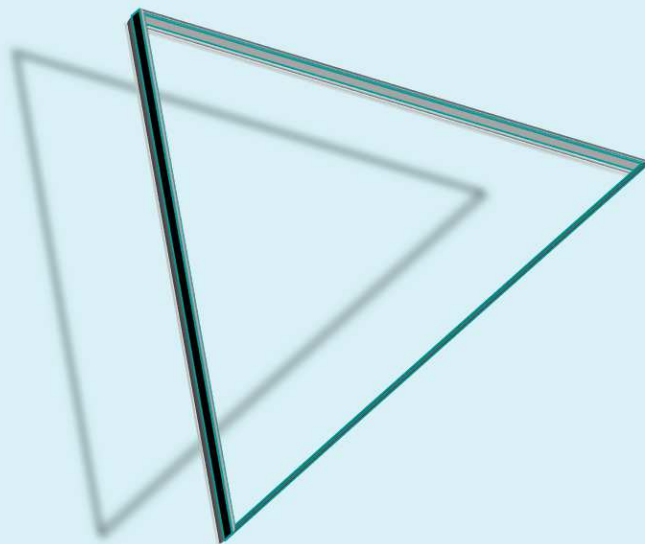
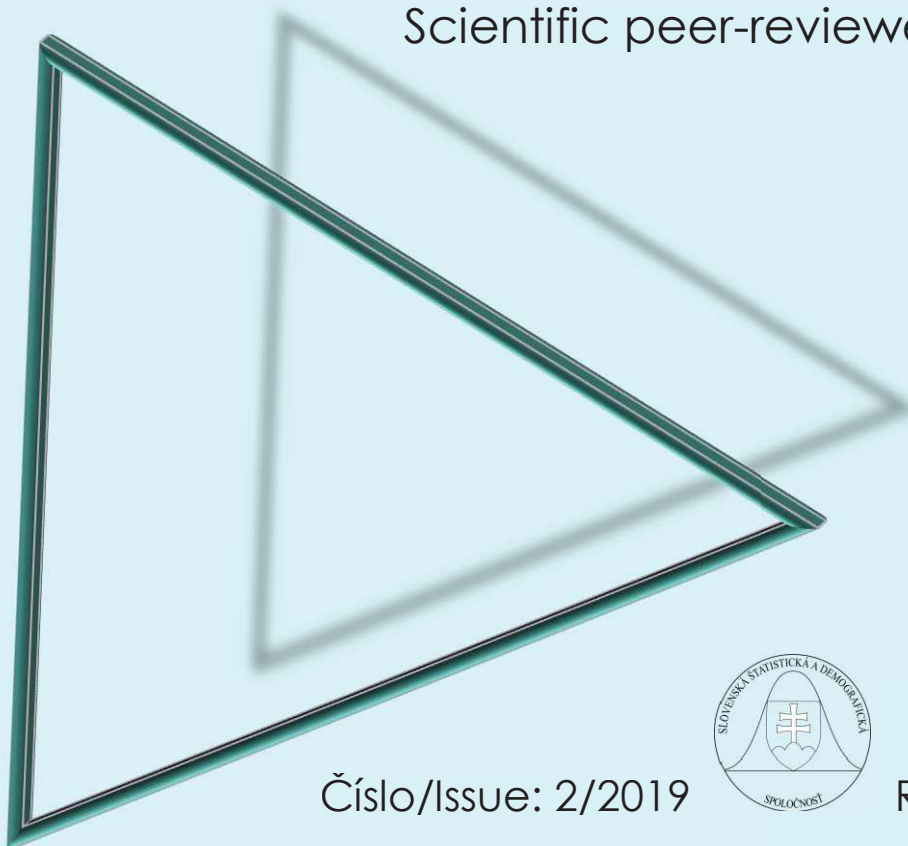


Slovenská štatistická a demografická spoločnosť
Slovak Statistical and Demographic Society

FORUM STATISTICUM SLOVACUM



Recenzovaný vedecký časopis
Scientific peer-reviewed journal



Číslo/Issue: 2/2019



Ročník/Volume: XV

FORUM STATISTICUM SLOVACUM

recenzovaný vedecký časopis Slovenskej štatistickej a demografickej spoločnosti
scientific peer-reviewed journal of the Slovak Statistical and Demographic Society

Ročník/Volume: XV (2019)

Číslo/Issue: 2

Editori / Editors

Iveta Stankovičová (Fakulta managementu Univerzity Komenského v Bratislave, Slovensko)

Martin Boďa (Ekonomická fakulta Univerzity Mateja Bela v Banskej Bystrici, Slovensko)

Redakčná rada / Editorial Board

Branislav Bleha (Prírodovedecká fakulta Univerzity Komenského v Bratislave, Slovensko)

Boris Burcin (Prírodovedecká fakulta Univerzity Karlovej v Prahe, Česko)

Joanna Dębicka (Fakulta manažmentu, informatiky a financií Ekonomickej univerzity vo Wroclawi, Poľsko)

Estefanía Mourelle Espasandín (Fakulta ekonómie a podnikania Univerzity v La Coruña, Španielsko)

Stanislav Katina (Ústav matematiky a štatistiky Masarykovej univerzity v Brne, Česko)

Jana Kubanová (Fakulta ekonomicko-správna Univerzity Pardubice, Česko)

Dagmar Kusendová (Prírodovedecká fakulta Univerzity Komenského v Bratislave, Slovensko)

Viera Labudová (Fakulta hospodárskej informatiky Ekonomickej univerzity v Bratislave, Slovensko)

Jitka Langhamrová (Fakulta informatiky a štatistiky Vysokej školy ekonomickej v Prahe, Česko)

Ivan Lichner (Ekonomický ústav Slovenskej akadémie vied, Slovensko)

Tomáš Löster (Fakulta informatiky a štatistiky Vysokej školy ekonomickej v Prahe, Česko)

Janka Medová (Fakulta prírodných vied Univerzity Konštantína Filozofa v Nitre, Slovensko)

Silvia Megyesiová (Podnikovohospodárska fakulta v Košiciach Ekonomickej univerzity v Bratislave, Slovensko)

Oľga Nanásiová (Fakulta elektrotechniky a informatiky Slovenskej technickej univerzity v Bratislave, Slovensko)

Viliam Páleník (Ekonomický ústav Slovenskej akadémie vied, Slovensko)

Marek Radvanský (Ekonomický ústav Slovenskej akadémie vied, Slovensko)

Hana Řezanková (Fakulta informatiky a štatistiky Vysokej školy ekonomickej v Prahe, Česko)

Ľubica Sipková (Fakulta hospodárskej informatiky Ekonomickej univerzity v Bratislave, Slovensko)

Mária Stachová (Ekonomická fakulta Univerzity Mateja Bela v Banskej Bystrici, Slovensko)

Anna Tirpáková (Fakulta prírodných vied Univerzity Konštantína Filozofa v Nitre, Slovensko)

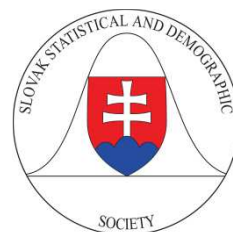
Vladimír Úradníček (Ekonomická fakulta Univerzity Mateja Bela v Banskej Bystrici, Slovensko)

Mária Vojtková (Fakulta hospodárskej informatiky Ekonomickej univerzity v Bratislave, Slovensko)

Tomáš Želinský (Ekonomická fakulta Technickej univerzity v Košiciach, Slovensko)



Vydavateľ: Slovenská štatistická a demografická spoločnosť, Miletičova 3, 824 67 Bratislava, Slovensko. **Publisher:** Slovak Statistical and Demographic Society, Miletičova 3, 824 67 Bratislava, Slovakia. **Adresa redakcie/Editorial office:** Miletičova 3, 824 67 Bratislava, Slovakia. **IČO/Company ID:** 00178764. **DIČ/Tax ID:** 2021504276. **Mailový kontakt/E-mail contact:** adm.ssds@ssds.sk **Web site/Webové sídlo:** <http://www.ssds.sk/>



Registráciu vykonalo Ministerstvo kultúry Slovenskej republiky. **Dátum registrácie:** 27. júla 2005. **Evidenčné číslo:** EV 3287/09. **Tematická skupina:** B1. **Periodicita:** minimálne dvakrát ročne. **ISSN:** 1336-7420.

Registered by the Ministry of Culture of the Slovak Republic. **Date of registration:** 27 July 2005. **Registration No:** EV 3287/09. **Topic group:** B1. **Periodicity:** at least two issues per year. **ISSN:** 1336-7420.

Tvorba atraktivních statistických grafů v prostředí SAS

Creating attractive statistical graphs in the SAS environment

Roman Pavelka

Štatistický úrad Slovenskej republiky, Odbor metód štatistických zisťovaní, Miletičova 3, 824 67 Bratislava, Slovenská republika

Statistical Office of the Slovak Republic, Statistical Surveys and Methodology Department, Miletičova 3, 824 67 Bratislava, Slovak Republic

roman.pavelka@statistics.sk

Abstrakt: Příspěvek popisuje nové možnosti statistické grafiky z programového systému SAS při vytváření statistických grafů na vysoké grafické úrovni. Mohutným nástrojem pro tvorbu všech statistických grafů jsou procedury (programové moduly) statistické grafiky, z nichž základem pro tvorbu statistické grafiky je především procedura SGPLOT. Procedura statistické grafiky SGPLOT (ve verzi systému SAS 9.4) umožňuje uživatelům systému SAS vytvářet všechny typy statistických grafů - od jednoduchých sloupcových grafů až po komplexní business grafy s prostorovými efekty, které dokáží zprostředkovat data jednoduchým a přímým způsobem s minimálním rozptýlením uživatele. Statistická grafika SAS při vhodném využití zabudovaných nástrojů vizualizace dat umí upoutat pozornost čtenáře uprostřed moře informací. Procedury statistické grafiky SAS umožňují vkládat do výstupních grafů takové vizuální prvky, které zvýrazní nejdůležitější informace a zatraktivní pořizovaný graf. Takové vizuální prvky vkládají do grafu ozdobnou příchutí, aby přilákaly oko čtenáře a povzbudily je ve sledování statistické grafiky. Příspěvek proto pojednává o nových možnostech tvorby statistické grafiky pomocí procedury SGPLOT.

Abstract: This paper describes new possibilities of statistical graphics from the SAS program system for creating high-level statistical graphs. Powerful tools for creating all statistical graphs are procedures (program modules) of statistical graphics. The statistical graphics' SGPLOT procedure (in SAS 9.4) allows SAS users to create all types of statistical graphs - from simple bar graphs to complex business graphs with spatial effects, which can convey data in a simple and straightforward manner with minimal user distraction. Using SAS's built-in data visualization tools, SAS statistical graphics can capture the reader's attention in the middle of a sea of information. SAS statistical graphics procedures allow adding visual elements to the output charts that highlight the most important information and make the graph more attractive. Such visual elements add a decorative flavour to the chart to attract the eye of the reader and encourage them to observe statistical graphics. The paper therefore discusses new possibilities of creating statistical graphics using the SGPLOT procedure.

Klíčové slová: procedura SGPLOT, SAS, statistická grafika, vizualizace.

Key words: SGPLOT procedure, SAS, statistical graphics, visualization.

1 Úvod

Nejdůležitějším nástrojem statistické grafiky v systému SAS se stala procedura SGPLOT, která je dokumentována skrze SAS Institute (2019). Tento programový modul dokáže vytvářet grafy, které názorně sdělují předávané informace

čtenářům jasným a efektivním způsobem. Pomocí uvedené procedury je možné vytvářet grafy od jednoduchého sloupcového grafu po komplexní business grafy. Většinou se jedná o takové statistické grafy, které vyhýbají zbytečným třpytům nebo ozdobám.

Příspěvek chce ukázat, jak vytvořit dekorativní statistické grafiky, přiblížit možnosti vizualizace dat efektním a neobvyklým způsobem. Všechny ilustrované statistické grafy budou vytvářeny pomocí syntaxe procedury SGPLOT a uživatelsky vytvořených datových souborů. Příkazy ke všem vytvářeným statistickým grafikám jsou zaznamenány v Příloze. Každý vytvořený graf je popsán nejen v textu, ale také i sekvencí příkazů, jimiž byl vytvořený graf naprogramován. Všechny předkládané statistické grafiky a datové vizualizace, které budou prezentovány v příspěvku, jsou plně funkční a použitelné pro další úpravy. Je však třeba správně nastavit cesty k obrazovým souborům, které jsou potřebné pro dekorativní grafiku. Příspěvek o statistické grafice vychází zejména z technické dokumentace k systému SAS dostupné na webové stránce <https://support.sas.com/en/documentation.html>. Dalšími zdroji informací jsou vybrané příspěvky komunity uživatelů systému SAS.

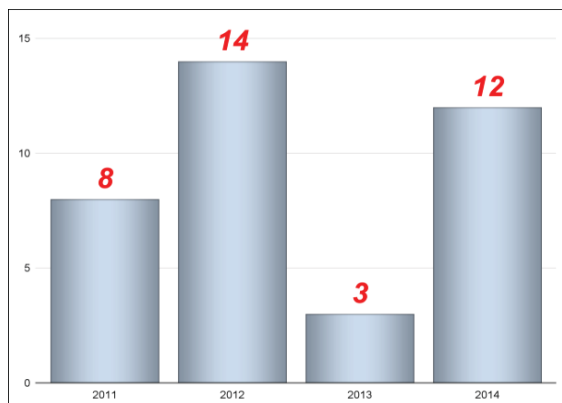
2 Dekorativní statistická grafika programu SAS

2.1 Jak účinně upoutat pozornost čtenářů

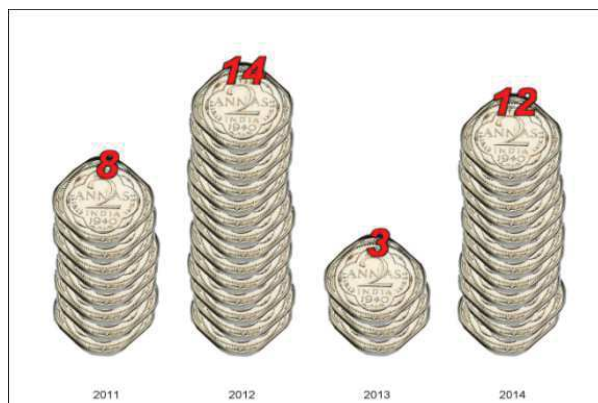
Nejdůležitější informace ze stránek časopisů nebo různých brožur jsou často skryty v balastu méně podstatných poznatků. Navíc různé články na stránce soupeří navzájem o pozornost čtenáře. Jak se dá vyniknout požadované informaci mezi ostatními údaji? A je to vůbec možné? Jedním ze způsobů, jak účinně upoutat pozornost čtenářů, je použití efektních dekorativních statistických grafik.

Dekorativní statistickou grafiku lze nejlépe definovat názorně -pomocí obrázků. Tradiční sloupcový graf ročních příjmů zobrazuje Obrázek 1. Ačkoliv graf využívá speciálního efektu k vytvoření sloupců válcového vzhledu, jedná se o zcela standardní graf. Graf z Obrázku 2 zobrazuje stejná data, ale využívá sloupec reálných mincí k prezentování ročních příjmů. Na stránce přeplněné textem a grafikou může graf na Obrázku 2 vyniknout a upoutat pozornost čtenáře daleko více než standardní graf z Obrázku 1.

V dekorativním grafu na Obrázku 2 je možné, že statistická grafika nezobrazuje zcela přesně velikost sledovaných příjmů, a tedy každý sloupec nemusí být zcela přesně úměrný datům. Na Obrázku 2 jsou tyto detaily sekundární, protože cílem grafu je zachytit pozornost čtenáře.



Obrázek 1 Sloupcový graf ročních příjmů (v milionech) (Zdroj: vlastní zpracování)



Obrázek 2 Sloupcový graf ročních příjmů (v milionech) (Zdroj: vlastní zpracování)

2.2 Klíčová funkce pro vytváření dekorativní statistické grafiky

Klíčovou vlastností procedury SGPLOT, která je použita ke tvorbě dekorativní statistické grafiky v článku, je její schopnost využívat obrázky k definování značek v bodovém grafu pomocí příkazu SYMBOLIMAGE. Příkaz byl poprvé zaveden do programu SAS verze 9.4. Obecný tvar tohoto příkazu v rámci syntaxe procedury SGPLOT je následující:

SymbolImage name=symbolname image="file-name" </ options>;

Podporovanými hodnotami parametrů jsou následující:

- HOFFSET = Horizontální posun k bodu umístění symbolu v souřadnicové soustavě,
- ROTATE = Úhel rotace symbolu ve stupních,
- SCALE = Škálovací faktor symbolu ke změnám velikosti značky a
- VOFFSET = Vertikální posun k bodu umístění symbolu v souřadnicové soustavě.

Příkaz lze použít k definování symbolu se jménem určeným parametrem **name=**, parametrem **image=** jméno obrazového souboru včetně cesty v daném systému souborů. Po definování symbolu jménem a odkazem na soubor obrázku se tento symbol pomocí přiřazeného názvu dá použít k zobrazení značek v bodovém grafu. Takové symboly se mohou využít i k reprezentaci skupin. Jako obrázky se zpravidla používají obrázky formátu JPG, PNG nebo PCX, i s transparentním (průhledným) pozadím. Použití průhledných obrázků jako symbolů pro značky umožňuje použít je jako tvary nebo masky nebo podoby (vzhledy).

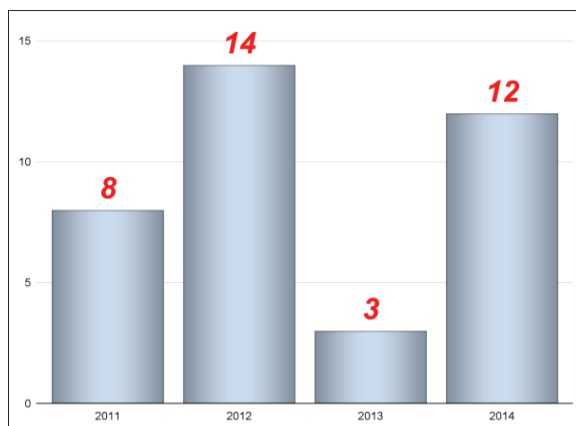
2.3 Použití obrazových souborů pro nastavení tvarů v grafu

Použijí-li se na místo značek transparentní obrázky, ve kterých jsou pixely mimo požadovaný tvar průhledné, potom tyto symboly mohou opticky vytvářet rozličné tvary v grafech. Příklady takových statistických grafů jsou uvedeny dále.

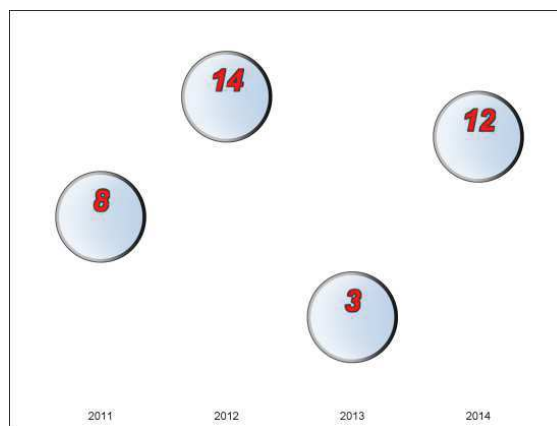
2.3.1 Příklad 1 - Graf hodnot ročních příjmů

Datový soubor ročních výnosů Bar, který bude použit do následujících grafů, je vytvořen pomocí příkazů uvedených v Příloze. Nejprve se vytvoří sloupcový graf znázorněný na Obrázku 3 pomocí příkazu VBAR procedury SGPLOT. Aby bylo možné použít obrázek mince jako značky k vytvoření sloupce mincí, je nutné namísto příkazu VBAR použít v příkazu procedury SGPLOT příkaz SCATTER (umožňuje použití příkazu SYMBOLIMAGE) pro tvorbu bodového grafu (viz Příloha).

Při použití příkazu SCATTER vzniká bodový graf a pro roční příjem se v grafu vytvoří pouze 1 značka, která odpovídá ročnímu příjmu (výšce sloupce) ve sloupcovém grafu na Obrázku 3. Příkazem TEXT se zobrazí číselné hodnoty ročních příjmů u každé značky. Příkazy XAXIS a YAXIS slouží ke zneviditelnění os grafu. Bodový graf je zobrazen na Obrázku 4.

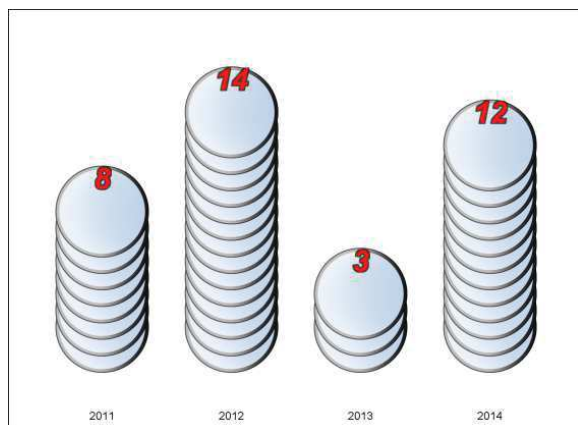


Obrázek 3 Sloupcový graf ročních příjmů (v milionech) (Zdroj: vlastní zpracování)



Obrázek 4 Bodový graf ročních příjmů (v milionech) (Zdroj: vlastní zpracování)

Jelikož se jedná o bodový graf (příkaz SCATTER v proceduře SGPLOT), je pro každý rok nutné vytvořit v grafu sloupec značek odpovídající ročním výnosům. První sloupec příjmů potřebuje osm značek atd. Proto je potřebné původní data upravit tak, aby bylo vytvořeno více pozorování ročně. Pro úpravu původních dat se využije sekvence příkazů. Upravená původní data se uloží do datového souboru Coins. Do vytvořených dat se zpracuje nová proměnná YVAL, jejíž hodnota odpovídá ročním příjmům a která slouží k popisu vzniklého sloupce. Takto upravená data slouží jako vstup procedury SGPLOT při tvorbě bodového grafu (s příkazem SCATTER). Vzniká tak graf, který je znázorněn na Obrázku 5. Následuje krok nahrazení každé značky mincí. Příkazem SYMBOLIMAGE se vytvoří nový symbol pro značku v podobě mince. Tento symbol se pak použije v parametru MARKERATTRS v příkazu SCATTER o velikosti = 110. Ve výsledném grafu je každá značka ve sloupci nahrazena symbolem mince, jak je znázorněno na Obrázku 6.



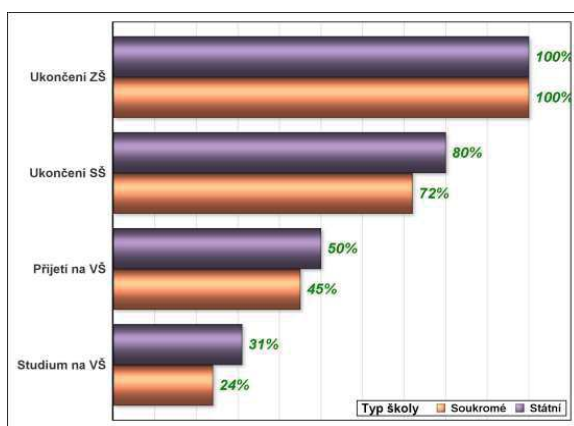
Obrázek 5 Graf sloupců značek příjmů (v milionech) (Zdroj: vlastní zpracování)



Obrázek 6 Graf sloupců symbolů příjmů (v milionech) (Zdroj: vlastní zpracování)

2.3.2 Příklad 2 - Graf porovnání úspěšnosti studentů soukromých a státních škol

Úkolem grafu je porovnání úspěšnosti studentů soukromého a státního školství. Z pohledu poskytnutí nejúčinnějšího způsobu provedení srovnání by byl nejlepší skupinový (klastrový) sloupcový graf znázorněný na Obrázku 7. Z něho je velmi snadné porovnat úspěšnost studentů soukromého školství se studenty škol státních, když jsou výsledky zobrazeny vedle sebe pro každou kategorii.



