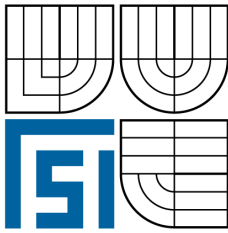


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA STROJNÍHO INŽENÝRSTVÍ

ÚSTAV MATEMATIKY

FACULTY OF MECHANICAL ENGINEERING

INSTITUTE OF MATHEMATICS

ANALÝZA A PŘEDPOVĚĎ ČASOVÝCH ŘAD POMOCÍ STATISTICKÝCH METOD SE ZAMĚŘENÍM NA METODU BOX-JENKINS

TIME SERIES ANALYSIS AND PREDICTION BY MEANS OF STATISTICAL METHODS – BOX
JENKINS

DIPLOMOVÁ PRÁCE

DIPLOMA THESIS

AUTOR PRÁCE

AUTHOR

RADOMÍR ZATLOUKAL

VEDOUCÍ PRÁCE

SUPERVISOR

RNDr. LIBOR ŽÁK, Ph.D.

Abstrakt

V prvním kroku budeme analyzovat dvě reálné časové řady, jedna z oblasti energetické a druhá z oblasti ekonomické. U energetické časové řady se bude jednat o spotřebu elektrické energie v USA a v ekonomickém případě půjde o vývoj indexu PX50. Pokusíme se prokázat zda platí či nikoliv hypotéza, že za pomoci testovacích funkcí, by bylo možné, na základě určitých kritérií, stanovit zastoupení náhodné složky v těchto dvou reálných řadách.

Summary

Two real time series, one discussing the area of energy, other discussing the area of economy. By the energetic area we will be dealing with the electric power consumption in the USA, by the economic area we will be dealing with the progress of index PX50. We will try to approve the validity of hypothesis that with some test functions we will be able to set down the accidental unit distribution in these two time series.

Klíčová slova

Časové řady, Box-Jenkins, ARIMA, ARMA, SARIMA, dekompozice

Keywords

Time series, Box-Jenkins, ARIMA, ARMA, SARIMA, decomposition

ZATLOUKAL, R. *Analýza a předpověď časových řad pomocí statistických metod se zaměřením na metodu Box-Jenkins*. Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2008. 51 s. Vedoucí diplomové práce RNDr. Libor Žák, Ph.D.

Prohlašuji, že jsem diplomovou práci *Analýza a předpověď časových řad pomocí statistických metod se zaměřením na metodu Box-Jenkins* vypracoval samostatně pod vedením RNDr. Libora Žáka, Ph.D.; s použitím materiálů uvedených v seznamu literatury.

Radomír Zatloukal

Děkuji svému školiteli RNDr. Liboru Žakovi, Ph.D. za vedení mé diplomové práce a velice vstřícnému postoji obohaceném zajímavými názory.

Radomír Zatloukal

Obsah

1	Úvod	3
2	Časové řady	4
2.1	Dělení	5
2.2	Problémy analýzy	5
3	Box-Jenkinsova metodologie	8
3.1	Stacionarita	8
3.2	Autokorelační a parciální autokorelační funkce	10
3.3	Základní procesy	11
3.4	Integrované a sezónní procesy	14
3.5	Identifikace procesu	17
3.6	Verifikace procesu	18
3.7	Předpověď procesu	21
4	Dekompozice časových řad	23
4.1	Složky řady	23
4.2	Typy dekompozice	24
4.3	Analýza trendové složky	26
4.4	Analýza sezónní složky	27
5	Experimentální část	30
5.1	Energetická řada	30
5.2	Ekonomická řada	34
5.3	Testovací funkce	36
6	Závěr	39
	Literatura	41

Kapitola 1

Úvod

Tato práce je věnována analýze časových řad, se zaměřením na Box-Jenkinsonovu metodologii. Je rozdělena do několika kapitol. V druhé kapitole stručně popíšeme základní vlastnosti časových řad jakožto i samotné definice. Ve třetí kapitole si ukážeme základní procesy Box-Jenkinsonovi metodologie, a samozřejmě i celkový postup při odhadování modelu včetně jeho verifikace a výpočtu předpovědí. V další kapitole si jen vysvětlíme základní postup při odhadování časových řad pomocí jiné metody, a to tzv. Dekompoziční. Následně bude navazovat další kapitola, která se bude již zabývat naším experimentem, kde naším cílem bude pokus o vyslovení hypotézi, zda-li se dá určit procentuální zastoupení náhodnosti v reálných časových řadách za pomoci testovací funkce. A v poslední kapitole, kterou bude tvořit závěr, stručně zhrneme všechny naše naměřené výsledky a stanovíme, zda naše hypotéza má či nemá smysl. Samozřejmě nebude chybět ani příloha, kde budou vyobrazena naše měření a data. K práci budeme nejčastěji využívat statistický program STATISTICA a samozřejmě budeme využívat dalších nástrojů jako je Matlab nebo Excel.

Kapitola 2

Časové řady

Než se začnu věnovat samotné problematice časových řad, musíme si říci, co to vlastně časové řady jsou. Jako první uvedu ryze matematickou definici.

DEFINICE 2.1 *Nechť je dán pravděpodobnostní prostor (Ω, \mathcal{A}, P) , indexová množina $T \subseteq \mathbb{R}$ a reálná funkce*

$$X : \Omega \times T \rightarrow \mathbb{R} \tag{2.0.1}$$

definovaná pro $\forall \omega \in \Omega$ a $\forall t \in T$. Jestliže je $X(\omega, t)$ pro $\forall t \in T$ borelovsky měřitelná vzhledem k \mathcal{A} , tj. pro $\forall B \in \mathcal{B}$ a pro $\forall t \in T$

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega, t) \in B\} \in \mathcal{A}, \tag{2.0.2}$$

kde \mathcal{B} je σ -algebrovských podmnožin, pak tuto funkci nazýváme náhodným procesem. Náhodný proces $X(\omega, t)$ při pevném $\omega \in \Omega$ se nazývá realizace (trajektorie) procesu. Pravděpodobnostní míru

$$P_X(B) = P(X^{-1}(B))$$

nazýváme rozdělení pravděpodobností náhodného procesu $X(\omega, t)$.

POZNÁMKA: U náhodných veličin místo $\{X(\omega), \omega \in \Omega\}$ píšeme pouze $\{X\}$ a u náhodných procesů místo $\{X(\omega, t), \omega \in \Omega, t \in T\}$ zapisujeme $\{X_t, t \in T\}$.

DEFINICE 2.2 *Pokud $T = \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ nebo $T \subset \mathbb{Z}$ mluvíme o procesu s diskrétním časem nebo o časové řadě či náhodné posloupnosti. Pokud*

$$T = \langle t_1, t_2 \rangle, \text{ kde } -\infty \leq t_1 < t_2 \leq \infty,$$

říkáme, že $\{X_t \mid t \in T\}$ je náhodný proces se spojitým časem.

Nejčastěji se časovými řadami zabývá ve vyhodnocování ekonomických ukazatelů. Ve většině publikacích se používá této slovní definice:

Časovou řadou rozumíme posloupnost hodnot ukazatelů měřených v určitých časových intervalech. Tyto intervaly jsou zpravidla rovnoměrné (ekvidistantní), a proto je můžeme zapsat následujícím způsobem:

$$x_1, x_2, \dots, x_n \quad \text{neboli} \quad x_t, t = 1, \dots, n,$$

2.1. Dělení

Časové řady rozdělujeme na základě několika ukazatelů. Toto rozdělení není striktně dané, je pouze intuitivní, a to na základě jejich vlastností, které nám pomáhají při její další analýze. Základní rozdělení je na tyto dva typy:

Deterministické: neobsahují žádný prvek náhody, při znalosti toho, jak jsou generovány, je můžeme dokonale a bezchybně předpovídat, např. funkce sinus.

Stochastické: jako realizace náhodného procesu, obsahují prvek náhodnosti (skoro všechny reálné časové řady).

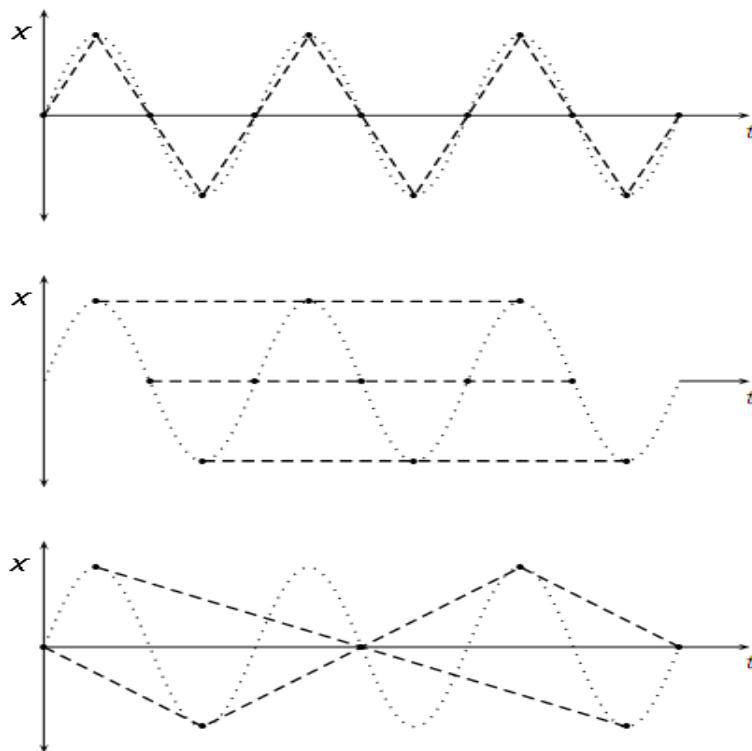
V praxi máme většinou řady stochastické, proto budeme v dalším textu časovými řadami rozumět stochastické časové řady. Dále je dělíme podle délky kroku na ekvidistantní (konstantní délka kroku) a neekvidistantní, s kterými pracovat není tak snadné, protože musíme provádět úpravy proměnlivého kroku. Je také dobré si uvědomit, že řady můžeme dělit na krátkodobé a dlouhodobé. U obou typů nás mohou zajímat rozdílné faktory. Zatímco u krátkodobých (např. měsíčních) nás často zajímají spíše sezónní vlivy, u dlouhodobých spíše existence dlouhodobých trendů. Důležité je třeba rozdělení ekonomických řad v naturálním a peněžním vyjádření. U peněžního musíme brát v úvahu i inflaci. Zjednodušeně záleží na dané řadě a konkrétním pozorování a na tom, co ve skutečnosti chceme sledovat.

2.2. Problémy analýzy

V životě je mnoho procesů nepravidelných a nespolehlivých, jejichž struktura se nám mění přímo před očima. Proto bychom neměli některé problémy při analýze časových řad opomíjet. Patří mezi ně tyto:

Problémy s volbou časových bodů pozorování. Při statistické analýze jsme v podstatě většinou omezeni na analýzu diskrétních časových řad. Ovšem většina ekonomických veličin se vyvíjí spojitě, je tedy nutné je nějakým způsobem diskretizovat.

Volba vhodného způsobu diskretizace a vhodné délky období je poměrně zásadní. Jednak ovlivňuje množství informace obsažené v datech, jednak špatná volba může snadno vést k zavádějícím výsledkům. Nicméně i špatná volba bodů měření může náš odhad průběhu sezónnosti výrazně zkreslit, jak je vidět na obr.2.1.



Obrázek 2.1: Různé volby bodů

Problém s délkou časové řady. Délkou časové řady rozumíme počet pozorování. Právě počet pozorování výrazně ovlivňuje „množství informace“ obsažené v řadě. Není zde však žádná přímá závislost. Ztrojnásobení počtu pozorování nemusí znamenat ztrojnásobení informací. Kupříkladu Box-Jenkinsovské metody vyžadují určitou minimální délku aspoň 50 pozorování. Paradoxně také přílišná délka může být na závadu. Nejen, že zvyšuje výpočetní nároky, ale především se zvyšuje riziko, že charakter řady (struktura generujícího procesu) se v průběhu času změnil.

Problémy s kalendářem. V současné době se předpokládá různý počet víkendů a pracovních dní v měsíci, musíme ale předpokládat i některé pohyblivé svátky. To pochopitelně ovlivňuje většinu ekonomických ukazatelů. Dalším problémem mohou být neočekávané události jako výpadky energie, volby, změny směrnic apod.. Tyto problémy se řeší různými metodami očištění, například: zavedením standartního měsíce apod..

Problémy s oficiální statistikou. Pro správnou analýzu jsou potřebná „správná“ data. Bohužel je někdy nutné věnovat pozornost i správnosti dat publikovaných oficiální statistikou. Příkladně je potřeba dát si pozor na to, zda jde o odhady prvně publikované, závěrečné údaje, nebo jejich pozdější korekce.

Problém „ceteris paribus“. Je zřejmé, že modely časových řad platí pouze „za jinak stejných podmínek“. Proto musíme opustit starší modely, pokud se výrazně změny vnější podmínky a struktura procesu.

Kapitola 3

Box-Jenkinsova metodologie

Box-Jenkinsova metodologie (zkráceně B–J metodologie) ztotožňuje systematickou část časové řady s částí deterministickou a je založena na myšlence, že časová řada může být chápána jako řada stochastického charakteru. Zásluha Boxe a Jenkinse nespočívá v objevení principu, ale ve vytvoření konkrétního postupu, jak tyto principy prakticky využívat.

Výhody - je flexibilní a rychle se adaptuje na změnu v charakteru modelovaného procesu.

A v mnoha případech dává nejlepší výsledky (vzhledem k MSE-střední čtvercové chybě).

Nevýhody - musí být dostatečně dlouhé realizace. Také se ztrácí možnost jednoduché interpretace výsledných modelů. (Lze těžko přesvědčit zadavatele, že řadu lze modelovat pomocí náhodných šoků. Jediným argumentem jsou zde často jen kvalitní předpovědi získané pomocí těchto modelů).

3.1. Stacionarita

B–J metodologie je odvozena pouze pro stacionární časové řady. Nicméně, některé nestacionární řady lze pomocí snadných postupů převést na řady stacionární. Existují dvě koncepce stacionarity, silná (též zvaná striktní) a slabá.

DEFINICE 3.1 Časová řada $X := \{X_t, t \in \mathbb{Z}\}$ se nazývá **striktně stacionární**, jestliže každá distribuční funkce jejího konzistentního systému $\{F_t\}_{t \in M}$, nezávisí na posunutí:

$F_t(x_i) \equiv F_{t+h}(x_i)$ pro každé $t \in M$ a $h \in \mathbb{Z}$.

Kde $M := \{\mathbf{t} \mid \mathbf{t} = (t_1, t_2, \dots, t_n \in T^n, t_i \neq t_j, \text{ pro } i \neq j, n \in \mathbb{N})\}$

Jinými slovy striktní stacionarita předpokládá pravděpodobnostní rozdělení dvou odpovídajících si vektorů hodnot časové řady $(x_t, x_{t+1}, \dots, x_{t+k})$ a $(x_{t+h}, x_{t+h+1}, \dots, x_{t+h+k})$, je stejné bez ohledu na to, kde se v časové řadě vektor nachází (tj. pravděpodobnostní rozložení je invariantní¹ vůči posunům v čase). Tento předpoklad je ovšem velmi silný a v praxi lze velmi obtížně ověřit, a proto byl v analýze časových řad zaveden méně omezující pojem slabá stacionarita.

