

UNIVERZITA PALACKÉHO V OLOMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA  
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

## DIPLOMOVÁ PRÁCE

ARMA modely časových řad



Vedoucí diplomové práce:  
**RNDr. et PhDr. Ivo Müller, Ph.D.**  
Rok odevzdání: 2010

Vypracovala:  
**Bc. Jitka Šírová**  
AME, II. ročník

### **Prohlášení**

Prohlašuji, že jsem vytvořila tuto diplomovou práci samostatně za vedení RNDr. et PhDr. Ivo Müllera, Ph.D. a že jsem v seznamu použité literatury uvedla všechny zdroje použité při zpracování práce.

V Olomouci dne 2. dubna 2010

## **Poděkování**

Ráda bych na tomto místě poděkovala vedoucímu diplomové práce RNDr. et PhDr. Ivu Müllerovi, Ph.D. za obětavou spolupráci i za čas, který mi věnoval při konzultacích a při opakovaném pročítání vznikajícího textu.

# Obsah

<b>1</b>	<b>Úvod</b>	<b>5</b>
<b>2</b>	<b>Základní pojmy</b>	<b>7</b>
2.1	Časová řada . . . . .	7
2.2	Autokorelační vlastnosti časových řad . . . . .	7
2.2.1	Stacionarita . . . . .	7
2.2.2	Autokovariance a autokorelace . . . . .	8
<b>3</b>	<b>Stacionární procesy</b>	<b>14</b>
3.1	Lineární proces . . . . .	14
3.2	Proces klouzavých součtů . . . . .	16
3.3	Autoregresní proces . . . . .	27
3.4	Smíšený proces . . . . .	36
<b>4</b>	<b>Výstavba modelů</b>	<b>42</b>
4.1	Identifikace modelu . . . . .	42
4.2	Odhad parametrů . . . . .	48
4.2.1	Podmíněná metoda nejmenších nelineárních čtverců . . . . .	48
4.2.2	Nepodmíněná metoda nejmenších nelineárních čtverců . . . . .	50
4.2.3	Minimalizace součtu nelineárních čtverců . . . . .	52
4.3	Ověřování modelu . . . . .	57
4.3.1	Metoda přeparametrizování modelu . . . . .	57
4.3.2	Normalita nesystematické složky . . . . .	58
4.3.3	Metoda odhadnutých reziduí . . . . .	59
4.3.4	Portmanteau test . . . . .	60
4.3.5	Durbinův-Watsonův test . . . . .	61
4.3.6	Oprava modelu při zamítnutí $ARMA(p,q)$ . . . . .	63
4.3.7	Výběr modelu . . . . .	63
<b>5</b>	<b>Rozšířené ARMA modely</b>	<b>66</b>
5.1	Integrované modely (ARIMA) . . . . .	66
5.1.1	Proces náhodné procházky ("random walk") . . . . .	68
5.1.2	Procesy nestacionární v rozptylu . . . . .	68
5.1.3	Výstavba modelu ARIMA – Shrnutí . . . . .	69
5.2	Sezónní modely . . . . .	69
5.2.1	Výstavba modelu $SARIMA(p, d, q) \times (P, D, Q)_s$ – Shrnutí . . . . .	70
<b>6</b>	<b>Konstrukce předpovědí</b>	<b>72</b>

<b>7</b>	<b>Boxova-Jenkinsova metodologie v praxi</b>	<b>76</b>
7.1	Modelování původní řady pouze pomocí Boxova-Jenkinsova přístupu	76
7.2	Modelování řady reziduí získané po odstranění trendu . . . . .	83
7.3	Modelování po částech . . . . .	89
7.4	Srovnání . . . . .	94
<b>8</b>	<b>Závěr</b>	<b>95</b>

# 1 Úvod

Při práci s časovými řadami často zkoumáme její systematické složky (trendovou, sezónní a cyklickou), přičemž předpokládáme nekorelovanost jednotlivých pozorování. Náhodná složka však může být tvořena korelovanými náhodnými veličinami. V takovém případě můžeme využít Boxovu-Jenkinsovu metodologii, jejíž postup spočívá právě v modelování časových řad tvořených korelovanými náhodnými veličinami.

Modely získanými Boxovou-Jenkinsovou metodologií lze stochasticky modelovat nejen náhodnou složku, ale i trend a sezónnost. Zároveň rychle reagují na změnu charakteru dat. Další výhodou je, že se časová řada analyzuje pomocí předem daného algoritmu a subjektivní zásahy jsou minimální.

Ačkoli má tento přístup i své nevýhody, kterými jsou především požadovaná délka řady alespoň 50 pozorování, časová náročnost výpočtů, pro které je třeba statistický software, a obtížná interpretace modelů, patří Boxova-Jenkinsova metodologie mezi analytické metody s dobrými praktickými výsledky.

V dalším textu se předpokládá znalost základních pojmů z oblasti pravděpodobnosti a matematické statistiky.

## Použité symboly

$\mathbb{N}(\mathbb{R})[\mathbb{Z}]$	množina přirozených (reálných)[komplexních] čísel
$\mathbf{A}_k$	čtvercová matice typu $k \times k$
$ \mathbf{A} $	determinant matice $\mathbf{A}$
$X$	náhodná veličina $X$
$F(X)$	distribuční funkce náhodné veličiny $X$
$E(X)$	střední hodnota náhodné veličiny $X$
$\text{var}(X)$	rozptyl náhodné veličiny $X$
$\text{cov}(X, Y)$	kovariance náhodných veličin $X$ a $Y$
$\text{cor}(X, Y)$	korelační koeficient náhodných veličin $X$ a $Y$
$\bar{X}$	aritmetický průměr náhodné veličiny $X$
$H_0$	nulová hypotéza
$X \sim \mathcal{AN}(0, 1)$	náhodná veličina $X$ má asymptoticky normální rozdělení se střední hodnotou 0 a rozptylem 1
$\varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2)$	proces $\varepsilon_t$ je proces bílého šumu se střední hodnotou 0 a rozptylem $\sigma_\varepsilon^2$
$\chi_s^2(\alpha)$	$(1 - \alpha)$ -kvantil $\chi^2$ rozdělení s $s$ stupni volnosti
$\frac{\partial f(x,y)}{\partial x}$	parciální derivace funkce $f(x, y)$ podle proměnné $x$

## 2 Základní pojmy

Stěžejním pramenem teorie, kterou obsahuje druhá až patá kapitola, je literatura [3].

### 2.1 Časová řada

**Definice 2.1.** *Systém náhodných veličin*

$$\{X_t\}_{t \in T}, \quad T \subset \mathbb{R}, \quad (1)$$

*nazýváme náhodný proces.*

**Definice 2.2.** *Časovou řadu chápeme jako:*

1. *náhodnou posloupnost, tedy jako náhodný proces s diskrétním časem, tj.  $\{X_t\}_{t \in T}$ , kde  $T$  je spočetná množina ( např.  $T \subset \mathbb{Z}$  );  
nebo*
2. *konečnou verzi 1., tedy jako náhodnou posloupnost rozsahu  $n$ .*

**Definice 2.3.** *Existuje-li pro každé  $t \in T$  střední hodnota  $E(X_t)$ , potom funkci  $\mu_t = E(X_t)$  nazýváme střední hodnotou procesu.*

*Je-li střední hodnota nulová, nazýváme proces centrováný.*

### 2.2 Autokorelační vlastnosti časových řad

#### 2.2.1 Stacionarita

Řekneme-li o nějakém náhodném procesu, že je stacionární, znamená to, že chování tohoto procesu v čase je v jistém smyslu stochasticky ustálené. Rozlišujeme striktní a slabou stacionaritu.

**Definice 2.4.** *Náhodný proces se nazývá striktně stacionární, jestliže všechna jeho konečněrozměrná rozdělení jsou invariantní vůči posunům v čase, tj.*

$$F_{t_1, \dots, t_n}(x_{t_1}, \dots, x_{t_n}) = F_{t_1+h, \dots, t_n+h}(x_{t_1+h}, \dots, x_{t_n+h}) \quad (2)$$

$$\forall n \in \mathbb{N}, \quad \forall h \in \mathbb{R}, \quad \forall t_i \in T : t_{i+h} \in T, \quad i = 1, \dots, n.$$



**Definice 2.5.** *Náhodný proces se nazývá slabě stacionární, jestliže má konstantní střední hodnotu, konstantní rozptyl a kovarianční strukturu druhého řádu invari-  
antní vůči posunům v čase.*

Je-li proces striktně stacionární, je také slabě stacionární. Opačný vztah obecně neplatí.

**Definice 2.6.** *Jestliže jakékoli sdružené rozdělení pravděpodobností náhodného procesu je normální, potom tento proces nazýváme normální (gaussovský).*

Pokud je normální proces slabě stacionární, potom je také striktně stacionární, protože normální rozdělení je jednoznačně určeno prvními dvěma momenty. V dalším textu budeme předpokládat, že pracujeme s náhodnými veličinami, které mají normální rozdělení, a proto bude používán jen pojem stacionarita.

## 2.2.2 Autokovariance a autokorelace

Předpokládáme stacionární časovou řadu.

**Definice 2.7.** *Funkci*

$$\gamma_k := \text{cov}(X_t, X_{t+k}) = E(X_t - \mu)(X_{t+k} - \mu), \quad k \in \mathbb{Z}, \quad (3)$$

*nazýváme autokovarianční funkce časové řady.*

**Poznámka 2.1.** *Speciálně pro  $k = 0$  je  $\gamma_0 = \text{var}(X_t) = \sigma_X^2$ , což je rozptyl časové řady.*

**Definice 2.8.** *Funkci*

$$\rho_k := \frac{\gamma_k}{\gamma_0} = \frac{\text{cov}(X_t, X_{t+k})}{\sqrt{\text{var}(X_t)\text{var}(X_{t+k})}} = \text{cor}(X_t, X_{t+k}), \quad k \in \mathbb{Z}, \quad (4)$$

*nazýváme autokorelační funkce časové řady.*

Díky stacionaritě jsou autokovarianční i autokorelační funkce sudé, tedy  $\gamma_k = \gamma_{-k}$  a  $\rho_k = \rho_{-k}$ , proto se můžeme omezit jen na  $k \in \mathbb{N}$ .

Předpona *auto-* se používá proto, že kovarianci a korelaci počítáme pomocí veličin z téže časové řady.

Graf autokorelační funkce  $\varrho_k$  pro jednotlivá  $k$  se nazývá *korelogram*.

### Odhady autokovarianční a autokorelační funkce

Při práci s daty  $X_1, \dots, X_n$ , která tvoří časovou řadu ve smyslu 2. definice 2.1, budeme používat odhad

$$\bar{X} = \hat{E}(X_t) = \sum_{t=1}^n \frac{X_t}{n} \quad (5)$$

pro střední hodnotu,

$$c_k = \hat{\gamma}_k = \sum_{t=1}^{n-k} \frac{(X_t - \bar{X})(X_{t+k} - \bar{X})}{n}, \quad k = 0, 1, \dots, n-1, \quad (6)$$

pro autokovarianční funkci a

$$r_k = \hat{\varrho}_k = \frac{c_k}{c_0}, \quad k = 0, 1, \dots, n-1, \quad (7)$$

pro autokorelační funkci. Pro praktické výpočty bývá požadováno  $n > 50$  a  $k < \frac{n}{4}$ .

**Definice 2.9.** Časovou řadu  $X = \{X_t\}$  nazveme bílý šum se střední hodnotou  $\mu$  a rozptylem  $\sigma_X^2$ , jestliže  $\mu_t \equiv \mu$  a platí

$$\gamma_k = \begin{cases} \sigma_X^2, & \text{pro } k = 0, \\ 0, & \text{pro } k \neq 0, \end{cases} \quad (8)$$

Píšeme  $X \sim \mathcal{WN}(\mu, \sigma_X^2)$ .

**Poznámka 2.2.** Stacionární časová řada, která není bílým šumem, se někdy nazývá také barevný šum.

### Hledání identifikačního bodu

Podle průběhu autokorelační funkce identifikujeme příslušný model (viz. kapitola 4.1). Přitom je třeba nalézt tzv. *identifikační bod*  $k_0$ . To je taková hodnota, že

$$\varrho_k = 0, \quad k > k_0. \quad (9)$$

Protože teoretickou (skutečnou) autokorelační funkci  $\varrho_k$  neznáme, ale pouze ji odhadujeme pomocí funkce  $r_k$ , potřebujeme zjistit, jak blízko nule musí být odhadnutá autokorelační funkce  $r_k$ , abychom mohli s předem danou spolehlivostí tvrdit, že teoretická autokorelační funkce  $\varrho_k = 0$ . Testujeme proto hypotézu

$$H_0 : \quad \varrho_k = 0, \quad k > k_0.$$

Nejprve spočítáme pomocí Bartlettovy aproximace směrodatnou odchylku autokorelační funkce  $r_k$ . Za předpokladu normality zkoumaného náhodného procesu a pokud je  $k_0$  skutečný identifikační bod, tak platí:

$$\sigma(r_k) = \sqrt{\text{var}(r_k)} \stackrel{H_0}{\doteq} \sqrt{\frac{1}{n} \left(1 + 2 \sum_{j=1}^{k_0} r_j^2\right)}, \quad k > k_0. \quad (10)$$

Zvolíme  $k_0 = 0$  a pokud  $|r_k| > 2\sqrt{\text{var}(r_k)}$  pro nějaké  $k > k_0$ , zvětšíme  $k_0$  o jedničku ( $k_0 := k_0 + 1$ ) a znovu porovnáme. Pokud  $|r_k| < 2\sqrt{\text{var}(r_k)}$  pro všechna  $k > k_0$ , potom je  $k_0$  hledaný identifikační bod a hypotézu  $H_0 : \varrho_k = 0$  pro  $k > k_0$  nezamítáme.

**Poznámka 2.3.** V předchozím odstavci využíváme toho, že

$$\frac{r_k}{\sqrt{\text{var}(r_k)}} \sim \mathcal{AN}(0, 1), \quad (11)$$

a pravidlo, že normální náhodná veličina s nulovou střední hodnotou překročí v absolutní hodnotě dvojnásobek své směrodatné odchylky s pravděpodobností přibližně 0,05.

### Parciální autokorelační funkce

**Definice 2.10.** *Parciální autokorelační funkcí  $\varrho_{kk}$  rozumíme parciální korelační koeficient náhodných veličin  $X_t$  a  $X_{t+k}$  při pevných hodnotách  $X_{t+1}, \dots, X_{t+k-1}$ .*

Platí

$$\varrho_{kk} = \frac{|\mathbf{A}_k^*|}{|\mathbf{A}_k|}, \quad (12)$$

kde

$$\mathbf{A}_k = \begin{pmatrix} 1 & \varrho_1 & \dots & \varrho_{k-1} \\ \varrho_1 & 1 & \dots & \varrho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \varrho_{k-1} & \varrho_{k-2} & \dots & 1 \end{pmatrix}$$

a matice  $\mathbf{A}_k^*$  vznikne z matice  $\mathbf{A}_k$  změnou posledního sloupce

$$\mathbf{A}_k^* = \begin{pmatrix} 1 & \varrho_1 & \dots & \varrho_1 \\ \varrho_1 & 1 & \dots & \varrho_2 \\ \vdots & \vdots & \ddots & \vdots \\ \varrho_{k-1} & \varrho_{k-2} & \dots & \varrho_k \end{pmatrix}$$

Pro první dvě hodnoty parciální autokorelační funkce  $\varrho_{kk}$  tedy platí

$$\varrho_{11} = \varrho_1$$

$$\varrho_{22} = \frac{\begin{vmatrix} 1 & \varrho_1 \\ \varrho_1 & \varrho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \varrho_1 \\ \varrho_1 & 1 \end{vmatrix}} = \frac{\varrho_2 - \varrho_1^2}{1 - \varrho_1^2}.$$

Odhady hodnot funkce  $\varrho_{kk}$  počítáme pomocí rekurentních vzorců

$$r_{11} = r_1 \quad (13)$$

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j}, \quad k > 1, \quad (14)$$

kde  $r_{k,j} = r_{k-1,j} - r_{kk} r_{k-1,k-j}$  pro  $j = 1, \dots, k-1$ . Ani v tomto případě neznáme hodnoty skutečné parciální autokorelační funkce  $\varrho_{kk}$ , proto testujeme nulovost

jejích hodnot. K tomu použijeme následující hypotézu

$$H_0 : \quad \varrho_{kk} = 0, \quad k > k_0.$$

Vycházíme z Quenouilleovy aproximace, podle které je

$$\sigma(r_{kk}) \stackrel{H_0}{\doteq} \sqrt{\frac{1}{n}}, \quad k > k_0. \quad (15)$$

Dále postupujeme stejně jako u předchozí hypotézy pro autokorelační funkci.

### Příklad 2.1.

V tabulce 1 je po řádcích uveden kurz české koruny vůči australskému dolaru po dobu 11 pracovních týdnů v době od 30. června 2008 do 12. září 2008.

14,597	14,425	14,494	14,427	14,554	14,392	14,324	14,236	14,345
14,313	14,235	14,356	14,272	14,242	14,161	14,125	14,131	14,551
14,452	14,354	14,391	14,417	14,539	14,475	14,404	14,366	14,164
14,17	14,191	14,337	14,296	14,048	13,987	14,31	14,378	14,561
14,404	14,425	14,335	14,317	14,34	14,322	14,335	14,49	14,49
14,469	14,287	14,298	14,302	14,095	14,375	14,239	14,167	14,051
14,003								

Tabulka 1: Kurz CZK/AUD v období od 30.6.2008 do 12.9.2008

Pomocí (7) a (14) spočítáme odhady  $r_k$  a  $r_{kk}$  autokorelační funkce  $\varrho_k$  a parciální autokorelační funkce  $\varrho_{kk}$  řady z tabulky 1. Prvních jedenáct hodnot funkcí  $r_k$  a  $r_{kk}$  je uvedeno v tabulce 2. Nyní budeme hledat identifikační bod funkce  $\varrho_k$ .

k	1	2	3	4	5	6	7	8	9	10	11
$r_k$	0,53	0,25	0,02	0,02	0,05	-0,07	-0,16	-0,24	-0,35	-0,31	-0,22
$r_{kk}$	0,53	-0,06	-0,12	0,09	0,04	-0,19	-0,09	-0,09	-0,26	-0,05	0,02

Tabulka 2: Hodnoty odhadnuté autokorelační a parciální autokorelační funkce

Nejprve otestujeme, zda není  $k_0 = 0$ . Podle (10) je

$$\sigma(r_k) \doteq \sqrt{\frac{1}{n}} = \sqrt{\frac{1}{55}} = 0,135 \quad \text{pro } k > 0.$$

Hodnota  $|r_1| = 0,53$  je skoro dvakrát větší než hodnota  $2\sigma_{r_k} = 0,27$ . Budeme v testování pokračovat a posuneme identifikační bod na  $k_0 = 1$ . V tomto případě je

$$\sigma(r_k) \doteq \sqrt{\frac{1}{n}(1 + 2r_1^2)} = \sqrt{\frac{1}{55}(1 + 2 \cdot 0,53^2)} = 0,169 \quad \text{pro } k > 1.$$

Hodnotu  $2\sigma(r_k) = 0,306$  přesahuje hodnota  $|r_9| = 0,35$ , a tedy ani bod  $k = 1$  není identifikační. Pokud by byl identifikační bod  $k_0 = 2$ , pak by platilo

$$\sigma(r_k) \doteq 2\sqrt{\frac{1}{n}(1 + 2(r_1^2 + r_2^2))} = 2\sqrt{\frac{1}{55}(1 + 2(0,53^2 + 0,25^2))} = 0,1751 \quad \text{pro } k > 2. \quad (16)$$

Dvojnásobek směrodatné odchylky, tj. hodnotu 0,3502, nepřekračuje žádná z hodnot odhadnuté autokorelační funkce  $r_k$ ,  $k > 2$ , a proto  $k_0 = 2$ .

Pro parciální autokorelační funkci identifikační bod nalezneme rychleji. Podle (15) je

$$\sigma(r_{kk}) \doteq \sqrt{\frac{1}{n}} = \sqrt{\frac{1}{55}} = 0,135 \quad \text{pro } k > k_0. \quad (17)$$

Proto budeme všechny hodnoty  $r_{kk}$  z tabulky 2 porovnávat s  $2\sigma(r_{kk}) = 0,27$ . Tu překračuje pouze hodnota  $r_{11}$  a tedy identifikační bod je  $k_0 = 1$ .

### 3 Stacionární procesy

Tato kapitola je věnována základním modelům Boxovy-Jenkinsovy metodologie. Nejprve se budeme zabývat obecným lineárním procesem, dále potom procesy z něho odvozenými. Jsou to: proces klouzavých součtů MA, autoregresní proces AR a smíšený proces ARMA.

#### 3.1 Lineární proces

Stacionární proces, který neobsahuje žádnou deterministickou složku, lze vždy vyjádřit jako lineární kombinaci řady nekorelovaných náhodných veličin ve formě lineárního procesu<sup>1</sup>.

**Definice 3.1.** Časovou řadu ve tvaru

$$X_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \dots, \quad t \in \mathbb{Z}, \quad (18)$$

kde  $\varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2)$  a  $\psi_j$  jsou parametry, nazýváme lineárním procesem.

Jiný způsob zápisu lineárního procesu je pomocí operátoru zpětného posunutí  $B$ , který je dán vztahem

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

a který můžeme aplikovat několikanásobně jako

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

Potom lze psát

$$X_t = \psi(B)\varepsilon_t, \quad (19)$$

kde

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots = 1 + \sum_{j=1}^{\infty} \psi_j B^j. \quad (20)$$

$\psi(B)$  nazýváme operátor lineárního procesu.

---

<sup>1</sup>Dokázal švédský statistik Herman Ole Andreas Wold ve své práci A Study in the Analysis of Stationary Time Series (1938)

Před uvedením věty 3.1 je třeba definovat ještě pojem konvergence podle kvadratického středu.

**Definice 3.2.** *Platí-li*

$$\lim_{t \rightarrow t_0} E(Y_t - Y)^2 = 0, \quad (21)$$

*konverguje řada  $Y_t$ ,  $t \rightarrow t_0$ ,  $t \in T$  podle kvadratického středu k náhodné veličině  $Y$ , tj.*

$$Y = \text{l.i.m.}_{t \rightarrow t_0} Y_t. \quad (22)$$

**Věta 3.1.** *Nechť  $\psi(B)$  je reálná funkce, která konverguje pro  $|B| \leq 1$  a nechť  $\varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2)$ . Potom  $\psi(B)\varepsilon_t$  konverguje podle kvadratického středu, takže existuje náhodná veličina*

$$X = \text{l.i.m.}_{t \rightarrow \infty} \psi(B)\varepsilon_t.$$

**Důkaz:** viz. Anděl, 1976, str. 58.

**Důsledek 3.1.** *Lineární proces je stacionární a centrováný.*

Za jistých podmínek lze lineární proces přepsat (invertovat) na tvar

$$X_t = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \dots + \varepsilon_t. \quad (23)$$

**Poznámka 3.1.** *Procesu ve formě (23) se také říká Boxova-Jenkinsova reprezentace.*

Použijeme-li operátor zpětného posunutí, zapisujeme proces (23) ve tvaru

$$\pi(B)X_t = \varepsilon_t, \quad (24)$$

kde

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots = 1 - \sum_{j=1}^{\infty} \pi_j B^j. \quad (25)$$

Pokud lze lineární proces zapsat ve tvaru (24), potom se nazývá *invertibilní*.



**Věta 3.2** (Podmínka invertibility). *Nechť  $\pi(B)$  konverguje pro*

$$|B| \leq 1. \quad (26)$$

*Potom je proces  $\{X_t\}$  invertibilní.*

Pro parametry  $\psi(B)$  a  $\pi(B)$  platí vztah

$$\psi(B)\pi(B) = 1, \quad (27)$$

tedy

$$(1 + \sum_{j=1}^{\infty} \psi_j B^j)(1 - \sum_{j=1}^{\infty} \pi_j B^j) = 1 + (\psi_1 - \pi_1)B + (\psi_2 - \psi_1\pi_1 - \pi_2)B^2 + \dots = 1.$$

### 3.2 Proces klouzavých součtů

**Definice 3.3.** *Proces vyjádřený ve tvaru*

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \quad t \in \mathbb{Z}, \quad \varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2), \quad (28)$$

*se nazývá proces klouzavých součtů řádu  $q$ . Značíme ho  $MA(q)$ .*

Proces (28) lze vyjádřit též pomocí operátoru zpětného posunutí  $B$  ve formě

$$X_t = \theta(B)\varepsilon_t, \quad t \in \mathbb{Z}. \quad (29)$$

V tomto případě je operátor zpětného posunutí dán vztahem

$$\theta(B) = 1 + \sum_{j=1}^q \theta_j B^j \quad (30)$$

a nazýváme ho *operátor klouzavých součtů*.

Proces klouzavých součtů (28) vychází přímo z lineárního procesu (18), změna je pouze v tom, že má konečný počet parametrů. Z toho plyne, že všechny procesy klouzavých součtů jsou stacionární pro libovolnou volbu parametrů, neboť

$$\psi(B) = \theta(B) < \infty \quad \text{pro } |B| \leq 1, \quad \forall \theta_1, \dots, \theta_q \quad (31)$$

a stacionarita je důsledkem věty 3.1.

Pro první dva momenty tohoto procesu platí

$$E(X_t) = 0$$

a

$$\sigma_X^2 = \gamma_0 = (1 + \theta_1^2 + \dots + \theta_q^2)\sigma_\varepsilon^2. \quad (32)$$

Autokorelační funkce  $\varrho_k$  má tvar

$$\begin{aligned} \varrho_k &= \frac{\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}, & k = 1, \dots, q, \\ \varrho_k &= 0, & k > q. \end{aligned} \quad (33)$$

Identifikační bod autokorelační funkce procesu (28) je  $k_0 = q$ . Při známých  $\varrho_k$  je možné spočítat parametry  $\theta_i$ ,  $i = 1, 2, \dots, q$ . Parciální autokorelační funkce  $\varrho_{kk}$  je omezená geometricky klesající posloupností nebo sinusoidou s geometricky klesající amplitudou. Identifikační bod  $k_0$  této funkce neexistuje. Podle věty (3.2) je proces (28) invertibilní, pokud řada

$$\pi(B) = \theta^{-1}(B) \quad (34)$$

konverguje pro  $|B| \leq 1$ .  $\theta(B)$  (polynom řádu  $q$ ) rozložíme na kořenové činitele následovně

$$\theta(B) = \prod_{i=1}^q (1 - h_i^{-1}B), \quad (35)$$

kde  $h_i$ ,  $i = 1, 2, \dots, q$ , jsou kořeny polynomu  $\theta(B)$ . Potom řada  $\theta^{-1}(B)$ ,  $|B| \leq 1$ , konverguje pro  $|h_i| > 1$ ,  $i = 1, 2, \dots, q$ . Tím je dokázána následující věta.

**Věta 3.3.** *Proces klouzavých součtů je invertibilní, leží-li kořeny  $h_i$ ,  $i = 1, 2, \dots, q$ , polynomu  $\theta(B)$  v komplexní rovině vně jednotkového kruhu.*

### Proces MA(1)

V tomto případě je  $q = 1$ . Model časové řady má tvar

$$X_t = \varepsilon_t + \theta_1\varepsilon_{t-1}, \quad t \in \mathbb{Z}, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2). \quad (36)$$

**Věta 3.4.** *Pro invertibilní proces MA(1) platí*

$$|\theta_1| < 1. \quad (37)$$

**Důkaz:** Polynom  $\theta(B) = 1 + \theta_1 B$  má pouze jeden kořen  $h_1 = -\frac{1}{\theta_1}$ , pro který musí podle věty 3.3 platit  $|h_1| > 1$ . ■

Operátor zpětného posunutí je zde ve tvaru

$$\pi(B) = \theta^{-1}(B) = \frac{1}{1 + \theta_1 B} = \frac{1}{1 - (-\theta_1 B)} = \sum_{j=0}^{\infty} (-\theta_1 B)^j = 1 - \sum_{j=1}^{\infty} (-1)^{j+1} \theta_1^j B^j$$

a proces (36) lze při splnění (37) psát ve tvaru

$$X_t = \theta_1 X_{t-1} - \theta_1^2 X_{t-2} + \theta_1^3 X_{t-3} - \theta_1^4 X_{t-4} + \dots + \varepsilon_t. \quad (38)$$

Autokorelační funkce procesu MA(1) nabývá dle (33) hodnot

$$\begin{aligned} \varrho_1 &= \frac{\theta_1}{1 + \theta_1^2}, \\ \varrho_k &= 0 \quad \text{pro } k > 1. \end{aligned} \quad (39)$$

Její identifikační bod je  $k_0 = 1$ .

