

MASARYKOVA UNIVERZITA  
PŘÍRODOVĚDECKÁ FAKULTA  
ÚSTAV MATEMATIKY A STATISTIKY

# Diplomová práce

BRNO 2015

VAVŘINEC HAVLÍČEK



MASARYKOVA UNIVERZITA  
PŘÍRODOVĚDECKÁ FAKULTA  
ÚSTAV MATEMATIKY A STATISTIKY

---



# ARFIMA modely - procesy s dlouhou pamětí

Diplomová práce

Vavřinec Havlíček

Vedoucí práce: RNDr. Marie Forbelská, Ph.D.

Brno 2015

# Bibliografický záznam

**Autor:** Bc. Vavřínek Havlíček  
Přírodovědecká fakulta, Masarykova univerzita  
Ústav matematiky a statistiky

**Název práce:** ARFIMA modely - procesy s dlouhou pamětí

**Studijní program:** Aplikovaná matematika

**Studijní obor:** Statistika a analýza dat

**Vedoucí práce:** RNDr. Marie Forbelská, Ph.D.

**Akademický rok:** 2014/2015

**Počet stran:** ix + 74

**Klíčová slova:** Náhodné procesy; časové řady; procesy s dlouhou pamětí; Autoregresní frakcionálně integrovaný procesy klouzavých součtů (ARFIMA); program R; balíček `forecast`; balíček `fracdiff`

# Bibliographic Entry

**Author:** Bc. Vavřínek Havlíček  
Faculty of Science, Masaryk University  
Department of Mathematics and Statistics

**Title of Thesis:** ARFIMA - time series with long memory

**Degree Programme:** Applied Mathematics

**Field of Study:** Statistics and Data Analysis

**Supervisor:** RNDr. Marie Forbelská, Ph.D.

**Academic Year:** 2014/2015

**Number of Pages:** ix + 74

**Keywords:** Stochastic process; Time Series; Long-Memory Processes; Autoregressive Fractionally Integrated Moving Average (ARFIMA) processes; The R Project for Statistical Computing; package `forecast`; package `fracdiff`

# Abstrakt

Tato diplomová práce je úvodem do problematiky modelování časových řad s dlouhou pamětí. K modelům jsou použity autoregresní frakcionálně integrované procesy klouzavých součtů. Jsou definovány jejich základní vlastnosti a vysvětleny metody odhadů jejich parametrů. Na reálné časové řadě jsou uskutečněny předpovědi pomocí polynomického a periodického trendu. Ty jsou následně porovnány s ARFIMA modely.

# Abstract

This thesis is an introduction to the problematics associated with modeling the time series with long memory. For this modeling are used Autoregressive Fractionally Integrated Moving Average (ARFIMA) processes. The basic properties are defined and the estimation methods for the parameters are explained. The real time series data is forecasted by polynomial and periodic trend. This forecast is compared with the ARFIMA models.



Masarykova univerzita

Přírodovědecká fakulta



## ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: **Bc. Vavřinec Havlíček**

Studijní program - obor: **Matematika - Statistika a analýza dat**

Ředitel Ústavu matematiky a statistiky PřF MU Vám ve smyslu Studijního a zkušebního řádu MU určuje diplomovou práci s tématem:

### **ARFIMA modely – procesy s dlouhou pamětí**

#### **ARFIMA - time series with long memory**

*Oficiální zadání:* Pro modelování časových řad s dlouhou pamětí se používají tzv. ARFIMA modely, tj. autoregresní frakcionálně integrované procesy klouzavých součtů. Cílem diplomové práce je seznámit se s problematikou těchto modelů, popsat jejich vlastnosti a shrnout metody odhadů neznámých parametrů modelu. V programovacím prostředí R na reálných (popř. simulovaných) datech demonstrovat vyložené metody.

#### *Doporučená literatura*

*FORBELSKÁ, Marie. Stochastické modelování jednorozměrných časových řad. 1. vyd. Brno: Masarykova univerzita, 2009. 251 s. 4761/Př-3/09-17/31. ISBN 978-80-210-4812-6.*

*PFAFF, Bernhard. Analysis of integrated and cointegrated time series with R. 2nd ed. New York, N.Y. Springer, 2008. xx, 188 s. ISBN 9780387759661.*

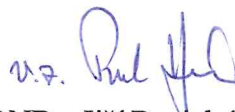
*SHUMWAY, Robert H. a David S. STOFFER. Time series analysis and its applications :with R examples. 3rd ed. New York, N.Y. Springer, 2006. xiii, 575. ISBN 0-387-29317-5.*

*Vedoucí diplomové práce:* RNDr. Marie Forbelská, Ph.D. 

*Datum zadání diplomové práce:* listopad 2012

*Datum odevzdání diplomové práce:* dle harmonogramu ak. roku 2013/2014

V Brně dne 14. 11. 2012



prof. RNDr. Jiří Rosický, DrSc.  
ředitel Ústavu matematiky a statistiky

Zadání diplomové práce převzal dne:

*3. prosince 2012*

Podpis studenta

*Havlíček*

# Poděkování

Děkuji RNDr. Marie Forbelské, Ph.D. za odborné vedení mé práce, za cenné rady a za čas, který strávila čtením a konzultací tohoto textu.

# Prohlášení

Prohlašuji, že jsem svou diplomovou práci vypracoval samostatně a výhradně s využitím citovaných pramenů.

Brno 6. ledna 2015

.....  
Vavřinec Havlíček

# Obsah

<b>1</b>	<b>Teoretické základy analýzy časových řad</b>	<b>1</b>
1.1	Základní pojmy . . . . .	1
1.2	Stacionární náhodné procesy . . . . .	3
1.2.1	Herglotzova věta . . . . .	4
1.2.2	Periodogram . . . . .	5
1.2.3	Test R. A. Fishera . . . . .	8
1.2.4	Siegelův test . . . . .	9
1.2.5	Iterativní metoda . . . . .	9
1.2.6	Neparametrické odhady spektrální hustoty . . . . .	10
1.2.7	Definice ARMA . . . . .	14
1.3	Modelování nestacionarity . . . . .	18
1.3.1	ARIMA procesy: diferenčně stacionární náhodné procesy . . . . .	18
1.3.2	Regresní modely . . . . .	19
<b>2</b>	<b>Procesy s dlouhou pamětí</b>	<b>22</b>
2.1	Definování dlouhé paměti pomocí autokovariancí . . . . .	22
2.2	Fracionálně diferencovaný bílý šum . . . . .	22
2.3	Odhady parametrů . . . . .	24
2.4	ARFIMA procesy . . . . .	26
2.4.1	Stacionarita, kauzalita a invertibilita . . . . .	27
2.4.2	Spektrální hustota . . . . .	27
2.4.3	Whittleův odhad . . . . .	28
<b>3</b>	<b>Příklad reálné časové řady modelované pomocí ARFIMA modelů</b>	<b>30</b>
3.1	Popis zkoumaných dat . . . . .	30
3.2	Polynomický trend s ARMA reziduální částí . . . . .	33
3.3	Periodický trend s ARMA reziduální částí . . . . .	38
3.4	Balíčky <code>forecast</code> a <code>fracdiff</code> . . . . .	47
3.5	Srovnání s dalšími balíčky . . . . .	56
3.5.1	Balíček <code>fracdiff</code> . . . . .	57
3.5.2	Balíček <code>FGN</code> . . . . .	57
3.5.3	Balíček <code>arfima</code> . . . . .	58
3.5.4	Balíček <code>longmemo</code> . . . . .	59



3.5.5 Balíček afmtools .....	61
<b>Závěr</b>	<b>64</b>
<b>Přílohy</b>	<b>65</b>
Pomocná funkce tperiod .....	65
Pomocná funkce fisher .....	66
Pomocná funkce damsleth .....	68
Pomocná funkce fishper .....	69
Pomocná funkce na výpočet kombinačního čísla .....	70
Pomocná funkce trigpol1 .....	70
Pomocná funkce trigoval .....	70
Pomocná funkce trigofit .....	71
Metoda skrytých period .....	71
Pomocná funkce trigpol2 .....	72
<b>Seznam použité literatury</b>	<b>73</b>

# Úvod

Dlouhé paměti si u časových řad jako první všiml britský hydrolog Harold Edwin Hurst. Studoval minimální hladiny řeky Nilu za účelem předpovědi kolísání meziročního toku vody. Své výsledky publikoval v roce 1951 ve svém článku [16]. Cílem jeho práce bylo určit objem přehrady dostatečně veliký na to, aby zajistil konstantní odtok vody i přes vysoké nepravidelnosti v přítocích do přehrady. Hurst si při svých výzkumech všiml, že množství vody v řece silně závisí na pozorováních ze vzdálené minulosti. Tím se závislost mezi jednotlivými naměřenými hodnotami lišila od standardních modelů. Konkrétně autokorelační funkce klesala pomaleji než exponenciálně. Později byl pro tento jev zaveden termín Hurstův fenomén nebo též dlouhá paměť jako protiklad ke krátké paměti určené exponenciálním poklesem autokorelační funkce [22].

Diplomová práce je rozdělena do tří kapitol a doplněna o přílohy. V první kapitole představíme teorii modelování jednorozměrných časových řad na základě literatury [9]. V druhé kapitole uvedeme teorii k časovým řadám s dlouhou pamětí a definujeme tzv. ARFIMA modely, tj. autoregresní frakcionálně integrované procesy klouzavých součtů. Popíšeme jejich vlastnosti a shrneme metody odhadů neznámých parametrů modelu. Vycházíme převážně z článku [1], z knihy [19] a článku [5]. V třetí kapitole ukážeme na reálných datech vlastnosti dlouhé paměti v programovacím prostředí R [21]. Srovnáme balíčky umožňující předpovědi na základě ARFIMA modelů. V příloze jsou uvedeny pomocné funkce užité při analýze výše uvedených dat.

# Kapitola 1

## Teoretické základy analýzy časových řad

V této kapitole začneme základními pojmy pro modelování jednorozměrných časových řad, tj. nadefinováním *náhodného procesu*, jeho momentů, *striktní* a *slabé stacionarity* a dalších základních definic z [9]. V části 1.2 pokračujeme uvedením Herglotzovy věty, definováním *periodogramu*, Fisherovým a Siegelovým testem, iterativní metodou, neparametrickými odhady spektrální hustoty a definicí *ARMA procesu*, taktéž z [9]. Na závěr kapitoly definujeme *diferenčně stacionární náhodné procesy* a uvedeme základní regresní modely, tj. polynomický a periodický trend.

### 1.1 Základní pojmy

**1.1 Definice.** Nechť je dán pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$ , indexová množina  $T \subseteq \mathbb{R}$  a reálná funkce  $X : \Omega \times T \rightarrow \mathbb{R}$  definovaná pro  $\omega \in \Omega$  a  $t \in T$ .

Jestliže pro  $t \in T$  je  $X(\omega, t)$  borelovsky měřitelná funkce vzhledem k  $\mathcal{A}$  (tj. pro  $B \in \mathcal{B}$ ,  $t \in T$  platí  $X^{-1}(B) = \{\omega \in \Omega : X(\omega, t) \in B\} \in \mathcal{A}$ , kde  $\mathcal{B}$  je  $\sigma$ -algebra borelovských podmnožin), pak tuto funkci nazýváme ( $n$ -rozměrným) *náhodným procesem*.

Náhodný proces  $X(\omega, t)$  při pevném  $\omega \in \Omega$  se nazývá *realizace (trajektorie) procesu*. Pravděpodobnostní míru  $P_X(B) = P(X^{-1}(B))$  nazýváme *rozdělení pravděpodobností náhodného procesu*  $X(\omega, t)$ .

