

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

MZDOVÉ RODOVÉ ROZDIELY A ICH VÝVOJ

DIPLOMOVÁ PRÁCA

2014

Bc. Mária ŽIŽÁKOVÁ

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

MZDOVÉ RODOVÉ ROZDIELY A ICH VÝVOJ

DIPLOMOVÁ PRÁCA

Študijný program: Ekonomická a finančná matematika
Študijný odbor: 9.1.9. Aplikovaná matematika
Školiace pracovisko: Katedra aplikovanej matematiky a štatistiky
Vedúci práce: Mgr. Ing. Michal Páleník, PhD.



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Mária Žižáková
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský

Názov: Mzdové rodové rozdiely a ich vývoj
Gender wage differences and their development

Cieľ: Cieľom diplomovej práce je kvantifikovať a popísať mzdové rodové rozdiely v ekonomike používajúc individuálne mikrodáta. Hlavný dôraz bude kladený na vývoj počas obdobia finančnej krízy a otázkou, či problémy ekonomiky prispeli k prehĺbeniu rozdielov alebo nie.

Vedúci: Mgr. Ing. Michal Páleník, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.
Dátum zadania: 25.01.2013

Dátum schválenia: 04.02.2013 prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

študent

vedúci práce

PodĎakovanie: Touto cestou sa chcem poďakovať svojmu vedúcemu diplomovej práce Mgr. Ing. Michalovi Páleníkovi, PhD. za ochotu, pomoc a podnetné pripomienky. Ďakujem tiež Pavlovi Hudákovi za pomoc s prekladom abstraktu, svojej rodine a priateľom za ich trpezlivosť a podporu.

Abstrakt v štátnom jazyku

ŽIŽÁKOVÁ, Mária: Mzdové rodové rozdiely a ich vývoj [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Mgr. Ing. Michal Páleník, PhD., Bratislava, 2014, 91 s.

Jedným z kľúčov k dosiahnutiu pokroku v oblasti znižovania mzdových rozdielov medzi mužmi a ženami je skúmanie vplyvu rôznych faktorov a zmien v ekonomike na mzdové rodové rozdiely. Naša diplomová práca je odozvou na stratégiu Európskej komisie, ktorá sa snaží o vyrovňovanie mzdových rozdielov v Európe a tiež reaguje na zvýšený záujem laickej verejnosti o danú problematiku. V práci sa zameriavame na analýzu vývoja mzdových rodových rozdielov na území Slovenska v období od roku 2005 do roku 2013 s dôrazom na obdobie krízy. Využívame pritom dáta zo Sociálnej poisťovne SR a teóriu spojenú s analýzou časových radov. Pomocou modelov ako ARIMA model alebo Holt-Wintersov model a na základe dát pred rokom 2009 vytvoríme predpovede na obdobie finančnej krízy. Následne naše predpovede porovnáme s reálnymi hodnotami mzdových rozdielov medzi mužmi a ženami. Okrem toho, že sa pozrieme na vývoj mzdových rodových rozdielov, priblížime si aj správanie miezd mužov a žien či zmeny v jednotlivých zárobkových skupinách. Uvedené analýzy najprv predvedieme na priemerných dátach a neskôr ich pomocou softvéru R aplikujeme na dáta v jednotlivých okresoch a vekových skupinách. Nakoniec sa pozrieme na súvislosti medzi jednotlivými skupinami pomocou zhlukovej analýzy.

Kľúčové slová: mzdové rodové rozdiely (MRR), analýza časových radov, predikcie, zhluková analýza

Abstract

ŽIŽÁKOVÁ, Mária: Gender wage differences and their development [Master thesis], University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Ing. Michal Páleník, PhD., Bratislava, 2014, 91 p.

One of the keys for reducing the difference of the wage gap between men and women is to explore the effects of various factors including the state of the economy on the gender pay gap. Our thesis topic is a response to the call from the European Commission regarding identifying a strategy seeking to eliminate the gender pay gap within Europe. Additionally, it serves as a response to the increasing public awareness of this disparity and its subsequent resolution. Our work analyzes the occurrence and development of gender pay gap in Slovakia during the period of economic crisis from 2005 to 2013. We utilized the data from the Social Insurance Agency of the Slovak Republic, applying the analysis of time series data for extracting meaningful statistics, utilizing models such as ARIMA and Holt - Winters model, based on data from before 2009, we will create the forecasts to be applied for the period of the economic crisis. Subsequently, we will compare our predictions with the current real wage differences between men and women. Not only will we evaluate the evolution of the gender pay gap, but we will also evaluate the behavior of wages of men and women or of various income groups. Firstly we will perform analysis on the average data, secondly we will utilize software R to analyze the various regions and different age groups. Lastly, we will perform cluster analysis, evaluating the relevance of the data differences in the regions and age groups.

Keywords: gender pay gap (GPG), time series analysis, prediction, cluster analysis

Obsah

Zoznam obrázkov	10
Zoznam tabuliek	11
Úvod	12
1 Úvod do problematiky	14
1.1 Čo rozumieme pod pojmom MRR?	15
1.2 Príčiny mzdových rodových rozdielov	16
1.3 Výhody znižovania mzdového rozdielu	18
1.4 Iniciatívy proti mzdovým rozdielom	19
1.5 Situácia na Slovensku	21
1.5.1 Závery z dostupných štatistických údajov	24
2 Ekonometrické metódy	25
2.1 Lineárna a kvantilová regresia	25
2.2 Dekompozičné - rozkladové metódy	26
3 Časové rady a ich analýza	29
3.1 Kľúčové pojmy	29
3.1.1 Stacionarita a ergodicita procesu	29
3.1.2 Autokorelačná a parciálna autokorelačná funkcia	31
3.1.3 Biely šum	33
3.1.4 Lineárny proces a Woldova dekompozícia	34
3.1.5 Predpovede časových radov	35
3.2 Dekompozícia časových radov	35
3.2.1 Trend v časových radoch	36
3.2.2 Sezónosť v časových radoch	37
3.3 Stacionárne procesy	38
3.3.1 Autoregresný proces	39
3.3.2 Proces kľavých priemerov	40
3.3.3 Zmiešaný proces	41

3.4	Nestacionárne procesy	42
3.4.1	Integrovaný zmiešaný proces ARIMA	42
3.4.2	Sezónne ARIMA procesy	43
3.4.3	Výstavba modelu	43
3.4.4	Vytváranie predpovedí v ARIMA modeloch	47
4	Analýza - 1. časť	48
4.1	Popis využívaných dát	48
4.2	Popis využívaného programu R	48
4.3	Modelovanie zvolených časových radov	49
4.3.1	Priemerné MRR	49
4.3.2	Priemerné mzdy mužov a žien	54
4.3.3	Analýza v jednotlivých platových skupinách	57
4.4	Zhrnutie získaných výsledkov	59
5	Analýza - 2. časť	61
5.1	Deskriptívna analýza	62
5.2	Zhluková analýza	63
5.2.1	Korelačná analýza	63
5.2.2	Štandardizácia dát	65
5.2.3	Metóda k-priemery	65
5.2.4	Výsledky zhlukovej analýzy pre okresy	67
6	Analýza - 3. časť	70
6.1	Deskriptívna analýza	70
6.1.1	Výsledky zhlukovej analýzy pre okresy a vekové skupiny	71
6.1.2	Zhrnutie	74
	Záver	75
	ZOZNAM POUŽITEJ LITERATÚRY	77
	Príloha A - Zákony týkajúce sa diskriminácie	82
	Príloha B - Príkazy v programe R	84

Príloha C - Tabuľka výsledkov zhlukovej analýzy pre okresy	90
Príloha D - Prílohy na CD	90

Zoznam obrázkov

1	Mzdové rodové rozdiely v SR a priemerné mzdové rodové rozdiely v krajinách EÚ	21
2	Mzdové rodové rozdiely na Slovensku podľa veku.	23
3	Mzdové rodové rozdiely na Slovensku podľa SK ISCO.	23
4	Mzdové rodové rozdiely na Slovensku podľa dosiahnutého vzdelania. . .	24
5	Postup výberu modelu ARIMA.	45
6	Vývoj priemerných mzdových rozdielov na Slovensku v období 2005 - 2008	50
7	ACF a PACF sezónne diferencovaného radu MRR	51
8	Predpoveď ex-post priemerných MRR na základe modelu SARIMA(0, 0, 0) × (1, 0, 0) ₁₂	53
9	Odhad vývoja priemerných MRR na základe modelu SARIMA(0, 0, 0) × (1, 0, 0) ₁₂	54
10	Vývoj priemernej mzdy muža a odhadnuté lineárne trendové priamky. .	55
11	Vývoj priemernej mzdy ženy a odhadnuté lineárne trendové priamky. .	55
12	Priemerné počty všetkých mužov (čiernou) a žien (červenou).	57
13	Priemerné počty mužov (čiernou), žien (červenou) zarábajúcich od 300 do 3000 eur.	57
14	Priemerné počty mužov (čiernou), žien (červenou) zarábajúcich nad 3000 eur.	58
15	Priemerné počty mužov (čiernou), žien (červenou) zarábajúcich pod 300 eur.	58
16	Závislosť funkcie vzdialeností od počtu zhlukov pre okresy.	66
17	Mapa Slovenska vyfarbená podľa priemerných MRR okresov (modrá - najmenšie MRR, červená - najväčšie MRR).	68
18	Mapa Slovenska vyfarbená podľa zatriedenia do zhlukov (modrá - najnižšia hodnota centroidu MRR, červená - najvyššia hodnota centroidu MRR).	68
19	Mapa Slovenska vyfarbená podľa priemernej zmeny MRR okresov (modrá - najväčší pokles MRR, červená - najväčší nárast MRR).	69

20	Mapa Slovenska vyfarbená podľa zatriedenia do zhlukov (modrá - najzápornejšia hodnota centroidu zmeny MRR, červená - najkladnejšia hodnota centroidu zmeny MRR).	69
21	Závislosť funkcie vzdialeností od počtu zhlukov pre okresy a vekové skupiny.	72
22	Tabuľka hodnôt centroidov, a priemerného veku a okresu v ôsmich zhlukoch.	73
A.23	Tabuľka vybraných charakteristík pre okresy.	91

Zoznam tabuliek

1	Mzdové rodové rozdiely v SR pre roky 2001 - 2012.	21
2	Mzdové rodové rozdiely v SR pre roky 2005 - 2012.	22
3	Tvar autokorelačnej a parciálnej autokorelačnej funkcie procesov AR(p), MA(q) a ARMA(p,q)	44
4	Výsledky jednotlivých ukazovateľov ako MRR, vývoja platov a počtov osôb v rôznych platových skupinách.	60
6	Tabuľka korelácií medzi jednotlivými charakteristikami okresov.	64

Úvod

Rodová nerovnosť je stále pretrvávajúcim fenoménom, ktorého dôsledky sa dotýkajú rôznych oblastí nášho života. Aj napriek pozitívnemu trendu je stále viditeľná a zrejme aj na Slovensku. Keďže rodová nerovnosť nie je prejavom spravodlivej spoločnosti, podnikajú sa kroky na jej znižovanie, či už na úrovni Európskej únie, ale aj priamo na Slovensku. Okrem spravodlivejšej spoločnosti stratégie európskej Komisie pre rovnosť mužov a žien kladú dôraz aj na prínos rodovej rovnosti k hospodárskemu rastu a trvalo udržateľnému rozvoju. Jedným z hlavných bodov týchto stratégií je vyrovňovanie mzdových rozdielov medzi mužmi a ženami. Kľúčom k dosiahnutiu pokroku v tejto oblasti je aj skúmanie vplyvu rôznych faktorov a zmien v ekonomike na mzdové rodové rozdiely.

Existuje množstvo publikácií, ktoré sa zameriavajú na analyzovanie vplyvu špecifík zamestnancov, ako sú dosiahnutý stupeň vzdelania alebo typ povolania na výšku miezd mužov a žien. Rôznymi metódami rozdeľujú mzdové rodové rozdiely na časť, ktorá je vysvetlená na základe rozdielnych charakteristík, ktorými môžeme opísať ženy a mužov a na časť, ktorú označujeme ako diskrimináciu. Tieto analýzy však nič nehovoria o tom, ako sa reálne vyvíjali mzdy mužov a žien či veľkosť mzdových rodových rozdielov za určité dlhšie obdobie. V našej diplomovej práci sa zameriame práve na vývoj mzdových rodových rozdielov v období od roku 2005 do roku 2013. Keďže v tomto období bola Slovenská ekonomika postihnutá finančnou krízou, môžeme očakávať určité zmeny vo vývoji miezd a v zamestnanosti, a tým aj zmeny v mzdových rodových rozdieloch. Na základe analýzy časových radov sa pokúsime tieto zmeny, vzniknuté na Slovensku v období finančnej krízy, kvantifikovať a porovnať v jednotlivých okresoch či vekových skupinách využitím zhlukovej analýzy.

V prvej kapitole uvedieme teóriu spojenú s mzdovými rozdielmi medzi mužmi a ženami. Bližšie sa pozrieme na ich príčiny a dôsledky, dôležitosť danej témy, ale aj na štatistické výsledky, ktoré sú pre verejnosť dostupné. Druhá kapitola je zameraná na prehľad regresných a dekompozičných metód, ktoré sa v praxi používajú na odhad vplyvu pohlavia na výšku mzdy a na vysvetlenie mzdových rozdielov na základe rozdielnych charakteristík mužov a žien. Tento prehľad je informatívny a v našej práci tieto metódy nebudeme ďalej využívať. V tretej kapitole zhrnieme teóriu využívanú

pri analýze časových radov so zameraním na základné charakteristiky časových radov, dekompozíciu časových radov, Holt-Wintersovu metódu a Box-Jenkinsovú metodológiu. Táto teória poskytuje aparát, na základe ktorého budeme skúmať vývoj mzdových rodových rozdielov pred a počas hospodárskej krízy na Slovensku a vývoj miezd.

Ďalšia časť diplomovej práce - štvrtá, piata a šiesta kapitola budú praktické. V prvej časti štvrtej kapitoly opíšeme využívané dáta. Následne opíšeme zmeny priemerných mzdových rodových rozdielov, pričom budeme porovnávať reálnu situáciu, ktorá nastala počas krízy, s nami očakávaným vývojom týchto rozdielov, odhadnutým na základe dát pred krízou a vybraného modelu pre daný časový rad. Podobne budeme analyzovať aj vývoj priemernej mzdy mužov a žien na základe odhadnutia trendu vývoja miezd pred krízou a v období krízy. Nakoniec rozdelíme dáta v každom mesiaci na tri skupiny podľa zárobku jednotlivých zamestnancov a pozrieme sa na to, ako sa vývoj líšil v jednotlivých skupinách. Keďže analýza priemerných hodnôt nie je vždy zhodná s vývojom v niektorých menších skupinách vzorky, v predposlednej kapitole naše postupy aplikujeme na časové rady priemerných miezd a priemerných mzdových rodových rozdielov v rôznych okresoch Slovenska. Pomocou zhlukovej analýzy budeme hľadať podobnosti v správaní sa vybraných charakteristík jednotlivých okresov. V šiestej kapitole budeme postupovať rovnako ako v piatej, s tým rozdielom, že hodnoty v jednotlivých okresoch navyše rozdelíme aj podľa vekových skupín. Všetky uvedené výpočty a analýzy budú konštruované vo voľne dostupnom softvéri R. Obsahom práce sú aj prílohy, v ktorých čitateľ nájde všetky využívané príkazy aj s komentármi (Príloha B) a tiež CD s používanými skriptami z programu R a výslednými tabuľkami v programe MS Excel ako Príloha D.

1 Úvod do problematiky

V posledných desaťročiach sa stále častejšie rozpráva o rozdielnom ohodnocovaní práce žien a mužov. Táto téma je diskutovaná na rôznych internetových stránkach či televíznych kanáloch a často vyvoláva búrlivé debaty v pracovných kolektívoch. Veď kto by sa nestretol s nápismi na titulných stranách novín či časopisov, ktoré by sa týkali presne tohoto problému? Napríklad: "Rozdiel medzi platmi mužov a žien je 21% "[32] alebo "Súboj pohlaví: Ženy nezarábajú menej ako muži"[37].

Okrem príspevkov laickej verejnosti bolo na Slovensku i v zahraničí napísaných množstvo odborných štúdií a publikácií, v ktorých sa autori snažia rôznymi metódami kvantifikovať mzdové rozdiely medzi ženami a mužmi, či vysvetliť príčiny vzniku týchto rozdielov, ukázať potrebu ich znižovania a možné riešenia problému mzdovej rodovej nerovnosti.

Literatúru, týkajúcu sa danej problematiky a využívanú v našej práci, môžeme rozdeliť na tri väčšie skupiny:

- Prvou skupinou sú publikácie, ktoré obsahujú štúdie a deskriptívne analýzy o rodových nerovnostiach. Venujú sa ich príčinám, návrhom na zlepšenie a podávajú informácie o súčasnom stave problematiky.

Jedným z hlavných cieľov Európskej stratégie pre rovnosť medzi mužmi a ženami na obdobie 2010 - 2015 [14] je znižovanie mzdovej priepasti medzi pohlaviami. Je teda pochopiteľné, že sa orgány Európskej únie touto otázkou zaoberajú a vydali množstvo publikácií, či už všeobecne informujúcich o tomto probléme [37], [34], ale aj takých, ktoré konkrétne predstavujú možnosti, ako sa s príčinami stále veľkého mzdového rodového rozdielu vysporiadať [14], [7], [16].

Na Slovensku sa danou problematikou začali podrobnejšie zaoberať autori v období po vstupe Slovenska do Európskej únie. Viaceré práce, ktoré uvedieme, sú teda ovplyvnené práve vyššie spomenutými dokumentmi Európskej komisie. V práci vydanéj Záujmovým združením žien v roku 2002 - Na ceste do Európskej únie [9] je venovaná jedna kapitola práve ženám na trhu práce. Podobných prác bolo vytvorených v priebehu ďalších rokov viacero [2], [31], [26], [8].

Dôležité je spomenúť, že sa problematike rodovej nerovnosti venuje aj Štatistický

úrad Slovenskej republiky. Počnúc rokom 2010 každoročne vydáva publikácie, ktorých cieľom je priblížiť stav rodovej rovnosti na Slovensku a poskytnúť ľuďom dostupné informácie, charakterizujúce postavenie žien v našej spoločnosti [35]. Ďalej je spracovávaná súhrnná správa o stave rodovej rovnosti na Slovensku, ktorá hovorí o situácii žien a mužov počas obdobia hospodárskej a dlhovej krízy [18].

Cieľom prvej kapitoly je teda:

1. Na základe spomínanej literatúry zoznámiť čitateľa s pojmom mzdová rodová nerovnosť, predstaviť jej príčiny, poukázať na potrebu zmenšovania priepasti v ohodnocovaní medzi mužmi a ženami a ukázať možnosti ako proti nej bojovať.
 2. Predstaviť verejnosti dostupné štatistiky a na základe nich ponúknuť hypotézy, aký vplyv mala na mzdové rodové rozdiely na Slovensku finančná kríza. Týmto hypotézam sa budeme podrobnejšie venovať v ďalších kapitolách a skúmať ich na našich dátach.
- Existuje množstvo literatúry, ktorá ponúka ekonometrické modely, vhodné na skúmanie príjmových nerovností medzi mužmi a ženami. Sú to rôzne regresné metódy alebo dekompozičné metódy. Pomocou nich a na základe štatistických dát autori odhadujú vplyv jednotlivých faktorov na veľkosť mzdových rodových rozdielov. Metódam a prehľadu existujúcej literatúry sa budeme podrobnejšie venovať v kapitole 2.
 - Podstatnou časťou literatúry, z ktorej budeme čerpať, sú publikácie, ponúkajúce teóriu k analýze časových radov. Na základe tejto teórie a vybraných údajov budeme analyzovať vývoj mzdových rodových rozdielov a miezd mužov a žien. Tejto teórii bude venovaná kapitola 3.

1.1 Čo rozumieme pod pojmom mzdový rodový rozdiel?

Najčastejšie vyskytujúcim sa slovným spojením v našej diplomovej práci bude pojem mzdový rodový rozdiel (MRR), v anglickom jazyku gender pay gap (GPG). V praxi rozlišujeme medzi absolútnym a relatívnym MRR. Absolútny MRR budeme chápať

ako rozdiel medzi priemernou hrubou mzdou mužov a priemernou hrubou mzdou žien. Relatívny MRR je vyjadrený ako podiel absolútneho MRR a priemernej hrubej mzdy mužov v percentách (v diplomovej práci budeme ďalej pod pojmom MRR myslieť relatívny mzdový rodový rozdiel).

Teda:

$$MRR = \frac{(\text{mzda muža} - \text{mzda ženy})}{(\text{mzda muža})} * 100 \quad (1)$$

Zaujímavé je tiež spomenúť že pod pojmom neupravený MRR myslíme rozdiel medzi mzdami mužov a žien bez ohľadu na rôzne vplyvajúce faktory, ktoré môžu značne ovplyvniť postavenie žien v pracovnom kolektíve. Týmito faktormi môžu byť napríklad rozdielne vzdelanie, počet odpracovaných rokov, prax, typ práce, počet detí a iné. Ale aj keď tieto faktory zvážime, veľká časť mzdového rozdielu môže ostať nevysvetlená.

1.2 Príčiny mzdových rodových rozdielov

Mzdová rodová nerovnosť je vysvetľovaná niekoľkými významnými faktormi, ktoré spolu úzko súvisia.

- Diskriminácia v odmeňovaní na pracoviskách

- Priama diskriminácia

Pod priamou diskrimináciou rozumieme iné ohodnotenie žien a mužov za rovnakú prácu. Priama diskriminácia sa vyskytuje menej často vďaka obmedzeniam právnymi predpismi. Avšak na základne prieskumu uvedeného v [34] si v roku 2008 myslelo stále 36% opýtaných ľudí žijúcich v EÚ, že diskriminácia podľa pohlavia je veľmi rozšíreným javom. (Právne predpisy, ktoré obmedzujú diskrimináciu žien na pracoviskách, sú spomínané neskôr v podkapitole 1.3.)

- Nepriama diskriminácia - podceňovanie práce žien

Pod nepriamou diskrimináciou rozumieme iné ohodnotenie žien a mužov za prácu rovnakej hodnoty. Práca rovnakej hodnoty sa môže úplne líšiť od práce, ku ktorej je prirovnávaná, ale na základe metód ohodnocovania práce

má rovnakú hodnotu. Poznáme rôzne metódy hodnotenia práce, niektoré z nich sú uvedené v [3]. Príkladom nepriamej diskriminácie môže byť nižší plat upratovačky v obchodnom dome ako dokladača tovaru s lepšou fyzickou silou. Častokrát je tiež lepšie ohodnotená práca muža, ktorý robí s kapitálom, ako ženy, ktorá sa stará o ľudí.

- Rodová segregácia, stereotypy a tradície

Napriek tomu, že sa pomer žien a mužov na školách s technickým zameraním pomaly vyrovnáva, stále možno vidieť prevahu mužov na oddeleniach informačných technológií či technických oddeleniach, v kontraste s obsadenosťou pracovných miest so sociálnym zameraním prevažne ženami.

Muži sú tiež stále považovaní za živiteľov rodiny, preto sú často lepšie ohodnocovaní, aj keď v súčasnej dobe je veľa prípadov, kedy zabezpečuje chod domácnosti žena aj po finančnej stránke.

- Zosúladenie pracovného a súkromného života zamestnancov

V živote ženy je prirodzenou úlohou materstvo. Žena si často volí úlohu matky pred kariérnym postupom. Túto skutočnosť zohľadňujú aj zamestnávateľia, ktorí zväčša automaticky počítajú s tým, že mladá žena časom prevezme zodpovednosť za rodinu a deti. Bude čerpať rodičovskú dovolenku alebo ostane doma s chorým dieťaťom či starým rodičom. Práve na základe tohto dôvodu zamestnávateľia preferujú prijať do zamestnania muža.

- Málo žien na vedúcich pozíciách

Toto tvrdenie vyplýva z rôznych štatistických výsledkov. Napríklad v práci [31] hovorí autorka o probléme takzvaného sklenného stropu. Tento strop, podľa nej, vymedzuje množstvo žien ktoré dosahujú najvyššie pozície a to aj v typicky ženských sektoroch. Uvádza tiež príklady vysoko ohodnotených zamestnaní, v ktorých sa ukazuje nízky podiel ženskej pracovnej sily, čo vedie k veľkému MRR. Príkladmi zamestnaní s najvyšším rozdielom v obsadenosti pracovných pozícií mužmi a ženami sú: leteckí dispečeri/ky, operátori/ky leteckej stanice, riaditelia/ky a prezidenti/ky veľkých organizácií, vedúci/e, riaditelia/ky v pod-

nikoch poskytujúcich obchod a služby, odborní zamestnanci/kyne v bankovníctve a peňažníctve.

- Väčšia ochota žien vziať aj nízko platené miesto

"Súčasná hospodárska kríza len posilňuje pretrvávajúce tohto stereotypu." [2]

1.3 Výhody znižovania mzdových rodových rozdielov

Barošová vo svojej prednáške týkajúcej sa rodových mzdových rozdielov na trhu práce [2] tvrdí: "Bojovať proti mzdovým rodovým rozdielom je dôležité nielen z dôvodov vytvorenia spravodlivejšej (rovnejšej) spoločnosti, ale aj z podnikateľského hľadiska. Odstraňovanie rozdielov v odmeňovaní medzi ženami a mužmi prináša totiž výhody tak zamestnávateľom, ako aj zamestnancom."

Skutočne by sa zníženie mzdového rozdielu medzi mužmi a ženami mohlo podieľať na vylepšení ekonomickej situácie na Slovensku a v EÚ? O akých výhodách pre zamestnancov či zamestnávateľov hovorí? Pre koho by mohlo byť zníženie MRR pozitívnou zmenou?

- Pre spoločnosť

Prínosom je vytvorenie rovnoprávnejšej spoločnosti, v ktorej by ženy mali nárok na rovnaké platové ohodnotenie ako muži. Tým pádom by sa zamedzilo nebezpečenstvu chudoby žien počas ich pracovného života, ale tiež po odchode do dôchodku.

- Pre firmy a zamestnancov

- Zvýšením súťaživosti v podniku.
- Vytváraním kvalitného pracovného prostredia, kde sú všetci zamestnanci hodnotení podľa ich prínosu spoločnosti a nie podľa rodových preferencií.
- Získavaním dôvery zamestnancov.
- Podávaním lepšieho obrazu o firme smerom k verejnosti.

- Pre ekonomiku či hospodársky rast

Ženský potenciál by mohol byť základom pre rast a oživenie ekonomiky. Ich zručnosti sú dôležitý faktor pre ľudský pokrok a ich nevyužívaním len strácame dôležitý zdroj. Lepšie ohodnotenie žien je dôležitou ekonomickou investíciou do budúcnosti a môže byť kľúčom k dosiahnutiu rastu zamestnanosti.

1.4 Iniciatívy proti mzdovým rodovým rozdielom

Rodová rovnosť je jedným zo základných zásad rovnosti medzi ľuďmi. Poznáme viaceré medzinárodné dokumenty, v ktorých je požiadavka nediskriminácie hlboko zakotvená. Príkladmi sú: Všeobecná deklarácia ľudských práv, Dohovor o odstránení všetkých foriem diskriminácie žien či Charta základných práv Európskej únie. Štáty vydávajú ústavy a zákony, podľa ktorých sú si všetci ľudia rovní. Hovorí o tom aj ústava Slovenskej republiky [38] v dvanástom článku v druhom odseku:

"Základné práva a slobody sa zaručujú na území Slovenskej republiky všetkým bez ohľadu na pohlavie, rasu, farbu pleti, jazyk, vieru a náboženstvo, politické, či iné zmýšľanie, národný alebo sociálny pôvod, príslušnosť k národnosti alebo etnickej skupine, majetok, rod alebo iné postavenie. Nikoho nemožno z týchto dôvodov poškodzovať, zvyhodňovať alebo znevýhodňovať." Legislatíva tiež v roku 2004 vydala tzv. antidiskriminačný zákon [40]. Tento zákon v súlade so zásadou rovnakého zaobchádzania zakazuje v pracovnoprávných vzťahoch diskrimináciu osôb z dôvodu ich pohlavia. S antidiskriminačným zákonom úzko súvisia aj niektoré časti zákonníka práce [39], kde sa hovorí o povinnosti zamestnávateľa zaobchádzať so zamestnancami v súlade so zásadou rovnakého zaobchádzania. Zakazuje diskrimináciu z dôvodu manželského či rodinného stavu, rodu alebo iného postavenia (kompletné znenie Antidiskriminačného zákona § 6 a Zákonníka práce § 13 v Prílohe A). Teda nie je pochyb o tom, že všetky civilizované krajiny zamietajú diskrimináciu a to aj v pracovnoprávných vzťahoch. Ak však chceme znižovať rozdiely medzi mzdami mužov a žien, treba urobiť aj iné ako právne opatrenia.

