

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

BAKALÁŘSKÁ PRÁCE

Analýza periodické složky v časové řadě



Katedra matematické analýzy a aplikací matematiky

Vedoucí bakalářské práce: **RNDr. et PhDr. Ivo Müller Ph.D.**

Vypracovala: **Žaneta Janktová**

Studijní program: B1103 Aplikovaná matematika

Studijní obor: Matematika-ekonomie se zaměřením na bankovníctví

Forma studia: prezenční

Rok odevzdání: 2017

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Žaneta Janktová

Název práce: Analýza periodické složky v časové řadě

Typ práce: Bakalářská

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: RNDr. et PhDr. Ivo Müller Ph.D.

Rok obhajoby práce: 2017

Abstrakt: Bakalářská práce představí základní modely vhodné pro zpracování periodických časových řad a konstrukci předpovědí, a to nejen teoreticky, ale i prakticky na konkrétních datech.

Klíčová slova: časové řady, analýza, trend, sezónní složka, periodicitu, periodogram, predikce

Počet stran: 65

Počet příloh: 0

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Žaneta Janktová

Title: Analysis of periodic component in time series

Type od thesis: Bachelor's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: RNDr. et PhDr. Ivo Müller Ph.D.

The year of presentation: 2017

Abstract: My bachelor theses presents the basic models suitable for processing of time series and designing of predictions not only in theory but also in real life on specific data sets (examples).

Key words: time series, analysis, trend, seasonal component, periodicity, periodogram, prediction

Number of pages: 65

Number of appendices: 0

Language: Czech

Poděkování

Ráda bych poděkovala především svému vedoucímu bakalářské práce panu RNDr. et PhDr. Ivo Müllerovi, Ph.D. za cenné rady, připomínky a za čas, který mi věnoval při konzultacích. Dále bych chtěla poděkovat své rodině a přátelům, kteří mě po celou dobu studia podporovali.

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně pod vedením RNDr. et PhDr. Ivo Müllera Ph.D. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

V Olomouci dne

.....

Podpis

OBSAH

ÚVOD	7
1 Časová řada.....	8
1.1 Základní dělení ČR	8
1.2 Analýza časové řady	9
2 Analytický popis trendu	12
2.1 Popis některých trendových funkcí.....	12
3 Analýza periodické složky	16
3.1 Pojmy a značení	16
3.2 Model skrytých period	17
3.2.1 Periodogram.....	17
3.2.2 Test R. A. Fishera	18
4 Modelování sezónnosti.....	20
4.1 Model konstantní sezónnosti.....	20
4.2 Model proporcionální sezónnosti.....	21
5 Míry vhodnosti modelu	23
6 Předpovědi v ČR.....	25
PRAKTICKÁ ČÁST.....	28
7 Sňatkovost	28
7.1 Model skrytých period (MSP) a předpovědi.....	30
7.2 Testování parametru lineárního regresního modelu	35
7.3 Model konstantní sezónnosti (MKS) a předpovědi	36
7.4 Model proporcionální sezónnosti (MPS) a předpovědi	39
8 Nezaměstnanost	43
8.1 Model skrytých period a předpovědi	44
8.2 Model konstantní sezónnosti a předpovědi.....	48
8.3 Test přítomnosti sezónnosti	50
9 Denní teploty	51
9.1 Trend	51
9.2 Model skrytých period	52
ZÁVĚR	56
LITERATURA	58
PŘÍLOHA	60

ÚVOD

Analýza časových řad, především předpovídání jejich budoucího chování, je velmi důležitou oblastí současné statistiky. Pomocí metod, které se užívají při analýze časových řad, je možné velmi dobře popsat většinu skutečných časových řad z ekonomické praxe. Na základě skutečných hodnot lze konstruovat model, který popisuje danou časovou řadu, a dále pak využít tento model pro predikci budoucího vývoje systému. V dnešní době je analýza časových řad ulehčena především tím, že máme k dispozici velké množství dat z různých oblastí. Největším zdrojem v ČR je Český statistický úřad, vydávající množství dat, která jsou volně k dispozici na internetových stránkách, případně v publikacích.

Cílem této bakalářské práce je bližší seznámení s periodickou složkou v časové řadě. Nejprve jsem měla nastudovat základní metody vhodné pro zpracování časových řad, které obsahují periodickou složku. Tyto metody a modely následně vyzkoušet na konkrétní datové sady a zkonstruovat předpovědi pro další období. Na závěr případně porovnat či určit, který model je pro vybrané datové sady nejvhodnější, a to mít samozřejmě podložené konkrétními výpočty.

Celá práce je rozdělena do dvou velkých částí, tj. teoretická a praktická část. Celkově obsahuje devět kapitol a několik podkapitol pro lepší přehlednost. V teoretické části se zabývám základními pojmy z oblasti časových řad, následně jsou blíže popsány jednotlivé modely, které jsou pak využity v praktické části. Vybrala jsem pro svou práci ty, se kterými jsme se seznámili během studia v rámci předmětu Časové řady. Ze sezónních modelů je to model skrytých period, model konstantní sezónnosti a model proporcionální sezónnosti. Součástí teorie jsou i míry vhodnosti modelu a předpovědi v časových řadách, které jsem stále využívala při praktických výpočtech, proto zde nesmějí chybět.

Pro praktickou část jsem vybrala tři datové sady, jež jsou rozděleny samostatně do kapitol. Na každou datovou sadu jsem využila některé metody pro analýzu, a následně konstrukci předpovědí, přičemž jsem vycházela z metod a postupů uvedených v teoretické části.

1 Časová řada

Časová řada je posloupnost věcně a prostorově srovnatelných pozorování (dat). Data, která tvoří časovou řadu, jsou v čase uspořádána ve směru minulost – přítomnost. Jako příklad třeba z oblasti makroekonomie lze uvést časovou řadu hrubého domácího produktu za jednotlivá čtvrtletí roku, nezaměstnanost obyvatel, nebo například objem výroby nějakého podniku za jednotlivé roky. V případě, kdy jsou sledovanými daty hodnoty některého ekonomického ukazatele, můžeme hovořit o ekonomických časových řadách.

Jednotlivá pozorování jsou přiřazena časovým okamžikům. Pozorování y v čase t tedy značíme y_t , kde $t = 1, 2, \dots, n$ pro ekvidistantní časové řady, pro neekvidistantní řady se používá značení y_{t_i} , kde $t_i = t_1, t_2, \dots, t_n$. Pojem ekvidistantní, resp. neekvidistantní časová řada je vysvětlen v následující podkapitole 1.1. V celé kapitole 1 jsem využila především literatury [1], [2], [3] a [4]

1.1 Základní dělení ČŘ

Rozlišujeme *intervalové* a *okamžikové* časové řady. *Intervalová* časová řada odpovídá na otázku „kolik čeho vzniklo, nebo zaniklo“. Vzniká agregací za určité období (interval) a hodnoty závisí na délce intervalu. Jednotlivá pozorování lze sčítat, součty dávají smysl. Naopak součty pozorování *okamžikových* časových řad nedávají smysl. Vznikají diskretizací spojité veličiny, nebo jsou diskrétní ze své podstaty. Říkají nám „kolik čeho existuje“. Většinou se hodnoty vztahují ke konci nebo počátku nějakého období.

Dále lze uvažovat *krátkodobé* časové řady s intervaly mezi pozorováními kratšími než jeden rok, tj. čtvrtletní, měsíční, týdenní atd., a *dlouhodobé* časové řady, kdy jsou hodnoty sledovány v ročních či delších časových úsecích.

Dle vzdálenosti mezi jednotlivými časovými okamžiky dělíme časové řady na již zmíněné *ekvidistantní* časové řady, které se vyznačují konstantní časovou vzdáleností mezi jednotlivými hodnotami, a *neekvidistantní* časové řady, jež mají různou časovou vzdálenost mezi jednotlivými hodnotami.

1.2 Analýza časové řady

Hlavním cílem analýzy časové řady je většinou konstrukce odpovídajícího modelu, na základě kterého můžeme vytvořit předpovědi, a to pomocí vhodného použití souboru metod. Některým z těchto metod se budu později v této práci detailněji věnovat.

Známe-li model, můžeme předpovídat budoucí vývoj časové řady. Například bychom mohli predikovat poptávku po určitém výrobku na následující roky. Konstrukce modelu nám také umožní do jisté míry řídit a optimalizovat činnost příslušného systému vhodnou volbou vstupních parametrů a počátečních podmínek. [1]

Existuje velké množství různých způsobů modelování časových řad. Výběr metody pro analýzu časové řady závisí na několika faktorech, z nichž nejdůležitější jsou účel analýzy, typ časové řady a zkušenosti statistika, který analýzu provádí, s čímž souvisí také dostupná výpočetní technika a programové vybavení.

Nyní bych stručně nastínila hlavní myšlenku tří přístupů k modelování časových řad, a to klasického dekompozičního přístupu, Boxovy – Jenkinsovy metodologie a spektrální analýzy.

- **Klasický dekompoziční přístup**

Jak už je z názvu patrné, jedná se o rozložení (dekompozici) časové řady na jednotlivé složky. Rozklad provádíme, abychom lépe a snadněji rozpoznali pravidelné chování řady.

Časovou řadu lze rozložit na čtyři složky. Základní složkou je *trend*, který značíme písmenem T_t . Další složkou je *periodická*, která zahrnuje *cyklickou* a *sezónní* složku, které budeme zapisovat C_t (cyklická) a S_t (sezónní). Poslední částí je *reziduální* složka, kterou značíme ϵ_t . Index t ve všech uvedených složkách značí opět jednotlivé časové okamžiky $t = 1, 2, \dots, n$.

Trendová složka určuje základní tendence vývoje (klesající, rostoucí). Měla by zachycovat dlouhodobé změny v chování časové řady.

Sezónní složka popisuje periodické změny v časové řadě, kdy délka periody nepřesahuje jeden rok. Sezónní výkyvy jsou pravidelně se opakující jevy, které jsou způsobeny především takovými faktory, jako je střídání ročních období, lidské zvyky atd.

Cyklická složka na rozdíl od sezónní složky zachycuje dlouhodobé kolísání kolem trendu. Perioda i amplituda je poněkud proměnlivá, proto se někteří autoři vyhýbají nazvat tuto složku cyklickou (nebo dokonce periodickou) a hovoří spíše o fluktuacích okolo trendu, v nichž se střídá fáze růstu s fází poklesu. [1] *Cyklická složka* je proměnlivější než sezónní složka, je velmi obtížné identifikovat příčiny vedoucí k jejímu vzniku. Příkladem by mohl být hospodářský cyklus nebo cyklus v módě.

Reziduální (náhodná) složka zbývá v časové řadě po očištění od trendové a periodické (sezónní a cyklické) složky. Rezidua jsou tedy drobné nesystematické odchylky. Má oproti předcházejícím složkám jiný charakter. Předpokládá se obvykle, že jde o náhodou veličinu, tzn. lze ji popsat rozdělením pravděpodobnosti, které je v ideálním případě normální. V některých pramenech se označuje jako tzv. bílý šum.

Tvar rozkladu může být dvojího typu. *Aditivní dekompozice* je dána součtem jednotlivých složek časové řady. Pro jednotlivé časové okamžiky $t = 1, 2, \dots, n$ můžeme zapsat jako

$$y_t = T_t + S_t + C_t + E_t, \quad (1.1)$$

multiplikativní dekompozice je dána jako součin všech složek a má tvar

$$y_t = T_t S_t C_t E_t. \quad (1.2)$$

- **Přístup Boxe a Jenkinse**

Tato metodologie uvažuje jako hlavní prvek konstrukce modelu časové řady reziduální neboli náhodnou složku. Náhodná složka se podle Boxe a Jenkinse považuje za systém náhodných veličin korelovaných v čase, tzn. základním nástrojem je právě korelační analýza. Pro modelování tohoto typu se vyžaduje mít alespoň 50 pozorování.

Jako příklad mohu uvést tzv. model klouzavých součtů (značíme MA), který je jedním z nejjednodušších modelů, s nimiž tato metodologie pracuje. Jiným typem jsou tzv. autoregresní modely (značíme AR) a smíšené modely (ARMA). Existují také integrované

modely ARIMA a tzv. sezónní modely, do kterých zahrnujeme také trendovou a sezónní složku. Tyto modely jsou daleko flexibilnější než dekompoziční modely, tzn. mnohem lépe a rychleji reagují na změněný charakter časové řady.

- **Spektrální analýza**

Jedná se o zcela odlišný přístup analýzy časových řad. Zkoumaná řada se považuje za (nekonečnou) směs sinusových a kosinusových křivek s různými frekvencemi a amplitudami. Na rozdíl od výše uvedených případů, kdy se jednalo o analýzu v časové doméně, zde přecházíme do frekvenční domény. Základním nástrojem je Fourierova transformace, která slouží pro převod signálů z časové oblasti do oblasti frekvenční. Pomocí speciálních statistických nástrojů, jako je např. periodogram nebo spektrální hustota, můžeme při spektrální analýze získat obraz o intenzitě zastoupení jednotlivých frekvencí v časové řadě (tj. o spektru řady). Následně lze vystopovat významné frekvence, které pak lze použít pro odhad koeficientů periodických složek odpovídajících těmto frekvencím.

2 Analytický popis trendu

Pro jednoduchost zatím nebudeme uvažovat periodicitu, máme tedy neperiodickou časovou řadu, kterou můžeme vyjádřit jako součet *trendové složky* T_t a *náhodné složky* ε_t , jednotlivá pozorování v čase t značíme $y_t = T_t + \varepsilon_t$, kde $t = 1, 2, \dots, n$. Někdy lze popsat trend analyticky některými jednoduchými matematickými křivkami. Cílem je odhadnout neznámé parametry dané trendové funkce.

V této kapitole jsem čerpala především z literatury [1] a [4].

2.1 Popis některých trendových funkcí

Typ nejvhodnější matematické křivky pro danou časovou řadu určujeme na základě předběžného rozboru, nejčastěji pomocí grafického záznamu řady, nebo na základě předpokládaných vlastností trendové složky. [1]

- **Konstantní trend**

Tento typ trendu je nejjednodušší funkcí. Funkce neroste ani neklesá, je konstantní. Jako a_0 označíme absolutní člen, trend je roven v každém časovém okamžiku $t = 1, 2, \dots, n$ právě této hodnotě. Lze zapsat jako $T_t = a_0$, kde parametr $a_0 \in R$. Pro jednotlivá pozorování v čase t pak platí vztah

$$y_t = a_0 + E_t, \quad (2.1)$$

pro $t = 1, 2, \dots, n$ v případě ekvidistantní řady, $t = t_1, t_2, \dots, t_n$ pro neekvidistantní řady.

Abychom mohli odhadnout hodnoty konstantního trendu, je potřeba určit odhad parametru a_0 (odhad značíme \hat{a}_0). Můžeme ho zjistit pomocí tzv. *metody nejmenších čtverců*. Tato metoda spočívá v tom, že chceme minimalizovat reziduální součet čtverců. Tzn. hledáme takové koeficienty, při kterých bude výraz

$$S_e = \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (2.2)$$

minimální. V rovnici je y_t pozorování v čase t , \hat{y}_t značí odhadnuté hodnoty časové řady.

V případě konstantního trendu (2.1) tedy hledáme $\min_{a_0} \sum_{t=1}^n (y_t - a_0)^2$.

Provedeme derivaci podle a_0

$$\frac{\partial}{\partial a_0} = -2 \sum_{t=1}^n (y_t - a_0).$$

Derivaci nyní položíme rovnu 0 a tím získáme tzv. *normální rovnici*

$$\sum_{t=1}^n y_t - na_0 = 0.$$

Řešením této rovnice je odhad parametru a_0 ve tvaru

$$\hat{a}_0 = \frac{1}{n} \sum_{t=1}^n y_t = \bar{y}. \quad (2.3)$$

Pro vyrovnané hodnoty $\hat{y}_t, t = 1, 2, \dots, n$, pak platí $\hat{y}_t = \hat{a}_0$.

- **Lineární trend**

Lineární trend je ve tvaru *přímky* a lze jej zapsat jako $T_t = a_0 + a_1 t$, kde parametry $a_0, a_1 \in R, t \in R$. Pro jednotlivá pozorování v čase t pak platí vztah

$$y_t = a_0 + a_1 t + \varepsilon_t, \quad (2.4)$$

pro $t = 1, 2, \dots, n$ v případě ekvidistantní řady, $t = t_1, t_2, \dots, t_n$ pro neekvidistantní řady.

Odhady parametrů \hat{a}_0 a \hat{a}_1 dostaneme opět *metodou nejmenších čtverců* jako v případě konstantního trendu. Jelikož v praktické části pracuji často s lineárním trendem, je nutné opět určit odhady parametrů \hat{a}_0 a \hat{a}_1 .

V případě lineárního trendu (2.4) tedy hledáme $\min_{a_0, a_1} \sum_{t=1}^n (y_t - a_0 - a_1 t)^2$.

Provedeme derivaci postupně podle \hat{a}_0 a \hat{a}_1

$$\frac{\partial}{\partial a_0} = -2 \sum_{t=1}^n (y_t - a_0 - a_1 t),$$

$$\frac{\partial}{\partial a_1} = -2 \sum_{t=1}^n t (y_t - a_0 - a_1 t).$$

Derivace položíme rovny nule a dostaneme následující *soustavu normálních rovnic*

$$a_0 n + a_1 \sum_{t=1}^n t = \sum_{t=1}^n y_t,$$

$$a_0 \sum_{t=1}^n t + a_1 \sum_{t=1}^n t^2 = \sum_{t=1}^n t y_t.$$

Řešením této soustavy jsou odhady \hat{a}_0 a \hat{a}_1 parametrů a_0, a_1

$$\hat{a}_1 = \frac{\sum_{t=1}^n t y_t - \bar{t} \sum_{t=1}^n y_t}{\sum_{t=1}^n t^2 - n \bar{t}^2}, \quad (2.5.)$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{t}, \quad (2.6)$$

kde $\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$, $\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i$ pro neekvidistantní řady, $\bar{t} = \frac{n+1}{2}$ pro ekvidistantní řady. Pro vyrovnané hodnoty \hat{y}_t , $t = 1, 2, \dots, n$, pak platí $\hat{y}_t = \hat{a}_0 + \hat{a}_1 t$.

- **Kvadratický trend**

Tento typ trendu se někdy nazývá parabolický trend. Jak je již z názvů zřejmé, křivka, která zobrazuje kvadratický trend, je *parabola*. Můžeme jej zapsat ve tvaru $T_t = a_0 + a_1 t + a_2 t^2$, kde $a_0, a_1, a_2 \in R$, $t \in R$. Po zahrnutí náhodné složky pak pro časovou řadu, která je dána kvadratickým trendem, platí vztah

$$y_t = a_0 + a_1 t + a_2 t^2 + E_t, \quad (2.7)$$

kde $t = 1, 2, \dots, n$ pro ekvidistantní řady, $t = t_1, t_2, \dots, t_n$ pro neekvidistantní řady.

Opět nás budou zajímat odhady \hat{a}_0 , \hat{a}_1 a \hat{a}_2 , které můžeme zjistit pomocí *metody nejmenších čtverců* jako v předchozích případech. Minimalizujeme výraz (2.2), vypočítáme postupně derivace podle a_0, a_1, a_2 a položíme je rovny nule. Získáme soustavu normálních rovnic, jejímž řešením jsou dle [4] odhady $\hat{a}_0, \hat{a}_1, \hat{a}_2$

$$\hat{a}_2 = \frac{(\sum_{t=1}^n t^2 y_t - \bar{y} \sum_{t=1}^n t^2) - (n+1)(\sum_{t=1}^n t y_t - \bar{y} \sum_{t=1}^n t)}{\frac{n}{180}(n^2-1)(n^2-4)}, \quad (2.8)$$

$$\hat{a}_1 = \frac{\sum_{t=1}^n t y_t - \frac{n(n+1)}{2} \bar{y}}{\frac{n}{12}(n^2-1)} - (n+1) \hat{a}_2, \quad (2.9)$$

$$\hat{a}_0 = \bar{y} - \frac{n+1}{2} \hat{a}_1 - \frac{1}{6} (n+1)(2n+1) \hat{a}_2. \quad (2.10)$$

Pro vyrovnané hodnoty pak platí $\hat{y}_t = \hat{a}_0 + \hat{a}_1 t + \hat{a}_2 t^2$, pro $t = 1, 2, \dots, n$.

- **Další typy trendů**

Existuje několik dalších typů trendů. Dalšími dosti využívanými jsou *exponenciální trend*, *modifikovaný exponenciální trend*, *logistický trend*, *Gompertzova křivka*. Tyto ve své

práci blíží popisovat nebudu, jelikož v praktické části jsem si vystačila pouze s konstantním, lineárním a kvadratickým trendem. Bližší popis jednotlivých trendů včetně odvození jednotlivých vzorců lze nalézt v literatuře [1, str. 30-41].

3 Analýza periodické složky

Periodickou složkou rozumíme cyklickou a sezónní složku časové řady. Jak jsem již zmínila v kapitole 1.2, sezónní složka zobrazuje krátkodobé kolísání kolem trendu, cyklická složka naopak dlouhodobé. V praxi se často setkáváme s časovými řadami, kde mohou být přítomny obě složky současně.

Modely, ve kterých uvažujeme navíc periodickou složku, jsou podstatně flexibilnější než ty, které zahrnují pouze trendovou složku. Je možné provést *izolaci sezónní složky* a následně její rozbor. Ten nám například u ekonomických časových řad může rozšířit naše znalosti o zákonitostech chování určitého ekonomického jevu. Stejně tak lze provést rozbor *izolované cyklické složky*. Další možností je očištění (eliminace) časové řady od sezónnosti, tzv. *sezónní očišťování*.

Konstrukce samotného modelu může být poněkud složitější než klasické určení trendu, avšak na základě takto sestaveného modelu periodické časové řady můžeme získat mnohem přesnější předpovědi pro budoucí období. V následujících podkapitolách si představíme některé z těchto modelů. V této kapitole jsem použila především zdroje [1], [2], [4] a [5].

3.1 Pojmy a značení

Základní veličina, která popisuje periodické chování, je *perioda* T . Je to doba, po které se pohyb znovu opakuje, tzn. T udává trvání jednoho cyklu (délku periody). Odtud se pak získá *frekvence* $\frac{1}{T}$, která udává počet cyklů za časovou jednotku. Při analýze časových řad budeme pracovat s *úhlovou frekvencí* $\omega = \frac{2\pi}{T}$. Udává se v radiánech za uvažovanou jednotku času, kterou je časový interval mezi dvěma sousedními pozorováními. Čím je frekvence funkce větší, tím častěji se v jejím průběhu střídají jednotlivé cykly [1].

3.2 Model skrytých period

Časovou řadu, respektive její hodnoty, lze popsat směsí sinusovek a kosinusovek o různých amplitudách a frekvencích. V modelu y_t uvedeném níže, si lze představit, že se jedná o součet H harmonických členů, tzv. harmonik.