**Poznámka.** Stejně jako u náhodné veličiny  $X(\omega)$ ,  $\omega \in \Omega$ , kterou jsme zapisovali jen  $X$ , budeme při zapisování náhodného procesu  $\{X(\omega, t) | \omega \in \Omega, t \in T\}$  využívat zkrácený tvar  $\{X_t | t \in T\}$ .

**1.2 Příklad.** Náhodná procházka  $\{X_t | t = 1, 2, \dots\}$  je definována

$$X_0 = 0, \quad X_t = X_{t-1} + \varepsilon_t,$$

kde  $\varepsilon_t$  jsou nezávislé stejně rozdělené náhodné veličiny s nulovou střední hodnotou a konstantním rozptylem, tj.

$$\varepsilon_t \sim \mathcal{L}(0, \sigma^2).$$

**1.3 Definice.** Existuje-li pro  $t \in T$  střední hodnota  $EX_t$ , pak funkci

$$\mu_t = EX_t$$

nazýváme *střední hodnotou* náhodného procesu  $\{X_t | t \in T\}$ .

**1.4 Definice.** Náhodný proces  $\{X_t | t \in T\}$  nazýváme *procesem druhého řádu*, jestliže pro  $t \in T$  platí

$$EX_t^2 < \infty.$$

Náhodný proces má konečné druhé momenty.

Z definice náhodného procesu druhého řádu a ze Schwarzovy nerovnosti plyne, že pro tyto procesy vždy existuje střední hodnota  $EX_t = \mu_t$  a rozptyl  $DX_t = EX_t^2 - (EX_t)^2 = \sigma_t^2$ .

**1.5 Definice.** Náhodný proces  $\{X_t | t \in T\}$  nazýváme *stacionární ve střední hodnotě*, pokud pro  $t \in T$  je střední hodnota konstantní, tj.  $EX_t = \mu$ . Pokud  $EX_t = 0$ , náhodný proces nazýváme *centrovaným*.

**1.6 Definice.** Uvažujme náhodný proces  $\{X_t | t \in T\}$ , který má konečné druhé momenty. Pak funkci

$$\gamma(s, t) = C(X_s, X_t) = E(X_s - EX_s)(X_t - EX_t)$$

nazveme *autokovarianční funkcí*.

Kovarianční funkce udává míru lineární závislosti mezi jakoukoli dvojicí náhodných veličin  $X_s$  a  $X_t$ .

**1.7 Definice.** Náhodný proces  $\{X_t | t \in T\}$  se nazývá *kovariančně stacionární*, pokud pro  $t, s \in T$  platí  $\gamma(s, t) = \gamma(0, |s-t|)$ , což budeme také psát ve formě  $\gamma(s, t) = \gamma(s-t)$ , tj. autokovarianční funkce závisí na svých argumentech pouze prostřednictvím jejich rozdílů.

**1.8 Definice.** Náhodný proces  $\{X_t | t \in T\}$  se nazývá *slabě stacionární*, je-li kovariančně stacionární a současně stacionární ve střední hodnotě.

**Poznámka.** Je-li náhodný proces striktně stacionární, pak je také stacionární.

**1.9 Definice.** Nechť je náhodný proces  $\{X_t | t \in T\}$  stacionární. Položme  $\gamma(0) = \sigma^2$ . Potom funkci

$$\rho(t) = \frac{\gamma(t)}{\sigma^2} = \frac{\gamma(t)}{\gamma(0)}$$

nazýváme *autokorelační funkcí* (ACF) stacionárního náhodného procesu.

**1.10 Definice.** Nechť  $\{Y_t | t \in \mathbb{Z}\} \in L^2(\Omega, A, P)$  je stacionární proces. Pak *parciální autokorelační funkce* (PACF) je definována vztahem

$$\begin{aligned} \alpha(1) &= R(Y_t, Y_{t+1}) \\ \alpha(k) &= R(Y_t - \hat{Y}_t, Y_{t-k} - \hat{Y}_{t-k}) \quad \text{pro } |k| > 1 \end{aligned}$$

kde  $\hat{Y}_t$ , resp.  $\hat{Y}_{t-k}$  jsou nejlepší lineární predikce  $Y_t$  (resp.  $Y_{t-k}$ ) pomocí  $Y_{t-k+1}, \dots, Y_{t-1}$ .

## 1.2 Stacionární náhodné procesy

V této podkapitole definujeme *bílý šum*, *IID*, *operátor zpětného posunutí*, *lineární proces s lineárním filtrem* a jejich charakteristiky (*kovariance a spektrální hustota*), *ARMA proces* a jeho *kauzalitu* a *invertibilitu* na základě literatury [9].

**1.11 Definice.** Náhodný proces  $\{\varepsilon_t | t \in T\}$  nazýváme *bílým šumem*, pokud jsou  $\varepsilon_t$  nekorelované náhodné veličiny s nulovou střední hodnotou, tedy  $E\varepsilon_t = 0$ ,  $D\varepsilon_t = \sigma^2$ ,  $C(\varepsilon_t, \varepsilon_s) = 0$  pro  $(s \neq t)$ . Píšeme  $\varepsilon_t \sim WN(0, \sigma^2)$ . Pokud jsou navíc náhodné veličiny  $\varepsilon_t$  i nezávislé, zapisujeme je symbolem *IID* (*independent identical defined*) a píšeme  $\varepsilon_t \sim IID(0, \sigma^2)$ .

**1.12 Věta.** *Náhodné procesy  $\varepsilon_t \sim WN(0, \sigma^2)$  a  $\varepsilon_t \sim IID(0, \sigma^2)$  jsou stacionárními náhodnými procesy.*

*Důkaz.* Zřejmý. □

**1.13 Definice.** Necht  $\{Y_t | t \in \mathbb{Z}\}$  je posloupnost náhodných veličin. *Operátor zpětného posunutí* je definován pomocí výrazu

$$BY_t = Y_{t-1},$$

přičemž jej lze aplikovat několikanásobně jako

$$B^j Y_t = Y_{t-j}.$$

**1.14 Definice.** Řekneme, že náhodný proces  $\{X_t | t \in T\}$  je *spojitý podle středu v bodě*  $t_0 \in T$ , jestliže při  $t \rightarrow t_0$  konvergují  $X_t$  k  $X_{t_0}$  podle kvadratického středu, tj. když

$$E|X_t - X_{t_0}|^2 \rightarrow 0 \quad \text{pro } t \rightarrow t_0.$$

V tom případě píšeme

$$X_{t_0} = \text{l.i.m.}_{t \rightarrow t_0} X_t$$

(zkratka pochází z anglického názvu *limit in the mean*).

Je-li proces  $\{X_t | t \in T\}$  spojité v každém bodě množiny  $T$ , říkáme stručně, že je **spojitý**.

Než zavedeme pojem lineárního procesu, vyslovme větu, která zabezpečuje jeho korektnost.

**1.15 Věta.** *Necht  $\{\varepsilon_t, t \in \mathbb{Z} \sim WN(0, \sigma_\varepsilon^2)\}$  je bílým šumem, dále mějme posloupnost reálných čísel  $\{\psi_j\}_{j=0}^\infty$  takovou, že  $\sum_{j=0}^\infty \psi_j^2 < \infty$ . Pak řada  $\sum_{j=0}^\infty \psi_j \varepsilon_j$  konverguje podle kvadratického středu, tj. existuje náhodná veličina  $Y \in L^2(\Omega, A, P)$  a platí*

$$Y = \text{l.i.m.}_{N \rightarrow \infty} \sum_{j=0}^N \psi_j \varepsilon_j.$$

*Důkaz.* Viz [9] strana 138. □

**1.16 Definice.** Mějme  $\{\varepsilon_t, t \in \mathbb{Z} \sim WN(0, \sigma_\varepsilon^2)\}$  a posloupnost reálných čísel  $\{\psi_j\}_{j=0}^\infty$  takovou, že  $\sum_{j=0}^\infty \psi_j^2 < \infty$ , pak *lineární proces* je definován vztahem

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}.$$

**1.17 Definice.** Necht  $\{X_t | t \in \mathbb{Z}\}$  je stacionární náhodná posloupnost a  $\{\psi_j\}_{j=-\infty}^\infty$  je absolutně konvergentní posloupnost reálných čísel (tj.  $\sum_{j=-\infty}^\infty |\psi_j| < \infty$ ). Pak  $Y_t = \sum_{j=-\infty}^\infty \psi_j X_{t-j}$  nazveme *lineárním filtrem* procesu  $\{X_t, t \in \mathbb{Z}\}$ .

### 1.2.1 Herglotzova věta

Významnou vlastností stacionárních náhodných procesů je možnost vyjádřit autokovarianční funkci jako (nespočetný) součet harmonických funkcí s různými frekvencemi a amplitudami.

#### 1.18 Věta. (Herglotzova věta)

Je-li  $\{Y_t | t \in \mathbb{Z}\}$  stacionární náhodná posloupnost, pak se její autokovarianční funkce  $\gamma(t)$  dá vyjádřit ve tvaru

$$\gamma(t) = \int_{-\pi}^{\pi} e^{it\lambda} dF(\lambda),$$

kde  $F(\lambda)$  je neklesající, zprava spojitá funkce taková, že

$$F(-\pi) = 0 \quad a \quad F(\pi) = \gamma(0).$$

Přitom  $F(\lambda)$  je jediná.

Vzorci

$$\gamma(t) = \int_{-\pi}^{\pi} e^{it\lambda} dF(\lambda), \quad \text{resp.} \quad \gamma(t) = \int_{-\infty}^{\infty} e^{it\lambda} dF(\lambda),$$

se říká *spektrální rozklad autokovarianční funkce*. Funkce  $F(\lambda)$  se nazývá *spektrální distribuční funkce*.

Je-li  $F(\lambda)$  absolutně spojitá, pak existuje taková funkce  $f(\lambda)$ , že pro náhodné stacionární posloupnosti, resp. pro stacionární náhodné procesy, platí

$$F(\lambda) = \int_{-\pi}^{\lambda} f(x) dx, \quad \text{resp.} \quad F(\lambda) = \int_{-\infty}^{\lambda} f(x) dx. \quad (1.1)$$

Jelikož  $F(\lambda)$  je neklesající, je  $f(\lambda)$  skoro všude nezáporná. Je-li třeba, pozměníme ji na množině míry nula tak, aby byla všude nezáporná. Tím se integrál (1.1) nezmění. Funkce  $f(\lambda)$  se nazývá *spektrální hustota*. Existuje-li spektrální hustota, pak můžeme psát

$$\gamma(t) = \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda) d\lambda, \quad \text{resp.} \quad \gamma(t) = \int_{-\infty}^{\infty} e^{it\lambda} f(\lambda) d\lambda. \quad (1.2)$$

## 1.2.2 Periodogram

V dalším budeme předpokládat, že  $\{Y_t | t \in \mathbb{Z}\}$  je *centrovaná stacionární náhodná posloupnost*.

