

# **Bakalářská práce**

Model časové řady počtu nezaměstnaných  
s ohledem na možné předpovědi

# Prohlášení

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 30. května 2013

.....

Podpis

## **Abstrakt:**

Tato bakalářská práce se zabývá časovou řadou počtu nezaměstnaných s ohledem na možné předpovědi. Po výběru vhodného předpovědního modelu pro časovou řadu počtu nezaměstnaných jsou zkonstruovány předpovědi pro rok 2013 pro jednotlivé kraje České republiky.

**Klíčová slova:** časová řada, dekompozice časových řad, předpověď

## **Abstract:**

The bachelor thesis focuses on the time series of the unemployed with regard to the probable predictions. The predictions of all the regions in the Czech Republic have been constructed for the year 2013 after selecting an appropriate prediction model of the time series of the unemployed.

**Key words:** time series, decomposition of the time series, prediction

# Obsah

Úvod .....	1
1 Metodiky pro měření nezaměstnanosti .....	2
1.1 Český statistický úřad .....	2
1.2 Ministerstvo práce a sociálních věcí .....	3
1.2.1 Registrovaná nezaměstnanost .....	3
1.2.2 Podíl nezaměstnaných osob .....	4
1.3 Rozdíly mezi metodikami .....	4
2 Časové řady .....	5
2.1 Typy časových řad .....	5
2.2 Problémy s údaji v časové řadě .....	6
3 Modelování časových řad .....	7
3.1 Dekompozice časových řad .....	7
3.2 Modelování trendové složky .....	9
3.2.1 Klasické přístupy .....	9
3.2.2 Adaptivní přístupy .....	10
3.2.2.1 Klouzavé průměry .....	10
3.2.2.2 Exponenciální vyrovnání .....	10
3.2.3 Jiné metody pro popis trendu .....	14
3.2.4 Volba vhodného trendu .....	16
3.3 Modelování sezónní složky .....	18
3.3.1 Model sezónnosti v aditivním tvaru dekompozice .....	18
3.3.2 Model sezónnosti v multiplikativním tvaru dekompozice .....	19
3.3.3 Sezónní očišťování .....	20
4 Extrapolace časových řad .....	20
4.1 Klasifikace extrapolačních předpovědí .....	21
4.2 Extrapolace u dekompozice časových řad .....	22
4.2.1 Extrapolace u klasických přístupů .....	22
4.2.2 Extrapolace u aditivních přístupů .....	23
4.2.3 Extrapolace zvolených metod z Boxovo-Jenkinsovy metodologie .....	24
4.3 Výběr vhodného předpovědního modelu .....	25

5	Časová řada počtu nezaměstnaných 2005-2012 .....	26
5.1	Konstrukce předpovědních modelů .....	26
5.1.1	Trend pomocí dvojitého exponenciálního vyrovnání .....	28
5.1.2	Autoregresní model.....	32
5.1.3	Trend pomocí Arima model .....	34
5.2	Porovnání předpovědních modelů .....	38
6	Předpovědi pro rok 2013 u jednotlivých krajů ČR .....	40
	Závěr .....	47
	Literatura .....	48
	Internetové zdroje.....	49
	Přílohy.....	50
	A Tištěné .....	50
	A.1 Předpovědi na rok 2013 pro jednotlivé kraje .....	50
	A.2 Předpovědi na rok 2013 – graficky znázorněné .....	52
	B Obsažené na CD.....	59

# Seznam tabulek

Tabulka 1.1: Porovnání měr nezaměstnanosti .....	4
Tabulka 3. 3.1: Orientační testy pro výběr trendové složky.....	17
Tabulka 5. 5.1: Hodnoty <i>M. S. E.</i> jednotlivých typů dekompozice .....	27
Tabulka 5. 5.2: Sezónní faktory multiplikativní dekompozice.....	27
Tabulka 5. 5.3: Předpovědi dvojitého exponenciálního vyrovnání .....	31
Tabulka 5. 5.4: Parametry autoregresního modelu.....	32
Tabulka 5. 5.5: Předpovědi autoregresního modelu.....	33
Tabulka 5. 5.6: Předpovědi modelu ARIMA (1, 1, 0).....	38
Tabulka 5. 5.7: Rozklad <i>M. S. E.</i> .....	39
Tabulka 6. 6.1: Předpovědi ČR pro rok 2013.....	41
Tabulka 6. 6.2: Bodová předpověď pracovní síly ČR .....	43
Tabulka 6. 6.3: 95% a 90% intervalová předpověď registrované nezaměstnanosti .....	43
Tabulka 6. 6.4: 80% a 60% intervalová předpověď registrované nezaměstnanosti.....	44
Tabulka 6. 6.5: Porovnání předpovědí pro ČR a součtu krajů.....	45
Tabulka A. 0.1: Předpovědi pro kraje.....	50

## Seznam obrázků

Obrázek 5.1: Porovnání sezónně očištěné řady s původní řadou .....	28
Obrázek 5.2: Regresní odhady počátečních parametrů z prvních šesti pozorování.....	29
Obrázek 5.3: Regresní odhady dvojitého exponenciálního vyrovnaní ze všech pozorování ....	30
Obrázek 5.4: Předpovědi dvojitého exponenciálního vyrovnaní .....	31
Obrázek 5.5: Předpovědi autoregresního modelu.....	33
Obrázek 5.6: ACF řady počtu neumístěných uchazečů.....	34
Obrázek 5.7: PACF řady počtu neumístěných uchazečů.....	34
Obrázek 5.8: ACF 1. diference řady počtu neumístěných uchazečů.....	35
Obrázek 5.9: PACF 1. diference řady počtu neumístěných uchazečů.....	35
Obrázek 5.10: ACF modelu ARIMA (1, 1, 0).....	36
Obrázek 5.11: PACF modelu ARIMA (1, 1, 0).....	36
Obrázek 5.12: Výsledky modelu ARIMA (1, 1, 0) z programu MINITAB .....	37
Obrázek 5.13: Předpovědi modelu ARIMA (1, 1, 0) .....	38
Obrázek 6.1: Kontrolní graf trendu počtu nezaměstnaných .....	40
Obrázek 6.2: Předpovědi ČR pro rok 2013.....	41
Obrázek 6.3: Pracovní síla ČR .....	42
Obrázek 6.4: Srovnání předpovědí součtu krajů a ČR .....	45
Obrázek A.1: Předpovědi pro kraj Praha .....	52
Obrázek A.2: Předpovědi pro Středočeský kraj .....	52
Obrázek A.3: Předpovědi pro Jihočeský kraj.....	53
Obrázek A.4: Předpovědi pro Plzeňský kraj .....	53
Obrázek A.5: Předpovědi pro Karlovarský kraj.....	54
Obrázek A.6: Předpovědi pro Ústecký kraj .....	54
Obrázek A.7: Předpovědi pro Liberecký kraj.....	55
Obrázek A.8: Předpovědi pro Královehradecký kraj.....	55
Obrázek A.9: Předpovědi pro Pardubický kraj.....	56
Obrázek A.10: Předpovědi pro kraj Vysočina .....	56
Obrázek A.11: Předpovědi pro Jihomoravský kraj.....	57
Obrázek A.12: Předpovědi pro Zlínský kraj.....	57
Obrázek A.13: Předpovědi pro Olomoucký kraj .....	58
Obrázek A.14: Předpovědi pro Moravskoslezský kraj .....	58

# Úvod

Cílem této bakalářské práce je predikce časových řad počtu nezaměstnaných jednotlivých krajů a celé České republiky na cca jeden rok dopředu. Práce se skládá z šesti částí. V první části se seznámíme s jednotlivými metodikami, které se používají pro měření míry nezaměstnanosti v České republice. V druhé části popíšeme pojem časové řady, jejich klasifikace a možné problémy. Třetí část popisuje modelování časových řad, které je zaměřeno na metodu dekompozice časových řad. V čtvrté části řešíme problém extrapolace časových řad a výběr vhodného předpovědního modelu. V páté a šesté části se nachází praktická část bakalářské práce, kde je vybrán vhodný předpovědní model a následně jsou zkonstruovány předpovědi časových řad počtu nezaměstnaných jednotlivých krajů na rok 2013.



# 1 Metodiky pro měření nezaměstnanosti

V České republice jsme se setkávali do roku 2013 s dvěma přístupy, jak měřit míru nezaměstnanosti. Zástupci těchto metodik jsou Český statistický úřad (ČSÚ), jenž publikuje obecnou míru nezaměstnanosti, a Ministerstvo práce a sociálních věcí (MPSV), jehož míra nezaměstnanosti se nazývá registrovaná nezaměstnanost. Avšak od 1. ledna 2013 MPSV nahradilo registrovanou nezaměstnanost za nový ukazatel, který se nazývá Podíl nezaměstnaných osob.

## 1.1 Český statistický úřad

ČSÚ vychází při měření míry nezaměstnanosti z definic Mezinárodní organizace práce (ILO) a z metodiky Eurostatu. K vyhodnocení situace na trhu práce používá Výběrové šetření pracovních sil (VŠPS). VŠPS probíhá kontinuálně (po referenčních týdnech) v náhodně vybraném vzorku domácností, které se nacházejí v 25 tisíců obvykle zahrnutých bytů do šetření na území celé České republiky (ČR). Předmětem šetření jsou všechny bydlící osoby v těchto domácnostech, které setrvávají, či mají v úmyslu zůstat na území ČR po dobu 1 roku a více, přičemž nezáleží, zdali mají zde pobyt trvalý, přechodný, dlouhodobý nebo nehlášený. Jelikož VŠPS probíhá pouze na zlomku určité části populace, jsou její výsledky pak převáženy na celkovou populaci ČR dle předběžné statistiky obyvatelstva podle území, pětiletých věkových skupin a pohlaví respondentů. Výsledky VŠPS jsou vždy průměrné hodnoty za sledované čtvrtletí (13 referenčních týdnů), které jsou publikovány nejpozději tři měsíce po skončení čtvrtletí (např. výsledky VŠPS za 4. čtvrtletí roku 2012 byly publikovány 2. 4. 2013). (ČSÚ, 2012a)

Pomocí VŠPS určuje ČSÚ počet nezaměstnaných a celkovou pracovní sílu, jenž slouží k sestrojení ukazatele míry nezaměstnanosti. Do počtu nezaměstnaných se zahrnují všechny osoby od patnácti let, které v referenčním týdnu splňovaly zároveň tyto 3 podmínky ILO:

- *nebyly zaměstnané,*
- *byly připraveny k nástupu do práce, tj., během referenčního období byly k dispozici okamžitě nebo nejpozději do 14 dnů pro výkon placeného zaměstnání nebo sebe zaměstnání,*
- *v průběhu posledních 4 týdnů hledaly aktivně práci (prostřednictvím úřadu práce nebo soukromé zprostředkovatelny práce, přímo v podnicích, využíváním inzerce, podnikáním kroků pro založení vlastní firmy, podáním žádosti o pracovní povolení a licence nebo jiným způsobem).*(ČSÚ, 2012c)

Mezi nezaměstnané zahrnujeme také osoby, které nesplňují poslední podmínku, ale pouze v případě, že novou práci již našly a jsou schopny do ní nastoupit nejpozději do 14 dnů.

Celková pracovní síla je dána součtem zaměstnaných a nezaměstnaných osob. Do skupiny zaměstnaných zařazujeme všechny osoby od patnácti let věku, které v průběhu referenčního

týdne byly odměněny platem, mzdou či jiným způsobem za svoji vyvinutou pracovní aktivitu. Měly by během tohoto období odpracovat alespoň 1 hodinu anebo mít formální vztah k zaměstnání. (ČSÚ, 2012c)

Nyní si můžeme definovat míru nezaměstnanosti ČSÚ. Obecná míra nezaměstnanosti ČSÚ je vyjádřena jako podíl počtu nezaměstnaných na celkové pracovní síle (v procentech). (ČSÚ, 2012b)

## **1.2 Ministerstvo práce a sociálních věcí**

### **1.2.1 Registrovaná nezaměstnanost**

Na rozdíl od ČSÚ metodika MPSV je postavena na základě počtu nezaměstnaných osob registrovaných na úřadech práce, a proto její míra nezaměstnanosti se nazývá registrovaná nezaměstnanost. Tato míra nezaměstnanosti je definována jako podíl počtu dosažitelných neumístěných uchazečů o zaměstnání<sup>1</sup> na celkové pracovní síle. Zde celková pracovní síla se skládá ze součtu 3 skupin: počtu zaměstnaných z VŠPS, počtu dosažitelných neumístěných uchazečů o zaměstnání a počtu cizinců podle evidence MPSV a Ministerstva průmyslu a obchodu.<sup>2</sup>(ČSÚ, 2012b)

Pod pojmem dosažitelní neumístění uchazeči se skrývají osoby, kterým nebrání žádná objektivní překážka k bezprostřednímu nastoupení do nového zaměstnání. Za objektivní překážku je například považováno vzetí do vazby, být ve výkonu trestu, pobírání peněžité pomoci v mateřství a hmotného zabezpečení po dobu mateřské dovolené, pracovní neschopnost, zařazení do rekvalifikačních kursů nebo vykonávání krátkodobého zaměstnání. (MPSV, 2004)

Výše zmíněná metodika byla zavedena v červenci roku 2004 jako následek vstupu ČR do Evropské Unie v květnu téhož roku. Do výpočtu ukazatele míry nezaměstnanosti se kvůli této skutečnosti do pracovní síly a mezi uchazeče o zaměstnání začali zahrnovat i případní uchazeči o zaměstnání ze zemí Evropského hospodářského prostoru (tj. členské státy Evropské Unie, Norsko, Island, Lichtenštejnsko) a Švýcarska. (MPSV, 2004)

Pro srovnání zmíníme i starší metodiku, jež byla používána od roku 1997 a byla definována jako podíl přesného počtu registrovaných, neumístěných uchazečů o zaměstnání (občany ČR) vedených na úřadě práce v okrese jejich trvalého bydliště na pracovní síle. Pracovní síla v sobě zahrnovala počet zaměstnaných v národním hospodářství s jediným nebo hlavním pracovním poměrem a počet registrovaných, neumístěných uchazečů o zaměstnání. (MPSV, 2004)

---

<sup>1</sup> Tento pojem je vymezen v zákoně o zaměstnanosti 435/2004 Sb., avšak dochází zde k menší kolizi terminologie - zde pojmenován pouze jako uchazeč o zaměstnání.

<sup>2</sup> Jedná se o pracující cizince registrovaných na úřadech práce (evidence MPSV) nebo s platným povolením k zaměstnávání či živnostenským oprávněním (evidence MPO).

V porovnání se starou metodikou se změnil jak číselník, tak i jmenovatel u výpočtu registrované nezaměstnanosti. Nahrazením číselníku ve staré metodice za počet dosažitelných neumisťovaných uchazečů o zaměstnání lépe odpovídá definici nezaměstnaných pro výpočet obecné míry nezaměstnanosti podle metodiky ILO (ILO, 2011). Také v nové metodice máme větší hodnotu jmenovatele – pracovní síly, jež byla zvýšena o počet pracujících cizinců v ČR. Touto změnou v metodice MPSV byl snížen rozdíl mezi obecnou mírou nezaměstnanosti a registrovanou nezaměstnaností<sup>3</sup>.

## 1.2.2 Podíl nezaměstnaných osob

Od ledna 2013 registrovanou nezaměstnanost nahradil nový ukazatel Podíl nezaměstnaných osob. MPSV tak učinilo na základě dohody s ČSÚ s několika důvodů:

- *srovnávání nebo záměna registrované nezaměstnanosti MPSV a obecné míry nezaměstnanosti ČSÚ a jejich nesprávná interpretace,*
- *pracovní síla registrované nezaměstnanosti je tvořena kombinací z více zdrojů – dosažitelní uchazeči o zaměstnání z evidence úřadu práce, zaměstnané osoby z VŠPS, počet cizinců podle evidence MPSV a Ministerstva průmyslu a obchodu,*
- *údaje o zaměstnanosti z VŠPS na úrovni okresů nejsou dostatečně reprezentativní, detailní výsledky trpí vyšší chybivostí, pro nižší územní celky zcela chybí. (MPSV, 2012)*

Nově zavedený ukazatel Podíl nezaměstnaných osob je definován jako podíl dosažitelných uchazečů o zaměstnání ve věku 15 – 64 let ke všem obyvatelům ve stejném věku (MPSV, 2012). Rozdíl oproti registrované nezaměstnanosti nacházíme v pracovní síle, jež dříve zahrnovala pouze ekonomicky aktivní osoby. Nyní v novém ukazateli do pracovní síly patří všichni obyvatelé ve věku 15 – 64 let bez ohledu na jejich ekonomickou aktivitu.

## 1.3 Rozdíly mezi metodikami

Po představení metodik používaných k určení míry nezaměstnanosti vyvstává otázka, který z těchto přístupů lépe popisuje trh práce. Vývoj ukazatelů nezaměstnanosti je obdobný avšak v hodnotách se liší, jak můžeme sledovat v tabulce 1. 1.

**Tabulka 1.1: Porovnání měr nezaměstnanosti**

	1. čtvrt. 2012	2. čtvrt. 2012	3. čtvrt. 2012	4. čtvrt. 2012
<b>obecná míra nezaměstnanost</b>	7.1	6.7	7	7.2
<b>registrovaná nezaměstnanost</b>	8.9	8.1	8.4	9.4
<b>podíl nezaměstnaných osob</b>	7.0	6.4	6.6	7.4

Zdroj: ČSÚ, MPSV

<sup>3</sup> MPSV vydalo v roce 2004 výsledky registrované nezaměstnanosti pro obě metodiky. Podle staré metodiky v roce 2004 byla průměrná hodnota registrované nezaměstnanosti 10,2 % a podle nové metodiky pouze 9,2 %. Průměrná obecná míra nezaměstnanosti (ČSÚ) pro rok 2004 byla 8,3 %.

Příčiny diferencí mezi jednotlivými mírami nezaměstnanosti vyplývají z odlišného pojetí měření míry nezaměstnanosti. Obecná míra nezaměstnanosti se využívá k mezinárodnímu srovnání a registrovaná nezaměstnanost a Podíl nezaměstnaných osob odráží národní specifika trhu práce.

Rozdíly nacházíme už v samotném sběru dat, periodicitě publikování míry nezaměstnanosti, v rozdílné definici nezaměstnaných osob, které jsou pak poměřovány s pracovní silou, která se také liší náplní u jednotlivých ukazatelů nezaměstnanosti. Podrobnější informace o těchto rozdílech mezi jednotlivými metodikami, i o nich samotných, se můžeme dočíst více na internetových stránkách ČSÚ<sup>4</sup> a MPSV<sup>5</sup>.

## 2 Časové řady

Většina ekonomických ukazatelů je nám k dispozici v podobě chronologicky uspořádaných údajů, které jsou časově, věcně a prostorově srovnatelné. Tato posloupnost údajů uspořádaná v čase zpravidla ve směru minulost – přítomnost je tedy dána ve formě časové řady, kterou lze zapsat ve tvaru:

$$y_t, \quad t = 1, \dots, n,$$

kde  $y$  značí pozorovaný ukazatel,  $t = 1, \dots, n$  vyjadřuje časovou proměnnou a  $n$  je počet pozorování časové řady neboli také délka časové řady (Kozák, 1994).

### 2.1 Typy časových řad

Ekonomické časové řady lze rozdělit podle různých hledisek do jednotlivých typů. Podle časového hlediska rozhodného pro zjišťování údajů můžeme rozlišovat<sup>6</sup>:

- **časové řady intervalové (obratové, za období)** – údaje jsou zjišťovány za určitý interval (např. výnosy za měsíc),
- **časové řady okamžikové (stavové, k datu)** – údaje se vztahují k určitému okamžiku (např. počet dosažitelných nezaměstnaných k poslednímu dni v měsíci).

Podle periodicity, s jakou jsou údaje sledovány, rozlišujeme:

- **časové řady dlouhodobé** – údaje jsou sledovány v roční periodicitě,
- **časové řady krátkodobé** – údaje jsou zaznamenány ve čtvrtletních, měsíčních, týdenních aj. periodách.

---

<sup>4</sup> <http://www.czso.cz/>

<sup>5</sup> <http://www.mpsv.cz/cs/>, <http://portal.mpsv.cz/sz/stat>

<sup>6</sup> Klasifikace časových řad je čerpána z (Hindls, 2002).

Další členění může být podle druhu sledovaných ukazatelů, které nám klasifikuje:

- **časové řady původních hodnot,**
- **časové řady odvozených charakteristik** - časová řada jako funkce vycházející z časových řad původních hodnot.

Podle způsobu vyjádření ukazatelů rozlišujeme:

- **časové řady naturálních ukazatelů** – hodnoty ukazatele jsou vyjádřeny v naturálních jednotkách,
- **časové řady peněžních ukazatelů.**

## 2.2 Problémy s údaji v časové řadě<sup>7</sup>

Před analýzou časové řady bychom se měli ujistit, zda údaje v časové řadě jsou skutečně srovnatelné z časového<sup>8</sup>, věcného<sup>9</sup> a prostorového<sup>10</sup> hlediska.

Komplikace s časovou srovnatelností údajů nastává zejména u intervalových časových řad, kde hodnota závisí na délce intervalu. Intervaly by měly být ekvidistantní tj. stejně dlouhé, neboť by u případného srovnávání docházelo ke zkreslování. Problémy s různou délkou intervalu nám přináší kalendář např. ve smyslu rozdílné délky kalendářních měsíců, různý počet pracovních dní v měsíci, pohyblivé svátky aj.

Nejčastějším řešením pro zajištění časové srovnatelnosti je operace, která se nazývá očišťování časových řad od důsledku kalendářních variací. Operace k získání údajů očištěné na kalendářní dny je ve tvaru

$$y_t^{(0)} = y_t \frac{\bar{k}_t}{k_t},$$

kde  $y_t$  je hodnota očišťovaného ukazatele v příslušném dílčím období roku (měsíce, čtvrtletí)  $t$ ,  $k_t$  je „délka“ (počet kalendářních dní) příslušného dílčího období roku a  $\bar{k}_t$  je průměrná „délka“ (průměrný počet kalendářních dní) dílčího období.

U věcné srovnatelnosti bychom se měli přesvědčit, že v průběhu pozorování se nezměnil způsob zjišťování údajů anebo obsahové vymezení. Příkladem změny obsahového vymezení nám může posloužit změna používané metodiky k určení registrované nezaměstnanosti MPSV, která byla zavedena v červenci roku 2004 (MPSV, 2004). „Kompenzace“ vlivu rozdílů ve věcné srovnatelnosti“ se pak děje podle konkrétní povahy změny.

<sup>7</sup> Při zpracování této kapitoly jsem čerpala z knihy (Hindls, 2002).

<sup>8</sup> „Zpravidla se předpokládá, že časová řada je uspořádaná ekvidistantně, tj. že časová vzdálenost mezi sousedními pozorováními časové řady je shodná.“ (Kozák, 1994)

<sup>9</sup> „Věcnými znaky (hledisky) se rozumí věcná (metodická) náplň pojmu, která bývá zpravidla přesně vymezena určitou závaznou metodickou směrnici.“ (Kozák, 1975)

<sup>10</sup> Prostorové hledisko v sobě zahrnuje územní uspořádání, k němuž se empirická pozorování vztahují (Kozák, 1975).

Pod prostorově srovnatelnými údaji chápeme možnost používat údaje, vztahující se ke stejným geografickým územím. Ne vždy se musí jednat pouze o geografický problém, neboť může vzniknout také odlišný „ekonomický prostor“.

## 3 Modelování časových řad

Při modelování časových řad máme dva hlavní cíle. Prvním cílem je určit vhodný model (mechanismus), který popisuje dostatečně „dobře“ chování časové řady v období „interpolace“. Pokud tento model nalezneme, nabízí se ho použít ke konstrukci predikcí budoucích hodnot neboli k „extrapolaci“ časové řady, což je druhým cílem analýzy časových řad. Do základních přístupů modelování časových řad patří dekompozice časových řad, Boxovo-Jenkinsova metodologie, lineární dynamické modely a spektrální analýza<sup>11</sup>. Ve své práci se budu zabývat prvním uvedeným přístupem – dekompozicí časových řad.

### 3.1 Dekompozice časových řad<sup>12</sup>

Dekompozice časových řad vychází z myšlenky, že časové řady lze rozložit na několik složek, u kterých se nám podaří snadněji identifikovat chování složky než v původní nerozložené řadě. Časová řada se obvykle rozkládá na tyto složky:

- trendová složka (trend)  $T_t$ ,
- sezónní složka  $S_t$ ,
- cyklická složka  $C_t$ ,
- reziduální (náhodná) složka  $\varepsilon_t$ .

**Trendová složka** odráží hlavní tendenci dlouhodobého vývoje hodnot analyzovaného ukazatele. Může být rostoucí, klesající, nemonotónní nebo i konstantní.

**Sezónní složka** popisuje pravidelně se opakující změnu od trendu, která se vyskytuje v průběhu základního období. Příčiny sezónního kolísání můžeme najít především ve střídání ročních období anebo ve společenských zvyklostech (např. dle kalendáře).

**Cyklická složka** vyjadřuje kolísání okolo trendu, u kterého se střídají fáze růstu a poklesu. Má nepravidelný charakter tj. různou momentální periodu a amplitudu. Je podobná sezónní složce, ale na rozdíl od ní se obvykle projevuje v delším období. Někteří statistici cyklickou složku nepovažují za samostatnou složku a zahrnují ji do trendu (tzv. střednědobý trend).

**Reziduální (náhodnou) složku** získáváme po eliminaci trendové, sezónní a (případně) cyklické složky z časové řady. Zahrnuje v sobě náhodné chování, které nejsme schopni

---

<sup>11</sup> Všechny zmíněné přístupy jsou podrobně popsány například (Cipra, 1986).

