení N , zN a NIG není zamítnuta nulová hypotéza a u všech ostatních je zamítnuta, pro jednotlivé akcie u všech rozdělení je nulová hypotéza buď zamítnuta, nebo ne. Konkrétně u firem ERSTE, ČEZ, KB, MONETA, VIG všechna rozdělení prošla testem jako vhodná, naopak u firem KOFOLA a PM není vhodné ani jedno ze sledovaných rozdělení.

Od obou sledovaných statistik byly získány odlišné indicie k rozhodnutí o tom, která rozdělení jsou nelepší pro popsání výnosů z jednotlivých akcií na českém trhu. Jelikož však KS-test ve většině případů nezamítl hypotézu o původu náhodného výběru a hodnoty AIC normálního a zešíkmeného rozdělení jsou významně nižší než u ostatních, lze za 2 nejlepší rozdělení určit zobecněné studentovo a zešíkmené, zobecněné studentovo rozdělení. Pojd' me si nakonec zobrazit histogramy jednotlivých rozdělení spolu s nafitovaným, odhadnutým rozdělením s nejlepším AIC .



Obrázek 1: Skutečné vs odhadnuté hodnoty rozdělení výnosů jednotlivých firem

Zdroj: vlastní zpracování

Na obrázku 1 je vidět, že ve většině případů odhadnutá rozdělení sedí na skutečné hodnoty, nebo se to alespoň na pohled zdá. Tento odhad byl potvrzen již dříve KS-testem. Jedinými výjimkami, jak již samotný test v předešlé části napověděl jsou výnosy firem *KOFOLA* a *PM*. Zde je již na první pohled patrné, že zvolená rozdělení nesedí na hodnoty výnosů. V obou případech je odhadnuté rozdělení mnohem špičatější, čímž je jeho hustota mnohem větší okolo empirických hodnot s nejvyšší hustotou. Je to způsobeno nízkou likviditou těchto akcií, díky čemuž se některé dny akcie neobchodovaly a výsledný výnos tak byl nulový, odhadnutá rozdělení jsou tak díky velkému počtu 0 špičatá. Ve výsledku se však vychýlily odhadnuté hodnoty od reálných až příliš, což se odráží právě i v rozdílu distribučních funkcí, tedy i v hodnotě KS . Pro tyto 2 konkrétní firmy by bylo zřejmě lepší využít nějaké jiné pravděpodobnostní rozdělení s vysokou

špičatostí, jako například Laplaceovo rozdělení. Další variantou by pak bylo ignorovat nulové výnosy, nebo zadefinovat rozdělení, jehož hustota je s určitou pravděpodobností rovna 0, jinak má například zešikmené, zobecněné studentovo rozdělení.

3 Závěr

V článku byla představena některá pravděpodobnostní rozdělení, která se nejčastěji využívají k popisu výnosů akcií. Cílem bylo otestovat, jaké rozdělení mají výnosy akcií na českém trhu. Vybráno k tomu bylo 8 největších firem obchodovaných na pražské burze, které jsou zároveň obsaženy ve výpočtu indexu PX. K testování vhodnosti rozdělení byly využity statistiky *AIC* a *KS*.

Nejlépeších výsledků dle *AIC* dosahovaly zejména zobecněné studentovo rozdělení a zešikmené, zobecněné studentovo rozdělení. Oproti výsledkům M. Erlinga se tedy na českém trhu vyplatí využít o jeden parametr komplexnější rozdělení, než je zešikmené studentovo rozdělení. Výsledky statistik *KS* jsou na českém trhu zkreslené díky velmi nízké likviditě, důsledkem čehož je velký výskyt nulových hodnot výnosů.

Použitá literatura

1. NEUGEBAUER, Jakub; BOROVIČKA, Adam; BERANOVÁ, Lucie. *Optimalizace portfolia za různých měr rizika a rozdělení výnosů s využitím genetického algoritmu*. 2022. Dostupné také z: <https://insis.vse.cz/zp/78193/priloha/23885>. PhD Thesis. Diplomová práce (Ing.)—Vysoká škola ekonomická v Praze. Fakulta informatiky a statistiky, 2022.
2. ELING, Martin. Fitting asset returns to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics*. 2014, roč. 59, s. 45–56.
3. ADCOCK, Chris J. Asset pricing and portfolio selection based on the multivariate extended skew-Student-t distribution. *Annals of Operations Research*. 2010, roč. 176, č. 1, s. 221–234.
4. DAVIS, Carter. *The skewed generalized t distribution tree package vignette*. Chicago, 2015.

Kontakt

Ing. Jakub Neugebauer
Vysoká škola ekonomická, Fakulta informatiky a statistiky
nám. Winstona Churchilla 1938/4, 120 00 Praha 3-Žižkov
Tel: +420 725 516 362
email: jakub.neugebauer@vse.cz

MODEL VÝBERU PORTFÓLIA S VYUŽITÍM MIER RIZIKA CVAR A PRIEMERNÝ DRAW DOWN

PORTFOLIO SELECTION MODEL USING CVAR AND AVERAGE DRAW DOWN RISK MEASURES

Juraj Pekár, Ivan Brezina, Marian Reiff

Abstrakt

Na určenie optimálnej stratégie investora možno použiť rôzne rozhodovacie modely výberu portfólia, ktoré umožňujú diverzifikovať aktíva s cieľom zníženia celkového rizika uvažovanej investície a maximalizácii výnosu. Pri minimalizácii rizika možno využiť viacero mier rizika, ktoré sa dopĺňajú a poskytujú väčšie zabezpečenie investície rizika tak, aby adekvátne odzrkadľovali krátkodobé a dlhodobé prepady zodpovedajúceho aktíva. Príspevok sa orientuje na vytvorenie modelu výberu portfólia rešpektujúceho riziko pri investovaní, pričom boli zvolené miery rizika Conditional Value at Risk (CVaR) a priemerný DrawDown (AvDD). Výsledkom riešenia modelu je pri určených parametroch modelu získanie množiny takých efektívnych riešení, ktoré odzrkadľujú rôzne kombinácie stanovených hodnôt očakávaných výnosov a rizík.

Kľúčové slová: výber portfólia, CVaR, AvDD

Abstract

To determine the optimal strategy of the investor, various decision-making models of portfolio selection can be used, which allow for the diversification of assets intending to reduce the overall risk of the considered investment and maximize the return. When minimizing the risk, several measures can be used, which complement and provide greater risk assurance in such a way that they adequately reflect the short-term and long-term declines of the corresponding risk. The paper focuses on the creation of a portfolio selection model that respects investment risk, while the risk measures of conditional value at risk (CVaR) and average drawdown (AvDD) were chosen. The result of solving the model is, for the specified parameters of the model, obtaining a set of such effective solutions that reflect various combinations of determined values of expected returns and risks.

Keywords: Portfolio Selection, CVaR, AvDD

ÚVOD

Modely výberu portfólia slúžia k diverzifikácii aktív s cieľom minimalizácie rizika a maximalizácie výnosu. V prípade, keď treba zabezpečiť minimalizáciu strát za jedno obdobie, ale taktiež minimalizovať priemerný prepád kapitálu za celé analyzované obdobie možno formulovať úlohu s tromi kritériami, ktoré vyjadrujú ciele:

- maximalizovať očakávaný výnos,
- minimalizovať krátkodobé straty pomocou vybranej miery rizika Conditional Value at Risk (CVaR),
- minimalizovať možné dlhodobé prepady kapitálu využitím zvolenej miery rizika priemerný DrawDown (AvDD).

Výsledkom riešenia modelu zostaveného pri uvedených cieľoch a pri určených parametroch modelu bude množina efektívnych riešení odzrkadľujúcich rôzne kombinácie stanovených hodnôt očakávaných výnosov a rizík (Pekár 2015).

V prvej časti príspevku prezentujeme spôsob výpočtu vstupných parametrov modelu ako aj miery očakávaných výnosov a rizík využitých pri konštrukcii modelu. Následne definujeme model, ktorý slúži pri jeho určených parametroch ku generovaniu efektívnych portfólií pri zohľadnení uvedených troch kritérií.

1 OČAKÁVANÝ VÝNOS A MIERY RIZIKA

Pri konštrukcii nami formulovaného modelu je dôležité vypočítať vstupné dáta (vektor výnosov, vektor kumulovaného výnosov a spôsob získania hodnoty očakávaného výnosu aktíva).

Uvažujme s n aktívami a nech P_{jt} zodpovedá cene j -teho aktíva, $j = 1, 2, \dots, n$, v čase $t = 1, 2, \dots, T$, kde T je dĺžka sledovaného obdobia.