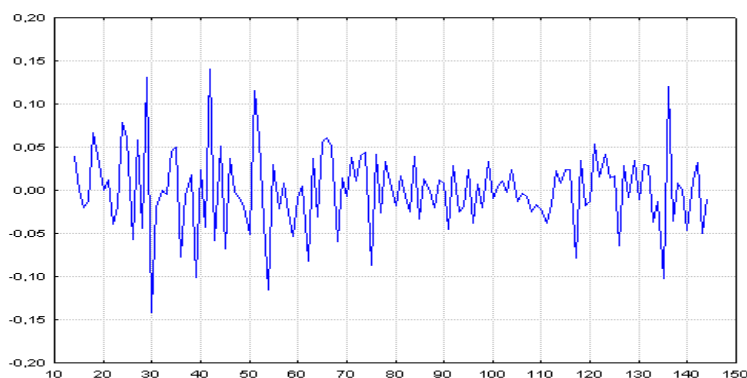
DEFINICE 3.2 Časová řada $X := \{X_t \mid t \in \mathbb{Z}\}$ se nazývá **slabě stacionární**, jestliže jsou splněny následující podmínky:

1. X má konečné druhé momenty: $\sigma_X^2(t) < \infty$ pro každé $t \in \mathbb{Z}$
2. $\gamma_X(r, s) = \gamma_X(r + h, s + h)$ každé $r, s, h \in \mathbb{Z}$. h je libovolné.
3. $\mu_X(x_i) \equiv \mu_X$ je konstantní funkce.

Jestliže jsou splněny pouze body 1 a 2, pak se nazývá X **kovariačně stacionární**.

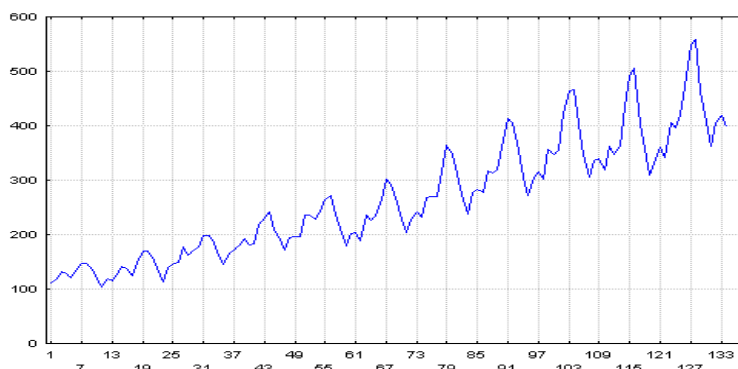
V bodě dva je $\gamma_X(r, s)$ autokovariační funkce, kterou si za okamžik nadefinujeme a v bodě tři se jedná o střední hodnotu časové řady.

Tedy slabě stacionární proces má v každém okamžiku stejnou střední hodnotu a stejný rozptyl. Kovariance dvou hodnot procesu ve dvou okamžicích závisí pouze na jejich vzdálenosti ne na jejich konkrétním umístění v časové řadě (t). Dále budeme stacionaritou rozumět, stacionaritu slabou. Pro řady s touto vlastností je snadné popsat vzájemnou závislost jednotlivých pozorování pomocí autokorelační, resp. parciální autokorelační funkce. Na obrázku 3.1 a 3.2 můžeme vidět srovnání stacionární a nestacionární řady.



Obrázek 3.1: Stacionární řada

¹Invariance-neměnnost



Obrázek 3.2: Nestacionární řada

3.2. Autokorelační a parciální autokorelační funkce

Autokorelační funkce ACF

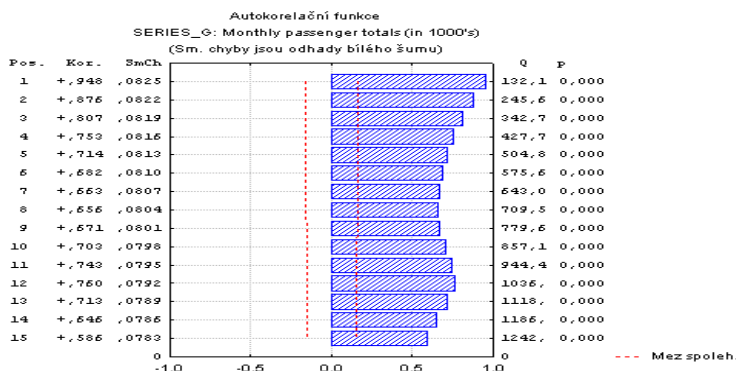
V případě stacionárního stochastického procesu X_t lze vyjádřit autokovarianční funkci mezi veličinami X_t a X_{t-k} jako

$$\gamma_k = \text{cov}(X_t, X_{t-k}) = E(X_t - \mu)(X_{t-k} - \mu), \quad k = 0, 1, \dots, n-1,$$

je zřejmé, že hodnota γ_0 je právě rozptylem hodnot časové řady σ^2 . Tuto funkci je snadné normovat, čímž získáme autokorelační funkci ρ_k

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma^2} = \frac{\text{cov}(X_t, X_{t-k})}{\sqrt{D(X_t)}\sqrt{D(X_{t-k})}}, \quad k = 0, 1, \dots, n-1$$

Vzhledem ke stacionaritě hodnota $\rho_0 = 1$ a ostatní hodnoty autokorelační funkce se pohybují v intervalu $\langle -1, 1 \rangle$ a její graf nazýváme korelogram, obr. 3.3. Jejich hodnoty



Obrázek 3.3: Korelogram

jsou pro nás neznámé, ale dokážeme je odhadnout přímo z pozorovaných dat. Čili odhad

r_k autokorelační funkce ρ_k získáme ze vztahu

$$r_k = \frac{\sum_{t=k+1}^n (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad k = 1, 2, \dots, n-1$$

Hodnoty autokorelační funkce zobrazují vztah mezi dvěma pozorováními tak, jak je ovlivněn pozorováními a náhodnými vlivy mezi těmito pozorováními. Např. hodnota autokorelační funkce mezi hodnotami y_1 a y_5 zahrnuje vliv hodnot y_2, \dots, y_4 . Tedy její hodnota nám podává informaci o síle lineární závislosti.

Parciální autokorelační funkce PACF

Abychom abstrahovali vliv vnitřních pozorování, zavádí se parciální autokorelační funkce. Její hodnoty ρ_{kk} jsou dány vztahem

$$\rho_{kk} = \frac{\det \mathbf{P}_k^*}{\det \mathbf{P}_k} \quad (3.2.1)$$

kde $\det \mathbf{P}_k$ je determinant matice

$$\mathbf{P}_k = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{pmatrix}$$

a $\det \mathbf{P}_k^*$ je determinant matice

$$\mathbf{P}_k^* = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_2 \\ \rho_2 & \rho_1 & 1 & \dots & \rho_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & \rho_k \end{pmatrix}$$

Jak je vidět, matice se liší pouze posledním sloupcem. Odhad r_{kk} parciální autokorelační funkce ρ_{kk} získáme snadno dosazením odhadnutých hodnot r_k za hodnoty ρ_k do vztahu 3.2.1.

3.3. Základní procesy

B–J procesy se skládají ze dvou základních procesů. Procesu AR a procesu MA. Z těchto základních procesů se pak konstruují další složené modely. Než se začneme věnovat samotné analýze, nadefinujeme si tzv. operátor zpětného posunutí. Tento operátor nemá žádný matematický význam, slouží nám pouze ke zjednodušení zápisu.

DEFINICE 3.3 *Nechť $\{X_t \mid t \in \mathbb{Z}\}$ je náhodný proces. Operátor zpětného posunutí je definován pomocí výrazu*

$$BX_t = X_{t-1},$$

přičemž jej lze aplikovat několikanásobně jako

$$B^j X_t = X_{t-j}$$

Autoregresní proces AR(p)

Autoregresní proces (Autoregressive process, AR(p)) řádu p označuje takový proces, kdy je hodnota časové řady v čase t tvořena lineární kombinací minulých hodnot této řady, tedy

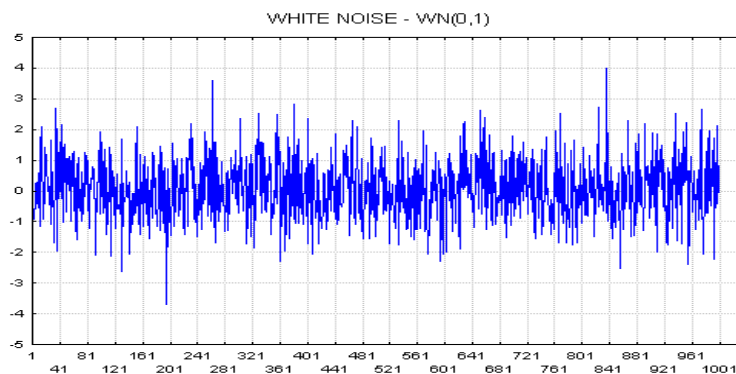
$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + a_t,$$

pomocí zpětného operátoru lze zkráceně zapsat

$$\phi_p(B)x_t = a_t,$$

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

Dále x_t jsou pozorované hodnoty řady v čase t , ϕ_i jsou neznámé parametry a a_t je hodnota náhodné veličiny v čase t . Předpokládáme, že tato náhodná veličina má charakter bílého šumu (White Noise obr. 3.4), tj. že je tvořena procesem $\{a_t, t \in T\}$, jehož hodnoty mají nulovou střední hodnotu $Ea_t = 0$, v čase konstantní rozptyl $Da_t = \sigma^2$ a jsou vzájemně nekorelované $\text{cov}(a_t, a_h) = 0$, ($h \neq t$) (mají nulové ACF a PACF).



Obrázek 3.4: Gausovský bílý šum

Parametr ϕ je možné chápat jako indikátor "paměti" procesu. Čím je v absolutní hodnotě bližší jedné, tím je paměť procesu delší a naopak, čím je bližší nule, tím je paměť kratší. Je-li roven nule, autokorelační funkce je nulová, proces nemá žádnou paměť a jedná se o proces bílého šumu.

Stále ještě nebyl zaveden pojem invertibilita. Proces je invertibilní, když ho lze převést na odpovídající (nekonečný) proces $AR(\infty)$. Je zřejmé, že AR procesy jsou vždy invertibilní. Podmínky stacionarity se dají přeformulovat na podmínku za pomoci polynomální rovnice $\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$. Tato podmínka je splněna, leží-li kořeny rovnice $(1 - \phi_1 B - \dots - \phi_p B^p) = 0$ vně jednotkového kruhu.

Hodnoty autokorelační funkce lze vyjádřit na základě diferenční rovnice

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}, \quad k = 1, 2, \dots$$

ACF tvoří kombinace exponenciálně klesajících pohybů (v případě reálných kořenů rovnice $\rho_p(B) = 0$) a exponenciálně klesajících sinusoidních pohybů (v případě komplexních kořenů). Hodnoty PACF pro zpoždění $k = 1, 2, \dots, p$ jsou různé od nuly, pro další zpoždění se potom rovnají nule.

Proces klouzavých součtů $MA(q)$

Proces klouzavých součtů (Moving Averages, $MA(q)$) řádu q označuje takový proces, kdy hodnota vysvětlované veličiny v čase t je tvořena lineární kombinací současných a minulých hodnot náhodné veličiny a_t . Tato veličina má charakter bílého šumu.

$$x_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_q,$$

nebo také opět pomocí zpětného operátoru

$$x_t = \theta_q(B) a_t.$$

MA procesy jsou vždy stacionární. Invertibilní je, pokud jsou všechny kořeny polynomu $(1 - \theta_1 B - \dots - \theta_q B^q) = 0$ vně jednotkové kružnice. Pak ACF je tvaru

$$\rho_k = \frac{-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}, \quad k = 1, 2, \dots, q,$$

$$\rho_k = 0, \quad k > q.$$

Hodnoty ACF jsou tedy pro zpoždění $k = 1, 2, \dots, q$ různé od nuly, pro další zpoždění se potom rovnají nule. PACF tvoří kombinace exponenciálně klesajících pohybů (v případě reálných kořenů rovnice $\theta_q(B) = 0$) a exponenciálně klesajících sinusoidních pohybů (v případě komplexních kořenů).

Autoregresní proces klouzavých součtů $ARMA(p, q)$

Autoregresní proces klouzavých součtů (nebo také smíšený proces) má tvar

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_q.$$

Tento proces je spojením či zobecněním základních procesů AR a MA. Je zřejmé, že ARMA($p, 0$) dává proces AR(p), stejně tak je to s procesem MA(q).

Tvar ACF bude obdobný jako v případě procesu AR(p), tj. bude mít formu kombinace exponenciálně klesajících pohybů a exponenciálně klesajících sinusoidních pohybů. Tento tvar však bude následovat až po prvních $q - p$ hodnotách, jestliže $q > p$. Hodnoty $\rho_0, \rho_1, \dots, \rho_{q-p}$ tento tvar mít nebudou. Pro $k > p - q$ se PACF bude v případě, že $p > q$ chovat stejně jako u procesu MA(q). Pro $k \leq p - q$ je však tvar této funkce odlišný.

3.4. Integrované a sezónní procesy

Dosud jsme se zabývali pouze stacionárními procesy. Zejména v ekonomické praxi se však můžeme setkat s časovými řadami tvořenými nestacionárními stochastickými procesy. Nejedná se jen o řady ekonomické, ale i energetické, s kterými budeme také pracovat. Nestacionarita procesu může být způsobena buď v čase se měnící střední hodnotou, nebo v čase se měnícím rozptylem procesu.

Protože doposud uvedené procesy lze aplikovat pouze na stacionární řady (dají se aplikovat i na nestacionární řady, bohužel s podstatně významnou chybou), zavádí se integrované procesy.

Proces náhodné procházky

Uvažujme proces

$$X_t = X_{t-1} + a_t,$$

který se nazývá proces náhodné procházky (Random Walk Process) zobrazeného na obr.3.5 . Jedná se o zvláštní případ procesu AR(1), kde $\phi_1 = 1$, pomocí zpětného operátoru posunutí jej lze vyjádřit jako

$$(1 - B)X_t = a_t. \tag{3.4.2}$$

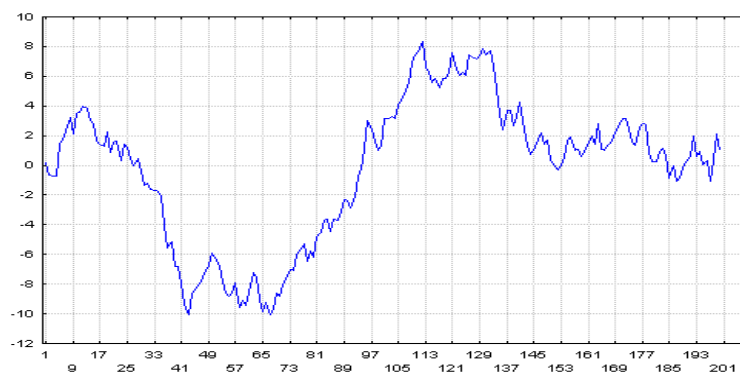
Vzhledem k tomu, že

$$X_t = (X_{t-2} + a_{t-1}) + a_t = (X_{t-3} + a_{t-2}) + a_{t-1} + a_t = \dots,$$

lze proces přepsat do formy

$$X_t = a_t + a_{t-1} + a_{t-2} + \dots = \sum_{i=0}^{\infty} a_{t-i}.$$

Náhodná procházka je nestacionární proces. Zdrojem její nestacionarity je stochastický trend $\sum_{j=1}^t a_j$. Je tedy zřejmé, že proces náhodné procházky je tvořen kumulováním náhodných veličin tvořících proces bílého šumu. Proces náhodné procházky se také nazývá



Obrázek 3.5: Náhodná procházka

integrováný proces. Protože jeho první diference je proces bílého šumu, nazývá se integrováný proces řádu jedna a označuje $I(1)$ (stacionární procesy, jako je například proces bílého šumu, či procesy AR, MA a ARMA, se označují $I(0)$). V případě, že je proces 3.4.2 modifikován do tvaru

$$(1 - B)^d X_t = a_t, \quad d = 2, 3, \dots$$

nazývá se integrovaným procesem řádu d a označuje se $I(d)$.

Proces ARIMA

Vykazuje-li po transformaci integrovaného procesu pomocí diference d -tého řádu výsledný proces takové autokorelace a parciální autokorelace, že jej lze vyjádřit ve formě stacionárního a invertibilního modelu $ARMA(p, q)$, potom se původní integrováný proces vyjádřený ve formě

$$\phi_p(B)(1 - B)^d x_t = \theta_q(B)a_t$$

nazývá *autoregresní integrováný proces klouzavých průměrů řádu p, d, q* a označuje se jako $ARIMA(p, d, q)$. Vlastnosti tohoto typu procesu jsou obdobné jako vlastnosti náhodné procházky. Tyto procesy se někdy nazývají pouze integrované procesy řádu d .