Obrázek 7 Graf srovnání úspěšnosti (v %) studentů (Zdroj: vlastní zpracování)



Obrázek 8 Graf srovnání úspěšnosti (v %) studentů (Zdroj: vlastní zpracování)

I když se na Obrázku 7 jedná o velmi názorný graf, nepatří do skupiny grafů, které upoutají pozornost čtenáře. Z hlediska atraktivity grafického provedení je možné využít grafu z Obrázku 8, který využívá k vyjádření hodnot symboly s různými barvami. Graf srovnání úspěšnosti (v %) studentů podle Obrázku 8 je vytvořen pomocí příkazů SAS v Příloze.

Hodnoty sledované úspěšnosti (odezvy) jsou zaznamenány v procentech, od 0% do 100%. Proto je možné vytvořit pozorování pro každých 10% hodnoty úspěšnosti (odezvy), a to zvlášť pozorování úspěšnosti (odezvy) pro soukromé

a zvlášť pro státní školství na všech sledovaných úrovních. K nakreslení grafu jsou použity symboly znázorněné na Obrázku 8. Pokud je hodnota pro úspěšnost násobkem 10%, zobrazí se s plnobarevným symboly (v příkazech jako symbol A1 nebo B1). Pokud úspěšnost představuje pouze část 10%, použije se částečně zbarvený symbol (v příkazech jako symbol A2 a B2). Pro hodnoty úspěšnosti (odezvy), která je větší než zjištěná hodnota a menší než 100%, se uplatní šedý symbol C1.

2.4 Použití obrazových souborů pro nastavení tvarů v grafu a jako masky

Pokud jsou vnější pixely neprůhledné (obvykle barvy, které odpovídají pozadí grafu) a vnitřní pixely jsou průhledné, lze obrázek použít jako masku. Obrázek 9 ukazuje obrazový soubor šálku, ve kterém jsou všechny pixely uvnitř průhledné. Když je tento obrázek vložen na červeném pozadí, jsou viditelné pouze červené pixely uprostřed, zatímco ostatní pixely jsou překresleny neprůhlednými pixely obrazu šálku, jak je znázorněno na obrázku 10.

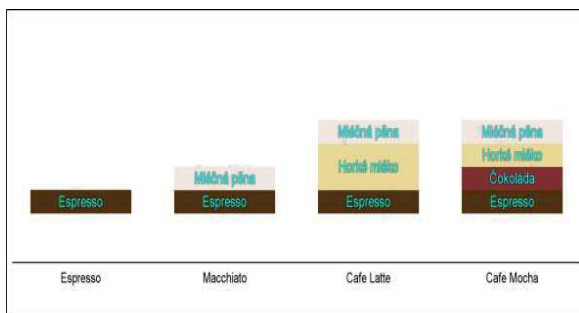


Obrázek 9 Všechny pixely uvnitř šálku jsou průhledné) (Zdroj: vlastní zpracování)

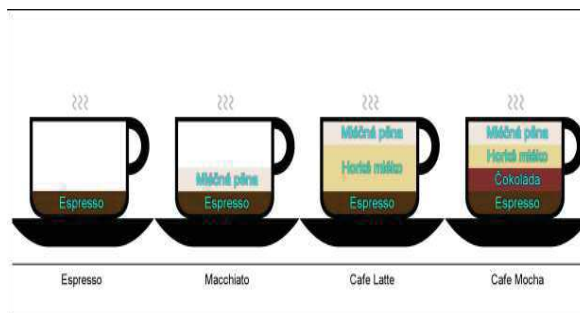


Obrázek 10 Obrázek šálku je vložen na červené pozadí (Zdroj: vlastní zpracování)

Tuto techniku můžeme použít k tvarování prvků grafu. Příkladem tohoto mohou být následující obrázky, ve kterém znázorněny recepty na čtyři různé populární kávové nápoje.



Obrázek 11 Graf složení kávových nápojů (Zdroj: vlastní zpracování)



Obrázek 12 Graf receptů kávových nápojů (Zdroj: vlastní zpracování)

Poznámka: Příkazy SAS potřebné pro tvorbu grafů na Obrázku 11 i na Obrázku 12 jsou v Příloze.

2.5 Použití obrazových souborů pro nastavení tvaru, jako masky a podoby v grafu

Prvním krokem je vytvoření sloupcového grafu obratu obchodů s Coca-Colou (v mil \$) znázorněném na Obrázku 13. Dalším krokem je tvarování všech sloupce do podoby láhve od Coca-Coly (viz Obrázek 14). Jako maska je použit obrázek klasické láhve od minerální vody, která se vloží do původního grafu jako další (překryvný) graf. Pro podobu láhve od Coca-Coly se použije obrazový soubor zprava z Obrázku 16. Tento obrázek je opačně transparentní než použitá maska. Vnější pixely jsou průhledné a pixely tvořící tvar láhve se zvýrazněním a odrazy ve skle jsou neprůhledné. K navrstvení tohoto obrázku na předchozí tvarovaný graf se použije další bodový graf. Obrázek láhve od Coca-Coly se na původní tvarovaný graf navrství s částečnou průhledností. To umožňuje, aby se barvy tvarovaných sloupců i barvy láhve ukázaly jako smíchané dohromady. Výsledný dekorativní graf je znázorněn na obrázku 15.

2.6 Symboly statistické grafiky jako uživatelsky definované znaky

Další důležitou funkcí procedury SGPLOT, kterou je možné využít ke tvorbě efektních statistických grafů systému SAS, je příkaz SYMBOLCHAR. Příkaz umožňuje nahradit původní značku ve statistickém grafu uživatelsky definovaným znakem. Uživatel do příkazu vkládá znak v kódové stránce Unicode¹ v hexadecimálním tvaru (Unicode 12.1 Character Code Charts, 2019). Příkaz byl poprvé zaveden do programu SAS ve verzi 9.4. Obecný tvar tohoto příkazu v rámci syntaxe procedury SGPLOT je následující:

SymbolChar name=symbolname char="hex-value" </ options>;

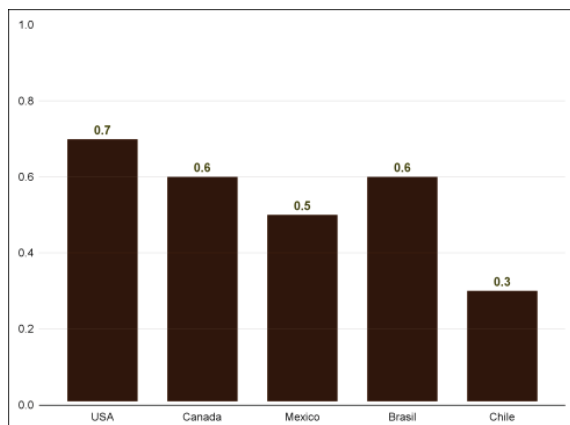
Podporovanými hodnotami argumentů jsou následující:

- HOFFSET = Horizontální posun k bodu umístění symbolu v souřadnicové soustavě,
- ROTATE = Úhel rotace symbolu ve stupních,
- SCALE = Škálovací faktor symbolu ke změnám velikosti značky a
- VOFFSET = Vertikální posun k bodu umístění symbolu v souřadnicové soustavě.

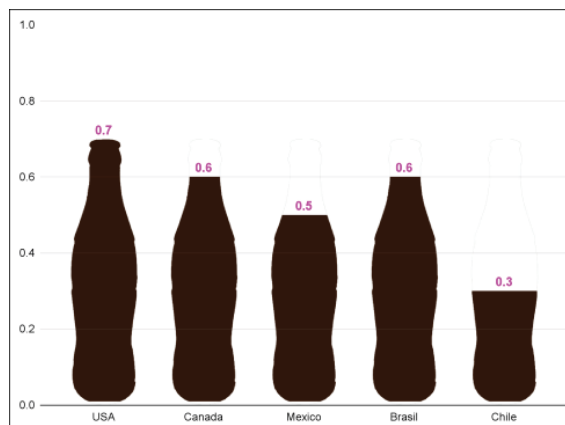
Parametry příkazu SYMBOLCHAR jsou stejné jako hodnoty parametrů u příkazu SYMBOLIMAGE. Znak použitý v uvedeném příkazu nahrazuje původní značku grafu a může být dále upravován v rámci procedury SGPLOT podle

¹ Unicode je mezinárodní standard, kterého cílem je definovat kódovací schéma schopné reprezentovat většinu znaků používaných v psaných jazycích spolu s dalšími symboly. Definuje konzistentní kódování pro reprezentaci a zpracovávání textů použitelné pro většinu v současnosti používaných písem.

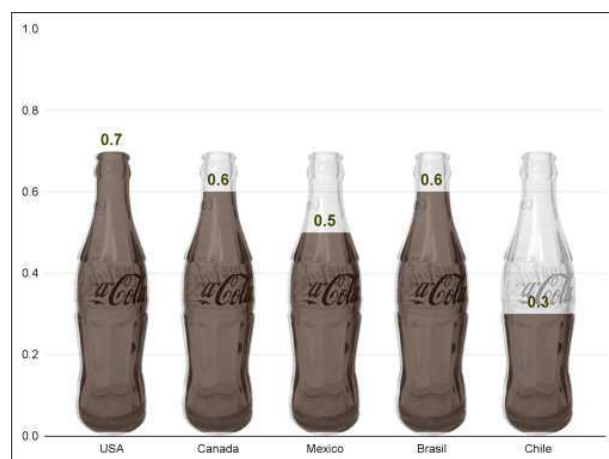
požadavků uživatele. Jako znaky se nepoužívají pouze písmena a číslice, ale mohou se využít různé emotikony² či piktogramy³.



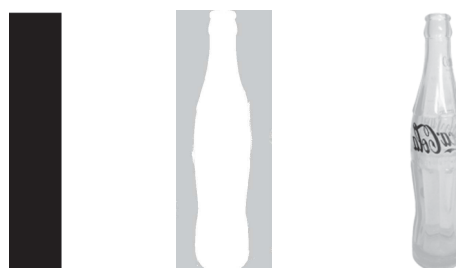
Obrázek 13 Sloupkový graf obrátů obchodu s Coca-Colou (Zdroj: vlastní zpracování)



Obrázek 14 Graf obrátů obchodu s Coca-Colou (tvarované sloupce maskou) (Zdroj: vlastní zpracování)



Obrázek 15 Graf obrátů obchodu s vodou ve tvaru láhve od Coca-Coly (Zdroj: vlastní zpracování)



Obrázek 16 Původní sloupec, soubor masky a soubor lahve od Coca-Coly (Zdroj: vlastní zpracování)

Poznámka: Příslušný kód pro všechny výše uvedené grafy je zaznamenán v Příloze.

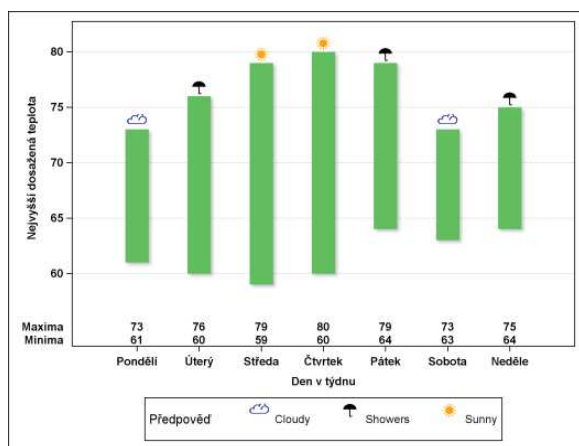
2.7 Emotikony a piktogramy jako symboly pro značky ve statistickém grafu

Příklad grafu s piktogramy je ilustrován na Obrázku 17. Piktogramy představují symboly pro předpověď počasí pro jednotlivé dny v týdnu. Pro znázornění piktogramu v grafu na Obrázku 17 byl v rámci syntaxe procedury SGPlot použit příkaz SYMBOLCHAR, do kterého byly vloženy hexadecimální znaky v kódové

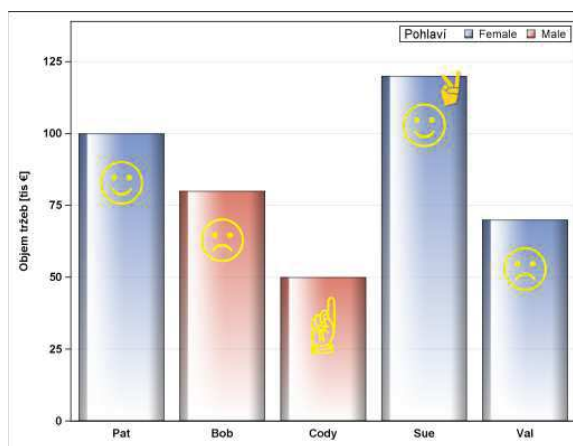
² Emotikon (též emotikona) je text připomínající tvarem a seskupením znaků výraz tváře, který vyjadřuje pisatelovu náladu, postoj či emoce. Obvykle je složený z interpunkčních a speciálních znaků, písmen nebo číslic.

³ Piktogram je grafický znak (ideogram) znázorňující pojem nebo sdělení obrazově. Většinou jde o malý a srozumitelný náčrt věci.

stránce Unicode vybraných piktogramů. Pro každý piktogram byl použit samostatný příkaz SYMBOLCHAR. Podobným způsobem vznikl také i graf obratu s emotikony z Obrázku 18. Zdrojové kódy k oběma grafům jsou uvedeny v dalším textu Přílohy.



Obrázek 17 Graf s piktogramy předpovědi počasí (Zdroj: vlastní zpracování)



Obrázek 18 Graf obratu s emotikony (Zdroj: vlastní zpracování)

3 Závěr

Cílem tohoto příspěvku je informovat odbornou veřejnost o méně známých možnostech statistické grafiky programového systému SAS. Největší pozornost byla zaměřena na funkce procedury SGPLOT, která podporuje tvorbu dekorativních statistických grafů. Tato procedura umožňuje jakýkoliv obrázek, piktogram nebo emotikon použít jako symbol pro značky ve statistickém grafu. Pokud se jako značky používají průhledné obrázky, mohou se použít jako tvary, masky nebo vzhledy. Vytvoření velmi efektních statistických grafů pomocí syntaxe procedury SGPLOT a nezbytných datových souborů lze dosáhnout kreativní kombinací značek vystupujících jako tvary, masky nebo vzhledy.

Poznání procedur statistické grafiky programového systému SAS a pochopení jejích možností představuje významný prvek pro zefektivnění práce většiny datových analytiků i ostatních uživatelů, čemuž má napomoci i předkládaný příspěvek.

4 Literatura

Matange, S. (2018). Decorative InfoGraphs using SAS. [online]. Cary, NC: SAS Institute, 2018. [cit. 21-10-2019]. Dostupné na: http://www.misug.org/uploads/8/1/9/1/8191072/phebb_bar_info_graphs.pdf.

SAS Institute (2019). SAS® 9.4 ODS Graphics: Procedures Guide. [online]. 6. vyd. Cary, NC: SAS Institute, 2019, 1816 s., ISBN 978-1-62960-799-3 [cit. 21-10-2019]. Dostupné na: <https://>

documentation.sas.com/api/docsets/grstatproc/9.4/content/grstatproc.pdf?locale=en#nameddest=titlepage.

Unicode 12.1 Character Code Charts 2019. [online]. 29. 7. 2019. [cit. 21- 10-2019]. Dostupné na: <http://unicode.org/charts>.

Příloha

```

/*--Macro na převod RGB hodnoty na Hexadecimální--*/
%macro RGBHex(rr, gg, bb);
    %sysfunc(compress(CX%sysfunc(putn(&rr, hex2.)) %sysfunc(putn(&gg,
    hex2.)) %sysfunc(putn(&bb, hex2.))))
%mend RGBHex;

ODS listing CLOSE;
/* XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX Proměnné XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX */
%let path=D:\SASData\odbor320\O_PAVELKA\SAS ODS GRAPHICS EXAMPLE\Images;

%LET coinIndian=&path.\2_annas_transparent.jpg;
%LET coinEuros=&path.\2_Euros_transparent.jpg;

%let file1=&path.\A1.png;
%let file2=&path.\B1.png;
%let file3=&path.\C1.png;
%let file4=&path.\A2.png;
%let file5=&path.\B2.png;

%let file6=&path.\Coffee_800_t4_Cleaned.png;
%let file7=&path.\CoffeeSteam_800.png;

%let In=&path.\Coke_280_In.png;
%let Out=&path.\Coke_300_Out_Gray.png;

ods graphics / reset width=640px height=480px attrpriority=none;

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 1 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
data Bar;
    input Year Resp;
    datalines;
2011 8
2012 14
2013 3
2014 12
;

proc Sgplot data=Bar noborder noautolegend;
    vbar year / response=resp dataskin=pressed datalabel
    datalabelattrs=(family="Arial" size=24pt color=red style=italic
    weight=bold);
    xaxis display=(noline noticks nolabel) integer;
    yaxis display=(noline noticks nolabel) integer grid;
run;

```

```

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 2 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
data Coins;
    set bar;
    YVal=resp+1;
    do Val=1 to Resp by 1; output; YVal=.; end;
run;

proc Sgplot data=Coins noborder noautolegend;
    symbolimage name=Indian image="&coinIndian" / rotate=0 hoffset=0
    voffset=0 scale=1.00;
    scatter x=year y=val / markerattrs=(symbol=Indian size=110) jitter
    jitterwidth=0.05;
    text x=year y=YVal text=resp / strip position=top backlight=0.75
    textattrs=(family="Arial" size=32pt color=red style=italic
    weight=bold);
    xaxis display=(noticks noline nolabel) offsetmin=0.15 offsetmax=0.15
    INTEGER;
    yaxis display=none offsetmin=0.2 offsetmax=0.2;
run;

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 3 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
proc Sgplot data=Bar noborder noautolegend;
    vbar year / response=resp dataskin=pressed datalabel
    datalabelattrs=(family="Arial" size=24pt color=red style=italic
    weight=bold);
    xaxis display=(noline noticks nolabel) integer;
    yaxis display=(noline noticks nolabel) integer grid;
run;

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 4 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
proc Sgplot data=bar noborder noautolegend;
    scatter x=year y=resp / markerattrs=(symbol=circlefilled size=100)
    filledoutlinedmarkers dataskin=gloss;
    text x=year y=resp text=resp / strip position=top backlight=0.75
    textattrs=(family="Arial" size=24pt color=red style=italic
    weight=bold);
    xaxis display=(noticks noline nolabel) offsetmin=0.15 offsetmax=0.15
    integer;
    yaxis display=none offsetmin=0.2 offsetmax=0.2;
run;

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 5 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
proc Sgplot data=Coins noborder noautolegend;
    scatter x=year y=val / markerattrs=(symbol=circlefilled size=100)
    filledoutlinedmarkers dataskin=gloss;
    text x=year y=Yval text=resp / strip position=top backlight=0.75
    textattrs=(family="Arial" size=24pt color=red style=italic
    weight=bold);
    xaxis display=(noticks noline nolabel) offsetmin=0.15 offsetmax=0.15
    integer;
    yaxis display=none offsetmin=0.2 offsetmax=0.2;
run;

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 6 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
proc Sgplot data=Coins noborder noautolegend;
    symbolimage name=Euros image="&coinEuros" / rotate=0 hoffset=0
    voffset=0 scale=1.00;
    scatter x=year y=val / markerattrs=(symbol=Euros size=110) jitter
    jitterwidth=0.05;

```



```

text x=year y=YVal text=resp / strip position=top backlight=0.75
textattrs=(family="Arial" size=32pt color=red style=italic
weight=bold);
xaxis display=(noticks noline nolabel) offsetmin=0.15 offsetmax=0.15
INTEGER;
yaxis display=none offsetmin=0.2 offsetmax=0.2;

run;

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 7 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
data Readiness;
    format value percent.;
    input 'Typ školy'n $1-9 Level $10-22 Value;
    datalines;
Soukromé Ukončení ZŠ 1.00
Soukromé Ukončení SŠ 0.72
Soukromé Přijetí na VŠ 0.45
Soukromé Studium na VŠ 0.24
Státní Ukončení ZŠ 1.00
Státní Ukončení SŠ 0.80
Státní Přijetí na VŠ 0.50
Státní Studium na VŠ 0.31
;
run;

proc Sgplot data=Readiness;
    styleattrs datacolors=(%rgbhex(254, 145, 104) %rgbhex(130, 109, 146));
    hbarparm category=level response=value / group='Typ školy'n
    groupdisplay=cluster dataskin=sheen
        datalabel datalabelattrs =(Color=Green Family=Arial Size=12
        Style=Italic Weight=Bold)
    outlineattrs=(color=black);
    xaxis values=(0 to 1.1 by 0.10) labelattrs=(size=16pt weight=bold
    color=gray33) valueattrs=(size=10pt weight=bold color=gray33)
        offsetmax=0 grid minor minorcount=1 display=(noticks NOLABEL
    NOVALUES);
    yaxis labelattrs=(size=16pt weight=bold color=gray33)
    valueattrs=(size=10pt weight=bold color=gray33) display=(noticks
    NOLABEL);
    keylegend / border location=inside position=bottomright opaque
    titleattrs=(weight=bold) valueattrs=(weight=bold);

run;

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 8 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
/*--Replicate observations by value and School--*/
data Readiness3;
    format valueLblA ValueLblB percent.;
    set Readiness;
    x120=1.2; x0=-0.2; x130=1.3; x270=2.7;
    start_School=0; start_State=2.5;
    xmiss=1;
    if 'Typ školy'n eq 'Soukromé' then
    /*--Replicate values for "Soukromé"--*/
        do; valueLblA=value; do Val=0.1 to 1 by 0.1; Group='C1'; if Val
    <= value then Group='A1';
            else if Val > value and Val < value+0.099 then Group='A2';
            output; valueLblA=.; end;
        end;
    else

```

```

/*--Replicate values for "Státní"--*/
do; valueLb1B=value; do i=0.1 to 1 by 0.1; Group='C1'; if i <=
value then Group='B1';
else if i > value and i < value+0.099 then Group='B2'; val=i+1.5;
output; valueLb1B=.;
end;
end;
run;

/*--Append data for headers--*/
data Readiness4;
set Readiness3 end=last;
output;
if last then
do; call missing (val, level); xlbl=0.5; text='Soukromé
školství'; output; xlbl=2.05; text='Státní školství'; output; end;
run;

proc Sgplot data=Readiness4 noborder noautolegend;
symbolimage name=A1 image="&file1";
symbolimage name=B1 image="&file2";
symbolimage name=C1 image="&file3";
symbolimage name=A2 image="&file4";
symbolimage name=B2 image="&file5";
styleattrs datasymbols=(A1 A2 C1 B1 B2/* C1*/)
datacontrastcolors=(%rgbhex(254, 145, 104) %rgbhex(130, 109, 146));
xaxistable text / x=xlbl colorgroup=text location=inside position=top
nolabel valueattrs=(weight=bold size=10);
scatter y=level x=val / group=group markerattrs=(size=40);
text x=x0 y=level text=valueLb1A / textattrs=(color=%rgbhex(254, 145,
104) size=9);
text x=x270 y=level text=valueLb1B / textattrs=(color=%rgbhex(130, 109,
146) size=9);
text x=x130 y=level text=level / textattrs=(color=gray size=8)
splitchar='.' splitpolicy=splitalways;
xaxis display=none max=2.7 offsetmin=0.05 offsetmax=0.0;
yaxis display=none reverse splitchar='.' offsetmin=0.1
fitpolicy=splitalways splitjustify=center;
run;

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 11 XXXXXXXXXXXXXXXXXXXXXXXX';
ods graphics / width=640px height=240px attrpriority=none;
data AttrMap;
length id value fillcolor $15;
id='Coffee'; value='Espresso'; fillcolor='cx4f2f00'; output;
id='Coffee'; value='Čokoláda'; fillcolor='cx7f2f2f'; output;
id='Coffee'; value='Horké mléko'; fillcolor='cream'; output;
id='Coffee'; value='Mléčná pěna'; fillcolor='cxefe7e0'; output;
run;

data coffee;
input Name $1-20 Group $21-40 Value low high;
mid=(low+high)/2; y=0; y=y+0.013; ys=1.2;
datalines;
Espresso Espresso 0.25 0.0 0.25
Macchiato Espresso 0.25 0.0 0.25
Macchiato Mléčná pěna 0.25 0.25 0.5
Cafe Latte Espresso 0.25 0.0 0.25
Cafe Latte Horké mléko 0.50 0.25 0.75
Cafe Latte Mléčná pěna 0.25 0.75 1.0

