

EKONOMICKÁ UNIVERZITA V BRATISLAVE

FAKULTA HOSPODÁRSKEJ INFORMATIKY

Tendencie vývoja zamestnanosti na trhu práce v Slovenskej republike

Diplomová práca

2020

FRANTIŠEK KUCHÁR

EKONOMICKÁ UNIVERZITA V BRATISLAVE

FAKULTA HOSPODÁRSKEJ INFORMATIKY

Evidenčné číslo: 1003005/I/2020/36097107977319428

Tendencie vývoja zamestnanosti na trhu práce v Slovenskej republike

Diplomová práca

Študijný program: Štatistické metódy v ekonómii

Študijný odbor: 8. - ekonómia a manažment

Školiace pracovisko: Katedra štatistiky FHI

Vedúci záverečnej práce: Ing. Katarína Moravčíková, PhD.

Bratislava 2020

FRANTIŠEK KUCHÁR



Ekonomická univerzita v Bratislave
Fakulta hospodárskej informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. František Kuchár
Študijný program: štatistické metódy v ekonómii (Jednoodborové štúdium, inžiniersky II. st., denná forma)
Študijný odbor: 8. - ekonómia a manažment
Typ záverečnej práce: Inžinierska záverečná práca
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Tendencie vývoja zamestnanosti na trhu práce v Slovenskej republike

Anotácia: Práca sa zaoberá analýzou vývoja zamestnanosti na trhu práce v Slovenskej republike. Na analýzu sú využité dáta zo stránky Štatistického úradu SR.

Vedúci: Ing. Katarína Moravčíková, PhD.
Katedra: KŠ FHI - Katedra štatistiky FHI
Vedúci katedry: doc. Ing. Mária Vojtková, PhD.

Dátum zadania: 09.10.2018

Dátum schválenia: 24.10.2018

doc. Ing. Mária Vojtková, PhD.
vedúci katedry

Čestné vyhlásenie

Čestne vyhlasujem, že som diplomovú prácu vypracoval samostatne, a že som uviedol všetku použitú literatúru.

Dátum:

.....

podpis študenta

POĎAKOVANIE

Touto cestou sa chcem poďakovať svojej školiteľke Ing. Kataríne Moravčíkovej, PhD. za odbornú pomoc a usmernenie pri písaní mojej práce.

ABSTRAKT

KUCHÁR, František: Tendencie vývoja zamestnanosti na trhu práce v Slovenskej republike. – Ekonomická univerzita v Bratislave. Fakulta hospodárskej informatiky, Katedra štatistiky FHI – Vedúci záverečnej práce: Ing. Katarína Moravčíková, PhD. – Bratislava: FHI EU, 2020, 77 s.

Cieľom diplomovej práce je analyzovanie vývoja zamestnanosti na trhu práce v Slovenskej republike za obdobie 1994-2018. Práca je rozdelená do štyroch kapitol. Obsahuje 31 grafov, 6 tabuliek a 12 obrázkov. V prvej kapitole sa venujeme analýze súčasného stavu zamestnanosti a vymedzeniu pojmov trh práce, zamestnanec, ekonomicky aktívne obyvateľstvo, uchádzač o zamestnanie, dodatočná potreba pracovných síl. V druhej kapitole je stanovený hlavný cieľ a pomocné ciele. V tretej kapitole sú popísané metódy potrebné na splnenie cieľov. Štvrtá kapitola pozostáva z grafickej analýzy časového radu. Potom nasleduje tvorba modelu pomocou Box-Jenkinsovej metodológie. Na záver tejto časti sa venujeme výberu najvhodnejšieho modelu a testovaniu jeho kvality.

Kľúčové slová: zamestnanosť, trh práce, Box-Jenkins, ARIMA, stacionarita

ABSTRACT

KUCHÁR, František: Employment trends of labour market in Slovakia. – The University of Economics in Bratislava. Faculty of Economic Informatics, Department of Statistics – The supervisor: Ing. Katarína Moravčíková, PhD. – Bratislava: FEI UE, 2020, 77 p.

The aim of the thesis is to analyze the development of employment on the labour market in the Slovak Republic during 1993-2018. The thesis is divided into four chapters. It contains 31 graphs, 6 tables and 12 pictures. In the first chapter, we analyze the current state of employment and define the terms labour market, employee, economically active population, job seeker, additional need for labor. The second chapter sets the main objective and auxiliary objectives. The third chapter describes the methods used to meet the objectives. The fourth chapter consists of graphical analysis of time series. This is followed by modeling using the Box-Jenkins methodology. The end of this section is focused on selecting the most appropriate model and tests the robustness.

Keywords: employment, labour market, Box-Jenkins, ARIMA, stationarity

Obsah

Úvod.....	11
1. Súčasný stav riešenej problematiky doma a v zahraničí.....	12
1.1 Trh práce	12
1.2 Metodika podľa ÚPSVaR	13
1.3 Metodika podľa ŠÚ SR.....	14
1.4 Analýza súčasného stavu zamestnanosti.....	16
1.4.1 Dodatočná potreba pracovných síl.....	17
1.4.2 Správanie podnikov na trhu práce.....	23
1.4.2 Prílev absolventov na trh práce.....	24
2. Cieľ práce.....	27
3. Metodika práce	28
3.1 Stacionarita časového radu.....	28
3.2 Autokorelačná a parciálna autokorelačná funkcia	29
3.3 Testovanie rezíduí	30
3.4 Box–Jenkinsové modely	36
3.4.1 Autoregresný proces AR.....	37
3.4.2 Proces kĺzavých priemerov MA.....	39
3.4.3 Integrované procesy ARIMA.....	40
3.4.4 Sezónne procesy SARIMA	44
3.5 Fázy tvorby modelu.....	46
4. Výsledky práce	50
4.1 Deskriptívna analýza časového radu.....	50
4.2 Identifikácia modelu.....	55
4.3 Tvorba modelov	57
4.3.1 ARIMA(0,1,0)x(1,1,1) _{4c}	57
4.3.2 ARIMA(0,1,0)x(2,1,1) _{4c}	58
4.3.3 ARIMA(2,1,3)x(0,1,0) ₄	60
4.3.4 ARIMA(2,1,3)x(1,1,1) ₄	62
4.3.5 ARIMA(2,1,3)x(2,1,1) _{4c}	63
4.4 Výber modelu.....	65
4.5 Splnenie predpokladov o náhodnej zložke modelu.....	67
4.6 Zápis modelu.....	70
4.7 Simulácia ex-post	71
4.8 Prognóza ex-ante.....	72
Záver	74
Zoznam použitej literatúry	76

Zoznam tabuliek, obrázkov a grafov

Tabuľka č. 1 – Správanie ACF a PACF.....	40
Tabuľka č. 2 – Rozšírený DF test	55
Tabuľka č. 3 – AIC/BIC kritérium	66
Tabuľka č. 4 – Prognóza ex-post od Q1/2017 do Q4/2018	71
Tabuľka č. 5 – Prognóza ex-ante na rok 2019	72
Tabuľka č. 6 – Porovnanie skutočných hodnôt a prognózy.....	73
Graf č. 1 – Počet zamestnancov v tis. osôb vo vybraných odvetviach 2015 - 2019.....	18
Graf č. 2 – Vývoj vekovej štruktúry v hospodárstve SR	19
Graf č. 3 – Dodatočná potreba pracovných síl do roku 2024	20
Graf č. 4 – Vývoj dodatočnej potreby pracovných síl 2019-2024.....	20
Graf č. 5 – Dodatočná potreba v odvetviach národného hospodárstva v r. 2019-2024.....	22
Graf č. 6 – Podiel vysokoškolákov na trhu práce 2002 oproti 2017.....	25
Graf č. 7 – Počet pracujúcich v súkromnom sektore Q1/1994 – Q4/2018	50
Graf č. 8 – 1. nesezónna diferenciácia	51
Graf č. 9 – ACF časového radu pracujúcich v súkromnom sektore	52
Graf č. 10 – PACF časového radu pracujúcich v súkromnom sektore	52
Graf č. 11 – ACF logaritmovaného časového radu 1. diferenciácií	53
Graf č. 12 – 1. nesezónna diferenciácia logaritmovaného ČR.....	53
Graf č. 13 – 1. nesezónna diferenciácia a 1. sezónna diferenciácia logaritmovaného ČR ..	54
Graf č. 14 – 2. nesezónna diferenciácia a 1. sezónna diferenciácia logaritmovaného ČR ..	54
Graf č. 15 – ACF po diferenciácií logaritmovaného ČR	55
Graf č. 16 – PACF po diferenciácií logaritmovaného ČR.....	56
Graf č. 17 – ACF modelu $ARIMA(0,1,0) \times (1,1,1)_{4c}$	58
Graf č. 18 – PACF modelu $ARIMA(0,1,0) \times (1,1,1)_{4c}$	58
Graf č. 19 – ACF modelu $ARIMA(0,1,0) \times (2,1,1)_{4c}$	59
Graf č. 20 – PACF modelu $ARIMA(0,1,0) \times (2,1,1)_{4c}$	60
Graf č. 21 – ACF modelu $ARIMA(2,1,3) \times (0,1,0)_4$	61
Graf č. 22 – PACF modelu $ARIMA(2,1,3) \times (0,1,0)_4$	61
Graf č. 23 – ACF modelu $ARIMA(2,1,3) \times (1,1,1)_4$	63
Graf č. 24 – PACF modelu $ARIMA(2,1,3) \times (1,1,1)_4$	63
Graf č. 25 – ACF modelu $ARIMA(2,1,3) \times (2,1,1)_{4c}$	65

Graf č. 26 – PACF modelu $ARIMA(2,1,3) \times (2,1,1)_{4c}$	65
Graf č. 27 – ACF rezíduí modelu	67
Graf č. 28 – PACF rezíduí modelu	67
Graf č. 29 – Graf rezíduí modelu	68
Graf č. 30 – Histogram rezíduí modelu	69
Graf č. 31 – Prognóza ex-ante na rok 2019	73
Obrázok č. 1 – Teoretické grafy ACF a PACF	38
Obrázok č. 2 – Porovnanie diferenciácií logaritmovaného ČR pracujúcich v SS	54
Obrázok č. 3 – Model $ARIMA(0,1,0) \times (1,1,1)_{4c}$	57
Obrázok č. 4 – Model $ARIMA(0,1,0) \times (2,1,1)_{4c}$	59
Obrázok č. 5 – Model $ARIMA(2,1,3) \times (0,1,0)_4$	60
Obrázok č. 6 – Model $ARIMA(2,1,3) \times (1,1,1)_4$	62
Obrázok č. 7 – Model $ARIMA(2,1,3) \times (2,1,1)_{4c}$	64
Obrázok č. 8 – Porovnanie výsledných modelov	66
Obrázok č. 9 – Testy nezávislosti náhodných zložiek	68
Obrázok č. 10 – F-test homoskedasticity náhodných zložiek	69
Obrázok č. 11 – Kolmogorov-Smirnov test normality rezíduí modelu	70
Obrázok č. 12 – Miery presnosti prognózy ex-post	71

Úvod

Sledovanie zamestnanosti v Slovenskej republike je jednou z najdôležitejších úloh štátu. Priaznivý vývoj tohto ukazovateľa je základom rozvoja ekonomiky. Neustále prebieha analýza dát a hľadanie príčin nárastu alebo poklesu. Pre štát je výhodné, aby čo najviac ľudí pracovalo, a teda nemusel vynakladať veľa výdavkov na pomoc nezamestnaným občanom. Často môže zamestnanosť napovedať o spokojnosti či životnej úrovni občanov daného štátu.

Cieľom tejto práce je oboznámiť čitateľa so skúmanou problematikou. V prvej kapitole sme pristúpili k popísaniu trhu práce a jeho jednotlivých subjektov. Vymedzili sme hlavné pojmy potrebné na pochopenie správania trhu práce. Do tejto časti sme zahrnuli zoznámenie sa s metodikami Ústredia práce, sociálnych vecí a rodiny (ÚPSVaR), Štatistického úradu Slovenskej republiky (ŠÚ SR) a ich odlišnosťami v definíciách a výpočtoch. Taktiež sme podrobne analyzovali zamestnanosť na trhu práce, aby sme mali obraz o tom, ako to v skutočnosti vyzerá na Slovensku. K tomu nám pomohol aj Národný projekt Prognózy vývoja na trhu práce v SR II, ktorý vypracovala spoločnosť TREXIMA. Vymedzili sme nové pojmy, s ktorými sme sa v minulosti nestretávali, ako napr. dodatočná potreba pracovnej sily.

Následne je nevyhnutné určiť si hlavný cieľ, ktorého splnenie dosiahneme pomocou čiastkových cieľov. V tretej kapitole sme pomocou slovenskej a zahraničnej literatúry popísali problematiku analýzy časových radov, presnejšie Box-Jenkinsovu metodológiu. Hlavnou výhodou použitia tejto metódy je fakt, že dokáže úspešne popísať dynamické systémy, s ktorými sa v súčasnej ekonomike často stretávame. Prognostické metódy sú v súčasnosti dôležitým nástrojom pre pochopenie vývoja dát. Na základe metód v tejto kapitole sme schopný skonštruovať model, pomocou ktorého môžeme predpovedať budúci vývoj. Je to veľmi dôležité pre poskytnutie kvalitných a zodpovedných rozhodnutí pre trh práce.

V poslednej kapitole pristúpime k praktickej časti tvorby modelu a na základe analýzy vývoja minulosti, sa pokúsime vytvoriť modely pre sledovaný ukazovateľov. Následne spomedzi všetkých možností vyberieme najviac vyhovujúci model a porovnáme ho so skutočnosťou. V závere stručne zhodnotíme všetky získane informácie a ich využitie v praxi.

1. Súčasný stav riešenej problematiky doma a v zahraničí

Súčasný ekonomický rast Slovenska významne ovplyvňuje dianie na trhu práce. Pre správnu analýzu je potrebné si preštudovať všetky potrebné pojmy na pochopenie problematiky. Je dôležité poznať skutočný stav na Slovensku a v zahraničí. Taktiež pre analytika je dôležité poznať očakávania z externého prostredia, a tak lepšie pochopiť správanie trhu práce.

1.1 Trh práce

Trh práce chápeme ako veľmi citlivé a zraniteľné miesto, kde sa stretáva na jednej strane ponuka po práci, a na tej druhej dopyt po práci. Do skupiny ponuka po práci zaraďujeme obyvateľov, pričom cenou tejto práce je často mzda, resp. iná odmena za vykonanú prácu. Na strane dopytu to sú všetci tí, čo potrebujú zamestnancov, teda zamestnávateľa. Pre nich je dôležitá celková cena práce, ktorá obsahujú viaceré zložky ako napr. hrubá mzda, prípadné odvody a aj naturálne a iné odmeny, ktoré môžu zamestnanci získať. Všetky vzťahy medzi týmito dvomi stranami ovplyvňujú úroveň ceny práce.

Základnou črtou trhu práce je, že sa na ňom predáva samotná práca. Tu chápeme ako určitú pracovnú aktivitu viazanú na individuálne osoby, ktorá vykonáva uvedomelú činnosť za ohodnotenie. Aj napriek tomu, že sa tento trh javí ako trhový mechanizmus, neplatia preň rovnaké princípy ako pri ostatných voľných trhoch. Napríklad nestretneme sa tú s automatickou rovnováhou ponuky a dopytu pri voľnej súťaži. Dôležitým faktom je aj opatrné zaobchádzanie s cenou práce, ktorá sa nedokáže okamžite meniť, ako pri finančných trhoch. Taktiež samotná cena práce sa nedokáže prispôbiť vzniknutej prípadnej rovnováhe medzi dopytom a ponukou po práci, a to hlavne z rôznych legislatívnych dôvodov, ako je napr. určenie minimálnej mzdy, výsledky kolektívneho vyjednávania či rôzne nastavené mzdové tarify.

A preto je trh práce spomedzi všetkých trhov veľmi špecifickým miestom, vyznačujúcim sa nižšou flexibilitou ako iné trhy. Za zmienku stojí oboznámenie sa s dôležitým pojmom tzv. flexi-istoty. Tento pojem berie do úvahy na jednej strane pružnosť trhu práce a na tej druhej zase určité istoty pre pracovníkov. (Lubyová, Sodomová, 2016)

Podľa zákonníka práce považujeme za zamestnanca, každú osobu, ktorá je v pracovnom pomere alebo v obdobnom pracovnom vzťahu a vykonáva pre zamestnávateľa závislú činnosť.

1.2 Metodika podľa ÚPSVaR

Pri výpočte ekonomického aktívneho obyvateľstva dochádza k určitej nekonzistencii, keďže sa tento ukazovateľ počíta z priemerného počtu pracujúcich za minulý rok podľa ŠÚ SR a z priemerného počtu evidovaných nezamestnaných za minulý rok podľa metodiky ÚPSVaR. Takže sa tu miešajú dve rôzne metodiky a táto hodnota sa používa pre celý rok. Pre lepšie pochopenie sme uviedli vzorec:

$$EAO_{\dot{U}P,t} = ZAM_{\dot{S}\dot{U},t-1} + NEZ_{\dot{U}P,t-1},$$

kde

$EAO_{\dot{U}P,t}$ – ekonomicky aktívne obyvateľstvo podľa ÚPSVaR v čase t,

$ZAM_{\dot{S}\dot{U},t-1}$ – priemerný počet pracujúcich podľa ŠÚ SR za obdobie t-1,

$NEZ_{\dot{U}P,t-1}$ – priemerný počet evidovaných nezamestnaných podľa ÚPSVaR za obdobie t-1.

Dôležitým pojmom pri metodike ÚPSVaR je uchádzač o zamestnanie. Stáva sa ním každý občan, ktorý je schopný pracovať, aktívne si hľadá zamestnanie, chce pracovať a dôležitou podmienkou je podanie písomnej žiadosti o zaradenie do evidencie uchádzačov o zamestnanie na príslušnom úrade práce. Táto osoba musí spĺňať aj nasledujúce podmienky:

- nie je zamestnanec,
- neprevádzkuje samostatnú zárobkovú činnosť,
- nevykonáva zárobkovú činnosť v zahraničí,
- nie je v pracovnoprávnom pomere podľa dohody o práci vykonávanej mimo pracovného pomeru, pričom tu je výnimka, podľa ktorej môže mať túto dohodu, ak jeho pracovný pomer nepresiahne dokopy 40 dní v kalendárnom roku a mesačná odmena nie je vyššia ako životné minimum na jednu plnoletú osobu. (ÚPSVaR, 2020)

Do tejto skupiny sa nezaraďujú záujemcovia o zamestnanie, t. j. osoby, ktoré hľadajú iné zamestnanie alebo si chcú zlepšiť popri prípade zmeniť kvalifikáciu pomocou rôznych kurzov a školení. Taktiež sem nezaraďujeme osoby pripravujúce sa na výkon povolania, dôchodcov, dočasne práceneschopných a občanov, ktorí požiadali o vyradenie z tejto evidencie, nespolupracujú s úradmi práce, vykonávajú prácu nelegálneho typu alebo im bolo umožnené pracovať v zahraničí.

Mieru evidovanej nezamestnanosti následne na základe dohovoru s Medzinárodnou organizáciou práce, počíta ako podiel disponibilných uchádzačov o zamestnanie, ktorí môžu okamžite nastúpiť do zamestnania a počtu ekonomicky aktívnych osôb podľa metodiky ÚPSVaR. Taktiež sa rozlišuje registrovaná miera nezamestnaností, ktorej výpočet je obdobný ako pri evidovanej s tým rozdielom, že sa nepoužíva disponibilný ale celkový počet uchádzačov o zamestnanie.

1.3 Metodika podľa ŠÚ SR

Na Slovensku monitoruje ŠÚ SR zamestnanosť a nezamestnanosť pomocou výberového zisťovania pracovných síl (VZPS) v domácnostiach. Táto metóda získavania údajov bola spustená v roku 1993 a plne korešponduje s pravidlami Medzinárodnej organizácie práce (ILO). Výberové zisťovanie pozostáva zo štvrtročného náhodného výberu bytov na území Slovenska.

Vzorku pri zisťovaní tvorí 10 250 bytov, ktoré predstavujú 0,6 % trvalo obývaných bytov. Dôležitým krokom je aj pokrytie celého územia Slovenska, preto sa výber robí vo všetkých 79 okresoch. ŠÚ SR získava informácie o každej osobe, ktorá žije v domácnosti zvoleného bytu. Takýmto spôsobom sa každý štvrtrok zahrnie do VZPS približne 23 tisíc osôb s vekom vyšším ako 15 rokov. (ŠÚ SR, 2020)

Do výberu sa nezahrňa tzv. inštitucionálne obyvateľstvo, t. j. obyvatelia vo väzniciach alebo kláštoroch. Každý štvrtrok prebieha obmena 20 % vzorky, pričom každý vybraný byt je zahrnutý do vzorky na najbližších päť období.

Na základe výsledkov sa následne počíta ekonomicky aktívne obyvateľstvo. Podľa VZPS tu zaraďujeme osoby staršie ako 15 rokov, ktoré pracujú alebo sú nezamestnané. Je potrebné podotknúť, že počas rokov 1997 až 2005 do tejto skupiny patrili aj vojaci vykonávajúci vojenskú službu.

Za pracujúcich ŠÚ SR považuje všetky osoby staršie ako 15 rokov, ktoré počas sledovaného týždňa:

- vykonali aspoň 1 hodinu práce za mzdu, plat alebo inú odmenu (taktiež v prípade, že išlo o prácu na plný alebo iný pracovný čas, či už to bola dočasná alebo stála práca),
- pracovali v zahraničí menej ako jeden rok, pričom za prácou pravidelne dochádzajú,
- majú prácu, ale kvôli rôznym dôvodom nemohli pracovať (čerpanie dovolenky, choroba, štrajk, výluka, materská dovolenka, s výnimkou osôb na rodičovskej dovolenke a osôb na dlhodobom neplatenom voľne),
- zaraďujeme sem aj osoby, ktoré sú členmi domácnosti podnikateľov a pomáhajú bez toho aby poberali mzdu alebo odmenu.

Za nezamestnanú osobu ŠÚ SR považuje všetky osoby od 15 do 74 rokov, ktoré nemajú žiadnu platenú prácu v sledovanom týždni. Za posledné štyri týždne aktívne pristupujú k hľadaniu práce alebo už prácu majú dohodnutú, no nastúpia na výkon práce najneskôr do troch mesiacov. Patria sem aj osoby, ktoré dokážu začať pracovať do dvoch týždňov. Za aktívne hľadanie považujeme spôsob nájdenia práce pomocou úradu práce alebo iných sprostredkovateľov práce, návštevy pohovorov alebo príprava sa na podnikateľskú činnosť.

Na základe vyššie popísaných poznatkov medzi ekonomicky neaktívne obyvateľstvo podľa VZPS zaraďujeme osoby mimo trhu práce, t. j.:

- osoby do 15 rokov,
- osoby staršie ako 15 rokov, ktoré sú nezamestnané a nehľadajú si prácu alebo nedokážu začať pracovať do 14 dní,
- osoby pripravujúce sa na výkon práce, dôchodcovia, osoby na rodičovskej dovolenke,
- a zaujímavou skupinou sú tzv. odradení, ktorí si nehľadajú prácu, lebo jednoducho neveria, že nejakú nájdu aj keď chcú pracovať. (ŠÚ SR, 2020)

1.4 Analýza súčasného stavu zamestnanosti

Slovensko už dlhšiu dobu patrí pod EÚ, a preto nesmie zabudnúť sledovať aj ekonomické strategické ciele tohto zoskupenia krajín. Trh práce zaraďuje EÚ do oblasti s vysokou prioritou. Prijatím stratégie Európa 2020, EÚ zdôraznila potrebu zvyšovania zamestnanosti a produktivity. V roku 2010 sa nastavil jeden z hlavných cieľov pre rok 2020, že v priemere traja zo štyroch ekonomicky aktívnych občanov EÚ bude zamestnaných. To znamená v percentuálnom vyjadrení, že miera zamestnanosti má mať hodnotu 75 %. Nevyhnutnou potrebou na dosiahnutie tohto cieľa sú investície do vzdelávania. (Stratégia Európa 2020, 2010)

Členské štáty EÚ majú za úlohu zvyšovať relevantnosť vysokoškolského vzdelania pre potreby trhu práce. Poskytovanie kvalitného vzdelania počas celého pracovného života, zvyšuje schopnosť prispôbovať sa zmenám pracovných činností a odolnosť hospodárstva voči výkyvom. A preto bol v roku 2017 prijatý Európsky pilier sociálnych práv. Medzi základné zásady patria:

- rovnaká príležitosť a prístup na trh práce,
- spravodlivé mzdové oceňovanie a pracovné podmienky,
- sociálna ochrana a začlenenie.