Mieru výnosu j -teho aktíva, $j = 1, 2, \dots, n$, medzi obdobia t a $t-1$ možno vyjadriť ako relatívnu mieru výnosu r_{jt} pre to určité časové obdobie (j), ktorú možno vypočítať ako

$$r_{jt} = \frac{P_{jt} - P_{j,t-1}}{P_{j,t-1}}. \quad (1)$$

Mieru kumulovaného výnosu j -teho aktíva, $j = 1, 2, \dots, n$, v čase t možno následne vyjadriť ako

$$y_{jt} = \frac{P_{jt} - P_{j1}}{P_{j1}}. \quad (2)$$

Očakávaný výnos j -teho aktíva, $j = 1, 2, \dots, n$, na základe hodnoty kumulovaného výnosu (2) je určený vzťahom

$$E_j = (1 + y_{jT})^{\frac{1}{T}} - 1. \quad (3)$$

Hodnotu miery rizika DrawDown $DD_j(t)$ j -teho aktíva (portfólia), $j = 1, 2, \dots, n$, v čase t , $t = 1, 2, \dots, T$, možno vypočítať ako rozdiel maximálnej hodnoty vo v -tom období ($v = 1, 2, \dots, t$) a hodnoty v čase t . Podľa (Brezina a kol. 2018) ho možno vypočítať

$$DD_j(t) = \max_{v=1,2,\dots,t} y_{jv} - y_{jt}. \quad (4)$$

Okrem uvažovaného ukazovateľa $DD_j(t)$ sa budeme ďalej zaoberať aj odvodenou mierou rizika (Brezina a kol. 2018), priemerným DrawDown ($AvDD$). Pretože je prirodzené, že investora predovšetkým zaujíma pravdepodobnosť, že investícia do aktíva nedosiahne očakávaný výnos, práve miera rizika priemerný DrawDown umožňuje sledovať priemerný prepád v hodnote zodpovedajúceho aktíva. Tento ukazovateľ je založený na historických hodnotách cien aktív a popisuje priemerné zníženie hodnoty aktíva, k akému počas sledovaného obdobia (T) prišlo. Priemerný DrawDown $AvDD_j(T)$ pre j -te aktívum, $j = 1, 2, \dots, n$, možno vypočítať

$$AvDD_j(T) = \sum_{t=1}^T \frac{DD_j(t)}{T}. \quad (5)$$

Okrem ukazovateľa dlhodobého prepádu kapitálu ($AvDD$) je v modeli využitý aj ukazovateľ krátkodobého prepádu kapitálu, ktorý je reprezentovaný v súčasnosti najpoužívanejšou mierou (Pekár a kol., 2022) rizika Conditional Value at Risk ($CVaR$). Pri výpočte tejto hodnoty sa vychádza z miery rizika VaR , ktorá je vo všeobecnosti definovaná ako maximálna

možná strata za dané časové obdobie na určitom intervale spoľahlivosti α (Yamai a Yoshida 2002, Alexandre 2008). Potom možno na základe hodnoty VaR definovať hodnotu $CVaR$ ako priemer zo strát vyšších ako je hodnota VaR . Formuláciu vzťahu na výpočet $CVaR$ (počas sledovaného obdobia T pre j -te aktívum), $j = 1, 2, \dots, n$, možno definovať (Rockafellar, Uryasev 2000, Pekár 2015) nasledujúcim spôsobom

$$CVaR_{\alpha}(\mathbf{r}_j) = VaR_{\alpha} - \frac{1}{\alpha} E\left[\left[\mathbf{r}_j + VaR_{\alpha} \mid -\right]\right], \quad (6)$$

kde \mathbf{r}_j reprezentuje vektor výnosov j -teho aktíva počas sledovaného obdobia T ,
 α predstavuje definovanú hladinu významnosti.

2 OPTIMALIZÁCIA PORTFÓLIA

V tejto časti uvedieme konštrukciu modelu založeného na miere očakávaného výnosu a mierach rizika $CVaR$ a $AvDD$, pričom v prvom rade definujeme premenné modelu a parametre využité pri jeho konštrukcii. Riešením sformulovaného modelu budú efektívne portfóliá, ktoré zohľadňujú, na rozdiel od klasických modelov výberu portfólia, dve miery rizika odzrkadľujúce krátkodobé a dlhodobé prepady kapitálu.

Pri konštrukcii modelu využijeme premenné $z_t \geq 0$ ($t = 1, 2, \dots, T$), ktoré nadobúdajú hodnotu rozdielu medzi VaR a výnosom portfólia v stave t , ak je výnos nižší ako VaR , alebo sa budú rovnať nule na základe hladiny významnosti α . Uvažujme ďalej, že vektor $\mathbf{u} = (u_0, u_1, \dots, u_T)$ voľných premenných predstavuje maximálnu hodnotu kumulovaného výnosu u_t v časovom období $0, 1, \dots, T$. Nech VaR_{α} je premenná, ktorá reprezentuje prahovú hodnotu pre výnos portfólia na základe hladiny významnosti α . Okrem uvedených voľných premenných vystupujú v modeli taktiež premenné w_1, w_2, \dots, w_n , ktoré nadobudnú optimálne hodnoty váh jednotlivých aktív ($j = 1, 2, \dots, n$).

Okrem premenných modelu treba definovať vstupné parametre dané investorom. Okrem výnosov, kumulovaných výnosov a očakávaných výnosov jednotlivých aktív tak treba stanoviť hodnoty parametrov $AvDD_p$, ktorá predstavuje priemernú povolenú hodnotu straty kapitálu portfólia (teda priemernú hodnotu strát určenú investorom, pod ktorú nesmie priemerná strata klesnúť, p označuje, že ide o vstupný parameter) a hodnoty parametrov E_p označujúcich minimálny očakávaný výnos portfólia (teda hodnotu určenú investorom, ktorú musí vypočítané portfólio dosiahnuť, p opäť označuje, že ide o vstupný parameter). Kombináciou uvedených dvoch parametrov možno získať množinu efektívnych portfólií, ktoré poskytujú investorovi návrhy investičných stratégií pre rôzne hodnoty požadovaných výnosov a rizika vyjadreného pomocou $AvDD$.

Potom možno trojkriteriálnu úlohu (sú v nej rešpektované všetky tri formulované ciele) transformovať na minimalizáciu rizika $CVaR$ s ohľadom na ohraničenia maximálnej povolenej straty kapitálu a minimálnej požadovanej hodnoty očakávaného výnosu, na úlohu matematického programovania:

$$\begin{aligned}
\min CVaR(w_1, w_2, \dots, w_n, z_1, z_2, \dots, z_T, u_0, u_1, \dots, u_T) &= VaR_\alpha + \frac{1}{\alpha T} \sum_{t=1}^T z_t \\
z_t + \sum_{j=1}^n r_{jt} w_j + VaR_\alpha &\geq 0, t = 1, 2, \dots, T \\
\sum_{j=1}^n E_j w_j &\geq E_p \\
\frac{1}{T} \sum_{t=1}^T \left(u_t - \sum_{j=1}^n y_{jt} w_j \right) &\leq AvDD_p \\
u_t &\geq \sum_{j=1}^n y_{jt} w_j, t = 1, 2, \dots, T \\
u_t &\geq u_{t-1}, t = 1, 2, \dots, T \\
u_0 &= 0 \\
\sum_{j=1}^n w_j &= 1 \\
w_1, w_2, \dots, w_n &\geq 0, z_1, z_2, \dots, z_T \geq 0
\end{aligned} \tag{7}$$

ZÁVER

Príspevok sa zaoberá modifikáciou modelu výberu portfólia na báze kritérií maximalizácia očakávaného výnosu, minimalizácia rizika na základe mier $CVaR$ a priemerný DrawDown $AvDD$, v ktorom na rozdiel od modelov známych z literatúry uvažujeme s tromi kritériami. Uvedený model zohľadňuje dve miery rizika, čo do väčšej miery zabezpečí zníženie celkového rizika investície. Testovanie uvedeného modelu bolo realizované v programovom jazyku Python. Výsledkom celého procesu je získanie efektívnych riešení pre uvedený model v programovom prostredí Python (module MIP), ktorý dokumentuje možnosť výpočtu riešenia, ktoré je podkladom pre rozhodnutie investora.

Príspevok bol spracovaný v rámci riešenia grantovej úlohy – VEGA 1/0339/20 „Využitie skrytého Markovovho modelu vo finančnom modelovaní“.

Použitá literatúra

- [1] ALEXANDE, C. 2008. *Value at Risk Models, Market Risk Analysis*, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,.
- [2] BREZINA, I., PEKÁR, J., 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* (pp. 263–278). London: Risk Books.
- [4] GROSSMAN, S. J. and ZHOU, Z. 1993. Optimal investment strategies for controlling drawdowns. *Mathematical Finance*, 3, s. 241–276 .
- [5] PEKÁR, J. 2015. Modely matematického programovania na výber portfólia. Bratislava: Vydavateľstvo EKONÓM.

- [6] PEKÁR, J., BREZINA, I., 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, s. 421-433.
- [8] PEKÁR, J., BREZINA, I. and REIFF, M. 2022. Determining the Investor's Strategy During the COVID-19 Crisis Based on CVaR Risk Measure. In: *Strategic Management and Decision Support Systems in Strategic Management SM 2022*, Subotica : University of Novi Sad s. 7-13.
- [9] ROCKAFELLAR, T. R. – URYASEV, S. 2000 Optimization of Conditional Value-at-Risk, *The Journal of Risk*, 2(3) s. 21-42.

Kontaktné údaje

prof. Mgr. Juraj Pekár, PhD.
Ekonomická univerzita v Bratislave,
Fakulta hospodárskej informatiky
Katedra operačného výskumu a ekonometrie
Dolnozemska 1/b, 852 35 Bratislava
Slovenská republika
email: juraj.pekar@euba.sk

prof. Ing. Ivan Brezina, CSc.
Ekonomická univerzita v Bratislave,
Fakulta hospodárskej informatiky
Katedra operačného výskumu a ekonometrie
Dolnozemska 1/b, 852 35 Bratislava
Slovenská republika
email: ivan.brezina@euba.sk

doc. Ing. Marian Reiff, PhD.
Ekonomická univerzita v Bratislave,
Fakulta hospodárskej informatiky
Katedra operačného výskumu a ekonometrie
Dolnozemska 1/b, 852 35 Bratislava
Slovenská republika
email: marian.reiff@euba.sk

OPAKOVANÉ HRY V TEÓRII HIER

REPEATED GAMES IN GAME THEORY

Allan Jose Sequeira Lopez

Abstrakt

V teórii hier nájdeme viacero typov hier. Opakovaná hra je taká, v ktorej sa akcie vykonávajú viackrát a výsledok hry je časom ovplyvnený; To znamená, že každý hráč má možnosť získať si reputáciu svojím správaním a má tiež možnosť študovať stratégiu konkurentov v priebehu hry. V tomto článku predstavujeme opakovanú hru na analyzujeme možnosť ich aplikácie v ekonomickej oblasti.

Kľúčové slová: hra, opakovanie, stratégie

Abstract

In the game theory we can find multiple types of games. A repeated game is one in which actions are performed more than once and the outcome of the game is affected over time. It means that each player can earn a reputation for his behavior and has an option to study the strategy of competitors during the game. In this article we present repeated games and we analyze the possibility of applying them to the economic field.