```

Cafe Mocha	Espresso	0.25	0.0	0.25
Cafe Mocha	Čokoláda	0.25	0.25	0.50
Cafe Mocha	Horké mléko	0.25	0.50	0.75
Cafe Mocha	Mléčná pěna	0.25	0.75	1.0

;

run;

```
proc Sgplot data=Coffee noborder noautolegend nocycleattrs datrmap=attrmap
  pad=(bottom=10pct);
  highlow x=name low=low high=high / group=group type=bar nooutline
  barwidth=0.7 intervalbarwidth=0.5 name='a' attrid=Coffee;
  text x=name y=mid text=group / backlight=0.4 textattrs=(color=cyan
  SIZE=10);
  xaxis display=(nolabel noticks) offsetmin=0.12 offsetmax=0.12;
  yaxis display=none min=0 max=1 offsetmin=0.20 offsetmax=0.42;
run;
```

```
title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 12 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
```

```
proc Sgplot data=Coffee noborder noautolegend nocycleattrs datrmap=attrmap
  pad=(bottom=10pct);
  symbolimage name=Cup image="&file6" / voffset=0.15;
  symbolimage name=Steam image="&file7" / voffset=0.00;
  highlow x=name low=low high=high / group=group type=bar nooutline
  barwidth=0.7 intervalbarwidth=0.5 name='a' attrid=Coffee;
  scatter x=name y=y / markerattrs=(symbol=Cup size=190);
  scatter x=name y=ys / markerattrs=(symbol=Steam size=110);
  text x=name y=mid text=group / backlight=0.4 textattrs=(color=cyan
  SIZE=10);
  xaxis display=(nolabel noticks) offsetmin=0.12 offsetmax=0.12;
  yaxis display=none min=0 max=1 offsetmin=0.20 offsetmax=0.42;
run;
```

```
title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 13 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
```

```
data coke;
  format high 3.1;
  input Region $1-10 Product $11-20 Value Low High;
  Mid=(low+high)/2; y=0; ylbl=0.1; low=low+0.01;
  datalines;
```

USA	Coke	0.7	0.0	0.699
Canada	Coke	0.6	0.0	0.6
Mexico	Coke	0.5	0.0	0.5
Brasil	Coke	0.6	0.0	0.6
Chile	Coke	0.3	0.0	0.3

;

run;

```
ods graphics / width=640px height=480px attrpriority=none;
```

```
/*--Bars Only--*/
```

```
proc Sgplot data=coke noborder noautolegend nocycleattrs;
  highlow x=Region low=low high=high / type=bar nooutline barwidth=0.7
  fillattrs=(color=cx2f0f00) highlabel=high labelattrs=(size=10
  weight=bold color=cx3f3f00);
  refline 0.2 0.4 0.6 0.8 / transparency=0.8;
  xaxis display=(nolabel noticks) offsetmin=0.12 offsetmax=0.12;
  yaxis display=(nolabel noticks noline) min=0 max=1 offsetmin=0
  offsetmax=0 values=(0 to 1 by 0.2);
run;
```

```

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 14 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
/*--Bars with Masks--*/
proc Sgplot data=coke noborder noautolegend nocycleattrs;
    symbolimage name=In image="&In" / voffset=0.5 hoffset=0.01;
    highlow x=Region low=low high=high / type=bar nooutline barwidth=0.7
    fillattrs=(color=cx2f0f00) highlabel=high labelattrs=(size=10
    weight=bold color=cxFF00FF);
    scatter x=Region y=y / markerattrs=(symbol=In size=300);
    refline 0.2 0.4 0.6 0.8 / transparency=0.8;
    xaxis display=(nolabel noticks) offsetmin=0.12 offsetmax=0.12;
    yaxis display=(nolabel noticks noline) min=0 max=1 offsetmin=0
    offsetmax=0 values=(0 to 1 by 0.2);
run;

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 15 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
/*--Bars with Masks and Skins--*/
proc Sgplot data=coke noborder noautolegend nocycleattrs;
    symbolimage name=In image="&In" / voffset=0.5 hoffset=0.01;
    symbolimage name=Out image="&Out" / voffset=0.5;
    highlow x=Region low=low high=high / type=bar nooutline
    barwidth=0.7 fillattrs=(color=cx2f0f00);
    scatter x=Region y=y / markerattrs=(symbol=In size=300);
    scatter x=Region y=y / markerattrs=(symbol=Out size=300)
    transparency=0.5;
    highlow x=Region low=low high=high / type=bar nooutline barwidth=0.7
    fillattrs=(transparency=1.0) highlabel=high labelattrs=(size=12
    weight=bold color=cx3f3f00);
    refline 0.2 0.4 0.6 0.8 / transparency=0.8;
    xaxis display=(nolabel noticks) offsetmin=0.12 offsetmax=0.12;
    yaxis display=(nolabel noticks noline) min=0 max=1 offsetmin=0
    offsetmax=0 values=(0 to 1 by 0.2);
run;

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 17 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX';
data temperatures;
    input Day : $9. High Low Forecast $7.;
    datalines;
Pondělí 73 61 Cloudy
Úterý 76 60 Showers
Středa 79 59 Sunny
Čtvrtek 80 60 Sunny
Pátek 79 64 Showers
Sobota 73 63 Cloudy
Neděle 75 64 Showers
;
run;

data myattrmap;
    input id $ value $ markercolor $ markersymbol $;
    datalines;
myid Cloudy blue Cloud
myid Sunny orange Sun
myid Showers black Umbrella
;
run;

```

```

proc Sgplot data=temperatures dattrmap=myattrmap;
  symbolchar      name=Cloud      char='2601'x      /      voffset=.35
  textattrs=(weight=bold);
  symbolchar      name=Sun        char='2600'x      /      voffset=.35
  textattrs=(weight=bold);
  symbolchar      name=Umbrella   char='2602'x      /      voffset=.35
  textattrs=(weight=bold);
  styleattrs datacontrastcolors=(black);
  highlow x=day low=low high=high / lineattrs=(color=biyg thickness=20pt)
  dataskin=sheen;
  scatter x=day y=high / markerattrs=(size=25pt) attrid=myid
  group=forecast name='a';
  xaxis label='Den v týdnu' labelattrs=(size=9pt weight=bold)
  valueattrs=(size=9pt weight=bold);
  xaxistable high / label='Maxima' labelattrs=(size=9pt weight=bold)
  valueattrs=(size=9pt weight=bold) location=inside;
  xaxistable low / label='Minima' labelattrs=(size=9pt weight=bold)
  valueattrs=(size=9pt weight=bold) location=inside;
  yaxis grid gridattrs=(pattern=MediumDash) label='Nejvyšší dosažená
  teplota' labelattrs=(size=9pt weight=bold) valueattrs=(size=9pt
  weight=bold);
  keylegend 'a' / title='Předpověď';
run;

```

```

title 'XXXXXXXXXXXXXXXXXXXXXXXXXXXX Obrázek 18 XXXXXXXXXXXXXXXXXXXXXXXXXXXX';
data bars;
  length Status $6;
  input Cat $ Sex $ Sales;
  loc=Sales-10;
  if Sales GE 100 then Status='Great'; else if Sales GE 60 then
  Status='Good'; else Status='Poor';
  if Sales=120 then do; Ribbon=1; rloc=110; end;
  datalines;
Pat   Female   100
Bob   Male     80
Cody  Male     50
Sue   Female   120
Val   Female   70
;
run;

```

```

data myattrmap;
  input id $ value $ markercolor $ markersymbol $;
  datalines;
myid Great white Great
myid Good white Good
myid Poor white Poor
;
run;

```

```

proc Sgplot data=bars dattrmap=myattrmap noautolegend;
  symbolchar name=Great char='263A'x / voffset=-0.15;
  symbolchar name=Good char='2639'x / voffset=-0.15;
  symbolchar name=Poor char='261D'x / voffset=-0.15;
  symbolchar name=Ribbon char='270C'x / voffset=0.30 rotate=15;
  vbarparm category=cat response=Sales / group=sex dataskin=gloss
  groupdisplay=cluster filltype=gradient name='x';
  scatter x=cat y=loc / markerattrs=(color=yellow size=80pt) attrid=myid
  group=status;

```

```
scatter x=cat y=rloc / group=ribbon discreteoffset=0.25
markerattrs=(symbol=Ribbon size=80 color=gold) dataskin=crisp ;
xaxis display=(nolabel noticks) valueattrs=(size=9pt weight=bold)
offsetmin=0.1 offsetmax=0.1;
yaxis offsetmin=0 offsetmax=0.1 grid gridattrs=(pattern=MediumDash)
label='Objem tržeb [tis €]' labelattrs=(size=9pt weight=bold)
valueattrs=(size=9pt weight=bold);
keylegend 'x' / title='Pohlaví' position=topright location=inside
noopaque;

run;

ods graphics / reset;
title;
```


Impact of population ageing on health services: a case study for Slovakia

Vplyv starnutia populácie na zdravotné služby: prípadová štúdia pre Slovensko

Richard Priesol

Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Mlynská dolina F1, 842 48 Bratislava, Slovenská republika

Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Mlynská dolina F1, 842 48 Bratislava, Slovak Republic

richard.priesol@gmail.com

Abstract: *Population ageing presents a serious problem for the economy of Slovakia. This negative demographic trend leads to an increase of the population in the post-productive age and a decline in the productive population. This situation has a negative impact not only on the social and pension systems but also on other sectors of the economy; for example, on health services. In this paper we explain an impact of the population ageing on the total production in the economy, with further implications for the public sector, through a computable general equilibrium model. Since a lower public income leads also to a lower income in the sector of health services, we can evaluate an impact of the population ageing on the financial situation in this highly important economic sector. We conduct a case study for Slovakia based on a model simulation of different demographic projections from the baseline year 2016 to the target year 2030, including the population ageing scenario as well as potential improvements in the productive population.*

Abstrakt: *Starnutie populácie predstavuje vážny problém pre ekonomiku Slovenska. Tento negatívny demografický trend vedie k nárastu populácie v poproduktívnom veku a poklesu v produktívnej populácii. Táto situácia negatívne vplyva nielen na sociálny a dôchodkový systém, ale aj na ostatné sektory ekonomiky, napríklad zdravotné služby. V tomto článku za pomoci modelu všeobecnej spočítateľnej rovnováhy vysvetľujeme vplyv starnúcej populácie na celkovú ekonomickú produkciu, vrátane ďalších implikácií na verejný sektor. Keďže nižší verejný dôchodok vedie tiež k nižšiemu dôchodku v sektore zdravotných služieb, môžeme vyhodnotiť vplyv starnutia populácie na finančnú situáciu tohto významného ekonomického sektora. Realizujeme prípadovú štúdiu pre Slovensko založenú na modelovej simulácii rozličných demografických projekcií zo základného roku 2016 do cieľového roku 2030 zahrňujúceho scenár starnutia populácie, ako aj potenciálne zlepšenia v produktívnej populácii.*

Kľúčové slová: *starnutie populácie, zdravotné služby, všeobecná teória rovnováhy.*

Key words: *population ageing, health services, general equilibrium theory.*

1 Introduction

Population ageing is generally characterized as an upward structural change in the median population age, due to a decline in the fertility rates or an increase in the life expectancy. This negative demographic trend has several serious

implications for the economic development and structure. First, more people in the post-productive age put higher pressure on the public expenditures, mostly on the social and pension systems and the health services. Second, less people in the productive age lead to a decline in the public income from taxes and contributions and thus additional pressure on the government budget. Finally, the population ageing leads to a decline in the labour supply factor and thus suppresses the potential production in the economy. In the end, these implications lead either to higher taxes and contributions or higher public deficit to fund additional government spending.

Structural economic changes should be made to absorb the impact of demographic factors, for example an increase in the retirement age or a linkage between the retirement age and the life expectancy. On the other hand, we should ask what are the drivers of this negative demographic trend, what measures could be adopted to maintain sustainable demographic situation and finally, what is the true impact of these improvements. In this article, we propose a simple general equilibrium model to simulate the impact of demographic factors on the total production in the economy through the production function theory. Subsequently, we evaluate an impact of the population ageing on the public finance, with a focus on the health services as one of the most affected sectors in the economy. Finally, we compare the actual demographic projection with the potential improvements in the productive population and their impact on the domestic economy. The article is based on a case study for Slovakia from the baseline year 2016 to the target year 2030.

Confronting the theoretical concept with the statistical data of a population structure, we observe significant differences between the actual demographic situation compared to the historical dataset for the economy of Slovakia (Figure 1). While the productive population from 15 to 64 years is declining since 2012, the population in the post-productive age is rising through the history with the strongest dynamics in the last years. Recent changes in the population structure are thus in line with the process of population ageing.

There are two explanations of this development. First, declining fertility rates lead to a downswing in the productive population observed in the last years and expected also in the future (Figure 2). Second, while the mortality and fertility rates are on a similar level, the population problem is characterized by the population ageing rather than a decline in the total population (Figure 3). The paper is organized as follows. First, we propose a literature review with focus on the international and domestic research of the population ageing. Second, we define a simple general equilibrium model suitable for the demographic projections, further enriched for the public expenditures on the health services.

Finally, we propose the details of model simulations and the discussion of simulation results.

This research was supported by project APVV-15-0722 *Social trap – costs and the way out*.

2 Related literature

There is a number of papers evaluating an impact of the population ageing on the economic situation in particular countries. We should mention the work of Ignaciuk et al. (2008) arguing that the standard connection between the labour and population dynamics widely used in the macroeconomic models is oversimplified, especially for various demographic projections. Subsequently, they propose an alternative solution based on the endogenous connection between the demographic and macroeconomic models and evaluate different structural changes in the population structure, for example a decline in the fertility and mortality rates.

Authors argue that the process of population ageing and subsequent changes in the dependency ratio have a substantial effect on the economic growth, especially in the European countries. According to this, they propose a number of policy implications to compensate this negative demographic trend, for example an increase in the fertility and participation rates or a modification of the retirement age. However, despite positive simulation results, these changes are not able to fully compensate the impact of the population ageing. Similarly, the paper by McGregor et al. (2008) proposes a link between the demographic and macroeconomic models to evaluate an impact of the population ageing on the economy of Scotland.

Population ageing becomes a relevant topic also for the economy of China, as proposed by Peng (2009). The paper argues that this negative demographic trend leads to a decline in the economic growth according to the explicit decrease of the labour supply through the population projection and the implicit decrease

of the capital formation through the investment channel. However, the living standards of households improve and a decline in the savings is overcome by a decline in the investments, thus creating a surplus on the capital market and a positive current account balance. Furthermore, the model by Peng and Mai (2013) analyses an impact of the retirement age extension on the economic growth, under the population ageing scenario in the economy of China.

On the other hand, the paper by Turner (2009) evaluates an impact of the population ageing on the economy of Great Britain. The author argues that the

standard assumptions based on the fertility and immigration factors as the optimal compensation of this negative demographic trend under the fixed retirement age are overstated and essentially wrong. Subsequently, he proposes an alternative solution in the form of a welfare optimizing model incorporating additional factors like increase in the life expectancy, increase in the retirement age, capital inheritance and population density to capture the evolution of the dependency ratio and possible implications for the public policies.

We should mention also the work of Szreter (1993), describing the process of demographic transition from a developing economy characterized by higher fertility and mortality rates to a developed economy with advanced technologies and education system characterized by rather low fertility and mortality rates. Furthermore, the paper by Rosero-Bixby and Casterline (1993) analyses a decline in the fertility rates through the diffusion effects that depart from the standard theory based on the economic and social progress as the main explanation of the fertility dynamics, thus proposing more complex look on the population projection.

Finally, the work by Páleník et al. (2014) analyses the problem of population ageing in the context of Slovakia. They evaluate implications of the population ageing for the potential production as well as an impact on the macroeconomic development through a general equilibrium model of an open economy. The model distinguishes between the productive and post-productive population as well as the individual consumption sectors. Population ageing leads to a decline in the total employment and consumption of the productive households partially compensated by the post-productive households and the government.

3 Macroeconomic model

Evaluation of demographic implications for the economy of Slovakia is performed by a standard Computable general equilibrium (CGE) model of an open economy. We apply a regular two-level production function to model the production chain of the value added and the intermediate consumption, similarly to Shen and Whalley (2013). In the first step, we incorporate two production factors, the labour factor (L_t) and the capital factor (K_t), to explain the value added in the economy under the Cobb-Douglas function form. In the second step, we model the total production in the economy as a function of the value added (VA_t) and the intermediate consumption (IC_t) under the Leontief function form, thus assuming complementarity of the production factors. Corresponding prices of the production factors are determined by the first order conditions, in line with a standard assumption of a zero profit.

$$VA_t = CobbDouglas[L_t, K_t] \quad (1)$$

$$Y_t = Leontief[VA_t, IC_t] \quad (2)$$

Export and import of commodities are captured by the Armington model of international trade with the Constant elasticity of substitution (CES) function incorporating the imported commodities (M_t) to the total consumption (Q_t) and the Constant elasticity of transformation (CET) function incorporating the exported commodities (X_t) to the total production (Y_t). Both equations are closed by the domestic component (D_t). Furthermore, export and import prices are determined by the external factors, including external deflators and exchange rates, while the domestic price is chosen as the model numeraire. Trade elasticities are estimated by a time-series econometric approach based on the first order conditions, similarly to Gallaway et al. (2003). Estimation results are in line with the standard literature, assuming relatively high substitutability between the domestic and external markets. For further information about the estimation process and results see Priesol (2018). Again, the system of trade equations is closed by the first order conditions.

$$Q_t^{-\gamma_1} = \alpha_1^{-\gamma_1} \beta_1 D_t^{-\gamma_1} + \alpha_1^{-\gamma_1} (1 - \beta_1) M_t^{-\gamma_1} \quad (3)$$

$$Y_t^{\gamma_2} = \alpha_2^{\gamma_2} \beta_2 D_t^{\gamma_2} + \alpha_2^{\gamma_2} (1 - \beta_2) X_t^{\gamma_2} \quad (4)$$

Calibration of the model is based on constant depreciation and saving rates and constant ratios between the structural variables (public deficit and current account) and the gross domestic product. We thus assume that the agents in the economy maintain either their personal preferences or their domestic market share. Private consumption (C_t) and public consumption (G_t) are determined by the budget restrictions of households and government, while the gross capital formation (I_t) results from the market clearing condition. The model is closed by the exogenous intersectoral transfers. Finally, share of the public spending on the health services (H_t) is captured by the ratio between the natural social transfers and the public consumption in the reference year.

$$Q_t = IC_t + I_t + C_t + G_t \quad (5)$$

Budget of households is driven by the income from labour and capital and intersectoral transfers from corporates (property income), government (social transfers) and external world (external households). Private consumption is then determined by the savings of households. Income of corporates is mostly driven by the operating surplus, with the expenditures distributed between savings, households, government and external world. Budget of government is driven by taxes and contributions from the labour and capital income together with the

indirect taxes from consumption, import and production. Public deficit then determines the disposable income distributed between social transfers and public consumption. Finally, the external sector is captured by the trade balance and intersectoral transfers and closed with the current account.

Dynamization of the model variables is performed through the labour and capital production factors, with the first one captured by an expected projection of the employment dynamics and the latter one by a capital accumulation process based on the perpetual inventories method (PIM). Furthermore, dynamization of the model prices is captured by an expected projection of the price numeraire, specifically the producer price index (PPI). Although the population ageing leads to a decline in the productive population, significant part of this decline is compensated by higher participation rate (pension workers), better migration balance (external workers) and lower unemployment (requalification).

Relationship between the population and employment decline is obtained from the Ageing Report of the European Commission (2018). While we operate with the official population projection for Slovakia, the employment projection is derived from the participation and unemployment rates of the European Union, due to additional domestic factors not influenced by the population ageing, for example a different view of the structural participation and unemployment in the medium-term horizon.

Model calibration is based on the Social accounting matrix (SAM) in the reference year 2016. Historical data are obtained from the Statistical Office of the Slovak Republic, specifically from the sectoral national accounts proposing the necessary information for a construction of the calibration matrix (the nu1028rs table in the DataCube database available on the internet at <http://datacube.statistics.sk>). Projection of the inflation rate is based on the historical values from the latest year to capture the most actual price development. Estimation of the model is based on a set of 13 structural equations, including the production and trade functions, the first order conditions and the market clearing condition, that are solved by a standard trust-region algorithm in Matlab.

4 Discussion of results

In line with the historical data of the population structure, we simulate a decline in the productive population, in contrast to the neutral demographic scenario, to capture an impact of the population ageing on the labour supply. Furthermore, we simulate a positive population projection at the levels of 1% (conservative scenario) and 2% (optimistic scenario). While leaving other

parameters of the model unchanged, we could thus evaluate an impact of the demographic improvement on the economy of Slovakia through the dynamic simulations. Model evaluation is based on the 15-years simulation horizon from the baseline year 2016 to the target year 2030. Results are proposed in the Appendix.

When compared to the neutral demographic scenario, the population ageing results in a negative projection of the productive population and total employment affecting the sector of health services by 0.8% in 2030. In contrast, the conservative scenario leads to a potential improvement by almost 2.2% of the population ageing projection and the optimistic scenario results in the improvement by more than 3.6% in 2030. However, if we consider that the additional public income is transferred to the sector of health services, according to higher requirements on the public health driven by the post-productive population, we obtain more than 8.1% additional income under the conservative scenario and almost 13.5% under the optimistic scenario.

Furthermore, we compare the results with the international case studies of Peng and Mai (2013) and McGregor et al. (2008). While the first one calculates an impact of the retirement age extension on the economic growth, the second one analyses different macroeconomic projections under the population ageing scenario. Improvement of the domestic employment by 1.4% leads to an increase of the economic performance by 1.3%, according to Peng and Mai (2013). On the other hand, the baseline simulation by McGregor et al. (2008) assumes that a decline in the domestic employment by 6.1% leads to worse economic performance by approximately 5.4%. These numbers are mostly in line with the proposed results, in terms of the elasticity between the domestic employment and the gross domestic product, since an increase in the total employment by 1.0% implied by the neutral demographic scenario leads to better economic performance by almost 0.8%.⁴

Finally, we compare the results with the model of Páleník et al. (2014), calibrated on the domestic dataset. We assume that the population ageing leads to a decline in the total employment by approximately 1.0% in the target year 2030, in line with 0.9% estimated by Páleník et al. (2014). Dropout in the domestic employment leads to a decline in the final consumption by approximately 0.8% in the target year 2030, in contrast to 0.4% proposed by Páleník et al. (2014), driven mostly by a dynamization of the model and thus higher cumulative impact on the economy compared to the static specification.