Sezónní procesy

Prozatím jsme se zabývali nesezónními procesy, nyní si ukážeme sezónní procesy. Myšlenka těchto procesů je následující, když jsme se zabývali vzájemnou závislostí mezi veličinami $\dots, x_t, x_{t+1}, x_{t+2}, \dots$ u výše popsaných procesů a tento proces obsahuje ještě sezónní kolísání, lze očekávat i závislost mezi sobě odpovídajícími veličinami v jednotlivých sezónách. Tedy mezi veličinami $\dots, x_{tS}, x_{t+1S}, x_{t+2S}, \dots$, kde S je délka sezónní periody (např. u měsíčních časových řad $s=12$). Protože podmínky pro stacionaritu a invertibilitu jsou obdobné, popíšeme si základní sezónní procesy SAR, SMA, SARMA, SARIMA, ve zkrácené formě.

Sezónní autoregresní proces SAR(P)

$$X_t = \Phi_1 X_{t-S} + \Phi_2 X_{t-2S} + \dots + \Phi_P X_{t-PS} + a_t,$$

pomocí operátoru zpětného posunutí jej lze zapsat jako

$$(1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P B^{PS})X_t = a_t, \quad \text{resp.} \quad \Phi_P(B^S)X_t = a_t.$$

Sezónní proces klouzavých průměrů SMA(Q)

Pomocí operátoru zpětného posunutí

$$X_t = (1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS})a_t = \Theta_Q(B^S)a_t.$$

Smíšené sezónní a nesezónní procesy SARMA(p, q)(P, Q)

Pomocí operátoru zpětného posunutí lze tento model vyjádřit jako

$$\phi_p(B)\Phi_P(B^S)X_t = \theta_q(B)\Theta_Q(B^S)a_t,$$

kde

$$\begin{aligned}\phi_p(B) &= (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p), \\ \theta_q(B) &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q), \\ \Phi_P(B^S) &= (1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P B^{PS}), \\ \Theta_Q(B^S) &= (1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS}).\end{aligned}$$

Proces SARIMA(p, d, q)(P, D, Q) $_S$

Předpokládejme, že proces obsahuje oba typy závislostí výše popsaných. Závislost uvnitř period je zachycena modelem ARIMA

$$\phi_p(B)(1 - B)^d x_t = \theta_q(B)b_t.$$

Proces $\{b_t\}$ obsahuje pouze sezónní závislosti a může být popsán modelem

$$\Phi_P(B^S)(1 - B^S)^D b_t = \Theta_Q(B^S)a_t,$$

kde $\Phi_P(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^{PS}$ a $\Theta_Q(B^S) = 1 - \Theta_1 B^S - \dots - \Theta_Q B^{QS}$. Prostřednictvím členu $(1 - B^S)$ se konstruují sezónní diference. Jestliže se procesy spojí, získá se proces

$$\Phi_P(B^S)\phi_p(B)(1 - B)^d(1 - B^S)^D x_t = \theta_q(B)\Theta_Q(B^S)a_t,$$

který je označován jako SARIMA(p, d, q)(P, D, Q) $_S$, kde p je řád procesu AR, q řád procesu MA, d řád prosté diference, P řád sezónního procesu SAR, Q řád sezónního procesu SMA, D řád sezónní diference a S je délka procesu.

3.5. Identifikace procesu

Počátečním stádiem tvorby modelu je obecně rozhodnutí o struktuře tohoto modelu. Situace je obvykle (aspoň teoreticky) jednoduchá u ekonometrických nebo stavových modelů, jejichž struktura je dána podkladovou ekonomickou teorií. V případě dekompozičních modelů, které ještě zmíníme, může být také někdy ekonomická teorie využita, navíc je možné charakter trendu a ostatních složek „odhadnout“ z grafu řady. Pro B–J modely bez vysvětlujících vstupů, tj. modely AR, MA, ARMA, ARIMA apod., to však neplatí. K dispozici obvykle není žádná podkladová hypotéza o závislostech mezi současnou a minulou hodnotou vysvětlované veličiny, natož pak o vztahu vysvětlované veličiny a minulých hodnot náhodné složky.

Přesto je právě volba počtu zpoždění v autoregresní složce p a složce klouzavých součtů q klíčová pro tvorbu modelu. Jde o nelehkou úlohu, jenž je v mnoha případech závislá na citu a zkušenosti analytika. Identifikace je přitom pouze první fází konstrukce modelů, neboť identifikovaný model je třeba ještě ověřit a upravit. Podívejme se nyní na jednotlivé kroky identifikace modelu.

Proces ARIMA

Prvním krokem při identifikaci časové řady je prozkoumání jejího grafu. V mnoha případech je možné na první pohled rozpoznat nestacionaritu řady, třeba přítomností trendu, popřípadě další problémy jako jsou odchýlená pozorování, sezónnost apod. V této fázi jde především o subjektivní zhodnocení situace. Nicméně, již na základě tohoto zhodnocení je možné vhodnou transformací převést na řadu stacionární (stacionarizovat) časovou řadu, či provést jiné úpravy jako je linearizace časové řady, resp. její stabilizace z hlediska rozptylu pomocí logaritmické transformace. Tento typ transformace je však vhodné provádět před vlastním diferencováním časové řady. Důvodem je skutečnost, že diferencováním je možné získat i záporné hodnoty.

Druhým krokem je odhadnout výpočet odhadů ACF a PACF původní časové řady. Na jejich základě je možné potvrdit, že časovou řadu je třeba stacionarizovat (v případě, že hodnoty výběrové ACF a PACF v prvním zpoždění jsou velmi blízké jedné a ostatní hodnoty výběrové ACF klesají pomalu).

Po stacionarizaci časové řady se snažíme určit, jaký proces časovou řadu generuje. Používají se výběrové ACF a PACF pro identifikaci modelů AR a MA (nalezení hodnot p, q). Tato identifikace je založena na principu podobnosti výběrové ACF a PACF s teoretickými ACF a PACF, kterými je každý B–J proces jednoznačně určen. Jak je vidět v tabulce 3.1 a na obr.3.6.

Proces SARIMA

Model	ACFE	PACF
AR(p)	exponenciální a/nebo exponenciálně sinusoidní pokles	$\phi_{kk} = 0$ pro $k > p$
MA(q)	$\rho_k = 0$ pro $k > p$	Omezená exponenciálním a/nebo exponenciálně sinusoidním poklesem
ARMA(p, q)	Od zpoždění ($q - p$) pro $q > p$ exponenciální nebo exponenciálně sinusoidní pokles	Od zpoždění ($p - q$) pro $p > q$ omezená exponenciálním nebo exponenciálně sinusoidní pokles

Tabulka 3.1: Tvary ACF a PACF procesů AR, MA, ARMA

Obsahuje-li časová řada také sezónní složku, je třeba identifikovat model typu SARIMA. Princip je stejný jako v předchozím případě .

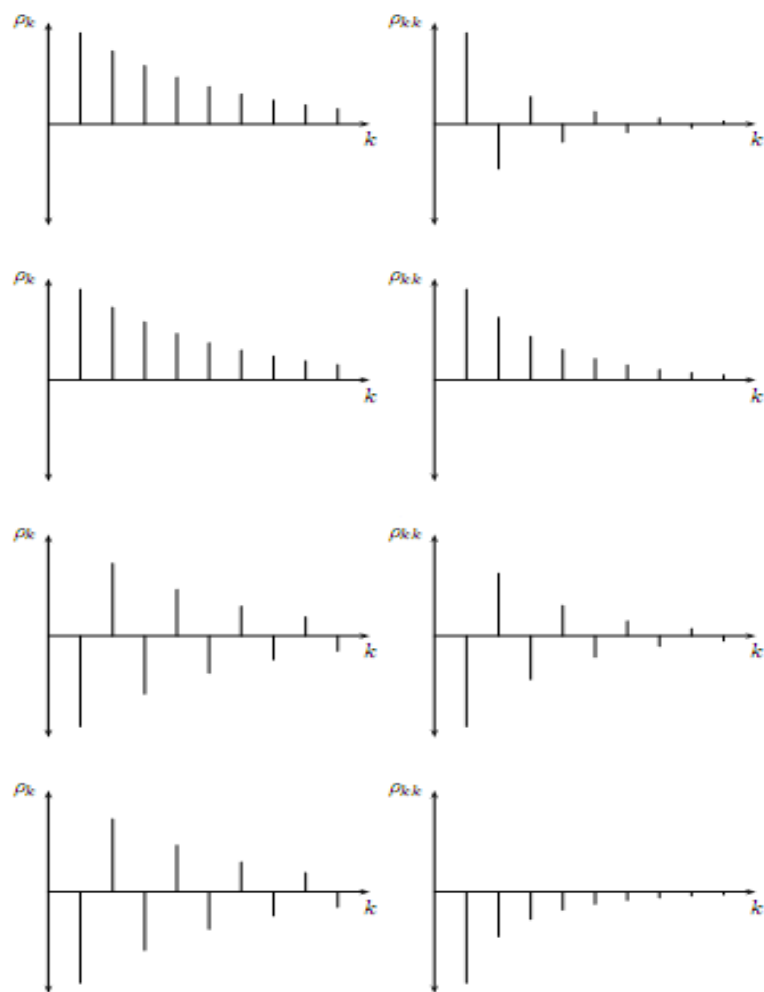
Nejprve je třeba časovou řadu stacionarizovat. Pokud je to nutné, linearizuje se řada pomocí logaritmické transformace. Poté se řada diferencuje, je třeba přitom mít na paměti, že sezónní difference zahrnuje rovněž diferenci prostou. Stacionarizaci pomocí sezónní difference indikuje tvar výběrové ACF a PACF. Tyto funkce jsou charakteristické vysokými hodnotami v nesezónních frekvencích.

V další fázi se vypočítá výběrová ACF a PACF pro stacionarizovanou časovou řadu, s jejich pomocí se potom určí typ sezónních modelů SAR, SMA, či SARMA (v sezónních frekvencích mají tyto funkce statisticky významně odlišné hodnoty od nuly, tyto hodnoty však nejsou tak vysoké, aby bylo možné časovou řadu považovat za nestacionární). Po identifikaci a určení modelu sezónní složky je třeba vypočítat výběrovou ACF a PACF tentokrát pro rezidua sezónního modelu a na jejich základě posoudit, zda by bylo vhodné model doplnit o složku AR, MA či ARMA.

Protože ACF a PACF jsou nutně náhodně vychýleny proti teoretickým hodnotám, umožňují proto jen přibližný odhad typu procesu a řádu procesu v jednotlivých složkách. Obvykle se vyzkouší víc alternativních modelů a z nich se vyberou ty, které mají všechny parametry statisticky významné a z nich dále ten model, který poskytuje nejlepší vysvětlení.

3.6. Verifikace procesu

Výše jsme zmínili jednu z mnoha cest, jak odhadnutý model označit za vyhovující. Je mnoho testů, kterými se odhanutý model ověřuje. My si zde uvedeme pouze tři z nich.



Obrázek 3.6: Typické tvary ACF a PACF pro ARMA(p, q)

Testování reziduí

Jednou ze dvou základních podmínek je, aby rezidua modelu měla charakter bílého šumu. Je tedy třeba otestovat normalitu reziduí a jejich nekorelovanost. My si uvedeme pouze korelační testy. Přítomnost autokorelace v reziduích lze otestovat odhadem autokorelační funkce, resp. výběrové autokorelační funkce

$$r_k = \frac{\sum_t \hat{a}_t \hat{a}_{t-k}}{\sum_t \hat{a}_t^2}.$$

Pokud hodnoty této funkce leží uvnitř intervalu $\pm 2/\sqrt{T}$ (95% interval spolehlivosti), můžeme hypotézu o nekorelovanosti přijmout. V opačném případě to je silný důvod pro zamítnutí modelu jako celku. A je vhodné pokusit se identifikovat model s větším počtem parametrů AR, MA, ... atd.

Druhá možnost ověření nekorelovanosti reziduí je tzv. *Portmanteau test*. Je-li model vyhovující, potom statistika

$$Q = n(n+2) \sum_{k=1}^K \frac{r_k^2}{n-k},$$

kde K je vhodně zvolené číslo a doporučuje se blízké \sqrt{n} , kde n je počet pozorování. Pro model ARMA(p, q) má Q statistika přibližně χ^2 rozdělení s $K - p - q$ stupni volnosti. Pokud Q překročí kritickou hodnotu $\chi_{K-p-q}^2(\alpha)$, pak na hladině významnosti α zamítneme hypotézu, že rezidua jsou nekorelovaná.

Testování parametrů modelu

Posledním zde uvedeným testem je test jednotlivých parametrů modelu, tj. parametr ϕ_i , pro $i = 1, 2, \dots, p$ a θ_i , pro $i = 1, 2, \dots, q$. Testování se provádí na základě t-testů, resp. pomocí statistik

$$T_{\phi_i} = \frac{\hat{\phi}_i}{\hat{S}_{\phi_i}}, \quad \text{pro } i = 1, 2, \dots, p, \quad (3.6.3)$$

$$T_{\theta_i} = \frac{\hat{\theta}_i}{\hat{S}_{\theta_i}}, \quad \text{pro } i = 1, 2, \dots, q,$$

kde \hat{S}_{ϕ_i} a \hat{S}_{θ_i} jsou odhady směrodatných chyb odhadů jednotlivých parametrů.

Poměrně často se stává, že identifikace modelu není jednoznačně určena, nebo že při zkoušení alternativních modelů je zjištěno, že data lze modelovat několika alternativními modely. A jak jsme již zmínili, existují další kritéria, podle kterých se rozhodujeme o vhodnosti daného modelu. Kupříkladu historicky nejstarším kritériem je AIC (Akaike Information Criterion) nebo BIC (Bayesian Extension of Information Criterion). Jinou možností je vybrat model, který poskytuje nejlepší předpovědi.

3.7. Předpověď procesu

Předpověď s minimální střední čtvercovou chybou

Úkolem je předpovědět budoucí hodnoty v čase X_{T+h} . V této souvislosti se T nazývá prahem predikce a h se označuje jako horizont predikce. Lineární proces v čase $(T+h)$ lze podle principu „ceteris paribus“ zapsat ve formě

$$X_{T+h} = a_{T+h} + \psi_1 + \dots + \psi_{h-1}a_{T+1} + \psi_h a_T + \psi_{h+1}a_{T-1} + \dots = \psi(B)a_{T+h},$$

kde $\psi(B) = \phi_p(B)^{-1}\Delta^{-1}\theta_q(B)$. Předpověď hodnoty X_{T+h} pro $h \geq 1$ je konstruován v čase T , takže jsou známy hodnoty X_T, X_{T-1}, \dots , které jsou jejich lineární kombinací. Předpověď s minimální čtvercovou chybou lze zapsat ve formě

$$X_T(h) = \psi_h^* a_T + \psi_{h+1}^* a_{T-1} + \psi_{h+2}^* a_{T-2} + \dots$$

Střední čtvercová chyba (Mean Square Error) předpovědi je

$$MSE[X_T(h)] = E[X_{T+h} - X_T(h)]^2 = \sigma_a^2 \sum_{j=0}^{h-1} \psi_j^2 + \sigma_a^2 \sum_{j=0}^{\infty} (\psi_{h+j} - \psi_{h+j}^*)^2,$$

minimální je tehdy, jestliže $\psi_{h+j}^* = \psi_{h+j}$, $j = 0, 1, \dots$. Z toho plyne, že

$$X_T(h) = \psi_h a_T + \psi_{h+1} a_{T-1} + \psi_{h+2} a_{T-2} + \dots$$

Za předpokladu, že

$$E(a_{T+j} | X_T, X_{T-1}, \dots) = \begin{cases} a_{T+j}, & j \leq 0 \\ 0, & j > 0, \end{cases}$$

lze psát

$$E(X_{T+h} | X_T, X_{T-1}, \dots) = \psi_h a_T + \psi_{h+1} a_{T-1} + \psi_{h+2} a_{T-2} + \dots$$

Z toho vyplývá, že

$$X_T(h) = E(X_{T+h} | X_T, X_{T-1}, \dots),$$

tj. předpověď hodnoty v čase $(T+h)$ s minimální čtvercovou chybou je podmíněná střední hodnota náhodné veličiny X_{T+h} . Předpovědní chyba má formu

$$e_T(h) = X_{T+h} - X_T(h) = \sum_{j=0}^{h-1} \psi_j a_{T+h-j}.$$

Její střední hodnota je tedy nulová, tj. $E[e_T(h)] = 0$. Rozptyl předpovědi se rovná její střední čtvercové chybě a má formu

$$D[X_{T+h} | X_T, X_{T-1}, \dots] = D[e_T(h)] = \sigma_a^2 \sum_{j=0}^{h-1} \psi_j^2.$$

Pro normální (gaussovský) stochastický proces je $(1 - \alpha)100\%$ předpovědní interval vymezen hranicemi

$$X_T(h) \pm u_{1-\alpha/2} \left[\sum_{j=0}^{h-1} \psi_j^2 \right]^{1/2} \sigma_a,$$

kde $u_{1-\alpha/2}$ je $(1 - \alpha/2)100\%$ kvantil normovaného normálního rozdělení. Pro stacionární procesy, tj. v případě, že $d = 0$, existuje $\lim_{h \rightarrow \infty} \sum_{j=0}^{h-1} \psi_j^2$. To znamená, že interval spolehlivosti se limitně blíží ke dvěma horizontálně paralelním přímkám. Pro nestacionární procesy, tj. v případě, že $d = 1, 2, \dots$, $\lim_{h \rightarrow \infty} \sum_{j=0}^{h-1} \psi_j^2$ neexistuje. Proto se také interval spolehlivosti s rostoucím horizontem předpovědi h rozšiřuje neomezeně.