**Lemma 3.1.** *Pro invertibilní proces platí*

$$|\varrho_1| < \frac{1}{2}. \quad (40)$$

**Důkaz:** Z předchozího textu víme, že  $\varrho_1 = \frac{\theta_1}{1 + \theta_1^2}$ . S ohledem na (37) budeme hledat extrémy funkce  $f(x) = \frac{x}{1+x^2}$  pro  $|x| < 1$ . První derivaci této funkce

$$\frac{df(x)}{dx} = \frac{1 - x^2}{(1 + x^2)^2}$$

položíme rovnou nule. Extrémy nastávají na hranici definičního oboru v bodech  $x = -1$  a  $x = 1$ . První derivace  $f'(x)$  je na celém definičním oboru kladná,

funkce  $f(x)$  je tedy rostoucí a tudíž se v bodě  $x = -1$  jedná o globální minimum s funkční hodnotou  $-\frac{1}{2}$  a v bodě  $x = 1$  o globální maximum s funkční hodnotou  $\frac{1}{2}$ . Z toho plyne, že  $|\varrho_1| = \left| \frac{\theta_1}{1+\theta_1^2} \right| < \frac{1}{2}$ . ■

Za předpokladu, že známe  $\varrho_1$ , vypočítáme hodnotu parametru  $\theta_1$ . Upravíme vztah (39)

$$\varrho_1 \theta_1^2 - \theta_1 + \varrho_1 = 0.$$

Kořeny této rovnice jsou

$$t_{1,2} = \frac{1 \pm \sqrt{1 - 4\varrho_1^2}}{2\varrho_1}.$$

Kořen  $t_1 = \frac{1 + \sqrt{1 - 4\varrho_1^2}}{2\varrho_1}$  nespĺňuje pro  $|\varrho| < \frac{1}{2}$  podmínku invertibility (37), protože

$$|t_1| = \frac{|1 + \sqrt{1 - 4\varrho_1^2}|}{2|\varrho_1|} > |1 + \sqrt{1 - 4\varrho_1^2}| > 1.$$

Kořen  $t_2$  podmínce invertibility vyhovuje a dostáváme

$$\theta_1 = \frac{1 - \sqrt{1 - 4\varrho_1^2}}{2\varrho_1}. \quad (41)$$

Lze ukázat, že parciální autokorelační funkce procesu MA(1) má tvar

$$\varrho_{kk} = \frac{(-1)^{k-1} \theta_1^k (1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}. \quad (42)$$

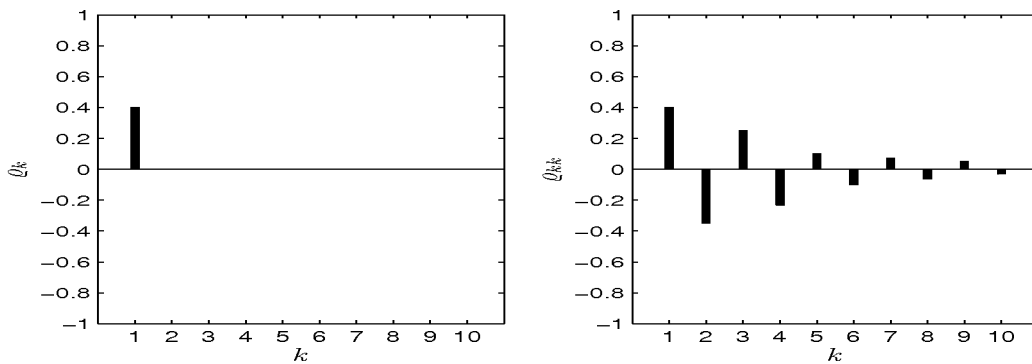
**Lemma 3.2.** *Parciální autokorelační funkce procesu MA(1) je omezená geometricky klesající posloupností*

$$|\varrho_{kk}| < |\theta_1|^k. \quad (43)$$

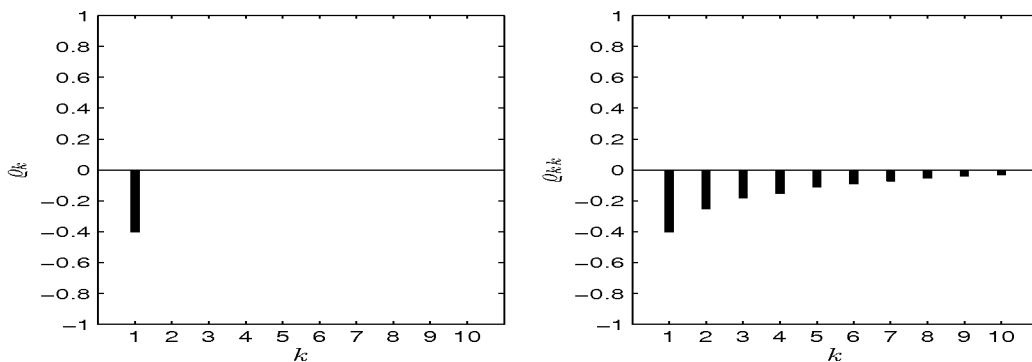
**Důkaz:** Z věty 3.4 víme, že  $|\theta_1| < 1$  a tedy  $\theta_1^{2(k+1)} < \theta_1^2$ , proto je  $\frac{1 - \theta_1^2}{1 - \theta_1^{2(k+1)}} < 1$  a pro absolutní hodnotu parciální autokorelační funkce potom platí (viz. vztah (42))

$$|\varrho_{kk}| = |\theta_1|^k \left| \frac{1 - \theta_1^2}{1 - \theta_1^{2(k+1)}} \right| < |\theta_1|^k. \quad \blacksquare$$

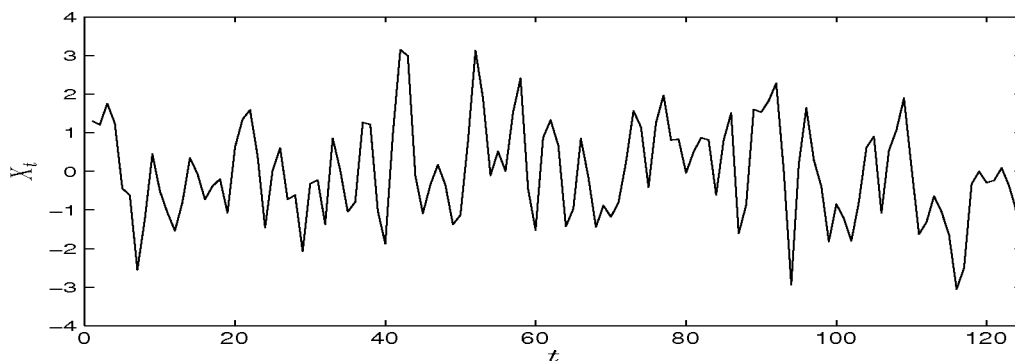
Na obrázcích 1 a 2 jsou znázorněny možné tvary  $\varrho_k$  a  $\varrho_{kk}$  procesu MA(1) pro kladnou a zápornou hodnotu parametru  $\theta_1$ . Obrázek 3 zachycuje časovou řadu 124 hodnot simulovanou na základě procesu MA(1) ve formě  $X_t = \varepsilon_t + 0,8\varepsilon_{t-1}$ . Na obrázku 4 jsou potom odhady autokorelační a parciální autokorelační funkce této řady. Na obrázku 5 je zobrazena řada generovaná pomocí stejných  $\varepsilon_t$ , ale s parametrem s opačným znaménkem, grafy odhadů její autokorelační a parciální autokorelační funkce si můžeme prohlédnout na obrázku 6.



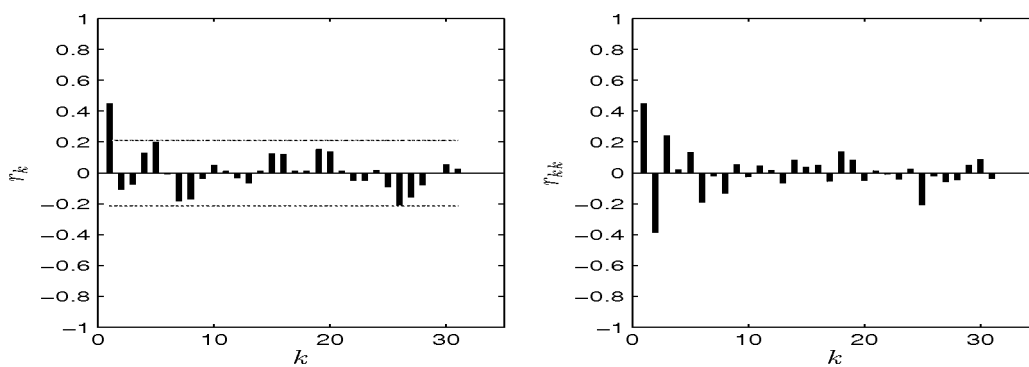
Obrázek 1:  $\varrho_k$  a  $\varrho_{kk}$  pro parametr  $\theta_1 > 0$



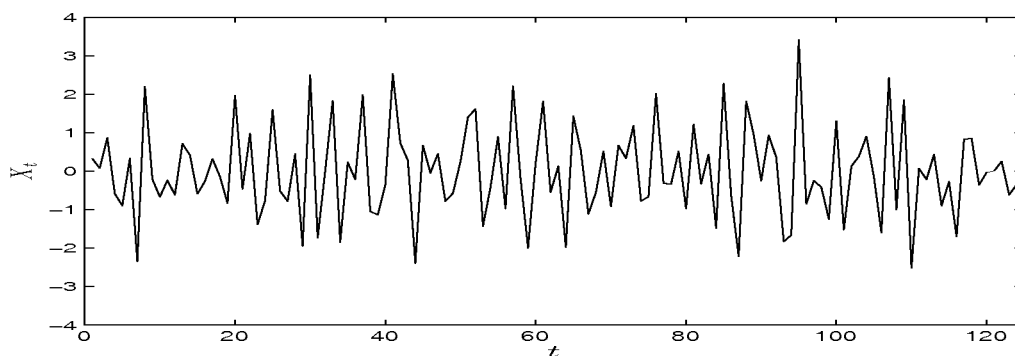
Obrázek 2:  $\varrho_k$  a  $\varrho_{kk}$  pro parametr  $\theta_1 < 0$



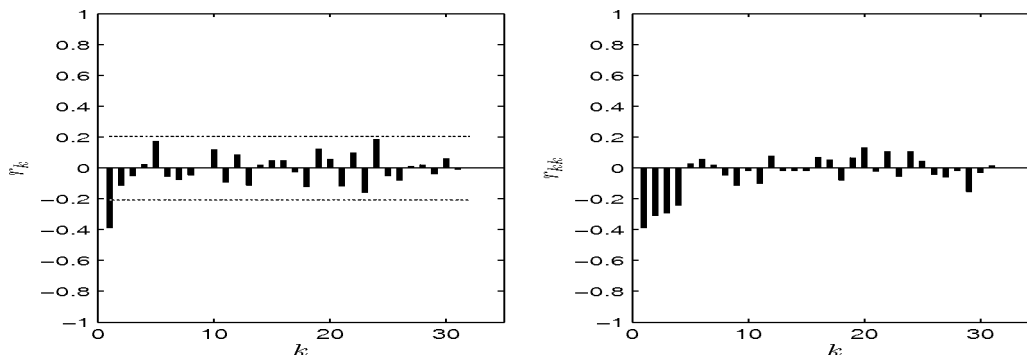
Obrázek 3: Simulovaný proces  $X_t = \varepsilon_t + 0,8\varepsilon_{t-1}$



Obrázek 4: Odhad autokorelační a parciální autokorelační funkce procesu  $X_t$



Obrázek 5: Simulovaný proces  $X_t = \varepsilon_t - 0,8\varepsilon_{t-1}$



Obrázek 6: Odhad autokorelační a parciální autokorelační funkce procesu  $X_t$

### Proces MA(2)

Tento proces má tvar

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \quad t \in \mathbb{Z}, \quad \varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2). \quad (44)$$

Pomocí operátoru klouzavých součtů lze psát

$$X_t = \theta(B)\varepsilon_t = (1 + \theta_1 B + \theta_2 B^2)\varepsilon_t. \quad (45)$$

**Věta 3.5.** *Pro invertibilní proces MA(2) platí následující vztahy*

$$i) \quad |\theta_1| < 2, \quad |\theta_2| < 1, \quad (46)$$

$$ii) \quad \theta_2 + \theta_1 > -1, \quad (47)$$

$$iii) \quad \theta_2 - \theta_1 > -1. \quad (48)$$

**Důkaz:** Z věty 3.3 plyne, že pro kořeny polynomu  $\theta(B)$  platí  $|h_1| > 1$  a  $|h_2| > 1$ .

Polynom rozepíšeme  $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 = \left(1 - \frac{B}{h_1}\right)\left(1 - \frac{B}{h_2}\right) = 1 - \left(\frac{1}{h_1} + \frac{1}{h_2}\right)B + \frac{1}{h_1 h_2} B^2$  a porovnáme koeficienty u  $B$ . Dostaneme, že  $\theta_1 = -\left(\frac{1}{h_1} + \frac{1}{h_2}\right)$

a  $\theta_2 = \frac{1}{h_1 h_2}$ . Potom je

$$i) \quad |\theta_1| = \left|\frac{1}{h_1} + \frac{1}{h_2}\right| \leq \left|\frac{1}{h_1}\right| + \left|\frac{1}{h_2}\right| < 2$$

$$|\theta_2| = \left|\frac{1}{h_1 h_2}\right| = \left|\frac{1}{h_1}\right| \left|\frac{1}{h_2}\right| < 1$$

$$ii) \quad \theta_2 + \theta_1 = \frac{1}{h_1 h_2} - \frac{1}{h_1} - \frac{1}{h_2}. \text{ Budeme hledat minimum funkce } f(x, y) =$$

$= f(y, x) = \frac{1}{xy} - \frac{1}{x} - \frac{1}{y}$  pro  $|x| > 1, |y| > 1$ . Parciální derivace

$$\frac{\partial f(x, y)}{\partial x} = -\frac{1}{x^2} \frac{1}{y} + \frac{1}{x^2} = \frac{1}{x^2} \left(1 - \frac{1}{y}\right),$$

$$\frac{\partial f(x, y)}{\partial y} = -\frac{1}{y^2} \frac{1}{x} + \frac{1}{y^2} = \frac{1}{y^2} \left(1 - \frac{1}{x}\right)$$

položíme rovny nule. Rovnost platí v prvním případě, pokud je  $x = \pm\infty$  nebo  $y = 1$ , ve druhém případě obdobně, tedy pokud  $y = \pm\infty$  nebo  $x = 1$ . Vypočítáme hodnoty funkce  $f(x, y)$  v těchto bodech:

$$f(1, 1) = f(1, -\infty) = f(1, \infty) = f(-\infty, 1) = f(\infty, 1) = -1,$$

$$f(-\infty, -\infty) = f(-\infty, \infty) = f(\infty, -\infty) = f(\infty, \infty) = 0.$$

Odtud plyne vztah  $\theta_2 + \theta_1 > -1$ .

*iii)* Dokážeme obdobně jako *ii)* :  $\theta_2 - \theta_1 = \frac{1}{h_1 h_2} + \frac{1}{h_1} + \frac{1}{h_2}$ . Budeme hledat minimum funkce  $f(x, y) = f(y, x) = \frac{1}{xy} + \frac{1}{x} + \frac{1}{y}$  pro  $|x| > 1, |y| > 1$ . Parciální derivace

$$\frac{\partial f(x, y)}{\partial x} = -\frac{1}{x^2} \frac{1}{y} - \frac{1}{x^2} = -\frac{1}{x^2} \left(1 + \frac{1}{y}\right),$$

$$\frac{\partial f(x, y)}{\partial y} = -\frac{1}{y^2} \frac{1}{x} - \frac{1}{y^2} = -\frac{1}{y^2} \left(1 + \frac{1}{x}\right)$$

položíme rovny nule. To nastane v bodech  $x = \pm\infty$  nebo  $y = -1$ , popřípadě  $y = \pm\infty$  nebo  $x = -1$ . Vypočítáme hodnoty funkce  $f(x, y)$  :

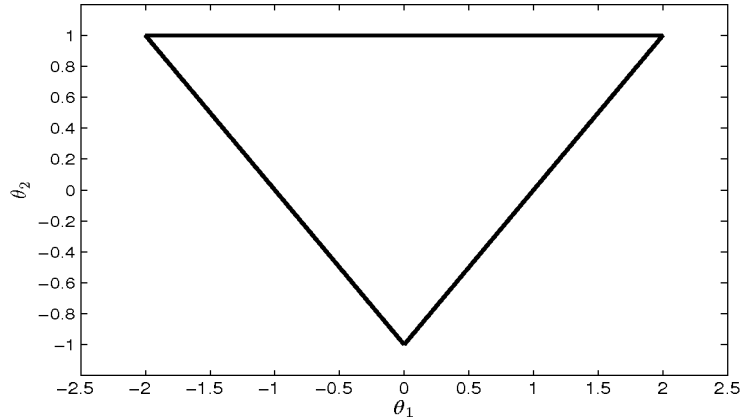
$$f(-1, -1) = f(-1, -\infty) = f(-1, \infty) = f(-\infty, -1) = f(\infty, -1) = -1,$$

$$f(-\infty, -\infty) = f(-\infty, \infty) = f(\infty, -\infty) = f(\infty, \infty) = 0.$$

A tedy platí  $\theta_2 - \theta_1 > -1$ . ■

Na obrázku 7 je zobrazena dvourozměrná oblast invertibility procesu MA(2), která je určena vztahy (46), (47) a (48).





Obrázek 7: Oblast invertibility procesu MA(2)

Autokorelační funkce je dána vztahy

$$\begin{aligned}\varrho_1 &= \frac{\theta_1(1 + \theta_2)}{1 + \theta_1^2 + \theta_2^2}, \\ \varrho_2 &= \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}, \\ \varrho_k &= 0, \quad k > 2.\end{aligned}\tag{49}$$

Její identifikačním bodem je  $k_0 = 2$ .

**Lemma 3.3.** *Pro hodnoty  $\varrho_1, \varrho_2$  libovolného invertibilního procesu (44) platí*

$$i) \quad \varrho_2 + \varrho_1 > -\frac{1}{2},\tag{50}$$

$$ii) \quad \varrho_2 - \varrho_1 > -\frac{1}{2}.\tag{51}$$

**Důkaz:** V důkazu využijeme skutečnosti, že pro  $a^2 + b^2 + c^2 > 0$ , kde  $(a, b, c) \neq (0, 0, 0)$ , platí vztah

$$\frac{ab + ac + bc}{a^2 + b^2 + c^2} \geq -\frac{1}{2}.\tag{52}$$

Tento vztah je odvozen z toho, že obecně platí  $(a + b + c)^2 \geq 0$  a tedy také  $a^2 + b^2 + c^2 + 2(ab + ac + bc) \geq 0$ , odtud dostaneme  $ab + ac + bc \geq -\frac{a^2 + b^2 + c^2}{2}$ , vydělíme  $a^2 + b^2 + c^2 > 0$  a získáme vztah (52).

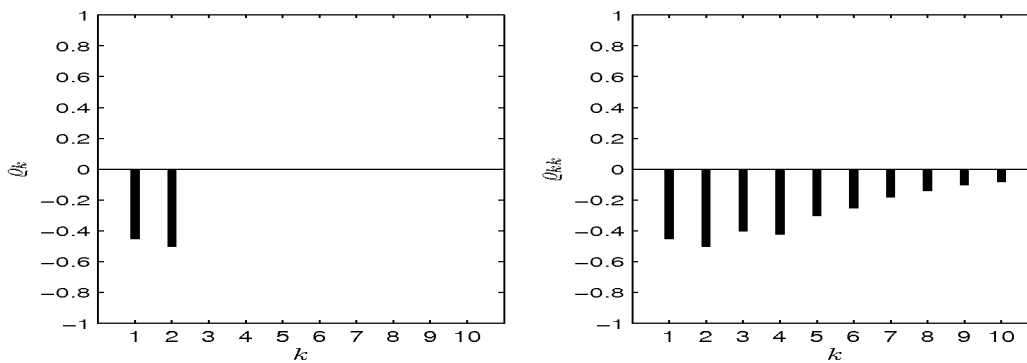
*i)* Nejprve sečteme  $\varrho_2 + \varrho_1 = \frac{\theta_2}{1+\theta_1^2+\theta_2^2} + \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2} = \frac{\theta_1+\theta_2+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2}$ . Zvolíme  $a = 1, b = \theta_1, c = \theta_2$  a dostáváme přímo tvar (52). Rovnost  $1 + \theta_1 + \theta_2 = 0$  určuje hranici oblasti invertibility ( $\theta_2 = -\theta_1 - 1$ ), nerovnost (50) tedy platí uvnitř celé oblasti invertibility.

*ii)* Dokážeme obdobně:  $\varrho_2 - \varrho_1 = \frac{\theta_2}{1+\theta_1^2+\theta_2^2} - \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2} = \frac{-\theta_1+\theta_2-\theta_1\theta_2}{1+\theta_1^2+\theta_2^2}$ . Tentokrát zvolíme  $a = 1, b = -\theta_1, c = \theta_2$ . Rovnost  $1 - \theta_1 + \theta_2 = 0$  určuje hranici oblasti invertibility ( $\theta_2 = \theta_1 - 1$ ), nerovnost (51) tedy také platí uvnitř celé oblasti invertibility. ■

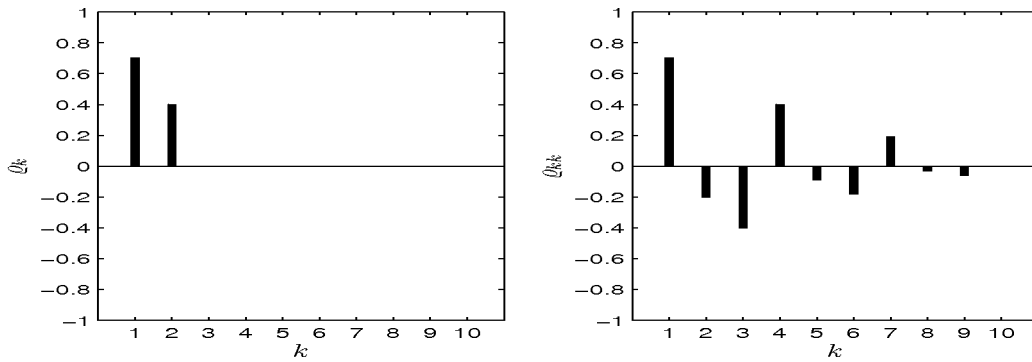
Výpočet parametrů  $\theta_1$  a  $\theta_2$  ze známých hodnot  $\varrho_1$  a  $\varrho_2$  se provádí iteračně nebo pomocí speciálních diagramů.

Parciální autokorelační funkce  $\varrho_{kk}$  je omezena geometricky klesající posloupností (v případě reálných kořenů polynomu  $\theta(B)$ ) nebo sinusoidou s geometricky klesající amplitudou (v případě komplexních kořenů polynomu  $\theta(B)$ ).

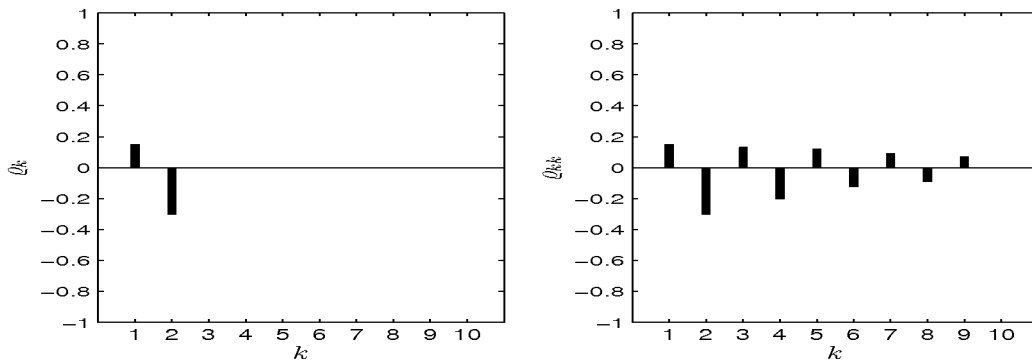
Na obrázcích 8-11 jsou znázorněny možné tvary  $\varrho_k$  a  $\varrho_{kk}$  procesu MA(2) pro různé hodnoty parametrů  $\theta_1$  a  $\theta_2$ . Obrázek 12 zachycuje časovou řadu 118 hodnot simulovanou na základě procesu MA(2) ve formě  $X_t = \varepsilon_t + 0,8\varepsilon_{t-1} - 0,4\varepsilon_{t-2}$ . Na obrázku 13 jsou potom odhady autokorelační a parciální autokorelační funkce této řady.



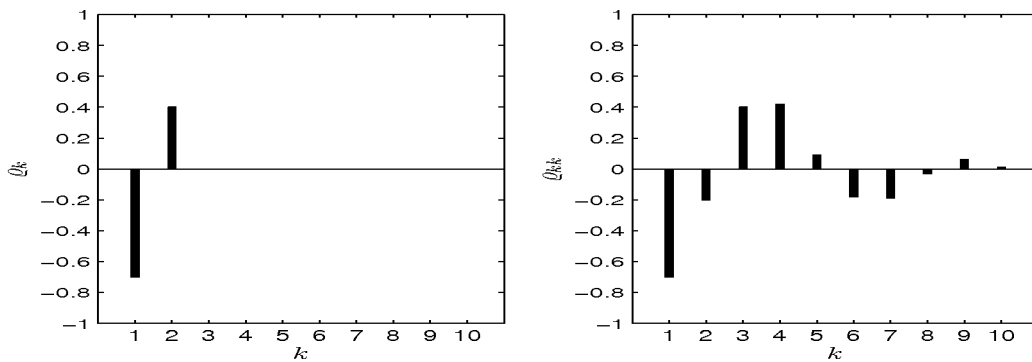
Obrázek 8:  $\varrho_k$  a  $\varrho_{kk}$  pro parametry  $\theta_1 < 0, \theta_2 < 0$



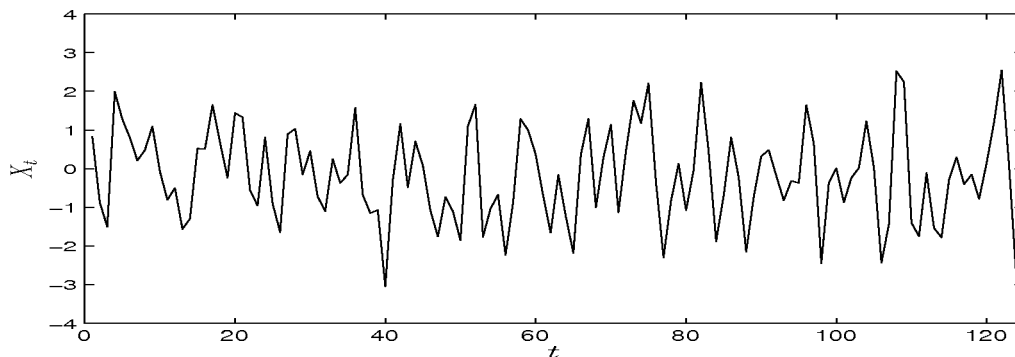
Obrázek 9:  $\rho_k$  a  $\rho_{kk}$  pro parametry  $\theta_1 > 0$ ,  $\theta_2 > 0$



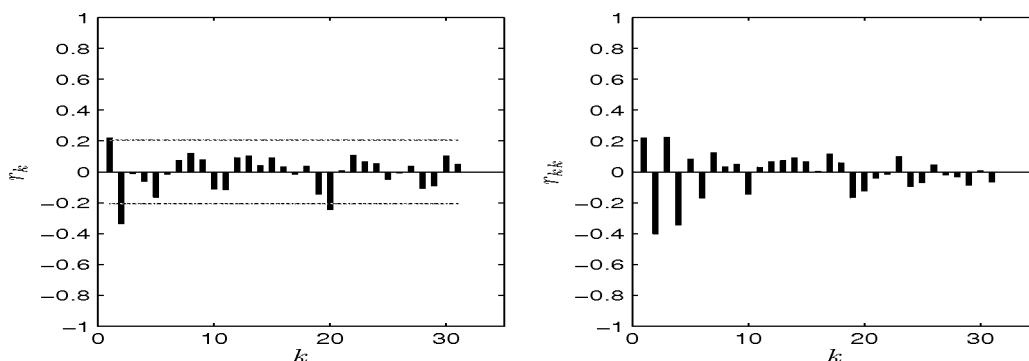
Obrázek 10:  $\rho_k$  a  $\rho_{kk}$  pro parametry  $\theta_1 > 0$ ,  $\theta_2 < 0$



Obrázek 11:  $\rho_k$  a  $\rho_{kk}$  pro parametry  $\theta_1 < 0$ ,  $\theta_2 > 0$



Obrázek 12: Simulovaný proces  $X_t = \varepsilon_t + 0,8\varepsilon_{t-1} - 0,4\varepsilon_{t-2}$



Obrázek 13: Odhad autokorelační a parciální autokorelační funkce procesu  $X_t$

### 3.3 Autoregresní proces

**Definice 3.4.** Časovou řadu vyjádřenou vztahem

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2), \quad (53)$$

kde  $\varphi_1, \dots, \varphi_p$  jsou parametry, nazýváme autoregresním procesem řádu  $p$ . Značíme  $AR(p)$ .

Proces (53) můžeme zapsat také pomocí operátoru zpětného posunutí  $B$  jako

$$\varphi(B)X_t = \varepsilon_t, \quad (54)$$

kde

$$\varphi(B) = 1 - \sum_{i=1}^p \varphi_i B^i \quad (55)$$

je autoregresní operátor.

**Věta 3.6.** (*Podmínka stacionarity*) Autoregresní proces je stacionární, pokud leží kořeny  $g_j$ ,  $j = 1, 2, \dots, p$ , polynomu  $\varphi(B)$  v komplexní rovině vně jednotkového kruhu.

**Důkaz:** Důkaz se provede na stejném principu jako pro větu 3.3. ■

Autoregresní proces (53) odpovídá centrovanému lineárnímu procesu zapsanému v invertovaném tvaru (24) s konečným počtem parametrů. Tento proces je tedy invertibilní pro libovolné parametry  $\varphi_i$ ,  $i = 1, 2, \dots, p$ .

Vyhovuje-li proces podmínce stacionarity, lze jej psát v neinvertovaném tvaru

$$X_t = \varphi^{-1}(B)\varepsilon_t. \quad (56)$$

Hodnoty autokorelační funkce získáme řešením soustavy diferenčních rovnic

$$\varrho_k = \varphi_1\varrho_{k-1} + \varphi_2\varrho_{k-2} + \dots + \varphi_p\varrho_{k-p}, \quad k = 1, 2, \dots \quad (57)$$

Nyní ukážeme, že tyto rovnice jsou odvozeny postupným násobením vztahu (53) veličinami  $X_{t-k}$ ,  $k = 1, 2, \dots, p$ , a následným přechodem ke středním hodnotám. Víme, že střední hodnota tohoto procesu je nulová a že budoucí hodnoty chyb jsou nekorelované s minulými hodnotami časové řady. Z toho plyne  $E(X_{t-k}\varepsilon_t) = 0$ ,  $k = 1, 2, \dots$ , a tedy

$$\gamma_k = \varphi_1\gamma_{k-1} + \varphi_2\gamma_{k-2} + \dots + \varphi_p\gamma_{k-p}, \quad k = 1, 2, \dots \quad (58)$$

Tyto rovnosti nakonec vydělíme  $\gamma_0$ . Získáme soustavu (57). Dále odvodíme tvar autokorelační funkce pro různé kořeny polynomu  $\varphi(B)$ . K výpočtu parametrů  $\varphi_i$ ,  $i = 1, 2, \dots, p$ , při známých hodnotách autokorelační funkce se používá prvních  $p$  rovnic soustavy (57), tedy tzv. *Yuleova-Walkerova soustava rovnic*, tj.

$$\begin{aligned} \varrho_1 &= \varphi_1 + \varphi_2\varrho_1 + \dots + \varphi_p\varrho_{p-1}, \\ \varrho_2 &= \varphi_1\varrho_1 + \varphi_2 + \dots + \varphi_p\varrho_{p-2}, \\ &\vdots \\ \varrho_p &= \varphi_1\varrho_{p-1} + \varphi_2\varrho_{p-2} + \dots + \varphi_p. \end{aligned} \quad (59)$$

Vztah pro rozptyl procesu  $AR(p)$  odvodíme tak, že vynásobíme vztah (53) veličinou  $X_t$ , přejdeme ke střední hodnotě, vydělíme  $\gamma_0$  a po úpravě dostaneme

$$\sigma_X^2 = \gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \varphi_1 \varrho_1 - \dots - \varphi_p \varrho_p}. \quad (60)$$

Parciální autokorelační funkce  $\varrho_{kk}$  má identifikační bod  $k_0 = p$  a pro  $k > p$  jsou hodnoty této funkce rovny nule.