Je veľmi dôležité zabezpečiť možnosť rekvalifikácie počas celého života a mladým ľuďom umožniť prístup ku kvalitnému vzdelávaciemu systému a rôznym stážam.

Aktuálne sa na Slovensku zvyšuje miera zamestnanosti, a to aj vďaka zvýšenej tvorbe počtu voľných pracovných miest, čo spôsobuje následný pokles počtu uchádzačov o zamestnanie. Zaujímavý je hlavne prílev pracovníkov z tretích krajín, ktorý bol zapríčinený hlavne rozhodnutím vlády o zjednodušení spôsobu zamestnania migrujúcich pracovníkov.

Hospodárstvo SR rastie solídnym tempom, pričom hlavným dôvodom ekonomického rastu je u nás automobilový priemysel. To môžeme považovať za vynikajúcu výhodu, no musíme sa na to pozrieť aj z druhej strany. Závislosť SR na výsledok jedného typu priemyslu je hazard, a to hlavne kvôli citlivosti na externé vplyvy.

Firmy sú menej optimistickéjšie, čo sa reflektuje aj na úrovni zamestnanosti v priemysle. Hlavným limitom je u nás nedostatok kvalifikovanej sily, ktorej každým

rokom ubúda. V súčasnosti sa tento problém snaží vyriešiť rezort práce, sociálnych vecí a rodiny, ktorý zrealizoval projekt „Pripravený na prácu“. Tento projekt je veľkou príležitosťou pre evidovaných nezamestnaných, ktorí majú chuť pracovať v automobilovom priemysle, pričom platnosť projektu je do 31. marca 2021 s predpokladanými nákladmi 30,9 milióna eur. (ÚPSVaR, 2019)

Počas trvania projektu je umožnené 20 000 uchádzačom o zamestnanie sa rekvalifikovať. V prvotnej fáze spoznávajú základnú odbornú terminológiu v automobilom, strojárskom a elektrotechnickom priemysle. Veľkou výhodou sú hlavne praktické ukážky a nácvik, ktorý tak urýchľuje rekvalifikáciu. Taktiež pripravujú uchádzačov na pohovor, pomáhajú im s vytvorením štandardizovaného životopisu a zo základmi lean manažmentu, či s ďalším odborným vzdelaním. Vďaka všetkým týmto krokom dokáže Slovensko zvýšiť kvalifikovanosť zamestnancov a obsadiť ich pre pozície ako sú napr. mechatronik, elektrotechnik automatizačnej techniky, atď.

Projekt realizuje konzorcium Deutsch-Slowakische Akademien (DSA), dcérska spoločnosť nemeckého vzdelávania koncernu EBG a spoločnosť Newport Group (NG), pričom spolupracujú aj rôznymi duálnymi akadémiami, ktoré spolupracujú so spoločnosťami Siemens či Volkswagen.

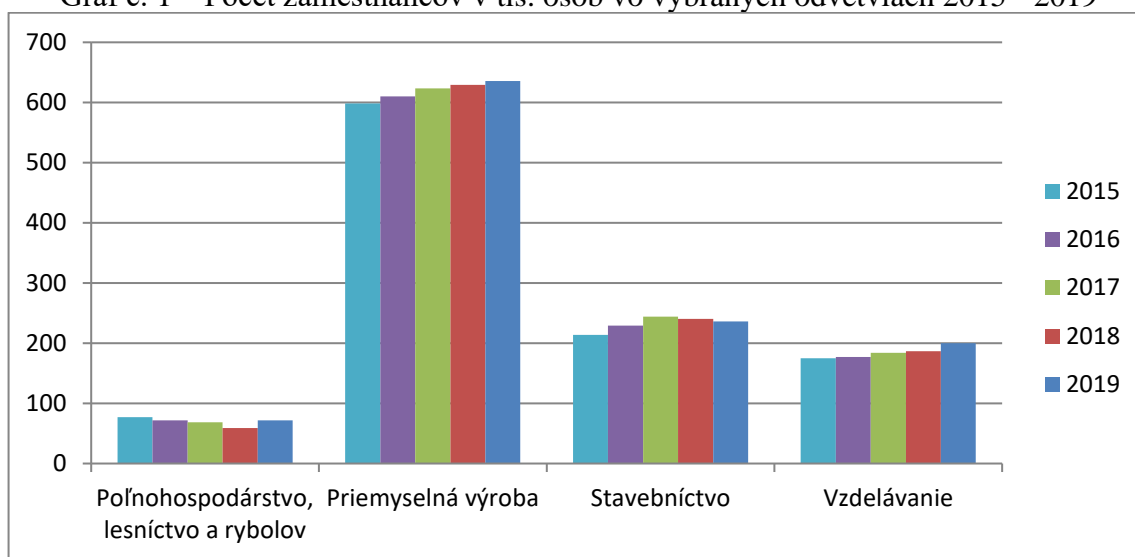
1.4.1 Dodatočná potreba pracovných síl

Za posledných päť rokov sa na území SR zlepšilo vyhodnocovanie potrieb trhu práce. Bolo to zapríčinené hlavne realizáciou Národného projektu Prognózy vývoja na trhu práce v SR. Tento projekt definoval dôležitý faktor na trhu práce, a to dodatočnú potrebu zamestnancov. Tento ukazovateľ vyjadruje počet pracovných príležitostí v nadchádzajúcich rokoch, ktoré kvôli rôznym okolnostiam nie je možné obsadiť z aktuálneho počtu zamestnaných, pričom výpočet bral do úvahy počet vzniknutých, zaniknutých a uvoľnených pracovných miest z dôvodu odchodu pracovníkov z trhu práce.

V roku 2016 Slovensko zaznamenalo historicky najvyšší nárast zamestnaných osôb, a to 55 tisíc osôb. Nasledujúci rok sme si stále držali slušnú úroveň počtu zamestnaných, presnejšie sme zaznamenali nárast o 52 tisíc osôb. Spomalenia sa však dostavilo v roku 2018, keď počet zamestnancov vzrástol o približne 46 tisíc osôb. Za poklesom stojí fakt, že aj pri vytváraní nových pracovných miest, nedokážeme nájsť

dostatočne kvalifikovaných ľudí na tieto pozície. Na základe grafu č. 1, zobrazujúci vývoj počtu zamestnancov vo vybraných odvetviach, je hlavným ťahúňom stále priemysel, ktorý tvoril približne 20 % z celkového nárastu počtu zamestnaných osôb. K zvýšeniu prispeli aj oblasť Vzdelávanie a Stavebníctvo, kde sme zaznamenali nárast o približne 7 a 5 tisíc osôb. (MPSVaR, 2019)

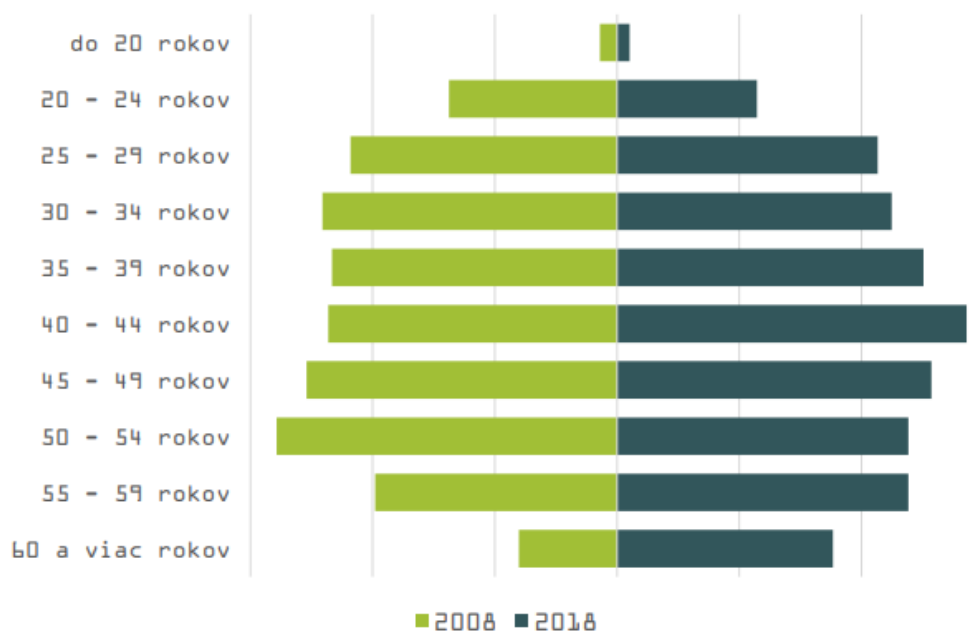
Graf č. 1 – Počet zamestnancov v tis. osôb vo vybraných odvetviach 2015 - 2019



Zdroj: Vlastné spracovanie dát zo ŠÚ SR

K priaznivejšej situácii na trhu práce neprispieva oblasť poľnohospodárstva, kde už od roku 2015 dochádza k postupnému poklesu počtu zamestnaných. V súčasnosti sa však stretávame so spoločným problémom vo všetkých oblastiach. Zlá alokácia absolventov do potrebných odvetví a zvyšovanie disponibilného počtu uchádzačov o zamestnanie spôsobuje spomalenie ekonomiky. Do úvahy musíme zobrať aj fakt, že populácia Slovenska starne, čo potvrdzuje aj graf č. 2 porovnávajúci vekovú štruktúru v roku 2008 a 2018.

Graf č. 2 – Vývoj vekovej štruktúry v hospodárstve SR

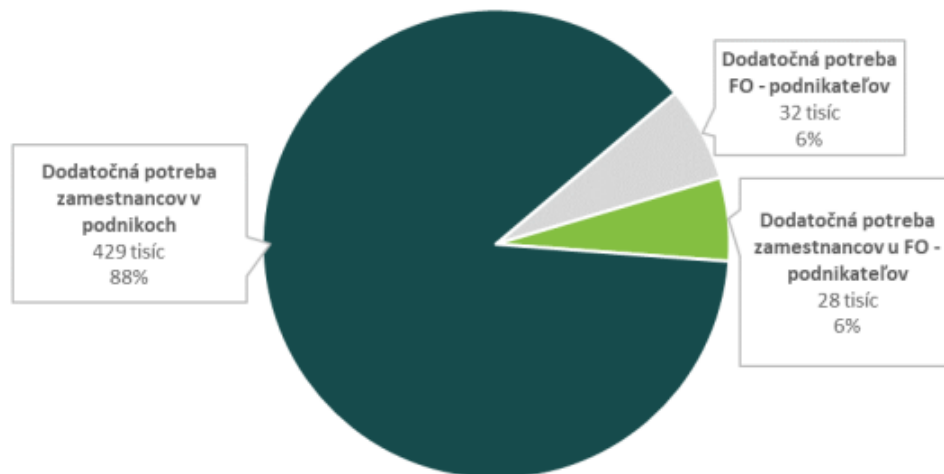


Zdroj: TREXIMA, 2018

Na trh práce sa dostávajú tzv. miléniové deti, ktorých ročníky považujeme z hľadiska početnosti za najslabšie. Zamestnávateľom robí vrásky na tvári aj odchod najsilnejších ročníkov, tzv. povojnové ročníky. Aj napriek klesajúcej miere nezamestnanosti nedokážu obsadzovať všetky potrebné pozície pre zvyšujúci sa počet objednávok. Pomôcť môže prílev cudzincov z tretích krajín do oblasti s veľkými problémami so zamestnanosťou a uľahčenie procesu ich zamestnania.

Spoločnosť TREXIMA, ktorá dlhodobo zbiera údaje od zamestnávateľov následným analyzovaním dospela k záveru, že do roku 2024 bude na Slovensku dodatočná potreba pracovných síl dosahovať úroveň približne 490 tisíc osôb. Túto potrebu následne rozdelili do troch základných skupín, ktoré zobrazuje graf č. 3.

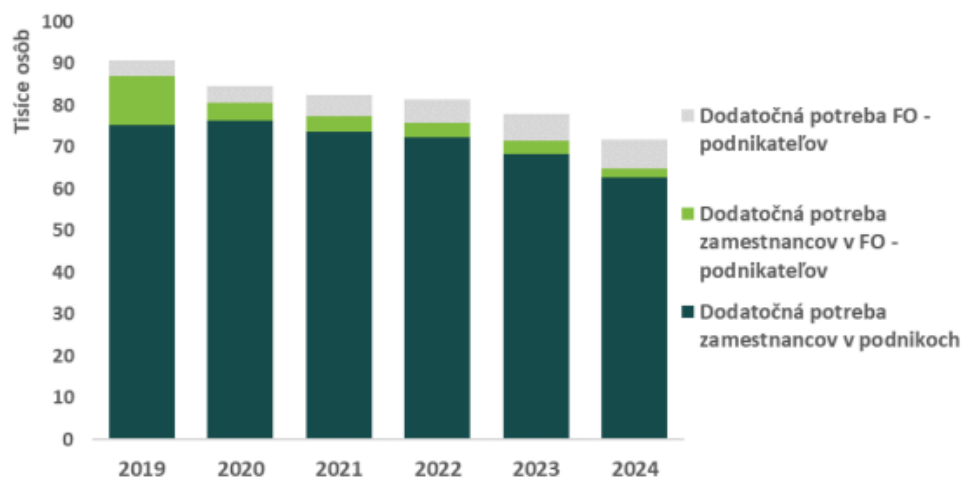
Graf č. 3 – Dodatočná potreba pracovných síl do roku 2024



Zdroj: TREXIMA, 2018

Predpokladá sa postupný pokles dodatočnej potreby, spôsobený postupným spomalením expanzie ekonomiky od roku 2020. Prognózu od spoločnosti TREXIMA na jednotlivé roky zobrazuje nasledujúci graf č. 4.

Graf č. 4 – Vývoj dodatočnej potreby pracovných síl 2019-2024



Zdroj: TREXIMA, 2018

Náhradu pracovných síl považujeme za jeden z hlavných faktorov pri potrebe fyzických osôb – podnikateľov. Za fyzické osoby – podnikateľov považujeme osoby vykonávajúce podnikavú činnosť na vlastnú zodpovednosť a riadia sa podľa osobitných predpisov. Patria sem živnostníci, osoby so slobodným povoláním a samostatne hospodáriaci roľníci.

V súčasnosti sa predpokladá postupný pokles počtu fyzických osôb – podnikateľov približne o 6 tisíc osôb do roku 2024, ale tento jav môžeme pozorovať už posledných 10 rokov. Štruktúra tejto zložky pracovného trhu sa však vyznačuje staršími ročníkmi, čo zapríčini v najbližších 5 rokoch početný odchod z trhu práce. Potreba fyzických osôb – podnikateľov na Slovensku je definovaná na základe dopytu spoločnosti po tovaroch a službách. (MPSVaR, 2019)

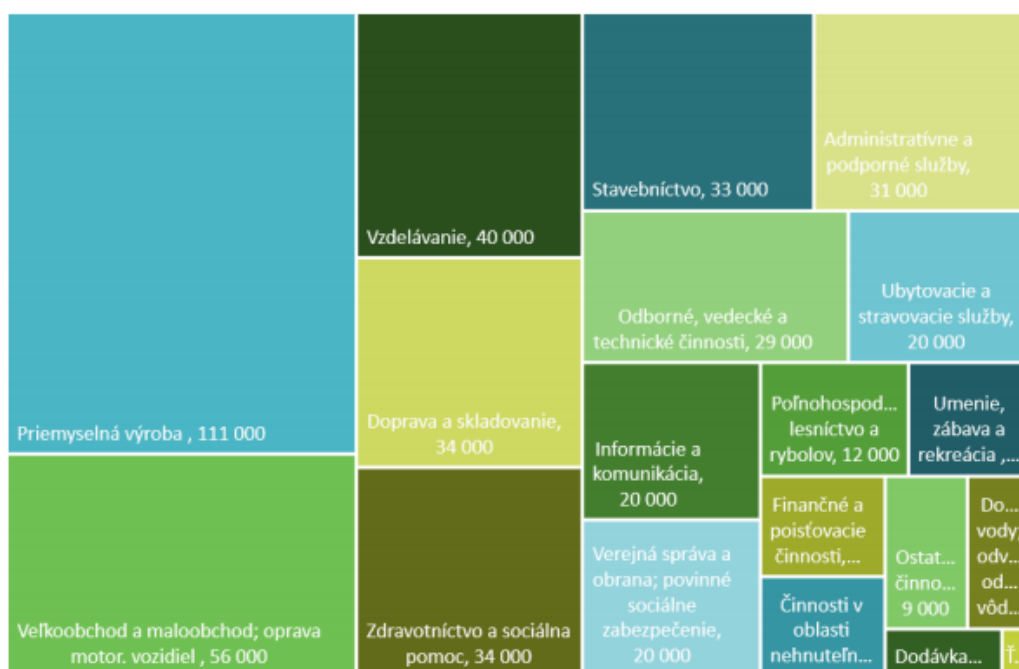
Nedostatočným počtom fyzických osôb – podnikateľov má za následok rast cien tovarov a služieb. Avšak prázdne miesto na trhu práce sa pri tejto situácii rýchlo vyplní konkurenciou. To znamená, že pri odchode z trhu práce nebude nahradený inou fyzickou osobou – podnikateľom, ale rozšírením produkcie iných subjektov.

Prílev cudzincov do našich radov zamestnaných nemá vždy pozitívny vplyv. Nesprávnym alokovaním do potrebných oblastí môžeme dospieť k stavu, keď pracovníci zo zahraničia spôsobia pokles miezd a zamestnanosti domáceho obyvateľstva. Zamestnávateľom na rozdiel od štátu ide o zisk, a preto v prípade priaznivých podmienok na zamestnanie cudzincov nastane nezáujem zamestnávateľov hľadať v radoch nezamestnaných ľudí Slovenska.

Štát musí dať pozor na to aby sa zahraniční pracovníci nestali substituálnou, ale komplementárnou zložkou pracovného trhu. Obsadením týchto osôb do oblastí so skutočným nedostatkom z rád Slovákov, spôsobí zvýšenie počtu zamestnaných, úroveň produkcie a hlavne sa zvýši výška daňových príjmov štátu, čo je výhodne pre Slovensko. Pri správnych rozhodnutiach v decíznej sfére cudzinci tak neohrozia domáce pracovné zložky.

Dodatočnú potrebu na úrovni odvetví národného hospodárstva máme zobrazenú na nasledujúcom grafe č. 5. Najdôležitejšími odvetviami budú priemysel a obchod, pretože približne tretina vytvorených pracovných ponúk v budúcnosti bude v týchto dvoch odvetviach.

Graf č. 5 – Dodatočná potreba v odvetviach národného hospodárstva v r. 2019-2024



Zdroj: TRIXIMA, 2018

Na Slovensku existuje množstvo podskupín zamestnaní. TRIXIMA predpokladá, že v rokoch 2019 až 2024 bude pätinu potrieb trhu práce tvoriť dopyt po špecialistoch. V tejto kategórii sa nachádza až 90 podskupín zamestnaní. Na druhej strane približne len 30 podskupín obsahujú administratívne, pomocné a nekvalifikované pozície. Taktiež nižšie množstvo podskupín v prípade rôznych manuálnych prác. Z toho vyplýva že systém na Slovensku v skupinách s postačujúcou nízkou kvalifikáciou koncentruje potreby do niekoľkých zamestnaní.

Až 56 % celkovej potreby je v prípade nekvalifikovaných pracovníkov koncentrovaná do troch podskupín, a preto pri porovnávaní na úrovni zamestnaní sú v popredí nízko kvalifikované pozície než špecializované, keďže sú viac početnejšie. Napr. pri špecialistoch je do najpočetnejších zamestnaní koncentrovaná len 21 % potreby. Nesmieme ale zabudnúť, že dopyt na trhu práce bude vždy vyšší pri špecialistoch, len je tento dopyt rozdelený do viacerých podskupín, preto jeho relatívne vyjadrenie na jednotlivé zamestnania nie je tak vysoké. (MPSVaR, 2019)

Najhoršie na tom budú upratovačky, vodiči nákladných automobilov a kamiónov, montážni pracovníci v strojárskvej výrobe, kde dodatočná potreba do roku 2024 presiahne hranicu 10 tisíc osôb. Tieto pozície sú na tom zle, ak sa na to pozrieme z absolútneho hľadiska. Dôležité je aj relatívne posúdenie, pretože aj keď je absolútna potreba rovnaká

relatívne vyjadrenie môže upozorniť na vyššie zaťaženie. Napríklad ak máme dvoch zamestnávateľov, z ktorých jeden zamestnáva 10 tisíc osôb a ten druhý 200, v prípade potreby 100 osôb je dodatočná potreba v relatívnom vyjadrení u prvého len 1 %, no u menšieho zamestnávateľa je to až 50 %. Pri podskupinách ako sú napr. vývojári softvéru, programátori, správcovia databáz, opatrovatelia starších ľudí bude podiel viac ako 50 %.

1.4.2 Správanie podnikov na trhu práce

Spoločnosť TREXIMA v roku 2018 vykonala terénny prieskum, v ktorom oslovila spolu 1418 podnikov s viac ako 620 tisíc pracovníkmi, čo tvorilo približne 30 % zamestnaných v podnikoch. Výberovú vzorku tvoril mix podnikov zo všetkých odvetví ekonomických činností vo všetkých regiónoch. Zber údajov prebiehal pomocou osobných rozhovorov a elektronicky. Pričom sa výberová vzorka vytvorila na základe stratifikovaného výberu bez opakovania.

Následným spracovaním údajov, dospeli k zaujímavým číslam, ktoré nám pomôžu pochopiť správanie sa trhu práce. Podniky na Slovensku potrebujú v priemere až 128 dní na obsadenie voľného miesta. Oproti roku 2016, kde bolo potrebných 86 dní, ide o výrazný nárast. Najväčším problémom je nájsť dostačujúco kvalifikovaných ľudí na riadiace pozície. Ďalším problémom je neochota mladých ľudí vykonávať prácu ako operátor či montér strojov a zariadení. Často sa stretávame aj s neadekvátnymi požiadavkami na ohodnotenie týchto prác.

Najkratšie potrebné obdobie na obsadenie pracovnej pozície je v službách a obchode, kde dosahujeme hranicu 80 dní. Taktiež sa rýchlo obsadzujú administratívne pozície, a to v priemere za 110 dní. Aj tento fakt nasvedčuje súčasnému trendu nezáujmu o manuálnu prácu. (TREXIMA, 2019)

Dôležité je aj zdôrazniť, že voľné pracovné miesto nám nezvyšuje automaticky dodatočnú potrebu trhu práce. V súčasnosti na trhu práce dochádza k vysokej fluktuácii, ktorá je príčinou až 62 % voľných miest. Takýto zamestnanec však ostáva stále na trhu práce len zmenil zamestnávateľa. Krásnym príkladom je oblasť ubytovacích a stravovacích služieb, kde dochádza k fluktuácii 90 %.

Hlavnou príčinou sú mzdy, ktoré sú najnižšie spomedzi všetkých odvetví. Tento jav pozorujeme aj v IT odvetví. IT pracovníci, vďaka svojej kvalifikovanosti vedia akú majú hodnotu a môžu si vybrať zamestnávateľov. Takýto pracovníci opúšťajú pracovné miesta,

no stále sú súčasťou trhu práce, pričom takéto voľné pracovné miesta nie je potrebné doplniť napr. z pracovných síl zo zahraničia.