<sup>12</sup> Při zpracování této kapitoly jsem čerpala z (Cipra, 2008) a (Kozák, 1994).

detailně popsat, chyby v měření údajů časové řady a také chyby, kterých se dopouštíme při vlastní analýze řady. Obvykle předpokládáme, že reziduální složka má charakter bílého šumu. Termín (zde) bílý šum obvykle značí posloupnost  $\{\varepsilon_t\}$  nezávislých náhodných veličin s nulovou střední hodnotou a konstantním (konečným) rozptylem  $\sigma^2 > 0$ . Někdy se dokonce předpokládá, že reziduální složka má normální rozdělení čili, že má charakter bílého normálního šumu.

Rozklad na všechny složky není nutný, záleží na charakteru zkoumaného ukazatele. Například neperiodické časové řady obsahují pouze trendovou a reziduální složku (nemusí mít sezónní nebo cyklickou složku), neboť v jejich chování se nemusí vyskytovat explicitní dílčí (viditelné, postižitelné) periodické kolísání tedy sezónní a cyklická složka. U krátkodobých časových řad zas předpokládáme zanedbání cyklické složky, neboť může mít periodu delší.

Dekompozici můžeme provést dvěma základními způsoby (a jejich modifikacemi). Buď uvažujeme, že časová řada je dána součtem jednotlivých složek anebo je to model časové řady, v němž jsou složky v součtovém tvaru. První způsob se nazývá **aditivní tvar**:

$$y_t = T_t + S_t + C_t + \varepsilon_t, \quad t = 1, 2, \dots, n,$$

kde jednotlivé složky jsou uvažovány ve svých skutečných absolutních hodnotách a měřeny ve stejných jednotkách jako původní časová řada  $y_t$ .

Druhý způsob je označován jako **multiplikativní tvar**:

$$y_t = T_t S_t C_t \varepsilon_t, \quad t = 1, 2, \dots, n,$$

kde většinou trendová složka je zachována ve své absolutní hodnotě, a ostatní složky jsou vůči ní uvažovány v relativních hodnotách. Trend je měřen v jednotkách časové řady  $y_t$  a ostatní složky jsou bezrozměrné.

Multiplikativní tvar snadno převedeme na aditivní pomocí logaritmické transformace:

$$\ln(y_t) = \ln(T_t) + \ln(S_t) + \ln(C_t) + \ln(\varepsilon_t) \quad t = 1, 2, \dots, n,$$

to vše za předpokladu, že  $T_t, S_t, C_t, \varepsilon_t > 0$ .

Můžeme se také setkat se způsobem dekompozice, v němž se hodnoty systematické a reziduální složky sčítají, kdežto jednotlivé elementy systematické složky (trend, sezónní složka, cyklická složka) jsou v součtovém tvaru. Takto provedená dekompozice je ve **smíšeném tvaru**, např.:

$$y_t = T_t S_t C_t + \varepsilon_t, \quad t = 1, 2, \dots, n,$$

kde trendová a reziduální složka je zachována ve svých měrných jednotkách a cyklická a sezónní složka je uvažována v relativních hodnotách vůči trendu.

## 3.2 Modelování trendové složky

Po seznámení se s jednotlivými složkami si můžeme představit časovou řadu jako mechanismus, kde podstatnou roli hraje trendová složka, proto se nejdříve budeme zabývat přístupy, jak můžeme trendovou složku eliminovat.

Existují dva základní přístupy k eliminaci trendu<sup>13</sup>:

- klasické přístupy (matematické, analytické přístupy)
- adaptivní přístupy

### 3.2.1 Klasické přístupy

Tyto přístupy v sobě zahrnují snahu popsat trend pomocí některých jednoduchých matematických funkcí<sup>14</sup>. Odhady parametrů těchto trendových funkcí získáme většinou pomocí metody nejmenších čtverců, která se používá v případě, že zvolená trendová funkce je lineární v parametrech. Jedná se tedy o lineární regresní model, kde v roli vysvětlující proměnné je čas. Výhodou metody nejmenších čtverců je to, že je metodika zpracována, obecně dostupná a dochází k minimalizaci „čtverce náhodné složky“.

Nejpoužívanější matematické funkce v ekonomické praxi jsou<sup>15</sup>:

- **lineární trend**

$$T_t = \beta_0 + \beta_1 t \quad t = 1, 2, \dots, n$$

- **parabolický trend**

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 \quad t = 1, 2, \dots, n$$

- **exponenciální trend**

$$T_t = \beta_0 \beta_1^t, \quad \text{kde } \beta_1 > 0, \quad t = 1, 2, \dots, n$$

Exponenciální trend není v parametrech lineární, proto ho převedeme pomocí linearizující logaritmické transformace na lineární trend (pro nezáporné časové řady):

$$\ln(T_t) = \ln(\beta_0) + t \ln(\beta_1).$$

Po použití metody nejmenších čtverců získáme odhady parametrických funkcí  $\ln(\hat{\beta}_0)$  a  $\ln(\hat{\beta}_1)$ , po jejichž odlogaritmováním získáme hodnoty odhadnutých parametrů  $\hat{\beta}_0$  a  $\hat{\beta}_1$ . Samozřejmě převod mění „měření kvality“.

- **modifikovaný exponenciální trend**

$$T_t = \xi + \beta_0 \beta_1^t, \quad \text{kde } \beta_1 > 0, \quad t = 1, 2, \dots, n$$

K odhadům parametrů můžeme použít několik metod – metodu částečných součtů, metodu dílčích průměrů nebo metodu vybraných bodů.

<sup>13</sup> Přístupy k eliminaci trendu převzaty z (Cipra, 2008).

<sup>14</sup> Tento pojem pro popsání trendu je převzatý z (Hindls, 2000).

<sup>15</sup> Převzato z (Hindls, 2002).



- **Gompertzova křivka**

$$T_t = \xi \beta_0^{\beta_1^t} \quad t = 1, 2, \dots, n$$

Základem postupu k odhadu parametrů je logaritmická transformace

$$\ln(T_t) = \ln(\xi) + \beta_1^t \ln(\beta_0).$$

Poté aplikujeme metodu částečných součtů.

- **Logistický trend** je vyjadřován v různých tvarech. Nejčastěji je prezentován ve tvaru

$$T_t = \frac{\xi}{1 + \beta_0 \beta_1^t} \quad \text{kde } \xi > 0, \beta_0 > 1, 0 < \beta_1 < 1, t = 1, 2, \dots, n.$$

Používané metody k odhadu parametrů jsou metoda částečných součtů, metoda vybraných bodů a metoda diferenčních odhadů parametrů.

### 3.2.2 Adaptivní přístupy

Na rozdíl od předcházejícího způsobu eliminace trendu, kdy byl trend popsán pomocí jedné matematické funkce v celém období „interpolace“, v níž tak parametry zůstávaly neměnné v čase, přicházejí adaptivní přístupy s konstrukcí trendové složky, v které se nepředpokládá stabilita analytického tvaru ani parametrů v čase, tedy jsou to modely s časově proměnlivými parametry. Měnivost parametrů v čase je docílena předpokladem, že v krátkých úsecích časové řady je vyrovnání pomocí matematických funkcí možné, přestože se obvykle matematické funkce v jednotlivých úsecích liší v parametrech. Nejznámější adaptivní přístupy jsou exponenciální vyrovnání a metoda klouzavých průměrů.

#### 3.2.2.1 Klouzavé průměry<sup>16</sup>

Podstatou metody klouzavých průměrů je nahrazení původních hodnot časové řady  $y_t$  pro  $t = 1, 2, \dots, n$  vypočtenými klouzavými průměry. Klouzavé průměry spočítáme z kratších úseků o délce  $m = 2p + 1$  hodnot, na které byla původní časová řada rozdělena. Tyto kratší úseky postupně (vždy o jednu hodnotu) posouváme směrem od začátku do konce časové řady, a tak postupně získáváme posloupnost klouzavých průměrů. Důležitou roli v této metodě hraje stanovení délky úseku, ze kterého jsou spočteny klouzavé průměry. Obvykle ji volíme na základě věcného charakteru ekonomické jevu. V praxi jsou většinou u neperiodických časových řad voleny klouzavé části menší délky, např.  $m = 5, 7, 9$ . U periodických časových řad se délka klouzavé části shoduje s délkou periody (sezóny).

Výpočetně jednoduchým typem klouzavých průměrů jsou **prosté klouzavé průměry** (Cipra, 1986). Prosté klouzavé průměry získáme vyrovnáním jednotlivých klouzavých částí o délce  $m = 2p + 1$  pro  $m < n$  a  $p \geq 1$  lineárním trendem s využitím metody nejmenších čtverců pro odhad parametrů lineárního trendu. Výsledkem je vzorec pro vyrovnané hodnoty časové řady v podobě aritmetického průměrů

<sup>16</sup> Při zpracování této kapitoly jsem čerpala z knihy (Hindls, 2000).

$$\bar{y}_t = \frac{1}{2p+1} \sum_{i=-p}^p y_{t+i} = \frac{y_{t-p} + y_{t-p+1} + \dots + y_{t+p}}{2p+1},$$

kde  $t = p + 1, p + 2, \dots, n - p$  jsou střední body jednotlivých klouzavých částí. Přitom prvních a posledních  $p$  hodnot zůstává nevyrovnáno.

U periodických časových řad (zejména u krátkodobých časových řad) se setkáváme se situací, kdy volíme za délku klouzavé části sudé číslo. V tomto případě střední body klouzavých částí již nejsou celá čísla a není možné je přiřadit přímo k původním pozorováním dané časové řady. Tato problematika se někdy řeší pomocí centrovaných klouzavých průměrů ve tvaru

$$\bar{y}_t = \frac{1}{4p} (y_{t-p} + 2y_{t-p+1} + \dots + 2y_{t+p-1} + y_{t+p}).$$

Kromě prostých a centrovaných klouzavých průměrů se používají složitější **vážené klouzavé průměry**, jež se získávají vyrovnáním klouzavé části polynomickým trendem s řádem vyšším než 1. Opět obvykle použijeme metodu nejmenších čtverců, obdržíme poměrně složité vzorce pro výpočet vyrovnaných hodnot. Jelikož se málokdy v praxi takové klouzavé průměry používají, tak se jimi zde nebudeme dále zabývat. Více se o nich můžeme dozvědět v literatuře, zabývající se dekompozicí časových řad, např. (Cipra, 1986).

### 3.2.2.2 Exponenciální vyrovnání<sup>17</sup>

Hlavním aspektem metody exponenciálního vyrovnání je uvažování vlivu „stáří“ pozorování. Vyrovnaná hodnota pomocí metody exponenciálního vyrovnání vychází ze všech předchozích pozorování, přičemž jejich váha ( $w$ ) do minulosti klesá podle exponenciální funkce:

$$w_k = \alpha^k, \quad 0 < \alpha < 1, \quad k = 1, 2, \dots, n - 1,$$

kde  $\alpha$  se nazývá vyrovnávající konstanta a  $k$  interpretuje „stáří“ pozorování od časového okamžiku  $n$ , který představuje pozorování v přítomném čase. Předpokládá se tedy, že pozorování bližší časovému okamžiku  $n$ , budou mít větší vliv na budoucí vývoj časové řady než pozorování ze vzdálenější minulosti.

Odhady parametrů modelu exponenciálního vyrovnání získáme někdy pomocí metody nejmenších čtverců

$$\sum_{k=0}^{n-1} (y_{n-k} - T_{n-k})^2 \alpha^k \rightarrow \min,$$

kde  $T_{n-k}$  představuje hodnotu trendové složky v čase  $n - k$ .

Pokud trendovou složku můžeme považovat v krátkých úsecích časové řady za konstantní, používáme tzv. **jednoduché exponenciální vyrovnání**. Jestliže by měly tyto krátké úseky

<sup>17</sup> Při zpracování této kapitoly jsem čerpal z knihy (Hindls, 2000) a (Cipra, 1986).

přibližně lineární trend, půjde o **dvojitě exponenciální vyrovnání** a pokud lze trend považovat zhruba za kvadratický, využívá se **trojitě exponenciální vyrovnání**. Naším úkolem je tedy nalézt odhady parametrů  $\beta_0, \beta_1, \dots, \beta_k$  podle zvoleného typu exponenciálního vyrovnání. Tyto odhady budou záviset na okamžiku jejich sestavení, neboť jsou konstruované ze všech dostupných dat. Jelikož bychom po každém přidání nového pozorování museli nalézt nové odhady těchto parametrů, což by bylo výpočetně náročné, používají se k vlastnímu vyrovnání tzv. „vyrovnávací statistiky“, které vycházejí pouze z počátečních odhadů parametrů  $\beta_0, \beta_1, \dots, \beta_k$ . Pomocí těchto „vyrovnávacích statistik“ počítáme vyrovnané hodnoty a předpovědi časové řady rekurentně.

Důležitou roli při konstrukci této metody hraje nalezení nejlepší vyrovnávací konstanty  $\alpha$ . Tento problém se řeší např. pomocí simulací modelu pro různé  $\alpha$ . Poté vybereme nejlepší hodnotu vyrovnací konstanty  $\alpha$ , u které nalezneme nejmenší hodnotu zvoleného kritéria např. střední čtvercová chyba modelu (M. S. E.). Pokud hodnota vyrovnací konstanty  $\alpha$  bude blízká jedničce, vliv minulých pozorování bude slábnout pozvolna. Naopak pokud hodnota vyrovnací konstanty  $\alpha$  bude blízká nule, vliv minulých pozorování bude slábnout velmi rychle.

Nyní si popíšeme obecný postup při exponenciálním vyrovnání:

- (1) Z prvních  $n_1 = 6$  hodnot, jak uvádí monografie (Cipra, 1986), spočteme počáteční odhady parametrů  $\beta_0, \beta_1, \dots, \beta_k$  příslušného trendu (parametr  $\beta_0$  pro konstantní trend u jednoduchého exponenciálního vyrovnání, parametry  $\beta_0, \beta_1$  pro lineární trend u dvojitě exponenciálního vyrovnání a parametry  $\beta_0, \beta_1, \beta_2$  pro kvadratický trend u trojitě exponenciálního vyrovnání).
- (2) Pomocí odhadů těchto parametrů a námi zvolené vyrovnávací konstanty  $\alpha$  spočteme počáteční hodnoty příslušných vyrovnávacích statistik.
- (3) Pomocí vyrovnávacích statistik rekurentně spočteme vyrovnané hodnoty časové řady.
- (4) Nalezneme nejlepší vyrovnávací konstantu  $\alpha$  pomocí například minimalizace střední čtvercové chyby modelu.
- (5) Po nalezení nejlepší vyrovnávací konstanty  $\alpha$  provedeme celý tento vyrovnávací postup (bod 1 – 3) znovu s tím rozdílem, že počáteční odhady parametrů  $\beta_0, \beta_1, \dots, \beta_k$  příslušného trendu odhadneme ze všech pozorování.

Po seznámení se s postupem exponenciálního vyrovnání si uvedeme příslušející vzorce bez odvození ke každému typu vyrovnání. Odvození těchto vzorců můžeme nalézt např. v (Cipra, 1986).

### Jednoduché exponenciální vyrovnání

Trendovou složku považujeme v krátkých úsecích časové řady za konstantní:

$$T_{n-k} = \beta_0, \quad k = 0, 1, \dots, n - 1.$$

Počáteční odhad parametru  $\beta_0$  je ve tvaru

$$b_0 = \frac{1}{n_1} \sum_{t=1}^{n_1} y_t = \hat{y}_0$$

který u jednoduchého exponenciálního vyrovnání také představuje počáteční vyrovnávací statistiku, kterou budeme značit  $\hat{y}_0$ . Následně vyrovnané hodnoty získáme podle rekurentního vztahu

$$\hat{y}_{n-k} = (1 - \alpha)y_{n-k} + \alpha\hat{y}_{n-k-1}, \quad k = 0, 1, \dots, n - 1,$$

kde  $\hat{y}_{n-k}$  je vyrovnaná hodnota v čase  $n - k$ .

### Dvojitě exponenciální vyrovnání

Trendovou složku považujeme v krátkých úsecích časové řady za lineární:

$$T_{n-k} = \beta_0 - \beta_1 k, \quad k = 0, 1, \dots, n - 1.$$

Počáteční odhady parametrů  $\beta_0$  a  $\beta_1$  jsou regresní odhady přímky proložené prvními  $n_1$  pozorování. Z takto získaných odhadů vypočítáme počáteční hodnoty vyrovnávacích statistik

$$S_0 = b_0(0) - \frac{\alpha}{1 - \alpha} b_1(0),$$

$$S_0^{[2]} = b_0(0) - \frac{2\alpha}{1 - \alpha} b_1(0),$$

kde  $b_0(0)$  a  $b_1(0)$  jsou počáteční odhady parametrů  $\beta_0$  a  $\beta_1$ , počáteční hodnota vyrovnávací konstanty  $\alpha$  je námi zvolená. Vyrovnávací statistiky jsou v následujícím tvaru

$$S_{n-k} = (1 - \alpha)y_{n-k} + \alpha S_{n-k-1}, \quad k = 0, 1, \dots, n - 1,$$

$$S_{n-k}^{[2]} = (1 - \alpha)S_{n-k} + \alpha S_{n-k-1}^{[2]}, \quad k = 0, 1, \dots, n - 1,$$

kde  $S_{n-k}$  je jednoduchá vyrovnávací statistika a  $S_{n-k}^{[2]}$  je tzv. dvojitá vyrovnávací statistika. Vyrovnanou hodnotu potom vypočteme pomocí vztahu

$$\hat{y}_{n-k} = 2S_{n-k} - S_{n-k}^{[2]}, \quad k = 0, 1, \dots, n - 1.$$

### Trojité exponenciální vyrovnání

Trendovou složku považujeme v krátkých úsecích časové řady za kvadratickou:

$$T_{n-k} = \beta_0 + \beta_1 k + \beta_2 k^2, \quad k = 0, 1, \dots, n - 1.$$

Počáteční odhady parametrů  $\beta_0$ ,  $\beta_1$  a  $\beta_2$  jsou regresní odhady získané metodou nejmenších čtverců z prvních  $n_1$  hodnot, z nichž se vypočtou počáteční hodnoty vyrovnávacích statistik

$$S_0 = b_0(0) - \frac{\alpha}{1-\alpha} b_1(0) + \frac{\alpha(1+\alpha)}{2(1-\alpha)^2} b_2(0),$$

$$S_0^{[2]} = b_0(0) - \frac{2\alpha}{1-\alpha} b_1(0) + \frac{2\alpha(1+2\alpha)}{2(1-\alpha)^2} b_2(0),$$

$$S_0^{[3]} = b_0(0) - \frac{3\alpha}{1-\alpha} b_1(0) + \frac{3\alpha(1+3\alpha)}{2(1-\alpha)^2} b_2(0),$$

kde  $b_0(0)$ ,  $b_1(0)$  a  $b_2(0)$  jsou počáteční odhady parametrů  $\beta_0$ ,  $\beta_1$  a  $\beta_2$ . Opět počáteční hodnota vyrovnávací konstanty  $\alpha$  je námi zvolená. Vedle jednoduché a dvojitě vyrovnávací statistiky tato metoda používá ještě navíc trojitou vyrovnávací statistiku ve tvaru

$$S_{n-k}^{[3]} = (1-\alpha)S_{n-k}^{[2]} + \alpha S_{n-k-1}^{[3]}, \quad k = 0, 1, \dots, n-1.$$

Vyrovnané hodnoty poté spočteme pomocí vztahu

$$\hat{y}_{n-k} = 3S_{n-k} - 3S_{n-k}^{[2]} + S_{n-k}^{[3]}, \quad k = 0, 1, \dots, n-1.$$

### 3.2.3 Jiné metody pro popis trendu

Při dekompozici časové řady můžeme kombinovat mezi sebou různé přístupy k jednotlivým složkám v časové řadě. Pro vyjádření trendu vyzkoušíme také některé z přístupu Boxovo-Jenkinsovy metodologie a to autoregresní procesy a model ARIMA.

Boxovo-Jenkinsova metodologie považuje za základní prvek modelování časové řady náhodnou složku. Základním požadavkem Boxovo-Jenkinsovy metodologie na časovou řadu je stacionarita. V analýze časových řad si vystačíme se slabou stacionaritou tzv. stacionární proces má konstantní střední hodnotu, konstantní rozptyl a jeho kovarianční struktura je v čase neměnná v rámci momentů do druhého řádu tj. pro každé  $s$  a  $t$  platí

$$\text{cov}(y_s, y_t) = \text{cov}(y_{s+h}, y_{t+h})$$

pro libovolné  $h$  (Cipra, 1986). Časové řady se většinou stacionarizují pomocí první či druhé difference.

#### Autoregresní model

Autoregresní model je motivován Box-Jenkinsovou metodologií, který lze řešit pomocí metody nejmenších čtverců. Trend pomocí autoregresního modelu řádu  $p$  můžeme zapsat ve tvaru<sup>18</sup>:

$$T_t = a_0 + \sum_{i=1}^p a_i T_{t-i}$$

Pomocí metody nejmenších čtverců řešíme soustavu  $Ax = b$ :  $x = (A^T A)^{-1} A^T b$

---

<sup>18</sup> Převzato z (Enders, 1995)

$$\text{kde } A = \begin{pmatrix} 1T_{p-1} & T_{p-2} \cdots T_0 \\ 1T_{p-2} & T_{p-3} \cdots T_1 \\ \vdots & \vdots \\ 1T_{t-1} & T_{t-2} \cdots T_{t-p} \end{pmatrix}, \quad x = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix}, \quad b = \begin{pmatrix} T_p \\ T_{p+1} \\ \vdots \\ T_t \end{pmatrix}.$$

Takto zapsaný proces se řeší pomocí mnohonásobné lineární regrese, předpokládáme, že vysvětlující proměnné jsou vzájemně lineárně nezávislé. Pokud tento předpoklad není splněn, hovoří se o tzv. multikolinearitě, jež způsobuje nadhodnocení směrodatných chyb odhadů parametrů  $a_0, \dots, a_p$ . To má pak za následek zkreslení výsledků testů vztahujících se k těmto parametrům. (Kozák, 1994)

### Model ARIMA<sup>19</sup>

Pro nestacionární řady, které lze převést na stacionární pomocí diferencování, jsou v rámci Box-Jenkinsovy metodologie určeny procesy typu ARIMA. Model ARIMA má tři parametry  $p, d, q$ . Skládá se z autoregresního procesu řádu  $p$  (značí se AR ( $p$ )), z integrovaného procesu řádu  $d$  (značí se I( $d$ ) - parametr  $d$  představuje řád použité diference, abychom časovou řadu stacionarizovali) a z procesu klouzavých průměrů řádu  $q$  (značí se MA ( $q$ )). Model lze zapsat ve tvaru

$$\Phi_p(B)(1 - B)^d T_t = \Theta_q(B)e_t$$

kde  $\Phi_p(B)$  značí (operátorovou) část autoregresního procesu řádu  $p$ ,  $(1 - B)^d$  představuje část integrovaného procesu řádu  $d$  a  $\Theta_q(B)$  značí část procesu klouzavých průměrů řádu  $q$  v modelu ARIMA (vše s využitím operátoru časového posuvu  $p$ . Pokud  $d = 0$  model ARIMA se redukuje na ARMA( $p, q$ ), pokud  $p = 0$  redukuje se na IMA( $d, q$ ) a v případě  $q = 0$  na ARI( $p, d$ ).

Identifikace modelu ARIMA probíhá ve třech krocích. V prvním kroku subjektivně zhodnotíme situaci z grafu. Posoudíme stacionaritu časové řady, výskyt odlehlých pozorování. Určíme řád diference ke stacionarizaci časové řady. Pak vypočteme autokorelační funkci (ACF) a parciální autokorelační funkci (PACF):

#### autokorelační funkce:

$$\rho_k = \frac{\text{cov}(T_t, T_{t-k})}{\sqrt{D(T_t)}\sqrt{D(T_{t-k})}} = \frac{\gamma_k}{\gamma_0},$$

kde  $\gamma_k = \text{cov}(T_t, T_{t-k})$  je autokovarianční funkce mezi veličinami  $T_t$  a  $T_{t-k}$ , a vzhledem ke stacionaritě procesu  $D(T_t) = D(T_{t-k}) = \gamma_0$ . ACF je symetrická kolem  $k = 0$ , proto se pro její popis můžeme omezit na  $k \geq 0$ . Nabývá hodnot v intervalu  $\langle -1; 1 \rangle$  a její graf se nazývá korelogram.

<sup>19</sup> Při zpracování této kapitoly jsem čerpala (Kozák, 1994).

**parciální autokorelační funkce:**

$$\rho_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & \rho_1 & 1 \end{vmatrix}},$$

kde  $||$  označuje determinant matice. Parciální autokorelace nám dává informaci o korelaci veličin  $y_t$  a  $y_{t-k}$  očištěné o vliv veličin ležících mezi nimi. Odvození PACF nalezneme například v monografii (Arlt, 1999).

Z grafu ACF a PACF si můžeme ověřit, zda je řada diferencovaná dostatečně a následně použít pro identifikaci parametru  $p$  a  $q$  (pro identifikaci modelu AR a MA), která je založena na principu porovnávání odhadnutých ACF a PACF s teoretickými ACF a PACF.

Po identifikaci modelu následuje jeho ověřování, které je založeno na testech, které zkoumají, zda jsou rezidua nekorelovaná a následně testování významnosti parametrů modelu pomocí t-testu. Boxova-Jenkinsova metodologie má softwarovou podobu, která obsahuje jak výpočty ACF a PACF tak i potřebné testy na ověřování modelu, jenž je dnes běžně dostupná (STATISTICA, STRATGRAPHICS jiné).

### 3.2.4 Volba vhodného trendu

Následně se objevuje otázka, jak zvolit vhodný model trendové složky anebo jak se rozhodnout mezi více modely, tj. který lépe popisuje trendovou složku.

Někdy nám napoví ekonomický ukazatel a hlavně jeho ekonomická podstata, od kterého máme z ekonomického hlediska nějaké očekávání ohledně jeho chování. Při věcně ekonomické analýze můžeme posoudit, o jakou funkci jde.