**1.19 Definice.** Necht  $Y_1, \dots, Y_n$  jsou pozorování náhodné posloupnosti  $\{Y_t | t \in \mathbb{Z}\}$ . Pak *periodogram* definujeme vztahem

$$I_n(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^n Y_t e^{-it\omega} \right|^2 \quad \omega \in [-\pi, \pi]. \quad (1.3)$$

**1.20 Lemma.** *Položme*

$$A_n(\omega) = \sqrt{\frac{2}{n}} \sum_{t=1}^n Y_t \cos t\omega \quad B_n(\omega) = \sqrt{\frac{2}{n}} \sum_{t=1}^n Y_t \sin t\omega,$$

*pak platí*

$$I_n(\omega) = \frac{1}{4\pi} [A_n^2(\omega) + B_n^2(\omega)].$$

*Důkaz.*

$$\begin{aligned} I_n(\omega) &= \frac{1}{2\pi n} \left| \sum_{t=1}^n Y_t e^{-it\omega} \right|^2 = \frac{1}{2\pi n} \left| \sum_{t=1}^n Y_t \cos t\omega - i \sum_{t=1}^n Y_t \sin t\omega \right|^2 = \\ &= \frac{1}{2\pi n} \left[ \left( \sum_{t=1}^n Y_t \cos t\omega \right)^2 + \left( \sum_{t=1}^n Y_t \sin t\omega \right)^2 \right] = \frac{1}{4\pi} [A_n^2(\omega) + B_n^2(\omega)]. \end{aligned}$$

□

**Poznámka.** Někteří autoři definují periodogram poněkud jinak:

$$I_n^*(\omega) = \frac{2}{n} \left| \sum_{t=1}^n Y_t e^{-it\omega} \right|^2 = [A_n^2(\omega) + B_n^2(\omega)] = 4\pi I_n(\omega).$$

**1.21 Lemma.** *Pokud označíme pro  $k = 0, 1, \dots, n-1$*

$$C_k = \frac{1}{n-k} \sum_{t=1}^{n-k} Y_t Y_{t+k}$$

$$C_k^* = \frac{1}{n} \sum_{t=1}^{n-k} Y_t Y_{t+k},$$

*pak platí*

$$I_n(\omega) = \frac{1}{2\pi} \left[ C_0 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) C_k \cos k\omega \right] = \frac{1}{2\pi} \left[ C_0^* + 2 \sum_{k=1}^{n-1} C_k^* \cos k\omega \right].$$

Důkaz.

$$\begin{aligned}
 I_n(\omega) &= \frac{1}{2\pi n} \left[ \left( \sum_{t=1}^n Y_t \cos t\omega \right)^2 + \left( \sum_{t=1}^n Y_t \sin t\omega \right)^2 \right] \\
 &= \frac{1}{2\pi n} \left[ \left( \sum_{t=1}^n Y_t \cos t\omega \right) \left( \sum_{s=1}^n Y_s \cos s\omega \right) + \left( \sum_{t=1}^n Y_t \sin t\omega \right) \left( \sum_{s=1}^n Y_s \sin s\omega \right) \right] \\
 &= \frac{1}{2\pi n} \sum_{t=1}^n \sum_{s=1}^n Y_t Y_s (\cos t\omega \cos s\omega + \sin t\omega \sin s\omega) \\
 &= \frac{1}{2\pi n} \sum_{t=1}^n \sum_{s=1}^n Y_t Y_s \cos \omega(s-t)
 \end{aligned}$$

Zavedeme-li dále substituci  $k = s - t$ , pak  $-n + 1 \leq k \leq n - 1$  a

$$\begin{array}{ccc}
 1 & \leq & t & \leq & n & \text{týká se kladných } k \\
 1 & \leq & s=t+k & \leq & n & \Rightarrow \max(1, \underbrace{1-k}) \leq t \leq \min(n, \overbrace{n-k}). \\
 1-k & \leq & t & \leq & n-k & \text{týká se záporných } k
 \end{array}$$

Pak platí

$$I_n(\omega) = \frac{1}{2\pi n} \sum_{k=-n+1}^{n-1} \cos k\omega \sum_{t=\max(1, 1-k)}^{\min(n, n-k)} Y_t Y_{t+k}.$$

Nyní vezměme zvlášť případy, kdy  $k = 0$  a ostatní, přičemž využijme faktu, že funkce  $\cos$  je sudou funkcí. Dostaneme proto

$$\begin{aligned}
 I_n(\omega) &= \frac{1}{2\pi} \underbrace{\frac{1}{n} \sum_{t=1}^n Y_t^2}_{C_0} + \frac{1}{2\pi} \sum_{k=-n+1}^{-1} \frac{n-|k|}{n} \cos k\omega \underbrace{\frac{1}{n-|k|} \sum_{t=1-k}^n Y_t Y_{t+k}}_{C_{-k}=C_k} \\
 &\quad + \frac{1}{2\pi} \sum_{k=1}^{n-1} \frac{n-k}{n} \cos k\omega \underbrace{\frac{1}{n-k} \sum_{t=1}^{n-k} Y_t Y_{t+k}}_{C_k} = \\
 &= \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \left( 1 - \frac{|k|}{n} \right) C_k \cos k\omega = \frac{1}{2\pi} \left[ C_0 + 2 \sum_{k=1}^{n-1} \left( 1 - \frac{k}{n} \right) C_k \cos k\omega \right]
 \end{aligned}$$

$$\begin{aligned}
 I_n(\omega) &= \frac{1}{2\pi} \underbrace{\frac{1}{n} \sum_{t=1}^n Y_t^2}_{C_0^*} + \frac{1}{2\pi} \sum_{k=-n+1}^{-1} \cos k\omega \underbrace{\frac{1}{n} \sum_{t=1-k}^n Y_t Y_{t+k}}_{C_{-k}^*=C_k^*} \\
 &\quad + \frac{1}{2\pi} \sum_{k=1}^{n-1} \cos k\omega \underbrace{\frac{1}{n} \sum_{t=1}^{n-k} Y_t Y_{t+k}}_{C_k^*} = \frac{1}{2\pi} \left( C_0^* + 2 \sum_{k=1}^{n-1} C_k^* \cos k\omega \right).
 \end{aligned}$$

□



**Poznámka.** K numerickému výpočtu hodnot periodogramu se často používají právě předchozí vzorce.

**Poznámka.** Pro teoretické účely bývá výhodnější tato varianta

$$I_n(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) C_k \cos k\omega = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} C_k^* \cos k\omega.$$

Pro náhodnou posloupnost  $\{Y_t | t \in T \subseteq \mathbb{Z}\}$  platí

$$f(\omega) = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \gamma(t) \cos t\omega.$$

Veličiny  $(1 - \frac{k}{n}) C_k$ , (resp.  $C_k^*$ ) můžeme považovat za jakýsi odhad  $\gamma(k)$  a periodogram se tudíž dá považovat za empirický odhad spektrální hustoty.

Vlastnosti tohoto odhadu udává následující věta.

**1.22 Věta.** Jestliže  $\{Y_t | t \in T \subseteq \mathbb{Z}\}$  je stacionární náhodná posloupnost s nulovou střední hodnotou a se spojitou spektrální hustotou  $f(\omega)$ , pak má periodogram  $I_n(\omega)$  následující vlastnosti:

$$\lim_{n \rightarrow \infty} EI_n(\omega) = f(\omega) \quad \omega \in [-\pi, \pi],$$

$$\lim_{n \rightarrow \infty} DI_n(\omega) = \begin{cases} f^2(\omega) & \omega \neq 0, \omega \in (-\pi, \pi), \\ 2f^2(\omega) & \omega = 0, \pm\pi. \end{cases}$$

*Důkaz.* Viz [9] strana 114 až 117. □

**Poznámka.** Z předchozí věty vyplývá

1. Periodogram  $I_n(\omega)$  je asymptoticky nestranným odhadem spektrální hustoty.
2. Periodogram  $I_n(\omega)$  není konzistentním odhadem spektrální hustoty, neboť jeho rozptyl nekonverguje k nule, vzrůstá-li neomezeně délka posloupnosti  $n$ .

Budeme-li předpokládat, že náhodný proces lze rozložit jako součet periodické deterministické funkce a bílého šumu

$$Y_t = \frac{\alpha_0}{2} + \sum_{j=1}^p (\alpha_j \cos t\omega_j + \beta_j \sin t\omega_j) + \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2),$$

pak bude mít periodogram pro velká  $n$  v bodech  $\omega = \omega_1, \dots, \omega_p$  výrazně velké hodnoty. Jinde jeho hodnoty budou relativně malé, budou kolísat kolem hodnoty  $\sigma^2/2\pi$ .

Odtud plyne, že významná lokální maxima v průběhu periodogramu by měla identifikovat periodickou strukturu uvažovaného modelu tak, že vyznačí neznámé frekvence  $\omega_1, \dots, \omega_p$ .

Vhodným statistickým testem pak rozhodneme, které hodnoty periodogramu můžeme opravdu považovat za významně velké ve srovnání s hodnotami ostatními.

### 1.2.3 Test R. A. Fishera

Ronald Aylmer Fisher (viz [8]) odvodil test, kterým se dá zjistit významnost nejvyšších hodnot periodogramu. Uvažujme posloupnost nezávislých náhodných veličin  $Y_1, \dots, Y_n$ .

Nulová hypotéza:

$$H_0 : Y_t = \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2) .$$

Předpokládejme, že  $n = 2m + 1$  je liché číslo (je-li  $n$  sudé, obvykle se vynechá první člen) a uvažujme hodnoty periodogramu  $I_n(\omega_k)$  v bodech  $\omega_k = \frac{2\pi}{n}k$  ( $k = 1, \dots, m$ ). Srovnajme je podle velikosti a označme  $V_1, \dots, V_m$ . Test se provádí na základě tzv. *Fisherovy statistiky*

$$W = \frac{V_1}{V_1 + \dots + V_m},$$

která nabývá hodnot mezi nulou a jedničkou.

Velké hodnoty (blízké 1) budou tvořit kritický obor nulové hypotézy proti alternativě

$$H_1 : Y_t = \sum_{j=1}^p \rho_j \cos(t\omega_j - \theta_j) + \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2)$$

R. A. Fisher odvodil distribuční funkci statistiky  $W$  za platnosti hypotézy  $H_0$  : (za předpokladu, že uvažujeme gaussovský bílý šum)

$$1 - F_{W|H_0}(x) = P(W > x|H_0) = m(1-x)^{m-1} - \binom{m}{2}(1-2x)^{m-1} + \dots ,$$

kde  $0 < x < 1$  a na pravé straně sčítáme tak dlouho, dokud jsou členy  $(1-kx)$  kladné, což lze také zapsat takto

$$P(W > x|H_0) = \sum_{k=1,2,\dots} \binom{m}{k} [\max(0, 1-kx)]^{m-1} = \sum_{k=1,2,\dots} \binom{m}{k} [(1-kx)_+]^{m-1} .$$

Hypotézu  $H_0 : Y_t = \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2)$  zamítáme na hladině významnosti  $\alpha$ , pokud

$$1 - F_{W|H_0}(w) = P(W > w|H_0) = \alpha_{W|H_0} \leq \alpha,$$

kde  $w$  je skutečná hodnota Fisherovy statistiky při daných hodnotách konkrétní časové řady a  $\alpha_{W|H_0}$  je tzv. *p-value*.

#### Modifikace Fisherova testu pro testování dalších periodicit

V případě, že pomocí Fisherova testu zjistíme signifikantní periodicitu určité frekvence, postupujeme při testování dalších periodicit následujícím způsobem.

Vynecháme  $V_1$ , na základě veličin  $V_2, \dots, V_m$  položíme

$$W^{(2)} = \frac{V_2}{V_2 + \dots + V_m}$$

a stanovíme  $P(W^{(2)} > w^{(2)})$  podle stejného vzorce, kde místo  $m$  dosadíme  $m - 1$ . Vyjde-li i tato druhá největší hodnota významná, opět se vynechá, a  $m$  se zmenší o další jedničku.

Když takto stanovíme frekvence  $\omega_1, \dots, \omega_p$ , získáme model se známými frekvencemi a zbylé neznámé koeficienty odhadneme metodou nejmenších čtverců (viz odstavec Lineární model pro známé frekvence v části 1.3.2).