### Proces $AR(1)$

Tento proces je vyjádřený tvarem

$$X_t = \varphi_1 X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2). \quad (61)$$

**Věta 3.7.** *Pro stacionární proces  $AR(1)$  platí*

$$|\varphi_1| < 1. \quad (62)$$

**Důkaz:** Polynom  $\varphi(B) = 1 + \varphi_1 B$  má pouze jeden kořen  $g_1 = -\frac{1}{\varphi_1}$ , pro který musí platit podle věty (3.6)  $|g_1| > 1$ , a tedy  $|\varphi_1| < 1$ . ■

Pokud je splněna podmínka (62), lze proces (61) přepsat na tvar

$$X_t = \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_1^2 \varepsilon_{t-2} + \dots \quad (63)$$

Pro autokorelační funkci procesu  $AR(1)$  platí

$$\varrho_1 = \varphi_1 \quad (64)$$

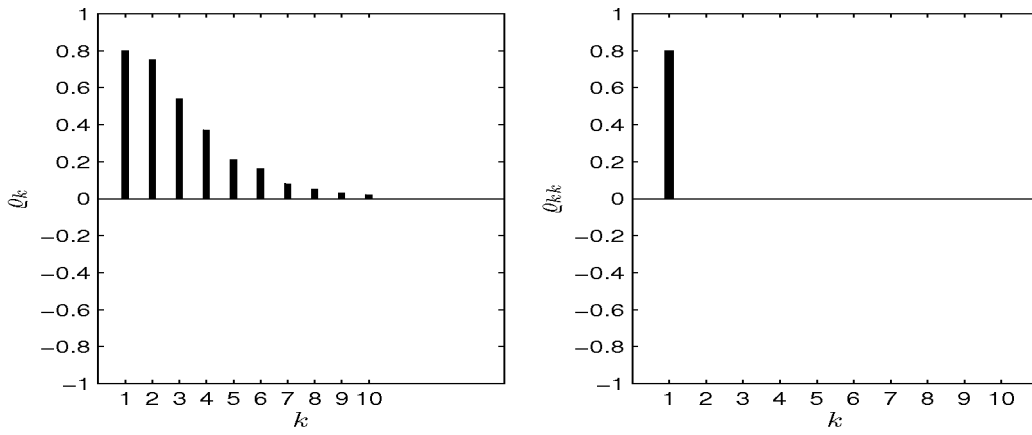
$$\varrho_k = \varphi_1 \varrho_{k-1} = \varphi_1^2 \varrho_{k-2} = \dots = \varphi_1^k, \quad k \geq 1. \quad (65)$$

Je to tedy klesající geometrická posloupnost. Parciální autokorelační funkce je dána vztahy

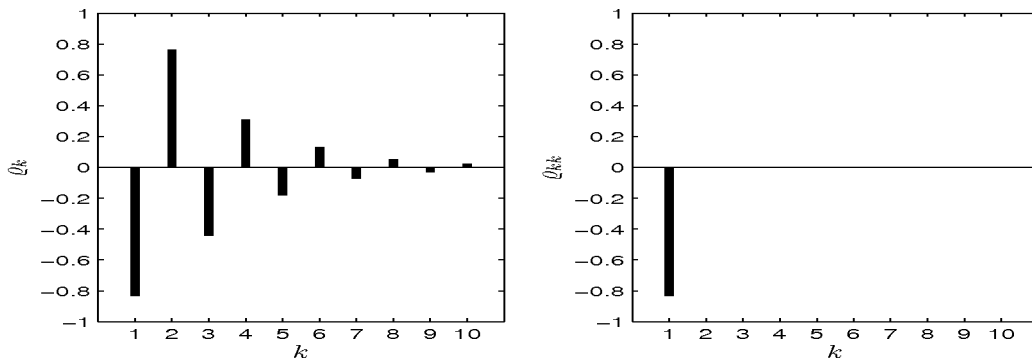
$$\varrho_{11} = \varphi_1, \quad (66)$$

$$\varrho_{kk} = 0, \quad k \geq 2, \quad (67)$$

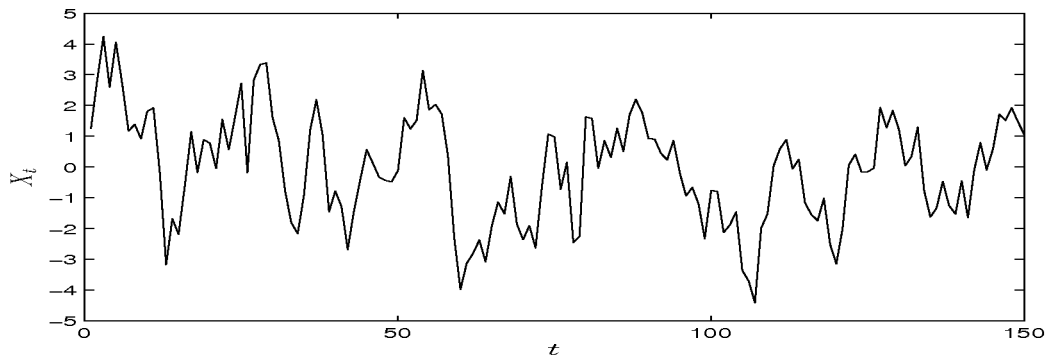
jejím identifikačním bodem je  $k_0 = 1$ .



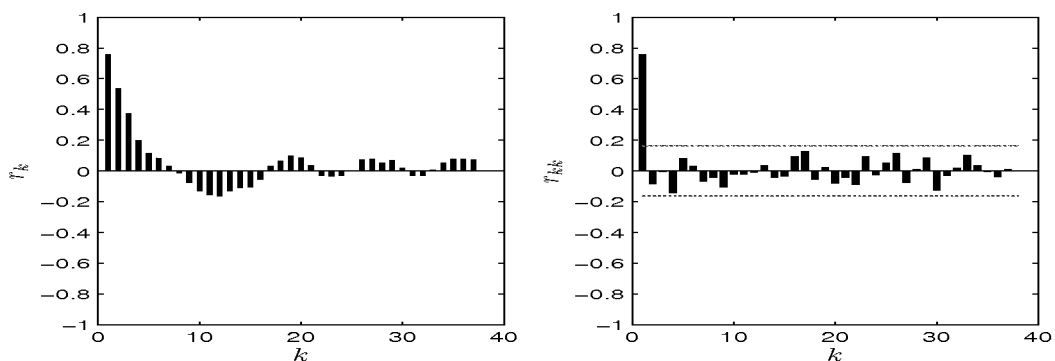
Obrázek 14:  $\rho_k$  a  $\rho_{kk}$  pro parametr  $\varphi_1 > 0$



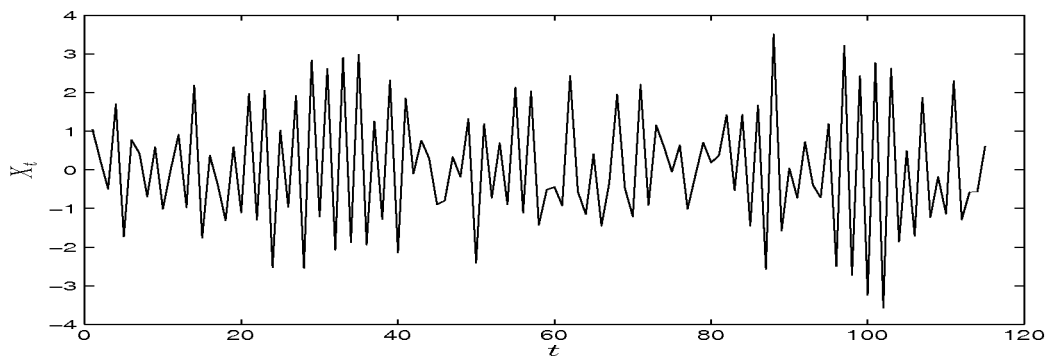
Obrázek 15:  $\rho_k$  a  $\rho_{kk}$  pro parametr  $\varphi_1 < 0$



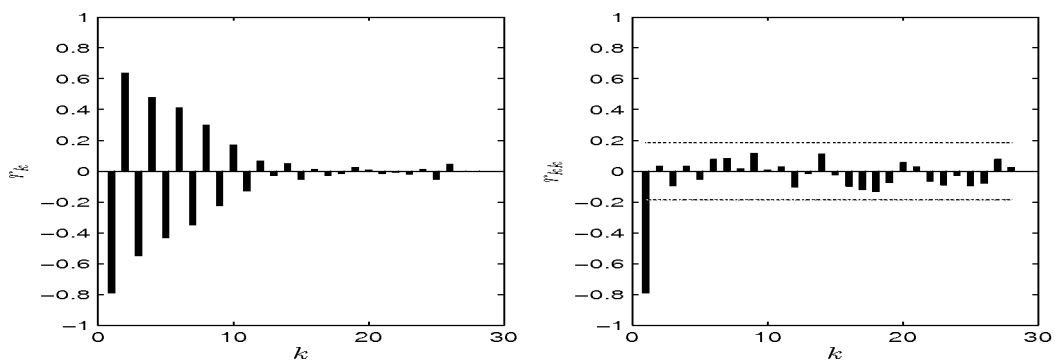
Obrázek 16: Simulovaný proces  $X_t = 0,8X_{t-1} + \varepsilon_t$



Obrázek 17: Odhad autokorelační a parciální autokorelační funkce procesu  $X_t$



Obrázek 18: Simulovaný proces  $X_t = -0,8X_{t-1} + \varepsilon_t$



Obrázek 19: Odhad autokorelační a parciální autokorelační funkce procesu  $X_t$



Na obrázcích 14 a 15 jsou znázorněny možné tvary  $\varrho_k$  a  $\varrho_{kk}$  procesu AR(1) pro kladnou a zápornou hodnotu parametru  $\varphi_1$ . Obrázek 16 zachycuje časovou řadu 115 hodnot simulovanou na základě procesu AR(1) ve formě  $X_t = 0,8X_{t-1} + \varepsilon_t$ . Na obrázku 17 jsou potom odhady autokorelační a parciální autokorelační funkce této řady. Řada simulovaná se záporným parametrem je zobrazena na obrázku 18, odhady její autokorelační a parciální autokorelační funkce jsou znázorněny grafy na obr. 19.

### Proces AR(2)

Tento proces je ve tvaru

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2), \quad (68)$$

**Věta 3.8.** *Pro stacionární proces AR(2) platí vztahy*

$$i) \quad |\varphi_1| < 2, \quad |\varphi_2| < 1, \quad (69)$$

$$ii) \quad \varphi_2 + \varphi_1 < 1, \quad (70)$$

$$iii) \quad \varphi_2 - \varphi_1 < 1. \quad (71)$$

**Důkaz:** Důkaz provedeme stejně jako u věty 3.5, ale místo  $\theta_1$  dosadíme  $\varphi_1 = \frac{1}{g_1} + \frac{1}{g_2} = -\theta_1$  a místo  $\theta_2$  parametr  $\varphi_2 = \frac{1}{g_1 g_2} = -\theta_2$ , kde  $g_1, g_2$  jsou kořeny polynomu  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$ . ■

Dvourozměrnou oblast stacionarity danou vztahy

$$\varphi_1 + \varphi_2 < 1, \quad \varphi_2 - \varphi_1 < 1, \quad -1 < \varphi_2 < 1, \quad (72)$$

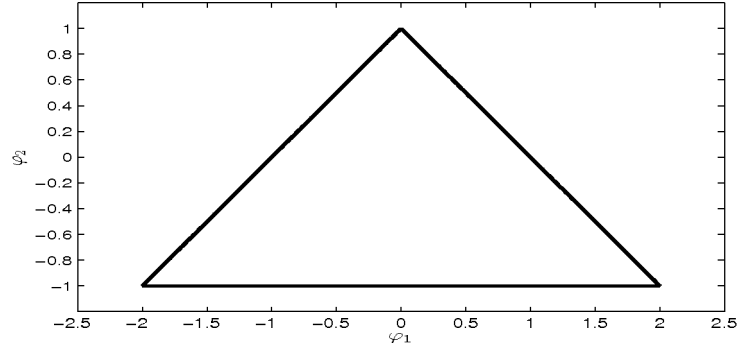
ukazuje obrázek 20. Tvar autokorelační funkce  $\varrho_k$  procesu (68) je odvozen v literatuře ([1], str. 60-61).

Jestliže jsou  $z_1$  a  $z_2$  navzájem různé kořeny charakteristického polynomu

$$z^2 - \varphi_1 z - \varphi_2 \quad (73)$$

příslušnému k soustavě diferenčních rovnic

$$\varrho_k - \varphi_1 \varrho_{k-1} - \varphi_2 \varrho_{k-2}, \quad (74)$$



Obrázek 20: Oblast stacionarity procesu AR(2)

potom má autokorelační funkce  $\varrho_k$  tvar

$$\varrho_k = \frac{(1 - z_2^2)z_1^{k+1} - (1 - z_1^2)z_2^{k+1}}{(z_1 - z_2)(1 + z_1z_2)}. \quad (75)$$

Pro kořeny charakteristického polynomu (73) platí

$$z_1 = \frac{1}{g_1}, \quad z_2 = \frac{1}{g_2}, \quad (76)$$

kde  $g_1, g_2$  jsou kořeny polynomu  $\varphi(B) = \varphi(\frac{1}{z}) = 1 - \varphi_1\frac{1}{z} - \varphi_2(\frac{1}{z})^2$ . Z věty 3.6 plyne, že pro stacionární proces musí platit  $|z_1| < 1$ ,  $|z_2| < 1$ .

Je-li  $z_1 = z_2$  dvojnásobný reálný kořen, pak je

$$\varrho_k = \left(1 + \frac{1 - z_1^2}{1 + z_1^2}k\right)z_1^k. \quad (77)$$

Jsou-li kořeny komplexně sdružené, tj. platí  $z_2 = \bar{z}_1 = ze^{i\omega}$ ,  $z = |z_2| = |z_1|$ , potom dostáváme řešení ve tvaru

$$\varrho_k = Az^k \sin(\omega k + B), \quad (78)$$

kde  $A = \frac{1}{\sin(B)}$  a  $B = \arctg\left(\frac{1+z^2}{1-z^2}\operatorname{tg}\omega\right)$ .

Řešením Yuleovy-Walkerovy soustavy rovnic

$$\begin{aligned} \varrho_1 &= \varphi_1 + \varphi_2\varrho_1, \\ \varrho_2 &= \varphi_1\varrho_1 + \varphi_2 \end{aligned} \quad (79)$$

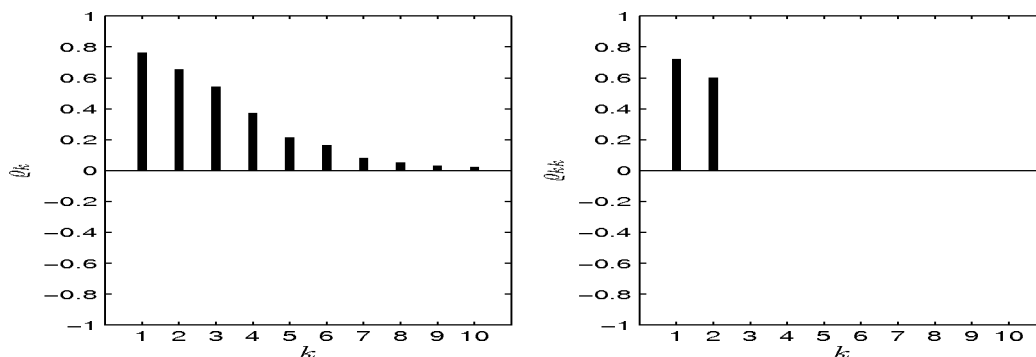
dostaneme vztahy

$$\varphi_1 = \frac{\varrho_1(1 - \varrho_2)}{1 - \varrho_1^2}, \quad \varphi_2 = \frac{(\varrho_2 - \varrho_1^2)}{1 - \varrho_1^2}. \quad (80)$$

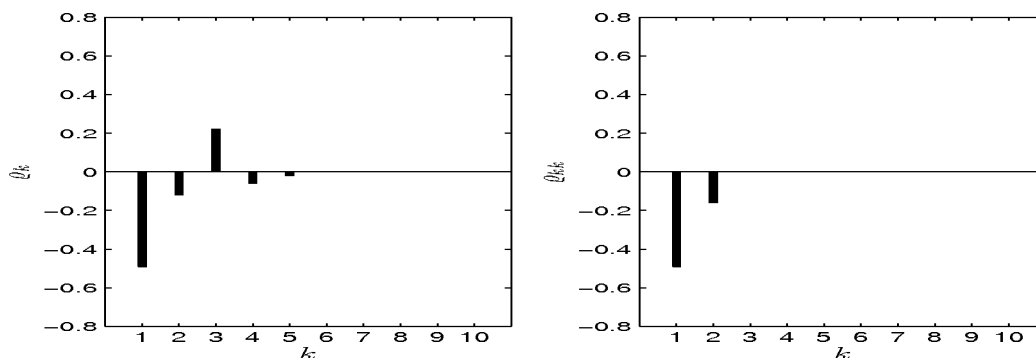
Pro autokorelační funkci procesu AR(2) lze odvodit

$$\begin{aligned} i) \quad & |\varrho_1| < 1, \\ ii) \quad & |\varrho_2| < 1, \\ iii) \quad & \varrho_1^2 < \frac{\varrho_2 + 1}{2}. \end{aligned} \quad (81)$$

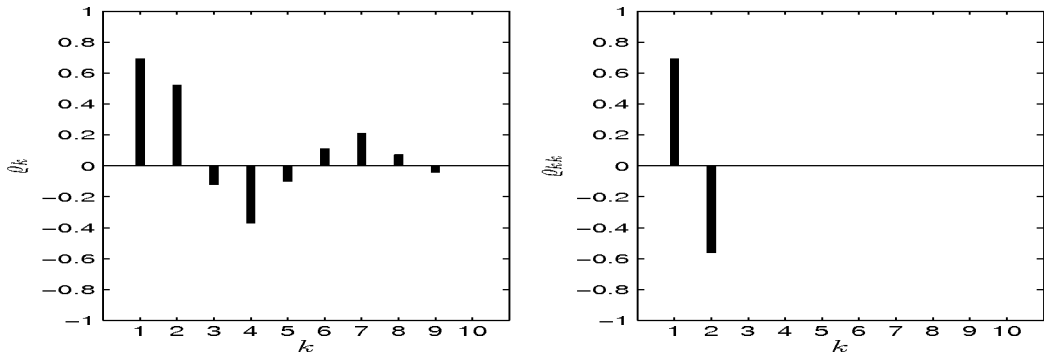
Parciální autokorelační funkce procesu (68) má identifikační bod  $k_0 = 2$ .



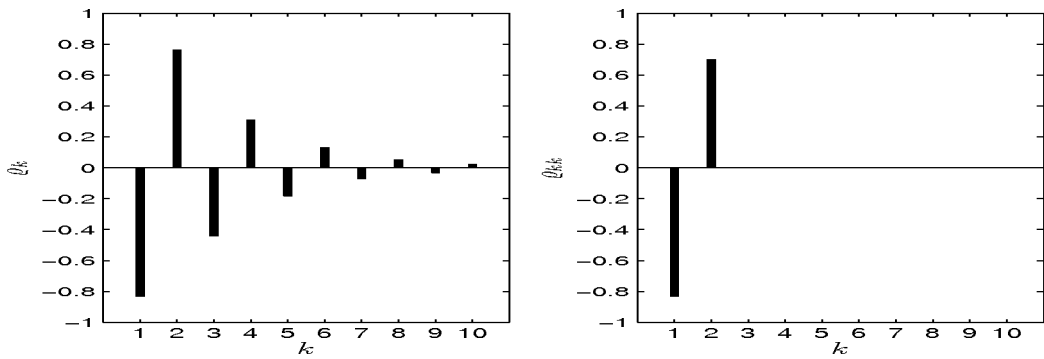
Obrázek 21:  $\varrho_k$  a  $\varrho_{kk}$  pro parametry  $\varphi_1 > 0, \varphi_2 > 0$



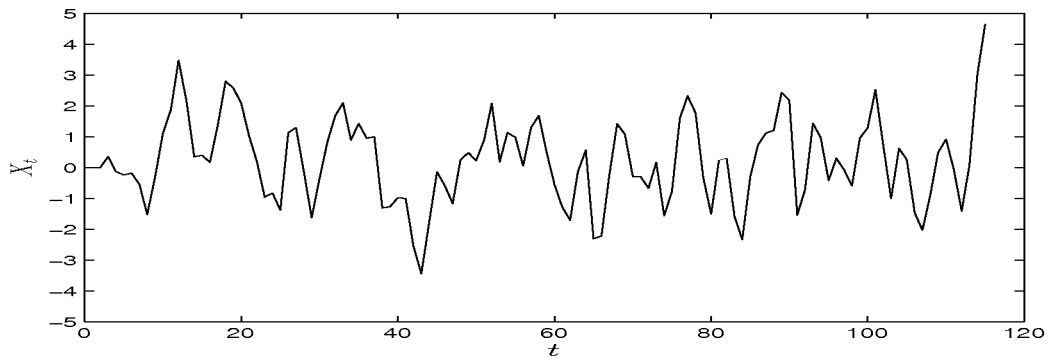
Obrázek 22:  $\varrho_k$  a  $\varrho_{kk}$  pro parametry  $\varphi_1 < 0, \varphi_2 < 0$



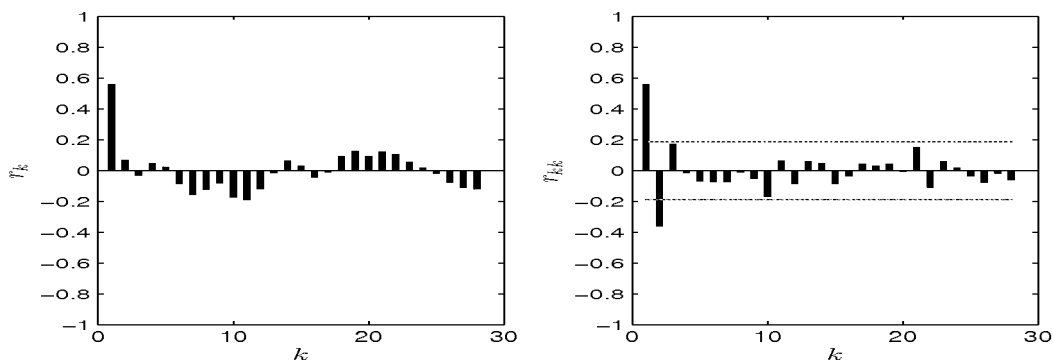
Obrázek 23:  $\rho_k$  a  $\rho_{kk}$  pro parametry  $\varphi_1 > 0, \varphi_2 < 0$



Obrázek 24:  $\rho_k$  a  $\rho_{kk}$  pro parametry  $\varphi_1 < 0, \varphi_2 > 0$



Obrázek 25: Simulovaný proces  $X_t = 0,8X_{t-1} - 0,4X_{t-2} + \varepsilon_t$



Obrázek 26: Odhad autokorelační a parciální autokorelační funkce procesu  $X_t$

Na obrázcích 21-24 jsou znázorněny možné tvary  $\varrho_k$  a  $\varrho_{kk}$  procesu AR(2) pro různé hodnoty parametrů  $\varphi_1$  a  $\varphi_2$ . Obrázek 25 zachycuje časovou řadu 106 hodnot simulovanou na základě procesu AR(2) ve formě  $X_t = 0,8X_{t-1} - 0,4X_{t-2} + \varepsilon_t$ . Na obrázku 26 jsou potom odhady autokorelační a parciální autokorelační funkce této řady.

### 3.4 Smíšený proces

**Definice 3.5.** *Smíšeným procesem řádu  $p$  a  $q$  rozumíme proces daný vztahem*

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad t \in \mathbb{Z}, \quad \varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2). \quad (82)$$

Značíme ho  $ARMA(p, q)$ .

Pomocí operátoru zpětného posunutí můžeme proces (82) zapsat následovně

$$\varphi(B)X_t = \theta(B)\varepsilon_t, \quad (83)$$

kde  $\theta(B)$  je *operátor klouzavých součtů* a  $\varphi(B)$  je *autoregresní operátor*. Podmínka stacionarity je stejná jako u autoregresního procesu (53) a podmínka invertibility je shodná s procesem klouzavých součtů (28), tj. kořeny polynomů (55) a (30) leží vně jednotkového kruhu. Střední hodnota procesu (82) je nulová. Autokorelační funkce splňuje soustavu diferenciálních rovnic

$$\varrho_k = \varphi_1 \varrho_{k-1} + \varphi_2 \varrho_{k-2} + \dots + \varphi_p \varrho_{k-p}, \quad k > q. \quad (84)$$

Pokud  $z_i$  jsou navzájem různé kořeny polynomu (55), má soustava (84) stejné řešení jako pro proces  $\text{AR}(p)$ , tentokrát pro  $k \geq \max\{0, q - p + 1\}$ . Pokud tedy je  $q \geq p$ , pak hodnoty  $\varrho_0, \varrho_1, \dots, \varrho_{q-p}$  se vymykají charakteristickému průběhu autokorelační funkce, která je jinak lineární kombinací geometricky klesajících posloupností a sinusoid s geometricky klesající amplitudou. Parciální autokorelační funkce procesu (82) se pro  $k \geq \max\{1, p - q + 1\}$  chová stejně jako u procesu  $\text{MA}(q)$  a tedy ani pro ni neexistuje identifikační bod.

### Proces $\text{ARMA}(1,1)$

Tento proces má tvar

$$X_t = \varphi_1 X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad \varepsilon_t \sim \mathcal{WN}(0, \sigma_\varepsilon^2). \quad (85)$$

Podmínka stacionarity je stejná jako u procesu  $\text{AR}(1)$ , tedy  $|\varphi_1| < 1$ , obdobně podmínka invertibility je shodná s podmínkou invertibility procesu  $\text{MA}(1)$ , tedy  $|\theta_1| < 1$ . Autokorelační funkce procesu (84) vyhovuje soustavě

$$\varrho_k = \varphi_1 \varrho_{k-1}, \quad k > 1. \quad (86)$$

Vynásobením vztahu (85) veličinou  $X_t$  a následným přechodem ke střední hodnotě získáme vztah následující pro rozptyl

$$\gamma_0 = \varphi_1 \gamma_1 + \sigma_\varepsilon^2 + \theta_1 \sigma_\varepsilon^2 (\varphi_1 + \theta_1). \quad (87)$$

Stejnou úpravou ale tentokrát s veličinou  $X_{t-1}$  dostaneme vztah pro kovarianci mezi veličinami  $X_t$  a  $X_{t-1}$

$$\gamma_1 = \varphi_1 \gamma_0 + \theta_1 \sigma_\varepsilon^2. \quad (88)$$

Dosazením (87) do (88) a opačně dosazením (88) do (87) tyto vztahy upravíme na tvary

$$\sigma_X^2 = \gamma_0 = \frac{1 + \theta_1^2 + 2\varphi_1\theta_1}{1 - \varphi_1^2} \sigma_\varepsilon^2, \quad (89)$$

$$\gamma_1 = \frac{(1 + \varphi_1\theta_1)(\varphi_1 + \theta_1)}{1 - \varphi_1^2} \sigma_\varepsilon^2. \quad (90)$$

Pomocí těchto vztahů můžeme spočítat hodnotu

$$\varrho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 + \varphi_1\theta_1)(\varphi_1 + \theta_1)}{1 + \theta_1^2 + 2\varphi_1\theta_1} \quad (91)$$

a

$$\varrho_2 = \varphi_1\varrho_1.$$

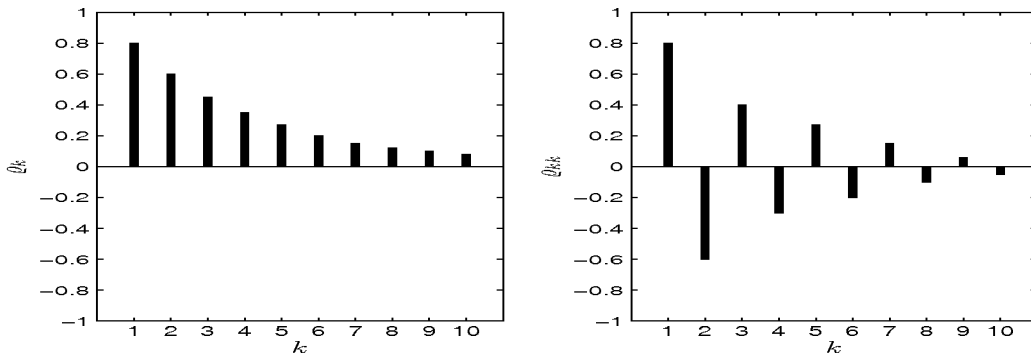
Odtud lze zpětně určit hodnoty parametrů  $\varphi_1$  a  $\theta_1$  takto

$$\begin{aligned} \varphi_1 &= \frac{\varrho_1}{\varrho_2}, \\ \theta_1 &= \frac{b \pm \sqrt{b^2 - 4}}{2}, \text{ kde } b = \frac{1 - 2\varrho_2 + \varphi_1^2}{\varrho_1 - \varphi_1} \end{aligned} \quad (92)$$

a znaménko + nebo - volíme tak, aby  $|\theta_1| < 1$ . Z předchozího vztahu lze odvodit, že pro stacionární a invertibilní proces musí platit

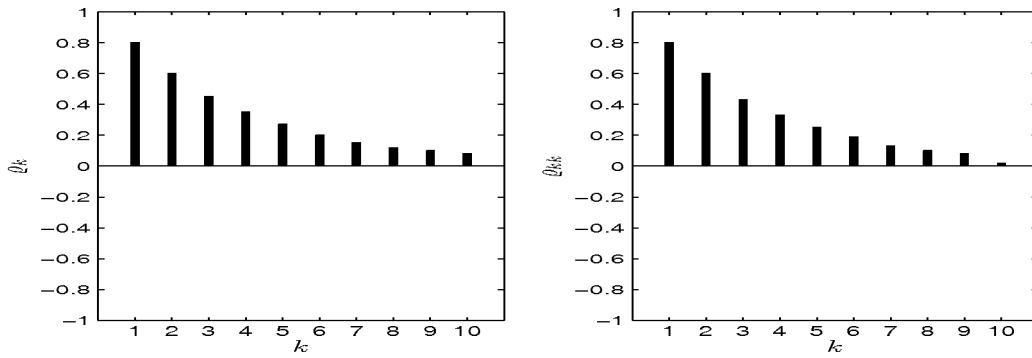
$$2\varrho_1^2 - |\varrho_1| < \varrho_2 < |\varrho_1|. \quad (93)$$

Parciální autokorelační funkce procesu (85) je omezena geometricky klesající posloupností stejně jako u procesu (36).

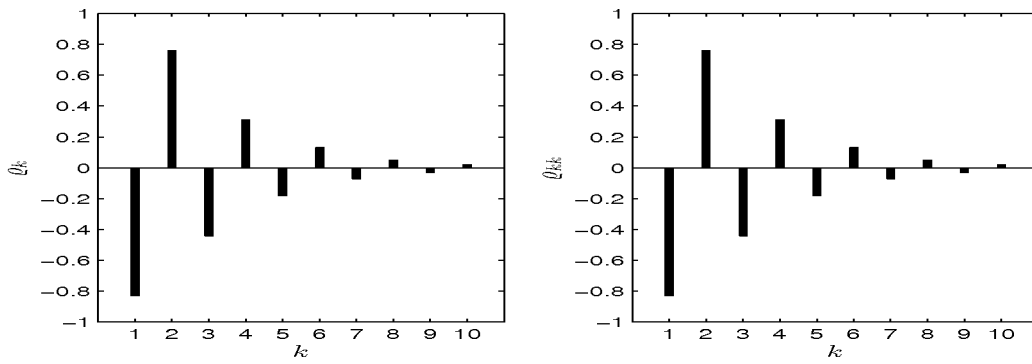


Obrázek 27:  $\varrho_k$  a  $\varrho_{kk}$  pro parametry  $\theta_1 > 0, \varphi_1 > 0$

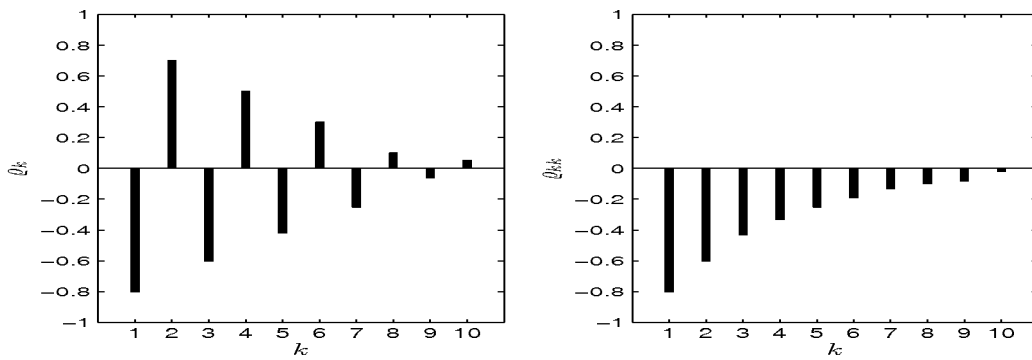
Na obrázcích 27-30 jsou znázorněny možné tvary  $\varrho_k$  a  $\varrho_{kk}$  procesu ARMA(1,1) pro různé hodnoty parametrů  $\theta_1$  a  $\varphi_1$ .