Podnikatelia v oblasti informačných a komunikačných technológií sú však pozitívne naladení a očakávajú zvýšenie objemu poskytnutých služieb. Podobne aj odvetvia ako napr. finančníctvo, poisťovníctvo, veľkoobchod a maloobchod veria v nárast vyprodukovaných tovarov a služieb. Na druhej strane postupným nezájmom o manuálnu prácu je kladne naladená len pätina zamestnávateľov v ťažbe a stavebníctve. Sem patria aj odvetvia poľnohospodárstvo, doprava, lesníctvo a rybolov, kde nárast predpovedá len tretina zamestnávateľov.

TREXIMA taktiež tvrdí, že zamestnávatelia majú problém s obsadením až polovice voľných miest zapríčinených nedostatkom osôb s potrebným odborom vzdelania. Zaujímavé je, že problém s obsadením nemajú iba v 8 % voľných miest. Ďalším faktorom neobsadenia je aj nízka atraktivita práce, približne 17 % prípadov.

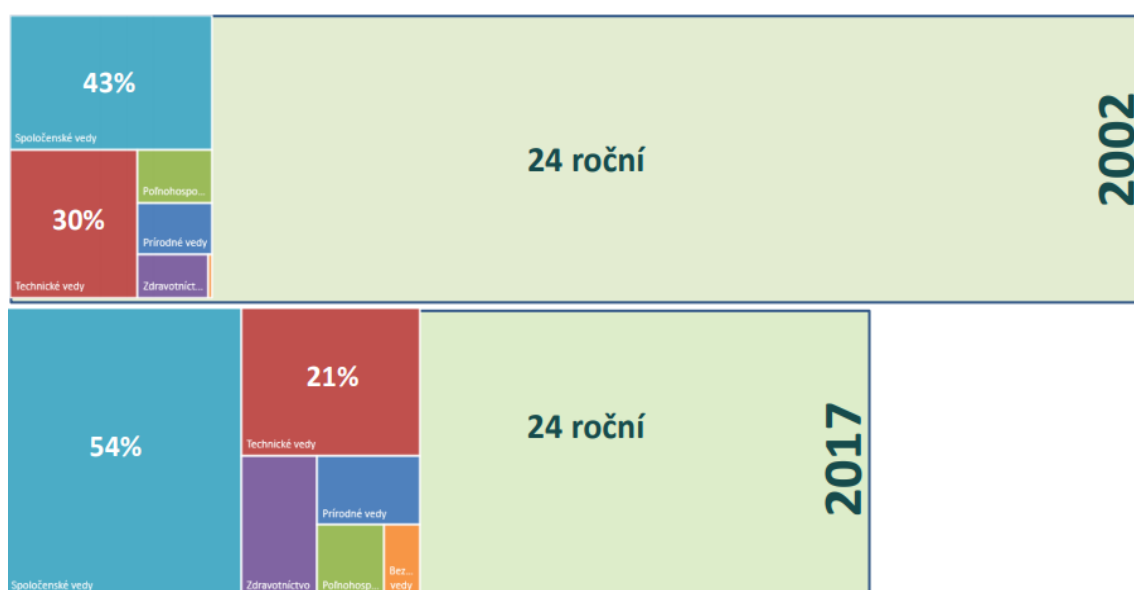
1.4.2 Prílev absolventov na trh práce

Za absolventov prichádzajúcich na trh práce, považujeme osoby, ktoré na Slovensku ukončia niektorý stupeň vzdelávania a nerozhodnú sa pokračovať ďalej v štúdiu.

Potreby zamestnávateľov výrazne závisia od systému vzdelávania a zloženia absolventov škôl. Tieto osoby tvoria významnú zložku ponuky práce. Bohužiaľ za klesajúcim počtom príchodu absolventov na trh práce stojí demografický vývoj pôrodnosti v rokoch 2002 až 2003. Počas tohto obdobia dosiahlo Slovensko historicky najmenej narodených detí. Preto predpokladáme pokles príchodu absolventov na trh práce v rokoch 2020 až 2021.

Na základe demografického vývoja pôrodnosti, sa predpokladá v rokoch 2019 až 2024 príchod 275 tisíc absolventov. Slovensko sa preto dostáva do situácie, kedy dopyt silno prevýši ponuku na trhu práce. V minulosti bola výhoda na strane zamestnávateľov a mohli si vybrať koho zamestnajú. Teraz však majú lepšie postavenie zamestnanci, ktorí si môžu vybrať zamestnávateľa a požadovať adekvátnejšie podmienky ohodnotenia. Preto musia zamestnávatelia kvôli ich nedostatku zvädzať konkurenčný boj. Všetky tieto okolnosti vyúsťujú do rastu miezd, ktoré môžu naspäť prilákať Slovákov pracujúcich v zahraničí. Pokles a štruktúru vysokoškolákov zobrazuje graf č. 6.

Graf č. 6 – Podiel vysokoškolákov na trhu práce 2002 oproti 2017



Zdroj: TRESIMA, 2018

Momentálne panuje znepokojujúca situácia ohľadom uplatnenia absolventov vysokých a stredných škôl z pohľadu odboru, ktorý vyštudovali. Počas štúdia len približne 10 % študentov predpokladá, že sa nebudú venovať po škole tomu, čo študujú. V skutočnosti ale nastáva jav, keď až 50 % absolventov vysokých škôl nachádza uplatnenie mimo svojho odboru. (MPSVaR, 2019)

Absolventi pracujúci mimo svoj vyštudovaný odbor znižujú využiteľnosť ekonomického potenciálu Slovenska. Vzniká tak nedôvera v systém školstva a frustrácia zamestnávateľov, ktorí musia vynakladať ďalšie dodatočné náklady na školenia osôb z iných študijných smerov než zamestnávateľ potrebuje.

Postupnou neochotou absolventov, či už dobrovoľne alebo vplyvom trhu práce, využiť vedomosti a zručnosti získané počas štúdia, zvyšujú neefektivitu vzdelávania a vynaložených finančných prostriedkov. V takýchto prípadoch sa často stretávame s fluktuáciou a so zvýšenou mierou výdavkov na aktívne opatrenia trhu práce. Tiež takéto osoby nadobúdajú nižšie mzdy a produktivitu práce než ich spolupracovníci, ktorí vyštudovali v prislúchajúcom odbore.

Na trhu práce môže nastať aj situácia, že absolventi vysokých škôl začínajú obsadzovať pozície, na ktoré je postačujúce stredoškolské vzdelanie. Jedným z dôvodov je nedostatočná kvalitatívna a kvantitatívna zhoda medzi trhom práce a systémom vzdelávania.

V roku 2012 takto pracovalo 23 % vysokoškolsky vzdelaných, v súčasnosti je to už 26 %. Príčinou je ťažšie uplatnenie mladých absolventov hneď po škole, hlavne ak počas štúdia nepracovali a nenadobudli tak prax. Situácia sa nemení ani po piatich rokoch po vyštudovaní, keď v roku 2012 pracovalo na stredoškolskej pozícií 30 % absolventov s vysokou školou. V súčasnosti sme presiahli hranicu 34 %. (MPSVaR, 2019)

Tento problém nie je len na Slovensku, ale týka sa aj iných krajín Európskej Únie. Kvalitné vzdelanie je jeden z hlavných predpokladov dosiahnutia žiadaného hospodárstva členských krajín. Musíme sa zamerať na využitie plného potenciálu vzdelávania, aby sa zabezpečoval rovnoprávny prístup ku kvalitnému vzdelávaniu alebo rôznym školeniam počas všetkých fáz života. Kľúčovou prioritou všetkých krajín EÚ je úspešné nastavenie procesov vzdelávania, ktoré pomôže znížiť nezamestnanosť absolventov a zvýši ich rýchlosť adaptovať sa na prácu v odbore.

2. Cieľ práce

Hlavným cieľom práce je analýza zamestnanosti na trhu práce v Slovenskej republike. Pre naše potreby sme si zvolili využiť ARIMA modelov. Keďže ekonomické časové rady sú často poznačené sezónnosťou, budeme musieť pristúpiť ku komplexnej analýze časového radu.

V tejto práci sme sa zamerali hlavne na vývoj počtu zamestnaných osôb v súkromnom sektore, získaných pomocou výberového zisťovania pracovných síl, ktoré uskutočňuje Štatistický úrad Slovenskej republiky. Analyzovať budeme štvrt'ročné údaje od roku 1994 po rok 2018, pomocou ktorých sa pokúsime predpovedať vývoj na ďalšie štyri obdobia.

Na analýzu a modelovanie bude použitý program Statgraphics Centurion XVI a pomocné programy ako napr. Excel a SAS Enterprise Guide.

Pre dosiahnutie hlavného cieľa sme museli pristúpiť k postupnému splneniu nasledujúcich pomocných cieľov:

- popísanie vývoja počtu zamestnaných osôb v súkromnom sektore,
- na základe predošlej analýzy vykonanie predbežných úsudkov o tvare modelu,
- kombináciou parametrov vytvoriť viacero modelov, spomedzi ktorých vyberieme najvhodnejší,
- overenie kvality modelu, štatistickej významnosti parametrov a preskúmanie heteroskedasticity hodnôt,
- vytvorenie finálnej prognózy na základe vybraného ARIMA modelu.

3. Metodika práce

Skôr ako pristúpime k analýze časového radu, je veľmi dôležité zhromaždiť čo najviac informácií o danej problematike a prejsť si štatistické poznatky o časových radoch. V tejto kapitole sa budeme zaoberať teoretickým popisom základných vlastností časového radu a metód, ktoré v práci využijeme.

Časový rad, môžeme chápať ako súbor pozorovaní, ktoré sú chronologicky zoradené vo vzostupnom poradí. V našej práci budeme používať dáta, pri ktorých je interval medzi jednotlivými pozorovaniami konštantný.

3.1 Stacionarita časového radu

Stacionarita je vlastnosť časového radu, ktorá u hodnôt vyvoláva tendenciu vracať sa ku konštante. Rozlišujeme dva základné typy stacionarity:

- Striktná stacionarita tvrdí, že správanie sa náhodného procesu je invariantné voči posunu v čase. Pravdepodobnostne rozdelenie náhodného vektoru $(x_{t_1}, \dots, x_{t_k})$ je z definície rovnaké ako rozdelenie náhodného vektoru $(x_{t_1+h}, \dots, x_{t_k+h})$ pre ľubovoľný časový posun h a ľubovoľné časy t_1, \dots, t_k .
- Slabá stacionarita na rozdiel od striktnej požaduje, aby mal príslušný náhodný proces konštantnú strednú hodnotu ($E(x_t) = \mu$), konštantný rozptyl ($var(x_t) = \sigma^2$) a pre kovarianciu platnosť vzťahu $cov(x_t, x_s) = cov(x_{t+h}, x_{s+h})$ pre každé časové okamihy t, s a pre každý časový posun h .

Avšak pre väčšinu ekonomických časových radov je typický opak, a teda nestacionárnosť. Preto sa musí časový rad hodnôt často transformovať. Medzi najpoužívanejšie transformácie patria:

- odpočítanie priemeru $y_t - \hat{y}$ zaručí $E y_t = 0$, ale nezaručí stacionaritu modelu, tzn., že transformácia je výhodná pri stacionárnych modeloch s $E y_t \neq 0$,
- odpočítanie trendu $y_t - a - b_t$ môže odstrániť nestacionaritu, transformácia dobre funguje pri lineárnom trende v časovej rade,
- 1. diferenciacia $D_t = y_t - y_{t-1}$ výhodné pri nestacionarite v strednej hodnote,
- 2. diferenciacia $D_t - D_{t-1}$ používa sa pri modeloch s nestacionaritou v rozptyle,
- logaritmovanie $\ln y_t$,

- diferencie logaritmov $\ln y_t - \ln y_{t-1}$ často používané na modelovanie finančných časových radov,
- normované diferencie $(y_t - y_{t-1})/y_t$.

Nestacionárnosť časového radu vieme zistiť aj na základe grafického posúdenia priebehu autokorelačnej funkcie ACF, pri ktorej dochádza k extrémnemu a pomalému klesaniu jednotlivých koeficientov. Často je však grafické posúdenie nepostačujúce a svoje rozhodnutie musíme podložiť výsledkom testu. V tomto prípade nám posluží Rozšírený Dickey-Fullerov test. O tomto teste si viac povieme pri integrovaných procesoch ARIMA (podkapitola 3.4.3).

3.2 Autokorelačná a parciálna autokorelačná funkcia

Typickým znakom, s ktorým sa stretávame pri časových radoch je silná korelovanosť. Na kvantitatívny popis tohto javu v štatistike využívame autokorelačné funkcie, resp. parciálne autokorelačné funkcie.

Mierou lineárnej závislosti časovo posunutých veličín a_t a a_{t-k} sú koeficienty autokorelácie rezíduí definované vzťahom

$$r_k = \hat{\rho}_k = \frac{\sum_{t=k+1}^T \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^T \hat{a}_t^2}.$$

Graf, v ktorom sú na vodorovnej osi časové posuny a na zvislej osi s rozpätím od -1 do 1 koeficienty autokorelácie rezíduí r_k , sa nazýva reziduálnou autokorelačnou funkciou (ACF). Taktiež sa v tomto grafe nachádzajú 95 % intervaly spoľahlivosti $(-2/\sqrt{T}, 2/\sqrt{T})$. V prípade, že žiadny autokorelačný koeficient r_k nepresiahne tieto hranice, môžeme predpokladať, že nesystematická zložka nie je autokorelovaná.

Pomocou grafu autokorelačnej funkcie rezíduí, dokážeme zistiť sezónnu zložku v rezíduách. V prípade štvrťročných dát sa predpokladá, že nám štvrtý koeficient autokorelácie prekročí hornú hranicu 95 % intervalu spoľahlivosti pre koeficienty autokorelácie rezíduí, a teda bude štatisticky významný.

Grafické posúdenie správania sa autokorelačnej funkcie je pre praktickú časť významným prvkom, ktorý nám napovedá, aký typ modelu zvoliť a jeho počet parametrov. Najdôležitejšie je identifikovať tzv. bod useknutia k_0 . Za touto hodnotou začína byť autokorelačná funkcia nulová.

Autokorelačná funkcia nám poskytuje informácie o sile lineárnej závislosti medzi veličinami y_t a y_{t-k} . Korelácia medzi náhodnými veličinami y_t a y_{t-k} však môže byť zapríčinená ich koreláciou s veličinami y_{t-1} , y_{t-2} , ..., y_{t-k+1} . Avšak parciálna autokorelačná funkcia (PACF) nám dáva informácie o korelácií veličín očistenú o vplyv veličín ležiacich medzi nimi. Parciálnu autokoreláciu s posunom k vyjadruje parciálny regresný koeficient φ_{kk} v autoregresii k -tého rádu

$$y_t = \varphi_{k1}y_{t-1} + \varphi_{k2}y_{t-2} + \dots + \varphi_{kk}y_{t-k} + e_t,$$

kde veličina e_t je nekorelovaná s veličinami y_{t-j} , $j = 1, 2, \dots, k$.

V prípade náhodného časového radu sa hodnoty autokorelačných koeficientov v každom časovom posune približujú nule. Ak máme niekoľko prvých autokorelačných koeficientov vysokých a potom ich hodnoty postupne klesajú k nule s pribúdajúcim posunom, vykazuje nám to prítomnosť trendu.

3.3 Testovanie rezíduí

Pri overovaní modelu sa musíme zaoberať aj rezíduami. V tejto podkapitole si popíšeme testy týkajúce sa podmienok o náhodných zložkách e_t modelu, ktorých platnosť je nevyhnutná pri získaní modelu pomocou metódy najmenších štvorcov. Keďže náhodné zložky nepoznáme, môžeme tieto predpoklady overiť na rezíduách \hat{e}_t .

V prípade krátkodobých radov, a keď nepoznáme pravdepodobnostné rozdelenie náhodných premenných, použijeme neparametrické test náhodnosti premenných. Súčasťou každého testovania je úspešne naformulovanie správnych hypotéz:

- $H_0: y_t = \beta_0 + e_t$, pričom β_0 je konštanta a e_t sú vzájomne lineárne závislé,
- $H_1: \text{non } H_0$, pričom β_0 nie je konštanta, alebo e_t sú vzájomne lineárne závislé.

Prvý spôsob na overenie hypotéz sa využíva znamienkový test sekvencií nad alebo pod mediánom (runs above and below median).

- prvým dôležitým krokom je usporiadanie časového radu podľa veľkosti a následne nájsť medián $y_{0,50}$,
- následne budeme porovnávať hodnoty časového radu s mediánom $(y_t - y_{0,50})$, pričom si všimame znamienka $+$ alebo $-$,

- množstvo plusov alebo mínusov za sebou vytvára sekvencie, tzv. runs alebo obdobia,
- následne si určíme testovaciu štatistiku R pomocou, ktorej testujeme nulovú hypotézu podľa nasledujúceho testovacieho kritéria pre časové rady ($T > 40$):

$$z = \frac{R - \mu_R}{\sigma_R} \sim N(0; 1),$$

kde R – počet sekvencií nad a pod mediánom,

$\mu_R = m + 1$ – očakávaný počet sekvencií,

$\sigma_R = \sqrt{\frac{m(m-1)}{2m-1}}$ – smerodajná odchýlka počtu sekvencií,

pričom $m = \frac{T}{2}$, ak T je párne;

$m = \frac{(T-1)}{2}$, ak T je nepárne,

- nulovú hypotézu zamietneme na hladine významnosti α , ak:

$$|z| > z_{1-\frac{\alpha}{2}},$$

pričom rad obsahuje trend alebo e_t sú vzájomne závislé.

Druhým dôležitým testom pri tejto problematike je test náhodnosti založený na bodoch obratu (runs up and down or turning points test od randomness).

Postup testovania je nasledovný:

- na začiatok si vytvoríme rad prvých absolútnych diferencií $(y_t - y_{t-1})$ pre $t = 2, 3, \dots, T$,
- následne si diferencie označíme znamienkami $+$ ($y_t > y_{t-1}$) a $-$ ($y_t < y_{t-1}$),
- sledujeme počet bodov obratov U , t. j. časový okamih keď je znamienko pri diferencii iné ako v nasledujúcom okamihu,
- vypočítame si testovaciu charakteristiku vhodnú pre časové rady ($T > 20$):

$$z = \frac{U - \mu_U}{\sigma_U} \sim N(0; 1),$$

kde U – počet bodov obratu,

$\mu_U = \frac{2(T-2)}{3}$ – očakávaný počet bodov obratu,

$\sigma_U = \sqrt{\frac{16T-29}{90}}$ – smerodajná odchýlka počtu bodov obratu,

- nulovú hypotézu zamietneme na hladine významnosti α , ak:

$$|z| > z_{1-\frac{\alpha}{2}}$$

pričom je prítomný trend alebo e_t sú vzájomne závislé.

„Zistenie, že rezíduá sú striedavo v kladných alebo záporných zoskupeniach znamená, že ak je rezíduum v čase t kladné, očakáva sa, že aj rezíduum v čase $t + 1$ bude kladné a naopak. Tieto cykly indikujú možnú pozitívnu autokoreláciu medzi susednými rezíduami.“ (Rublíková, 2007, s. 49)

Graficky na základe sledovania rezíduí v čase môžeme predvídať, či sa tu nachádza autokorelácia alebo nie. Presnejšie výsledky získame pomocou Durbin-Watsonovej štatistiky, pričom testujeme hypotézy:

- H_0 : rady e_t a e_{t-1} sú lineárne nezávislé, resp. $\rho_1(e) = 0$,
- H_1 : rady e_t a e_{t-1} sú pozitívne autokorelované, resp. $\rho_1(e) > 0$,

Durbin-Watsonovu testovaciu štatistiku vieme zapísať:

$$D-W = \frac{\sum_{t=2}^T (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^T \hat{e}_t^2} \in (0; 4).$$

O zamietnutí alebo prijatí nulovej hypotézy sa rozhodujeme na základe empirických pravidiel:

- ak $D-W = 2$, náhodné zložky majú náhodný charakter, a teda nie sú autokorelované,
- ak $D-W < 2$, náhodné zložky vykazujú systematický pohyb v čase, ide o pozitívnu autokorelovanosť, a preto je potrebné tento pohyb doplniť do modelu,
- ak $D-W > 2$, model nie je vhodný, keďže sú e_t negatívne autokorelované (v prípade ak $D-W \rightarrow 4$, ide o silnú negatívnu autokoreláciu).

Ďalším užitočným testom je Box-Pierceov test rezíduí, ktorý na hladine významnosti α testuje skupinu prvých m koeficientov autokorelácie a ich rovnosť nule. Hypotézy pre tento test vieme zapísať:

- $H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$,
- $H_1: \text{non } H_0$.

Na testovanie sa používa testovacia štatistika Q_m :

$$Q_m = T \sum_{k=1}^m r_k^2 \approx \chi^2(m).$$

Ak $Q_m = T \sum_{k=1}^m r_k^2 > \chi_{1-\alpha}^2(m)$, tak zamietame nulovú hypotézu na hladine významnosti α ,

kde $\chi_{1-\alpha}^2(m)$ je $(1 - \alpha)$ % percentil chi-kvadrát rozdelenia.

Ďalším z problémom pri analýze časových radov so sezónnou zložkou je heteroskedasticita pozorovaní. Heteroskedasticita spôsobuje rastúcu variabilitu pozorovaní v čase. Hlavným dôvodom je, že v prípade prítomnosti rastúceho trendu v časovom rade má sezónna zložka tendenciu taktiež rásť a naopak. Na posúdenie homoskedasticity, teda konštantnosti rozptylu rezíduí v čase, môžeme využiť grafické vyobrazenie alebo rôzne testy.

Graficky môžeme porovnávať rezíduá s časom alebo s odhadnutými, resp. vyrovnanými hodnotami. Typickým znakom homoskedastických rezíduí je podobnosť náhodným rezíduám, t. j. rezíduá kolíšu v určitom pásme okolo nuly s hranicami \pm smerodajná odchýlka rezíduí. Z grafu však nie je vždy jasné, či je heteroskedasticita štatisticky významná, a preto je adekvátne použiť aj parametrický test homoskedasticity rezíduí. Postup spočíva v tom, že sa rad rezíduí rozdelí na dve rovnako dlhé súbory, pričom sa predpokladá rovnosť rozptylov rezíduí prvého a druhého súboru:

- $H_0: \sigma_1^2 = \sigma_2^2$,
- $H_1: \sigma_1^2 \neq \sigma_2^2$.

Pri testovaní sa využíva nasledujúca štatistika:

$$F = \frac{\text{rozptyl rezíduí}_1}{\text{rozptyl rezíduí}_2} \approx F\left(\frac{T}{2} - 1; \frac{T}{2} - 1\right).$$

H_0 sa zamietá, ak $F > F_{1-\frac{\alpha}{2}}$ alebo $F < F_{\frac{\alpha}{2}}$, pričom $F_{1-\frac{\alpha}{2}}$ a $F_{\frac{\alpha}{2}}$ sú kvantily Fisherovho rozdelenia pravdepodobnosti s $\left(\frac{T}{2} - 1; \frac{T}{2} - 1\right)$ stupňami voľnosti.

V prípade potvrdenia heteroskedasticity rezíduí, musíme transformovať hodnoty časového radu, pomocou ktorej dosiahneme stabilitu variability pozorovaní. Tento fakt nemôžeme ignorovať, pretože by bolo znemožnené využitie metódy najmenších štvorcov. Pri heteroskedasticite časového radu variabilita náhodných premenných závisí od času, a preto môžeme tvrdiť, že sa variabilita rezíduí mení s rôznymi hodnotami \hat{y}_t . Medzi najčastejšie transformácie patria:

- $\sqrt{y_t}$, ak sa smerodajná odchýlka rezíduí mení proporcionálne s $\sqrt{\hat{y}_t}$, t. j. $\hat{\sigma}_t = k\sqrt{\hat{y}_t}$,
- $\log(y_t)$, ak smerodajná odchýlka rezíduí rastie proporcionálne s \hat{y}_t , t. j. $\hat{\sigma}_t = k\hat{y}_t$,
- $\frac{1}{y_t}$, ak sa smerodajná odchýlka rezíduí mení úmerne s druhou mocninou odhadov \hat{y}_t^2 , t. j. $\hat{\sigma}_t = k(\hat{y}_t)^2$.