### 1.2.4 Siegelův test

V případě více významných period není Fisherův test dostatečně silný. Vhodnější by měl být *Siegelův test*.

Zde se používá statistika  $T_\lambda$  tvaru

$$T_\lambda = \sum_{j=1}^m \left( \frac{I_n(\omega_j)}{\sum_{i=1}^m I_n(\omega_i)} - \lambda g_F \right)_+,$$

kde  $(x)_+ = \max(x, 0)$ ,  $g_F$  je  $100(1 - \alpha)\%$  kritická hodnota Fisherova testu (tj.  $P(W > g_F | H_0) = \alpha$ ) a  $0 < \lambda < 1$  je předem zvolená konstanta (doporučuje se volit  $\lambda = 0.6$ ).

Nulovou hypotézu zamítáme, pokud  $T_\lambda > t_\lambda$ , kde  $t_\lambda$  je kritická hodnota tohoto testu. Kritické hodnoty bývají tabelovány pro různá  $\lambda$  a  $m$ . Jako významné jsou uznány pouze ty periodicity, jejichž odpovídající sčítance přispěly do celkové hodnoty testovací statistiky  $T_\lambda$ .

Pokud počet dat odpovídá násobku délky periody, najde Fisherův (či Siegelův) test skutečnou periodu přesně. V případě, že však počet dat není násobkem délky periody nebo snad je dat méně než je délka periody, uvedené testy skutečnou periodu neobjeví.

### 1.2.5 Iterativní metoda

Následující iterativní metoda vychází z článku [6]. Metoda se snaží najít významné frekvence, které nejsou závislé na délce časové řady. Postupuje takto:

- (1) Pomocí Fisherova (nebo jiného) testu zjistíme významnou frekvenci  $\omega_0$ . Pokud taková neexistuje, končíme.
- (2) Minimalizujeme nelineární regresní model

$$Y_i = \mu + \alpha \cos t_i \omega + \beta \sin t_i \omega + \varepsilon_i$$

vzhledem k neznámým parametrům  $(\mu, \alpha, \beta, \omega)$ , přičemž jako *počáteční hodnoty* bereme

- $\omega_0$  frekvenci nabídnutou v kroku (1) Fisherovým či jiným testem,
- $\mu_0, \alpha_0, \beta_0$  získáme jako řešení *lineárního regresního modelu*

$$Y_i = \mu_0 + \alpha_0 \cos t_i \omega_0 + \beta_0 \sin t_i \omega_0 + \zeta_i.$$

Získané hodnoty označme  $\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\omega}$ .

(3) Z řady odstraníme vliv  $\hat{\omega}$ , tj.

$$Z_i = Y_t - \hat{\mu} - \hat{\alpha} \cos t_i \hat{\omega} - \hat{\beta} \sin t_i \hat{\omega} + \eta_i$$

a s řadou  $Z_i$  jdeme na bod (1).

Algoritmus při odhadu frekvencí vychází z významných frekvencí, které určil Fisherův či jiný test. Současně s odhadem dalších parametrů odhadnuté frekvence ještě zpřesňuje. Jeho nevýhodou je, že odhaluje frekvence postupně, tj. v každém kroku hledá pouze jedinou významnou frekvenci.

### 1.2.6 Neparametrické odhady spektrální hustoty

Neparametrické odhady spektrální hustoty (*Window Spectral Estimation*) *centrované stacionární náhodné posloupnosti*  $\{Y_t | t \in \mathbb{Z}\}$  jsou založeny na zlepšení vlastností periodogramu.

Periodogram je empirickým odhadem spektrální hustoty, který je asymptoticky nestranný, avšak nekonzistentní. Připomeňme, že platí

$$I_n(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^n Y_t e^{-it\omega} \right|^2 = \frac{1}{2\pi} \left[ C_0^* + 2 \sum_{k=1}^{n-1} C_k^* \cos k\omega \right].$$

Využijme dále vztahů

$$C_k^* = C_{-k}^*, \quad \text{kde} \quad C_k^* = \frac{1}{n} \sum_{t=1}^{n-k} Y_t Y_{t+k} \quad \text{pro} \quad k = 0, \pm 1, \pm 2, \dots, \pm(n-1)$$

a

$$\cos k\omega = \frac{1}{2} (e^{ik\omega} + e^{-ik\omega}).$$

Upravujme postupně

$$\begin{aligned} I_n(\omega) &= \frac{1}{2\pi} \left[ C_0^* + \sum_{k=1}^{n-1} C_k^* e^{ik\omega} + \sum_{k=1}^{n-1} C_k^* e^{-ik\omega} \right] \\ &= \frac{1}{2\pi} \left[ C_0^* + \sum_{s=-(n-1)}^{-1} C_{-s}^* e^{-is\omega} + \sum_{k=1}^{n-1} C_k^* e^{-ik\omega} \right] \\ &= \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} C_k^* e^{-ik\omega}. \end{aligned}$$

Periodogram (jakožto odhad spektrální hustoty) je založen na všech možných odhadech autokovarianční funkce v bodech  $k=0, \pm 1, \pm 2, \dots, \pm(n-1)$ , tj.

$$\begin{aligned}
 C_0^* &= \frac{1}{n} \underbrace{(Y_1^2 + \dots + Y_n^2)}_{n \text{ členů}} \\
 C_1^* = C_{-1}^* &= \frac{1}{n} \underbrace{(Y_1 Y_2 + \dots + Y_{n-1} Y_n + Y_3 Y_n)}_{n-1 \text{ členů}} \\
 &\vdots \\
 C_{n-3}^* = C_{-(n-3)}^* &= \frac{1}{n} \underbrace{(Y_1 Y_{n-2} + Y_2 Y_{n-1} + Y_3 Y_n)}_{3 \text{ členy}} \\
 C_{n-2}^* = C_{-(n-2)}^* &= \frac{1}{n} \underbrace{(Y_1 Y_{n-1} + Y_2 Y_n)}_{2 \text{ členy}} \\
 C_{n-1}^* = C_{-(n-1)}^* &= \frac{1}{n} \underbrace{Y_1 Y_n}_{1 \text{ člen}}
 \end{aligned}$$

a tedy je založen i na velmi málo kvalitních odhadech. K určitému zlepšení jistě dojde, pokud budeme používat jen  $m \ll n$  nejkvalitnějších odhadů. Mluvíme pak o *prostém useknutém periodogramu*

$$\hat{f}_n(\omega) = \frac{1}{2\pi} \sum_{k=-m}^m C_k^* \cos k\omega = \frac{1}{2\pi} \sum_{k=-m}^m C_k^* e^{-ik\omega},$$

což lze také zapsat takto

$$\hat{f}_n(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} w(k) C_k^* \cos k\omega = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} w(k) C_k^* e^{-ik\omega},$$

kde

$$w(k) = \begin{cases} 1 & |k| \leq m \\ 0 & |k| > m \end{cases}.$$

Označme *Fourierovu transformaci* funkce  $w(k)$

$$W(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} w(k) e^{-ik\omega} = \frac{1}{2\pi} \sum_{k=-m}^m e^{-ik\omega}$$

a řadu přeindexujeme tak, aby indexy šly od 1 do  $2m+1$ , tj. položíme  $s = k + m + 1$ , pak  $k = s - m - 1$  a

$$\begin{aligned}
 \text{(a) pro } \omega \neq 2k\pi \text{ je } W(\omega) &= \frac{1}{2\pi} \sum_{s=1}^{2m+1} e^{-i(s-m-1)\omega} \\
 &= \frac{1}{2\pi} e^{i(m+1)\omega} \sum_{s=1}^{2m+1} e^{-is\omega} \\
 &= \frac{1}{2\pi} e^{im\omega} \frac{1 - e^{-i(2m+1)\omega}}{1 - e^{-i\omega}} \\
 &= \frac{1}{2\pi} e^{im\omega} \frac{e^{-i\frac{2m+1}{2}\omega} \left( e^{i\frac{2m+1}{2}\omega} - e^{-i\frac{2m+1}{2}\omega} \right)}{e^{-i\frac{1}{2}\omega} \left( e^{i\frac{1}{2}\omega} - e^{-i\frac{1}{2}\omega} \right)} \\
 &= \frac{1}{2\pi} \frac{\sin(m+\frac{1}{2})\omega}{\sin\frac{1}{2}\omega} = D_m(\omega),
 \end{aligned}$$

kde  $D_m(\omega)$  je tzv. *Dirichletovo jádro*,

(b) pro  $\omega = 2k\pi$  je  $W(\omega) = 2m + 1$ .

Vzhledem k tomu, že lze psát

$$I_n(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} C_k^* e^{-ik\omega},$$

vidíme, že  $I_n(\omega)$  je Fourierovou transformací  $C_k^*$ , takže naopak lze pomocí *inverzní Fourierovy transformace* psát

$$C_k^* = \int_{-\pi}^{\pi} I_n(\theta) e^{ik\theta} d\theta.$$

Počítejme postupně

$$\begin{aligned} \hat{f}_n(\omega) &= \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} w(k) C_k^* e^{-ik\omega} \\ &= \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} w(k) \int_{-\pi}^{\pi} I_n(\theta) e^{ik\theta} d\theta e^{-ik\omega} \\ &= \int_{-\pi}^{\pi} I_n(\theta) \underbrace{\frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} w(k) e^{-ik(\omega-\theta)} d\theta}_{W(\omega-\theta)} \\ &= \int_{-\pi}^{\pi} I_n(\theta) W(\omega - \theta) d\theta. \end{aligned}$$

Jde o tzv. *vyhlazený periodogram* (*smoothed periodogram*).

Funkce  $W(\omega)$  se nazývá *spektrální okénko* (*spectral window*). Tato funkce má do jisté míry aproximovat Diracovu  $\delta$  funkci a platí pro ni

$$\int_{-\pi}^{\pi} W(\omega) d\omega = 1.$$

Takto počítat odhad spektrální hustoty by však bylo (vzhledem k málo hladkému průběhu periodogramu) nepohodlné, proto se obvykle odhad počítá podle vzorce

$$\hat{f}_n(\omega) = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} w(k) C_k^* e^{-ik\omega},$$

přičemž inverzní Fourierova transformace

$$w(k) = \int_{-\pi}^{\pi} W(\theta) e^{ik\theta} d\theta, \quad k = 0, \pm 1, \pm 2, \dots, \pm(n-1)$$

se nazývá *korelační okénko* (*covariance lag window* nebo *time-domain window*).

Typickými korelačními okénky jsou tzv. *useknutá okénka*, pro která existuje takové přirozené číslo  $m$  (bod useknutí, *truncation point*) tak, že  $w(k) = 0$  pro  $|k| > m$  ( $m$  se obvykle volí v rozmezí od  $\frac{n}{6}$  do  $\frac{n}{5}$ ).

## PŘÍKLADY KORELAČNÍCH A SPEKTRÁLNÍCH OKÉNEK

### Prostý useknutý odhad

$$w(k) = \begin{cases} 1 & 0 < |k| \leq m \\ 0 & |k| > m \end{cases}$$

$$W(\omega) = \frac{1}{2\pi} \frac{\sin(m + \frac{1}{2})\omega}{\sin \frac{1}{2}\omega}$$

### Bartletovo okénko

$$w(k) = \begin{cases} \left(1 - \frac{|k|}{m}\right) & 0 < |k| \leq m \\ 0 & |k| > m \end{cases}$$

$$W(\omega) = \frac{1}{2\pi m} \frac{\sin^2 m \frac{\omega}{2}}{\sin^2 \frac{\omega}{2}} = F_m(\omega)$$

$W(\omega)$  je v tomto případě Fejérovým jádrem.