⁴ While the model by Peng and Mai (2013) operates on a time horizon from the baseline year 2014 to the target year 2030, the model by McGregor et al. (2008) initiates at the baseline year 2000.

The difference is even higher for the gross domestic product, explained by a different approach to the model closure, based on fixed model ratios rather than fixed model levels, together with a different reference year.

Although the simulation results do not produce any specific policy implications, since we have no additional information about the structure of the population changes, we could propose some explanations of a reasonable structure of the demographic shift. According to the global factors beyond the structure of a particular economy, rising the replacement ratio in a form of higher fertility and immigration is not sustainable in the long-term horizon. On the other hand, more effective development of the labour market in a form of better participation and unemployment rates reaches the limits of the structural inactivity and unemployment.

We thus argue that the structural changes in the pension system are necessary to maintain sustainable economic development. For example, a linkage between the retirement age and the life expectancy should stabilize the dependency ratio and ensure the sustainability of the public finance, as proposed by Porubský and Novysedlák (2018). We should also mention that the incorporation of the health services into the model is performed from a demand side of the economy, specifically through the consumption of the public sector. However, this approach does not incorporate the sectoral and production point of view. To model a development of the economic sectors, we need to analyse a supply side of the economy and perform a sectoral disaggregation of the general equilibrium model.

5 Conclusions

In this article, we outlined negative implications of the population ageing for the economy of Slovakia under no policy change scenario. We then constructed a simple general equilibrium model of an open economy to evaluate the impact of this negative demographic trend on the production sector and implicitly also on the sector of health services. Subsequently, under two basic scenarios of the population improvement, we evaluated an impact of positive changes in the demographic projection on the public income and expenditures. Finally, we argued that the structural changes in the pension system should be made to maintain sustainable economic development under the population ageing scenario.

6 Bibliography

European Commission (2018). *The 2018 Ageing Report: Underlying Assumptions and Projection Methodologies*. Working paper, European Commission, 2018.

- Gallaway, M. P., McDaniel, C. A., Rivera, S. A. (2003). *Short-run and Long-run Industry-level Estimates of U.S. Armington Elasticities*. The North American Journal of Economics and Finance, 14 (1), 49-68.
- Hillberry, R., Hummels, D. (2012). *Trade Elasticity Parameters for a Computable General Equilibrium Model*. Handbook of Computable General Equilibrium Modeling, Elsevier, 2012.
- Ignaciuk A. M., Hilderink H., Peterson S. (2008). *Integrating Demography into a Regional CGE Model*. Working paper, Kiel Institute for the World Economy, 2008.
- Imbs, J., Mejean, I. (2010). *Trade Elasticities: A Final Report for the European Commission*. Working paper, European Commission, 2010.
- Koesler, S., Schymura, M. (2012). *Substitution Elasticities in a CES Production Framework: An Empirical Analysis on the Basis of Non-linear Least Squares Estimations*. Discussion paper. Centre for European Economic Research, 2012.
- Mcgregor, P., Lisenkova, K., Turner, K., Swales, K., Pappas, N. (2008). *Scotland the Grey: A Linked Demographic CGE Analysis of the Impact of Population Ageing and Decline*. Regional studies, 44 (10), 1351-1368.
- Páleník, V. et al. (2014). *Silver Economy: Potential on Slovakia*, Bratislava, Slovakia, 2014. ISBN 978-80-7144-234-9.
- Peng, X. (2009). *Demographic Shift, Population Ageing and Economic Growth in China: A Computable General Equilibrium Analysis*. Pacific Economic Review, 13 (5), 680-697.
- Peng, X., Mai, Y. (2013). *Population Ageing, Retirement Age Extension and Economic Growth in China – A Dynamic General Equilibrium Analysis*. Working Paper, Centre of Policy Studies, Monash University, 2013.
- Porubský, M., Novýsedlák, V. (2018). *Retirement Age: Myths and Facts*. Discussion Paper, Council for Budget Responsibility, 2018.
- Priesol, R. (2018). *Estimation of the Trade Elasticities in Slovakia*. Acta Aerarii Publici, 15(2), 87-98.
- Rosero-Bixby, L., Casterline, J.B. (1993). *Modelling Diffusion Effects in Fertility Transition*. Population studies, 47(1), 147-167.
- Shen, K., Whalley, J. (2013). *Capital-Labor-Energy Substitution in Nested CES Production Functions for China*. Working paper, National Bureau of Economic Research, 2013.
- Szreter, S. (1993). *The Idea of Demographic Transition and the Study of Fertility Change*. Population and Development Review, 19 (1), 659-702.
- Turner, A. (2009). *Population Ageing: What Should We Worry About?* Philosophical Transactions of the Royal Society, Vol. 364(1), 3023-3030.

Appendix

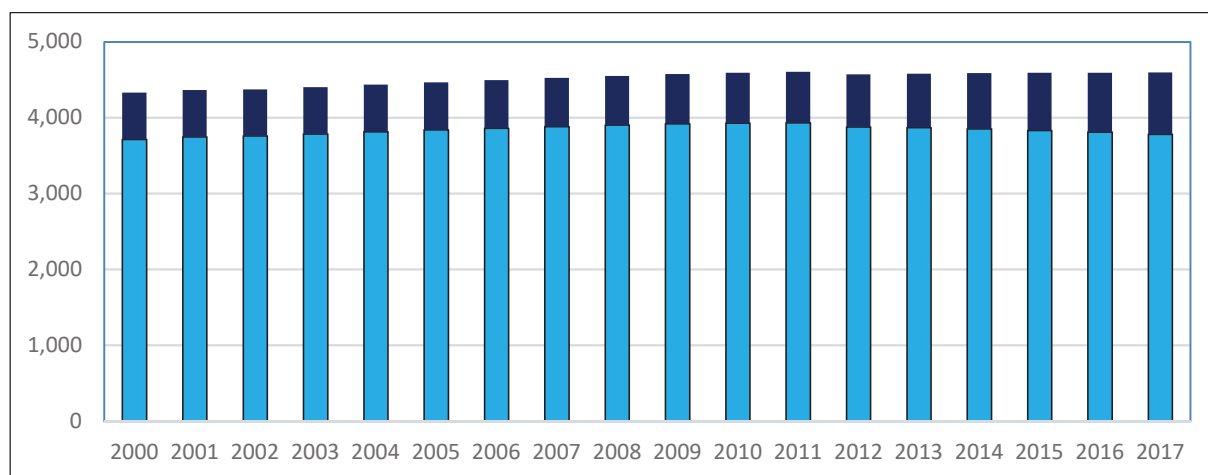


Fig. 1 Population structure in thousands of persons. Light blue describes the productive population from 15 to 64 years and dark blue the post-productive population over 64 years. Data are proposed on the annual basis from 2000 to 2017. (Source: Institute for the Financial Policy)

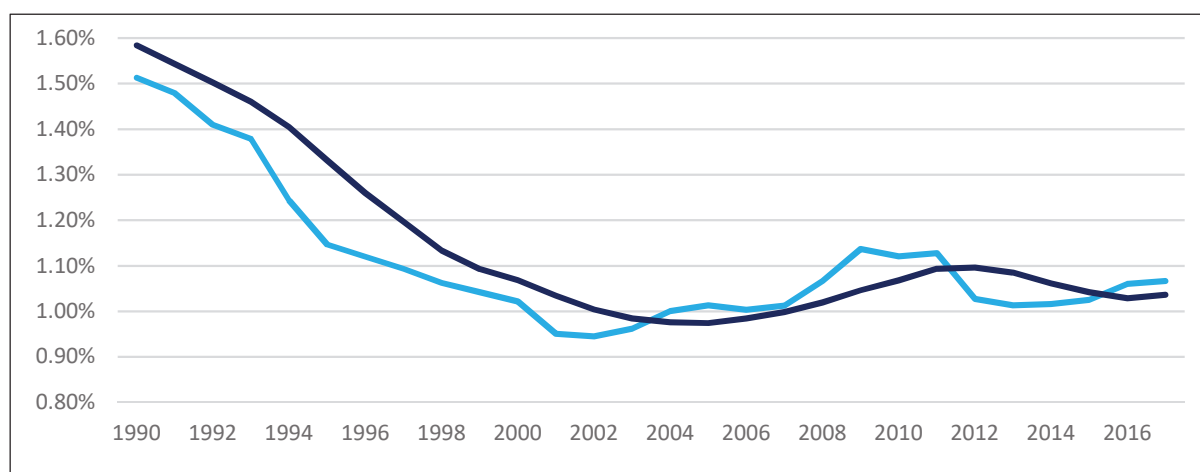


Fig. 2 Fertility rates in percentage points obtained as the ratio between the number of life births and total population. Light blue describes the annual fertility rates and dark blue the 5-year moving averages. Data are proposed on the annual basis from 1990 to 2017. (Source: Eurostat)

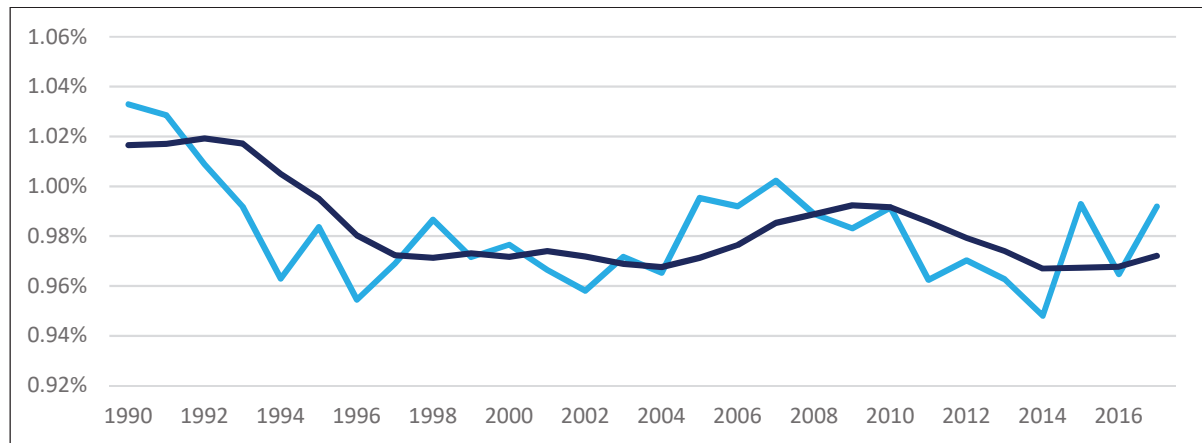


Fig. 3 Mortality rates in percentage points obtained as the ratio between the number of deaths and total population. Light blue describes the annual mortality rates and dark blue the 5-year moving averages. Data are proposed on the annual basis from 1990 to 2017. (Source: Eurostat)

Tab. 1 Simulation results under the population ageing scenario. Variables are stated in nominal terms in million euros. (Source: Author)

	L_t	K_t	Y_t	Q_t	C_t	G_t	I_t	X_t	M_t	H_t
2016	31 835	41 538	192 584	192 506	44 372	15 739	18 666	75 949	73 501	4 238
2017	33 121	43 216	200 363	200 282	46 165	16 375	19 420	79 017	76 470	4 410
2018	34 364	44 838	207 883	207 799	47 897	16 990	20 149	81 983	79 340	4 575
2019	35 574	46 416	215 200	215 113	49 583	17 588	20 858	84 868	82 133	4 736
2020	36 758	47 960	222 362	222 272	51 233	18 173	21 553	87 693	84 866	4 894
2021	37 923	49 481	229 412	229 319	52 858	18 749	22 236	90 473	87 557	5 049
2022	39 076	50 986	236 387	236 292	54 465	19 319	22 912	93 224	90 219	5 202
2023	40 223	52 481	243 323	243 224	56 063	19 886	23 584	95 959	92 866	5 355
2024	41 367	53 975	250 246	250 145	57 658	20 452	24 255	98 689	95 508	5 507
2025	42 514	55 471	257 183	257 079	59 256	21 019	24 928	101 425	98 156	5 660
2026	43 667	56 975	264 158	264 051	60 863	21 589	25 604	104 176	100 818	5 814
2027	44 829	58 492	271 189	271 079	62 483	22 163	26 285	106 948	103 501	5 968
2028	46 004	60 024	278 294	278 181	64 120	22 744	26 974	109 751	106 213	6 125
2029	47 193	61 576	285 489	285 374	65 778	23 332	27 671	112 588	108 959	6 283
2030	48 400	63 151	292 789	292 671	67 460	23 929	28 379	115 467	111 745	6 444

Tab. 2 Simulation results under the neutral demographic scenario. Variables are stated in nominal terms in million euros. *(Source: Author)*

	L_t	K_t	Y_t	Q_t	C_t	G_t	I_t	X_t	M_t	H_t
2016	31 835	41 538	192 584	192 506	44 372	15 739	18 666	75 949	73 501	4 238
2017	33 131	43 229	200 423	200 342	46 179	16 380	19 426	79 041	76 493	4 411
2018	34 388	44 868	208 023	207 939	47 930	17 001	20 163	82 038	79 393	4 578
2019	35 613	46 467	215 436	215 349	49 638	17 607	20 881	84 962	82 223	4 741
2020	36 816	48 036	222 711	222 621	51 314	18 201	21 586	87 831	84 999	4 901
2021	38 002	49 584	229 890	229 797	52 968	18 788	22 282	90 661	87 739	5 059
2022	39 179	51 120	237 008	236 912	54 608	19 370	22 972	93 469	90 456	5 216
2023	40 351	52 649	244 100	244 001	56 242	19 949	23 660	96 266	93 162	5 372
2024	41 524	54 179	251 193	251 092	57 876	20 529	24 347	99 063	95 870	5 528
2025	42 701	55 715	258 314	258 209	59 517	21 111	25 037	101 871	98 587	5 685
2026	43 886	57 261	265 483	265 376	61 169	21 697	25 732	104 698	101 324	5 843
2027	45 083	58 822	272 722	272 612	62 837	22 289	26 434	107 553	104 086	6 002
2028	46 294	60 402	280 048	279 934	64 524	22 887	27 144	110 442	106 882	6 163
2029	47 522	62 005	287 476	287 359	66 236	23 494	27 864	113 372	109 717	6 327
2030	48 769	63 632	295 021	294 901	67 974	24 111	28 595	116 347	112 597	6 493

Tab. 3 Simulation results under more conservative scenario of 1% population growth. Variables are stated in nominal terms in million euros. *(Source: Author)*

	L_t	K_t	Y_t	Q_t	C_t	G_t	I_t	X_t	M_t	H_t
2016	31 835	41 538	192 584	192 506	44 372	15 739	18 666	75 949	73 501	4 238
2017	33 150	43 253	200 535	200 454	46 204	16 389	19 437	79 085	76 536	4 413
2018	34 430	44 923	208 281	208 197	47 989	17 022	20 188	82 140	79 492	4 584
2019	35 685	46 561	215 874	215 787	49 739	17 643	20 924	85 134	82 390	4 751
2020	36 923	48 175	223 358	223 268	51 463	18 254	21 649	88 086	85 246	4 916
2021	38 149	49 775	230 775	230 681	53 172	18 860	22 368	91 010	88 077	5 079
2022	39 369	51 368	238 159	238 062	54 873	19 464	23 084	93 922	90 895	5 241
2023	40 590	52 960	245 542	245 442	56 574	20 067	23 799	96 834	93 713	5 404
2024	41 815	54 558	252 951	252 849	58 281	20 673	24 517	99 756	96 541	5 567
2025	43 048	56 168	260 413	260 307	60 000	21 283	25 241	102 699	99 388	5 731
2026	44 294	57 793	267 947	267 839	61 737	21 898	25 971	105 670	102 264	5 897
2027	45 554	59 438	275 575	275 464	63 494	22 522	26 710	108 678	105 175	6 065
2028	46 834	61 107	283 314	283 199	65 277	23 154	27 460	111 730	108 129	6 235
2029	48 134	62 803	291 179	291 061	67 089	23 797	28 223	114 832	111 131	6 408
2030	49 457	64 530	299 185	299 064	68 934	24 451	28 999	117 989	114 186	6 585

Tab. 4 Simulation results under more optimistic scenario of 2% population growth. Variables are stated in nominal terms in million euros. *(Source: Author)*

	L_t	K_t	Y_t	Q_t	C_t	G_t	I_t	X_t	M_t	H_t
2016	31 835	41 538	192 584	192 506	44 372	15 739	18 666	75 949	73 501	4 238
2017	33 168	43 277	200 647	200 565	46 230	16 398	19 448	79 129	76 578	4 416
2018	34 473	44 979	208 540	208 455	48 049	17 043	20 213	82 242	79 591	4 590
2019	35 758	46 656	216 312	216 224	49 839	17 678	20 966	85 307	82 557	4 761
2020	37 030	48 315	224 006	223 916	51 612	18 307	21 712	88 341	85 494	4 930
2021	38 295	49 966	231 662	231 568	53 376	18 933	22 454	91 360	88 416	5 098
2022	39 560	51 617	239 313	239 216	55 139	19 558	23 196	94 378	91 336	5 267
2023	40 829	53 272	246 990	246 890	56 908	20 186	23 940	97 405	94 266	5 436
2024	42 107	54 940	254 720	254 617	58 689	20 817	24 689	100 454	97 216	5 606
2025	43 397	56 624	262 527	262 421	60 488	21 455	25 446	103 533	100 195	5 778
2026	44 704	58 329	270 432	270 323	62 309	22 101	26 212	106 650	103 212	5 952
2027	46 031	60 059	278 456	278 343	64 158	22 757	26 989	109 814	106 275	6 128
2028	47 379	61 819	286 615	286 499	66 038	23 424	27 780	113 032	109 389	6 308
2029	48 753	63 612	294 926	294 807	67 953	24 103	28 586	116 310	112 561	6 491
2030	50 155	65 440	303 404	303 281	69 906	24 796	29 408	119 653	115 796	6 677

A proposal of a modified approach to classification evaluation

Návrh modifikovaného přístupu k hodnocení klasifikace

Hana Řezanková

Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky, nám. W. Churchilla 4,
130 67 Praha, Česká republika

University of Economics in Prague, Faculty of Informatics and Statistics, W. Churchill Sq. 4,
130 67 Prague, Czech Republic

hana.rezankova@vse.cz

Abstract: *In the paper, selected coefficients for evaluation of results obtained in classification tasks are discussed. Classification of objects into three or more known groups is considered. For this purpose, the following coefficients can be applied: the total success rate, the group success rate, the kappa coefficient of agreement, and kappa-n and lambda-gamma coefficients. In addition to these coefficients, we propose the SRGM (Success Rate for Groups / Mutability) measure which extends the group success rate by its penalties in the cases when the variability of agreements (observed and predicted assignments of objects into groups) and the variability of the target variable differ. The nominal variance (mutability) is applied as a variability measure. In comparison with the other five evaluation measures discussed in this paper, the SRGM measure evaluates the prediction results most comprehensively.*

Abstrakt: *V článku jsou diskutovány vybrané koeficienty, které mohou být použity pro hodnocení výsledků získaných v klasifikačních úlohách. Je uvažována klasifikace objektů do tří nebo více skupin. K tomuto účelu mohou být aplikovány následující koeficienty: celková míra úspěšnosti, skupinová míra úspěšnosti, koeficienty souhlasu kappa, kappa-n a lambda-gama. K těmto koeficientům je navíc navržena míra SRGM (Success Rate for Groups / Mutability), která rozšiřuje skupinovou míru úspěšnosti o penalizaci v případech, kdy se liší variabilita souhlasů (pozorovaného a predikovaného přiřazení objektů do skupin) a variabilita vysvětlované proměnné. Jako míra variability je aplikován nominální rozptyl (mutabilita). V porovnání s pěti hodnotícími měry diskutovanými v tomto článku míra SRGM hodnotí výsledky predikce nejkompaktněji.*

Key words: *classification task, success rate, measure of agreement.*

Klíčová slova: *klasifikační úloha, míra úspěšnosti, míra souhlasu.*

1 Introduction

One group of multivariate statistical methods for classification is created by methods for data analysis with the known assignment of objects into classes (groups). The assignment is expressed by a target nominal variable with codes of groups. The aim of the classification task is to estimate a model or rules for the assignment of new objects into the defined groups and to estimate the values of the target variable for new objects.

Different methods can be used in the classification task described above, e.g. discrimination analysis, logistic regression (binary or multinomial), and classification trees. Each approach is more suitable for a certain type of explanatory variables (quantitative or qualitative), but generally, the quantitative variable can be categorized and the qualitative variable can be transformed into a set of binary variables. Thus, several methods can be used for the analysis of the same data.

The analyst should compare results obtained by several methods and choose the most suitable model or rules for the prediction of unknown values of the target variable. Different approaches to such a comparison have been proposed and some of them are implemented in software systems. The simplest way is to use the percentage of successfully assigned objects (the success rate for prediction) according to a certain method.

If only two groups exist, there are many measures for evaluation and comparison of classification methods (see, e.g. Sokolova and Lapalme, 2009; Řezanková and Húsek, 2015). In addition, the area under the receiver operating characteristic (ROC) curve (AUC) is also often used as a criterion for choosing the best method in the case of two groups (see, e.g. Fawcett, 2006).

For the case of three or more groups, there are two basic approaches to the evaluation. The first one (two-step) consists in the application of the measures for two groups to the individual groups considering one selected group and all other groups joined into the second group. A final measure is based on these individual measures (Sokolova and Lapalme, 2009). The success rates can be computed in this way. The second approach is the computation of the evaluation coefficients in one step. The coefficients of agreement can be mentioned, e.g. Cohen's kappa (see Cohen, 1960).

In this paper, we discuss the selected coefficients for comparison of prediction results in the case of three or more groups. These coefficients are the total success rate, the group success rate, the kappa coefficient of agreement, and kappa- n and lambda-gamma coefficients. In addition to these coefficients, we propose a new coefficient with some better properties than the other five measures of interest have. This proposed *SRGM* (Success Rate for Groups / Mutability) measure extends the group success rate by its penalties in the cases when the variability of agreements (observed and predicted assignments of objects into groups) and the variability of the target variable differ. The nominal variance (mutability) is applied as a variability measure.

The total success rate and the group success rate can be computed using both the two-step and one-step approaches mentioned above. So, in this paper, all

coefficients are presented using the notation based on one classification table for all individual groups.

2 Approaches to classification evaluation

The classification table is often prepared for the purpose of prediction evaluation. Let us have such a table with frequencies (absolute N_{ij} or relative p_{ij}) of cases in which a certain object of the i th group was predicted into the j th group, where $i = 1, 2, \dots, n$ (the same is valid for the j index), n is the number of rows and also the number of columns, i.e. the number of groups. The total number of objects is denoted as N . The agreement occurs if the object of the i th group is predicted into the i th group. The counts of such cases are expressed by the frequencies on the main diagonal (absolute N_{ii} or relative p_{ii}), N_{i+}, N_{+i} are the corresponding marginal absolute totals, similarly p_{i+}, p_{+i} are the corresponding marginal relative totals.

The total *success rate* (SR) for predictions expresses the ratio of the number of the correctly predicted objects and the total number of the objects, i.e.

$$SR = \frac{\sum_{i=1}^n N_{ii}}{N} . \quad (1)$$

It means that the value of SR is the sum of the relative frequencies on the main diagonal and it takes on the values from the interval $\langle 0; 1 \rangle$. It is usually presented as a percentage. Sometimes (e.g., in some statistical software products), it is called *overall percentage* or *classification accuracy*. This measure can be also expressed by a formula based on true positive and false positive counts (used for the measure *precision*) applied to the individual groups separately. Sokolova and Lapalme (2009) call this measure as *precision _{μ}* .