Pri použití regresných modelov s konštantným členom β_0 , nesmieme zabudnúť overiť predpoklad o normalite náhodných premenných, resp. že majú normálne pravdepodobnostné rozdelenie s nulovou strednou hodnotou a konštantným rozptylom. Na základe rezíduí modelu vieme tento predpoklad overiť:

- graficky,
- chí-kvadrát testom dobrej zhody,
- neparametrickými testami: Kolmogotovo-Smirnovov test, Shapirov-Wilksov test,
- normovanou šikmosťou a špicatosťou,
- Jarqueovo-Beraovym testom, a pod.

Od dlhých časových radov sa očakáva, že vytvorený histogram rezíduí bude mať hodnoty sústredené okolo strednej hodnoty nula, a je viditeľná podobnosť s histogramom normálneho rozdelenia pravdepodobnosti. Taktiež sa využívajú empirické pravidlá:

- 99,7 % štandardizovaných rezíduí je z intervalu +3 až -3,
- 95 % štandardizovaných rezíduí je z intervalu +2 až -2,
- 68 % štandardizovaných rezíduí je z intervalu +1 až -1.

Grafické vyhodnotenie môžeme získať pomocou škatuľového grafu Box-whisker plot. Normálne rozdelenie je symetrické, keď modus = medián = priemer. Od rezíduí sa očakáva približná podobnosť s normálnym rozdelením, ak sú ich priemer a medián skoro

rovnaké, a preto sa v škatuľkovej grafe priemer(krížik) bude takmer zhodovať s mediánom(kolmá čiara).

Medzi najznámejší test normality patrí chí-kvadrát test dobrej zhody. Bohužiaľ hlavnou nevýhodou tohto testu je, že výsledky závisia od rozsahu súboru, pričom test možno použiť len pre dlhé časové rady s dĺžkou okolo sto pozorovaní. Preto sa odporúča využiť neparametrické testy ako je napr. Kolmogorovov-Smirnovov test.

Pri testovaní sa využívajú hypotézy:

- H_0 : y_t má normálne rozdelenie,
- H_1 : y_t nemá normálne rozdelenie.

„Kolmogorovov-Smirnovov test vychádza z predpokladu, že distribučná funkcia náhodnej premennej y z výberového súboru rozsahu $n - F_n(y)$ má podobné vlastnosti ako distribučná funkcia náhodnej premennej Y v základnom súbore pri platnosti nulovej hypotézy – $F_0(y)$.“(Rublíková, 2007, s. 65)

Ak časový rad má normálne rozdelenie, platí že rozdiely medzi $F_n(y)$ a $F_0(y)$ budú minimálne. Na overenie využívame testovaciu štatistiku:

$$\max[F_0(y) - F_n(y)],$$

pričom

$$F_n(y) = \frac{\text{počet údajov v súbore} \leq y}{n}.$$

Test sa skladá z nasledujúcich krokov:

- usporiadanie hodnôt výberového súboru podľa veľkosti,
- vypočítanie normovanej hodnoty premennej $z = \frac{y - \mu_y}{\sigma_y}$, kde μ_y a σ_y sú hodnoty pre platnosť H_0 ,
- určenie distribučnej funkcie $F_0(y)$ a $F_n(y) = \Phi(z)$,
- zistenie hodnoty testovacej štatistiky: $\max[F_0(y) - F_n(y)]$,
- porovnaním hladiny významnosti α s p-value rozhodneme o prijatí nulovej hypotézy, ak $\alpha < p\text{-value}$ a naopak zamietneme, ak $\alpha > p\text{-value}$.

3.4 Box–Jenkinsové modely

Vďaka nástupu výpočtovej techniky koncom 20. storočia, štatistickí mohli začať uplatňovať lineárne stochastické modely inak nazývané ako Boxova-Jenkinsovova metodológia, ktorá sa v súčasnosti používa ako nový smer dynamického modelovania ekonomických javov vo viacerých vyspelých krajinách. Taktiež by sme túto metódu mohli zaradiť k matematicky náročnejším metódam, a preto si v nasledujúcich podkapitolách popíšeme celý proces.

„Boxova-Jenkinsova metodológia ARIMA modelov (pre časové rady bez sezónnosti) alebo SARIMA modelov (pre časové rady so sezónnou zložkou) slúži na modelovanie stochastického vývoja časového radu, t. j. na modelovanie náhodnej zložky časového radu (po vylúčení trendu a sezónnosti z časového radu)“ (Rublíková, 2007, s. 195)

Pomocou týchto modelov teda vieme spracovať časové rady s navzájom závislými pozorovaniami, čo je ťažisko ich postupov. Nevyhnutnou podmienkou využitia týchto metód je stacionarita časového radu, pričom túto vlastnosť sme si popísali v podkapitole 3.1. Box-Jenkinsova metodológia vytvára čo najpresnejšie predpovedné modely pre každý súbor údajov. Tieto modely nám poskytujú viac usporiadanejší postup budovania, analyzovania a prognózovania modelov časových radov.

Táto metodika využíva najnovšie pozorovania časového radu ako predvolené hodnoty, a následne analyzuje chyby prognózy v minulosti, aby správne nastavila budúce obdobie.

Medzi jednu z hlavných výhod Box-Jenkinsovej metodológie považujeme význačnú flexibilitu v analyzovaní časového radu. Taktiež sa rýchlo prispôbuje zmenám a výborne funguje v rizikových analýzach a analýzach neistoty. Dokáže analyzovať rady, pri ktorých klasická analýza nie je dostačujúca. Za výhodu považujeme aj systematickosť danej metodológie, keďže sa postup skladá z viacerých krokov, ktoré majú vopred daný kľúč.

Aj táto metóda má však nevýhody, pričom hlavnou je zložitosť výpočtu, keďže zahŕňa niekoľko iteračných postupov. V dôsledku toho neexistuje žiadny automatizovaný proces pre modelovanie. V určitých prípadoch môže byť značnou nevýhodou aj minimálna potrebná dĺžka časového radu 50 pozorovaní.

Box-Jenkinsova metodológia využíva na prognózy autoregresné modely a kľzavé priemery. Podľa toho, akú techniku model používa, poznáme modely AR (autoregresný model), MA (kľzavé priemery), ARMA (autoregresný model a kľzavé priemery), ARIMA (integrovateľný zmiešaný model), ktoré môžu byť navyše rôzneho rádu. Táto možnosť nám prináša značnú flexibilitu, ale na druhú stranu vyžaduje veľa analytickej subjektivity, ku ktorej potrebujeme subjektívne rozhodovanie a väčšie skúsenosti.

Praktická aplikácia je teda časovo aj finančne náročnejšia, kvôli potrebe štatistického softvéru, a tak strácame príležitosť jednoduchšej interpretácie modelu.

3.4.1 Autoregresný proces AR

Najprirodzenejšou formuláciou modelu časových radov je lineárna regresia z minulých pozorovaní, t. j. regresia samotného časového radu. Toto vytvorilo termín autoregresia. V praxi sa tieto modely ukazujú ako veľmi dôležité a sú najobľúbenejšie.

Autoregresný proces funguje na princípe modelovania vývoja časového radu pomocou časovo oneskorených hodnôt. Teda pracuje so závislosťou medzi pôvodnými hodnotami časového radu a oneskorenými hodnotami. Tento proces nám pomáha modelovať lineárnu závislosť hodnôt časového radu y_t od viacerých (p) oneskorených hodnôt, a preto ho môžeme nazvať autoregresný proces p -teho stupňa.

Ak $E(y_t) = \mu \neq 0$, potom má model AR(p) tvar:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + e_t,$$

kde $\phi_0 = (\mu - \phi_1 \mu - \dots - \phi_p \mu) = \mu(1 - \sum_{j=1}^p \phi_j)$.

Ak $E(y_t) = 0$, model má jednoduchý tvar:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + e_t.$$

Pomocou spätného operátora B proces môžeme zapísať:

$$(1 - \phi_1 B - \dots - \phi_p B^p) y_t = e_t.$$

Aby sme mohli použiť model AR musí byť splnená podmienka stacionarity, t. j. proces neobsahuje systematickú zložku a vieme ho zapísať ako lineárnu kombináciu nezávislých rovnako rozdelených náhodných premenných. Podmienka je splnená ak

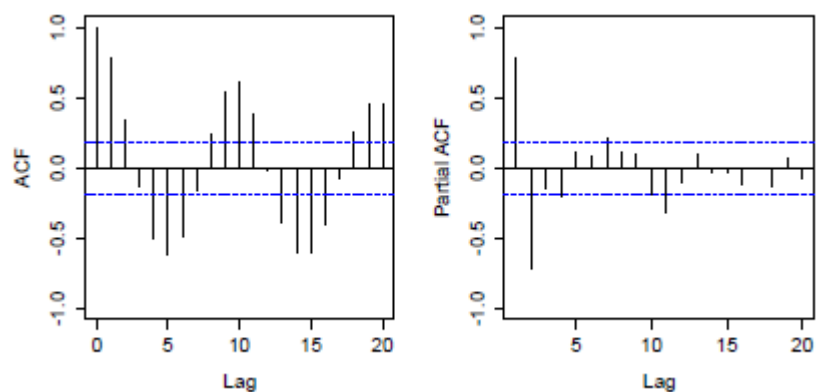
korene polynomickej rovnice $(1 - \phi_1 B - \dots - \phi_p B^p)y_t = 0$ ležia mimo jednotkového kruhu, resp. ak platí $\phi_1 + \phi_2 + \dots + \phi_p < 1$. Špeciálne pri modeli AR(1) nastáva situácia, keď $\rho_1 = \phi_1$, $-1 \leq \rho_1 \leq 1$. Pri modeli AR(2) máme musia platiť podmienky $\phi_1 + \phi_2 < 1$; $\phi_2 - \phi_1 < 1$; $-1 \leq \phi_2 \leq 1$. Platí, že AR(p) proces je vždy invertibilný.

Prispôsobenie procesu AR(p) údajom časového radu zahŕňa tri hlavné kroky. Najprv je potrebné identifikovať model a jeho poradie. Následne pristúpime k odhadu parametrov a na koniec sa overí kvalita odhadnutého modelu pomocou analýzy rezíduí, ktorú sme popísali podkapitole 3.3.

Prvý krok identifikácia modelu najprv vyžaduje overenie, či údaje ukazujú vlastnosti, vďaka ktorým je pravdepodobné, že boli vygenerované z AR(p). Najmä časové rady, ktoré sa snažíme modelovať, musia byť stacionárne. Potrebu procesu AR(p) nám naznačuje graf ACF, pri ktorom dochádza k približne exponenciálnemu vývoju v sínusoidách a graf PACF s rozpoznateľným skokom medzi koeficientmi pri určitom oneskorení p menšom ako 5 až 10.

Ak dôjde k silnému porušeniu ktorejkoľvek z týchto troch nedostatkov, je nepravdepodobné, že AR(p) poskytne uspokojivý výsledok a môžu existovať modely, ktoré sú pre daný problém vhodnejšie. Výber poradia modelu p sa potom spolieha na analýzu grafu PACF. Sledovaním jednotlivých koeficientov by sme najskôr vyskúšali najjednoduchší model, ktorý sa zdá byť hodnoverný. To znamená výber najmenšieho p, pri ktorom sa predpokladá prerušenie, t. j. už spomínaný skok, po ktorom nenasleduje žiadna alebo iba malo a slabo významná čiastočná autokorelácia. Pre lepšie pochopenie si popíšeme nasledujúce grafy na obrázku č. 1.

Obrázok č. 1 – Teoretické grafy ACF a PACF



Zdroj: Dettling, 2016

Z ľavého grafu ACF pozorujeme cyklické správanie, ktoré má exponenciálny vývoj s pomalým rozpadom. Na základe pravého grafu PACF nájdeme hodnotu, kde dochádza k danému skoku medzi hodnotami jednotlivých koeficientov autokorelácie, a teda potreby procesu AR(p). Z grafu to nie je až také jasné, pretože máme viacero možností p (2, 4, 7 alebo 11). Výber výšky parametra je diskutabilný, ale odporúča sa zvoliť najmenšieho kandidáta, a teda AR(2). Celkovo sa pri tvorbe modelu snažíme o čo najnižšiu výšku parametrov, aby model nebol príliš zložitý.

3.4.2 Proces kľzavých priemerov MA

V tejto časti si popíšeme modely kľzavých priemerov. Tieto je možné vnímať ako rozšírenie procesu bieleho šumu. Ako uvidíme, vedie to k procesu časových radov, ktorý je stacionárny. Ďalej uvidíme, že v mnohých ohľadoch sú modely s kľzavým priemerom komplementárne s autoregresívnymi modelmi. Niekedy je vhodnejšie namiesto vysokého stupňa AR procesu použiť model kľzavých priemerov q-teho stupňa MA(q).

Modely kľzavých priemerov MA(q) sa úspešne uplatňujú v mnohých aplikovaných oblastiach, najmä v ekonometrii. Samotné časové rady ekonomických ukazovateľov sú ovplyvnené rôznymi náhodnými udalosťami, ako sú štrajky, rozhodnutie vlády, referendá, nedostatok kľúčových komodít, atď.

Takéto udalosti budú mať nielen okamžitý účinok, ale môžu tiež ovplyvniť hodnotu (v menšej miere) vierohodnosť po sebe nasledujúcich období. Pretože podstatou myšlienky kľzavých priemerov je popísať vývoj hodnôt časového radu pomocou časového radu náhodných šokov. Je preto pravdepodobné, že sa kľzavé priemerné procesy objavujú v praxi. Niektoré z ich teoretických vlastností sú navyše komplementárne s procesmi AR.

Všeobecný zápis tohto procesu teda vyjadruje lineárnu závislosť hodnôt časového radu y_t od q oneskorených hodnôt náhodných šokov e_t . Pričom proces kľzavých priemerov vieme následne zapísať v tvare (ak $E(y_t) = \mu \neq 0$):

$$y_t = \theta_0 + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q},$$

kde $\theta_0 = \mu$.

Ak $E(y_t) = \mu = 0$, potom má model jednoduchší tvar:

$$y_t = e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}.$$

Pomocou spätného operátora B model $MA(q)$ vieme zapísať:

$$y_t = \theta_q(B)e_t.$$

Model MA je finálnou formou lineárneho procesu, takže všetky modely MA sú stacionárne. Avšak tento proces pre úspešnú interpretáciu musí byť aj invertibilný. Podmienku invertibility proces MA spĺňa, ak ho dokážeme vyjadriť vo forme reprezentácie $AR(\infty)$, ktorá konverguje. To znamená, že korene polynomickej rovnice $(1 - \theta_1 B - \dots - \theta_p B^p) = 0$ ležia mimo jednotkového kruhu, t. j. ak platí $\theta_1 + \theta_2 + \dots + \theta_q < 1$. Musíme však brať do úvahy, že nie každý invertibilný proces je stacionárny. A taktiež každý stacionárny proces nemusí byť invertibilným. Podmienka inveribility je ale významná hlavne pre konštrukciu predpovedí.

3.4.3 Integrované procesy $ARIMA$

$ARIMA$ modely sa zameriavajú na popis časových radov s existujúcim deterministickým trendom, ktorý môže byť odstránený diferencovaním, a kde môžeme tieto rozdiely popísať pomocou modelu $ARMA(p,q)$. Nasledujúca tabuľka nám popisuje správanie ACF a $PACF$ pri procesoch $AR(p)$, $MA(q)$ a $ARMA(p,q)$.

Tabuľka č. 1 – Správanie ACF a $PACF$

	autokorelačná funkcia	parciálna autokorelačná funkcia
$AR(p)$	nepreruší sa	preruší sa od p
$MA(q)$	preruší sa od q	nepreruší sa
$ARMA(p,q)$	nepreruší sa	nepreruší sa

Zdroj: Rublíková, 2008

Pod pojmom „preruší sa“ rozumieme situáciu, keď sa daná funkcia po určitom počte posunov vynuluje. Pri sledovaní autokorelačnej a parciálnej autokorelačnej funkcie v prípade AR a MA procesov môže nastať situácia, kedy sa ani jedna funkcia nepreruší. Ale stáva sa to iba vtedy keď máme $MA(\infty)$ proces, ktorý môže byť invertibilný na $AR(\infty)$. Bohužiaľ sa tieto procesy nemôžu použiť na empirické modelovanie, pretože sú charakterizované nekonečne veľa parametrami. Ale držíme sa pravidla, že v aplikačných časových radoch by sme mali mať počet odhadovaných parametrov čo najmenší.

Pri vyššie popísaných procesoch sme uvažovali, že pracujeme so stacionárnym časovým radom. V praxi sa ale často stretávame s tým, že reálne dáta nedokážu splniť

požiadavku nekorelovanosti chyby prognóz. V prípade, že Durbin-Watsonova štatistika poukazuje na prítomnú autokoreláciu, následná analýza rozptylu a regresia bude neplatná a taktiež zavádzajúca. Preto sa používa model, ktorý vie pracovať aj s takýmito časovými radmi, a to autoregresný integrovaný model kľzavých priemerov ARIMA.

Modelovanie pomocou ARIMA modelov je sofistikovaná a komplexná metodika, ktorá nekladie dôraz na konštrukciu jednorovnicového modelu, ale viacrovnicového. Oproti regresným modelom, kde je závislá premenná Y vysvetľovaná niekoľkými regresormi, je pri ARIMA metóde závislá premenná Y vysvetľovaná dátami tej istej premennej Y v minulosti spolu s chybovými členmi v súčasných alebo minulých pozorovaniach.

ARIMA modely môžeme považovať aj za tzv. záchranné modely v prípade, že naše údaje, ktoré sme chceli použiť na lineárne alebo logistické regresné modelovanie, obsahujú autokoreláciu. Najčastejšie sa tieto modely používajú v marketingových a sociálnych predpovediach, pričom sa skladajú z nasledujúcich členov:

- AR – autoregresný člen vyjadruje vzťah k predošlým pozorovaniam časového radu,
- MA – člen kľzavých priemerov v modeli využíva hodnoty chýb predošlých predpovedí,
- I – integračný člen nám popisuje výšku diferenciacie potrebnej na odstránenie nestacionarity časového radu.

Pri tvorbe modelov ARIMA je dôležité určiť rád integrovania resp. stupeň diferencie. Keďže ekonomické časové rady nie sú často stacionárne, najčastejšie sa používa prvý alebo druhý stupeň diferencovania. V prípade skúseného analytika môžeme použiť na odhad stupňa diferencovania jednoduchú grafickú analýzu časového radu. V prípade viditeľného stochastického lineárneho trendu použijeme prvú diferenciu, a pri odstránení stochastického kvadratického trendu musíme použiť druhú diferenciu. Zvyčajne sa porovnáva variabilita nediferencovaného pôvodného časového radu s hodnotami po diferenciacii.

Druhou možnosťou je posúdenie výberovej autokorelačnej funkcie a výberovej parciálnej autokorelačnej funkcie. Diferenciácia je potrebná ak sa:

- prvý koeficient VACF približuje k jednej a zvyšné koeficienty klesajú veľmi pomaly lineárnym trendom,

- a súčasne VPACF má prvú hodnotu blízku k jednej, no ostatné hodnoty klesajú exponenciálne v sínusových vlnách k nule.

Takýto časový rad je nestacionárny a najvhodnejší stupeň diferencovania volíme na základe najnižšej smerodajnej odchýlky.

Musíme si dať pozor na to, aby sme rad príliš nepoddiferencovali alebo neprediferencovali. Preto okrem subjektívneho posúdenia grafov použijeme testy jednotkového koreňa. Medzi najpoužívanejší test zaradíme Dickeyov-Fullerov test (DF test). Overuje sa hypotéza, že odhad parametra ϕ_1 z rovnice procesu AR(1):

$$y_t = \phi_1 y_{t-1} + e_t,$$

sa rovná jednej, t. j. existuje jednotkový koreň. Na testovanie sa využíva odhad parametrov regresnej rovnice:

$$\Delta y_t = (\phi_1 - 1)y_{t-1} + e_t = \delta y_{t-1} + e_t.$$

„Dickeyov-Fullerov test po odhade parametra δ a výpočte Studentovej štatistiky t_δ , porovnáva túto hodnotu s tabelovanou τ hodnotou. Nižšia záporná hodnota štatistiky t_δ znamená zamietnutie nulovej hypotézy, a teda stacionaritu radu. Pri nezamietnutí nulovej hypotézy o nestacionarite sa môže proces testovania opakovať s modelmi vyšších diferencií, až kým nie je nulová hypotéza zamietnutá. Podľa stupňa diferenciácie premennej pri zamietnutej nulovej hypotéze o jej parametri sa potom určí stupeň integrácie časového radu.“(Lukáčik a Pekár, 2006)

Pri tomto teste však kritické hodnoty závisia od typu modelu. Preto existujú tabelované kritické hodnoty pre model s absolútnym členom aj pre model s absolútnym členom a lineárnym trendom. Ak nastane situácia, že je náhodná zložka autokorelovaná, DF test prestáva fungovať. Preto bol model rozšírený z prvého stupňa na všeobecný proces p-teho stupňa, ktorý odstráni autokoreláciu náhodnej zložky.

Následne má rozšírený model tvar:

$$\Delta y_t = \delta y_{t-1} + \sum_{j=2}^p \phi_{j-1} \Delta y_{t-j+1} + e_t,$$

rozšírený model, ak obsahuje absolútny člen:

$$\Delta y_t = \mu + \delta y_{t-1} + \sum_{j=2}^p \phi_{j-1} \Delta y_{t-j+1} + e_t,$$

a model, ktorý obsahuje navyše aj lineárny časový trend:

$$\Delta y_t = \mu + \alpha_1 t + \delta y_{t-1} + \sum_{j=2}^p \phi_{j-1} \Delta y_{t-j+1} + e_t .$$

V praxi sa tento vylepšený test volá rozšírený Dickeyov-Fullerov test (augmented DF test; ADF). Na testovanie použijeme pôvodné tabelované kritické hodnoty a na voľbu optimálnej výšky autoregresného stupňa p nám posluži Akaikeho alebo Schwarzovho informačné kritérium (SC).

Postup ADF testu pozostáva s nasledujúcich siedmych krokov:

1. ako prvé odhadneme metódou najmenších štvorcov najvšeobecnejší model,
2. pre testovanie prítomnosti trendu použijeme hypotézu $\alpha_1=0$, ak $\delta=0$, v prípade potvrdenia tejto hypotézy, časový rad obsahuje deterministický trend a ideme na krok 4,
3. pri zamietnutí hypotézy vykonáme test o parametri δ , že nadobudne hodnotu nula – nevýznamnosť tohto parametra potvrdzuje, že časový rad je proces náhodnej prechádzky s trendom,
4. následne testujeme model s absolútnym členom, ale bez trendu,
5. pre testovanie prítomnosti posunu v tomto modeli využívame podmienenú hypotézu $\mu=0$, ak $\delta=0$ – prijatie hypotézy potvrdzuje, že rad je proces náhodnej prechádzky bez posunu a trendu.
6. v prípade zamietnutia hypotézy o nulovom parametri μ , testujeme parameter δ – nevýznamnosť δ napovedá, že ide o proces náhodnej prechádzky s posunom.
7. ak je δ významný v kroku tri alebo šesť, procedúra sa zastaví, pretože rad už je stacionárny, pričom pre zistenie trendu alebo posunu sa použije klasický t-test. (Lukáčik, 2006)

Diferencovaním odstránime nestacionaritu v strednej hodnote, ale v praxi sa môžeme stretnúť s prípadmi, keď ju musíme odstrániť v rozptyle. Pri takejto situácii musíme uvažovať, že procesy majú nelineárny charakter a lineárne modely ARIMA nebudú teda vhodné. Tu nám pomôže linearizujúca transformácia, ktorá musí byť vykonaná ešte pred diferencovaním. Linearizovanie logaritmicou transformáciou ($\log(y_t)$) zabezpečí náhodným zmenám stredných hodnôt podobu bieleho šumu, s normálnym rozdelením a konštantným rozptylom.