### Parzenovo okénko

$$w(k) = \begin{cases} 1 - 6 \left(\frac{k}{m}\right)^2 + 6 \left(\frac{|k|}{m}\right)^3 & |k| < \frac{m}{2} \\ 2 \left(1 - \frac{|k|}{m}\right)^3 & \frac{m}{2} < |k| \leq m \\ 0 & |k| > m \end{cases}$$

$$W(\omega) = \frac{3}{8\pi m^3} \left(\frac{\sin m \frac{\omega}{4}}{\frac{1}{2} \sin \frac{\omega}{2}}\right)^4 \left(1 - \frac{2}{3} \sin^2 \frac{\omega}{2}\right)$$

kde  $m$  je nějaké sudé číslo.

### Obecné Tukeyovo okénko

$$w(k) = \begin{cases} 1 - 2a + 2a \cos \frac{\pi k}{m} & |k| \leq m \\ 0 & |k| > m \end{cases}$$

$$W(\omega) = aD_m\left(\omega - \frac{\pi}{m}\right) + (1 - 2a)D_m(\omega) + aD_m\left(\omega + \frac{\pi}{m}\right)$$

kde  $a \in \left(0, \frac{1}{4}\right]$ .

Pokud  $a = \frac{1}{4}$ , pak se nazývá *Tukey-Hanningovo okénko*.

### Tukey-Hammingovo okénko

$$w(k) = \begin{cases} 0.54 + 0.46 \cos \frac{\pi k}{m} & |k| \leq m \\ 0 & |k| > m \end{cases}$$

$$W(\omega) = 0.23D_m\left(\omega - \frac{\pi}{m}\right) + 0.54D_m(\omega) + 0.23D_m\left(\omega + \frac{\pi}{m}\right)$$

**Daniellovo okénko** Na závěr ještě uvedeme jedno *neusekнутé korelační okénko*.

Mějme pro  $\delta \in (0, \pi)$  následující *spektrální okénko*

$$W(\omega) = \begin{cases} \frac{1}{2\delta} & |\omega| < \delta \\ 0 & |\omega| > \delta \end{cases},$$

které je vlastně hustotou náhodné veličiny s rovnoměrně spojitým rozdělením na intervalu  $(-\delta, \delta)$ . Pro  $k = \pm 1, \pm 2, \dots, \pm(n-1)$  počítejme nejprve odpovídající korelační okénko:

$$\begin{aligned} w(k) &= \int_{-\pi}^{\pi} W(\omega) e^{ik\omega} d\omega = \int_{-\delta}^{\delta} \frac{1}{2\delta} e^{ik\omega} d\omega \\ &= \frac{1}{2\delta} \left[ \frac{e^{ik\omega}}{ik} \right]_{-\delta}^{\delta} = \frac{1}{k\delta} \underbrace{\frac{1}{2i} (e^{ik\delta} - e^{-ik\delta})}_{\sin k\delta} = \frac{\sin k\delta}{k\delta}. \end{aligned}$$

Pro  $k = 0$  je zřejmě rovno jedné, celkově tedy

$$w_k = \begin{cases} 1 & k = 0 \\ \frac{\sin k\delta}{k\delta} & k = \pm 1, \pm 2, \dots \end{cases}.$$

## 1.2.7 Definice ARMA

**1.23 Definice.** *ARMA*<sup>1</sup> proces řádu  $p, q$  je definován vztahem

$$Y_t - \varphi_1 Y_{t-1} - \dots - \varphi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \text{ kde } \varepsilon_t \sim WN(0, \sigma_\varepsilon^2),$$

přičemž pomocí operátoru zpětného chodu lze psát

$$Y_t \sim ARMA(p, q) : \Phi(B)Y_t = \Theta(B)\varepsilon_t,$$

kde

$$\Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p \quad (\varphi_0 \equiv 1)$$

a

$$\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (\theta_0 \equiv 1).$$

Řekneme, že  $\{Y_t, t \in \mathbb{Z}\}$  je *ARMA*( $p, q$ ) se střední hodnotou  $\mu$ , jestliže  $\{Y_t - \mu\}$  je *ARMA*( $p, q$ ) proces.

Speciální případy *ARMA* procesů nazýváme:

*Autoregresní proces* (AR proces):  $Y_t \sim AR(p) \sim ARMA(p, 0)$ , tj.  $q = 0$

*Proces klouzavých součtů* (MA proces):  $Y_t \sim MA(q) \sim ARMA(0, q)$ , tj.  $p = 0$

---

<sup>1</sup>autoregressive moving average



**1.24 Definice.** ARMA proces  $Y_t \sim \text{ARMA}(p, q)$  se nazývá *kauzální*, jestliže existuje absolutně konvergentní posloupnost reálných čísel  $\Psi = \{\psi_j\}_{j=0}^{\infty}$ , (tj.  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ ) tak, že

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \text{ tj. zkráceně } Y_t \sim MA(\infty) : Y_t = \Psi(B)\varepsilon_t.$$

**1.25 Definice.** ARMA proces  $Y_t \sim \text{ARMA}(p, q)$  se nazývá *invertibilní*, jestliže existuje absolutně konvergentní posloupnost reálných čísel  $\pi = \{\pi_j\}_{j=0}^{\infty}$  (tj.  $\sum_{j=0}^{\infty} |\pi_j| < \infty$ ) tak, že

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}, \text{ tj. zkráceně } Y_t \sim AR(\infty) : \varepsilon_t = \pi(B)Y_t.$$

**Střední hodnota, rozptyl, autokovarianční a autokorelační funkce procesů ARMA(p, q)**

**Poznámka.** Je-li ARMA(p, q) proces kauzální, můžeme počítat *střední hodnotu*

$$EY_t = E \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \sum_{j=0}^{\infty} \psi_j E\varepsilon_{t-j} = 0.$$

**Poznámka.** Při odvození rozptylu pro kauzální ARMA(p, q) proces nejprve rovnici

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

vynásobme výrazem  $Y_t$  a spočtíme střední hodnoty obou stran, tj.

$$EY_t^2 = \varphi_1 EY_{t-1}Y_t + \dots + \varphi_p EY_{t-p}Y_t + E\varepsilon_t Y_t + \theta_1 E\varepsilon_{t-1}Y_t + \dots + \theta_q E\varepsilon_{t-q}Y_t. \quad (1.4)$$

Spočtíme pro  $i = 0, 1, \dots, q$

$$E\varepsilon_{t-i}Y_t = E\varepsilon_{t-i} \left( \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \right) = \sum_{j=0}^{\infty} \psi_j E\varepsilon_{t-i}\varepsilon_{t-j} = \psi_i \sigma_{\varepsilon}^2 \quad (\text{přičemž } \psi_0 = 1).$$

Po dosazení do rovnice (1.4) dostaneme

$$\gamma(0) - \varphi_1 \gamma(1) - \dots - \varphi_p \gamma(p) = \sigma_{\varepsilon}^2 (1 + \theta_1 \psi_1 + \dots + \theta_q \psi_q). \quad (1.5)$$

Podělme obě strany rovnice (1.5) výrazem  $\gamma(0)$ . Vzhledem k tomu, že  $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$ , dostaneme

$$1 - \varphi_1 \rho(1) - \dots - \varphi_p \rho(p) = \frac{\sigma_{\varepsilon}^2 (1 + \theta_1 \psi_1 + \dots + \theta_q \psi_q)}{\gamma(0)},$$

takže

$$DY_t = \gamma(0) = \frac{\sigma_{\varepsilon}^2 (1 + \theta_1 \psi_1 + \dots + \theta_q \psi_q)}{1 - \varphi_1 \rho(1) - \dots - \varphi_p \rho(p)}.$$

**Poznámka.** Při výpočtu autokovariance pro kauzální ARMA( $p, q$ ) proces rovnici

$$Y_t - \varphi_1 Y_{t-1} - \dots - \varphi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

vynásobíme výrazem  $Y_{t-k}$  a spočítáme střední hodnoty obou stran, takže dostaneme

$$\gamma(k) - \varphi_1 \gamma(k-1) - \dots - \varphi_p \gamma(k-p) = EY_{t-k} \varepsilon_t + \theta_1 EY_{t-k} \varepsilon_{t-1} + \dots + \theta_q EY_{t-k} \varepsilon_{t-q}. \quad (1.6)$$

Nejprve je třeba si uvědomit, že pro  $s \geq 0$  platí

$$E\varepsilon_t Y_{t-s} = E\left(\varepsilon_t \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-s-j}\right) = \sum_{j=0}^{\infty} \psi_j E\varepsilon_t \varepsilon_{t-s-j} = 0.$$

Spočtíme pro  $i = 0, 1, \dots, q$

$$\begin{aligned} E\varepsilon_{t-i} Y_{t-k} &= E\varepsilon_{t-i} \left( \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j-k} \right) = \sum_{j=0}^{\infty} \psi_j E\varepsilon_{t-i} \varepsilon_{t-j-k} = \\ &= \begin{cases} \sigma_\varepsilon^2 \psi_{i-k} & k \leq i \text{ přičemž } \psi_0 = 1 \Rightarrow k \leq q \\ 0 & k > i \text{ neboť } \psi_j = 0 \text{ pro } j < 0. \end{cases} \end{aligned}$$

Uvážíme-li, že  $\psi_j = 0$  pro  $j < 0$ , potom pro  $0 \leq k \leq \max(p, q+1)$  platí

$$\gamma(k) - \varphi_1 \gamma(k-1) - \dots - \varphi_p \gamma(k-p) = \sigma_\varepsilon^2 (\theta_k + \theta_{k+1} \psi_1 + \dots + \theta_q \psi_{q-k}) \quad (1.7)$$

a pro  $k > \max(p, q+1)$  platí

$$\gamma(k) - \varphi_1 \gamma(k-1) - \dots - \varphi_p \gamma(k-p) = 0.$$

Podělme obě strany rovnice (1.7) výrazem  $\gamma(0)$ . Dostaneme

$$\rho(k) - \varphi_1 \rho(k-1) - \dots - \varphi_p \rho(k-p) = \frac{\sigma_\varepsilon^2 (\theta_k + \theta_{k+1} \psi_1 + \dots + \theta_q \psi_{q-k})}{\gamma(0)},$$

resp.

$$\rho(k) - \varphi_1 \rho(k-1) - \dots - \varphi_p \rho(k-p) = 0.$$

Nechť např.

$$q+1 > p.$$

Pak máme více rovnic pro určení počátečních  $p$  podmínek. V tomto případě prvních  $q-p+1$  autokovariančních koeficientů je určeno z prvních  $q-p+1$  podmínek. Obecné řešení homogenní diferenční rovnice

$$\rho(k) - \varphi_1 \rho(k-1) - \dots - \varphi_p \rho(k-p) = 0 \quad \text{tj.} \quad \Phi(B)\rho(k) = 0$$

je tvaru

$$\rho(k) = \sum_{j=1}^m \left( \sum_{s=0}^{p_j-1} c_{js} k^s \right) \lambda_j^k,$$

kde

$$\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_m}$$

jsou různé kořeny s násobnostmi

$$p_1, \dots, p_m,$$

přičemž

$$p = p_1 + \dots + p_m$$

a  $c_{js}$  je právě  $p$  konstant, které jsou určeny počátečními podmínkami.