Další jednoduchá možnost volby trendu je na základě analýzy grafu časové řady. Zde hrozí jen nebezpečí subjektivity, neboť na základě grafické analýzy stejné časové řady může každý z nás dojít k různým závěrům o volbě vhodného typu trendové složky. Pro výběr vhodné matematické funkce u klasických přístupů existují jednoduché orientační testy, které vycházejí z charakteristických vlastností jednotlivých matematických funkcí. Následující tabulka 3. 1 zachycuje přehled těchto testů příslušejících k jednotlivým matematickým funkcím.

**Tabulka 3. 3.1: Orientační testy pro výběr trendové složky**

Trend	Orientační test
lineární	první diference $y_{t+1} - y_t$ jsou přibližně konstantní
parabolický	druhé diference $y_{t+2} - 2y_{t+1} + y_t$ jsou přibližně konstantní
exponenciální	podíly sousedních hodnot $\frac{y_{t+1}}{y_t}$ (resp. první diference logaritmů $\ln(y_{t+1}) - \ln(y_t)$ ) jsou přibližně konstantní
Modifikovaný exponenciální	podíly sousedních prvních diferencí $\frac{y_{t+2} - y_{t+1}}{y_{t+1} - y_t}$ jsou přibližně konstantní
logistický	(1) histogram prvních diferencí $y_{t+1} - y_t$ má tvar hustoty $N(0, 1)$ ; (2) podíly sousedních prvních diferencí převrácených hodnot $\left(\frac{1}{y_{t+2}} - \frac{1}{y_{t+1}}\right) / \left(\frac{1}{y_{t+1}} - \frac{1}{y_t}\right)$ jsou přibližně konstantní
Gompertzův	podíly sousedních prvních diferencí zlogaritmovaných hodnot $\frac{\ln(y_{t+2}) - \ln(y_{t+1})}{\ln(y_{t+1}) - \ln(y_t)}$ jsou přibližně konstantní

Zdroj: Tomáš Cipra - Finanční Ekonometrie

Částečné odstranění subjektivity při rozhodování a možnost porovnání různých modelů trendu nám přinášejí statistická kritéria<sup>20</sup>:

- Residuální součet čtverců (Residual Sum of Squares):

$$RSS = \sum_{t=1}^n (y_t - T_t)^2$$

- střední chyba odhadu (Mean Error):

$$M.E. = \frac{\sum (y_t - T_t)}{n}$$

- střední čtvercová chyba odhadu (Mean Squared Error):

$$M.S.E. = \frac{\sum (y_t - T_t)^2}{n}$$

- střední absolutní chyba odhadu (Mean Absolute Error):

$$M.A.E. = \frac{\sum |y_t - T_t|}{n}$$

- střední procentní chyba odhadu (Mean Percentage Error):

$$M.P.E. = \sum_{t=1}^n \left( \frac{y_t - T_t}{y_t} \right) \frac{100}{n}$$

<sup>20</sup> Statistická kritéria převzata z (Hindls, 2000).



- střední absolutní procentní chyba odhadu (Mean Absolute Percentage Error):

$$M.A.P.E. = \sum_{t=1}^n \left( \frac{|y_t - T_t|}{y_t} \right) \frac{100}{n}$$

Uvedená kritéria jsou jakousi mírou „přilnavosti“ modelu ke skutečnosti, neboť vycházejí z porovnání součtu (průměru) čtverců odchylek empirických a teoretických hodnot. Při výběru vhodného kritéria bereme v úvahu metodu, kterou jsme využili k odhadnutí modelu. Lepším modelem se stává ten model, který má menší hodnotu zvoleného kritéria. Nejčastěji používaným kritériem je průměrná odchylka čtverců (M. S. E.) Zde je třeba poznamenat, že subjektivitu výběru „grafického tvaru“ nahrazujeme často subjektivitou výběru kritéria. Tato kritéria se používají jak u klasických přístupů, tak i u aditivních přístupů.

### 3.3 Modelování sezónní složky

V této kapitole se budeme věnovat eliminaci sezónní složky, která se objevuje v krátkodobých časových řadách, tj. údaje jsou zaznamenány ve čtvrtletních, měsíčních, týdenních a jiných periodách. Sezónní složku vytváří tzv. sezónní faktory  $I_1, I_2, \dots, I_r$  ( $r$  počet sezón, např.  $r = 12$  v případě měsíčního pozorování).

Také bychom si měli doplnit potřebnou symboliku pro zjišťování sezónnosti. Dříve námi zavedená posloupnost časové proměnné  $t = 1, \dots, n$  tvoří posloupnost po sobě jdoucích sezón v jednotlivých (např.) letech. Proto při modelování sezónní složky budeme používat pořadové označení čísel let jako  $i = 1, 2, \dots, m$ , přičemž každý rok obsahuje  $r$  sezón, kterým přísluší index  $j = 1, 2, \dots, r$ . Tento vztah lze zapsat:

$$t = t_{ij} = (i - 1)r + j \quad \text{pro } i = 1, 2, \dots, m, j = 1, 2, \dots, r.$$

Model tedy můžeme zapsat například ve tvaru aditivní dekompozice:

$$y_{ij} = T_{ij} + S_{ij} + \varepsilon_{ij}, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, r,$$

kde  $y_{ij}$  značí pozorovaný ukazatel v roce  $i$  a v sezóně  $j$ ,  $T_{ij}$  značí trend v roce  $i$  a v sezóně  $j$ ,  $S_{ij}$  představuje sezónní složku v roce  $i$  a v sezóně  $j$  a  $\varepsilon_{ij}$  je hodnota náhodné složky v roce  $i$  a v sezóně  $j$ .

#### 3.3.1 Model sezónnosti v aditivním tvaru dekompozice<sup>21</sup>

V tomto modelu předpokládáme, že sezónnost je pravidelná, takže sezónní faktory pro jednotlivé roky zůstávají konstantní. Model v aditivním tvaru dekompozice po nově zavedené symbolice můžeme zapsat ve tvaru:

$$y_{ij} = T_{ij} + S_{ij} + \varepsilon_{ij}, \quad i = 1, 2, \dots, m; j = 1, 2, \dots, r.$$

<sup>21</sup> zpracováno dle (Cipra, 2008)

Jelikož při modelování sezónní složky již známe trendovou složku, časovou řadu můžeme „trendově očistit“

$$y_{ij} - T_{ij} = S_{ij} + \varepsilon_{ij}, \quad i = 1, 2, \dots, m; \quad j = 1, 2, \dots, r.$$

Poté lze z „trendově očištěné“ řady vyjádřit sezónní faktor jako aritmetický průměr ze všech těch hodnot, které odpovídají j-tému měsíci v roce ( $i = 1, 2, \dots, m$ ):

$$I_j = \frac{1}{m} \sum_{i=1}^m (y_{ij} - T_{ij}), \quad j = 1, 2, \dots, r.$$

Vypočtené sezónní faktory by měly splňovat normalizační pravidlo:

$$\sum_{j=1}^r I_j = 0, \quad j = 1, 2, \dots, r,$$

tzv. sezónní výkyvy se v rámci roku kompenzují, takže jejich roční součet se rovná nule.

Jestliže normalizační pravidlo není splněno, provedeme jejich centrování odečtením jejich aritmetického průměru

$$I_{j(\text{centrované})} = I_j - \bar{I}.$$

Tento způsob konstrukce sezónních faktorů se nazývá model konstantní sezónnosti.

### 3.3.2 Model sezónnosti v multiplikatívním tvaru dekompozice<sup>22</sup>

Multiplikatívní tvar dekompozice po nově zavedené symbolice můžeme zapsat ve tvaru

$$y_{ij} = T_{ij} S_{ij} \varepsilon_{ij}, \quad i = 1, 2, \dots, m; \quad j = 1, 2, \dots, r.$$

Z multiplikatívního tvaru dekompozice můžeme sledovat, že sezónní složka bude ovlivňována vývojovými změnami v charakteru trendové složky, neboť není samostatná složka, jako je to u aditivního tvaru dekompozice. Proto se v tomto případě sezónní faktory budou přímo úměrně měnit s dosaženou úrovní trendu, což někdy lépe vystihuje některé ekonomické skutečnosti (při procentním, relativním měření). Model, který si nyní popíšeme, se nazývá model proporcionální sezónnosti.

Časovou řadu opět „trendově očistíme“

$$\frac{y_{ij}}{T_{ij}} = S_{ij} \varepsilon_{ij}, \quad i = 1, 2, \dots, m; \quad j = 1, 2, \dots, r.$$

---

<sup>22</sup> čerpáno z (Cipra, 2008)

Sezónní faktory odhadneme podobným způsobem jako u aditivní dekompozice

$$I_j = \frac{1}{m} \sum_{i=1}^m \frac{y_{ij}}{T_{ij}}, \quad j = 1, 2, \dots, r,$$

kteřé by měly splňovat normalizační pravidlo ve tvaru

$$\prod_{j=1}^r I_j = 1.$$

Pokud toto pravidlo nespĺňují, provedeme jejich centrování vydělením jejich geometrickým průměrem

$$I_{j(\text{centrované})} = \frac{I_j}{\sqrt[r]{I_1 \cdot \dots \cdot I_r}}$$

### 3.3.3 Sezónní očišťování

Při určování sezónních faktorů jsme časovou řadu „trendově očišťovali“ a nyní si ukážeme, jak lze časovou řadu očistit od sezónní složky. Sezónní očišťování se využívá k tomu, abychom mohli porovnávat pozorování v časové řadě uvnitř roku i tehdy pokud jsou ovlivněny sezónností.

Mechanismus sezónního očišťování probíhá ve třech krocích: výpočet klouzavých průměrů, určení sezónních faktorů a nakonec očištění řady od sezónních faktorů jejich odečtením z časové řady nebo vydělením, záleží na typu dekompozice.

Výpočet klouzavých průměrů se v tomto mechanismu používá díky jeho schopnosti výrazně eliminovat periodické složky, jejichž perioda nepřesahuje délku klouzavého průměru. Avšak metod pro sezónní očišťování je mnoho, od těch nejjednodušších až po ty velmi komplikované. Pro eliminaci sezónní složky se například také používají různé speciální klouzavé průměry (Hendersonovy filtry), regresní metody, Wintersovo exponenciální vyrovnání a další<sup>23</sup>.

## 4 Extrapolace časových řad

Po rozpoznání mechanismu chování časové řady v minulosti, se můžeme zaměřit na již dříve zmiňovaný druhý cíl modelování časových řad – „extrapolaci“. Extrapolační metody časových řad patří do kvantitativních předpovědních metod, jež provádějí předpovědi na základě objektivních matematicko-statistických postupů. V těchto metodách se předpokládá, že zákonitosti vývoje v minulosti a přítomnosti jsou přeneseny do budoucnosti. Tedy provádíme

<sup>23</sup> Uvedené další přístupy k sezónnímu očišťování časové řady najdeme např. v (Hindls, 2002), (Cipra, 1986).

jakousi extrapolaci (prodloužení) minulých a současných pozorování řady do budoucnosti. Místo adekvátního termínu extrapolace se ale často setkáváme z konvekčních důvodů s pojmenováním této problematiky jako předpověď (Cipra, 2008).

## 4.1 Klasifikace extrapoláčních předpovědí<sup>24</sup>

*„Předpověď je kvantitativním odhadem pravděpodobnosti budoucí hodnoty časové řady pomocí minulého i současného pozorování reprezentovaným odhadnutým modelem.“* (Hušek, 1998) Rozlišujeme **bodové předpovědi** a **intervalové předpovědi**. Bodová předpověď představuje náš odhad budoucí hodnoty dané časové řady, která je ale vždy zatížená určitou chybou. Proto se sestrojuje předpovědní interval, kde středem tohoto intervalu je právě bodová předpověď. Předpovědní interval je obdobou intervalu spolehlivosti odhadu. To znamená, že  $(1 - \alpha)100\%$  předpovědní interval spolehlivosti představuje interval, v kterém bude ležet budoucí hodnota  $y_{n+h}$  s pravděpodobností  $(1 - \alpha)100\%$ .

Podle způsobu provedení předpovědi klasifikujeme předpověď ve vzorku nebo předpověď mimo vzorku.

**Předpověď ve vzorku** je předpověď pro pozorování, která jsme použili pro konstrukci předpovědního modelu. V předvídané budoucnosti zůstávají v platnosti modelové předpoklady (toto předpokládáme) a tak můžeme očekávat dobré předpovědní výsledky (samozřejmě za platnosti uvedených předpokladů).

**Předpověď mimo vzorek** je naopak předpověď pro pozorování, která nebyla zahrnuta při konstrukci předpovědního modelu. Nezahrnutí nastalo z důvodu neznámé hodnoty těchto pozorování, neboť se v době konstrukce modelu jednalo o budoucí hodnoty anebo jsme záměrně pozorování odstranili ze vzorku pro následné posouzení předpovědní schopnosti modelu. Takto záměrně uměle odstraněná data se nazývají zatajený vzorek.

Podle počtu budoucích pozorování, která předpovídáme ve stejném čase, rozlišujeme jednotlivou předpověď a multipředpověď (viz dále).

**Jednotlivá předpověď** představuje předpověď zkonstruovanou pro jediné pozorování. Obvykle se tato konstrukce týká hned následujícího pozorování tj. v čase  $n$  pro čas  $n + 1$ , ale může to být také pro pozorování vzdálenější např. v čase  $n + 4$ .

**Multipředpověď** je předpověď pro více pozorování současně. Představuje tedy vektor jednotlivých předpovědí, které byly sestrojeny ve stejném čase.

---

<sup>24</sup> Není-li uvedeno jinak, čerpáno z (Cipra, 2008).

## 4.2 Extrapolace u dekompozice časových řad

Extrapolace u metody dekompozice časových řad se provádí prodloužením jednotlivých systematických složek do budoucnosti. Předpověď tedy bude součtem nebo součinem předpovědi sezónní složky a trendové složky, záleží na typu dekompozice. U sezónní složky jsme předpokládali, že sezónní faktory pro jednotlivé roky jsou konstantní, proto v předpovědním horizontu<sup>25</sup> zůstanou stejné jako v předchozím období. U trendové složky si představíme předpovědi u všech námi popsanými přístupy v předchozí kapitole.

### 4.2.1 Extrapolace u klasických přístupů<sup>26</sup>

- **Lineární trend**

Bodová předpověď:

$$T_P = b_0 + b_1 P \quad \text{pro } P > n,$$

kde  $T_P$  představuje předpověď budoucí hodnoty v čase  $P$ ,  $b_0$  a  $b_1$  jsou odhady parametrů předpovědního modelu a  $n$  je počátek predikce.

Odpovídající  $(1 - \alpha)100\%$  intervalová předpověď pro hodnotu  $T_P$ :

$$\left( T_P - t_{1-\frac{\alpha}{2}}(n-2)sf_P, T_P + t_{1-\frac{\alpha}{2}}(n-2)sf_P \right),$$

kde  $s = \sqrt{\frac{\sum_{t=1}^n y_t^2 - \frac{(\sum_{t=1}^n y_t)^2}{n}}{n-2}}$ ,  $f_P = \sqrt{1 + \frac{1}{n} + \frac{(P-\bar{t})^2}{\sum_{t=1}^n (t^2 - n\bar{t}^2)}}$ ,  $t_{1-\frac{\alpha}{2}}(n-2)$  je  $(1 - \frac{\alpha}{2})100\%$

kvantil  $t$  rozdělení o  $n - 2$  stupních volnosti.

- **Parabolický trend**

Bodová předpověď:

$$T_P = b_0 + b_1 P + b_2 P^2 \quad \text{pro } P > n,$$

kde  $T_P$  představuje předpověď budoucí hodnoty v čase  $P$ ,  $b_0$ ,  $b_1$  a  $b_2$  jsou odhady parametrů předpovědního modelu.

Odpovídající  $(1 - \alpha)100\%$  intervalová předpověď pro hodnotu  $T_P$ :

$$\left( T_P - t_{1-\frac{\alpha}{2}}(n-3)sf_P, T_P + t_{1-\frac{\alpha}{2}}(n-3)sf_P \right),$$

kde  $t_{1-\frac{\alpha}{2}}(n-3)$  je  $(1 - \frac{\alpha}{2})100\%$  kvantil  $t$  rozdělení o  $n - 3$  stupních volnosti,

$$s = \sqrt{\frac{\sum_{t=1}^n y_t^2 - \frac{(\sum_{t=1}^n y_t)^2}{n}}{n-3}}, \quad f_P = \sqrt{1 + (1PP^2)(X'X)^{-1}(1PP^2)'}, \quad X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{bmatrix} \quad \text{a} \quad X'$$

značí transpozici matice  $X$ .

- **Exponenciální trend**

Bodová předpověď:

$$T_P = b_0 b_1^P \quad \text{pro } P > n,$$

kde  $T_P$  představuje předpověď budoucí hodnoty v čase  $P$ ,  $b_0$  a  $b_1$  jsou odhady parametrů předpovědního modelu.

<sup>25</sup> Předpovědní horizont je délka období, na které se předpovědi počítají.

<sup>26</sup> Při zpracování této kapitoly jsem čerpala z (Hindls, 2000)

- **Modifikovaný exponenciální trend**

Bodová předpověď:

$$T_P = \xi + b_0 b_1^P \quad \text{pro } P > n,$$

kde  $T_P$  představuje předpověď budoucí hodnoty v čase  $P$ ,  $b_0$ ,  $b_1$ ,  $\xi$  jsou odhady parametrů předpovědního modelu.

- **Logistický trend**

Bodová předpověď:

$$T_P = \frac{\xi}{1 + b_0 b_1^P} \quad \text{pro } P > n,$$

kde  $T_P$  představuje předpověď budoucí hodnoty v čase  $P$ ,  $b_0$ ,  $b_1$ ,  $\xi$  jsou odhady parametrů předpovědního modelu.

- **Gompertzův trend**

Bodová předpověď:

$$T_P = \xi b_0^{b_1^P} \quad \text{pro } P > n,$$

kde  $T_P$  představuje předpověď budoucí hodnoty v čase  $P$ ,  $b_0$ ,  $b_1$ ,  $\xi$  jsou odhady parametrů předpovědního modelu.

## 4.2.2 Extrapolace u aditivních přístupů<sup>27</sup>

- **Klouzavé průměry**

Bodová předpověď jednoduchého klouzavého průměru:

$$y_{n+h}(n) = \frac{y_{n-2p} + y_{n-2p+1} + \dots + y_n}{2p + 1} \quad \text{pro } h > 0,$$

kde  $y_{n+h}$  představuje předpověď budoucí hodnoty v čase  $n + h$ , která byla konstruovaná v čase  $n$ .

- **Exponenciální vyrovnání**

**Jednoduché exponenciální vyrovnání:**

Bodová předpověď:

$$\hat{y}_{n+h}(n) = \hat{y}_n,$$

kde  $\hat{y}_{n+h}(n)$  je předpověď hodnoty  $y_{n+h}$  v čase  $n$  a  $\hat{y}_n$  označuje vyrovnanou hodnotu v čase  $n$ .

Předpovědní interval má tvar

$$\left( \hat{y}_{n+h} - u_{\left(1-\frac{p}{2}\right)} d\Delta(n), \quad \hat{y}_{n+h} + u_{\left(1-\frac{p}{2}\right)} d\Delta(n) \right),$$

kde  $u_{\left(1-\frac{p}{2}\right)}$  je  $\left(1 - \frac{p}{2}\right) \cdot 100\%$  kvantil normálního rozdělení  $N(0, 1)$ ,  $d = 1,25$ ,

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

**Dvojitě exponenciální vyrovnání:**

Bodová předpověď  $\hat{y}_{t+h}(t)$  budoucí hodnoty  $y_{t+h}$  konstruovaná v čase  $t$  o  $h$  kroků dopředu je ve tvaru:

<sup>27</sup> Převzato z (Cipra, 1986).

$$\hat{y}_{t+h}(t) = b_0(t) + b_1(t)h = \left(2 + \frac{(1-\alpha)h}{\alpha}\right)S_t - \left(1 + \frac{(1-\alpha)h}{\alpha}\right)S_t^{[2]},$$

neboť  $b_0(t) = 2S_t - S_t^{[2]}$  a  $b_1(t) = \frac{1-\alpha}{\alpha}(S_t - S_t^{[2]})$  pro  $t = 1, \dots, n$ .

Předpovědní interval má tvar

$$\left( \hat{y}_{t+h} - u_{\left(1-\frac{p}{2}\right)}d\Delta(n), \quad \hat{y}_{t+h} + u_{\left(1-\frac{p}{2}\right)}d\Delta(n) \right),$$

kde  $u_{\left(1-\frac{p}{2}\right)}$  je  $\left(1 - \frac{p}{2}\right) \cdot 100\%$  kvantil normálního rozdělení  $N(0, 1)$ ,

$$d = 1,25 \left\{ \frac{1 + \frac{1-\alpha}{(1+\alpha)^3} \left( (1+4\alpha+5\alpha^2) + 2(1-\alpha)(1+3\alpha)h + 2(1-\alpha)^2h^2 \right)}{1 + \frac{1-\alpha}{(1+\alpha)^3} \left( (1+4\alpha+5\alpha^2) + 2(1-\alpha)(1+3\alpha) + 2(1-\alpha)^2 \right)} \right\}^{\frac{1}{2}},$$

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

**Trojité exponenciální vyrovnání:**

Bodová předpověď v čase  $t$  o  $h$  kroků dopředu:

$$\begin{aligned} \hat{y}_{t+h} = & \frac{1}{2\alpha^2} \{ (6\alpha^2 + (1+5\alpha)(1-\alpha)h + (1-\alpha)^2h^2)S_t - \\ & - (6\alpha^2 + 2(1+4\alpha)(1-\alpha)h + 2(1-\alpha)^2h^2)S_t^{[2]} + \\ & + (2\alpha^2 + (1+3\alpha)(1-\alpha)h + (1-\alpha)^2h^2)S_t^{[3]} \}. \end{aligned}$$

### 4.2.3 Extrapolace zvolených metod z Boxovo-Jenkinsovy metodologie

- **Autoregresní proces pomocí metody nejmenších čtverců**

Bodová předpověď trendu o jeden krok dopředu:

$$\hat{T}_{n+1} = a_0 + \sum_{i=1}^p a_i T_{n-i},$$

kde  $\hat{T}_{n+1}$  označuje předpověď hodnoty  $T_{n+1}$ .

- **Model ARIMA**

Připomeňme si, že model ARIMA je zapsán ve tvaru

$$\Phi_p(B)(1-B)^d T_t = \Theta_q(B)\varepsilon_t,$$

který lze také zapsat jako

$$T_t^* = \Phi_1 T_{t-1}^* + \dots + \Phi_p T_{t-p}^* + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \dots - \Theta_q \varepsilon_{t-q},$$

kde  $T_t^* = (1-B)^d T_t$ .

Bodová předpověď o krok dopředu:

$$T_{n+1}^* = \Phi_1 T_n^* + \dots + \Phi_p T_{n-p+1}^* + \varepsilon_{n+1} - \Theta_1 \varepsilon_n - \dots - \Theta_q \varepsilon_{n-q+1},$$

Předpověď  $T_{n+1}^*$  je podmíněnou střední hodnotou  $E(T_{n+1}^* | T_1^*, \dots, T_n^*)$ , tj.

$$\begin{aligned} T_{n+1}^* = E(T_{n+1}^* | T_1^*, \dots, T_n^*) = & \hat{\Phi}_1 T_n^* + \dots + \hat{\Phi}_p T_{n-p+1}^* - \hat{\Theta}_1 \varepsilon_n - \dots - \\ & - \hat{\Theta}_q \varepsilon_{n-q+1}, \end{aligned}$$

kde  $e_n, e_{n-1}, \dots$  jsou odhadnutá rezidua,  $\hat{\Phi}_1, \dots, \hat{\Phi}_p, \hat{\Theta}_1, \dots, \hat{\Theta}_q$  jsou odhady parametrů. Také je třeba poznamenat, že  $E(\varepsilon_{n+h} | T_1^*, T_2^*, \dots) = 0$  pro  $h = 1, 2, \dots$ , takže odhadem veličiny  $e_{n+1}$  je nula.

### 4.3 Výběr vhodného předpovědního modelu

Pro ověření předpovědních vlastností modelu se využívá metoda zatajovaného vzorku, o které jsme se zmínili v klasifikaci předpovědí v předpovědi mimo vzorek. Z konce analyzované řady oddělíme určitou část pozorování, kterou nezahrneme při určování modelu časové řady v období „interpolace“. Poté tento model podrobíme zkoušce ze schopnosti předpovídat tak, že ho prodloužíme do budoucnosti (kterou nám právě pokrývá náš zatajovaný vzorek). Předpovědi získané pomocí modelu porovnáme se zatajovaným vzorkem, čímž určíme schopnost modelu, jak dobře dokáže extrapolovat data.

Ověření vhodnosti použité předpovědní techniky anebo porovnání různých předpovědních modelů můžeme provést na základě kritérií, která hodnotí kvalitu předpovědí. Nejčastěji používaná míra kvality pro předpovědi  $\hat{y}_{n+1}, \dots, \hat{y}_{n+h}$  je střední čtvercová chyba předpovědi (Mean Squared Error)<sup>28</sup>:

$$M.S.E. = \frac{\sum_{t=n+1}^{n+h} (y_t - \hat{y}_t)^2}{h} = \frac{1}{h} \sum_{t=n+1}^{n+h} \varepsilon_t^2,$$

Kritérium implicitně předpokládá znalost skutečných budoucích hodnot. Toto kritérium můžeme dále rozložit na součet tří složek<sup>29</sup>:

$$\frac{\sum_{t=n+1}^{n+h} (y_t - \hat{y}_t)^2}{h} = (\bar{\hat{y}} - \bar{y})^2 + (s_{\hat{y}} - s_y)^2 + 2(1 - r_{\hat{y}y}) \cdot s_{\hat{y}} \cdot s_y,$$

kde  $\bar{\hat{y}}, \bar{y}$  jsou příslušné průměry,  $s_{\hat{y}}, s_y$  jsou příslušné výběrové směrodatné odchylky,  $r_{\hat{y}y}$  je výběrový korelační koeficient mezi  $\hat{y}$  a  $y$ . První sčítanec

$$(\bar{\hat{y}} - \bar{y})^2 \geq 0$$

je chyba nestejně polohy předpovídaných a skutečných hodnot. Hodnota této chyby bude tím vyšší, čím se bude více lišit průměr předpovědí od průměru předpovídaných hodnot.