However, the total percentage does not consider the differences among the percentages obtained for individual groups, when for one group the success rate can be close to 100% and for some other group close to 0%. For this reason, the use of other measures or coefficients is a better solution. In each group, the number of correctly predicted objects should be as high as possible. The success rates for the individual groups can be expressed in summary using the average. This *group success rate* or *success rate for groups* (SRG) is calculated as

$$SRG = \frac{1}{n} \sum_{i=1}^n \frac{N_{ii}}{N_{i+}} . \quad (2)$$

This measure can be also expressed as an average of the precisions calculated for the individual groups separately. Sokolova and Lapalme (2009) call this

measure as *precision_M*. If the sizes of all groups are the same, then $N_{i+} = N/n$ for $i = 1, 2, \dots, n$, and the values of *SR* and *SRG* are equal.

The observed sizes of individual groups can be expressed relatively by row marginal relative totals p_{i+} . The counts of agreements for individual groups can be expressed relatively within the total sum of agreements. For the good prediction, the variability of agreements should be the same (or similar) as the variability of the row marginal totals. The nominal variance (i.e. the mutability) proposed by Gini (1912), reviewed in (Light and Margolin, 1971) can be applied as a variability measure.

The mutability of the row marginal totals is defined as

$$M_{RMT} = \sum_{i=1}^n \left(\frac{N_{i+}}{N} \left(1 - \frac{N_{i+}}{N} \right) \right) = \sum_{i=1}^n (p_{i+} (1 - p_{i+})). \quad (3)$$

Generally, the mutability takes on the value zero for a constant. However, we consider prediction into three or more groups, it means that the mutability of the row marginal totals cannot be zero. It reaches its maximum, i.e. $(n-1)/n$, when all row marginal totals are equal. By dividing the maximal value, we obtained the normalized mutability from the interval $(0; 1)$.

Let us define – for the case where at least one $N_{ii} \neq 0$ – the mutability of agreements as

$$M_A = \sum_{i=1}^n \left(\frac{N_{ii}}{\sum_{j=1}^n N_{jj}} \left(1 - \frac{N_{ii}}{\sum_{j=1}^n N_{jj}} \right) \right). \quad (4)$$

It takes on values from the interval $\langle 0; (n-1)/n \rangle$. By dividing by its maximal possible value, we obtained the normalized mutability from the interval $\langle 0; 1 \rangle$. Zero indicates that the objects are correctly assigned only into one group, which means it makes no sense for the method to be used for prediction.

In this paper, we propose a new measure for the classification evaluation based on the *SRG* measures and the mutability ratio. We will denote it as the *SRGM* measure. If the mutability of agreements is the same as the mutability of the row marginal totals then the mutability ratio is one and *SRG* and *SRGM* are equal. All other cases are “penalized” by the mutability ratio which has to take on values less than one (the lower value is divided by the greater value). If objects are assigned only into one group, then the mutability is zero and *SRGM* is also zero (some objects are assigned correctly but the applied method cannot be used for prediction).

We can express the calculation of the *SRGM* measure in the following way:

$$SRGM = 0 \quad \text{for } \sum_{i=1}^n N_{ii} = 0, \quad (5)$$

$$SRGM = \left(\frac{1}{n} \sum_{i=1}^n \frac{N_{ii}}{N_{i+}} \right) \frac{M_A}{M_{RMT}} \quad \text{for } \sum_{i=1}^n N_{ii} \neq 0 \text{ and } M_A \leq M_{RMT}, \quad (6)$$

$$SRGM = \left(\frac{1}{n} \sum_{i=1}^n \frac{N_{ii}}{N_{i+}} \right) \frac{M_{RMT}}{M_A} \quad \text{for } \sum_{i=1}^n N_{ii} \neq 0 \text{ and } M_A > M_{RMT}. \quad (7)$$

This measure takes on values from the interval from 0 (no agreement or correct assignments only to one group) to 1 (complete agreement for all groups).

Řezanková and Húsek (2015) proposed the *RH coefficient* for prediction evaluation. The authors consider only a special case with the same sizes of groups when the values of *SR* and *SRG* are equal. The mutability of agreements is expressed in the normalized form with the values from the interval $\langle 0; 1 \rangle$. Then this measure is expressed by the formula

$$RH = \frac{\sum_{i=1}^n N_{ii}}{N} \frac{n}{n-1} \sum_{i=1}^n \left(\frac{N_{ii}}{\sum_{j=1}^n N_{jj}} \left(1 - \frac{N_{ii}}{\sum_{j=1}^n N_{jj}} \right) \right), \quad (8)$$

where the ratio $n/(n-1)$ normalizes the mutability. It is a special case of the general *SRGM* measure because, in the case of the same sizes of groups, the normalized mutability of the row marginal totals is one (in Eqs. (6) and (7), the ratio $n/(n-1)$ is omitted because there would be the same value in the numerator and denominator).

The well-known coefficient of agreement, which is used for the prediction evaluation, is the *kappa coefficient of agreement*. It was proposed by Cohen (1960) for the purpose when two psychologists independently categorize psychological test protocols into a certain number of categories. However, its use is wider. Its application for the prediction evaluation is often described in the literature (e.g. Powers, 2012). This coefficient is implemented in some software systems, usually as a part of the contingency table analysis for the square tables.

The basic features of kappa are the following. If all predicted groups correspond to observed groups, the value of this coefficient is one. If the counts in the classification table are the same as the frequencies expected under the condition of independence (according to the Pearson chi-square test considering marginal totals for predictions the same as they were really predicted), the value of this coefficient is zero. The values of kappa can be also negative. The lower limit depends on the marginal distributions, e.g. for two groups with the same group sizes it is -1 .

In this paper, the sample coefficients based on the observed frequencies in the classification table are considered. Using the relative or absolute frequencies, kappa (κ) is expressed as

$$\kappa = \frac{\sum_{i=1}^n p_{ii} - \sum_{i=1}^n p_{i+} p_{+i}}{1 - \sum_{i=1}^n p_{i+} p_{+i}} = \frac{\sum_{i=1}^n N_{ii} - \frac{1}{N} \sum_{i=1}^n N_{i+} N_{+i}}{N - \frac{1}{N} \sum_{i=1}^n N_{i+} N_{+i}}. \quad (9)$$

If all frequencies in the contingency table except for the diagonal ones are zero, then the sum of all diagonal relative frequencies $\sum p_{ii}$ is one and also kappa is one. If the diagonal frequencies correspond to the frequencies expected under statistical independence, then $\sum p_{ii} = \sum p_{i+} p_{+i}$ and kappa is zero. Many modifications of the kappa coefficient were proposed; for their survey, see e.g. (Warens, 2013).

One of the several modifications of the kappa coefficient is *kappa-n* (κ_n), based on free marginal proportions when the random assignment of objects into groups can be characterized by value $1/n$ for all relative marginal totals. *Kappa-n* is derived as

$$\kappa_n = \frac{\sum_{i=1}^n p_{ii} - n \frac{1}{n} \cdot \frac{1}{n}}{1 - n \frac{1}{n} \cdot \frac{1}{n}} = \frac{\sum_{i=1}^n p_{ii} - \frac{1}{n}}{1 - \frac{1}{n}} = \frac{\sum_{i=1}^n N_{ii} - N \cdot \frac{1}{n}}{N - N \cdot \frac{1}{n}}. \quad (10)$$

This coefficient was firstly proposed by Bennet et al. (1954) as the *S coefficient*. Several other authors later proposed this coefficient under different names. The well-known name *kappa-n* was introduced by Brennan and Prediger (1981). For the history of this coefficient, see e.g. (Zwick, 1988; Hsu and Field, 2003).

The lower limit of *kappa-n* depends on the number of groups. It corresponds to the situation when the sum of frequencies on the main diagonal is zero. In the case of two groups, it is -1 . However, in classification evaluation, row marginal proportions p_{i+} are known. Considering the known row marginal totals and the free column marginal total, we can write *kappa* as

$$\kappa_n = \frac{\sum_{i=1}^n p_{ii} - \frac{1}{n}}{1 - \frac{1}{n}} = \frac{\sum_{i=1}^n p_{ii} - \frac{1}{n} \sum_{i=1}^n p_{i+}}{1 - \frac{1}{n} \sum_{i=1}^n p_{i+}} \quad (11)$$

Both kappa and *kappa-n* have their negatives. Kappa considers the marginal totals for prediction the same as they were really predicted, it does not consider the free proportions. On the other hand, *kappa-n* is not dependent on the observed marginal totals.

Before Cohen's kappa, the *lambda-gamma* (λ_γ) coefficient was proposed by Goodman and Kruskal (1954). It is expressed as

$$\lambda_\gamma = \frac{\sum_{i=1}^n p_{ii} - \frac{1}{2} \max(p_{i+} + p_{+i})}{1 - \frac{1}{2} \max(p_{i+} + p_{+i})} = \frac{\sum_{i=1}^n N_{ii} - \frac{1}{2} \max(N_{i+} + N_{+i})}{N - \frac{1}{2} \max(N_{i+} + N_{+i})}. \quad (12)$$

For obtaining the values greater than zero, the sum of the frequencies on the main diagonal must be greater than half of the sum of the two largest marginal totals.

The value -1 characterizes a situation where non-zero frequencies are only either in the first row and in the first column or in the last row and in the last column of the classification table, and the frequency in the corner cell is equal or greater than other frequencies. However, the value -1 can be a result of some corner frequencies lower than other frequencies.

Table 1 shows a summary of the properties of the six coefficients described above. The coefficients achieve their maximum in the case of complete agreement. They achieve their minimum if the sum of agreement frequencies for all groups is zero.

3 Discussion of prediction evaluation

Let us consider a classification task with three groups with frequencies 12, 9 and 15. For SR equal to one, we obtain the classification table denoted as Table 2 (the frequencies of agreements are in bold). For this situation, the values of all coefficients mentioned in this paper are equal to one.

If the predicted marginal totals are free (the same group sizes are considered), as it is shown in Table 3, $SR = 0.333$, $SRG = 0.333$, $SRGM = 0.333$, $\kappa = 0$, $\kappa_n = 0$ and $\lambda_\gamma = -0.067$.

If a certain method assigns (predicts) all objects only into one group, then the success rate can be promising but kappa cannot be computed. The user can obtain the zero value by some software product because the observed and expected frequencies are the same. If all 36 objects considered in Tables 2 and 3 are predicted only into group 3, then $SR = 0.417$, $SRG = 0.333$, $SRGM = 0$, $\kappa = 0$, $\kappa_n = 0.125$ and $\lambda_\gamma = -1$.

Tab. 1 Summary of properties of the selected coefficients for classification evaluation (*Source: the author*)

Coefficient	Maximum	Minimum	Disadvantages	Advantages
<i>SR</i>	1	0	<i>SR</i> is commonly available in statistical software systems.	The success rates in individual groups are not taken into account.
<i>SRG</i>	1	0	The success rates in individual groups are taken into account.	The variability of the agreements and the variability of the row marginal totals are not taken into account.
<i>SRGM</i>	1	0	The variability of the agreements and the variability of the row marginal totals are taken into account.	<i>SRGM</i> is not available in the software.
kappa	1	It depends on the marginal distributions (–1 for two groups with the same group sizes).	Kappa is commonly available in statistical software systems.	Kappa considers the marginal totals for predictions the same as they were really predicted.
kappa-<i>n</i>	1	It depends on the number of groups (–1 for two groups).	Kappa- <i>n</i> is based on the free marginal proportions.	Kappa- <i>n</i> does not take into account the observed group sizes.
lambda-gamma	1	It depends on the marginal distributions.	The assumptions for value –1 are defined.	Lambda-gamma is not available in commercial software systems.

For an illustration of coefficient value changes depending on the changes of frequencies in the classification table, let us consider marginal frequencies both for rows and for columns the same as they are in Table 2, i.e., 12, 9 and 15. In Table 4, there are frequencies in individual cells (*i, j*) of the classification table for nine different situations. These situations are considered from the completely correct prediction to the completely wrong prediction. The values of coefficients *SR*, *SRG*, *SRGM*, kappa, kappa-*n*, and lambda-gamma are included.

Tab. 2 A classification table for the 100% success rate (*Source: the author*)

	Prediction into group 1	Prediction into group 2	Prediction into group 3	Marginal total
Observed group 1	12	0	0	12
Observed group 2	0	9	0	9
Observed group 3	0	0	15	15
Marginal total	12	9	15	36

Tab. 3 A classification table for the uniform assignment of objects into groups (*Source: the author*)

	Prediction into group 1	Prediction into group 2	Prediction into group 3	Marginal total
Observed group 1	4	4	4	12
Observed group 2	3	3	3	9
Observed group 3	5	5	5	15
Marginal total	12	12	12	36

The SR , SRG , $SRGM$ coefficients, take on values from the interval 0 to 1. Kappa, kappa- n , and lambda-gamma have their maximum also in 1 but the lower limits are different. For these coefficients, the definitions of lower limits may be interesting from a mathematical point of view, but from a practical point of view (for prediction evaluation) they have no meaning.

The SR coefficient equals 0.5 (the middle of the value interval) for Situation 4 in Table 4. The values of SRG and $SRGM$ are lower for this situation. Kappa- n equals 0 for Situation 6 and lambda-gamma equals 0 for Situation 5. Kappa cannot have a value equal 0 for the considered marginal totals because some expected values under statistical independence are not integers.

In the illustration example mentioned above, the values of all coefficients are ordered from 1 to the last one. However, in some other cases, the orders are not the same, see below. For practical use, the SRG and $SRGM$ coefficients should be greater than 0.5. There are only the first three situations in the example. In the case of the kappa coefficient, the situations with kappa greater than 0.4 are recommended. The first three situations also satisfy this condition.

In Table 5, there are frequencies in individual cells (i, j) of the classification table for different nine situations for the 36 objects. The values of SR and SRG coefficients are greater than 0.5. We can compare them with the values of the other coefficients studied in this paper.

Tab. 4 Frequencies for the individual cells (described by the row and column orders) of the classification table for 9 situations with marginal totals 12, 9 and 15 (both for rows and for columns) and values of the selected measures (*Source: the author*)

Cell	Sit. 1	Sit. 2	Sit. 3	Sit. 4	Sit. 5	Sit. 6	Sit. 7	Sit. 8	Sit. 9
(1,1)	12	10	8	6	5	4	2	0	0
(1,2)	0	1	2	3	3	4	5	6	6
(1,3)	0	1	2	3	4	4	5	6	6
(2,1)	0	1	2	3	4	4	4	3	0
(2,2)	9	7	5	3	2	1	0	0	0
(2,3)	0	1	2	3	3	4	5	6	9
(3,1)	0	1	2	3	3	4	6	9	12
(3,2)	0	1	2	3	4	4	4	3	3
(3,3)	15	13	11	9	8	7	5	3	0
Coefficient									
SR	1.000	0.833	0.667	0.500	0.417	0.333	0.194	0.083	0.000
SRG	1.000	0.826	0.652	0.478	0.391	0.304	0.167	0.067	0.000
SRGM	1.000	0.818	0.635	0.447	0.351	0.252	0.104	0.000	0.000
kappa	1.000	0.745	0.489	0.234	0.106	-0.021	-0.234	-0.404	-0.532
kappa-n	1.000	0.750	0.500	0.250	0.125	0.000	-0.208	-0.375	-0.500
lambda-gamma	1.000	0.714	0.429	0.143	0.000	-0.143	-0.381	-0.571	-0.714

In Situations 1, 2 and 3, the number of correctly predicted objects is 27, in Situation 4 it is 21 and for the rest situations, it is 19. The values of the *SR* coefficient are 0.75, 0.783 and 0.528 for these three different numbers. The values of the *SRG* coefficient are equal for Situations 3 and 4 because all objects from two groups and none object from the third group were correctly predicted. The different values for Situations 1 and 2 were obtained only for kappa and lambda-gamma because these coefficients take into account the column marginal totals, compared to the other studied coefficients.

The order of the values of the *SRGM* coefficient is similar to the order of the *SRG* values. Situation 5 was evaluated as the worst by both the *SRG* and *SRGM* coefficients. However, in some cases (Situations 3, 4 and 5), the *SRGM* values are lower by more than 0.1 due to the greater difference between the mutability of

agreements and the mutability of the row marginal totals, in comparison with other situations.

Tab. 5 Frequencies for the individual cells (described by the row and column orders) of the classification table for 9 situations with different marginal totals and values of the selected measures (*Source: the author*)

Cell	Sit. 1	Sit. 2	Sit. 3	Sit. 4	Sit. 5	Sit. 6	Sit. 7	Sit. 8	Sit. 9
(1,1)	12	12	12	12	12	10	8	6	6
(1,2)	0	0	0	0	0	1	3	3	0
(1,3)	0	0	0	0	0	1	1	3	6
(2,1)	0	0	4	0	0	0	0	0	0
(2,2)	9	9	0	9	1	2	3	5	8
(2,3)	0	0	5	0	8	7	6	4	1
(3,1)	0	9	0	15	0	2	4	6	6
(3,2)	9	0	0	0	9	6	3	1	4
(3,3)	6	6	15	0	6	7	8	8	5
Coefficient									
SR	0.750	0.750	0.750	0.583	0.528	0.528	0.528	0.528	0.528
SRG	0.800	0.800	0.667	0.667	0.504	0.507	0.511	0.530	0.574
SRGM	0.787	0.787	0.504	0.500	0.385	0.448	0.486	0.529	0.573
kappa	0.640	0.629	0.597	0.394	0.282	0.277	0.277	0.277	0.292
kappa-n	0.625	0.625	0.625	0.375	0.292	0.292	0.292	0.292	0.292
lambda-gamma	0.538	0.500	0.514	0.000	0.209	0.190	0.190	0.190	0.244

As concerns the coefficients of agreements, kappa- n takes on the values 0.625, 0.375 and 0.292, depending only on the number of correctly predicted objects, as well as the SR coefficient. The values of kappa and kappa- n are the same in the case when the column marginal totals are equal (Situation 9). The lowest value was found by lambda-gamma for Situation 4. For classification evaluation, the coefficients of agreement are not as suitable as the SRG and $SRGM$ measure; they evaluate the obtained results in a very different way.

4 Conclusion

In the paper, selected coefficients for result evaluation in classification tasks with three or more groups of objects were discussed – five already proposed

coefficients and one newly proposed coefficient. Although some coefficients of this type were proposed more than 50 years ago (the first of them more than 100 years ago), they are still discussed in the literature (Powers, 2012; Warrens, 2013). In recent years, the interest in these coefficients again increases with the increasing need to evaluate results obtained in classification tasks.

We can distinguish two groups of these coefficients: success rates and coefficients of agreement. The success rates discussed in this paper (the total success rate, the group success rates, and the new *SRGM* measure) take on the values from the interval 0 to 1. Coefficients of agreement (κ , κ_n , and λ - γ) have the upper limit also equal to 1 but they can also take on negative values. The lower limit depends on the marginal distributions and the value of zero is achieved in different situations.

The total success rate is well known and commonly available in the software. However, it does not take into account the success rates in individual groups. Hence, the average of the success rates for the individual groups (the group success rate) characterizes the predictions better. In the paper, a new measure is proposed which extends the group success rate by its penalties in the cases when the variability of agreements (observed and predicted assignments of objects into groups) and the variability of the target variable differ. The nominal variance (mutability) is applied as a variability measure. Thus, this *SRGM* measure evaluates the prediction results most comprehensively. Moreover, the correct assignments of objects only to one group are evaluated by the value zero.

Coefficients of agreement are represented by coefficients κ , κ_n and λ - γ . The κ coefficient of agreement is well-known and commonly available in the software. However, it considers the marginal totals for predictions the same as they were really predicted. κ_n is based on the free marginal proportions but it does not take into account the observed group sizes. λ - γ evaluates the correct assignments of objects only to one group by low values.

As shown in this paper, the application of coefficients of agreement is not suitable for classification evaluation in some cases. These coefficients were proposed for other purposes and they do not take into account the specifics of the prediction. For that reason, using success rates is a better way. The total success rate does not distinguish some different situations. Thus, we can recommend either the group success rate or the newly proposed *SRGM* measure for prediction evaluation.

This paper was supported by Grant IGA F4/44/2018 of the University of Economics in Prague

5 References

- Cohen, J. (1960). *A coefficient of agreement for nominal scales*. Educational and Psychological Management, 20(1), 37-46.
- Bennet, E. M., Alpert, R., Goldstein, A. C. (1954). *Communications through limited-response questioning*. Public Opinion Quarterly, 18(3), 303-308.
- Brennan, R.L., Prediger, D.J. (1981). *Coefficient kappa: Some uses, misuses, and alternatives*. Educational and Psychological Management, 41, 687-699.
- Fawcett, T. (2006). *An introduction to ROC analysis*. Pattern Recognition Letters, 27(8), 861-874.
- Gini, C. W. (1912). *Variability and mutability. Contribution to the study of statistical distributions and relations*. Studi Economico-Giuridici della R. Università de Cagliari, 1912.
- Goodman, L. A., Kruskal, W. H. (1954). *Measures of association for cross classifications*. Journal of the American Statistical Association, 49, 1732-1769.
- Hsu, L. M., Field, R. (2003). *Interrater agreement measures: Comments on Kappan, Cohen's kappa, Scott's π , and Aickin's α* . Understanding Statistics, 2(3), 205-219.
- Light, R. J., Margolin, B. H. (1971). *An analysis of variance for categorical data*. Journal of the American Statistical Association, 66, 534-544.
- Powers, D. M. W. (2012). *The problem with kappa*. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. Avignon: Association for Computational Linguistics, 2012, s. 345-355.
- Řezanková, H., Húsek, D. (2015). *Measures for Classification Results Evaluation. Technical Report No. 1220*. Praha: Institute of Computer Science, the Czech Academy of Science, 2015. Dostupné na: <http://hdl.handle.net/11104/0249608>.
- Sokolova, M., Lapalme, G. (2009). *A systematic analysis of performance measures for classification tasks*. Information Processing and Management, 45, 427-437.
- Warrens, M. J. (2013). *A comparison of Cohen's kappa and agreement coefficients by Corrado Gini*. International Journal of Research and Reviews in Applied Sciences, 16(3), 345-351.
- Zwick, R. (1988). *Another look at interrater agreement*. Psychological Bulletin, 103, 374-378.