Matematicky je model vyjádřen jako polynom ve tvaru

$$y_t = a_0 + \sum_{j=1}^H a_j \sin(\omega_j t) + \sum_{j=1}^H b_j \cos(\omega_j t), \quad (3.1)$$

kde $H = \frac{n}{2}$ pro n sudé, $H = \frac{(n-1)}{2}$ pro n liché, n je počet pozorování neboli délka základní periody, j indexuje periody (frekvence), úhlové frekvence jsou značeny $\omega_j = \frac{2\pi j}{n}$, $j = 1, \dots, H$, a a_0, a_j, b_j jsou parametry, jejichž hodnoty je nutné odhadnout.

Odhady parametrů provedeme *metodou nejmenších čtverců*. Chceme tedy minimalizovat výraz (2.2), čili hledáme

$$\min_{a_0, a_j, b_j} \sum_{t=1}^n (y_t - a_0 - \sum_{j=1}^H a_j \sin(\omega_j t) - \sum_{j=1}^H b_j \cos(\omega_j t))^2.$$

Zderivujeme postupně podle a_0, a_j, b_j a derivace položíme rovny nule. Získáme soustavu normálních rovnic, jejímž řešením jsou odhady parametrů

$$\hat{a}_0 = \frac{1}{n} \sum_{t=1}^n y_t = \bar{y}, \quad (3.2)$$

$$\hat{a}_j = \frac{2}{n} \sum_{t=1}^n y_t \sin(\omega_j t), \quad (3.3)$$

$$\hat{b}_j = \frac{2}{n} \sum_{t=1}^n y_t \cos(\omega_j t), \quad (3.4)$$

kde $j = 1, \dots, H$.

V praxi se do modelu nezahrnují všechny periody $\omega_1, \omega_2, \dots, \omega_H$, ale pouze ty významné. Tím se model výrazně zjednoduší. Pro nalezení významných period použijeme tzv. *periodogram* a následně *Fisherův test periodicity*, které jsou popsány níže.

3.2.1 Periodogram

Periodogram $I(\omega)$ časové řady y_1, \dots, y_n se definuje jako funkce proměnné ω ve tvaru

$$I(\omega_j) = \frac{1}{2} (\hat{a}_j^2 + \hat{b}_j^2), \quad (3.5)$$

kde $j = 1, \dots, H$; $\omega_j = \frac{2\pi j}{n}$ jsou úhlové frekvence a \hat{a}_j, \hat{b}_j jsou odhady (3.3) a (3.4). Vysoká hodnota funkce v určité frekvenci ω_j značí přítomnost periody určité délky [2]. Jelikož periodogram často zobrazuje periody, které nejsou až tak významné, nebo někdy také periody, které neexistují, zhodnocení na základě periodogramu je pouze subjektivní, proto je možné provést tzv. *Fisherův test významnosti period*.

3.2.2 Test R. A. Fishera

Test R. A. Fishera neboli Fisherův test významnosti period se používá pro identifikaci významných period, které poté můžeme zahrnout do modelu (3.1). Jedná se o klasické testování hypotéz. Nulová hypotéza bude ve tvaru $H_0: y_t = \varepsilon_t$, tzn. předpokládá se, že časová řada neobsahuje žádnou periodicitu a je rovna přímo náhodné složce. Tato hypotéza je testována oproti alternativě, že existuje alespoň jedna perioda, tj. $H_A: \exists j: a_j \neq 0 \vee b_j \neq 0$. Testová statistika vychází z hodnot periodogramu $I(\omega_j)$ dané řady y_t , kde $t = 1, \dots, n$. Jak jsem uvedla v kapitole 3.1., vysoká hodnota periodogramu ukazuje na velkou významnost periody, proto si hodnoty $I(\omega_j)$ nejprve vypočítáme dosazením do (3.5) a poté seřadíme od největší po nejmenší, tedy

$$I_{(1)} \geq I_{(2)} \geq \dots \geq I_{(H)}, \quad (3.6)$$

kde indexy $1, 2, \dots, H$ v závorkách neoznačují periody, ale značí pořadí. Vezmeme největší hodnotu $I_{(1)}$ a znormujeme

$$Y_1 = \frac{I_{(1)}}{\sum_{k=1}^H I_{(k)}}.$$

Tuto hodnotu Y_1 , která odpovídá určité periodě j , porovnáme s kritickou hodnotou Fisherova testu na hladině významnosti alfa $g_F^H(\alpha)$, kde $H = \left\lceil \frac{n}{2} \right\rceil$. Hranaté závorky značí funkci „celá část“, v tomto případě dolní celá část, nebo také $H = \frac{n}{2}$ pro n sudé, $H = \frac{(n-1)}{2}$ pro n liché. Hladinu významnosti α si libovolně volíme, nejčastěji 0,01 nebo 0,05. Nulovou hypotézu zamítáme na hladině α , pokud platí nerovnost

$$Y_1 > g_F^H(\alpha).$$

Zamítáme-li hypotézu, znamená to, že je daná perioda významná, a následně testujeme další velké hodnoty periodogramu. Otestujeme tedy druhou největší hodnotu $I_{(2)}$, přičemž neuvažujeme již otestovanou periodu. Testová statistika je nyní ve tvaru

$$Y_2 = \frac{I_{(2)}}{\sum_{k=2}^H I_{(k)}}.$$

Opět porovnáme s kritickou hodnotou $g_F^{H-1}(\alpha)$ a zjistíme, zda se jedná o významnou periodu či nikoliv. Cílem je najít takovou periodu, pro kterou bude platit

$$Y_k < g_F^{H-k-1}(\alpha).$$

V takovém případě hypotézu nelze zamítnout, což znamená, že daná perioda již není významná a žádná další v pořadí dle (3.6) taktéž není významná.

Výsledkem Fisherova testu významnosti period jsou získané významné periody, které zahrneme do modelu (3.1) a tím ho zjednodušíme. Zredukování počtu period nebude mít negativní dopad na kvalitu daného modelu, takže veličiny popisující vhodnost modelu (viz kapitola 6) se nijak výrazně nezhorší.

4 Modelování sezónnosti

Při modelování sezónní složky je nutné zohlednit, zda má časová řada aditivní (1.1) nebo multiplikativní (1.2) charakter. Časová řada vykazuje multiplikativní sezónní složku (někdy se zde mluví také o proporcionální sezónnosti), je-li amplituda sezónního pohybu úměrná průměrné úrovni řady (vyjádřené pomocí trendu). V případě, že amplituda sezónního pohybu nezávisí na průměrné úrovni řady, uvažujeme sezónní složku jako aditivní. V praxi se předpokládá, že před samotnou analýzou časové řady jsme již rozhodnutí, který z těchto dvou typů dekompozice použijeme. [1]

Doposud jsme používali indexování $t = 1, 2, \dots, n$, kde n je počet pozorování (délka časové řady), v případě modelování sezónní složky je nutné indexy přechíslovat. Zavedeme index $i = 1, \dots, m$, který bude pro jednotlivé roky, index $j = 1, \dots, r$ bude značit sezóny. Navíc předpokládáme, že platí rovnost $n = mr$.

Obecně model zahrnující trendovou, sezónní a náhodnou složku bude vypadat následovně

$$y_{ij} = T_{ij} + S_{ij} + \varepsilon_{ij}. \quad (4.1)$$

4.1 Model konstantní sezónnosti

Tento model použijeme v případě, kdy máme k dispozici časovou řadu s aditivní sezónní složkou, tzn. není závislá na průměrné úrovni řady. Dalším předpokladem, jak už z názvu metody vyplývá, je předpoklad konstantní sezónnosti. Průběh sezónnosti bude v každém roce stejný, j -tá sezónní složka tedy nezáleží na roku $i = 1, \dots, m$, což můžeme zapsat jako $S_{ij} = S_j = b_j$, kde b_j bude vystupovat jako parametr, a předpokládáme, že $\sum_{j=1}^r S_j = \sum_{j=1}^r b_j = 0$.

Budeme uvažovat časovou řadu, která je dána lineárním trendem (2.4). Do modelu (4.1) pak dosadíme za T_{ij} lineární trend, a to s centrovaným časem pro jednotlivé roky. Rovnice trendu pak bude ve tvaru

$$T_{ij} = a_0 + a_1(i - \bar{i}) + \varepsilon_{ij}, \quad (4.2)$$

kde $i = 1, \dots, m, j = 1, \dots, r, a_0$ a a_1 jsou parametry lineárního trendu, ε_{ij} je náhodná složka, $\bar{i} = \frac{m+1}{2}$.

Dosadíme-li trend (4.2) do modelu (4.1), získáme rovnici pro *model konstantní sezónnosti s lineárním trendem*

$$y_{ij} = a_0 + a_1(i - \bar{i}) + b_j + \varepsilon_{ij}, \quad (4.3)$$

kde a_0, a_1 a b_j jsou parametry, které chceme odhadnout. Pro odhady můžeme využít *metodu nejmenších čtverců*. Provedeme minimalizaci

$$\min_{a_0, a_1, b_j} \sum_{i=1}^m \sum_{j=1}^r (y_{ij} - a_0 - a_1(i - \bar{i}) - b_j)^2. \quad (4.4)$$

Derivujeme podle všech parametrů a derivace položíme rovny nule, abychom získali *soustavu normálních rovnic* (viz kapitola 2.1). Vyřešením těchto rovnic získáme hledané odhady parametrů ve tvaru

$$\hat{a}_0 = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^r y_{ij} = \bar{y}, \quad (4.5)$$

$$\hat{b}_j = \frac{1}{m} \sum_{i=1}^m y_{ij} - \hat{a}_0 = \hat{y}_{.j} - \bar{y}, \quad j = 1, \dots, r, \quad (4.6)$$

$$\hat{a}_1 = \frac{\sum_{i=1}^m i(\hat{y}_{i.} - \bar{y})}{\frac{m}{12}(m^2 - 1)}, \quad (4.7)$$

kde $\hat{y}_{.j} = \frac{1}{m} \sum_{i=1}^m y_{ij}$ čili jedná se o průměr j -té sezóny, $\hat{y}_{i.} = \frac{1}{r} \sum_{j=1}^r y_{ij}$ udává průměr v i -tém roce. Výsledný model získáme dosazením výše uvedených odhadů do modelu (4.3).

4.2 Model proporcionální sezónnosti

Model proporcionální sezónnosti se používá pro časové řady multiplikativního charakteru. Jak jsem již zmínila v úvodu této kapitoly, jedná se o časovou řadu, která se vyznačuje sezónní složkou úměrnou trendu v dané řadě. To znamená že $S_{ij} = c_j T_{ij}$, kde c_j je tzv. sezónní parametr. Po dosazení do (4.1) dostaneme model proporcionální sezónnosti ve tvaru

$$y_{ij} = T_{ij} + c_j T_{ij} + \varepsilon_{ij} = (1 + c_j) T_{ij} + \varepsilon_{ij}, \quad (4.8)$$

kde opět jsou roky $i = 1, \dots, m$ a sezóny $j = 1, \dots, r$. Hodnoty $(1 + c_j)$ jsou tzv. *sezónní indexy*. Pokud je $c_j > 0$, dochází k růstu oproti trendu, naopak při $c_j < 0$ se jedná o pokles.

Do modelu (4.8) je potřeba dosadit odhady parametru $(\widehat{1 + c_j})$. Pro výpočet těchto odhadů je potřeba určit odhad trendu \hat{T}_{ij} , kterým se daná časová řada vyznačuje (konstantní, lineární, kvadratický atd.). Odhad trendu je již popsán v kapitole 2.1. Opět *metodou nejmenších čtverců* získáme odhad $(\widehat{1 + c_j})$. Provedeme minimalizaci

$$\min_{(1+c_j)} \sum_{i=1}^m \sum_{j=1}^r (y_{ij} - (1 + c_j)T_{ij})^2.$$

Derivaci položíme rovnu nule a získáme *normální rovnice*

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

Odtud

$$(\widehat{1 + c_j}) = \frac{\sum_{i=1}^m y_{ij}\hat{T}_{ij}}{\sum_{i=1}^m \hat{T}_{ij}^2}. \quad (4.9)$$

Protože pro sezónní indexy platí předpoklad $\sum_{j=1}^r (1 + c_j) = r$, většinou je v praktických příkladech nutné provést standardizaci sezónních indexů na normovaný tvar

$$(\widehat{1 + c_j})^{(N)} = \frac{r(\widehat{1 + c_j})}{\sum_{k=1}^r (\widehat{1 + c_k})}, \quad (4.10)$$

kde r je počet sezón. Nyní už zbývá jen dosadit normované sezónní indexy do modelu (4.8) a získáme model proporcionální sezónnosti s lineárním trendem.

V kapitole 4 jsem čerpala především z literatury [1], [2], [4].

5 Míry vhodnosti modelu

Při sestavování jednotlivých modelů nás může zajímat, který z modelů popisuje konkrétní časovou řadu nejlépe a nejpřesněji. Proto si zavedeme některé míry, které se používají pro ocenění kvality zkonstruovaných modelů a také předpovědí. Tyto míry nám dají informaci o vhodnosti sestaveného modelu. Následně máme možnost jednotlivé modely porovnávat a rozhodnout se, který je pro nás lepší. Vzorce v této kapitole jsou převzaty z [1] a [4].

Základním ukazatelem je tzv. *reziduální součet čtverců*, můžeme jej označit *RSČ*, v některých literaturách se značí SSE (z anglického Sum of Squared Errors) a je dán vzorcem

$$RSČ = \sum_{t=1}^n \hat{\varepsilon}_t^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (5.1)$$

kde y_t jsou původní data, \hat{y}_t představují data odhadnutá modelem. Jak je možné vidět ze vzorce i z názvu, jedná se o sumu druhých mocnin reziduí $\hat{\varepsilon}_t$. Jelikož rezidua chci mít vždy co nejmenší, plyne odtud, že čím bude hodnota *RSČ* nižší, tím lépe model pasuje na data.

Dalším z hojně používaných ukazatelů je tzv. *index determinace*, značíme R^2 . Může nabývat hodnot z intervalu $\langle 0,1 \rangle$, přičemž čím blíže je hodnotě 1, tím model lépe vystihuje danou časovou řadu a je lepší. Naopak když se index determinace blíží 0, není model vhodný. Vypočítáme pomocí vzorce

$$R^2 = 1 - \frac{RSČ}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad (5.2)$$

kde $\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$, za *RSČ* dosadíme (5.1). Výraz ve jmenovateli vyjadřuje celkovou variabilitu v datech.

Ani jeden z ukazatelů nezohledňuje počet parametrů v modelu, proto se často pro hodnocení kvality užívá *modifikovaného indexu determinace*, značíme R_a^2 (*a* z anglického adjusted), který je dán vztahem

$$R_a^2 = R^2 - \frac{(1-R^2)(p-1)}{n-p}, \quad (5.3)$$

kde n značí počet pozorování, p je počet parametrů v modelu, R^2 je index determinace (5.2).

Dále lze určit tzv. *střední čtvercovou chybu*, značíme *MSE* (Mean Squared Error), jejíž hodnotu chceme minimální, ve tvaru

$$MSE = \sqrt{\sum_{t=1}^n \frac{\hat{\varepsilon}_t^2}{n}} = \sqrt{\sum_{t=1}^n \frac{(y_t - \hat{y}_t)^2}{n}}. \quad (5.4)$$

Existují některé další ukazatele popisující kvalitu modelu, avšak tyto základní ukazatele jsou v tuto chvíli dostačující, jde mi především o představení těch, které využiji v praktické části této práce. Další ukazatele podobného charakteru se týkají předpovědí, proto je uvedu na konci následující kapitoly 6.

6 Předpovědi v ČR

A na konec teoretické části se dostávám k jednomu z nejdůležitějších témat. Předpovědi (jinak také *predikce*, *prognózy* či *extrapolace*) jsou jedním z hlavních důvodů pro analýzu časových řad. Dle [2] extrapolacemi nazýváme kvantitativní odhady budoucích hodnot časové řady, které vznikají prodloužením z minulosti a přítomnosti do budoucnosti s horizontem h za předpokladu, že se tento vývoj nezmění.

Najdeme-li vhodný model pasující na zadaná data, není již složité zkonstruovat předpovědi. Model pak využijeme jako tzv. prognostický model pro výpočet předpovědí. Při konstrukci predikcí předpokládáme, že se charakter časové řady v čase nemění, což nám umožňuje určit spolehlivý statistický odhad. Existují dva druhy předpovědí, a to *bodové* a *intervalové*.

Bodová předpověď se určuje v čase n (přítomnost) o h období dopředu. Písmeno h je horizont předpovědi. Předpověď budoucí neznámé hodnoty v čase $T = n + h$ označujeme \hat{y}_{n+h}^P neboli \hat{y}_T^P . Index P označuje předpověď.

Jak jsem již zmínila, samotná konstrukce předpovědi vychází z konkrétního modelu. Pro analytické trendové funkce bude bodová předpověď pro období $n + h$ ve tvaru

$$\hat{y}_{n+h}^P = \hat{T}_{n+h}. \quad (6.1)$$

Pokud budu vycházet například z modelu skrytých period ve tvaru (3.1), bude bodová předpověď vypadat následovně

$$\hat{y}_{n+h}^P = \hat{y}_T^P = \hat{a}_0 + \sum_{j=1}^m \hat{a}_j \sin(\omega_j T) + \sum_{j=1}^m \hat{b}_j \cos(\omega_j T), \quad (6.2)$$

kde \hat{a}_0 , \hat{a}_j a \hat{b}_j jsou odhady parametrů a_0 , a_j a b_j vypočtené dle (3.2), (3.3), (3.4), $j = 1, \dots, m$ jsou pouze významné periody, které jsme odhalili pomocí Fisherova testu (viz kapitola 3.2.2).

Další bodová předpověď, kterou jsem použila v praktické části, je pro model konstantní sezónnosti ve tvaru (4.3), kde počítáme předpověď na j -tou sezónu o h roků vpřed

$$\hat{y}_{m+h,j}^P = \hat{a}_0 + \hat{a}_1 \left(m + h - \frac{m+1}{2} \right) + \hat{b}_j, \quad (6.3)$$

kde m je počet roků, h horizont předpovědi, j značí sezónu, \hat{a}_0 , \hat{a}_1 , \hat{b}_j jsou odhady parametrů vypočítané dle (4.5), (4.6), (4.7).

Stejně tak mohu vytvářet predikce pro model proporcionální sezónnosti (4.8) a vypočítat budoucí hodnoty pro j -tou sezónu o h roků dopředu. Předpovědní model bude ve tvaru

$$\hat{y}_{m+h,j}^P = \hat{T}_{m+h,j}(\widehat{1 + c_j}). \quad (6.4)$$

Jak si lze všimnout, pro konstrukci předpovědi v modelu proporcionální sezónnosti je pouze nutné sestavit předpověď pro trendovou funkci (jako v (6.1)).

Je nutné se ještě zmínit o chybách v předpovědích. Chyba předpovědi \hat{y}_T^P skutečné hodnoty y_T je definována jako

$$\Delta_{n+h} = \Delta_T = y_T - \hat{y}_T^P. \quad (6.5)$$

Její hodnotu je samozřejmě možné určit až v budoucnosti v čase T , kdy poznáme skutečnou hodnotu y_T , kterou jsme přirozeně při konstrukci předpovědi neznali. Často se však konstruují předpovědi známých hodnot časové řady, které se pak mohou porovnat a následně lze posoudit kvalitu předpovědního modelu. Zdrojem chyb je především reziduální složka, jelikož je náhodného charakteru a představuje nepředvídatelné fluktuace v datech. Velikost chyby je však ovlivněna také tím, jak jsme zvládli předpovědět ostatní systematické složky řady. Velké chyby v předpovědi mohou znamenat značný podíl reziduální složky v řadě, nebo nevhodnost použité předpovědní techniky. [1]

Pro ocenění kvality zkonstruovaných předpovědí se používají následující míry. *Průměrnou čtvercovou chybu předpovědi MSEP* (z anglického Mean Squared Error of Prediction) určíme jako

$$MSEP = \sqrt{\frac{1}{m} \sum_{h=1}^m \Delta_{n+h}^2}. \quad (6.6)$$

S použitím vzorce (6.5) vypočítáme *relativní chybu* jako

$$\delta_{n+h} = \delta_T = \frac{\Delta_{n+h}}{y_{n+h}}, \quad (6.7)$$

y_{n+h} je skutečná hodnota časové řady v čase $n + h$.

Druhým typem předpovědí jsou *intervalové*. Jelikož v praktické části jsem s intervalovými předpověďmi nepracovala, nebudu se zabývat jednotlivými vzorci dopodrobna. Předpovědní interval je úplnou analogií intervalu spolehlivosti z matematické statistiky. $(1 - \alpha)100\%$ – ní interval udává dolní a horní mez, mezi nimiž bude ležet

příslušná skutečná hodnota y_{n+h} sledované časové řady s pravděpodobností $(1 - \alpha)$. Většinou se jedná o interval symetrický okolo bodové předpovědi \hat{y}_{n+h}^P , tj. dle [2]

$$\hat{y}_{n+h}^P \pm t_{1-\frac{\alpha}{2}}(n-p-1)\hat{\sigma}, \quad (6.8)$$

kde $t_{1-\frac{\alpha}{2}}(n-p-1)$ je $(1 - \frac{\alpha}{2})$ 100%-ní kvantil Studentova rozdělení s $n - (p + 1)$ stupni volnosti, kde $p + 1$ je počet odhadnutých parametrů, $\hat{\sigma}$ je směrodatná chyba předpovědi v horizontu h (směrodatná odchylka reziduí).

PRAKTICKÁ ČÁST

V této části budu aplikovat výše uvedené metody a nástroje na konkrétní data. U jednotlivých časových řad budu zjišťovat, jakým trendem se daná řada vyznačuje, jak to vypadá s periodicitou, tj. pomocí nástrojů spektrální analýzy mohu určovat, které periody jsou významné, případně zda časová řada obsahuje sezónnost atd. Na základě analýzy budu moci určit, který model je pro danou časovou řadu nejvhodnější. Najdu-li vhodný model, mohu konstruovat předpovědi budoucích hodnot časové řady a určit kvalitu těchto předpovědí.