Keywords: game, repetition, strategies

1 ÚVOD

Existuje niekoľko typov hier. Pre hry s nulovým súčtom platí, že to, čo jeden hráč získa, stratí druhý. V hrách s nenulovým súčtom naopak platí, že čo jeden hráč získa, nie je nevyhnutne to, čo stratí ten druhý. Ďalej poznáme hry, v ktorých sa berú do úvahy iba čisté stratégie a tiež hry, v ktorých sú povolené zmiešané stratégie (t. j. rozdelenie pravdepodobnosti čistých stratégií). Ďalším typom hier sú kooperatívne hry, v ktorých môžu agenti komunikovať, aby rokovali o dohode, alebo nekooperatívne hry, v ktorých sa jednotlivci rozhodujú nezávisle.

Tento príspevok v rámci nekooperatívnych hier skúma možnosti spolupráce medzi jednotlivcami pri opakovaní hry; to znamená ide o hry, v ktorých hráči interagujú viac ako raz. Opakovaná hra je taká, v ktorej sa akcie vykonávajú viackrát a výsledok hry je časom ovplyvnený. To znamená, že každý hráč má možnosť získať reputáciu.

Opakované hry môžu mať konečný časový horizont, kde sa hra opakuje konečný a známy počet krát, alebo môžu byť hry s nekonečným časovým horizontom, čo znamená, že hra sa opakuje počas nekonečných období alebo v ekvivalentných termínoch, hra sa opakuje konečný počet krát, ale trvanie hry nie je pre hráčov známe (Möbius, 2007).

1.1 Konečný časový horizont

Uvažujeme o hre, ktorá sa hrá iba raz, pričom v tejto hre existujú možnosti spolupráce, ale nezodpovedajú situácii Nashovej rovnováhy (Dlouhý, Fiala, 2009). Zaujímá nás, či je možné udržať spoluprácu ako rovnováhu, keď sa hra opakuje konečný a známy počet krát. Keďže hráči poznajú trvanie hry, retroaktívne indukčné uvažovanie nám umožní potvrdiť, že nebude možné udržať rovnováhu pri opakovanej hre okrem herného obdobia, ktoré nasleduje po období

krátkodobého rovnovážneho správania (toho, ktoré zodpovedá hre, ktorá sa hrá iba raz). Začnime tým, že sa pozrieme na posledné obdobie. Keďže hra sa skončila a nie je žiadne pokračovanie, v poslednom období bude každý hráč v krátkodobom horizonte sledovať svoje optimálne správanie (v poslednom období zostáva hrať iba hra, ktorá sa hrá iba raz) a nebude spolupracovať so súperom. V predchádzajúcom období sa každý hráč musí rozhodnúť, či bude alebo nebude spolupracovať s predpokladom, že v poslednom období nebude spolupracovať, takže hráči sa budú riadiť optimálnym správaním aj v krátkodobom horizonte. Preto sa Nashova rovnováha bude v krátkodobom horizonte opakovať toľkokrát, koľkokrát sa hra opakuje¹. V jednom bode dokonalej rovnováhy v podhrách² sa bude každý hráč správať obdobie po období tak, ako by sa správal v krátkodobom horizonte. Preto spolupráca medzi hráčmi sa neudrží ako rovnováha, ak je časový horizont konečný (Sequeira Lopez, 2020).

1.2 Nekonečný časový horizont

Keď sa situácia opakuje v priebehu času, správanie ľudí môže ovplyvniť rozhodnutia, ktoré urobia v budúcnosti. To je presne to, čo sa stane, keď sa hra opakuje nekonečne veľa období. Keď hráč porovnáva rôzne typy správania, musí porovnať diskontovanú súčasnú hodnotu, ktorú by získal s každým z nich. Označme δ diskontný faktor, ktorý je definovaný ako:

$$\delta = \frac{1}{1+r}$$

kde r je typ diskontu. Ak $\delta = 0$ (čo zodpovedá $r = \infty$), hráča zaujíma iba prítomnosť a ak $\delta = 1$ (čo zodpovedá $r = 0$), hráča zaujíma rovnako budúcnosť aj prítomnosť. V tomto type hier je niekedy možné udržiavať ako rovnováhu správanie, ktoré nie je v krátkodobom horizonte rovnovážne, a to zahrnutím hrozieb trestu v prípade porušenia dohody o spolupráci (Abreu, 1988).

Na udržanie spolupráce v opakovaných hrách môžu hráči v priebehu hry použiť rôzne stratégie, ako napríklad *Trigger* stratégia (ang. *Trigger strategy*) alebo stratégia „oko za oko“ (ang. *Tit for tat strategy*).

Trigger stratégia spočíva v tom, že hráč spolupracuje v prvom období a v každom ďalšom, ktoré nasleduje po období spolupráce a nespupracuje počas zvyšku hry, ak sa predtým niektorý hráč odchyľil od spolupráce. Dalo by sa povedať, že hráč, ktorý nespupracuje, je vystavený nekonečnému trestu.

Stratégia *Tit for tat* sa stretáva s prísloviem *oko za oko, zub za zub*. Hráč využívajúci túto stratégiu bude najskôr spolupracovať a potom zopakuje predchádzajúcu akciu protivníka. To znamená, že ak sa hráč odchyľil od spolupráce v jednom období, súper ho potrestá v nasledujúcom období, ale po uplynutí trestu je možné sa vrátiť k spolupráci.

V tomto príspevku sa budeme zaoberať iba *Trigger* stratégiami, pretože sú najpoužívanejšie pre analýzu ekonomických problémov.

2 EKONOMICKÉ APLIKÁCIE VÄZŇOVEJ DILEMY

Väznova dilema je základným problémom teórie hier. Hoci ho pôvodne formulovali matematici Merri M. Flood a Melvin Dresher v 50. rokoch, až Albert W. Tucker z Princetonskej univerzity, odborník na lineárne programovanie a teóriu hier, dokončil

¹ Nashova rovnováha je kombináciou stratégií, v ktorých je výsledok pre každého z hráčov lepší alebo rovný výsledku, ktorý by dosiahli, pričom zostáva konštantná hra ostatných, čo hrajú inú stratégiu.

² Ťah alebo kombinácia stratégií, čo je Nashova rovnováha, predstavuje dokonalú rovnováhu v podhrách, ak príslušné časti rovnovážnych stratégií každého z hráčov sú tiež rovnovážne pre každú z podhier.

formuláciu a pomenovanie tejto dilemy. Jeho cieľom bolo ilustrovat' zložitost' analýzy situácií, v ktorých víťazstvo jedného hráča nie je nevyhnutne porážkou iného. To znamená, že je to príklad hier s nenulovým súčtom. Odvtedy sa teória hier aplikuje na rôzne oblasti, ako je filozofia, biológia, psychológia alebo ekonómia.

Väzňova dilema sa analyzuje v situáciách, keď pri rozhodovaní dochádza k individuálnym a kolektívnym konfliktom záujmov. Pomocou tejto dilemy pokračujeme v overovaní, že spoluprácu nie je možné udržať ako rovnováhu, keď sa hra opakuje konečný a známy počet krát, ale je možné udržať spoluprácu ako rovnováhu, keď je časový horizont nekonečný.

Matica rozhodnutí predstavuje roky väznenia v hre a jej výsledky sú nasledovné:

	Hráč 2	
Hráč 1	C	NC
C	(-6, -6)	(0, -10)
NC	(-10, 0)	(-1, -1)

(C – prizná sa, NC – neprizná sa)

Vzhľadom na to, že čísla zastúpené v matici rozhodnutí sú roky väznenia hráčov a že hráči sú racionálni, ich cieľom bude maximalizovať ich zisk a byť odsúdení na menej rokov väznenia.

Keď sa hra hrá iba raz, (C, C) je Nashova rovnováha v dominantných stratégiách. To znamená, že pred akoukoľvek stratégiou druhého hráča je najlepšou odpoveďou každého z hráčov priznať sa. Hráči by však dosiahli lepšie výsledky, ak by boli schopní spolupracovať hraním stratégie (NC, NC), keďže táto kombinácia stratégií je Pareto optimálna a tiež maximalizuje celkové zisky.

Ďalej sa pozrieme na možnosti spolupráce, keď sa hra opakuje, rozlišujúc medzi konečným časovým horizontom a nekonečným časovým horizontom. Ak aplikujeme väzňovu dilemu na opakovanú hru s konečným časovým horizontom, nastane nasledovné:

Počet opakovaní hry je daný, nazvime ho T . Ak $T = 3$, znamená to, že ak sa hra opakuje tretíkrát, hráči už vedia, že po treťom opakovaní už hra končí. V poslednom období dospejeme k rovnakému záveru ako v klasickej väzňovej dileme a obaja hráči sa rozhodnú priznať. Stáva sa to preto, že hráči vedia, že hra končí v období 3 a nemajú záujem o zlepšenie vzťahu v budúcnosti, pretože neexistuje. Preto bude každý hráč hrať svoju dominantnú krátkodobú stratégiu, ktorou je „priznať sa“.

Ak sa namiesto posledného obdobia zaradíme do obdobia $T - 1$, hráči budú pokračovať v implementácii dominantnej stratégie hry (C, C), pretože vedia, že hra sa skončila a nemá zmysel spolupracovať na pokuse o zlepšenie v budúcnosti, pretože v poslednom období nebude žiadna spolupráca. Preto bez ohľadu na to, koľkokrát sa hra opakuje, v poslednom období bude každý hráč hrať svoju dominantnú stratégiu „priznať sa“ a jedine v dokonalej Nashovej rovnováhe v subhrách sa bude hrať stratégia (C, C) v každom období.

Z uvedeného vyplýva, že ak aplikujeme väzňovu dilemu na hru s konečným časovým horizontom, znamená to, že keď sa hra opakuje konečný a známy počet krát, nebude možná ani spolupráca medzi hráčmi, hoci spolupráca maximalizuje celkové zisky.