Európska komisia je hlavnou aktérkou v boji pre znižovanie mzdových rodových rozdielov. Vo svojej stratégii rovnosti žien a mužov 2010 - 2015 [14], ktorá nadväzuje na plán uplatňovania rovnosti žien a mužov 2006 - 2010, sa zaoberá piatimi bodmi a ponúka rôzne opatrenia na pokrok v týchto oblastiach. Druhým zo spomínaných

bodov je "Rovnaká odmena za rovnakú prácu a prácu rovnakej hodnoty." Kľúčom k reálnemu pokroku v znižovaní mzdovej rodovej nerovnosti sú podľa Európskej komisie tieto opatrenia:

- Lepšia informovanosť ľudí o mzdových rodových rozdieloch.
- Podporovanie iniciatívy rovnakého odmeňovania na pracoviskách.
- Podporovanie žien pri výbere netradičných profesií.
- Skúmanie vplyvu faktorov na mzdové rodové rozdiely.

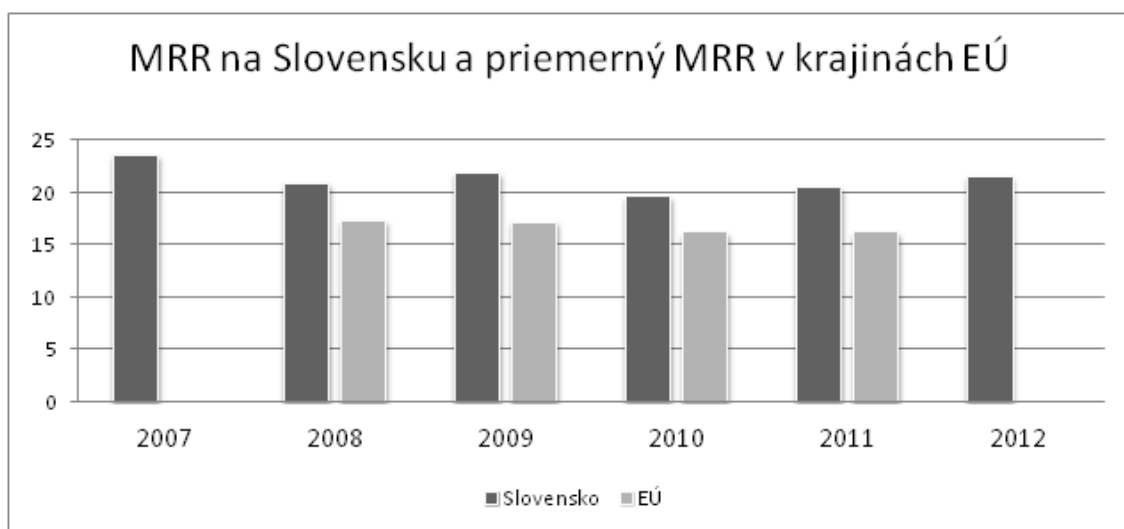
Ďalšími diskutovanými opatreniami sú takzvané "Kvóty na ženy". Tieto podľa určitých pravidiel a pod hrozbou sankcií určujú, aký podiel žien na rôznych pracovných pozíciách vo firme musí pracovať. Príkladom je návrh nového pravidla v EU, teda 40 percentný podiel žien na postoch nevýkonných riadiacich pracovníkov vo verejne obchodovateľných spoločnostiach v Európe a to do roku 2020. Toto pravidlo má zvýšiť šance žien na vyššie ohodnotenú pracovnú miesta a rozbiť takzvaný sklený kariérny strop pre ženy.

Aj na vnútroštátnych úrovniach fungujú opatrenia, ktoré sa snažia zmenšiť mzdové nerovnosti medzi mužmi a ženami. Príkladom je povinnosť zamestnávateľov zverejňovať resp. neutaľovať platy zamestnancov či dni zamerané na informovanie verejnosti o mzdových rodových rozdieloch. Niektoré členské štáty Európskej únie pristúpili na zavedenie tzv. "otcovských kvót", ktoré stanovujú počet mesiacov rodičovskej dovolenky, ktoré môže čerpať iba otec dieťaťa.

Otázkou teda ostáva, či sa rôznymi opatreniami mzdové nerovnosti vôbec dajú výraznejšie znížiť alebo úplne odstrániť. Názory veľkej časti laickej verejnosti sa zhodujú v tom, že žena má ostať ženou, k čomu patrí starostlivosť o deti a rodinu a iné výsady, ktoré sú s tým spojené. Často preto nemôže v takej miere zastupovať muža na pracovných pozíciách, ktoré vyžadujú neustálu angažovanosť v odbore. V práci [21] autor uvádza názor, že typicky ženské, nižšie platené povolania môžu mať výhody nemzdového charakteru, ako napríklad pružný pracovný čas. Rozdielna koncentrácia žien v jednotlivých typoch zamestnaní môže byť teda spôsobovaná rozdielnymi, pre rod špecifickými preferenciami a voľbami.

1.5 Situácia na Slovensku

Podľa údajov Štatistického úradu Európskych spoločností - Eurostatu [21] patrí Slovensko od roku 2007 medzi krajiny s najväčším MRR. Najhoršia situácia je v Estónsku, nasleduje Rukúsko, Česká republika, Nemecko a na piatom mieste je Slovenská republika s MRR cca. 21 %, čo je o 3 - 5 percentuálnych bodov viac, ako je priemer v krajinách EÚ (viď Obr. 1).



Obr. 1: Mzdové rodové rozdiely v SR a priemerné mzdové rodové rozdiely v krajinách EÚ zo štatistických údajov uverejnených na stránke EUROSTATu [42].

V Tabuľke 1 môžeme vidieť údaje o MRR vypočítané z priemerných hrubých hodinových miezd mužov a žien na Slovensku za posledných 12 rokov (s výnimkou roku 2006 - chýbajúci údaj). Podľa tejto tabuľky sa nám zdá, že MRR v období finančnej krízy začali klesať a ustálili sa na hodnote cca 21 %, v porovnaní s obdobím pred krízou, kedy sa táto hodnota pohybovala okolo 26 %. To by znamenalo nezanedbateľný pokles MRR na Slovensku v období krízy zhruba o 5 percentuálnych bodov.

Tabuľka 1: Mzdové rodové rozdiely v SR pre roky 2001 - 2005 zdroj: [5], pre roky 2007 - 2012 zdroj: [42].

2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
26,5	26,0	25,9	26,1	26,7		23,6	20,9	21,9	19,6	20,5	21,5

Pravdivosť tohto tvrdenia však môže byť ohrozená rôznymi zdrojmi dát do roku 2005 a od roku 2007. Rôzne zdroje môžu napríklad používať na vytváranie štatistík rôzne veľké vzorky zamestnaných osôb. Ďalej môže byť problém v rozdielnom vypočítavaní priemernej mzdy. V [5] využívajú autori údaje zo Štatistického zisťovania o cene práce (ISCP) a priemerná mzda je počítaná iba z 3. kvartálu, v ktorom je najmenšia pravdepodobnosť ovplyvnenia platu odmenami alebo 13. platom. Eurostat využíva mikrodáta SES (Structure of Earnings Survey) a nezahŕňa do výpočtu priemernej mzdy nepravidelné prémie, odmeny a iné mzdové zložky.

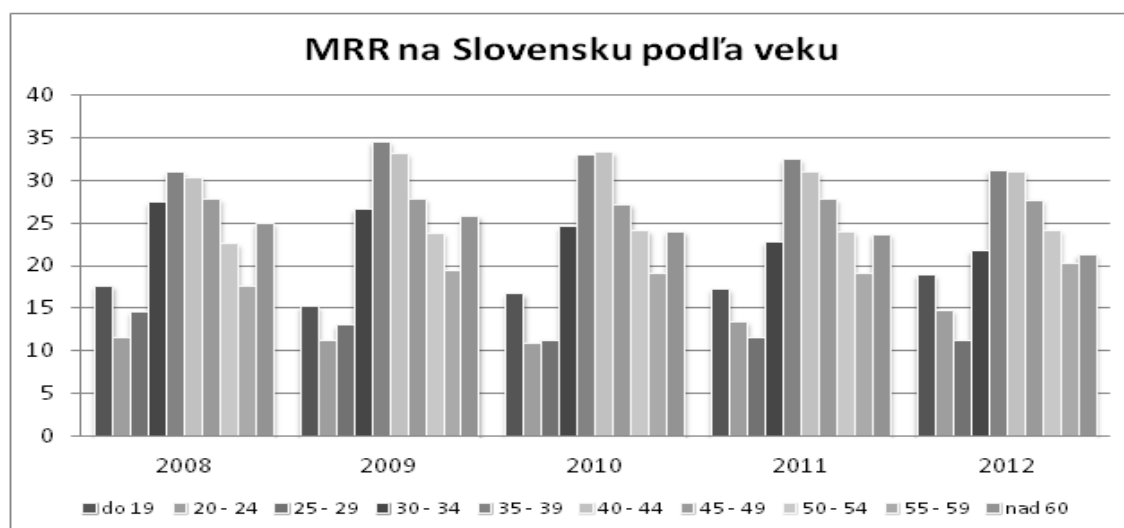
V Tabuľke 2 môžeme vidieť údaje o MRR, vypočítané z priemerných hrubých mesačných miezd mužov a žien na Slovensku za posledných 8 rokov, ktorých zdrojom je Štatistický úrad SR [43]. Údaje nás podporujú v predpoklade, že v období finančnej krízy klesli MRR v tomto prípade o cca 2 - 3 percentuálne body.

Tabuľka 2: Mzdové rodové rozdiely v SR z priemernej hrubej mesačnej mzdy pre roky 2005 - 2012 zdroj [43].

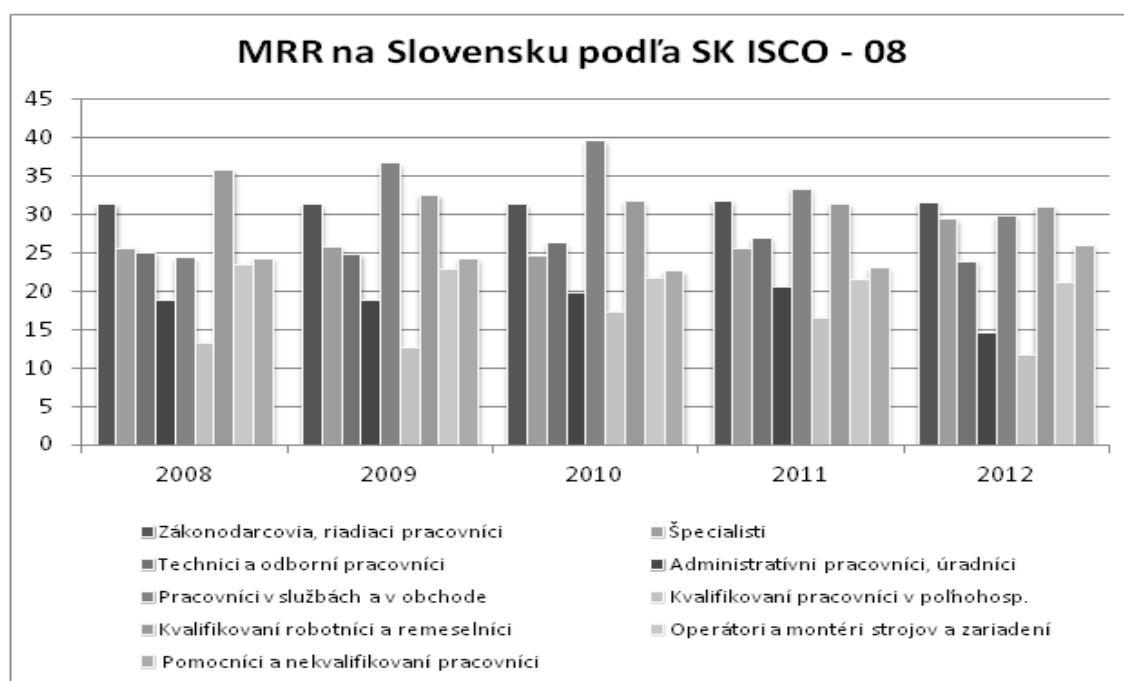
2005	2006	2007	2008	2009	2010	2011	2012
28,4	26,9	25,8	24,0	25,3	24,7	24,2	24,1

Na Obr. 2, Obr. 3 a Obr. 4 môžeme vidieť MRR vypočítané z priemernej hrubej mesačnej mzdy, rozdelené na 10 skupín podľa veku, na 9 skupín podľa odvetvového členenia SK ISCO - 08 a na 10 skupín podľa dosiahnutého vzdelania. Je vidno, že MRR sa od roku 2008 menili s pribúdajúcimi rokmi v jednotlivých skupinách len mierne. Napríklad z vekových skupín sa najviac znížil MRR v skupine 30 - 34 ročných a to o 5,6 percentuálnych bodov, ostatné mzdové rozdiely sa menili s maximálnym rozdielom MRR priemerne o 3 percentuálne body. Väčšie rozdiely vidno medzi jednotlivými skupinami. Zaujímavé je všimnúť si výrazne nižší MRR pri mladších 20 - 29 ročných zamestnancoch v kontraste s 30 - 49 ročnými zamestnancami. Na nízkych MRR 20 - 29 ročnej skupiny a postupnom znižovaní MRR v nasledujúcej skupine 30 - 34 ročných, môžeme vidieť pozitívny trend postupného znižovania MRR medzi novou pracovnou silou. Pri výpočte MRR uvádzaných na Obr. 2, Obr. 3 a Obr. 4 sme vychádzali z výsledkov výberového Štatistického zisťovania o štruktúre miezd, ktoré boli uverejnené

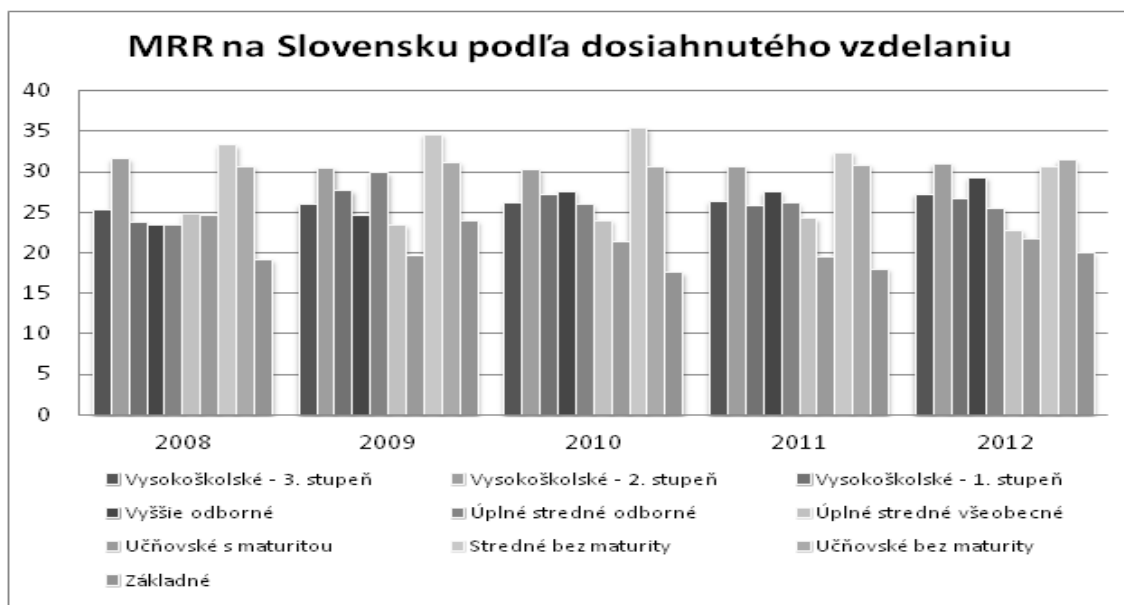
v [35]. Použili sme priemernú hrubú mesačnú mzdu jednotlivých skupín a podľa vzorca (1) sme dospeli k uvedeným výsledkom.



Obr. 2: Mzdové rodové rozdiely na Slovensku podľa veku zdroj: [35] a vlastné výpočty.



Obr. 3: Mzdové rodové rozdiely na Slovensku podľa SK ISCO - 08 zdroj: [35] a vlastné výpočty.



Obr. 4: Mzdové rodové rozdiely na Slovensku podľa dosiahnutého vzdelania zdroj: [35] a vlastné výpočty.

1.5.1 Závěry vychádzajúce zo štatistických údajov pre Slovensko

1. Je opodstatnené na Slovensku hovoriť o vysokom MRR vzhľadom ku krajinám EÚ. Výrazný platový rozdiel sa ukázal aj pri delení zamestnancov na rôzne skupiny vo väčšine z nich.
2. Z údajov v Tabuľkách 1 a 2 môžeme predpokladať pokles MRR na Slovensku v období krízy.
3. Vidno nižšie mzdové rozdiely medzi novou generáciou zamestnaných mužov a žien. Tento trend sa teda môže s pribúdajúcimi rokmi dostávať aj do ďalších vekových skupín. Vytvára to teda predpoklad následného znižovania celkových MRR.

Pri týchto záveroch je dôležité pripomenúť fakt, že napriek často uvádzaným štatistikám o neočistenom mzdovom rodovom rozdieli, on samotný nie je až taký smerodajný pri objasňovaní, akým spôsobom sa so ženami na trhu práce zaobchádza, prípadne či sa ich diskriminácia dotýka alebo nie.

2 Ekonometrické metódy skúmania mzdových rodových rozdielov

Existuje množstvo štúdií, ktoré sa zaoberajú skúmaním mzdových rodových rozdielov na základe ekonometrických metód. Na skúmanie tohto javu sa najčastejšie využívajú regresné analýzy, či už lineárne alebo kvantilové, ktoré poskytujú informáciu o vplyve pohlavia na výšku mzdy. Na ne nadväzujú metódy rozkladu. Cieľom týchto metód je rozdeliť mzdové rozdiely medzi dvoma špecifickými skupinami (v našom prípade mužmi a ženami) na časť, ktorá je vysvetlená na základe ich rôznych charakteristík a na časť, ktorá ostáva nevysvetlená (tej potom hovoríme diskriminácia). V nasledujúcej kapitole ponúkame prehľad používaných metód a literatúry, na základe ktorej autori najčastejšie skúmajú mzdové rodové rozdiely. V našej diplomovej práci sa však chceme pozrieť hlavne na časový vývoj mzdových rozdielov pred a počas finančnej krízy. Zameriame sa na analyzovanie časových radov (miezd mužov a žien a vývoja MRR) a tak pojedáme analýzu mzdových rodových rozdielov z iného pohľadu. Teóriu, ktorá je jadrom tejto kapitoly, uvádzame len informatívne pre komplexnejší pohľad na problematiku mzdových rodových rozdielov.

2.1 Lineárna a kvantilová regresia

Lineárna regresia

Základným nástrojom analyzovania vplyvu pohlavia na výšku mzdy je nepochybne klasický lineárny regresný model v tvare:

$$\ln Y = \alpha + \beta X + \gamma D + \epsilon. \quad (2)$$

Kde Y je výška mzdy, X sú kvantitatívne premenné, D je kvalitatívna premenná tzv. dummy premenná, ktorá nadobúda hodnotu 1 ak človek patrí do špecifikovanej skupiny a 0 inak a ϵ je vektor chýb. Interpretácia parametra γ je nasledovná: $\gamma = E[Y|D = 1, X] - E[Y|D = 0, X]$. Tj. rozdiel strednej hodnoty príjmu mužov a žien pri danom X . Nevýhodou lineárnej regresie je, že ponúka iba priemerné výsledky. Aby sa vyhlo tomuto problému, začala sa používať kvantilová regresia, ktorá skúma

efekt pohlavia na výšku mzdy pozdĺž celého mzdového rozdelenia.

Kvantilová regresia

Kvantilová regresia bola zavedená v ekonometrickej literatúre Koenkerom a Bassettom v publikácii [24], ako alternatíva ku klasickej lineárnej regresii. Kvantilovú regresiu môžeme použiť na preskúmanie správania sa nezávislých premenných v každom kvantile skúmanej premennej, čo nám dáva príležitosť pre ucelenejší pohľad na priebeh a vzťahy medzi premennými, ako použitie metód klasickej lineárnej regresie.

2.2 Dekompozičné - rozkladové metódy

Prvou a najznámejšou dekompozičnou metódou je metóda navrhnutá Blinderom a Oaxacom v roku 1973 [4], [29]. Autori využívajú odhad mzdovej rovnice zvlášť pre ženy a mužov. Rovnice majú semi-logaritmickú funkcionálnu formu a vyzerajú nasledovne:

$$\ln W_i = \beta_i X_i + \epsilon_i, \quad i = M, W, \quad (3)$$

kde W_i je mzda, β_i je vektor koeficientov, X_i vektor charakteristík, $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ je vektor nezávislých rovnako rozdelených (iid) chýb a $M = \text{Man}$, $W = \text{Woman}$.

Rozdiel medzi priemernou mzdou mužov a žien môže byť potom rozložený nasledovne:

$$\overline{\ln W_M} - \overline{\ln W_W} = \overline{X_M}(\widehat{\beta_M} - \beta^*) + \overline{X_W}(\beta^* - \widehat{\beta_W}) + (\overline{X_M} - \overline{X_W})\beta^*. \quad (4)$$

Kde $\widehat{\beta_i}$ sú koeficienty odhadnuté metódou najmenších štvorcov z rovnice (3) a β^* je hodnota, ktorá by mala byť odhadnutá pre každý z koeficientov v prípade neexistujúcej diskriminácie. Prvý člen ukazuje výhodu mužov, druhý znevýhodnenie žien a tretí je vysvetlenou zložkou mzdového rozdielu medzi mužmi a ženami. V dekompozícii Oaxaca a Blindera v prípade diskriminácie namierenej voči ženám určujeme $\beta^* = \widehat{\beta_M}$, teda dostávame upravenú rovnicu v tvore:

$$\overline{\ln W_M} - \overline{\ln W_W} = \overline{X_W}(\widehat{\beta_M} - \widehat{\beta_W}) + (\overline{X_M} - \overline{X_W})\widehat{\beta_M}. \quad (5)$$

Prvý člen rovnice (5) predstavuje vysvetlený rozdiel prisudzovaný rozdielom vo vlastnostiach mužov a žien. Druhý člen pravej strany rovnice je nevysvetlený mzdový rozdiel medzi mužmi a ženami.

Po uverejnení štúdií Oaxaca a Blindera rástol počet výskumov mzdových rozdielov. Autori zväčša používali základnú myšlienku Oaxaca - Blinderovej metódy, ktorú v niektorých ohľadoch rozšírili alebo upravili pre svoje potreby. Autori sa napríklad rozchádzajú v názore na vhodný výber β^* . Niektorí odporúčajú výber regresných koeficientov, vypočítaných pre spoločné dáta mužov a žien, iní váhenie mužských a ženských koeficientov na základe ich početnosti vo vzorke. Ďalším rozšírením metódy je jej upravenie Juhncom, Murphym a Pierceom [20], ktoré uľahčuje dekompozíciu mzdových rozdielov medzi dvoma bodmi v čase alebo dvoma krajinami.

Tieto prístupy však poskytujú len obmedzený obraz o mzdových rozdieloch medzi mužmi a ženami, pretože odhadujú mzdové nerovnosti v priemere a nie sú reprezentatívne pre celé mzdové rozdelenia. Autori sa preto začali zaoberať hľadaním všeobecnejšej metódy na rozklad mzdového rozdielu. V publikácií [10] prezentujú semiparametrickú procedúru na analýzu vplyvu inštitucionálneho a pracovného trhu na mzdové rozdelenie v USA. Efekt týchto faktorov je odhadovaný použitím jadrových odhadov hustoty. Iný prístup bol prezentovaný v [11], kde sa prijímajú a rozširujú techniky odhadu funkcie hazardu pre distribučné funkcie príjmov a miezd. Kvantilovú regresiu využívali autori napríklad v práci [25] a na túto metódu nadviazal Melly [27], rozkladajúc mzdové rozdiely v rôznych kvantiloch nepodmieneného rozdelenia. V štúdiu [11] autori využívajú tzv. RIF - regresiu, ktorá je podobná štandardnej regresii s výnimkou nahradenia závislej premennej centrovanou funkciou vplyvu (Recentered influence function - RIF).

Neparametrické dekompozičné párovacie metódy, ktoré porovnávajú resp. párujú platy mužov a žien s rovnakými charakteristikami sú intuitívnou alternatívou k metódam založených na regresii. Napríklad Nopo [28] uvádza rozklad mzdových rozdielov na štyri súčtové zložky. Prvá zložka je vysvetľovaná ako rozdiel medzi dvoma skupinami mužov, tými, ktorých charakteristiky môžu byť priradené k ženským charakteristikám a tými, ktorí nemôžu byť priradení. Druhá zložka predstavuje to isté ako prvá s tou

odlišnosťou, že sa pozeráme na rozdiel medzi dvoma ženskými skupinami. Ďalšie dva členy sú podobné ako pri Oaxacovej - Blinderovej dekompozícii. Teda jeden je vysvetľovaný ako rozdiel v charakteristikách medzi k sebe priradenými mužmi a ženami s podobnými vlastnosťami. A posledný je nevysvetlený rozdiel.

Touto problematikou sa zaoberalo aj veľa iných autorov, ktorí modifikovali uvádzané metódy. Podrobnejší prehľad týchto dekompozičných metód je uvedený napríklad v [15]. Aplikáciou lineárnej regresie a metódy Oaxaca - Blindera, kvantilovej regresie a Mellyho dekompozičnej metódy na slovenské dáta sa zaoberala v diplomovej práci v roku 2013 Klaudíniová [22].

3 Časové rady a ich analýza

Analýza časových radov sa stáva jednou z najdôležitejších oblastí v rozvoji súčasnej štatistiky. Hlavným dôvodom rastúceho významu tejto disciplíny je, že sa úspešne vyrovnáva s popisom dynamických systémov, s ktorými prichádzame bežne do styku. Hlavnou úlohou analýzy časových radov je nájsť vhodný model, ktorý najlepšie a pritom jednoducho aproximuje pozorované údaje a na základe tohto modelu predpovedať budúce hodnoty.

Časový rad tvorí usporiadaná množina údajov - pozorovaní $(x_1, \dots, x_T)$. Táto množina je jednou realizáciou stochastického procesu $\{X_t, t \in I\}$, ktorý je neprázdnu množinou náhodných premenných, definovaných na rovnakom pravdepodobnostnom priestore $(\Omega, \mathcal{A}, P)$. Počet pozorovaní resp. dĺžku časového radu budeme označovať písmenom T .

Každý stochastický proces $\{X_t\}$ je úplne charakterizovaný T - rozmernou distribučnou funkciou. Zvyčajne sa však zameriavame iba na prvé dva momenty:

1. Strednú hodnotu: $E(X_t) \ t = 1, 2, \dots, T$
2. Varianciu (rozptyl): $D(X_t) \ t = 1, 2, \dots, T$

A kovarianciu (mieru závislostí): $cov(X_t, X_s) \ t, s = 1, 2, \dots, T$. Keďže ide o kovarianciu medzi náhodnými premennými toho istého procesu, označujeme ich ako autokovariancie.

Tretia kapitola je voľne spracovaná na základe [6], [17], [33], [30] a [36] a poskytuje teóriu, využívanú v praktickej časti diplomovej práce.

3.1 Kľúčové pojmy, ktoré sa využívajú pri analýze časových radov

3.1.1 Stacionarita a ergodicita procesu

Stacionaritu poznáme v dvoch formách, striktnú (alebo silnú) stacionaritu a slabú stacionaritu (tiež známu aj ako kovariančnú). V metodológii Boxa-Jenkinsa sa dajú modelovať iba stacionárne časové rady (postačuje slabá stacionarita časového radu

preto ďalej pod pojmom stacionárny časový rad budeme myslieť slabo stacionárny časový rad). Z nestacionárneho procesu však vieme urobiť stacionárny za pomoci rôznych transformácií, viac v podkapitole 3.4.