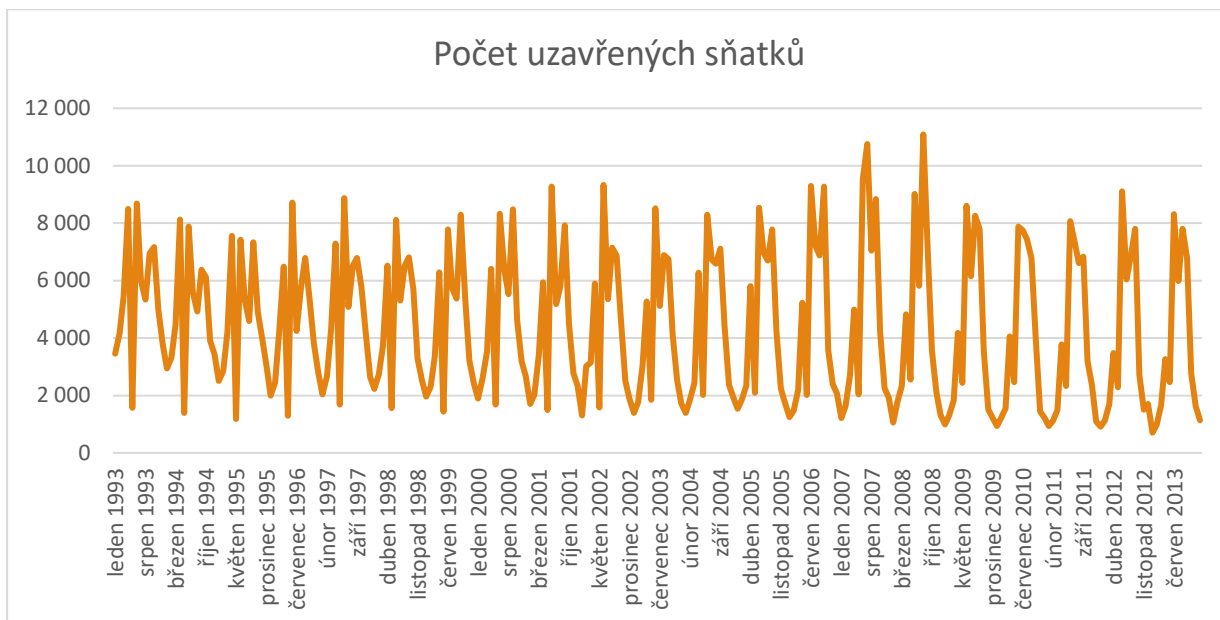
7 Sňatkovost

Pro praktickou část mé práce jsem potřebovala najít časovou řadu, která obsahuje periodickou složku. Z internetových stránek Českého statistického úřadu [6] jsem si vybrala data demografického typu, a to konkrétně měsíční časovou řadu, která zobrazuje sňatkovost v České republice od ledna roku 1993 do června roku 2016. V řadě jsou tedy uvedeny počty uzavřených sňatků na území ČR za jednotlivé měsíce.

Délku časové řady jsem si úmyslně zkrátila, abych na závěr mohla porovnat předpovědi se skutečnými hodnotami a zjistila tak, jak jsou predikce kvalitní. Řadu jsem zkrátila o 30 pozorování a ponechala jsem si pro analýzu data od ledna 1993 do prosince 2013, tzn. rozmezí 21 roků, délka zkrácené časové řady je $n = 252$. Všechna pozorování jsou uvedena v příloze č. 1. Data nejprve pro představu vykreslím do grafu č. 1.

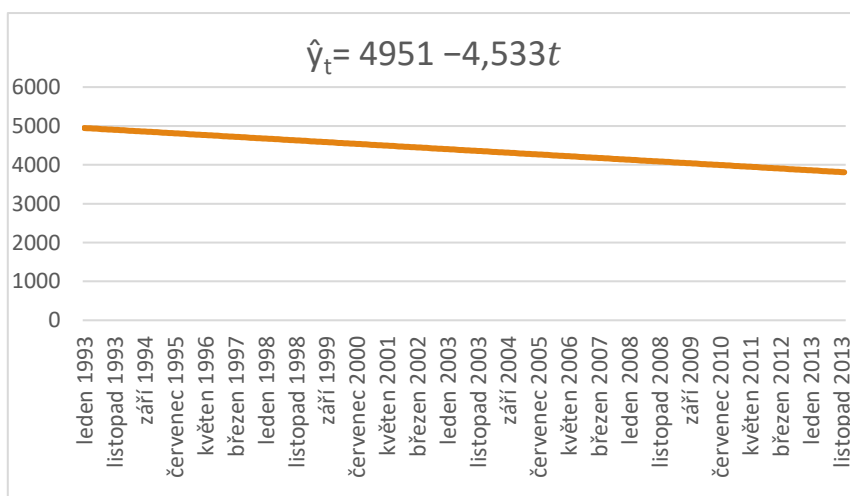
Na první pohled je zřejmé, že data obsahují periodickou složku. Když se blíže podíváme, zjistíme, že počet sňatků se v nejslabších měsících pohybuje kolem jednoho až dvou tisíc, naopak v nejsilnějších měsících se dostáváme až k deseti tisícům uzavřených sňatků. Z dat je možné pozorovat, že silný nárůst začíná vždy již v dubnu, poté každoročně výrazně klesne v květnu a následně opět rapidně roste a tyto vysoké počty se drží až do října. Poté zájem o uzavírání sňatků opět klesá. Jako překvapivý se může jevit každoroční pokles v květnu. Květen, měsíc lásky, zdálo by se, že je jako stvořený pro svatby. Avšak Češi jsou stále v zajetí pověry „Svatba v máji, do roka máry“. Pověra je založena na historické skutečnosti, která dnes samozřejmě neplatí, avšak jak statistiky ukazují, Češi jsou zřejmě pověřiví. Více se můžete dočíst na webu [7]. Nejoblíbenější jsou tedy teplé letní měsíce,

méně oblíbené jsou pak zimní měsíce, kdy počet klesl mnohokrát i pod hranici tisíce uzavřených sňatků.



Graf č. 1 - Vývoj dat od ledna 1993 do prosince 2013

Z grafu to možná není na první pohled zřejmé, ale dlouhodobě časová řada mírně klesá. Je dána lineárním trendem ve tvaru (2.4). Odhady parametrů a_0 a a_1 jsem vypočítala dosazením do vzorců (2.5) a (2.6) a získala jsem hodnoty $\hat{a}_0 = 4951$, $\hat{a}_1 = -4,533$. Dosazením odhadů získáme rovnici lineárního trendu ve tvaru $\hat{y}_t = 4951 - 4,533t$; $t = 1, \dots, 252$. Vyrovnané hodnoty $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{252}$ získáme dosazením proměnné t do rovnice. V grafu č. 2 je zobrazena lineární trendová funkce pro danou časovou řadu.



Graf č. 2 - Lineární trend pro sňatkovost

7.1 Model skrytých period (MSP) a předpovědi

Jako další na data použijeme model skrytých period (3.2), ve kterém budeme uvažovat konstantní trend. Nejprve si vypočteme úhlové frekvence $\omega_j = \frac{2\pi j}{n}$, pro každé $j = 1, \dots, H$, kde $H = \frac{n}{2} = \frac{252}{2} = 126$. Získáme tedy hodnoty $\omega_1, \omega_2, \dots, \omega_{126}$, které jsou uvedeny v příloze č. 2. Nyní už je možné vypočítat odhady parametrů a_0, a_j, b_j , které jsou potřebné pro sestavení modelu skrytých period. Výpočet jsem provedla s použitím vzorců (3.2), (3.3) a (3.4). Dosazením do vzorce získáme hodnotu

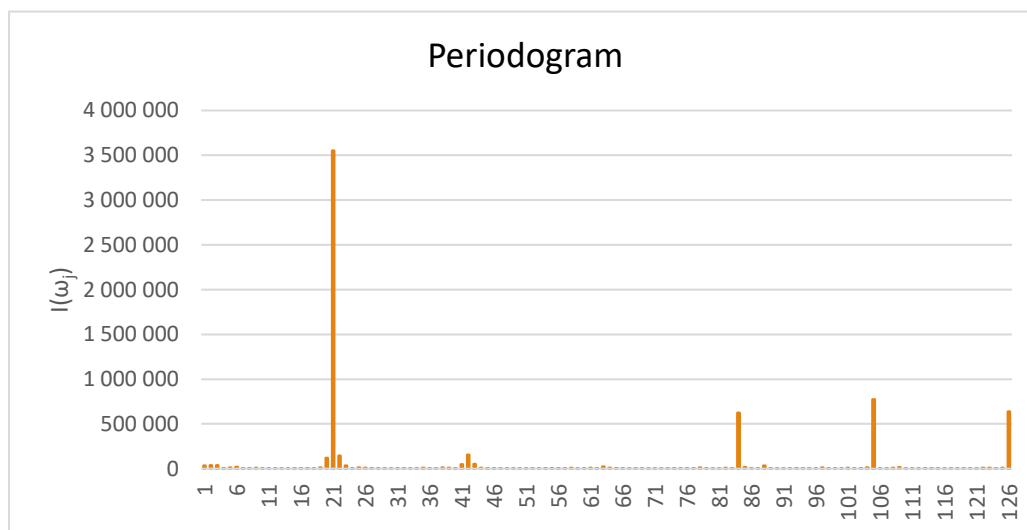
$$\hat{a}_0 = \frac{1}{252} \sum_{t=1}^n y_t = 4378.$$

Pro další výpočty je nutné nejprve určit součiny $\omega_1 t, \omega_2 t, \dots, \omega_{126} t$ pro každé $t = 1, 2, \dots, 252$. Následně už můžeme určit

$$\hat{a}_j = \frac{2}{252} \sum_{t=1}^{252} y_t \sin(\omega_j t) \text{ a } \hat{b}_j = \frac{2}{252} \sum_{t=1}^{252} y_t \cos(\omega_j t).$$

Po dosazení získám odhady $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_{126}, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_{126}$, jejichž hodnoty jsou uvedeny v příloze č. 2.

Nyní je třeba zjistit, které z period jsou významné. Není nutné zahrnovat do modelu všech 126 period. Obvykle se postupuje tak, že použijeme periodogram (viz kapitola 3.2.1), který slouží k identifikaci významných period. Vypočítáme dle vzorce (3.5) hodnoty periodogramu $I(\omega_j)$ pro každé $j = 1, \dots, 126$. Tyto hodnoty jsou uvedeny taktéž v příloze č. 2. Hodnoty periodogramu jsem zakreslila do grafu č. 3.



Graf č. 3 - Periodogram pro sňatkovost

Z grafu lze vidět, které periody jsou významné. Avšak vizuální posouzení je subjektivní, proto využijeme ještě Fisherův test významnosti period (viz kapitola 3.2.2). Testujeme nulovou hypotézu, že časová řada neobsahuje periodicitu, tzn. $H_0: y_t = \varepsilon_t$, oproti opačné alternativně. Testovat budeme na hladině $\alpha = 0,01$.

Nejprve seřadíme vypočítané hodnoty $I(\omega_j)$ podle velikosti. Největší hodnotu označíme $I_{(1)}$, další $I_{(2)}$ a tak dále, až nejmenší $I_{(126)}$. Nyní vezmeme největší hodnotu $I_{(1)}$, což je v našem případě hodnota 3 550 122, která odpovídá $j = 21$ (tzn. roční perioda) a znormujeme dle vzorce (3.7). Získáme hodnotu testové statistiky $Y_1 = 0,52693112$. Tu nyní porovnáme s kritickou hodnotou Fisherova testu, některé tyto hodnoty jsou uvedeny v následující tabulce 1.

Tabulka 1 – Kritické hodnoty Fisherova testu

H	$g_F^H(0,01)$
119	0,0766
120	0,0759
121	0,0754
122	0,0750
123	0,0745
124	0,0740
125	0,0735
126	0,0731

Odtud vidíme, že $Y_1 > g_F^{126}(0,01)$, což znamená, že hypotézu nelze zamítnout a perioda je významná. Proto otestujeme další v pořadí tak, že dosud největší hodnotu vynecháme a o tuto hodnotu snížíme celkový součet (jmenovatel) a navíc hodnotu H nahradíme hodnotou $H - 1$. Pokračujeme dále analogicky, dokud nám nevyjde, že kritická hodnota Fisherova testu je větší než hodnota testové statistiky. V takovém případě hypotézu zamítáme. Hodnoty Fisherova testu jsou uvedeny v následující tabulce 2.

Tabulka 2 - Fisherův test pro sňatkovost

		j		Hodnoty testové statistiky			$g_F^H(0,01)$	
$I_{(1)}$	3 550 122	21	=»	$Y_{(1)}$	0,527	>	0,073	- zamítám H_0
$I_{(2)}$	772 905	105	=»	$Y_{(2)}$	0,243	>	0,074	- zamítám H_0
$I_{(3)}$	636 917	126	=»	$Y_{(3)}$	0,264	>	0,074	- zamítám H_0
$I_{(4)}$	622 222	84	=»	$Y_{(4)}$	0,350	>	0,074	- zamítám H_0
$I_{(5)}$	153 114	42	=»	$Y_{(5)}$	0,133	>	0,075	- zamítám H_0
$I_{(6)}$	142 915	22	=»	$Y_{(6)}$	0,143	>	0,075	- zamítám H_0
$I_{(7)}$	118 391	20	=»	$Y_{(7)}$	0,138	>	0,076	- zamítám H_0
$I_{(8)}$	52 180	43	=»	$Y_{(8)}$	0,070	<	0,077	- H_0 nelze zamítnout

Dle Fisherova testu existuje v modelu sedm významných period. Délku periody určíme vždy jako n/j , tedy nejvýznamnější perioda je roční ($252/21 = 12$ měsíců). Dále zde máme významnou dvouměsíční periodu ($252/126 = 2$ měsíce) a o kousek delší ($252/105 = 2,4$ měsíců), významná je také čtvrtletní perioda ($252/84 = 3$ měsíce) a půlroční ($252/42 = 6$ měsíců). Jako významné se ukázaly i o kousek kratší než roční perioda ($252/22 = 11,45$ měsíců) a o kousek delší ($252/20 = 12,6$ měsíců).

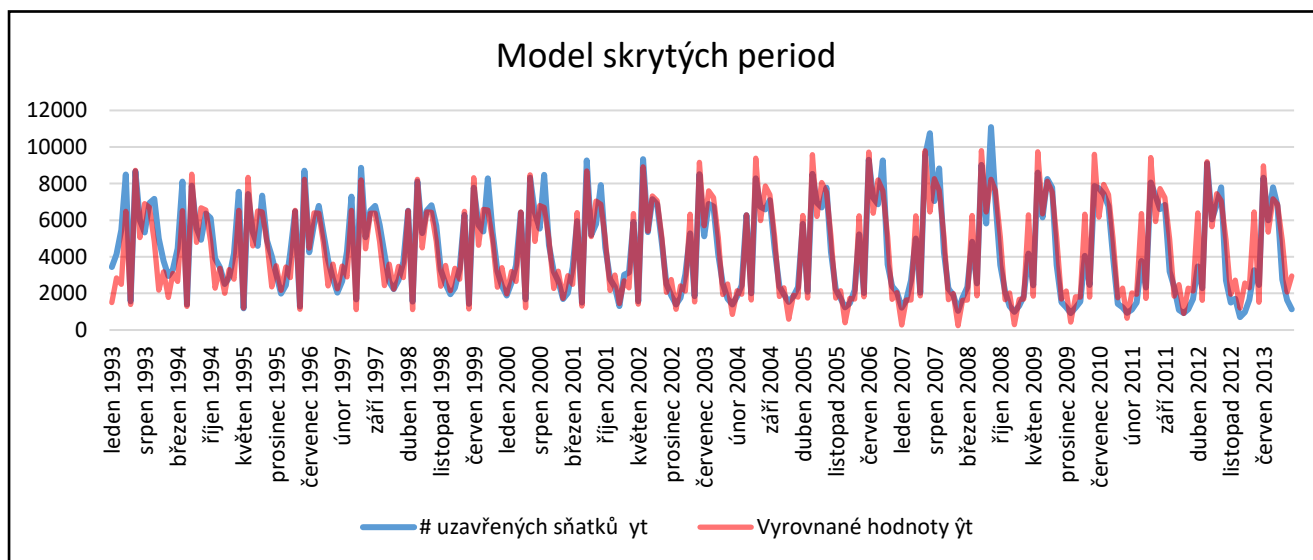
Do modelu skrytých period ve tvaru (3.2) tedy zahrneme pouze těchto sedm indexů: $j = 21, 105, 126, 84, 42, 22$ a 20 . Po dosazení získáme vyrovnané hodnoty \hat{y}_t , které jsou uvedeny v příloze č. 3. Skutečné i vyrovnané hodnoty jsem vykreslila do grafu č. 4.

Jak je vidět z grafu, model velmi hezky pasuje na danou časovou řadu sňatků. Pro určení kvality modelu jsem využila index determinace dle vzorce (5.2) a modifikovaný index determinace dle vzorce (5.3)

$$R^2 = 1 - \frac{266\,925\,354}{6390650211} = 0,958,$$

$$R_a^2 = 0,958 - \frac{(1 - 0,958)(15 - 1)}{252 - 15} = 0,956.$$

Obě hodnoty jsou z intervalu $\langle 0,1 \rangle$ a navíc jsou velmi blízké 1, což nám říká, že je daný model kvalitní.

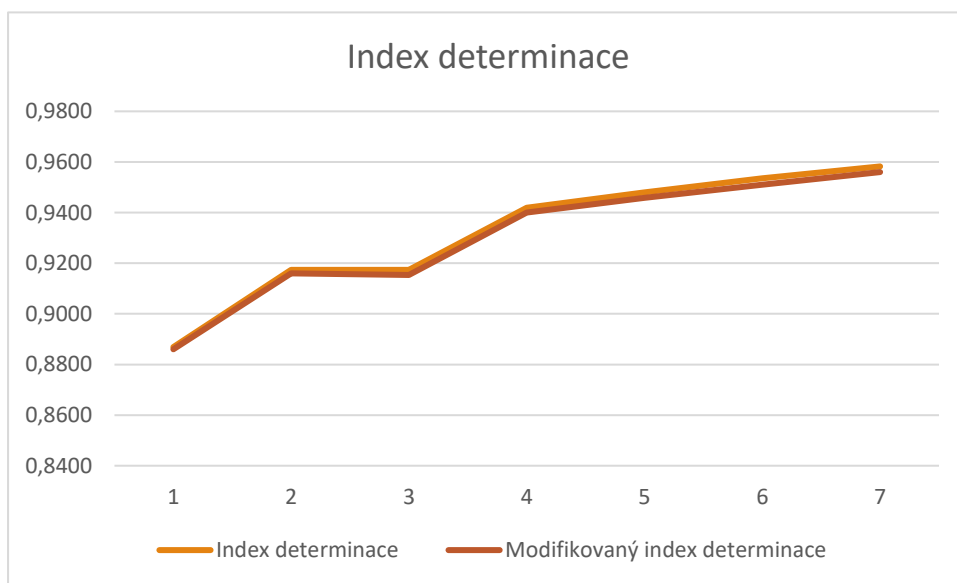


Graf č. 4 - Vyrovnané hodnoty – MSP pro sňatkovost

Dále jsem se rozhodla prozkoumat, jaký vliv má počet parametrů v modelu na kvalitu modelu (danou indexem determinace R^2). Postupně jsem odebírala z modelu periody j (od těch nejméně významných, tzn. jako první jsem odebrala $j = 20$, poté $j = 22, 42$ atd.) a po každém odebrání jsem vypočítala znovu normální a modifikovaný index determinace. Jak se mění R^2 a R_a^2 v závislosti na počtu parametrů je možné vidět v následující tabulce 3 a grafu č. 5.

Tabulka 3 – Index determinace a počet period

# period	R^2	R_a^2
1	0,8869	0,8860
2	0,9174	0,9160
3	0,9174	0,9153
4	0,9419	0,9400
5	0,9479	0,9458
6	0,9536	0,9510
7	0,9582	0,9560



Graf č. 5 – Index determinace v závislosti na počtu period v modelu skrytých period

Jak je vidět i z grafu, při zahrnutí více než čtyř period se index determinace nijak výrazně nezlepšuje. Proto by bylo možné pro zjednodušení uvažovat pouze čtyři nejvýznamnější periody místo původních sedmi. Já osobně jsem použila při této analýze všech sedm period, abych při konstrukci předpovědí získala co nejpřesnější hodnoty.

Předpovědi MSP

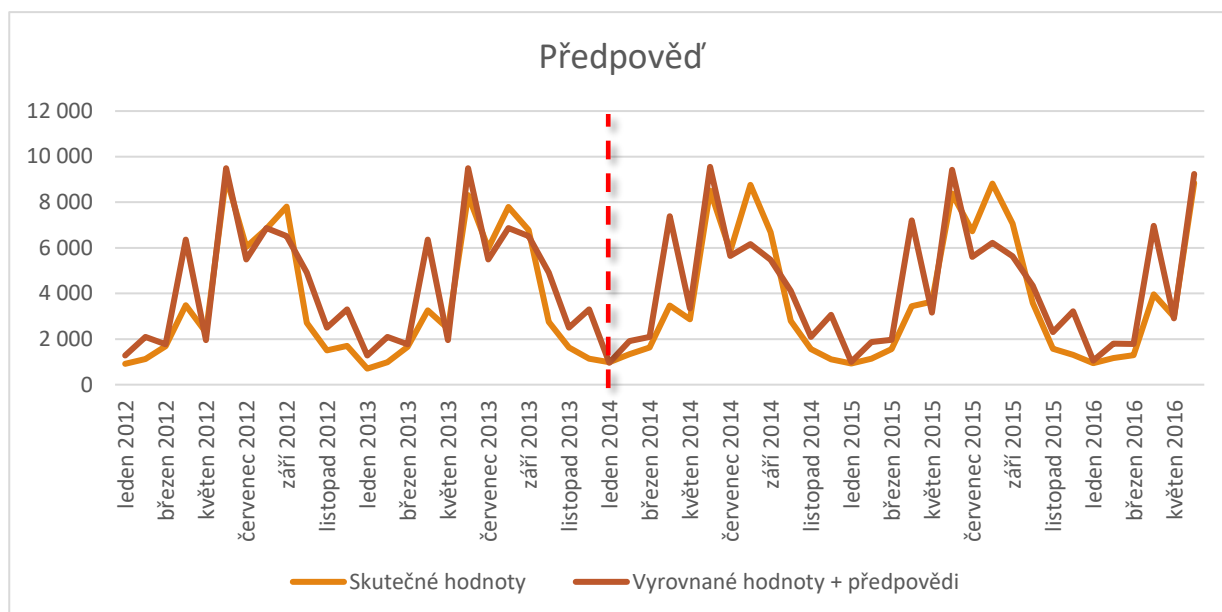
Jak jsem již zmínila v úvodu praktické části, časovou řadu jsem si úmyslně zkrátila a ponechala si 30 hodnot pro porovnání s vytvořenými predikcemi. Ve skutečnosti tedy mám k dispozici dalšího dva a půl roku (tj. do června 2016). Pro toto období nyní zkonstruuji bodové předpovědi \hat{y}_{n+h}^P dle vzorce (6.2), $h = 1, \dots, 30$; tzn. vypočítám hodnoty $\hat{y}_{253}^P, \hat{y}_{254}^P, \dots, \hat{y}_{282}^P$.

V tabulce 4 jsou uvedeny tyto předpovědi vedle skutečných hodnot $y_{253}, y_{254}, \dots, y_{282}$.

Na závěr můžeme určit chyby předpovědi dle (6.5), relativní chyby dle (6.7), jež jsou také v tabulce 4. Průměrná čtvercová chyba předpovědi s použitím vzorce (6.6) je $MSEP = \sqrt{2668378} = 1634$. Pro lepší představu jsem předpovědi zobrazila v grafu č. 6 (za červenou čárkovanou čarou).