Druhý sčítanec

$$(s_{\hat{y}} - s_y)^2 \geq 0$$

je chyba nestejně variability předpovídaných a skutečných hodnot. Hodnota této chyby bude tím vyšší, čím více se bude lišit směrodatná odchylka předpovědí od směrodatné odchylky předpovídaných hodnot.

<sup>28</sup> Extrapoláční kritérium bylo převzato z (Cipra, 2008).

<sup>29</sup> Rozklad M. S. E. převzat (Kozák, 1975)



Třetí sčítanec

$$2(1 - r_{\hat{y}y}) \cdot s_{\hat{y}} \cdot s_y$$

zahrnuje v sobě chybu zbývající nesystematické části předpovědi, proto předpokládáme u dobré předpovědi, že tato chyba v ní bude výrazně převládat. Třetí sčítanec představuje chybu neúplné kovariance.

Jestliže výše uvedené sčítance vydělíme hodnotou M. S. E., dostáváme relativní chyby polohy, variability a neúplné kovariance. Dalším ukazatelem relativních chyb předpovědí je Theilův koeficient:

$$T^2 = \frac{\sum_{t=n+1}^{n+P} \left( \frac{\hat{y}_t - y_t}{y_t} \right)^2 \cdot y_t^2}{\sum_{t=n+1}^{n+P} y_t^2}$$

Koeficient opět implicitně předpokládá znalost skutečných budoucích hodnot. Theilův koeficient představuje váženou průměrnou čtvercovou relativní chybu předpovědi a jeho odmocnina značí průměrnou relativní chybu předpovědi bez ohledu na znaménko. Pokud se průměrná relativní chyba pohybuje v rozmezí 3 – 5 % je chyba předpovědi považována za malou a posuzovaný předpovědní model může být dobrým nástrojem pro konstrukci předpovědí.

Čím budou uvedená extrapoláčnÍ kritéria nižší, tím lepší předpovědní model získáváme. Zde však musíme zdůraznit, že tato kritéria vypovídají pouze o statistické přesnosti předpovědí, nikoli o ekonomické či finanční adekvátnosti.

## 5 Časová řada počtu nezaměstnaných 2005-2012

Pro konstrukci předpovědi časové řady počtu nezaměstnaných osob byla vybrána časová řada počtu dosažitelných neumístěných uchazečů o zaměstnání MPSV<sup>30</sup> kvůli její obsáhlosti (měsíční pozorování). Podle klasifikace časových řad uvedené v 2. kapitole je časová řada počtu dosažitelných neumístěných uchazečů o zaměstnání řadou okamžikovou, krátkodobou a původních hodnot. Kvůli věcné srovnatelnosti časové řady bylo zvoleno období pro modelování od roku 2005, neboť během roku 2004 se změnila metodická náplň tohoto ukazatele počtu nezaměstnaných osob (MPSV, 2004). Zdrojová data byla stažena ze stránek MPSV<sup>31</sup> a jsou k nahlédnutí v Excel souboru s názvem data.xlsx.

### 5.1 Konstrukce předpovědních modelů

Pro ověření předpovědních schopností modelů použijeme metodu zatajovaného vzorku (rok 2012) na časovou řadu počtu nezaměstnaných osob a sestrojíme předpověď mimo vzorek.

<sup>30</sup> Dále v textu řadu počtu dosažitelných neumístěných uchazečů o zaměstnání MPSV budeme zkráceně označovat za řadu počtu nezaměstnaných osob.

<sup>31</sup> <http://portal.mpsv.cz/sz/stat/nz/mes>

Nejdříve časovou řadu 2005 - 2011 o délce 84 pozorování sezónně očistíme pomocí centrovaných klouzavých průměrů délky 12 a tak eliminujeme sezónní složku. Vyzkoušeny byly všechny typy dekompozice, jejichž sestavení nalezneme v Excel souboru s názvem KP\_2005-2011.xlsx. Nejvýstižnější dekompozice je multiplikativní podle interpolačního kritéria M. S. E., jehož výsledky jsou uvedené v následující tabulce.

**Tabulka 5. 5.1:** Hodnoty M. S. E. jednotlivých typů dekompozice

	<b>MSE</b>
<b>aditivní</b>	146 836 070
<b>multiplikativní</b>	125 904 072
<b>smíšená</b>	130 513 868

Takto získané sezónní faktory v přirozených logaritmech jsou uvedené v následující tabulce.

**Tabulka 5. 5.2:** Sezónní faktory multiplikativní dekompozice

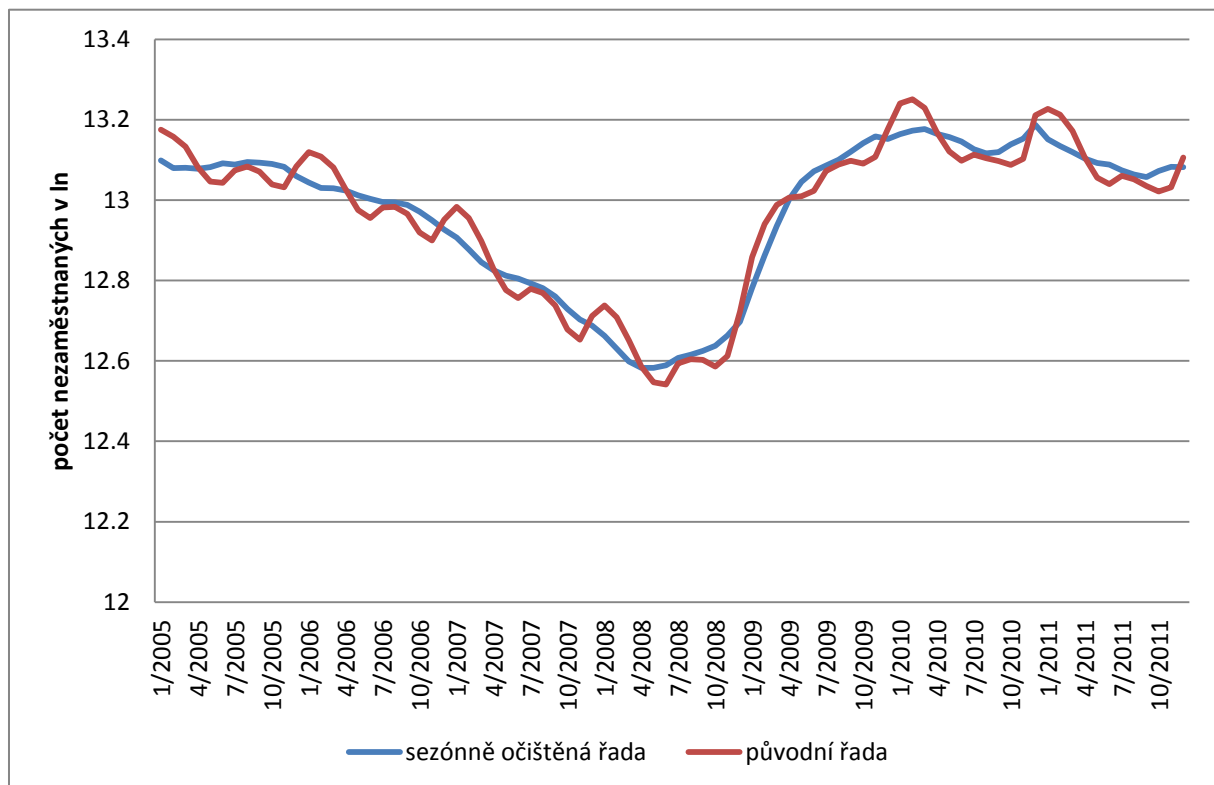
<b>měsíc</b>	<b>ln(sezónní faktory)</b>
<b>leden</b>	0.076
<b>únor</b>	0.078
<b>březen</b>	0.052
<b>duben</b>	0.003
<b>květen</b>	-0.036
<b>červen</b>	-0.048
<b>červenec</b>	-0.014
<b>srpen</b>	-0.011
<b>září</b>	-0.022
<b>říjen</b>	-0.051
<b>listopad</b>	-0.051
<b>prosinec</b>	0.024

Sezónní faktory jsou v přirozených logaritmech, neboť jsme si před sezónním očišťováním model v multiplikativní dekompozici převedli na aditivní pomocí logaritmické transformace:

$$\ln(y_t) = \ln(T_t) + \ln(S_t) + \ln(C_t) + \ln(\varepsilon_t) \quad t = 1, 2, \dots, n.$$

Porovnat původní časovou řadu a sezónně očištěnou řadu můžeme na následujícím obrázku:

**Obrázek 5.1:** Porovnání sezónně očištěné řady s původní řadou



Nyní zbývá eliminovat trendovou složku. Pro popis trendu jsme zvolili tři možnosti - dvojitě exponenciální vyrovnání z adaptivních přístupů, autoregresní procesy a model ARIMA z Box - Jenkinsovy metodologie. Klasické přístupy pro dlouhodobé předpovědi nejsou vhodné. Po eliminaci trendu sestrojíme bodovou a intervalovou předpověď pro zatajovaný vzorek (rok 2012). Bodová předpověď je součtem hodnot trendové a sezónní složky v budoucím okamžiku  $P$ , tj. ve tvaru

$$\ln(\hat{y}_P) = \ln(T_P) + \ln(S_P) \quad P > n,$$

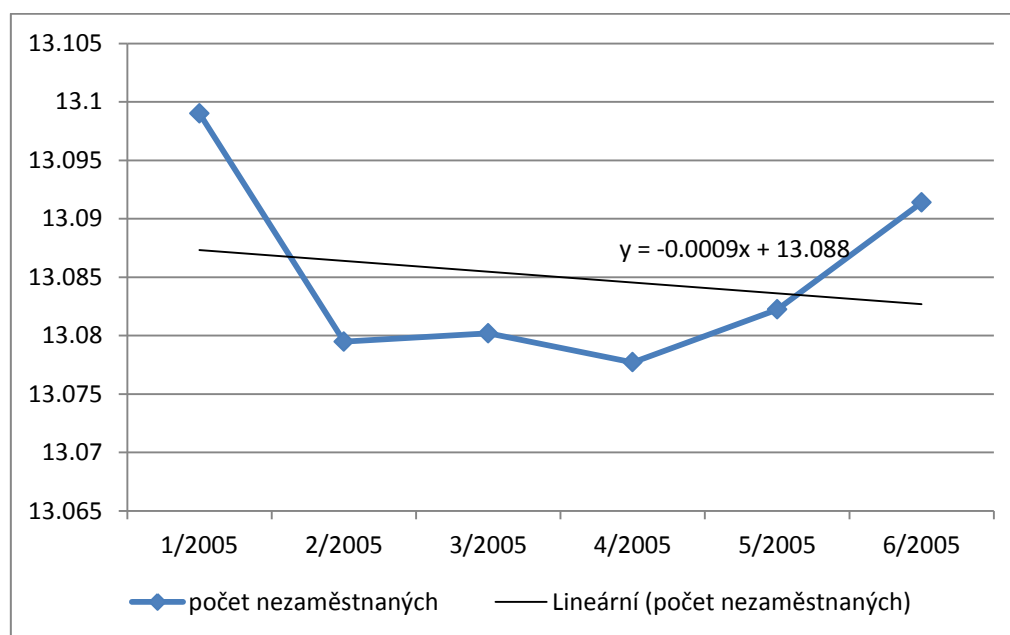
kterou následně odlogaritmujeme. Sezónní složku  $S_P$  již pro jednotlivé měsíce známe z tabulky 5. 1. 2, neboť sezónní faktory zůstávají pro jednotlivé roky konstantní.

### 5.1.1 Trend pomocí dvojitěho exponenciálního vyrovnání

Při konstrukci metody dvojitěho exponenciálního vyrovnání jsme postupovali podle obecného postupu, který byl popsán v kapitole 3.2.2.2 Exponenciální vyrovnání. Tato metoda byla provedena v prostředí Excelu.

- (1) Určili jsme počáteční odhady parametrů  $\beta_0$  a  $\beta_1$ , které jsou regresní odhady přímky proložené prvními  $n_1 = 6$  pozorování.

**Obrázek 5. 2:** Regresní odhady počátečních parametrů z prvních šesti pozorování



Z výše uvedeného obrázku získáváme počáteční odhady  $b_0$  a  $b_1$  parametrů  $\beta_0$  a  $\beta_1$ :

$$b_0(0) = 13,088, \quad b_1(0) = -0,0009.$$

(2) Vypočteme počáteční vyrovnávací statistiky, kde za počáteční  $\alpha$  jsme například zvolili hodnotu 0,7.

$$S_0 = b_0(0) - \frac{\alpha}{1-\alpha} b_1(0) = 13,088 - \frac{0,7}{1-0,7} (-0,0009) = 13,09,$$

$$S_0^{[2]} = b_0(0) - \frac{2\alpha}{1-\alpha} b_1(0) = 13,088 - \frac{2 \cdot 0,7}{1-0,7} (-0,0009) = 13,092.$$

(3) Pomocí vyrovnávacích statistik rekurentně spočteme vyrovnané hodnoty časové řady:

$$\hat{y}_{n-k} = 2S_{n-k} - S_{n-k}^{[2]}, \quad k = 0, 1, \dots, n-1, n = 1, \dots, 84,$$

a vyrovnané hodnoty předpovědí o 1 krok dopředu:

$$\hat{y}_{n-k}(n-k-1) = b_0(n-k-1) + b_1(n-k-1) = \left(2 + \frac{(1-\alpha)}{\alpha}\right) S_{n-k-1} - \left(1 + \frac{(1-\alpha)}{\alpha}\right) S_{n-k-1}^{[2]}, \quad k = 0, 1, \dots, n-1, n = 1, \dots, 84,$$

kde  $b_0(n-k-1) = 2S_{n-k-1} - S_{n-k-1}^{[2]}$  a  $b_1(n-k-1) = \frac{1-\alpha}{\alpha} (S_{n-k-1} - S_{n-k-1}^{[2]})$ .

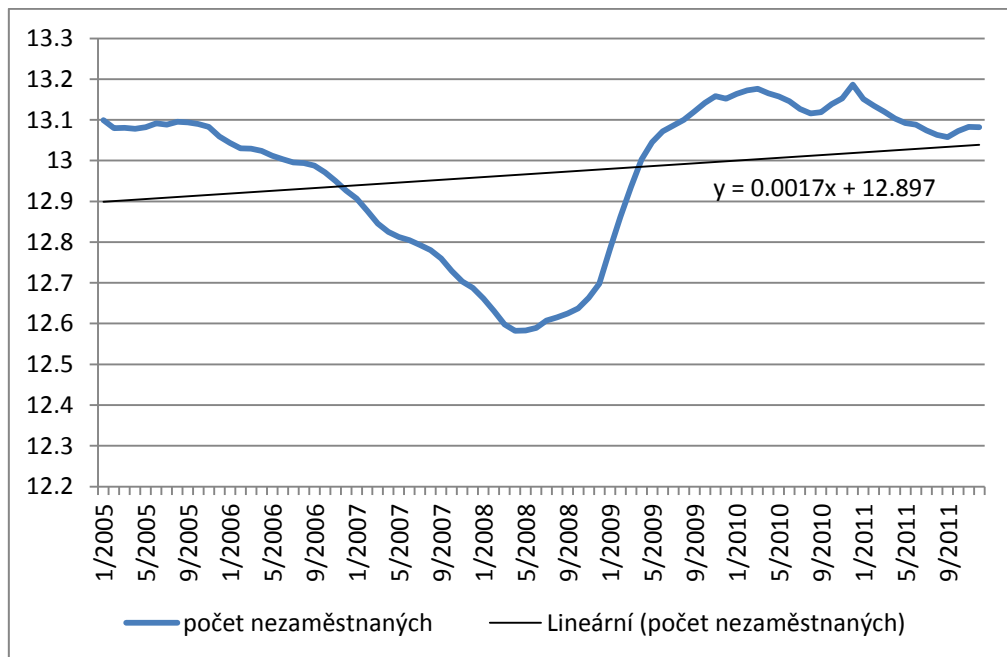
(4) Pro nalezení nejlepší vyrovnávací konstanty  $\alpha$  jsme použili v Excelu řešitel, kterým jsme minimalizovali námi zvolené kritérium M. A. E., zkonstruované pro období 2005 - 2011 (bez zatajovaného vzorku) díky změně parametru vyrovnávací konstanty  $\alpha$ . Kritérium M. A. E. je ve tvaru

$$M.A.E. = \frac{\sum |y_{n-k} - \hat{y}_{n-k}(n-k-1)|}{n}$$

Po spuštění řešitele v Excelu byla nalezena nejlepší vyrovnávací konstanta  $\alpha = 0,113$  pro minimalizovanou hodnotu kritéria  $M.A.E. = 0,009$ .

- (5) Nalezneme počáteční odhady parametrů  $\beta_0$  a  $\beta_1$ , které odhadneme tentokrát ze všech pozorování.

**Obrázek 5. 3:** Regresní odhady dvojitého exponenciálního vyrovnání ze všech pozorování



Z výše uvedeného obrázku můžeme vyčíst z rovnice přímky proležené všemi pozorování počáteční odhady  $b_0$  a  $b_1$  parametrů  $\beta_0$  a  $\beta_1$ :

$$b_0(0) = 12,897, \quad b_1(0) = 0,0017.$$

Poté vypočteme opět počáteční vyrovnávací statistiky a z nich rekurentně vyrovnané hodnoty.

Bodové předpovědi spočteme pomocí vzorce:

$$\hat{y}_{84+h} = b_0(84) + b_1(84)h, \quad h = 1, 2, \dots, 12,$$

kde  $b_0(84) = 2 S_{84} - S_{84}^{[2]}$  a  $b_1(84) = \frac{1-\alpha}{\alpha} (S_{84} - S_{84}^{[2]})$ . Intervalové předpovědi jsou pak ve tvaru

$$\left( \hat{y}_{84+h} - u_{\left(1-\frac{p}{2}\right)} d\Delta(n), \quad \hat{y}_{84+h} + u_{\left(1-\frac{p}{2}\right)} d\Delta(n) \right),$$

kde  $u_{\left(1-\frac{p}{2}\right)}$  je  $\left(1 - \frac{p}{2}\right) \cdot 100\%$  kvantil normálního rozdělení  $N(0, 1)$ ,

$$d = 1,25 \left\{ \frac{1 + \frac{1-\alpha}{(1+\alpha)^3} ((1+4\alpha+5\alpha^2) + 2(1-\alpha)(1+3\alpha)h + 2(1-\alpha)^2h^2)}{1 + \frac{1-\alpha}{(1+\alpha)^3} ((1+4\alpha+5\alpha^2) + 2(1-\alpha)(1+3\alpha) + 2(1-\alpha)^2)} \right\}^{\frac{1}{2}},$$

$$\Delta(n) = \sum_{t=1}^n \frac{|y_{n-k} - \hat{y}_{n-k}((n-k-1))|}{n} \text{ pro } n = 1, \dots, 84, k = 0, \dots, n-1.$$

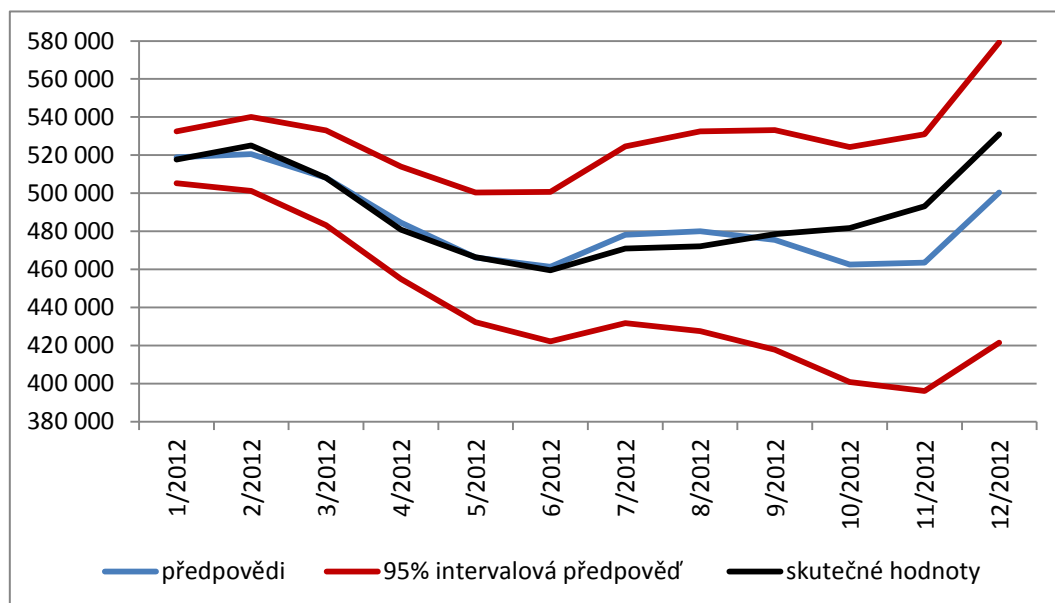
Za  $\alpha$  volíme hodnotu nalezené nejlepší vyrovnávací konstanty.

V následující tabulce se nacházejí bodové předpovědi a 95% intervalová předpověď pro zatajovaný vzorek 2012, které jsou také graficky znázorněny na obrázku 5.4.

**Tabulka 5. 5.3: Předpovědi dvojitého exponenciálního vyrovnání**

	předpovědi	dolní mez	horní mez	skutečné
<b>1/2012</b>	518 832	505 157	532 508	517 723
<b>2/2012</b>	520 601	501 164	540 039	525 119
<b>3/2012</b>	508 178	483 284	533 071	508 158
<b>4/2012</b>	484 535	455 025	514 044	480 818
<b>5/2012</b>	466 297	432 287	500 306	466 352
<b>6/2012</b>	461 409	422 178	500 639	459 497
<b>7/2012</b>	478 206	431 758	524 654	470 964
<b>8/2012</b>	480 056	427 615	532 497	472 120
<b>9/2012</b>	475 495	417 799	533 192	478 548
<b>10/2012</b>	462 556	400 840	524 272	481 737
<b>11/2012</b>	463 521	396 084	530 957	493 208
<b>12/2012</b>	500 370	421 548	579 191	530 994

**Obrázek 5. 4: Předpovědi dvojitého exponenciálního vyrovnání**



Všechny výše uvedené výpočty metody dvojitého exponenciálního vyrovnání nalezneme v Excelu s názvem dvojité\_vyrovnaní.xlsx.

## 5.1.2 Autoregresní model

Trend předpokládáme ve tvaru

$$T_t = a_0 + \sum_{i=1}^{p=12} a_i T_{t-i}$$

Pomocí metody nejmenších čtverců řešíme soustavu  $Ax = b$ :  $x = (A^T A)^{-1} A^T b$

$$\text{kde } A = \begin{pmatrix} 1T_{12-1} & T_{12-2} \cdots & T_0 \\ 1T_{12-2} & T_{12-3} \cdots & T_1 \\ \vdots & \vdots & \vdots \\ 1T_{84-1} & T_{84-2} \cdots & T_{84-12} \end{pmatrix}, \quad x = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{12} \end{pmatrix}, \quad b = \begin{pmatrix} T_{12} \\ T_{12+1} \\ \vdots \\ T_{84} \end{pmatrix}.$$

Tento model byl opět řešen v Excelu. K řešení této mnohonásobné regrese jsme použili doplněk Excelu – Analýza dat → Regrese a získali jsme odhady parametrů  $a_0, \dots, a_{12}$ :

**Tabulka 5. 5.4:** Parametry autoregresního modelu

<i>parametr</i>	<i>Odhad v ln</i>
<b>a0</b>	0.293
<b>a1</b>	1.765
<b>a2</b>	-0.617
<b>a3</b>	-0.348
<b>a4</b>	0.143
<b>a5</b>	0.048
<b>a6</b>	0.154
<b>a7</b>	-0.149
<b>a8</b>	-0.095
<b>a9</b>	0.106
<b>a10</b>	0.010
<b>a11</b>	-0.046
<b>a12</b>	0.006

Poté jsme zkonstruovali předpovědi a intervalové předpovědi pro rok 2012 podle následujících vzorců.