Definícia 3.1 (Striktná stacionarita). *Hovoríme, že stochastický proces $\{X_t\}$ je striktné stacionárny, ak združená distribučná funkcia vektora $(x_{t_1}, x_{t_2}, \dots, x_{t_k})$ sa nemení pri posunutí v čase. Teda:*

$$P\{x_{t_1} \leq c_1, \dots, x_{t_k} \leq c_k\} = P\{x_{t_1+h} \leq c_1, \dots, x_{t_k+h} \leq c_k\}$$

pre všetky $k=1,2,\dots$, pre všetky časové body $t_1, t_2, \dots, t_k$, pre všetky $c_1, c_2, \dots, c_k$ a časové posunutia $h = 0, \pm 1, \pm 2, \dots$

Definícia 3.2 (Slabá stacionarita). *Stochastický proces $\{X_t\}$ je slabo stacionárny ak platí:*

1. Má konštantnú strednú hodnotu $\mu_t = E(X_t) = \mu$.
2. Kovariancie $\gamma(s, t) = \text{cov}(x_t, x_s)$ sú invariantné voči posunu v čase, teda závisia iba na časovom vzdialenosti $|s - t|$ a nie na konkrétnom uniestnení v časovej rade.
 $\gamma(s, t) = \text{cov}(x_s, x_t) = \text{cov}(x_{s+h}, x_{t+h}) = \gamma(|s - t|)$ pre ľubovoľné h, s a t .

Z druhého bodu tiež vyplýva konštantnosť variancie $D(X_t) = \sigma^2$.

Ergodický náhodný proces je taký proces, ktorého momenty vieme určiť z jednej realizácia s dostatočne dlhou dobou trvania. Teda ak máme jednu realizáciu procesu $\{X_t\}$ dĺžky T , označme ju ako: $\{x_1^{(1)}, x_2^{(1)}, \dots, x_T^{(1)}\}$, vieme vypočítať výberovú strednú hodnotu ako:

$$\bar{y} \equiv \frac{1}{T} \sum_{t=1}^T x_t^{(1)}.$$

Ak táto stredná hodnota konverguje pre $T \rightarrow \infty$ podľa pravdepodobnosti k $E(X_t)$ hovoríme, že proces je ergodický pre strednú hodnotu. Podobne hovoríme, že kovariančne stacionárny proces je ergodický pre druhý výberový moment ak:

$$\frac{1}{(T-j)} \sum_{t=1+j}^T (x_t - \mu)(x_{t-j} - \mu) \xrightarrow{p} \gamma_j.$$

3.1.2 Autokorelačná a parciálna autokorelačná funkcia

Základným prostriedkom, ktorý informuje o charaktere stochastického procesu je autokorelačná a parciálna autokorelačná funkcia. Na základe tvaru týchto funkcií vieme identifikovať model konkrétnych časových radov, ktoré analyzujeme.

Autokorelačná funkcia

Autokorelačnú funkciu (označujeme ACF) ρ_τ stacionárneho radu $\{X_t\}$ definujeme ako

$$\rho_\tau = \text{cor}(x_t, x_{t+\tau}) = \frac{\text{cov}(x_t, x_{t+\tau})}{\sqrt{\text{var}(x_t)\text{var}(x_{t+\tau})}} = \frac{\gamma_\tau}{\gamma_0} \quad \text{pre} \quad \tau = 0, \pm 1, \pm 2, \dots$$

Autokorelačná funkcia má v prípade stochastického procesu tieto vlastnosti:

1. Je párna, teda $\gamma_\tau = \gamma_{-\tau}$, preto pre popis funkcie stačí uvažovať $\tau \geq 0$.
2. $\gamma_0 = 1$ a $|\gamma_\tau| \leq 1$.

Ak navyše predpokladáme ergodicitu príslušného procesu, vieme na základe dát $(x_1, \dots, x_T)$ odhadnúť príslušnú strednú hodnotu, varianciu a disperzie a na základe toho autokorelačnú funkciu v tvare:

$$\begin{aligned} \hat{\mu} &= \frac{1}{T} \sum_{t=1}^T x_t, & \hat{\gamma}_0 &= \frac{1}{T} \sum_{t=1}^T (x_t - \hat{\mu})^2, \\ \hat{\gamma}_\tau &= \frac{1}{T} \sum_{t=1}^{T-\tau} (x_t - \hat{\mu})(x_{t+\tau} - \hat{\mu}) & \text{pre} \quad \tau = 0, 1, \dots, T-1, \\ \hat{\varrho}_\tau &= \frac{\hat{\gamma}_\tau}{\hat{\gamma}_0} & \text{pre} \quad \tau = 0, 1, \dots, T-1. \end{aligned}$$

Parciálna autokorelačná funkcia

Parciálnu autokorelačnú funkciu (označujeme PACF) stacionárneho procesu x_t definujeme pre všetky $h = 1, 2, \dots$, ako:

$$\varrho_{11} = \text{cor}(x_t, x_{t+1}) = \varrho_1$$

a

$$\varrho_{hh} = \text{cor}(x_t - \hat{\mu}, x_{t-h} - \hat{\mu}), \quad h \geq 2.$$

Pričom obidva členy $(x_t - \hat{\mu})$ $(x_{t-h} - \hat{\mu})$ a sú nekorelované s $\{x_{t-1}, \dots, x_{t-h+1}\}$. Jednotlivé koeficienty ϱ_{hh} vieme vypočítať na základe nasledujúceho vzťahu:

$$\varrho_{hh} = \frac{\begin{vmatrix} 1 & \varrho_1 & \varrho_2 & \dots & \varrho_{h-2} & \varrho_1 \\ \varrho_1 & 1 & \varrho_1 & \dots & \varrho_{h-3} & \varrho_2 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \varrho_{h-1} & \varrho_{h-2} & \varrho_{h-3} & \dots & \varrho_1 & \varrho_h \end{vmatrix}}{\begin{vmatrix} 1 & \varrho_1 & \varrho_2 & \dots & \varrho_{h-2} & \varrho_{h-1} \\ \varrho_1 & 1 & \varrho_1 & \dots & \varrho_{h-3} & \varrho_{h-2} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \varrho_{h-1} & \varrho_{h-2} & \varrho_{h-3} & \dots & \varrho_1 & 1 \end{vmatrix}}$$

kde $||$ označuje determinant príslušnej matice. Odhady $\hat{\varrho}_{hh}$ sa počítajú rekurentne podľa vzorcov:

$$\hat{\varrho}_{11} = \hat{\varrho}_1, \quad \hat{\varrho}_{hh} = \frac{\hat{\varrho}_h - \sum_{j=1}^{h-1} \hat{\varrho}_{h-1,j} \hat{\varrho}_{h-j}}{1 - \sum_{j=1}^{h-1} \hat{\varrho}_{h-1,j} \hat{\varrho}_j} \quad \text{pre} \quad h > 1$$

$$\text{kde } \hat{\varrho}_{hj} = \hat{\varrho}_{h-1,j} - \hat{\varrho}_{hh} \hat{\varrho}_{h-1,h-j} \quad \text{pre} \quad j = 1, 2, \dots, h-1.$$

Ako už bolo uvedené na začiatku podkapitoly 3.1.2, chovanie autokorelačnej a parciálnej autokorelačnej funkcie je v metodológii Boxa a Jenkinsa dôležitým ukazovateľom. Pre identifikáciu modelu je dôležité nájsť také k , za ktorým začína byť autokorelačná či parciálna autokorelačná funkcia nulová, alebo zistiť, že také k neexistuje. Keďže pre konkrétny analyzovaný časový rad nepoznáme skutočné hodnoty ϱ_τ a ϱ_{hh} , chceme vedieť, ako blízko k nule musia byť naše odhady $\hat{\varrho}_\tau$ a $\hat{\varrho}_{hh}$, aby sme o ich skutočných hodnotách mohli predpokladať, že sú nulové.

Ak predpokladáme normalitu uvažovaného procesu, za účelom výpočtu smerodajnej odchýlky odhadu $\hat{\varrho}_\tau$ autokorelačnej funkcie môžeme použiť Bartlettovu aproximáciu: ak $\gamma_\tau = 0$ pre $\tau > k$, potom pre smerodajnú odchýlku platí:

$$\sigma(\hat{\varrho}_\tau) = \sqrt{\text{var}(\hat{\varrho}_\tau)} \approx \sqrt{\frac{1}{T} \left(1 + 2 \sum_{j=1}^k \gamma_j^2 \right)}, \quad \text{pre} \quad \tau > k.$$

Teraz môžeme testovať hypotézu o nulovosti ϱ_τ nasledovne:

$$H_0 : \quad \varrho_\tau = 0 \quad \text{vs.} \quad H_1 : \quad \varrho_\tau \neq 0$$

Hypotézu H_0 zamietame na 5% hladine významnosti ak

$$|\hat{\varrho}_\tau| > 2\sigma_k$$

Podobne sa dá testovať aj nulovosť hodnôt parciálnej autokorelačnej funkcie ϱ_{hh} na základe Quenouilleovej aproximácie. V tomto prípade pre smerodajnú odchýlku odhadu $\hat{\varrho}_{hh}$, za podmienky $\varrho_{hh} = 0$ pre $h > k$, platí:

$$\sigma(\hat{\varrho}_{hh}) \approx \sqrt{\frac{1}{T}}, \quad \text{pre} \quad h > k.$$

3.1.3 Biely šum

Definujeme tiež proces, ktorý bude mať dôležitú úlohu pri aplikáciách a je príkladom stacionárneho procesu. Pri Box-Jenkinsovej metodológii budeme od rezíduí modelov, určených pre daný časový rad, požadovať vlastnosti bieleho šumu.

Definícia 3.3 (Biely šum). *Proces $\{w_t\}$ je známy ako biely šum (white noise) ak*

1. $E(w_t) = 0$
2. $E(w_t^2) = \text{var}(w_t) = \sigma^2$
3. $E(w_t w_s) = \text{cov}(w_t, w_s) = 0 \quad \text{pre} \quad t \neq s$

Ďalej platí:

- autokorelačná funkcia má tvar:

$$\varrho_\tau = \begin{cases} 1 & \text{ak } \tau = 1 \\ 0 & \text{ak } \tau \neq 1 \end{cases}$$

- parciálna autokorelačná funkcia má tvar:

$$\varrho_{hh} = \begin{cases} 1 & \text{ak } h = 1 \\ 0 & \text{ak } h \neq 1 \end{cases}$$

3.1.4 Lineárny proces a Woldova dekompozícia

Stacionárne procesy uvažované v 3.3 sa dajú zapísať v tvare:

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j w_{t-j}, \quad \text{kde } w_{t-j} \text{ je biely šum,} \quad \psi_0 = 1 \quad \text{a} \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty.$$

To, že sa tak dá zapísať každý stacionárny proces, ktorý neobsahuje žiadnu deterministickú zložku, dokázal Wold v roku 1938. Tento proces je lineárnou kombináciou bieleho šumu a parametrov ψ_j a voláme ho lineárny proces.

Veta 3.4 (Woldova dekompozícia). *Každý stacionárny proces $\{X_t\}$ s nulovou strednou hodnotou môže byť vyjadrený v tvare*

$$X_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} + \kappa_t, \quad \text{kde} \quad \psi_0 = 1 \quad \text{a} \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty.$$

Člen w_{t-j} je biely šum a reprezentuje chybu urobenú pri odhade X_t . Hodnota κ_t je nekorelovaná s w_{t-j} pre všetky j , napriek tomu κ_t môže byť odhadovaná veľmi dobre na základe minulých hodnôt Y :

$$\kappa_t = \hat{E}(\kappa_t | X_{t-1}, t-2, \dots).$$

Člen κ_t je označovaný ako deterministický komponent X_t , a $\sum_{j=0}^{\infty} \psi_j^2$ ako nedeterministický komponent.

Lineárny proces vo všeobecnosti závisí na nekonečnom množstve parametrov, čo nám pri metodológii Boxa-Jenkinsa nevyhovuje, pretože chceme modely časových radov konštruovať čo najúspornejšie. Autoregresný proces, či proces kľzavých priemerov, vynuluje všetky parametre až na konečný počet. Tieto dva procesy sa dajú ďalej kombinovať do tzv. ARMA procesov alebo v prípade nestacionarity či sezónneho časového radu do ARIMA a SARIMA procesov.

Tieto procesy sa často zapisujú s využitím tzv. **operátora spätného posunutia** L (Lag operator), ktorý definujeme ako:

$$Lx_t = x_{t-1} \quad \text{pričom ho môžeme aplikovať viac krát v tvare:} \quad L^j x_t = x_{t-j}.$$

3.1.5 Predpovede časových radov

Pod pojmom **predpoveď** budeme myslieť bodovú predpoveď pre analyzovaný časový rad v jednotlivých obdobiach. Vzťahy pre jej výpočet uvádzame v nasledujúcich dvoch podkapitolách v rámci niektorých metód.

3.2 Dekompozícia časových radov

Pod pojmom dekompozícia časového radu máme na mysli rozklad časového radu na štyri zložky:

- trendovú zložku (T_t)
- sezónnu zložku (S_t)
- cyklickú zložku (C_t)
- náhodnú (reziduálnu, zbytkovú) zložku (E_t)

Tento rozklad sa zväčša vykonáva preto, že nám môže pomôcť pri identifikovaní pravidelného chovania sa daného časového radu. Zaujímavé výsledky tiež môže priniesť štúdium jej oddelených zložiek. Trendová zložka opisuje dlhodobé zmeny v chovaní časového radu - dlhodobý rast či pokles (na základe tejto zložky budeme v našej práci sledovať trend vývoja miezd). Sezónna zložka popisuje periodické zmeny v časovom rade. Perióda môže byť napríklad jeden rok a môže byť daná napríklad zmenou ročných období alebo výškou odmien zamestnancov v jednotlivých mesiacoch. Cyklická zložka hovorí o pohyboch okolo trendu, kde sa strieda fáza rastu s fázou poklesu. Dĺžka cyklov je premenlivá a môže trvať niekoľko rokov. Náhodná časť ostane v rade po odstránení ostatných troch zložiek. Je tvorená pohybmi, ktoré nemajú systematický charakter. Pokrýva tiež chyby merania údajov a chyby spôsobené pri analýze časových radov. Pri reziduálnej zložke sa predpokladá, že je bielim šumom a niekedy dokonca, že sa jedná o biely šum s normálnym rozdelením.

Pri dekompozícií časového radu rozoznávame aditívnu a multiplikatívnu dekompozíciu:

- Aditívna dekompozícia má tvar: $y_t = T_t + S_t + C_t + E_t$.

- Multiplikatívna dekompozícia má tvar: $y_t = T_t S_t C_t E_t$.

3.2.1 Trend v časových radoch

Poznáme viaceré metódy, pomocou ktorých môžeme eliminovať trend z časového radu.

- Popísanie trendu matematickými krivkami
- Metóda kľzavých priemerov
- Exponenciálne vyrovnanie

Popis trendu matematickými krivkami

Pri tejto metóde sa snažíme opísať trend časového radu analyticky, jednoduchými krivkami ako sú priamka alebo exponenciála. Pre takúto krivku sa potom dá jednoducho vypočítať predpoveď budúcich hodnôt. Najvhodnejšiu krivku pre analyzovaný časový rad určujeme na základe grafickej ukážky alebo predpokladu o správaní sa časového radu. Môžeme teda uvažovať rôzne prípady trendových kriviek. Konštantný trend, lineárny trend, kvadratický trend, exponenciálny trend či logistický trend a iné.

Metóda kľzavých priemerov

Metóda kľzavých priemerov patrí medzi adaptívne prístupy k trendovej zložke, čo znamená, že dokáže zaznamenávať zmeny v charaktere trendu časového radu. Ďalej sa ale predpokladá, že v kratších úsekoch časového radu je vyrovnanie pomocou kriviek možné s tým rozdielom, že má v rozdielnych úsekoch iné parametre. Tieto krátke úseky, ktorých dĺžka je vopred stanovená, sa ďalej snažíme vyrovnávať polynomickými krivkami. S voľbou dĺžky úsekov sú však spojené problémy a určuje sa skôr subjektívne, čo nie vždy vedie k presným výsledkom.

Exponenciálne vyrovnanie

Metóda exponenciálneho vyrovnania je taktiež adaptívna metóda, ktorá odstraňuje nevýhodu metódy kľzavých priemerov. Výpočet každej vyrovnanej hodnoty je založený na všetkých predošliach meraniach. Nástrojom je modifikovaná metóda najmenších štvorcov, a to taká, že váhy jednotlivých štvorcov sa znižujú smerom do minulosti

exponenciálne. Pre vyrovnanú hodnotu časového radu v čase t $\hat{y}_t$ pri exponenciálnom vyrovnaní minimalizujeme výraz tvaru:

$$(y_t - \hat{y}_t)^2 + (y_{t-1} - \hat{y}_{t-1})^2\alpha + (y_{t-2} - \hat{y}_{t-2})^2\alpha^2 + \dots$$

cez všetky známe hodnoty a α je vyrovnávajúca konštanta z rozmedzia 0 a 1. Poznáme jednoduché exponenciálne vyrovnanie, dvojité či trojité exponenciálne vyrovnanie.

Viac o uvedených metódach, napríklad o odhadoch vhodných parametrov sa môže čitateľ dozvedieť napríklad v [6].

3.2.2 Sezónnosť v časových radoch

Sezónne očisťovanie alebo analýza sezónnej zložky je dôležitou súčasťou analýzy časových radov. Cieľom očisťovania časových radov od sezónnej zložky je následné efektívnejšie štúdium dlhodobých tendencií, ktorým je skúmaný rad podriadený. Sezónna zložka môže mať rozdielny charakter a to aditívny (nezávisí na priemernej úrovni radu, vyjadreného pomocou trendu) alebo multiplikatívny (ak je amplitúda sezónneho pohybu úmerná priemernej úrovni radu). Existujú rôzne metódy prístupu k sezónnej zložke. My sa bližšie pozrieme na tzv. Holt-Wintersovu metódu.

Holt-Wintersova metóda

Holt - Wintersova metóda je zovšeobecnením metódy exponenciálneho vyrovnávania pre časové rady, obsahujúce okrem trendovej zložky aj sezónnu zložku. Pri tejto metóde sa vykonáva adaptívny odhad trendovej a navyše aj sezónnej zložky. Rozoznávame dve varianty metódy, a to aditívnu a multiplikatívnu Holt-Wintersovu metódu. Pri oboch druhoch metódy pracujeme s tromi vyrovnávajúcimi konštantami α , β a γ z intervalu (0,1). Predkladáme, že trendová zložka má lineárny charakter, teda: $T_t = \beta_0 + \beta_1 t$. Písmenom L označujeme počet pozorovaní nameraných v jednej sezóne (roku), odhady β_0 , β_1 a odhad sezónnej zložky S_t , ktoré sú vytvorené v čase t označíme $b_0(t)$, $b_1(t)$ a $s_t(t)$. Ak teda rovnica $a_0(t) = b_0(t) + b_1(t)t$ odhaduje úroveň trendu potom môžeme v ďalšej časti používať iba odhady $a_0(t)$, $b_1(t)$ a $s_t(t)$.

Aditívna Holt-Wintersova metóda

Pri tejto metóde predpokladáme aditívnu dekompozíciu časového radu a pre $a_0(t)$,

$b_1(t)$ a $s_t(t)$ platí:

$$a_0(t) = \alpha(y_t - s_{t-L}(t-L)) + (1-\alpha)(a_0(t-1) + b_1(t-1)), \quad (6)$$

$$b_1(t) = \beta(a_0(t) - a_0(t-1)) + (1-\beta)b_1(t-1), \quad (7)$$

$$s_t(t) = \gamma(y_t - a_0(t)) + (1-\gamma)s_{t-L}(t-L). \quad (8)$$

Predpoveď hodnoty $y_{t+\tau}$, konštruovaná v čase t , má pre aditívnu Holt-Wintersovu metódu tvar: $y_{t+\tau}(t) = a_0(t) + b_1(t)\tau + s_{t+\tau-L}(t+\tau-L)$, pričom, ak $\tau > L$, musíme položiť odhad sezónnej zložky nasledovne: $s_{t+\tau-2L}(t+\tau-2L)$ atď., resp. použijeme najaktuálnejší odhad sezónnej zložky.

Multiplikatívna Holt-Wintersova metóda

Pri tejto Holt-Wintersovej metóde predpokladáme multiplikatívnu dekompozíciu časového radu a pre $a_0(t)$, $b_1(t)$ a $s_t(t)$ platia nasledujúce rekurentné vzťahy:

$$a_0(t) = \alpha \frac{y_t}{s_{t-L}(t-L)} + (1-\alpha)(a_0(t-1) + b_1(t-1)), \quad (9)$$

$$b_1(t) = \beta(a_0(t) - a_0(t-1)) + (1-\beta)b_1(t-1), \quad (10)$$

$$s_t(t) = \gamma \frac{y_t}{a_0(t)} + (1-\gamma)s_{t-L}(t-L). \quad (11)$$

Predpoveď hodnoty $y_{t+\tau}$, konštruovaná v čase t , má pre multiplikatívnu Holt-Wintersovu metódu tvar: $y_{t+\tau}(t) = (a_0(t) + b_1(t)\tau)s_{t+\tau-L}(t+\tau-L)$, pričom, ak $\tau > L$ musíme položiť odhad sezónnej zložky nasledovne: $s_{t+\tau-2L}(t+\tau-2L)$ atď., resp. použijeme najaktuálnejší odhad sezónnej zložky.

Príklad počiatkových podmienok, ktoré môžeme vziať, aby sme mohli použiť uvedené vzťahy (6),(7) a (8) pre aditívnu metódu a (9), (10) a (11) pre multiplikatívnu metódu, je uvedený v [6]. Pri Holt-Wintersových metódach je potrebné vhodne zvoliť vyrovňavajúce konštanty α , β a γ . Pričom sa stačí obmedziť na hodnoty $0 < \alpha, \beta, \gamma \leq 0,30$.

3.3 Box-Jenkinsova metodológia - stacionárne procesy

V nasledujúcej podkapitole uvedieme teoretické základy Boxovej-Jenkinsovej metodológie, ktorá poskytuje užitočnú triedu modelov, opisujúcich jednorozmerné časové rady. Pri používaní tejto metodológie sa kladie dôraz na reziduálnu zložku (v porovnaní s

dekompozičnými metódami, kde ju berieme ako zbytkovú časť), ktorá môže byť tvorená závislými (korelovanými) náhodnými veličinami. Box-Jenkinsova metodológia teda môže spracovávať časové rady s navzájom závislými pozorovaniami. Modely, ktoré uvedieme, sú modelmi stochastickými a dokážu modelovať aj trend či sezónnosť. Navyše analýza sa v rámci tejto metodológie robí systematicky v troch hlavných krokoch (identifikácia modelu, odhad parametrov a overovanie modelu). Ďalšou výhodou modelov, s ktorými táto metodológia pracuje, sú dobré výsledky v porovnaní s inými modelmi, či jednoduché modelovanie časového radu v rôznych štatistických programoch. Nevýhodou je, že pre vytvorenie spoľahlivého modelu potrebujeme časový rad aspoň s 50 pozorovaniami, čo sa nie vždy dá jednoducho zabezpečiť.

3.3.1 Autoregresný proces (AR - proces)

Autoregresný proces rádu p , označovaný ako $AR(p)$ je definovaný ako:

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t \quad (12)$$

Ak navyše definujeme autoregresný operátor ako:

$$\phi(L) = 1 - \phi_1(L) - \phi_2(L^2) - \dots - \phi_p(L^p),$$

vieme prepísať rovnicu (12) do tvaru $\phi(L)x_t = c + w_t$.

Daný proces je stacionárny, ak všetky korene $\phi(L)$ ležia mimo jednotkového kruhu v komplexnej rovine a uvažujeme stochastické začiatkové podmienky.

Strednú hodnotu autoregresného procesu vieme nájsť z rovnice (12) a predpokladu stacionarity procesu nasledovne:

$$E(x_t) = \mu = c + \phi_1 \mu + \phi_2 \mu + \dots + \phi_p \mu, \quad \Rightarrow \quad \mu = \frac{c}{(\phi_1 + \phi_2 + \dots + \phi_p)}. \quad (13)$$

Na základe (13) vieme prepísať rovnicu (12) do tvaru:

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \dots + \phi_p(x_{t-p} - \mu) + w_t. \quad (14)$$

Autokovariancie vieme nájsť vynásobením oboch strán rovnice (14) členom $(x_{t-j} - \mu)$ a vytvorením stredných hodnôt:

$$\gamma_j = \begin{cases} \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p} & \text{pre } j = 1, 2, \dots \\ \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma^2 & \text{pre } j = 0 \end{cases} \quad (15)$$

Tento systém sa dá riešiť pre $j = 0, 1, 2, \dots, p$ ako sústava $p+1$ rovníc o $p+1$ neznámych $(\gamma_0, \gamma_1, \dots, \gamma_p)$ a pre $j > p$ rekurentne z diferenčnej rovnice:

$$\gamma_t - \phi_1 \gamma_{t-1} - \phi_2 \gamma_{t-2} - \dots - \phi_p \gamma_{t-p} = 0.$$

Predelením prvých $p+1$ rovností (15) členom γ_0 dostávame vzťah pre varianciu a počiatočné podmienky pre autokorelačnú funkciu (tzv. Yule-Walkerove rovnice) autoregresného procesu:

- $\sigma_y^2 = \frac{\sigma^2}{1 - \phi_1 \gamma_1 - \phi_2 \gamma_2 - \dots - \phi_p \gamma_p}$
- $\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p}$ pre $j = 1, 2, \dots, p$

Ďalej v prípade autoregresného modelu p -teho rádu môžeme identifikovať rád modelu na základe PACF vďaka jej nulovosti $\rho_{hh} = 0$ pre $h > p$.

3.3.2 Proces kľzavých priemerov (MA - proces)

Proces kľzavých priemerov (moving average process) q -teho rádu $MA(q)$ je charakterizovaný rovnicou:

$$x_t = \mu + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q}. \quad (16)$$

Ak navyše definujeme operátor kľzavých priemerov ako:

$$\theta(L) = 1 + \theta_1(L) + \theta_2(L^2) + \dots + \theta_q(L^q),$$

vieme prepísať rovnicu (12) do tvaru $x_t - \mu = \theta(L)w_t$.

Stacionarita procesu je zabezpečená pre ľubovoľné parametre θ_j kde $j = 1, 2, \dots, q$. Podmienkou invertibility sú korene operátora kľzavých priemerov $\theta(L)$ mimo jednotkového kruhu v komplexnej rovine. To, že je $MA(q)$ proces invertovateľný znamená, že ho vieme previesť (invertovať) na stacionárny proces $AR(\infty)$.

Z rovnice (16) za predpokladu nulovosti strednej hodnoty bieleho šumu dostávame strednú hodnotu $MA(q)$ procesu $E(x_t) = \mu$. Na základe nekorelovanosti bieleho šumu sú variancia a autokovariancie $MA(q)$ procesu nasledovné:

$$\gamma_j = \begin{cases} (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)\sigma^2 & \text{pre } j = 0 \\ (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \dots + \theta_q\theta_{q-j})\sigma^2 & \text{pre } j = 1, 2, \dots, q \\ 0 & \text{pre } j = 1, 2, \dots, q. \end{cases} \quad (17)$$

Predelením rovnice (17) varianciou procesu dostávame vzťah pre autokorelačnú funkciu:

$$\rho_j = \begin{cases} \frac{(\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \dots + \theta_q\theta_{q-j})}{(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)} & \text{pre } j = 1, 2, \dots, q \\ 0 & \text{pre } j \geq q. \end{cases}$$

Vidíme, že identifikačným bodom pre MA(q) proces je $k = q$ a tento bod vieme nájsť na základe autokorelačnej funkcie.

3.3.3 Zmiešaný proces (ARMA - proces)

Zmiešaný proces rádu p a q, ktorý označujeme ako ARMA(p,q) je definovaný ako:

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q} \quad (18)$$

a vzniká kombináciou autoregresného procesu a procesu kľzavých priemerov.

Pomocou autoregresného operátora a operátora kľzavých priemerov môžeme ARMA(p,q) proces zapísať ako:

$$\phi(L)x_t = c + \theta(L)w_t,$$

pričom požadujeme, aby $\phi(L)$ a $\theta(L)$ nemali rovnaké korene.