3.4.4 Sezónne procesy SARIMA

Po zoznámení sa s modelmi ARIMA je úplne prirodzené popísať si informácie o sezónnosti časových radov, keďže pomocou diferencií vieme vplyvy sezónnosti odstrániť. „Myšlienka sezónneho procesu je nasledujúca: ako v prípade procesu ARIMA predpokladáme vzájomnú závislosť medzi veličinami ... $y_{t-3}, y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2}, y_{t+3}, \dots$. Pretože tento proces obsahuje ešte sezónne kolísanie, môžeme očakávať i závislosť medzi navzájom zodpovedajúcimi veličinami v jednotlivých sezónach, t. j. medzi veličinami ... $y_{t-2s}, y_{t-1s}, y_{ts}, y_{t+1s}, y_{t+2s}, \dots$, kde s je dĺžka sezónnej periódy (pri mesačných časových radoch je to 12, pri štvrtročných 4 atď.).“ (Rublíková, 2003, s. 134)

Proces SAR(P)_s má nasledujúci vzťah:

$$y_t = \Phi_1 y_{t-s} + \Phi_2 y_{t-2s} + \dots + \Phi_p y_{t-ps} + e_t.$$

Pomocou spätného operátora ho vieme zapísať:

$$(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}) y_t = e_t,$$

resp.

$$\Phi_p(B^s) y_t = e_t.$$

Stabilitu procesu zabezpečuje rovnica $\Phi_p(B^s) y_t = 0$, t. j. korene rovnice ležia mimo jednotkového kruhu. Stacionaritu procesu dosiahneme, ak je možné ho prepísať do sezónneho lineárneho procesu:

$$Y_t = \Psi(B^s) e_t.$$

Na základe zistených vedomostí vieme popísať proces SMA(Q)_s:

$$y_t = e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s} - \dots - \Theta_Q e_{t-Qs},$$

a pomocou spätného operátora:

$$y_t = (1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs}) e_t = \Theta_Q(B^s) e_t.$$

Obdobne ako pri procese MA(q) aj SMA(Q) je vždy stacionárny. Pri tomto procese bude autokorelačná funkcia nenulové hodnoty nadobúdať len v sezónnych posunoch $s, 2s,$

$3s, \dots, Qs$. Podmienku invertibility spĺňa, ak korene rovnice $\Theta_Q(B^S) = 0$ ležia mimo jednotkového kruhu.

Zmiešaný proces SARMA(p,q)(P,Q)_s vieme zapísať matematicky pomocou spätného operátora B :

$$\phi_p(B)\Phi_P(B^S)y_t = \theta_q(B)\Theta_Q(B^S)e_t,$$

kde

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) - \text{nesezónny operátor procesu AR}(p),$$

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^q) - \text{nesezónny operátor procesu MA}(q),$$

$$\Phi_P(B^S) = (1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P B^{PS}) - \text{sezónny operátor procesu AR}(P),$$

$$\Theta_Q(B^S) = (1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS}) - \text{sezónny operátor procesu MA}(Q),$$

Stacionarita nastáva, ak jeho rovnica autoregresných procesov $\phi_p(B)\Phi_P(B^S) = 0$ má korene mimo jednotkového kruhu a invertibilitu dosiahneme, ak má rovnica $\theta_q(B)\Theta_Q(B^S) = 0$ korene mimo jednotkového kruhu.

V prípade sezónneho časového radu musíme sledovať splnenie podmienky stacionarity aj z pohľadu sezónnosti. Problém nastáva pri stacionarizácii sezónnej úrovne, pretože dosiahnutie stacionarizácie v celkovej úrovni nám nemusí automaticky stacionarizovať časový rad hodnôt aj v sezónnej úrovni.

Sezónnu diferenciu časového radu y_t s periódou dĺžky s vyjadríme:

$$\nabla^s y_t = y_t - y_{t-s} = (1 - B^s)y_t.$$

Následne SARIMA(p, d, q)(P, D, Q)_s s konštantou vieme zapísať v tvare:

$$\phi_p(B)\Phi_P(B^S)\nabla^d \nabla_s^D y_t = \theta_q(B)\Theta_Q(B^S)e_t.$$

Zistiť potrebu sezónnej diferencie, resp. či budeme musieť použiť aj diferenciu nesezónnu, je nevyhnutným a veľmi dôležitým prvým krokom pri sezónnych časových radoch. Zvyčajne sa nepoužíva druhá alebo vyššia sezónna diferencia, a taktiež dokopy sa v modeli nepoužívajú viac ako dve diferencie spolu. Ak pri grafickej analýze zistíme, že koeficienty autokorelácie so sezónnym oneskorením s postupne klesajú je potrebné použiť

sezónnu diferenciu. Pri použití sezónnej diferenciacie stupňa D sa časový rad skraca o $D*s$ pozorovaní. V prípade, že použijeme aj nesezónnu diferenciaciu, následne si časový rad skrátime o $d + D*s$ hodnôt.

Na základe získaných vedomostí z predošlých časti práce, dokážeme zapísať model SARIMA(p,d,q)(P,D,Q) $_s$ pomocou spätného operátora B nasledujúcim vzťahom:

$$\Phi_p(B^s)\phi_p(B)(1-B)^d(1-B^s)^D y_t = \theta_q(B)\Theta_Q(B^s)e_t.$$

Tento model môžeme chápať ako rozšírený, resp. špeciálny typ modelu ARIMA. Preto sú všetky postupy identifikácie, odhadovania parametrov, predpovedania a kontroly obdobné ako pri modeloch ARIMA.

3.5 Fázy tvorby modelu

Tvorba modelov pomocou Box-Jenkinsových modelov pozostáva zo štyroch základných krokov. Prvým a najdôležitejším krokom je identifikácia. Pri tejto časti pozorne preskúmame graf časového radu. Niektoré javy nemožno pozorovať hneď na prvý pohľad, ako napr. prítomnosť trendu. Ale v tejto počiatočnej fáze ide hlavne o subjektívne posúdenie situácie. Nevyhnutnou podmienkou použitia tejto metodológie je stacionarita časového radu. Preto už v tomto kroku vieme zistiť potrebu rôznych transformácií.

V prípade nestacionarity v rozptyle je vhodné použiť linearizáciu časového radu, teda logaritmickú transformáciu. Nesmieme zabudnúť, že je vhodné použiť túto techniku ešte pred samotným diferencovaním. Hlavným dôvodom je, že použitím diferenciacie môžeme dostať aj záporné hodnoty.

Na základe grafickej analýzy ACF a PACF pre stacionárny časový rad uskutočňujeme výber vhodných parametrov do výsledného modelu. Teoretické tvary pre rôzne stupne procesov sme si popísali v podkapitole 3.4. Nevýhodou je ale fakt, že rozhodnutie závisí na skúsenosti a cite analytika.

Postupným pridaním parametrov do modelu zvyšujeme jeho presnosť, no strácame zrozumiteľnosť. Preto je dôležité nájsť zlatú strednú cestu medzi presnosťou a jednoduchosťou modelu. Následne môžeme pristúpiť k samotnému odhadu modelu.

V tejto fáze si odhadneme pomocou softvéru všetky predpokladané modely, ktoré berieme do úvahy na základe poznatkov z prvej fázy, a to identifikácie. Na odhad

parametrov sa zvyčajne využíva metóda najmenších štvorcov. V praxi sa môžeme stretnúť aj s využitím metódy momentov alebo metódy maximálnej vierohodnosti. Každopádne to závisí hlavne od softvéru, ktorý analytik využíva.

Pri praktických analýzach sa často stáva, že postupným kombinovaním parametrov sa nevieme rozhodnúť medzi veľkým množstvom modelov. V tomto prípade si musíme dať pozor nato, aby sme modely nepreparametrizovali. Aby sme takejto situácii predišli, môžeme použiť tzv. penalizačné miery kvality modelu, ako sú napr. Akaikeovo informačné kritérium (AIC) a Swarz-Bayesovo kritérium (SBC).

Tieto doplnkové kritéria sú založene na porovnávaní rezíduí všetkých dostupných modelov pomocou sumárnych štatistík. Pričom tvary týchto kritérií sú:

$$AIC = \log \hat{\sigma}_{p,k}^2 + \frac{2(p+q)}{n},$$
$$SBC = \log \hat{\sigma}_{p,q}^2 + \frac{(p+q) \log n}{n},$$

kde

$\hat{\sigma}_{p,q}^2$ – odhadnutý rozptyl bieleho šumu procesu ARMA(p,q),

n – dĺžka časového radu.

V literatúre sa často môžeme stretnúť s iným označením ako SBC, napr. BIC. Pri porovnávaní modelov vyberiem ten, ktorý má hodnoty týchto kritérií najnižšie. Otázne je, ale podľa ktorého kritéria sa rozhodnúť. V praxi sa častejšie využíva hodnota SBC a má prioritu pred AIC. Záleží však opäť na analytikovi, ktorej penalizačnej miere dá prednosť.

SBC nám poskytuje silne konzistentné odhady parametrov p a q , ale na druhú stranu s veľkým rozptylom. Pri požití AIC je to však naopak, odhady sú nekonzistentné a taktiež sa potvrdilo, že táto miera nadhodnocuje stupeň autoregresie. Výhodou AIC kritéria je malý rozptyl oproti SBC.

Tieto miery sú teda používané na penalizáciu zbytočne vysokých hodnôt p a q . Kritérium bude nižšie, čím je nižšia smerodajná chyba, a čím je nižší počet vysvetľujúcich premenných, a teda parametrov nami vytvoreného modelu. V tomto kroku nemôžeme zabudnúť na podmienku, aby bol časový rad stacionárny a model invertibilný.

Po odhadnutí modelu nasleduje fáza verifikácie. Pri tejto fáze je dôležité skontrolovať kvalitu modelu, a to najmä rezíduá. Našou úlohou je odhadnúť model tak, aby jeho rezíduá mali viac-menej náhodné správanie. To znamená zabezpečenie nekorelovanosti v čase. Všetky potrebné metodiky sme zmienili v podkapitole 3.3, kde sme si popísali všetky využité testy rezíduí v práci.

Poslednou fázou je samotné prognózovanie na základe vytvoreného modelu. Hlavným cieľom ARIMA modelov je prognostická aplikácia. Preto sa kvalita modelu uskutočňuje aj po prognóze, a to hlavne za pomoci mier presnosti prognóz. Úlohou analytika je odhadnúť model tak, aby sa čo najmenej odlišoval od skutočnosti. Preto chybu prognózy vieme zapísať nasledujúcim vzťahom:

$$e_t = y_t - \hat{y}_t.$$

Medzi miery presnosti patrí viacero priemerných charakteristík, ktoré si postupne popíšeme. Ako prvú si rozoberieme priemernú chybu označovanú aj ako ME, t. j. mean error. Táto miera nadobúda zápornú hodnotu v prípade, že nadhodnocuje skutočnosť a kladú hodnotu, ak podhodnocuje reálne dáta. Túto mieru považujeme za mieru skreslenia a interpretujeme ju na základe jej znamienka, pričom matematicky zápis je nasledovný:

$$ME = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t) = \frac{1}{n} \sum_{t=1}^n e_t.$$

Priemerná absolútna chyba, označovaná aj MAE podľa anglického znenia mean absolute error, vyjadruje priemernú absolútnu odchýlku reálnych hodnôt časového radu od tých predikovaných modelom. Táto miera sa vyjadruje v rovnakých jednotkách ako hodnoty časového radu a matematický zápis je nasledovný:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| = \frac{1}{n} \sum_{t=1}^n |e_t|.$$

Ďalšou mierou je rozptyl chýb, ktorú v literatúre označujú ako priemernú štvorcovú chybu MSE podľa anglického výrazu mean square error. Oproti priemernej absolútnej odchýlky je rozptyl viac citlivý na veľké chyby, ktorým pripisuje vyššiu váhu. Túto mieru často využívame pri výbere z viacerých modelov, keď volíme model s najnižšou hodnotou MSE. Výpočet prebieha na základe nasledujúceho vzorca:

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 = \frac{1}{n} \sum_{t=1}^n e_t^2.$$

Pričom často sa stretávame v programoch s hodnotou RMSE (root mean square error), čo je vlastne smerodajná odchýlka chýb prognóz a vypočítame ju ako:

$$RMSE = \sqrt{MSE}.$$

Predchádzajúce miery sú vyjadrené v rovnakých jednotkách ako časový rad, ale poznáme aj miery, ktoré sa vyjadrujú percentuálne. Patrí sem priemerná percentuálna chyba MPE (mean percentage error), ktorú vypočítame podľa vzťahu:

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(y_t - \hat{y}_t)}{y_t} * 100 \% = \frac{1}{n} \sum_{t=1}^n \frac{e_t}{y_t} * 100 \%.$$

Tak ako MAE ide o mieru skreslenia a interpretuje sa na základe znamienka, čiže ak nadobúda zápornú hodnotu potom systematicky nadhodnocuje realitu a naopak.

Poslednou používanou mierou je priemerná absolútna percentuálna chyba, ktorá sa označuje taktiež podľa anglického znenia MAPE (mean absolute percentage error). Táto miera napovedá akú priemernú veľkosť chýb v percentách dosahuje odhadnutý model na celom úseku prognózovania:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t} * 100 \% = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{y_t} * 100 \%.$$

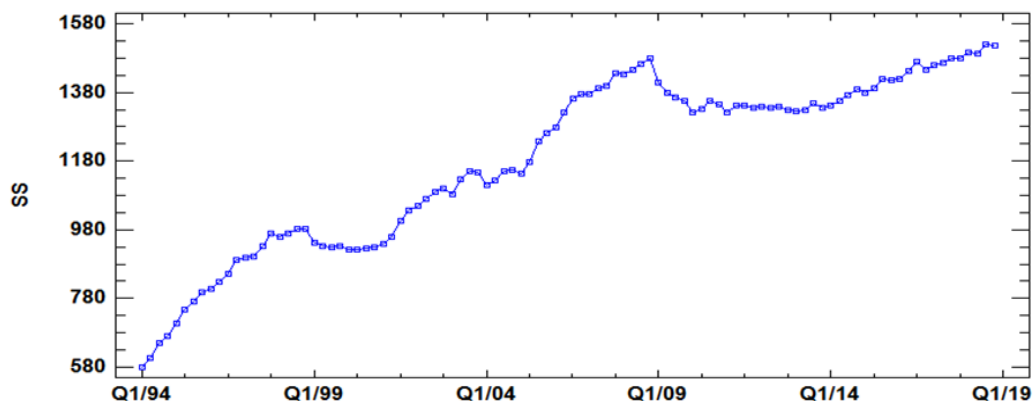
4. Výsledky práce

V tejto časti práce budeme analyzovať časový rad počtu pracujúcich v súkromnom sektore za obdobie Q1/1994 – Q4/2018 v Slovenskej republike. Máme presne 100 pozorovaní, čo je postačujúce pre použitie Box-Jenkinsových modelov. Tieto štvrťročné údaje sme získali z databázy DATACUBE Štatistického úradu SR. Údaje sú v tisíckach osôb pracujúcich v súkromnom sektore, ktoré Štatistický úrad zisťuje pomocou VZPS. Následne otestujeme stacionaritu časového radu. V prípade, že tento predpoklad nespĺňame pristúpime k využitiu potrebných transformácií na dosiahnutie stacionarity. Po splnení tohto požiadavku môžeme modelovať pomocou Box-Jenkinsovej metodológie. Pokúsime sa nájsť najvhodnejší model na prognózu na základe získaných vedomostí z prvej a tretej kapitoly.

4.1 Deskriptívna analýza časového radu

Skôr ako pristúpime k samotnému modelovaniu, musíme poznať správanie vývoja časového radu hodnôt. Je dôležité zistiť prítomnosť sezónnosti, trendu a stacionárnosti hodnôt. Na grafe č. 7 máme znázornený časový rad hodnôt počtu pracujúcich osôb v súkromnom sektore v SR, ktorý sme si zobrazili v programe Statgraphics Centurion XVI, pričom sme si ho pomenovali SS ako skratku súkromného sektora.

Graf č. 7 – Počet pracujúcich v súkromnom sektore Q1/1994 – Q4/2018



Zdroj: Vlastné spracovanie dát ŠÚ SR

Vývoj počtu zamestnancov vieme rozdeliť do viacerých období. Po vzniku Slovenskej republiky nastalo obdobie veľkých reforiem a došlo k zrušeniu väčšiny dodávateľských vzťahov medzi Slovenskom a Českom. Pozitívom bolo prijatie daňovej

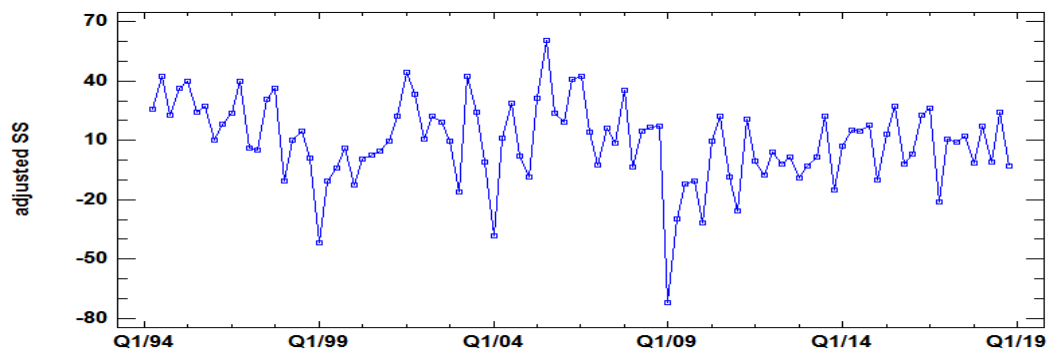
reformy, ktorá nasmerovala Slovensko k trhovému typu ekonomiky. Významným krokom bolo spomalenie privatizácie, čo spôsobilo rozmach zamestnanosti v súkromnom sektore. Napomáhala tomu aj aktívna politika zamestnanosti, ktorá dotovala zamestnávateľov pri zamestnaní absolventov. Táto expanzia pokračovala do roku 1999, keď došlo k miernemu poklesu zapríčinený zavretím veľkých podnikov. Slovenská vláda, ale v roku 2000 zriadila Garančný fond, ktorý pomáhal zamestnancom z veľkých podnikov.

Dôležitým krokom bolo prijatie Národného plánu zamestnanosti v roku 1999, ktorý zlepšoval schopnosť získať zamestnanie, podporoval rozvoj podnikania. Vďaka týmto krokom začali zahraniční investori prichádzať na Slovensko a v roku 2002 sme začali ďalšiu etapu miernu expanzie počtu zamestnaných v súkromnom sektore. Výraznejšia expanzia sa dostavila po vstupe do EÚ, keď Slovensko prijalo Národný akčný plán, ktorého heslom bolo „pracovať sa oplatí“. Začalo sa množstvo reforiem hlavne sa pristúpilo k novelizácii zákonníka práce.

Do konca roka 2008 rástol počet zamestnaných veľmi priaznivým tempom, no potom nastala kríza, ktorá zasiahla všetkých. V roku 2009 došlo k výraznému prepúšťaniu zamestnancov v súkromnej sfére. Následky tejto krízy sa začali zmierňovať v roku 2011, keď sa celá EÚ dostala z najhoršieho. V nasledujúcich rokoch nedošlo k extrémnym nárastom alebo poklesom a ich počet rástol miernym tempom.

Po identifikácii vývoja zamestnaných osôb v súkromnom sektore môžeme pristúpiť k využitiu štatistických metód, ktoré sme si popísali v kapitole č. 3. Na základe grafu č. 7 usudzujeme, že nie je stacionárny v strednej hodnote, keďže hodnoty nekolíšu okolo konštantného priemeru a majú rastúci trend. Taktiež sa domnievame, že časový rad nie je stacionárny v rozptyle, to zistíme pomocou grafu 1. nesezónnej diferenciacie (graf č. 8).

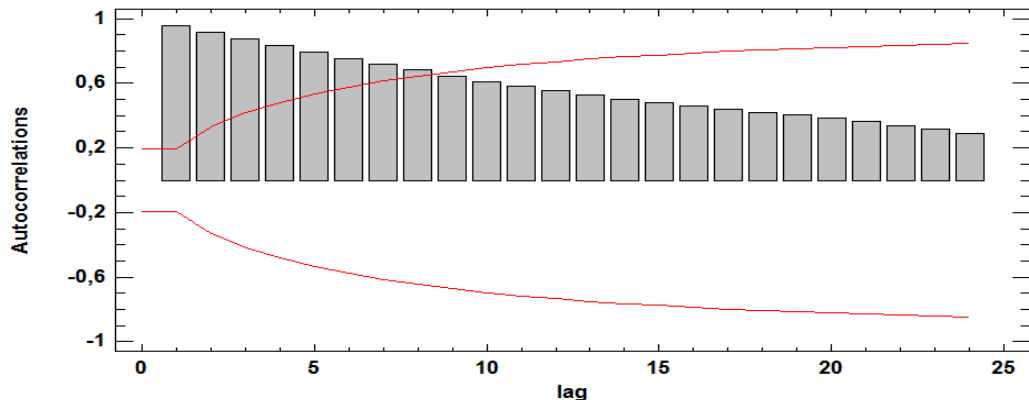
Graf č. 8 – 1. nesezónna diferenciacia



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Z grafu prvej diferenciácie pozorujeme, že variabilita po druhom štvrtroku 2009 je nižšia ako v predošlých obdobiach. Rad je potrebné preto stacionarizovať v rozptyle pomocou logaritmickej transformácie. Nestacionaritu vieme zistiť aj nasledujúcim grafom č. 9, kde je zobrazená autokorelačná funkcia (ACF) pôvodného časového radu.

Graf č. 9 – ACF časového radu pracujúcich v súkromnom sektore

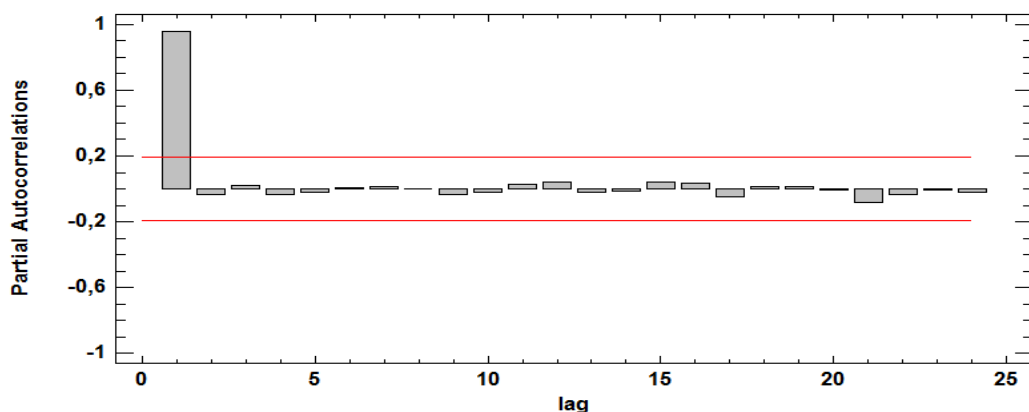


Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

V grafe autokorelačnej funkcie pôvodných hodnôt časového radu pozorujeme nestacionárnosť časového radu v strednej hodnote. Napovedá nám to veľmi pomalý, schodovitý pokles jednotlivých autokorelačných koeficientov, ktorý nám spôsobuje prítomnosť trendu. Taktiež nám zobrazuje štatistickú nevýznamnosť prvých ôsmich autokorelačných koeficientov, keďže prekračujú 95 % hranicu intervalu spoľahlivosti.