Tabulka 4 - Bodové předpovědi MSP

období	h	$n + h$	y_{n+h}	\hat{y}_{n+h}^P	Δ_{n+h}	$\delta_{n+h}(\%)$
leden 2014	1	253	981	961	20	2
únor 2014	2	254	1 333	1915	-582	-44
březen 2014	3	255	1 628	2096	-468	-29
duben 2014	4	256	3 477	7385	-3908	-112
květen 2014	5	257	2 872	3338	-466	-16
červen 2014	6	258	8 504	9553	-1049	-12
červenec 2014	7	259	5 861	5652	209	4
srpen 2014	8	260	8 766	6162	2604	30
září 2014	9	261	6 686	5479	1207	18
říjen 2014	10	262	2 794	4121	-1327	-47
listopad 2014	11	263	1 565	2092	-527	-34
prosinec 2014	12	264	1 108	3064	-1956	-177
leden 2015	13	265	932	995	-63	-7
únor 2015	14	266	1 145	1865	-720	-63
březen 2015	15	267	1 565	1967	-402	-26
duben 2015	16	268	3 444	7208	-3764	-109
květen 2015	17	269	3 644	3159	485	13
červen 2015	18	270	8 379	9421	-1042	-12
červenec 2015	19	271	6 731	5608	1123	17
srpen 2015	20	272	8 814	6223	2591	29
září 2015	21	273	7 069	5636	1433	20
říjen 2015	22	274	3 583	4335	-752	-21
listopad 2015	23	275	1 572	2307	-735	-47
prosinec 2015	24	276	1 313	3221	-1908	-145
leden 2016	25	277	937	1049	-112	-12
únor 2016	26	278	1 165	1794	-629	-54
březen 2016	27	279	1 296	1787	-491	-38
duben 2016	28	280	3 956	6962	-3006	-76
květen 2016	29	281	2 974	2913	61	2
červen 2016	30	282	8 849	9241	-392	-4



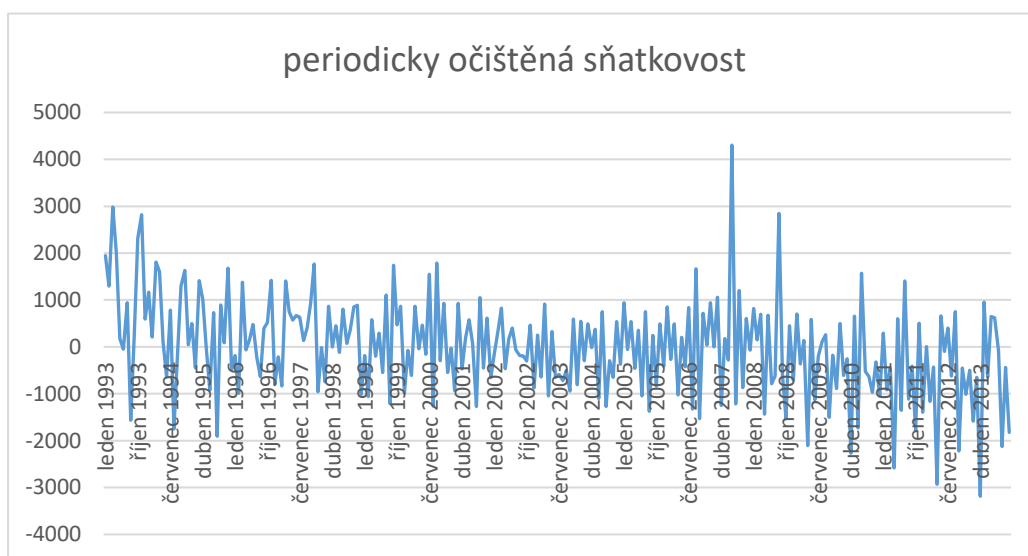
Graf č. 6 – MSP – bodová předpověď pro sňatkovost

7.2 Testování parametru lineárního regresního modelu

Dále jsem se rozhodla otestovat významnost poklesu lineárního trendu, jelikož se zdá, že pokles za dané časové období není nijak výrazný. Záměrně jsem nejdříve udělala MSP, abych nyní mohla časovou řadu očistit od periodicity (graf č. 7), tzn. odečíst od skutečných hodnot vyrovnané hodnoty z MSP a získat tak očištěné hodnoty $y_t^{(o)} = y_t - \hat{y}_t$, $t = 1, \dots, 252$. Pro tuto očištěnou časovou řadu odhadneme trend, který bude ve tvaru $\hat{y}_t^{(o)} = \hat{a}_0 + \hat{a}_1 t$. Využiji vzorce (2.5) a (2.6) a získám lineární trend $\hat{y}_t^{(o)} = 617 - 4,88t$. Je možné si všimnout, že oproti původnímu trendu neočištěné časové řady sňatků, který měl předpis $\hat{y}_t = 4951 - 4,533t$, se výrazně změnil absolutní člen, ale směrnice se liší jen nepatrně.

Testujeme na hladině $\alpha = 0,05$ nulovou hypotézu, že pokles lineárního trendu není významný, tzn. $H_0: a_1 = 0$, oproti alternativě $H_A: a_1 \neq 0$. Testová statistika bude dle [8] ve tvaru $|T| = \frac{\hat{a}_1}{\sqrt{\hat{\sigma}^2 q_{11}}}$, kde je $\hat{\sigma}^2$ nestranným odhadem rozptylu náhodné složky, $\hat{\sigma}^2 = \frac{1}{n-p} \sum_{t=1}^n (y_t^{(o)} - \hat{y}_t^{(o)})^2 = \frac{235208641}{250} = 940835$, p je počet parametrů (v případě lineárního modelu $p = 2$), matice $Q = [X^T X]^{-1}$, kde X je matice $n \times 2$, v prvním sloupci má jedničky, v druhém sloupci hodnoty $1, \dots, n$; pak prvek $q_{11} = 0,016$. Dosazením do vzorce

vypočítáme hodnotu testové statistiky $|T| = -\frac{4,533}{\sqrt{940835 \cdot 0,016}} = 0,037$, kterou porovnáme s kritickou hodnotou studentova t rozdělení o $n - p$ stupních volnosti. Nulovou hypotézu zamítneme na hladině α , pokud hodnota testové statistiky bude větší než kritická hodnota $t_{n-p}(\alpha)$. Dle [9] je hodnota $t_{250}(0,05) = 1,960$. Jelikož $T < t_{250}(0,05)$ nulovou hypotézu nelze zamítnout, parametr a_1 je statisticky nevýznamný, je tedy možné příslušnou proměnnou vynechat, protože nijak výrazně nepřispívá ke zpřesnění odhadu $\hat{y}_t^{(o)}$ a je v modelu zbytečná (proto stačí v MSP ponechat konstantní trend, tak jak jsme to provedli v předchozí podkapitole).



Graf č. 7 – Řada očištěná od periodické složky

7.3 Model konstantní sezónnosti (MKS) a předpovědi

Pro sestavení modelu konstantní sezónnosti ve tvaru (4.3) je nejprve nutné zavést indexy $i = 1, \dots, 21$ pro roky a $j = 1, \dots, 12$ pro sezóny. Dle (4.5), (4.6) a (4.7) vypočteme odhady parametrů

$$a_0 = 4378,$$

$$a_1 = -58,53,$$

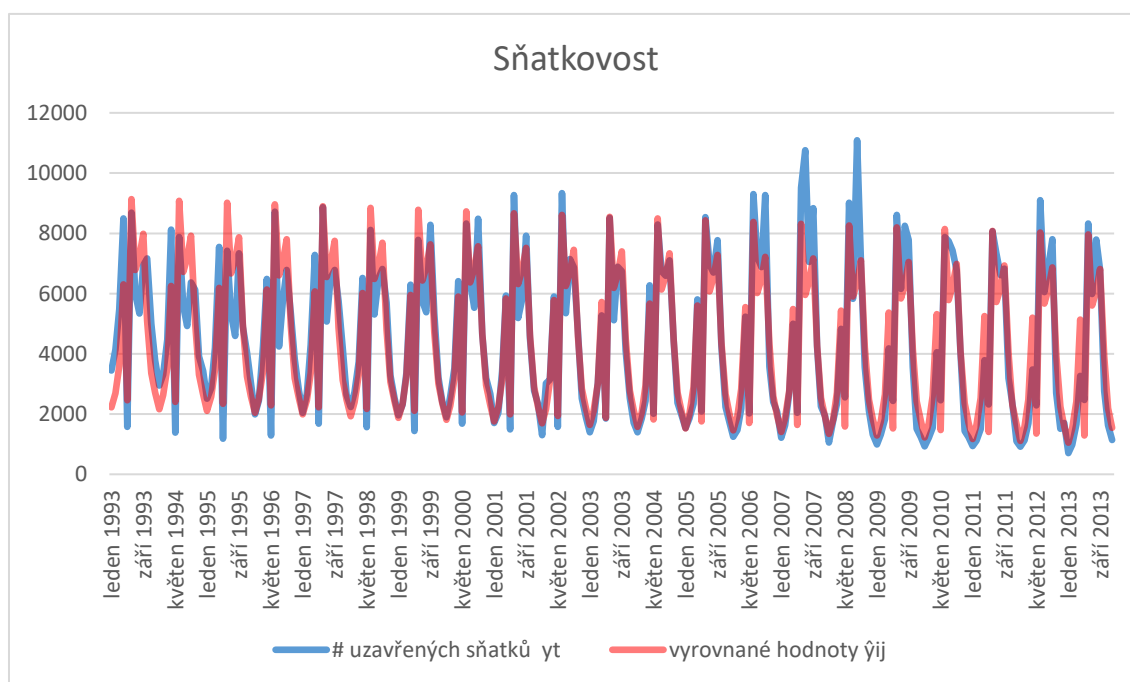
odhady b_j jsou uvedeny v následující tabulce 5.

Tabulka 5 – Odhady parametrů MKS

j	b_j
1	-2743
2	-2274
3	-1373
4	1344
5	-2502
6	4172
7	1815
8	2271
9	3018
10	126
11	-1602
12	-2253

Změna zde nastává v tom, že počítáme s centrovaným časem ($i - \bar{i}$), kde $\bar{i} = 11$. Dosazením odhadů do modelu (4.3) získáme vyrovnané hodnoty, které jsou uvedené v příloze č. 3, skutečné a vyrovnané hodnoty jsou zakresleny v grafu č. 8.

Tak jako u předchozího modelu skrytých period, je vidět, že daný model dobře pasuje na tuto časovou řadu. Opět jsem vypočítala index determinace dle (5.2) a získala hodnotu $R^2 = 0,8669$, dle (5.3) pak modifikovaný index determinace $R_a^2 = 0,8596$. Tyto hodnoty už jsou poněkud horší než pro model skrytých period, ale stále jsou blízko hodnotě 1, proto je i v tomto případě možné konstruovat na základě tohoto modelu bodové předpovědi.



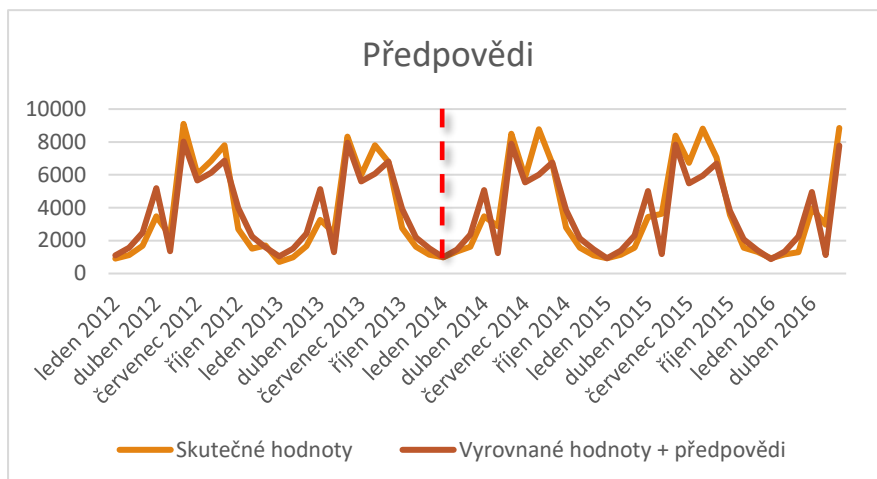
Graf č. 8 - Vyrovnané hodnoty – model konstantní sezónnosti

Předpovědi MKS

Budeme-li při konstrukci předpovědi nyní vycházet z modelu konstantní sezónnosti, budeme dosazovat do vzorce (6.3). Získáme bodové předpovědi uvedené včetně chyb předpovědí v tabulce 6. Abychom mohli například porovnávat modely pro časovou řadu sňatků, určíme průměrnou čtvercovou chybu předpovědi s použitím vzorce (6.6), $MSEP = \sqrt{1390229} = 1179$. Pro porovnání, u modelu skrytých period byla tato hodnota 8490, z čehož je možné usoudit, že předpovědi, při kterých jsme vycházeli z modelu konstantní sezónnosti, jsou mnohem přesnější než ty, při kterých jsme vycházeli z modelu skrytých period. V grafu č. 9 jsou opět za červenou čarou vykresleny předpovědi.

Tabulka 6 – Předpovědi – MKS

<i>období</i>	<i>h, j</i>	$y_{m+h,j}$	$\hat{y}_{m+h,j}^P$	$\Delta_{m+h,j}$	$\delta_{m+h,j}$
leden 2014	h=1,j=1	981	991	-10	-1,1
únor 2014	h=1,j=2	1333	1460	-127	-9,5
březen 2014	h=1,j=3	1628	2361	-733	-45,1
duben 2014	h=1,j=4	3477	5079	-1602	-46,1
květen 2014	h=1,j=5	2872	1232	1640	57,1
červen 2014	h=1,j=6	8504	7906	598	7,0
červenec 2014	h=1,j=7	5861	5550	311	5,3
srpen 2014	h=1,j=8	8766	6005	2761	31,5
září 2014	h=1,j=9	6686	6752	-66	-1,0
říjen 2014	h=1,j=10	2794	3860	-1066	-38,2
listopad 2014	h=1,j=11	1565	2133	-568	-36,3
prosinec 2014	h=1,j=12	1108	1481	-373	-33,7
leden 2015	h=2,j=1	932	933	-1	-0,1
únor 2015	h=2,j=2	1145	1402	-257	-22,4
březen 2015	h=2,j=3	1565	2303	-738	-47,1
duben 2015	h=2,j=4	3444	5020	-1576	-45,8
květen 2015	h=2,j=5	3644	1174	2470	67,8
červen 2015	h=2,j=6	8379	7848	531	6,3
červenec 2015	h=2,j=7	6731	5491	1240	18,4
srpen 2015	h=2,j=8	8814	5946	2868	32,5
září 2015	h=2,j=9	7069	6694	375	5,3
říjen 2015	h=2,j=10	3583	3802	-219	-6,1
listopad 2015	h=2,j=11	1572	2074	-502	-31,9
prosinec 2015	h=2,j=12	1313	1423	-110	-8,4
leden 2016	h=3,j=1	937	874	63	6,7
únor 2016	h=3,j=2	1165	1343	-178	-15,3
březen 2016	h=3,j=3	1296	2244	-948	-73,2
duben 2016	h=3,j=4	3956	4961	-1005	-25,4
květen 2016	h=3,j=5	2974	1115	1859	62,5
červen 2016	h=3,j=6	8849	7789	1060	12,0



Graf č. 9 – Předpovědi – model konstantní sezónnosti

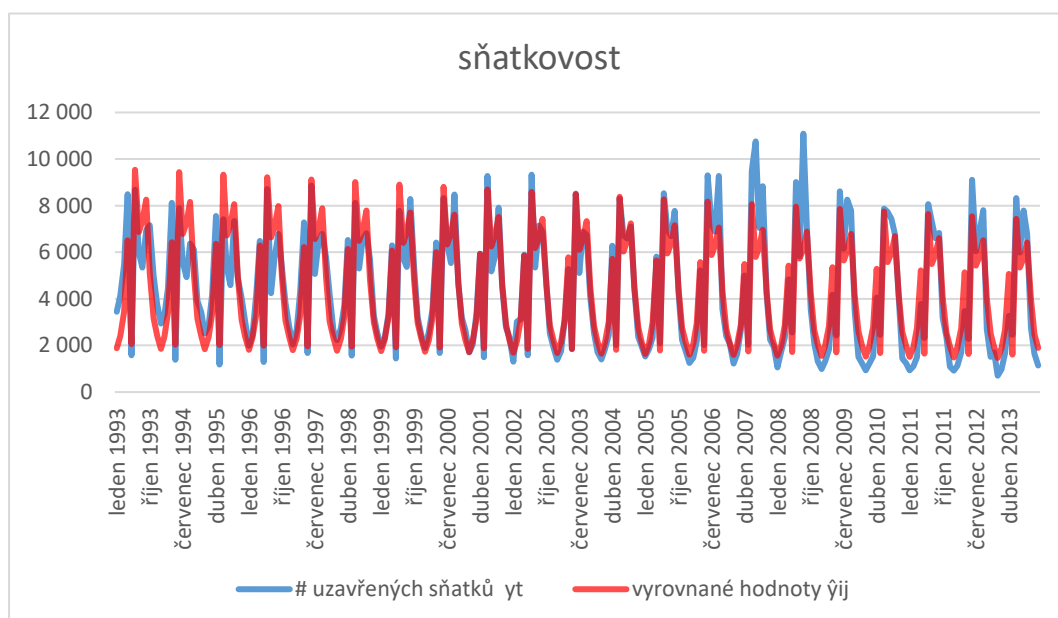
7.4 Model proporcionální sezónnosti (MPS) a předpovědi

Jako poslední zkusíme aplikovat na danou časovou řadu model proporcionální sezónnosti (4.8). Stejně jako u modelu konstantní sezónnosti zde používáme indexy $i = 1, \dots, 21$ pro roky a $j = 1, \dots, 12$ pro sezóny. Dále je nutné určit odhady parametrů sezóny ($\widehat{1 + c_j}$) dle vzorce (4.9) a následně dle vztahu (4.10) znormujeme. Pro výpočet jsme požili trend odhadnutý již na začátku kapitoly 7, a to $\hat{y}_t = 4951 - 4,533t$; $t = 1, \dots, 252$. Dostaneme hodnoty uvedené v následující tabulce 7. Z tabulky je nyní zřejmé, které měsíce jsou oblíbené a které méně. Duben, červen, červenec, srpen, září a říjen mají nejvyšší hodnoty, což značí, že jsou to nejoblíbenější měsíce. Neboli čím vyšší hodnota, tím oblíbenější doba pro uzavírání sňatků na území ČR.

Po dosažení odhadů do modelu získáme vyrovnané hodnoty, které jsou uvedeny v příloze č. 3 a celou časovou řadu zobrazuje graf č. 10. Index determinace v tomto případě vychází $R^2 = 0,845$, modifikovaný $R_a^2 = 0,837$. V porovnání s modelem skrytých period a modelem konstantní sezónnosti mají indexy nyní nejnižší hodnotu, což značí, že model proporcionální sezónnosti je nejméně vhodný.

Tabulka 7 – Parametry sezóny

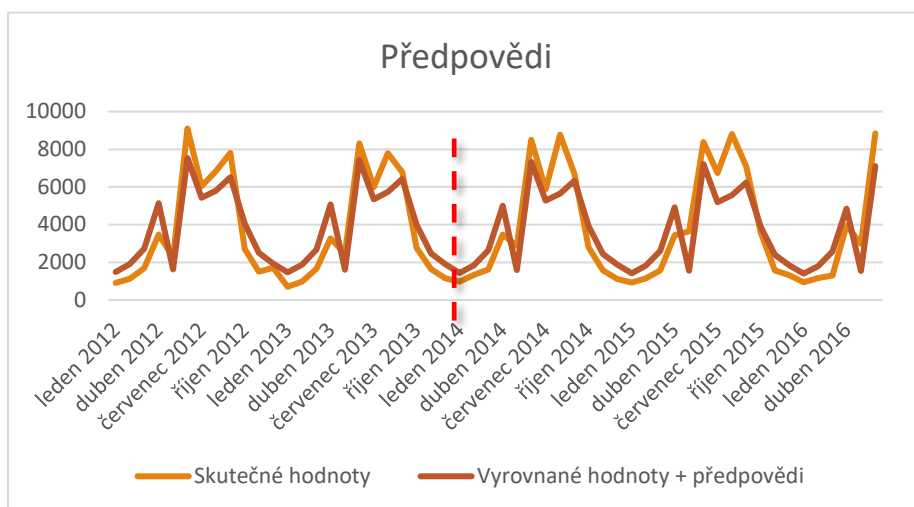
j	$(\widehat{1+c_j})$	$(\widehat{1+c_j})^{(N)}$
1	0,38036261	0,38004607
2	0,488235623	0,48782931
3	0,698123356	0,69754238
4	1,32004839	1,31894984
5	0,418906072	0,41855746
6	1,938648365	1,93703502
7	1,396475293	1,39531314
8	1,496927215	1,49568147
9	1,681602004	1,68020257
10	1,044223714	1,04335471
11	0,649088237	0,64854806
12	0,497353858	0,49693996
Σ	12,01	12



Graf č. 10 – Vyrovnané hodnoty – model proporcionální sezónnosti

Předpovědi MPS

I pro tento model zkusíme zkonstruovat bodové předpovědi, využijeme vzorce (6.4). Jediné, co je nutné při konstrukci předpovědí u tohoto typu modelu udělat, je vypočítat budoucí hodnoty trendu. Předpovědi i s chybami předpovědí jsou uvedeny v následující tabulce 8. V grafu č. 11 jsou opět za červenou čarou vykresleny předpovědi.



Graf č. 11 – Bodové předpovědi – MPS – sňatkovost

Tabulka 8 – Předpovědi MPS

období	h, j	$y_{m+h,j}$	$\hat{y}_{m+h,j}^P$	$\Delta_{m+h,j}$	$\delta_{m+h,j}$
leden 2014	h=1,j=1	981	1443	-462	-47
únor 2014	h=1,j=2	1333	1851	-518	-39
březen 2014	h=1,j=3	1628	2644	-1016	-62
duben 2014	h=1,j=4	3477	4995	-1518	-44
květen 2014	h=1,j=5	2872	1584	1288	45
červen 2014	h=1,j=6	8504	7323	1181	14
červenec 2014	h=1,j=7	5861	5271	590	10
srpen 2014	h=1,j=8	8766	5645	3121	36
září 2014	h=1,j=9	6686	6336	350	5
říjen 2014	h=1,j=10	2794	3931	-1137	-41
listopad 2014	h=1,j=11	1565	2441	-876	-56
prosinec 2014	h=1,j=12	1108	1869	-761	-69
leden 2015	h=2,j=1	932	1422	-490	-53
únor 2015	h=2,j=2	1145	1824	-679	-59
březen 2015	h=2,j=3	1565	2606	-1041	-67
duben 2015	h=2,j=4	3444	4923	-1479	-43
květen 2015	h=2,j=5	3644	1561	2083	57
červen 2015	h=2,j=6	8379	7218	1161	14
červenec 2015	h=2,j=7	6731	5195	1536	23
srpen 2015	h=2,j=8	8814	5564	3250	37
září 2015	h=2,j=9	7069	6244	825	12
říjen 2015	h=2,j=10	3583	3874	-291	-8
listopad 2015	h=2,j=11	1572	2406	-834	-53
prosinec 2015	h=2,j=12	1313	1842	-529	-40
leden 2016	h=3,j=1	937	1402	-465	-50
únor 2016	h=3,j=2	1165	1798	-633	-54
březen 2016	h=3,j=3	1296	2568	-1272	-98
duben 2016	h=3,j=4	3956	4852	-896	-23
květen 2016	h=3,j=5	2974	1538	1436	48
červen 2016	h=3,j=6	8849	7113	1736	20

Na závěr ještě vypočítáme MSEP pro model proporcionální sezónnosti dle (6.6),
 $MSEP = \sqrt{1738799} = 1319$.