Pri ARMA procese je podmienka stacionarity rovnaká ako pri AR(p) procese. Daný ARMA proces je teda stacionárny, ak všetky korene $\phi(L)$ ležia mimo jednotkového kruhu v komplexnej rovine a uvažujeme stochastické začiatkové podmienky. Podmienka invertibility je zase totožná s podmienkou MA(q) procesu. Podmienkou invertibility sú teda korene operátora kľzavých priemerov $\theta(L)$ mimo jednotkového kruhu v komplexnej rovine.

Stredná hodnota procesu je rovnaká ako pri AR(p) procese. Variancia, autokovariancie a autokorelácie sa vypočítajú podobne ako pri predošlých procesoch.

Dôležitým faktom je, že pri ARMA procese neplatí, že $ACF(k) = 0$ pre $k > p$ a $PACF(k) = 0$ pre $k > q$ teda ACF ani PACF sa nevynuluje pri konečnom počte členov.

3.4 Box-Jenkinsova metodológia - nestacionárne procesy

Dôležitý predpoklad pri ARMA procesoch je ich stacionarita. V praxi však časové rady majú často nestacionárny charakter, spôsobený meniacou sa strednou hodnotou, alebo rozptylom procesu. Box-Jenkinsova metodológia sa vysporiadava s nestacionaritou časových radov pomerne jednoducho na základe vytvorenia diferencií pôvodného časového radu, čím niektoré z nich môžeme previesť na stacionárne.

3.4.1 Integrovaný zmiešaný proces ARIMA

Nestacionarita časového radu vyplýva z existencie tzv. jednotkového koreňa autoregresného operátora. Vo všeobecnosti, ak má proces jednotkový koreň násobnosti d (ostatné mimo jednotkového kruhu), jeho d -te diferenciácie sú stacionárne a môžeme ich modelovať ako ARMA proces. Diferenčný operátor Δ môžeme vyjadriť pomocou operátora spätného posunutia nasledovne:

$$\Delta = 1 - L, \quad \text{teda} \quad \Delta x_t = x_t - x_{t-1} = (1 - L)x_t, \quad \Delta^2 x_t = (1 - L)^2 x_t = x_t - 2x_{t-1} + x_{t-2}, \dots$$

Model ARIMA(p, d, q) teda môžeme napísať ako:

$$\phi(L)(1 - L)^d x_t = c + \theta(L)w_t.$$

Takéto rady sa často nazývajú časové rady s jednotkovými koreňmi. Pri praktickej výstavbe modelu nájdeme rád diferencovania d radu x_t , skonštruujeme diferencovaný rad $y_t = \Delta x_t$ a budujeme pre neho klasický ARMA proces. Pred diferencovaním sa niekedy používa aj transformácia časového radu, ktorej cieľom je daný časový rad linearizovať tak, aby bolo možné ho popísať procesom ARIMA (často sa využíva logaritmická, mocninová alebo iné transformácie).

Určovanie rádu diferencovania

Pre určenie rádu diferencovania môžeme použiť rôzne prístupy:

- Využitie grafického záznamu analyzovanej časovej rady a rady prvých a druhých diferencií.
- Štúdium odhadnutej autokorelačnej funkcie časového radu. Ak jednotlivé autokorelácie klesajú pomaly (lineárne), je to znak nestacionárnej časovej rady.

- Posúdením veľkosti variancie pôvodného radu a diferencovaných radov. Ukazuje sa, že diferencovaním časového radu sa variancie znižujú, pokiaľ nie je dosiahnutá stacionarita, následne začnú opäť rásť.
- Využitím Dickey-Fullerovho testu či vylepšeného ADF testu (Augmented Dickey-Fullerovho testu), ktorým môžeme testovať hypotézu, že časový rad má jednotkovú koreň. Viac o tomto teste napríklad v [17].

3.4.2 Sezónne ARIMA procesy

Sezónnosť je tak isto ako trend (pri predošlom ARIMA procese), modelovaná v metodológii Boxa-Jenkinsa stochasticky a to sezónnymi ARIMA modelmi (SARIMA modelmi). Kvôli trendu robíme diferencie $x_t - x_{t-1}$ a kvôli sezónnosti diferencie $x_t - x_{t-s}$, kde s je perióda sezónnosti dát. Sezónny autoregresný operátor rádu P definujeme ako:

$$\Phi(L^s) = 1 - \Phi_1(L^s) - \Phi_2(L^{2s}) - \dots - \Phi_P(L^{Ps}).$$

Sezónny operátor kľzavých priemerov rádu Q definujeme nasledovne:

$$\Theta(L^s) = 1 + \Theta_1(L^s) + \Theta_2(L^{2s}) + \dots + \Theta_Q(L^{Qs}).$$

A sezónny diferenčný operátor:

$$\Delta_s = 1 - L^s.$$

Pomocou definovaných operátorov môžeme nasledovne definovať multiplikatívny sezónny model rádu $(p, d, q) \times (P, D, Q)_s$:

$$\phi(L)\Phi(L^s)\Delta_s^d\Delta_s^Dx_t = \theta(L)\Theta(L^s)w_t.$$

3.4.3 Výstavba modelu

Celý postup vytvárania modelu je zhrnutý v nasledujúcich troch bodoch a na Obr. 10.

1. Identifikácia ARMA modelu

Analýzu časového radu, využívajúc Boxovu-Jenkinsovu metodológiu, začíname overovaním stacionarity modelov, či sezónneho charakteru modelu. Ďalej nasleduje skúmanie priebehu odhadnutých autokorelačných funkcií ρ_τ a parciálnych autokorelačných funkcií ρ_{hh} . Porovnaním vlastností ACF a PACF procesov

AR(p), MA(q) a ARMA(p,q) uvedených v Tabuľke 3 s našimi odhadmi funkcií pre skúmaný časový rad vieme približne identifikovať vhodný model. Vhodné je tiež prijať alternatívy navrhnutého modelu, ktoré sa neskôr môžu prejavíť ako lepšie.

Tabuľka 3: Tvar autokorelačnej a parciálnej autokorelačnej funkcie procesov AR(p), MA(q) a ARMA(p,q) zdroj:[6].

	AR(p)	MA(p)	ARMA(p,q)
ρ_τ	neexistuje k, ρ_τ je v tvare U	$k = q$	neexistuje k, ρ_τ je v tvare U po prvých $q - p$ hodnotách
ρ_{hh}	$k = p$	neexistuje k, ρ_{hh} je v tvare U	neexistuje k, ρ_{hh} je v tvare U po prvých $p - q$ hodnotách

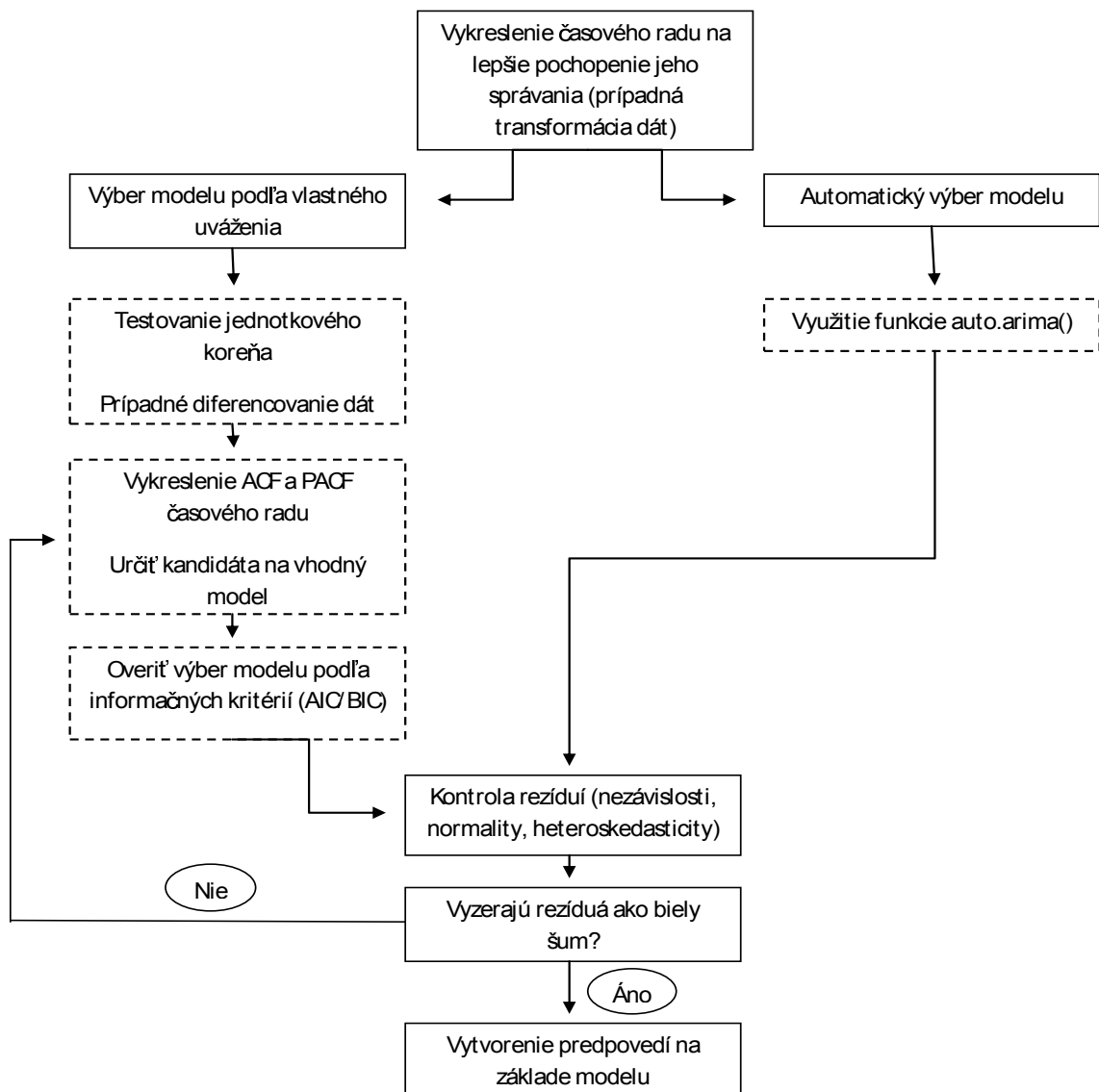
2. Odhad parametrov modelov

Pre odhad parametrov Box-Jenkinsových modelov sa používajú iteračné metódy, zväčša nelineárna metóda najmenších štvorcov, ktorú podporuje väčšina štandardných štatistických alebo ekonometrických programových balíkov. Popis metódy najmenších nelineárnych štvorcov je uvedený napríklad v [6]. Inou vhodnou odhadovacou metódou je napríklad metóda maximálnej vierohodnosti.

3. Metódy overovania adekvátnosti modelov

Predtým, ako využijeme model na prognózovanie budúcich hodnôt skúmaného javu, treba sa zamyslieť nad tým, či nami zvolený model dostatočne popisuje analyzovaný časový rad alebo vybrať z viacerých prípustných modelov ten najvhodnejší. Na to slúžia rôzne metódy, informačné kritéria a koeficienty. Tiež treba overiť, či sú splnené požiadavky, ktoré sú kladené na rezíduá modelov (normalita, nezávislosť, ...), podmienka stacionarity a invertovateľnosti procesu.

Koeficient determinácie (R-squared) hovorí o kvalite zvoleného modelu. Nadobúda hodnoty v intervale $\langle 0, 1 \rangle$, pričom hodnota 1 znamená ideálny model.



Obr. 5: Postup výberu modelu ARIMA.

Vieme ho vypočítať na základe vzťahu:

$$R^2 = 1 - \frac{RSS}{TSS},$$

kde $RSS = \epsilon'\epsilon$ je reziduálna suma štvorcov, a $TSS = (x - \bar{x})'(x - \bar{x})$ je celková suma štvorcov.

Upravený koeficient determinácie (Adjusted R-squared) navyše penalizuje pridávanie nadbytočných parametrov do modelu nasledovne

$$R_A^2 = 1 - \frac{T-1}{T-k}(1 - R^2).$$

Ďalej si uvedieme dve najpoužívannejšie informačné kritériá Akaikeho a Bayesovo informačné kritérium, ktoré je niekedy označované ako Schwarzovo kritérium. Kritériá sa líšia v penalizačnej funkcii.

Definícia 3.5 (AIC). Akaikeho informačné kritérium (AIC) definujeme ako

$$AIC = -2\log L_k + 2k,$$

kde $\log L_k$ je hodnota logaritmickej funkcie maximálnej vierohodnosti, ktorú sa snažíme maximalizovať a k je počet parametrov uvažovaného modelu.

Definícia 3.6 (BIC / SC). Bayesovo informačné kritérium (BIC) známe tiež ako Schwarzovo kritérium (SC) definujeme ako

$$BIC = -2 \log L_k + k \log T,$$

kde $\ln L_k$ je hodnota logaritmickej funkcie maximálnej vierohodnosti, ktorú sa snažíme maximalizovať, k je počet parametrov uvažovaného modelu a T je dĺžka analyzovaného časového radu.

Pre AIC a BIC kritériá platí, že čím je menšia ich hodnota tým je model lepší.

Ďalším efektívnym spôsobom, ako určiť správnosť modelu, je vytvorenie **predpovede ex-post**. Zmenšením rozsahu údajov o časť m , ktoré použijeme na porovnanie zhody s predpoveďou na základe $T - m$ údajov.

Na ohodnotenie kvality predikcie oproti reálnym hodnotám sa často používa hodnota **strednej štvorcovej chyby RMSE** (Root Mean Square Error) v nasledujúcom tvare:

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (x_t - \bar{x}_t)^2}{T}},$$

kde x_t sú reálne hodnoty zaznamenané v čase t a $\bar{x}_t$ sú hodnoty odhadnuté pomocou nami zvoleného modelu pre čas t .

Kontrola rezíduí modelu je vykonávaná na základe rôznych štatistických testov. Od rezíduí modelu požadujeme, aby mali vlastnosti bieleho šumu, teda nemôže v

nich zostať žiadna autokorelácia. Táto podmienka sa dá testovať viacerými spôsobmi, jedným z nich je **Ljung-Boxova štatistika**. Pre časový rad rezíduí pomocou nej testujeme hypotézu, že autokorelačné koeficienty od prvého až po k -ty rád sú nulové. Ak sa hypotéza nezamieta, model je v poriadku. Model by mal tiež spĺňať požiadavku normality rozdelenia pravdepodobností jeho rezíduí. Podľa [1] je však fakt, že pri modelovaní jednorozmerných časových radov táto podmienka často nie je splnená, bežný a neodstrániteľný preto nebudeme striktné dbať na splnenie tejto podmienky pri analýze našich časových radov.

3.4.4 Vytváranie predpovedí v ARIMA modeloch

Základný vzťah pre výpočet predpovedí pre proces tvaru:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q},$$

je

$$\bar{x}_{t+k} = \phi_1 [x_{t+k-1}] + \phi_2 [x_{t+k-2}] + \dots + \phi_p [x_{t+k-p}] + [w_{t+k}] + \theta_1 [w_{t+k-1}] + \theta_2 [w_{t+k-2}] + \dots + \theta_q [w_{t-q}].$$

Pričom si treba uvedomiť, že platí:

$$[x_{t+j}] = \bar{x}_{t+j}(t) \quad \text{pre } j > 0,$$

$$[x_{t+j}] = x_{t+j} \quad \text{pre } j \leq 0$$

a

$$[w_{t+j}] = 0 \quad \text{pre } j > 0,$$

$$[w_{t+j}] = w_{t+j} = x_{t+j} - \bar{x}_{t+j}(t+j-1) \quad \text{pre } j \leq 0.$$

Aby sme mohli zahájiť rekurentný výpočet, musíme položiť $w_1 = w_2 = \dots = w_q = 0$.

4 Analýza priemerných mzdových rodových rozdielov

Analýza časových radov je hodnotným zdrojom informácií o priebehu skúmaného javu, v našom prípade vývoji mzdových rodových rozdielov, a často poskytuje základ pre úspešnú prognózu na nejaké ďalšie obdobie. V nasledujúcej kapitole opíšeme dáta, s ktorými budeme pracovať, a postup pri výbere vhodných modelov k daným časovým radom. Ďalej uvedieme výsledky pre priemerné hodnoty mzdových rodových rozdielov a miezd mužov a žien.

4.1 Popis využívaných dát

V diplomovej práci budeme využívať dáta zo zdrojov Sociálnej poisťovne SR. Ide o administratívne dáta o individuálnych príjmoch obyvateľov. Dáta sú dostupné za obdobie od roku 2005 do roku 2013. Vysvetľujúcimi premennými sú obdobie (mesiac a rok), okres (trvalý pobyt osoby), rok narodenia a pohlavie. Hlavnými premennými sú výška a druh príjmu. V našej diplomovej práci budeme využívať iba dáta o osobách, ktoré vykazujú príjem zo zamestnaneckého pomeru. Nebudeme brať do úvahy zisky zo živností, dôchodové príjmy alebo príjmy osôb s uzavretou dohodou o vykonaní práce. Uvedené dáta zo sociálnej poisťovne majú výhodu veľkej reprezentatívnosti údajov vzhľadom k tomu, že obsahujú všetky príjmy všetkých osôb, ktoré sú v systéme. Nejde teda o výberové vzorky. Nevýhodou je malé množstvo vysvetľujúcich premenných a teda nepoužiteľnosť dekompozičných metód opísaných v kapitole 2. Kvôli tejto slabine našich dát sme sa rozhodli bližšie pozrieť na vývoj mzdových rodových rozdielov na základe analýzy vybraných časových radov. Dôležité je tiež spomenúť, že sme uvedené dáta, s ktorými budeme, pracovať očistili o infláciu na základe indexov spotrebiteľských cien (CPI) k decembru 2013, uvedených na stránke ŠÚSR [43] a že sa jedná o hrubé mesačné príjmy. Výsledné tabuľky, zostavené na základe uvedených dát a využívané v nasledujúcich kapitolách, uvádzame na CD v Prílohe D.

4.2 Popis využívaného programu R

S rozvojom teórie analýzy časových radov dochádza tiež k vzniku nových programov, ktoré sprístupňujú metódy stále väčšiemu okruhu záujemcov o danú problematiku. V

našej práci budeme využívať voľne dostupný program R, v ktorom sú implementované funkcie, užitočné pri analyzovaní časových radov. Všetky príkazy, spájajúce sa s nasledujúcimi tromi praktickými kapitolami, sú uvedené v prílohe B a na CD v prílohe D. Sú označené jednotlivými číslami prislúchajúcich podkapitol.

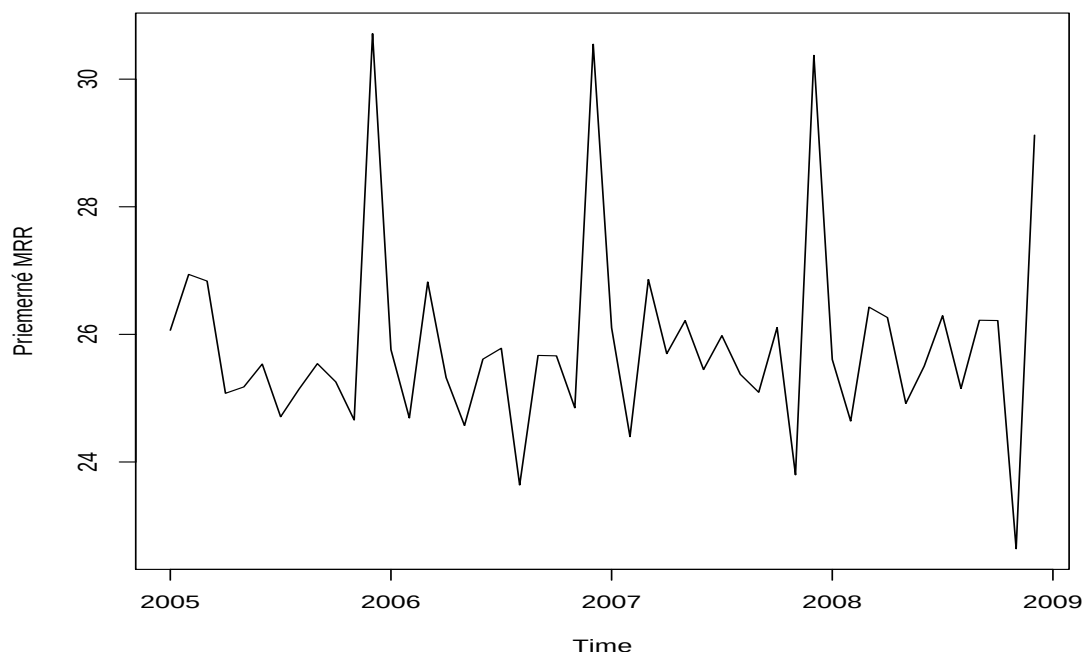
4.3 Modelovanie zvolených časových radov

V nasledujúcej časti budeme analyzovať časový rad priemerných MRR na Slovensku z obdobia 2005 - 2008 (MRR sme vypočítali na základe dát opísaných v podkapitole 4.1 podľa vzorca (1) a výsledky sú uvedené v prílohe D). Vyberieme vhodný model pre tento časový rad a na základe neho budeme predpovedať správanie mzdových rodových rozdielov na Slovensku od roku 2009, teda na obdobie po vypuknutí finančnej krízy na Slovensku. Naše predpovede neskôr porovnáme s reálnymi hodnotami. Budeme tak analyzovať vplyv určitého ekonomického šoku (finančnej krízy) na vývoj mzdových rodových rozdielov. Ďalej sa pozrieme na priemerné mzdy slovenského muža a ženy a na prípadné odlišnosti v ich vývoji počas krízy. Nakoniec rozdelíme dáta v každom období (mesiaci) na tri skupiny podľa zárobku jednotlivých osôb. V prvej skupine budú osoby zarábajúce mesačne menej ako 300 eur, druhá skupina bude zostavená z osôb zarábajúcich 300 - 3000 eur, v tretej skupine budú osoby zarábajúce mesačne nad 3000 eur. Pozrieme sa ako sa menili počty v jednotlivých skupinách v období od roku 2005 do roku 2013.

4.3.1 Priemerné MRR

V tejto podkapitole sa zameriame na výber vhodného modelu pre dáta mzdových rodových rozdielov z obdobia 2005 - 2008 a následné predpovedanie hodnôt MRR od roku 2009. Vhodný ARIMA model vyberieme na základe postupu, popísaného v podkapitole 3.4.3. Uvedieme postup, ktorý zahŕňa výber modelu podľa vlastného uváženia, ale predvedieme aj funkcie, ktoré ponúka štatistický program R na automatické určenie rádu ARIMA modelu. A overíme vhodnosť vybraných modelov.

Na obrázku 6 môžeme vidieť mesačný vývoj mzdových rodových rozdielov od roku 2005 do roku 2008. Vidíme, že MRR sa v tomto období pohybovali okolo hodnoty 25,9 %. Na prvý pohľad nerozoznávame prítomnosť trendu v časovej rade. Ďalej vieme

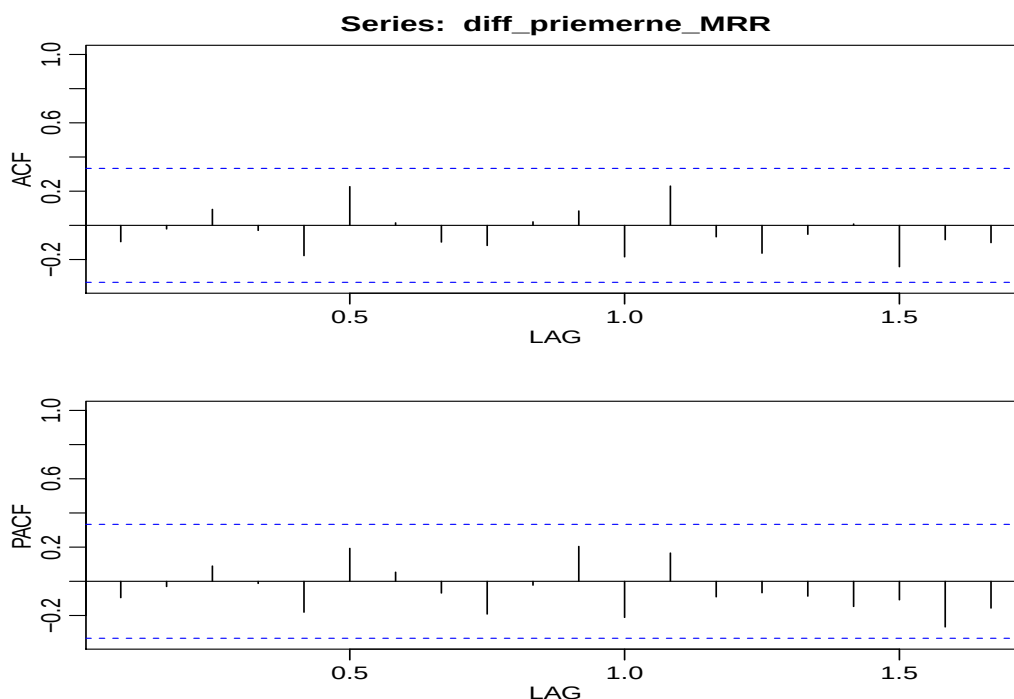


Obr. 6: Vývoj priemerných mzdových rozdielov v čase od roku 2005 do 2008 na Slovensku.

identifikovať sezónnosť, čo je spôsobené zrejme rôznou výškou odmien na konci rokov u mužov a žien. Na to, aby sme z uvedeného časového radu dostali stacionárny, ho budeme musieť sezónne diferencovať, s periódou dvanástich mesiacov. Overíme, či nie je v diferencovanom rade jednotkový koreň pomocou ADF testu (Augmented Dickey–Fuller test), ktorý nepotvrdil hypotézu o prítomnosti jednotkového koreňa v našom sezónne diferencovanom rade MRR. Takto sme si pripravili časový rad na ďalší krok a pri identifikácii modelu budeme teda využívať sezónne diferencovaný rad MRR.

Vieme, že z vykresleného priebehu odhadnutej autokorelačnej a parciálnej autokorelačnej funkcie sa dá odhadnúť vhodný model pre náš upravený časový rad. Z vykresleného korelogramu na obrázku 7 vidíme, že identifikačné body k sú v oboch prípadoch rovné nule. Teda na prvý pohľad sa môže zdať najvhodnejším modelom pre rad priemerných MRR model $SARIMA(0, 0, 0) \times (0, 1, 0)_{12}$.

V programe R je implementovaná funkcia `auto.arima()`, pomocou ktorej automatizujeme odhaľovanie jednotkového koreňa a identifikujeme najlepší model vzhľadom ku niektorému z informačných kritérií (AIC, BIC, AICc) pre daný časový rad. Vstupom



Obr. 7: ACF a PACF sezónne diferencovaného radu MRR

tejto funkcie je analyzovaný časový rad, maximálne hodnoty p , q , P a Q , d a D (v prípade nezadania rádu diferencovania a sezónneho diferencovania ich rád sa určí na základe jedného z testov, ktorý odhaľuje prítomnosť jednotkového koreňa (kpss test, adf test, pp test (Phillips-Perron test)), či prítomnosť sezónneho jednotkového koreňa (ch test (Canova-Hansen test), ocsb test (Osborn-Chui-Smith-Birchhall test))), ďalej môžeme určiť, či sa jedná o stacionárny rad, prípadne rad so sezónnou zložkou, ale aj informačné kritérium, podľa ktorého hodnoty sa určí najlepší model pre náš časový rad. Pomocou tejto funkcie nám ako najvhodnejší model vyšiel $SARIMA(0, 0, 0) \times (0, 1, 0)_{12}$ (s využitím kpss testu a ocsb testu), no pri zmene argumentov funkcie (využitím kpss testu a ch testu) je jej výstupom model $SARIMA(0, 0, 0) \times (1, 0, 0)_{12}$ a pri adf teste a ocsb teste model $SARIMA(0, 2, 2) \times (0, 1, 0)_{12}$.

V ďalšej fáze overíme vhodnosť týchto troch modelov a v prípade splnenia požiadaviek, ktoré sú kladené na rezíduá modelu, porovnáme ich predikčné vlastnosti vynechaním posledných šiestich údajov časového radu.