Obrázek 28:  $q_k$  a  $q_{kk}$  pro parametry  $\theta_1 < 0 < \varphi_1$

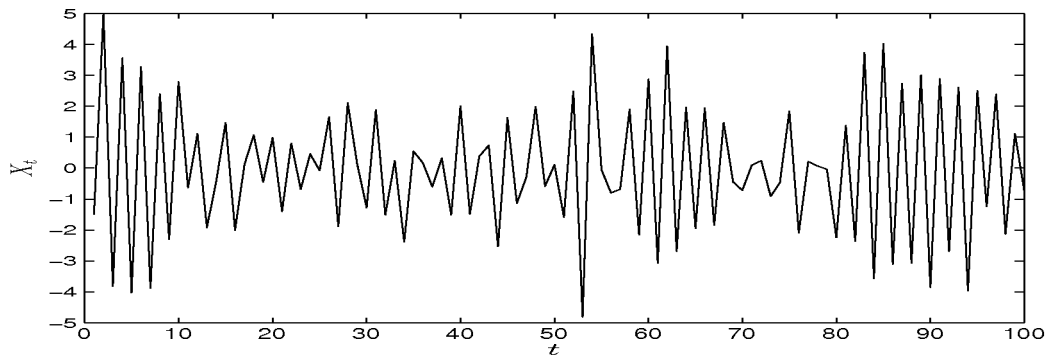


Obrázek 29:  $q_k$  a  $q_{kk}$  pro parametry  $\varphi_1 < 0 < \theta_1$

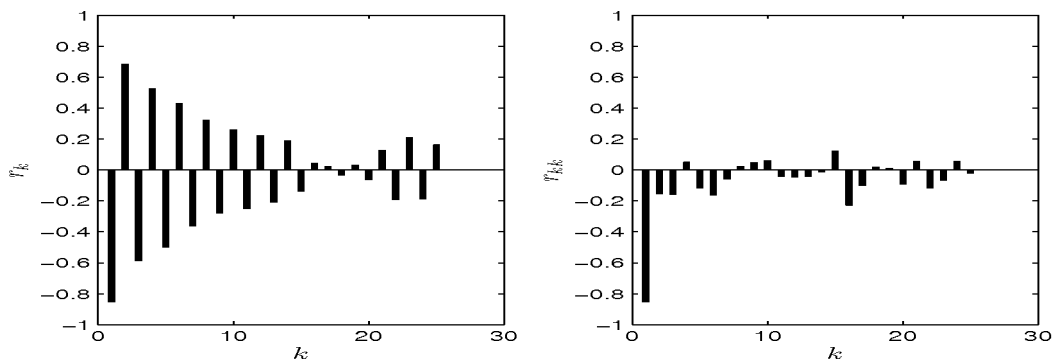


Obrázek 30:  $q_k$  a  $q_{kk}$  pro parametry  $\varphi_1 < 0, \theta_1 < 0$





Obrázek 31: Simulovaný proces  $X_t = -0,8X_{t-1} + \varepsilon_t - 0,4\varepsilon_{t-1}$



Obrázek 32: Odhad autokorelační a parciální autokorelační funkce procesu  $X_t$

Obrázek 31 zachycuje časovou řadu 101 hodnot simulovanou na základě procesu ARMA(1,1) ve formě  $X_t = -0,8X_{t-1} + \varepsilon_t - 0,4\varepsilon_{t-1}$ . Na obrázku 32 jsou potom odhady autokorelační a parciální autokorelační funkce této řady. Celkem snadno lze nalézt podobnost mezi těmito odhady a průběhy funkcí na obr. 30.

**Poznámka 3.2.** V Boxově-Jenkinsově metodologii se snažíme nalézt modely s co nejmenším počtem parametrů. Ve většině praktických případů vystačíme s modely popsány v této kapitole, tj. modely, ve kterých je  $p + q \leq 2$ .

**Poznámka 3.3.** Pokud není časová řada centrována, provedeme centrování

$$X_t \rightsquigarrow X_t - \mu.$$

*Například model  $AR(p)$  centrujeme následovně*

$$X_t - \mu = \varphi_1(X_{t-1} - \mu) + \dots + \varphi_p(X_{t-p} - \mu) + \varepsilon_t, \quad (94)$$

*což lze psát zkráceně ve tvaru*

$$X_t = \delta + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t, \quad (95)$$

*kde*

$$\delta = \mu(1 - \varphi_1 - \dots - \varphi_p). \quad (96)$$

*Hodnotu  $\mu$  odhadujeme buď aritmetickým průměrem dané časové řady, nebo společně s parametry  $\varphi_i, \theta_j, i = 1, \dots, p, j = 1, \dots, q$ .*

## 4 Výstavba modelů

### 4.1 Identifikace modelu

Tato fáze je nejdůležitější a zároveň nejobtížnější při analýze časové řady. V této fázi určujeme jaký typ a řád modelu zvolit pro danou časovou řadu.

Ze všeho nejdříve je vhodné:

- Pořídít grafický záznam dat, pomocí kterého můžeme ohodnotnit stacionaritu (kolísání kolem konstantní úrovně, stejný rozptyl).
- Centrovat stacionární řadu, pokud její střední hodnota není nulová.

Nejsme-li si jistí nulovostí střední hodnoty, lze testovat hypotézu  $H_0 : \mu = 0$ . Nejčastěji se používá test založený na srovnání aritmetického průměru  $\bar{X}$  s dvojnásobkem odhadnuté směrodatné odchylky tohoto průměru  $\sigma_{\bar{X}}$ . Je-li  $|\bar{X}| > 2\sigma_{\bar{X}}$ , zamítneme hypotézu  $H_0$  na hladině 95%. Odhady směrodatných odchylek pro jednotlivé modely jsou uvedeny v tabulce 3.

Tyto odhady jsme získali pomocí následující aproximace

$$\text{var}(\bar{X}) \doteq \frac{\gamma_0}{n} \left(1 + 2 \sum_{k=1}^n \varrho_k\right) =: \sigma_{\bar{X}}^2, \quad (97)$$

odvozené takto

$$\begin{aligned} \text{var}(\bar{X}) &= \text{var}\left(\frac{1}{n} \sum_{t=1}^n X_t\right) = \frac{1}{n^2} \sum_{t=1}^n \sum_{s=1}^n \text{cov}(X_t, X_s) = \frac{1}{n^2} \left( \sum_{t=1}^n \text{var}(X_t) + \right. \\ &\quad \left. + 2 \sum_{t=1}^{n-1} \text{cov}(X_t, X_{t+1}) + 2 \sum_{t=1}^{n-2} \text{cov}(X_t, X_{t+2}) + \dots + 2\text{cov}(X_1, X_n) \right) = \\ &= \frac{1}{n^2} (n\gamma_0 + 2(n-1)\gamma_1 + 2(n-2)\gamma_2 + \dots + 2\gamma_n) \doteq \frac{1}{n} (\gamma_0 + 2\gamma_1 + \\ &\quad + 2\gamma_2 + \dots + 2\gamma_n) = \frac{1}{n} (\gamma_0 + 2 \sum_{k=1}^n \gamma_k) = \frac{\gamma_0}{n} \left(1 + 2 \sum_{k=1}^n \varrho_k\right). \end{aligned} \quad (98)$$

Model	Odhad $\sigma_{\bar{X}}$
MA(1)	$\sqrt{\frac{c_0(1+2r_1)}{n}}$
MA(2)	$\sqrt{\frac{c_0(1+2r_1)}{n}}$
AR(1)	$\sqrt{\frac{c_0(1+r_1)}{n(1-r_1)}}$
AR(2)	$\sqrt{\frac{c_0(1+r_1)(1-2r_1^2+r_2)}{n(1-r_1)(1-r_2)}}$
ARMA(1,1)	$\sqrt{\frac{c_0}{n} \left(1 + \frac{2r_1^2}{r_1-r_2}\right)}$

Tabulka 3: Odhady směrodatných odchylek pro průměr

Výpočet  $\sigma_{\bar{X}}^2$  pro konkrétní modely potom probíhá takto:

MA(1)

Pro hodnoty autokorelační funkce platí (viz. kap. 3.2)  $\varrho_k = 0$ ,  $k = 2, 3, \dots$

Rozptyl je tedy dán vztahem

$$\sigma_{\bar{X}}^2 = \frac{\gamma_0}{n} (1 + 2\varrho_1). \quad (99)$$

MA(2)

Obdobně jako u procesu MA(1) zde platí  $\varrho_k = 0$ ,  $k = 3, 4, \dots$

Pro rozptyl potom platí

$$\sigma_{\bar{X}}^2 = \frac{\gamma_0}{n} (1 + 2(\varrho_1 + \varrho_2)). \quad (100)$$

### AR(1)

Víme, že (viz. kap. 3.3)  $\varrho_k = \varphi_1^k$ ,  $k = 1, 2, \dots$ . Sečteme-li na obou stranách

přes  $k$ , dostaneme  $\sum_{k=1}^{\infty} \varrho_k = \sum_{k=1}^{\infty} \varphi_1^k = \frac{\varphi_1}{1 - \varphi_1} = \frac{\varrho_1}{1 - \varrho_1}$ . Dosazením do (97)

získáme vztah

$$\sigma_{\bar{X}}^2 = \frac{\gamma_0}{n} \left( 1 + 2 \frac{\varrho_1}{1 - \varrho_1} \right) = \frac{\gamma_0(1 + \varrho_1)}{n(1 - \varrho_1)}. \quad (101)$$

### AR(2)

Zde platí

$$\varrho_1 = \varphi_1 + \varphi_2 \varrho_1, \quad \varphi_1 = \varrho_1 \frac{1 - \varrho_2}{1 - \varrho_1^2},$$

$$\varrho_2 = \varphi_1 \varrho_1 + \varphi_2, \quad \varphi_2 = \frac{\varrho_2 - \varrho_1^2}{1 - \varrho_1^2},$$

$$\varrho_k = \varphi_1 \varrho_{k-1} + \varphi_2 \varrho_{k-2}, \quad k = 3, 4, \dots$$

Sečteme-li rovnice na levé straně přes  $k$ , získáme následující rovnost

$$\sum_{k=1}^{\infty} \varrho_k = \varphi_1 \left( \sum_{k=1}^{\infty} \varrho_k + 1 \right) + \varphi_2 \left( \sum_{k=1}^{\infty} \varrho_k + \varrho_1 + 1 \right),$$

po úpravě dostaneme  $\sum_{k=1}^{\infty} \varrho_k = \frac{\varphi_1 + \varphi_2 + \varphi_2 \varrho_1}{1 - \varphi_1 - \varphi_2}$ , dosadíme do (97) a máme

pro rozptyl následující vztah

$$\sigma_{\bar{X}}^2 = \frac{\gamma_0}{n} \frac{(1 + \varrho_1)((1 + \varrho_2 - 2\varrho_1^2))}{((1 - \varrho_1)(1 - \varrho_2))}. \quad (102)$$

### ARMA(1,1)

Pro tento proces platí  $\varrho_k = \varphi_1 \varrho_{k-1}$ ,  $k = 2, 3, \dots$ , což je rekurentně zadaná

geometrická řada, její součet tedy je  $\sum_{k=1}^{\infty} \varrho_k = \frac{\varrho_1}{1 - \varphi_1}$ . Za  $\varphi_1$  dosadíme  $\frac{\varrho_2}{\varrho_1}$

a rozptyl je potom ve tvaru

$$\sigma_{\bar{X}}^2 = \frac{\gamma_0}{n} \left( 1 + 2 \frac{\varrho_1^2}{\varrho_1 - \varrho_2} \right). \quad (103)$$

Vlastní identifikace je založena na odhadech autokorelační funkce  $r_k$  a parciální autokorelační funkce  $r_{kk}$ . Ty porovnáme s průběhy teoretické autokorelační funkce  $\varrho_k$  a teoretické parciální autokorelační funkce  $\varrho_{kk}$ , které jsou uvedeny v tabulce 4. Symbolem  $\smile$  je označena křivka ve tvaru lineární kombinace klesajících geometrickým posloupností a sinusoid s geometricky klesající amplitudou.

	MA( $p$ )	AR( $q$ )	ARMA( $p, q$ )
$\varrho_k$	$k_0 = q$	neexistuje $k_0$ , $\varrho_k$ má tvar $\smile$	neexistuje $k_0$ , $\varrho_k$ má tvar $\smile$ po prvních $q - p$ hodnotách
$\varrho_{kk}$	neexistuje $k_0$ , $\varrho_{kk}$ omezená $\smile$	$k_0 = p$	neexistuje $k_0$ , $\varrho_{kk}$ omezená $\smile$ po prvních $p - q$ hodnotách

Tabulka 4: Tvar autokorelační a parciální autokorelační funkce

Pro praktické výpočty je třeba mít časovou řadu o délce minimálně  $n = 50$  a odhad autokorelační funkce  $r_k$  a parciální autokorelační funkce  $r_{kk}$  tvořený asi  $\frac{n}{4}$  hodnotami.

Pokud se rozhodneme pro nějaký model, vypočítáme hodnoty parametrů podle tabulky 5. Tyto odhady parametrů nelze použít jako výsledné. Použijeme je ale jako počáteční hodnoty iterací v následující fázi. Musíme také zkontrolovat, jestli jsou splněna omezení pro daný model. Ta jsou uvedena ve třetím sloupci téže tabulky.

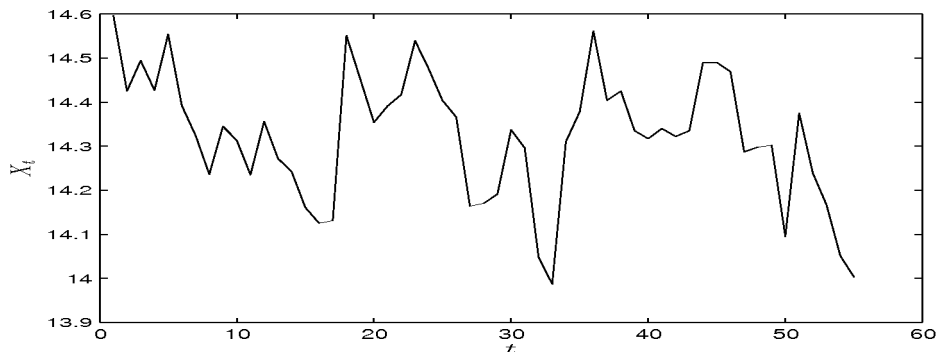
Model	Počáteční odhady	omezení na $r_k$	
MA(1)	$\hat{\theta}_1 = \frac{1 - \sqrt{1 - 4r_1^2}}{2r_1},$ $\hat{\sigma}_\varepsilon^2 = \frac{c_0}{1 + \hat{\theta}_1^2}$	$ r_1  < \frac{1}{2}$	
MA(2)	iteračně nebo pomocí grafického diagramu nebo $\hat{\theta}_1 = \hat{\theta}_2 = 0, 1$ $\hat{\sigma}_\varepsilon^2 = \frac{c_0}{1 + \hat{\theta}_1^2 + \hat{\theta}_2^2}$	$r_2 + r_1 > -\frac{1}{2},$ $r_2 - r_1 > -\frac{1}{2},$	
AR(1)	$\hat{\varphi}_1 = r_1,$ $\hat{\sigma}_\varepsilon^2 = c_0(1 - \hat{\varphi}_1 r_1)$	$ r_1  < 1$	
AR(2)	$\hat{\varphi}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2},$ $\hat{\varphi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2},$ $\hat{\sigma}_\varepsilon^2 = c_0(1 - \hat{\varphi}_1 r_1 - \hat{\varphi}_2 r_2)$	$ r_2  < 1,$ $r_1^2 < \frac{r_2 + 1}{2}$	
ARMA(1,1)	$\hat{\varphi}_1 = \frac{r_2}{r_1},$ $\hat{\sigma}_\varepsilon^2 = \frac{c'_0}{1 + \hat{\theta}_1^2},$ kde $c'_0 = \frac{1}{n} \sum_{t=1}^n n(y'_t - \bar{y}')^2$	$\hat{\theta}_1 = \frac{\hat{b} \pm \sqrt{\hat{b}^2 - 4}}{2},$ kde $\hat{b} = \frac{1 - 2r_2 + \varphi_1^2}{r_1 - \hat{\varphi}_1},$ znaménko volíme tak, aby $ \hat{\theta}_1  < 1,$ $y'_t = y_t - \hat{\varphi}_1 y_{t-1}$	$2r_1^2 -  r_1  < r_2 <  r_1 $

Tabulka 5: Tabulka počátečních odhadů a omezení na autokorelační funkci

#### Příklad 4.1.

Na obrázku 33 je znázorněna řada  $X_t$  kurzu CZK/AUD, jejíž hodnoty jsou uvedeny v tabulce 1. Řada není centrovaná, ale střední hodnota je zřejmě konstatní, rozptyl se s rostoucím časem nezvětšuje.

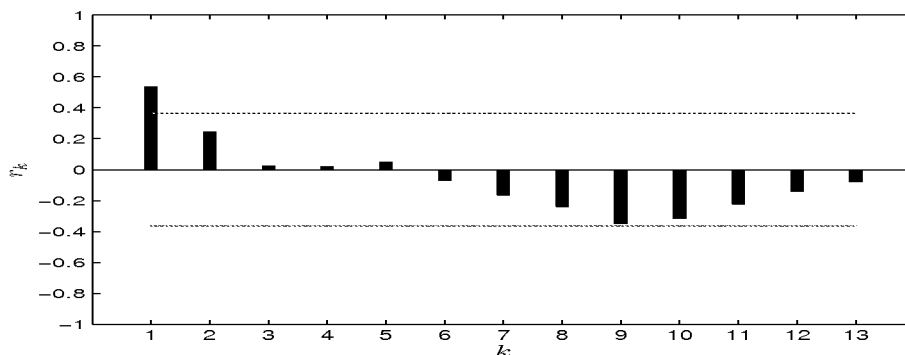
Nejdříve řadu centrujeme. Střední hodnotu odhadneme aritmetickým prů-



Obrázek 33: Graf hodnot z tabulky 1

měrem jako  $\bar{X} = 14,323$ , modelovat potom budeme centrovanou řadu

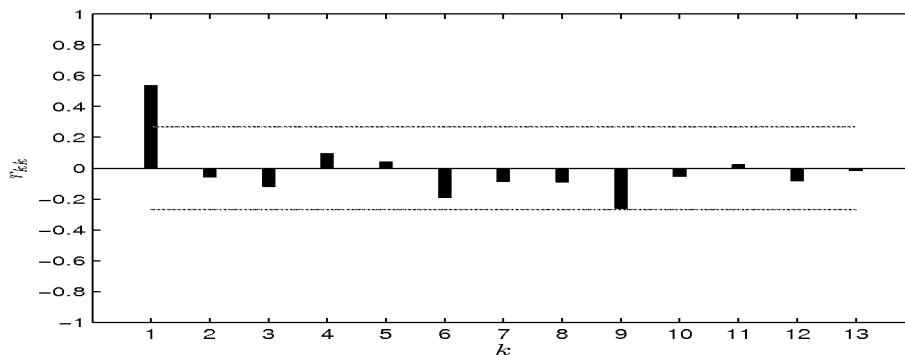
$$X_t^* = X_t - 14,323, \quad t = 1, 2, \dots, 55. \quad (104)$$



Obrázek 34: Odhad autokorelační funkce  $\varrho_k$

Na obr. 34 a 35 jsou uvedeny odhady autokorelační funkce  $\varrho_k$  a  $\varrho_{kk}$ . U obou grafů je znázorněn pás daný dvojnásobkem odpovídající směrodatné odchylky pomocí (16), resp. (17). Identifikační body těchto funkcí jsme určili v příkladu 2.1. Připomeňme, že v případě autokorelační funkce je to bod  $k_0 = 2$ , pro parciální autokorelační funkci potom bod  $k_0 = 1$ . Z grafu odhadnuté autokorelační funkce  $r_k$  (obr. 34) lze ovšem soudit, že funkce  $\varrho_k$  je omezena sinusoidou s klesající amplitudou. Naproti tomu odhadnutá parciální autokorelační funkce  $r_{kk}$  má výrazně nenulovou pouze první hodnotu, kterou následují nízké hodnoty. Tyto skutečnosti





Obrázek 35: Odhad parciální autokorelační funkce  $q_{kk}$

ukazují na to, že řadu bychom mohli modelovat procesem AR(1). Průběh obou funkcí lze porovnat s obr.(14) na straně 30.

Podmínka procesu AR(1),  $|r_1| < 1$ , je splněna. Počáteční odhady jsou

$$\begin{aligned}\hat{\varphi}_1 &= r_1 = 0,5346, \\ \hat{\sigma}_\varepsilon &= c_0(1 - \hat{\varphi}_1 r_1) = 0,0209(1 - 0,5346^2) = 0,0149.\end{aligned}\quad (105)$$

## 4.2 Odhad parametrů

Po identifikaci (nalezení hodnot  $p, q$ ) modelu  $\varphi(B)X_t = \theta(B)\varepsilon_t$ , kde  $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$  a  $\theta(B) = \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ , je třeba odhadnout vektor parametrů

$$(\boldsymbol{\varphi}, \boldsymbol{\theta}) = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q).$$

### 4.2.1 Podmíněná metoda nejmenších nelineárních čtverců

Označme  $(\hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{\theta}})$  vektor těch parametrů, které v oblasti stacionarity a invertibility minimalizují funkci

$$S_P(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \sum_{t=1}^n \varepsilon_t^2(\boldsymbol{\varphi}, \boldsymbol{\theta}), \quad (106)$$

kde

$$\varepsilon_t(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \theta^{-1}(B)\varphi(B)X_t \quad (107)$$

jsou odhadnuté hodnoty bílého šumu. Název metody pochází z toho, že (107) je *nelineární* funkce parametrů  $\varphi$  a  $\theta$ .

Předpokládáme centrováný model ( $E(X_t) = 0$ ).

**Poznámka 4.1.** *Pokud není model centrováný, musíme odhadovat i hodnotu parametru  $\mu$  s počáteční hodnotou  $\bar{X}$ .*

K výpočtu reziduí  $\varepsilon_t(\varphi, \theta)$  nepoužíváme vztah (107), ale rekurentní vztah

$$\varepsilon_t(\varphi, \theta) = X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} - \theta_1 \varepsilon_{t-1}(\varphi, \theta) - \dots - \theta_q \varepsilon_{t-q}(\varphi, \theta). \quad (108)$$

Je nutné vhodně zvolit počáteční hodnoty  $X_{1-p}, \dots, X_0, \varepsilon_{1-q}, \dots, \varepsilon_0$ . Místo neznámých hodnot lze dosadit nuly (za předpokladu, že proces  $X_t$  je stacionární a  $\varepsilon_t$  je proces bílého šumu, jsou to střední hodnoty těchto veličin).

**Poznámka 4.2.** *Pokud leží kořeny polynomu  $\varphi(B)$  blízko jednotkové kružnice, může se skutečně naměřená hodnota  $X_1$  značně lišit od nuly. V tomto případě by byla hodnota  $\hat{\varepsilon}_1$  vysoká a došlo by ke zkreslení procedury.*

Modifikace této metody spočívá v tom, že nahradíme nulami veličiny  $\varepsilon_{p-q+1}, \dots, \varepsilon_p$  a hodnoty veličin  $\varepsilon_{p+1}, \dots, \varepsilon_n$  vypočítáme dle vztahu (108). Nevýhodou této modifikace je, že přijdeme o prvních  $p$  sčítanců.

#### Příklad 4.2.

Podmíněnou metodou nejmenších nelineárních čtverců odhadneme parametr  $\varphi_1$  modelu AR(1) pro data z příkladu 2.1 na straně 12. Zvolíme  $\hat{\varepsilon}_1 = 0$  a další odhady reziduí spočítáme pro dané  $\hat{\varphi}_1$  pomocí (108), takže například pro  $\hat{\varphi}_1 = -0,5$  bude

$$\hat{\varepsilon}_2 = X_2 + 0,5X_1 = 0,102 + 0,5 \cdot 0,274 = 0,239$$

$$\hat{\varepsilon}_3 = X_3 + 0,5X_2 = 0,171 + 0,5 \cdot 0,102 = 0,222$$

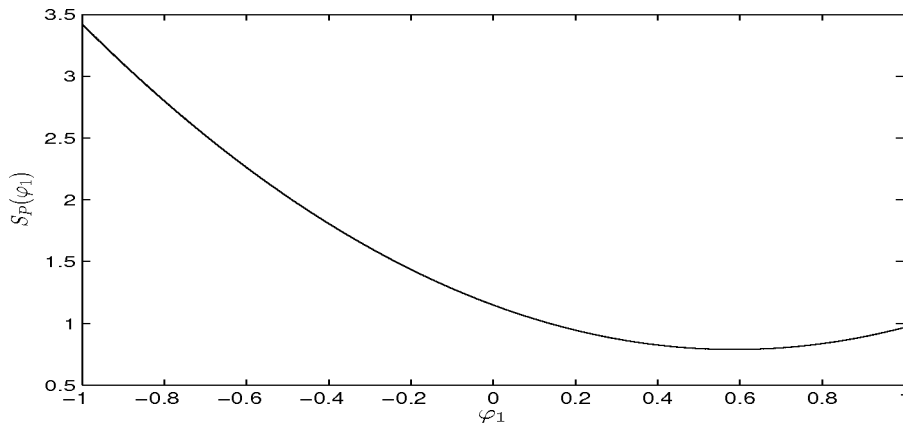
atd. až

$$\hat{\varepsilon}_3 = X_{55} + 0,5X_{54} = -0,32 + 0,5(-0,272) = -0,456$$

Součet čtverců odpovídající odhadu  $\hat{\varphi}_1 = -0,5$  bude

$$S_P(-0,5) = 0,239^2 + 0,222^2 + \dots + (-0,456)^2 = 2,0236. \quad (109)$$

Na obr. 36 je zobrazen průběh funkce  $S_P(\varphi_1)$  na intervalu  $\langle -1, 1 \rangle$ .



Obrázek 36: Součty čtverců pro parametr  $\varphi_1$  na intervalu  $\langle -1, 1 \rangle$

Minimální hodnotu funkce  $S_P(\varphi_1) = 0,7879$  dává odhad  $\hat{\varphi}_1 = 0,587$ .

#### 4.2.2 Nepodmíněná metoda nejmenších nelineárních čtverců

V tomto případě hledáme parametry, které minimalizují funkci

$$S_N(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \sum_{t=-\infty}^n [E(\varepsilon_t | X_1, \dots, X_n)]^2, \quad (110)$$

kde  $E(\varepsilon_t | X_1, \dots, X_n)$  je podmíněná střední hodnota veličiny  $\varepsilon_t$  při pevných hodnotách  $X_1, \dots, X_n$  (dále ji budeme značit krátce  $[\varepsilon_t]$ ), jsou-li  $\boldsymbol{\varphi}$  a  $\boldsymbol{\theta}$  hodnoty parametrů v uvažovaném modelu.

**Poznámka 4.3.** V nepodmíněné metodě se používají podmíněné střední hodnoty. Ty jsou ale počítány pomocí známých hodnot  $X_1, \dots, X_n$ , kdežto v podmíněné metodě se tyto střední hodnoty počítají v závislosti na neznámých počátečních hodnotách.

Abychom mohli zkonstruovat výraz (110), potřebujeme získat odhady hodnot  $X_{-\infty}, \dots, X_{-1}, X_0$ . K tomu použijeme tzv. *zpětné extrapolování*. Princip této metody spočívá v tom, že proces ARMA( $p, q$ ) lze zapsat také pomocí *operátoru posunutí vpřed* jako

$$\varphi(F)X_t = \theta(F)\eta_t, \quad (111)$$

kde  $\varphi(F) = 1 - \varphi_1 F - \dots - \varphi_p F^p$  a  $\theta(F) = 1 + \theta_1 F + \dots + \theta_q F^q$ .  $F$  je inverzní operátor k operátoru  $B$ , tj.  $FX_t = X_{t+1}$ . Protože je proces  $X_t$  stacionární, má proces (111) stejnou autokovarianční strukturu jako proces (83), tudíž  $\{\eta_t\}$  je také proces bílého šumu se stejným rozptylem jako proces bílého šumu  $\varepsilon_t$  a lze psát

$$X_t = \varphi_1 X_{t+1} + \varphi_2 X_{t+2} + \dots + \varphi_p X_{t+p} + \eta_t + \theta_1 \eta_{t+1} + \dots + \theta_q \eta_{t+q}. \quad (112)$$

Jedná se o tzv. *zpětný model*, protože generuje časovou řadu ve směru proti postupu času, kdežto model (82) je tzv. *přímý model*.

Nejprve vypočítáme podmíněné střední hodnoty  $[\eta_t]$ ,  $t = 1, 2, \dots, n-p$ , takto

$$[\eta_t] = [X_t] - \varphi_1 [X_{t+1}] - \dots - \varphi_p [X_{t+p}] - \theta_1 [\eta_{t+1}] - \dots - \theta_q [\eta_{t+q}], \quad (113)$$

kde  $[X_t] = X_t$  pro  $t = 1, 2, \dots, n$ ,  $[\eta_t] = 0$  pro  $t = n-p+1, \dots, n-p+q$  a také pro  $t = 0, -1, -2, \dots$ . Nyní můžeme vypočítat podmíněné střední hodnoty

$$[X_0] = \varphi_1 [X_1] + \dots + \varphi_p [X_p] + \theta_1 [\eta_1] + \dots + \theta_q [\eta_q], \quad (114)$$

$$[X_{-1}] = \varphi_1 [X_0] + \dots + \varphi_p [X_{p-1}] + \theta_2 [\eta_1] + \dots + \theta_q [\eta_{q-1}], \quad (115)$$

$$[X_{-2}] = \varphi_1 [X_{-1}] + \dots + \varphi_p [X_{p-2}] + \theta_3 [\eta_1] + \dots + \theta_q [\eta_{q-2}], \quad (116)$$

⋮

V praktických výpočtech samozřejmě nepokračujeme až k  $[X_{-\infty}]$ , ale protože zpětné extrapolace  $[X_{-i}]$  se s rostoucím  $i$  blíží ke střední hodnotě procesu, ukončíme extrapolování ve chvíli, kdy se přiblížíme se zvolenou přesností nule. Jednou z variant, kdy zastavit, je situace, kdy tři po sobě jdoucí hodnoty  $[X_{-i}]$  jsou menší než 5%  $\sigma_X$ . Potom považujeme hodnoty  $\varepsilon_{-i-1}, \varepsilon_{-i-2}, \dots$  za nezávislé na hodnotách  $X_1, \dots, X_n$  a položíme je rovny nule. Nakonec pomocí (108) spočítáme  $[\varepsilon_t]$ ,

$t = -i, \dots, n$ . Ty dosadíme do vztahu (110). Pro  $t > n$  je  $[\varepsilon_t] = 0$ .