Výpočet předpovědi

Při výpočtu se vychází ze skutečnosti, že předpověď $X_T(h)$ je dána podmíněnou střední hodnotou $E(X_{T+h} | X_T, X_{T-1}, \dots)$. Z tohoto důvodu lze pro výpočet předpovědi použít přímo daný model ve formě diferenční rovnice. Předpokládá se, že

$$\alpha_{p+d}(B) = \phi_p(B)(1 - B)^d = (1 - \alpha_1 B - \dots - \alpha_{p+d} B^{p+d}).$$

Model ve formě ?? může být přepsán do podoby

$$(1 - \alpha_1 B - \dots - \alpha_{p+d} B^d) X_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t,$$

pro $t = T + h$ potom

$$X_{T+h} = \alpha_1 X_{T+h-1} + \alpha_2 X_{T+h-2} + \dots + \alpha_{p+d} X_{T+h-p-d} + a_{T+h} - \theta_1 a_{T+h-1} - \dots - \theta_q a_{T+h-q}$$

Tedy předpověď s horizontem h konstruovaná v čase T má tvar

$$X_T(h) = \alpha_1 X_T(h-1) + \dots + \alpha_{p+d} X_T(h-p-d) + a_T(h) - \theta_1 a_T(h-1) - \dots - \theta_q a_T(h-q), \quad (3.7.4)$$

kde

$$X_T(j) = E(X_{T+j} | X_T, X_{T-1}, \dots) \quad \text{pro } j \geq 1,$$

$$X_T(j) = X_{T+j} \quad \text{pro } j \leq 0,$$

$$a_T(j) = 0 \quad \text{pro } j \geq 1,$$

$$a_T(j) = X_{T+j} - X_{T+j-1}(1) = a_{T+j} \quad \text{pro } j \leq 0.$$

Pro odhadnuté parametry daného modelu, v našem případě pro proces ARIMA, se předpověď konstruuje na základě modelu 3.7.4, do kterého se místo známých hodnot parametrů dosadí jejich odhady. Intuitivně můžeme také konstatovat, že předpovědní interval je pro odhadnuté parametry širší.

Kapitola 4

Dekompozice časových řad

Klasická analýza časových řad vychází z předpokladu, že časovou řadu je možné rozložit na čtyři složky.

4.1. Složky řady

Trendová Složka (T) vyjadřuje dlouhodobou tendenci vývoje zkoumaného jevu. Je výsledkem faktorů, které dlouhodobě působí stejným směrem, např. technologie výroby, podmínky na trhu apod.

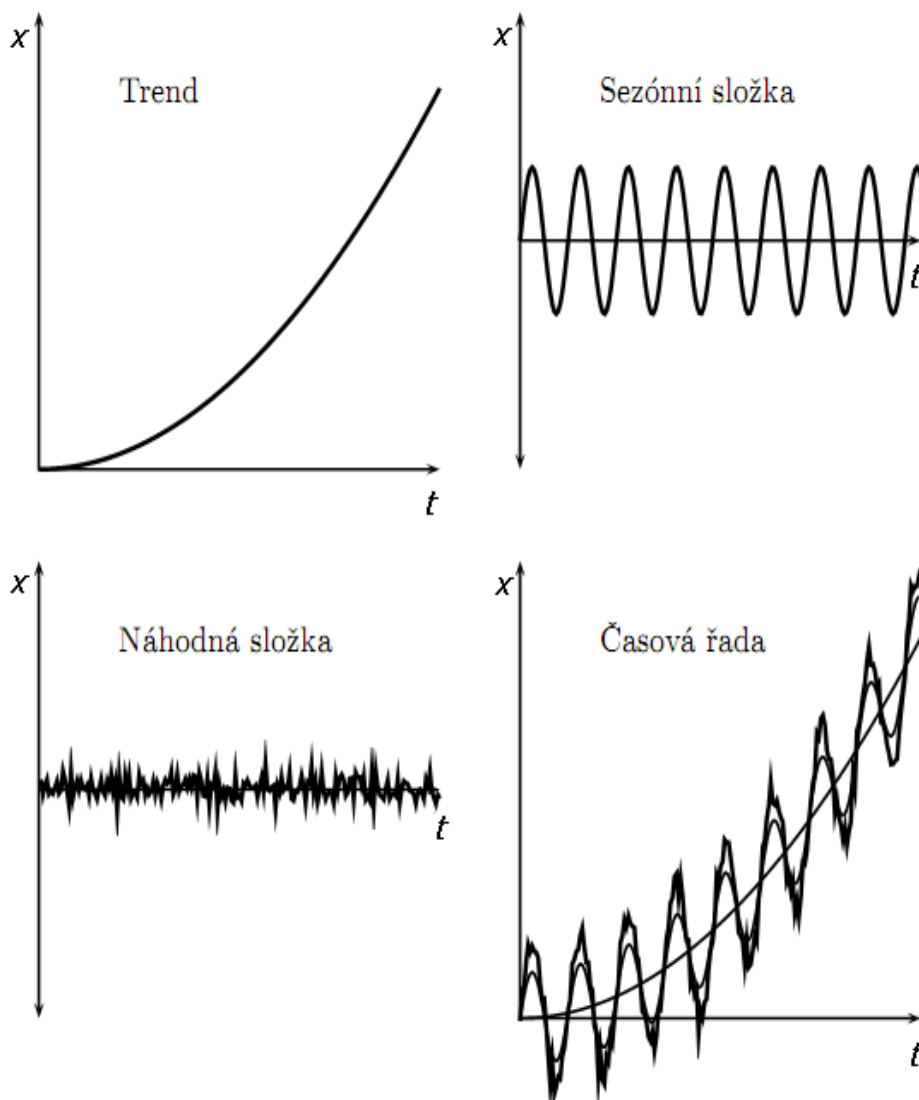
Cyklická složka (C) vyjadřuje kolísání okolo trendu, ve kterém se střídají fáze růstu a poklesu. Jednotlivé cykly (periody) se vytvářejí za období delší než jeden rok a mají nepravidelný charakter, tj. různou délku a amplitudu. Cykly jsou v ekonomických časových řadách způsobeny ekonomickými a neekonomickými faktory. V posledních letech se věnuje pozornost zejména technologickým, inovačním či demografickým cyklům. Zároveň je to nejspornější část časové řady. Protože se dekompoziční metody používají především na krátkodobé a střednědobé předpovědi, bývá cyklická složka někdy zanedbána, tj. zahrnuta do trendu.

Sezónní složka (S) vyjadřuje pravidelné kolísání okolo trendu v rámci kalendářního roku. Sezónní výkyvy se opakují každoročně ve stejných obdobích (délka periody je jeden rok) a vznikají v důsledku střídání ročních období nebo vlivem různých institucionalizovaných zvyků, jako jsou např. svátky, dovolené apod.

Nesystematická složka (I) vyjadřuje nahodilé a jiné nesystematické výkyvy, ale také chyby měření apod.

Trendová a cyklická složka mohou být přítomné v časových řadách ročních údajů, ale také v krátkodobých časových řadách, tj. v řadách s intervalem sledování kratším než jeden

rok, např. čtvrtletních, měsíčních, týdenních apod. Sezónní složka se vyskytuje pouze v krátkodobých časových řadách, obvykle v měsíčních a čtvrtletních. Nesystematická složka je přítomná v každé časové řadě.



Obrázek 4.1: Jednotlivé složky časové řady

4.2. Typy dekompozice

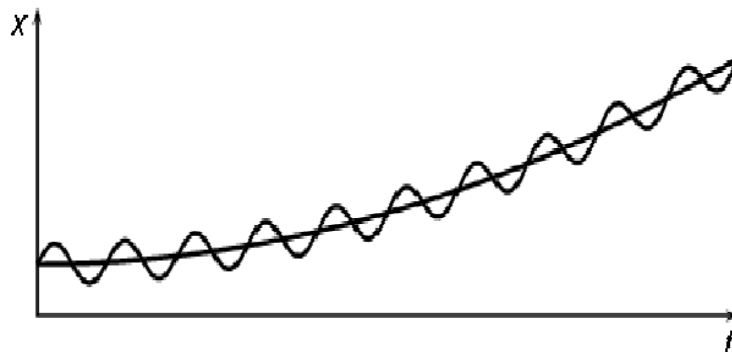
Při dekompozici se vychází ze tří různých modelů časové řady: aditivního, multiplikativního a smíšeného. Tyto modely specifikují, jakým způsobem jsou jednotlivé složky časové řady „skloubeny“ dohromady.

- **Aditivní model** předpokládá, že hodnoty časové řady se dají určit jako součet

hodnot jednotlivých složek, tedy že

$$x_t = T_t + C_t + S_t + I_t$$

V tomto modelu je každá ze složek uváděna v absolutní hodnotě a má stejnou



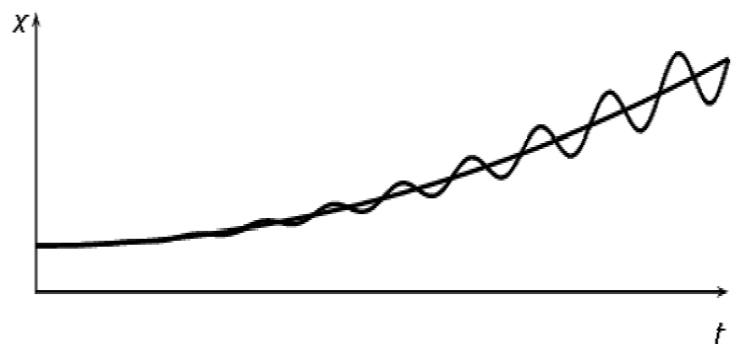
Obrázek 4.2: Časová řada s aditivní sezónní složkou

jednotku jako celý sledovaný ukazatel X_t .

- **multiplikativní model** předpokládá, že hodnoty časové řady se dají určit jako součin hodnot jednotlivých složek, tj.

$$x_t = T_t \cdot C_t \cdot S_t \cdot I_t$$

U multiplikativních časových řad přisuzujeme absolutní hodnotu (a tedy i jednotku)



Obrázek 4.3: Časová řada s multiplikativní sezónní složkou

pouze trendu; ostatní složky považujeme za bezrozměrné koeficienty.

- **Smíšený model** je vlastně jen kombinací obou předchozích přístupů. Některé složky mohou být v součtu, jiné v součinu. Typickým příkladem může být třeba takovýto model časové řady:

$$x_t = T_t \cdot C_t \cdot S_t + I_t$$

Smíšené časové řady se identifikují nejobtížněji.

Při dekompozici se snažíme nejdříve identifikovat trend a potom teprve sezónní složku. Někdy se však postupuje i opačně. K identifikaci trendu se používají nejčastěji především čtyři následující metody:

- naivní modely
- proložení matematickou křivkou
- vyrovnání metodou klouzavých průměrů
- exponenciální vyrovnání

4.3. Analýza trendové složky

My se budeme zabývat pouze exponenciálním vyrovnáváním. To patří mezi tzv. adaptivní metody vyrovnávání trendu, které se přizpůsobují změnám v charakteru analyzované veličiny z důvodu zpracování dané řady po částech nebo použití „zapomínání“ starých hodnot.

My si zde ukážeme podstatu na tzv. dvojitým exponenciálním vyrovnání, které je vhodné použít pro krátké období s lineárním charakterem. Odhadují se pak dva parametry b_0 a b_1 na základě minimalizace výrazu

$$S^2 = \sum_{\tau=0}^{t-1} \gamma^\tau (x_{t-\tau} - b_0 - b_1(-\tau))^2 \rightarrow \min,$$

kde $\gamma \in (0, 1)$ je koeficient zapomínání. Řešením normálních rovnic získáme vztahy

$$b_0 - \frac{\gamma}{1-\gamma} b_1 = (1-\gamma) \sum_{\tau=0}^{\infty} \gamma^\tau x_{t-\tau}, \quad (4.3.1)$$

$$\gamma b_0 - \frac{\gamma(\gamma+1)}{1-\gamma} b_1 = (1-\gamma)^2 \sum_{\tau=0}^{\infty} \tau \gamma^\tau x_{t-\tau}. \quad (4.3.2)$$

Pro zjednodušení můžeme pravou část první normální rovnice pojmenovat jednoduchá vyrovnávací statistika (odpovídá totiž jednoduchému exponenciálnímu vyrovnání) a označit ji symbolem S_t

$$S_t = (1-\gamma) \sum_{\tau=0}^{\infty} \gamma^\tau x_{t-\tau} = (1-\gamma)x_t + \gamma S_{t-1}. \quad (4.3.3)$$

Dále zavedeme dvojitou vyrovnávací statistiku, která je vlastně jednoduchým exponenciálním vyrovnáním jednoduché vyrovnávací statistiky S_t

$$S'_t = (1-\gamma) \sum_{\tau=0}^{\infty} \gamma^\tau S_{t-\tau} = (1-\gamma)S_t + \gamma S'_{t-1}. \quad (4.3.4)$$

Pokud tyto statistiky dosadíme do 4.3.2, pak lze snadno získat odhady parametrů v čase t

$$\hat{b}_0(t) = 2S_t - S'_t, \quad \hat{b}_1(t) = \frac{1-\gamma}{\gamma} (S_t - S'_t).$$

Předpověď na čas $t + \gamma$ na základě informací v čase t má tvar

$$\hat{y}_{t+\tau}(t) = \hat{b}_0(t) + \hat{b}_1(t)\tau = \left(2 + \frac{(1-\gamma)\tau}{\gamma}\right) S_t - \left(1 + \frac{(1-\gamma)\tau}{\gamma}\right) S'_t.$$