Prvou metódou overenia modelu je vykreslenie ACF rezíduí. Hodnota autokore-

lačnej funkcie je štatisticky významná ak prekoná interval spoľahlivosti, ktorý je vyznačený na korelograme. Na základe týchto korelogramov rezíduí našich troch modelov vieme určiť nevhodnosť modelu $\text{SARIMA}(0, 2, 2) \times (0, 1, 0)_{12}$, pretože sa neukázalo, že jeho rezíduá sú bielym šumom. Ďalej overíme nezávislosť rezíduí modelov $\text{SARIMA}(0, 0, 0) \times (0, 1, 0)_{12}$ a $\text{SARIMA}(0, 0, 0) \times (1, 0, 0)_{12}$ pomocou Box-Piercovovej Q štatistiky. P-hodnota nám v oboch prípadoch vyšla väčšia ako 0,05, teda nezamietame nulovú hypotézu o nezávislosti rezíduí. Nakoniec overíme normalitu rozdelenia pravdepodobností pomocou Shapiro-Wilkovho testu, ktorý je vhodný pre dáta s menším počtom pozorovaní. Hypotézu H_0 a teda normalitu dát nezamietame iba pre model $\text{SARIMA}(0, 0, 0) \times (1, 0, 0)_{12}$, s ktorým teda budeme ďalej pracovať. Podmienka stacionarity a invertibility procesu je automaticky kontrolovaná. Program R vypíše chybové hlásenie v prípade nesplnenia jednej z podmienok. V našom prípade sú aj tieto podmienky splnené. Parametre modelu sme určili metódou maximálnej vierohodnosti nasledovne:

$$x_t - 25,86 = 0,86(x_{t-12} - 25,86) + w_t,$$

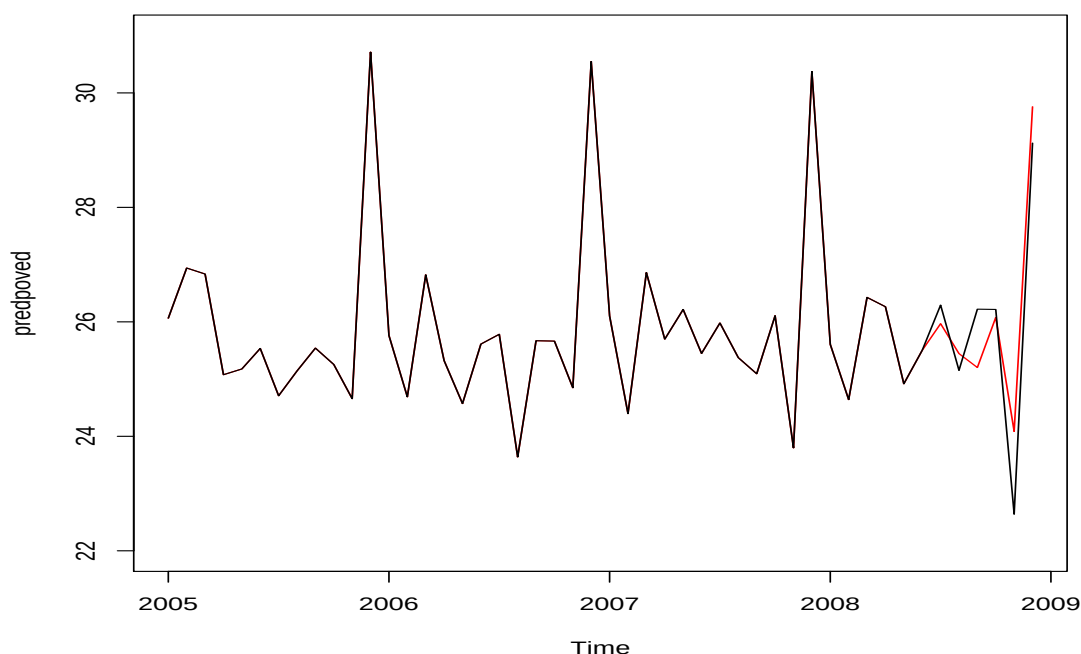
čo ľahko prepíšeme do tvaru:

$$x_t = 3,62 + 0,86(x_{t-12}) + w_t.$$

Ako sme už uviedli v podkapitole 3.4.3, ďalšou možnosťou, ako overiť vhodnosť modelu, môže byť vytvorenie predpovede ex-post. Zmeníme vzorku dát na obdobie od roku 2005 do júna 2008.

Na základe nich a nami vybraného modelu budeme robiť predikcie na druhú polovicu roku 2008. Predikcie na základe SARIMA modelu červenou a pôvodné dáta čiernou farbou vidíme na obrázku 8. Predpovede s reálnymi hodnotami porovnáme na základe RMSE, pričom budeme požadovať čo najnižšiu hodnotu. RMSE pre tieto predikcie vyšla rovná 0,79.

V programe R je taktiež implementovaná funkcia `HoltWinters()`, ktorej vstupným údajom je analyzovaný časový rad a tiež informácia o tom, či chceme využiť aditívnu alebo multiplikatívnu metódu. V prípade využitia Holt-Wintersovej metódy pomocou vyššie uvedenej funkcie nám vyšla RMSE pre multiplikatívnu dekompozíciu s vyrov-



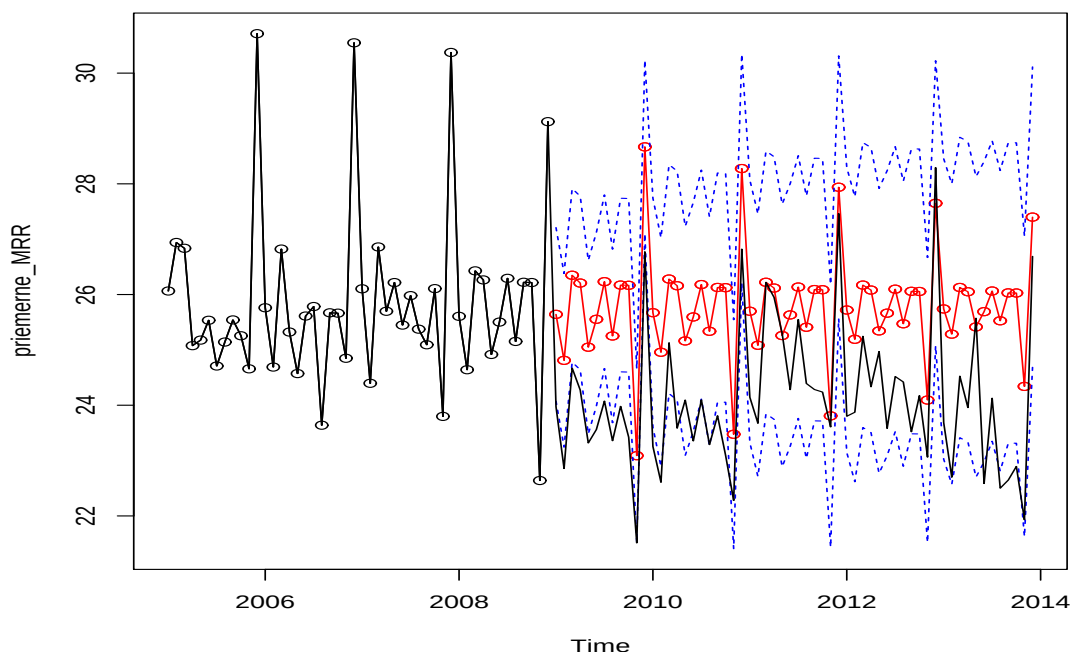
Obr. 8: Skušobná predpoveď ex-post priemerných MRR na základe modelu $\text{SARIMA}(0,0,0) \times (1,0,0)_{12}$

návajúcimi konštantami $\alpha = 0,0083$, $\beta = 0$, $\gamma = 0,6919$ rovná 0,916. Pri aditívnej metóde s vyrovnávajúcimi konštantami $\alpha = 0,0102$, $\beta = 0$, $\gamma = 0,6935$, sa RMSE rovná 0,922. V oboch prípadoch je na základe tejto hodnoty pre rad priemerných MRR lepšie využiť na predpovedanie ich vývoja SARIMA model opísaný vyššie.

Posledným krokom tejto analýzy je na základe nami vybraného modelu, teda $\text{SARIMA}(0,0,0) \times (1,0,0)_{12}$, urobiť predikcie na obdobie od roku 2009 do roku 2013. Tieto odhady červenou spolu s ich intervalmi spoľahlivosti modrou a s vykreslenými pôvodnými dátami čiernou vidíme na obrázku 9. Priemerný rozdiel skutočných a odhadnutých hodnôt je $-1,66$ percentuálneho bodu. To znamená, že v období krízy nastali na Slovensku zmeny, ktoré viedli k zníženiu priemerných mzdových rodových rozdielov.

Ďalej sa teda budeme zaoberať skúmaním dôvodov zníženia MRR. Príčin môže byť viacero, napríklad:

1. Pomalšie zvyšovanie miezd mužov ako žien v období krízy.



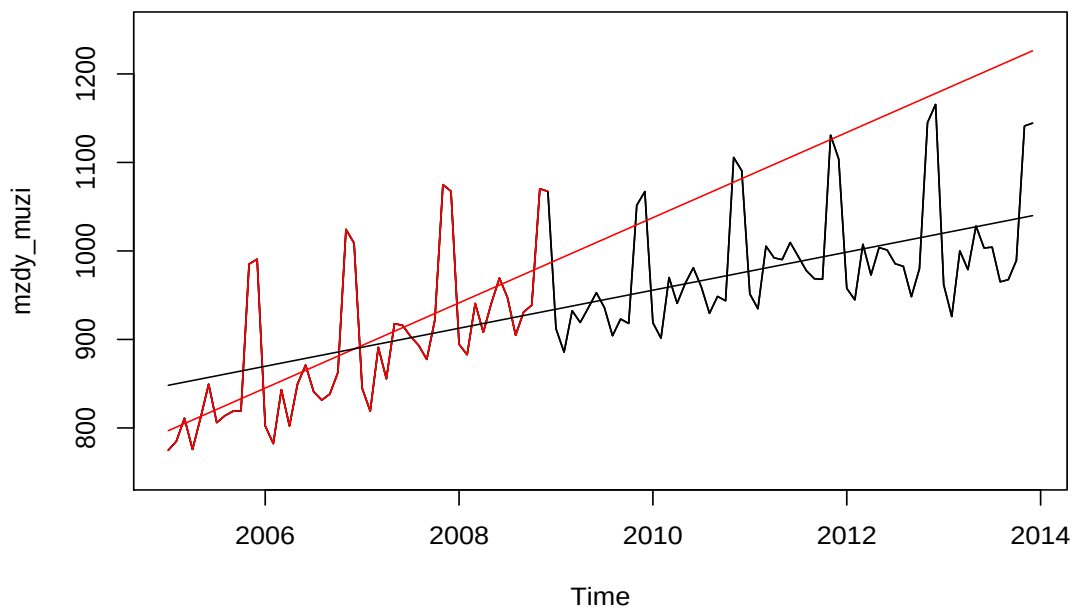
Obr. 9: Odhad vývoja priemerných MRR na základe modelu $SARIMA(0, 0, 0) \times (1, 0, 0)_{12}$ v období od roku 2009 do 2013

2. Dôsledkom finančnej krízy by mohlo byť znižovanie počtov zamestnancov s vysoko nadpriemerným platom (my sme zobrali hranicu 3000 eur), pričom predpokladáme, že v skupine ľudí s platom vyšším ako 3000 eur sú prevažne muži.

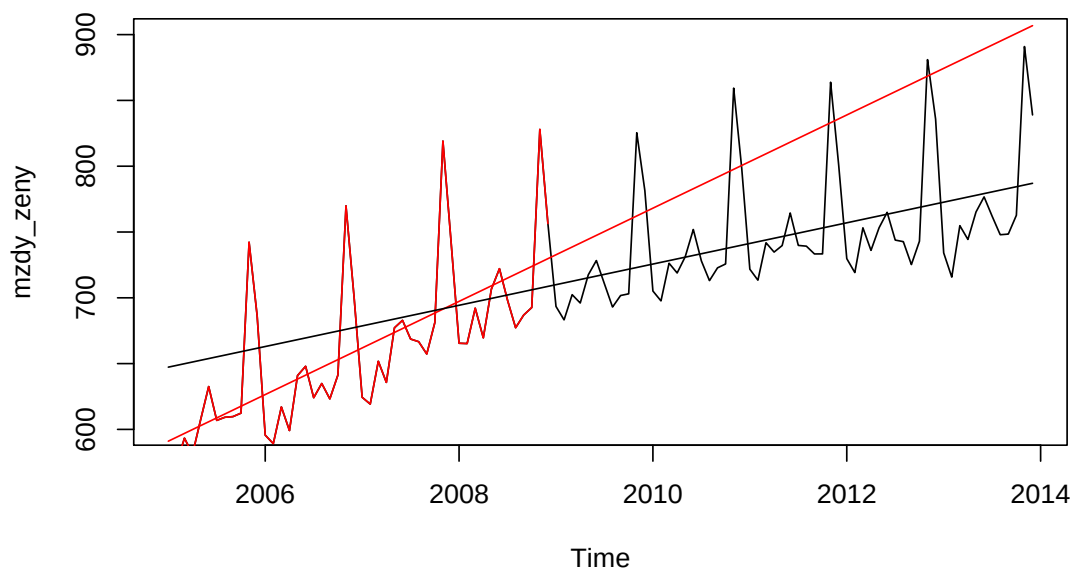
Obidve možnosti by mohli byť zapríčinené segregáciou na trhu práce a tým, že ženy obsadzujú miesta, ktorých sa kríza až tak nedotkla. V tom prípade by sa mohlo ukázať, že napriek tomu, že ženy majú prevažne nižší plat, ich pracovné miesta sú menej ohrozené prípadnými šokmi v ekonomike. V ďalšej časti skúsime overiť platnosť našich predpokladov o príčinách mzdových rodových rozdielov.

4.3.2 Priemerné mzdy mužov a žien

V nasledujúcej časti sa bližšie pozrieme na vývoj priemerných miezd mužov a žien. Zaujímavé výsledky môže priniesť štúdium oddelených zložiek nami zvolených časových radov. Pomocou trendovej zložky radov vieme porovnávať rýchlosť rastu v našom prípade časového radu výšky miezd. Na Obr. 10 a Obr. 11 je zaznamenaný vývoj prie-



Obr. 10: Vývoj priemernej mzdy muža a odhadnuté lineárne trendové priamky.



Obr. 11: Vývoj priemernej mzdy ženy a odhadnuté lineárne trendové priamky.

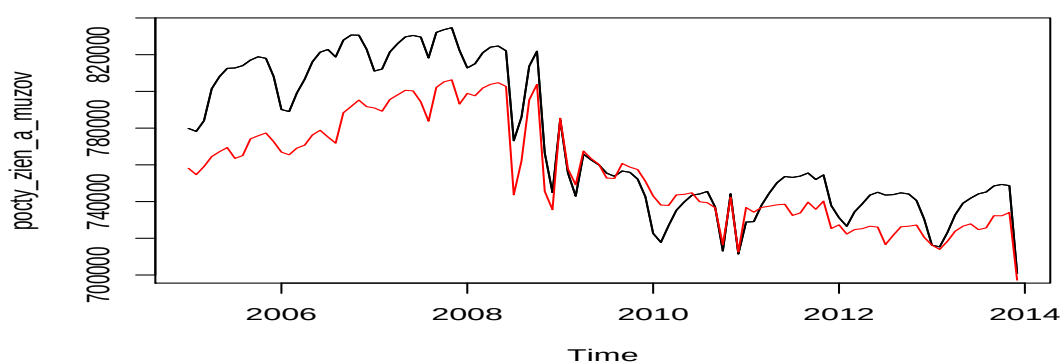
mernej mzdy muža a ženy. Z týchto obrázkov môžeme vidieť, že v období od roku 2005 do roku 2008 bola rýchlosť rastu miezd v oboch prípadoch približne rovnaká, pričom po roku 2008 sa začala znižovať. Takéto správanie je pozorovateľné na mzdách mužov aj žien, pričom sa môže zdať, že u žien je opísaná zmena trendu menej výrazná. Bližšie sa teda pozrieme na naše tušenie pomocou odhadov lineárnych trendových priamok. Pre každý z radov odhadneme nasledovné dve lineárne trendové priamky metódou minimalizácie sumy štvorcov odchýlok odhadov od skutočných pozorovaní:

1. Priamka znázornená červenou farbou je odhadnutá na základe dát z rokov 2005 - 2008, teda prislúcha k mzdám z obdobia pred krízou, ktoré sú znázornené červenou farbou.
2. Priamka znázornená čiernou farbou je odhadnutá na základe dát od začiatku roka 2009, teda prislúcha k mzdám z obdobia krízy, ktoré je znázornené čiernou farbou.

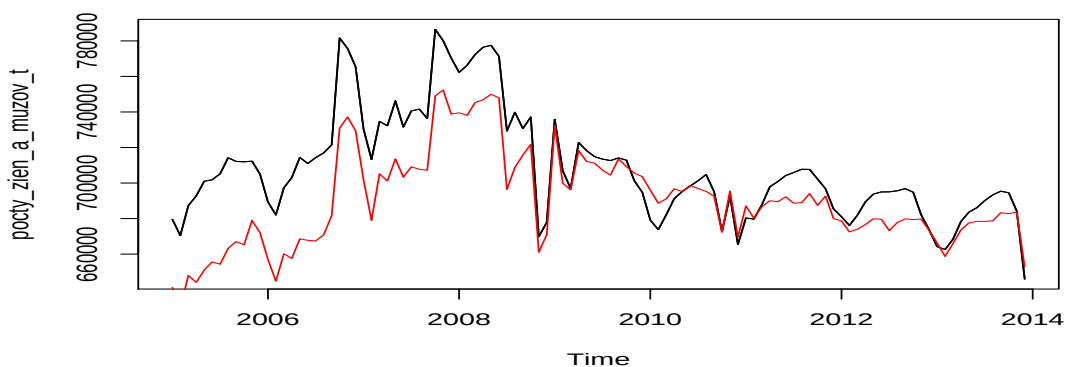
Z obrázkov nemusí byť celkom jasné, ktorého časového radu sa vývojový trend zmenil viac. Za porovnávaciu hodnotu zoberieme rozdiel hodnôt červenej a čiernej trendovej priamky v poslednom odhadnutom bode 1.12.2013, ako percento z priemerného platu mužov a žien. U mužov nám vyšiel koncový rozdiel jednej a druhej trendovej priamky -188,34, čo vieme interpretovať nasledovne. Ak by sa vyvíjali mzdy mužov takou rýchlosťou ako pred krízou, mal by priemerný muž na konci roka 2013 plat zhruba o 188,34 eur vyšší v porovnaní s realitou. Tento rozdiel vyjadrený v percentách z priemernej mzdy muža za obdobie 2005 - 2013 je rovný 19,95%. U žien nám vyšiel koncový rozdiel jednej a druhej trendovej priamky -122,71 čo vieme interpretovať nasledovne. Ak by sa vyvíjali mzdy žien takou rýchlosťou ako pred krízou, mala by priemerná žena na konci roka 2013 plat zhruba o 122,71 eur vyšší v porovnaní s reálnou priemernou mzdou žien. Tento rozdiel vyjadrený v percentách z priemernej mzdy ženy za obdobie 2005 - 2013 je rovný 17,31%. Je teda vidieť, že rýchlosť rastu priemernej mzdy mužov sa znížila výraznejšie ako rýchlosť rastu priemernej mzdy žien na Slovensku.

4.3.3 Analýza počtov v jednotlivých platových skupinách

Na Obr. 12 je zobrazený vývoj počtu žien a mužov za analyzované obdobie. Na Obr. 13, 14 a 15 sú tieto počty rozdelené do troch skupín podľa výšky zárobku jednotlivca v danom mesiaci. Prvá skupina je najväčšia a zahŕňa osoby zarábajúce od 300 do 3000 eur. Druhá skupina je skupinou osôb zarábajúcich nad 3000 eur a tretia skupina - osoby zarábajúce pod 300 eur.



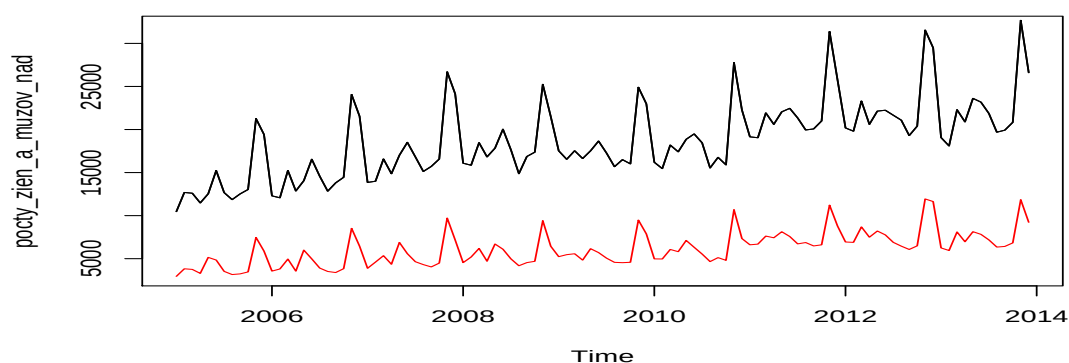
Obr. 12: Priemerné počty všetkých mužov (čiernou) a žien (červenou).



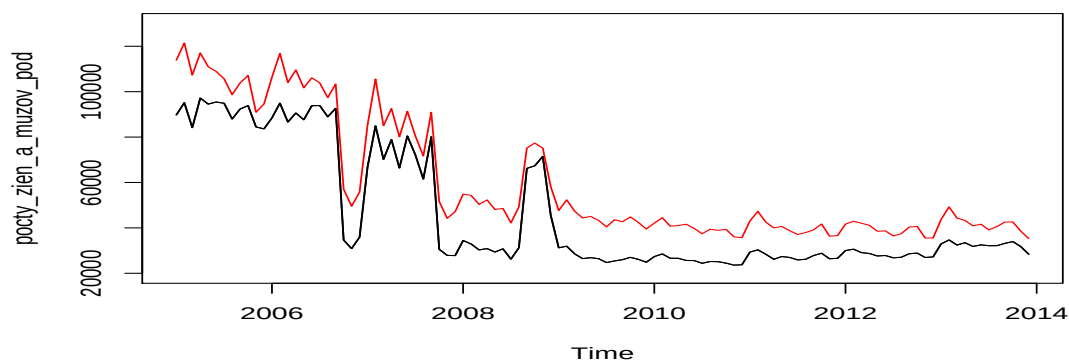
Obr. 13: Priemerné počty mužov (čiernou), žien (červenou) zarábajúcich 300 - 3000 eur.

Z Obr. 12 a 13 môžeme identifikovať podobný vývoj celkových počtov a počtov v najväčšej platovej skupine (od 300 do 3000 eur), čo sa dá očakávať. V oboch prípadoch je jasné zníženie počtov mužov a žien prechodom z roku 2008 na rok 2009, pričom počty

mužov sa znížily výraznejšie a dostali sa zruba na úroveň počtu žien. Priemerný počet mužov (v rokoch 2005 - 2008) bol 811 594 a v období 2009 - 2013 sa znížil na 741 543. To je zníženie zruba o 8,63 % z pôvodného priemerného počtu mužov. Priemerný počet žien (v rokoch 2005 - 2008) bol 781 214 a v období 2009 - 2013 sa znížil na 735 816. To je zníženie zruba o 5,81 % z pôvodného priemerného počtu žien. Pozorujeme teda vyrovnanie podielu mužského a ženského zastúpenia medzi zamestnancami.



Obr. 14: Priemerné počty mužov (čiernou), žien (červenou) zarábajúcich nad 3000 eur.



Obr. 15: Priemerné počty mužov (čiernou), žien (červenou) zarábajúcich pod 300 eur.

Pre výberové skupiny, ako je skupina s platom vyšším ako 3000 eur a nižším ako 300 eur však nevieme identifikovať nejakú výraznú zmenu vo vývoji príchodom finančnej krízy v rozložení počtov v jednotlivých skupinách medzi mužmi a ženami. Na Obr. 14 vidíme veľký rozdiel v počtoch mužov a žien zarábajúcich nad 3000 eur. Ženy v

celom sledovanom období tvoria menej ako 30 % celkového počtu osôb v tejto platovej kategórii a počas krízy nenastal nijaký výrazný zlom. Na Obr. 15 je pozorovateľné výrazné zníženie počtu mužov a žien, ktorí zarábajú menej ako 300 eur. Toto zníženie je približne rovnaké v prípade mužov a žien.

4.4 Zhrnutie získaných výsledkov

V predošlých častiach kapitoli sme pozreli na vývoj priemerných miezd mužov a žien či počtov zamestnaných osôb v jednotlivých obdobiach v rôznych platových skupinách. Táto analýza miezd a počtov nás doviedla k presvedčeniu, že zníženie mzdových rodových rozdielov nie je zapríčinené zlepšením postavenia žien na trhu práce, skôr naopak zhoršením postavenia mužov v období finančnej krízy.

V Tabuľke 4, ktorá je rozdelená do štyroch skupín podľa výšky platu sa bližšie pozrieme na odhadnuté a vypočítané charakteristiky v jednotlivých platových skupinách a zároveň aj na ich jednotky, ktoré v ďalších textoch nebudeme uvádzať vždy. Tabuľka budeme tvorená nasledovnými charakteristikami:

- Priemerné MRR - priemerné MRR za obdobie od roku 2005 do 2013 vyjadrené v %, vypočítané na základe vzorca 1.
- Zmena MRR - priemerný rozdiel odhadnutých a skutočných hodnôt vypočítaný ako v podkapitole 4.3.1 v percentuálnych bodoch.
- Percentuálna zmena MRR - podiel Zmeny MRR z Priemerných MRR v percentách.
- Zmena mzdy (mužov a žien) - zmena vo vývoji mzdy vypočítaná ako v podkapitole 4.3.2 v eurách.
- Zmena mzdy (mužov a žien) v % - zmena vo vývoji mzdy ako percentuálna hodnota z priemernej mzdy pred rokom 2009 vypočítaná ako v podkapitole 4.3.2 v eurách.
- Počet (mužov a žien) - priemerný počet zamestnaných osôb alebo osôb pracujúcich na dohodu za obdobie 2005 - 2013.
- Počet (mužov a žien) po v % - podiel priemerného počtu zamestnancov z obdobia 2009 - 2013 z priemerného počtu zamestnancov z obdobia 2005 - 2013 v percentách.

V uvedenej Tabuľke 4 sú niektoré zaujímavé čísla, ktoré ešte neboli spomenuté, zvýraznené červenou farbou. Prvým zaujímavým faktom je to, že priemerné mzdové

Tabuľka 4: Výsledky jednotlivých ukazovateľov MRR, vývoja platov a počtov osôb v rôznych platových skupinách.

	Všetci zamestnanci	Zarábajúci 300 - 3000 €	Zarábajúci viac ako 3 000 €	Zarábajúci menej ako 300 €
Priemerné MRR	24,90	16,34	15,81	-1,75
Zmena MRR	-1,66	-0,46	-5,87	-7,19
Zmena MRR %	-6,67	-2,82	-37,13	-410,86
Zmena mzdy muža v €	-186,264	-101,43	-399,82	16,67
Zmena mzdy ženy v €	-119,80	-91,12	446,78	17,20
Zmena mzdy muža v %	-19,72	-11,96	-6,35	7,70
Zmena mzdy ženy v %	-16,88	-12,84	8,47	7,84
Počet mužov	772 677	708 118	18 739	45 821
Počet mužov po v %	91,37	95,20	127,95	41,56
Počet žien	755 993	690 000	6 107	59 886
Počet žien po v %	94,19	99,27	140,11	49,23

rodové rozdiely sa znížili zmenšením výberu na zarábajúcich 300 - 3000 eur na hodnotu 16,3 % a počas krízy neutrpeli podľa našich odhadov až také výrazné zníženie ako bolo odhadnuté pre všetkých zamestnancov. Jedine pre osoby s platom nižším ako 300 eur nám vyšla priemerná hodnota MRR záporná, čo značí vyššie priemerné platy žien ako mužov v tejto skupine. Mzdové rozdiely medzi mužmi a ženami sa najvýraznejšie znížili taktiež pre štvrtú platovú skupinu. Dôležité je pripomenúť význam hodnoty zmeny mzdy ženy v treťom stĺpci, ktorá je rovná 446,78 eur. Takáto výrazná hodnota nám vyšla na základe toho, že pred rokom 2009 mzdy žien zarábajúcich viac ako 3000 eur klesali a počas finančnej krízy sa ustálili približne na hodnote, aká bola začiatkom roku 2009. Teda sa zmenil trend z klesajúceho na konštantný, čo spôsobilo vyššie priemerné platy na konci roku 2013 ako sme očakávali o 446,8 eur.