Graf č. 10 nám zobrazuje parciálnu funkciu koeficientov autokorelácie. Pozorujeme, že prvý koeficient je veľmi blízky 0 a následné koeficienty majú tvar sinusoidných vln.

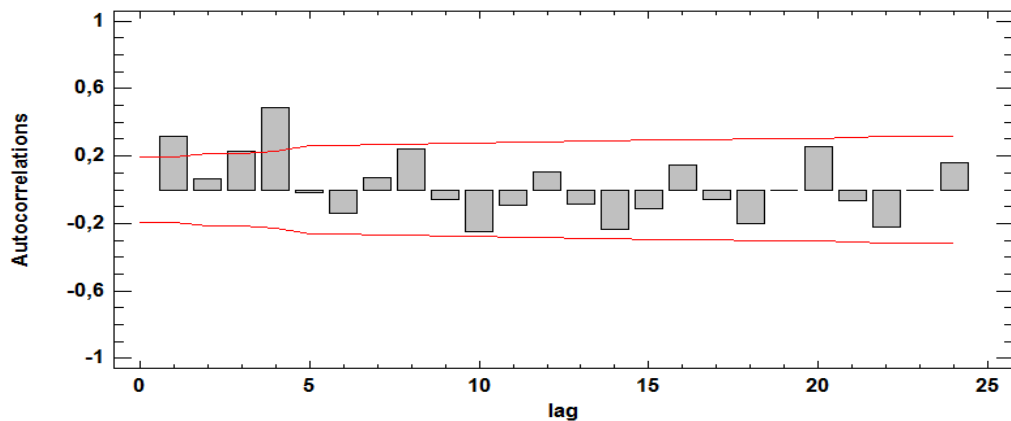
Graf č. 10 – PACF časového radu pracujúcich v súkromnom sektore



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Aj tento pohľad na graf nám potvrdil nestacionárnosť časového radu. Zo zistených informácií s istotou tvrdíme, že budeme časový rad logaritmovať, aby sme odstránili nestacionaritu vo variabilite a následne diferencovať, čím odstránime nestacionaritu v strednej hodnote. Keďže používame štvrťročne údaje, je potrebné overiť aj potrebu sezónnej diferenciácie. K tomu nám pomôže graf autokorelačnej funkcie prvých diferenciácií logaritmovaného časového radu, ktorý je zobrazený na grafe č. 11.

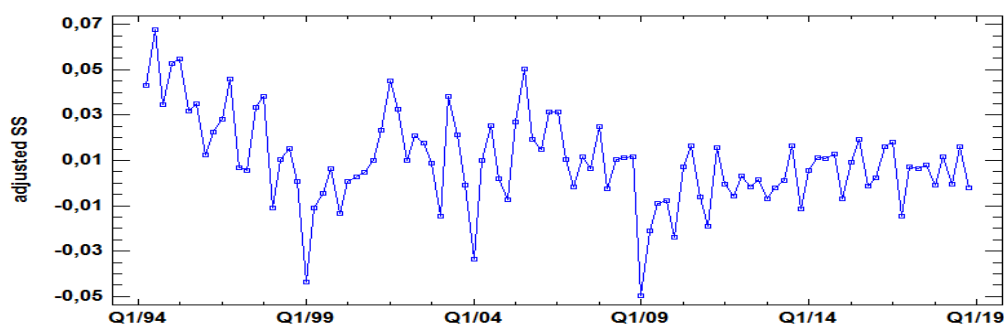
Graf č. 11 – ACF logaritmovaného časového radu 1. diferenciácií



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

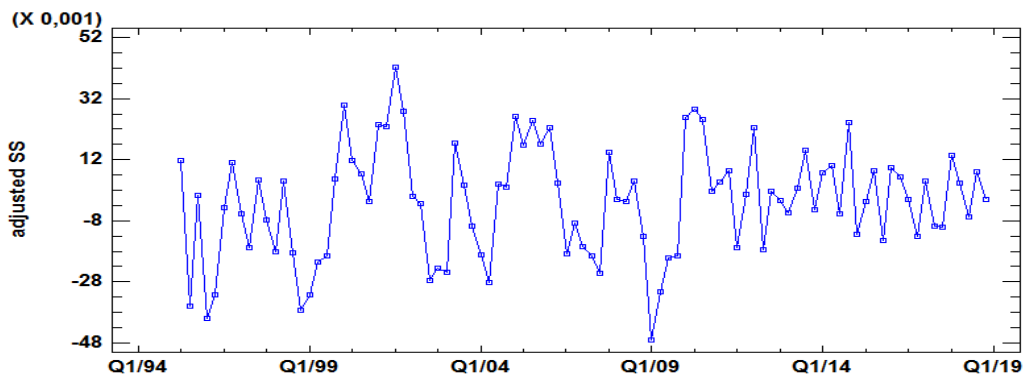
Na základe grafickej analýzy vieme povedať, že štvrtý autokorelačný koeficient je na 5 % hladine významnosti štatisticky významný, a preto bude potrebná aj sezónna diferenciácia. Ešte pred samotným diferencovaním použijeme logaritmickú transformáciu hodnôt. Následne sa pokúsime diferencovať časový rad, až kým nedosiahneme požadovanú stacionárnosť. Porovnáme si nasledujúce 3 grafy, kde použijeme 1. nesezónnu diferenciáciu (graf č. 12), 1. nesezónnu diferenciáciu a 1. sezónnu diferenciáciu (graf č. 13), 2. nesezónnu diferenciáciu a 1. sezónnu diferenciáciu (graf č. 14).

Graf č. 12 – 1. nesezónna diferenciácia logaritmovaného ČR



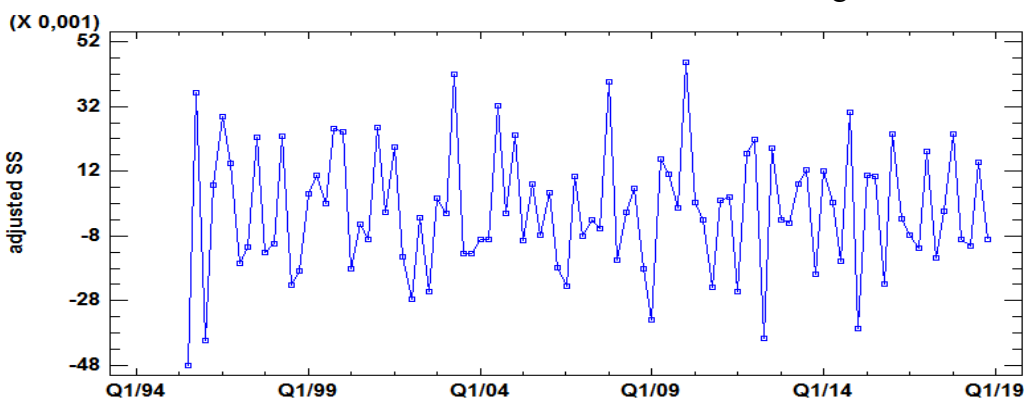
Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Graf č. 13 – 1. nesezónna diferenciácia a 1. sezónna diferenciácia logaritmovaného ČR



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Graf č. 14 – 2. nesezónna diferenciácia a 1. sezónna diferenciácia logaritmovaného ČR



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Vizuálne najstacionárnejšie vyzerá graf č. 14 po použití 2. nesezónnej diferenciácie a 1. sezónnej diferenciácie. Pri tomto kroku si ale musíme dať pozor nato, aby náš časový rad nebol prediferencovaný. Na to nám poslúži nasledujúce porovnanie, kde vyberieme vhodnú diferenciáciu podľa najnižšej hodnoty štandardnej chyby odhadu, teda RMSE.

Obrázok č. 2 – Porovnanie diferenciácií logaritmovaného ČR pracujúcich v SS

Models

- (A) ARIMA(0,1,0) with constant
Math adjustment: Natural log
- (B) ARIMA(0,1,0)x(0,1,0)4 with constant
Math adjustment: Natural log
- (C) ARIMA(0,2,0)x(0,1,0)4 with constant
Math adjustment: Natural log

Estimation Period

Model	RMSE	MAE	MAPE	ME	MPE
(A)	21,6119	16,2351	1,46144	-2,05584	-0,0194272
(B)	21,003	16,9332	1,44638	0,161255	-0,0161912
(C)	22,9775	18,4383	1,55073	-0,238879	-0,018492

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Po zvážení jednotlivých hodnôt RMSE z výstupu na obrázku č. 2 pre jednotlivé typy diferenciácií, použijeme 1. nesezónnu diferenciáciu a 1. sezónnu diferenciáciu. Na zistenie prítomnosti stacionarity v časovom rade sme využívali hlavne grafickú analýzu. V súčasnosti ale máme aj dôkladnejšie techniky ako je napr. rozšírený Dickey-Fullerov test, ktorým zistíme či naše transformácie pretvorili náš pôvodný časový rad na stacionárny. V tabuľke č. 2 máme zobrazené výsledky tohto testu získaného v programe SAS EG 5.1

Tabuľka č. 2 – Rozšírený DF test

Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-52.8563	<.0001	-6.05	<.0001		
Single Mean	0	-53.4682	0.0009	-6.07	<.0001	18.42	0.0010
Trend	0	-55.0036	0.0003	-6.18	<.0001	19.09	0.0010

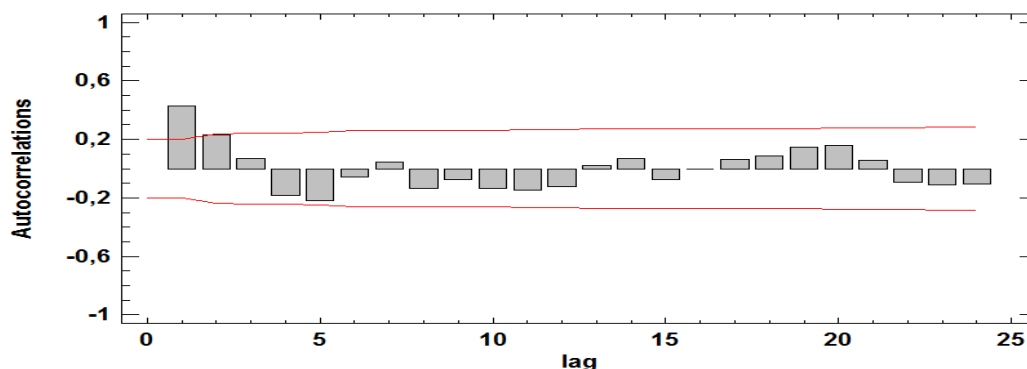
Zdroj: Vlastné spracovanie v programe SAS EG 5.1

Na základe p-hodnôt testu, môžeme potvrdiť správnosť voľby stupňa diferenciácie, keďže je pri každej testovacej štatistike p-hodnota nižšia ako nami zvolená 5 % hladina významnosti, a teda sme môžeme zamietnuť nulovú hypotézu o nestacionarite.

4.2 Identifikácia modelu

Teraz pristúpime k jednej z najťažších úloh Box-Jenkinsových modelov, a to výber parametrov nášho modelu. Na základe zobrazenia autokorelačnej funkcie (graf č. 15) a parciálnej autokorelačnej funkcie (graf č. 16) časového radu dokážeme zvoliť vhodné parametre pre tvorbu modelu.

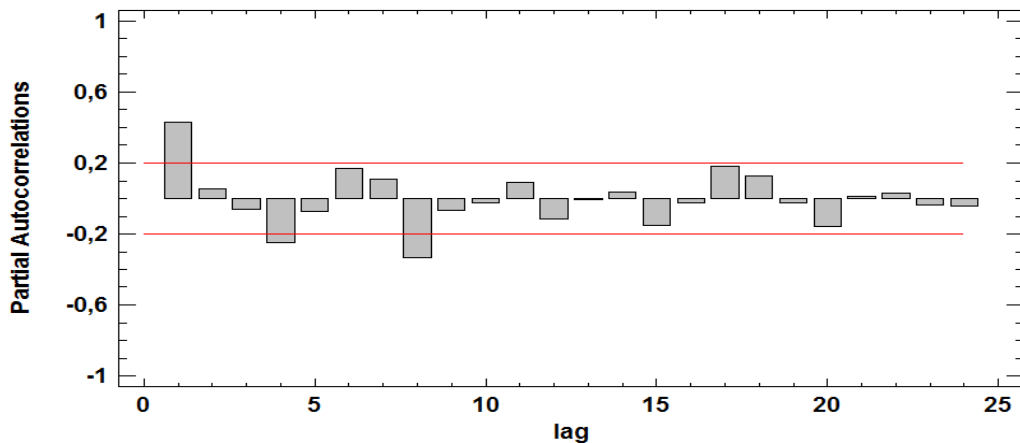
Graf č. 15 – ACF po diferenciácií logaritmovaného ČR



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Na základe grafu č. 15 zobrazujúci autokorelačnú funkciu diferencovaného časového radu, rozhodneme o predpokladanej výške procesov MA resp. SMA. Výrazný pokles v tretom stĺpci nám naznačuje použitie procesu MA s parametrom 3. Aj napriek tomu, že piaty autokorelačný koeficient nepresiahol hranicu štatistickej nevýznamosti, kvôli následnému veľkému poklesu v šiestom autokorelačnom koeficiente uvažujeme aj o procese SMA s parametrom 1.

Graf č. 16 – PACF po diferenciácií logaritmovaného ČR



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Graf č. 16, na ktorom sme si znázornili parciálnu autokorelačnú funkciu diferencovaného časového radu, nám pomôže identifikovať procesy AR resp. SAR. Štatistická významnosť prvého koeficientu a následný pokles nám naznačuje AR proces s parametrom 2. Podobnú situáciu máme aj v štvrtom koeficiente, po ktorom nasleduje pokles, uvažujeme teda aj o procese SAR s parametrom 1.

V nasledujúcich krokoch sa pokúsime vhodnou kombináciou parametrov nájsť model na prognózovanie, poprípade viacero modelov a vyberieme ten najvhodnejší.

4.3 Tvorba modelov

V predošlej podkapitole sme predpokladali, aké parametre použijeme. Teraz si postupným vkladaním parametrov a sledovaním autokorelačnej funkcie a parciálnej autokorelačnej funkcie vytvoríme vhodné modely na prognózu. Keďže náš časový rad tvoria štvrťročné údaje začneme so sezónnymi procesmi SAR a SMA. V prípade, že konštanta v modeli nebude významná, odstránime ju.

4.3.1 ARIMA(0,1,0)x(1,1,1)_{4c}

Na začiatok zistíme ako vplývajú sezónne parametre na náš časový rad. Na základe výstupu z obrázka č. 3 pozorujeme, že bodové odhady parametrov sú štatisticky významne, čo nám potvrdila p-hodnota, ktorá je nižšia ako nami zvolená 5 % hladina významnosti.

Obrázok č. 3 – Model ARIMA(0,1,0)x(1,1,1)_{4c}

Forecast Summary

Math adjustment: Natural log
Nonseasonal differencing of order: 1
Seasonal differencing of order: 1
Forecast model selected: ARIMA(0,1,0)x(1,1,1)₄ with constant
Number of forecasts generated: 0
Number of periods withheld for validation: 0

	<i>Estimation</i>	<i>Validation</i>
<i>Statistic</i>	<i>Period</i>	<i>Period</i>
RMSE	18,0975	
MAE	13,9973	
MAPE	1,19948	
ME	-2,81454	
MPE	-0,277008	

ARIMA Model Summary

<i>Parameter</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t</i>	<i>P-value</i>
SAR(1)	0,420172	0,0948231	4,43111	0,000026
SMA(1)	0,953357	0,0210941	45,1954	0,000000
Mean	-0,000986658	0,000393149	-2,50963	0,013834
Constant	-0,000572092			

Backforecasting: yes

Estimated white noise variance = 0,000257209 with 92 degrees of freedom

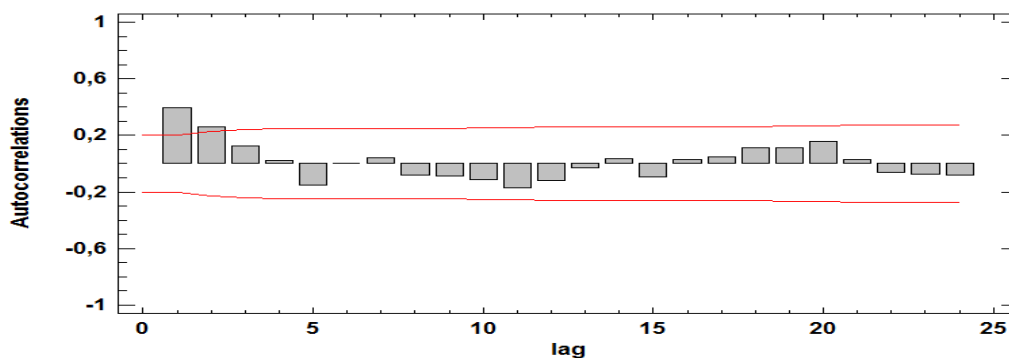
Estimated white noise standard deviation = 0,0160378

Number of iterations: 8

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

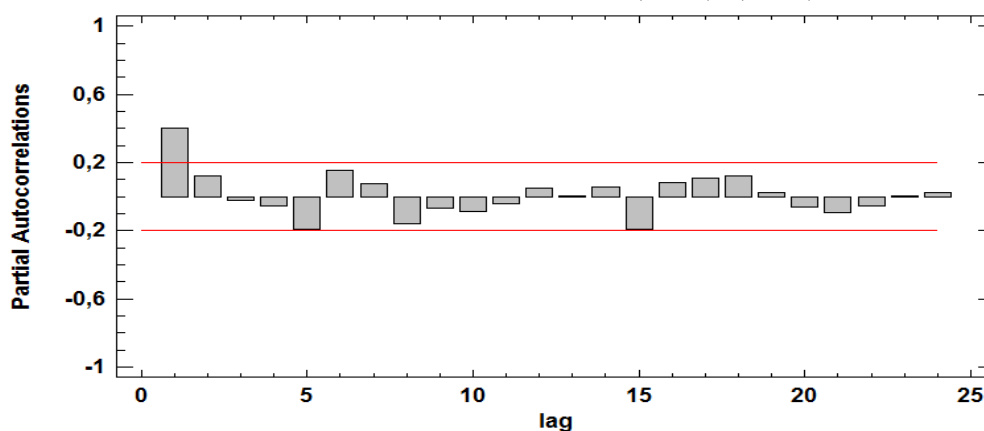
Taktiež dôležitým krokom je otestovanie podmienok stacionarity a invertibility parametrov. V našom prípade to splňame, keďže absolútne hodnoty parametrov sú menšie ako jedna. Pozrime sa bližšie na grafy autokorelačnej a parciálnej autokorelačnej funkcie a zistíme následok vloženia týchto parametrov do modelu.

Graf č. 17 – ACF modelu $ARIMA(0,1,0) \times (1,1,1)_{4c}$



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Graf č. 18 – PACF modelu $ARIMA(0,1,0) \times (1,1,1)_{4c}$



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Z grafov č. 17 a 18 indikujeme potrebu pridať do modelu proces AR s parametrom 2 a proces MA s parametrom 3. Výšku parametrov sme zvolili na základe skokov, ktoré vznikajú medzi autokorelačnými koeficientmi nad hranicou spoľahlivosti a následnými autokorelačnými koeficientmi v hraniciach intervalu spoľahlivosti. Mali by sme sa však pozrieť bližšie na proces SAR, ktorý môžeme identifikovať v grafe parciálnej autokorelačnej funkcie. Momentálne má tento proces parameter 1, no kvôli vysokému rozdielu koeficientov v piatom a šiestom autokorelačnom koeficiente, zvážime aj model s parametrami na úrovni 2.

4.3.2 $ARIMA(0,1,0) \times (2,1,1)_{4c}$

Nasledujúci obrázok č. 4 nám popisuje ďalšiu možnú kombináciu parametrov modelu, presnejšie zo zvýšením stupňa sezónnej autoregresnej časti SAR na 2. Takýmito postupnými iteračnými krokmi sa pokúsime nájsť rôzne modely a vyberieme najvhodnejší.

Obrázok č. 4 – Model ARIMA(0,1,0)x(2,1,1)_{4c}

Forecast Summary

Math adjustment: Natural log
 Nonseasonal differencing of order: 1
 Seasonal differencing of order: 1
 Forecast model selected: ARIMA(0,1,0)x(2,1,1)₄ with constant
 Number of forecasts generated: 0
 Number of periods withheld for validation: 0

	<i>Estimation</i>	<i>Validation</i>
<i>Statistic</i>	<i>Period</i>	<i>Period</i>
RMSE	18,0581	
MAE	13,7913	
MAPE	1,17115	
ME	-2,23343	
MPE	-0,230954	

ARIMA Model Summary

<i>Parameter</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t</i>	<i>P-value</i>
SAR(1)	0,471437	0,10205	4,61966	0,000013
SAR(2)	-0,157125	0,103716	-1,51495	0,133251
SMA(1)	0,948102	0,022306	42,5043	0,000000
Mean	-0,0010534	0,000340296	-3,09555	0,002611
Constant	-0,000722306			

Backforecasting: yes

Estimated white noise variance = 0,000257175 with 91 degrees of freedom

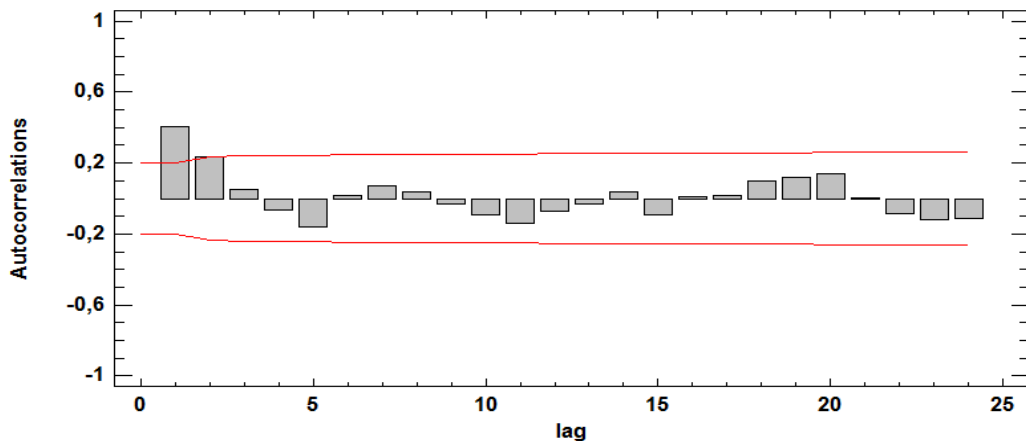
Estimated white noise standard deviation = 0,0160367

Number of iterations: 10

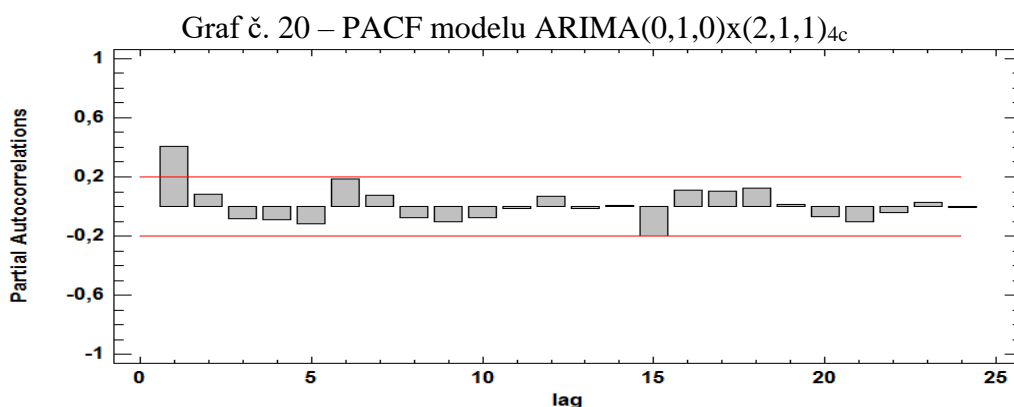
Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

P- hodnota parametra SAR(2) je vyššia ako nami zvolaná hladina významnosti 5 %, teda tento parameter nie je štatisticky významný. V nasledujúcich modeloch však budeme s týmto parametrom počítať, keďže pridaním ďalších parametrov sa môže stať SAR(2) štatisticky významný. Taktiež sú splnené aj podmienky stacionarity a invertibility. Pre tento model si zobrazíme graf č. 19, na ktorom môžeme sledovať zmenu autokorelačnej funkcie po pridaní sezónnych parametrov.