Pro řady, které mají délku alespoň 75, je rozdíl mezi podmíněnou a nepodmíněnou metodou zanedbatelný.

### Příklad 4.3.

Nepodmíněnou metodu nejmenších nelineárních čtverců ukážeme na stejných datech a pro stejný parametr jako v předchozím příkladu. V tabulce 6 jsou uve-

$t$	$[X_t]$	$[\eta_t]$	$[\varepsilon_t]$
-7	-0,001	0	0
-6	-0,002	0	-0,001
-5	0,004	0	0,003
-4	-0,009	0	-0,006
-3	0,017	0	0,013
-2	-0,034	0	-0,026
-1	0,069	0	0,051
0	-0,137	0	-0,103
1	0,274	0,325	0,206
2	0,102	0,188	0,239
3	0,171	0,223	0,222
4	0,104	0,220	0,190
$\vdots$	$\vdots$	$\vdots$	$\vdots$
54	-0,272	-0,432	-0,350
55	-0,320	0	-0,456

Tabulka 6: Princip zpětného extrapolování v modelu AR(1) pro  $\varphi_1 = -0,5$

deny hodnoty podmíněných středních hodnot  $[X_t]$ ,  $[\eta_t]$  a  $[\varepsilon_t]$  pro  $t = -7, -6, \dots, 55$ . Zpětné extrapolování ukončíme pro  $t = -7$ , protože  $X_{-5}, X_{-6}, X_{-7} < 0,007 = 0,05\sigma_X$ . Nepodmíněnou metodou nepodmíněných nelineárních čtverců dostaneme pro parametr  $\varphi_1 = -0,5$  součet

$$S_N(-0,5) = (-0,001)^2 + 0,003^2 + (-0,006)^2 + \dots + (-0,456)^2 = 2,0048. \quad (117)$$

#### 4.2.3 Minimalizace součtu nelineárních čtverců

Hledáme takové parametry  $\varphi, \theta$ , které minimalizují výraz  $S(\varphi, \theta)$  získaný jednou z výše uvedených metod.

K tomu, abychom pro odhad parametrů mohli použít metodu nejmenších čtverců, musí být splněny podmínky klasického lineárního modelu (podmínky o reziduální složce  $\varepsilon_t$ ). Uvažujeme-li model  $Y_t = Z_t + \varepsilon_t$ , jsou to tyto podmínky

- $E(\varepsilon_t) = 0$ , střední hodnota je nulová,
- $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ , rozptyl je konstantní,
- $E(\varepsilon_t \varepsilon_s) = 0$  pro  $t \neq s$ , náhodné chyby jsou nezávislé,
- $E(Z_t \varepsilon_t) = 0$ , rezidua a vysvětlující proměnné jsou nezávislé.

Tyto podmínky splňuje například proces AR(1) tvaru  $X_t = \varphi_1 X_{t-1} + \varepsilon_t$ , pro který má odhad parametru  $\varphi_1$  získaný metodou nejmenších čtverců formu

$$\hat{\varphi}_1 = \frac{\sum_{t=2}^n X_{t-1} X_t}{\sum_{t=2}^n X_{t-1}^2} \quad (118)$$

Ale u procesu AR( $p$ ) se již jedná o mnohonásobnou lineární regresi. Zde musíme uvažovat ještě podmínku, že vysvětlující proměnné jsou vzájemně nezávislé (nekorelované). Často u regresních modelů tohoto typu není splněna ani podmínka na nezávislost chyb.

Metodu nejmenších čtverců nelze použít pro odhad parametrů procesu ARMA, protože není lineární v parametrech. V tomto případě se používá například tzv. *Gaussův-Newtonův iterační algoritmus*. Za počáteční odhad  $\boldsymbol{\beta}^0$  bereme vektor  $(\hat{\boldsymbol{\varphi}}, \hat{\boldsymbol{\theta}})$  získaný při identifikaci modelu pomocí vzorců z tabulky 5. Pomocí Taylova rozvoje v bodě  $\boldsymbol{\beta}^0$  linearizujeme funkci  $\varepsilon_t(\boldsymbol{\varphi}, \boldsymbol{\theta}) = \varepsilon_t(\boldsymbol{\beta})$ , která je nelineární v parametrech. Dostaneme vztah

$$\varepsilon_t(\boldsymbol{\beta}) = \varepsilon_t(\boldsymbol{\beta}^0) - \sum_{i=1}^{p+q} (\beta_i - \beta_i^0) x_{ti}(\boldsymbol{\beta}^0), \quad t = p + q, \dots, n, \quad (119)$$

kde

$$x_{ti}(\boldsymbol{\beta}^0) = -\frac{\partial \varepsilon_t(\boldsymbol{\beta}^0)}{\partial \beta_i}. \quad (120)$$

Získáme lineární regresní model

$$\varepsilon(\boldsymbol{\beta}^0) = \mathbf{X}(\boldsymbol{\beta}^0)(\boldsymbol{\beta} - \boldsymbol{\beta}^0) + \varepsilon(\boldsymbol{\beta}), \quad (121)$$

kde

$$\varepsilon(\boldsymbol{\beta}) = (\varepsilon_{p+1}(\boldsymbol{\beta}), \dots, \varepsilon_n(\boldsymbol{\beta})) \quad (122)$$

a

$$\mathbf{X}(\boldsymbol{\beta}^0) = x_{ti}(\boldsymbol{\beta}^0), \quad t = p+1, \dots, n, \quad i = 1, \dots, p+q. \quad (123)$$

Metodou nejmenších čtverců určíme odhad optimálního přírůstku parametrů

$$\widehat{\boldsymbol{\beta} - \boldsymbol{\beta}^0} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon(\boldsymbol{\beta}^0). \quad (124)$$

Odhad (124) vylepšíme

$$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^0 + (\widehat{\boldsymbol{\beta} - \boldsymbol{\beta}^0}) \quad (125)$$

a postup opakujeme, dokud nedosáhneme požadované přesnosti.

Pomocí této iterační metody dovedeme také odhadnout kovarianční matici výsledného odhadu  $\hat{\boldsymbol{\beta}}$  vztahem

$$\text{var}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}_\varepsilon^2 [\mathbf{X}^T(\hat{\boldsymbol{\beta}}) \mathbf{X}(\hat{\boldsymbol{\beta}})]^{-1}, \quad (126)$$

kde  $\hat{\sigma}_\varepsilon^2 = \frac{S(\hat{\boldsymbol{\beta}})}{n}$  resp.  $\hat{\sigma}_\varepsilon^2 = \frac{S(\hat{\boldsymbol{\beta}})}{n-(p+q)}$  je odhad rozptylu procesu  $\varepsilon_t$ . Pomocí vztahu (126) odvodíme aproximace pro odhady směrodatných odchylek odhadů parametrů jednotlivých procesů.

Výpočet  $\text{var}(\hat{\boldsymbol{\beta}})$  pro autoregresní modely:

AR(1)

Model je dán vztahem  $X_t = \varphi_1 X_{t-1} + \varepsilon_t$ ,  $t = 1, \dots, n$ , rezidua jsou ve tvaru

$$\varepsilon_t = X_t - \varphi_1 X_{t-1}, \text{ a parciální derivace potom jsou } x_{t1} = -\frac{\partial \varepsilon_t}{\partial \varphi_1} = X_{t-1}, \text{ tj.}$$

$$\text{matice } \mathbf{X}_{n \times 1} = (X_0, X_1, \dots, X_{n-1})^T, \quad \mathbf{X}^T \mathbf{X} = \sum_{t=0}^{n-1} X_t^2 = n \cdot c_0.$$

Víme, že (viz. tabulka 5)  $\sigma_\varepsilon^2 = \gamma_0(1 - \varphi_1 \varrho_1)$ , kde  $\varrho_1 = \varphi_1$ . Dosadíme do (126)

$$\text{a dostaneme } \text{var}(\hat{\varphi}_1) = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1} = c_0 (1 - \hat{\varphi}_1^2) \frac{1}{nc_0} = \frac{1 - \hat{\varphi}_1^2}{n}. \quad (127)$$

AR(2)

Zde máme model tvaru  $X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varepsilon_t$ ,  $t = 1, \dots, n$ , rezidua jsou tedy  $\varepsilon_t = X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2}$ , parciální derivace jsou tentokrát

$$x_{t1} = -\frac{\partial \varepsilon_t}{\partial \varphi_1} = X_{t-1}, \quad x_{t2} = -\frac{\partial \varepsilon_t}{\partial \varphi_2} = X_{t-2}, \quad \text{ty tvoří matici}$$

$$\begin{aligned} \mathbf{X}_{n \times 2} &= \begin{pmatrix} X_0 & X_{-1} \\ X_1 & X_0 \\ \vdots & \vdots \\ X_{n-1} & X_{n-2} \end{pmatrix}, \quad \mathbf{X}^T \mathbf{X} = \begin{pmatrix} \sum_{t=0}^{n-1} X_t^2 & \sum_{t=0}^{n-1} X_t X_{t-1} \\ \sum_{t=0}^{n-1} X_t X_{t-1} & \sum_{t=0}^{n-1} X_t^2 \end{pmatrix} = \\ &= \begin{pmatrix} nc_0 & nc_1 \\ nc_1 & nc_0 \end{pmatrix} = nc_0 \begin{pmatrix} 1 & -r_1 \\ -r_1 & 1 \end{pmatrix}. \end{aligned}$$

Inverze této matice  $(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{nc_0(1-r_1^2)} \begin{pmatrix} 1 & r_1 \\ r_1 & 1 \end{pmatrix}$  má shodné diagonální prvky a proto platí  $\text{var}(\hat{\varphi}_1) = \text{var}(\hat{\varphi}_2)$ . Ze soustavy (79) odvodíme,

že  $\varrho_1 = \frac{\varphi_1}{1-\varphi_2}$ , a  $\varrho_2 = \frac{\varphi_2 + \varphi_1^2 - \varphi_2^2}{1-\varphi_2}$ . V tomto případě je rozptyl bílého

$$\begin{aligned} \text{šumu } \sigma_\varepsilon^2 &= \gamma_0(1 - \varphi_1 \varrho_1 - \varphi_2 \varrho_2) = \gamma_0 \left( 1 - \varphi_1 \frac{\varphi_1}{1-\varphi_2} - \varphi_2 \frac{\varphi_2 + \varphi_1^2 - \varphi_2^2}{1-\varphi_2} \right) = \\ &= \frac{\gamma_0}{1-\varphi_2} (1 + \varphi_2) [(1-\varphi_2)^2 - \varphi_1^2]. \quad \text{Položíme } r_1 = \hat{\varrho}_1 \text{ a dosadíme do (126)} \end{aligned}$$

$$\begin{aligned} \text{var}(\hat{\varphi}_1) &= \hat{\sigma}^2 [(\mathbf{X}^T \mathbf{X})^{-1}]_{11} = \frac{c_0}{1-\hat{\varphi}_2} (1 + \hat{\varphi}_2) [(1-\hat{\varphi}_2)^2 - \hat{\varphi}_1^2] \frac{1}{1-\hat{\varphi}_1^2} \cdot \frac{1}{nc_0} = \\ &= \frac{1-\hat{\varphi}_2}{n} (1 + \hat{\varphi}_2) = \frac{1-\hat{\varphi}_2^2}{n}. \end{aligned} \quad (128)$$

Při počítání rozptylu vektoru  $\boldsymbol{\beta}$  pro modely klouzavých součtů vycházíme z invertovaného tvaru. Přehled směrodatných odchylek je uveden v tabulce 7.



Model	Směrodatné odchyly odhadnutých parametrů
MA(1)	$\sigma_{\hat{\theta}_1} \sim \sqrt{\frac{1-\hat{\theta}_1^2}{n}}$
MA(2)	$\sigma_{\hat{\theta}_1} = \sigma_{\hat{\theta}_2} \sim \sqrt{\frac{1-\hat{\theta}_2^2}{n}}$
AR(1)	$\sigma_{\hat{\varphi}_1} \sim \sqrt{\frac{1-\hat{\varphi}_1^2}{n}}$
AR(2)	$\sigma_{\hat{\varphi}_1} = \sigma_{\hat{\varphi}_2} \sim \sqrt{\frac{1-\hat{\varphi}_2^2}{n}}$
ARMA(1,1)	$\sigma_{\hat{\varphi}_1} \sim \sqrt{\frac{(1-\hat{\varphi}_1^2)(1+\hat{\varphi}_1\hat{\theta}_1)^2}{n(\hat{\varphi}_1+\hat{\theta}_1)^2}}, \quad \sigma_{\hat{\theta}_1} \sim \sqrt{\frac{(1-\hat{\theta}_1^2)(1+\hat{\varphi}_1\hat{\theta}_1)^2}{n(\hat{\varphi}_1+\hat{\theta}_1)^2}}$

Tabulka 7: Přibližné směrodatné odchyly odhadnutých parametrů

#### Příklad 4.4.

Protože jsme pro data z příkladu 2.1 identifikovali model AR(1), můžeme pro odhad parametru  $\varphi_1$  použít metodou nejmenších čtverců, tj.

$$\hat{\varphi}_1 = \frac{\sum_{t=2}^{55} X_{t-1}X_t}{\sum_{t=2}^{55} X_{t-1}^2} = 0,587. \quad (129)$$

Stejný odhad jsme dostali i podmíněnou metodou nejmenších nelineárních čtverců. Ke shodnému výsledku dospějeme, i když budeme odhad  $\varphi_1$  hledat pomocí Gaussova-Newtonova iteračního algoritmu s počátečním odhadem  $\varphi_1^0 = r_1 = 0,5346$  (viz. tabulka 5). Matice  $\mathbf{X}(\beta^0) = \mathbf{X}(\varphi_1)$  je typu  $54 \times 1$  a  $x_{t1}(\varphi_1) = -\frac{\partial \varepsilon_t(\varphi_1^0)}{\partial \varphi_1} = X_t$ ,  $t = 1, 2, \dots, 54$ . Prvky vektoru  $\varepsilon(\beta^0) = \varepsilon(\varphi_1^0)$  spočítáme pomocí vztahu (108). Optimální přírůstek parametru potom je

$$\widehat{\varphi_1 - \varphi_1^0} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon(\varphi_1^0) = -0,0523.$$

Vylepšený odhad bude

$$\varphi_1^1 = \varphi_1^0 + (\widehat{\varphi_1 - \varphi_1^0}) = 0,5346 + 0,0524 = 0,587, \quad (130)$$

pro něj je optimální přírůstek prakticky nulový, protože

$$\widehat{\varphi_1 - \varphi_1^1} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon(\varphi_1^1) = -2,7756 \cdot 10^{-17}.$$

Směrodatná odchylka odhadu parametru  $\varphi_1$  je podle vztahu z tabulky 7

$$\sigma_{\hat{\varphi}_1} = \sqrt{\frac{1 - \hat{\varphi}_1^2}{n}} = \sqrt{\frac{1 - 0,587^2}{55}} = 0,1092. \quad (131)$$

Přibližný 95% interval spolehlivosti potom je

$$\langle \hat{\varphi}_1 - 2\sigma(\hat{\varphi}_1), \hat{\varphi}_1 + 2\sigma(\hat{\varphi}_1) \rangle = \langle 0,3686; 0,8053 \rangle. \quad (132)$$

### 4.3 Ověřování modelu

V této fázi porovnáváme napozorovanou časovou řadu se zkonstruovaným modelem. Posuzujeme, zda jsme vybrali vhodný model. Metodami uvedenými v dalším textu hodnotíme model z různých hledisek ( $\text{var}(\hat{\beta}), \hat{\sigma}_\varepsilon^2$ ) a s různou účinností.

#### 4.3.1 Metoda přeparametrizování modelu

Je-li nějaká z hodnot  $\text{var}(\hat{\beta})$  nebo  $\hat{\sigma}_\varepsilon^2$  velká, zvýšíme počet parametrů v modelu (např. MA(1)  $\rightarrow$  MA(2)). U rozšířeného modelu vypočítáme odhady  $\hat{\beta}$ , jejich rozptyl a reziduální rozptyl  $\hat{\sigma}_\varepsilon^2$ . Pokud se rozptyly výrazně zmenšily a pokud jsou odhady nově zavedených parametrů nenulové, použijeme rozšířený model.

**Poznámka 4.4.** *Nově zavedený parametr považujeme za nenulový, je-li jeho absolutní hodnota větší než dvojnásobek jeho směrodatné odchylky.*

### 4.3.2 Normalita nesystematické složky

Normalitu nesystematické složky předpokládáme při testování parametrů modelu, při testování autokorelace, při konstrukci předpovědí a v neposlední řadě pro interpretaci parametrů. Jak bylo uvedeno ve druhé kapitole, předpokládáme normalitu dat také proto, že v tomto případě stačí ověřit, že je proces slabě stacionární, abychom zjistili, že je také striktně stacionární.

Jednou z možností testování normality je *Jarqueův-Berův test*. Je založen na současném testování šikmosti a špičatosti. Testujeme hypotézu

$$\varepsilon_t \stackrel{H_0}{\doteq} \mathcal{N}(0, \sigma_\varepsilon^2). \quad (133)$$

Statistika

$$JB = SK^2 + K^2, \quad (134)$$

kde

$$SK = \sqrt{\frac{n\hat{m}_3^2}{6\hat{m}_2^3}} \quad (135)$$

je testové kritérium pro šikmost rozdělení,

$$K = \frac{\hat{m}_4}{\hat{m}_2^2 - 3} \sqrt{\frac{n}{24}} \quad (136)$$

je testové kritérium pro špičatost rozdělení a

$$\hat{m}_j = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^j. \quad (137)$$

Za platosti nulové hypotézy má statistika  $JB$  rozdělení  $\chi_2^2$ .

**Poznámka 4.5.** Příčinou zamítnutí nulové hypotézy může být kromě nenormality  $\varepsilon_t$  také jejich heteroskedasticita (tzn.  $\varepsilon_t$  nemají konstantní rozptyl).

#### Příklad 4.5.

Otestujeme normalitu reziduí z předchozího příkladu. Dosadíme do vztahu (134), tj.

$$\begin{aligned} JB = SK^2 + K^2 &= \frac{n\hat{m}_3^2}{6\hat{m}_2^3} + \left(\frac{\hat{m}_4}{\hat{m}_2^2 - 3}\right)^2 \frac{n}{24} = \frac{55 \cdot 1,17 \cdot 10^{-7}}{6 \cdot 2,18 \cdot 10^{-6}} + \frac{55}{25} \left(\frac{5,76 \cdot 10^{-4}}{1,68 \cdot 10^{-4} - 3}\right) = \\ &= 0,4931 + 8,43 \cdot 10^{-8} \doteq 0,4931, \end{aligned} \quad (138)$$

což je menší než kritická hodnota  $\chi_2^2(0,05) = 5,99$ . Lze tedy předpokládat, že rezidua mají normální rozdělení.

#### 4.3.3 Metoda odhadnutých reziduí

Odhadnutými rezidui rozumíme odhadnuté hodnoty bílého šumu vypočtené ve druhé fázi výstavby modelu, tedy

$$\hat{\varepsilon}_t = \varepsilon_t(\hat{\varphi}, \hat{\theta}) = \hat{\theta}^{-1}(B)\hat{\varphi}(B)X_t \quad (139)$$

$$= X_t - \hat{\varphi}_1 X_{t-1} - \dots - \hat{\varphi}_p X_{t-p} - \hat{\theta}_1 \varepsilon_{t-1} - \dots - \hat{\theta}_q \varepsilon_{t-q}. \quad (140)$$

Tato metoda je založena na zkoumání odhadnuté autokorelační funkce těchto reziduí. Při výpočtu  $r_k(\hat{\varepsilon})$  využíváme toho, že řada  $\hat{\varepsilon}_t$  má nulovou střední hodnotu.

$$r_k(\hat{\varepsilon}) = \frac{\sum_{t=1}^{n-k} \hat{\varepsilon}_t \hat{\varepsilon}_{t+k}}{\sum_{t=1}^n \hat{\varepsilon}_t^2}, \quad k = 1, 2, \dots \quad (141)$$

Pokud jsou rezidua  $\varepsilon_t$  nezávislá a mají normální normované rozdělení ( $E(\varepsilon_t) = 0$ ,  $\sigma_\varepsilon = 1$ ), potom lze při větším  $n$  aproximovat odhadnuté hodnoty  $r_k(\varepsilon)$  nezávislými náhodnými veličinami s rozdělením  $\mathcal{N}(0, \frac{1}{n})$ . Odhadnuté hodnoty  $r_k(\hat{\varepsilon})$  mají také normální rozdělení a nulovou střední hodnotu, ale nejsou nezávislé a mají jiný rozptyl. Proto byla odvozena úprava, která pro proces AR(1) dává vztah

$$\text{var}[r_k(\hat{\varepsilon})] \sim \frac{1 - \varphi_1^{2(k-1)}(1 - \varphi_1^2)}{n}, \quad k = 1, 2, \dots, \quad (142)$$

speciálně pro  $k = 1$  je tedy

$$\text{var}[r_1(\hat{\varepsilon})] \sim \frac{\varphi_2^2}{n}. \quad (143)$$

Pro rozptyl prvních dvou odhadnutých hodnot autokorelační funkce  $r_1(\hat{\varepsilon})$ ,  $r_2(\hat{\varepsilon})$  procesu AR(2), pak platí

$$\text{var}[r_1(\hat{\varepsilon})] \sim \frac{\varphi_2^2}{n}, \quad (144)$$

$$\text{var}[r_2(\hat{\varepsilon})] \sim \frac{\varphi_2^2 + \varphi_1^2(1 + \varphi_2)^2}{n}. \quad (145)$$

Proces ARMA( $p, q$ ) daný tvarem (83), můžeme v tomto případě považovat za proces AR( $p + q$ ) ve tvaru

$$\varphi(B)\theta(B)X_t = \varepsilon_t, \quad (146)$$

protože odhadnutá rezidua jsou v obou případech přibližně stejná. Pro proces MA(1) tedy platí

$$\text{var}[r_k(\hat{\varepsilon})] \sim \frac{1 - \theta_1^{2(k-1)}(1 - \theta_1^2)}{n}, \quad k = 1, 2, \dots \quad (147)$$

Pokud  $|r_k(\hat{\varepsilon})| > 2\sqrt{\text{var}[r_k(\hat{\varepsilon})]}$ , jsou rezidua zřejmě autokorelovaná a model není dobrý.

#### 4.3.4 Portmanteau test

Tento test je další možností, jak zjistit, jestli jsou odhadnutá rezidua neautokorelovaná. Je to metoda méně přesná než předchozí, ale její výhodou je výpočetní jednoduchost. Testujeme hypotézu

$$H_0 : r_k(\hat{\varepsilon}) \sim \mathcal{N}\left(0, \frac{1}{n}\right) \text{ nezávislé}, \quad (148)$$

kde  $K \doteq \sqrt{n}$ . Použijeme tzv. *portmanteau statistiku*

$$Q = n \sum_{k=1}^K r_k^2(\hat{\varepsilon}). \quad (149)$$

Za platnosti nulové hypotézy má statistika  $Q$  pro velká  $n$  přibližně rozdělení  $\chi_{K-p-q}^2$ . Pokud je

$$Q > \chi_{K-p-q}^2(\alpha), \quad (150)$$

zamítneme daný model na hladině  $\alpha$  jako nevhodný.

Pro rozsahy používané v praxi je doporučováno použít jiné statistiky, které dávají lepší výsledky při menším  $n$  a které se od statistiky  $Q$  liší pouze výběrem konstant ve vzorci. Vhodná je například statistika

$$Q^* = n(n+2) \sum_{k=1}^K \frac{r_k^2(\hat{\varepsilon})}{n-k}. \quad (151)$$

#### 4.3.5 Durbinův-Watsonův test

Tento test slouží k ověření nezávislosti reziduí. Testová statistika je zde

$$DW = \frac{\sum_{t=2}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2}. \quad (152)$$

Hodnoty statistiky  $DW$  se pohybují v intervalu od nuly do čtyř. Pokud jsou rezidua nezávislá, pohybuje se tato statistika okolo čísla 2, v případě přímé závislosti reziduí, se hodnota  $DW$  blíží zprava nule, pokud se jedná o nepřímou závislost, blíží se zleva čtyřem.

#### Příklad 4.6.

Pro řadu  $X_t$ , s níž jsme pracovali v předchozích příkladech, jsme zkonstruovali model

$$X_t - 14,323 = 0,587(X_{t-1} - 14,323) + \varepsilon_t, \quad (153)$$

resp.

$$X_t = 5,9154 + 0,587X_{t-1} + \varepsilon_t. \quad (154)$$

Rozptyl reziduí  $\hat{\varepsilon}$  je  $\hat{\sigma}_{\hat{\varepsilon}}^2 = 0,0131$ .

Zda jsou rezidua v tomto modelu neautokorelovaná, ověříme nejprve portman-teau testem. Použijeme odhady  $\hat{\varepsilon}_t$ ,  $t = 2, 3, \dots, 55$ , které jsme získali při hledání odhadu parametru  $\varphi_1$ . Pro rezidua platí  $E(\hat{\varepsilon}_t) = 0$ . Podle vztahu (141) vypočítáme hodnoty autokorelační funkce reziduí. Potom je např.

$$r_1(\hat{\varepsilon}) = \frac{\sum_{t=2}^{54} \hat{\varepsilon}_t \hat{\varepsilon}_{t+1}}{\sum_{t=2}^{55} \hat{\varepsilon}_t^2} = 0,0455.$$

$$r_2(\hat{\varepsilon}) = \frac{\sum_{t=2}^{53} \hat{\varepsilon}_t \hat{\varepsilon}_{t+2}}{\sum_{t=2}^{55} \hat{\varepsilon}_t^2} = 0,0071,$$

atd. Prvních osm hodnot funkce  $r_k(\hat{\varepsilon})$ , které jsou třeba k výpočtu, je uvedeno v tabulce 8.

$k$	1	2	3	4	5	6	7	8
$r_k$	0,0455	0,0071	-0,1704	-0,0617	0,1280	0,0306	-0,0286	-0,0584

Tabulka 8: Prvních osm hodnot funkce  $r_k(\hat{\varepsilon})$

Modifikovaná portmanteau statistika pro  $K = 8$  je

$$Q^* = n(n+2) \sum_{k=1}^K \frac{r_k^2(\hat{\varepsilon})}{n-k} = 55 \cdot 57 \sum_{k=1}^8 \frac{r_k^2(\hat{\varepsilon})}{55-k} = 3,4764. \quad (155)$$

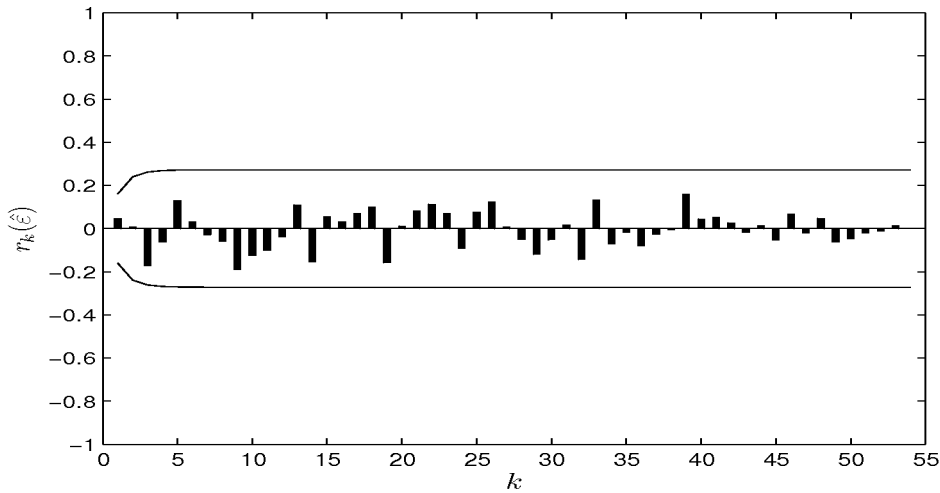
Kritická hodnota  $\chi^2$ -rozdělení s  $K - p - q = 8 - 1 - 0 = 7$  stupni volnosti na hladině významnosti  $\alpha = 0,05$  je

$$\chi_7^2(0,05) = 14,07,$$

tedy  $Q^* < \chi_7^2(0,05)$  a hypotézu o neautokorelovanosti reziduí nelze zamítnout. Tuto hypotézu potvrzuje i metoda odhadnutých reziduí. Na obr. 37 jsou zachyceny hodnoty autokorelační funkce  $r_k(\hat{\varepsilon})$  a také hodnoty dvojnásobků směrodatných odchylek jednotlivých hodnot této funkce spočítaných dle (142). Žádná z hodnot  $r_k(\hat{\varepsilon})$  nepřekročila v absolutní hodnotě dvojnásobek své směrodatné odchylky, z toho plyne, že rezidua  $\hat{\varepsilon}_t$  nejsou autokorelovaná. Stejný výsledek dostaneme i Durbinovým-Watsonovým testem, jelikož

$$DW = \frac{\sum_{t=2}^{55} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{55} \hat{\varepsilon}_t^2} = 1,8732 \quad (156)$$

je hodnota blízko čísla 2, a tedy nelze zamítnout neautokorelovanost reziduí.



Obrázek 37: Graf hodnot funkce  $r_k(\hat{\varepsilon})$  a dvojnásobku jejich směrodatných odchylek

#### 4.3.6 Oprava modelu při zamítnutí ARMA( $p, q$ )

Pokud při analýze reziduí zjistíme, že jsou korelovaná, lze model opravit následovně. Pro odhadnutá rezidua  $\text{varepsilon}_t$  sestavíme model ARMA( $p^*, q^*$ ) ve tvaru

$$\varphi(B)^* \hat{\varepsilon}_t = \theta(B)^* \eta_t, \quad \eta_t \sim \mathcal{WN}(0, \sigma_\eta^2) \quad (157)$$

Tento model sloučíme s původním modelem  $\varphi(B)X_t = \theta(B)\varepsilon_t$  a získáme nový ve tvaru

$$\varphi(B)^* \varphi(B)X_t = \theta(B)^* \theta(B)\eta_t, \quad (158)$$

kde  $\varphi(B)^* \varphi(B)$  je operátor řádu  $(p + p^*)$  a  $\theta(B)^* \theta(B)$  je operátor řádu  $(q + q^*)$ . Vytvořili jsme model ARMA( $p + p^*, q + q^*$ ).

#### 4.3.7 Výběr modelu

Problém výběru modelu nastává v případě, že jsme identifikovali dva nebo více vhodných modelů. Můžeme se rozhodovat na principu porovnání odhadnutých reziduí. Vybíráme model s menším reziduálním rozptylem a s menším počtem parametrů. Z tohoto principu bylo odvozeno několik kritérií.