Pokud porovnáme všechny 3 modely (model skrytých period, model konstantní sezónnosti a model proporcionální sezónnosti), zjistíme, že při konstrukci předpovědi nejlépe obstál MKS s hodnotou $MSEP = 1179$, na dalším místě je MPS s chybou 1319 a nejhůře při předpovídání dopadl model skrytých period s hodnotou 1634.

Avšak pokud porovnáme indexy determinace, vychází jako nejvhodnější MSP, dále MKS a nakonec MPS. Porovnání je uvedeno v tabulce 9 (zde p indexuje počet parametrů).

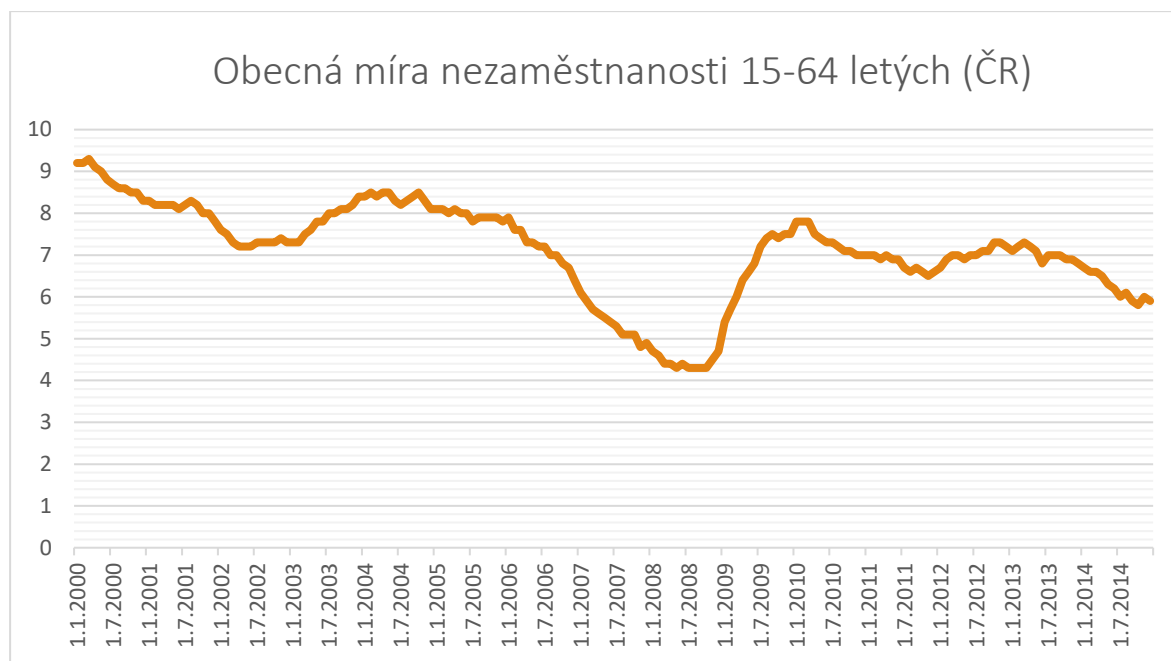
Tabulka 9 – Porovnání modelů

	p	R^2	R_a^2	MSEP
MSP	15	0,958	0,956	1634
MKS	14	0,867	0,86	1179
MPS	12	0,845	0,837	1319

Kdybychom chtěli určit, který z modelů je nejlepší, bude záležet na tom, co je pro nás nejdůležitější. Pokud vyžadujeme co nejpřesnější předpověď, zvolili bychom jistě model konstantní sezónnosti, jehož hodnoty předpovědí vyšly nejlépe. Navíc lze brát v úvahu obtížnost jednotlivých metod. V tomto směru hraje roli i počet parametrů v modelu, jelikož čím nižší počet parametrů, tím jednodušší model, a mít co nejjednodušší model je taktéž jeden z požadavků při analýze časových řad. Model skrytých period má sice nejlepší indexy determinace, avšak chyba předpovědi je v tomto případě největší, a navíc byl docela náročný na výpočty, bylo potřeba vytvořit několik velmi rozsáhlých tabulek, abychom mohly určit parametry modelu atd. Dalo by se tedy říct, že každá metoda je v některém směru vítěz. Při použití MPS jsme pracovali s nejméně parametry, dle MKS jsme získali nejlepší předpovědi a dle MSP vycházejí nejlépe indexy determinace.

8 Nezaměstnanost

Druhou sadu dat jsem převzala z veřejné databáze časových řad ARAD, která je součástí informačního servisu České národní banky [10]. Konkrétně se jedná o obecnou míru nezaměstnanosti 15 – 64 letých na území ČR. Jde o měsíční časovou řadu, jednotlivá pozorování jsou vždy uváděna na konci každého měsíce. Zvolila jsem časovou řadu o délce $N = 200$, kterou jsem o 20 pozorování zkrátila, abych následně mohla porovnat predikce se skutečnými hodnotami. Pro analýzu jsem tedy použila řadu o délce $n = 180$, a to leden 2000 až prosinec 2014. Všechna data jsou uvedena v příloze č. 4, a vykresleny jsou v následujícím grafu č. 12.



Graf č. 12 – Nezaměstnanost v ČR

Na první pohled není zřejmá periodicitu, proto můžeme zkusit na data použít jednotlivé modely a prozkoumat, zda má časová řada nějaké významné periody či nikoliv.

8.1 Model skrytých period a předpovědi

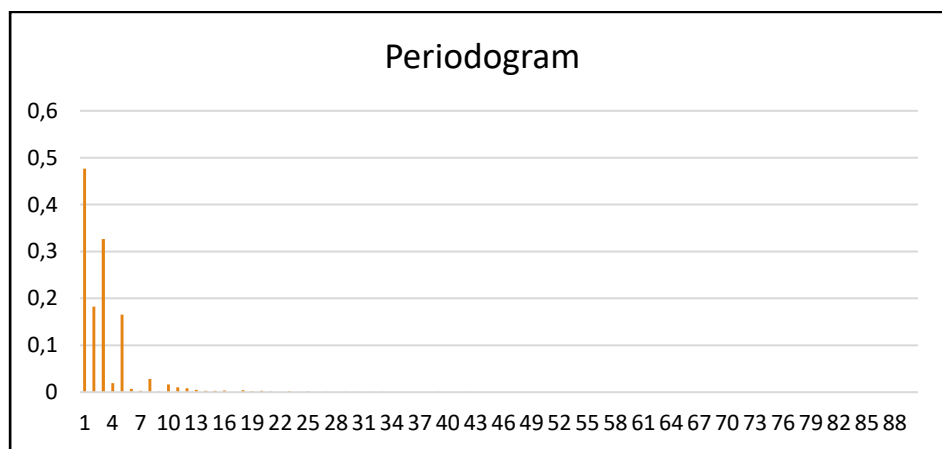
Pro model (3.2) budeme potřebovat úhlové frekvence $\omega_j = \frac{2\pi j}{n}$ pro každé $j = 1, \dots, H$, kde $H = \frac{n}{2} = \frac{180}{2} = 90$. Tyto frekvence $\omega_1, \omega_2, \dots, \omega_{90}$ využijeme pro výpočet odhadů parametrů a_0, a_j, b_j . Nejjednodušší odhad parametru a_0 dle vzorce (3.2) vypadá následovně

$$\hat{a}_0 = \frac{1}{252} \sum_{t=1}^n y_t = 7,1089.$$

Odhady $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_{90}, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_{90}$ jsou uvedeny v příloze č. 5 a získáme je dosazením do vzorců (3.3) a (3.4), tj.

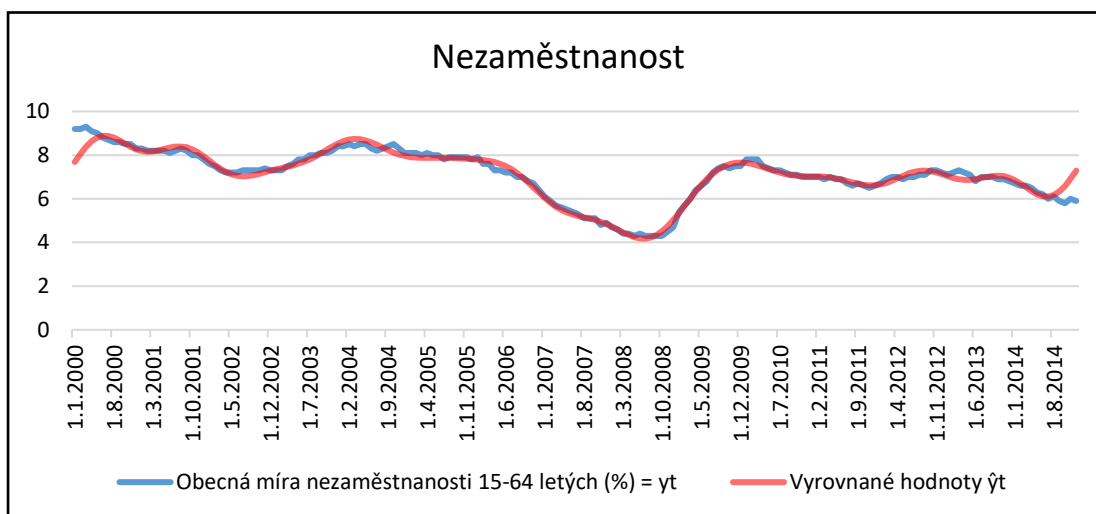
$$\hat{a}_j = \frac{2}{180} \sum_{t=1}^{180} y_t \sin(\omega_j t) \text{ a } \hat{b}_j = \frac{2}{180} \sum_{t=1}^{180} y_t \cos(\omega_j t).$$

Pro identifikaci významných period využijeme opět periodogram. V příloze č. 5 jsou vypočítané hodnoty $I(\omega_j)$ pro každé $j = 1, \dots, 90$, dle vzorce (3.5), zakresleny jsou v grafu č. 13.



Graf č. 13 – Periodogram – nezaměstnanost

Z periodogramu lze opět vizuálně posoudit, které periody jsou významné, avšak to je pouze subjektivní, proto je třeba opět provést objektivní statistický test, a to Fisherův test významnosti period. Výsledky ukázaly, že významné periody jsou $j = 1, 2, 3, 4, 5, 8, 10, 11, 12$, kde nejvýznamnější vyšla hodnota $j = 1$, která odpovídá periodě o délce 180 měsíců. Těchto devět period zahrneme do modelu skrytých period. Dosazením do modelu (3.2) získáme vyrovnané hodnoty \hat{y}_t , které jsou uvedeny v příloze č. 5. Následující graf č. 14 zobrazuje skutečné hodnoty časové řady a vyrovnané hodnoty.



Graf č. 14 – Model skrytých period

I přesto, že dle periodogramu a Fisherova testu časová řada vykazuje pouze velmi dlouhé periody, z grafu se zdá, že model na danou řadu pasuje dobře. Kvalitu modelu jsem určila pomocí indexu determinace R^2 ze vzorce (5.2) a modifikovaného indexu determinace R_a^2 ze vzorce (5.3)

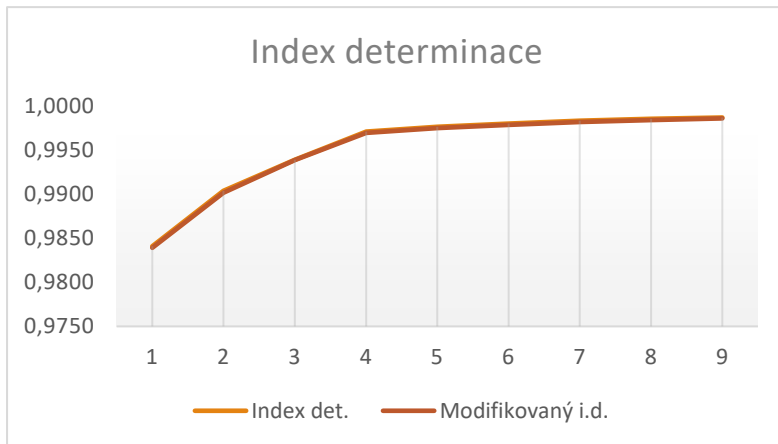
$$R^2 = 1 - \frac{12,11850485}{9330,74} = 0,9987,$$

$$R_a^2 = 0,9987 - \frac{(1-0,9987)(19-1)}{180-19} = 0,9986 .$$

Obě hodnoty jsou velmi blízké 1, což sice značí velice dobrou kvalitu modelu, avšak jak se za chvíli ukáže, není vhodný pro konstrukci předpovědí. Stejně jako u časové řady sňatků jsem i nyní vyzkoušela odebrat některé periody a zjistit, co se bude dít s klasickým a modifikovaným indexem determinace neboli jak se model zhorší, když zahrneme méně period (viz tabulka 10 a graf č. 15).

Tabulka 10 – Index determinace a počet period

p	R^2	R_a^2
1	0,9841	0,9839
2	0,9904	0,9902
3	0,9939	0,9939
4	0,9971	0,9970
5	0,9976	0,9975
6	0,9980	0,9979
7	0,9983	0,9982
8	0,9985	0,9984
9	0,9987	0,9986

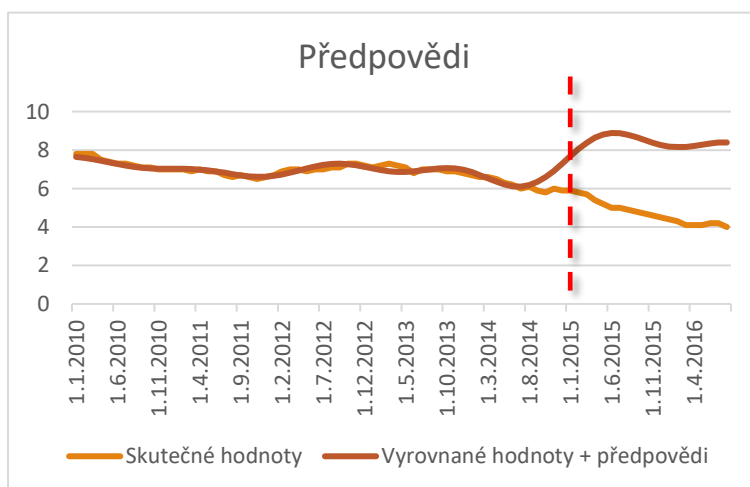


Graf č. 15 - Index determinace v závislosti na počtu period v MSP

Z grafu vyplývá, že do modelu není nutné zahrnovat všech devět period. Již při zahrnutí čtyř period je index determinace velmi blízka 1. Já jsem však opět zahrнула všech devět period, abych docílila co nejpřesnějších prognóz.

Předpovědi MSP

Na začátku jsem si časovou řadu zkrátila o 20 pozorování. Původní časová řada je od ledna 2000 do srpna 2016. Nyní mám možnost zkonstruovat bodové předpovědi a porovnat je se skutečnými hodnotami. Vypočítám tedy dle vzorce (6.2) předpovědi \hat{y}_{n+h}^P , kde $h = 1, \dots, 20$, tzn. vypočítám hodnoty $\hat{y}_{181}^P, \hat{y}_{182}^P, \dots, \hat{y}_{200}^P$. Hodnoty jsou vedle skutečných hodnot a včetně chyb předpovědí uvedeny v tabulce 11 a zakresleny v grafu č.16.



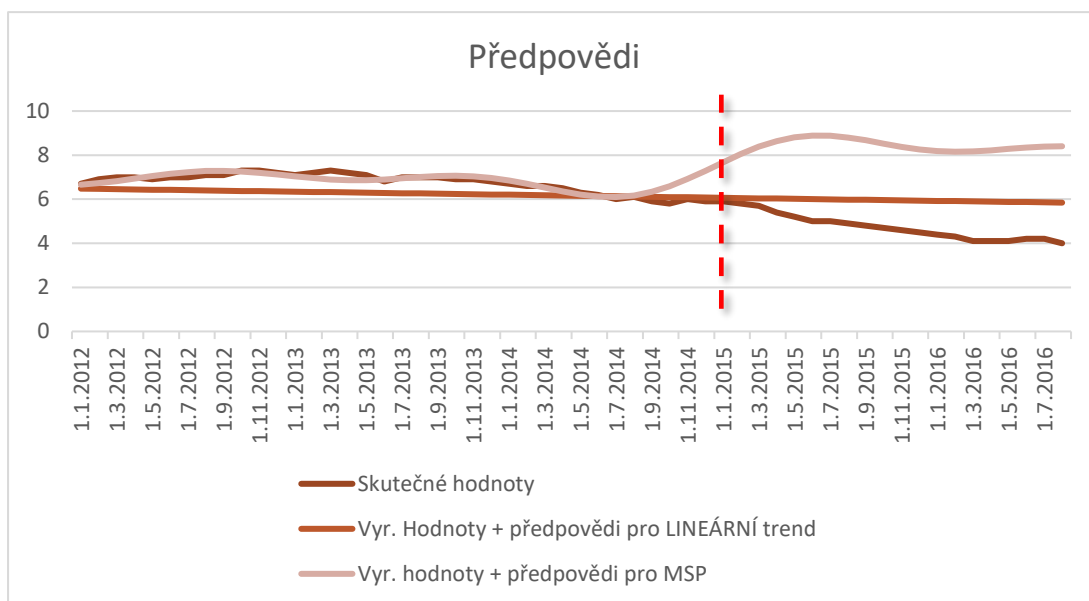
Graf č. 16 – Bodové předpovědi MSP – nezaměstnanost

Tabulka 11 – Předpovědi MSP – nezaměstnanost

<i>období</i>	<i>h</i>	<i>n + h</i>	y_{n+h}	\hat{y}_{n+h}^P	Δ_{n+h}	$\delta_{n+h}(\%)$
31.1.2015	1	181	5,9	7,7	-1,8	-30,2
28.2.2015	2	182	5,8	8,1	-2,3	-38,9
31.3.2015	3	183	5,7	8,4	-2,7	-47,0
30.4.2015	4	184	5,4	8,6	-3,2	-59,9
31.5.2015	5	185	5,2	8,8	-3,6	-69,4
30.6.2015	6	186	5	8,9	-3,9	-77,7
31.7.2015	7	187	5	8,9	-3,9	-77,5
31.8.2015	8	188	4,9	8,8	-3,9	-79,6
30.9.2015	9	189	4,8	8,7	-3,9	-80,7
31.10.2015	10	190	4,7	8,5	-3,8	-81,4
30.11.2015	11	191	4,6	8,4	-3,8	-82,2
31.12.2015	12	192	4,5	8,3	-3,8	-83,6
31.1.2016	13	193	4,4	8,2	-3,8	-86,0
29.2.2016	14	194	4,3	8,2	-3,9	-89,6
31.3.2016	15	195	4,1	8,2	-4,1	-99,2
30.4.2016	16	196	4,1	8,2	-4,1	-100,3
31.5.2016	17	197	4,1	8,3	-4,2	-102,0
30.6.2016	18	198	4,2	8,3	-4,1	-98,7
31.7.2016	19	199	4,2	8,4	-4,2	-99,7
31.8.2016	20	200	4	8,4	-4,4	-109,9

Dle vzorce (6.6) jsme si vypočítali i průměrnou čtvercovou chybu předpovědi $MSEP = \sqrt{13,822} = 3,718$. Jak je vidět i v grafu, bodové předpovědi jsou velmi nepřesné a zajisté není vhodné na podobný typ časové řady využívat tento model a dle něj konstruovat předpovědi.

Do modelu skrytých period jsme zahrnuli 9 period, které vyšly dle Fisherova testu jako nejvýznamnější, tzn. celkově jsme v modelu (3.1) měli 19 parametrů. Snaha zachytit co nejvíce výkyvů v řadě a popsat ji velkým množstvím vln čili uvažovat velké množství parametrů, se někdy ukáže jako zbytečné. Tomuto jevu se říká tzv. „*overfitting*“ (přeurčení modelu). Budeme-li uvažovat pro danou časovou řadu jednoduchý lineární trend, uvidíme, že je to vhodnější a predikce jsou přesnější. S využitím (2.5) a (2.6) jsem určila rovnici trendu $\hat{y}_t = 8,154 - 0,012t, t = 1, \dots, 180$, pro který jsem vypočítala dle (6.1) předpovědi $\hat{y}_{n+h}^P, h = 1, \dots, 20$; a následně vypočítala $MSEP = \sqrt{1,738} = 1,318$. Predikce při užití lineárního trendu jsou tedy kvalitnější než při užití mnohem složitějšího MSP. (viz graf č. 17)



Graf č. 17 – Bodové předpovědi pro nezaměstnanost při užití lineárního trendu nebo MSP

Nyní bychom mohli vyzkoušet model konstantní a proporcionální sezónnosti, avšak už teď je jisté, že závěr bude podobný jako v případě modelu skrytých period. Navíc tyto dva modely je možné použít pouze pro časové řady, které obsahují sezónnost (periody kratší než jeden rok).

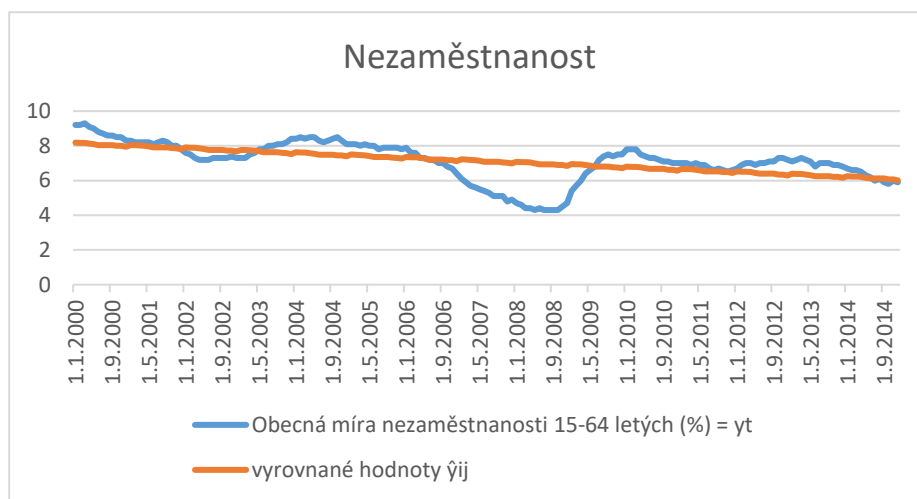
8.2 Model konstantní sezónnosti a předpovědi

Pro zajímavost můžeme vyzkoušet, jak by model vypadal. Při konstrukci modelu bychom postupovali stejně jako v případě časové řady sňatků v kapitole 7. Indexy, které značí roky, budou $i = 1, \dots, 15$ a pro měsíce $j = 1, \dots, 12$. Dle (4.5), (4.6) a (4.7) vypočteme odhady parametrů $\hat{a}_0 = 7,109$, $\hat{a}_1 = -0,138$ a \hat{b}_j jsou v tabulce 12.

Tabulka 12 – Odhady parametrů

j	\hat{b}_j	\hat{b}_j^2
1	0,11	0,012
2	0,10	0,010
3	0,09	0,008
4	0,06	0,004
5	0,02	0,001
6	-0,03	0,001
7	-0,03	0,001
8	-0,03	0,001
9	-0,03	0,001
10	-0,07	0,005
11	-0,08	0,006
12	-0,13	0,017

Dosadíme tyto odhady do modelu (4.3) a získáme vyrovnané hodnoty $\hat{y}_t, t = 1, \dots, 180$. Skutečné i vyrovnané hodnoty jsou v grafu č. 18.