Trend definition as a support tool for managing a portfolio of investment strategies

Definícia trendu ako podporný nástroj pre riadenie portfólia investičných stratégií

Tomáš Virdzek

Univerzita Mateja Bela v Banskej Bystrici, Ekonomická fakulta, Výskumné a inovačné centrum, Tajovského 10, 975 90 Banská Bystrica, Slovenská republika

Matej Bel University in Banská Bystrica, Faculty of Economics, Research and Innovation Centre, Tajovského 10, 975 90 Banská Bystrica, Slovak Republic

tomas.virdzek@umb.sk

Peter Kubaška

Univerzita Mateja Bela v Banskej Bystrici, Ekonomická fakulta, Katedra kvantitatívnych metód a informačných systémov, Tajovského 10, 975 90 Banská Bystrica, Slovenská republika

Matej Bel University in Banská Bystrica, Faculty of Economics, Department of Quantitative Methods and Information Systems, Tajovského 10, 975 90 Banská Bystrica, Slovak Republic

peter.kubaska@umb.sk

Abstract: *Investment strategies may have a positive impact from the investor's perspective on multisource financing of business activities and directly or indirectly affect the financial health of a company. However, the dynamic development of the financial market can cause that any investment strategy may show also very unpredictable behavior in terms of performance parameters. Many authors use trend definition in their trend following trading systems. In our paper, we understand respective trend definition as a decision-making algorithm tool for a company which is actively managing a portfolio of investment strategies (trading systems) to maximize profit and minimize drawdowns in its equity curve. The goal of this paper is to test and evaluate the ability of respective trend definition to bring higher performance parameters versus a passive diversification of capital.*

Abstrakt: *Investičné stratégie môžu mať z investorovho pohľadu pozitívny vplyv na viaczdrojové financovanie podnikateľských aktivít a môžu tak priamo alebo nepriamo ovplyvniť aj finančné zdravie spoločnosti. Dynamický vývoj finančného trhu môže spôsobiť, že akákoľvek investičná stratégia vie z hľadiska výkonnostných parametrov vykazovať aj nepredvídateľné správanie. Veľa autorov používa definíciu trendu vo svojich trendových obchodných systémoch. V článku však chápeme príslušnú definíciu trendu ako rozhodovací algoritmus pre spoločnosť, ktorá aktívne riadi portfólio viacerých investičných stratégií (obchodných systémov) s cieľom maximalizovať zisk a minimalizovať prepady v krivke jeho výkonnosti. Cieľom príspevku je otestovať a vyhodnotiť schopnosť príslušnej definície trendu dosiahnuť aktívnou diverzifikáciou vyššie parametre výkonnosti oproti pasívnej diverzifikácii kapitálu.*

Key words: *trend definition, active diversification, passive diversification, portfolio management.*

Kľúčové slová: *definícia trendu, aktívna diverzifikácia, pasívna diverzifikácia, riadenie portfólia.*

1 Introduction

The financial market brings investment opportunities not only to individuals but also to businesses. While an individual investor invests his savings, the company can invest its free cash. Corporate cash holding is one of the most essential issues and strategies of corporate financial management, which not only relates to corporate operation and development, but also relates to the corporate governance. Cash is the most liquid asset, but on the other hand, it is also one of the least profitable assets. Especially when we talk about today's historically lowest interest rates. The benefit of holding cash will decrease in case the company has „redundant“ cash on balance. Whether is this situation short or long term the fact is that each company can invest it on financial market. Opportunities on the capital market create space for creation investment strategies. Tomasini and Jeakle (2011) characterize the investment strategy as a detailed set of rules that automatically and without any human intervention define buy or sell signals. And today, there is not only the possibility of investing cash in conventional financial instruments (stocks, bonds, ETF, mutual funds etc.), but thanks to advances in the financial industry also in new financial services. These solutions allow investors to invest in complex investment strategies (trading systems). PAMM (percent allocation money management) systems and other similar financial services bring them a lot of new financial opportunities. The biggest advantage of these solutions is that, unlike hedge funds, they are already available at very low investments and allow investors to spread their capital into an unlimited investment strategy portfolio. The main disadvantage of these solutions may be the fact that some strategies, in terms of their risk settings, may bring the investor a total loss of invested money. Therefore, diversification plays a key role in this case. Some investors spread their capital between various managers who prefer different investment strategies and financial instruments. This approach can be understood as a passive one. On the other hand, there can be also an active approach to diversification, which in the biggest simplicity means that in some inappropriate circumstances, some investment strategy is excluded from the final portfolio. This approach requires, of course, some effective system or tool. Each investor would like to have portfolio of investment strategies that are forever only in uptrend. The dynamic development of the financial market causes that this is impossible in reality. However, from our point of view, the idea of the need for an ever-growing trend is interesting. Therefore, our article deals with three different ways of defining the trend, that we are using as a tool for actively managed investment strategies portfolio.

2 Diversification

The principle of diversification consists in redistributing the investor's capital into the various financial instruments. The basic capital distribution approach is naive diversification. The principle of this kind of diversification means, that the investor evenly distributes his capital into the financial instruments. This approach is also understood as passive, because naive diversification concept does not require a deeper analysis of assets included in the portfolio. Conversely, the concept of active diversification requires a deeper analysis of assets and selects only potentially better assets into the portfolio. The goal of an active approach is to overcome the passive approach. The main question is whether these active approaches are capable to overcome the passive diversification of capital.

DeMiguel et al. (2009) compared 14 different diversification models with 1/N approach. All these models were constructed with regular portfolio rebalancing. Author's results point to the fact that none of 14 diversification models was consistently better than naive diversification in terms of Sharpe ratio. This fact was due to author's opinion caused by estimation error in parameters. Their results are confirmed by Allena et al. (2016). By analyzing portfolios created from European indices, these authors did not confirm the higher performance of actively diversified portfolios over the naive diversification performance. Authors Hwang et al. (2018) also tested the performance of naive diversification compared to the optimal portfolio. Like DeMiguel et al. (2009), they demonstrated that the naive diversification overcame by its performance the optimal portfolio. They argue that the portfolio created on the base of naive diversification is characterized by increased exposure in the left tail. It means that such a portfolio tends to have not only a negative coefficient of skewness, but also an increased positive coefficient of kurtosis compared to normal distribution. Statistical data has rather concaved properties, which again confirms the fact, that naive diversification tends to overcome the optimal portfolio. The last factor that results in an increase in performance is associated with the number of assets held in the portfolio. A higher number of assets in the portfolio results in an increase the characteristics in the naive diversification. According to Tu and Zhou (2011) Markowitz's modern theory of portfolio and all its extensions lag behind the principle, what is in line with the results of previous authors. The authors understood the concept of total portfolio diversification as a combination of principle with four other models. Ultimately, the authors achieved better results of the risk-return profile compared to the principle.

3 Methodology

3.1 Investment strategies portfolios

For the purposes of our research, we have used our own generator that artificially simulates the development of investment strategies. This generator randomly sorts the daily returns of Dow Jones Industrial Average in 5294 business days. These data were obtained from the web site www.finance.yahoo.com. Based on this generator we created 10 investment strategies. These strategies are divided into two groups. First group (IS1, IS2, IS3, IS4, IS5) represents only growth investment strategies, which bring to investor high final return. Second group (IS6, IS7, IS8, IS9, IS10) represents those investment strategies, which reached significant loss in a term of Total return. IS6-IS10 also represents a risky part of investment. Figure 1 shows performance of selected investment strategies (IS1-IS10) in 5294 business days.

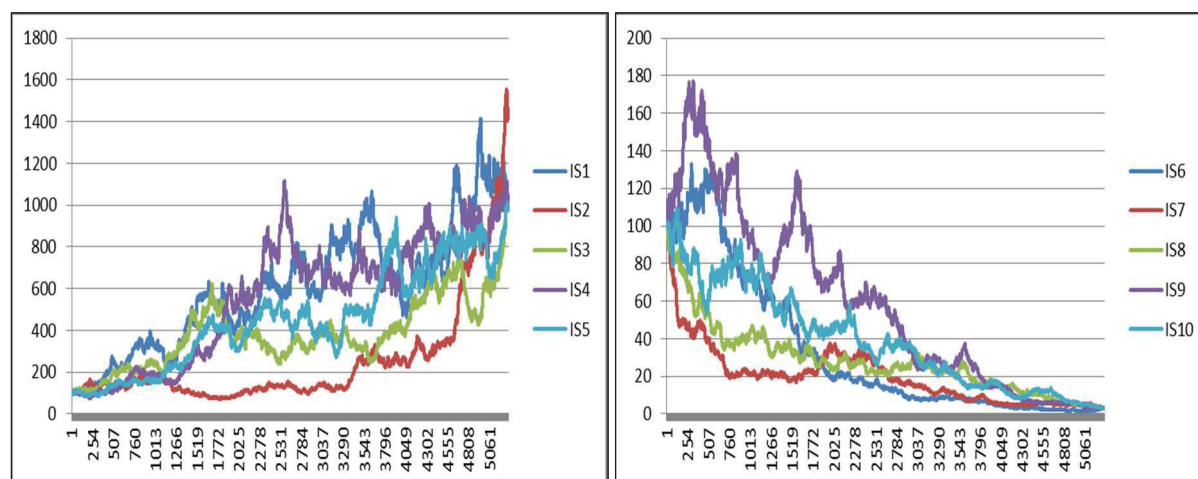


Fig. 1 Investment strategies (growth/loss) (Source: the authors)

From these investment strategies we created final portfolios. Each portfolio consists always of 5 investment strategies. From the point of view of combinations, we could create exactly 252 portfolios. Our first boundary portfolio consists only of 5 growth and the second only of 5 loss investment strategies. We created another 50 final portfolios using random selection without repetition. It means that each created portfolio consists of 5 different investment strategies. Each strategy has 20% weight in the portfolio (systemic benchmark = passive diversification).

3.2 Performance evaluation indicators

We evaluate performance of simulated final portfolios by Maximum drawdown, Total return and Recovery factor. We focus on return indicator (Total

return), on risk indicator (Maximum drawdown) and on whole performance indicator (Recovery factor).

Maximum drawdown

Maximum drawdown is the largest % difference between the local maximum and minimum over the observed period. It expresses the largest percentage decline in the value of a fund, financial instrument or investment strategy over time. Pospisil and Vecer (2008) understand the maximum drawdown as a contingent claim that is valuable and so the investor (or the manager) can hedge against it by using hedging tools. If P_t is the close price of an investment strategy at time t , then maximum price in time t is defined as:

$$P_t = \max\{P_u, u \in [0, t]\}. \quad (1)$$

Then maximum drawdown (MDD_t) at the close price of an investment strategy is defined as:

$$MDD_t = \max\{M_u - P_u, u \in [0, t]\}. \quad (2)$$

Total return

Investment strategy return is an important quantitative indicator of overall strategy performance. Fernandez (2018) points to the way, in which was easy to outperform the US S&P 500 index performance. The author states that if a rational investor invests in the long-term, he cares about the state of his investment at the end of the investment horizon (e.g. retirement). Furthermore, he defines a rational investor as one that diversifies the portfolio in order to prevent concentration of risk in individual investments. Rational investors do not care about using the "best model" on the market. The author says that investors are not even interested in maximizing Sharp's ratio or minimizing the volatility of their portfolios. The only thing the investor is interested in is the final return when the investment horizon is met. He demonstrates his thoughts on 137 students. He shows to these students 2 types of investment for 10 years. The first investment reaches an average annual return of 16% with a Sharpe ratio of 0.4 and the second investment reaches an average annual return of 13%, but the Sharpe ratio is higher, reaching 1.3. All the students surveyed preferred the first investment.

In this paper we use discrete returns.

Recovery factor

Recovery factor is calculated as the ratio of the final return (%) and the absolute value of the maximum drawdown over observed time (%). This indicator shows how many return units the investor earned for 1 unit of maximum risk in form of the largest % loss he had to “survive” as an investor in the past. It compares the return per unit of risk measured by the maximum drawdown in the investment strategy. The recovery factor (RF) is defined as:

$$RF = \frac{\text{total return of investment strategy}}{|MDD|}, \quad (3)$$

where MDD is defined by (2).

3.3 Trend definition

We have decided to test 3 kinds of trend definitions: moving average, moving minimums and ideal equity curve. First two are very good known for each trading system developer and the third is our own idea.

1. Moving average – if the investment strategy equity curve > moving average of the respective period, then we can say, that this investment strategy is in uptrend.
2. Moving minimums - if the investment strategy equity curve > moving minimums of the respective period, then we can say, that this investment strategy is in uptrend.
3. Ideal equity curve - if the investment strategy equity curve > line of the respective slope (exponential curve), then we can say, that this investment strategy is in uptrend.

Moving average (MA) and breakout system (BO)

We use the Moving average system as a tool for selecting investment strategies into the portfolio. We decided to use three different moving average periods ($X = 50$, $X = 100$ and $X = 200$). The algorithm for selecting investment strategies into the final portfolio is based on the cutting principle. If the closing value of the investment strategy is higher than its moving average, investment strategy is selected into the final portfolio.

Second tool for managing investment strategy portfolio is Breakout system. This tool generates entry or exit signals based on closing the strategy out of defined range. We define the range based on moving maximums and minimum in investment strategy values. Entry signal is defined if the closing value of

respective investment strategy is higher than its moving maximum (Mmax) and exit signal is defined if the closing value of investment strategy is below its moving minimum (Mmin). We decided to test the same three different periods as in case of moving averages ($X = 50$, $X = 100$ and $X = 200$). Compared to Moving average strategy, entry or exit signals are given by 2 variables (Mmax and Mmin). But we set them both always at the same level.

Table 1 shows the condition for entering/exiting investment strategies in the final portfolio.

Tab. 1 Buy and sell conditions for the MO and BO system (*Source: the authors*)

	Condition for ENTRY	Condition for EXIT
Moving Average System	Close price > MA (X)	Close price < MA (X)
Breakout System	Close price > Mmax (X)	Close price < Mmin (X)

Ideal equity curve (IEC)

Third tool is our own idea for managing the investment strategies portfolio. It is a curve that represents the condition of a hypothetical and ideal investment strategy over time. A hypothetical ideal investment strategy may be that which only grows at the same rate and without any volatility (exponential curve), or that which never grows or falls and has no volatility (straight line). We decided to test three different "slopes" - parameters (0%, 0.0297%/day = 7.7% per year, 0.0382%/day = 10% per year). These parameters were estimated by counting changes displayed by the MSCI World Index available on the internet at <https://www.msci.com/end-of-day-data-search>. An example of an EIC is displayed in Figure 2.

Setting IEC slope to 0 is a specific situation when investor does not gain any returns, but he also never gains negative returns. From this point of view, IEC is therefore "ideal" and only theoretical, so investor cannot invest in it. However, the IEC concept may be helpful in investment strategies selecting process. The investor will never know the future performance of investment strategies portfolio, but he can have minimum performance requirements on these investment strategies. If these investment strategies do not comply with his requirements, they are excluded from the final portfolio. The additional capital can be invested in those investment strategies which comply with requirements.

If three approaches (tools) are based on the idea that, if some investment strategy is excluded from the final portfolio, weights of the other investment strategies will proportionally rise (e.g. if one investment strategy is excluded = then weight of other strategies in portfolio will rise up to 25%). In case that all

investment strategies are exited from the portfolio, we calculate only with holding cash.

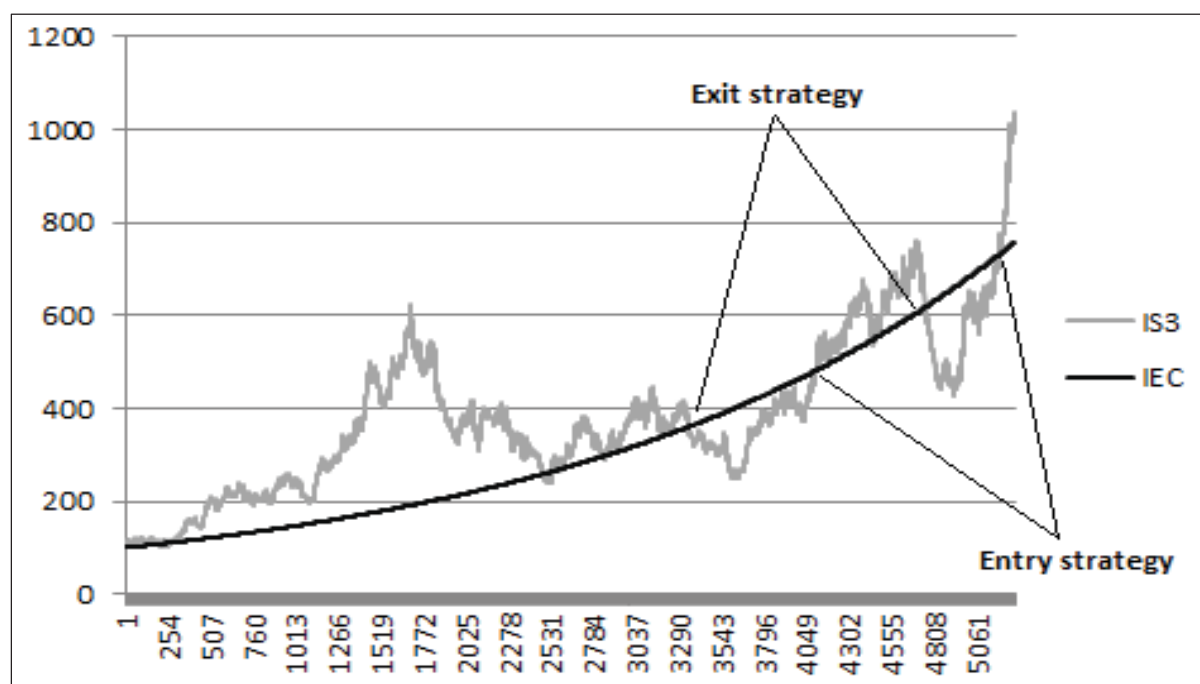


Fig. 2 Ideal equity curve (Source: the authors)

4 Results and discussion

Table 2 summarizes the most important results of our research. The figures in the table (# 1, # 2) represent the number of cases in which the use of portfolio investment management tools has achieved a better performance parameter than the passive diversification.

The figures in the third part of the table (# 3) reflect the frequency of cases where:

- a) Final portfolios achieved total return > 0 (profit) while passive diversification reached total return > 0 (profit);
- b) Final portfolios reached total return < 0 (loss) while passive diversification reached total return > 0 (profit);
- c) Final portfolios achieved total return > 0 (profit) while passive diversification reached total return < 0 (loss);
- d) Final portfolios achieved total return < 0 (loss) while passive diversification reached total return < 0 (loss).

The first part of the table (# 1) compares the results of the final portfolios with the passive approach based on the values of the monitored performance indicators. These results point to the fact that the use of tools based on the moving parameter concept on the one hand contributes to higher returns (TR) than passive diversification but on the other hand also to higher maximum

drawdowns (especially when using a lower period). In this comparison, the IEC tool seems to be the most capable of ensuring a lower maximum drawdown and a higher overall return on invested capital.

At first sight, the use of MA, BO and IEC systems seem to be a suitable management tool. In the first part of the table (# 1) we compared the performance of passive diversification and active portfolio management approach. However, there are also situations in which the active approach is better than passive diversification but shows low values of the TR and RF indicators, or very high maximum drawdowns values (for example - BO 50 system: TR= -88%; passive diversification: TR= -95%).

Tab. 2 Testing portfolios results (*Source: the authors*)

#1: Performance indicator comparison (52 simulations)									
BO, MA length	50			100			200		
IEC parameter ("slope")	IEC=0			IEC= 0,0297			IEC=0,038		
Performance indicators	TR	MDD	RF	TR	MDD	RF	TR	MDD	RF
MA	46	10	42	49	14	50	52	20	48
BO	46	27	43	51	27	44	51	29	49
IEC	51	39	51	51	38	51	51	37	51
#2: Performance indicator comparison (52 simulations)									
BO, MA length	50			100			200		
IEC parameter	IEC=0			IEC= 0.0297			IEC=0.038		
Performance indicators	TR	MDD	RF	TR	MDD	RF	TR	MDD	RF
MA	26	null	18	37	null	35	40	null	36
BO	33	null	30	41	null	34	44	null	42
IEC	50	null	50	50	null	50	50	null	50
#3: Total return comparison (52 simulations)									
Length/parameter	50	100	200	50	100	200	0	0.0297	0.038
Strategy	MA	MA	MA	BO	BO	BO	IEC	IEC	IEC
Strategy + Naive +	26	26	26	26	26	26	26	26	26
Strategy - Naive +	0	0	0	0	0	0	0	0	0
Strategy + Naive -	2	11	14	13	16	19	25	25	25
Strategy - Naive -	24	15	12	13	10	7	1	1	1

For this reason, in the second part of the table (#2) we added the non-negative condition of the TR and the RF indicator. Thus, figures in the second part of the table represent the number of cases (out of 52 simulations) in which the use of the active management tool (MA, BO, IEC) is more advantageous than the benchmark, but the investment also has to achieve positive values of TR and RF indicator. The maximum drawdown takes values from the interval $<-1.0>$ and therefore the non-negativity condition is unfounded. The most significant change in the figures versus the first part of the table (#1) is observed in MA and BO systems. Especially with the MA tool, we can see an increase in the number of cases when final portfolio (strategy) claims better TR and RF values than benchmark by using a longer period for MA. Kisela et al. (2015) also deal with the management of individual equity curves of investment strategies in their research. The authors' study looked at the use of various technical analysis tools, one of which was the moving average. Their research indicates that the use of moving averages to manage an individual investment strategy is linked to the optimization process. However, results from the second part of the table (#2) show that the IEC tool was the only one capable to achieve a positive return in 50 cases or positive value of the RF indicator while at the same time overcoming naive diversification.

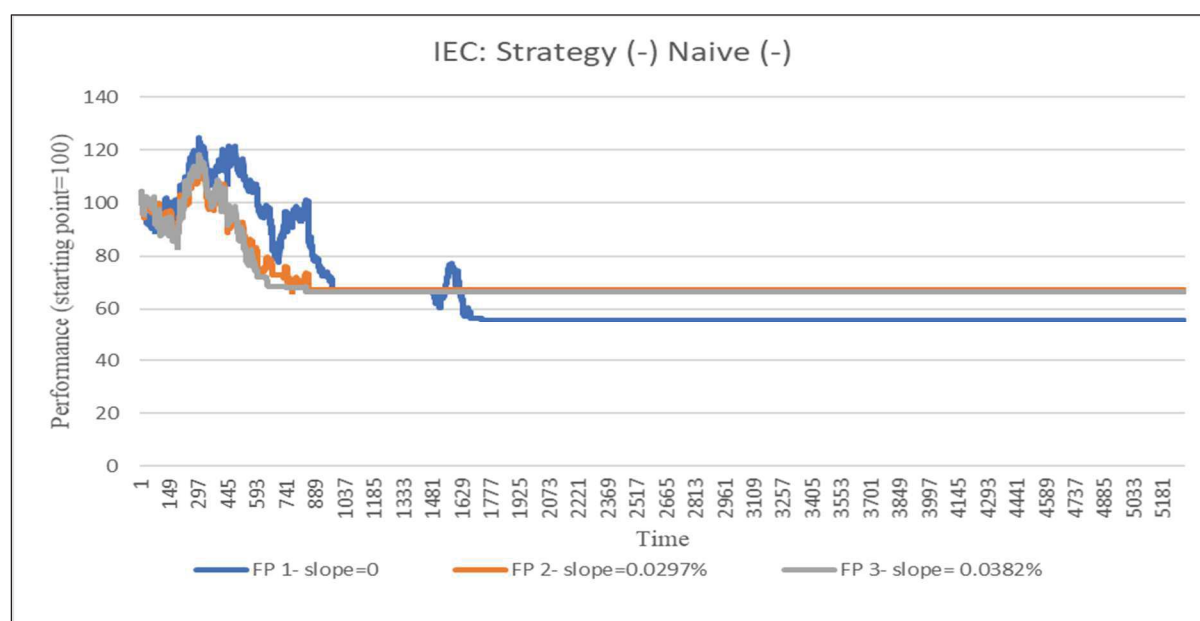


Fig. 3 Performance of IEC final portfolios (FP) in case of result: Strategy (-) Naive (-) (Source: the authors)

An interesting finding from the last part of the table (#3) is the fact that none of the simulations had a situation where final portfolio (strategy) would made loss and passive approach made profit (Strategy - Naive +). Our attention was logically most interested in the situation when the IEC tool made loss same as

the passive approach. Each investor would like to avoid tools that bring him a low possibility of outperforming the passive approach plus with loss result. And there was only one situation for each slope, when the IEC tool made loss same as the passive approach (Figure 3). So, the rest of the simulations ended with making IEC tool a profit for the investor (51 simulations). The fact is that this situation is different when using MA or BO systems.

Our last Table 3 summarizes some additional figures. We tested our performance parameters (indicators) from the point of view of all 52 simulations.