Počáteční hodnoty S_0 a S'_0 získáme dosazením do vztahů, kde parametry \hat{b}_0 a \hat{b}_1 jsou regresní odhady parametrů získané tak, že prvních 6 nebo $n/2$ pozorování proložíme přímkou. Koeficient zapomínání γ hledáme simulací, která se snaží najít nejvhodnější hodnotu, pro kterou dává model nejmenší součet čtverců chyb předpovědi. Čili srovnává se skutečná hodnota s předpovědí z minulého kroku. Předpovědní $(1 - \alpha)100\%$ interval je tvaru

$$(\hat{x}_{t+\tau}(t) - u(\alpha/2)d_\tau\Delta(n), \quad \hat{x}_{t+\tau}(t) + u(\alpha/2)d_\tau\Delta(n)), \quad (4.3.5)$$

kde $u(\alpha/2)$ je $(1 - \alpha)100\%$ kvantil normovaného normálního rozdělení $N(0, 1)$, $\Delta(n)$ je průměrná absolutní chyba předpovědi o jeden krok dopředu

$$\Delta(n) = \sum_{t=1}^n \frac{|x_t - \hat{x}_t(t-1)|}{n}$$

a

$$d_\tau = 1.25 \sqrt{\frac{1 + \frac{1-\gamma}{(1+\gamma)^3}((1 + 4\gamma + 5\gamma^2) + 2(1-\gamma)(1 + 3\gamma)\tau + 2(1-\gamma)^2\tau^2)}{1 + \frac{1-\gamma}{(1+\gamma)^3}((1 + 4\gamma + 5\gamma^2) + 2(1-\gamma)(1 + 3\gamma) + 2(1-\gamma)^2)}}.$$

4.4. Analýza sezónní složky

Nyní se zaměříme na odhad sezónní složky. Představíme si Wintersovu metodu, která je zobecněním exponenciálního vyrovnání. Předpokládejme, že trend odpovídá výše popsanému dvojitému exponenciálnímu vyrovnání. Označme model následujícím způsobem. Máme k dispozici n pozorování, rok je tvořen m sezónními obdobími. Odhady sezónní složky S_t v čase t označme $s_t(t)$. Úroveň trendu označíme $a_0(t)$

$$\hat{a}_0(t) = \hat{b}_0(t) + \hat{b}_1(t).$$

Narozdíl od exponenciálního vyrovnání, které používá pouze jeden koeficient zapomínání, Wintersova metoda používá tři různé koeficienty $\gamma, \delta, \eta \in (0, 1)$. Adaptační vzorce multiplicative modelu sezónnosti pro výpočet odhadovaných hodnot z jejich minulých hodnot mají tvar

$$\hat{a}_0(t) = \gamma \frac{x_t}{s_{t-m}(t-m)} + (1-\gamma)(\hat{a}_0(t-1) + \hat{b}_1(t-1)),$$

$$\begin{aligned}\hat{b}_1(t) &= \delta(a_0(t) - \hat{a}_0(t-1)) + (1-\delta)\hat{b}_1(t-1), \\ s_t(t) &= \eta \frac{x_t}{\hat{a}_0(t)} + (-\eta)s_{t-m}(t-m).\end{aligned}$$

Tyto vztahy mají intuitivní odůvodnění. Odhad úrovně trendu $\hat{a}_0(t)$ v čase t je opravou předpovědi $\hat{a}_0(t-1) + \hat{b}_1(t-1)$ nové hodnoty podle současného měření $x_t/s_{t-m}(t-m)$. Zde používáme minulý odhad sezónní složky, protože nový ještě není k dispozici. Koeficient zapomínání γ slouží jako váha mezi minulou a současnou hodnotou. Stejným způsobem je možné odůvodnit i adaptační vztahy pro výpočet $\hat{b}_1(t)$ a $s_t(t)$.

Předpověď na čas $t + \tau$ na základě informací dostupných v čase t je tvaru

$$\hat{x}_{t+\tau}(t) = (\hat{a}_0(t) + \hat{b}_1(t)\tau)s_{t+\tau-m}(t + \tau - m). \quad (4.4.6)$$

V případě, že τ je velké číslo, takže sezónní faktor pro čas $t + \tau - m$ ještě není k dispozici, použijeme poslední známou hodnotu odpovídajícího sezónního faktoru, tedy hodnotu z času $t + \tau - 2m, t + \tau - 3m$ atd. Vyrovnanou hodnotu získáme ze vztahu 4.4.6 dosazením za $\tau = 0$, tedy ze vztahu

$$\hat{x}_t(t) = \hat{a}_0(t)s_t(t).$$

Přibližný interval spolehlivosti předpovědi má na hladině významnosti α tvar 4.3.5, kde

$$\begin{aligned}\Delta(n) &= \frac{\sum_{t=1}^n \left| \frac{x_t}{s_{t-m}(t-m)} - \hat{a}_0(t-1) - \hat{b}_1(t-1) \right|}{n}, \\ d_\tau &= 1.25 \sqrt{\frac{1 + \frac{\theta}{(1+\nu)^3} ((1 + 4\nu + 5\nu^2) + 2\theta(1 + 3\nu)\tau + 2\theta^2\tau^2)}{1 + \frac{\theta}{(1+\theta)^3} ((1 + 4\nu + 5\nu^2) + 2\theta(1 + 3\nu) + 2\theta^2)}}\end{aligned}$$

kde $\theta = \max(\gamma, \delta, \eta)$ a $\nu = 1 - \theta$.

Počáteční hodnoty parametrů spočítáme dle

$$\begin{aligned}\hat{b}_1(0) &= \frac{\bar{x}_r - \bar{x}_1}{(r-1)m}, \\ \hat{a}_0(0) &= \bar{x}_1 - \frac{m}{2}\hat{b}_1(0), \\ s_t(t) &= \frac{m}{\sum_{i=1}^m \bar{s}_i} \bar{s}_{t+m}, \quad t = 1 - m, 2 - m, \dots, 0,\end{aligned}$$

kde r je počet let, pro která máme k dispozici pozorování, \bar{y}_j je aritmetický průměr hodnot pozorování y_t v roce j ,

$$\begin{aligned}\bar{s}_i &= \frac{1}{r} \sum_{j=0}^{r-1} s_{i+mj}^*, \quad i = 1, \dots, m, \\ s_t^* &= \frac{x_t}{\bar{x}_j - ((m+1)/2 - i)\hat{b}_1(0)}, \quad t = 1, \dots, mr,\end{aligned}$$

kde j označuje v posledním vztahu číslo roku a index i číslo sezónního období, které odpovídá času t . Kupříkladu pro $t = m + 1$ je $j = 2$ a $i = 1$.

Pro vlastní vyrovnání je zapotřebí ještě stanovit hodnoty koeficientů zapomínání γ, δ a η . Určujeme je obdobným způsobem, který byl popsán výše.

Kapitola 5

Experimentální část

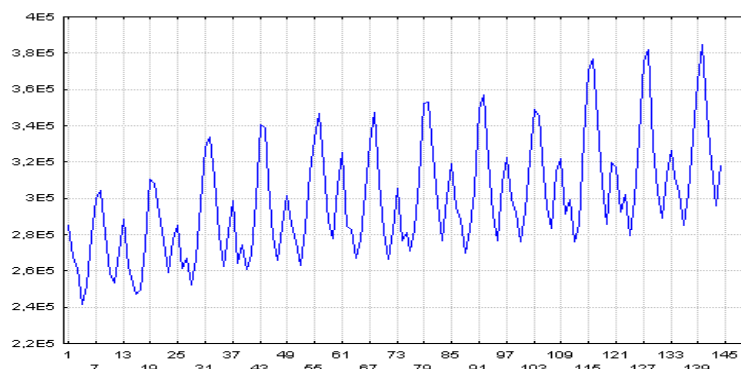
Již od pradávna se snaží lidstvo odhadovat, co by mohlo nastat v blízké budoucnosti. Ať už v podobě věštění z karet, numerologie, nebo v posledních letech snahou o vytvoření kvalitních modelů, které by na základě dat z minulosti mohly odhadnout, jaká budou data v blízké budoucnosti. Proto je naším cílem pokusit se tak trochu o nemožné. A to, zda-li je možné odhadnout z dané časové řady zastoupení náhodné složky. Vezmeme si příklad, pokud budeme mít časovou řadu jako na obr.5.1 , která je řadou energetickou, vyjadřující spotřebu elektrické energie v USA za posledních 11 let. Je zcela viditelné, že tato řada vykazuje jistou pravidelnost a i intuitivně můžeme říci, že náhodnost v této řadě bude jistě malá. Jako druhou řadu jsme si vybrali ekonomickou. Jedná se o vývoj indexu PX50, viz obr.5.6, která by mohla mít větší zastoupení náhodné složky.

My se zde budeme zabývat pouze základními B–J modely, přestože dnes se už pracuje s jejich upravenými modely, jako např. GARCH, STGARCH, ... atd.. Tento postup je zcela zřejmý, pokud naše teorie nebude platit pro základní modely, je zcela zřejmé, že nebude platit ani pro odvozené.

Vypočítané hodnoty predikce daným modelem budeme porovnávat s našimi testovacími funkcemi, resp. s jejich vypočtenými predikcemi.

5.1. Energetická řada

Jak jsme již zmínili, z reálných řad jsme si vybrali dvě řady. Jedna je z oblasti ekonomie a druhá je z oblasti energetiky. V případě energetické řady byla zvolena spotřeba elektrické energie v USA, která je vyjádřena měsíční spotřebou v miliónech kilowatthodin za období od začátku roku 1996 až po konec roku 2007, obr.5.1. Pro zjednodušení v další části textu, pojmenujeme tuto časovou řadu „USA“.



Obrázek 5.1: Spotřeba energie v USA

Stacionarizace

Prvním krokem naší analýzy je zjištění, zda řada je řadou stacionární či nestacionární. Pro potvrzení stacionarity se používají nejrůznější testy, ovšem na první pohled je z grafu zřejmé, že se jedná o řadu nestacionární. Navíc je i vidět výraznou sezónnost. Vzhledem k tomu, že sezónní výkyvy mají rostoucí tendenci, bude vhodné řadu stabilizovat logaritmickou transformací danou vztahem

$$x_t^* = \ln x_t,$$

kde x_t jsou hodnoty původní řady. Transformovaná řada x_t^* zobrazená na obr.A.1 vykazuje při pohledu na její autokorelační funkci obr.A.2 a parciální autokorelační funkci obr.A.3 nám jen potvrzují sezónní charakter. Vzhledem k relativně blízkým (prvním) hodnotám jedné aplikujeme první diference. Také je vidět v bodech 12 a 24 nutné použití sezónních diferencí. Tyto diference jsou dány vztahem

$$x_t^{**} = x_t^* - x_{t-p}^*,$$

kde p udává místo potřebné difference, tedy nabývá p hodnot 1 a poté 12.

Po této transformaci a následných diferencích už dostáváme aspoň na první pohled řadu stacionární obr.A.4 . Také její ACF a PACF ukazují, že jsme se zbavili všech silných závislostí. Pro ověření stacionarity byl použit KPSS test, který je založen na testování nulté hypotézy H_0 :řada je stacionární. Testovaná statistika má hodnotu

$$T = 0.3697 < 0.463 = k_{0,05} < 0.739 = k_{0,01},$$

kde k_α je kritická hodnota testu na hladině významnosti α .

Odhad a verifikace modelu

Nyní přistoupíme k dalšímu kroku, kterým je odhad parametrů modelu. Při zkoumání vhodnosti různých modelů jsme nakonec došli k modelu SARIMA(1, 1, 1)(1, 1, 1)₁₂.

Toto zkoumání probíhá následovně. Nejprve si pro již stacionární řadu vyzkoušíme proces AR(1). Na základě hodnot ACF a PACF pro rezidua zjistíme, že hodnoty těchto funkcí neleží v intervalu spolehlivosti, a tedy jsou zde pozorované jisté závislosti. Proto zkusíme přidat proces MA(1) a opět vyhodnotíme ACF a PACF. Přitom zde nesledujeme pouze reziduální ACF a PACF, ale nesmíme také opomenout statistickou významnost odhadnutých parametrů, které byly důkladně popsány v kapitole číslo 3. Dospěli jsme tedy k našemu odhadnutému modelu, kde jeho odhadnuté parametry můžeme vidět na obr.A.6.

Pokračujeme v testování statistické významnosti parametrů, které je založeno na t-testech, viz. str.20. Například pro parametr $p = 1$ je

$$T_{\phi_1} = \frac{\hat{\phi}_1}{\hat{S}_{\phi_1}} = \frac{0,557061}{0,087961} = 6,33302 > t = 1,96,$$

kde t je 95% kvantil rozdělení t , tj. 1,96. V programu STATISTICA je užito následné označení

$$\begin{array}{ll} \hat{\phi}_1 & \dots\dots \text{Param.} \\ \hat{S}_{\phi_1} & \dots\dots \text{Asympt.SmCh} \\ T_{\phi_1} & \dots\dots \text{Asympt.t.} \end{array}$$

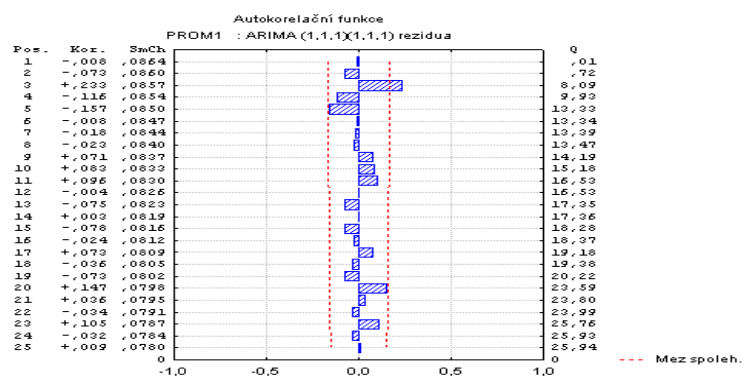
Výsledek nám indikuje statistickou významnost tohoto parametru. Pro ostatní parametry se tyto hodnoty počítají stejným způsobem. Uvedené výsledky jsou zobrazeny na obr.A.6. Statisticky významné hodnoty jsou uvedeny červenou barvou. Další možností testování statistické významnosti je Portmanteaův test, který ověřuje nekorelovanost reziduí. Na obr.5.2 a obr.5.3 jsou vidět výsledné hodnoty Q , ACF a PACF reziduí. Je vidět, že hodnoty Q Portmanteava testu nepřesahují kritickou hodnotu

$$\chi^2(23) = 35,172,$$

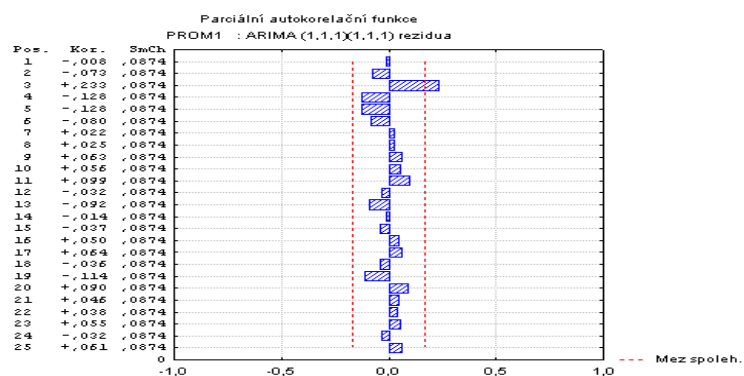
kde číslo 23 udává počet stupňů volnosti, viz.str.20. Tento test nám tedy potvrzuje správně vybraný model. Jak je vidět na tvarech ACF a PACF hodnoty jsou též statisticky významné, sice hodnota v bodě 3, značí jistou závislost, ale přesto tento odhad modelu je pro danou řadu tím nejlepším možným.