5 Analýza mzdových rodových rozdielov v rôznych okresoch

Nasledujúca kapitola je zameraná na analýzu mzdových rodových rozdielov zvlášť pre jednotlivé okresy. V našich dátach rozlišujeme 74 okresov, v ktorých sú zahrnuté všetky okresy s nasledujúcimi výnimkami. Skupina 0 obsahuje údaje o obyvateľoch nepriradených k žiadnemu z okresov, skupina 900 obsahuje údaje o osobách žijúcich mimo Slovenska, pracujúcich na Slovenku, skupina 101 zahŕňa okresy Bratislavu 1 - 5, tak isto skupina 802 zahŕňa Košice 1 - 4. Pre všetky okresy sme vypočítali priemerné MRR, priemerný počet žien a mužov a percentuálnu hodnotu podielu pracujúcich žien a mužov po 2009 a pred 2009, ako aj percentuálnu hodnotu podielu pracujúcich žien a mužov s platom vyšším ako 3 000 eur po 2009 a pred 2009. Pomocou zostavených cyklov v programe R, uvedených v prílohe B a D, sme odhadli charakteristiky zmena MRR a rozdiel miezd žien a mužov v %. Čo si čitateľ môže pod uvedenými charakteristikami predstaviť, sme uviedli v podkapitole 4.4.

Bližšie si ale opíšeme cyklus, ktorým sme sa snažili zautomatizovať výber vhodného modelu pre časové rady mzdových rodových rozdielov jednotlivých skupín a pomocou ktorého sme dostali väčšinu hodnôt zmien MRR. Uvedený cyklus vykonáva všetky operácie uvedené v 4.3.1. Pomocou hodnoty RMSE vyberá najlepší model z troch navrhnutých. Prvým je model SARIMA, vybraný funkciou `auto.arima()` (s využitím kpss testu a ocsb testu na identifikáciu jednotkového koreňa), a ďalšie dva modely sú multiplikatívny a aditívny model vypočítaný Holt-Wintersovou metódou. Pri nespĺňaní podmienky nezávislosti rezíduí u vybraného SARIMA modelu sa následne vyberie lepší z dvoch Holt-Wintersových modelov. Podmienku normality rezíduí sme v cykle nebrali do úvahy pre dôvod uvedený v podkapitole 3.4.3. Automaticky sa tiež vykresľujú predpovedané hodnoty, a tak môžeme kontrolovať správnosť vybraného modelu aj vizuálne. Pri spustení cyklu pre tabuľku priemerných mzdových rodových rozdielov všetkých okresov sme ako výstup dostali vektor rozdielov odhadnutých MRR a skutočných hodnôt priemerných MRR za obdobie 2009 - 2013. Dôležité je spomenúť, že pre 6 okresov sme sa pozreli bližšie na vybraný model a následne vyhodnotili najlepší model bez pomoci cyklu, práve kvôli tomu, že sa nám vykreslený priebeh predpovedaných hodnôt nezdal

najlepším možným pre konkrétne dáta.

Ďalej treba uviesť, že pred tým, ako sme odhadovali trendové priamky pre priemerné mzdy mužov a žien v jednotlivých okresoch, sme v analyzovaných radoch našli outlierov a nahradili ich hraničnými hodnotami. Hranice sme určili výpočtom 1. kvartilu Q_1 a 3. kvartilu Q_3 a hodnoty medzikvartilového intervalu $IQR = Q_3 - Q_1$. Dolnú hranicu sme položili rovnú $Q_1 - 1,5 * IQR$ a hornú hranicu $Q_3 + 1,5 * IQR$. Tento postup vyhodnocovania a nahradenia outlierov budeme používať aj pri clustrových analýzach. Pri analýze radov MRR chceme zachovať ich často výraznú sezónnosť, preto nahrádzanie outlierov vynechávame.

5.1 Deskriptívna analýza

Po odhadnutí zmien MRR pre jednotlivé okresy pomocou vyššie uvedeného cyklu a prepočítaní týchto hodnôt na percentuálny podiel z prislúchajúcich priemerných MRR sme vypočítali aj priemernú hodnotu týchto percentuálnych zmien, váženú počtom zamestnancov v okresoch. Táto hodnota nám vyšla -6,22%. Podobne sme urobili pre odhadnuté hodnoty rozdielu miezd žien v % a rozdielu miezd mužov v %. Uvedené hodnoty nám vyšli -17,91 % pre mzdy žien a -20,66 % pre mzdy mužov. Tieto vážené priemerné hodnoty odhadnutých zmien MRR a miezd sme porovnali s hodnotami odhadnutými v podkapitolách 4.3.1 a 4.3.2, kde nám vyšla percentuálna zmena MRR -6,67 %, zmena mzdy ženy -16,88 % a zmena mzdy muža -19,72 %. Keďže sa tieto hodnoty odhadnutých zmien MRR či miezd líšia nanejvýš zhruba o jeden percentuálny bod, môže to byť znak správnosti našich odhadov.

V nasledujúcej Tabuľke 5 sú uvedené okresy s najnižšími a najvyššími MRR, zmenami MRR a zmenami miezd mužov a žien a ich hodnoty. Hodnoty odhadnuté pre ostatné okresy, ako aj priemery a farebné odlíšenie maxima a minima jednotlivých charakteristík, sú uvedené v prílohe C a na CD v prílohe D.

Mzdové rodové rozdiely podľa našich odhadov klesli v porovnaní s očakávaným vývojom v 48 okresoch, naopak vzrástli v 26 okresoch. V prípade spojenia medzi zmenou mzdových rodových rozdielov a zmenou vývoja miezd žien a mužov sa naše očakávania naplnili v 62 okresoch. Teda platilo, že ak MRR rástli, rástli mzdy mužov počas krízy rýchlejšie ako mzdy žien počas kríz, v porovnaní s obdobím pred krízou a naopak. V

Tabuľka 5: Okresy s najnižšími a najvyššími MRR, zmenami MRR a zmenami miezd mužov a žien a ich hodnoty.

MRR	Zmena MRR v %	Zmena mzdy ženy v %	Zmena mzdy muža v %
Bardejov 11,57	S. Ľubovňa -49,70	Myjava -26,29	S. Ľubovňa -31,95
Púchov 35,88	Lučenec 26,94	Osoby mimo SR 12,40	Osoby mimo SR 23,09

případe spojenia medzi zmenou mzdových rodových rozdielov a zmenou v zamestnanosti v skupine osôb zarábajúcich nad 3000 eur sa očakávania naplnili v 55 okresoch. Teda pre nich platí, že ak MRR rástli, percentuálny podiel mužov, zarábajúcich nad 3000 eur, v období krízy ku obdobiu pred krízou bol väčší ako u žien. Teda výraznejšie vzrástol počet mužov v tejto skupine ako žien vzhľadom k obdobiu pred krízou. A naopak. Tento ukazovateľ však nič nehovorí o tom, ako sa menili mzdy mužov a žien v tejto skupine, teda je pochopiteľné, že táto kauzalita nám nevyšla vo všetkých okresoch.

5.2 Zhluková analýza

Na hľadanie súvislostí medzi jednotlivými okresmi sa najlepšie hodí zhluková analýza. Keďže zhluk predstavuje skupinu objektov, ktorých vzdialenosť, to znamená nepodobnosť, je menšia ako vzdialenosť objektov do zhliku nepatriacich, vieme pomocou nej nájsť okresy s podobnými charakteristikami a lepšie tak sledovať možné príčiny správania sa mzdových rodových rozdielov. Príkazy využívané pri zhlukovej analýze a jej výstupy sú uvedené v prílohe B a D aj s popismi.

5.2.1 Korelačná analýza pre charakteristiky okresov

Dôležitou časťou zhlukovej analýzy je obmedziť vstupy korelovaných premenných do jej algoritmov. Preto sa najskôr pozrieme na to, či nie sú v nami vybraných charakteristikách jednotlivých okresov také, ktoré spolu úzko súvisia, teda sú výrazne korelované. Pomocou funkcie `cor()`, ktorá automaticky vypočítava Pearsonov korelačný koeficient, ak nie je zadané inak, si necháme vypísať korelačnú maticu, ktorej hodnoty sú uvedené v Tabuľke 5. Taktiež si necháme vykresliť graf, na ktorom sú zobrazené všetky dvojice charakteristík. (P.MRR - priemerné MRR, P.z.MRR - percentuálna zmena MRR,

R.m.ž.% - rozdiel miezd žien v %, R.m.m.% - rozdiel miezd mužov v %, P.p.ž - priemerný počet zamestnaných žien, P.ž. - percentuálny podiel počtov zamestnaných žien pred 2009 a po 2009, P.p.m - priemerný počet zamestnaných mužov, P.m. - percentuálny podiel počtov zamestnaných mužov pred 2009 a po 2009, P.ž.3000 - percentuálny podiel zamestnaných žien pred 2009 a po 2009 s platom vyšším ako 3000 eur, P.m.3000 - percentuálny podiel zamestnaných mužov pred 2009 a po 2009 s platom vyšším ako 3000 eur). Veľká korelácia nám vyšla pre charakteristiky P.p.ž a P.p.m, pre P.ž a P.m a pre P.ž.3000 a P.m.3000. Tieto hodnoty sú v Tabuľke 6 zvýraznené červenou farbou. Kvôli týmto veľkým koreláciám sme sa rozhodli nahradiť počty mužov a žien ich súčtom, teda priemerným počtom zamestnaných osôb za obdobie 2005 - 2013 - P.počet, hodnoty podielu pracujúcich žien a podielu pracujúcich mužov po 2009 a pred krízou hodnotou ich rozdielu - R.podielu ž. a m., taktiež v prípade žien a mužov zarábajúcich nad 3000 eur - R.podielu nad. Opätovným vypísaním korelácií sme overily vhodnosť týchto úprav. Ďalej budeme preto pracovať s týmito charakteristikami, v ktorých nám najvyššia korelácia vyšla 0,58 medzi P.z.MRR a R.m.m.%.

Tabuľka 6: Tabuľka korelácií medzi jednotlivými charakteristikami okresov.

P.MRR	0,22	-0,30	-0,06	0,37	0,31	0,39	0,45	-0,28	-0,22
0,22	P.z.MRR	0,02	0,58	0,02	0,32	0,06	0,24	0,01	0,18
-0,30	0,02	R.m.ž.%	0,56	-0,18	0,05	-0,15	0,03	0,35	0,34
-0,06	0,58	0,56	R.m.m.%	-0,12	0,17	-0,07	0,14	0,39	0,45
0,37	0,02	-0,18	-0,12	P.p.ž.	0,23	0,99	0,31	-0,24	-0,13
0,31	0,32	0,05	0,17	0,23	P.ž.	0,24	0,89	0,16	0,41
0,39	0,06	-0,15	-0,07	0,99	0,24	P.p.m.	0,33	-0,23	-0,12
0,45	0,24	0,03	0,14	0,31	0,89	0,33	P.m.	-0,02	0,29
-0,28	0,01	0,35	0,39	-0,24	0,16	-0,23	0,02	P.ž.3000	0,67
-0,22	0,18	0,34	0,45	-0,13	0,41	-0,12	0,29	0,67	P.m.3000

Pri analýze korelácií sme sa zvlášť pozreli aj na vzťah medzi percentuálnymi zmenami MRR a rozdielom rozdielov miezd žien a rozdielom miezd mužov. V tomto prípade vyšiel Pearsonov korelačný koeficient pomerne vysoký, rovný -0,67. Táto skutočnosť poukazuje na významný vzťah medzi týmito dvoma charakteristikami, čo sa dalo pred-

pokladať a uviedli sme to aj v podkapitole 4.3.1 a v predošlej deskriptívnej analýze. Taktiež sme sa pozreli na koreláciu medzi zmenami MRR a hodnotami rozdielov podielov pracujúcich žien a podielov pracujúcich mužov po 2009 a pred krízou, zarábajúcich nad 300 eur. Tu nám korelačný koeficient vyšiel rovný -0,23, čím sa nepotvrdila výrazná spojitosť medzi zmenou mzdových rodových rozdielov a zmenou počtu zamestnancov v skupine žien a mužov zarábajúcich nad 3000 eur.

5.2.2 Štandardizácia dát

V predchádzajúcej kapitole sme skúmali jednotlivé charakteristiky nezávisle od iných štatistických znakov súboru dát. Ak však sú rôzne charakteristiky našich okresov uvádzané v rôznych jednotkách a mierkach, je potrebné pred využitím algoritmov zhukovej analýzy zabezpečiť odstránenie závislosti na jednotkách a parametroch polohy alebo rozptylu. Takáto forma úpravy sa nazýva štandardizácia dát. Metód štandardizácie dát je niekoľko a každá má svoje výhody aj nevýhody. My sme si vybrali predelenie hodnôt jednotlivých charakteristík rozdielom ich maxima a minima. Teda:

$$stand.data_{i,j} = \frac{x_i}{x_{i,max} - x_{i,min}}$$

kde i sú vybrané charakteristiky a j sú okresy. Teda v našom prípade $i=1,2,\dots,6$ a $j=1,2,\dots,74$.

5.2.3 Metóda k-priemery

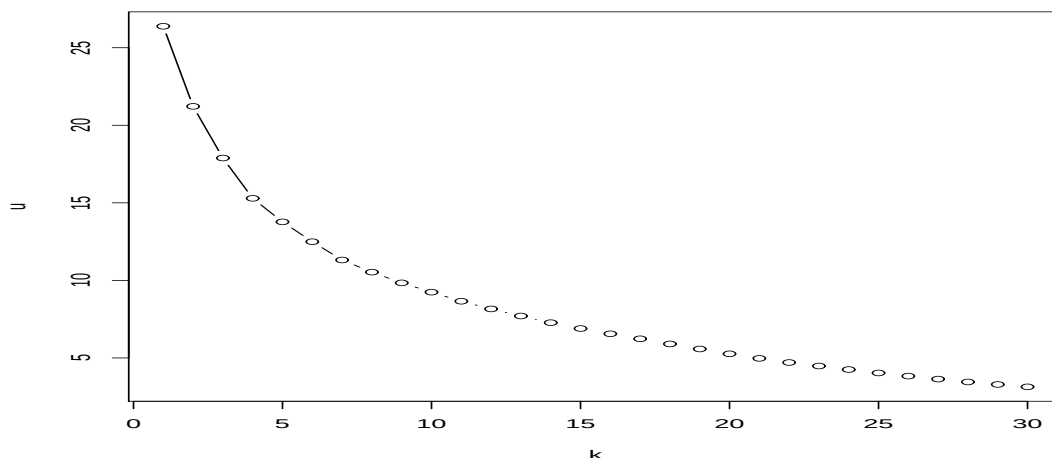
Metóda k-priemery alebo inak metóda k-means patrí medzi nehierarchické algoritmy a triedi dáta do k zhukov na základe ich vlastností. Túto metódu po prvý krát publikoval MacQueen v roku 1967. Algoritmus pracuje tak, že zaradí v našom prípade okresy do skupín tak, aby bol súčet euklidovských vzdialeností okresov v zhuku od centroidu (ťažiska) minimálny. Centroid zhuku dostaneme priemerom súradníc všetkých okresov v danom zhuku.

Optimálne roztriedenie hľadáme na základe tzv. Lloydovho algoritmu, ktorý funguje nasledovne:

1. Náhodne rozdelí okresy do k zhukov.
2. Vypočíta centroid pre každý zhuk.

3. Vypočíta vzdialenosť okresov od všetkých centroidov a vytvorí nový zhuk podľa minimálnej vzdialenosti okresu od daného centroidu.
4. Postup sa opakuje, až kým rozdelenie zhukov nie je totožné s predošlým.

Počet zhukov k určujeme na začiatku a je to jeden z problémov zhukovej analýzy. Existuje viacero kritérií, no nie je celkom jasné, ktoré z nich je optimálne. V našej práci určíme k na základe analýzy funkcie vzdialeností (celková suma štvorcov vo vnútri zhuku) pre rôzne počty zhukov. Pre $k = 1, 2, \dots, 30$ vypočítame optimálne zhukovania a hodnoty funkcie vzdialeností $u_1, u_2, \dots, u_{30}$. Závislosť u od počtu zhukov je zobrazená na Obr. 16. Na základe "laktového kritéria" vyberieme ten bod, v ktorom si všimneme najväčší zlom (vytvára sa tam tzv. "zaoblený lakeť"), ako optimálny počet zhukov. Na základe Obr. 16 sme ako optimálne k vybrali $k = 8$.



Obr. 16: Závislosť funkcie vzdialeností od počtu zhukov pre okresy.

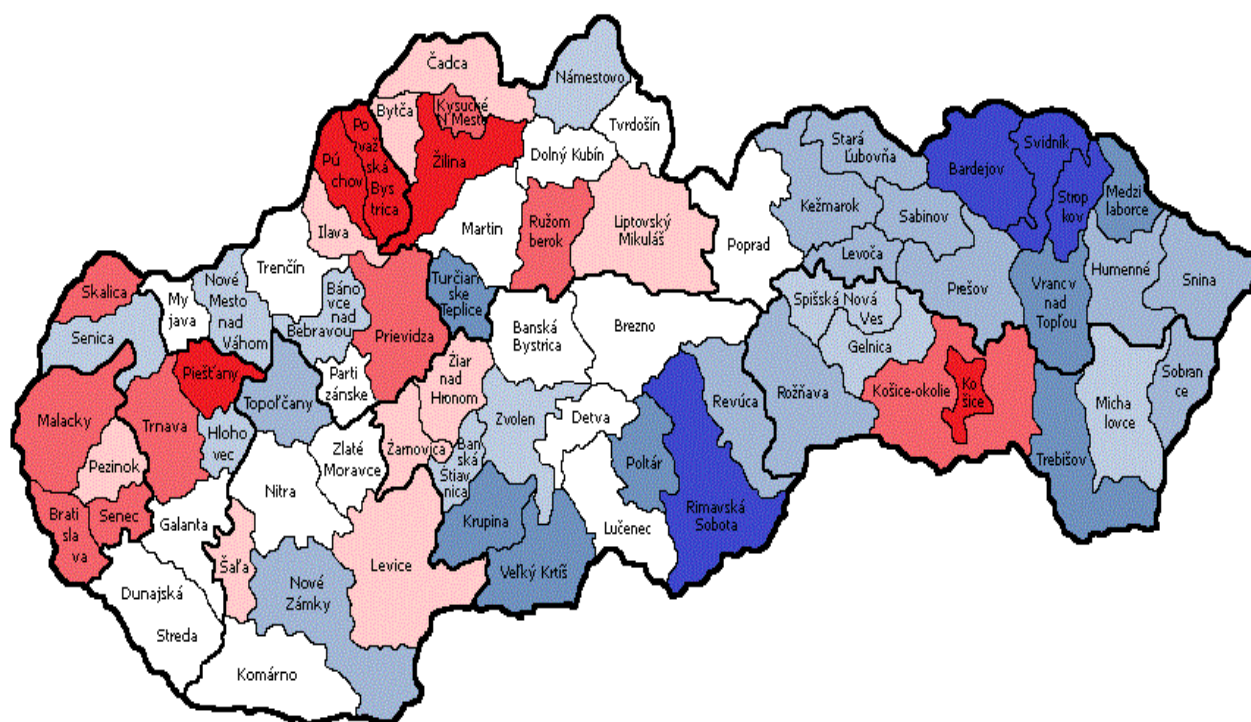
Využitím funkcie `kmeans()` vypočítame centroidy a tak rozdelíme okresy do ôsmich zhukov. Vstupmi tejto funkcie sú nekorelované štandardizované dáta k jednotlivým okresom, k , maximálny počet iterácií a tzv. `nstart`, ktorý hovorí o tom, koľko krát prebehne Lloydov algoritmus s iným náhodným usporiadaním do počiatočných zhukov. My sme zvolili maximálny počet iterácií aj `nstart` najväčšie možné, ktoré umožňuje program R teda: `iter.max=1000000000` a `nstart=1000000`, kvôli zabezpečeniu konvergenencie.

5.2.4 Výsledky zhukovej analýzy pre okresy

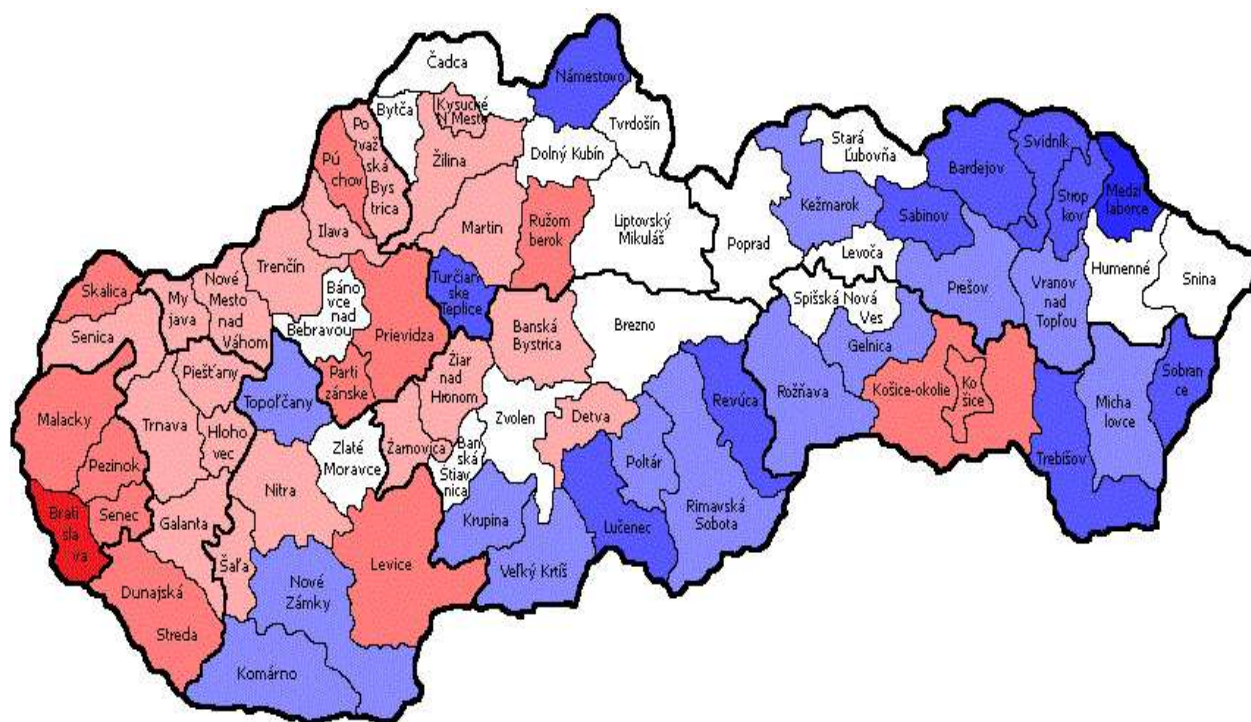
Pomocou zhukovej analýzy sme rozdelili okresy Slovenska do ôsmich skupín s najpodobnejším správaním vybraných charakteristík. Takéto roztriedenie prinieslo niekoľko zaujímavých zistení. Tieto zistenia a jednotlivé zhuky sú opísané nižšie:

- Šiesty zhuk je tvorený iba jednou skupinou " okresom " 900. Charakteristiky skupiny osôb žijúcich mimo Slovenska teda nie sú podobné zo žiadnymi charakteristikami skupín osôb žijúcich v rôznych okresoch na Slovensku.
- Do piateho zhuku boli pomocou metódy k-priemery zaradené všetky okresy krajských miest, teda Bratislava, Trnava, Trenčín, Nitra, Žilina, Banská Bystrica, Prešov a Košice. Okrem toho navyše Prievidza a Nové Zámky. Výrazným znakom tejto skupiny okresov je najvyšší priemerný počet zamestnancov a vo všetkých desiatich okresoch MRR klesali v období finančnej krízy z pôvodnej nadpriemernej hodnoty MRR.
- Väčšina okresov východnej polovice Slovenska bola zaradená do 1., 2. a 8. zhuku. Všetky okresy zaradené v týchto zhukoch majú nízke priemerné MRR, ale líšia sa v charakteristike percentuálnej zmeny MRR počas krízy. V prvom zhuku sa zmenili MRR najviac vzhľadom k priemerným MRR, naopak v druhom zhuku sa zmenili najmenej v porovnaní s ostatnými zhukmi.
- Ostatné okresy, ktoré neboli spomenuté vyššie, sú okresy prevažne západnej a severnej časti Slovenska s vyššími priemernými MRR. Tieto sme taktiež rozdelili do troch zhukov: 3., 4., a 7.. V dvoch z nich priemerná hodnota zmeny MRR vyšla kladne, v jednom záporne.

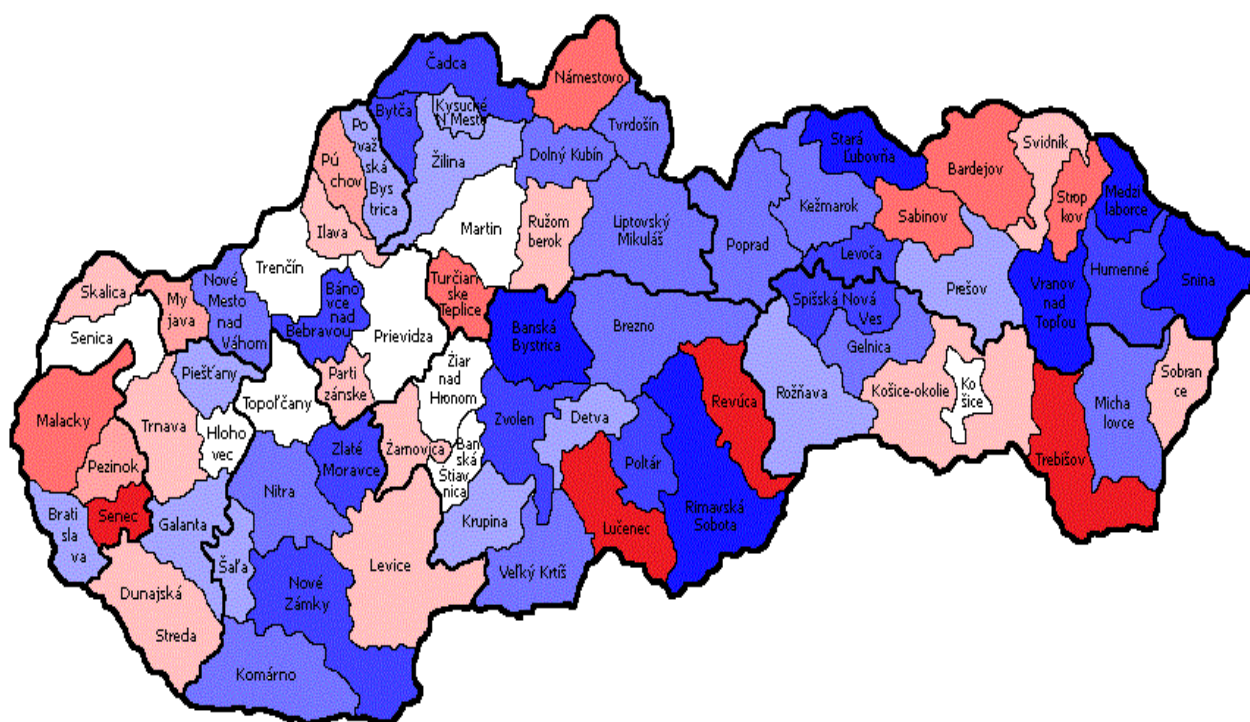
Na lepšie predstavenie si podobnosti medzi jednotlivými zhukmi sme vytvorili farebné mapy Slovenska. Obr. 17 je vyfarbený na základe hodnôt priemerných MRR v jednotlivých okresoch. Obr. 18 je vyfarbený podľa hodnoty MRR centroidu zhuku (teda priemernej hodnoty MRR v zhuku). Druhé dva Obr. 19 a 20 sú vyfarbené na základe zmeny MRR rovnakým štýlom. Modrá farba znamená najnižšie MRR a najväčšie zníženie MRR, červená farba naopak najvyššie MRR a najväčšie zvýšenie MRR.