- Kritérium AIC (Akaike Information Criterion)

$$\text{AIC}(m) = n \ln \hat{\sigma}_\varepsilon^2 + 2m, \quad (159)$$

- Kritérium BIC (Bayesian Extension of Information Criterion)

$$\text{BIC}(m) = n \ln \hat{\sigma}_\varepsilon^2 - (n - m) \ln \left( 1 - \frac{m}{n} + m \ln n + m \ln \left[ \frac{\hat{\sigma}_X^2 - \hat{\sigma}_\varepsilon^2}{\hat{\sigma}_\varepsilon^2 m} \right] \right), \quad (160)$$

- Kritérium SBC (Schwartzovo bayesovské kritérium)

$$\text{SBC}(m) = n\hat{\sigma}_\varepsilon^2 + m \ln n, \quad (161)$$

- FPE (Final Prediction Error)

$$\text{FPE}(m) = \frac{1 + \frac{m}{n}}{1 - \frac{m}{n}} \hat{\sigma}_\varepsilon^2, \quad (162)$$

kde  $n$  značí počet pozorování,  $m$  počet parametrů v modelu,  $\hat{\sigma}_\varepsilon^2$  odhad rozptylu reziduí a  $\hat{\sigma}_X^2$  odhad rozptylu časové řady  $X_t$ . Vybereme ten model, který minimalizuje dané kritérium.

**Poznámka 4.6.** *Kritérium BIC bylo navrženo, protože kritérium AIC má tendence nadhodnocovat řád autoregrese.*

**Poznámka 4.7.** *Kritérium FPE používá systém MATLAB.*

Druhý přístup spočívá v porovnání přesnosti předpovědí. Časovou řadu rozdělíme na dvě části. Z první je model identifikován, pomocí druhé určíme přesnost předpovědí.

**Příklad 4.7.**

V tabulce 9 jsou uvedeny základní charakteristiky modelu AR(1) pro data z příkladu 2.1. Jsou zde pro úplnost uvedena i porovnávací kritéria, ale protože jsme identifikovali pouze jeden model, jejich hodnoty znát nepotřebujeme.

	$\hat{\varphi}_1 = 0,587$
Reziduální součet	0,7128
Směrodatná odchylka reziduí	0,1146
Směrodatná odchylka odhadu $\hat{\varphi}_1$	0,1092
Durbinův-Watsonův test	1,8732
Portmanteau test $Q^*$	3,4754
Minimální hladina významnosti $Q^*$ pro $K = 20$	0,835
Jarqueův-Berův test	0,4931
Minimální hladina významnosti J.-B. testu	0,78
AIC	-236,3150
BIC	-233,7954
SBC	4,9294
FPE	0,0136

Tabulka 9: Základní charakteristiky modelu AR(1) z příkladu 2.1

## 5 Rozšířené ARMA modely

### 5.1 Integrované modely (ARIMA)

Pomocí integrovaných modelů můžeme modelovat i časové řady, které vykazují nestacionaritu ve velikosti úrovně nebo jejím sklonu, přičemž tyto změny mohou mít i nesystematický náhodný charakter. Modely ARIMA modelují stochasticky kromě náhodné složky i složku trendovou.

V předchozím textu bylo uvedeno, že procesy  $AR(p)$  a  $ARMA(p,q)$  jsou stacionární, pokud leží všechny kořeny polynomu  $\varphi(B)$  vně jednotkového kruhu. Není-li tato podmínka splněna, proces není stacionární, ale obsahuje tzv. *stochastický trend*.

Pro konstrukci integrovaných modelů není požadována stacionarita tak jako u předchozích modelů, ale předpokládáme tzv. *homogenní nestacionaritu*.

**Definice 5.1.** *Časová řada, která obsahuje stochastický trend, je homogenně nestacionární, je-li převeditelná na řadu stacionární přechodem k prvním nebo vyšším diferencím.*

**Věta 5.1.** *Leží-li jeden kořen polynomu  $\varphi(B)$  na jednotkové kružnici a ostatní jsou vně, je proces stacionární po první diferenci a obsahuje tzv. lineární stochastický trend.*

**Věta 5.2.** *Leží-li na jednotkové kružnici dva kořeny polynomu  $\varphi(B)$  a ostatní vně této kružnice, je proces stacionární po druhé diferenci a obsahuje tzv. parabolický stochastický trend.*

**Poznámka 5.1.** *Obecně: počet kořenů polynomu  $\varphi(B)$ , které leží na jednotkové kružnici, určuje řád procesu.*

**Poznámka 5.2.** *Jestliže leží alespoň jeden z kořenů  $\varphi(B)$  uvnitř jednotkové kružnice, jedná se o nestacionaritu explozivního typu.*

**Definice 5.2.** *Integrovaný smíšený model ARIMA( $p, d, q$ ) je dán vztahem*

$$\varphi(B)w_t = \theta(B)\varepsilon_t, \tag{163}$$

kde

$$w_t = \Delta^d X_t \quad (164)$$

je  $d$ -tá diference procesu  $X_t$ ,  $\varepsilon_t$  je proces bílého šumu a model (163) je stacionární model typu (82) pro proces  $w_t$ .

Platí  $\Delta X_t = X_t - X_{t-1}$ . Diferenční operátor  $\Delta$  lze tedy vyjádřit pomocí operátoru zpětného posunutí takto

$$\Delta = 1 - B. \quad (165)$$

Model (163) můžeme zapisovat také ve tvaru

$$\nu(B)X_t = \theta(B)\varepsilon_t, \quad (166)$$

kde

$$\nu(B) = \varphi(B)(1 - B)^d \quad (167)$$

je *zobecněný autoregresní operátor*, což je polynom řádu  $p + d$ .

### Určení řádu diferencování

- Opticky posoudíme stacionaritu časové řady pomocí grafického záznamu. Máme-li pochybnosti o stacionaritě, pořídíme ještě graf první nebo i druhé difference (v praxi se málokdy použije větší řád diferencování než 2).

Jsme-li na pochybách, lze použít jednu z následujících metod:

- Určíme odhad rozptylu dané řady  $\sigma_X^2$  a odhady rozptylů jejích diferencí  $\sigma_{\Delta X}^2, \sigma_{\Delta^2 X}^2, \dots$ . Za  $d$  zvolíme hodnotu, která dává nejmenší rozptyl. Při postupném diferencování se totiž odhadnuté hodnoty rozptylu zmenšují, dokud není dosaženo stacionarity, potom se začínají opět zvyšovat. Použijeme-li vyšší řád diferencování, nastává tzv. *přediferencování*, což je nežádoucí.
- Nebo můžeme zkoumat odhadnutou autokorelační funkci. Pokud klesají hodnoty  $r_k$  pomalu (přibližně lineárně a ne geometricky), je nejspíše aspoň jeden kořen autoregresního operátoru velmi blízký jedné a je nutné dále diferencovat.

### 5.1.1 Proces náhodné procházky ("random walk")

Tento proces je zvláštním případem procesu (163), kde  $p = 0$ ,  $q = 0$ . Jeho charakteristickou vlastností je, že po  $d$ -té diferenci z něho vznikne proces bílého šumu. Můžeme jej zapsat jako

$$(1 - B)^d X_t = \varepsilon_t. \quad (168)$$

### 5.1.2 Procesy nestacionární v rozptylu

Mění-li se rozptyl v čase, již se nejedná o homogenní nestacionaritu. Někdy na čase může záviset i kovariance.

Stacionarity nedosáhneme pouhým diferencováním, ale musíme proces vhodně transformovat. V mnohých případech s rozptylem stabilizujeme i kovarianci.

Obecná transformace, kterou navrhli Box a Cox, má následující tvar

$$X_t^{(\lambda)} = \begin{cases} \frac{X_t^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \ln X_t, & \lambda = 0, \end{cases} \quad (169)$$

kde  $\lambda$  je tzv. *transformační parametr*. Ten je odhadován pomocí procedury založené na metodě maximální věrohodnosti (viz. Anděl). Hledáme takové  $\lambda$ , které vede k transformaci, která minimalizuje reziduální součet čtverců časové řady. Tento postup je dosti komplikovaný, častěji se používá transformace

$$X_t^{(\lambda)} = \begin{cases} X_t^\lambda, & \lambda \neq 0, \\ \ln X_t, & \lambda = 0. \end{cases} \quad (170)$$

Hodnotu  $\lambda$  volíme tak, že danou řadu rozdělíme na úseky o délce 4 – 12 pozorování, v každém úseku vypočteme aritmetický průměr a rozdíl mezi největší a nejmenší hodnotou, příslušné dvojice hodnot vyneseme do grafu. Pokud zakreslené body jsou konstantní, rozptyl se s časem nemění a řadu není třeba transformovat. Jestliže body leží přibližně na rostoucí přímce, použijeme logaritmickou transformaci. Roste-li rozptyl exponenciálně, je možné řadu transformovat funkcí  $\frac{1}{\sqrt{X_t}}$ . Tato metoda je sice méně přesná, ale jednodušší a pro praktické účely obvykle vyhovuje.

### 5.1.3 Výstavba modelu ARIMA – Shrnutí

- Pokud je řada nestacionární v rozptylu, vhodně ji transformujeme.
- Diferencujeme na stacionární řadu ve tvaru(164).
- Pro tuto řadu sestrojíme model ARMA( $p,q$ ) tvaru (163).

## 5.2 Sezónní modely

V této části se budeme zabývat stochastickým modelováním sezónní složky. Předpokládáme vzájemnou závislost mezi veličinami  $\dots, X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}, \dots$ , ale vzhledem k sezónnímu kolísání očekáváme i závislost mezi veličinami  $\dots, X_{t-2s}, X_{t-s}, X_t, X_{t+s}, X_{t+2s}, \dots$ , kde  $s$  je délka sezónní periody. Nejprve zkonstruujeme model pro časovou řadu odpovídající jedné sezóně (například pro měsíční pozorování zkonstruujeme model pro lednová data) v tomto tvaru

$$\phi(B^s)\Delta_s^D X_t = \Theta(B^s)\eta_t, \quad (171)$$

kde

$$\phi(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{Ps} \quad (172)$$

je tzv. *sezónní autoregresní operátor řádu  $P$* ,

$$\Theta(B^s) = 1 + \phi_1 B^s + \phi_2 B^{2s} + \dots + \phi_p B^{Qs} \quad (173)$$

je tzv. *sezónní operátor klouzavých součtů řádu  $Q$*  a

$$\Delta_s = (1 - B^s) \quad (174)$$

je tzv. *sezónní diferenční operátor*,  $D$  je řád sezónního diferencování. Model (171) je v podstatě ARIMA model popisující průběh pozorování pro danou sezónu. Předpokládáme, že tento model lze použít i pro ostatní sezóny. Je ovšem třeba nalézt model i pro řadu tvořenou náhodnými složkami  $\eta_t$ , poněvadž ty by měly být mezi jednotlivými sezónami korelované (například hodnoty pro únor jsou ovlivněny hodnotami v lednu), a tedy netvoří bílý šum. ARIMA ( $p, d, q$ ) tvaru

$$\varphi(B)\Delta^d \eta_t = \theta(B)\varepsilon_t, \quad (175)$$

kde  $\varepsilon_t$  je proces bílého šumu. Oba modely (171) a (175) můžeme sloučit do jediného.

**Definice 5.3.** *Sezónní smíšený model SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_s$  je dán vztahem*

$$\varphi(B)\phi(B^s)\Delta^d\Delta_s^D X_t = \theta(B)\Theta(B^s)\varepsilon_t. \quad (176)$$

### 5.2.1 Výstavba modelu SARIMA( $p, d, q$ ) $\times$ ( $P, D, Q$ ) $_s$ – Shrnutí

Výstavba modelu spočívá opět ve třech fázích. Odhad parametrů i verifikace jsou stejné jako u modelů ARIMA, rozdíl je pouze v první fázi identifikaci.

- Určíme řád diferencování

$$\left\langle \begin{array}{l} \text{běžného} \dots d \\ \text{sezónního} \dots D \end{array} \right.$$

Obvykle je  $d \leq 1, D \leq 1$ . O řádu diferencování většinou rozhodneme na základě odhadnuté autokorelační funkce  $r_k$ . Je vhodné prozkoumat aspoň  $4s$  hodnot této funkce.

- Má-li tato funkce lokální maxima v bodech  $s, 2s, 3s, \dots$ , pak je nutné sezónně diferencovat bez ohledu na její průběh mezi těmito body.
- Pokud mezi body  $s, 2s, 3s, \dots$ , klesá autokorelační funkce pomalu (lineárně), je třeba provést běžné diferencování.
- Nebo můžeme hledat řád diferencování  $d, D$  tak, že najdeme minimální hodnotu z odhadů rozptylů  $\hat{\sigma}_X^2, \hat{\sigma}_{\Delta X}^2, \hat{\sigma}_{\Delta_s X}^2, \hat{\sigma}_{\Delta_s \Delta_s X}^2$ .

Když známe řád diferencování, zkonstruujeme řadu

$$w_t = \Delta^d \Delta_s^D X_t \quad (177)$$

a pro tuto identifikujeme model (176). K tomu použijeme odhadnutou autokorelační funkci  $r_k$  a odhadnutou parciální autokorelační funkci  $r_{kk}$  řady (177). Pokud

- jsou hodnoty funkce  $r_k$  významně nenulové v bodech  $1, \dots, q, s - q, \dots, s + q, 2s - q, \dots, 2s + q, \dots, Qs - q, \dots, Qs + q$ ,
- jsou hodnoty funkce  $r_k$  mezi těmito body přibližně nulové,
- funkce  $r_{kk}$  v úsecích mezi body  $s, 2s, 3s, \dots$  v absolutní hodnotě geometricky klesá,
- funkce  $r_{kk}$  klesá v bodech  $s, 2s, 3s, \dots$ ,

potom řada (177) vyhovuje modelu  $\text{SARIMA}(0, 0, q) \times (0, 0, Q)_s$  a  $X_t$  vyhovuje modelu  $\text{SARIMA}(0, d, q) \times (0, D, Q)_s$ .

Pokud

- funkce  $r_k$  v úsecích mezi body  $s, 2s, 3s, \dots$  v absolutní hodnotě geometricky klesá,
- funkce  $r_k$  klesá v bodech  $s, 2s, 3s, \dots$ ,
- jsou hodnoty funkce  $r_{kk}$  významně nenulové v bodech  $1, \dots, p, s, \dots, s + p, 2s, \dots, 2s + p, \dots, Ps, \dots, Ps + p$ ,
- jsou hodnoty funkce  $r_{kk}$  mezi těmito body přibližně nulové,

potom (177) vyhovuje modelu  $\text{SARIMA}(p, 0, 0) \times (P, 0, 0)_s$  a  $X_t$  vyhovuje modelu  $\text{SARIMA}(p, d, 0) \times (P, D, 0)_s$ .

**Poznámka 5.3.** *Modely  $\text{SARIMA}(p, d, q) \times (P, D, Q)_s$  s vesměs nenulovými parametry jsou obtížně identifikovatelné.*



## 6 Konstrukce předpovědi

Označme  $\hat{X}_{t+h}(t)$ , předpověď budoucí veličiny  $X_{t+h}$  v čase  $t$  o  $h$  období do-  
předu. Číslo  $h$  se nazývá *horizont* předpovědi a  $t$  *práh* předpovědi.

Předpokládejme proces ARMA( $p, q$ ) tvaru

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (178)$$

Konstruujeme předpověď veličiny  $X_{t+h}$  pro  $h \geq 1$  v čase  $t$ , známe tedy hodnoty  
veličin  $X_t, X_{t-1}, X_{t-2}, \dots$ , resp. veličin  $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ . Předpověď veličiny  $X_{t+h}$   
bude lineární kombinací těchto veličin. Snažíme se získat předpověď s minimální  
čtvercovou chybou danou vztahem

$$E[X_{t+h} - \hat{X}_{t+h}(t)]^2 = E[\varepsilon_{t+h}(t)]^2. \quad (179)$$

Tuto předpověď lze zapsat ve tvaru

$$\hat{X}_{t+h}(t) = \psi_h^* \varepsilon_t + \psi_{h+1}^* \varepsilon_{t-1} + \psi_{h+2}^* \varepsilon_{t-2} + \dots \quad (180)$$

Použijeme-li pro budoucí veličinu  $X_{t+h}$  tvar lineárního procesu, který jsme pou-  
žívali v předchozích kapitolách, budeme psát

$$X_{t+h} = \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \psi_2 \varepsilon_{t+h-2} + \dots + \psi_{h-1} \varepsilon_{t+1} + \psi_h \varepsilon_t + \dots \quad (181)$$

a budeme hledat koeficienty  $\psi_h^*, \psi_{h+1}^*, \dots$ , které minimalizují riziko předpovědi

$$(1 + \psi_1^2 + \dots + \psi_{h-1}^2 + \sum_{i=h}^{\infty} (\psi_i - \psi_i^*)^2) \sigma_\varepsilon^2, \quad (182)$$

Toto riziko je minimální, pokud jsou

$$\psi_i^* = \psi_i, \quad i = h, h+1, \dots \quad (183)$$

Předpověď má potom tvar

$$\hat{X}_{t+h}(t) = \sum_{i=0}^{\infty} \psi_{h+i} \varepsilon_{t-i}. \quad (184)$$

Jiný způsob zápisu předpovědi je pomocí podmíněných středních hodnot při pevných hodnotách  $X_t, X_{t-1}, \dots$ , tedy jako

$$\hat{X}_{t+h}(t) = E(X_{t+h}|X_t, X_{t-1}, \dots). \quad (185)$$

Chybu předpovědi  $\hat{X}_{t+h}(t)$  definovanou jako

$$e_{t+h}(t) = X_{t+h} - \hat{X}_{t+h}(t) \quad (186)$$

lze zapsat jako konečný lineární proces ve tvaru

$$e_{t+h}(t) = \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \dots + \psi_{h-1} \varepsilon_{t+1}. \quad (187)$$

Tato veličina má nulovou střední hodnotu a rozptyl

$$\text{var}[e_{t+h}(t)] = (1 + \psi_1^2 + \dots + \psi_{h-1}^2) \sigma_\varepsilon^2. \quad (188)$$

### Výpočet předpovědi

Praktický výpočet předpovědi  $\hat{X}_{t+h}$  se provádí rekurentně na základě odhadů parametrů daného modelu. Uvažujme budoucí veličinu  $X_{t+h}$  ve tvaru

$$X_{t+h} = \varphi_1 X_{t+h-1} + \dots + \varphi_p X_{t+h-p} + \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \dots + \theta_q \varepsilon_{t+h-q}. \quad (189)$$

Předpověď této veličiny lze zapsat jako

$$\begin{aligned} \hat{X}_{t+h}(t) &= \hat{\varphi}_1 E(X_{t+h-1}|X_t, X_{t-1}, \dots) + \dots + \hat{\varphi}_p E(X_{t+h-p}|X_t, X_{t-1}, \dots) + \\ &+ E(\varepsilon_{t+h}|X_t, X_{t-1}, \dots) + \hat{\theta}_1 E(\varepsilon_{t+h-1}|X_t, X_{t-1}, \dots) + \dots + \\ &+ \hat{\theta}_q E(\varepsilon_{t+h-q}|X_t, X_{t-1}, \dots) \end{aligned} \quad (190)$$

Dále ve výpočtu využijeme skutečnost, že

$$E(X_{t+j}|X_t, X_{t-1}, \dots) = \begin{cases} \hat{X}_{t+j} & \text{pro } j > 0, \\ X_{t+j} & \text{pro } j \leq 0 \end{cases} \quad (191)$$

a

$$E(\varepsilon_{t+j}|X_t, X_{t-1}, \dots) = \begin{cases} 0 & \text{pro } j > 0, \\ \varepsilon_{t+j} = X_{t+j} - \hat{X}_{t+j}(t+j-1) & \text{pro } j \leq 0. \end{cases} \quad (192)$$

Nejprve vypočítáme předpovědi o jeden krok dopředu  $\hat{X}_{q+1}(q), \hat{X}_{q+2}(q+1), \dots$ , ty využijeme pro výpočet předpovědí o dva kroky dopředu atd.

Do vztahu(190) dosadíme za podmíněné střední hodnoty  $E(\varepsilon_s|X_t, X_{t-1}, \dots)$  pro  $s > t$ , nuly, pro  $s \leq t$  je nahradíme výrazy  $X_s - \hat{X}_s(s-1)$ , podmíněné střední hodnoty  $E(X_s|X_t, X_{t-1}, \dots)$  nahradíme pro  $s > t$  předpověďmi  $\hat{X}_s(t)$ , které jsou již známé a pro  $s \leq t$  potom použijeme hodnoty  $X_s$ .

Počáteční podmínka pro zahájení rekurentního výpočtu je

$$\varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_q = 0. \quad (193)$$

Pro model ARMA ve tvaru (178) je 95% interval spolehlivosti dán jako

$$\langle \hat{X}_{t+h}(t) - 2\sigma[e_{t+h}(t)], \hat{X}_{t+h}(t) + 2\sigma[e_{t+h}(t)] \rangle, \quad (194)$$

kde směrodatná odchylka  $\sigma[e_{t+h}(t)]$  je po dosazení do vzorce (188) ve tvaru

$$\sigma_\varepsilon \sqrt{1 + \sum_{i=1}^{h-1} \psi_i^2} \quad (195)$$

**Poznámka 6.1.** *Nepracujeme se skutečnými parametry, ale pouze s jejich odhady. Vzorce použité v této kapitole ale naštěstí nejsou příliš citlivé na chyby vzniklé při výpočtu těchto odhadů.*

**Poznámka 6.2.** *Předpovědi a intervaly spolehlivosti pro modely ARIMA a SARIMA se konstruují analogicky jako pro proces ARMA.*

### Příklad 6.1.

V závěru kapitoly zkonstruujeme předpovědi pro řadu  $X_t$ , pro níž jsme v předchozím textu identifikovali model (154), tj.

$$X_t = 5,916 + 0,587X_{t-1} + \varepsilon_t.$$

V případě procesu AR(1) je to velice snadné. Dosadíme do (190), potom bude předpověď o jeden krok dopředu

$$\hat{X}_{56}(55) = 5,9154 + \hat{\varphi}_1 X_{55} = 5,9154 + 0,587 \cdot 14,003 = 14,1352, \quad (196)$$

předpověď o dva kroky dopředu

$$\hat{X}_{57}(55) = 5,9154 + \hat{\varphi}_1 \hat{X}_{56}(55) = 5,9154 + 0,587 \cdot 14,1352 = 14,2127 \quad (197)$$

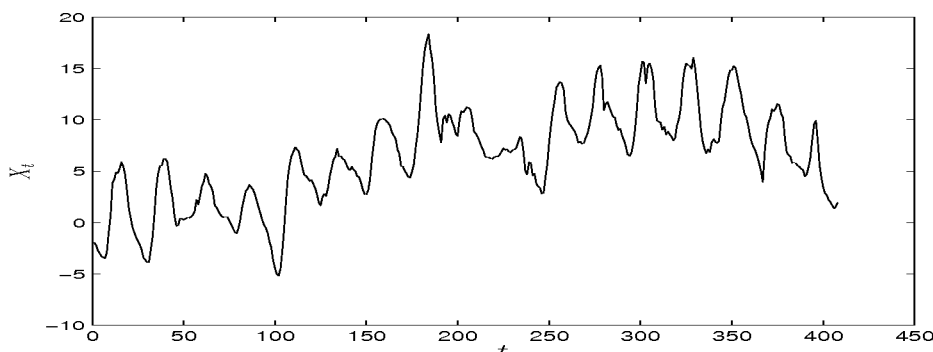
atd. až předpověď o osmnáct kroků dopředu se bude rovnat střední hodnotě procesu, tj.

$$\hat{X}_{73}(55) = 5,9154 + \hat{\varphi}_1 \hat{X}_{72}(55) = 5,9154 + 0,587 \cdot 14,3229 = 14,323. \quad (198)$$

## 7 Boxova-Jenkinsova metodologie v praxi

Na závěr uvedenou teorii ukážeme na reálných datech. Zkoumanou časovou řadu  $X_t$  tvoří teploty vzduchu ze stanice Staré Město u Uherského Hradiště měřené v hodinových intervalech od 20. ledna do 5. února roku 2009 (data pochází z měření Českého hydrometeorologického ústavu). Celkem se jedná o 408 měření během 17 dnů. V následujících podkapitolách jsou ukázány tři způsoby, jak je možné hledat model pro tuto řadu.

### 7.1 Modelování původní řady pouze pomocí Boxova-Jenkinsova přístupu

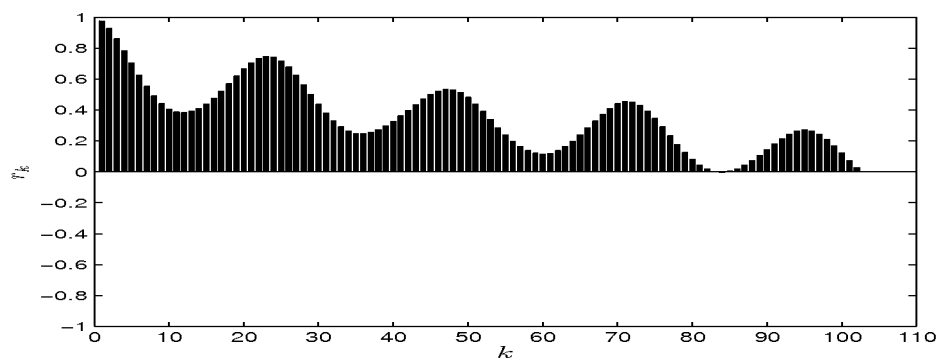


Obrázek 38: Naměřené teploty

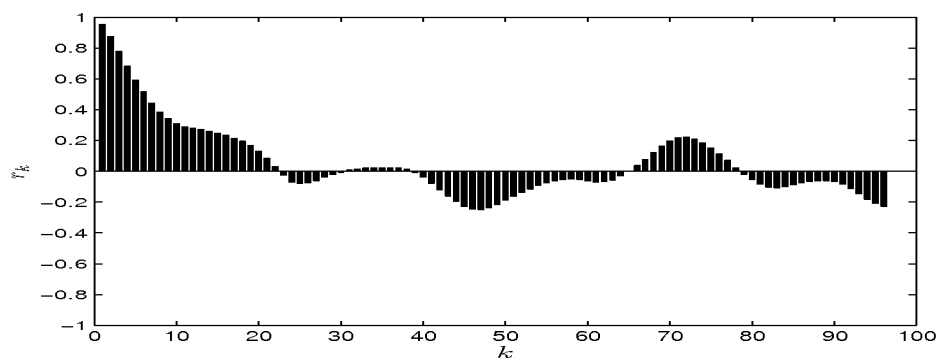
Z obrázku 38 je patrné, že řada  $X_t$  je nestacionární. Z následujících grafů je zřejmé, že bude třeba provést kombinované diferencování. Na obrázku 39 je graf odhadnuté autokorelační funkce  $r_k$  řady  $X_t$ . Tato funkce má v bodech, které jsou násobky 24, lokální maxima, resp. minima. To ukazuje na nutnost diferencovat sezónně. Graf odhadnuté autokorelační funkce  $r_k$  řady diferencí

$$X_t^* = X_t - X_{t-24} \quad (199)$$

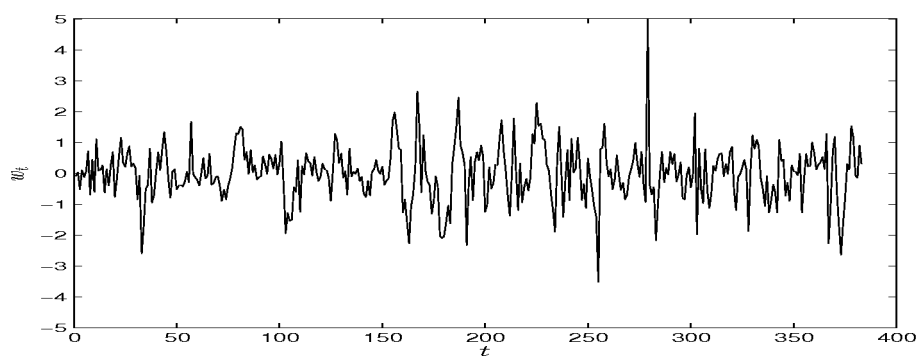
si můžeme prohlédnout na obrázku 40. První hodnota této funkce je velmi blízká jedné a lze říci, že pomalu klesá mezi násobky čísla 24, a tedy je třeba provést i běžné diferencování.



Obrázek 39:  $r_k$  pro řadu naměřených teplot

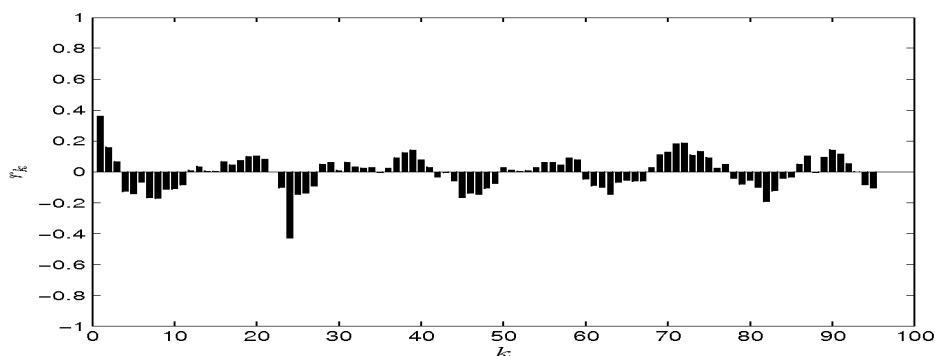


Obrázek 40:  $r_k$  řady po sezónním diferencování

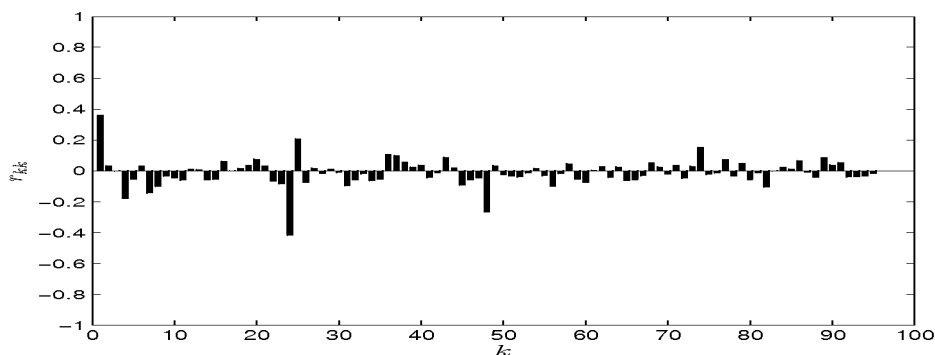


Obrázek 41: Řada  $w_t$

Po kombinovaném diferencování dostáváme řadu  $w_t$ , kde  $t = 1, 2, \dots, 383$ . Z obrázku 41 lze usoudit, že střední hodnota řady  $w_t$  je konstatní a nulová (aritmetický průměr  $\bar{w} = -0,0106$ ) a rozptyl se s časem nemění, tj. proces je stacionární a centrováný.