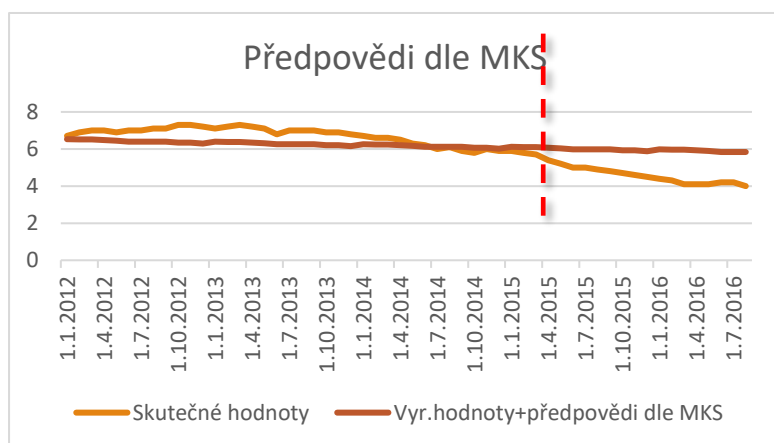


Graf č. 18 – Model konstantní sezónnosti

Jak je z grafu zřejmé, model není vhodný pro tuto časovou řadu. Jak je vidět už v tabulce 12, sezónní indexy jsou příliš malé, téměř zanedbatelné, což se i v grafu projevilo tak, že je křivka vyrovnaných hodnot pouze lehce hrbolatá, místo toho, aby zde byly nějaké vlny. Nepřítomnost sezónnosti lze dokázat i pomocí statistického testu, který jsem provedla v následující kapitole 8.3.

Předpovědi MKS

Jelikož jsou parametry b_j v MKS tak malé, je logické, že předpovědi (počítané z 6.3) dle tohoto modelu budou podobné jako predikce pro obyčejný lineární trend, které jsme určili v předchozí podkapitole. Zobrazeny jsou za červenou čarou v následujícím grafu č. 19.



Graf č. 19 – Předpovědi dle MKS

Střední čtvercová chyba předpovědi vychází $MSEP = \sqrt{1,753} = 1,324$, což je ještě o něco větší než chyba u předpovědi v případě, kdy jsme uvažovali lineární trend (zde to bylo 1,318).

8.3 Test přítomnosti sezónnosti

Dle [4] pro model konstantní a proporcionální sezónnosti ve tvaru (4.3) a (4.8) můžeme testovat nulovou hypotézu $H_0: b_j = 0, j = 1, \dots, 12$, tzn. že není přítomna sezónnost. Testujeme oproti alternativě $H_A: \exists j: b_j \neq 0, j = 1, \dots, 12$. Budeme testovat na hladině $\alpha = 0,05$ a testová statistika je ve tvaru

$$F = \frac{\frac{m \sum_{j=1}^r \hat{b}_j^2}{(r-1)}}{\frac{RS\check{C}}{(n-p)}}$$

kde r je počet sezón ($r = 12$), m počet let ($m = 15$), n je počet pozorování ($n = 180$), $RS\check{C}$ neboli reziduální součet čtverců vypočtený dle vzorce (5.1) ($RS\check{C} = 169,18$), p je počet volných parametrů, v případě MKS s lineárním trendem je $p = r + 1$.

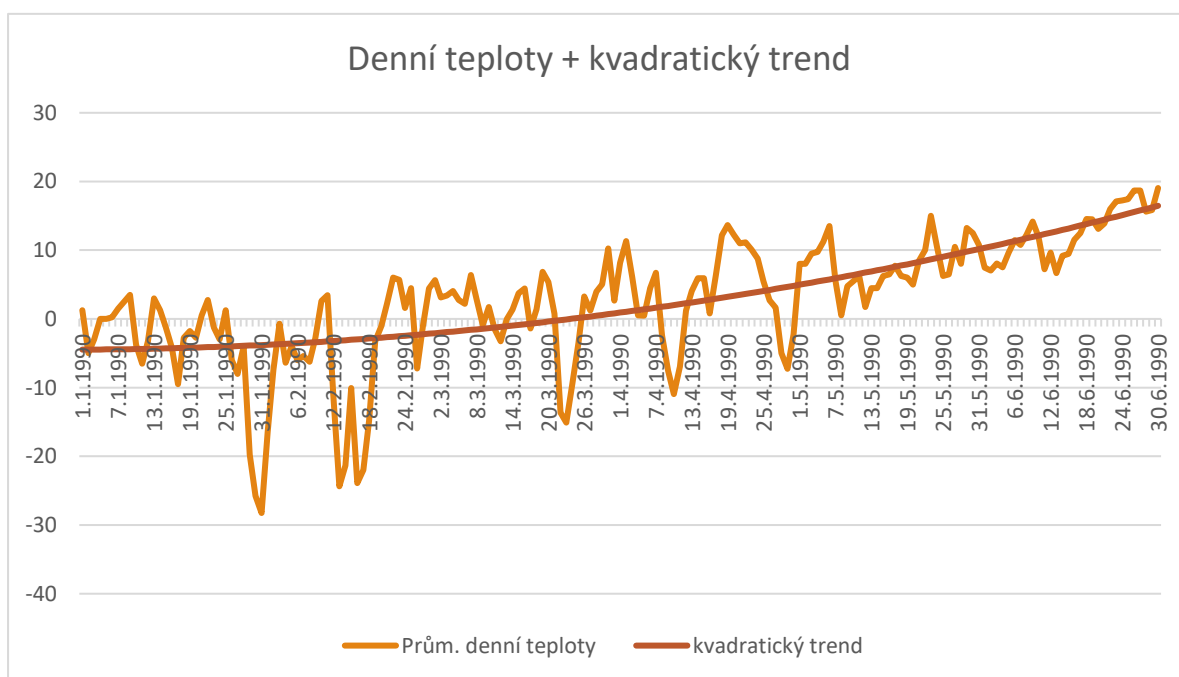
Nulovou hypotézu zamítneme na hladině α , pokud platí $F > F_{r-1, n-p}(\alpha)$. Parametry b_j mám již určené z modelu konstantní sezónnosti (viz tabulka 12). Po dosazení získám $F = 0,088$ a porovnávám s kritickou hodnotou Fisher-Snedecorova rozdělení $F_{[v_1, v_2]}$ o v_1 a v_2 stupních volnosti. V našem případě je to hodnota $F_{11, 167} = 0,4158$, což znamená, že $F < F_{11, 167}(0,05)$ a nulovou hypotézu nelze zamítnout. Nelze zamítnout hypotézu, že není přítomna sezónnost, což znamená, že je zbytečné konstruovat model konstantní a proporcionální sezónnosti.

Na závěr kapitoly 8 můžeme porovnat jednotlivé modely pro časovou řadu nezaměstnanosti. Ukázalo se, že v této časové řadě není přítomna sezónnost, proto jsou sezónní modely zbytečné. Jako nejlepší se ukázala volba jednoduchého lineárního trendu, předpovědi byly v tomto případě nejpřesnější (chyba $MSEP$ jen 1,318), hned v závěsu byl model konstantní sezónnosti (chyba 1,324), což je způsobeno tím, že parametry b_j byly v tomto modelu tak malé, že se dost podobal lineárnímu trendu. Nejhuře dopadl model skrytých period s chybou předpovědi 3,718.

9 Denní teploty

Třetí datovou sadu jsem převzala z internetových stránek [11]. Vybrala jsem časovou řadu průměrných denních teplot v oblasti Oldman River v Kanadě, a to především proto, že se jedná o denní časovou řadu, zatímco předešlé časové řady byly měsíční. Pro analýzu jsem použila data z období 1.leden 1990 až 30.červen 1990, tzn. délka řady je $n = 181$, původní časová řada je o 30 pozorování delší ($N = 211$), opět jsem ponechala těchto 30 hodnot pro porovnání predikcí se skutečnými hodnotami. Analyzovaná časová řada je zobrazena v grafu č. 20 včetně trendu, pozorování jsou uvedena v příloze č. 6.

9.1 Trend



Graf č. 20 – Průměrné denní teploty, Oldman River, Canada + kvadratický trend

Jak je možné z grafu vidět, trend nejprve lehce klesá a poté začíná růst, což je logické, blíží se léto a denní teploty rostou. Proto je možné pro danou časovou řadu uvažovat kvadratický trend ve tvaru (2.7). Nyní jednoduchým dosazením do vzorců (2.8), (2.9) a (2.10) určíme odhady parametrů

$$\hat{a}_0 = -4,4686,$$

$$\hat{a}_1 = 0,0021,$$

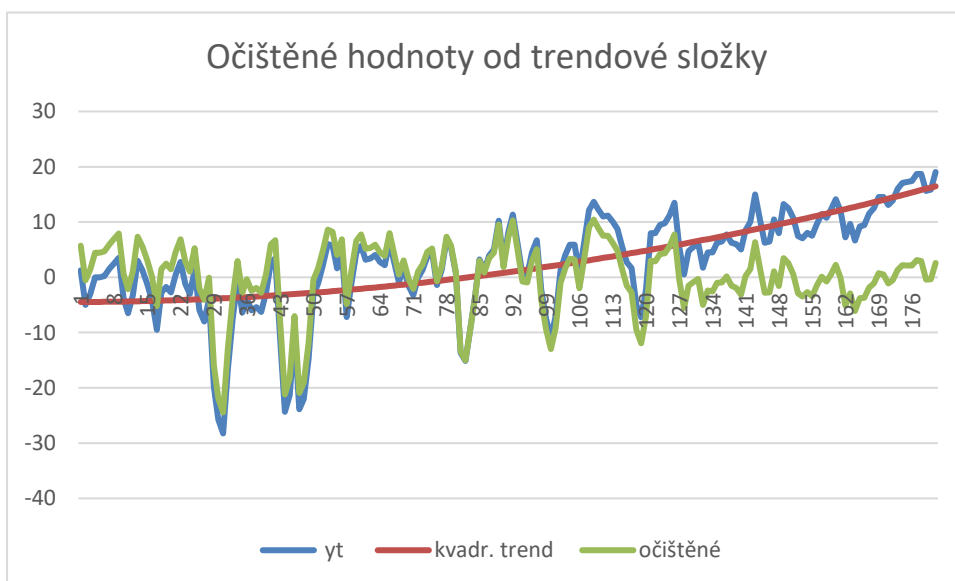
$$\hat{a}_2 = 0,0006.$$

Vyrovnané hodnoty $\hat{T}_t = -4,4686 + 0,0021t + 0,0006t^2$ jsou zakresleny také v grafu č. 18. Hodnoty \hat{a}_1 i \hat{a}_2 jsou dost malé, můžeme proto zkusit provést statistický test významnosti parametru.

Testujeme nejprve hypotézu $H_0: a_2 = 0$ oproti opačné alternativě. Testová statistika je ve tvaru $|T| = \frac{\hat{a}_2}{\sqrt{\hat{\sigma}^2 q_{11}}}$, kde $\hat{\sigma}^2 = \frac{1}{n-p} \sum_{t=1}^n (y_t - \hat{y}_t)^2 = \frac{7208}{178} = 40,5$, p je počet parametrů (pro kvadratický trend $p = 3$), matice $Q = [X^T X]^{-1}$, X je nyní matice $3 \times n$, její prvek $q_{11} = 0,051$. Dosadíme do $|T| = \frac{0,0006}{\sqrt{40,5 * 0,051}} = 0,00042$ a porovnáme s kritickou hodnotou $t_{178}(0,05) = 1,960$. Hypotézu nelze zamítnout, jelikož $T < t_{178}(0,05)$. Parametr a_2 tedy není až tak významný, proto by bylo možné proměnnou t^2 vynechat a uvažovat jednodušší lineární trend, což se na závěr, jak uvidíme, ukáže jako nejlepší volba.

9.2 Model skrytých period

Před konstrukcí modelu skrytých period očistíme řadu od trendové složky (zde jsem uvažovala kvadratický trend). Vypočítáme očištěné hodnoty $y_t^{(o)}$ tak, že od hodnot zadané časové řady odečteme vyrovnané hodnoty trendu, tj. $y_t^{(o)} = y_t - \hat{T}_t$ a dostaneme hodnoty uvedené v příloze č. 6 a grafu č. 21.



Graf č. 21 – Řada očištěná od kvadratického trendu

A právě s těmito hodnotami budeme pracovat při sestavování modelu skrytých period. Následně dle periodogramu a Fisherova testu určíme, které periody jsou nejvýznamnější a zakreslíme do grafu výsledný model.

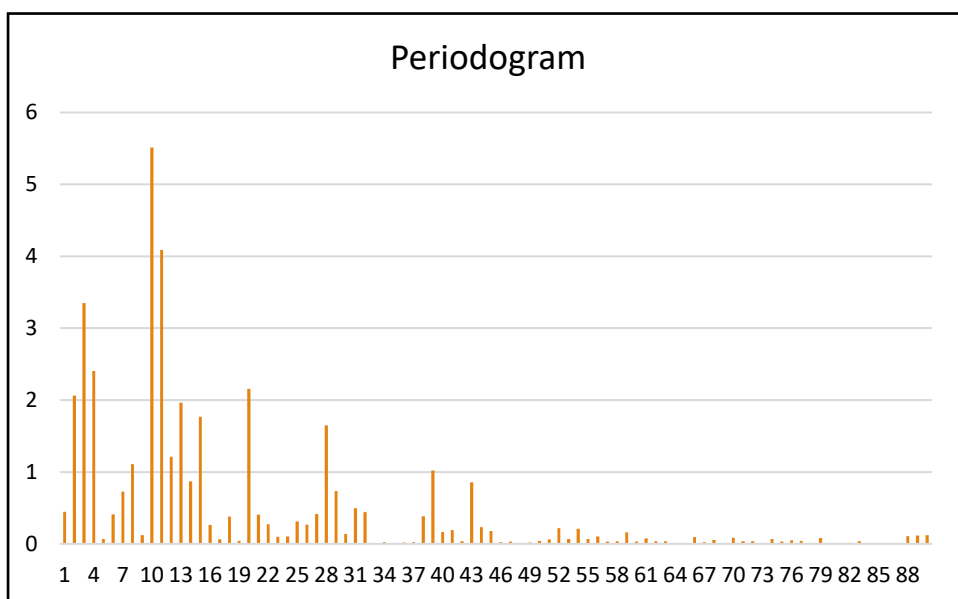
Nejprve určíme úhlové frekvence $\omega_j = \frac{2\pi j}{n}$, $j = 1, \dots, H$, $H = \frac{181-1}{2} = 90$. Výsledné $\omega_1, \omega_2, \dots, \omega_{90}$ použijeme pro výpočet odhadů a dle vzorců (3.2), (3.3) a (3.4) vypočítáme odhady parametrů a_0, a_j, b_j .

Absolutní člen $\hat{a}_0 = 0$, jelikož se jedná o očištěnou časovou řadu, některé odhady \hat{a}_j a \hat{b}_j jsou uvedeny v následující tabulce 13.

Tabulka 13 – Odhady parametrů – MSP – teploty

j	\hat{a}_j	\hat{b}_j
1	-0,66	-0,68
2	-0,36	2,00
3	1,44	2,15
4	1,98	-0,94
5	-0,37	-0,03
6	-0,26	0,87
7	0,55	-1,07
8	-1,46	0,31
9	-0,41	-0,27
10	1,52	2,95

Dosažením do vzorce (3.5) určíme hodnoty periodogramu a zakreslíme do grafu č. 22.



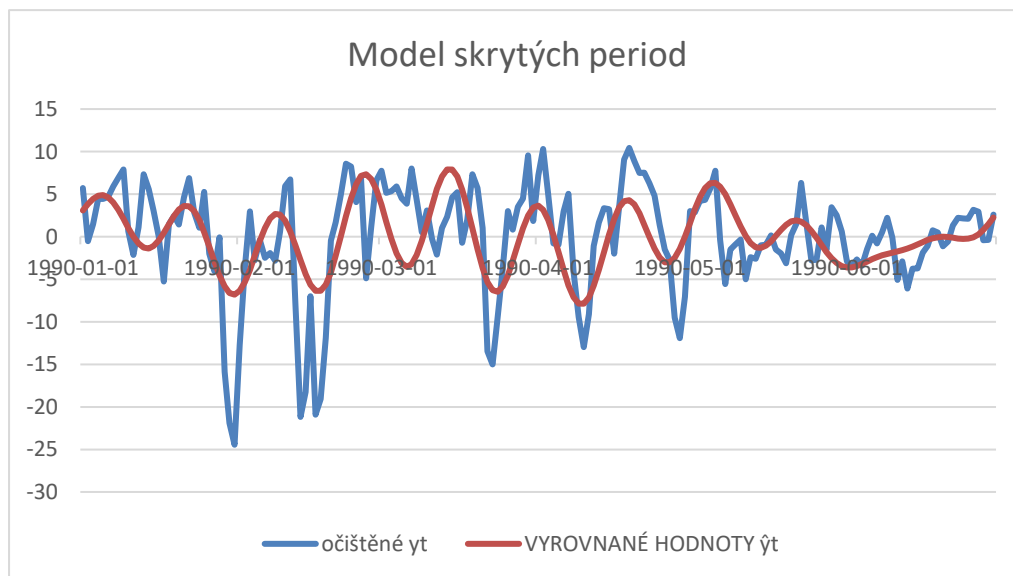
Graf č. 22 – Periodogram – teploty

Dle periodogramu se zdá, že je významných hned několik period, proto provedeme i Fisherův test významnosti period, postupujeme analogicky jako v předchozích kapitolách. Ačkoliv z grafu se zdá, že významných period je hned několik, dle testu (opět na hladině 0,01) se ukazuje, že významné jsou pouze indexy $j = 10, 11$ a 3 , jež odpovídají délkám period 18,1 měsíců neboli 18 měsíců a 3 dny, dále 16 měsíců a 13 dnů a nakonec 60 měsíců a 10 dnů.

Do modelu zahrneme jen tyto 3 periody a získáme vyrovnané hodnoty \hat{y}_t , $t = 1, \dots, 181$. Několik hodnot jsem pro představu uvedla do tabulky 14, všechny jsou zobrazeny v grafu č. 23.

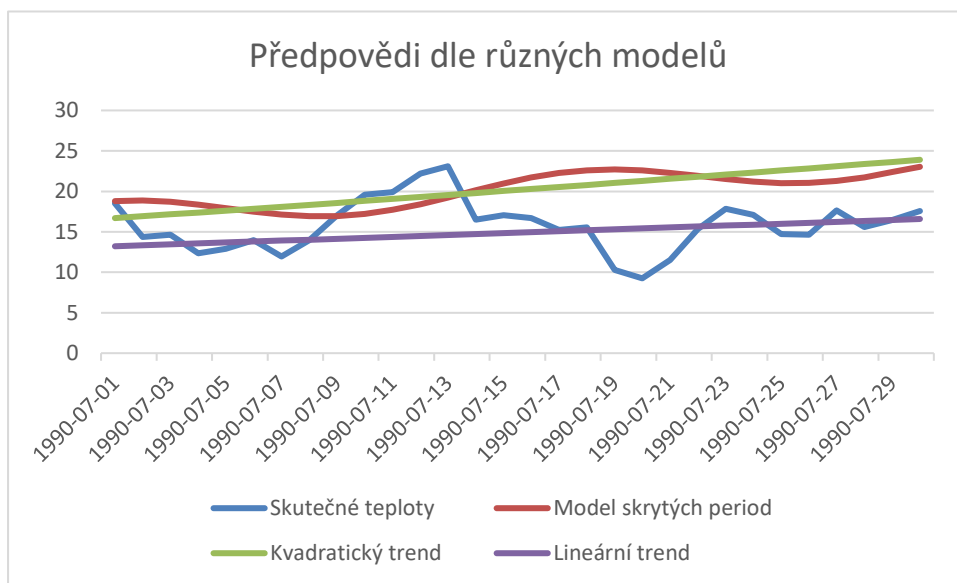
Tabulka 14 – Některé hodnoty MSP pro denní teploty

Datum	očistěné $y_t^{(o)}$	MSP
1990-01-01	5,72	3,08
1990-01-02	-0,54	3,85
1990-01-03	1,46	4,45
1990-01-04	4,45	4,82
1990-01-05	4,44	4,88
1990-01-06	4,68	4,61
1990-01-07	5,92	4,02
1990-01-08	6,91	3,15
1990-01-09	7,90	2,12
1990-01-10	0,63	1,03
1990-01-11	-2,13	0,02
1990-01-12	1,10	-0,78
1990-01-13	7,33	-1,27
1990-01-14	5,57	-1,37
1990-01-15	3,05	-1,09



Graf č. 23 – Vyrovnané hodnoty MSP

Již z grafického znázornění je patrné, že model skrytých period nekopíruje danou časovou řadu moc dobře. Důvodem bude nejspíš velký podíl náhodné složky v řadě. Kvalitu jsme vyjádřili i pomocí indexu determinace R^2 dle (5.2) a ten nám to jen potvrdil. Vypočítali jsme, že $R^2 = 0,3252$. Hodnota je příliš nízká a potvrzuje nevhodnost modelu pro danou časovou řadu. Co se týká predikcí dle tohoto modelu, opět se jako lepší varianta se jeví ta, že bychom predikce zkonstruovali pouze pro lineární, případně konstantní trend tak jako v předchozí kapitole u nezaměstnanosti. Vyzkoušela jsem tedy vypočítat predikce na následujících 30 dní ($h=30$) nejprve dle uvedeného MSP včetně kvadratického trendu (periodická i trendová složka dohromady) a střední čtvercová chyba předpovědi vyšla $MSEP = 5,94$. Další možnost je konstruovat predikce dle výše uvedeného kvadratického trendu ve tvaru $\hat{T}_t = -4,4686 + 0,0021t + 0,0006t^2$, pak predikce (vzorec (6.1)) vycházejí o něco přesnější, s chybou $MSEP = 5,79$. Pro zajímavost jsem určila i lineární trend ve tvaru $\hat{T}_t = -7,950 + 0,116t$ (pomocí vzorců (2.5), (2.6)). A jako nejpresnější se ukázaly právě predikce dle tohoto lineárního trendu, chyba vyšla nejmenší, $MSEP = 3,36$. Všechny tři možnosti jsou zakresleny v posledním grafu č. 24.



Graf č. 24 – Průměrné denní teploty – predikce budoucích hodnot

ZÁVĚR

Práce je rozdělena na dvě hlavní části, a to na část teoretickou a praktickou. V teoretické části jsem představila hlavní pojmy z oblasti časových řad, dále jednotlivé modely vhodné pro zpracování periodických časových řad, které jsem následně aplikovala v praktické části na konkrétní data, včetně ukazatelů kvality modelů a predikcí časových řad.

Cílem práce však bylo nejen prohloubení teoretických znalostí, ale zároveň i praktické využití modelů na vybrané datové sady. V této bakalářské práci jsem využila troje data. Časovou řadu, která udává počet sňatků uzavřených na území České republiky, nezaměstnanost v ČR a průměrné denní teploty z oblasti Oldman River v Kanadě.