Shrneme-li tedy výše uvedené, za předpokladu, že  $\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_m}$  jsou různé kořeny s násobnostmi  $p_1, \dots, p_m$ , přičemž  $p = p_1 + \dots + p_m$ , pak řešení homogenní diferenční rovnice  $\Phi(B)\rho(k) = 0$  je tvaru

$$\rho(k) = \sum_{j=1}^m \left( \sum_{s=0}^{p_j-1} c_{js} k^s \right) \lambda_j^k,$$

kde  $c_{js}$  jsou konstanty, které jsou určeny počátečními podmínkami. Dále položíme  $\lambda_j = r_j e^{i\theta_j}$ . Pak máme

$$\rho(k) = \sum_{j=1}^m \left( \sum_{s=0}^{p_j-1} c_{js} k^s \right) r_j^k e^{ik\theta_j}.$$

Vzhledem k tomu, že platí

$$|\lambda_j| = r_j < 1,$$

dostáváme odtud, že  $\rho(k)$  klesá pro  $k \rightarrow \infty$  exponenciálně k nule, tj.

$$\rho(k) \xrightarrow[k \rightarrow \infty]{} 0,$$

což je velmi důležitá identifikační vlastnost autoregresních  $AR(p)$  procesů.

### 1.26 Věta. (Spektrální hustota ARMA(p, q) procesů)

Nechť  $\Phi(B)Y_t = \Theta(B)\varepsilon_t$  je kauzální a invertibilní ARMA(p, q) proces, přičemž  $\Phi(z)$  a  $\Theta(z)$  nemají společné kořeny. Pak spektrální hustota ARMA(p, q) procesu je rovna

$$f_Y(\omega) = \frac{\sigma_\varepsilon^2 |\Theta(e^{-i\omega})|^2}{2\pi |\Phi(e^{-i\omega})|^2} \quad \text{pro } \omega \in [-\pi, \pi].$$

Důkaz. Viz [9] strana 166. □

## 1.3 Modelování nestacionarity

### 1.3.1 ARIMA procesy: diferenčně stacionární náhodné procesy

Zavedme nejprve tzv. *diferenční operátor*:

$$\begin{aligned}\Delta Y_t &= Y_t - Y_{t-1} = (1 - B)Y_t \\ \Delta^2 Y_t &= \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) \\ &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &= Y_t - 2Y_{t-1} + Y_{t-2} = (1 - B)^2 Y_t \\ &\vdots \\ \Delta^d Y_t &= (1 - B)^d Y_t\end{aligned}$$

Nestacionární proces se stochastickým trendem nazýváme *integrovaným smíšeným modelem*

$$ARIMA(p, d, q).$$

Formálně jej zapišeme pomocí operátoru zpětného chodu takto:

$$ARIMA(p, d, q) : \Phi(B)(1 - B)^d Y_t = \Theta(B)\varepsilon_t$$

a položíme-li

$$W_t = (1 - B)^d Y_t,$$

pak  $W_t$  je *stacionární* ARMA(p, q).

#### Zvláštní případy ARIMA(p, d, q)

$p$	$d$	$q$	Zkratka	Název
0			IMA(d, q)	Integrovaný proces klouzavých součtů
0	0		MA(q)	Proces klouzavých součtů
		0	ARI(p, d)	Integrovaný autoregresní proces
	0	0	AR(p)	Autoregresní proces
0		0	I(d)	Integrovaný proces
0	1	0	I(1)	Náhodná procházka

NÁHODNÁ PROCHÁZKA:  $Y_t = Y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim VW(0, \sigma_\varepsilon^2)$

Proces náhodné procházky je limitním případem procesu  $AR(1)$ , kde  $\varphi_1 = 1$ .

Pro náhodnou procházku platí:

- Hodnoty  $ACF = \rho(k)$  budou klesat velmi pomalu (lineárně).
- Hodnoty  $PACF = \alpha(k)$  jsou logicky velmi podobné procesu  $AR(1)$ .

V praxi se používá obecnější tvar:

$$Y_t = \beta + Y_{t-1} + \varepsilon_t \quad \beta \in \mathbb{R}.$$

Potom, pokud budeme postupně upravovat, dostaneme

$$Y_t = \beta + Y_{t-1} + \varepsilon_t = Y_{t-2} + 2\beta + \varepsilon_t + \varepsilon_{t-1} = \dots = \underbrace{Y_0 + \beta \cdot t}_{\text{deterministický lineární trend}} + \sum_{j=1}^t \varepsilon_j.$$

Operátor

$$\nu(B) = \Phi(B)(1 - B)^d$$

se někdy nazývá *zobecněný autoregresní operátor*.

Pokud  $\nu(B)$  chápeme jako polynom v proměnné  $B$ , pak vzhledem ke kauzalitě modelu

$$(1 - B)^d W_t = \Theta(B)\varepsilon_t$$

má  $\nu(B)$  právě  $p$  kořenů ležících vně jednotkového kruhu a  $d$  kořenů rovných 1.

V praxi se nejprve diferencováním časové řady získá stacionární řada  $W_t$  a pro ni se vybuduje proces  $ARMA(p, q)$ . Pokud jsme původně měli

$$Y_1, \dots, Y_n,$$

po diferencování zůstanou

$$W_{d+1}, \dots, W_n.$$

## 1.3.2 Regresní modely

**1.27 Definice.** Řekneme, že náhodný proces  $\{Y_t | t \in T\}$  splňuje *lineární regresní model*, pokud pro jeho střední hodnotu platí

$$\forall t \in T : \quad EY_t = \mu_t = f(t) = \sum_{j=0}^p \beta_j g_j(t),$$

kde  $g_0, \dots, g_p$  jsou známé funkce definované na  $T$ ,  $\beta = (\beta_0, \dots, \beta_m)'$  je vektor neznámých **regresních koeficientů**.

Z velkého okruhu *trendových funkcí*, které vedou k lineárnímu regresnímu modelu, se zaměříme na

- *polynomický trend*:  $f(t) = \beta_0 + \beta_1 t + \dots + \beta_p t^p$
- *periodický trend*:  $f(t) = \mu + \sum_{j=1}^p (\alpha_j \cos \lambda_j t + \beta_j \sin \lambda_j t)$

Neznámé parametry modelu budeme hledat pomocí metody nejmenších čtverců.

## Polynomický trend

V případě *polynomického trendu*, matice plánu je tvaru

$$\mathbf{X} = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^p \end{pmatrix}.$$

## Periodický trend

Pokud střední hodnota náhodného procesu  $EY_t = \mu_t = f(t)$  je periodickou s periodou  $T_0$ , pak *frekvencí* rozumíme veličinu  $\lambda = \frac{2\pi}{T_0}$ .

Uvažujme model:

$$Y_i = f(t_i) + \varepsilon_i \quad E\varepsilon_i = 0; \quad D\varepsilon_i = \sigma^2; \quad C(\varepsilon_i, \varepsilon_j) = 0; \quad i \neq j; \quad i, j = 1, \dots, n$$

kde (a)  $f(t_i) = \mu + \sum_{j=1}^p (\alpha_j \cos \lambda_j t_i + \beta_j \sin \lambda_j t_i)$

nebo (b)  $f(t_i) = \mu + \sum_{j=1}^p \gamma_j \cos(\lambda_j t_i + \omega_j) \quad \gamma_j = \sqrt{\alpha_j^2 + \beta_j^2}, \quad \omega_j = \arctan \frac{\beta_j}{\alpha_j}.$

Jde o *nelineární regresní model* vzhledem k  $(3p + 1)$  neznámých parametrů:

(a)  $\alpha_1, \dots, \alpha_p \quad \beta_1, \dots, \beta_p \quad \mu \quad \lambda_1, \dots, \lambda_p$

(b)  $\gamma_1, \dots, \gamma_p \quad \mu \quad \lambda_1, \dots, \lambda_p \quad \omega_1, \dots, \omega_p$

Odhad vektoru *neznámých parametrů* pomocí *metody nejmenších čtverců* minimalizuje výraz

(a)  $S(\mu, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p, \lambda_1, \dots, \lambda_p) = \sum_{i=1}^n (Y_i - f(t_i))^2$

(b)  $S(\mu, \gamma_1, \dots, \gamma_p, \omega_1, \dots, \omega_p, \lambda_1, \dots, \lambda_p) = \sum_{i=1}^n (Y_i - f(t_i))^2$

Numericky lze systém nelineárních rovnic řešit např. pomocí Gauss-Newtonovy metody.

V prostředí R lze v základním statistickém balíčku `stats` najít funkci `nls()`, kterou je možno použít k odhadu parametrů nelineárních regresních modelů jako je tento.

## Lineární model pro známé frekvence

Situace se zjednoduší, pokud frekvence  $\lambda_1, \dots, \lambda_p$  jsou známé.

Pak model (a) je lineární a matice plánu je tvaru:

$$\mathbf{X}_{n \times (2p+1)} = \begin{pmatrix} 1 & c_{11} & s_{12} & \cdots & c_{p1} & s_{p1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & c_{1n} & s_{1n} & \cdots & c_{pn} & s_{pn} \end{pmatrix}, \quad \text{kde} \quad \begin{matrix} c_{ji} = \cos \lambda_j t_i \\ s_{ji} = \sin \lambda_j t_i \end{matrix} \quad \text{pro} \quad \begin{matrix} j = 1, \dots, p \\ i = 1, \dots, n \end{matrix}$$

Za předpokladu, že

$$\begin{aligned} n &= 2m + 1 \\ t_i &= i \\ \lambda_j &= \frac{2\pi j}{n} \text{ pro některá } j \in \{1, \dots, m\} \end{aligned}$$

matice plánu (viz [9]) je tvaru

$$\mathbf{X}'\mathbf{X}_{(2p+1) \times (2p+1)} = \begin{pmatrix} n & 0 & \cdots & 0 \\ 0 & \frac{n}{2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & \frac{n}{2} \end{pmatrix}$$

a odtud velmi snadno z normálních rovnic dostaneme odhady neznámých parametrů ve tvaru

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{t=1}^n Y_t \\ \hat{\alpha}_j &= \frac{2}{n} \sum_{t=1}^n Y_t \cos \lambda_j t & j = 1, \dots, p. \\ \hat{\beta}_j &= \frac{2}{n} \sum_{t=1}^n Y_t \sin \lambda_j t \end{aligned}$$

Neznámé parametry modelu (b) získáme ze vztahů

$$\begin{aligned} \hat{\gamma}_j &= \sqrt{\hat{\alpha}_j^2 + \hat{\beta}_j^2} \\ \hat{\omega}_j &= \arctan \frac{\hat{\beta}_j}{\hat{\alpha}_j} \end{aligned} \quad j = 1, \dots, p.$$

# Kapitola 2

## Procesy s dlouhou pamětí

V této kapitole nejprve definujeme nejjednodušší model *dlouhé paměti*. Dále v části 2.2 uvedeme *frakcionálně diferencovaný bílý šum* a následně v části 2.3 základní metody odhadů jeho parametrů na základě článku [1]. Na závěr kapitoly v části 2.4 zobecníme *frakcionálně diferencovaný bílý šum* na *autoregresní frakcionálně integrované procesy klouzavých součtů*. Budeme vycházet z knihy [19] a z druhé části článku [5].