Obrázek 42:  $r_k$  řady  $w_t$



Obrázek 43:  $r_{kk}$  řady  $w_t$

Graf na obr. 42 znázorňuje průběh odhadnuté autokorelační funkce  $r_k$ . Odtud můžeme usoudit, že funkce  $\varrho_k$  je mezi body  $k = 0, 24, 48, \dots$  omezena sinusoidou s klesající amplitudou. Hodnota  $r_{24} = -0,4289$  je výrazně větší než ostatní. Z grafu odhadnuté parciální autokorelační funkce  $r_{kk}$  na obr. 43 je dobře vidět, že výrazně nenulové jsou hodnoty  $r_{1,1} = 0,3601$ ,  $r_{24,24} = -0,4177$ ,  $r_{25,25} = 0,2077$  a  $r_{48,48} = -0,2664$ , za kterými následují nulové hodnoty. Relativně vysokých hodnot nabývají také hodnoty  $r_{4,4} = -0,1786$ ,  $r_{7,7} = -0,1444$  a  $r_{74,74} = 0,1520$ , to

ovšem může být způsobeno chybami při měření (zamyslíme-li se nad průběhem teplot během dne, nedává smysl, že by teplota v jistou hodinu závisela na teplotě o čtyři hodiny dříve, ale na teplotách o dvě a tři hodiny dříve ne) a také chybou vnesenou v průběhu odhadování  $r_k$  a  $r_{kk}$ . Zvolíme  $Q = 1$  a  $p = 1$ . Pro řadu  $w_t$  jsme identifikovali model SARIMA  $(1, 0, 0) \times (0, 0, 1)_{24}$  a pro původní řadu  $X_t$  model SARIMA  $(1, 1, 0) \times (0, 1, 1)_{24}$ .

Parametry budeme hledat pomocí podmíněné metody nejmenších nelineárních čtverců (viz. podkapitola 4.2.1). Vypočítáme odhady hodnot bílého šumu dle vztahu (107). Počáteční odhady  $\hat{\varepsilon}_t$  pro  $t \leq p$ , tj. pro  $t \leq 1$  položíme rovny nule. Například pro parametry  $\hat{\varphi}_1 = 0,4$  a  $\hat{\Theta}_1 = -0,8$  potom dostaneme

$$\begin{aligned}\hat{\varepsilon}_2(0, 4; -0, 8) &= w_2 - 0,4w_1 + 0,8\hat{\varepsilon}_{-22}(0, 4; 0, 8) = 0,02 - 0,4 \cdot (-0,10) + 0,8 \cdot 0 = \\ &= 0,0549\end{aligned}$$

$$\begin{aligned}\hat{\varepsilon}_3(0, 4; -0, 8) &= w_3 - 0,4w_2 + 0,8\hat{\varepsilon}_{-21}(0, 4; 0, 8) = -0,52 - 0,4 \cdot 0,02 + 0,8 \cdot 0 = \\ &= -0,5270,\end{aligned}$$

⋮

$$\begin{aligned}\hat{\varepsilon}_{26}(0, 4; -0, 8) &= w_{26} - 0,4w_{25} + 0,8\hat{\varepsilon}_2(0, 4; 0, 8) = -0,63 - 0,4 \cdot 0,2 + 0,8 \cdot 0,0549 = \\ &= 0,5940,\end{aligned}$$

⋮

$$\begin{aligned}\hat{\varepsilon}_{383}(0, 4; -0, 8) &= w_{383} - 0,4w_{382} + 0,8\hat{\varepsilon}_{359}(0, 4; 0, 8) = 0,30 - 0,4 \cdot 0,90 + 0,8 \cdot \\ &\cdot 0,4824 = 0,3259,\end{aligned}$$

Součet čtverců pro tuto dvojici parametrů je

$$S_P^w(0, 4; -0, 8) = 0,0549 + (-0,5270)^2 + \dots + 0,5949^2 + \dots + 0,3259^2 = 173,22. \quad (200)$$

V tabulce 10 jsou uvedeny hodnoty součtů čtverců  $S_P^w(\varphi_1, \Theta_1)$  na mříži hodnot parametrů

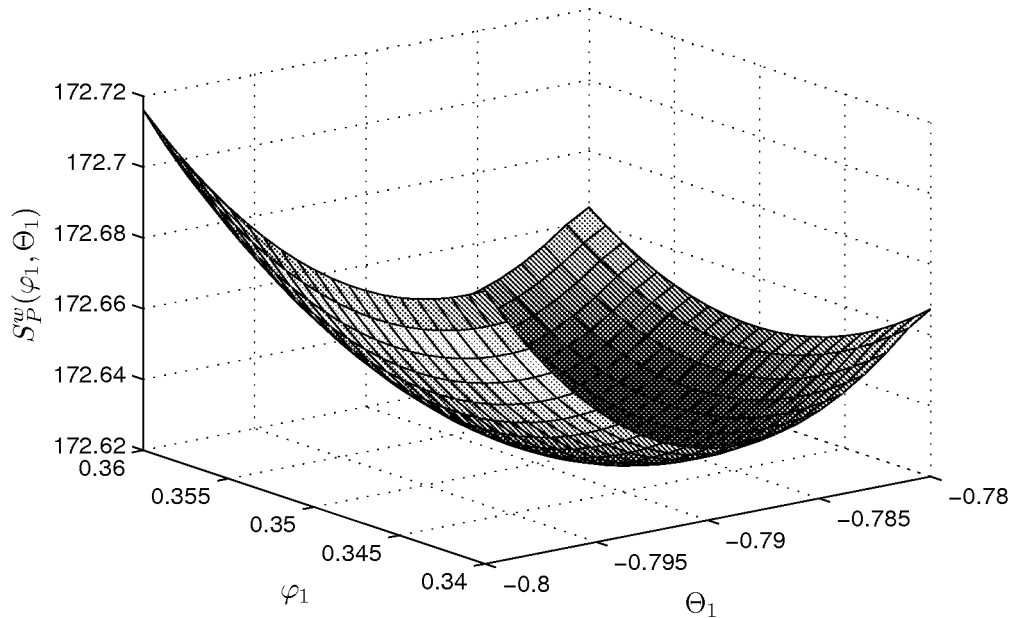
$$\varphi_1 = -0,75; -0,5; \dots; 0,75 \text{ a } \Theta_1 = -0,75; -0,5; \dots; 0,75.$$



$\Theta_1$	-0,75	-0,5	-0,25	0,00	0,25	0,50	0,75
$\varphi_1$							
-0,75	413,94	466,71	546,62	659,72	822,14	1 075,25	1 698,09
-0,50	317,25	357,18	418,15	505,19	630,87	828,44	1 312,62
-0,25	245,25	275,18	322,14	389,87	489,26	648,57	1 035,12
-0,00	197,93	220,71	258,27	313,75	397,32	535,63	865,57
0,25	175,30	193,76	226,65	276,85	355,04	489,64	803,99
0,50	177,35	194,35	227,28	279,16	362,44	510,58	850,37
0,75	204,08	222,47	260,17	320,67	419,50	598,45	1 004,71

Tabulka 10: Součty  $S_P^w(\varphi_1, \Theta_1)$

Minimální hodnotu funkce  $S_P^w(\varphi_1, \Theta_1)$  v tab. 10 dává dvojice parametrů  $\hat{\varphi}_1 = 0,25$ ,  $\hat{\Theta}_1 = -0,75$ . Ale součet (200) je ještě menší. Globální minimum funkce  $S_P^w(\varphi_1, \Theta_1)$  se nachází tedy na oblasti dané body  $[0; -1]$ ,  $[0; -0,75]$ ,  $[0,5; -1]$ , a  $[0,5; -0,75]$ . Budeme-li zde hustotu sítě postupně zjemňovat až na 0,001, získáme minimum



Obrázek 44: Součty  $S(\varphi_1, \Theta_1)$  na mříži parametrů  $\varphi_1 = 0,34; 0,341; \dots; 0,36$  a  $\Theta_1 = -0,8; -0,799; \dots; -0,78$

$$S_P^w(\varphi_1; \Theta_1) = 172,62$$

pro odhad parametrů  $\hat{\varphi}_1 = 0,3495$  a  $\hat{\theta}_1 = -0,7880$ . Graf funkce  $S_P^w(\varphi_1, \Theta_1)$

na mříži hodnot parametrů

$$\varphi_1 = 0,34; 0,341; \dots; 0,36 \text{ a } \Theta_1 = -0,8; -0,799; \dots; -0,78$$

zobrazuje obrázek 44.

Pro řadu  $w_t$  jsme identifikovali model SARIMA  $(1, 0, 0) \times (0, 0, 1)_{24}$  tvaru

$$w_t = 0,3495w_{t-1} - 0,788\varepsilon_{t-24} + \varepsilon_t. \quad (201)$$

Zda je správný, ověříme pomocí Portmanteau testu (viz. podkapitola 4.3.4). Odhady reziduí

$$\hat{\varepsilon}_t = w_t - 0,3495w_{t-1} + 0,788\hat{\varepsilon}_{t-24}, \quad t = 2, 3, \dots, \quad (202)$$

již známe. Nyní vypočítáme autokorelační funkci těchto reziduí

$$r_k(\hat{\varepsilon}) = \frac{\sum_{t=2}^{382} \hat{\varepsilon}_t \hat{\varepsilon}_{t+1}}{\sum_{t=2}^{383} \hat{\varepsilon}_t^2}. \quad (203)$$

Prvních dvacet hodnot této funkce je uvedeno v tabulce 11.

$k$	1	2	3	4	5	6	7	8	9	10
$r_k(\hat{\varepsilon})$	-0,01	0,02	0,05	-0,11	-0,05	0,04	-0,12	-0,12	-0,05	-0,05
$k$	11	12	13	14	15	16	17	18	19	20
$r_k(\hat{\varepsilon})$	-0,10	0,07	0,00	0,01	-0,01	0,01	-0,04	0,06	0,06	0,01

Tabulka 11: Prvních dvacet hodnot funkce  $r_k(\hat{\varepsilon})$

Zvolíme  $K = 20$ , potom je hodnota modifikované Portmanteau statistiky

$$Q^* = n(n+2) \sum_{k=1}^{20} \frac{r_k^2(\hat{\varepsilon})}{383-k} = 383 \cdot 385 \cdot 2,03 \cdot 10^{-4} = 29,90. \quad (204)$$

Stupně volnosti  $\chi^2$ -rozdělení jsou  $K - p - Q = 20 - 1 - 1 = 18$ .

V tabulce 12 jsou uvedeny základní charakteristiky řady  $w_t$  a výsledky testů diagnostické kontroly modelů.

Parametry modelu	$\hat{\varphi}_1 = 0,3495,$ $\hat{\Theta}_1 = -0,788$
Reziduální součet	172,62
Směrodatná odchylka reziduí	0,67
Směrodatná odchylka odhadu $\hat{\varphi}_1$	0,079
Směrodatná odchylka odhadu $\hat{\Theta}_1$	0,052
Durbinův-Watsonův test	2,015
Portmanteau test $Q^*$	29,90
Minimální hladina významnosti $Q^*$ pro $K = 20$	0,038
Jarqueův-Berův test	2,84
Minimální hladina významnosti J.-B. testu	0,24
AIC	-301,23
BIC	-293,12
SBC	184,51
FPE	0,46

Tabulka 12: Základní charakteristiky modelu SARIMA  $(1, 0, 0) \times (0, 0, 1)_{24}$  pro řadu  $w_t$

Nesystematickou složku  $\hat{\varepsilon}_t$  můžeme považovat za proces bílého šumu. Jarqueův-Berův test neprokázal, že by tato rezidua neměla normální rozdělení a ačkoli dle Portmanteau testu nelze hypotézu o neautokorelovanosti veličin  $\hat{\varepsilon}_t$  zamítnout až na hladině významnosti 3,8%, Durbinovým-Watsonovým testem nebyla autokorelovanost prokázána.

Pro původní řadu  $X_t$  získáváme model SARIMA  $(1, 1, 0) \times (0, 1, 1)_{24}$  ve formě

$$(1 - 0,3495B)(1 - B)(1 - B^{24})X_t = (1 - 0,788B^{24})\varepsilon_t, \quad (205)$$

resp.

$$X_t = 1,3495X_{t-1} - 0,3495X_{t-2} + X_{t-24} - 1,3495X_{t-25} + 0,3495X_{t-26} + \varepsilon_t - 0,788\varepsilon_{t-24}. \quad (206)$$

Přibližné 95%ní intervaly spolehlivosti parametrů jsou

$$\begin{aligned} \langle \hat{\varphi}_1 - 2\sigma(\hat{\varphi}_1), \hat{\varphi}_1 + 2\sigma(\hat{\varphi}_1) \rangle &= \langle 0,1913; 0,5077 \rangle, \\ \langle \hat{\Theta}_1 - 2\sigma(\hat{\Theta}_1), \hat{\Theta}_1 + 2\sigma(\hat{\Theta}_1) \rangle &= \langle -0,892; -0,684 \rangle. \end{aligned} \quad (207)$$

## 7.2 Modelování řady reziduí získané po odstranění trendu

Model (205), resp. (206), je již dosti složitý a jak bylo zmíněno v teoretické části, modely typu SARIMA  $(p, d, 0) \times (0, D, Q)_s$  se v praxi prakticky nepoužívají. Jinou možností, jak získat rezidua, která bychom mohli modelovat pomocí jednoduššího procesu, je proložit naměřená data trendovou křivkou. Pro charakter teplot znázorněných na obr. 38 lze taková rezidua získat vyrovnáním dat modelem konstantní sezónnosti se schodovitým trendem (viz např. [3]). Zde jsou jednotlivá měření dána vztahem

$$X_{ij} = T_i + S_j + \nu_{ij}, \quad (208)$$

kde  $T_i$ ,  $i = 1, \dots, 17$  jsou konstanty odhadnuté jako aritmetické průměry jednotlivých dnů ( $\hat{T}_i = \bar{X}_{i\bullet}$ ) a  $S_j$ ,  $j = 1, \dots, 24$ , jsou sezóny odhadnuté jako mezidenní průměr  $j$ -té hodiny minus celkový průměr ( $\hat{S}_j = \bar{X}_{\bullet j} - \bar{X}$ ). Předpokládáme, že

$$\sum_{j=1}^{24} S_j = 0. \quad (209)$$

Odhady  $T_i$  jsou uvedeny v tabulce 13, odhady  $S_j$  potom v tabulce 14.

$i$	1	2	3	4	5	6	7	8	9
$T_i$	0,73	0,97	1,83	1,01	2,05	4,63	7,07	10,07	9,34
$i$	10	11	12	13	14	15	16	17	
$T_i$	6,69	8,34	10,63	10,95	11,10	10,98	8,13	4,71	

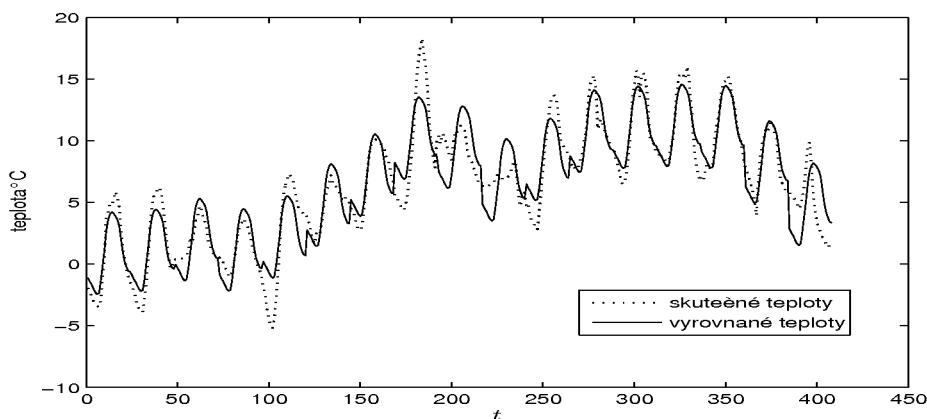
Tabulka 13: Odhady parametrů trendu  $T_i$

$j$	1	2	3	4	5	6	7	8	9	10	11	12
$S_j$	-1,85	-2,16	-2,39	-2,72	-3,02	-3,18	-3,07	-2,13	-0,86	0,40	1,90	2,83
$j$	13	14	15	16	17	18	19	20	21	22	23	24
$S_j$	3,31	3,50	3,38	3,18	2,90	2,40	1,38	0,29	-0,51	-0,93	-1,28	-1,35

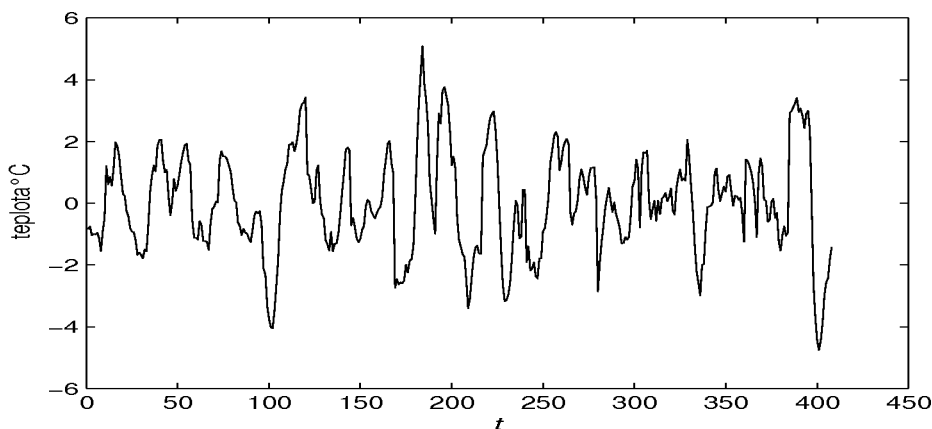
Tabulka 14: Odhady parametrů sezón  $S_j$

Porovnat skutečné a teoretické (vyrovnané) teploty můžeme na grafu na obr. 45. Rezidua  $\nu_t$ ,  $t = (i - 1)24 + j$ , kde  $i = 1, 2, \dots, 17$ ,  $j = 1, 2, \dots, 24$ , jsou náhodné

veličiny tvořící časovou řadu, pro kterou budeme hledat vhodný model.



Obrázek 45: Porovnání naměřených a vyrovnaných hodnot



Obrázek 46: Rezidua  $\nu_t$

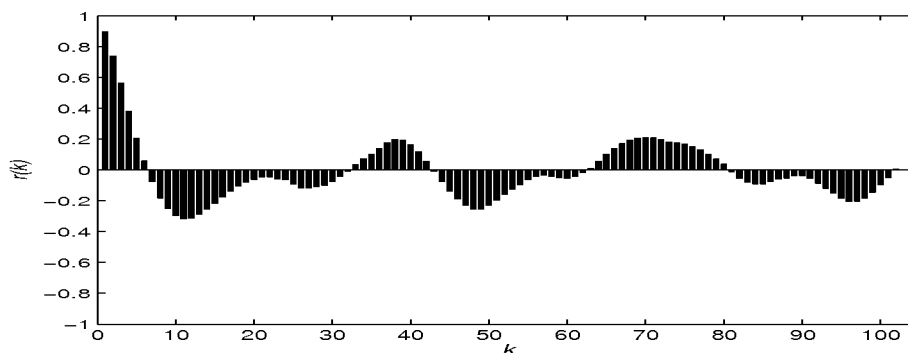
Z grafu na obrázku 46 lze předpokládat, že tento proces je stacionární a centrováný ( $\bar{\nu} = -2,05 \cdot 10^{-15}$ ). Pomocí vztahů (7) a (14) spočítáme odhady  $r_k$  a  $r_{kk}$ . Grafy těchto odhadů jsou znázorněny na obrázcích 47 a 48. Prvních deset hodnot funkcí  $r_k$  a  $r_{kk}$  funkcí je uvedeno v tabulce 15.

Pro autokorelační funkci  $\varrho_k$  neexistuje identifikační bod, protože  $2\sigma(r_k) = 2\sqrt{\frac{1}{n}} = 2\sqrt{\frac{1}{408}} = 0,1 < r_1 = 0,90$  a tedy  $k_0 \neq 0$ , stejně tak  $k_0 \neq 1$ , neboť

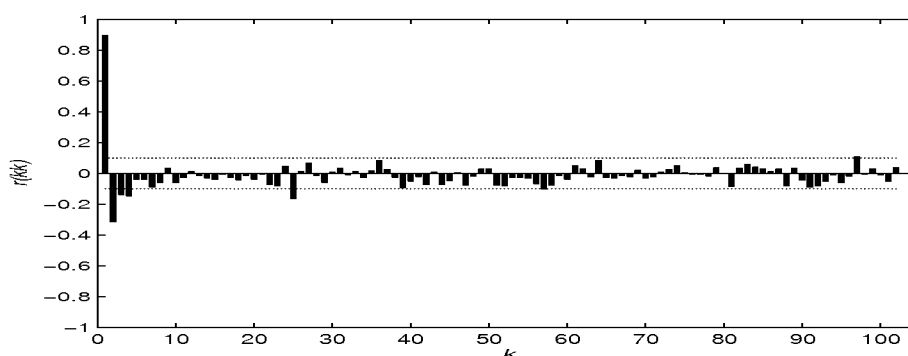
$k$	1	2	3	4	5	6	7	8	9	10
$r_k$	0,90	0,74	0,56	0,38	0,21	0,06	-0,07	-0,18	-0,25	-0,30
$r_{kk}$	0,90	-0,31	-0,14	-0,14	-0,04	-0,04	-0,09	-0,06	0,03	-0,06

Tabulka 15: Odhad prvních deseti hodnot  $\varrho_k$  a  $\varrho_{kk}$

$2\sigma(r_k) = 2\sqrt{\frac{1}{n}(1 + 2\sum_{j=1}^{k_0} r_j^2)} = 2\sqrt{\frac{1+2r_1^2}{408}} = 2\sqrt{\frac{1+2\cdot 0,9^2}{408}} = 0,16 < r_2 = 0,74$  a  
 $k_0 \neq 2$ , jelikož  $2\sigma(r_k) = 2\sqrt{\frac{1+2r_1^2+2r_2^2}{408}} = 2\sqrt{\frac{1+2(0,9^2+0,74^2)}{408}} = 0,19 < r_3 = 0,56$ .  
 Podle průběhu odhadnuté funkce  $r_k$ , lze soudit, že je autokorelační funkce  $\varrho_k$  omezena sinusoidou s klesající amplitudou.



Obrázek 47: Odhad autokorelační funkce  $\varrho_k$  řady  $\nu_t$



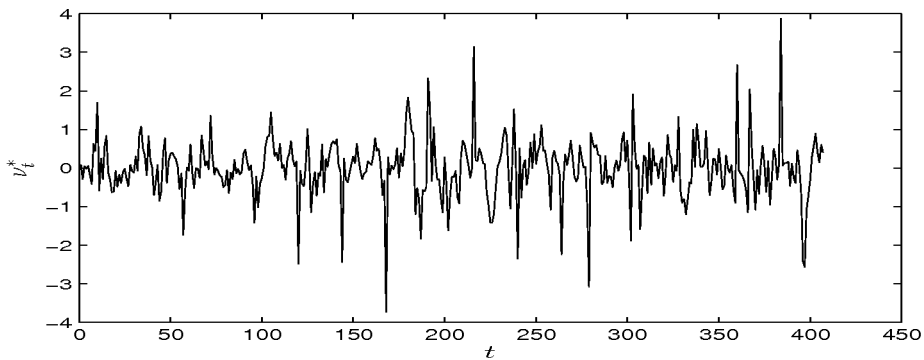
Obrázek 48: Odhad parciální autokorelační funkce  $\varrho_{kk}$  řady  $\nu_t$

Pro parciální autokorelační funkci  $\varrho_{kk}$  identifikační bod  $k_0$  nalezneme. Aproximace směrodatné odchylky je  $\sigma(r_{kk}) = \sqrt{\frac{1}{n}} = \sqrt{\frac{1}{408}} = 0,0495$ . Z obr. 48 je

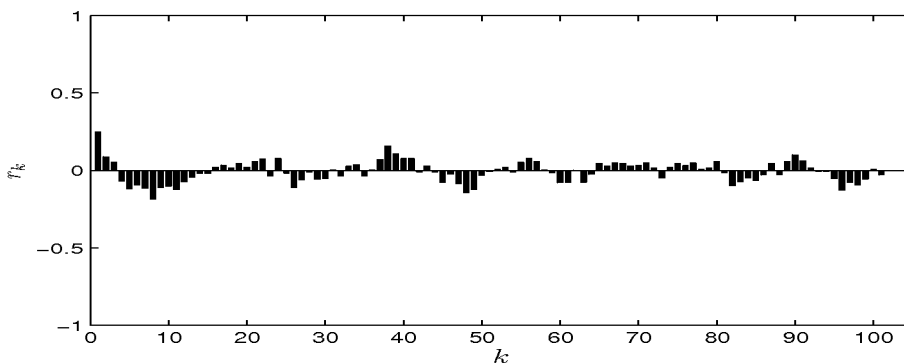
vidět, že absolutní hodnoty  $r_{11}, r_{22}, r_{33}, r_{44}$  jsou větší než  $2\sigma(r_{kk}) = 0,099$  a tedy  $k_0 = 4$ . Větší je také  $|r_{25,25}|$  a nepatrně i  $r_{56,56}$ . To může být způsobeno tím, že pracujeme s odhady a ty mohou být mezi sebou korelované. Identifikovali jsme proces AR(4) a pro něj by výpočty byly dosti složité. Zkusíme, jestli pro řadu diferencí řady  $\nu_t$ , tj. řadu

$$\nu_t^* = \nu_t - \nu_{t-1}, \quad (210)$$

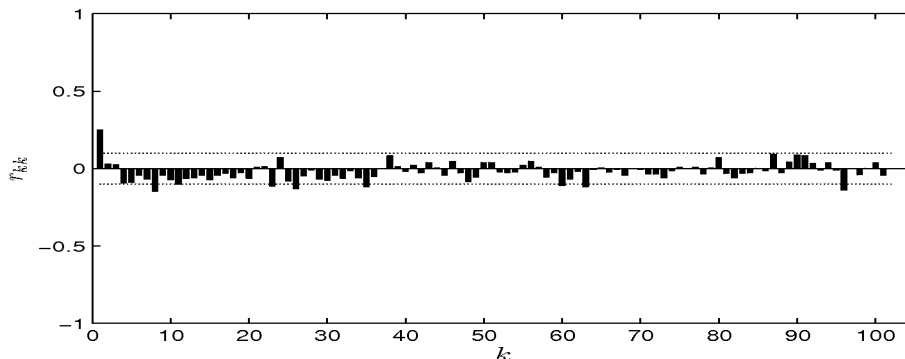
nenajdeme jednodušší model. Řada  $\nu_t^*$  má menší rozptyl  $\sigma_{\nu^*} = 0,5734$  než nediferencovaná řada  $\nu_t$ , jejíž rozptyl je  $\sigma_\nu = 2,7674$ . Řada druhých diferencí má rozptyl 0,8659 a dále diferencovat by tedy nemělo smysl. Z grafu řady  $\nu_t^*$  na obrázku 49 lze předpokládat, že řada je stacionární.



Obrázek 49: Řada prvních diferencí  $\nu_t^*$



Obrázek 50: Odhad parciální autokorelační funkce  $\varrho_k$  řady  $\nu_t^*$



Obrázek 51: Odhad parciální autokorelační funkce  $\varrho_{kk}$  řady  $\nu_t^*$

Podle grafů odhadnutých funkcí  $r_k$  a  $r_{kk}$  na obrázcích 50 a 51 lze říci, že autokorelační funkce  $\varrho_k$  řady  $\nu_t^*$  je omezena sinusoidou s klesající amplitudou a parciální autokorelační funkce  $\varrho_{kk}$  má identifikační bod  $k_0 = 1$  (hodnota  $r_{1,1}$  je nasledována šesti velmi malými hodnotami). Řadu  $\nu_t^*$  budeme tedy modelovat jako AR(1).

Metodou nejmenších čtverců dostaneme odhad

$$\hat{\varphi}_1 = \frac{\sum_{t=2}^{407} \nu_{t-1}^* \nu_t^*}{\sum_{t=2}^{407} (\nu_{t-1}^*)^2} = 0,2466. \quad (211)$$

Máme tedy model

$$\nu_t^* = 0,2466\nu_{t-1}^* + \varepsilon_t. \quad (212)$$

Pro původní řadu potom je

$$X_t = T_i + S_j + \nu_t, \quad t = (i-1)24 + j, \quad (213)$$

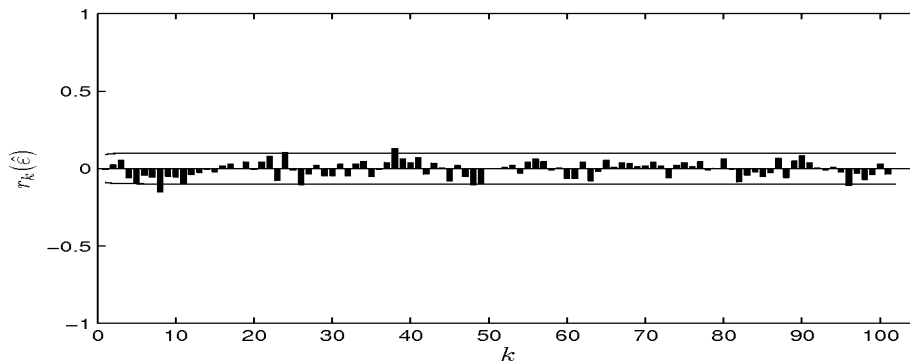
kde  $\nu_t$  je proces ARIMA(1,1,0) ve formě

$$\nu_t = 1,2466\nu_{t-1} - 0,2466\nu_{t-2} + \varepsilon_t. \quad (214)$$

Hodnoty  $T_i$ ,  $i = 1, 2, \dots, 17$  a  $S_j$ ,  $j = 1, 2, \dots, 24$ , jsou uvedeny v tabulkách 13 a 14.

Na obr. 52 je zobrazen graf odhadnuté autokorelační funkce  $r_k(\hat{\varepsilon})$  a dvojnásobky směrodatných odchylek hodnot této funkce. Na základě tohoto grafu





Obrázek 52: Hodnoty  $r_k(\hat{\varepsilon})$  pro řadu  $\nu_t^*$

bychom mohli zamítnout neautokorelovanost reziduí na hladině významnosti 95% (podle metody odhadnutých reziduí). Model ale nelze zcela zamítnout, protože prvních sedm hodnot této funkce testem projde, větší hodnoty pro vyšší  $k$  mohou být způsobeny tím, že pracujeme s odhady reziduí.

V tabulce 16 jsou uvedeny základní charakteristiky tohoto modelu, výsledky testů reziduí a hodnoty kritérií.

Parametry modelu	$\hat{\varphi}_1 = 0,2466$
Reziduální součet	218,63
Směrodatná odchylka reziduí	0,73
Směrodatná odchylka odhadu $\hat{\varphi}_1$	0,08
Durbinův-Watsonův test	2,014
Portmanteau test $Q^*$	25,72
Minimální hladina významnosti $Q^*$ pro $K = 20$	0,135
Jarqueův-Berův test	3,00
Minimální hladina významnosti J.-B. testu	0,22
AIC	-250,91
BIC	-248,60
SBC	224,64
FPE	0,54

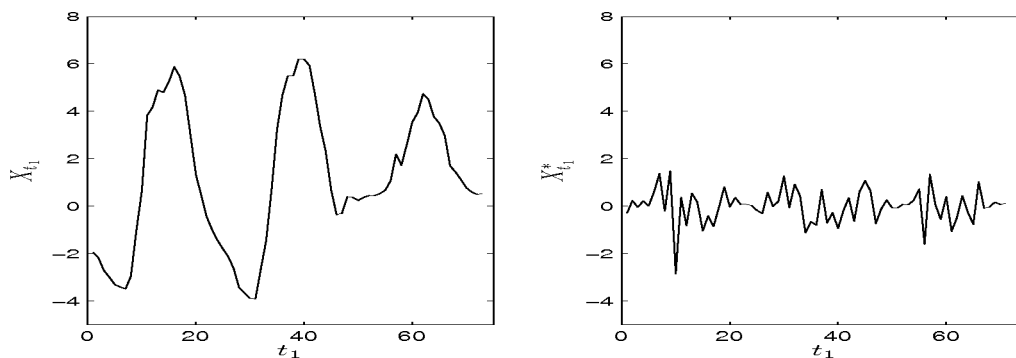
Tabulka 16: Základní charakteristiky modelu AR(1) pro řadu  $\nu_t^*$

Ani v tomto případě testy neprokázaly, že rezidua  $\hat{\varepsilon}_t$  netvoří proces bílého šumu.