Tab. 3 Testing portfolios results (*Source: the authors*)

Length/slope	50	100	200	50	100	200	0	0.0297	0.038
Strategy	MA	MA	MA	BO	BO	BO	IEC	IEC	IEC
MDD Max	-0.96	-0.97	-0.94	-0.92	-0.90	-0.91	-0.69	-0.61	-0.60
MDD Avg.	-0.62	-0.61	-0.57	-0.55	-0.56	-0.52	-0.38	-0.37	-0.42
RF Min	-0.99	-0.99	-0.96	-0.96	-0.92	-0.95	-0.8	-0.78	-0.77
RF Avg.	5.06	7.32	9.91	5.03	7.31	10.79	28.25	27.53	18.97
TR Min	-0.95	-0.95	-0.91	-0.89	-0.83	-0.86	-0.44	-0.33	-0.34
TR Avg.	1.99	3.12	3.64	2.00	3.23	4.44	8.97	8.98	6.70

This table brings us some new findings. For example:

- IEC was always the best in case of Maximal Drawdown (each slope).
- IEC was always the best in case of average Maximal Drawdown (each slope).
- IEC was always the best in case of minimal Recovery Factor (each slope).
- IEC was always the best in case of average Recovery Factor (each slope).
- IEC was always the best in case of minimal Total Return (each slope).
- IEC was always the best in case of average Total Return (each slope).
- The second best in terms of active management was BO system and the worst MA system (nearly in all periods).
- Setting the “signal” period (MA, BO) just like slope (IEC) plays key role for better final results (performance indicators), thus from our results “more is better” (MA, BO period) and “less is more” (IEC slope).

5 Conclusion

Today's historically lowest interest rates make from free cash one of the least profitable assets. The benefit of cash holdings will be reduced if the company has "redundant" cash. Whether this situation is short-term or long-term, the fact is that every company can invest it in the financial market, just as individual

investors do with their savings. Today, financial market brings opportunities not only in form of conventional financial instruments like stocks, bonds, ETF etc., but due to advances in the financial industry also in new financial services. The biggest advantage of these solutions is that, unlike hedge funds, they are already available at very low investments and allow investors to spread their capital into an unlimited investment strategy portfolio. These strategies may often bring a high return in a relatively short time, but also a very risky environment. Thus, diversification plays a key role for final portfolio creating. There are two main approaches to diversification - passive and active. In active diversification an investor or company needs some tool that will signalize some inappropriate circumstances in strategy performance and thus some investment strategy have to be excluded from the final portfolio. We have tested 3 such tools via different trend definitions.

We simulated 10 investment strategies in 5294 business days, based on Dow Jones Industrial Average returns. From these 10 investment strategies (IS1 – IS10) we randomly choose 52 simulations. Each of this simulation was created by 5 different investment strategies. In these simulations we tested 3 tools via different trend definitions: moving average system, breakout system and ideal equity curve. We compared final portfolios performance with passive approach by three indicators: return indicator (Total return), risk indicator (Maximum drawdown) and whole performance indicator (Recovery factor).

At first sight, the use of MA, BO and IEC systems seemed to be a suitable management tool. One interesting finding was the fact that none of the simulations had a situation where final portfolio (strategy) would make loss and passive approach profit (Strategy - Naive +). However, there were also situations (in case of MA and BO) in which the active approach was better than passive diversification but showed low values of the TR and RF indicators, or very high maximum drawdowns values. Results from our research showed that the IEC tool was the only one capable to achieve a positive TR in 50 cases or positive value of the RF indicator while at the same time overcoming naive diversification. There was only one situation for each slope, when the IEC tool made loss same as the passive approach (Strategy - Naive -). In another comparison and in case of all 52 simulations, IEC was always the best tool from the point of view of Maximal Drawdown, average Maximal Drawdown, minimal Recovery Factor, average Recovery Factor, minimal Total Return and average Total Return. We have found out that setting the “signal” period (MA, BO) just like slope (IEC) plays key role for better final results (performance indicators), thus from our results “more is better” (MA, BO period) and “less is more” (IEC slope).

From our point of view, the trend definition via IEC has the most effective impact on final performance parameters. Surely and of course, we used in our research 5 plus 5 absolutely different investment strategies on one clear purpose. To demonstrate, that only very simple method (IEC) can be very useful and effective in active diversification of investment strategies portfolio. In fact, no investor or company knows in advance what direction of investment will occur, but IEC was able in absolute most cases bring to investor positive TR, RF and also better results than naive (passive) diversification.

The paper arose in fulfilment of the tasks associated with grant scheme VEGA 1/0767/18 *SMART model - a decision support tool in management of enterprises* provided by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

6 References

- Allen, D., McAleer, M., Powell, R., Singh, A. (2016). *Down-Side Risk Metrics as Portfolio Diversification Strategies across the Global Financial Crisis*. Journal of Risk and Financial Management, 9(2), 1-18.
- Demiguel, V., Garlappi, L., Uppal, R. (2009). *Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?* Review of Financial Studies, 14(3), 644-664.
- Fernandez, P., Acin, P. F. (2018). *It Has Been Very Easy to Beat the S&P500 in 2000-2018: Several Examples*. Research paper. [cit. 20-05-2019] Dostupné na https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3184501.
- Hwang, I., Xu, S., In, F. (2018). *Naive versus optimal diversification: Tail risk and performance*. European Journal of Operational Research, 265(1), 372-388.
- Kisela, P., Virdzek, T., Vajda, V. (2015). *Trading the Equity Curves*. Procedia Economics and Finance, 32, 50-55.
- Pospisil, L., Vecer, J. (2008). *PDE methods for maximum drawdown*. The Journal of Computational Finance, 12(2), 59-76.
- Tomasini, E., Jaekle, U. (2011). *Trading Systems: A new approach to system development and portfolio optimisation*. Petersfield, UK: Harriman House, 2011.
- Tu, J., Zhou, G. (2011). *Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies*. Journal of Financial Economics, 127, 204-215.

Výpočtová štatistika 2019

Informácia o udalosti SŠDS

Computational Statistics 2019

Information about an event by SŠDS

V dňoch 5. až 6. decembra 2019 sa uskutočnil 28. ročník medzinárodného vedeckého seminára Výpočtová štatistika 2019. Miesto konania je tradične Prírodovedecká fakulta UK, Prezentačné centrum AMOS, pavilón B1, Mlynská dolina, Bratislava 4. Prvý deň, štvrtok 5. decembra 2019, sa niesol v znamení prezentácií príspevkov účastníkov seminára.

Spolu odznelo 10 príspevkov z rôznych oblastí štatistiky aj demografie. Napríklad G. Szűcs (FMFI UK v Bratislave) sa venoval oceňovaniu spojených dôchodkov v prípade závislosti životov poistených osôb, P. Jindrová (FES Univerzita Pardubice) riešila modelovanie extrémnych poistných škôd v dôsledku prírodných katastrof. Príspevky z oblasti demografie na tému strednej dĺžky života v dobrom zdraví a jej regionálnym rozdielom v ČR predniesli D. Morávek, T. Fiala, J. Langhamrová a J. Vrabcová (všetci z VŠE v Prahe). Ďalej odzneli príspevky s tematikou určenia veľkosti vzorky pre aplikovaný výskum (I. Stankovičová, FM UK v Bratislave), definovanie finančnej tiesne podniku pomocou metód panelového zhliadovania (M. Stachová, UMB v B. Bystrici), kvalita života a ľudský rozvoj v krajinách EÚ (A. Mojsejová, TU Košice), použitia štatistických mier v odvetvových štatistikách (M. Boďa, UMB v B. Bystrici), modelovanie mzdových rozdelení 4-parametrickým lognormálnym rozdelením (D. Bílková, VŠE v Prahe) a ďalšie.

Obrázok 1 ukazuje atmosféru počas seminára.



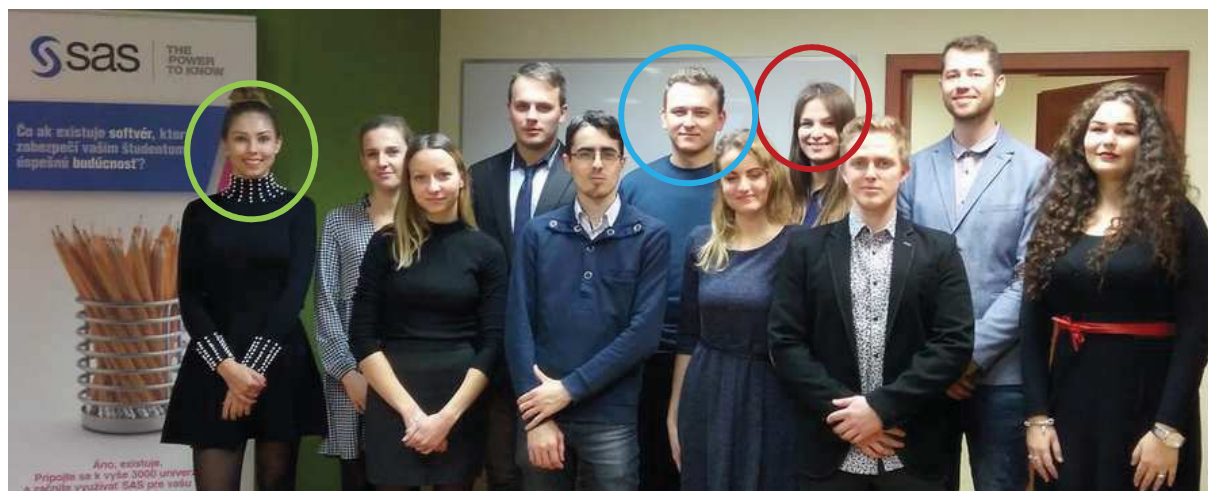
Obr. 1: Pohľad na účastníkov vedeckého seminára VS 2019 v Prezentačnom centre AMOS.

Popoludní vystúpilo 12 študentov v rámci Prehliadky prác mladých štatistikov a demografov 2019 a prezentovali výsledky svojej práce počas štúdia. Tento rok sa ich vystúpenia venovali hlavne demografickým a geografickým témam, lebo až 9 študentov z celkového počtu 12 prihlásených študentov bolo z Prírodovedeckej fakulty UK v Bratislave. Zúčastnení študenti získali knižné odmeny a 3 víťazné práce aj finančné odmeny.

Odborná porota vedená Martinom Boďom vybrala nasledovné tri víťazné práce z prednesených prác študentov:

1. Nikoleta Slavíková (PriF UK): Sťahovanie v uliciach mesta Skalica v rokoch 2010-2018 (červený kruh).
2. Dominika Zorvanová (FMFI UK): Aplikácia data miningu v marketingu (zelený kruh).
3. Ondrej Oravec (PriF UK): Využitie nepriamo štandardizovaných hrubých mier pri tvorbe Webbovej typológie celkového prírastku obyvateľstva v obciach SR v období 1988 – 2016 (modrý kruh).

Obrázok 2 predstavuje súťažiacich a identifikuje víťazov.



Obr. 2: Študenti - účastníci Prehliadky prác mladých štatistikov a demografov 2019.

V piatok 6. decembra 2019 pre študentov vysokých škôl bolo zorganizované zásluhou SŠDS pásmo prednášok pod názvom „Analytika očami profesionálov“. Účasť študentov bola bohatá (cca. 130 študentov). Prišli študenti z Univerzity Komenského v Bratislave (FM UK, FMFI UK, PriF UK), ale aj z Ekonomickej univerzity v Bratislave (FHI) a Slovenskej poľnohospodárskej univerzity v Nitre.

Študentom prednášali odborníci z praxe, ktorí majú skúsenosti v oblasti analýzy údajov. Témy prednášok boli nasledovné:

1. Milan Schnorrer a Marek Kováčik (Pricewise): Príklady použitia zákazníckych dát v maloobchode, resp. retaile.

2. Jaroslav Burian (PřF UP Olomouc): Štatistické metódy v územnom plánovaní.
3. Samuel Hudec (Mall.cz): Optimalizácia marketingových nákladov.
4. Mária Kulová (ŠÚ SR): Nový koncept Sčítania obyvateľov, domov a bytov v roku 2021.

Fotosnímka z priebehu podujatia je obsahom obrázku 3.



Obr. 3: Pohľad do auditória počas pásma prednášok pre študentov „Analytika očami profesionálov“.

Informácie o prednášajúcich

Mgr. Milan Schnorrer je zakladateľom a partnerom úspešnej slovenskej spoločnosti Pricewise. Vzdelanie získal na Fakulte managementu UK v Bratislave (Bc. stupeň) a v Holandsku na University of Groningen (Mgr. stupeň, odbor Marketing), kde dnes pôsobí ako hosťujúci profesor. Študijný odbor, ktorý študoval v Holandsku bol zameraný na analýzu údajov o zákazníkoch a využívanie marketingových modelov na skúmanie správania spotrebiteľov. Dnes je expertom s 10 ročnými skúsenosťami na priamu komunikáciu v oblasti analýzy údajov o zákazníkoch a má rozsiahle skúsenosti s navrhovaním a správou zákazníckych programov.

Mgr. Marek Kováčik absolvoval Fakultu matematiky, fyziky a informatiky Univerzity Komenského v Bratislave, odbor aplikovaná matematika. Počas bakalárskeho štúdia začal pracovať ako business analytik v spoločnosti Pricewise. Od absolvovania štúdia vedie v Pricewise tím, ktorý pomáha klientom ako NAY, MALL.CZ alebo Curaprox lepšie pracovať s dátami a využívať ich na zvyšovanie tržieb.

doc. RNDr. Jaroslav Burian, Ph.D. od roku 2007 pracuje na Katedre geoinformatiky PřF UP Olomouc, kde vyučuje geoinformatiku v odboroch humánna geografia, verejná správa a územné plánovanie. Jeho vedeckým zamarením je predovšetkým oblasť implementácie geoinformačných technológií

do priestorového plánovania a analýzy, modelovanie a simulácie mestského prostredia pomocou kvantitatívnych metód. Od roku 2009 vyvíja software Urban Planner, určený pre modelovanie scenárov potenciálu územia, ktorý bol nasadený v niekoľkých mestách a krajoch. Popri pedagogickej a výskumnej činnosti spolupracuje s verejnou správou a komerčným sektorom. Od roku 2019 pôsobí ako externý spolupracovník na Katedre geoinformatiky (UK v Prahe) a na Ústave informatiky a aplikovanej matematiky (MVŠO). Aktívne pôsobí v Českej asociácii pro geoinformace v odbornej skupine „GIS a územní plánování“.

Mgr. Samuel Hudec, PhD. absolvoval Fakultu prírodných vied na Univerzite Mateja Bela v Banskej Bystrici, kde získal Mgr. titul v odbore Finančná matematika a Matematická štatistika (2013). Doktorandské štúdium ukončil v auguste 2019 tiež na tejto univerzite v odbore pravdepodobnosť a matematická štatistika. V súčasnosti pracuje ako dátový vedec - analytik a modelár v spoločnosti Internet Mall, a.s. v Prahe. Táto spoločnosť pokrýva e-commerce trh prevažne v strednej Európe.

Mgr. Mária Kulová absolvovala Fakultu sociálnych vied, Univerzity sv. Cyrila a Metoda v Trnave. Pracuje na Štatistickom úrade SR na sekcii sociálnych štatistík a demografie. V súčasnosti je vedúcou oddelenia sčítania obyvateľov, domov a bytov a prierezových štatistík. Vo svojej práci sa venuje najmä prípravným štatistickým činnostiam, ktoré súvisia s cenom.

Iveta Stankovičová
predsedníčka SŠDS

Spomienka na doc. Ing. Jozefa Chajdiaka, CSc.

(* 18-03-1952 † 22-10-2019)

Smútočný oznam

A memory of doc. Ing. Jozef Chajdiak, CSc.

(* 18-03-1952 † 22-10-2019)

Obituary



V októbri 2019 sme sa naposledy rozlúčili s jednou z najvýraznejších osobností Slovenskej štatistickej a demografickej spoločnosti (SŠDS), s **doc. Ing. Jozefom Chajdiakom, CSc.** Zomrel vo veku 67 rokov po ťažkej a dlhotrvajúcej chorobe. Posledná rozlúčka sa konala 25. októbra 2019 v Bratislavskom krematóriu.

Niektoré fakty z jeho života

Jozef Chajdiak sa narodil 18.3.1952 v Bratislave. V rokoch 1970 – 1975 študoval na Fakulte riadenia Vysokej školy ekonomickej (FR VŠE) v Bratislave špecializáciu Ekonomicko-matematické výpočty. V rokoch 1975 – 1978 úspešne absolvoval internú vedeckú ašpirantúru na Moskovskom inštitúte riadenia, kde získal vedeckú hodnosť kandidáta ekonomických vied v špecializácii Matematické metódy a použitie výpočtovej techniky v plánovaní a riadení národného hospodárstva a jeho odvetví.

V rokoch 1978 - 2006 pracoval na Katedre štatistiky FR VŠE (teraz Katedre štatistiky Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave) ako odborný asistent (1978 - 1984), od roku 1984 ako docent a od roku 2003 ako mimoriadny profesor. V rokoch 1982 až 1990 bol zástupcom vedúceho katedry. V oblasti pedagogickej činnosti viedol rôzne predmety z oblasti štatistických metód a ekonomickej štatistiky: Štatistika, Ekonomická štatistika, Hospodárska štatistika, Ekonomicko-štatistické analýzy a podobne.

V rokoch 2008 - 2018 pracoval ako docent na Ústave manažmentu STU v Bratislave. V roku 1991 založil spoločnosť STATIS a spolupracoval s praxou na riešení úloh z oblasti ekonomických analýz a aplikácie kvantitatívnych metód na podporu rozhodovania. Pracoval hlavne v oblasti finančno-štatistických analýz podnikov, aplikácie štatistických metód v procesoch riadenia kvality a prognózovania ekonomického a demografického vývoja Slovenska resp. nižších územných celkov.

Jozef Chajdiak vstúpil do SŠDS v roku 1980 a zároveň sa stal aj členom Výboru SŠDS. V rokoch 1984 – 2010 zastával funkciu vedeckého tajomníka a v období 2010 – 2014 bol predsedom SŠDS. Významnou mierou prispel k „oživeniu“ činnosť tejto Spoločnosti. Prispel k tomu, že SŠDS aj v súčasnosti aktívne pôsobí na Slovensku. Spolu s kolegom RNDr. Jankom Luhom, CSc. založil v roku 2005 vedecký časopis Forum Statisticum Slovaccum (FSS) a bol desať rokov jeho šéfredaktorom. Bol iniciátorom, organizátorom alebo spoluorganizátorom desiatok podujatí a autorom mnohých príspevkov o histórii Spoločnosti. Organizoval resp. spoluorganizoval Slovenskú štatistickú konferenciu, Slovenskú demografickú konferenciu, založil týždňovú školu štatistiky EKOMSTAT (r. 1988), medzinárodný seminár Výpočtová štatistika (r. 1991), konferenciu Pohľady na ekonomiku Slovenska (r. 2002), Prehliadku prác mladých štatistikov a demografov a ďalšie akcie.

Docent Chajdiak bol známy slovenský odborník v oblasti metód na podporu rozhodovania, štatistiky a ekonómie. Je autorom 46 knižných publikácií, 22 skript a vyše 400 publikovaných časopiseckých alebo konferenčných príspevkov. Predstavuje priekopníka zavádzania výpočtovej techniky do výučby aj do praxe pri riešení štatistických úloh a ich aplikácií. V slovenských podmienkach organizoval osvojenie systémov SAS a SPSS na sálových počítačoch a systému Statgraphics na osobných počítačoch. Aktívne pracoval na šírení postupov práce s tabuľkovými procesormi (Lotus, SuperCalc, Quattro). Neskôr venoval pozornosť popularizácii riešenia štatistických úloh pomocou funkcií, nástrojov (PivotTable) a doplnkov riešení (Data Analysis) v Exceli (publikácie: Štatistika v Exceli; Štatistické úlohy a ich riešenie v Exceli; Štatistika v Exceli 2007; Štatistika jednoducho v Exceli).

Vo svojej vedecko-výskumnej činnosti sa venoval meraniu a analýze makroekonomických indikátorov hospodárskeho rastu ekonomiky Slovenska, finančným analýzám v podnikoch, aplikácii štatistických a demografických metód v oblasti rozhodovania. Bol až neskutočne aktívny pri vypracovávaní a podávaní rôznych interných grantových projektov a grantov na základe výziev Ministerstva školstva SR. Mnohé jeho projekty boli prijaté a teda bol ich úspešným riešiteľom. K spolupráci na projektoch pozýval aj mnohých svojich spolupracovníkov.

Niekoľko reminiscencií členov SŠDS a kolegov na Jozefa Chajdiaka

Docent Chajdiak patrí medzi osobnosti, ktoré značnou mierou prispeli k rozvoju štatistiky na Slovensku. Náleží mu významné miesto v histórii Slovenskej štatistickej a demografickej spoločnosti. Síce opustil tento svet, ale jeho nezmazateľná stopa v histórii SŠDS zostala a bude pretrvávať. Nielen že stál pri zrode mnohých akcií SŠDS, ale sa ich aj aktívne zúčastňoval. Vždy tam prichádzal s úsmevom a skvelou náladou. Svojím elánom a optimizmom dokázal „nakaziť“ všetkých účastníkov konferencie či seminára.

Do práce v SŠDS a v oblasti štatistiky vedel prilákať svojím príkladom a aktívnou prácou aj mnohých mladých kolegov a študentov. Aj ja som mu vďačná, že som po skončení školy mohla ostať pracovať na Katedre štatistiky VŠE v Bratislave. Je to aj jeho zásluha, že dnes som odborníčka na štatistické metódy v SASe a že som dlhé roky aktívnou členkou v SŠDS. Keď mu ťažká choroba obmedzila okruh jeho aktivít, prevzala som vedenie Spoločnosti a pokračujem v jeho práci a aktivitách.

Jozef bol veľký bojovník náročný na seba, ale aj na spolupracovníkov. Jeho odbornosť a ľudské kvality nám budú v SŠDS trvale chýbať. Všetci sme ho poznali ako mimoriadne pracovitého a činnorodého človeka so zmyslom pre osobitý humor. Vždy naňho budeme spomínať a bude nám chýbať.

Jozef, česť Tvojej pamiatke!

Iveta Stankovičová
predsedníčka SŠDS

Poznal som Jožka ešte zo študentských čias na Fakulte riadenia VŠE v Bratislave, kde sa sme obaja snažili presadiť v študentskej vedeckej a odbornej činnosti, čo predznamenal naše neskoršie pracovné záujmy. Neskôr, v sedemdesiatych rokoch, sme spolu bojovali s ruštinou a riešením našich dizertačných tém na Moskovskom inštitúte riadenia, čo sme aj úspešne zvládli. Študovalo tam veľa študentov z Čiech, Moravy a Slovenska, s ktorými nás spojili a až doteraz spájajú priateľstvá. Jožko sa tu mimo náročného štúdia prejavoval aj ako vyspelý šachista, nadšený športovec a turista, ale aj ako výborný kuchár a kamarát, ktorý nikdy nepokazil žiadnu zábavu. To, že vie zvládnuť nielen vedeckú, ale aj tvrdú fyzickú prácu dokázal aj svojou prácou na stavbe Bajkalsko-amurskej magistrály. V osemdesiatych rokoch sme opäť pôsobili ako kolegovia na Fakulte riadenia VŠE, on na Katedre štatistiky a ja na katedre ASR. Naše profesionálne dráhy sa rozišli v deväťdesiatych rokoch, keď on zostal na VŠE a ja som odišiel do praxe, aby sme sa zišli opäť v r. 2008 na Ústave manažmentu STU.