Ak aplikujeme väzňovu dilemu na opakovanú hru s nekonečným časovým horizontom, stane sa nasledovné: Hra sa opakuje donekonečna, čo znamená, že hráči nevedia, kedy sa hra skončí, ale poznajú minulú históriu hry. Hráči musia porovnať očakávanú výplatu pri rôznych

stratégiách. Za týmto účelom porovnávajú diskontovanú súčasnú hodnotu zárobkov. Označíme δ diskontný faktor (pozri 1.2)

Každý hráč má jednu dominantnú krátkodobú stratégiu, stratégia hry „prizná sa“, ako sme videli vyššie. Teraz však uvidíme, či pri využití *Trigger* stratégie existuje nejaká kombinácia stratégií, ktorá je dlhodobou rovnováhou. Je to otázka spolupráce v každom období, ak ste predtým spolupracovali, čo znamená, že nikto z hráčov sa neprizná, ak vo všetkých predchádzajúcich obdobiach bola kombinácia stratégií hráčov (NC, NC). V opačnom prípade nebude po zvyšok hry žiadna spolupráca. Napríklad, ak sa hráč 1 odchýli od stratégie, hráč 2 bude hrať „prizná sa“ vo všetkých nasledujúcich obdobiach. Dalo by sa povedať, že hráč, ktorý sa odchýlil od spolupráce, je potrestaný po zvyšok hry.

Aby sa spolupráca udržala ako rovnováha a aby hráči nemali motiváciu odchyľovať sa, je potrebné skontrolovať, či kombinácia stratégií (*Trigger* stratégia, *Trigger* stratégia) je Nashovou rovnováhou opakovanej hry. Na tento účel sa porovnávajú platby, ktoré by hráči dostali v prípade spolupráce, a platby, ktoré by dostali v prípade odchýlenia sa od stratégie.

Ak sa hráči riadia stratégiami spolupráce, zodpovedajúce platby zohľadňujúce maticu platieb sú:

$$-1 + (-1)\delta + (-1)\delta^2 + \dots = -1(1 + \delta + \delta^2 + \dots) \quad (1)$$

Zjednodušene ich platby by boli dané:

$$-\frac{1}{1-\delta}$$

Predpokladajme, že jeden z hráčov nespupracuje v prvom období. Keďže druhý hráč sleduje *Trigger* stratégiu, v žiadnom čase v budúcnosti už nebude spolupracovať a kombinácia stratégií (krátkodobých) prvého obdobia bude (C, NC) a v ostatných obdobiach (C, C). Pozrime sa, aké budú boli platby:

$$0 + (-6)\delta + (-6)\delta^2 + \dots = 0 + (-6)(\delta + \delta^2 + \dots)$$

Zjednodušene ich platby by boli dané:

$$\frac{\delta(-6)}{1-\delta}$$

Teraz porovnáme platby získané v každom prípade, aby sme zistili, či zisk dosiahnutý porušením dohody o spolupráci kompenzuje stratu zisku počas zvyšku hry. To znamená, že kontrolujeme, či hráči nemajú motiváciu odchyliť sa.

Aby sa spolupráca udržala ako rovnováha, musí platiť:

$$-\frac{1}{1-\delta} \geq \frac{\delta(-6)}{1-\delta}$$

Vyrovnaním ziskov vypočítame kritický diskontný faktor:

$$-\frac{1}{1-\delta} = \frac{\bar{\delta}(-6)}{1-\bar{\delta}} \rightarrow -\frac{1(1-\bar{\delta})}{(1-\bar{\delta})} = \bar{\delta}(-6)$$

Presunutím $\bar{\delta}$ naľavo dostaneme hodnotu kritického diskontného faktora, ktorý je daný vzťahom:

$$\bar{\delta} = \frac{1}{6} = 0,1667$$

Keďže hra je symetrická a pre jednoduchosť sa domnievame, že obaja hráči majú rovnaký diskontný faktor, pokiaľ $\delta \geq \bar{\delta} = \frac{1}{6}$ ani jeden hráč nebude mať motiváciu odchyliť sa. Preto je možné zachovať spoluprácu ako rovnováhu prostredníctvom *Trigger* stratégií.

Záver

Zistili sme, že prostredníctvom *Trigger* stratégie je možné udržať ako rovnováhu spoluprácu v hre, ktorá sa nekonečne opakuje. Stáva sa to vďaka tomu, že v prípade nedodržania dohody o spolupráci je hráč potrestaný po zvyšok hry a výhody, ktoré získa v krátkodobom horizonte (v období nespolupráce), nekompensujú stratu počas zvyšku hry.

Príspevok bol spracovaný v rámci riešenia grantovej úlohy VEGA 1/0427/20 Viackriteriálne modely teórie hier v ekonómii a politológii

Použitá literatúra

1. ABREU, D. 1988. "On the Theory of Infinitely Repeated Games with Discounting". *Econometrica*, Vol. 56, No 2, p. 383.
2. DLOUHÝ, M. - FIALA, P. 2009. Úvod do teórie her. Praha : Vysoká škola ekonomická - Oeconomica, 2009. ISBN: 978-80-245-1609-7.
3. MÖBIUS, M. M. 2007. Lecture XIII: Repeated Games. Harvard University.
4. SEQUEIRA LOPEZ, A. J. 2020. Teória hier a evolučne stabilná stratégia. AIESA 2020. medzinárodná vedecká konferencia. AIESA – Budovanie spoločnosti založenej na vedomostiach : recenzovaný zborník - 19. medzinárodná vedecká konferencia organizovaná pod záštitou rektora Ekonomickej univerzity v Bratislave. Bratislava : Publisher Letra Edu, 2020. ISBN 978-80-89962-68-6.

Kontaktné údaje

Ing. Allan Jose Sequeira Lopez, PhD.
Ekonomická univerzita v Bratislave, Fakulta aplikovaných jazykov
Dolnozemska 1/b, 852 35 Bratislava
Tel: (+421) 918528099
email: allan.lopez@euba.sk

3D PROBLÉM OBCHODNÍHO CESTUJÍCÍHO

3D TRAVELLING SALESMAN PROBLEM

Ing. Vojtěch Vávra

Abstrakt

Tento článek se zabývá především problémem obchodního cestujícího ve 3D. V úvodní části je definován problém. V druhé části jsou popsány heuristiky a metaheuristiky, které byly použity k řešení představeného problému. V experimentální části byly implementovány vybrané heuristické a metaheuristické metody v prostředí jazyka R i s několika doplňky, jako je například vizualizace řešení. V rámci vyhodnocení jsou metody porovnávány mezi sebou a nejlepší doporučeny podle výpočetního času i hodnoty účelové funkce k řešení uvedeného problému.

Klíčová slova: 3D-TSP, problém obchodního cestujícího, heuristiky, metaheuristiky, jazyk R, mezihvězdné cestování

Abstract

This article deals primarily with travelling salesman problem in 3D. In introductory part we deliver formulation of the proposed problem. In the second part we deliver heuristic and metaheuristic methods for solving the proposed problem. In the experimental part were implemented selected heuristic and metaheuristic methods in the environment of language R with several addition for example visualization of the solution. In conclusion all used methods are compared and best recommended according to computational time and value of the objective function for solving proposed problem.

Keywords: 3D-TSP, travelling salesman problem, heuristics, metaheuristics, language R, star tour

1 ÚVOD

Práce se zaměřuje na aplikaci vybraných heuristik a metaheuristik k řešení problému obchodního cestujícího (TSP) v trojrozměrném prostoru (3D-TSP). Dalším z cílů práce je zhodnotit efektivitu vybraných metod a zvolit nejvhodnější pro řešení 3D verzi TSP. Dále se v práci identifikují kvality a nedostatky jednotlivých metod.

Druhotným úkolem práce je otestovat prostředí jazyka R, zda je vhodné k případné výuce metod pro řešení úlohy obchodního cestujícího a dalších obdobných dopravních problémů pro Katedru ekonometrie na FIS VŠE.

1.1 Definice řešeného problému

Dle teorie grafů je TSP model definován na grafu $G = (V, E)$, ve kterém uzly (nodes, případně vrcholy, V , vertex) představují místa, která mají být navštívena, a hrany (E , edges) znázorňují komunikační síť mezi všemi uzly.

Pro TSP hledáme Hamiltonův cyklus na grafu G tak, aby přesun po hranách E , které jsou cenově c_{ij} ohodnoceny (vzdálenost, přepravní náklady, doba přejezdu...), byl minimální.

Hamiltonův cyklus lze definovat jako uzavřenou cestu na grafu, ta obsahuje všechny uzly grafu právě jednou.

Oproti běžnému dvou dimenzionálnímu TSP však budeme uvažovat tři dimenzionální TSP a budeme ho označovat jako 3D-TSP. Vstupními daty jsou souřadnice hvězd (x, y, z) a vzdálenost mezi jednotlivými hvězdami je následně počítána Euklidovskou metrikou.

Jako řešení úlohy pak budeme označovat x , které bude dáno hodnotami $\{1, 2, \dots, n\}$, kde n označuje počet hvězd v řešeném dílčím problému. Jednotlivé hodnoty x pak budou označovat index hvězdy, která má být právě navštívena.

2 METODY

Metody aplikované v článku byly rozděleny do dvou skupin – heuristiky a metaheuristiky.

Heuristiky jsou algoritmy, které se používají pro řešení daného problému. Ve většině případů je jejich výpočetní složitost nízká, což snižuje i výpočetní dobu, ale šance nalezení optimálního řešení je mizivá. Používají se tedy k nalezení přípustného řešení blízkého optimálnímu řešení. Jako zdroj uvádíme (Pelikán, 2001) v kombinaci s (Reinelt, 2003).