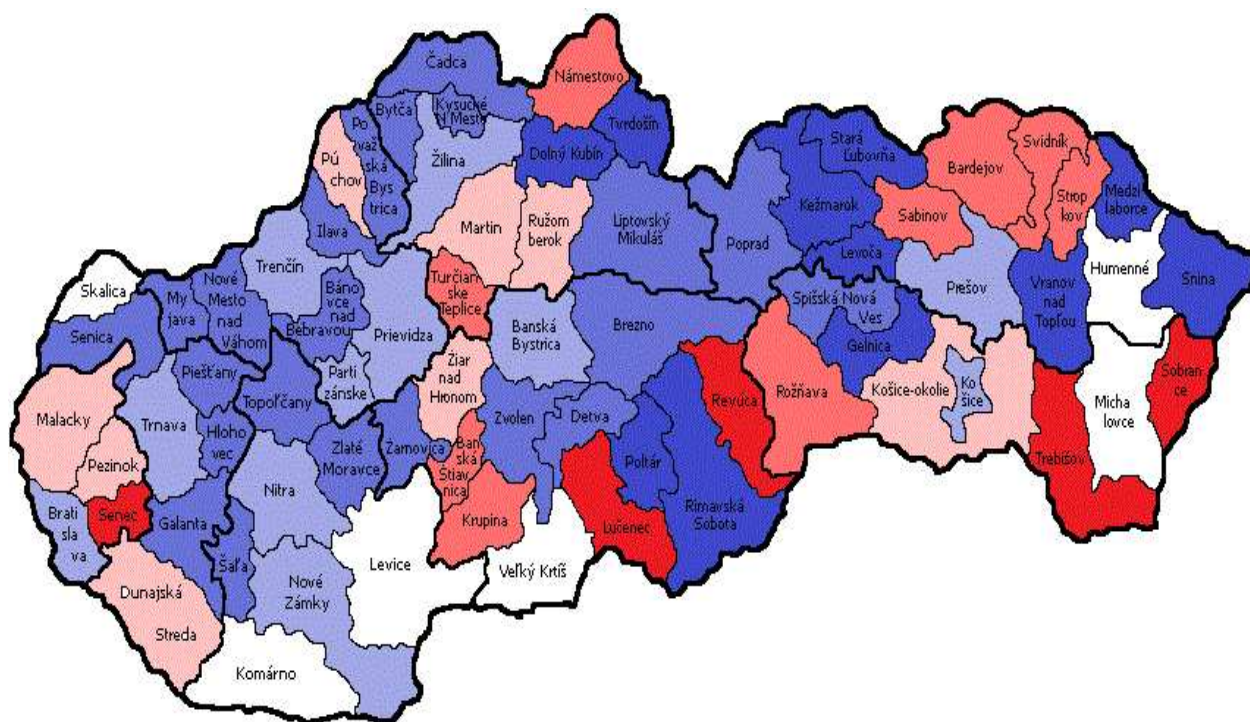
Obr. 17: Mapa Slovenska vyfarbená podľa priemerných MRR okresov (modrá - najmenšie MRR, červená - najväčšie MRR).



Obr. 18: Mapa Slovenska vyfarbená podľa zatriedenia do zhukov (modrá - najnižšia hodnota centroidu MRR, červená - najvyššia hodnota centroidu MRR).



Obr. 19: Mapa Slovenska vyfarbená podľa priemernej zmeny MRR okresov (modrá - najväčší pokles MRR, červená - najväčší nárast MRR).



Obr. 20: Mapa Slovenska vyfarbená podľa zatriedenia do zhukov (modrá - najzápornejšia hodnota centroidu zmeny MRR, červená - najkladnejšia hodnota centroidu zmeny MRR).

6 Analýza mzdových rodových rozdielov v rôznych okresoch a vekových skupinách

Nasledujúca kapitola je zameraná na analýzu mzdových rodových rozdielov zvlášť pre jednotlivé okresy a vekové skupiny. Osoby sme rozdelili do 74 skupín podľa miesta trvalého bydliska a každú z nich následne rozdelili do ôsmich skupín podľa veku. Prvou skupinou sú zamestnanci od 20 do 25 rokov, ďalej 25 - 30 rokov, 30 - 35 rokov, 35 - 40 rokov, 40 - 45 rokov, 45 - 50 rokov a posledná skupina od 55 do 60 rokov. Teda vynechávame osoby mladšie ako 20 rokov a staršie ako 60 rokov. Takýmto rozdelením a následným vypočítaním či odhadnutím jednotlivých charakteristík sme dostali tabuľky obsahujúce vektory MRR, percentuálnych zmien MRR, priemerných miezd mužov a žien a počtov zamestnancov pre 592 skupín.

6.1 Deskriptívna analýza

Po odhadnutí zmien MRR pre jednotlivé skupiny pomocou cyklu uvedeného v 5 na výpočet zmien MRR a prepočítaní týchto hodnôt na percentuálny podiel z prislúchajúcich priemerných MRR sme vypočítali aj priemernú hodnotu týchto percentuálnych zmien váženú počtom zamestnancov v skupinách. Táto hodnota nám vyšla -1,63 %. Podobne sme urobili pre odhadnuté hodnoty rozdielu miezd žien v % a rozdielu miezd mužov v %. Tieto vážené priemerné hodnoty nám vyšli -17,83 % pre mzdy žien a -19,76 % pre mzdy mužov. Predpokladaná kauzalita pre tieto priemerné hodnoty teda zostala zachovaná. Keďže v priemere sa vývoj miezd žien znižoval pomalšie ako vývoj miezd mužov znižovali sa aj MRR. Táto kauzalita je platná pre viac ako 80 % okresov a vekových skupín, teda platí pre 475 skupín. Predpoklad súvislosti medzi zmenou mzdových rozdielov a zvýšením početnosti žien v skupine s platom vyšším ako 3000 eur sa výrazne nenaplnil. V 342 skupinách platil a v ostatných 250 nie. Pozrime sa aj na korelácie, ktoré potvrdili naše výsledky, kde korelačný koeficient medzi zmenou MRR a rozdielom rozdielov miezd žien a rozdielom miezd mužov vyšiel rovný -0,65. A korelačný koeficient medzi zmenou MRR a rozdielom zmeny početnosti žien a mužov v skupine zarábajúcich nad 3000 eur iba -0,08. Mzdové rodové rozdiely podľa našich odhadov klesli v porovnaní s očakávaným vývojom v 366 skupinách, naopak vzrástli v

226 skupinách, čo potvrdzuje domnienku o znižovaní MRR počas krízy.

V nasledujúcej tabuľke sú uvedené okresy s najnižšími a najvyššími MRR, zmenami MRR a zmenami miezd mužov a žien a ich hodnoty. Hodnoty odhadnuté pre ostatné okresy sú uvedené iba na CD v prílohe D (z dôvodu veľkosti tabuľky).

Tabuľka 7: Skupina s najnižšími a najvyššími MRR, zmenami MRR a zmenami miezd mužov a žien a ich hodnoty.

	Najnižšie
MRR	okres Stropkov 20 - 25 roční -4,84
Zmena MRR v %	okres Medzilaborce 20 - 25 roční -3040,07
Zmena mzdy ženy v %	žijúci mimo SR 45 až 50 roční -77,65
Zmena mzdy muža v %	nezaradení do okresu 45 až 50 roční -94,01
	Najvyššie
MRR	žijúci mimo SR 50 až 55 roční 50,38
Zmena MRR v %	okres Stropkov 25 - 30 roční 5181,62
Zmena mzdy ženy v %	žijúci mimo SR 50 až 55 roční 198,84
Zmena mzdy muža v %	žijúci mimo SR 30 až 35 roční 112,55

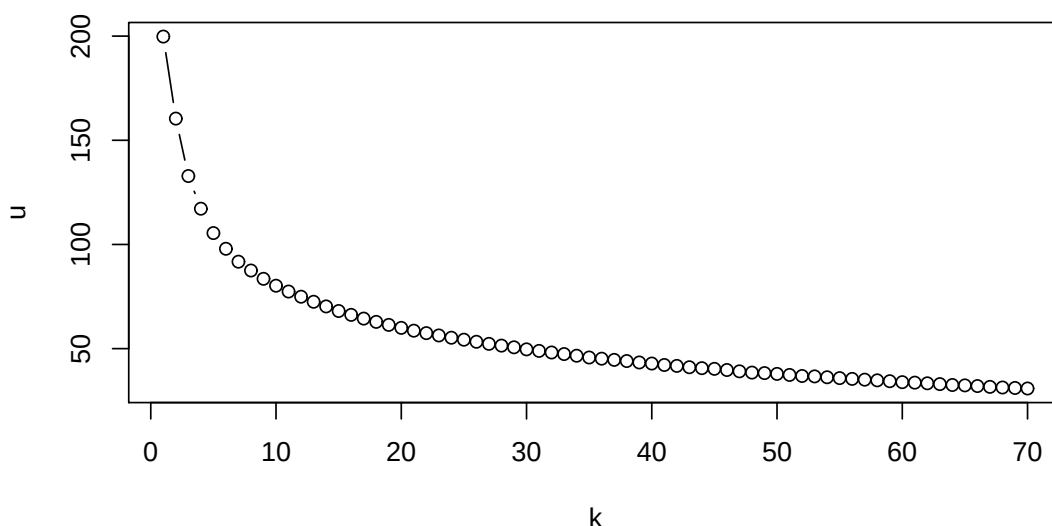
Je zaujímavé si všimnúť, že najčastejšie, až štyri krát z ôsmich, vybočujú z priemeru práve hodnoty pre osoby, ktoré žijú mimo Slovenska. Ďalej treba pripomenúť význam hodnôt -3040,07 a 5181,62. V obidvoch prípadoch sa jedná o skupiny s veľmi malými priemernými MRR (0,96 a 0,36). V prvom prípade sa navyše MRR ešte výrazne znížili o 29,18 percentuálnych bodov, čo je zníženie o 3040,07 % z ich priemernej hodnoty. V druhom prípade sa MRR zvýšili o 18,56 percentuálnych bodov, čo je zvýšenie o 5181,62 % z priemernej hodnoty MRR.

6.1.1 Výsledky zhlukovej analýzy pre okresy a vekové skupiny

Ako sme už uviedli v kapitole 5, vstupom do algoritmov zhlukovej analýzy musia byť nekorelované štandardizované dáta. Preto sme skontrolovali hodnoty korelačných koeficientov aj pre všetky okresy a vekové skupiny. Zobrali sme teda charakteristiky, ktoré nám v predošlej kapitole vyšli ako nekorelované a overili sme ich vhodnosť, vzhľadom na nekorelovanosť, aj pre okresy a vekové skupiny. V tomto prípade nám všetky kore-

lačné koeficienty vyšli menšie ako 0,5, teda budeme aplikovať metódu k -priemery na tieto charakteristiky (Priemerné MRR, Percentuálna zmena MRR, Rozdiel miezd žien v %, Rozdiel miezd mužov v %, Priemerný počet zamestnancov v skupinách, Rozdiel percentuálneho podielu počtov zamestnaných žien pred 2009 a po 2009 a podielu počtov zamestnaných mužov pred 2009 a po 2009 a Rozdiel percentuálneho podielu počtov zamestnaných žien pred 2009 a po 2009 a podielu počtov zamestnaných mužov pred 2009 a po 2009 zarábajúcich nad 3000 eur). Štandardizácia dát prebehla zhodne so štandardizáciou uvádzanou v kapitole 5, teda predelením charakteristík rozdielom ich maxima a minima.

Na základe Obr. 17 sme ako optimálne k vybrali aj v tomto prípade $k = 8$.



Obr. 21: Závislosť funkcie vzdialeností od počtu zhlukov pre okresy a vekové skupiny.

Pomocou metódy k -priemery sme rozdelili okresy Slovenska v rôznych vekových skupinách do ôsmich zhlukov s najpodobnejším správaním vybraných charakteristík. Roztriedenie do zhlukov prinieslo niekoľko zaujímavých zistení:

- Podobne ako pri zhlukovej analýze pre rôzne okresy aj tomto prípade sa skoro všetky skupiny okresov krajských miest vo všetkých vekových kategóriách sústreďovali do jedného zhluku (na Obr.22 štvrtý zhluk). Charakteristické pre tento zhluk

je znižovanie pomerne vysokých priemerných MRR, vysoký priemerný počet zamestnancov a výraznejšie sa zvyšujúci počet zamestnaných žien oproti mužom aj v priemernej skupine, aj v skupine zarábajúcich nad 3000 eur.

- Dva zhluky vznikly prevažne zo skupín s vekovou kategóriou 20 - 25 ročných. V oboch skupinách sú MRR nízke v porovnaní s ostatnými skupinami. V jednom z týchto dvoch zhlukov, ktorý zahŕňa prevažne okresy východného Slovenska však MRR navyše stále klesali (na Obr.22 prvý zhluk). Narozdiel od zhuku, ktorý zahŕňa prevažne okresy Západného Slovenska, v ktorom MRR počas krízy rástli (na Obr.22 druhý zhluk).
- Zaujímavý je pre nás aj zhluk tvorený prevažne skupinami s vekovov kategóriou 55 - 60 ročných, ktorého zaujímavosťou je najvýraznejšie sa zvyšujúci počet zamestnaných žien oproti mužom aj v priemernej skupine, aj v skupine zarábajúcich nad 3000 eur. Tento fakt môže byť spôsobený výraznejšie sa zvyšujúcim dôchodkovým vekom žien za posledné roky (na Obr.22 posledný zhluk).

Vek	Okres		P. MRR	P. z. MRR	R.m.ž. v %	R. m.m. v %	P.počet	R.podielu ž. a m.	R.podielu nad
22,64	558,27	Východ	10,23	-58,78	-25,73	-34,01	1404,83	-0,23	-12,60
23,31	492,71	Západ	12,27	38,35	-33,78	-27,83	1712,17	-0,19	-5,00
34,49	464,00	Západ	27,38	-7,21	-21,66	-25,39	2088,08	-6,86	0,31
39,01	437,02	Západ	26,93	-10,78	-18,25	-23,85	5148,68	0,03	20,97
41,25	544,43	Východ	24,34	-31,02	-11,74	-28,35	1709,05	3,04	91,01
42,06	502,22	Západ	25,26	10,85	-9,88	-3,03	2439,17	1,04	38,72
42,45	595,91	Východ	19,21	-4,68	-11,39	-11,68	1256,10	5,79	-83,70
53,77	536,16	Východ	20,10	-8,81	-9,46	-13,91	1757,98	21,52	181,80

Obr. 22: Tabuľka hodnôt centroidov, a priemerného veku a okresu v ôsmich zhlukoch.

Ostatné čísla a výsledky môžeme vidieť na Obr. 22. Červenou farbou sú zvýraznené maximálne hodnoty, modrou farbou minimálne hodnoty charakteristík. Hodnotu priemerného okresu zhuku sme vypočítali ako priemernú hodnotu čísel okresov v zhuku (každý okres má svoje číslo počnúc okresom 101 - Bratislava, končiac okresom 811 - Trebišov a tieto čísla sa zvyšujú od západu na východ Slovenska, pričom do tohto

priemerovania sme nezahrnuli skupinu 0 a 900). Priemerná hodnota čísel okresov nám vyšla 507,29, teda zhluky s vyššou priemernou hodnotou okresu považujeme za zhluky tvorené skôr okresmi východného Slovenska a naopak s nižšou hodnotou za zhluky tvorené skôr okresmi západného Slovenska. Ostatné výsledky a hodnoty k jednotlivým skupinám môže čitateľ nájsť v prílohe D na CD, z dôvodu rozsiahlosti tabuliek.

6.1.2 Zhrnutie

Výsledkom aplikačnej časti diplomovej práce je zistenie, že priemerné mzdové rodové rozdiely počas finančnej krízy klesali ako pri priemerných hodnotách pre zamestnancov na Slovensku tak aj pri väčšine skupín rozdelených podľa okresov trvalého bydliska a vekovej skupiny. Taktiež sa potvrdil predpoklad o výraznejšom znižovaní miezd mužov (na rozdiel od očakávaných) v období finančnej krízy, ako aj postupné vyrovňovanie zastúpenia žien a mužov na trhu práce. Teda môžeme tvrdiť, že zamestnaných žien sa finančná kríza nedotkla tak výrazne ako zamestnaných mužov. Ukázalo sa, že napriek prevažne nižšiemu platu žien ich pracovné miesta sú stálejšie a menej ohrozené prípadnými šokmi v ekonomike.

Novým zistením bolo pre nás odhalenie podobností v správaní sa charakteristík, týkajúcich sa mzdových rodových rozdielov. Ako bolo vyššie uvedené, okresy krajských miest boli zaradené do jedného zhluku, tak isto okresy východného Slovenska či skupiny s vekovou kategóriou 20 - 25 rokov alebo naopak skupiny s vekovou kategóriou 55 - 60 rokov.

Všetky výsledky pre jednotlivé skupiny sa nedajú v práci uviesť pre veľký rozsah výsledných tabuliek. Preto odporúčame čitateľovi bližšie sa pozrieť na odhadnuté či vypočítané charakteristiky, priemerné aj v jednotlivých skupinách, v tabuľkách MS Excelu na CD v prílohe D.

Záver

Rodová nerovnerovnosť je stále pretrvávajúcím fenoménom na Slovensku a jej dôsledky sa dotýkajú bežného života každého z nás. Naša diplomová práca preto reaguje na potrebu informovať verejnosť o stave a vývoji mzdových rozdielov medzi mužmi a ženami. Taktiež reaguje aj na výzvu Európskej komisie prikladať dôležitosť skúmaniu vplyvu zmien v ekonomike na vývoj mzdových rodových rozdielov.

Diplomová práca je rozdelená na teoretickú a praktickú časť. Obsahom teoretickej časti je zhrnutie teórie spojennej s mzdovými rodovými rozdielmi, objasnenie ich príčin a dôsledkov a zosumarizovanie pre verejnosť dostupných štatistických údajov. Pre dotvorenie obrazu o problematike mzdových rodových rozdielov sme ďalej uviedli prehľad regresných a dekompozičných metód, ktoré sa v praxi používajú na odhad vplyvu pohlavia na výšku mzdy a na vysvetlenie mzdových rozdielov na základe rozdielnych charakteristík mužov a žien. Tieto metódy sme ale ďalej nevyužívali z dôvodu výberu dát s malým množstvom vysvetľujúcich premenných, ktoré sú nevhodné pre tieto metódy. Nakoniec sme zhrnuli základnú teóriu využívanú pri analýze časových radov. Zamerali sme sa pritom na základné charakteristiky a na dekompozíciu časových radov, Holt-Wintersovu metódu a Box-Jenkinsovú metodológiu. Táto teória poskytuje aparát, ktorý sme využívali v aplikačnej časti.

Praktická časť je tvorená tromi kapitolami. V prvom rade sme sa zamerali na analyzovanie časového radu priemerných mzdových rodových rozdielov. Na základe najlepšieho modelu (spomedzi SARIMA a Holt-Wintersových modelov) a hodnôt mzdových rodových rozdielov pred rokom 2009 sme odhadli ich vývoj na obdobie 2009 - 2013. Tieto odhadnuté hodnoty sme porovnali s reálnymi hodnotami mzdových rodových rozdielov v tomto období. Priemerný rozdiel odhadnutých a reálnych hodnôt nám vyšiel -1,66 čo ukazuje na znižovanie mzdových rodových rozdielov počas finančnej krízy. Bližšie sme sa tiež pozreli na vývoj priemerných miezd mužov a žien či počtov zamestnaných osôb v jednotlivých obdobiach v rôznych platových skupinách. Táto analýza miezd a počtov nás doviedla k presvedčeniu, že zníženie mzdových rodových rozdielov nie je zapríčinené zlepšením postavenia žien na trhu práce, ale zhoršením postavenia mužov v období finančnej krízy.

V ďalších dvoch kapitolách sme sa zamerali na analýzu v rôznych skupinách za-

mestnancov. V piatej kapitole sme rozdelili zamestnancov do skupín podľa okresu ich trvalého bydliska. Na základe cyklov, ktoré sme zostavili v programe R na výber najlepšieho modelu pre časové rady MRR jednotlivých okresov, sme odhadli ich zmenu. Pokles MRR bol zaznamenaný v 48 okresoch na rozdiel od nárastu iba v 26 okresoch. Taktiež sa potvrdilo presvedčenie o výraznejšom zhoršení situácie mužov oproti ženám počas finančnej krízy. Ďalej sme využitím zhlukovej analýzy hľadali súvislosti medzi okresmi. Najzaujímavejšími výsledkami tejto analýzy bolo zaradenie všetkých okresov krajských miest do jedného zhluku, ako aj zhluk, ktorý tvorila iba jedna skupina - osoby žijúce mimo SR, ktoré pracujú na Slovensku, či zhluky tvorené prevažne okresmi východného Slovenska alebo okresmi západného Slovenska.

V poslednej kapitole sme rozdelili osoby podľa okresu ich trvalého bydliska a tiež podľa veku na 592 skupín. Podobne ako v piatej kapitole, aj tu sme aplikovali cykly na výpočet zmien MRR či zmien miezd. Podľa našich odhadov sa MRR znížili v 366 skupinách a zvýšili v 226 skupinách. Zhluková analýza nám poskytla zaujímavé výsledky. Potvrdilo sa podobné správanie v okresoch krajských miest vo väčšine vekových skupín. Taktiež vznikli dva zhluky obsahujúce najmladšie vekové skupiny a jeden zhluk tvorený prevažne najstaršou vekovou skupinou. Podobné správanie vývoja MRR a miezd sme identifikovali v skupinách prevažne východného Slovenska a tiež v skupinách prevažne západného Slovenska.

Cieľom diplomovej práce bolo analyzovať vývoj mzdových rodových rozdielov s dôrazom na vývoj počas obdobia finančnej krízy. Po ukončení modelovania a zhrnutí výsledkov môžeme povedať, že bol cieľ v rámci možností dostupnosti dát a rozsahu práce splnený. Prínos práce by mohol byť hlavne pre laickú verejnosť, zaujímajúcu sa o danú problematiku. Taktiež pre mňa bolo písanie práce obohatením a príležitosťou pracovať s väčším počtom dát a nie len teoreticky v programe R.

Možností, ako rozšíriť naše analýzy vývoja mzdových rozdielov medzi mužmi a ženami je niekoľko. Jednou z nich je vylepšenie odhadov predpovedí mzdových rodových rozdielov či už modelovaním ich časových radov pomocou iných modelov alebo využitím dát z dlhšieho obdobia pred krízou. Ďalšími možnosťami je využitie iného štatistického programu pre urýchlenie výpočtov uvedených odhadov, alebo rozšírenie analýz v rôznych platových skupinách.

ZOZNAM POUŽITEJ LITERATÚRY

- [1] Artl, J., Artlová, M.: *Ekonomické časové řady*, Professional Publishing, 1. vydanie, 2009
- [2] Barošová, M.: *Rodový mzdový rozdiel na trhu práce*, Inštitút pre výskum práce a rodiny, 2009, dostupné na internete (24.4.2014):
http://www.sspr.gov.sk/IVPR/images/IVPR/prezentacie/Barosova/Rodovy_mzdovy_rozdiel.pdf
- [3] Bassett, L. C.: *Job Evaluation Systems: Increasing Return on Payroll Dollars*, [online], dostupné na internete (24.4.2014):
<http://www.bassettgroup.com/articles/job%20evaluation%20systems.htm>
- [4] Blinder, A. S.: *Wage discrimination: Reduced form and structural estimates*, Journal of Human Resources 8 (4), 1973, 436-455
- [5] Bútorová, Z.: *Ženy, muži a vek v štatistikách trhu práce*, Inštitút pre verejné otázky, Bratislava, 2007, dostupné na internete (24.4.2014):
http://www.ivo.sk/buxus/docs/publikacie/subory/Zeny_muzi_a_vek_v_statistikach_trhu_prace.pdf
- [6] Cipra, T.: *Analýza časových řad s aplikacemi v ekonomii*, SNTL - Nakladatelství technické literatury, Praha, 1986
- [7] Council of the european union: *The Gender pay gap in the Member States of the European Union: quantitative and qualitative indicators*, 2010, dostupné na internete (24.4.2014):
<http://register.consilium.europa.eu/pd/en/10/st16/st16516-ad02.en10.pdf>
- [8] Cviková, J., Debrecéniová, J., Kobová, Ľ.: *Rodová rovnosť*, Občan a demokracia, Bratislava, 2007
- [9] Cviková, J., Juráňová, J.: *Na ceste do Európskej únie Sprievodkyňa nielen pre ženy*, Aspekt, Bratislava, 2002, ISBN 80-85549-39-5

- [10] Di Nardo, J., Fortin, N. M., Lemieux, T.: *Labor market institutions and the distribution of wages, 1973-1992: A semi-parametric approach*, *Econometrica* 64 (5), 1996, 1011-1044
- [11] Donald, S. G., Green, D. A., Paarsch, H. J.: *Differences in wage distributions between Canada and the United States: An application of a flexible estimator of distribution functions in the presence of covariates*, *Review of Economic Studies* 67 (4), 2000, 609-633
- [12] European Commission: *Opinion on The effectiveness of the current legal framework on Equal pay for equal work or work of equal value in tackling the gender pay gap*, 2009
- [13] European Commission: *Tackling the gender pay gap in the European Union*, Publications Office of the European Union, Luxembourg, 2013, dostupné na internete (24.4.2014):
http://ec.europa.eu/justice/gender-equality/files/gender_pay_gap/130422_gpg_brochure_en.pdf
- [14] Európska komisia: *Oznámenie komisie európskemu parlamentu, rade, európskemu hospodárskemu a sociálnemu výboru a výboru regiónov, Stratégia rovnosti žien a mužov 2010-2015*, Brusel, KOM (2010) 491 v konečnom znení, dostupné na internete (24.4.2014):
http://ec.europa.eu/justice/gender-equality/gender-pay-gap/eu-action/index_en.htm
- [15] Fortin, N., Lemieux, T., Firpo, S.: *Decomposition Methods in Economics*, NBER Working Paper No. 16045, 2010
- [16] Foubert, P.: *The The Gender Pay Gap in Europe from a Legal Perspective*, 2010, dostupné na internete (24.4.2014):
http://ec.europa.eu/justice/gender-equality/files/gender_pay_gap/genderpaygapfromlegalperspective-nov2010_en.pdf
- [17] Hammliton, J. D.: *Time Series Analysis*, Princeton University Press, New Jersey, 1994

- [18] Holubová, B., Filadelfiová, J.: *SÚHRNNÁ SPRÁVA O STAVE RODOVEJ ROVNOSTI NA SLOVENSKU ZA ROK 2012*, Inštitút pre výskum práce a rodiny, 2013
- [19] Inštitút zamestnanosti: *Rodové mzdové rozdiely*, 2013, dostupné na internete (24.4.2014):
<http://www.iz.sk/sk/stanoviska/rodove-mzdove-rozdiely>
- [20] Juhn, Ch., Murphy, K. M., Pierce, B.: *Wage inequality and the rise in returns to skill*, Journal of Political Economy 101 (3), 1993, 410-442
- [21] Jurajda, Š.: *Monitoring relatívneho postavenia žien na slovenskom trhu práce*, CERGE-EI, Praha, 2005
- [22] Klaudíniová, M.: *Analýza rodového mzdového rozdielu v Slovenskej republike*, Diplomová práca, Univerzita Komenského, 2013
- [23] Komisia európskych spoločenstiev: *Oznámenie komisie rade, európskemu parlamentu, európskemu hospodárskemu a sociálnemu výboru a výboru regiónov, Boj proti rozdielom v odmeňovaní žien a mužov*, Brusel, KOM (2007) 424 v konečnom znení
- [24] Kroenker, R., Bassett, G.: *Regression Quantiles*, Econometrica 46 (1), 1978, 33-50
- [25] Machado, J. A., Mata, J.: *Counterfactual decomposition of changes in wage distribution using quantile regression*, Journal of Applied Econometrics 20, 2005, 445-465
- [26] Marošiová, L., Šumšalová, S.: *Matky na trhu (Práce a života)*, Inštitút pre verejné otázky, 2006, dostupné na internete (24.4.2014):
http://www.ivo.sk/buxus/docs/publikacie/subory/Matky_na_trhu.pdf
- [27] Melly, B.: *Decomposition of differences in distribution using quantile regression*, Journal of Applied Labour Economics 12, 2005, 577-90
- [28] Nopo, H.: *Matching as a Tool to Decompose Wage Gaps*, IZA Discussion Paper No. 981, 2004

- [29] Oaxaca, R.: *Male-female wage differentials in urban labor markets*, International Economic Review 14 (3), 1973, 693-709, dostupné na internete (24.4.2014): <http://www-bcf.usc.edu/~ridder/Lnotes/Undeconometrics/Transparanten/Wagedecomp.pdf>
- [30] Pankratz, A.: *Forecasting with dynamic regression models*, John Wiley & Sons Inc, New York, 1991
- [31] Pietruchová, O.: *Rodový mzdový rozdiel*, Ministerstvo práce, sociálnych vecí a rodiny Slovenskej republiky, odbor rodovej rovnosti a rovnosti príležitostí, 2012, dostupné na internete (24.4.2014): http://rodovarovnost.files.wordpress.com/2012/03/rodovy_-mzdovy_rozdiel.pdf
- [32] Profesia: *Rozdiel medzi platmi mužov a žien je 21%*, [online], 2013, dostupné na internete (24.4.2014): <http://www.profesia.sk/cms/newsletter/marec-2013/rozdiel-medzi-platmi-muzov-a-zien-je-21/44225>
- [33] Shumway, R., H., Stoffer, D., S.: *Time Series Analysis and Its Applications*, Third edition, Springer Science+Business Media, New York, 2011
- [34] Special Eurobarometer 296: *Discrimination in the European Union: Perceptions, Experiences and Attitudes*, 2008, dostupné na internete (24.4.2014): http://ec.europa.eu/public_opinion/archives/ebs/ebs_296_en.pdf
- [35] Štatistický úrad Slovenskej republiky: *Rodová rovnosť 2013 - Gender equality 2013*, 2013, dostupné na internete (24.4.2014): <http://portal.statistics.sk/files/KrajskeSpravy/KE/gender2013/cely/gender-2013.pdf>
- [36] Štulajter, F.: *Random processes and time series*, Comenius University Bratislava, Bratislava, 2001
- [37] Trend: *Ženy nezarábajú menej ako muži*, [online], 2013, dostupné na internete (24.4.2014):

<http://ekonomika.etrend.sk/ekonomika-slovensko/zeny-nezarabaju-menej-ako-muzi.html>

[38] Ústavný zákon 460/1992 Zb. Ústava Slovenskej republiky

[39] Zákon č. 311/2001 Z. z. Zákonník práce

[40] Zákon č. 365/2004 Z.z. o rovnakom zaobchádzaní v niektorých oblastiach a o ochrane pred diskrimináciou a o zmene a doplnení niektorých zákonov (antidiskriminačný zákon)

[41] rodová rovnosť v odmeňovaní, [online], dostupné na internete (16.1.2014): <http://rodovarovnost.wordpress.com/>

[42] Štatistický úrad Európskych spoločností, [online], dostupné na internete (16.1.2014): <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>

[43] Štatistického úradu Slovenskej republiky, [online], dostupné na internete (16.1.2014): <http://slovak.statistics.sk/>

Príloha A

Zákon 365/2004 Z.z.(o rovnakom zaobchádzaní v niektorých oblastiach a o ochrane pred diskrimináciou a o zmene a doplnení niektorých zákonov (antidiskriminačný zákon)), § 6 Zásada rovnakého zaobchádzania v pracovnoprávnych vzťahoch a obdobných právnych vzťahoch

(1) V súlade so zásadou rovnakého zaobchádzania sa v pracovnoprávnych vzťahoch, obdobných právnych vzťahoch a v právnych vzťahoch s nimi súvisiacich zakazuje diskriminácia osôb z dôvodu ich pohlavia, náboženského vyznania alebo viery, rasového pôvodu, národnostného alebo etnického pôvodu, zdravotného postihnutia, veku alebo sexuálnej orientácie.