### 7.3 Modelování po částech

Z obrázku 38 vidíme, že průměrná denní teplota byla v prvních třech dnech konstantní (označíme tento úsek  $X_{t_1}$ , jeho délka je  $n_1 = 72$ ), po dobu dalších pěti dní rostla (úsek  $X_{t_2}$ , o délce  $n_2 = 120$ ), v následujících dvou dnech (úsek  $X_{t_3}$  o délce  $n_3 = 48$ ) nastala výrazná změna, teploty během dne méně kolísaly a jejich průměrná hodnota klesala, potom se mírně oteplilo, ale ke konci začala teplota opět klesat (úsek  $X_{t_4}$  o délce  $n_4 = 168$ ).

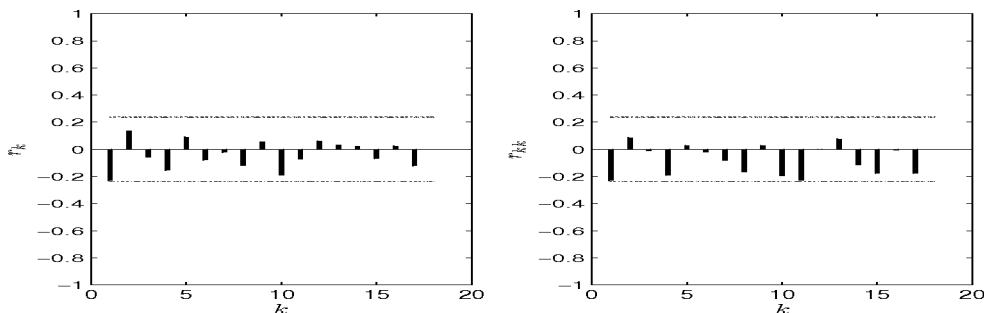
Pokud budeme řadu teplot zkoumat po částech, tj. každý z úseků  $X_{t_1}$ ,  $X_{t_2}$ ,  $X_{t_3}$  a  $X_{t_4}$  budeme modelovat zvlášť, dospějeme k různým modelům. Z grafů na obrázcích 53, 55, 57 a 59 je zřejmé, že řady tvořené jednotlivými úseky nejsou stacionární. Před samotnou identifikací modelu je bude tedy třeba ještě upravit. Pro všechny čtyři úseky bylo použito běžné diferencování, odstranění trendu vyrovnaním modelem konstantní sezónnosti se schodovitým trendem a kombinace těchto dvou úprav. Zde je vždy uvedena ta varianta, pro kterou byl výsledný model nejjednodušší ( $p + q \leq 2$ ).



Obrázek 53: Řada  $X_{t_1}$  a stacionární řada  $X_{t_1}^*$

V případě úseku  $X_{t_1}$  (viz. obr. 53 a obr. 54) dostáváme po druhé diferenci řadu  $X_{t_1}^*$ , kterou identifikujeme jako proces bílého šumu s rozptylem  $\hat{\sigma}_{X_1} = 0,5042$ . Jedná se tedy o proces ARIMA  $(0, 2, 0)$ , který zapíšeme jako

$$X_{t_1} = X_{t_1-1} + X_{t_1-2} + \varepsilon_{t_1}. \quad (215)$$

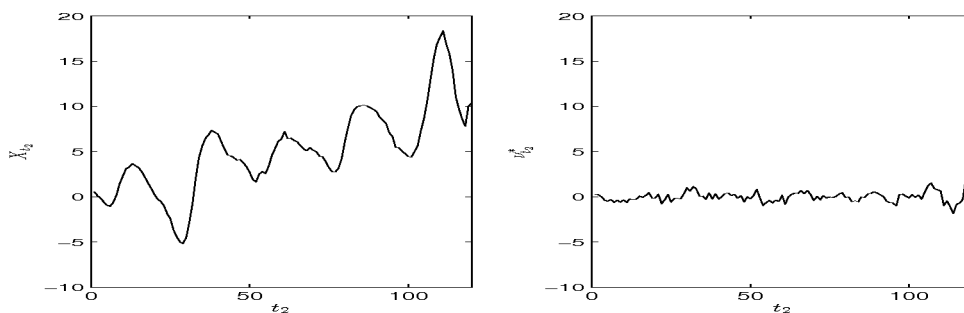


Obrázek 54: Grafy odhadnutých funkcí  $r_k$  a  $r_{kk}$  pro řadu  $X_{t_1}^*$

Výsledky testů normality a autokorelace pro proces (215) jsou uvedeny v tabulce 17.

Reziduální součet	35,78
Směrodatná odchylka reziduí	0,71
Durbinův-Watsonův test	2,46
Portmanteau test $Q^*$	9,69
Minimální hladina významnosti $Q^*$ pro $K = 8$	0,2
Jarqueův-Berův test	7,18
Minimální hladina významnosti J.-B. testu	0,028

Tabulka 17: Základní charakteristiky pro řadu  $X_{t_1}^*$

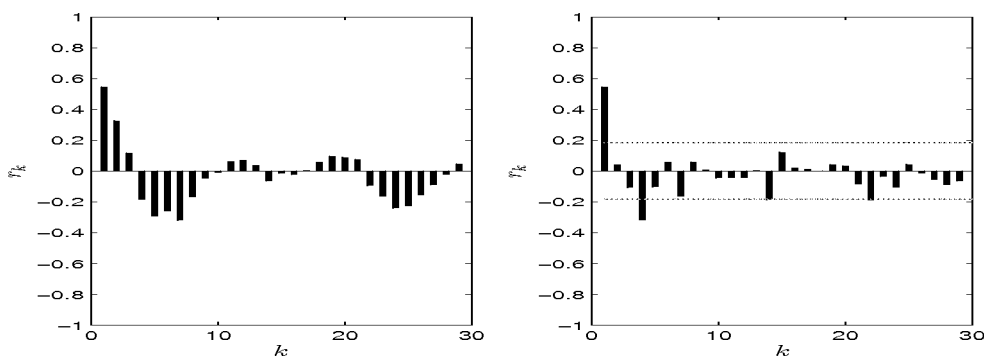


Obrázek 55: Řada  $X_{t_2}$  a stacionární řada  $\nu_{t_2}^*$

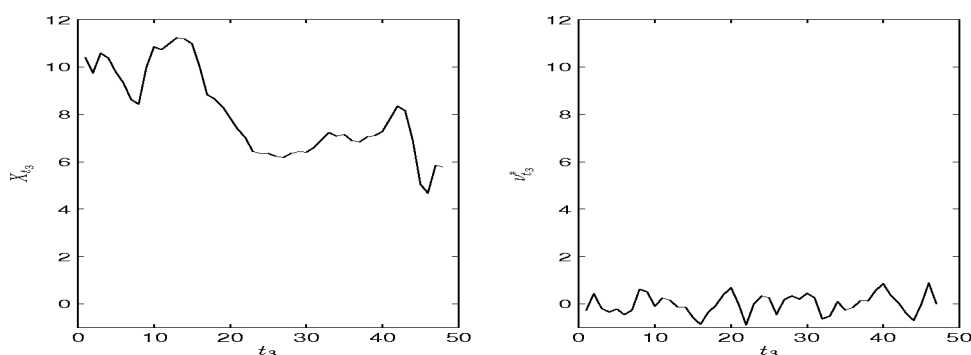
Pro řadu  $\nu_{t_2}^*$ , kterou tvoří první diference reziduí  $\nu_{t_2}$  získaných vyrovnáním úseku  $X_{t_2}$  modelem konstantní sezónnosti se schodovitým trendem, byl nalezen model AR(1) ve tvaru

$$\nu_{t_2}^* = 0,5572\nu_{t_2-1}^* + \varepsilon_{t_2}. \quad (216)$$

Grafický záznam řad  $X_{t_2}$  a  $\nu_{t_2}^*$  znázorňuje obrázek 55, odhady  $r_k$  a  $r_{kk}$  autokorelační a parciální autokorelační funkce řady  $\nu_{t_2}^*$  jsou zobrazeny na obr. 56.



Obrázek 56: Grafy odhadnutých funkcí  $r_k$  a  $r_{kk}$  pro řadu  $\nu_{t_2}^*$



Obrázek 57: Řada  $X_{t_3}$  a stacionární řada  $\nu_{t_3}^*$

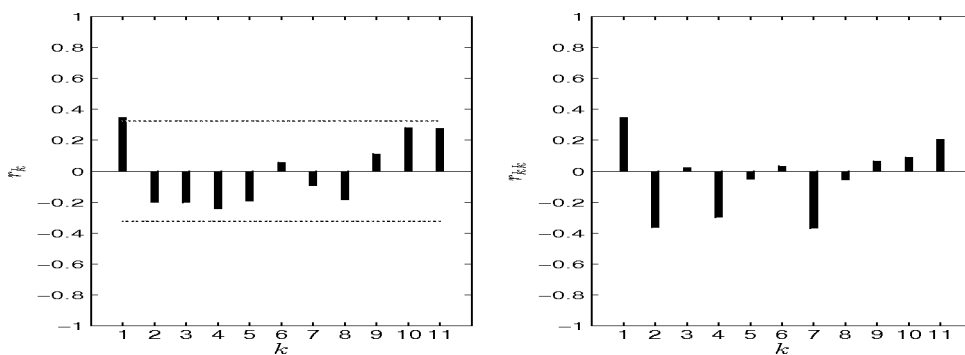
Na obrázcích 57 a 59 jsou zobrazeny řady  $X_{t_3}$ ,  $X_{t_4}$  a stacionární řady  $\nu_{t_3}^*$  a  $\nu_{t_4}^*$  získané z těchto řad odstraněním trendu a následným diferencováním stejně jako v případě řady  $X_{t_2}$ . Grafy odhadů  $r_k$ ,  $r_{kk}$  autokorelační a parciální autokorelační funkce pro řadu  $\nu_{t_3}^*$  znázorňuje obr. 58, pro řadu  $\nu_{t_4}^*$  potom obr. 60.

Pro řadu  $\nu_{t_3}^*$  byl identifikován model MA(1) ve tvaru

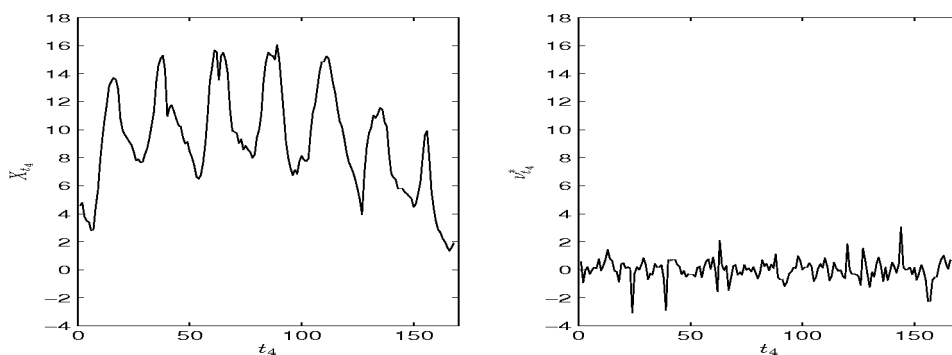
$$\nu_{t_3}^* = 0,604\varepsilon_{t_3-1}^* + \varepsilon_{t_3}, \quad (217)$$

pro řadu  $\nu_{t_4}^*$  model AR(1) ve formě

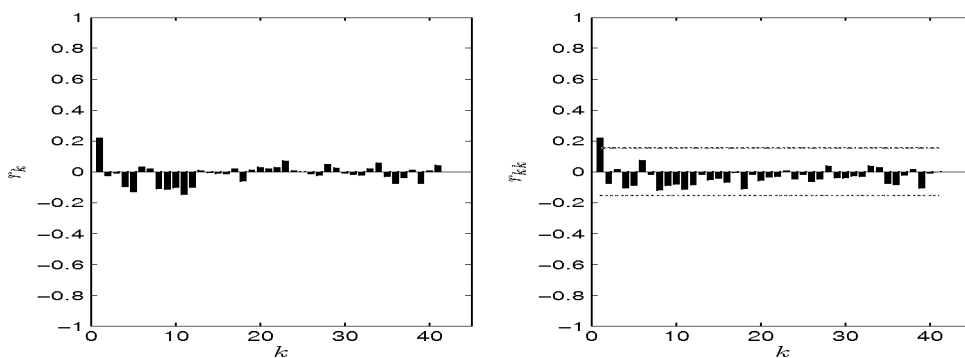
$$\nu_{t_4}^* = 0,2197\nu_{t_4-1}^* + \varepsilon_{t_4}. \quad (218)$$



Obrázek 58: Grafy odhadnutých funkcí  $r_k$  a  $r_{kk}$  pro řadu  $\nu_{t_3}^*$



Obrázek 59: Řada  $X_{t_4}$  a stacionární řada  $\nu_{t_4}^*$



Obrázek 60: Grafy odhadnutých funkcí  $r_k$  a  $r_{kk}$  pro řadu  $\nu_{t_4}^*$

V tabulce 18 jsou uvedeny základní charakteristiky modelů analyzovaných časových řad  $\nu_{t_2}^*$  (druhý sloupec),  $\nu_{t_3}^*$  (třetí sloupec) a  $\nu_{t_4}^*$  (čtvrtý sloupec), výsledky testů normality a neautokorelovanosti a porovnávací kritéria.

Model, délka řady	AR(1), 119	MA(1), 47	AR(1), 167
Odhady parametrů modelu	$\hat{\varphi}_1 = 0,5572$	$\hat{\theta}_1 = 0,604$	$\hat{\varphi}_1 = 0,2197$
Reziduální součet	25,02	6,12	86,80
Směrodatná odchylka reziduí	0,46	0,36	0,73
Směrodatná odchylka odhadu parametru	0,076	0,12	0,089
Durbinův-Watsonův test	2,08	1,97	1,94
Portmanteau test $Q^*$	19,74	8,22	13,62
Minimální hladina významnosti $Q^*$	0,011	0,22	0,32
Jarqueův-Berův test	4,15	0,002	5,71
Minimální hladina význ. J.-B. testu	0,125	0,998	0,055
AIC	-181,56	-93,81	-105,27
BIC	-178,63	-90,99	-104,18
SBC	30,23	10,10	92,97
FPE	0,22	0,14	0,53

Tabulka 18: Základní charakteristiky modelů pro řady  $\nu_{t_2}^*$ ,  $\nu_{t_3}^*$  a  $\nu_{t_4}^*$

$i_k$	1	2	3	4	5	6	7
$T_{i_2}$	0,91	2,29	4,66	7,11	10,28	—	—
$T_{i_3}$	9,34	6,69	—	—	—	—	—
$T_{i_4}$	8,34	10,63	10,95	11,10	10,98	8,13	4,71

Tabulka 19: Odhady parametrů trendů  $T_{i_2}$ ,  $T_{i_3}$ ,  $T_{i_4}$

Pro původní úseky řady  $X_t$  dostáváme modely

$$X_{t_k} = T_{i_k} + S_{j_k} + \nu_{t_k}, \quad t_k = (i_k - 1)24 + j_k, \quad (219)$$

kde  $k = 2, 3, 4$ ,  $i_2 = 1, \dots, 5$ ,  $i_3 = 1, 2$ ,  $i_4 = 1, \dots, 7$ ,  $j_2 = j_3 = j_4 = 1, \dots, 24$ , rezidua  $\nu_{t_k}$  jsou procesy v následujících tvarech

$$\nu_{t_2} = 1,5572\nu_{t_2-1} - 0,5572\nu_{t_2-2} + \varepsilon_{t_2}, \quad (220)$$

$$\nu_{t_3} = \nu_{t_3-1} - 0,604\varepsilon_{t_3-1} + \varepsilon_{t_3}, \quad (221)$$

$$\nu_{t_4} = 1,2197\nu_{t_4-1} - 0,2197\nu_{t_4-2} + \varepsilon_{t_4}, \quad (222)$$

parametry trendů  $T_{i_k}$  obsahuje tabulka 19 a parametry sezón  $S_{j_k}$  tabulka 20.

$j_k$	1	2	3	4	5	6	7	8	9	10	11	12
$S_{j_2}$	-2,30	-2,53	-2,83	-3,16	-3,34	-3,43	-3,41	-2,68	-1,28	-0,16	2,05	2,97
$S_{j_3}$	0,36	-0,03	0,36	0,34	0,10	-0,15	-0,40	-0,33	0,59	0,94	0,92	0,92
$S_{j_4}$	-1,82	-2,22	-2,31	-2,56	-2,86	-3,16	-3,11	-1,68	-0,51	0,85	2,51	3,55
$j_k$	13	14	15	16	17	18	19	20	21	22	23	24
$S_{j_2}$	3,61	3,84	4,14	4,11	3,79	2,95	1,52	0,53	-0,42	-1,18	1,40	-1,35
$S_{j_3}$	1,01	1,10	1,02	0,64	0,30	0,48	0,22	-0,64	-1,80	-2,16	-1,88	-1,95
$S_{j_4}$	3,84	3,71	3,26	2,86	2,78	2,28	1,05	-0,17	-1,00	-1,35	-1,79	-2,15

Tabulka 20: Odhady parametrů sezón  $S_{j_2}$ ,  $S_{j_3}$ ,  $S_{j_4}$

## 7.4 Srovnání

Na závěr se pokusíme porovnat jednotlivé modely. Nelze jednoznačně určit, který model je nejlepší, protože po úpravách jsme modelovali různé řady různých délek. V tabulce 21 jsou uvedeny reziduální rozptyly pro jednotlivé řady, přičemž hodnota rozptylu v posledním sloupci této tabulky je vypočítána jako vážený průměr reziduálních rozptylů řad  $X_{t_1}^*$ ,  $\nu_{t_2}^*$ ,  $\nu_{t_3}^*$ ,  $\nu_{t_4}^*$ , tj.

$$0,72^2 \frac{70}{403} + 0,46^2 \frac{119}{403} + 0,36^2 \frac{47}{403} + 0,73^2 \frac{167}{403} = 0,386 \quad (223)$$

Rozptyly se liší velmi málo, z tohoto hlediska jsou tedy modely srovnatelné.

Řada	$w_t$	$\nu_t^*$	$X_{t_1}^*$	$\nu_{t_2}^*$	$\nu_{t_3}^*$	$\nu_{t_4}^*$
Model	SARIMA(1, 0, 0) $\times$ (0, 0, 1) <sub>24</sub>	AR(1)	$\mathcal{WN}$	AR(1)	MA(1)	AR(1)
Rozptyl	0,4507	0,5372	0,3860			

Tabulka 21: Reziduální rozptyly pro jednotlivé modely

Výpočetně je nejsložitější model SARIMA(1, 1, 0)  $\times$  (0, 1, 1)<sub>24</sub>, ale je nejvhodnější z hlediska předpovědí, protože dokáže rychle reagovat na změnu dat. Postup v dalších dvou přístupech je stejně výpočetně náročný, samozřejmě musíme vzít v úvahu, že v modelování po částech ho provádíme čtyřikrát. Zajímavá je podobnost mezi modely (214) pro řadu všech pozorování  $X_t$  a (222) pro koncový úsek  $X_{t_4}$ . Předpovědi by byly pro oba přístupy obdobné. Pro přístup, kdy zkoumáme řadu po částech, nalezneme uplatnění především v případě, že chceme získat co nejvíce informací o chování řady.

## 8 Závěr

V diplomové práci je popsána a na reálných datech představena teorie potřebná k pochopení principů Boxovy-Jenkinsovy metodologie. Teoretická část začíná druhou kapitolou, kde jsou popsány nástroje korelační analýzy, které jsou používány v Boxově-Jenkinsově metodologii, jde především o autokorelační a partiální autokorelační funkci. Třetí kapitola je věnována stacionárním procesům. Nejprve je popsán obecný lineární model, potom i jeho speciální případy: proces klouzavých součtů MA, autoregresní proces AR a smíšený proces ARMA. Problémem identifikace se zabývá čtvrtá kapitola, ve které jsou popsány i metody odhadů parametrů pro jednotlivé procesy a také způsoby verifikace modelu. Rozšířené ARMA modely jsou náplní páté kapitoly. Jedná se o integrované modely ARIMA a sezónní modely SARIMA. Konstrukce předpovědí je vysvětlena v předposlední kapitole, která uzavírá teoretickou část. Poslední kapitola ukazuje praktický postup při hledání modelu řady teplot. Jsou zde popsány tři přístupy. Tento příklad je názornou ukázkou toho, že musíme vyzkoušet několik modelů, než najdeme ten vhodný. Při výběru modelu hraje také roli, za jakým účelem jej hledáme, zda pro konstrukci předpovědí budoucího vývoje řady nebo pro získání co nejvíce informací o jejím dosavadním chování.

Model získaný pouze použitím Boxova-Jenkinsova přístupu je výpočetně nejnáročnější, ale nejvhodnější pro konstrukci předpovědí. Přístup, kde nejprve odstraníme trend a model hledáme až pro zbylá rezidua, je jednodušší, ale chyby v tomto modelu mají největší rozptyl. Postupem, kdy zkoumáme řadu po částech, získáváme pro každou z těchto částí jiný model. Tímto přístupem je nejpřesněji popsán průběh teplot v daném období.

Tato práce se zabývá modelováním jednorozměrných časových řad. Na tuto teorii navazuje rozsáhlejší teorie vícerozměrných časových řad, která mimo jiné zkoumá vztah mezi jednotlivými řadami. Identifikace a specifikace těchto vztahů je důležitá například pro provádění hospodářské politiky, protože některé časové řady (např. časová řada zásoby peněz) jsou pod kontrolou centrálních institucí a lze je používat jako nástroje řízení a korigování ekonomického vývoje.



## Seznam obrázků

1	$\varrho_k$ a $\varrho_{kk}$ pro parametr $\theta_1 > 0$ . . . . .	20
2	$\varrho_k$ a $\varrho_{kk}$ pro parametr $\theta_1 < 0$ . . . . .	20
3	Simulovaný proces $X_t = \varepsilon_t + 0,8\varepsilon_{t-1}$ . . . . .	21
4	Odhad autokorelační a parciální autokorelační funkce procesu $X_t$ .	21
5	Simulovaný proces $X_t = \varepsilon_t - 0,8\varepsilon_{t-1}$ . . . . .	21
6	Odhad autokorelační a parciální autokorelační funkce procesu $X_t$ .	22
7	Oblast invertibility procesu MA(2) . . . . .	24
8	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\theta_1 < 0, \theta_2 < 0$ . . . . .	25
9	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\theta_1 > 0, \theta_2 > 0$ . . . . .	26
10	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\theta_1 > 0, \theta_2 < 0$ . . . . .	26
11	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\theta_1 < 0, \theta_2 > 0$ . . . . .	26
12	Simulovaný proces $X_t = \varepsilon_t + 0,8\varepsilon_{y-1} - 0,4\varepsilon_{t-2}$ . . . . .	27
13	Odhad autokorelační a parciální autokorelační funkce procesu $X_t$ .	27
14	$\varrho_k$ a $\varrho_{kk}$ pro parametr $\varphi_1 > 0$ . . . . .	30
15	$\varrho_k$ a $\varrho_{kk}$ pro parametr $\varphi_1 < 0$ . . . . .	30
16	Simulovaný proces $X_t = 0,8X_{t-1} + \varepsilon_t$ . . . . .	30
17	Odhad autokorelační a parciální autokorelační funkce procesu $X_t$ .	31
18	Simulovaný proces $X_t = -0,8X_{t-1} + \varepsilon_t$ . . . . .	31
19	Odhad autokorelační a parciální autokorelační funkce procesu $X_t$ .	31
20	Oblast stacionarity procesu AR(2) . . . . .	33
21	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\varphi_1 > 0, \varphi_2 > 0$ . . . . .	34
22	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\varphi_1 < 0, \varphi_2 < 0$ . . . . .	34
23	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\varphi_1 > 0, \varphi_2 < 0$ . . . . .	35
24	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\varphi_1 < 0, \varphi_2 > 0$ . . . . .	35
25	Simulovaný proces $X_t = 0,8X_{t-1} - 0,4X_{t-2} + \varepsilon_t$ . . . . .	35
26	Odhad autokorelační a parciální autokorelační funkce procesu $X_t$ .	36
27	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\theta_1 > 0, \varphi_1 > 0$ . . . . .	38
28	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\theta_1 < 0 < \varphi_1$ . . . . .	39
29	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\varphi_1 < 0 < \theta_1$ . . . . .	39
30	$\varrho_k$ a $\varrho_{kk}$ pro parametry $\varphi_1 < 0, \theta_1 < 0$ . . . . .	39
31	Simulovaný proces $X_t = -0,8X_{t-1} + \varepsilon_t - 0,4\varepsilon_{t-1}$ . . . . .	40
32	Odhad autokorelační a parciální autokorelační funkce procesu $X_t$ .	40
33	Graf hodnot z tabulky 1 . . . . .	47
34	Odhad autokorelační funkce $\varrho_k$ . . . . .	47
35	Odhad parciální autokorelační funkce $\varrho_{kk}$ . . . . .	48
36	Součty čtverců pro parametr $\varphi_1$ na intervalu $\langle -1, 1 \rangle$ . . . . .	50
37	Graf hodnot funkce $r_k(\hat{\varepsilon})$ a dvojnásobku jejich směrodatných odchylek . . . . .	63
38	Naměřené teploty . . . . .	76
39	$r_k$ pro řadu naměřených teplot . . . . .	77
40	$r_k$ řady po sezónním diferencování . . . . .	77

41	Řada $w_t$ . . . . .	77
42	$r_k$ řady $w_t$ . . . . .	78
43	$r_{kk}$ řady $w_t$ . . . . .	78
44	Součty $S(\varphi_1, \Theta_1)$ na mříži parametrů $\varphi_1 = 0, 34; 0, 341; \dots; 0, 36$ a $\Theta_1 = -0, 8; -0, 799; \dots; -0, 78$ . . . . .	80
45	Porovnání naměřených a vyrovnaných hodnot . . . . .	84
46	Rezidua $\nu_t$ . . . . .	84
47	Odhad autokorelační funkce $\varrho_k$ řady $\nu_t$ . . . . .	85
48	Odhad parciální autokorelační funkce $\varrho_{kk}$ řady $\nu_t$ . . . . .	85
49	Řada prvních diferencí $\nu_t^*$ . . . . .	86
50	Odhad parciální autokorelační funkce $\varrho_k$ řady $\nu_t^*$ . . . . .	86
51	Odhad parciální autokorelační funkce $\varrho_{kk}$ řady $\nu_t^*$ . . . . .	87
52	Hodnoty $r_k(\hat{\varepsilon})$ pro řadu $\nu_t^*$ . . . . .	88
53	Řada $X_{t_1}$ a stacionární řada $X_{t_1}^*$ . . . . .	89
54	Grafy odhadnutých funkcí $r_k$ a $r_{kk}$ pro řadu $X_{t_1}^*$ . . . . .	90
55	Řada $X_{t_2}$ a stacionární řada $\nu_{t_2}^*$ . . . . .	90
56	Grafy odhadnutých funkcí $r_k$ a $r_{kk}$ pro řadu $\nu_{t_2}^*$ . . . . .	91
57	Řada $X_{t_3}$ a stacionární řada $\nu_{t_3}^*$ . . . . .	91
58	Grafy odhadnutých funkcí $r_k$ a $r_{kk}$ pro řadu $\nu_{t_3}^*$ . . . . .	92
59	Řada $X_{t_4}$ a stacionární řada $\nu_{t_4}^*$ . . . . .	92
60	Grafy odhadnutých funkcí $r_k$ a $r_{kk}$ pro řadu $\nu_{t_4}^*$ . . . . .	92

## Seznam tabulek

1	Kurz CZK/AUD v období od 30.6.2008 do 12.9.2008 . . . . .	12
2	Hodnoty odhadnuté autokorelační a parciální autokorelační funkce	12
3	Odhady směrodatných odchylek pro průměr . . . . .	43
4	Tvar autokorelační a parciální autokorelační funkce . . . . .	45
5	Tabulka počátečních odhadů a omezení na autokorelační funkci .	46
6	Princip zpětného extrapolování v modelu AR(1) pro $\varphi_1 = -0,5$ .	52
7	Přibližné směrodatné odchylky odhadnutých parametrů . . . . .	56
8	Prvních osm hodnot funkce $r_k(\hat{\varepsilon})$ . . . . .	62
9	Základní charakteristiky modelu AR(1) z příkladu 2.1 . . . . .	65
10	Součty $S_p^w(\varphi_1, \Theta_1)$ . . . . .	80
11	Prvních dvacet hodnot funkce $r_k(\hat{\varepsilon})$ . . . . .	81
12	Základní charakteristiky modelu SARIMA $(1, 0, 0) \times (0, 0, 1)_{24}$ pro řadu $w_t$ . . . . .	82
13	Odhady parametrů trendu $T_i$ . . . . .	83
14	Odhady parametrů sezón $S_j$ . . . . .	83
15	Odhad prvních deseti hodnot $\varrho_k$ a $\varrho_{kk}$ . . . . .	85
16	Základní charakteristiky modelu AR(1) pro řadu $\nu_t^*$ . . . . .	88
17	Základní charakteristiky pro řadu $X_{t_1}^*$ . . . . .	90
18	Základní charakteristiky modelů pro řady $\nu_{t_2}^*$ , $\nu_{t_3}^*$ a $\nu_{t_4}^*$ . . . . .	93
19	Odhady parametrů trendů $T_{i_2}$ , $T_{i_3}$ , $T_{i_4}$ . . . . .	93
20	Odhady parametrů sezón $S_{j_2}$ , $S_{j_3}$ , $S_{j_4}$ . . . . .	94
21	Reziduální rozptyly pro jednotlivé modely . . . . .	94

## Literatura

- [1] Anděl, J.: *Statistická analýza časových řad*. SNTL, Praha, 1976.
- [2] Artl, J.: *Moderní metody modelování ekonomických časových řad*. Grada Publishing, spol s.r.o., Praha, 1999.
- [3] Cipra, T.: *Analýza časových řad s aplikacemi v ekonomii*. SNTL/ALFA, Praha, 1986.
- [4] Hindls, R., Hronová, S., Seger, J.: *Statistika pro ekonomy*. Professional Publishing, Praha, 2006.
- [5] Kozák, J., Hindls, R., Artl, J.: *Úvod do analýzy ekonomických časových řad*. Kapitola 4. VŠE v Praze, Praha, 1994.
- [6] Vašíček, T.: *Úvod do analýzy časových řad* [online]. Dostupné z <http://www.econ.muni.cz/~qasar/vyuka/emm2/skriptaemmii.pdf> [cit. 2009-11-21].
- [7] Veselý, T.: *Úvod do časových řad* [online]. Dostupné z <http://www.math.muni.cz/~vesely/educ/sptext.pdf> [cit. 2009-11-13].