Metaheuristiky jsou přístupy, které se používají pro řešení obecných problémů. Po drobných úpravách je lze použít na dané problémy jako je například diskretní TSP. Optimální řešení mohou nalézt, ale spíše naleznou řešení blízká optimálnímu (suboptimální). V porovnání s exaktními metodami jsou metaheuristiky rychlejší, ale méně přesné. Jako zdroj pro první 2 použité uvádíme zdroj (Pelikán, 2001) v kombinaci s (Reinelt, 2003) a pro následující uvádíme zdroj k jednotlivým zmínkám.

2.1 Heuristiky

Algoritmus nejbližšího souseda (Nearest neighbor, NN) se často zaměřuje s hladovým algoritmem, i když jeho hamižnost je významná a je zařazován do třídy hladových algoritmů. Jeho krása spočívá v jednoduchosti a počáteční efektivitě. Ta je ovšem ke konci algoritmu penalizována používáním delších a delších hran, které vedou k neefektivitě algoritmu. Smysl algoritmu spočívá v nalezení nejkratší hrany, kam se můžeme přesunout, a její následné použití k přesunu.

Hladový algoritmus (Greedy algorithm) se podobá algoritmu nejbližšího souseda. Hlavním rozdílem je, že nevybíráme pouze z hran, kam se můžeme přesunout z posledního navštíveného bodu, ale vždy vybíráme nejkratší přípustnou hranu na grafu. Zde se komplikuje algoritmus tím, že se musí kontrolovat, zda nedojde k vytvoření podcyklu, což by znemožnilo vytvoření Hamiltonova cyklu.

Metoda výhodnostních čísel (Clarke, Wright's savings method) spočívá ve výpočtu výhodnostní matice $S = s_{ij}$, která je složena z nejkratších hran mezi vrcholy v grafu. Vzorec pro výpočet matice S je dán vzorcem:

$$s_{ij} = d_{ij} + d_{1j} - d_{1i}, \quad \text{kde } i, j = 2, 3, \dots, n \quad \text{a } i \neq j.$$

Metoda minimální kostry (minimum spanning tree method, metoda MST) slouží pro symetrické úlohy. Výhodou tohoto algoritmu je, že vytváří menší skupinky uzlů, které spojí

nejkratší cestou. Problém ovšem nastává, když začne propojovat tyto skupinky. Kvůli tomu dochází k neefektivitě řešení. Tato metoda je ovšem díky pevnému základu vhodná pro následné úpravy pomocí algoritmů zlepšujících řešení. Součástí algoritmu je i úloha převodu Eulerova cyklu na Hamiltonův cyklus. To lze řešit Hierholzerovým algoritmem.

Metoda výměn (Lin Kernighen method) je určena pro zlepšování již nalezeného řešení. Vylepšení řešení je dosaženo pomocí výměny 2 hran za jiné 2 hrany (2-opt). Lze provést i výměny 3 a 3 hran (3-opt). Pro výpočet tedy již předpokládáme uspořádaný Hamiltonův cyklus C popsany jednotlivými uzly a množinou použitých hran E . Dnes je tato metoda považována za jednu z nejefektivnějších a specializuje se na ni Keld Helsgaun. Metoda končí, pokud již byly použity všechny dvojice hran a nebylo dosaženo lepšího řešení. Nebo lze také použít omezený počet výměn. V práci byla implementována verze 2-opt LK heuristiky.

2.2 Metaheuristiky

Metoda prahové akceptace (Threshold acceptance, TA) se lepší hodnota účelové funkce hned neoznačí jako nejlepší, nýbrž pouze tehdy, pokud bude vylepšeno o $(f(x') - f(x))$ předem stanovenou hodnotu. Tuto hodnotu označujeme jako práh T (threshold). Pokud budeme tento práh postupně snižovat parametrem $r \in (0,1)$, pak zabráníme zacyklení a povolíme přechod do lokálního minima. Pokud však bude $r = 0$, jedná se o lokální hledání.

Metoda simulovaného žhání oceli (simulated annealing method, SA, SIAM) rozšiřuje metodu lokálního hledání o stochastickou složku. Je založena na fyzikálních principech. Laicky řečeno „kováme, dokud je železo žhavé“. Zchladlá část oceli se již netvaruje a postupně celá vychladne. Pro metodu si potřebujeme stanovit hodnotu teploty $T > 0$, při které stále kováme. Dalším zmíněným prvkem je ochlazování dané $r \in (0,1)$. Určíme také dobu R , po kterou budeme ocel žíhat. Rozdíly v hodnotách účelových funkcí inspirovaného a inspirujícího řešení pak určuje, zda bude hned nově nalezené řešení použito nebo pokud bude horší, zda bude přesun do tohoto řešení možný. To určujeme pravděpodobností přechodu následovně:

$$P = \exp\left(\frac{-(f(x') - f(x))}{T}\right).$$

Genetický algoritmus (Genetic algorithm, GA) je postaven na základních principech evoluce – přirozeného výběru. Přežijí silní a slabí se vytratí. Genetická informace jedince je uchovávána v chromozomu. Jednotlivci tvoří populaci, která obsahuje celkem N jedinců. Ti jsou každou generací lepší a nestane se, že by slabší jedinec nahradil silnějšího jedince z populace. Platí však, že silnější jedinec nahradí slabšího v populaci. Nejlepší jedinec v populaci se označuje jako fitness. To, jak je který jedinec silný, se vypočítá pomocí účelové funkce, kterou pro případ TSP minimalizujeme. Populace prochází postupně generacemi (počet generací dán parametrem G). Noví potomci vznikají několika možnostmi:

- křížení matky a otce (případně i více jedinců),
- mutace (inverze),
- reprodukce.

Pro výběr otce a matky bylo v práci použito turnajové pravidlo. Otec s matkou jsou pak křížení pomocí metody částečně zmapovaného křížení (Partially mapped crossover, PMX). Jakmile vznikne nové dítě, může být ještě zmutováno a k tomu bylo v práci použito prohození

(Swap), náhodná (Scramble) mutace a inverze části genetického kódu. Nakonec jsou z populace vyřazeni nejslabší jedinci, aby byl zachován počet jedinců v populaci. Jako zdroj použitého algoritmu uvádíme (Potvin, 1996).

Optimalizace mravenčí kolonií (Ant colony optimization, ACO) vychází z myšlenky spolupráce v rámci roje. Mravenci (artificial ants = umělí mravenci, agenti) si při cestě za potravou předávají informace pomocí feromonu. Ten mravenci řekne, kudy lze dojít k potravě a které cesty již využili ostatní mravenci. Mravenec se pak sám rozhodne, jestli použije cestu, kterou už použil jiný mravenec, nebo třeba použije cesty od dvou předchůdců a zkombinuje je. Také se může rozhodnout najít úplně novou cestu. Vše se rozhoduje podle síly feromonu (čím více mravenců půjde jednou cestou, tím bude feromon silnější) a nastavením parametrů. Feromon však časem vyprchává, což dává příležitost mravenci nalézt nové neprozkoumané cesty. Díky tomu můžeme překonat lokální minima a nalézt lepší řešení, ne však nutně optimální řešení. Algoritmus se snaží simulovat chování mravenců a lze jej přizpůsobit pro větší problémy, protože ACO má 2 zásadní problémy. Za prvé má vysoké nároky na paměť ukládání hodnot feromonů. A za druhé přepočítání, který uzel má být navštíven jako další, je výpočetně náročný. Jako zdroj použitého algoritmu uvádíme (Chitty, 2017).

Optimalizace včelí kolonií (Bee colony optimization, BCO) vychází opět z chování zvířat a opět půjde o způsob sběru potravy. Včela když objeví nový zdroj potravy, přiletí zpět k úlu a tzv. včelím tanečkem (waggle dance) v podobě osmičky sděluje úlu, kterým směrem a jak daleko se nová potrava nachází. Díky tomu se k potravě dostanou další včely. Potravu pro úlohu TSP představuje hodnota účelové funkce v podobě celkové uražené vzdálenosti. Čím je trasa kratší, tím dokáže včela přinést více nektaru. Včelí taneček může včela provést pouze tehdy, pokud její trasa byla lepší než v posledním řešení. Každá včela si tedy musí pamatovat hodnotu své nejlepší účelové funkce během výpočtu. Včela, která tančí, doporučuje svoji cestu ostatním včelkám a je nahrazena jinou včelou z úlu, aby hledala novou cestu místo ní. Při hledání cesty si včela vybere včelu Θ , podle které se inspiruje v prohledávání své nové trasy. Jako zdroj použitého algoritmu uvádíme (Wong et al., 2008).