Předpovědi trendové složky byly zkonstruovány pomocí vzorce:

$$T_{n+1} = a_0 + \sum_{i=1}^{12} a_i T_{n-i}.$$

Intervalová předpověď byla zkonstruovaná pomocí vzorce:

$$\left( T_P - t_{1-\frac{\alpha}{2}}(84 - 13)sf_P, T_P + t_{1-\frac{\alpha}{2}}(84 - 13)sf_P \right),$$

kde  $t_{1-\frac{\alpha}{2}}(84 - 13)$  je  $\left(1 - \frac{\alpha}{2}\right) 100\%$  kvantil t rozdělení o  $84 - 13$  stupních volnosti,

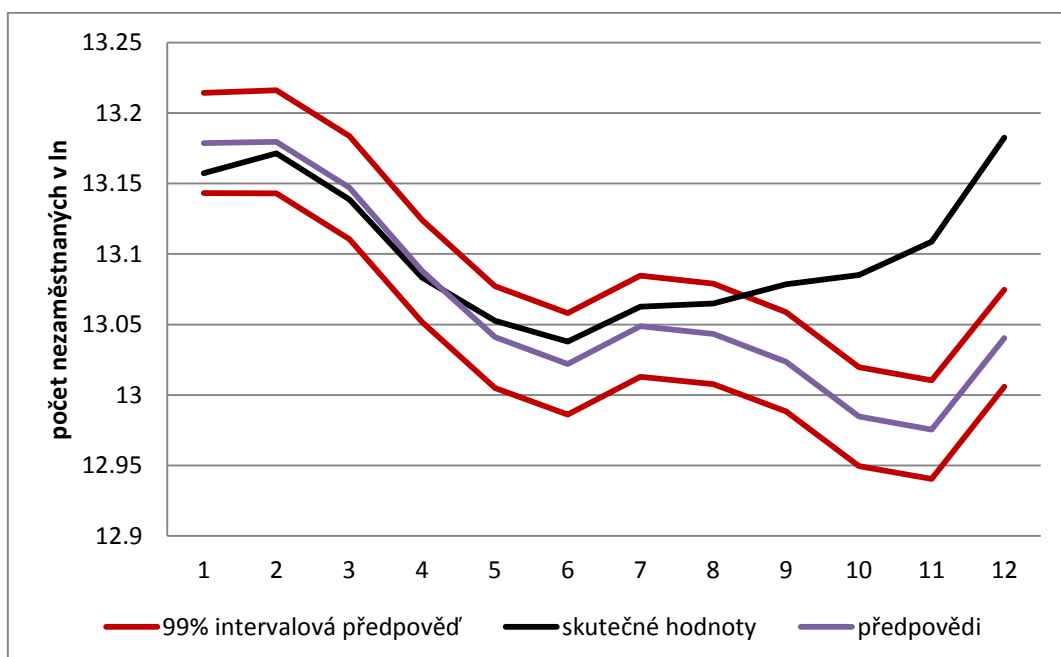
$s = \sqrt{\frac{\sum_{t=1}^{84} y_t^2 - \frac{(\sum_{t=1}^{84} y_t)^2}{84}}{84 - 13}}$ ,  $f_P = \sqrt{1 + a_P(A^T A)^{-1} a_P^T}$ , kde P je horizont předpovědi a vektor  $a_P$  ve tvaru  $a_P = [1 \quad y_{t-1+P} \quad y_{t-2+P} \quad \dots \quad y_{t-12+P}]$ .

Bodové předpovědi jsou zaznamenány v následující tabulce.

**Tabulka 5. 5.5: Předpovědi autoregresního modelu**

	předpovědi	skutečné
<b>1/2012</b>	528952	517 723
<b>2/2012</b>	529395	525 119
<b>3/2012</b>	512515	508 158
<b>4/2012</b>	483105	480 818
<b>5/2012</b>	460912	466 352
<b>6/2012</b>	452281	459 497
<b>7/2012</b>	464534	470 964
<b>8/2012</b>	461998	472 120
<b>9/2012</b>	452957	478 548
<b>10/2012</b>	435684	481 737
<b>11/2012</b>	431653	493 208
<b>12/2012</b>	460604	530 994

**Obrázek 5. 5: Předpovědi autoregresního modelu**



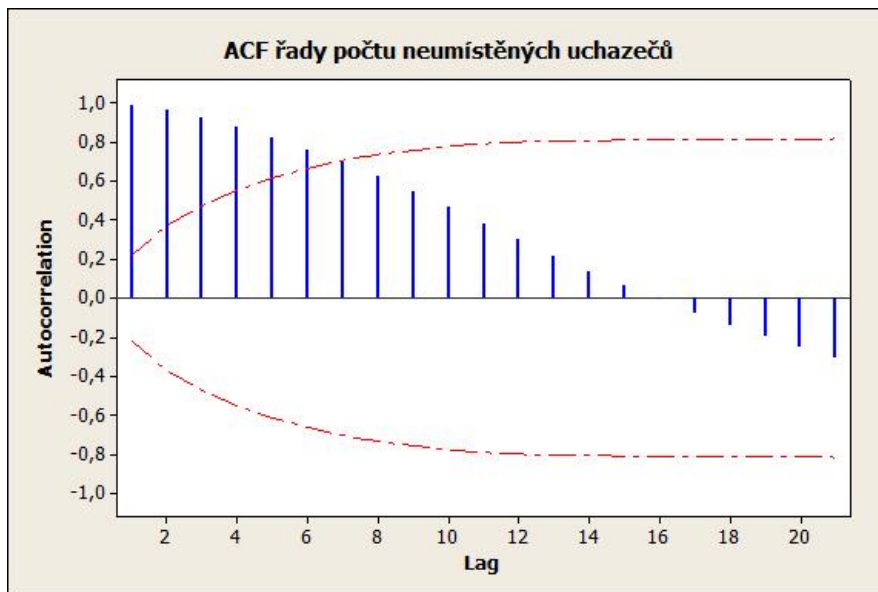


Z obrázku 5.5 pozorujeme, že intervalové předpovědi do srpna jsou uspokojivé, od září skutečné hodnoty bohužel do tohoto intervalu už nepatří. Všechny výpočty autoregresního modelu nalezneme v Excelu s názvem autoregresní\_model.xlsx.

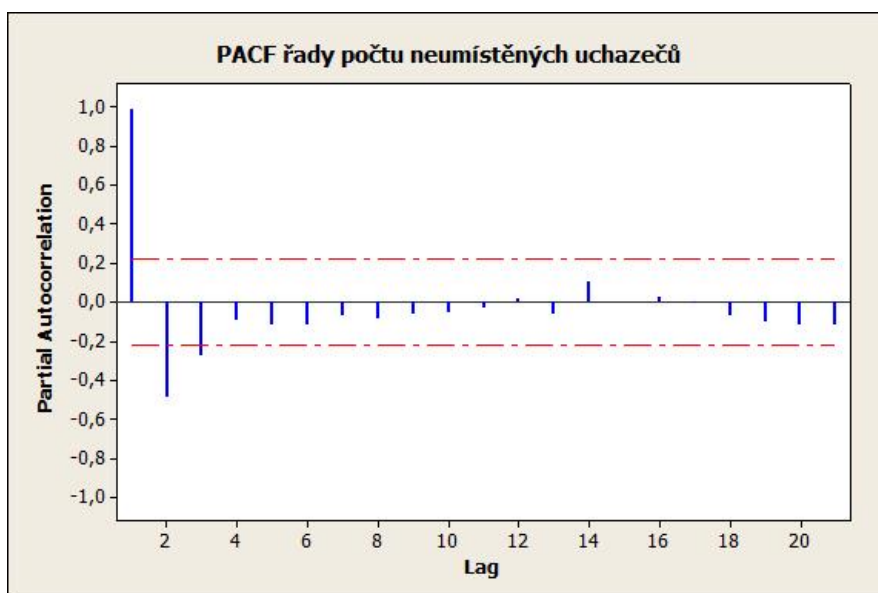
### 5.1.3 Trend pomocí Arima model

Arima model byl zpracován ve statistickém softwaru Minitab. Nejdříve ověříme stacionaritu časové řady pomocí grafu ACF a PACF.

**Obrázek 5.6:** ACF řady počtu neumístěných uchazečů

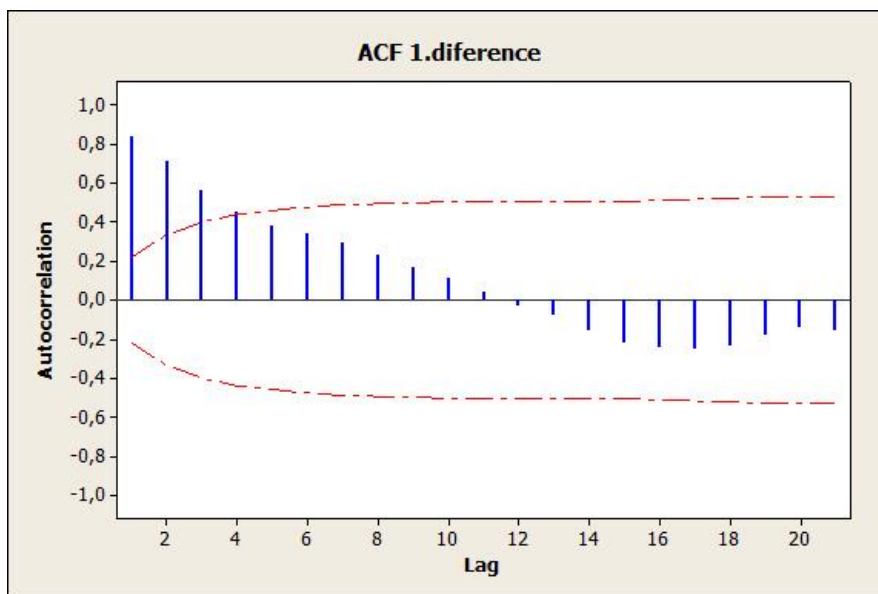


**Obrázek 5.7:** PACF řady počtu neumístěných uchazečů

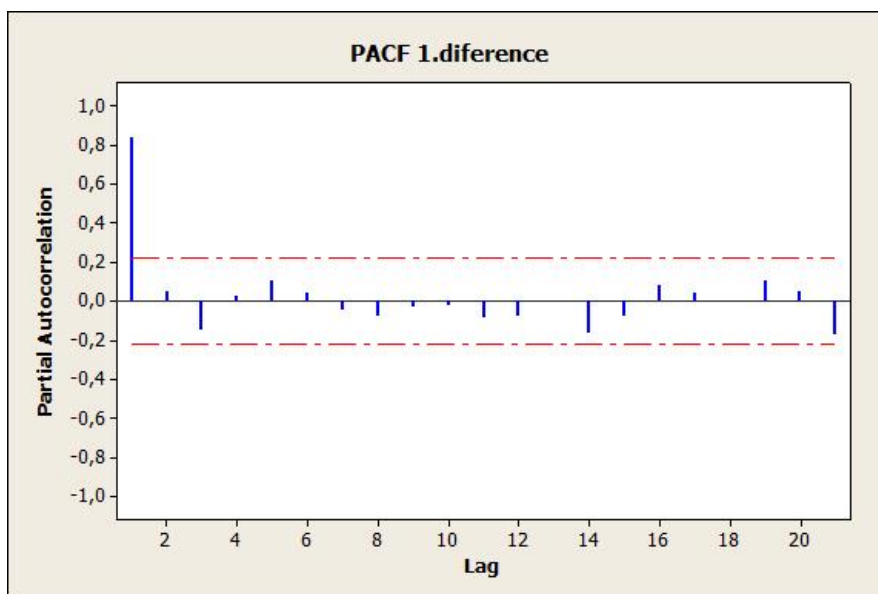


Z grafu ACF a PACF řady počtu neumístěných uchazečů pozorujeme, že řada je nestacionární. Řadu budeme stacionarizovat pomocí prvních diferencí.

**Obrázek 5.8:** ACF 1. difference řady počtu neumístěných uchazečů

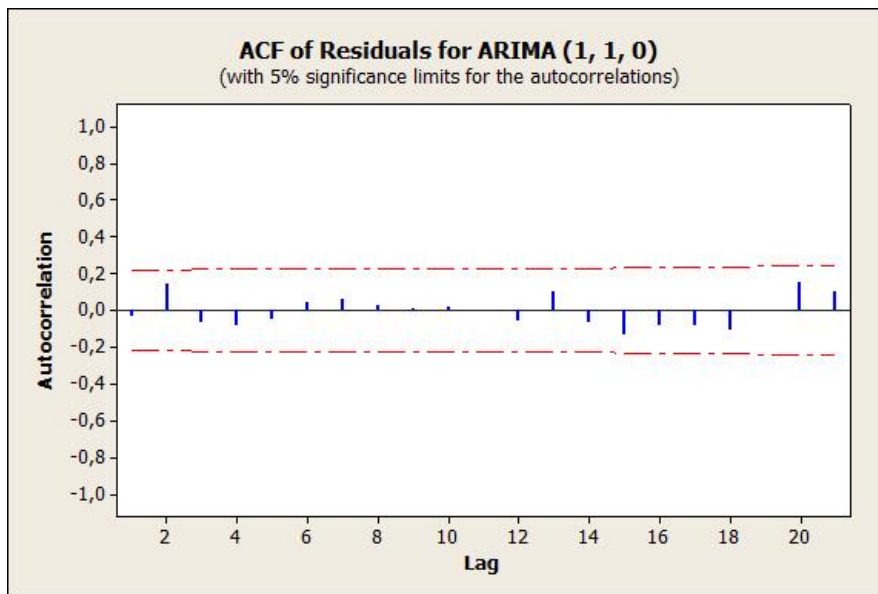


**Obrázek 5.9:** PACF 1. difference řady počtu neumístěných uchazečů

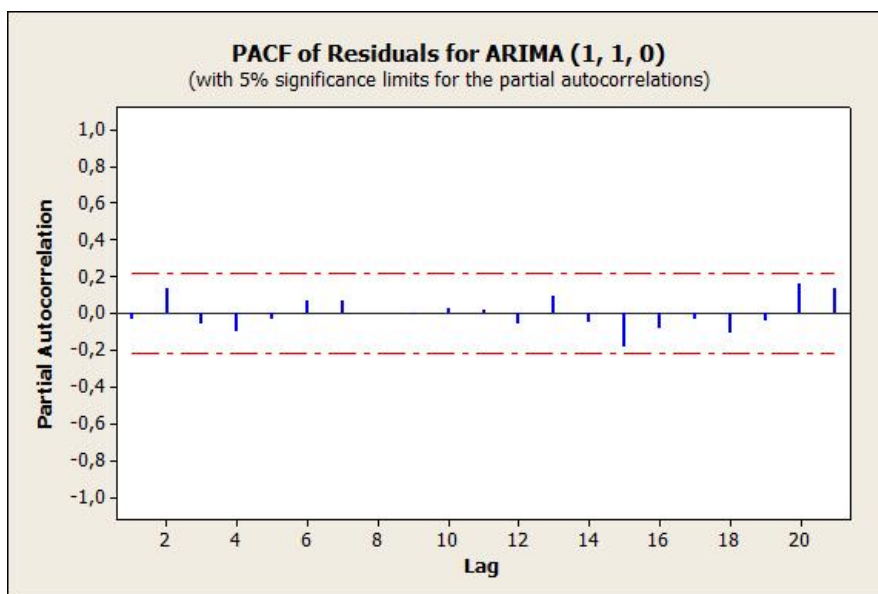


Je zřejmé, že na stacionarizovanou řadu lze použít model AR (1), neboť PACF má statisticky významnou pouze první hodnotu. Identifikovali jsme model ARIMA (1, 1, 0) a nyní ho ověříme. Prozkoumáme grafy ACF a PACF tohoto modelu:

**Obrázek 5. 10:** ACF modelu ARIMA (1, 1, 0)



**Obrázek 5. 11:** PACF modelu ARIMA (1, 1, 0)



ACF a PACF reziduí tohoto modelu indikují, že rezidua mají charakter bílého šumu, což také dokazuje Chí-kvadrát statistika, která určuje, zda jsou rezidua korelovaná či ne. P-hodnoty jsou vyšší než  $\alpha = 0,05$ , tak přijímáme hypotézu  $H_0$ , že rezidua jsou nekorelovaná.

Výsledky Chí-kvadrát statistiky a příslušné p-hodnoty jsou obsaženy v následujícím obrázku.

**Obrázek 5. 12:** Výsledky modelu ARIMA (1, 1, 0) z programu MINITAB

Final Estimates of Parameters					
Type	Coef	SE Coef	T	P	
AR 1	0,8418	0,0599	14,06	0,000	
Differencing: 1 regular difference					
Number of observations: Original series 84, after differencing 83					
Residuals: SS = 0,0139951 (backforecasts excluded)					
MS = 0,0001707 DF = 82					
Modified Box-Pierce (Ljung-Box) Chi-Square statistic					
Lag	12	24	36	48	
Chi-Square	3,8	24,9	28,4	33,8	
DF	11	23	35	47	
P-Value	0,976	0,358	0,778	0,925	

Z tabulky také pozorujeme, že parametr části AR (1) je statisticky významný, neboť p-hodnota a je menší než  $\alpha = 0,05$ .

Model ARIMA je obecně ve tvaru

$$\Phi_p(B)(1 - B)^d y_t = \Theta_q(B)e_t.$$

Část autoregresní procesu  $\Phi_p(B) = 1 - \Phi_1 B - \dots - \Phi_p B^p$  v našem případě je ve tvaru  $\Phi_1(B) = 1 - 0,8418B$  a část integrovaného procesu je ve tvaru  $(1 - B)^1$ . Část procesu klouzavých průměrů řádu q se v našem modelu ARIMA nenachází.

Model ARIMA (1, 1, 0) má tedy tvar

$$(1 - 0,8418B)(1 - B)y_t = e_t$$

nebo lze také vyjádřit jako

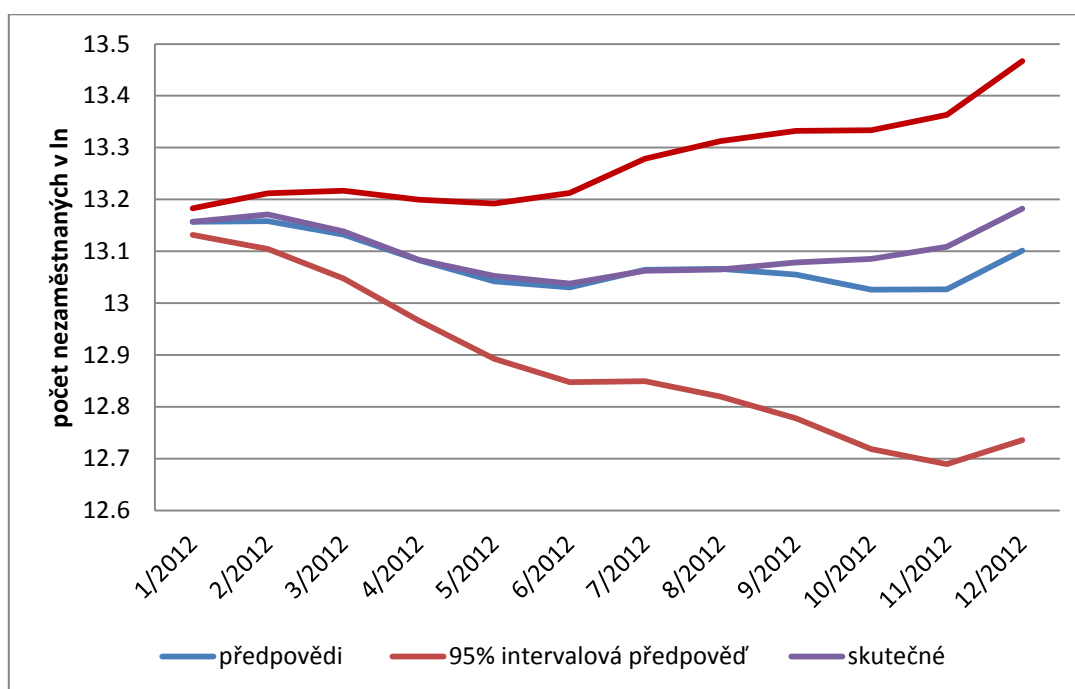
$$y_t = 1,8418y_{t-1} - 0,8418y_{t-2} + e_t.$$

Poté jsme zkonstruovali bodové předpovědi, které jsou zaznamenány v následující tabulce, a 95% intervalové předpovědi, které jsou graficky znázorněny na obrázku 5.6.

**Tabulka 5. 5.6: Předpovědi modelu ARIMA (1, 1, 0)**

	předpovědi	skutečné
1/2012	517590	517 723
2/2012	518265	525 119
3/2012	504886	508 158
4/2012	480482	480 818
5/2012	461518	466 352
6/2012	455904	459 497
7/2012	471698	470 964
8/2012	472718	472 120
9/2012	467479	478 548
10/2012	454030	481 737
11/2012	454295	493 208
12/2012	489626	530 994

**Obrázek 5. 13: Předpovědi modelu ARIMA (1, 1, 0)**



## 5.2 Porovnání předpovědních modelů

Předpovědní modely mezi sebou porovnáme pomocí extrapolačního kritéria M. S. E. a Theilova koeficientu:

$$M. S. E. = \frac{\sum_{t=84+h}^{84+h} (y_t - \hat{y}_t)^2}{h} = \frac{1}{P} \sum_{t=84+1}^{84+h} \varepsilon_t^2,$$

$$T^2 = \frac{\sum_{t=84+1}^{84+h} \left( \frac{\hat{y}_t - y_t}{y_t} \right)^2 \cdot y_t^2}{\sum_{t=84+1}^{84+h} y_t^2}, \quad h = 1, \dots, 12,$$

kde  $\hat{y}_{84+1}, \dots, \hat{y}_{84+h}$  jsou námi konstruované předpovědi pro rok 2012 a  $y_{84+1}, \dots, y_{84+h}$  jsou skutečné hodnoty roku 2012. Hodnoty těchto kritérií můžeme pozorovat v následující tabulce:

	<b>dvojité exp. vyrovnání</b>	<b>autoregresní</b>	<b>ARIMA (1, 1, 0)</b>
<b>M. S. E.</b>	195 913 476	992 809 654	350 900 922
<b>odmocnina Theilova koef.</b>	2.9%	6.4%	3.8%

Tabulka 7: Výsledky M. S. E. a Theilova koeficientu

Nejnižší hodnotu extrapolačního kritéria M. S. E. a průměrné relativní chyby předpovědi bez ohledu na znaménko má model s trendem, který je konstruován pomocí dvojitěho exponenciálního rozdělení. U kritéria M. S. E. můžeme provést ještě rozklad na chybu polohy, variability a neúplné kovariance, který byl uveden ve 4. kapitole. Hodnoty tohoto rozklad jsou zaznamenány v následující tabulce:

**Tabulka 5. 5.7: Rozklad M. S. E.**

	<b>dvojité exp. vyrovnání</b>		<b>autoregresní model</b>		<b>ARIMA (1, 1, 0)</b>	
	<b>absolutní</b>	<b>relativní</b>	<b>absolutní</b>	<b>relativní</b>	<b>absolutní</b>	<b>relativní</b>
<b>chyba polohy</b>	29 506 043	15%	308 143 712	31%	129 862 525	37%
<b>chyba variability</b>	5 650 327	3%	79 487 855	8%	431 660	0%
<b>chyba kovariance</b>	160 757 106	82%	605 178 087	61%	220 606 737	63%
<b>MSE</b>	195 913 476	100%	992 809 654	100%	350 900 922	100%

Z rozkladu M. S. E. opět pozorujeme, že nejvyšší hodnota chyby neúplné kovariance, která v sobě zahrnuje chybu zbývající nesystematické části předpovědi, je u dvojitěho exponenciálního rozdělení. Pro lepší představu odmocníme chybu polohy a chybu variability, neboť tyto sčítanci v M. S. E. jsou kvadráty. To znamená například u dvojitěho exponenciálního vyrovnání, že námi zkonstruované předpovědi se liší od skutečných hodnot v průměru o 5432 nezaměstnaných osob a směrodatná odchylka našich předpovědí je menší než směrodatná odchylka skutečných hodnot o 2377 nezaměstnaných osob.

Po porovnání předpovědních modelů pro konstrukci předpovědí pro ČR a její kraje byl vybrán model s trendem pomocí dvojitěho exponenciálního vyrovnání.

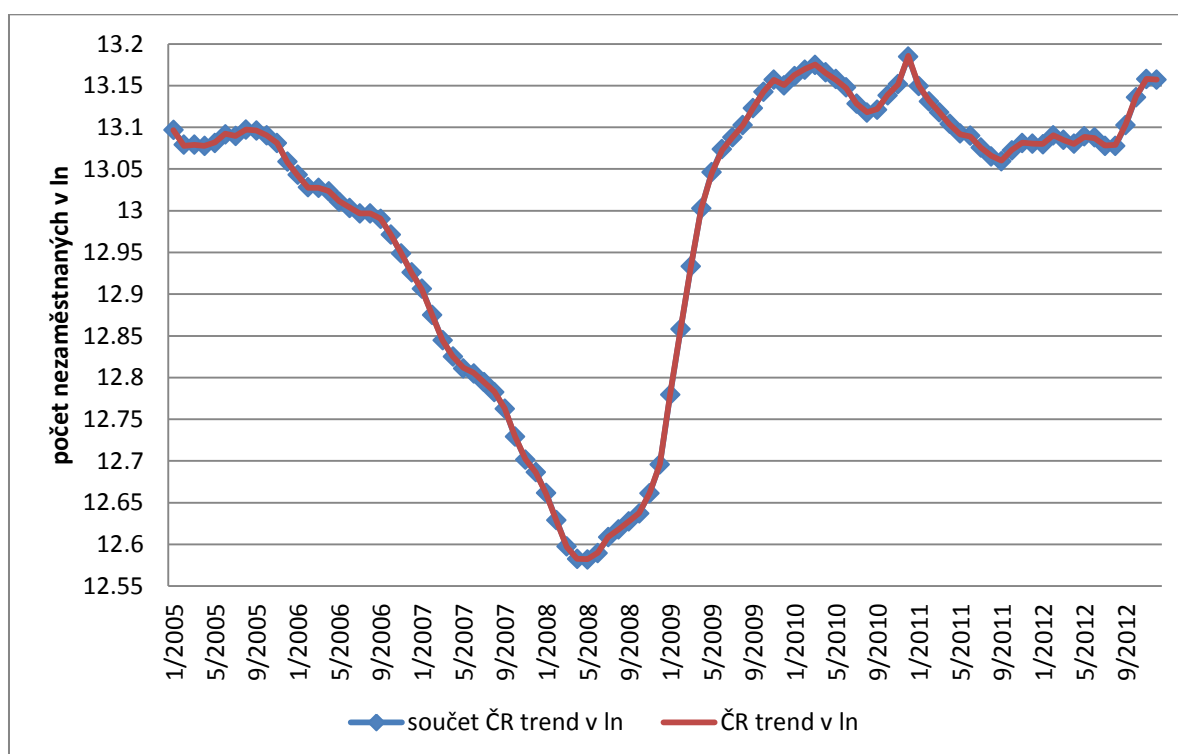
## 6 Předpovědi pro rok 2013 u jednotlivých krajů

Námi zvolený model aplikujeme na časové řady nezaměstnaných osob 2005 - 2012 jednotlivých krajů ČR a samotné ČR. Poté zkontrolujeme správnost výpočtu trendů, neboť součet trendů počtu nezaměstnaných jednotlivých krajů by měl být shodný s trendem počtu nezaměstnaných ČR. Trend u dvojitého exponenciálního vyrovnání je ve tvaru:

$$\hat{y}_{n-k} = 2S_{n-k} - S_{n-k}^{[2]}, \quad k = 0, 1, \dots, n-1, n = 1, \dots, 96.$$

Na obrázku 6.1 pozorujeme, že součet trendů počtu nezaměstnaných jednotlivých krajů je až na malé odchylky totožný s trendem počtu nezaměstnaných ČR. Tato shoda je způsobena hodnotou vyrovnávací konstanty  $\alpha$ , která se blíží k nule a tak metoda rychle reaguje na změnu charakteru dat, tj. vyrovnané hodnoty jsou blízké skutečným hodnotám.