Na první časové řadě sňatků v kapitole 7 se podařilo představit všechny tři sezónní modely nejlépe. Jednalo se konkrétně o model skrytých period, model konstantní a proporcionální sezónnosti. Tato řada má hned na první pohled výraznou periodickou složku, proto tyto modely dokázaly dobře zachytit průběh časové řady, a také predikovat budoucí vývoj. V závěru kapitoly 7 jsem provedla i porovnání modelů a ukázalo se, že každý model má určitou přednost. Při rozhodování by záleželo na tom, co je pro nás prioritní. Zda rychlost a jednoduchost při výpočtech, nejpřesnější predikce, či nejnižší index determinace (ukazatel kvality modelu). Já osobně bych volila nejspíš model konstantní sezónnosti, jelikož nebyl tak náročný na výpočty jako model skrytých period. Dává nejpřesnější předpovědi ze všech tří, a dle indexu determinace je druhý nejkvalitnější.

Ukázalo se, že časová řada nezaměstnanosti a časová řada průměrných denních teplot mají něco společného. Obě mají výraznou náhodnou složku, kterou lze při vizuálním posouzení zaměnit za periodickou. Analýza však ukázala, že opravdu není nutné na takové typy časových řad využívat složité modely s desítkami parametrů. Postačí obyčejné proložení konstantním, lineárním, případně kvadratickým trendem a predikce jsou potom také přesnější. Konkrétně pro časovou řadu nezaměstnanosti jsem vytvořila model skrytých period, na základě kterého jsem následně konstruovala předpovědi. Prognóza se v tomto případě úplně odchýlila od skutečných hodnot. Mnohem přesnější vyšly předpovědi, při kterých jsem uvažovala časovou řadu danou pouze lineárním trendem. Podobně tomu bylo i v případě datové sady denních teplot. Kapitola 8 a 9 tedy ukazuje, že ne vždy je možné sezónní modely na časové řady použít, i když by se tak mohlo zdát.

V praktické části jsem si oživila i jiné části statistiky, dostala jsem se i k testování hypotéz. V kapitole 7 jsem testovala významnost poklesu lineárního trendu, kterým je dána časová řada sňatků. Ukázalo se, že pokles není významný. V další kapitole jsem testovala přítomnost sezónnosti, která se právě u řady nezaměstnanosti v ČR neprokázala. V poslední kapitole jsem opět prováděla test významnosti parametrů regresního modelu. Dospěla jsem k závěru, že testovaný parametr není statisticky významný.

Psaní práce pro mne bylo přínosné, prohloubila jsem si znalosti získané v rámci studia. Navíc jsem se výrazně zlepšila v práci s MS Excel, který jsem používala pro výpočty v praktické části. Pro práci s takto obsáhlými daty je použití softwaru nezbytné. Věřím, že spoustu těchto nově nabytých znalostí využiji nejen v dalším studiu, ale především v praxi.

LITERATURA

- [1] T. Cipra, *Analýza časových řad s aplikacemi v ekonomii*, Praha: SNTL, 1986.
- [2] J. Arlt, M. Arltová a E. Rubíková, *Analýza ekonomických časových řad s příklady*, Praha: VŠE v Praze, Fakulta informatiky a statistiky, 2002.
- [3] J. A. M. Arlt, *Ekonomické časové řady*, Praha: Grada Publishing, 2007.
- [4] I. Müller, *Časové řady (texty k přednáškám)*, Olomouc: Univerzita Palackého, 2016.
- [5] I. Křivý, *Analýza časových řad*, Ostrava: Ostravská univerzita, 2012.
- [6] Český statistický úřad (online) - <https://www.czso.cz/> [26.10.2016].
- [7] <http://www.dama.cz/zabava/svatba-v-maji-povera-a-skutecnost-14900> [24.2.2017].
- [8] J. Neubauer, *Testování hypotéz o parametrech regresního modelu - Statistika II.*, Brno: Katedra ekonometrie FVL UO Brno.
- [9] J. Chajdiak, E. Rublíková a M. Gudába, *Štatistické metódy v praxi*, Bratislava: Statis, 1997.
- [10] Česká národní banka (online) - <https://www.cnb.cz/cs/index.html> [26.10.2016].
- [11] <https://datamarket.com/> [13.2.2017].
- [12] J. Prokopová, *Statistická analýza dat Energetického regulačního úřadu* - diplomová práce, Olomouc: PřF UPOL, 2016.

SEZNAM GRAFŮ

Graf č. 1 - Vývoj dat od ledna 1993 do prosince 2013	29
Graf č. 2 - Lineární trend pro sňatkovost	29
Graf č. 3 - Periodogram pro sňatkovost.....	30
Graf č. 4 - Vyrovnané hodnoty – MSP pro sňatkovost	32
Graf č. 5 – Index determinace v závislosti na počtu period v modelu skrytých period	33
Graf č. 6 – MSP – bodová předpověď pro sňatkovost	35
Graf č. 7 – Řada očištěná od periodické složky	36
Graf č. 8 - Vyrovnané hodnoty – model konstantní sezónnosti	37
Graf č. 9 – Předpovědi – model konstantní sezónnosti	39
Graf č. 10 – Vyrovnané hodnoty – model proporcionální sezónnosti	40
Graf č. 11 – Bodové předpovědi – MPS – sňatkovost	41
Graf č. 12 – Nezaměstnanost v ČR	43
Graf č. 13 – Periodogram – nezaměstnanost.....	44
Graf č. 14 – Model skrytých period	45
Graf č. 15 - Index determinace v závislosti na počtu period v MSP	46
Graf č. 16 – Bodové předpovědi MSP – nezaměstnanost	46
Graf č. 17 – Bodové předpovědi pro nezaměstnanost při užití lineárního trendu nebo MSP48	
Graf č. 18 – Model konstantní sezónnosti.....	49
Graf č. 19 – Předpovědi dle MKS	49
Graf č. 20 – Průměrné denní teploty, Oldman River, Canada + kvadratický trend	51
Graf č. 21 – Řada očištěná od kvadratického trendu.....	52
Graf č. 22 – Periodogram – teploty	53
Graf č. 23 – Vyrovnané hodnoty MSP	54
Graf č. 24 – Průměrné denní teploty – predikce budoucích hodnot.....	55

PŘÍLOHA

Příloha č. 1 - Sňatkovost ČR

období	t	Y _t	období	t	Y _t	období	t	Y _t	období	t	Y _t
leden 1993	1	3 450	duben 1998	64	6 522	červenec 2003	127	5 117	říjen 2008	190	3 554
únor 1993	2	4 141	květen 1998	65	1 561	srpen 2003	128	6 888	listopad 2008	191	2 105
březen 1993	3	5 475	červen 1998	66	8 114	září 2003	129	6 740	prosinec 2008	192	1 316
duben 1993	4	8 493	červenec 1998	67	5 302	říjen 2003	130	4 107	leden 2009	193	995
květen 1993	5	1 574	srpen 1998	68	6 522	listopad 2003	131	2 535	únor 2009	194	1 319
červen 1993	6	8 681	září 1998	69	6 812	prosinec 2003	132	1 717	březen 2009	195	1 832
červenec 1993	7	5 995	říjen 1998	70	5 697	leden 2004	133	1 395	duben 2009	196	4 179
srpen 1993	8	5 339	listopad 1998	71	3 284	únor 2004	134	1 847	květen 2009	197	2 439
září 1993	9	6 940	prosinec 1998	72	2 523	březen 2004	135	2 441	červen 2009	198	8 609
říjen 1993	10	7 164	leden 1999	73	1 957	duben 2004	136	6 274	červenec 2009	199	6 151
listopad 1993	11	5 002	únor 1999	74	2 312	květen 2004	137	2 015	srpen 2009	200	8 255
prosinec 1993	12	3 779	březen 1999	75	3 361	červen 2004	138	8 292	září 2009	201	7 787
leden 1994	13	2 946	duben 1999	76	6 291	červenec 2004	139	6 737	říjen 2009	202	3 548
únor 1994	14	3 309	květen 1999	77	1 437	srpen 2004	140	6 590	listopad 2009	203	1 513
březen 1994	15	4 463	červen 1999	78	7 783	září 2004	141	7 106	prosinec 2009	204	1 235
duben 1994	16	8 121	červenec 1999	79	5 739	říjen 2004	142	4 431	leden 2010	205	931
květen 1994	17	1 388	srpen 1999	80	5 377	listopad 2004	143	2 374	únor 2010	206	1217
červen 1994	18	7 880	září 1999	81	8 287	prosinec 2004	144	1 945	březen 2010	207	1546
červenec 1994	19	5 581	říjen 1999	82	5 349	leden 2005	145	1 532	duben 2010	208	4053
srpen 1994	20	4 926	listopad 1999	83	3 213	únor 2005	146	1 851	květen 2010	209	2459
září 1994	21	6 372	prosinec 1999	84	2 417	březen 2005	147	2 341	červen 2010	210	7875
říjen 1994	22	6 119	leden 2000	85	1 890	duben 2005	148	5 806	červenec 2010	211	7742
listopad 1994	23	3 912	únor 2000	86	2 590	květen 2005	149	2 090	srpen 2010	212	7434
prosinec 1994	24	3 423	březen 2000	87	3 522	červen 2005	150	8 538	září 2010	213	6774
leden 1995	25	2 512	duben 2000	88	6 413	červenec 2005	151	6 961	říjen 2010	214	4043
únor 1995	26	2 857	květen 2000	89	1 677	srpen 2005	152	6 697	listopad 2010	215	1437
březen 1995	27	4 196	červen 2000	90	8 328	září 2005	153	7 774	prosinec 2010	216	1235
duben 1995	28	7 552	červenec 2000	91	6 392	říjen 2005	154	4 265	leden 2011	217	935
květen 1995	29	1 182	srpen 2000	92	5 533	listopad 2005	155	2 239	únor 2011	218	1 113
červen 1995	30	7 424	září 2000	93	8 480	prosinec 2005	156	1 735	březen 2011	219	1 498
červenec 1995	31	5 325	říjen 2000	94	4 631	leden 2006	157	1 246	duben 2011	220	3 782
srpen 1995	32	4 593	listopad 2000	95	3 194	únor 2006	158	1 465	květen 2011	221	2 324
září 1995	33	7 335	prosinec 2000	96	2 671	březen 2006	159	2 185	červen 2011	222	8 069
říjen 1995	34	4 900	leden 2001	97	1 701	duben 2006	160	5 227	červenec 2011	223	7 334
listopad 1995	35	4 039	únor 2001	98	2 030	květen 2006	161	2 018	srpen 2011	224	6 612
prosinec 1995	36	3 041	březen 2001	99	3 419	červen 2006	162	9 300	září 2011	225	6 831
leden 1996	37	1 991	duben 2001	100	5 936	červenec 2006	163	7 215	říjen 2011	226	3 196
únor 1996	38	2 448	květen 2001	101	1 493	srpen 2006	164	6 878	listopad 2011	227	2 351
březen 1996	39	4 241	červen 2001	102	9 272	září 2006	165	9 275	prosinec 2011	228	1 092
duben 1996	40	6 483	červenec 2001	103	5 185	říjen 2006	166	3 578	leden 2012	229	915
květen 1996	41	1 287	srpen 2001	104	5 782	listopad 2006	167	2 401	únor 2012	230	1 127
červen 1996	42	8 715	září 2001	105	7 920	prosinec 2006	168	2 072	březen 2012	231	1 689
červenec 1996	43	4 243	říjen 2001	106	4 518	leden 2007	169	1 217	duben 2012	232	3 479
srpen 1996	44	5 764	listopad 2001	107	2 778	únor 2007	170	1 630	květen 2012	233	2 286
září 1996	45	6 783	prosinec 2001	108	2 340	březen 2007	171	2 700	červen 2012	234	9 103
říjen 1996	46	5 327	leden 2002	109	1 303	duben 2007	172	4 998	červenec 2012	235	6 045
listopad 1996	47	3 827	únor 2002	110	3 018	květen 2007	173	2 038	srpen 2012	236	6 837
prosinec 1996	48	2 787	březen 2002	111	3 137	červen 2007	174	9 514	září 2012	237	7 804
leden 1997	49	2 043	duben 2002	112	5 901	červenec 2007	175	10 759	říjen 2012	238	2 706
únor 1997	50	2 668	květen 2002	113	1 579	srpen 2007	176	7 048	listopad 2012	239	1 507
březen 1997	51	4 300	červen 2002	114	9 330	září 2007	177	8 839	prosinec 2012	240	1 708
duben 1997	52	7 288	červenec 2002	115	5 346	říjen 2007	178	4 230	leden 2013	241	704
květen 1997	53	1 681	srpen 2002	116	7 143	listopad 2007	179	2 256	únor 2013	242	980
červen 1997	54	8 867	září 2002	117	6 864	prosinec 2007	180	1 928	březen 2013	243	1 661
červenec 1997	55	5 077	říjen 2002	118	4 707	leden 2008	181	1 054	duben 2013	244	3 264
srpen 1997	56	6 516	listopad 2002	119	2 518	únor 2008	182	1 765	květen 2013	245	2 464
září 1997	57	6 787	prosinec 2002	120	1 886	březen 2008	183	2 336	červen 2013	246	8 320
říjen 1997	58	5 752	leden 2003	121	1 391	duben 2008	184	4 830	červenec 2013	247	5 990
listopad 1997	59	4 185	únor 2003	122	1 764	květen 2008	185	2 550	srpen 2013	248	7 797
prosinec 1997	60	2 640	březen 2003	123	3 038	červen 2008	186	9 017	září 2013	249	6 786
leden 1998	61	2 231	duben 2003	124	5 276	červenec 2008	187	5 823	říjen 2013	250	2 761
únor 1998	62	2 731	květen 2003	125	1 855	srpen 2008	188	11 089	listopad 2013	251	1 632
březen 1998	63	3 728	červen 2003	126	8515	září 2008	189	7 018	prosinec 2013	252	1 140

Příloha č. 2 – sňatkovost – MSP – úhlové frekvence, odhady parametru a_j a b_j , hodnoty periodogramu

j	ω_j	\hat{a}_j	\hat{b}_j	I_j
1	0,02	260,87	-7,01	34051
2	0,05	264,40	-57,35	36597
3	0,07	199,09	186,08	37133
4	0,10	58,54	75,73	4581
5	0,12	143,03	29,81	10672
6	0,15	196,29	1,81	19266
7	0,17	68,47	22,71	2602
8	0,20	18,77	-40,02	977
9	0,22	109,06	46,23	7016
10	0,25	68,87	-12,91	2455
11	0,27	47,44	12,63	1205
12	0,30	61,98	-14,04	2019
13	0,32	0,96	29,78	444
14	0,35	1,76	9,14	43
15	0,37	15,11	7,04	139
16	0,40	10,76	17,67	214
17	0,42	-66,39	31,03	2685
18	0,45	-76,91	30,02	3408
19	0,47	29,35	152,51	12060
20	0,50	-302,33	381,29	118391
21	0,52	-1666,76	-2078,98	3550122
22	0,55	487,24	-220,06	142915
23	0,57	245,48	-67,48	32407
24	0,60	65,47	-76,99	5107
25	0,62	135,28	-90,76	13269
26	0,65	102,77	-53,67	6721
27	0,67	43,19	-74,57	3713
28	0,70	82,55	-60,26	5223
29	0,72	83,25	-35,54	4097
30	0,75	75,15	-33,15	3373
31	0,77	81,39	-62,01	5235
32	0,80	73,02	-56,76	4277
33	0,82	44,78	-48,63	2185
34	0,85	47,07	-83,95	4632
35	0,87	101,41	-67,55	7423
36	0,90	37,36	-68,17	3022
37	0,92	19,40	-77,35	3180
38	0,95	105,21	-124,25	13253
39	0,97	108,81	-60,24	7734
40	1,00	10,00	-92,92	4367
41	1,02	109,64	-282,86	46014
42	1,05	245,72	-495,83	153114
43	1,07	-116,27	301,40	52180
44	1,10	-55,72	108,86	7478
45	1,12	-4,45	7,39	37
46	1,15	1,55	24,34	298
47	1,17	12,82	0,30	82
48	1,20	51,93	8,01	1381
49	1,22	24,48	26,25	644
50	1,25	21,73	10,16	288
51	1,27	47,28	-26,68	1474
52	1,30	-16,77	27,32	514
53	1,32	8,97	-15,90	167
54	1,35	36,11	-52,25	2017
55	1,37	38,46	48,78	1929
56	1,40	-12,55	-14,62	186
57	1,42	17,71	-101,23	5280
58	1,45	102,22	-56,18	6802
59	1,47	-15,86	28,99	546
60	1,50	-11,49	-53,96	1522
61	1,52	-14,28	-114,60	6669
62	1,55	31,21	-86,98	4270
63	1,57	-177,94	-110,26	21910

j	ω_j	\hat{a}_j	\hat{b}_j	I_j
64	1,60	-85,55	97,32	8395
65	1,62	28,79	52,15	1774
66	1,65	12,14	16,85	216
67	1,67	36,99	56,36	2272
68	1,70	-4,83	67,47	2288
69	1,72	-14,92	69,82	2549
70	1,75	-40,03	4,22	810
71	1,77	-0,77	45,75	1047
72	1,80	-50,63	54,32	2757
73	1,82	-29,30	-25,36	751
74	1,85	49,20	17,75	1368
75	1,87	1,02	48,62	1183
76	1,89	-99,71	-38,38	5708
77	1,92	57,02	-67,77	3922
78	1,94	88,28	112,95	10275
79	1,97	-6,77	105,11	5547
80	1,99	21,33	27,69	611
81	2,02	43,00	68,10	3243
82	2,04	38,03	115,77	7425
83	2,07	-44,39	82,35	4376
84	2,09	671,16	891,06	622222
85	2,12	39,46	-175,98	16263
86	2,14	21,84	-100,48	5286
87	2,17	92,42	22,03	4513
88	2,19	-201,29	-163,48	33620
89	2,22	-76,77	-73,99	5685
90	2,24	0,74	-62,09	1928
91	2,27	8,98	-14,19	141
92	2,29	-12,00	-30,05	524
93	2,32	25,92	-97,47	5086
94	2,34	60,40	-11,89	1895
95	2,37	-5,42	-44,06	985
96	2,39	-75,13	-74,46	5594
97	2,42	113,88	-88,50	10401
98	2,44	57,37	81,22	4944
99	2,47	-47,37	-9,17	1164
100	2,49	5,55	-78,51	3097
101	2,52	87,55	-79,70	7008
102	2,54	-64,30	15,00	2180
103	2,57	10,99	-65,54	2208
104	2,59	139,31	-35,61	10338
105	2,62	-706,53	-1023,05	772905
106	2,64	-42,95	38,47	1662
107	2,67	-49,96	29,10	1672
108	2,69	-114,07	36,77	7182
109	2,72	158,25	-87,16	16320
110	2,74	43,62	-48,64	2134
111	2,77	-19,54	56,84	1806
112	2,79	-69,18	-17,65	2548
113	2,82	1,22	-31,54	498
114	2,84	-23,66	41,20	1129
115	2,87	-45,80	55,80	2605
116	2,89	-17,28	-69,80	2586
117	2,92	79,82	-33,17	3736
118	2,94	2,95	8,16	38
119	2,97	-92,29	-5,03	4271
120	2,99	12,08	-97,01	4778
121	3,02	-32,70	57,40	2182
122	3,04	-31,64	135,53	9685
123	3,07	-75,76	-94,72	7356
124	3,09	-47,95	11,46	1215
125	3,12	-134,69	-29,82	9516
126	3,14159	0,00	1128,64	636917

Příloha č. 3 – Vyrovnané hodnoty – sňatkovost

t	MSP	MKS	MPS
1	1505	2220	1880
2	2844	2689	2411
3	2492	3590	3444
4	6489	6308	6507
5	1395	2461	2063
6	8727	9135	9538
7	5052	6779	6865
8	6906	7234	7352
9	6695	7981	8251
10	4850	5089	5119
11	2183	3362	3179
12	3187	2710	2434
13	1782	2162	1859
14	3097	2631	2384
15	2655	3532	3406
16	6519	6249	6435
17	1289	2403	2040
18	8516	9077	9433
19	4795	6720	6789
20	6674	7175	7270
21	6549	7923	8159
22	4824	5031	5062
23	2282	3303	3144
24	3378	2652	2407
25	2012	2103	1839
26	3303	2572	2358
27	2782	3473	3368
28	6539	6190	6363
29	1201	2344	2017
30	8349	9018	9328
31	4597	6661	6713
32	6500	7117	7189
33	6443	7864	8068
34	4811	4972	5005
35	2358	3244	3108
36	3518	2593	2379
37	2174	2045	1818
38	3442	2514	2331
39	2864	3415	3331
40	6546	6132	6292
41	1138	2286	1995
42	8239	8959	9222
43	4475	6603	6637
44	6399	7058	7107
45	6387	7806	7977
46	4810	4914	4949
47	2405	3186	3073
48	3594	2535	2352
49	2253	1986	1797
50	3503	2455	2305
51	2892	3356	3293
52	6540	6073	6220
53	1107	2227	1972
54	8198	8901	9117
55	4440	6544	6561
56	6380	7000	7026
57	6387	7747	7885
58	4823	4855	4892
59	2420	3127	3038
60	3599	2476	2325
61	2244	1928	1777
62	3479	2397	2278
63	2864	3298	3255

t	MSP	MKS	MPS
64	6521	6015	6148
65	1109	2169	1949
66	8229	8842	9012
67	4495	6486	6485
68	6445	6941	6945
69	6442	7689	7794
70	4848	4797	4835
71	2400	3069	3003
72	3534	2418	2298
73	2146	1869	1756
74	3374	2338	2252
75	2783	3239	3217
76	6491	5956	6076
77	1146	2110	1926
78	8328	8784	8906
79	4634	6427	6409
80	6588	6883	6863
81	6548	7630	7702
82	4882	4738	4778
83	2348	3010	2967
84	3403	2359	2271
85	1968	1811	1735
86	3196	2279	2225
87	2656	3181	3179
88	6453	5898	6005
89	1213	2052	1904
90	8487	8725	8801
91	4846	6369	6333
92	6796	6824	6782
93	6694	7571	7611
94	4923	4680	4721
95	2268	2952	2932
96	3219	2301	2244
97	1726	1752	1715
98	2962	2221	2199
99	2493	3122	3141
100	6410	5839	5933
101	1304	1993	1881
102	8692	8667	8695
103	5112	6310	6257
104	7051	6766	6701
105	6869	7513	7520
106	4968	4621	4665
107	2167	2893	2897
108	2997	2242	2217
109	1442	1694	1694
110	2692	2162	2172
111	2310	3064	3103
112	6365	5781	5861
113	1412	1935	1858
114	8925	8608	8590
115	5408	6252	6181
116	7331	6707	6619
117	7056	7454	7428
118	5011	4562	4608
119	2054	2835	2861
120	2758	2184	2190
121	1141	1635	1673
122	2410	2104	2146
123	2123	3005	3065
124	6324	5722	5789
125	1527	1876	1835
126	9165	8550	8485

t	MSP	MKS	MPS
127	5708	6193	6105
128	7609	6649	6538
129	7239	7396	7337
130	5050	4504	4551
131	1940	2776	2826
132	2523	2125	2163
133	849	1577	1653
134	2141	2045	2119
135	1949	2947	3027
136	6289	5664	5718
137	1639	1818	1813
138	9391	8491	8379
139	5985	6135	6030
140	7863	6590	6457
141	7401	7337	7245
142	5080	4445	4494
143	1833	2718	2791
144	2312	2067	2136
145	593	1518	1632
146	1909	1987	2093
147	1802	2888	2989
148	6263	5605	5646
149	1737	1759	1790
150	9582	8433	8274
151	6214	6076	5954
152	8068	6532	6375
153	7529	7279	7154
154	5100	4387	4438
155	1745	2659	2756
156	2145	2008	2109
157	395	1460	1611
158	1734	1928	2066
159	1696	2830	2951
160	6250	5547	5574
161	1813	1700	1767
162	9722	8374	8169
163	6376	6018	5878
164	8207	6473	6294
165	7611	7220	7063
166	5107	4328	4381
167	1683	2601	2720
168	2035	1950	2082
169	273	1401	1591
170	1633	1870	2040
171	1641	2771	2913
172	6249	5488	5502
173	1861	1642	1744
174	9798	8316	8063
175	6455	5959	5802
176	8268	6414	6212
177	7639	7162	6971
178	5101	4270	4324
179	1651	2542	2685
180	1994	1891	2055
181	238	1343	1570
182	1615	1811	2013
183	1641	2713	2875
184	6262	5430	5431
185	1876	1583	1721
186	9803	8257	7958
187	6446	5901	5726
188	8244	6356	6131
189	7611	7103	6880

t	MSP	MKS	MPS
190	5082	4211	4267
191	1654	2484	2650
192	2025	1833	2028
193	292	1284	1549
194	1679	1753	1986
195	1696	2654	2837
196	6287	5371	5359
197	1856	1525	1699
198	9738	8199	7852
199	6348	5842	5650
200	8139	6297	6050
201	7530	7045	6788
202	5052	4153	4211
203	1690	2425	2614
204	2124	1774	2001
205	432	1226	1529
206	1822	1694	1960
207	1801	2596	2799
208	6321	5313	5287
209	1804	1466	1676
210	9607	8140	7747
211	6170	5784	5574
212	7962	6239	5968
213	7402	6986	6697
214	5014	4094	4154
215	1757	2367	2579
216	2283	1716	1974
217	644	1167	1508
218	2031	1636	1933
219	1948	2537	2761
220	6362	5254	5215
221	1723	1408	1653
222	9423	8082	7642
223	5928	5725	5498
224	7727	6180	5887
225	7240	6928	6606
226	4971	4036	4097
227	1849	2308	2544
228	2488	1657	1947
229	910	1108	1487
230	2286	1577	1907
231	2122	2478	2723
232	6406	5196	5144
233	1622	1349	1630
234	9201	8023	7536
235	5644	5667	5422
236	7457	6122	5806
237	7057	6869	6514
238	4926	3977	4040
239	1957	2250	2509
240	2721	1599	1920
241	1206	1050	1467
242	2565	1519	1880
243	2309	2420	2685
244	6450	5137	5072
245	1510	1291	1608
246	8962	7965	7431
247	5343	5608	5346
248	7175	6063	5724
249	6870	6811	6423
250	4885	3919	3984
251	2072	2191	2473
252	2961	1540	1893