### 2.1 Definování dlouhé paměti pomocí autokovariancí

**2.1 Definice.** Nechť  $\gamma(h)$  je autokovarianční funkce pro případ, kdy jde o stacionární proces  $\{Y_t | t \in \mathbb{Z}\}$ . Řekneme, že *stacionární proces má dlouhou paměť*, jestliže platí

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| = \infty.$$

Analogicky řekneme, že *stacionární proces má krátkou paměť*, jestliže platí

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty.$$

### 2.2 Frakcionálně diferencovaný bílý šum

Nechť  $\{\varepsilon_t | t \in \mathbb{Z}\}$  je bílý šum se střední hodnotou  $E(\varepsilon_t) = 0$  a s rozptylem  $D(\varepsilon_t) = \sigma^2 > 0$ . Nechť  $B$  je operátor zpětného chodu, tj.  $BX_t = X_{t-1}$ ,  $B\varepsilon_t = \varepsilon_{t-1}$  atd. Pokud  $\{X_t | t \in \mathbb{Z}\}$  je lineární proces splňující podmínku

$$(1 - B)^d X_t = \varepsilon_t, \quad \text{kde } d \in \left(-\frac{1}{2}, \frac{1}{2}\right), \quad (2.1)$$

pak se  $\{X_t | t \in \mathbb{Z}\}$  nazývá (*jednoduchý*) *frakcionálně diferencovaný bílý šum*. V dalším textu budeme užívat zkratku FDWN z anglického názvu *fractionally differenced white noise*. Namísto (2.1) můžeme ekvivalentně použít zápis

$$X_t = (1 - B)^{-d} \varepsilon_t. \quad (2.2)$$



Proces  $\{X_t | t \in \mathbb{Z}\}$  má spektrální hustotu tvaru (viz článek [1]):

$$f(\lambda) = \frac{\sigma^2}{2\pi} |1 - e^{-i\lambda}|^{-2d} = \frac{\sigma^2}{2\pi} \left(4 \sin^2 \frac{\lambda}{2}\right)^{-d}. \quad (2.3)$$

Výraz (2.3) lze podrobněji derivovat pomocí metod vysvětlených v kapitole 9.1 v [2]. Zřejmě  $f(\lambda) \rightarrow \infty$  pro  $\lambda \rightarrow 0$ , pokud  $d > 0$ . Jelikož nás zajímají především modely tohoto druhu, budeme v dalším textu předpokládat  $d \in (0, 1/2)$ .

Dále budeme autokovarianční (popř. autokorelační) funkci značit také takto  $\gamma_k = \gamma(k)$  a  $\rho_k = \rho(k)$ .

**2.2 Věta.** Autokovarianční funkce  $\gamma_k$  a autokorelační funkce  $\rho_k$  procesu  $\{X_t | t \in \mathbb{Z}\}$  se spektrální hustotou (2.3) jsou dány vzorci

$$\gamma_k = \frac{(-1)^k \sigma^2 \Gamma(1 - 2d)}{\Gamma(k + 1 - d) \Gamma(-k + 1 - d)} \quad \text{pro } k = 0, \pm 1, \dots,$$

a

$$\rho_k = \frac{d(1+d) \dots (k-1+d)}{(1-d)(2-d) \dots (k-d)} \quad \text{pro } k = 1, 2, \dots, \quad (2.4)$$

kde  $\Gamma(x)$  je gamma funkce definovaná pro  $x \in \mathbb{R}$ .

*Důkaz.* Viz [15]. □

Předchozí vztahy mohou být také zapsány ve tvaru

$$\begin{aligned} \gamma_0 &= \sigma^2 \frac{\Gamma(1 - 2d)}{\Gamma^2(1 - d)}, \\ \gamma_{k+1} &= \frac{k + d}{k + 1 - d} \gamma_k \quad \text{pro } k \geq 0, \end{aligned} \quad (2.5)$$

$$\rho_0 = 1, \quad \rho_1 = \frac{d}{1-d}, \quad \rho_{k+1} = \frac{k+d}{k+1-d} \rho_k \quad \text{pro } k \geq 0. \quad (2.6)$$

Výše uvedené vzorce (2.5) a (2.6) se užívají k odhadům parametru  $d$  a k simulacím. Vzorec (2.4) je ekvivalentní s vyjádřením

$$\rho_k = \frac{\Gamma(1-d)}{\Gamma(d)} \frac{\Gamma(k+d)}{\Gamma(k+1-d)}.$$

Rozvojem  $\Gamma(k+d)$  a  $\Gamma(k+1-d)$  pomocí Stirlingova vzorce po pár krocích dostaneme

$$\rho_k \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \quad \text{pro } k \rightarrow \infty.$$

Z předchozího je zřejmé, že  $\sum |\rho_k| = \infty$ . Což znamená, že FDWN náleží mezi modely s dlouhou pamětí.

**2.3 Věta.** *Nechť  $d \in (0, 1/2)$  a  $|z| < 1$ . Pak*

$$(1 - z)^d = \sum_{j=0}^{\infty} a_j z^j, \quad (1 - z)^{-d} = \sum_{j=0}^{\infty} b_j z^j, \quad (2.7)$$

kde

$$a_j = \frac{\Gamma(j - d)}{\Gamma(-d)\Gamma(j + 1)} = \frac{(j - 1 - d)(j - 2 - d) \dots (1 - d)(-d)}{j!}, \quad (2.8)$$

$$b_j = \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)} = \frac{(j - 1 + d)(j - 2 + d) \dots (1 + d)d}{j!}. \quad (2.9)$$

Když  $j \rightarrow \infty$ , pak

$$j^{1+d} a_j \rightarrow \frac{1}{\Gamma(-d)}, \quad j^{j-d} b_j \rightarrow \frac{1}{\Gamma(d)}. \quad (2.10)$$

*Důkaz.* Vzorce (2.7), (2.8) a (2.9) vychází z Maclaurinovy formule. Užitím Strilingova vzorce dostaneme (2.10).  $\square$

**2.4 Věta.** *FDWN procesy lze vyjádřit pomocí  $AR(\infty)$  a  $MA(\infty)$ , tj. jsou ve tvaru*

$$\sum_{j=0}^{\infty} a_j X_{t-j} = \varepsilon_t$$

a

$$X_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j},$$

v daném pořadí, kde koeficienty  $a_j$  a  $b_j$  jsou dány vzorci (2.8) a (2.9) z předchozí věty.

*Důkaz.* Viz [15]. Tvrzení vychází z (2.1), (2.2) a (2.7), neboť

$$\sum_{j=0}^{\infty} |a_j| < \infty \quad \text{a} \quad \sum_{j=0}^{\infty} |b_j|^2 < \infty \quad \text{pro} \quad d \in \left(0, \frac{1}{2}\right).$$

$\square$

## 2.3 Odhady parametrů

Model (2.1) má dva parametry, tj.  $d$  a  $\sigma^2$ . Hlavním problémem je odhadnout parametr  $d$ , jelikož odhad  $\sigma^2$  následně provedeme pomocí vzorce (2.5).

Nejjednodušším postupem je momentová metoda. Nechť  $\hat{\gamma}_k$  a  $\hat{\sigma}_k$  jsou odhady pro  $\gamma_k$  a  $\rho_k$  v daném pořadí. Pomocí (2.6) je zřejmé, že  $d = \frac{\rho_1}{1 + \rho}$ . Odtud dostaneme odhad pro  $d$ , který je tvaru

$$\hat{d} = \frac{\hat{\rho}_1}{(1 + \hat{\rho}_1)}$$

a odhad pro  $\sigma^2$

$$\hat{\sigma}^2 = \hat{\gamma}_0 \frac{\Gamma^2(1 - \hat{d})}{\Gamma(1 - 2\hat{d})}.$$

Pokud jsou  $\hat{\gamma}_k$  a  $\hat{\rho}_k$  konzistentní odhady, pak platí totéž pro  $\hat{d}$  a  $\hat{\sigma}^2$ . Další možností je použít odhad  $\hat{\phi}_{kk}$  parciální autokorelační funkce ( $k = 1, 2, \dots, m \ll n$ , pro vhodné  $m$ , viz [1]). Z parciální korelační funkce

$$\alpha(k) = \phi_{kk} = \frac{d}{k - d} \quad \text{pro } k = 1, 2, \dots \quad (2.11)$$

procesu  $\{X_t\}$  (2.11) máme odhady

$$\hat{d}_k = \frac{k\hat{\phi}_{kk}}{1 + \hat{\phi}_{kk}},$$

které mohou posloužit na výsledný odhad

$$d^* = \frac{\hat{d}_1 + \dots + \hat{d}_m}{m}.$$

Tento postup dokáže ověřit vhodnost modelu. Ale je k tomu ještě zapotřebí dalších teoretických poznatků. Geweke a Porter-Hudak [12] navrhli následující metodu. Z (2.3) dostaneme

$$\ln f(\lambda) = \ln \frac{\sigma^2}{2\pi} - d \ln \left( 4 \sin^2 \frac{\lambda}{2} \right). \quad (2.12)$$

Nechť je dána realizace  $X_1, \dots, X_n$ . Uvažujme hodnoty periodogramu  $I(\lambda)$  v bodech  $\lambda = \lambda_j$ , kde

$$\lambda_j = \frac{2\pi j}{n} \quad \text{pro } j = 0, 1, \dots, n - 1.$$

Vzhledem k (2.12) dostaneme

$$\ln I(\lambda_j) = \ln \frac{\sigma^2}{2\pi} - d \ln \left( 4 \sin^2 \frac{\lambda_j}{2} \right) + \ln \frac{I(\lambda_j)}{f(\lambda_j)}. \quad (2.13)$$

Proměnné  $\ln \frac{I(\lambda_j)}{f(\lambda_j)}$  jsou asymptoticky nezávisle stejně rozdělené náhodné veličiny s asymptotickou střední hodnotou  $-C$  a s asymptotickým rozptylem  $\pi^2/6$ , kde  $C = 0,57721\dots$  je Eulerova konstanta. Odtud (2.13) může být považováno za regresní model

$$Y_j = a + dx_j + e_j \quad \text{pro } j = 0, 1, \dots, n - 1,$$

kde  $Y_j = \ln I(\lambda_j)$  a  $x_j = -\ln(4 \sin^2 \frac{\lambda_j}{2})$ . Pokud  $\tilde{a}$  a  $\tilde{d}$  jsou střední kvadratické odhady regresních parametrů, pak  $\tilde{d}$  může posloužit jako odhad našeho původního parametru  $d$ . Jelikož  $\tilde{a}$  je odhad pro  $\ln \frac{\sigma^2}{2\pi} + C$ , odhad pro  $\sigma^2$  je

$$\tilde{\sigma} = 2\pi e^{\tilde{a}-C}.$$

Pokud dvojice  $(Y_j, x_j)$  nemají lineární závislost, může to být náznakem, že FDWN není správným modelem na analyzovaná data.

V tomto případě můžeme ještě zkusit jiný postup založený na následující myšlence. Pokud aproximujeme v (2.3) hodnotu  $\sin x \doteq x$  pro malá  $x$ , pak

$$f(\lambda) \doteq \frac{\sigma^2 \lambda^{-2d}}{2\pi}.$$

Označme  $F(\lambda)$  spektrální distribuční funkci odpovídající hustotě  $f(\lambda)$ . Pak

$$F(x) - F(0) = \int_0^x f(\lambda) d\lambda \doteq \frac{\sigma^2}{2\pi} \frac{x^{1-2d}}{1-2d}.$$

Ale  $F(x) - F(0)$  může být odhadnuto pomocí  $\int_0^x I(\lambda) d\lambda$ . Jelikož

$$I(\lambda) = \frac{1}{2\pi} \left( C_0^* + 2 \sum_{k=1}^{n-1} C_k^* \cos k\lambda \right),$$

kde

$$C_k^* = \frac{1}{n} \sum_{t=1}^{n-k} X_t X_{t+k} \quad \text{pro } k = 0, 1, \dots, n-1,$$

takže máme

$$\int_0^x I(\lambda) d\lambda = \frac{1}{2\pi} \left( x C_0^* + 2 \sum_{k=1}^{n-1} \frac{C_k^*}{k} \sin kx \right).$$

Jelikož

$$\ln(F(x) - F(0)) = \ln \frac{\sigma^2}{2\pi(1-2d)} + (1-2d) \ln x,$$

můžeme opět použít odhady parametrů pomocí lineární regrese.