**Obrázek 6. 1:** Kontrolní graf trendu počtu nezaměstnaných



Poté sestrojíme předpovědi pro rok 2013 ve tvaru:

$$\hat{y}_{96+h} = b_0(96) + b_1(96)h, \quad h = 1, 2, \dots, 12,$$

kde  $b_0(96) = 2S_{96} - S_{96}^{[2]}$  a  $b_1(96) = \frac{1-\alpha}{\alpha}(S_{96} - S_{96}^{[2]})$ . Intervalové předpovědi jsou pak ve tvaru

$$\left( \hat{y}_{96+h} - u_{\left(1-\frac{p}{2}\right)} d\Delta(n), \quad \hat{y}_{96+h} + u_{\left(1-\frac{p}{2}\right)} d\Delta(n) \right),$$

kde  $u_{\left(1-\frac{p}{2}\right)}$  je  $\left(1 - \frac{p}{2}\right) \cdot 100\%$  kvantil normálního rozdělení  $N(0, 1)$ ,

$$d = 1,25 \left\{ \frac{1 + \frac{1-\alpha}{(1+\alpha)^3} ((1+4\alpha+5\alpha^2) + 2(1-\alpha)(1+3\alpha)h + 2(1-\alpha)^2 h^2)}{1 + \frac{1-\alpha}{(1+\alpha)^3} ((1+4\alpha+5\alpha^2) + 2(1-\alpha)(1+3\alpha) + 2(1-\alpha)^2)} \right\}^{\frac{1}{2}},$$

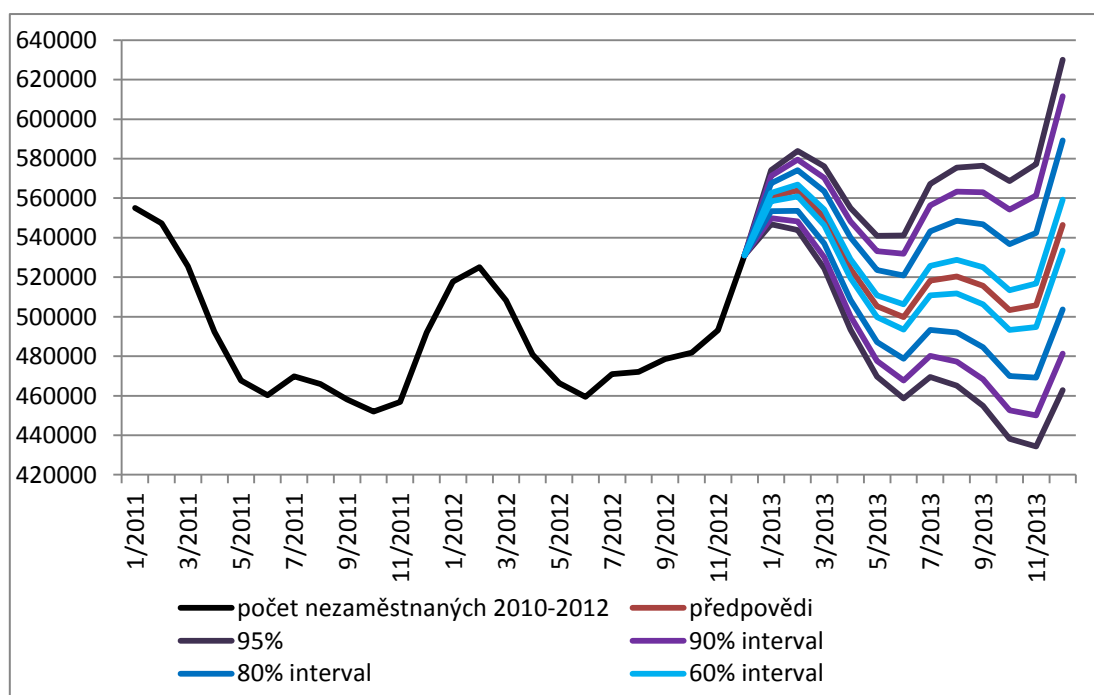
$$\Delta(n) = \sum_{t=1}^n \frac{|y_{n-k} - \hat{y}_{n-k}((n-k-1))|}{n}, \quad n = 1, \dots, 96, k = 1, \dots, n-1.$$

Bodové předpovědi a 95% intervalové předpovědi počtu nezaměstnaných osob ČR pro rok 2013 jsou zaznamenány v následující tabulce.

**Tabulka 6. 6.1: Předpovědi ČR pro rok 2013**

	95% interval		
	předpovědi	dolní mez	horní mez
<b>1/2013</b>	560 499	546 784	574 213
<b>2/2013</b>	563 887	543 915	583 859
<b>3/2013</b>	550 210	524 365	576 054
<b>4/2013</b>	524 319	493 522	555 115
<b>5/2013</b>	505 319	469 655	540 983
<b>6/2013</b>	499 849	458 644	541 055
<b>7/2013</b>	518 273	469 408	567 138
<b>8/2013</b>	520 328	465 117	575 538
<b>9/2013</b>	515 678	454 883	576 473
<b>10/2013</b>	503 412	438 150	568 673
<b>11/2013</b>	505 758	434 275	577 241
<b>12/2013</b>	546 434	462 835	630 033

**Obrázek 6. 2: Předpovědi ČR pro rok 2013**



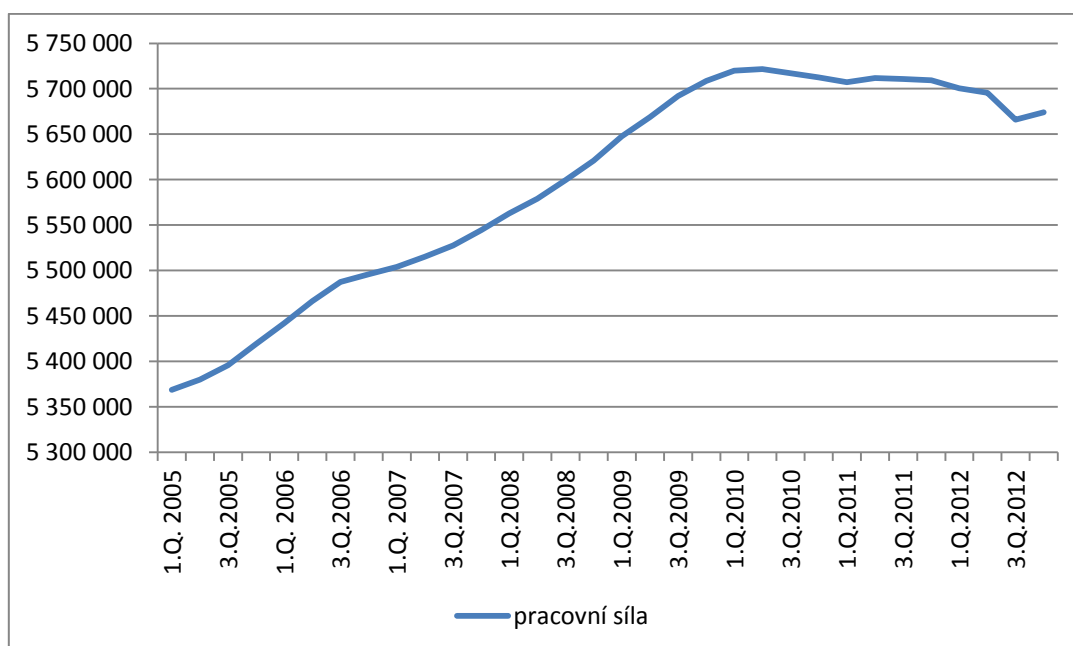


Z tabulky 6. 1 můžeme pozorovat, že od března začíná klesat počet nezaměstnaných osob. Tento pokles je způsoben sezónními faktory, neboť začínají sezónní práce ve stavebnictví, v zemědělství a dalších odvětvích. Také cestovní ruch začíná opět pomalu ožívat (pokud se nejedná o lyžařská střediska). Naopak nárůst počtu nezaměstnaných osob v červenci a v srpnu je způsoben vstupem absolventů škol na trh práce, kteří si po ukončení školy hledají své první zaměstnání. Následuje další pokles počtu nezaměstnaných osob v říjnu a v listopadu díky blížícím se vánočním svátkům a tím spojená zvýšená poptávka po určitém zboží. Nárůst počtu nezaměstnaných během zimních měsíců je způsoben hlavně krátkodobým propuštěním zaměstnanců na konci roku, kteří pracují v odvětvích, kde produkce je závislá na počasí (stavebnictví, zemědělství aj.) Tento koloběh se opakuje každý rok i v našich předpovědích, jak jsme pozorovali v tabulce 6. 1, a je schován v eliminované sezónní složce.

Na obrázku 6. 2 jsou postupně zobrazeny intervaly spolehlivosti 60%, 80%, 90% a 95%. Můžeme pozorovat, že čím je předpověď dál od prahu predikce ( $n = 96$ ), tím se zvětšuje naše nejistota budoucnosti a tím pádem je širší pás intervalové předpovědi. Podíváme se, jak tato šířka intervalové předpovědi bude mít vliv na intervalovou předpověď míry nezaměstnanosti.

Pro předpověď míry nezaměstnanosti potřebujeme znát hodnoty pracovní síly na rok 2013, jelikož míra nezaměstnanosti je dána podílem počtu nezaměstnaných osob na pracovní síle. Na obrázku 6.3 můžeme sledovat vývoj pracovní síly během let 2005 - 2012. Časová řada pracovní síly je sledována za čtvrtletí, tedy máme časovou řadu o délce 32 pozorování.

**Obrázek 6. 3: Pracovní síla ČR**



Podle analýzy grafu z obrázku 6. 3 usuzujeme, že časová řada pracovní síly obsahuje pouze trendovou a náhodnou složku. Pro konstrukci předpovědí vyzkoušíme, jak časovou řadu

pracovní síly vystihuje lineární a parabolický trend. Pomocí regrese zjišťujeme, že koeficient determinace (upravený) je pro lineární trend  $R^2 = 85\%$  a pro parabolický  $R^2 = 96\%$ . Na základě koeficientu determinace volíme trend parabolický, neboť vysvětluje variabilitu časové řady pracovní síly na 96%. Bodové předpovědi u parabolického trendu jsou ve tvaru:

$$T_p = b_0 + b_1P + b_2P^2 \quad \text{pro } P = 33, 34, 35, 36.$$

V tabulce 6.2 jsou zaznamenány hodnoty bodových předpovědí pracovní síly ČR na rok 2013.

**Tabulka 6. 6.2:** *Bodová předpověď pracovní síly ČR*

<b>1.Q. 2013</b>	5 692 935
<b>2.Q. 2013</b>	5 687 349
<b>3.Q. 2013</b>	5 680 752
<b>4.Q. 2013</b>	5 673 143

Nyní si sestojíme intervalovou předpověď míry nezaměstnanosti tak, že vydělíme dolní a horní mez intervalové předpovědi počtu nezaměstnaných osob získanými bodovými předpověďmi pracovní síly.

**Tabulka 6. 6.3:** *95% a 90% intervalová předpověď registrované nezaměstnanosti*

	95% interval			90% interval		
	dolní mez	horní mez	šířka	dolní mez	horní mez	šířka
<b>1/2013</b>	9.60%	10.09%	0.48%	9.66%	10.03%	0.38%
<b>2/2013</b>	9.55%	10.26%	0.70%	9.63%	10.18%	0.55%
<b>3/2013</b>	9.21%	10.12%	0.91%	9.31%	10.02%	0.71%
<b>4/2013</b>	8.68%	9.76%	1.08%	8.80%	9.64%	0.84%
<b>5/2013</b>	8.26%	9.51%	1.25%	8.40%	9.37%	0.98%
<b>6/2013</b>	8.06%	9.51%	1.45%	8.22%	9.35%	1.13%
<b>7/2013</b>	8.26%	9.98%	1.72%	8.45%	9.79%	1.34%
<b>8/2013</b>	8.19%	10.13%	1.94%	8.40%	9.92%	1.51%
<b>9/2013</b>	8.01%	10.15%	2.14%	8.24%	9.91%	1.67%
<b>10/2013</b>	7.72%	10.02%	2.30%	7.98%	9.77%	1.79%
<b>11/2013</b>	7.65%	10.17%	2.52%	7.93%	9.90%	1.96%
<b>12/2013</b>	8.16%	11.11%	2.95%	8.48%	10.78%	2.30%

**Tabulka 6. 6.4:** 80% a 60% intervalová předpověď registrované nezaměstnanosti

	80% interval			60% interval		
	dolní mez	horní mez	šířka	dolní mez	horní mez	šířka
<b>1/2013</b>	9.73%	9.96%	0.23%	9.81%	9.88%	0.07%
<b>2/2013</b>	9.74%	10.07%	0.33%	9.86%	9.95%	0.10%
<b>3/2013</b>	9.44%	9.88%	0.44%	9.60%	9.73%	0.13%
<b>4/2013</b>	8.94%	9.50%	0.55%	9.14%	9.30%	0.17%
<b>5/2013</b>	8.55%	9.22%	0.67%	8.78%	8.99%	0.20%
<b>6/2013</b>	8.40%	9.18%	0.78%	8.67%	8.91%	0.23%
<b>7/2013</b>	8.68%	9.57%	0.89%	8.99%	9.26%	0.27%
<b>8/2013</b>	8.66%	9.66%	1.01%	9.01%	9.31%	0.30%
<b>9/2013</b>	8.52%	9.64%	1.12%	8.91%	9.25%	0.34%
<b>10/2013</b>	8.25%	9.49%	1.24%	8.69%	9.06%	0.37%
<b>11/2013</b>	8.24%	9.59%	1.35%	8.71%	9.12%	0.41%
<b>12/2013</b>	8.90%	10.37%	1.47%	9.41%	9.85%	0.44%

V tabulkách 6. 3 a 6. 4 si můžeme všimnout nepřímé úměry mezi spolehlivostí a přesností odhadu. Nejpřesnější odhad máme u 60% intervalové předpovědi, neboť přípustná chyba (šířka intervalu) je nejmenší, ale máme pouze 60% spolehlivost, že budoucí pozorování budou do tohoto intervalu náležet. Naopak u 95% intervalové předpovědi budou budoucí pozorování ležet v tomto intervalu s pravděpodobností 95%, ale za cenu zvětšení šířky intervalu (zmenšení přesnosti odhadu), takže např. intervalová předpověď míry nezaměstnanosti pro prosinec 2013 je prakticky bezcenná, neboť šířka intervalu je velká (2,9 %).

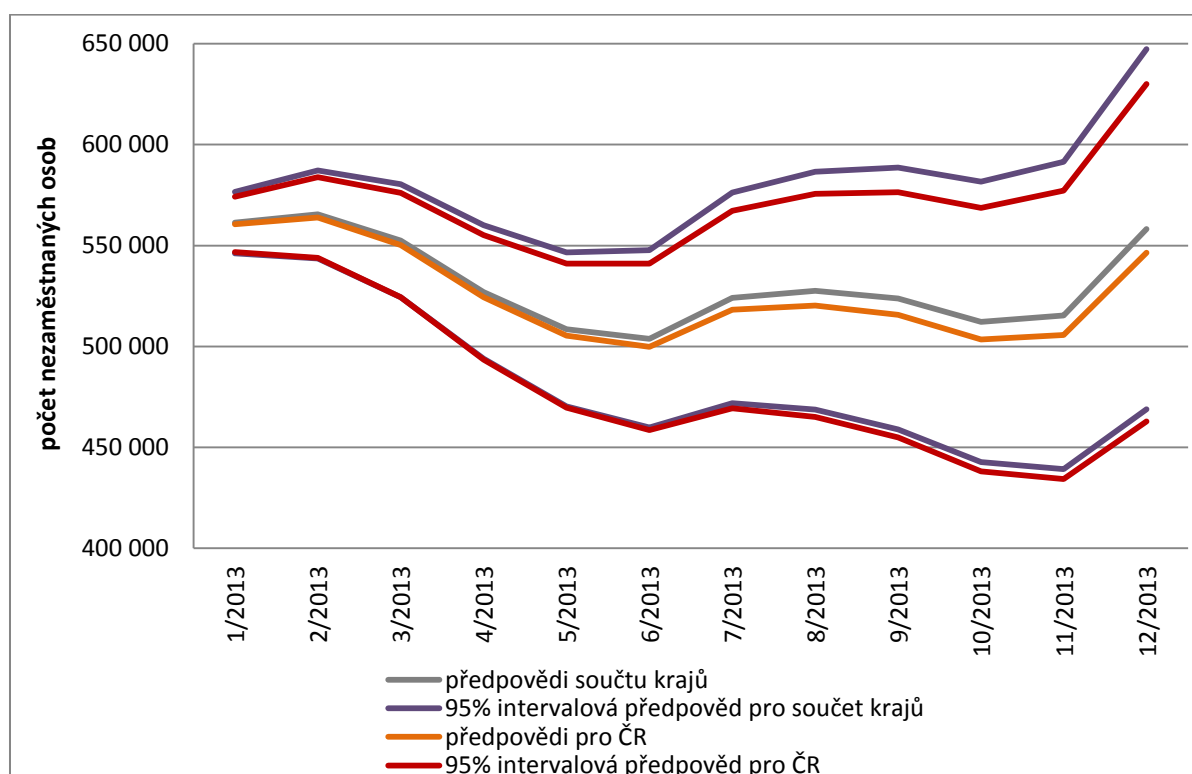
Předpovědi pro jednotlivé kraje nalezneme v příloze A. 1 a A. 2. Podle námi konstruovaných předpovědí by se nadále počet nezaměstnaných osob měl zvyšovat. Největší nárůst počtu nezaměstnaných po srovnání hodnot roku 2012 s bodovými předpověďmi na rok 2013 nastane v Moravskoslezském kraji. Na tomto nárůstu se podílí např. propouštění ke konci roku 2012 ze společnosti ArcelorMittal nebo ze společnosti Pegatron na Ostravsku (Česká televize, 2013). Můžeme si také povšimnout, že nezaměstnanost je nerovnoměrně rozložená v rámci ČR. Je to způsobeno odlišným potenciálem jednotlivých krajů a také nízkou mobilitou obyvatelstva se stěhovat za prací do jiného kraje.

Také zkontrolujeme, jestli předpovědi pro ČR jsou srovnatelné s předpověďmi dané součtem krajů.

Tabulka 6. 6.5: Porovnání předpovědí pro ČR a součtu krajů

ČR součet			ČR model		
předpovědi	dolní	horní	předpovědi	dolní	horní
561 403	546 189	576 618	560 499	546 784	574 213
565 445	543 628	587 262	563 887	543 915	583 859
552 409	524 436	580 381	550 210	524 365	576 054
526 962	493 875	560 048	524 319	493 522	555 115
508 471	470 341	546 601	505 319	469 655	540 983
503 799	459 858	547 741	499 849	458 644	541 055
524 100	471 928	576 271	518 273	469 408	567 138
527 614	468 670	586 557	520 328	465 117	575 538
523 807	458 980	588 633	515 678	454 883	576 473
512 188	442 692	581 683	503 412	438 150	568 673
515 342	439 255	591 429	505 758	434 275	577 241
558 128	468 934	647 323	546 434	462 835	630 033

Obrázek 6. 4: Srovnání předpovědí součtu krajů a ČR



Z tabulky 6. 5 a obrázku 6. 4 si můžeme všimnout, že odchylky mezi předpověďmi součtu krajů a ČR se zvětšují se vzdáleností předpovědi od prahu predikce. Je to dáno způsobem konstrukce předpovědí. Připomeňme si, že předpovědi byly konstruovány podle vzorce

$$\hat{y}_{96+h} = b_0(96) + b_1(96)h, \quad h = 1, 2, \dots, 12,$$

kde  $b_0(96) = 2 S_{96} - S_{96}^{[2]}$  a  $b_1(96) = \frac{1-\alpha}{\alpha} (S_{96} - S_{96}^{[2]})$ .

Již při kontrole shody trendu ČR a součtu krajů jsme pozorovali menší odchylky, které jsou způsobeny také tím, že jednotlivé kraje mají svá minima a maxima v různém období. Nyní se tyto odchylky budou zvyšovat, neboť odchylka parametrů  $b_0(96)$  součtu krajů a ČR bude sice stále konstantní, ale odchylka parametrů  $b_1(96)$  součtu krajů a ČR se bude  $h$ -násobně zvětšovat s krokem  $h$ .

Po uveřejnění nových pozorování je výhodou metody exponenciálního vyrovnání její adaptivnost. Nová pozorování adaptujeme pomocí vyrovnávacích statistik bez potřeby přepočtu celé metody, jelikož pro výpočet vyrovnané hodnoty nového pozorování  $\hat{y}_{n+h}$  stačí mít uložené v paměti pouze hodnoty vyrovnávacích statistik  $S_{n+h-1}$  a  $S_{n+h-1}^{[2]}$ .

## Závěr

Cílem této bakalářské práce byla predikce časových řad počtu nezaměstnaných jednotlivých krajů na cca jeden rok dopředu. Pro modelování časových řad jsem si zvolila přístup dekompozice, který obecně rozkládá časovou řadu na složku trendovou, sezónní, cyklickou a náhodnou. Z důvodu toho, že časová řada počtu nezaměstnaných je krátkodobá, tak se v ní nachází pouze složka trendová, sezónní a náhodná. Pro konstrukci předpovědí jsem si nejdříve zvolila tři předpovědní modely s různým typem modelování trendu – dvojitě exponenciální vyrovnání, autoregresní model a model ARIMA. Sezónní složka byla eliminována pomocí modelu proporcionální sezónnosti, který ve srovnání s modelem konstantní sezónnosti byl výstižnějším. Zvolené předpovědní modely jsem díky metodě zatajovaného vzorku podrobila zkoušce ze schopnosti předpovídat. Při porovnání předpovědních modelů mezi sebou se stal nejlepším model s trendem pomocí metody dvojitě exponenciálního vyrovnání. Tento model byl poté aplikován na časové řady počtu nezaměstnaných osob jednotlivých krajů ČR a následně byly sestrojeny předpovědi na rok 2013.

Ze sestrojených bodových předpovědí vyplývá, že by se nadále řada počtu nezaměstnaných osob ve všech krajích měla zvyšovat. Při porovnání s hodnotami z roku 2012 největší nárůst počtu nezaměstnaných nastane v Moravskoslezském kraji. Dalším typem předpovědi byla sestrojena intervalová předpověď, jejíž šířka se zvětšuje se vzdáleností od prahu predikce, což je způsobeno naší nejistotou budoucnosti. Avšak se zvětšující se šířkou intervalové předpovědi klesá vypovídající hodnota této předpovědi.

Pro modelování jednotlivých složek časové řady existuje mnoho přístupů a modifikací. Záleží na zkušenostech a softwarové vybavenosti statistika. Proto je možné, že by jiný statistik při jiném výběru přístupu k jednotlivým složkám našel lepší předpovědní model.

## Literatura

ARLT, Josef. *Moderní metody modelování ekonomických časových řad*. 1.vyd. Praha: Grada Publishing, 1999, 307 s. ISBN 80-716-9539-4.

ARLT, Josef. *Příklady z analýzy ekonomických časových řad*. 1.vyd. Praha: VŠE, 1997, 147 s. ISBN 80-707-9056-3.

CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii*. Praha: SNTL, 1986.

CIPRA, Tomáš. *Finanční ekonometrie*. 1. vyd. Praha: Ekopress, 2008. 538 s. ISBN 978-80-86929-43-9.

ENDERS, Walter. *Applied econometric time series*. New York: Wiley, c1995, xi, 433 p. ISBN 04-710-3941-1.

HINDLS, Richard. *Metody statistické analýzy pro ekonomy*. 2. přepracované vyd. Praha: Management Press, 2000, 259 s. ISBN 80-726-1013-9.

HINDLS, Richard. *Statistika pro ekonomy*. 1. vyd. Praha: Professional Publishing, 2002, 415 s. ISBN 80-864-1926-6.

HUŠEK, Roman. *Základy ekonometrické analýzy II: Speciální postupy a techniky*. 1.vyd. Praha: VŠE, 1998, 265 s. ISBN 80-707-9441-0.

KOZÁK, Josef. *Úvod do analýzy ekonomických časových řad*. 1.vyd. Praha: VŠE, 1994, 208 s. ISBN 80-707-9760-6.

KOZÁK, Josef a Jan SEGER. *Jednoduché statistické metody v prognostice*. 1.vyd. Praha: SNTL, 1975.