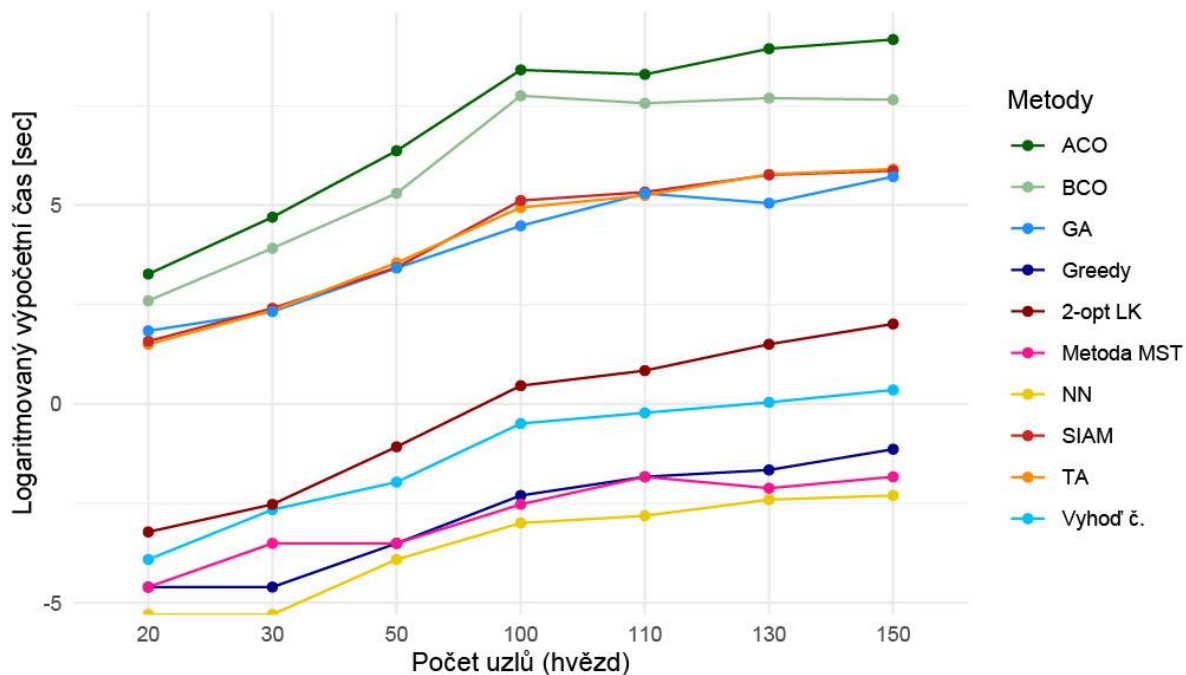
3 EXPERIMENTÁLNÍ VÝPOČTY A ZÁVĚR

Pořadí efektivity (výpočetní čas, hodnota účelové funkce) heuristik mezi sebou a metaheuristik mezi sebou uvádíme v **Tabulka 1**. Pro metaheuristiky bylo provedeno několik výpočtů a z nich vybrány ty nejlepší výsledky. To je z důvodu náhody, která se objevuje napříč jejich aplikací.

Tabulka 1: Pořadí heuristik a metaheuristik podle času a hodnoty účelové funkce.

Heuristiky			Metaheuristiky		
Metoda	Čas	Z	Metoda	Čas	Z
Nejbližší soused	2.	3.	TA	2.	2.
Hladový algoritmus	3.	4.	SIAM	3.	1.
Výhodnostní čísla	4.	2.	GA	1.	3.
Metoda MST	1.	5.	ACO	5.	4.
2-opt LK	5.	1.	BCO	4.	5.

Překvapivě dobrých výsledků dosáhla heuristika nejbližšího souseda. Nejlepších výsledků však dosáhla 2-opt LK heuristika s nejvyšší časovou náročností. Dosáhla tak dobrých řešení, že nakonec předčila i genetický algoritmus, který však pro menší úlohy dával lepší řešení.



Obrázek 1: Výpočetní čas v závislosti na počtu uzlů pro každou metodu.

Velmi rychlá a však špatných hodnot dosahující metoda minimální kostry by mohla být vylepšena použitím v teoretické části zmíněné Christofidovy metody, která by ovšem rapidně navýšila výpočetní čas. Rychlost algoritmu je však podmíněna použitím balíčku napsaného v jazyce C++ pro vytvoření minimální kostry. Nicméně je součástí Rka a tudíž naprosto legitimní v užití.

Metoda výhodnostních čísel dopadla vždy lépe, co se týče hodnoty účelové funkce, než hladový algoritmus. Obě tyto metody jsou založeny na podobném principu. Nicméně metoda výhodnostních čísel je výpočetně náročnější a výpočet jí trvá déle.

Metaheuristiky ACO a BCO nedosahovaly dobrých výsledků. Pro metodu ACO by šlo rychlost řešení vylepšit vyšší koncentrací feromonu na hranách použitých některou z heuristik. Pro metodu BCO jsme již aplikovali zlepšování pomocí jednoho kroku 2-opt heuristiky a inicializací tancujících včel, jejichž cesty byly dány heuristikou nejbližšího souseda. Mohli bychom řešení vylepšit aplikací 3-opt LK heuristiky.

Tabulka 2: Pořadí implementovaných metod podle času a hodnoty účelové funkce.

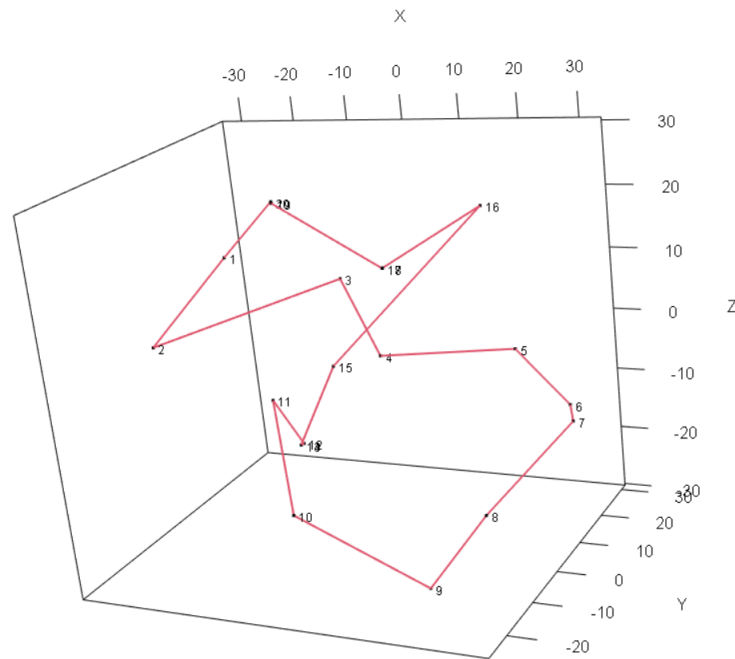
Metoda	Čas	Z
Nejbližší soused	2.	7.
Hladový algoritmus	3.	9.
Výhodnostní čísla	4.	6.
Metoda MST	1.	10.
2-opt LK	5.	3.
TA	7.	2.
SIAM	8.	1.
GA	6.	4.
ACO	10.	5.
BCO	9.	8.

Nejlepší mezi metaheuristikami byly TA a SIAM, které dosahovaly velmi podobných výsledků. Metoda SIAM dosahuje nejlepších řešení. Pro tuto metodu bychom mohli dále

implementovat 3-opt LK heuristiku a pro její zrychlení inicializovat řešení libovolnou rychlou heuristikou. Obdobné platí i pro TA metaheuristiku. Pro GA bychom mohli vytvořit další přístupy, podle kterých volit okolní řešení pro větší škálu řešení.

Závěrem uvádíme tabulku **Tabulka 2**, kde vypisujeme pořadí všech aplikovaných metod podle výpočetního času a podle hodnot účelových funkcí. Výsledky jednotlivých přístupů viz **Obrázek 1**. Příklad zobrazení výsledné cesty metodou GA viz **Obrázek 2**.

Prostředí jazyka R není příliš vhodné k výpočtům metaheuristik z důvodu pomalejšího procesu výpočtu smyček. V řádu nad 100 hvězd, pak byly výpočty zdouhavé. U heuristik pak tento problém nastává až u větších instancí problému v řádech několika tisíců a pouze u 2-opt.



Obrázek 2: Ukázka výsledné cesty ve 3D prostoru.

Použitá literatura

1. PELIKÁN, Jan. Diskrétní modely v operačním výzkumu. Praha: Professional Publishing, 2001.
2. REINELT, Gerhard. The traveling salesman: computational solutions for TSP applications. Springer, 2003.
3. POTVIN, Jean-Yves. Genetic algorithms for the traveling salesman problem. *Annals of Operations Research*, 1996, 63.3: 337-370.
4. CHITTY, Darren M. Applying ACO to large scale TSP instances. In: UK Workshop on Computational Intelligence. Springer, Cham, 2017. p. 104-118.
5. WONG, Li-Pei; LOW, Malcolm Yoke Hean; CHONG, Chin Soon. A bee colony optimization algorithm for traveling salesman problem. In: 2008 Second Asia International Conference on Modelling & Simulation (AMS). IEEE, 2008. p. 818-823.

Kontaktní údaje

Ing. Vojtěch Vávra

Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky
nám. W. Churchilla 1938/4, 130 67 Praha 3 – Žižkov

Tel: (+420) 731 986 046

email: yavv02@vse.cz

VYUŽITÍ LOGISTICKÉ REGRESE V CRM

UTILIZATION OF LOGISTIC REGRESSION IN CRM

Lukáš Veverka

Abstrakt

Vznik e-commerce způsobil rozsáhlý nárůst dat o zákaznících. Důvodem je jednoduché a zároveň anonymní sbírání dat v podobě cookies, díky kterým je možné zjistit chování zákazníka na vlastním webu, jeho zájmy a základní demografické údaje nebo účinnost on-line kampaní. Z těchto dat je poté možné sestavit základní systém řízení vztahů se zákazníky (CRM). Tato práce se bude zabývat analýzou CRM dat e-shopu s rámy na obrázky a určením cílové skupiny zákazníků vhodné pro remarketing (tzn. u kterých je pravděpodobné, že po zásahu reklamní kampaní v budoucnu znovu nakoupí).

Klíčové slova: CRM, Logistická regrese, Marketing, Ekonometrie

Abstract

The emergence of e-commerce has caused a large increase in customer data. The reason is the simple and at the same time anonymous collection of data in the form of cookies, thanks to which it is possible to find out the customer's behavior on website, interests and basic demographic data or the effectiveness of online campaigns. From this data, it is then possible to build a basic customer relationship management (CRM) model. This work will deal with the analysis of CRM data of an e-shop with picture frames and determine the target group of customers suitable for remarketing (ie, which are likely to buy again in the future after the intervention of advertising campaigns).