Předpověď modelu

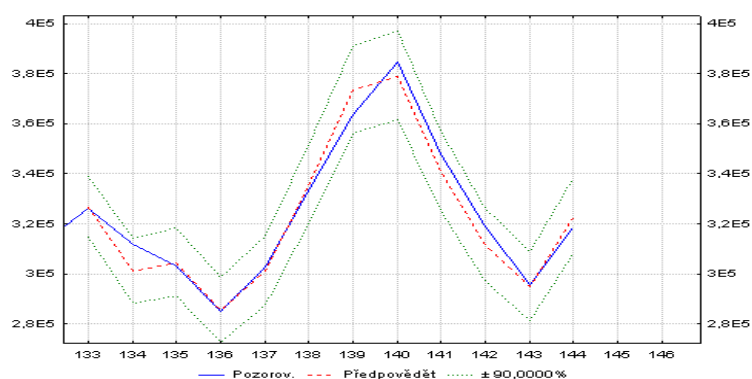
Následně si vypočítáme předpověď pro tuto řadu. Pro naše účely nebudeme předpokládat do budoucna. Předpovíme si poslední rok, abychom mohli srovnávat chybu našeho modelu s opravdovými reálnými hodnotami. Graf předpovědi vidíme na obr.5.4 a konkrétní hodnoty s intervaly spolehlivosti nalezneme na obr.A.5



Obrázek 5.2: Hodnoty ACF reziduí řady USA



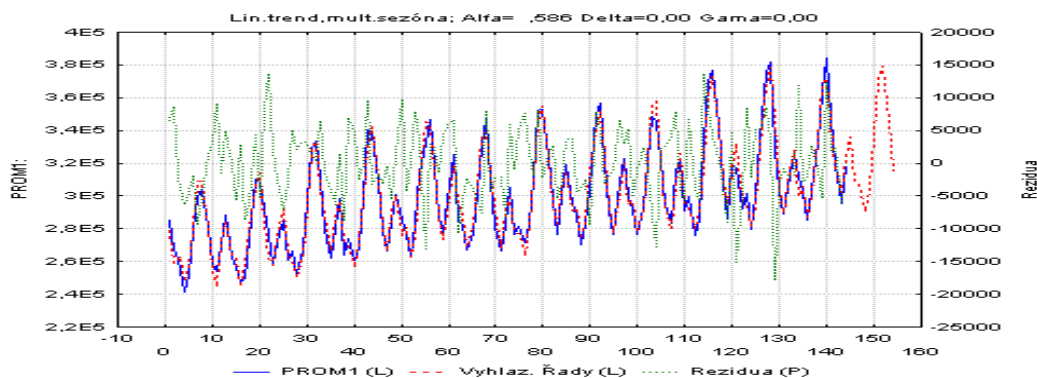
Obrázek 5.3: Hodnoty PACF reziduí řady USA



Obrázek 5.4: Předpověď řady USA

Exponenciální vyrovnávání

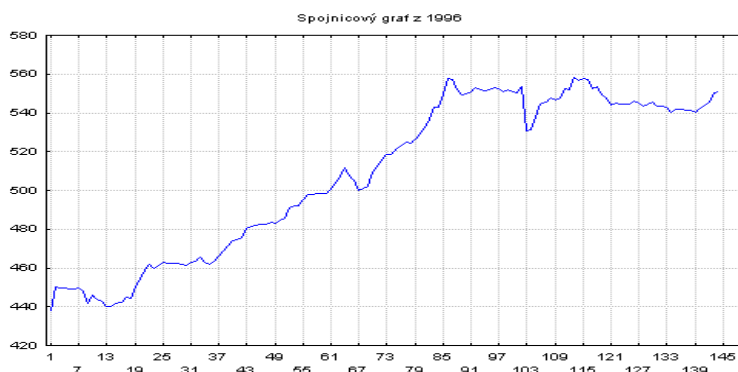
Postup popsany v kapitole o dekompozici časových řad a tvar energetické řady USA nás vede k využití multiplikativních vztahů Wintersovy metody, které nám odhadnou řadu v podobě obr.5.5. Využili jsme toho, že program STATISTICA při odhadech zmiňovanou metodou již sám automaticky odhaduje všechny parametry.



Obrázek 5.5: Předpověď řady USA

5.2. Ekonomická řada

Ekonomickou řadou jsme zvolili vývoj indexu PX50. Dlouho jsme hledali ekonomickou řadu, která by vyhovovala našim požadavkům, a taky byla predikovatelná jedním ze základních modelů. V dnešní době existuje velká škála modelů odvozených ze základních B–J procesů, kterými se také dnešní ekonomické řady, zejména vývoje indexu, predikují. Proto jsme museli vzít z naší řady pouze hodnoty pro první půlrok, obr.5.6.



Obrázek 5.6: Vývoj indexu PX50

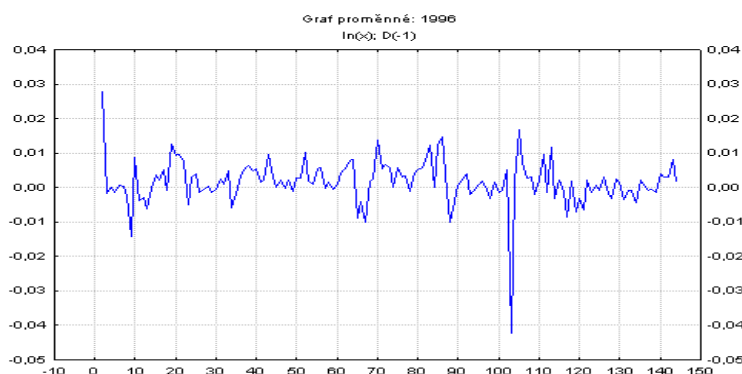
Stacionarizace

Opět se podíváme na identifikační proces vhodného modelu pro naši ekonomickou

řadu. Postup je zcela obdobný jako byl popsán výše. Z grafu je zcela zřejmé, že se nejedná o řadu stacionární. Je vidět lineární trend, a proto aplikujeme logaritmickou transformaci, která nám dává transformovanou řadu, obr.A.7 . Při pohledu na hodnoty funkcí ACF a PACF transformované řady obr.A.8 a obr.A.9 . Hodnoty v prvních hodnotách, které jsou blízké jedné nás evokují k použití první diference. Po těchto úpravách obdržíme již řadu stacionární obr.5.7, jak nám potvrzuje i KPSS test, který nabývá hodnot

$$T = 0.3570 < 0.463 = k_{0,05} < 0.739 = k_{0,01},$$

kde k_α je kritická hodnota testu na hladině významnosti α .



Obrázek 5.7: Řada PX50 po transformaci a diferenci

Odhad a verifikace modelu

Dále pokračujeme, jak bylo popsáno výše, tedy začínáme s analýzou ACF a PACF modelu ARIMA(0,1,0) a dostáváme se k odhadu modelu ARIMA(1,1,1). Následně odhadnuté parametry modelu prozkoumáme z hlediska statistické významnosti. Tedy statistiky pro parametry mají následující tvar

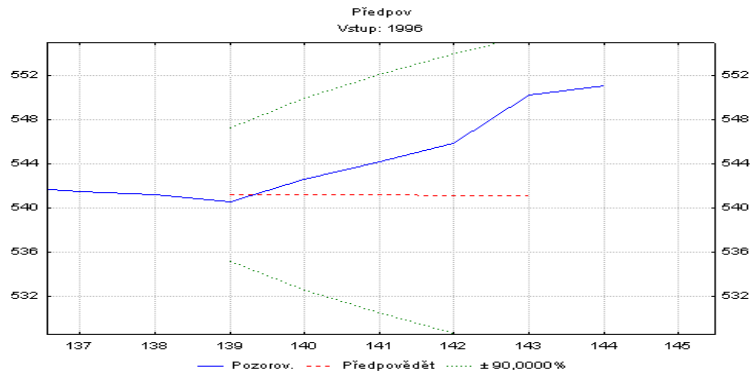
$$T_{\phi_1} = \frac{\hat{\phi}_1}{\hat{S}_{\phi_1}} = \frac{0,988653}{0,020494} = 48,24091 > t = 1,96,$$

$$T_{\theta_1} = \frac{\hat{\theta}_1}{\hat{S}_{\theta_1}} = \frac{0,958180}{0,035825} = 26,74640 > t = 1,96.$$

Na obr.A.12 můžeme vidět hodnoty odhadnutých parametrů. Také Portmanteauův test a hodnoty ACF a PACF reziduí ukazují obr.A.11 a obr.A.10, že náš odhadnutý model je ten správný. Kritická hodnota testu je $\chi^2(23) = 35,172$.

Předpověď modelu

Předpovězené hodnoty jsou opět použity pro poslední týden, abychom mohli vypočtené předpovědi opět porovnat se skutečnými hodnotami. Hodnoty předpovědí včetně jejich

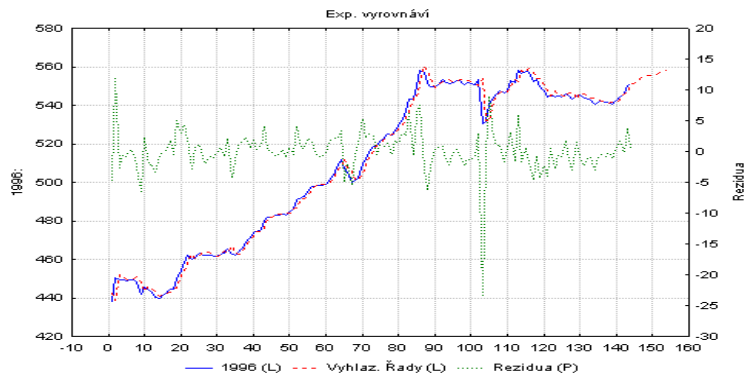


Obrázek 5.8: Předpověď řady PX50

intervalů spolehlivosti jsou vyobrazeny na obr.A.13.

Exponenciální vyrovnávání

Exponenciálním vyrovnáváním jsme dospěli k těmto hodnotám, k tomuto typu předpovědi bylo již napsáno mnoho, tudíž je zde ukázan pouze průběh předpovědi na obr.5.9. My budeme toto předpovězení potřebovat pouze k porovnání s B–J procesy, resp. odhadnutých modelů.



Obrázek 5.9: Exponenciální vyrovnání řady PX50

5.3. Testovací funkce

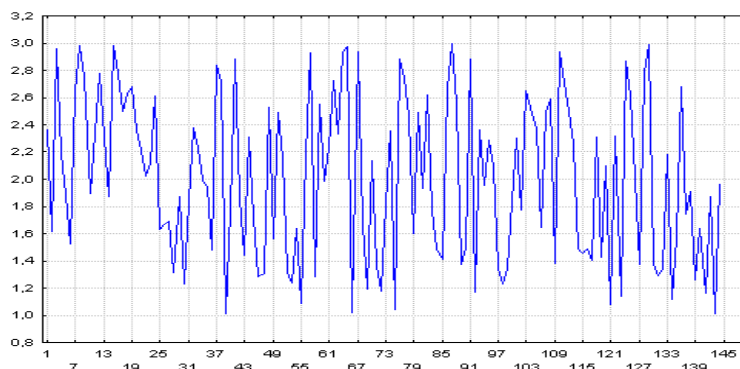
Pro naše účely jsme zvolili testovací funkci jako funkci sinu a náhodné složky ve tvaru

$$x(t) = (1 - p)s(t) + (p)(n_t), \quad \text{kde } t = 1, \dots, 144 \quad (5.3.1)$$

za předpokladu, že funkce sinu $s(t)$ je dána vztahem

$$s(t) = 2 + \sin\left(\frac{\pi t}{6}\right)$$

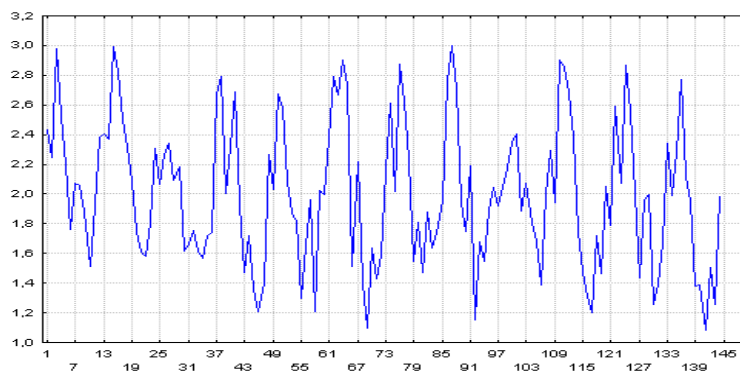
a funkce náhodné složky n_t tvoří výběr s rovnoměrným rozdělením v rozmezí od 1 do 3, jak je vidět na obr.5.10.



Obrázek 5.10: Náhodná složka

Parametr p nám udává v podobě $100p\%$ procentuální zastoupení náhodné složky. Pro naše účely bude naprosto postačující, aby tento parametr nabýval pouze hodnot $p = 0.1, 0.2, \dots, 1$. Kupříkladu testovací funkce s 50% zastoupením náhodného vlivu je znázorněna na obr.5.11, vzhled všech funkcí nalezneme v příloze na str.46.

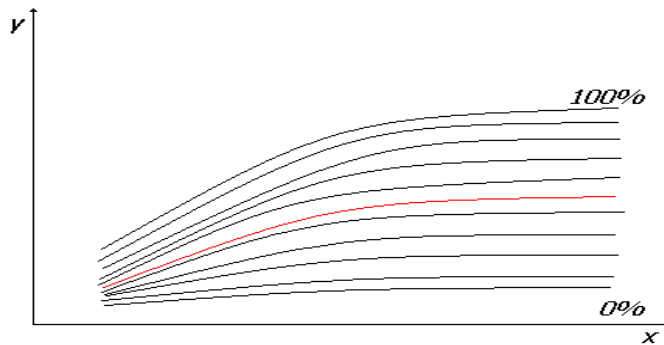
Takto zvolenou testovací funkci jsme volili s ohledem na podmínky B–J modelů, které jsme uváděli v kapitole o těchto modelech. Jak je vidět v definici vztahu 5.3.1, museli jsme tuto řadu posunout. Je to z důvodu, abychom se vyhlí případným komplikacím při výpočtech a finálnímu porovnávání chyb predikce, kde by nám mohlo hrozit dělení číslem blízkým nule, což by vedlo k zavádějícím výsledkům a následným špatným interpretacím našich závěrů.



Obrázek 5.11: Sinus s 50% zastoupením náhody

Při zkoumání a testování vhodnosti různých modelů na tyto řady se nám potvrdil náš předpoklad, že se změnami parametrů modelů ARIMA a SARIMA, které jsme na těchto řadách testovali, dostáváme zcela rozdílné výsledky, což není až tak překvapující zjištění.

Výsledné modely jsou uvedeny v tabulce na str.50 . My se bude chtít snažit ukázat, zda-li existují nějaké vazby, jinými slovy zda se nějakým způsobem dají tyto testovací funkce porovnat s reálnými řadami s rozumným výsledkem. Na obr.5.12 je vidět hlavní cíl našeho experimentu.



Obrázek 5.12: Myšlenka eperimentu

Kde osa x znázorňuje počet předpovídaných hodnot a osa y znázorňuje chybu našich předpovědí vůči reálné hodnotě danou vztahem

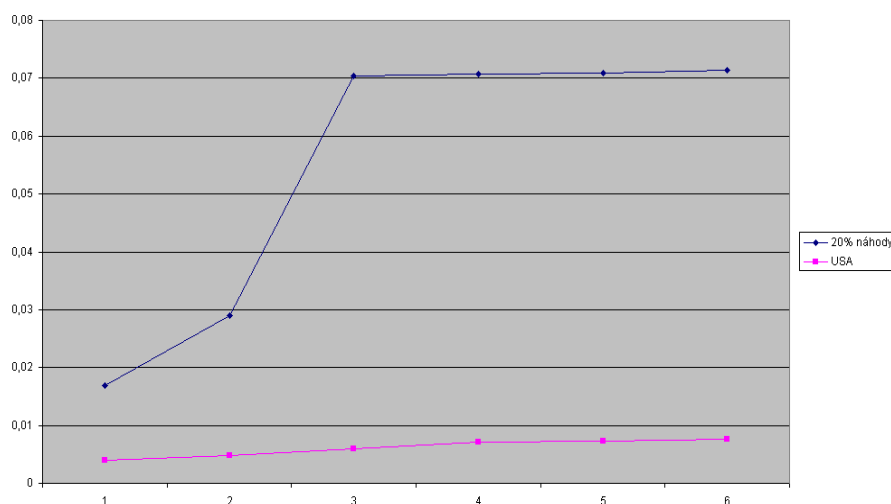
$$y_i = \frac{\sqrt{\sum_{j=0}^i \frac{x_j - x'_j}{x_j}}}{n}, \quad i = 1, \dots, n, \quad (5.3.2)$$

kde x'_i je předpovězená hodnota v i -tém bodě, resp. čase a x_i její reálná hodnota. Tyto hodnoty pro naše testovací funkce jsou zvýrazněny černou barvou a červenou barvou jsou zvýrazněny hodnoty reálné řady. V našem případě jde buď o řadu energetickou nebo ekonomickou. Na základě toho, kde bude hodnota „červené funkce“ ležet, bychom mohli vyslovit závěr, že daná řada obsahuje jisté zastoupení náhodnosti. Na našem ilustračním obrázku by se jednalo o 40%-50% zastoupení náhodné složky.