## Internetové zdroje

ČESKÝ STATISTICKÝ ÚŘAD. *I. Základní metodické přístupy* [online]. 2012a [cit. 2012-09-29]. Dostupné z: [http://www.czso.cz/csu/2012edicniplan.nsf/t/2400195FDB/\\$File/310112q1m01.pdf](http://www.czso.cz/csu/2012edicniplan.nsf/t/2400195FDB/$File/310112q1m01.pdf)

ČESKÝ STATISTICKÝ ÚŘAD. *II. Metodické vysvětlivky* [online]. 2012b [cit. 2012-09-29]. Dostupné z: [http://www.czso.cz/csu/2012edicniplan.nsf/t/EA0034DAC8/\\$File/311512m02%20cz.pdf](http://www.czso.cz/csu/2012edicniplan.nsf/t/EA0034DAC8/$File/311512m02%20cz.pdf)

ČESKÝ STATISTICKÝ ÚŘAD. *Zaměstnanost a nezaměstnanost podle výsledků VŠPS - Metodika: Metodické popisy ukazatelů VŠPS* [online]. 2012c [cit. 2012-09-29]. Dostupné z: [http://www.czso.cz/csu/redakce.nsf/i/zam\\_vsps](http://www.czso.cz/csu/redakce.nsf/i/zam_vsps)

Česká televize. *Počet nezaměstnaných v kraji přesáhl 80 tisíc* [online]. © 1996 – 2013 [cit. 2013-05-15]. Dostupné z: <http://www.ceskatelevize.cz/zpravodajstvi-ostrava/zpravy/210320-pocet-nezamestnanych-v-kraji-presahl-80-tisic/>

INTERNATIONAL LABOUR ORGANIZATION. *Labour force surveys - Source of statistics of the labour force and its components* [online]. 2011 [cit. 2013-04-04]. Dostupné z: [http://laborsta.ilo.org/applv8/data/SSM3\\_NEW/E/SSM3.html#C](http://laborsta.ilo.org/applv8/data/SSM3_NEW/E/SSM3.html#C)

MINISTERSTVO PRÁCE A SOCIÁLNÍCH VĚCÍ. *MPSV harmonizuje vykazování míry nezaměstnanosti s EU* [online]. 2004 [cit. 2012-09-29]. Dostupné z: <http://www.mpsv.cz/files/clanky/272/090804a.pdf>

MINISTERSTVO PRÁCE A SOCIÁLNÍCH VĚCÍ. *Změna metodiky ukazatele registrované nezaměstnanosti* [online]. 2012 [cit. 2013-04-04]. Dostupné z: [http://portal.mpsv.cz/sz/stat/nz/zmena\\_metodiky](http://portal.mpsv.cz/sz/stat/nz/zmena_metodiky)



# Přílohy

## A Tištěné

### A. 1 Předpovědi na rok 2013 pro jednotlivé kraje

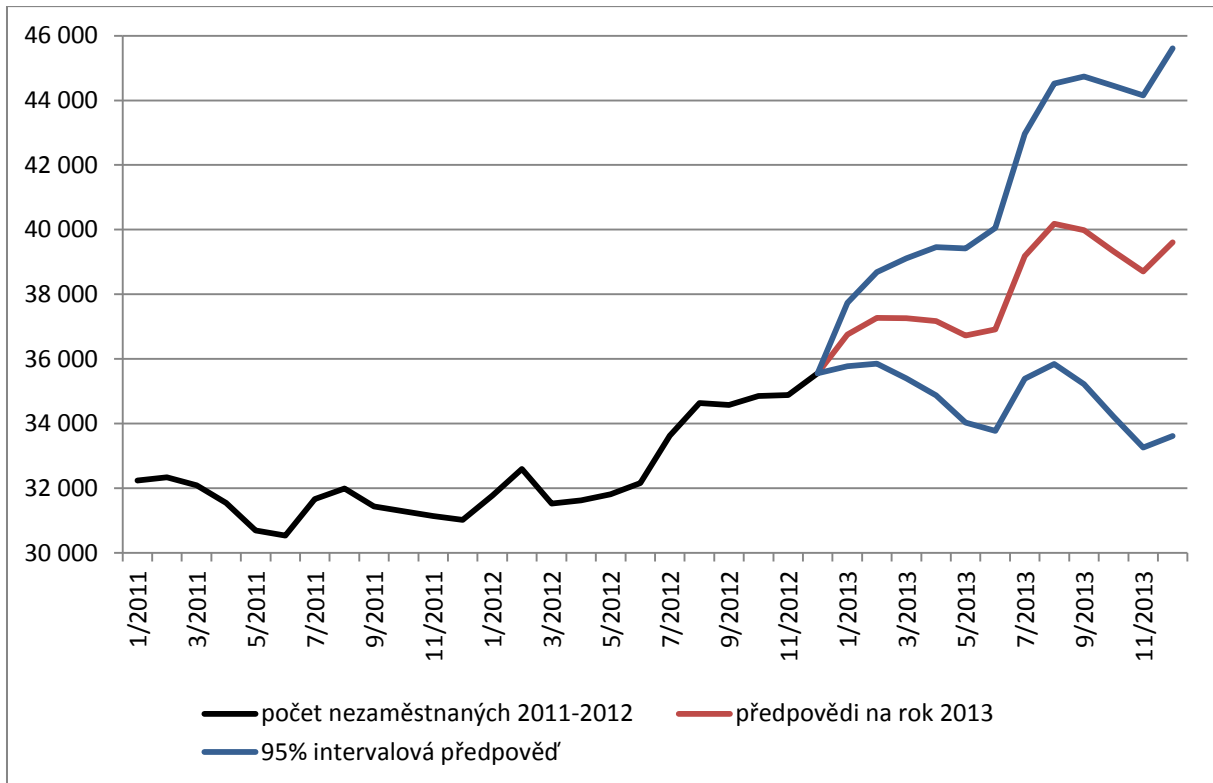
Tabulka A. 0.1: Předpovědi pro kraje

Praha			Středočeský kraj			Jihočeský		
	95% interval			95% interval			95% interval	
předpovědi	dolní mez	horní mez	předpovědi	dolní mez	horní mez	předpovědi	dolní mez	horní mez
36 761	35 790	37 732	56 085	54 686	57 484	30 190	29 401	30 978
37 268	35 874	38 663	56 051	54 051	58 051	30 150	29 001	31 298
37 258	35 414	39 102	54 389	51 751	57 028	28 551	27 021	30 081
37 170	34 866	39 474	51 467	48 175	54 760	26 058	24 138	27 977
36 726	33 956	39 496	49 502	45 548	53 456	24 192	21 878	26 505
36 916	33 677	40 155	48 887	44 266	53 508	23 532	20 822	26 242
39 179	35 470	42 889	50 616	45 326	55 906	24 241	21 133	27 349
40 183	36 001	44 364	50 790	44 829	56 751	24 208	20 701	27 715
39 980	35 325	44 635	50 288	43 654	56 922	23 846	19 939	27 752
39 334	34 205	44 462	48 769	41 462	56 076	23 216	18 909	27 523
38 708	33 105	44 310	48 464	40 483	56 445	23 592	18 885	28 300
39 613	33 536	45 690	51 637	42 981	60 293	26 510	21 402	31 619
Plzeňský kraj			Karlovarský kraj			Ústecký kraj		
	95% interval			95% interval			95% interval	
předpovědi	dolní mez	horní mez	předpovědi	dolní mez	horní mez	předpovědi	dolní mez	horní mez
24 630	23 946	25 314	18 906	18 464	19 348	61 824	60 866	62 783
24 673	23 692	25 655	18 840	18 200	19 480	61 718	60 367	63 069
23 885	22 588	25 183	18 162	17 312	19 012	60 034	58 265	61 804
22 759	21 137	24 381	17 430	16 366	18 495	57 798	55 599	59 997
21 912	19 963	23 862	16 762	15 480	18 044	55 840	53 206	58 474
21 919	19 639	24 198	16 450	14 949	17 951	54 863	51 790	57 935
23 177	20 567	25 788	16 643	14 923	18 363	55 433	51 920	58 945
23 491	20 548	26 434	16 479	14 539	18 419	54 722	50 768	58 677
23 184	19 908	26 459	16 356	14 195	18 516	53 945	49 549	58 342
22 587	18 978	26 197	16 065	13 683	18 446	52 403	47 563	57 243
22 765	18 822	26 708	16 156	13 554	18 759	52 104	46 820	57 387
24 934	20 658	29 211	17 120	14 297	19 943	54 818	49 090	60 546

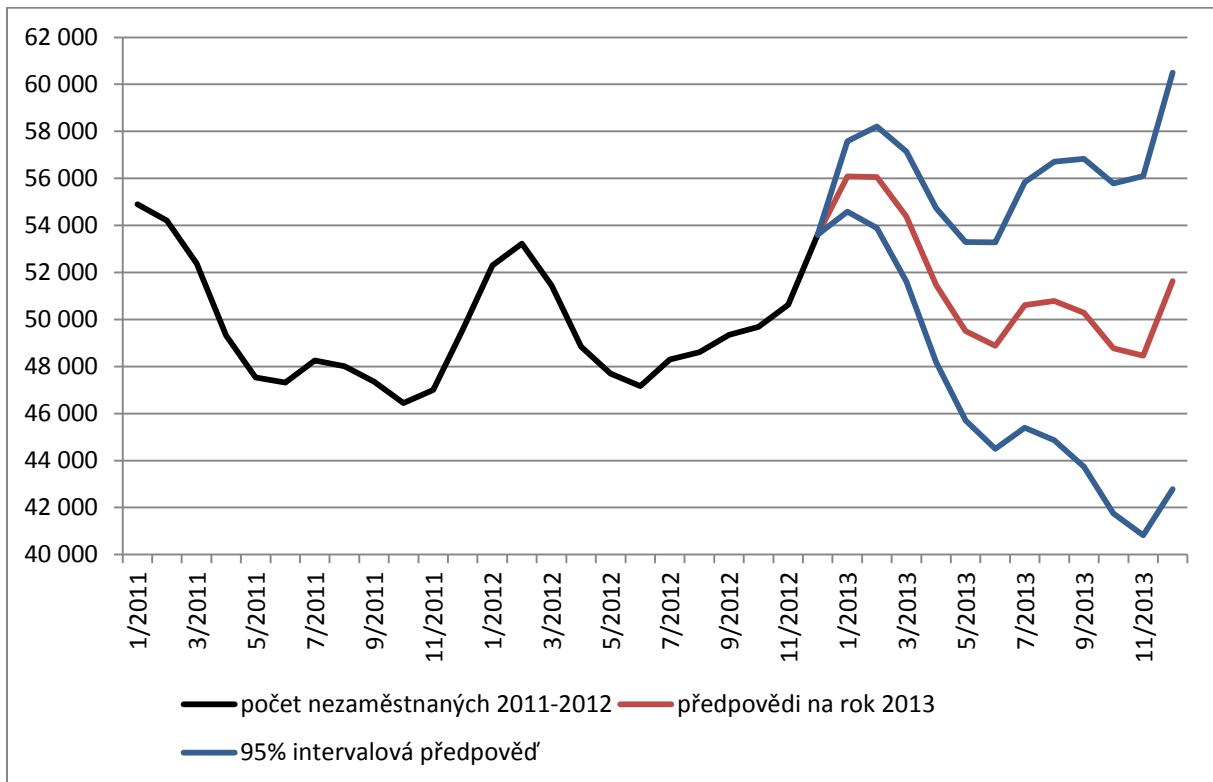
Liberecký kraj			Královohradecký kraj			Pardubický kraj		
	95% interval			95% interval			95% interval	
předpovědi	dolní mez	horní mez	předpovědi	dolní mez	horní mez	předpovědi	dolní mez	horní mez
24 418	23 755	25 082	26 295	25 577	27 013	26 808	25 998	27 618
24 541	23 630	25 452	26 601	25 608	27 595	27 124	25 980	28 268
24 335	23 159	25 511	26 204	24 916	27 492	26 226	24 726	27 726
23 679	22 231	25 128	25 109	23 518	26 700	24 495	22 630	26 361
23 005	21 279	24 730	24 229	22 330	26 127	23 364	21 129	25 599
22 711	20 706	24 716	24 197	21 988	26 405	23 163	20 555	25 770
23 612	21 326	25 898	25 888	23 368	28 409	24 635	21 653	27 616
23 500	20 932	26 068	26 280	23 447	29 113	25 066	21 709	28 423
23 208	20 357	26 058	26 168	23 022	29 315	25 135	21 402	28 869
22 673	19 539	25 807	25 880	22 420	29 341	24 604	20 494	28 715
22 505	19 087	25 922	26 574	22 800	30 349	25 149	20 662	29 637
23 780	20 079	27 482	29 572	25 482	33 661	28 599	23 734	33 464
Vysočina			Jihomoravský kraj			Zlínský kraj		
	95% interval			95% interval			95% interval	
předpovědi	dolní mez	horní mez	předpovědi	dolní mez	horní mez	předpovědi	dolní mez	horní mez
29 369	28 553	30 184	67 930	66 344	69 517	33 890	32 953	34 827
30 109	28 936	31 282	67 786	65 461	70 111	34 652	33 341	35 963
29 568	28 016	31 120	65 410	62 304	68 516	34 359	32 649	36 069
27 526	25 585	29 466	61 443	57 539	65 348	33 007	30 887	35 126
26 637	24 304	28 970	58 588	53 876	63 299	32 010	29 475	34 545
26 915	24 187	29 644	57 458	51 935	62 981	31 986	29 032	34 939
29 064	25 938	32 189	59 188	52 850	65 525	33 844	30 471	37 218
29 983	26 459	33 506	59 259	52 105	66 412	34 601	30 805	38 396
30 230	26 308	34 152	58 396	50 425	66 367	34 677	30 458	38 895
30 081	25 760	34 403	56 705	47 915	65 495	34 112	29 470	38 754
31 291	26 570	36 012	57 000	47 390	66 609	34 543	29 477	39 609
35 830	30 709	40 951	61 784	51 355	72 213	38 153	32 662	43 643
Olomoucký kraj			Moravskoslezský kraj					
	95% interval			95% interval				
předpovědi	dolní mez	horní mez	předpovědi	dolní mez	horní mez			
41 754	40 508	43 000	82 544	80 529	84 559			
41 956	40 123	43 788	83 976	81 127	86 826			
40 331	37 879	42 784	83 695	79 958	87 433			
37 615	34 529	40 701	81 405	76 756	86 053			
35 668	31 942	39 394	80 037	74 466	85 608			
34 702	30 333	39 072	80 101	73 600	86 601			
35 850	30 835	40 865	82 730	75 295	90 164			
35 527	29 864	41 189	83 525	75 155	91 896			
34 998	28 688	41 309	83 396	74 087	92 706			
33 824	26 864	40 784	81 933	71 684	92 182			
34 053	26 444	41 662	82 438	71 248	93 629			
37 581	29 321	45 840	88 197	76 065	100 329			

## A. 2 Předpovědi na rok 2013 – graficky znázorněné

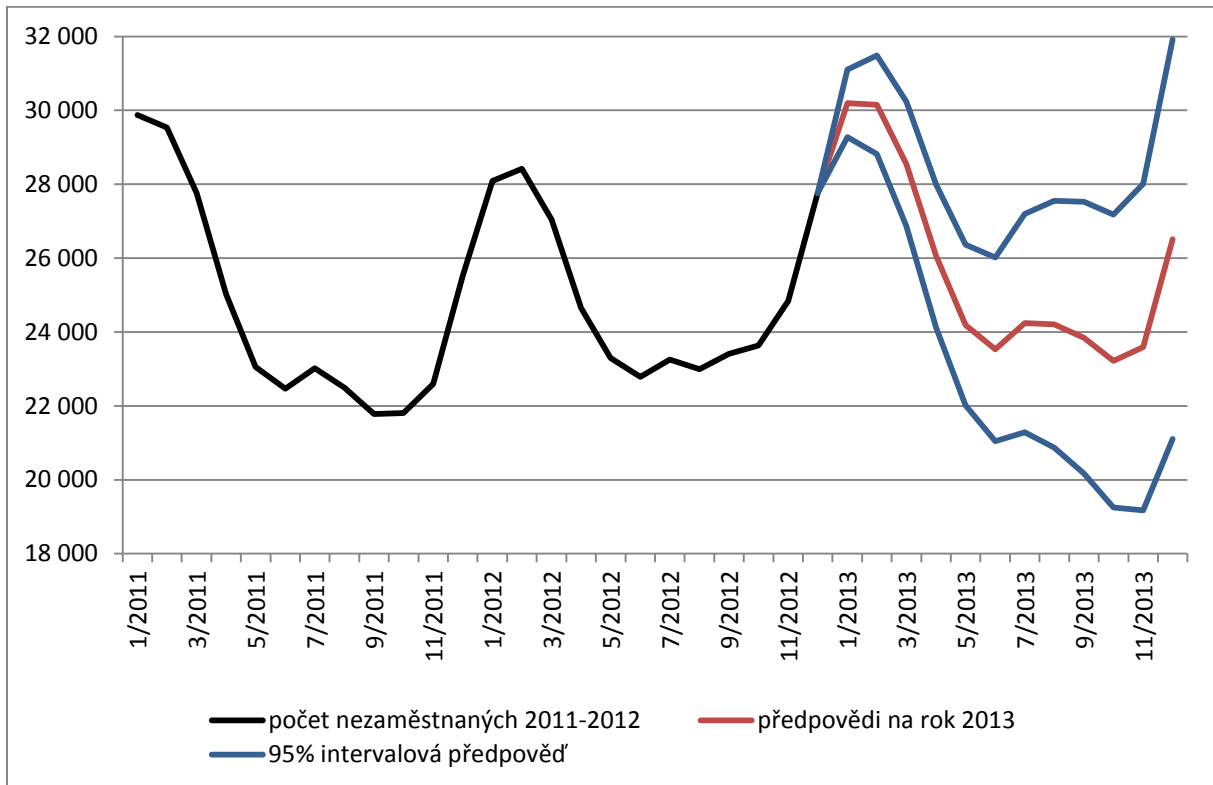
Obrázek A. 1: Předpovědi pro kraj Praha



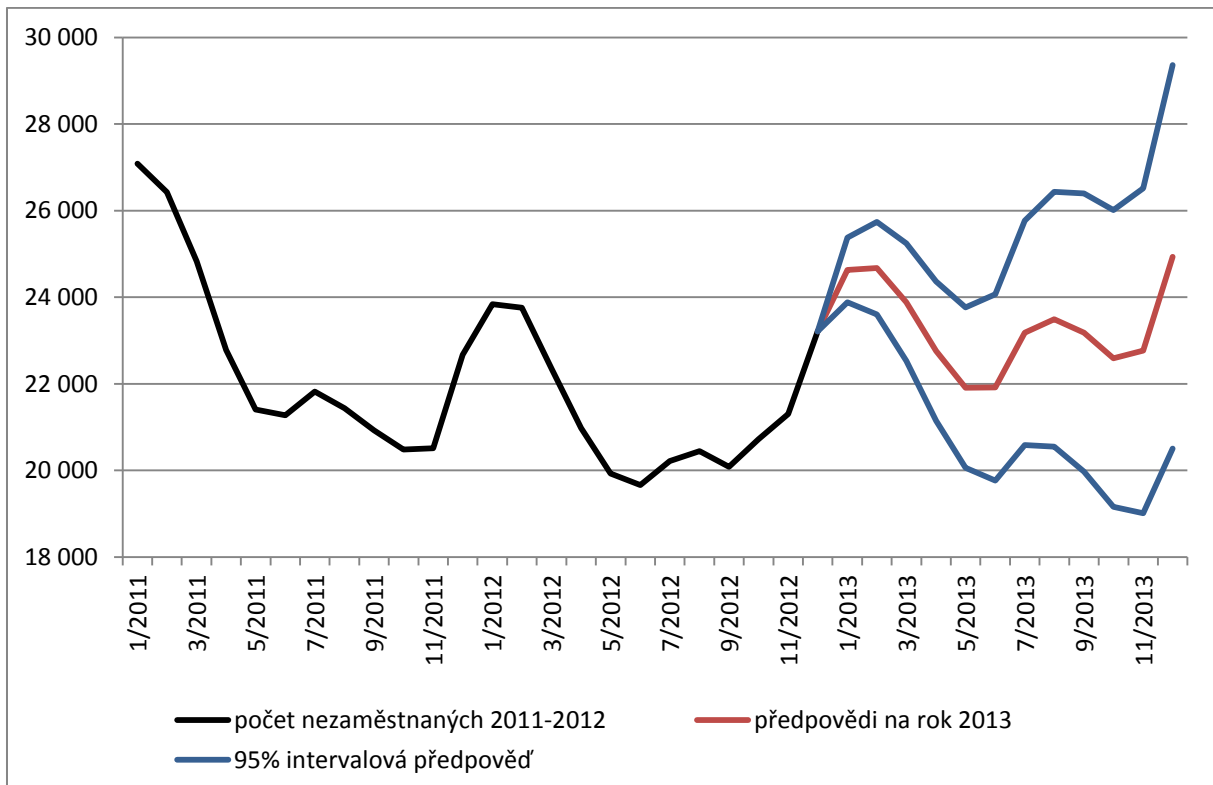
Obrázek A. 2: Předpovědi pro Středočeský kraj



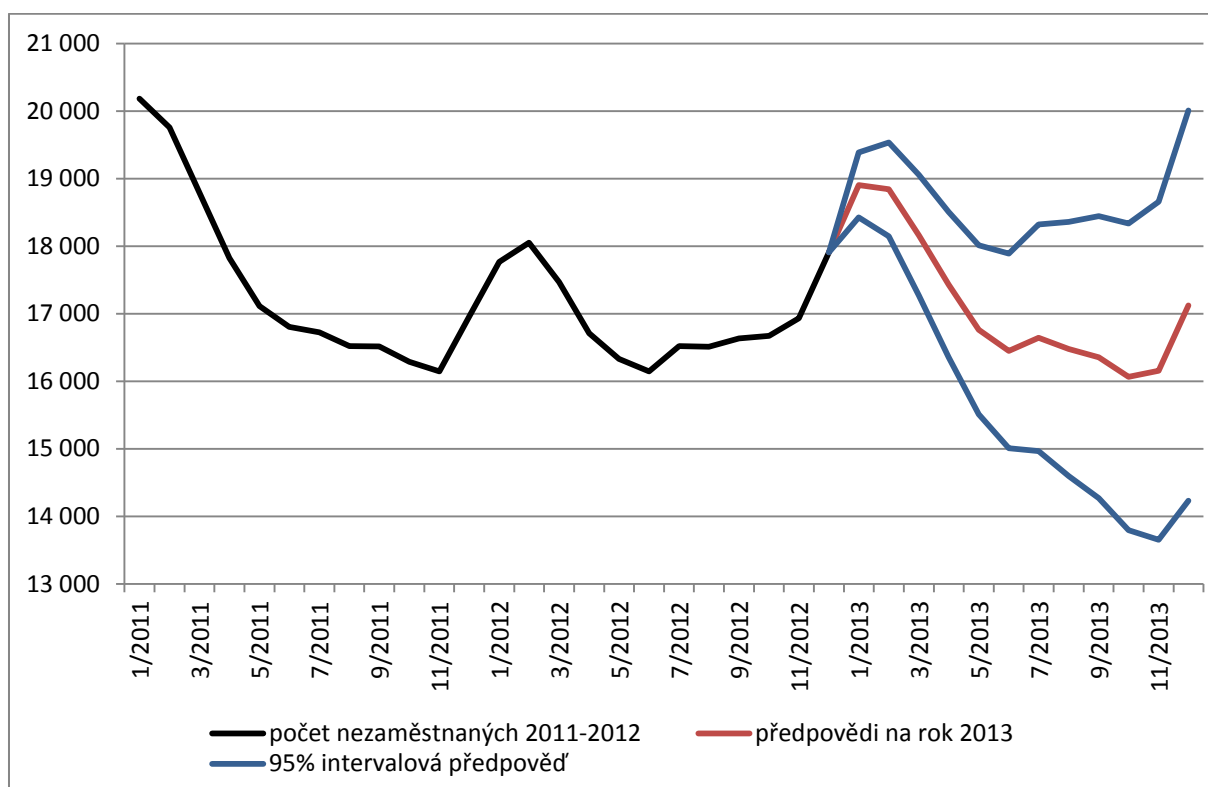
**Obrázek A. 3: Předpovědi pro Jihočeský kraj**



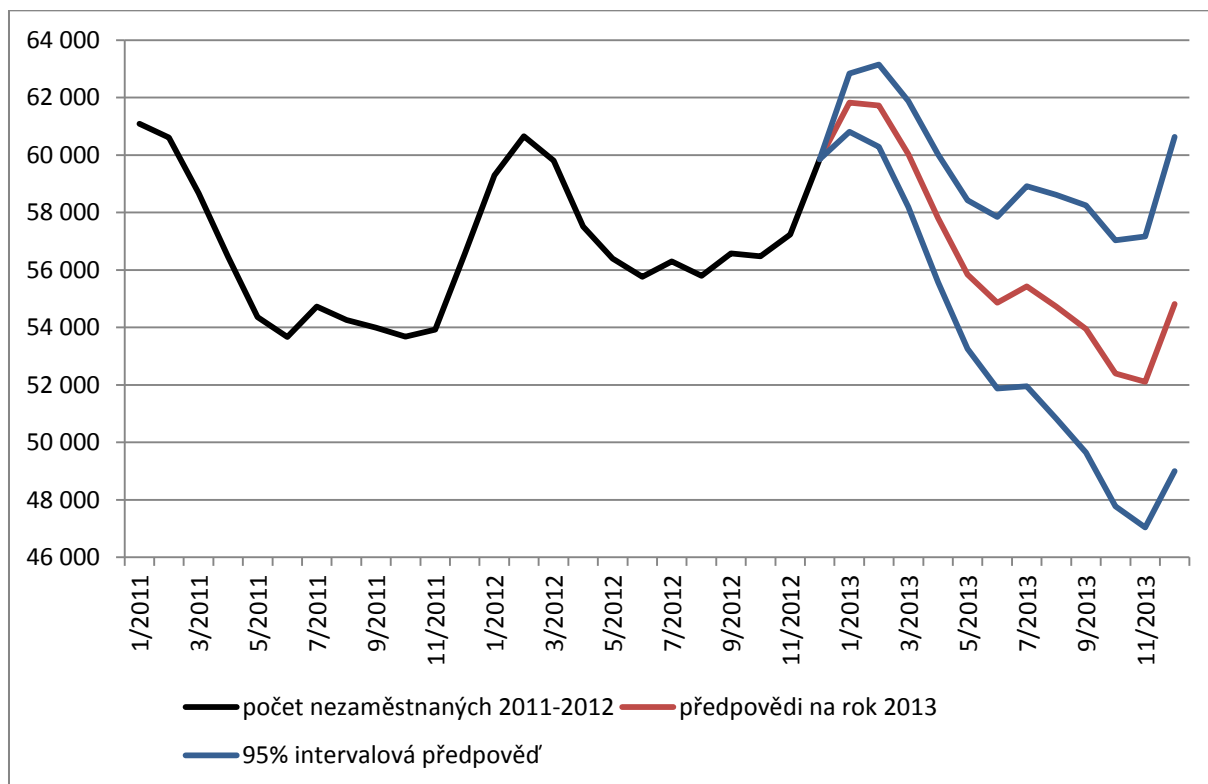
**Obrázek A. 4: Předpovědi pro Plzeňský kraj**