Graf č. 19 – ACF modelu ARIMA(0,1,0)x(2,1,1)_{4c}



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Z grafu č. 20 pozorujeme, že zvýšenie parametra SAR na úroveň 2 spôsobilo zníženie skoku medzi 5. a 6. koeficientom parciálnej autokorelačnej funkcie, a to spôsobilo zníženie závislosť rezíduí na predchádzajúcich pozorovaniach.

4.3.3 $ARIMA(2,1,3) \times (0,1,0)_4$

Teraz sa môžeme pokúsiť namodelovať model s procesmi AR a MA. Z výstupu na obrázku č. 5 pozorujeme štatistickú nevýznamnosť parametra AR(1), keďže je jeho p-hodnota mierne nad hranicou významnosti 5 %.

Obrázok č. 5 – Model $ARIMA(2,1,3) \times (0,1,0)_4$

Forecast Summary

Math adjustment: Natural log
 Nonseasonal differencing of order: 1
 Seasonal differencing of order: 1
 Forecast model selected: $ARIMA(2,1,3) \times (0,1,0)_4$
 Number of forecasts generated: 0
 Number of periods withheld for validation: 0

	Estimation	Validation
Statistic	Period	Period
RMSE	17,0722	
MAE	13,0662	
MAPE	1,10774	
ME	-1,21055	
MPE	-0,114468	

ARIMA Model Summary

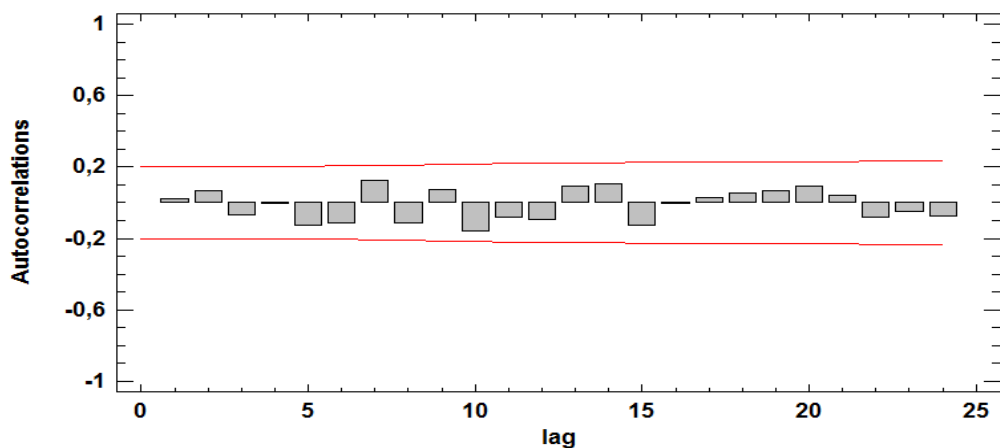
Parameter	Estimate	Std. Error	t	P-value
AR(1)	-0,194679	0,100296	-1,94105	0,055381
AR(2)	-0,476358	0,106893	-4,45638	0,000024
MA(1)	-0,588675	0,0731459	-8,04795	0,000000
MA(2)	-0,957502	0,0357272	-26,8004	0,000000
MA(3)	-0,611392	0,0873786	-6,99705	0,000000

Backforecasting: yes
 Estimated white noise variance = 0,000217762 with 90 degrees of freedom
 Estimated white noise standard deviation = 0,0147568
 Number of iterations: 12

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

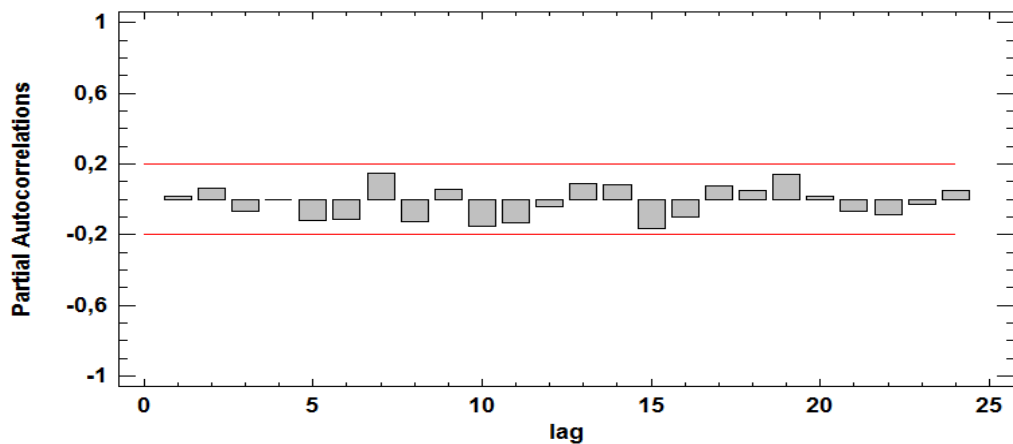
Budeme ale s týmto parametrom počítať aj ďalej, keďže následnou kombináciou so sezónnymi procesmi môže byť tento parameter významný. Dôležitým poznatkom je aj fakt, že nie je splnená podmienka invertibility, keďže absolútna hodnota súčtu parametrov procesu MA nie je nižšia ako 1. Avšak vloženie týchto parametrov spôsobilo, že všetky koeficienty autokorelačnej (graf č. 21) aj parciálnej autokorelačnej funkcie (graf č. 22) sú v rámci hraníc štatistickej významnosti. Tento výsledok zoberieme do úvahy pri ďalšom modelovaní.

Graf č. 21 – ACF modelu ARIMA(2,1,3)x(0,1,0)₄



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Graf č. 22 – PACF modelu ARIMA(2,1,3)x(0,1,0)₄



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

4.3.4 ARIMA(2,1,3)x(1,1,1)₄

Pomocou predošlých troch tzv. „pomocných modelov“ pristúpime ku kombinácii týchto modelov a porovnáme ich medzi sebou. Tento krok bude potrebovať množstvo iteračných krokov a postupov, ktoré sme už aplikovali v predošlých častiach. Preto často dôjdeme k podobným záverom, ktorých interpretácia sa bude možno podobat’.

Obrázok č. 6 – Model ARIMA(2,1,3)x(1,1,1)₄

Forecast Summary

Math adjustment: Natural log
 Nonseasonal differencing of order: 1
 Seasonal differencing of order: 1
 Forecast model selected: ARIMA(2,1,3)x(1,1,1)₄
 Number of forecasts generated: 0
 Number of periods withheld for validation: 0

	Estimation	Validation
Statistic	Period	Period
RMSE	15,8999	
MAE	11,8307	
MAPE	0,997952	
ME	-2,66817	
MPE	-0,216405	

ARIMA Model Summary

Parameter	Estimate	Std. Error	t	P-value
AR(1)	0,825463	0,13848	5,9609	0,000000
AR(2)	-0,62877	0,127934	-4,9148	0,000004
MA(1)	0,439012	0,147888	2,96855	0,003854
MA(2)	-0,68753	0,116838	-5,88447	0,000000
MA(3)	-0,281637	0,126166	-2,23227	0,028136
SAR(1)	0,477347	0,104356	4,57422	0,000016
SMA(1)	0,952279	0,0204442	46,5794	0,000000

Backforecasting: yes

Estimated white noise variance = 0,000193628 with 88 degrees of freedom

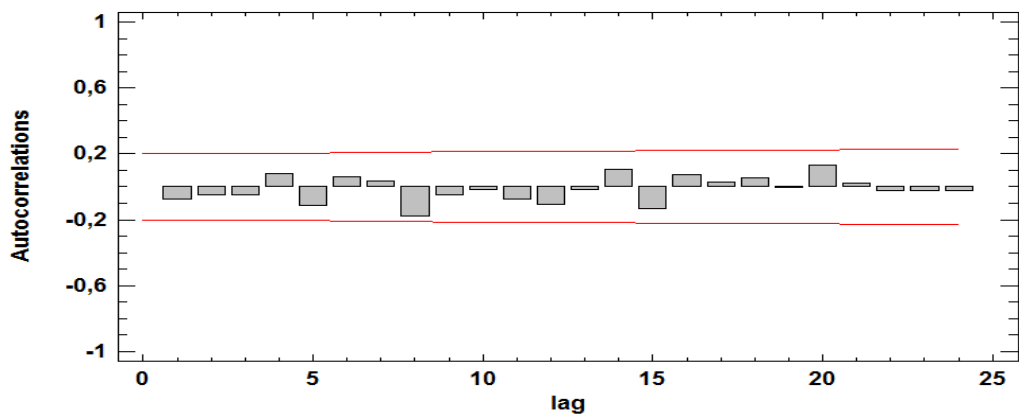
Estimated white noise standard deviation = 0,013915

Number of iterations: 11

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

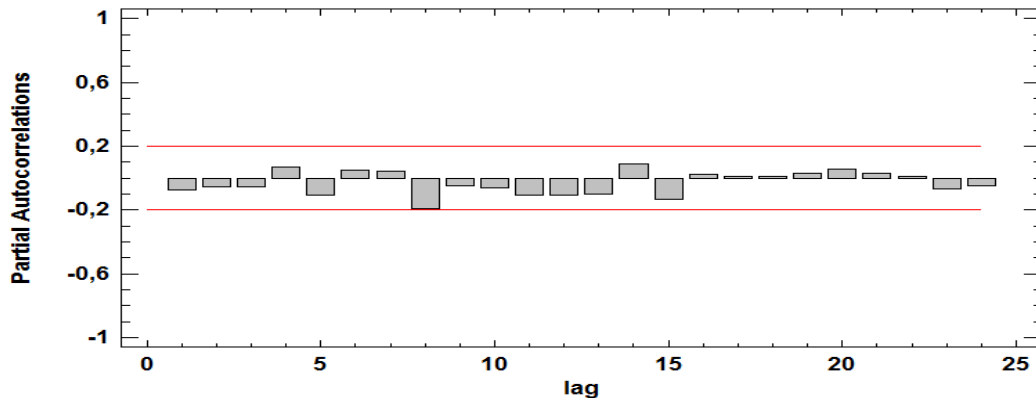
Pri porovnaní p-hodnoty na obrázku č. 6 s 5 % hladinou významnosti môžeme tvrdiť, že všetky parametre sú štatisticky významné. Stacionaritu procesu AR nám potvrdila platnosť nasledujúcich podmienok $\phi_2 + \phi_1 < 1$, $\phi_2 - \phi_1 < 1$, $-1 < \phi_2 < 1$. Pri procese SAR použijeme nasledujúcu podmienku $|\Phi_1| < 1$, ktorá nám potvrdila stacionárnosť. Nesmieme zabudnúť overiť podmienky invertibility pre procesy MA a SMA. O SMA môžeme potvrdiť invertibilnosť pri splnení podmienky $|\Theta_1| < 1$, čo sa v našom prípade potvrdilo. V prípade MA procesu sme potvrdili túto požiadavku, keďže musí byť absolútna hodnota súčtu všetkých parametrov menšia ako 1.

Graf č. 23 – ACF modelu $ARIMA(2,1,3) \times (1,1,1)_4$



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Graf č. 24 – PACF modelu $ARIMA(2,1,3) \times (1,1,1)_4$



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Všetky koeficienty ACF a PACF v grafoch č. 23 a 24 sú na hladine významnosti 5 % štatisticky významné, keďže neprekročili hranice znázornene červenou čiarou. To napovedá o nezávislosti rezíduí a vhodnosti modelu na ďalšiu prognózu.

4.3.5 $ARIMA(2,1,3) \times (2,1,1)_{4c}$

Predchádzajúci model spĺňa všetky základné požiadavky, ale ako sme už zistili Box-Jenkinsonova metodológia je založená na opakovaní krokov. Kombináciou iných parametrov môžeme následne dospieť k úspešnejšiemu modelu.

Obrázok č. 7 – Model ARIMA(2,1,3)x(2,1,1)_{4c}

Forecast Summary

Math adjustment: Natural log
 Nonseasonal differencing of order: 1
 Seasonal differencing of order: 1
 Forecast model selected: ARIMA(2,1,3)x(2,1,1)₄ with constant
 Number of forecasts generated: 0
 Number of periods withheld for validation: 0

	<i>Estimation</i>	<i>Validation</i>
<i>Statistic</i>	<i>Period</i>	<i>Period</i>
RMSE	15,1109	
MAE	11,3505	
MAPE	0,951812	
ME	-1,04351	
MPE	-0,109961	

ARIMA Model Summary

<i>Parameter</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>t</i>	<i>P-value</i>
AR(1)	0,8892	0,0876475	10,1452	0,000000
AR(2)	-0,749546	0,0760209	-9,85973	0,000000
MA(1)	0,500155	0,081743	6,11864	0,000000
MA(2)	-0,796184	0,049222	-16,1754	0,000000
MA(3)	-0,257737	0,0585681	-4,40063	0,000031
SAR(1)	0,557316	0,110397	5,04829	0,000002
SAR(2)	-0,32718	0,108295	-3,02118	0,003316
SMA(1)	0,934547	0,0230806	40,4905	0,000000
Mean	-0,000998866	0,000483162	-2,06735	0,041707
Constant	-0,000661598			

Backforecasting: yes

Estimated white noise variance = 0,000177377 with 86 degrees of freedom

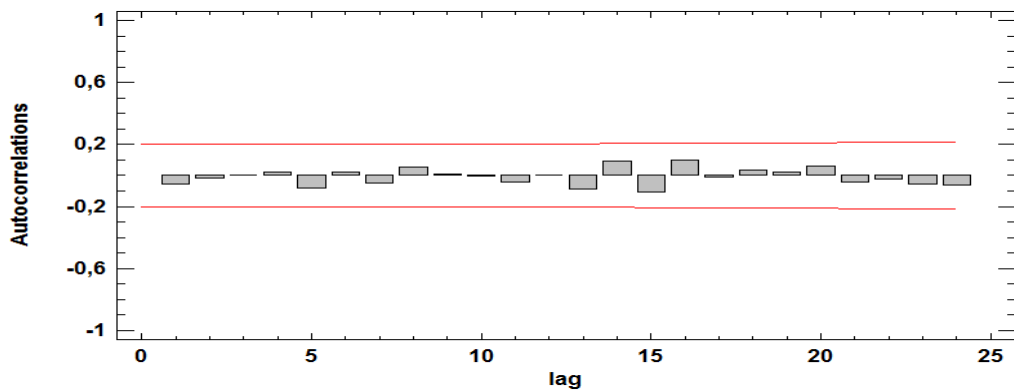
Estimated white noise standard deviation = 0,0133183

Number of iterations: 9

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

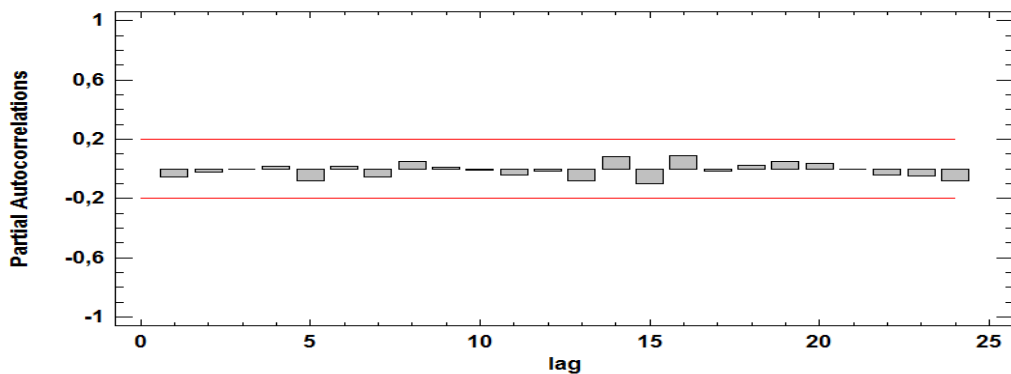
Na obrázku č. 7 sú všetky parametre štatisticky významné, keďže ich p-hodnoty sú nižšie ako nami zvolená 5 % hladina významnosti. Stacionaritu procesu AR nám potvrdila platnosť nasledujúcich podmienok $\phi_2 + \phi_1 < 1$, $\phi_2 - \phi_1 < 1$, $-1 < \phi_2 < 1$. Pri procese SAR použijeme obdobné podmienky ako pri AR, ktoré aj pri tomto procese potvrdili stacionárnosť. Tak ako pri predchádzajúcom modeli, zistíme invertibilitnosť našich procesov MA a SMA. Invertibilitu procesu SMA nám potvrdzuje podmienka $|\Theta_1| < 1$. V prípade MA procesu sme potvrdili túto požiadavku, keďže absolútna hodnota súčtu všetkých parametrov je menšia ako 1.

Graf č. 25 – ACF modelu $ARIMA(2,1,3) \times (2,1,1)_{4c}$



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Graf č. 26 – PACF modelu $ARIMA(2,1,3) \times (2,1,1)_{4c}$



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Tak ako pri predošlom modeli všetky koeficienty podľa grafu č. 25 a 26 sú na hladine významnosti 5 % štatisticky významne, čo nasvedčuje správne úsudku pri výbere parametrov jednotlivých procesov.

4.4 Výber modelu

Pomocou kombinácie parametrov ARIMA modelov sme dospeli k dvom vyhovujúcim modelom:

- $ARIMA(2,1,3) \times (2,1,1)_{4c}$,
- $ARIMA(2,1,3) \times (1,1,1)_4$.

V nasledujúcej časti zvolíme vhodnejší model na prognózu počtu pracujúcich v súkromnom sektore. Prognózovať budeme rok 2019, preto budeme potrebovať prognózu na ďalšie štyri obdobia. Výber modelu zväzíme taktiež na validačnom období o dĺžke osem. K zvoleniu správneho modelu nám dopomôžu aj úsudky o štatistickej významnosti

parametrov a o invertibilite, stacionarnosti procesov, ktoré sme získali v predchádzajúcej časti. Taktiež sa rozhodneme na základe priemerných charakteristík rezíduí a zväžíme aj hodnoty AIC resp. BIC kritéria.

Obrázok č. 8 – Porovnanie výsledných modelov

Models

(A) ARIMA(2,1,3)x(2,1,1)₄ with constant

Math adjustment: Natural log

(B) ARIMA(2,1,3)x(1,1,1)₄

Math adjustment: Natural log

Estimation Period

Model	RMSE	MAE	MAPE	ME	MPE
(A)	15,1109	11,3505	0,951812	-1,04351	-0,109961
(B)	15,8999	11,8307	0,997952	-2,66817	-0,216405

Model	RMSE	RUNS	RUNM	AUTO	MEAN	VAR
(A)	15,1109	OK	OK	OK	OK	OK
(B)	15,8999	OK	OK	OK	OK	OK

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Porovnaním RMSE jednotlivých modelov z obrázku č. 8, môžeme usudzovať že model A je vhodnejší na prognózu, keďže model s nižšou hodnotou RMSE poskytuje užšie intervaly spoľahlivosti. Podľa priemerných charakteristík rezíduí sa dopustíme v období odhadu chyby v priemere 0,95 % (MAPE) a hodnoty odhadu sa budú v priemere líšiť od skutočnosti o +/- 15,11 tisíc pracujúcich v súkromnom sektore. Model mierne nadhodnocuje údaje, čo nám potvrdila štatistika MPE, ktorá je záporná. Každý štvrťrok je nadhodnotený v priemere o 0,10 %. Všetky tieto štatistiky sú v prospech modelu A.

Prijatie modelu A nám potvrdili aj hodnoty AIC a BIC kritéria, ktoré sú nižšie ako v modeli B. Presné hodnoty sú v tabuľke č. 3.

Tabuľka č. 3 – AIC/BIC kritérium

	Model	AIC	BIC
(A)	ARIMA(2,1,3)x(2,1,1) _{4c}	5,61084	5,8453
(B)	ARIMA(2,1,3)x(1,1,1) ₄	5,67263	5,85499

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Na základe všetkých získaných vedomostí z výstupov a našich výpočtov, volíme na prognózovanie model ARIMA(2,1,3)x(2,1,1)_{4c}. Hoci náš model spĺňa podmienky stacionarity a invertibility, má štatisticky významne všetky parametre, musíme jeho vhodnosť a adekvátnosť na prognózovanie overiť pomocou analýzy rezíduí s cieľom

overiť podmienky bieleho šumu pre náhodné zložky. Presnejšie ich nezávislosť, homoskedasticitu a normalitu.

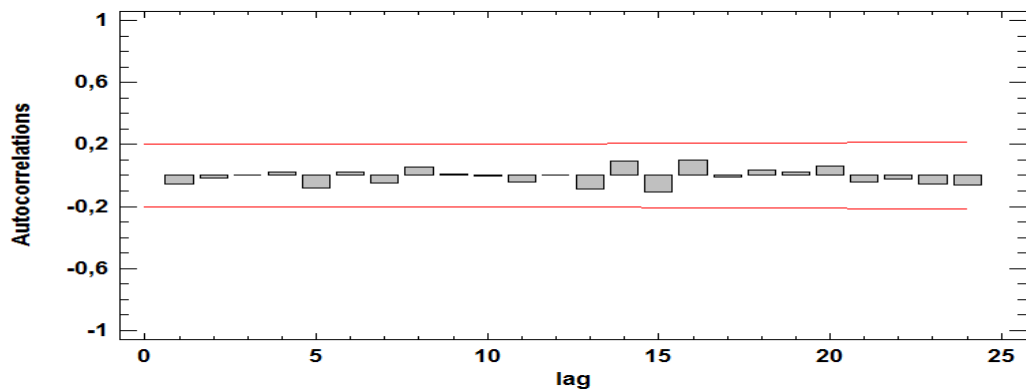
4.5 Splnenie predpokladov o náhodnej zložke modelu

Ako prvú podmienku otestujeme nezávislosť náhodných zložiek. Testovacie hypotézy zapíšeme do nasledujúceho vzťahu:

- $H_0: \rho_k(e_t) = 0$, teda náhodné premenné sú lineárne nezávislé,
- $H_1: \rho_k(e_t) \neq 0$, teda náhodné premenné sú lineárne závislé (autokorelované).

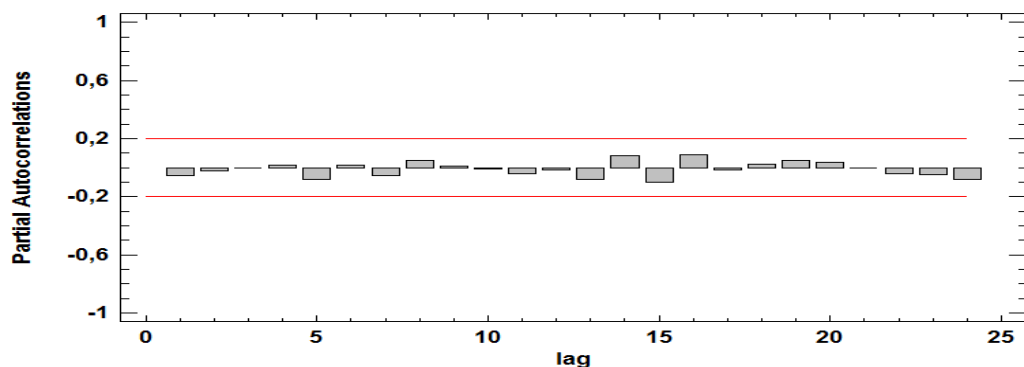
Graficky môžeme potvrdiť nezávislosť pomocou grafov autokorelačnej a parciálnej autokorelačnej funkcie rezíduí na nasledujúcich grafoch č. 27 a 28.