Příloha č. 4 – Obecná míra nezaměstnanosti 15-64 letých + vyrovnané hodnoty MSP

Období	t	Skutečné hodnoty	MSP	Období	t	Skutečné hodnoty	MSP	Období	t	Skutečné hodnoty	MSP	Období	t	Skutečné hodnoty	MSP
31.1.2000	1	9,2	7,68	31.10.2003	46	8,1	8,24	31.7.2007	91	5,3	5,21	30.4.2011	136	7	6,98
29.2.2000	2	9,2	8,05	30.11.2003	47	8,2	8,39	31.8.2007	92	5,1	5,15	31.5.2011	137	6,9	6,94
31.3.2000	3	9,3	8,38	31.12.2003	48	8,4	8,52	30.9.2007	93	5,1	5,09	30.6.2011	138	6,9	6,89
30.4.2000	4	9,1	8,64	31.1.2004	49	8,4	8,63	31.10.2007	94	5,1	5,03	31.7.2011	139	6,7	6,82
31.5.2000	5	9	8,81	29.2.2004	50	8,5	8,71	30.11.2007	95	4,8	4,95	31.8.2011	140	6,6	6,75
30.6.2000	6	8,8	8,88	31.3.2004	51	8,4	8,74	31.12.2007	96	4,9	4,84	30.9.2011	141	6,7	6,69
31.7.2000	7	8,7	8,88	30.4.2004	52	8,5	8,72	31.1.2008	97	4,7	4,73	31.10.2011	142	6,6	6,64
31.8.2000	8	8,6	8,80	31.5.2004	53	8,5	8,67	29.2.2008	98	4,6	4,59	30.11.2011	143	6,5	6,61
30.9.2000	9	8,6	8,67	30.6.2004	54	8,3	8,58	31.3.2008	99	4,4	4,46	31.12.2011	144	6,6	6,62
31.10.2000	10	8,5	8,52	31.7.2004	55	8,2	8,47	30.4.2008	100	4,4	4,34	31.1.2012	145	6,7	6,65
30.11.2000	11	8,5	8,38	31.8.2004	56	8,3	8,35	31.5.2008	101	4,3	4,24	29.2.2012	146	6,9	6,73
31.12.2000	12	8,3	8,26	30.9.2004	57	8,4	8,23	30.6.2008	102	4,4	4,19	31.3.2012	147	7	6,82
31.1.2001	13	8,3	8,18	31.10.2004	58	8,5	8,12	31.7.2008	103	4,3	4,18	30.4.2012	148	7	6,93
28.2.2001	14	8,2	8,15	30.11.2004	59	8,3	8,03	31.8.2008	104	4,3	4,23	31.5.2012	149	6,9	7,05
31.3.2001	15	8,2	8,17	31.12.2004	60	8,1	7,96	30.9.2008	105	4,3	4,35	30.6.2012	150	7	7,15
30.4.2001	16	8,2	8,21	31.1.2005	61	8,1	7,91	31.10.2008	106	4,3	4,53	31.7.2012	151	7	7,23
31.5.2001	17	8,2	8,28	28.2.2005	62	8,1	7,89	30.11.2008	107	4,5	4,76	31.8.2012	152	7,1	7,28
30.6.2001	18	8,1	8,34	31.3.2005	63	8	7,88	31.12.2008	108	4,7	5,03	30.9.2012	153	7,1	7,29
31.7.2001	19	8,2	8,39	30.4.2005	64	8,1	7,88	31.1.2009	109	5,4	5,34	31.10.2012	154	7,3	7,26
31.8.2001	20	8,3	8,40	31.5.2005	65	8	7,88	28.2.2009	110	5,7	5,67	30.11.2012	155	7,3	7,20
30.9.2001	21	8,2	8,35	30.6.2005	66	8	7,88	31.3.2009	111	6	6,01	31.12.2012	156	7,2	7,12
31.10.2001	22	8	8,26	31.7.2005	67	7,8	7,88	30.4.2009	112	6,4	6,33	31.1.2013	157	7,1	7,04
30.11.2001	23	8	8,13	31.8.2005	68	7,9	7,87	31.5.2009	113	6,6	6,63	28.2.2013	158	7,2	6,96
31.12.2001	24	7,8	7,95	30.9.2005	69	7,9	7,86	30.6.2009	114	6,8	6,90	31.3.2013	159	7,3	6,90
31.1.2002	25	7,6	7,76	31.10.2005	70	7,9	7,85	31.7.2009	115	7,2	7,14	30.4.2013	160	7,2	6,87
28.2.2002	26	7,5	7,56	30.11.2005	71	7,9	7,83	31.8.2009	116	7,4	7,33	31.5.2013	161	7,1	6,87
31.3.2002	27	7,3	7,38	31.12.2005	72	7,8	7,81	30.9.2009	117	7,5	7,48	30.6.2013	162	6,8	6,90
30.4.2002	28	7,2	7,23	31.1.2006	73	7,9	7,79	31.10.2009	118	7,4	7,58	31.7.2013	163	7	6,95
31.5.2002	29	7,2	7,11	28.2.2006	74	7,6	7,77	30.11.2009	119	7,5	7,64	31.8.2013	164	7	7,00
30.6.2002	30	7,2	7,05	31.3.2006	75	7,6	7,73	31.12.2009	120	7,5	7,66	30.9.2013	165	7	7,04
31.7.2002	31	7,3	7,03	30.4.2006	76	7,3	7,68	31.1.2010	121	7,8	7,64	31.10.2013	166	6,9	7,06
31.8.2002	32	7,3	7,04	31.5.2006	77	7,3	7,60	28.2.2010	122	7,8	7,60	30.11.2013	167	6,9	7,04
30.9.2002	33	7,3	7,09	30.6.2006	78	7,2	7,50	31.3.2010	123	7,8	7,53	31.12.2013	168	6,8	6,97
31.10.2002	34	7,3	7,15	31.7.2006	79	7,2	7,36	30.4.2010	124	7,5	7,45	31.1.2014	169	6,7	6,85
30.11.2002	35	7,4	7,22	31.8.2006	80	7	7,19	31.5.2010	125	7,4	7,36	28.2.2014	170	6,6	6,70
31.12.2002	36	7,3	7,29	30.9.2006	81	7	6,99	30.6.2010	126	7,3	7,28	31.3.2014	171	6,6	6,53
31.1.2003	37	7,3	7,35	31.10.2006	82	6,8	6,76	31.7.2010	127	7,3	7,20	30.4.2014	172	6,5	6,35
28.2.2003	38	7,3	7,41	30.11.2006	83	6,7	6,52	31.8.2010	128	7,2	7,13	31.5.2014	173	6,3	6,21
31.3.2003	39	7,5	7,47	31.12.2006	84	6,4	6,28	30.9.2010	129	7,1	7,08	30.6.2014	174	6,2	6,12
30.4.2003	40	7,6	7,53	31.1.2007	85	6,1	6,04	31.10.2010	130	7,1	7,05	31.7.2014	175	6	6,10
31.5.2003	41	7,8	7,61	28.2.2007	86	5,9	5,83	30.11.2010	131	7	7,03	31.8.2014	176	6,1	6,17
30.6.2003	42	7,8	7,70	31.3.2007	87	5,7	5,64	31.12.2010	132	7	7,03	30.9.2014	177	5,9	6,34
31.7.2003	43	8	7,81	30.4.2007	88	5,6	5,49	31.1.2011	133	7	7,02	31.10.2014	178	5,8	6,59
31.8.2003	44	8	7,94	31.5.2007	89	5,5	5,37	28.2.2011	134	7	7,02	30.11.2014	179	6	6,92
30.9.2003	45	8,1	8,08	30.6.2007	90	5,4	5,28	31.3.2011	135	6,9	7,01	31.12.2014	180	5,9	7,29

Příloha č. 5 – Nezaměstnanost – MSP – úhlové frekvence, odhady parametru a_j a b_j , hodnoty periodogramu

j	ω_j	\hat{a}_j	\hat{b}_j	I_j	j	ω_j	\hat{a}_j	\hat{b}_j	I_j
1	0,035	0,728	0,651	0,47676	46	1,606	0,009	-0,010	0,00009
2	0,070	-0,216	-0,564	0,18238	47	1,641	0,016	-0,016	0,00025
3	0,105	0,571	0,572	0,32657	48	1,676	0,016	-0,017	0,00027
4	0,140	-0,036	-0,192	0,01912	49	1,710	0,010	-0,029	0,00047
5	0,175	0,526	-0,234	0,16561	50	1,745	0,017	-0,013	0,00023
6	0,209	0,099	0,060	0,00669	51	1,780	0,005	-0,018	0,00017
7	0,244	0,065	-0,048	0,00324	52	1,815	0,005	-0,019	0,00020
8	0,279	0,232	0,043	0,02793	53	1,850	0,015	-0,017	0,00026
9	0,314	0,054	0,025	0,00175	54	1,885	0,009	-0,013	0,00013
10	0,349	0,169	-0,067	0,01653	55	1,920	0,015	-0,014	0,00021
11	0,384	0,145	0,019	0,01061	56	1,955	0,014	-0,006	0,00012
12	0,419	0,121	-0,045	0,00831	57	1,990	0,005	-0,023	0,00028
13	0,454	0,101	-0,009	0,00518	58	2,025	0,014	-0,016	0,00023
14	0,489	0,061	-0,053	0,00328	59	2,059	0,002	-0,024	0,00029
15	0,524	0,078	-0,011	0,00307	60	2,094	0,009	-0,024	0,00032
16	0,559	0,081	0,023	0,00357	61	2,129	-0,001	-0,012	0,00008
17	0,593	0,035	-0,006	0,00062	62	2,164	0,005	-0,015	0,00013
18	0,628	0,095	-0,007	0,00449	63	2,199	0,010	-0,023	0,00032
19	0,663	0,054	-0,035	0,00209	64	2,234	0,010	-0,013	0,00014
20	0,698	0,068	-0,039	0,00304	65	2,269	0,008	-0,014	0,00012
21	0,733	0,050	-0,013	0,00136	66	2,304	0,006	-0,021	0,00024
22	0,768	0,047	0,001	0,00110	67	2,339	-0,006	-0,018	0,00019
23	0,803	0,058	-0,031	0,00220	68	2,374	0,010	-0,024	0,00035
24	0,838	0,029	-0,035	0,00104	69	2,409	0,006	-0,012	0,00009
25	0,873	0,058	-0,008	0,00170	70	2,443	0,004	-0,029	0,00044
26	0,908	0,033	-0,013	0,00063	71	2,478	0,005	-0,015	0,00012
27	0,942	0,038	-0,032	0,00120	72	2,513	0,013	-0,027	0,00045
28	0,977	0,029	0,009	0,00047	73	2,548	0,009	-0,025	0,00037
29	1,012	0,011	-0,040	0,00084	74	2,583	0,005	-0,018	0,00018
30	1,047	0,030	-0,039	0,00122	75	2,618	0,002	-0,023	0,00027
31	1,082	0,033	-0,014	0,00062	76	2,653	0,008	-0,018	0,00020
32	1,117	0,032	-0,035	0,00111	77	2,688	0,004	-0,018	0,00018
33	1,152	0,035	-0,026	0,00095	78	2,723	0,003	-0,029	0,00043
34	1,187	0,017	0,003	0,00015	79	2,758	4,66E-04	-0,008	0,00003
35	1,222	0,028	-0,017	0,00052	80	2,793	0,009	-0,033	0,00059
36	1,257	0,024	-0,018	0,00045	81	2,827	0,008	-0,027	0,00040
37	1,292	0,027	0,000	0,00035	82	2,862	0,009	-0,013	0,00012
38	1,326	0,020	-0,021	0,00043	83	2,897	0,009	-0,016	0,00016
39	1,361	0,019	-0,032	0,00070	84	2,932	0,002	-0,010	0,00005
40	1,396	0,019	-0,028	0,00057	85	2,967	0,021	-0,023	0,00048
41	1,431	0,030	-0,010	0,00051	86	3,002	-0,017	-0,012	0,00021
42	1,466	0,034	-0,017	0,00073	87	3,037	0,005	-0,024	0,00029
43	1,501	0,013	-0,012	0,00015	88	3,072	0,008	-0,024	0,00031
44	1,536	0,005	-0,003	0,00002	89	3,107	0,004	-0,022	0,00025
45	1,571	0,020	-0,016	0,00032	90	3,142	-3,74E-15	-0,031	0,00048

Příloha č. 6 – Průměrné denní teploty, Oldman River, Canada

t	Datum	Yt	Yt ^(o)	t	Datum	Yt	Yt ^(o)	t	Datum	Yt	Yt ^(o)	t	Datum	Yt	Yt ^(o)
1	1990-01-01	1,25	5,7	46	1990-02-15	-10,05	-7,0	91	1990-04-01	8,15	7,2	136	1990-05-16	6,5	-0,9
2	1990-01-02	-5	-0,5	47	1990-02-16	-23,9	-20,9	92	1990-04-02	11,35	10,3	137	1990-05-17	7,75	0,2
3	1990-01-03	-3	1,5	48	1990-02-17	-22	-19,1	93	1990-04-03	6,05	4,9	138	1990-05-18	6,25	-1,5
4	1990-01-04	0	4,5	49	1990-02-18	-14,75	-11,9	94	1990-04-04	0,5	-0,8	139	1990-05-19	6	-1,9
5	1990-01-05	0	4,4	50	1990-02-19	-3,25	-0,5	95	1990-04-05	0,45	-0,9	140	1990-05-20	5	-3,1
6	1990-01-06	0,25	4,7	51	1990-02-20	-1,05	1,7	96	1990-04-06	4,5	3,0	141	1990-05-21	8,5	0,2
7	1990-01-07	1,5	5,9	52	1990-02-21	2,25	4,9	97	1990-04-07	6,7	5,1	142	1990-05-22	10	1,5
8	1990-01-08	2,5	6,9	53	1990-02-22	6	8,6	98	1990-04-08	-2,35	-4,1	143	1990-05-23	15	6,3
9	1990-01-09	3,5	7,9	54	1990-02-23	5,7	8,2	99	1990-04-09	-7,55	-9,4	144	1990-05-24	10,5	1,7
10	1990-01-10	-3,75	0,6	55	1990-02-24	1,6	4,1	100	1990-04-10	-10,95	-13,0	145	1990-05-25	6,25	-2,8
11	1990-01-11	-6,5	-2,1	56	1990-02-25	4,5	6,9	101	1990-04-11	-6,95	-9,1	146	1990-05-26	6,5	-2,7
12	1990-01-12	-3,25	1,1	57	1990-02-26	-7,2	-4,9	102	1990-04-12	1,2	-1,1	147	1990-05-27	10,5	1,1
13	1990-01-13	3	7,3	58	1990-02-27	-1,05	1,2	103	1990-04-13	4,05	1,6	148	1990-05-28	8	-1,6
14	1990-01-14	1,25	5,6	59	1990-02-28	4,45	6,6	104	1990-04-14	5,9	3,4	149	1990-05-29	13,25	3,5
15	1990-01-15	-1,25	3,0	60	1990-03-01	5,65	7,7	105	1990-04-15	5,9	3,2	150	1990-05-30	12,5	2,5
16	1990-01-16	-4,25	0,0	61	1990-03-02	3,15	5,2	106	1990-04-16	0,8	-2,0	151	1990-05-31	10,75	0,6
17	1990-01-17	-9,5	-5,2	62	1990-03-03	3,4	5,3	107	1990-04-17	6,3	3,4	152	1990-06-01	7,45	-2,9
18	1990-01-18	-2,75	1,5	63	1990-03-04	4,05	5,9	108	1990-04-18	12,15	9,1	153	1990-06-02	7,05	-3,5
19	1990-01-19	-1,75	2,5	64	1990-03-05	2,75	4,5	109	1990-04-19	13,65	10,4	154	1990-06-03	8,05	-2,7
20	1990-01-20	-2,75	1,4	65	1990-03-06	2,2	3,9	110	1990-04-20	12,25	8,9	155	1990-06-04	7,5	-3,4
21	1990-01-21	0,5	4,6	66	1990-03-07	6,4	8,0	111	1990-04-21	11	7,5	156	1990-06-05	9,65	-1,5
22	1990-01-22	2,75	6,9	67	1990-03-08	2,8	4,3	112	1990-04-22	11,15	7,5	157	1990-06-06	11,45	0,1
23	1990-01-23	-1,25	2,8	68	1990-03-09	-0,9	0,5	113	1990-04-23	10,05	6,3	158	1990-06-07	10,75	-0,8
24	1990-01-24	-3	1,1	69	1990-03-10	1,75	3,1	114	1990-04-24	8,75	4,8	159	1990-06-08	12,25	0,5
25	1990-01-25	1,25	5,3	70	1990-03-11	-1,5	-0,3	115	1990-04-25	5,45	1,4	160	1990-06-09	14,15	2,2
26	1990-01-26	-6	-2,0	71	1990-03-12	-3,25	-2,1	116	1990-04-26	2,7	-1,5	161	1990-06-10	12	-0,1
27	1990-01-27	-8	-4,0	72	1990-03-13	-0,05	1,0	117	1990-04-27	1,65	-2,7	162	1990-06-11	7,25	-5,1
28	1990-01-28	-4	-0,1	73	1990-03-14	1,35	2,3	118	1990-04-28	-5	-9,5	163	1990-06-12	9,65	-2,9
29	1990-01-29	-19,75	-15,9	74	1990-03-15	3,75	4,6	119	1990-04-29	-7,25	-11,9	164	1990-06-13	6,65	-6,1
30	1990-01-30	-25,75	-21,9	75	1990-03-16	4,45	5,2	120	1990-04-30	-2,25	-7,1	165	1990-06-14	9,2	-3,8
31	1990-01-31	-28,25	-24,5	76	1990-03-17	-1,4	-0,7	121	1990-05-01	8	3,0	166	1990-06-15	9,45	-3,7
32	1990-02-01	-16,5	-12,7	77	1990-03-18	1,4	2,0	122	1990-05-02	8	2,9	167	1990-06-16	11,5	-1,9
33	1990-02-02	-7,4	-3,7	78	1990-03-19	6,85	7,3	123	1990-05-03	9,5	4,2	168	1990-06-17	12,5	-1,1
34	1990-02-03	-0,7	3,0	79	1990-03-20	5,35	5,7	124	1990-05-04	9,75	4,3	169	1990-06-18	14,55	0,7
35	1990-02-04	-6,4	-2,8	80	1990-03-21	0,75	1,0	125	1990-05-05	11,25	5,7	170	1990-06-19	14,5	0,5
36	1990-02-05	-3,95	-0,4	81	1990-03-22	-13,65	-13,5	126	1990-05-06	13,5	7,7	171	1990-06-20	13,1	-1,1
37	1990-02-06	-6	-2,5	82	1990-03-23	-15,1	-15,0	127	1990-05-07	5,5	-0,4	172	1990-06-21	13,9	-0,6
38	1990-02-07	-5,4	-1,9	83	1990-03-24	-9,4	-9,4	128	1990-05-08	0,5	-5,6	173	1990-06-22	16	1,3
39	1990-02-08	-6,25	-2,8	84	1990-03-25	-3,5	-3,6	129	1990-05-09	4,75	-1,5	174	1990-06-23	17,1	2,2
40	1990-02-09	-2,8	0,6	85	1990-03-26	3,25	3,0	130	1990-05-10	5,5	-0,9	175	1990-06-24	17,25	2,1
41	1990-02-10	2,6	5,9	86	1990-03-27	1,2	0,8	131	1990-05-11	6,25	-0,3	176	1990-06-25	17,45	2,1
42	1990-02-11	3,45	6,7	87	1990-03-28	3,95	3,5	132	1990-05-12	1,75	-5,0	177	1990-06-26	18,7	3,1
43	1990-02-12	-11,05	-7,8	88	1990-03-29	5,05	4,5	133	1990-05-13	4,5	-2,4	178	1990-06-27	18,7	2,9
44	1990-02-13	-24,35	-21,2	89	1990-03-30	10,25	9,6	134	1990-05-14	4,5	-2,6	179	1990-06-28	15,6	-0,4
45	1990-02-14	-21,35	-18,2	90	1990-03-31	2,65	1,8	135	1990-05-15	6,25	-1,0	180	1990-06-29	15,85	-0,4
												181	1990-06-30	19,05	2,6