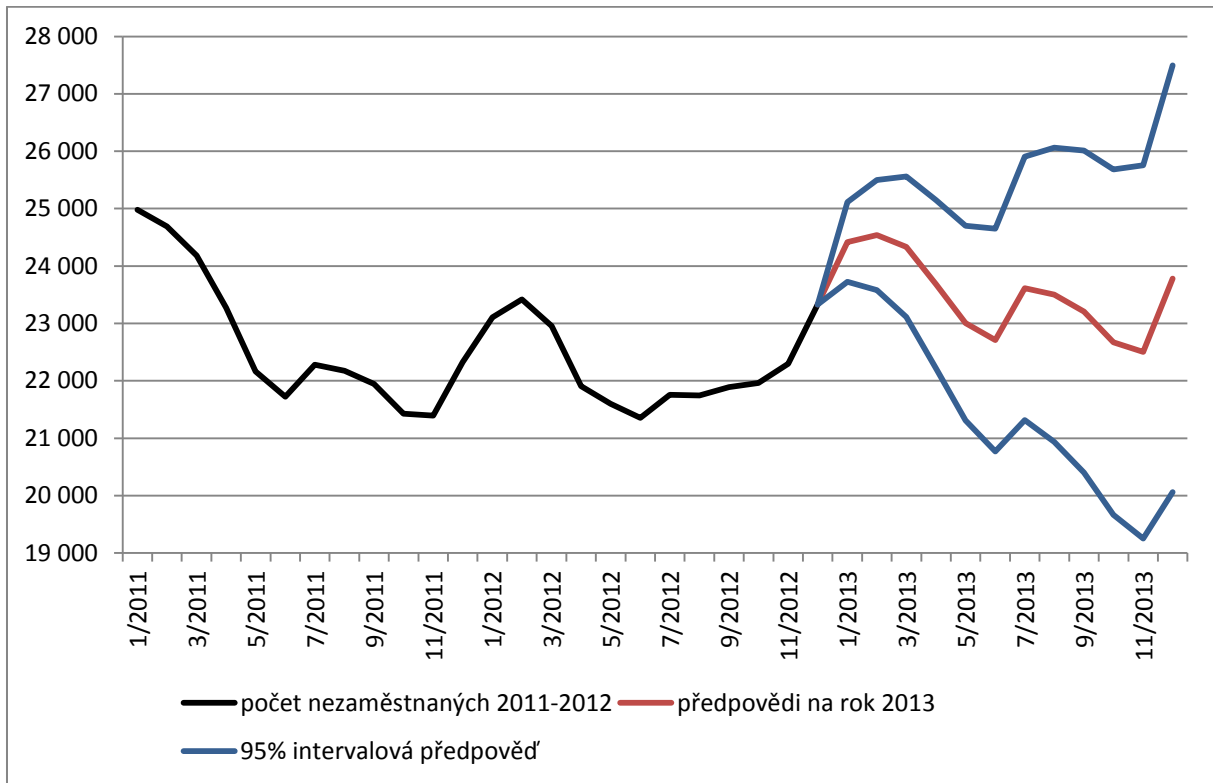
**Obrázek A. 5: Předpovědi pro Karlovarský kraj**



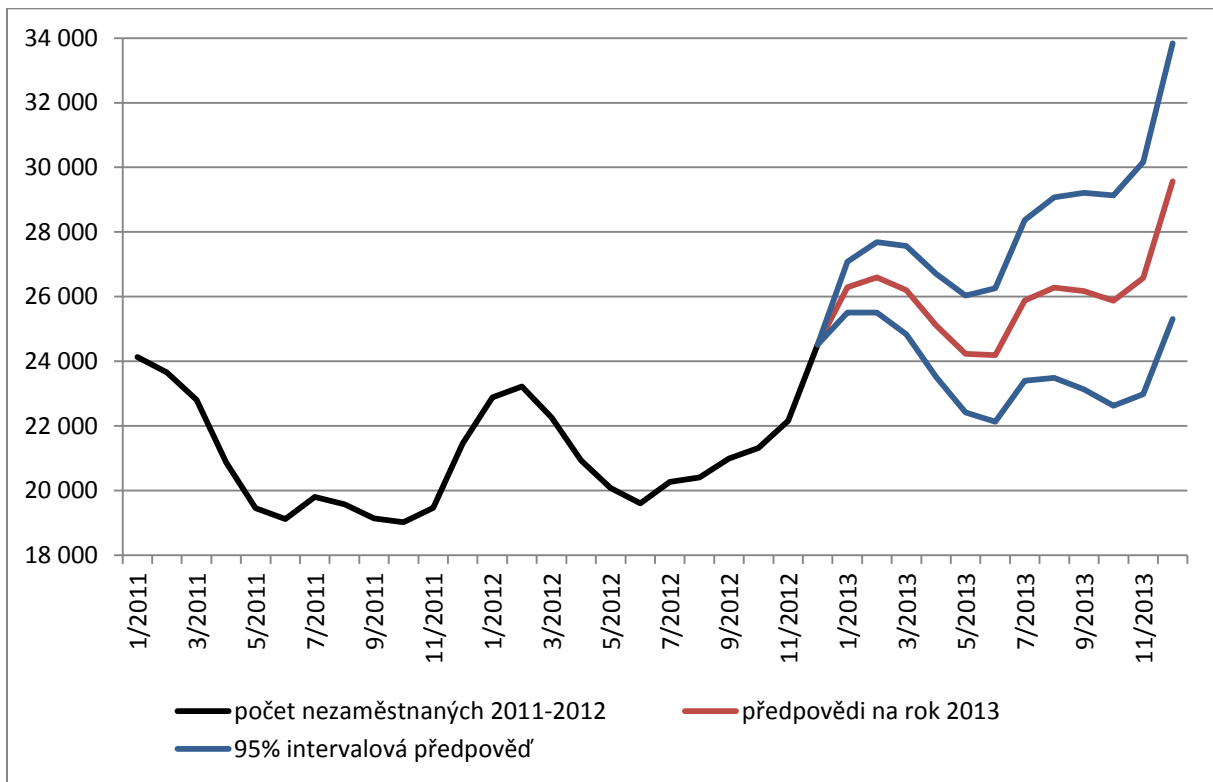
**Obrázek A. 6: Předpovědi pro Ústecký kraj**



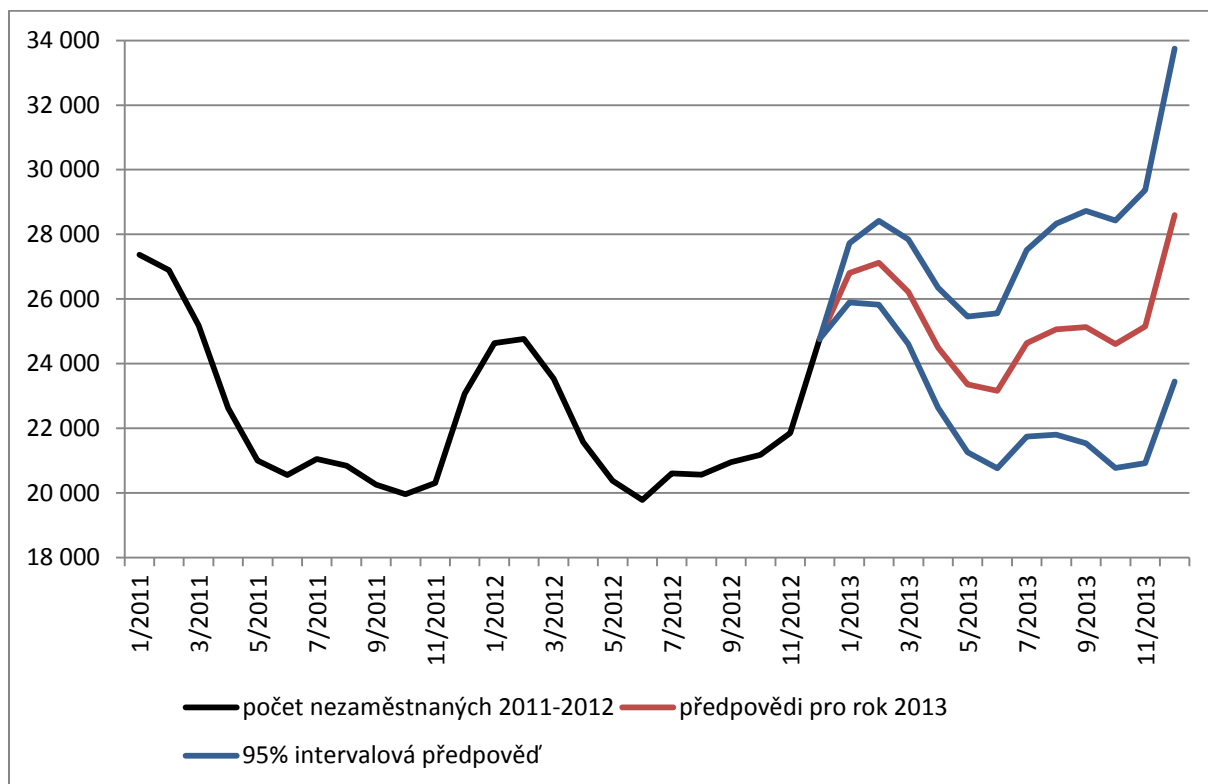
**Obrázek A. 7: Předpovědi pro Liberecký kraj**



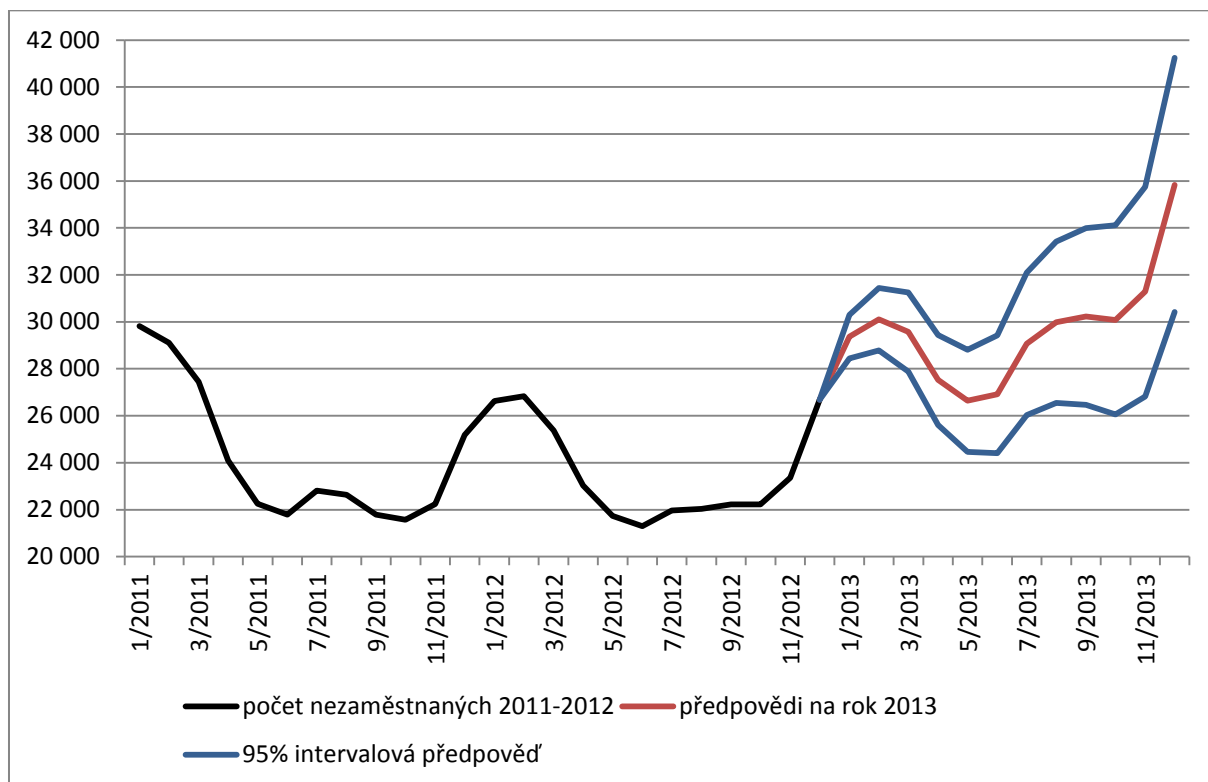
**Obrázek A. 8: Předpovědi pro Královehradecký kraj**



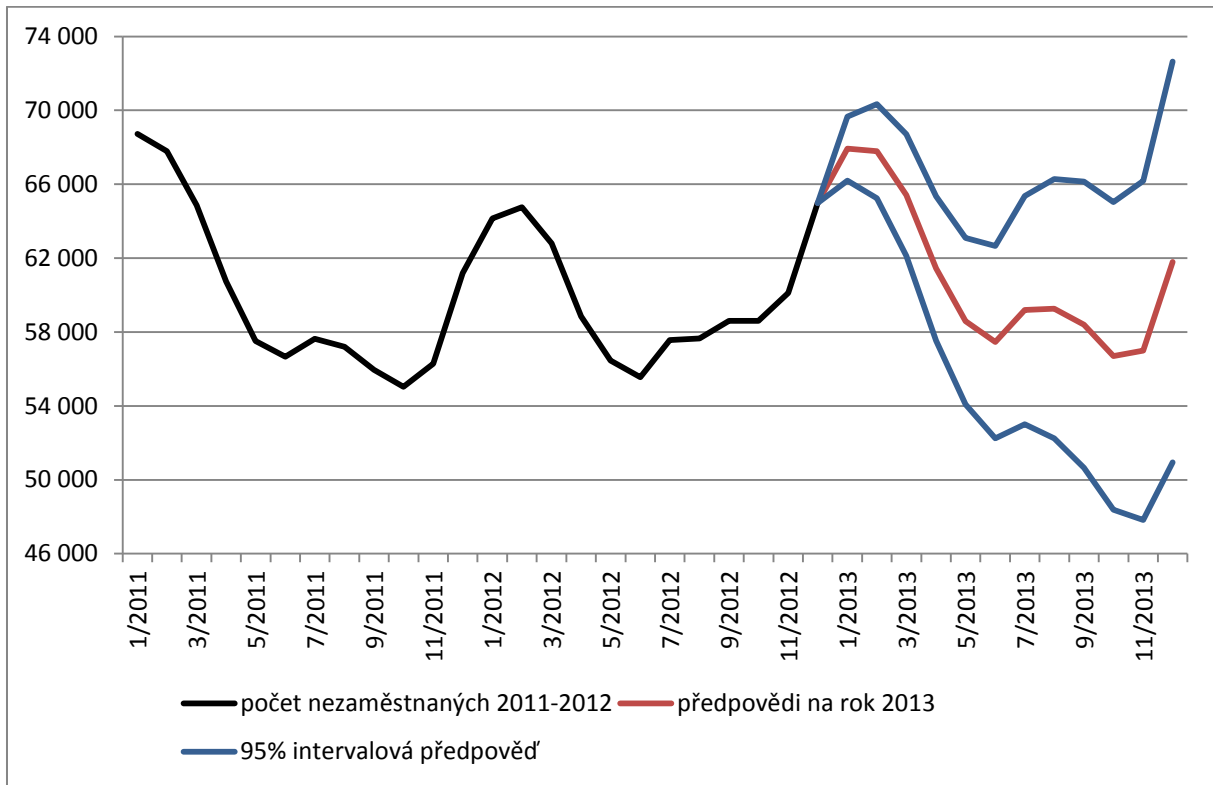
**Obrázek A. 9: Předpovědi pro Pardubický kraj**



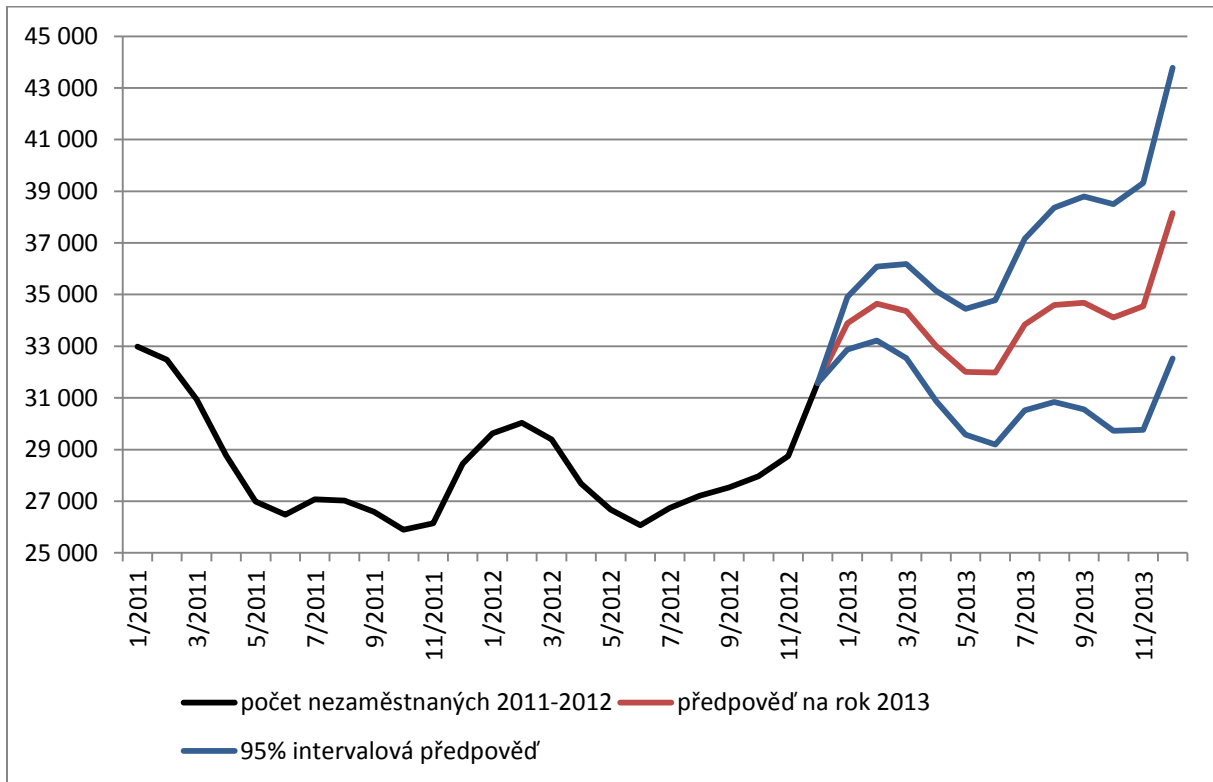
**Obrázek A. 10: Předpovědi pro kraj Vysočina**



**Obrázek A. 11: Předpovědi pro Jihomoravský kraj**

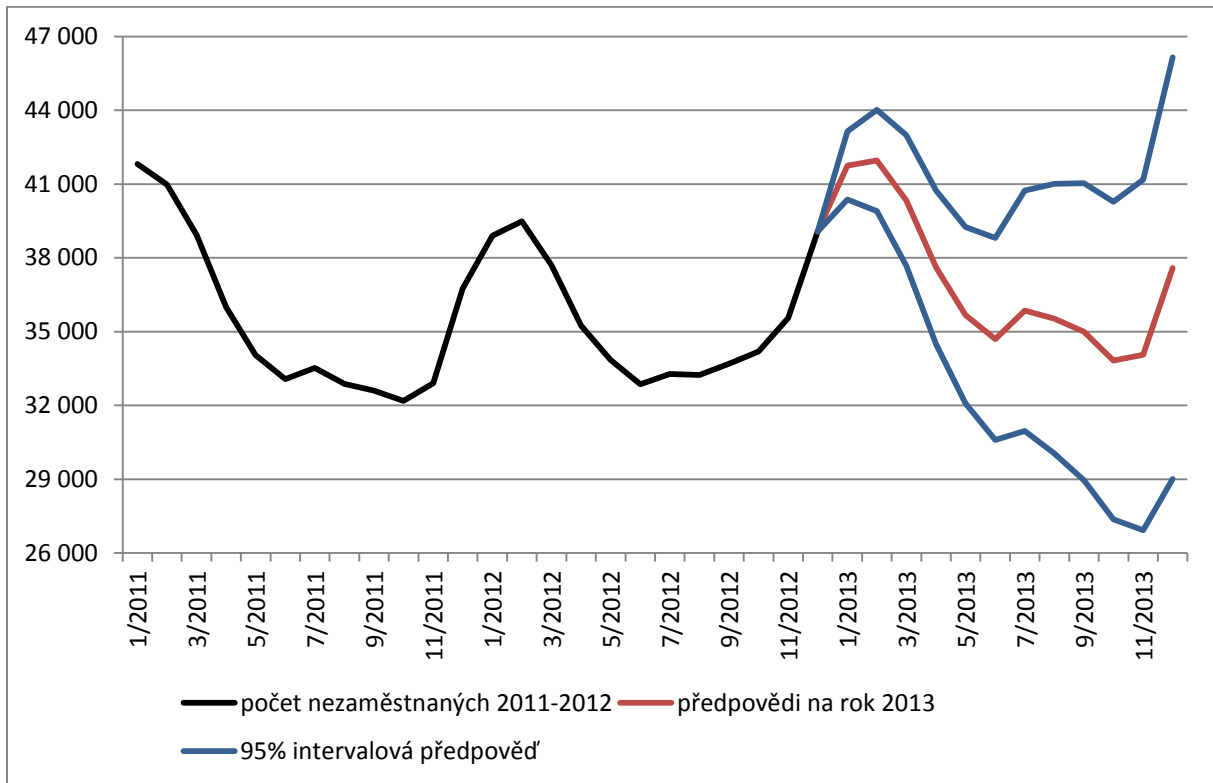


**Obrázek A. 12: Předpovědi pro Zlínský kraj**

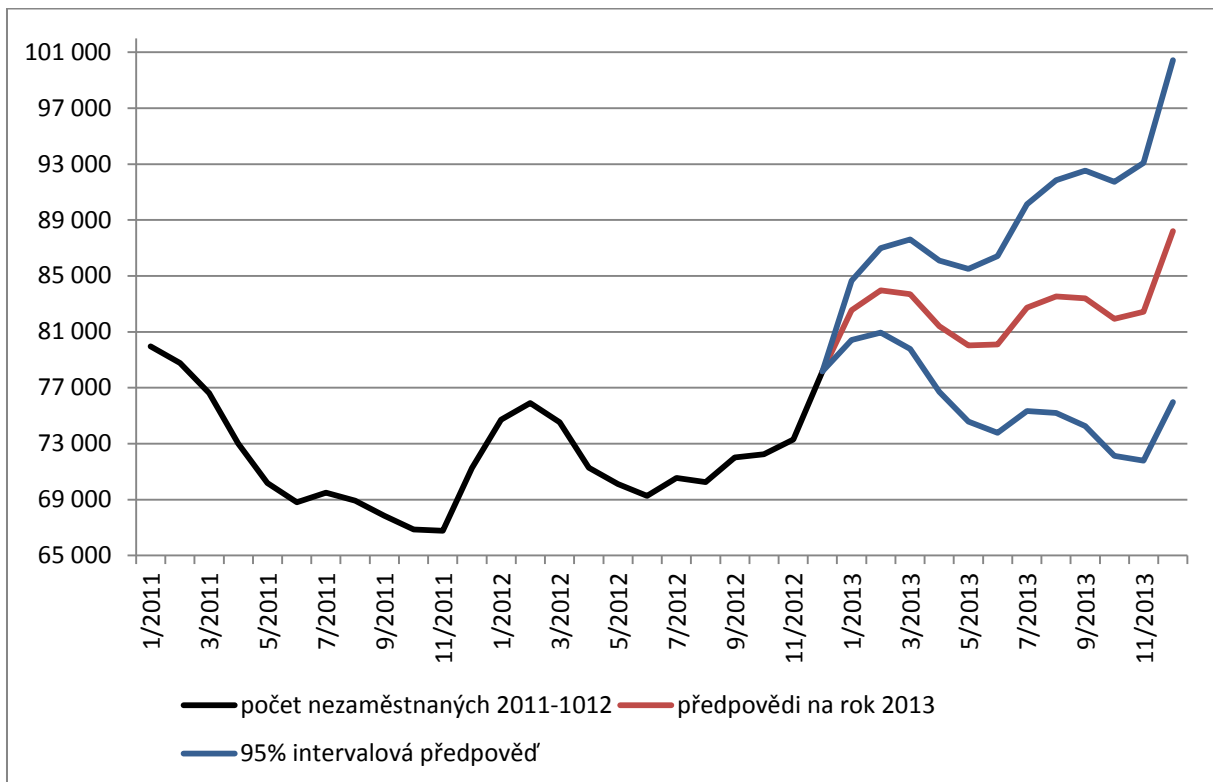




**Obrázek A. 13: Předpovědi pro Olomoucký kraj**



**Obrázek A. 14: Předpovědi pro Moravskoslezský kraj**



## **B Obsažené na CD**

**BP\_Svobodová\_A09B0139P.pdf** – soubor s kompletní bakalářskou prací v elektronické podobě

**data.xlsx** – Excel soubor obsahující zdrojová data

**předpovědní model** – složka obsahující předpovědní modely - autoregresní\_model.xlsx, dvojité\_vyrovnání.xlsx a sezónní očišťování KP\_2005-2011.xlsx

**kraje** – složka obsahující modelování časových řad všech krajů a celé ČR a kontrolu celku