Doc. Ing. Jozef Chajdiak, CSc. bol naším milým kolegom a priateľom počas svojho pracovného života na Ústave manažmentu STU. Bol som ten, ktorý ho

presvedčil prísť medzi nás na tento Ústav. Počas tohto obdobia sa prejavil ako zapálený učiteľ ekonomických predmetov a štatistiky, ktorý vedel upútať študentov, úspešný výskumník a autor desiatok vedeckých článkov, publikácií a učebníc, ktoré aj sám vydával. Popri tom pôsobil ako tajomník Slovenskej štatistickej a demografickej spoločnosti a tiež ako agilný a schopný organizátor vedeckých konferencií a seminárov zameraných na ekonomiku Slovenska a aplikáciu štatistických a demografických metód v riadení podnikov a ekonomiky. Vedel sa pritom obklopiť tímom ľudí, ktorí mu boli podobní. Poznali sme ho ako milého, kompetentného a mimoriadne pracovitého človeka so zmyslom pre osobitý humor, ochotného komukoľvek kedykoľvek pomôcť. Je škoda, že nešťastný úraz a neskôr neľahká choroba mu obmedzili okruh aktivít, ale napriek tomu bol Jozef stále bojovník náročný na seba a vedený snahou zvládať čo najviac z toho, čo zvládal, keď bol v plnej forme. Jeho odbornosť a ľudské kvality nám budú na Ústave manažmentu STU trvale chýbať.

Milý Jožko, lúčim sa s Tebou za kolegov z Ústavu manažmentu a tiež za česko-moravsko-slovenskú študentskú komunitu z Moskvy. Budeme Ťa postrádať a uchováme si Ťa v pamäti z čias Tvojej najlepšej formy. Nech Ti je zem slovenská ľahká.

Marián Zajko

riaditeľ Ústavu manažmentu STU v Bratislave

Pán docent Chajdiak bol jedným z prvých učiteľov, s ktorým som sa po svojom nástupe na Fakultu riadenia VŠE v roku 1981 stretol ako so svojím pedagógom nielen na pôde fakulty, ale aj v neformálnych rozhovoroch o význame štúdia, o živote, ale aj o priateľstve a spolupráci. Pokiaľ viem, rovnakú skúsenosť mali aj ďalší študenti, ktorí si ho ako pedagóga vážili a rešpektovali jeho náročnosť ako snahu odovzdať im maximálne množstvo svojich poznatkov. Verím, že docent Chajdiak ostal v spomienkach svojich študentov ako učiteľ, na ktorého prednášky chodili s dychtivým očakávaním preniknúť do tajov štatistiky a pokiaľ bude študijný program Štatistika naďalej úspešne pripravovať študentov, potiaľ mu budeme aj my všetci vzdávať hold ako jednému z jeho protagonistov.

Vieme, že nikto z nás nepozná deň, ani hodinu konca krivky životnej púte. Zostáva nám urobiť všetko preto, aby sme mali v tej chvíli čisté svedomie a tí, čo po nás zostanú, spomínali na nás len v tom najlepšom. Verím, pán docent, že Ty si to z veľkej časti dokázal. Chcel by som preto v mene vedenia Fakulty hospodárskej informatiky, v mene jej terajších aj bývalých pracovníkov, v mene Tvojich študentov, a v mene svojom vyjadriť presvedčenie, že pre nás všetkých odchádza s docentom Chajdiakom múdrosť a svojský prístup k riešeniu problémov.

Okrem toho, že sa s docentom Chajdiakom lúčim ako predstaviteľ inštitúcie, na ktorej dlhodobo v rokoch 1975 až 2006 pôsobil, chcel by som sa ním rozlúčiť aj za skupinu priateľov, s ktorými v dobrej atmosfére často rozoberal základné problémy ľudstva, nosné problémy štatistiky, ale aj politické problémy a názory na kvalitu vína a dobrého jedla. Lúčime sa s kolegom, na ktorého určite nezabudneme, nezabudneme na jeho „El pueblo, unido, jama sera vencido“. Dnešným dňom sa síce končí jedna životná krivka, ale v našich srdciach a myšliach ostanú výsledky jej riešenia ako stále živý odkaz a nasledovaniahodný príklad.

Češť Tvojej pamiatke!

Ivan Brezina

dekan Fakulty hospodárskej informatiky EU v Bratislave

Jozefa (spolu s Jankom Luhom) som spoznal ako „mladé ucho“ na jednej z demografických konferencií. Stačilo mi pár okamihov, aby som skonštatoval: *“Tak s týmito chlapmi bude sranda...”* A bola. Páčilo sa mi, že Jožo bol vždy „v stave“ (ako to on rád používal) akceptovať, prijať, zmeniť názor, vypočúť si iného. Nikdy, ale skutočne nikdy sa neurazil, aj keď mal niekto iný názor alebo mu povedal niečo kritické. Nikdy som ho nepočul zvýšiť hlas. „Nedotklivosť“ je jedna z najvzácnejších ľudských cností a Jožo bol jeden z mála ľudí, ktorý ňou oplýval.

Aj keď som už skonštatoval, že Jožo bol stelesnením dobrej nálady a humoru, dokázal (aj) vedenie spoločnosti robiť s ľahkosťou a nonšalantnosťou, a týkalo sa to tiež organizovania konferencií. Čo mi však utkvelo a bude tkvieť v pamäti asi najviac, je jeho neutíchajúca energia aj v dobe, kedy to už po zdravotnej stránke naozaj nemal ľahké a úprimne, asi iba málokto z nás by v takej situácii, napríklad vôbec prišiel na zasadnutie výboru Spoločnosti. Na to si spomeniem, keď príde nejaká ťažká chvíľa, spomeniem si, že Jožo by tú paličku z kúta opäť zobral a prišiel tam kam treba...

Branislav Bleha

podpredseda SŠDS pre demografiu

Když jsem se dozvěděla velmi smutnou zprávu o úmrtí Jozefa Chajdiaka, velmi mne to zasáhlo. Jozef patřil bezesporu spolu s Jankem Luhom k nezapomenutelným osobnostem Slovenské statistické a demografické společnosti. Opustil svoji milovanou statistickou partu v relativně nízkém věku, což je o to smutnější. Rozhodně nelze zapomenout na jeho milou, vstřícnou a optimistickou povahu. Když jsem se před mnoha lety poprvé účastnila konference (Fernstat či Výpočtová štatistika), měla jsem tu čest se s ním a jeho kolegou Jankem poprvé setkat. Ohromila mě Jozefova vstřícnost, obrovská

ochota a komunikativnost a nezapomenutelný smysl pro humor. Od té doby jsem si konference pořádané Slovenskou statistickou a demografickou společností bez jeho účasti ani neuměla představit. Byl zárukou nejenom dobré zábavy, ale také odborného přínosu. Vždy se snažil jednoduše vysvětlit složitosti statistiky a dařilo se mu to. I když byl v poslední době již nemocný, tak bylo vidět, že v něm stále dřímá mladý kluk, který se nebrání žádné legraci, a všichni jsme ho měli rádi. Do poslední chvíle žil pro svoji práci, kterou bezesporu ovládal a miloval. Pro mne byl a zůstane Jozef nezapomenutelnou osobností slovenské statistiky. Bude mi na akcích Slovenské statistické a demografické společnosti výrazně chybět.

Jitka Langhamrová

Fakulta informatiky a statistiky VŠE v Praze
predsedníčka Českej demografickej spoločnosti

Velmi mě zasáhla zpráva o úmrtí Joža Chajdiaka. Poznala jsem ho na začátku svého profesního působení v oblasti statistiky a s jeho osobností mám spojení celou řadu různých aktivit. Od společné přípravy výuky výpočetní statistiky na Vysoké škole ekonomické v Bratislavě a Vysoké škole ekonomické v Praze, při níž jsem Joža poznala, přes statistické konference konané na Slovensku, například tzv. semináře Výpočtová štatistika, až po časopis Informácie, později Forum Statisticum Slovaccum.

Protože v Čechách v mých mladých letech neexistovala statistická společnost, stala jsem se členkou Slovenské demografické a statistické společnosti, později existující pod názvem Slovenská štatistická a demografická spoločnosť. I v ní Jožo dlouhá léta zastával důležité funkce. Spolu s Jankem Luhou podporovali založení České statistické společnosti v roce 1990.

Jožo byl autorem mnoha knih, inspirativní pro výuku byly zejména knihy o statistice v Excelu. Významně přispěl k vytvoření podmínek pro co nejširší předávání zkušeností mezi statistiky, a to nejen na Slovensku, ale i v Čechách a dalších sousedních zemích. Navíc stále usměvavý, přátelský a humorný. Takový byl a stále je v mých vzpomínkách Jožo Chajdiak.

Hana Řezanková

Fakulta informatiky a statistiky VŠE v Praze

Jozefa som spoznal ako začínajúci asistent na Katedre štatistiky Vysokej školy ekonomickej v Bratislave po jeho návrate z vedeckej ašpirantúry na moskovskej univerzite. Od samého začiatku bol pre mňa odbornou autoritou a imponoval mi jeho inovatívny prístup k výučbe nie veľmi obľúbenej štatistiky medzi študentmi, hlavne vtahovaním prvkov výpočtovej techniky a snahou čo najviac pracovať s údajmi z reálnej praxe. Aj keď sa neskôr naše asistentské cesty rozišli, vedeli sme o sebe a pravidelne sme sa stretávali na rôznych akciách Slovenskej

štatistickej a demografickej spoločnosti, ktorých bol dlhé roky nadšeným organizátorom. Blízko k sebe sme mali aj ľudsky. Naše manželky boli spolužiačky, takže napríklad pri stretnutiach na pamätných fakultných plesoch bolo aj vďaka Jozefovi o zábavu vždy postarané.

Odchodom Jozefa do večnosti došlo k veľkej strate pre slovenskú štatistiku a demografiu, ale taký je život. Ja osobne si na neho vždy spomeniem s veľkou úctou a poklonou za jeho nesmiernu pracovitosť.

Mikuláš Cár
Národná banka Slovenska
s manželkou Blankou

Prvýkrát som meno Jozef Chajdiak počula ako študentka Vysokej školy ekonomickej v druhej polovici osemdesiatych rokov. Štatistiku nám neprednášal ani nemal s nami cvičenia, tak bolo v tom čase nemožné posúdiť, či chýry o jeho prísnosti, ktoré kolovali po chodbách školy boli pravdivé. No a pred cca 22 rokmi som ho spoznala osobne. Ako veľkého odborníka, vôbec nie prísneho, ale milého a priateľského človeka. Neboli to iba akcie, ktoré organizoval spolu s Jankom Luhom, ktoré nás spájali. Okrem viacerých kníh Štatistika v Exceli (s rôznymi obmenami v názve) písal aj publikácie o Finančnej analýze. K niektorým sme mu vo firme pripravovali podkladové údaje, neraz sme debatovali o konštrukcii pomerových finančných ukazovateľov – najmä ak sme na ňu nemali rovnaký názor. Išlo však o konštruktívny dialóg a verím, že nás obidvoch odborne obohacoval a posúval. Rozpísanú knihu žiaľ už nestihol dopísať a zostane mi síce pre ňu prázdne miesto na poličke, ale v mojom srdci spomienka na neho bude navždy zachovaná.

Jana Marková
hlavná analytička v CRIF – Slovak Credit Bureau, s. r. o.

Pre mňa sa názov našej spoločnosti stále personifikuje do dvoch osobností slovenskej štatistiky – Janka Luhu a Joža Chajdiaka. Obaja od roku 1990 pôsobili po dlhé roky ako funkcionári výboru SŠDS. Boli iniciátormi a neúnavnými organizátormi viacerých pravidelných podujatí, ktoré prežili až do dnešných dní. Stáli pri vzniku vedeckého časopisu Forum Statisticum Slovaca, ktorý aj redigovali a často i technicky pripravovali do tlače. Vo svojej práci v rámci spoločnosti sa snažili aktivizovať tiež mladších kolegov – štatistikov a demografov a pripravovať svojich pokračovateľov v organizačných štruktúrach SŠDS. To, že či rovnako obetavých, nadšených, kreatívnych a erudovaných, ukáže až budúcnosť. Som názoru, že práve zásluhou Janka a Jozefa môžeme s hrdosťou hovoriť

o úspešnej a bohatej histórii SŠDS, o jej veľmi dobrej kondícii v súčasnosti a snád' o ďalšom pozitívnom rozvoji v budúcnosti.

Necelý rok po odchode Janka Luhu sme sa v októbri tohto roku navždy rozlúčili aj s Jozefom. Tradične sa hovorí o odchode do večnosti. Pri týchto slovách sa mnohí z nás zahľadia do neba a predstavujú si večnosť niekde ďaleko, mimo priestor ľudských pohľadov. Málokto si uvedomuje, že to naše, ľudské nekonečno je v nás, v našich srdciach, spomienkach a príbehoch.

Na Jozefom parte bolo – okrem iného – napísané: *„Pracoval vždy do úmoru, kľudu si nedoprial. Srdce jeho zlaté bolo, každému len blaho prial.“*

V tomto príspevku si máme pripomenúť najmä Jozefa. Ale pripomínať si Jozefa bez Janka nie je pre mňa veľmi predstaviteľné. Preto aj záver rozšírim zo spomienky, ktorú som pred rokom napísal v prípade Janka Luhu: *Jozef a Janko, prišli sme o povesť, o peniaze a možno aj o zdravie. Ale bolo to krásne a bola to veľká sranda.* Ďakujem Vám obom za možnosť byť pri tom.

Vladimír Úradníček
podpredseda SŠDS pre akademickú štatistiku

Moje prvé stretnutie so Slovenskou štatistickou a demografickou spoločnosťou pred viac ako dvadsiatimi rokmi sa viaže s menami Jozef Chajdiak a Ján Luha. Bol som vtedy čerstvým predsedom Štatistického úradu SR a táto nerozlučná dvojica prišla za mnou diskutovať o možnostiach pokračovania prípadne rozšírenia spolupráce Spoločnosti s novým vedením úradu. Rýchlo sme sa dohodli a po krátkej dobe som sa stal aj priamou súčasťou vedenia Spoločnosti. Bol to práve Jozef Chajdiak, v tom čase vedecký tajomník Spoločnosti, ktorý ma dokázal získať pre spoluprácu. Teším sa, že som jeho ponuku prijal a práve naša tesná vzájomná spolupráca mi pomáhala. Mal som možnosť spolupracovať s Jozefom Chajdiakom, časť z tých úctyhodných 25 rokov, ktoré bol vedeckým tajomníkom Spoločnosti, a môžem zodpovedne povedať, že významne prispel k rozvoju jej činnosti.

Nemal som možnosť preniknúť do všetkých odborných oblastí, v ktorých sa Jozef pohyboval, ale vysoko som si vážil jeho nasadenie a angažovanosť pri príprave odborných podujatí Spoločnosti. Bol som priamo pri tom, keď sa v jeho hlave zrodil nápad pozvať na pôdu Spoločnosti prognostické tímy, aby pravidelne prezentovali svoje predikcie vývoja ekonomiky. Tak sa pred dvadsiatimi rokmi zrodila konferencia Pohľady na ekonomiku Slovenska. Jozef bol dušou tohto podujatia, komunikoval s prednášateľmi, generoval nápady, ako spraviť podujatie príťažlivejším. To je len jeden z viacerých príkladov, lebo niektoré podujatia, ktoré Jozef zakladal, už vtedy bežali a iné vznikli neskôr.

Jozef Chajdiak síce nestál pri vzniku našej Spoločnosti, ale v jej strednom veku bol jedným z tých, ktorí ju formovali do dnešnej podoby. Zaslúži si našu vďaku a trvalú spomienku.

Peter Mach
podpredseda SŠDS pre medzinárodné vzťahy

Doc. Ing. Jozefa Chajdiaka, CSc. som prvýkrát stretol v marci 1990, keď som po rozhovore s bývalým spolupracovníkom a priateľom z Ústavu teórie merania SAV, Jankom Luhom, rozšíril rady členov novo premenovanej Slovenskej štatistickej a demografickej spoločnosti (ŠSDS) o skupinu matematických štatistikov. Jozef bol vždy vedúcou osobnosťou SŠDS. Spoločnosť bola jeho srdcovou záležitosťou, aj akousi príležitosťou pre realizáciu jeho predstáv, vízií šírenia a organizácie štatistiky na Slovensku. Od Jozefa som pochopil, že štatistika bola preňho predovšetkým nástrojom na podporu rozhodovania v ekonómii, demografii, ale rovnako aj vo všetkých spoločenských aj prírodných vedách, v medicíne, farmácii,

Jozef mal svoju jasnú predstavu o postavení, členení a obsahu samostatného študijného odboru Štatistika, rozdielneho od súčasných študijných odborov Ekonómia a manažment alebo Matematika. Absolvoval som s ním veľa rozhovorov so štatistikmi z rôznych oblastí. Žiaľ tento smer jeho úsilia už nestihol dotiahnuť do úspešného konca. Možno, že aj pre jeho zhoršujúci sa zdravotný stav. Pritom bol Jozef bytostne presvedčený, že Štatistika si miesto medzi študijnými odbormi plným právom zasluhuje.

Jozef bol činnorodý, aktívny a veľmi schopný organizátor vedeckých aj odborných akcií, seminárov, zameraných na ekonomiku (hlavne Slovenska), na moderné aplikácie štatistických a demografických metód v riadení ekonomiky a podnikov. Akýkoľvek problém z tejto oblasti by som šiel pokonzultovať práve s ním. Jeho knižky, ktoré vydával vo „svojom“ vydavateľstve STATIS odporúčam každému, kto chce „pričuchnúť“ k štatistike a riešiť základné, ale aj pokročilé štatistické problémy. S úsmevom mi hovoril, že keď sa pokúšal získať informácie o možnosti dosiahnutia titulu DrSc., nazvali jeho knižky samizdatmi.

Jozef bol zapálený učiteľ štatistiky. Nie tej matematickej štatistiky, ale štatistiky aplikovanej, najmä v ekonómii. Bol predstaviteľom, akýmsi novátorom, zavádzania modernej výpočtovej techniky do výučby aj do praxe, a tiež pri riešení štatistických úloh a ich aplikácií. Pritom túto výpočtovú techniku aj výborne ovládal. Bol veľmi kritický ku svojej práci, ale aj k práci svojich kolegov. Mal zmysel pre „osobitný“ humor, čo mi bolo veľmi blízke. Jožko, budeš nám veľmi chýbať. Česť Tvojej pamiatke.

Gejza Wimmer
Matematický ústav SAV



Publikačná etika a vyhlásenie k nekalým praktikám

Redakčná rada Forum Statisticum Slovaca (ďalej len RR FSS) sa v plnom rozsahu stotožňuje s princípmi etiky publikovania deklarovanými Výborom pre publikačnú etiku. RR FSS prijíma štandardy publikačnej etiky a zavádza opatrenia proti akejkoľvek publikácii s nekalými praktikami. Autori predkladajú svoje práce do časopisu na publikovanie s vyhlásením, že predložené práce sú príspevky autorov a neboli kopírované alebo plagizované z iných prác. Editori posudzujú rukopisy na základe ich intelektuálneho obsahu, bez ohľadu na rasu, pohlavie, sexuálnu orientáciu, náboženské vyznanie, etnický pôvod, štátnu príslušnosť alebo politickú filozofiu autorov. RR FSS očakáva, že všetky strany podieľajúce sa na FSS dodržia publikačnú etiku. RR FSS nebude tolerovať plagiátorstvo alebo iné neetické správanie a neuverejní rukopis, ktorý nespĺňa tieto normy.

Zodpovednosť autorov: Autori potvrdzujú, že ich rukopis je ich pôvodnou prácou, že nebol doteraz uverejnený v rovnakej podobe a zároveň nie je v súčasnej dobe podaný na uverejnenie inde. Autori musia bezodkladne oznámiť RR FSS všetky strety záujmov. Autori musia uviesť všetky zdroje použité pri tvorbe svojho rukopisu. Autori musia bezodkladne nahlásiť všetky chyby, ktoré objavia vo svojom rukopise RR FSS.

Zodpovednosť recenzentov: Recenzenti musia oznámiť RR FSS všetky strety záujmov. Recenzenti musia uchovávať informácie týkajúce sa rukopisu ako dôverné. Recenzent musí upozorniť predsedu RR FSS na informácie, ktoré môžu byť dôvodom na zamietnutie vydania rukopisu. Recenzenti posudzujú rukopisy len na základe ich intelektuálneho obsahu.

Zodpovednosť RR FSS: RR FSS musí uchovávať informácie o podaných rukopisoch ako dôverné. RR FSS musí posúdiť rukopisy len na základe ich intelektuálneho obsahu. RR FSS rozhoduje o zverejnení predložených rukopisov. RR FSS odmietne príspevok, ktorý nie je v súlade s požiadavkami etiky publikovania.



Publication ethics and malpractice statement

The Editorial Board of Forum Statisticum Slovaca (hereinafter abbreviated as EB/FSS) is fully associated with the principles of ethics of publication declared by the Committee of Publication Ethics. The EB/FSS is committed to upholding the highest standards of publication ethics and takes all possible measures against any publication malpractices. Authors submitting their works to the journal for publication as original articles attest that the submitted works represent their authors' contributions and have not been copied or plagiarized in whole or in part from other works. An editor at any time evaluate manuscripts for their intellectual content without regard to race, gender, sexual orientation, religious belief, ethnic origin, citizenship, or political philosophy of the authors. The EB/FSS expects all parties participating in the publication of Forum Statisticum Slovaca commit to these publication ethics. The EB/FSS does not tolerate plagiarism or other unethical behaviour and will remove any manuscript that does not meet these standards.

Author responsibilities: Authors certify that their manuscripts are their original work unpublished previously elsewhere in the same form and not currently being considered for publication elsewhere. Authors must notify the EB/FSS of any conflicts of interest without any delay. Authors must identify all sources used in the creation of their manuscript. Authors must report any errors they discover in their manuscript to the EB/FSS without any delay.

Reviewer responsibilities: Reviewers must notify the EB/FSS of any conflicts of interest. Reviewers must keep information pertaining to the manuscript confidential. Reviewers must bring to the attention of the head of EB/FSS any information that may be reason to reject publication of a manuscript. Reviewers must evaluate manuscripts only for their intellectual content.

Editorial responsibilities: The EB/FSS must keep information pertaining to submitted manuscripts confidential. The EB/FSS must evaluate manuscripts only for their intellectual content. The EB/FSS is responsible for making publication decisions for submitted manuscripts. EB/FSS refuses the manuscript, which is not in accordance with the requirements of publishing ethics.

Obsah / Table of contents*Odborné a vedecké články / Instructive and regular papers*

Tvorba atraktívnych statistických grafov v prostredí SAS Creating attractive statistical graphs in the SAS environment	
Roman Pavelka.....	1
Impact of population ageing on health services: a case study for Slovakia Vplyv starnutia populácie na zdravotné služby: prípadová štúdia pre Slovensko	
Richard Priesol.....	18
A proposal of a modified approach to classification evaluation Návrh modifikovaného prístupu k hodnotení klasifikácie	
Hana Řezanková	31
Trend definition as a support tool for managing a portfolio of investment strategies Definícia trendu ako podporný nástroj pre riadenie portfólia investičných stratégií	
Tomáš Virdzek, Peter Kubaška	44

Zo života SŠDS / From the life of SŠDS

Výpočtová štatistika 2019 [informácia o konaní udalosti SŠDS] Computational Statistics 2019 [information about an event by SŠDS]	
Iveta Štankovičová.....	57
Spomienka na doc. Ing. Jozefa Chajdiaka, CSc. (*18-03-1952 † 22/10/2019) [smútočný oznam] A memory of doc. Ing. Jozef Chajdiak, CSc. (*18-03-1952 † 22/10/2019) [obituary]	
Iveta Štankovičová, Marián Zajko, Ivan Brežina, Branislav Bleha, Jitka Langhamrová, Hana Řezanková, Mikuláš Cár, Jana Marková, Vladimír Úradníček, Peter Mach, Gejza Wimmer.....	61



Cena / price: 25 €
 Ročné predplatné / annual subscription: 50 €
 Publikované / published: December 2019