Graf č. 27 – ACF rezíduí modelu



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Graf č. 28 – PACF rezíduí modelu



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Žiadny z koeficientov nepresahuje svoju 95 % hranicu spoľahlivosti, takže môžeme prijať nulovú hypotézu o nezávislosti rezíduí modelu. Aby sme nemuseli testovať každý koeficient autokorelácie v grafe a nepoužívali len grafickú analýzu, použijeme aj testy ako

sú napr. znamienkový test sekvencií nad alebo pod mediánom (runs above and below median), test náhodnosti založený na bodoch obratu (runs up and down) a Boxov-Piercov test, ktoré sme si popísali v kapitole 3. Všetky výsledky testov sú zobrazené na obrázku č. 9.

Obrázok č. 9 – Testy nezávislosti náhodných zložiek

Model: ARIMA(2,1,3)x(2,1,1)4 with constant
Math adjustment: Natural log

- (1) Runs above and below median
Median = -0,00172315
Number of runs above and below median = 52
Expected number of runs = 48,0
Large sample test statistic z = 0,725908
P-value = 0,467893
- (2) Runs up and down
Number of runs up and down = 59
Expected number of runs = 63,0
Large sample test statistic z = 0,859905
P-value = 0,38984
- (3) Box-Pierce Test
Test based on first 24 autocorrelations
Large sample test statistic = 6,62352
P-value = 0,979854

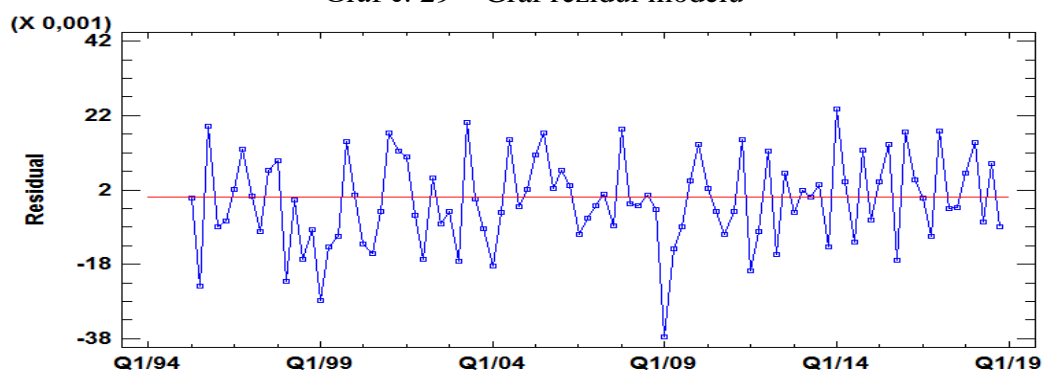
Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Pri všetkých testoch máme dostatočne vysokú p-hodnotu na prijatie nulovej hypotézy, teda koeficienty autokorelácie sa rovnajú 0, a teda rezíduá modelu sú nezávislé.

Pri tvorbe modelu nesmieme zabudnúť na heteroskedasticitu, ktorá sa často nachádza v modeloch so sezónnou zložkou. Nechceme aby nám variabilita časom rástla. Pre testovanie použijeme hypotézy:

- $H_0: \sigma_{\varepsilon 1}^2 = \sigma_{\varepsilon 2}^2$,
- $H_1: \sigma_{\varepsilon 1}^2 \neq \sigma_{\varepsilon 2}^2$.

Graf č. 29 – Graf rezíduí modelu



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Na grafe č. 29 pozorujeme vývoj rezíduí v určitom konštantnom pásme bez nejakého systematického priebehu, a teda variabilita časom nerastie. No aby sme si naše subjektívne rozhodnutie potvrdili, pristúpime k pokročilejším metódam a použijeme testy o zhode rozptylov. Sem zaraďujeme Bartlettov test, Leveneho test a Cochranov test. My sme ale v kapitole 3 popísali jednoduchý test homoskedasticity náhodných zložiek, a to F-test pre rezíduá modelu, ktorý je kvôli možnostiam programu prístupnejší ako predtým spomínané testy a vykonáme rozhodnutie o prijatí resp. zamietnutí hypotézy na základe tohto testu.

Obrázok č. 10 – F-test homoskedasticity náhodných zložiek

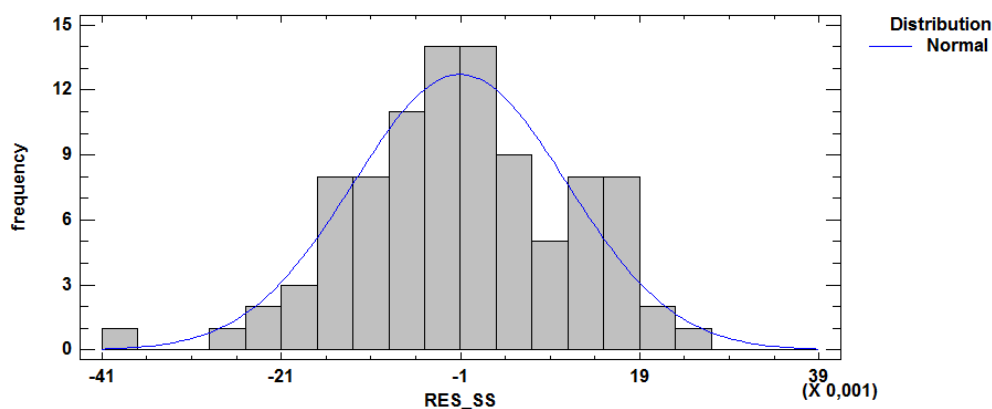
```
F-test to Compare Standard Deviations
Null hypothesis: sigma1 = sigma2
Alt. hypothesis: sigma1 NE sigma2
F = 1,05614 P-value = 0,853851
Do not reject the null hypothesis for alpha = 0,05.
```

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Na základe vysokej p-hodnoty z obrázku č. 10 potvrdzujeme prijatie nulovej hypotézy o homoskedasticite rezíduí na 5 % hladine významnosti.

Jednou z podmienok bieleho šumu je aj normalita náhodných zložiek. Tento predpoklad nám umožňuje vypočítať intervaly spoľahlivosti skutočných hodnôt časového radu, ale aj jeho krátkodobých prognóz. Tak ako pri predchádzajúcich testoch môžeme pristúpiť najprv ku grafickému overeniu pomocou histogramu.

Graf č. 30 – Histogram rezíduí modelu



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Z grafického zobrazenia histogramu rezíduí (graf č. 30) nemôžeme s určitosťou tvrdiť normalitu rezíduí, a preto pristúpime k sofistikovanejším metódam ako je napr. Chí-kvadrát test dobrej zhody, Shapiro-Wilkov test alebo Kolmogorov-Smirnov test. Keďže sme si v kapitole 3 bližšie popísali posledný spomínaný test, tak pomocou programu spustíme potrebnú procedúru a urobíme rozhodnutie pomocou tohto testu.

Obrázok č. 11 – Kolmogorov-Smirnov test normality rezíduí modelu

Kolmogorov-Smirnov Test	
	Normal
DPLUS	0,0404704
DMINUS	0,0620356
DN	0,0620356
P-Value	0,858068

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Aj pri tomto teste sme pomocou vysokej p-hodnoty z obrázku č. 11 potvrdili prijatie nulovej hypotézy o zhode rezíduí s teoretickým normálnym rozdelením. Po otestovaní všetkých podmienok bieleho šumu, môžeme potvrdiť vhodnosť modelu na prognózu. Teraz môžeme pristúpiť k matematickému zápisu nášho modelu.

4.6 Zápis modelu

Pomocou spätného operátora B vieme zapísať model na prognózovanie do nasledujúceho tvaru:

$$\begin{aligned}
 & [1 - (\phi_1 - \Phi_1)B + (\phi_1\Phi_1 - \phi_2)B^2 + \phi_2\Phi_1B^3 - \Phi_2B^8 + \phi_1\Phi_2B^9 - \phi_2\Phi_2B^{10}]z_t = \\
 & = \mu + (1 - \theta_1B - \theta_2B^2 - \theta_3B^3 - \Theta_1B^4 + \theta_1\Theta_1B^5 + \theta_2\Theta_1B^6 + \theta_3\Theta_1B^7)e_t,
 \end{aligned}$$

kde

$z_t = (1 - B)(1 - B^4)^{-1} \log(y_t)$ – stacionárny časový rad po transformáciách,

$\phi_1; \phi_2; \phi_3; \Phi_1; \Phi_2$ – parametre autoregresných procesov AR a SAR,

$\theta_1; \theta_2; \theta_3; \Theta_1$ – parametre procesov kĺzavých priemerov MA a SMA,

e_t – náhodné zložky,

μ – konštanta.

Kvôli obtiažnosti zápisu sme nezapísali model pomocou konkrétnych hodnôt parametrov a taktiež sme stacionárny časový rad nechali v tvare z_t , teda sme ho nerozpisovali pomocou hodnôt pôvodného časového radu y_t .

4.7 Simulácia ex-post

Pred prognózou ex-ante sme vykonali simuláciu ex-post a overili si ako náš model úspešne namodeloval hodnoty za posledných 8 období. Výsledné porovnanie sa nachádza v tabuľke č. 4.

Tabuľka č. 4 – Prognóza ex-post od Q1/2017 do Q4/2018

Obdobie	Skutočnosť	Ex-post	Chyba prognózy
Q1/17	1458,4	1432,54	25,86
Q2/17	1467,5	1466,85	0,65
Q3/17	1479,4	1478,25	1,15
Q4/17	1478	1468,04	9,96
Q1/18	1495,3	1473,85	21,45
Q2/18	1494,4	1504,91	-10,51
Q3/18	1518,5	1501,68	16,82
Q4/18	1515,3	1523,35	-8,05

Zdroj: Vlastné výpočty a program Statgraphic Centurion XVI

Pre zistenie kvality prognózy na validačnej množine 8 období použijeme miery presnosti, ktoré máme vo výstupe na obrázku č. 12.

Obrázok č. 12 – Miery presnosti prognózy ex-post

Validation Period					
Model	RMSE	MAE	MAPE	ME	MPE
(A)	14,5266	11,8063	0,793228	7,16547	0,484536

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Pri testovaní modelu na validačnom období 8 štvrťrokov sme zistili, že sme sa dopustili v priemere 0,79 % chyby (MAPE). Model podhodnocoval hodnoty časového radu každý štvrťrok v priemere o 0,48 % (MPE). Taktiež sme zistili na základe RMSE, že sa odhad odchyľuje od skutočnosti v priemere o 14,53 tisíc pracujúcich v súkromnom sektore.

4.8 Prognóza ex-ante

Po všetkom analýzach sme pristúpili k finálnej prognóze na ďalšie 4 obdobia, teda na celý rok 2019. Prognóza aj hranice 95 % intervalu spoľahlivosti sú v tabuľke č. 5.

Tabuľka č. 5 – Prognóza ex-ante na rok 2019

		<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
<i>Period</i>	<i>Forecast</i>	<i>Limit</i>	<i>Limit</i>
Q1/19	1503,13	1463,85	1543,46
Q2/19	1495,15	1428,91	1564,46
Q3/19	1507,66	1412,2	1609,58
Q4/19	1499,7	1376,4	1634,05

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Na základe výstupu prognózy ex-ante na rok 2019, môžeme konštatovať, že v prvom kvartáli 2019 klesne počet pracujúcich v súkromnom sektore na hodnotu 1503,13 tis osôb. Zároveň sa s pravdepodobnosťou 95 % očakáva, že sa počet pracujúcich bude nachádzať v intervale od 1463,85 do 1543,46 tisíc osôb.

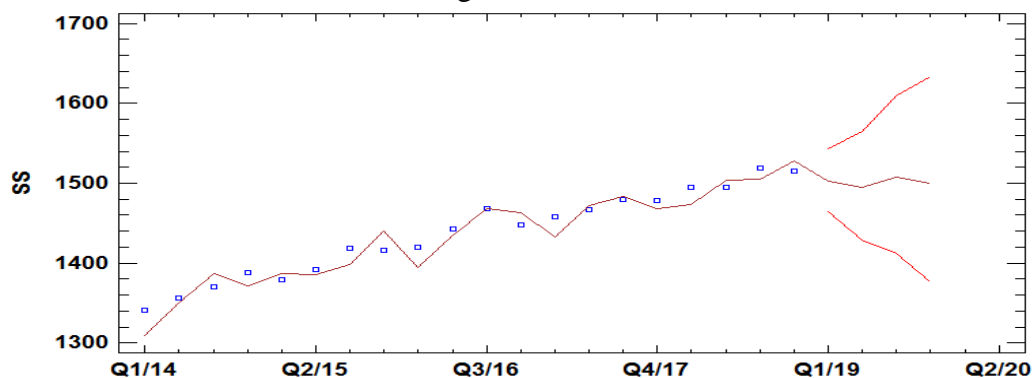
V druhom kvartáli 2019 sa počet pracujúcich v súkromnom sektore opäť zníži na úroveň 1495,15 tis. osôb. Zároveň sa s pravdepodobnosťou 95 % očakáva, že sa počet pracujúcich bude nachádzať v intervale od 1428,91 do 1564,46 tisíc osôb.

V treťom kvartáli 2019 dôjde k priaznivému nárastu počtu pracujúcich v súkromnom sektore, a to na hodnotu 1507,66 tis. osôb. Zároveň sa s pravdepodobnosťou 95 % očakáva, že sa počet pracujúcich bude nachádzať v intervale od 1412,2 do 1609,58 tisíc osôb.

V štvrtom kvartáli 2019 bude počet pracujúcich v súkromnom sektore vo výške 1499,7 tis. Zároveň sa s pravdepodobnosťou 95 % očakáva, že sa počet pracujúcich bude nachádzať v intervale od 1376,4 do 1634,05 tisíc osôb.

Pre lepšiu vizualizáciu sme si prognózované hodnoty vložili do grafu č. 31.

Graf č. 31 – Prognóza ex-ante na rok 2019



Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

V nasledujúcej tabuľke č. 6 máme porovnanie prognózy ex-ante na rok 2019 so skutočnými hodnotami, ktoré sme dokázali získať pomocou databázy ŠÚ SR DATACUBE.

Tabuľka č. 6 – Porovnanie skutočných hodnôt a prognózy

Obdobie	Skutočnosť	Prognóza	ME	MPE
Q1/19	1516,8	1503,13	13,67	0,90 %
Q2/19	1474,6	1495,15	- 20,55	- 1,39 %
Q3/19	1501,5	1507,66	- 6,16	- 0,41 %
Q4/19	1505,8	1499,7	6,1	0,41 %

Zdroj: Vlastné spracovanie v programe Statgraphic Centurion XVI

Náš model prognózoval prvý a štvrtý štvrťrok menej pracujúcich v súkromnom sektore ako naozaj bolo. Model nadhodnocoval hodnoty časového radu každý štvrťrok v priemere 0,12 % (MPE). To znamená, že model nadhodnotil každý štvrťrok v priemere o 1,74 (ME) tisíc zamestnaných osôb. Na základe nízkych priemerných hodnôt mier presnosti MPE a ME usudzujeme, že sme zvolili vhodný model na prognózu.

Záver

Je nesmierne dôležité poznať minulý, súčasný a aj budúci vývoj počtu zamestnaných osôb v súkromnom sektore. Rozvoj zamestnanosti je pre ekonomiku Slovenska jednou z možností ako podporiť rast domácej ekonomiky a udržať si pozíciu prosperujúcej krajiny EÚ. Schopnosť odhadnúť vývoj počtu zamestnaných osôb v súkromnom sektore pomáha krajine lepšie sa pripraviť na nadchádzajúce obdobie a hlavne pomáha prijať včas správne rozhodnutia, ktoré pomôžu zabrániť nepriaznivému vývoju ekonomiky Slovenska. Preto sme uznali za dôležité zanalyzovať vývoj počtu zamestnaných osôb v súkromnom sektore, aby mohli mať politické rozhodnutia aj kvantitatívne dôvody.

Cieľom diplomovej práce bolo pomocou využitia Box-Jenkinsovej metodológie navrhnúť vhodný model na prognózu počtu zamestnaných osôb v súkromnom sektore na všetky štvrťroky roku 2019.

Nevyhnutnou podmienkou využitia ARIMA modelov bola stacionárnosť časového radu hodnôt. Bohužiaľ väčšina ekonomických časových radov túto podmienku nespĺňa. Preto sme časový rad počtu zamestnaných osôb v súkromnom sektore v tisíckach osobách za obdobie Q1/94 až Q4/18 museli stacionarizovať pomocou potrebných transformácií, a to logaritmickou transformáciou, aby sme dosiahli stacionaritu v rozptyle a využitím sezónnej a nesezónnej diferenciácie sme dosiahli stacionaritu v strednej hodnote. Sledovaním vývoja autokorelačnej a parciálnej autokorelačnej funkcie transformovaného radu sme dospeli k výberu vhodného prognostického modelu.

Na základe štandardnej chyby RMSE sme vybrali vhodnejší model, a to $ARIMA(2,1,3) \times (2,1,1)_4$ s konštantou. Využitím modelu sme sa dopustili pri odhade roku 2019 v priemere 0,95 % chyby (MAPE), pričom hodnoty sa môžu odchyľovať v priemere o cca $\pm 15,11$ (RMSE) tisíc zamestnaných osôb v súkromnom sektore. Keďže je hodnota $MPE < 0$ model mierne systematicky nadhodnocuje skutočnosť, čo je v absolútnom vyjadrení nadhodnotenie každého štvrťroka v priemere o 1,04 tisíc zamestnaných osôb v súkromnom sektore. Všetky parametre modelu sú podľa testov štatisticky významne a spĺňajú podmienky invertibility a stacionarity. Analýzou rezíduí sme dospeli k záveru, že spĺňajú podmienky bieleho šumu.

Taktiež sme vykonali analýzu ex-post na validačnej množine osem období, t. j. rok 2018 a 2017. Zistili sme, že model sa dopustil v priemere 0,79 % chyby (MAPE) a podhodnocoval hodnoty časového radu každý štvrtrok v priemere o 0,48 % (MPE). Na základe RMSE môžeme tvrdiť, že sa odhad odchyľuje od skutočnosti v priemere o \pm 14,53 tisíc zamestnaných v súkromnom sektore. Prognóza ex-ante na posledný štvrtrok 2019 predpokladala pokles na úroveň 1499,7 tisíc osôb.

Tvorba modelu, ktorý najlepšie vystihuje vývoj je citlivá záležitosť a záleží na skúsenostiach analytika. My sme zvolili ARIMA modely, pomocou ktorých vieme pomerne rýchlo a jednoducho zistiť budúce hodnoty ukazovateľa. Výhodou je schopnosť modelovať časové rady s periodicitou alebo cyklickými pohybmi. Nevýhodou týchto modelov je analyzovanie iba jedného časového radu a fakt, že sa oddiaľujú od ekonomickej teórie oproti iným typom modelov. Je nesmierne dôležité model udržiavať aktuálny a reagovať na vývoj nových pozorovaní úpravou parametrov výsledného modelu.

Zoznam použitej literatúry

- ABUGABER, D. Using ARIMA for Time Series Analysis Dostupné na: <https://ademos.people.uic.edu/Chapter23.html>
- ARTL, J. – ARTLOVÁ M., – RUBLÍKOVÁ E. Analýza ekonomických časových řad s příklady: skriptá. Praha: Vysoká škola ekonomická v Prahe, 2002. 147 s.
- ARLT, J – ARLTOVÁ, M. Ekonomické časové řady. Praha : Grada Publishing, 2007. 285 s. ISBN 978-80-247-1319-9.
- COGHLAN, A. 2016. A Little Book of R For Time Series. e-book. 77 s.
- DETTLING, M. 2013. Applied Time Series Analysis.
- HYNDMAN, R. J. 2014. Forecasting: Principles & Practice University of Western Australia.
- CHATTERJEE, S. ARIMA/SARIMA vs LSTM with Ensemble learning Insights for Time Series Data. 2019 Dostupné na: <https://towardsdatascience.com/arima-sarima-vs-lstm-with-ensemble-learning-insights-for-time-series-data-509a5d87f20a>
- CHEEKOTY, S. Implementing time series ARIMA. 2019 Dostupné na: <https://towardsdatascience.com/implementing-time-series-arima-38fd77fc1d79>
- LUBYOVÁ, M. – SODOMOVÁ, E. Sociálna štatistika. Bratislava: Vydavateľstvo EKONÓM, 2016. 138 s. ISBN 978-80-225-4354-5.
- LUBYOVÁ, M. – ŠTEFÁNIK, M. a kol. Trh práce na Slovensku 2016+. Bratislava Ekonomický ústav Slovenskej akadémie vied. 2015. 241 s. ISBN 978-80-7144-255-4.
- LUKÁČIK, M. – PEKÁR, J. Kointegračná analýza v ekonometrii. Bratislava: Katedra operačného výskumu a ekonometrie FHI EU. 2006 STN ISO 690. Dostupné na: <http://www.fhi.sk/files/katedry/kove/veda-vyskum/prace/2006/Lukacik-Pekar2006.pdf>
- LUKÁČIKOVÁ, A. – LUKÁČIK, M. Ekonometrické modelovanie s aplikáciami. Bratislava : Vydavateľstvo EKONÓM, 2008. ISBN 978-80-225-2614-2.
- MAKLIN, C. ARIMA Model Python Example – Time Series Forecasting. 2019 Dostupné na: <https://towardsdatascience.com/machine-learning-part-19-time-series-and-autoregressive-integrated-moving-average-model-arima-c1005347b0d7>

MFSR. Revízia výdavkov zamestnanosti a odmeňovania vo verejnej správe. 2018. Dostupné na: https://izp.sk/wp-content/uploads/2019/06/revizia_vydavkov_sprava.pdf

MPSVR SR. Evidencia uchádzačov o zamestnanie, 2020. Dostupné na: https://www.slovensko.sk/sk/agendy/agenda/_evidencia-uchadzacov-o-zamestna/

MPSVR SR. Zákonník práce 2020. Dostupné na: <https://www.epi.sk/zz/rocnik/2020>

PRABHAKARAN, S. ARIMA Model – Complete Guide to Time Series Forecasting in Python. 2019 Dostupné na: <https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/>

RUBLÍKOVÁ, E a kol. Analýza časových radov : zbierka príkladov. 2. vyd. Bratislava Vydavateľstvo EKONÓM, 2003. 188 s. ISBN 80-225-1748-8.

RUBLÍKOVÁ, E. Analýza časových radov. Bratislava : Iura Edition, 2007. 207 s. ISBN 978-80-8078-139-2.

RUBLÍKOVÁ, E – PRÍHODOVÁ, I. Analýza vybraných časových radov - ARIMA modely. Bratislava : Vydavateľstvo EKONÓM, 2008. 214 s. ISBN 978-80-225-2540-4.

SHUMWAY, R. H. – STOFFER, D. S. 2011 Time Series Analysis and Its Applications Vydavateľstvo: Springer New York Dordrecht Heidelberg London.

TREXIMA. NP Prognózy vývoja na trhu práce v SR II. Dostupné na: <https://www.employment.gov.sk/sk/praca-zamestnanost/podpora-zamestnanosti/np-prognozy-vyvoja-trhu-prace-sr-ii/>

UPADHYAY, R. Step-by-Step Graphic Guide to Forecasting through ARIMA Modeling using R. 2015 Dostupné na: <http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/>

ÚPSVaR, Projekt „pripravený na prácu“. Dostupné na: https://www.upsvr.gov.sk/media/medialne-spravy/projekt-pripraveny-na-pracu-umozni-rekvalifikaciu-20-000-uchadzacom-o-zamestnanie.html?page_id=806672

VINK, R. Algorithm Breakdown: AR, MA and ARIMA models. 2018 Dostupné na: <https://www.ritchievink.com/blog/2018/09/26/algorithm-breakdown-ar-ma-and-arima-models/>

ZUCHINI, W. – NENADIÉ, O. Time Series Analysis with R.