Kapitola 6

Závěr

V této kapitole vyhodnotíme veškeré naměřené hodnoty a pokusíme se buď potvrdit naši teorii, popsanou v minulé kapitole, a nebo tuto teorii vyvrátit. Bylo zjištěno jisté pravidelné chování při stejných parametrech. To nás evokovalo k následnému teoretickému předpokladu. Odhadneme parametry modelu k reálným časovým řadám, samozřejmě za podmínky statistické významnosti a vytvoříme jejich predikci. Tuto predikci bychom mohli použít ke srovnání s predikcí testovacích funkcí, ovšem za předpokladu toho, že se nám podaří najít statisticky významný model o stejných parametrech pro některé z testovacích funkcí.



Obrázek 6.1: Graf testovací a reálné funkce USA

Na obr.A.24 je vidět přehled všech funkcí, jak testovacích tak reálných, vyjádřené naším krtitériem dané vztahem 5.3.2. Jak je z grafu zřejmé, existuje zde jistá zákonitost, ovšem při podrobnějším rozboru jsme zjistili, že některé funkce kolidují s naší předpokládanou teorií. Při pohledu na reálné řady je zajímavé, že jejich hodnoty si jsou tak blízké.

Příčinou je zřejmě také fakt, že všechny funkce nejsou odhadnuté stejným modelem (např. dvě funkce odhadnuté modelem ARMA(1,1)). U funkcí, které byly odhadnuty stejnými modely je vidět, že naše teorie by mohla platit. Je to vidět u řady USA, kterou jsme odhadli modelem SARIMA(1, 1, 1)(1, 1, 1)₁2 a testovacích funkcí zastoupených 20%, 30% a 50% náhodností. Výsledkem je vyslovení hypotézy, že podle našich výsledků by mělo být zastoupení náhodnosti v řadě USA menší než 20% obr.6.1. Této hypotéze nasvědčuje i samotný průběh řady USA, který, jak je vidět výše, je relativně zřejmý. Jasný průběh řady s relativně malým (dle naší hypotézy s méně než 20%) zastoupením náhodnosti a v podstatě s velkou pravidelností má opakující průběh.

Náš závěr je tedy kladný, a to, že naše hypotéza by mohla platit, leč za opravdu nelehkých podmínek, a to, že nalezneme testovací funkce, které jsou odhadnuty stejným modelem, který je samozřejmě statisticky významný. Vzhledem k náročnosti nalezení správného modelu a relativně velice dobrým výsledkům našeho experimentu bychom navrhovali pokračování, které by se pokusilo potvrdit naši hypotézu nalezením testovací funkce pro jednu konkrétní reálnou časovou řadu. Druhá možnost by také mohla být, že podobný test bychom mohli aplikovat na metodu dekompozice, protože jak ukazuje obrázek A.25, předpovědní chyby obou metod si jsou celkem blízké.

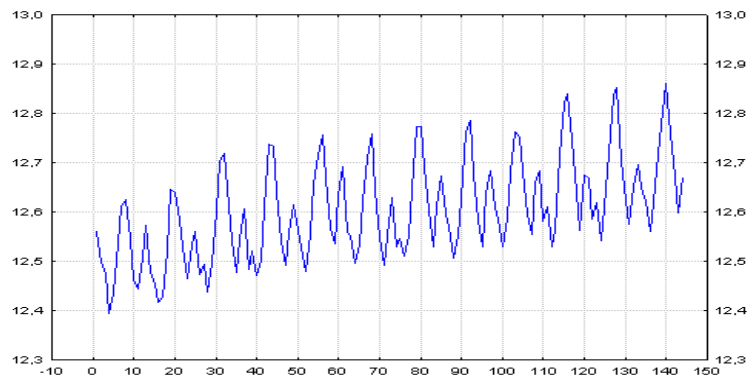
Literatura

- [1] George.E.P. Box, Gwilym.M. Jenkins, Gregory.C. Reinsel: *Time series analysis: Forecasting and control*. 3.vyd., Prentice Hall, Inc. (1994), ISBN 0-13-060774-6.
- [2] Douglas C. Montgomery, Lynwood A. Johnson, John S. Gardiner: *Forecasting and time series analysis*. 2.vyd., McGraw-Hill, Inc. (1990), ISBN 0-07-042858-1.
- [3] J. Anděl: *Matematická statistika*, 1.vyd., SNTL (1978).
- [4] J. Anděl: *Statistická analýza časových řad*, 1.vyd., SNTL (1976).
- [5] J. Arlt, M. Arltová: *Ekonomické časové řady*, 1.vyd., Grada Publishing,a.s. (1978), ISBN 978-80-247-1319-9.
- [6] T. Cipra: *Analýza časových řad s aplikacemi v ekonomii*, 1.vyd., SNTL (1986)

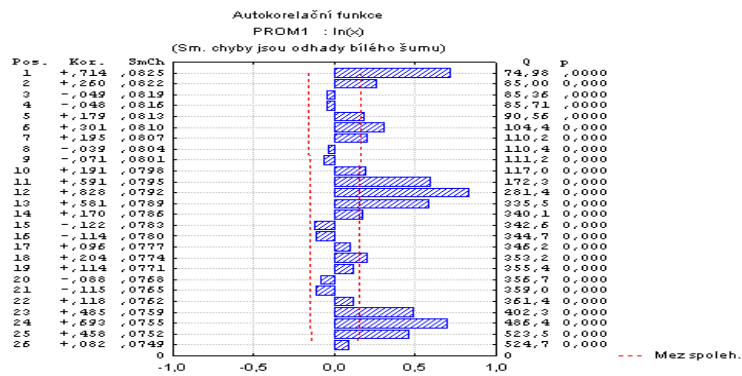
Příloha A

Grafy a data

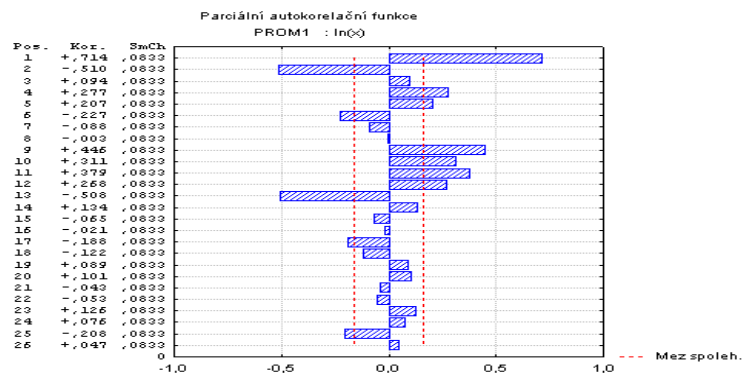
Zde jsou uvedeny veškeré grafy včetně dat, na které se odkazují ve své práci.



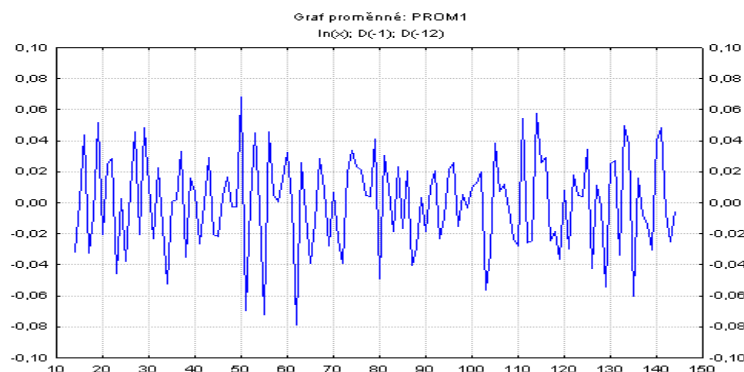
Obrázek A.1: Řada USA po logaritmské transformaci



Obrázek A.2: ACF transformované řady USA



Obrázek A.3: PACF transformované řady USA



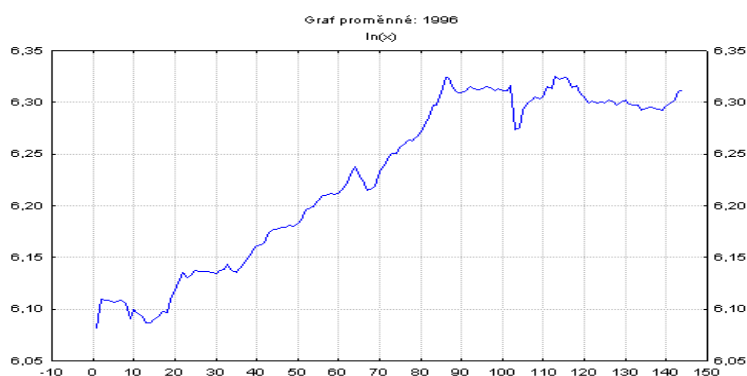
Obrázek A.4: Řada USA po transformaci a diferencích

Předpovědi; Model:(1,1,1)(1,1,1) Sezónní posun:12 (USA.sta) Vstup: PROM1 Začátek původ. : 1 Konec původ. : 132					
ČísloPřip.	Předpově	Dolní 90,0000%	Horní 90,0000%	Pozorov.	Residuál
133	326594,9	314658,5	338984,2	326182,2	-412,80
134	301062,4	288352,7	314332,3	311656,9	10594,50
135	304499,7	291072,5	318546,3	303013,0	-1486,71
136	285456,1	272658,6	298854,3	285056,0	-400,17
137	300934,5	287341,8	315170,2	302943,1	2008,64
138	335409,0	320197,9	351342,7	333198,3	-2210,70
139	373376,1	356399,4	391161,5	363568,7	-9807,33
140	378924,6	361662,6	397010,6	384648,1	5723,42
141	340958,8	325401,6	357259,7	348171,1	7212,33
142	311324,1	297098,9	326230,3	318883,7	7559,64
143	294900,2	281407,7	309039,7	295952,5	1052,31
144	322700,9	307917,6	338194,0	318431,8	-4269,07

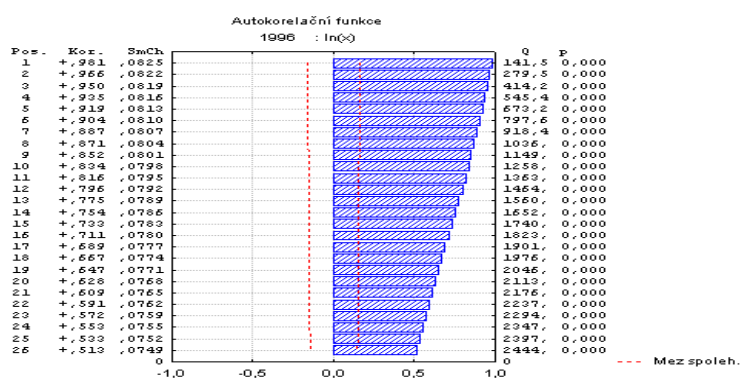
Obrázek A.5: Předpověď řady USA

Vstup: PROM1 (USA.sta)						
Transformace: ln(x),D(1),D(12)						
Model:(1,1,1)(1,1,1) Sezónní posun: 12 PČ Rezid. = ,00049						
Paramet.	Param.	Asympt. SmCh	Asympt. t(127)	p	Dolní 95% spol	Horní 95%
p(1)	0,557061	0,087961	6,33302	0,000000	0,383001	0,731120
q(1)	0,972202	0,032219	30,17477	0,000000	0,908447	1,035958
Ps(1)	0,426035	0,028753	0,20220	0,000080	0,387453	0,450281
Qs(1)	0,738990	0,084477	8,74783	0,000000	0,571826	0,906155

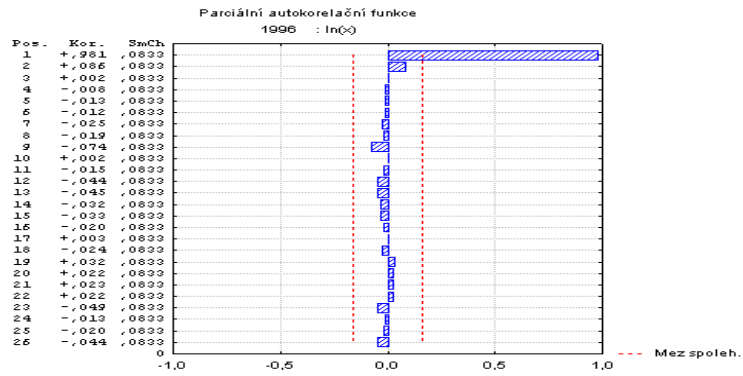
Obrázek A.6: Odhady parametrů modelu SARIMA řady USA



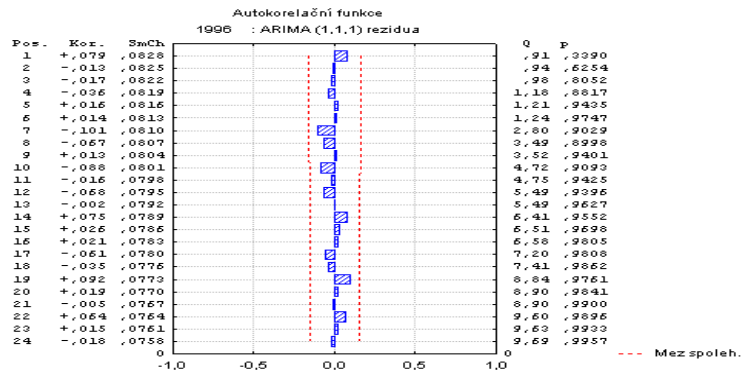
Obrázek A.7: Řada PX50 po logaritmické transformaci



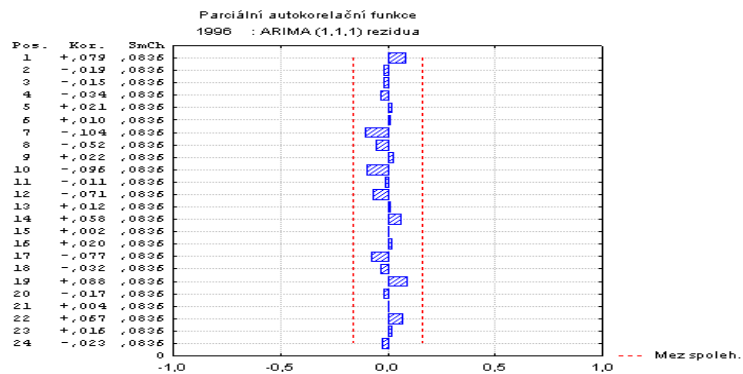
Obrázek A.8: ACF transformované PX50



Obrázek A.9: PACF transformované PX50



Obrázek A.10: ACF residuí PX50



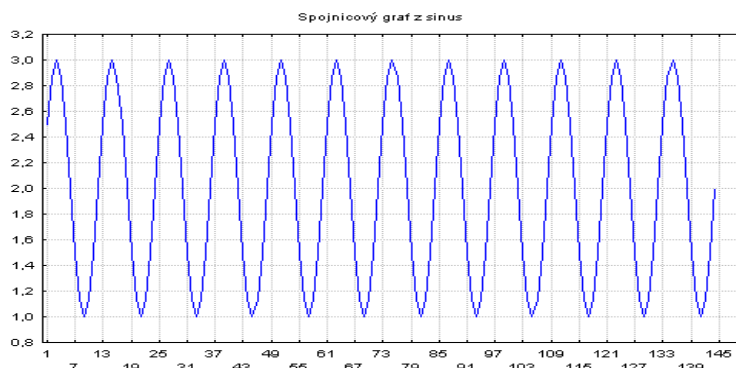
Obrázek A.11: PACF residuí PX50

Vstup: 1996 (PX50.sta)						
Transformace: ln(x),D(1)						
Model:(1,1,1)						
Paramet.	Param.	Asympt. SmCh	Asympt. t(141)	p	Dolní 95%	Horní 95%
p(1)	0,988653	0,020494	48,24091	0,00	0,948138	1,029168
q(1)	0,958180	0,035825	26,74640	0,00	0,887357	1,029003