Keywords: CRM, Logistic regression, Marketing, Econometrics

1 TEORETICKÝ ZÁKLAD

Již bylo dokázáno, že využívání CRM systémů má značný pozitivní vliv na efektivitu marketingu zaměřeného na jednotlivce (1). Z tohoto důvodu bude určeno, jaké faktory mají vliv na opakovaný nákup přes e-shop u daného výrobce ráků. Základní hypotézy tedy budou:

- H_1 : *To, zda zákazník nakoupí znovu, závisí na uděleném hodnocení v dotazníku spokojenosti.*
- H_2 : *Pohlaví má vliv na opakovaný nákup.*
- H_3 : *Starší lidé nakoupí znovu s větší pravděpodobností.*
- H_4 : *Příští nákup je ovlivněn výší útraty.*
- H_5 : *Opakovaný nákup závisí na počtu navštívených stránek po první koupi.*

2 DATA A METODOLOGIE

Data, která budou v této práci analyzována pochází ze CRM databáze e-shopu prodávajícího rámy a zboží s tím související. Jedná se o záznam dat z jednoho měsíce (listopad 2016) týkajících se útraty a spokojenosti určitého zákazníka, jeho základních demografických údajů a

chování na webu po nákupu. Podstatnou částí těchto dat je údaj, který definuje zda-li zákazník nakoupil znovu v tomto e-shopu maximálně do jednoho roku od provedení nákupu. Jedná se tedy o data průřezová.

V CRM databázi z uvedeného měsíce jsou záznamy o 138 zákaznících, přičemž všechna pozorování byla kompletní a neobsahovala chybějící hodnoty. Soubor nebyl ani redukován, protože neobsahoval žádná chybějící pozorování a odlehlá pozorování se vyskytla jen částečně u proměnné *Útrata*, kde to ovšem odpovídá realitě. V datech se vyskytují dohromady 3 binární proměnné, z nichž jedna bude vysvětlována.

- *Koupil znovu* – Pokud zákazník uskutečnil další nákup do jednoho roku, tak se rovná 1.
- *Spokojenost* – Vyjadřuje odpověď ve zpětné vazbě. Pokud byl zákazník celkově spokojený (tzn. odpověď ANO), tak proměnná nabývá hodnoty 1.
- *Muž* – Pokud byl zákazník muž, tak 1.

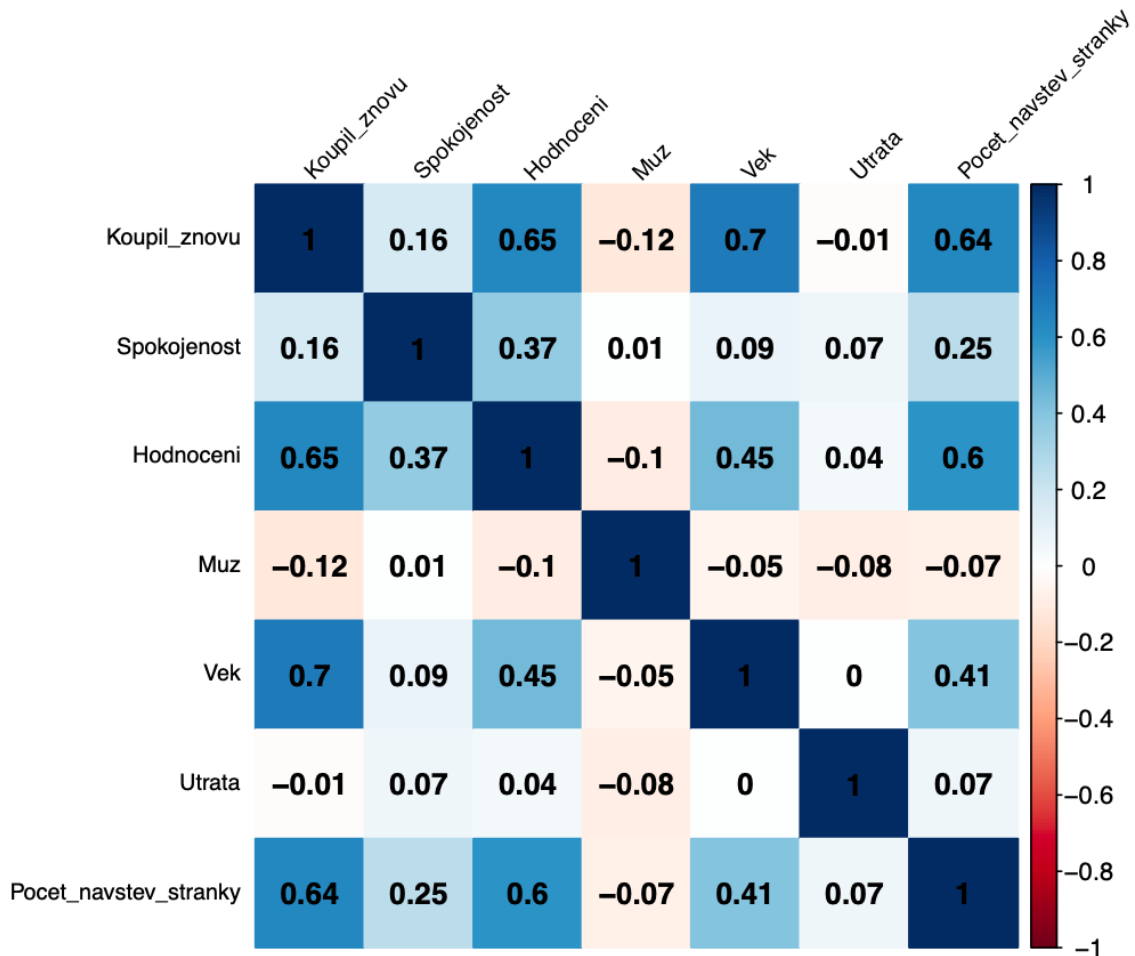
Dále jsou v datech kvantitativní proměnné – *Hodnocení*, *Věk*, *Útrata* (v Kč) a *Počet návštěv stránky*. Základní charakteristiky jednotlivých proměnných je možné nalézt v Tabulce 1. U proměnné *Hodnocení* je vhodné podotknout, že hodnoty jsou v rozsahu 0–100. *Počet návštěv stránky* je počet navštívených stránek po prvním nákupu (např. z důvodu recenze, přístupu k informacím o objednávkách nebo zájmu o další produkt).

Tabulka 1: Základní popis a charakteristika vstupních dat

Proměnná	Průměr	St. Dev.	Min	Max
Koupil znovu	0,341	0,476	0	1
Spokojenost	0,681	0,468	0	1
Hodnocení	69,500	18,527	28	99
Muž	0,551	0,499	0	1
Věk	39,754	12,812	18	73
Útrata	282,355	163,411	102	973
Počet návštěv stránky	3,181	1,801	1	8

Závislá proměnná tedy bude *Koupil znovu*. Vysvětlujícími proměnnými budou všechny ostatní proměnné uvedené v Tabulce 1. Vliv uvedených regresorů na opakovaný nákup bude zkoumán pomocí modelů binární volby (konkrétně logitu a probitu) (2).

Při analýze dat bylo využito několika balíčků v Rku. Pro rychlejší načtení .xlsx souborů je vhodné využít balíček *openxlsx* (3). Pro export do LATEXu byl použit balíček *stargazer* (4). Pro vytvoření obrázku 2 byl použit balíček *ggplot2*, který tvoří graficky atraktivnější výstupy (5). Funkce, která vytvoří přehlednou vizualizaci korelační matice (viz obrázek 1), je součástí balíčku *corrplot* (6). Pro výpočet průměrných průměrných mezních efektů posloužil balíček *mfx* (7).



Obrázek 1: Vizualizace korelační matice

zdroj: autor

3 VÝSLEDKY

Z vizualizace korelační matice na Obrázku 1, je možné si všimnout výraznější korelace mezi proměnnými *Koupil znovu*, *Hodnocení* a *Počet návštěv stránky*. Největší korelace (vyjma hlavní diagonály) se nachází mezi proměnnými *Koupil znovu* a *Věk*.

Uvažujme model logistické regrese ve tvaru:

$$\text{logit}(p) = \beta_0 + \beta_1 \text{Spokojenost} + \beta_2 \text{Hodnoceni} + \beta_3 \text{Muz} + \beta_4 \text{Vek} + \beta_5 \text{Utrata} + \beta_6 \text{Stranky} + u, \quad (1)$$

kde p vyjadřuje pravděpodobnost úspěchu (tedy opakovaného nákupu) a u je náhodná složka. Pro účely této práce bude část $x\beta$ označována zkráceně jako „skóre“.

Odhadneme tedy model pomocí funkce **glm** a získáme odhadnuté regresní koeficienty, které je možné nalézt v Tabulce 2. Pro zajímavost a porovnání byl vytvořen podobným způsobem i model probit. Výsledky jsou velmi podobné (hlavně ukazatel Log Likelihood a Akaikeho informační kritérium), protože průměrná pravděpodobnost úspěchu (opakovaného nákupu) v CRM datech je $p = 0,34$. Logit a probit se totiž výrazněji liší až u hodnot $p = 0,05$. Odhadnutý model poskytuje oporu pro H_1 , že opakovaný nákup pozitivně závisí na uděleném hodnocení ($p = 0,0023$). Pokud je tedy zákazník spokojený a udělí vysoké hodnocení, je

očekávatelné, že v případě potřeby dalšího produktu nabízeného v tomto e-shopu, se na něj znovu obrátí. Dále se díky modelu potvrzuje hypotéza H_3 , která tvrdí, že starší lidé nakoupí znovu s větší pravděpodobností než mladší ($p = 0,00014$). To může souviset buď s větší konzervativností starších lidí nebo s faktem, že e-shop má starší design, ve kterém se mladší lidé mohou hůře orientovat (aneb je méně *user-friendly*) a v dnešní uspěchané době to může vést k opuštění stránky bez dalšího nákupu. Poslední hypotézou, pro kterou odhadnutý model poskytuje oporu, je H_5 ($p = 0,0031$) – s větším počtem navštívených stránek po prvním nákupu roste pravděpodobnost, že zákazník nakoupí znovu. Pokud si například prohlédne 5 stránek, tak je velmi pravděpodobné, že na web nešel jen z důvodu recenzí, ale spíše se díval po dalších produktech. Model naopak vyvrátil hypotézy H_2 ($p = 0,413$) a H_4 ($p = 0,081$). Na pohlaví zákazníka tedy zjevně nezáleží a výše útraty také nemá signifikantní vliv na opakovaný nákup.