(2) Zásada rovnakého zaobchádzania podľa odseku 1 sa uplatňuje len v spojení s právami fyzických osôb ustanovenými osobitnými zákonmi v oblastiach

a) prístupu k zamestnaniu, povolaniu, inej zárobkovej činnosti alebo funkcii (ďalej len "zamestnanie") vrátane požiadaviek pri prijímaní do zamestnania a podmienok a spôsobu uskutočňovania výberu do zamestnania,

b) výkonu zamestnania a podmienok výkonu práce v zamestnaní vrátane odmeňovania, funkčného postupu v zamestnaní a prepúšťania,

c) prístupu k odbornému vzdelávaniu, ďalšiemu odbornému vzdelávaniu a účasti na programoch aktívnych opatrení na trhu práce vrátane prístupu k poradenstvu pre výber zamestnania a zmenu zamestnania⁹) (ďalej len "odborné vzdelávanie") alebo

d) členstva a pôsobenia v organizácii zamestnancov, organizácii zamestnávateľov a v organizáciách združujúcich osoby určitých profesií vrátane poskytovania výhod, ktoré tieto organizácie svojim členom poskytujú.

(3) Za diskrimináciu z dôvodu

a) pohlavia sa považuje aj diskriminácia z dôvodu tehotenstva alebo materstva, ako aj diskriminácia z dôvodu pohlavnej alebo rodovej identifikácie,

b) rasového pôvodu, národnostného alebo etnického pôvodu sa považuje aj diskriminácia z dôvodu vzťahu k osobe určitého rasového pôvodu, národnostného alebo etnického pôvodu,

c) náboženského vyznania alebo viery sa považuje aj diskriminácia z dôvodu vzťahu k osobe určitého náboženského vyznania alebo viery a aj diskriminácia fyzickej osoby

bez náboženského vyznania,

d) zdravotného postihnutia sa považuje aj diskriminácia z dôvodu predchádzajúceho zdravotného postihnutia alebo diskriminácia osoby, u ktorej by na základe vonkajších príznakov bolo možné predpokladať, že je zdravotne postihnutá.

Zákon 311/2001 Z.z.(zákoník práce), §13 Zákaz diskriminácie

(1) Zamestnancom patria práva vyplývajúce z pracovnoprávnych vzťahov bez akýchkoľvek obmedzení a priamej diskriminácie alebo nepriamej diskriminácie podľa pohlavia, manželského stavu a rodinného stavu, rasy, farby pleti, jazyka, veku, zdravotného stavu, viery a náboženstva, politického alebo iného zmýšľania, odborovej činnosti, národného alebo sociálneho pôvodu, príslušnosti k národnosti alebo etnickej skupine, majetku, rodu alebo iného postavenia okrem prípadu, ak to ustanovuje zákon alebo ak je na výkon prác vecný dôvod, ktorý spočíva v predpokladoch alebo požiadavkách a v povahe práce, ktorú má zamestnanec vykonávať.

(2) Na účely princípu rovnakého zaobchádzania podľa odseku 1 je nepriama diskriminácia navonok neutrálny pokyn, rozhodnutie alebo prax znevýhodňujúce podstatne väčšiu skupinu fyzických osôb, ak tento pokyn, rozhodnutie alebo prax nie sú vhodné a nevyhnutné a nemôžu sa ospravedlniť objektívnymi skutočnosťami.

(3) Výkon práv a povinností vyplývajúcich z pracovnoprávnych vzťahov musí byť v súlade s dobrými mravmi. Nikto nesmie tieto práva a povinnosti zneužívať na škodu druhého účastníka pracovnoprávneho vzťahu alebo spoluzamestnancov. Nesmie sa znížovať dôstojnosť zamestnancov na pracovisku.

(4) Zamestnanec má právo podať zamestnávateľovi sťažnosť v súvislosti s porušením práv a povinností ustanovených v odsekoch 1 až 3; zamestnávateľ je povinný na sťažnosť zamestnanca bez zbytočného odkladu odpovedať, vykonať nápravu, upustiť od takého konania a odstrániť jeho následky.

(5) Zamestnanec, ktorý sa cíti poškodený v dôsledku nedodržania podmienok ustanovených v odsekoch 1 až 3, môže sa domáhať svojich práv na súde vrátane primeranej náhrady nemajetkovej ujmy v peniazoch. Zamestnávateľ je povinný preukázať, že nedošlo k porušeniu princípu rovnakého zaobchádzania.

(6) Zamestnávateľ nesmie zamestnanca postihovať alebo znevýhodňovať preto, že zamestnanec uplatňuje svoje práva vyplývajúce z pracovnoprávnych vzťahov.

Príloha B

Príkazy pre výpočet zmien MRR - využívané v podkapitole 4.3.1.

```
library(tseries), library(tseries), library(tseries), library(tseries), library(tseries) #na-
čítanie balíčkov potrebných pri analýze časových radov
MRR<-read.csv("mrr_priemery.csv", header = TRUE, sep = ";", dec=".") #načíta-
nie dát z predpripraveného súboru .csv
priemerne_MRR<-MRR[1:48,1] # načítanie radu priemerných MRR v období pred
krízou
diff_priemerne_MRR<-diff(MRR[1:48,1],lag=12) #sezónne diferencovanie radu
summary(ur.df(diff_priemerne_MRR,type="none",lags=10,selectlags="BIC")) #ove-
renie neprítomnosti jednotkového koreňa
acf2(diff_priemerne_MRR, 20) # vypísanie prvých 20 ACF a PACF
aa<-auto.arima(priemerne_MRR, test=c("kpss"), seasonal.test=c("öcsb")) #odhadnu-
tie koeficientov najlepšieho modelu pre zadané vstupy
sarima(priemerne_MRR, 0,0,0,0,1,0,12) #výstup aj s ACF rezíduí, Q-Q grafu a P-
hodnôt Ljung-Box štatistiky
box_pierce<-Box.test(aa$residuals) # test nezávislosti rezíduí
shapiro.test(aa$residuals) #test o normalite rezíduí
aa<-auto.arima(priemerne_MRR, test=c("kpss"), seasonal.test=c("ch"))
sarima(priemerne_MRR, 0,0,0,1,0,0,12)
box_pierce<-Box.test(aa$residuals)
shapiro.test(aa$residuals)
aa<-auto.arima(priemerne_MRR, test=c("ädf"), seasonal.test=c("öcsb"))
sarima(priemerne_MRR, 0,2,2,0,1,0,12)
box_pierce<-Box.test(aa$residuals)
shapiro.test(aa$residuals)

#porovnanie predpovedí ex-post pomocou RMSE jednotlivých modelov
skuska_priemerne_MRR<-MRR[1:42,1] #skúšobný rad
a<-sarima.for(skuska_priemerne_MRR,6,0,0,0,1,0,0,12)
```

```

predpoved<-c(skuska_priemerne_MRR, a$pred)
predpoved= ts(predpoved, frequency=12,start=c(2005),end=c(2008,12))
plot(predpoved, col="red",ylim=c(22,31))
priemerne_MRR= ts(priemerne_MRR, frequency=12,start=c(2005),end=c(2008,12))
lines(priemerne_MRR)
RMSE_A<-sqrt(sum((priemerne_MRR[43:48]-predpoved[43:48])^2)/6) #RMSE_A chyba
predpovede vytvorenej pomocou ARIMA modelu
HW<- HoltWinters(priemerne_MRR, seasonal="multiplicative")
RMSE_HW<-sqrt(sum((priemerne_MRR[43:48]- HW$fitted[31:36,1])^2)/6) #RMSE
chyba predpovede vytvorenej pomocou Holt - Wintersovho multiplikatívneho modelu
HW<- HoltWinters(priemerne_MRR, seasonal="additive")
RMSE_HW<-sqrt(sum((priemerne_MRR[43:48]- HW$fitted[31:36,1])^2)/6) #RMSE
chyba predpovede vytvorenej pomocou Holt - Wintersovho aditívneho modelu

```

#porovnanie predpokladaného vývoja MRR na základe určeného modelu a skutočných hodnôt v percentuálnych bodoch

```

predpokladany_vyvoj<-sarima.for(priemerne_MRR,60,0,0,0,1,0,0,12) #predpokladaný
vývoj počas obdobia krízy
lines(MRR_ts)
MRR_ts <- ts(MRR[,1], frequency=12,start=c(2005))
predpokladany_vyvoj = as.numeric(predpokladany_vyvoj$pred)
rozdiel<-(MRR[49:108,1]- predpokladany_vyvoj)
priemerny_rozdiel<-mean(rozdiel) #priemerný rozdiel predpokladaného vývoja od sku-
točných hodnôt

```

Cyklus na výpočet zmien MRR - využívané v kapitole 5.

```

MRR<-read.csv("MRR_okr.csv", header = FALSE, sep = ";", dec=".") #načíta-
nie dát z predpripraveného súboru .csv
# definovanie pomocných matíc v požadovanej veľkosti
mrr<-MRR[1:48,] #z radov v tejto matici budeme odhadovať modely
mrr2<-MRR[1:42,] #z radov v tejto matici budeme overovať modely

```

PREDIKCIEVYSLEDOK<-MRR[49:108,] #v tejto matici budú uložené výsledné predikcie pre obdobie krízy

ROZDIEL<-MRR[49:108,] #neskôr v nej bude uložený rozdiel skutočných hodnôt počas krízy a odhadnutých (očakávaných) hodnôt

PRIEMER<-1:length(MRR) #vektor veľkosti počtu stĺpcov matice MRR

#nasledovný cyklus:

vyberie model z ARIMA modelov a Holt-Wintersových modelov

#na základe neho odhadne očakávané hodnoty a porovná ich s reálnymi hodnotami

vypočíta maticu ROZDIEL (teda rozdiel medzi očakávanými hodnotami mzdových rodových rozdielov a reálnymi hodnotami v období krízy)

vypočíta priemerný rozdiel za obdobie krízy a uloží ho do vektora PRIEMER

```

    for (i in 1 : ncol(MRR))
    {
mrr[,i]=MRR[1:48,i]
mrr[,i]<-ts(mrr[,i], frequency=12,start=c(2005))
AA<-auto.arima(mrr[,i])
box_pierce<-Box.test(AA$residuals)
mrr2= ts(mrr[,i], frequency=12,start=c(2005),end=c(2008,6))
SKUSKA<-sarima.for(mrr2,6,AA$arima[1], AA$arima[6], AA$arima[2], AA$arima[3],
AA$arima[7], AA$arima[4], AA$arima[5])
SKUSKAD<- MRR[43:48,i]
RMSE_A<-sqrt(sum((SKUSKAD-SKUSKA$pred)^2)/6)
HW_A<- HoltWinters(mrr[,i],seasonal="additive")
RMSE_HW_A<-sqrt(sum((SKUSKAD- HW_A$fitted[31:36,1])^2)/6)
HW_M<- HoltWinters(mrr[,i],seasonal="multiplicative")
RMSE_HW_M<-sqrt(sum((SKUSKAD- HW_M$fitted[31:36,1])^2)/6)
if((RMSE_A<RMSE_HW_A && RMSE_A<RMSE_HW_M) &&
box_pierce$p.value>0.05)
{

```

```

PRED<-forecast(AA, h=60)
PREDIKCIEVYSLEDOK [,i]<-c(PRED$mean)
ROZDIEL [,i]<-(MRR[49:108,i]- PREDIKCIEVYSLEDOK [,i])
PRIEMER[i]<-mean(ROZDIEL [,i])
}
else if (RMSE_HW_A<RMSE_HW_M )
{
PRED<- forecast.HoltWinters(HW_A, h=60)
PREDIKCIEVYSLEDOK [,i]<-c(PRED$mean)
ROZDIEL [,i]<-(MRR[49:108,i]- PREDIKCIEVYSLEDOK [,i])
PRIEMER[i]<-mean(ROZDIEL [,i])
}
else
{
PRED<- forecast.HoltWinters(HW_M, h=60)
PREDIKCIEVYSLEDOK [,i]<-c(PRED$mean)
ROZDIEL [,i]<-(MRR[49:108,i]- PREDIKCIEVYSLEDOK [,i])
PRIEMER[i]<-mean(ROZDIEL [,i])
}
MRR_ts <- ts(MRR[,i], frequency=12,start=c(2005))
plot(MRR_ts)
PREDIKCIEVYSLEDOK [,i]<- ts(PREDIKCIEVYSLEDOK [,i],frequency=12,
start=c(2009))
lines(PREDIKCIEVYSLEDOK [,i],col= "red")
}

```

Cyklus na výpočet zmien v raste miezd - využívané v podkapitole 4.3.2 a kapitole 5.

```

tab_mzdy<-read.csv("mzdy_zeny_nad.csv", header = FALSE, sep = ";", dec=".")
#načítanie dát z ôsmich predpripravených súborov .csv
rozdiel<-1:(ncol(tab_mzdy)) #do tohto vektora sa neskôr načítajú hodnoty rozdielu

```

koncových bodov lineárnej trendovej priamky z prvých 48 hodnôt a lineárnej trendovej priamky z posledných 60 hodnôt

```
percent_rozdiel<-1:(ncol(tab_mzdy)) #percentuálny rozdiel vzhľadom k priemernej
mzde jednotlivkej skupiny pred krízou
for (j in 1:(ncol(tab_mzdy)))
{
#prevedenie vektorov miezd na objekt časových radov
mzdy= ts(tab_mzdy[1:108,j], frequency=12,start=c(2005),end=c(2013,12))
mzdy1= ts(tab_mzdy[1:48,j], frequency=12,start=c(2005),end=c(2008,12))
x <- 1:48
fit <- glm(mzdy[1:48]~ x) #odhad koeficientov lineárnej trendovej priamky na základe
prvých 48 hodnôt
x2 <- 1:60 fit2 <- glm(mzdy[49:108]~ x2) #odhad koeficientov lineárnej trendovej
priamky na základe posledných 60 hodnôt
#definovanie nových vektorov
odhad<-mzdy
odhad2<-mzdy
for (i in 1:108)
odhad[i]<-fit$coefficients[1]+fit$coefficients[2]*i #odhady jednotlivých bodov priamky
na základe odhadnutých koeficientov prvej trendovej priamky
for (i in 1:108)
odhad2[i]<-fit2$coefficients[1]+fit2$coefficients[2]*(i-48) #odhady jednotlivých bodov
priamky na základe odhadnutých koeficientov druhej trendovej priamky
rozdiel[j]<-odhad2[108]-odhad[108] #načítanie rozdielu v eurách
percent_rozdiel[j]<-rozdiel[j]/(mean(mzdy[1:108])/100) #načítanie percentuálneho roz-
dielu vzhľadom k priemernej mzde skupiny do roku 2009
#vykreslenie grafov pre vizuálnu kontrolu
plot(mzdy,ylim=c(100,10000))
lines(mzdy)
lines(mzdy1,col="red")
lines(odhad, col="red")
```



```
lines(odhad2)
}
```

Cyklus na odstránenie outlierov z jednotlivých vektorov - využívané v kapitolách 5 a 6

```
#nájdenie outlierov a nahradenie ich pôvodnej hodnoty hraničnou hodnotou
for (j in 1:(ncol(tab_mzdy)))
{
IQR<-quantile(tab_mzdy[,j], probs=3/4)-quantile(tab_mzdy[,j], probs=1/4)
dolna<-quantile(tab_mzdy[,j], probs=1/4)-1.5*IQR
horna<-quantile(tab_mzdy[,j], probs=3/4)+1.5*IQR
for (i in 1:108)
{
if (tab_mzdy[i,j]<dolna)
tab_mzdy[i,j]<-dolna
if (tab_mzdy[i,j]>horna)
tab_mzdy[i,j]<-horna
}
}
}
```

Zhluková analýza - využívané v kapitolách 5 a 6.

```
okresy<-read.csv("tabulka_okr_vek.csv", header = TRUE, sep = ";", dec=",")
library(HSAUR),library(stats) #načítanie knižníc potrebných pri zhlukovej analýze v
programe R
#vykreslenie grafov každej dvojice charakteristík a vypočítanie korelácie medzi nimi
okresy1<-vysledok[,1:10]
plot(okresy1)
round(cor(vysledok1),2)

#nehierarchická analýza k-priemery:
```

```
# 1.štandardizácia dát
standard<-apply(okresy1,2,max)-apply(okresy1,2,min)
standard.dat<-sweep(vysledok1,2,standard,FUN="/")
standard.dat

#2.určenie počtu zhlukov na základe grafického zobrazenia a laktového kritéria
k=1:1:30
u=1:1:30
for(i in k)
{
standard.kmeans= kmeans(standard.dat, centers=i, iter.max=1000000000, nstart=1000000,
"Lloyd")
u[i]=standard.kmeans$tot.withinss
}
plot(k,u,type="b")

#3.Lloydov algoritmus, vypočítanie centroidov, priradenie jednotlivých okresov do
zhlukov
standard.kmeans<-kmeans(standard.dat,centers=8, iter.max=1000000000,nstart=1000000,
"Lloyd")
x=standard.kmeans$centers
table(standard.kmeans$cluster)
standard.kmeans$cluster

#návrät k pôvodnej mierke
z=rbind(standard,standard,standard,standard,standard,standard,standard,standard)
standard.kmeans$center*z
```

Príloha C

Príloha C - Tabuľka výsledkov zhlukovej analýzy pre okresy

Okres	P.MRR	P.z.MRR	R.m.ž. v %	R.m.m. v %	P. počet	R.podielu ž. a m.	R.podielu nad
0	28,59	8,34	-9,60	-28,87	3507,43	-11,75	-3,26
900	29,78	14,01	12,40	23,09	629,98	-84,43	139,03
101 Bratislava	29,07	-6,20	-17,86	-20,36	19041,04	-1,25	6,26
106 Malacky	27,66	14,10	-19,02	-14,44	2772,54	5,69	-1,69
107 Pezinok	25,76	5,61	-19,43	-13,21	2436,82	3,17	-0,87
108 Senec	28,59	24,34	-8,26	4,07	2514,04	2,38	-11,31
201 Dunajská Streda	23,53	0,13	-13,32	-9,96	4232,35	6,42	20,48
202 Galanta	23,00	-6,72	-24,57	-28,41	3687,77	4,96	-13,71
203 Hlohovec	22,19	-4,40	-21,01	-24,17	1811,10	4,51	20,18
204 Piešťany	30,72	-7,76	-21,32	-27,68	2469,59	1,94	33,64
205 Senica	22,02	-1,43	-24,25	-20,69	2209,56	0,24	22,60
206 Skalica	29,33	0,05	-17,53	-15,63	1854,33	-1,58	1,16
207 Trnava	29,99	3,87	-25,61	-23,19	5421,67	4,25	24,33
301 Bánovce n. Bebravou	21,49	-26,72	-20,12	-31,68	1386,84	2,63	16,88
302 Ilava	27,26	0,73	-24,66	-30,68	2436,03	2,35	18,42
303 Myjava	24,01	9,65	-26,29	-22,69	1171,31	4,55	51,67
304 Nové Mesto n. Váhom	21,85	-11,16	-22,61	-27,25	2571,04	2,08	26,38
305 Partizánske	24,83	1,11	-14,24	-16,92	1617,41	-0,75	-13,20
306 Považská Bystrica	30,56	-5,92	-23,57	-31,20	2294,91	5,20	7,00
307 Prievidza	27,66	0,00	-12,31	-10,40	6667,45	2,72	32,64
308 Púchov	35,88	8,90	-21,10	-12,89	1837,31	6,74	45,61
309 Trenčín	24,64	-2,57	-23,60	-27,28	4402,69	2,06	23,07
401 Komárno	17,42	-13,93	-14,50	-18,65	3495,55	3,84	-2,31
402 Levice	27,31	0,03	-13,18	-15,91	3721,16	0,93	-2,73
403 Nitra	24,73	-10,87	-19,84	-26,96	6216,26	3,13	23,82
404 Nové Zámky	18,48	-20,86	-17,36	-21,06	4825,54	3,17	10,87
405 Šaľa	27,36	-4,99	-23,09	-25,14	2038,72	2,42	8,35
406 Topoľčany	17,61	-3,00	-20,31	-20,27	2760,74	2,26	41,43
407 Zlaté Moravce	24,82	-20,80	-21,52	-24,82	1474,32	0,87	14,27
501 Bytča	26,56	-27,90	-17,26	-29,16	997,95	0,77	21,43
502 Čadca	26,41	-21,41	-18,23	-23,01	2645,84	3,94	35,26
503 Dolný Kubín	22,55	-17,54	-17,76	-20,85	1405,07	4,57	42,47
504 Kysucké Nové Mesto	29,88	-9,70	-21,08	-23,71	1182,15	0,64	32,53
505 Liptovský Mikuláš	25,02	-16,24	-20,50	-27,03	2681,22	5,40	19,39
506 Martin	23,95	-1,76	-17,96	-18,88	3679,87	5,51	28,97
507 Námestovo	20,15	12,55	-13,51	-25,51	1519,37	9,40	9,58
508 Ružomberok	27,80	2,43	-17,46	-15,85	1964,05	4,50	19,45
509 Turčianske Teplice	17,35	16,66	-22,11	-13,46	598,37	7,26	-26,39
510 Tvrdošín	22,81	-17,86	-15,25	-27,77	1098,19	3,23	40,31
511 Žilina	30,57	-7,93	-25,15	-29,74	5862,43	3,38	34,25
601 Banská Bystrica	22,84	-30,30	-21,64	-27,99	4536,32	3,79	31,03
602 Banská Štiavnica	21,54	-4,39	-20,55	-20,69	500,35	9,83	65,30
603 Brezno	23,26	-15,13	-20,25	-23,39	1994,92	5,43	1,63
604 Detva	22,91	-5,94	-21,64	-29,68	1051,85	3,48	44,06
605 Krupina	15,24	-9,29	-18,96	-23,99	737,35	9,60	-17,90
606 Lučenec	22,57	26,92	-13,31	-9,84	2124,99	3,52	9,75
607 Poltár	15,80	-26,75	-12,59	-16,83	604,28	4,39	58,59
608 Revúca	19,43	24,93	-18,86	-4,63	1002,20	4,83	26,88
609 Rimavská Sobota	12,84	-30,11	-11,58	-15,49	2070,25	5,76	11,25
610 Veľký Krtíš	17,25	-14,00	-14,32	-15,55	1200,01	-2,61	30,72
611 Zvolen	21,38	-21,95	-18,59	-26,80	2596,64	2,48	17,12
612 Žarnovica	25,03	2,59	-23,79	-21,57	814,10	3,26	-17,67
613 Žiar nad Hronom	25,48	-4,46	-19,81	-16,39	1663,24	4,82	18,86
701 Bardejov	11,57	12,86	-12,31	-19,54	2096,87	9,51	21,52
702 Humenné	19,98	-29,32	-12,83	-25,67	1927,09	1,45	6,42
703 Kežmarok	18,97	-14,18	-18,45	-26,33	1693,11	5,72	39,37
704 Levoča	18,64	-25,21	-20,21	-27,88	915,38	3,11	24,01
705 Medzilaborce	17,50	-48,00	-4,94	-22,17	294,70	2,14	171,85
706 Poprad	24,72	-17,99	-18,27	-25,76	3545,90	3,35	19,47
707 Prešov	19,69	-8,85	-19,11	-22,26	5218,38	3,56	36,61
708 Sabinov	19,94	19,29	-14,64	-18,75	1320,64	9,09	18,26
709 Snina	19,92	-33,11	-15,21	-24,53	1055,93	4,48	34,40
710 Stará Ľubovňa	18,57	-49,70	-18,77	-31,95	1190,87	7,40	49,50
711 Stropkov	13,43	13,35	-16,35	-12,35	559,85	8,03	91,02
712 Svidník	12,46	1,14	-13,92	-11,89	893,00	5,26	37,57
713 Vranov nad Topľou	15,68	-32,82	-12,46	-19,93	1966,35	4,40	42,09
801 Gelnica	20,18	-16,17	-14,00	-22,54	810,46	7,26	25,17
802 Košice	30,57	-1,55	-16,64	-17,23	7268,31	2,75	21,60
806 Košice-okolie	29,18	1,51	-18,11	-16,06	3464,38	8,63	31,72
807 Michalovce	21,21	-16,24	-9,55	-12,39	3025,67	2,09	-5,24
808 Rožňava	19,60	-7,12	-17,91	-14,96	1685,47	6,62	31,34
809 Sobrance	18,47	4,16	-9,01	-9,82	581,03	6,45	-8,22
810 Spišská Nová Ves	22,18	-20,49	-17,74	-26,88	2767,25	5,66	21,22
811 Trebišov	16,73	20,45	-8,23	-7,48	2766,61	4,53	26,54
Priemer	22,89	-6,77	-17,28	-20,10	2582,21	2,67	24,49
Váž priemer	24,5179	-6,2251	-17,90824	-20,667168		2,765021586	19,32122952
Min	11,57	-49,70	-26,29	-31,95	294,70	-84,43	-26,39
Max	35,88	26,92	12,40	23,09	19041,04	9,83	171,85

Obr. A.23: Tabuľka vybraných charakteristík pre okresy.