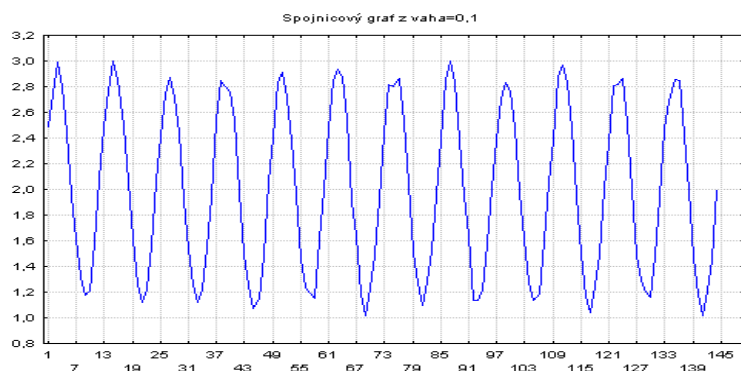
Obrázek A.12: Odhady modelu ARIMA(1,1,1) řady PX50

Předpovědi; Model:(1,1,1) Sezónní posun:12 (PX50.sta)					
Vstup: 1996					
Začátek původ. : 1 Konec původ. : 138					
ČísloPřip.	Předpově	Dolní 90,0000%	Horní 90,0000%	Pozorov.	Residuál
139	541,1831	535,1777	547,2559	540,5000	-0,683124
140	541,1664	532,5643	549,9075	542,6000	1,433560
141	541,1499	530,4768	552,0379	544,2000	3,050054
142	541,1336	528,6486	553,9135	545,8000	4,666361
143	541,1175	526,9785	555,6359	550,2000	9,082482

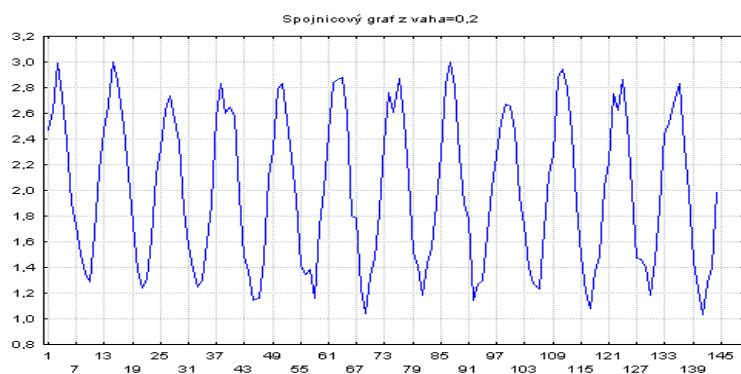
Obrázek A.13: Hodnoty předpovědí řady PX50



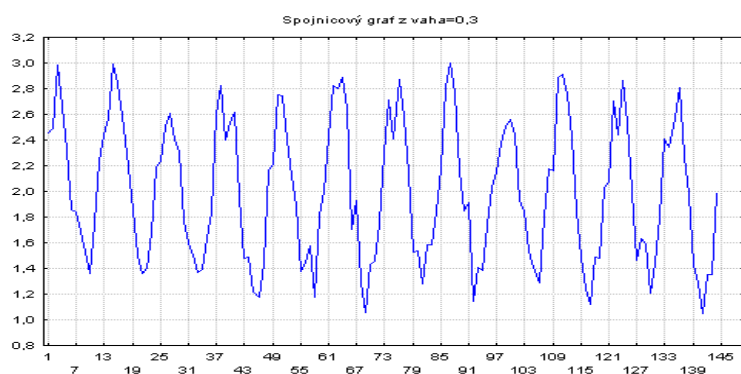
Obrázek A.14: Testovací funkce 0% náhodné složky



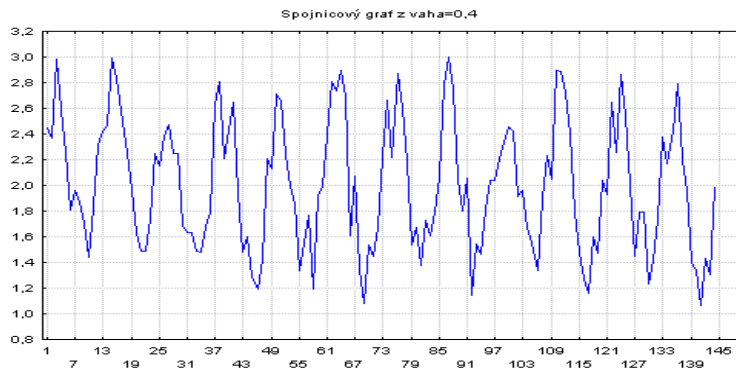
Obrázek A.15: Testovací funkce 10% náhodné složky



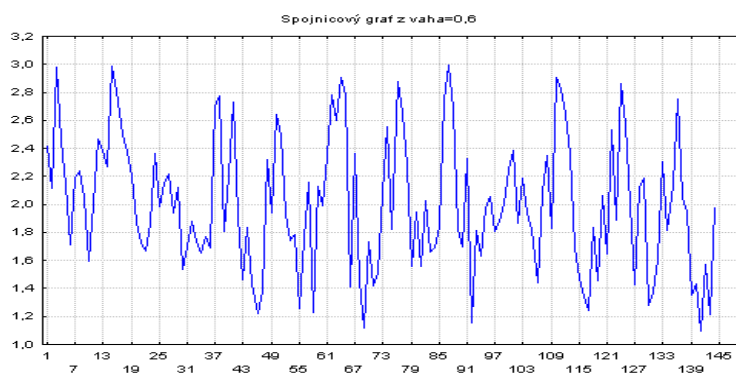
Obrázek A.16: Testovací funkce 20% náhodné složky



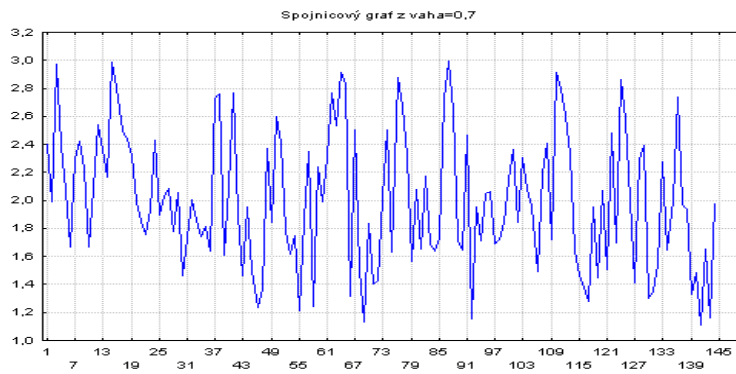
Obrázek A.17: Testovací funkce 30% náhodné složky



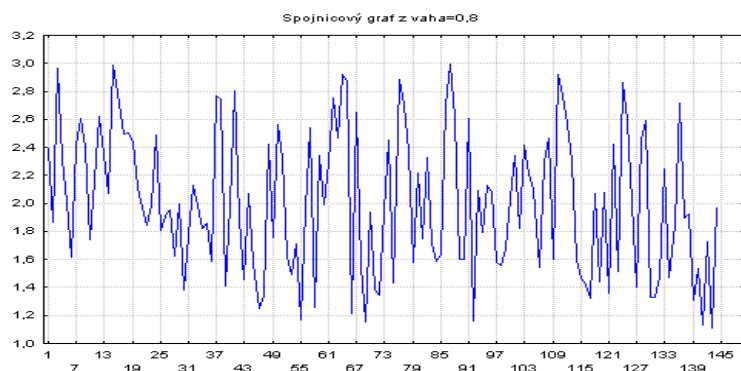
Obrázek A.18: Testovací funkce 40% náhodné složky



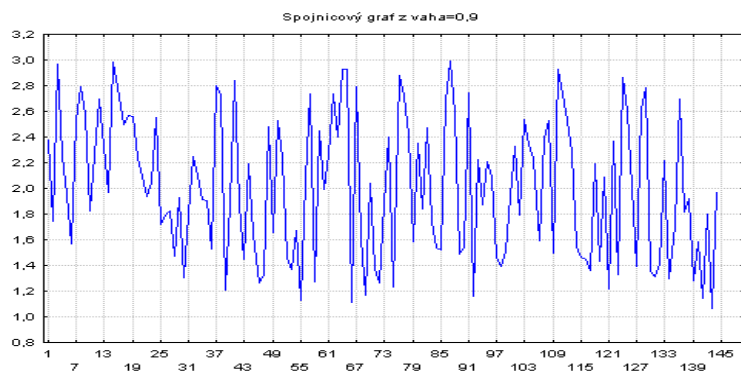
Obrázek A.19: Testovací funkce 60% náhodné složky



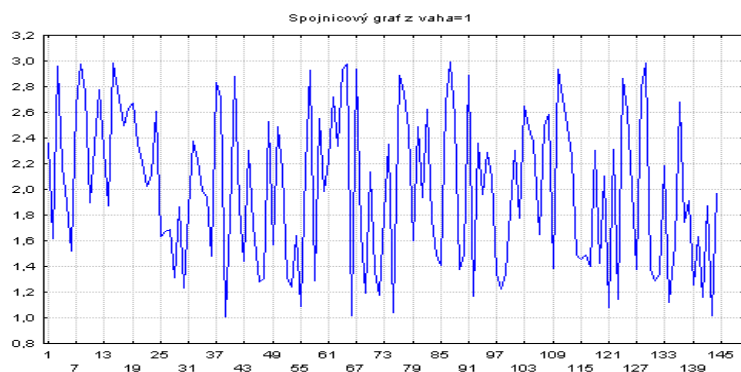
Obrázek A.20: Testovací funkce 70% náhodné složky



Obrázek A.21: Testovací funkce 80% náhodné složky



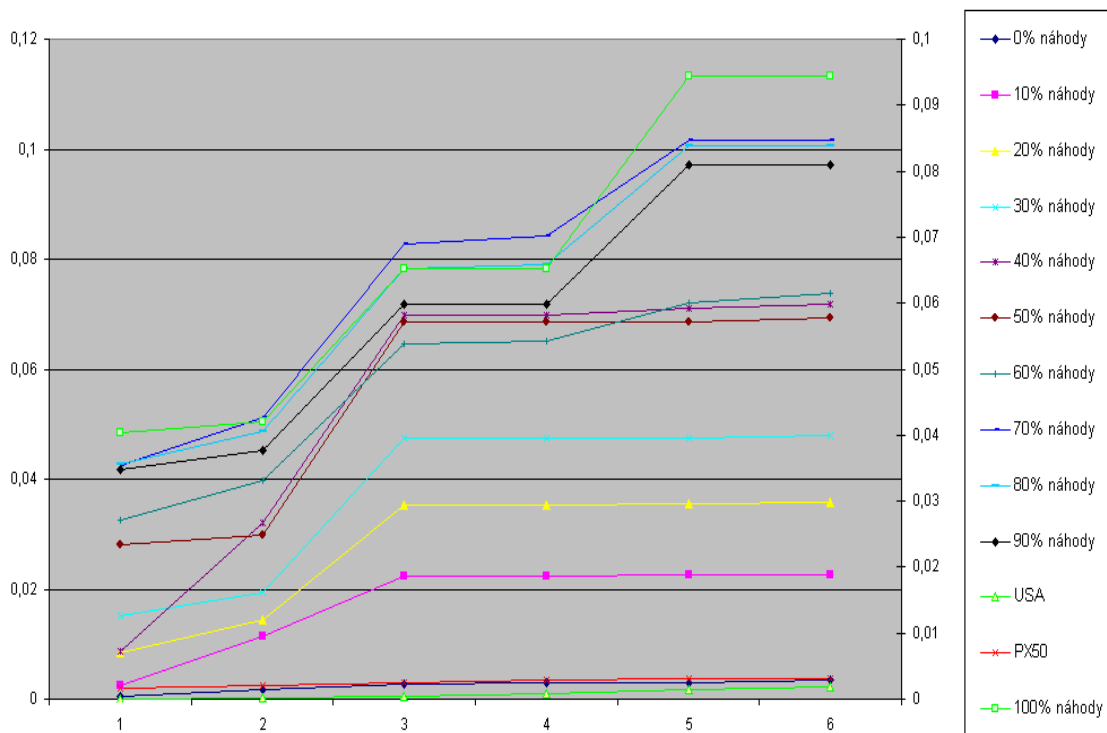
Obrázek A.22: Testovací funkce 90% náhodné složky



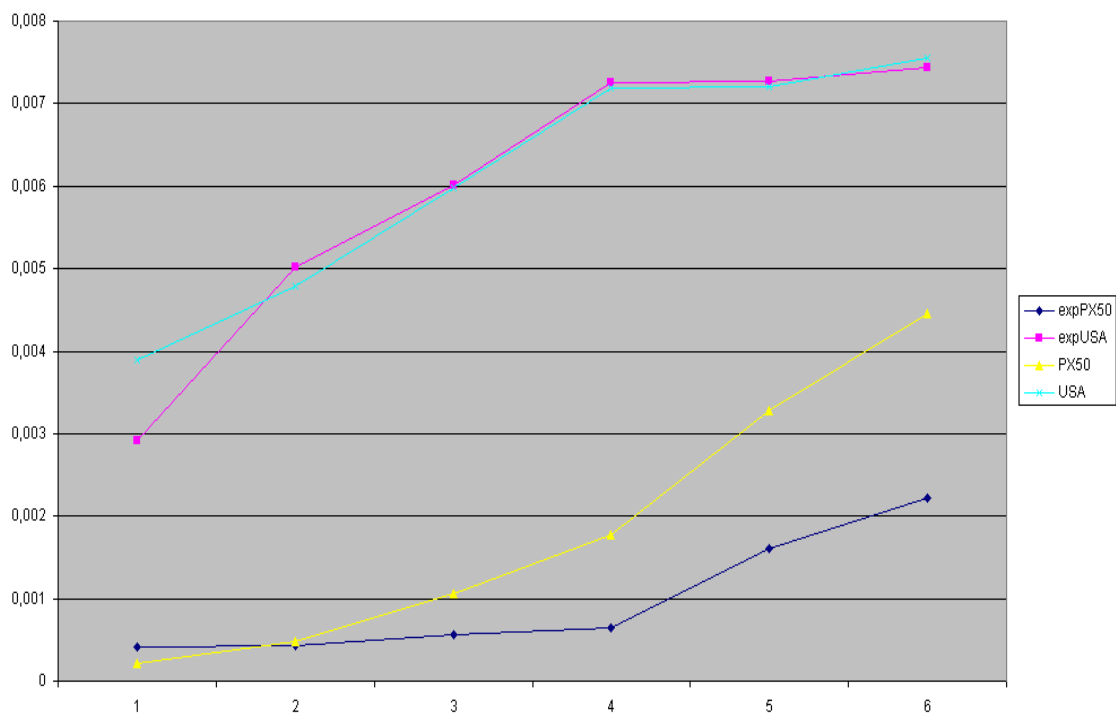
Obrázek A.23: Testovací funkce 100% náhodné složky

Funkce	Model
0% náhody	SARIMA(1,1,1)(1,1,0)
10% náhody	SARIMA(1,1,0)(1,1,0)
20% náhody	SARIMA(1,1,1)(1,1,1)
30% náhody	SARIMA(1,1,1)(1,1,1)
40% náhody	SARIMA(1,1,1)(1,1,0)
50% náhody	SARIMA(1,1,1)(1,1,1)
60% náhody	ARIMA(1,0,0)
70% náhody	ARIMA(1,0,1)
80% náhody	ARIMA(1,0,1)
90% náhody	ARIMA(1,0,1)
100% náhody	ARIMA(1,0,1)

Tabulka A.1: Odhadnuté modely pro testovací funkce



Obrázek A.24: Grafy testovacích a reálných funkcí



Obrázek A.25: Porovnání chyb předpovědí