Interpretovat výsledky nelze hned a je nutné ještě odhadnuté regresní parametry exponovat. Jejich hodnoty i s intervaly spolehlivosti je možné nalézt v Tabulce 3. Například hodnota exponovaného odhadnutého regresního parametru $\widehat{\beta}_2 = 1,197$ znamená, že s každým dalším uděleným bodem ve zpětné vazbě se poměr šancí (tzv. odds ratio) přenásobí hodnotou 1,197. To znamená, že pokud by poměr šancí byl $47/91 = 0,52$ (poměr šancí v celé CRM databázi), tak s jedním bodem hodnocení navíc by vzrostl na $(47 \times 1,197)/91 = 0,62$. Pravděpodobnost úspěchu by díky tomu tedy vzrostla z $47/(47+91) \times 100 = 34,1\%$ na hodnotu $(47 \times 1,197)/(47 \times 1,197 + 91) \times 100 = 38,2\%$ (tj. o 4,1 procentních bodů).

Tabulka 2: Porovnání modelů logit a probit a hodnoty AME pro logit

	Regresní koeficienty (β):		Průměrné mezní efekty (AME):
	<i>logit</i>	<i>probit</i>	<i>logit</i>
Spokojenost	-1,178 (1,163)	-0,592 (0,613)	-0,048 (0,045)
Hodnocení	0,180*** (0,059)	0,101*** (0,031)	0,007** (0,004)
Muž	-0,876 (1,070)	-0,419 (0,579)	-0,036 (0,043)
Věk	0,233*** (0,061)	0,133*** (0,033)	0,010** (0,005)
Útrata	-0,012* (0,007)	-0,007* (0,004)	-0,0005 (0,0003)
Počet návštěv stránky	1,269*** (0,429)	0,697*** (0,223)	0,052* (0,027)
Konstanta	-24,294*** (6,191)	-13,707*** (3,187)	—
Počet pozorování	138	138	138
Log Likelihood	-18,405	-18,232	—
Akaike Inf. Crit.	50,810	50,463	—

Poznámky:

(i) U AME se liší statistická významnost, protože se jedná o jiný ukazatel

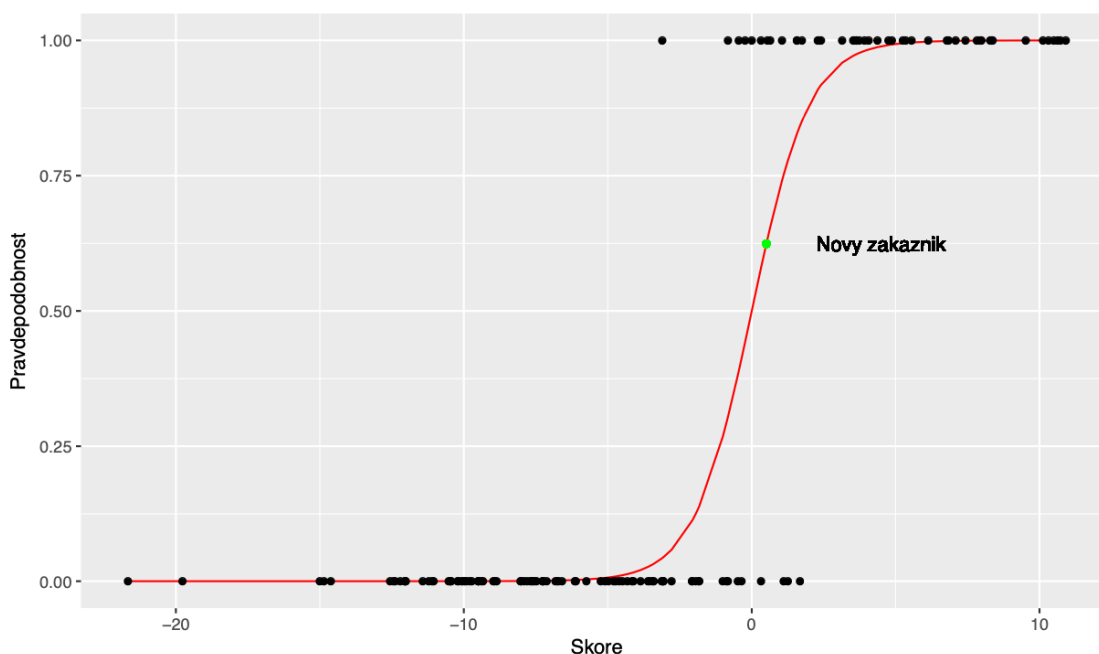
(ii) * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Změny pravděpodobností úspěchu lze také interpretovat pomocí průměrných mezních efektů (AME), kde AME je výběrový průměr mezních efektů pro jednotlivá pozorování. Hodnoty AME jsou také zaznamenány v Tabulce 2. Například u proměnné Věk hodnota AME v logitu říká, že s jedním rokem zákazníka navíc se pravděpodobnost úspěchu (tedy opakovaného nákupu) zvýší o zhruba $0,01 \times 100 = 1\%$.

Pokud si zobrazíme původní hodnoty proměnné *Koupil znovu* v závislosti na „skóre“ a vložíme do stejného grafu i logistickou funkci, tak dostaneme základní přehled, jak výsledný model dokáže určit pravděpodobnost dalšího nákupu. Na Obrázku 2 je vidět, že několik zákazníků i přes negativní skóre nakonec uskutečnilo další nákup. Stejně tak několik zákazníků i přes kladné skóre a teoreticky vysokou pravděpodobnost dalšího nákupu znovu nenakoupili.

Tabulka 3: Exponované parametry i s intervaly spolehlivosti

	Exp. parametry	2,5 %	97,5 %
Spokojenost	0,308	0,032	3,011
Hodnocení	1,197	1,066	1,343
Muž	0,417	0,051	3,394
Věk	1,263	1,119	1,424
Útrata	0,988	0,975	1,001
Počet návštěv stránky	3,559	1,534	8,258



Obrázek 2: Logistická závislost opakovaného nákupu na skóre

zdroj: autor

3.1 Predikce u nového zákazníka

Předpokládejme, že následující měsíc nakoupí v hodnotě 350,- Kč na e-shopu $x\hat{\beta}$, kde x je 55 let a ve zpětné vazbě uvede, že byla s nákupem spokojená a udělí 68 bodů. Následně si prohlédne 4 stránky. Jaká bude pravděpodobnost, že nakoupí znovu? Odpověď je poměrně jednoduchá – stačí dosadit uvedené údaje do modelu a získat tak její „skóre“ (tj. \hat{y} , kde x je vektor údajů o této zákaznici). Výsledné skóre je 0,507 a pokud jej vložíme do logistické funkce, dostaneme pravděpodobnost dalšího nákupu $1/(1+e^{-0,507}) \times 100 = 62,4\%$. Tento příklad je také zobrazen na Obrázku 2.

4 ZÁVĚR

Pomocí logistické regrese byly určeny faktory, ovlivňující opakovaný nákup. Největší vliv má počet navštívených stránek po provedení nákupu (např. prohlížení dalších produktů). Pokud si zákazník prohlédne o 1 stránku navíc, tak v průměru vzroste pravděpodobnost jeho dalšího nákupu o 5,2 %. Dalšími faktory je věk (starší lidé zopakují nákup s větší pravděpodobností než mladší) a udělené hodnocení v dotazníku spokojenosti (čím spokojenější zákazník, tím je větší pravděpodobnost, že opět nakoupí). Naopak pohlaví, vyjádření celkové spokojenosti pomocí ANO/NE a výše útraty nemá signifikantní vliv. Cílovou skupinou vhodnou pro remarketing je tedy skupina starších lidí, kteří udělili nadprůměrné hodnocení.

Zde se naskytá prostor pro využití webových analytických nástrojů a nastavení menší ceny pro remarketing zákazníků s vysokým skóre, protože pravděpodobnost dalšího nákupu je již vysoká. Naopak vyšší rozpočet by mohl být použit na remarketing zákazníků, kteří by mohli váhat a přejít ke konkurenci a zajistit tak jejich loajalitu. Vhodným remarketingovým nástrojem by v takovém případě mohl být slevový kupón a upozornění na probíhající akce.

Pro zpřesnění výsledků tohoto modelu by bylo možné použít strojové učení (tzv. *machine learning*), které de facto s každým dalším pozorováním opakuje výše uvedený model stále dokola a zpřesňuje tak odhady parametrů β . Tím se dá získat silný predikční nástroj k odhadu pravděpodobnosti dalšího nákupu a přesné identifikaci zákazníků, na které je nutné použít remarketingové nástroje.

Poděkování

Práce na tomto příspěvku byla podpořena Interní grantovou agenturou Vysoké školy ekonomické v Praze v rámci projektu F4/38/2022.

Použitá literatura

1. Mithas, S., Almirall, D., and Krishnan, M. S. (2006). Do CRM systems cause one-to-one marketing effectiveness? *Statistical Science*, 21(2):223–233.
2. Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press. ISBN: 978-0262232197.
3. Walker, A. (2017). *openxlsx: Read, Write and Edit XLSX Files*. R package version 4.0.17.
4. Hlavac, M. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Central European Labour Studies Institute (CELSI), Bratislava, Slovakia. R package version 5.2.1.
5. Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
6. Wei, T. and Simko, V. (2017). R package "corrplot": Visualization of a Correlation Matrix. (Version 0.84).
7. Fernihough, A. (2014). *mfx: Marginal Effects, Odds Ratios and Incidence Rate Ratios for GLMs*. R package version 1.1.

Kontaktní údaje

Ing. Lukáš Veverka

Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky
nám. W. Churchilla 1938/4, 130 67 Praha 3 - Žižkov

Tel: (+420) 723 298 590

email: lukas.veverka@vse.cz