## 2.4 ARFIMA procesy

Zobecněním FDWN procesů jsou autoregresní frakcionálně integrované procesy klouzavých součtů. V dalším textu budeme užívat zkratku ARFIMA z anglického názvu *autoregressive fractionally integrated moving average*. Definujeme je rovnicí

$$\Phi(B)Y_t = \Theta(B)(1-B)^{-d}\varepsilon_t, \tag{2.14}$$

kde  $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  a  $\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  jsou autoregresní operátor a operátor klouzavých součtů,  $\Phi(B)$  a  $\Theta(B)$  nemají žádné společné kořeny,  $B$  je operátor zpětného chodu a  $(1 - B)^{-d}$  je frakcionální diferenciální operátor daný binomickým rozvojem

$$(1 - B)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} B^j = \sum_{j=0}^{\infty} \eta_j B^j,$$

pro  $d \in (-1, 1/2)$  a  $\{\varepsilon_t | t \in \mathbb{Z}\}$  je posloupnost bílého šumu s nulovou střední hodnotou  $\sigma^2$ . Asymptotická aproximace pro

$$\eta_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}$$

pro velké  $j$  je

$$\eta_j \sim \frac{j^{d-1}}{\Gamma(d)},$$

kde  $\Gamma(x)$  je obvyklá gamma funkce.

**Poznámka.** ARFIMA(p, d, q) proces bývá někdy definován jako

$$\Phi(B)(1 - B)^d Y_t = \Theta(B)\varepsilon_t.$$

### 2.4.1 Stacionarita, kauzalita a invertibilita

**2.5 Věta.** Uvažujme ARFIMA proces definovaný pomocí (2.14) a předpokládejme, že polynomy  $\Phi(\cdot)$  a  $\Theta(\cdot)$  nemají žádné společné kořeny a že  $d \in (-1, 1/2)$ . Potom platí

- (a) Pokud kořeny  $\Phi(\cdot)$  leží vně jednotkové kružnice  $\{z : |z| = 1\}$ , pak existuje jediné stacionární řešení rovnice (2.14) dané  $Y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$ , kde  $\psi_j$  jsou koeficienty následujícího polynomu  $\psi(z) = (1 - z)^{-d} \Theta(z) / \Phi(z)$ .
- (b) Pokud kořeny  $\Phi(\cdot)$  leží vně uzavřeného jednotkového kruhu  $\{z : |z| \leq 1\}$ , pak je řešení  $\{Y_t\}$  kauzální.
- (c) Pokud kořeny  $\Theta(\cdot)$  leží vně uzavřeného jednotkového kruhu  $\{z : |z| \leq 1\}$ , pak je řešení  $\{Y_t\}$  invertibilní.

*Důkaz.* Viz [19] strana 44 až 46. □

### 2.4.2 Spektrální hustota

Uvažujme ARFIMA proces v  $MA(\infty)$  vyjádření

$$X_t = \psi(B)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}. \quad (2.15)$$

Spektrální hustota náhodného procesu (2.15) je absolutně spojitá funkce vzhledem k Lebesgueově míře na intervalu  $[-\pi, \pi]$ . Tato spektrální hustota procesu (2.14) může být zapsána jako

$$\begin{aligned} f(\lambda) &= \frac{\sigma^2}{2\pi} |\psi(e^{-i\lambda})|^2 \\ &= \frac{\sigma^2}{2\pi} |1 - e^{-i\lambda}|^{-2d} \frac{|\Theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2} \\ &= \frac{\sigma^2}{2\pi} \left(2 \sin \frac{\lambda}{2}\right)^{-2d} \frac{|\Theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2}, \end{aligned} \quad (2.16)$$

kde  $i$  značí imaginární jednotku.

### 2.4.3 Whittleův odhad

Předpokládejme, že  $Y_t \sim ARFIMA(p, d, q)$  s autokovarianční funkcí

$$\gamma_\theta(k-j) = \int_{-\pi}^{\pi} f_\theta(\lambda) e^{i\lambda(k-j)} d\lambda,$$

kde  $\theta = (p, d, q)$ . Dále označme konečný vektor  $\mathbf{Y} = (Y_1, \dots, Y_n)'$ . Logaritmičká věrohodnostní funkce procesu  $Y_t$  je dána (viz [19]) jako

$$L(\theta) = -\frac{1}{2n} \ln \det \mathbf{\Gamma}_\theta - \frac{1}{2n} \mathbf{Y}' \mathbf{\Gamma}_\theta^{-1} \mathbf{Y}, \quad (2.17)$$

kde  $\mathbf{\Gamma}_\theta = [\gamma_\theta(k-j)]$  je matice autokovariancí. K získání Whittleovy metody (2.17) uděláme dvě aproximace pro výrazy  $\ln \det \mathbf{\Gamma}_\theta$  a  $\mathbf{Y}' \mathbf{\Gamma}_\theta^{-1} \mathbf{Y}$ . Jelikož

$$\frac{1}{n} \ln \det \mathbf{\Gamma}_\theta \rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln (2\pi f_\theta(\lambda)) d\lambda,$$

pro  $n \rightarrow \infty$ , první výraz z (2.17) je aproximován

$$\frac{1}{2n} \ln \det \mathbf{\Gamma}_\theta \approx \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln (2\pi f_\theta(\lambda)) d\lambda.$$

Dále druhý výraz je aproximován

$$\begin{aligned} \frac{1}{2n} \mathbf{Y}' \mathbf{\Gamma}_\theta^{-1} \mathbf{Y} &\approx \sum_{\ell=1}^n \sum_{j=1}^n y_\ell \left( \frac{1}{8\pi^2 n} \int_{-\pi}^{\pi} f_\theta^{-1}(\lambda) \exp(i\lambda(\ell-j)) d\lambda \right) y_j \\ &= \frac{1}{8\pi^2 n} \int_{-\pi}^{\pi} f_\theta^{-1}(\lambda) \sum_{\ell=1}^n \sum_{j=1}^n y_\ell y_j \exp(i\lambda(\ell-j)) d\lambda \\ &= \frac{1}{8\pi^2 n} \int_{-\pi}^{\pi} f_\theta^{-1}(\lambda) \left| \sum_{j=1}^n y_j \exp(i\lambda j) \right|^2 d\lambda \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{I_n(\lambda)}{f_\theta(\lambda)} d\lambda, \end{aligned}$$

kde  $I_n(\lambda)$  je *periodogram* definovaný v rovnici (1.3). Z předchozího dostáváme

$$L(\theta) \approx -\frac{1}{4\pi} \left( \int_{-\pi}^{\pi} \ln(2\pi f_{\theta}(\lambda)) d\lambda + \int_{-\pi}^{\pi} \frac{I_n(\lambda)}{f_{\theta}(\lambda)} d\lambda \right), \quad (2.18)$$

pro  $n \rightarrow \infty$ . Diskrétní verze (2.18) je Riemannova aproximace integrálu a je tvaru

$$L(\theta) \approx -\frac{1}{2n} \left( \sum_{j=1}^n \ln f_{\theta}(\lambda_j) + \sum_{j=1}^n \frac{I_n(\lambda)}{f_{\theta}(\lambda)} \right),$$

kde  $\lambda_j = 2\pi j/n$  jsou Fourierovy frekvence.

## Kapitola 3

# Příklad reálné časové řady modelované pomocí ARFIMA modelů

### 3.1 Popis zkoumaných dat

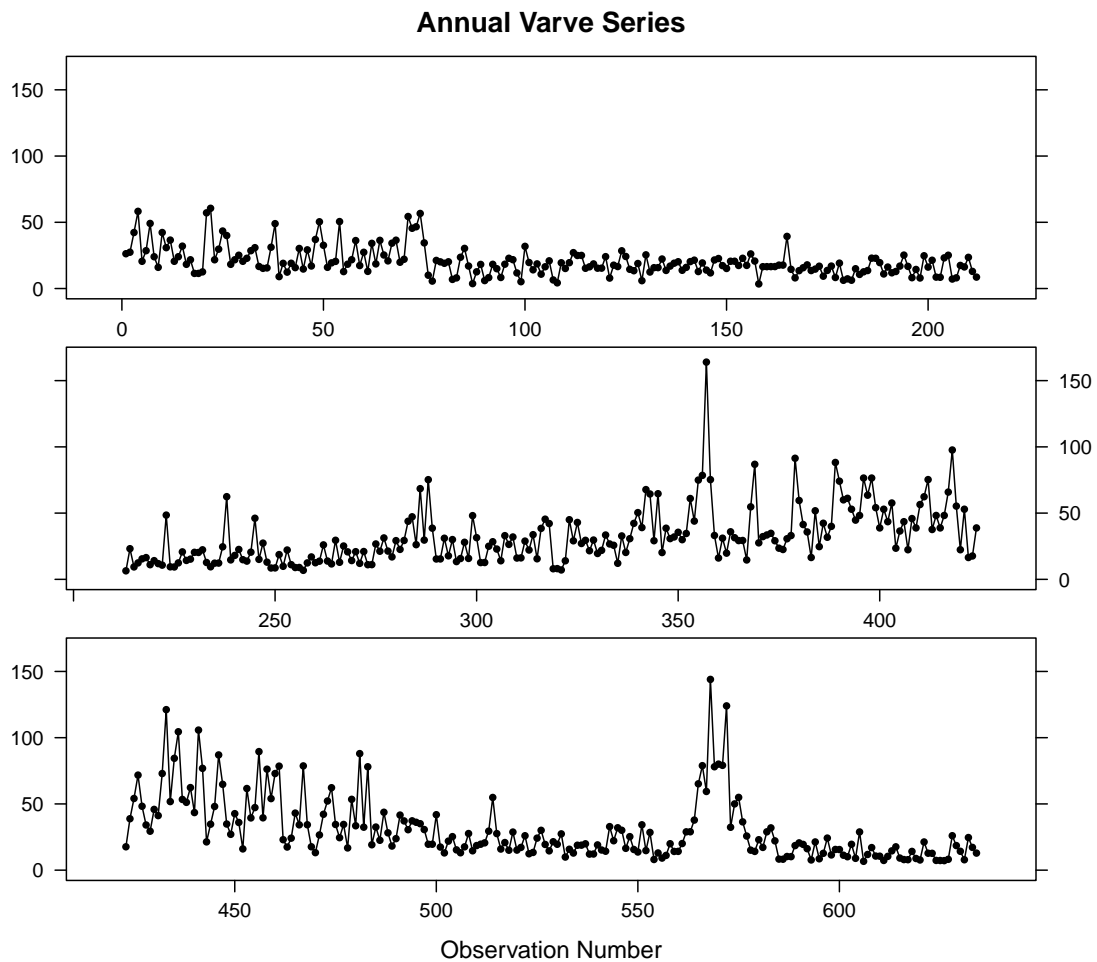
Budeme pracovat s datovým souborem *varve* z balíčku *astsa* [23]. Jde o geologické vrstvy (varvy) naměřené v Massachusetts. Nejstarší naměřená vrstva je stará 11 834 let a nejmladší 11 201 let. Data jsou roční, tedy 634 rokům odpovídá 634 naměřených hodnot. Varva je tenká vrstvička sedimentu (lamina) usazená během jednoho roku v jezerech napájených vodou z ledovce. Skládá se z písčité části usazené v létě a z části jílovité usazené v zimním období. V širším smyslu je varva jakákoliv vrstvička uložená během jednoho roku. Sediment složený z varv se nazývá varvit a jemné a pravidelné laminování se označuje jako varvitická textura [20].

V úvodu načteme potřebné balíčky a vstupní data graficky znázorníme.

```
>library(astsa)
>library(FitAR)
>library(fracdiff)
>library(forecast)

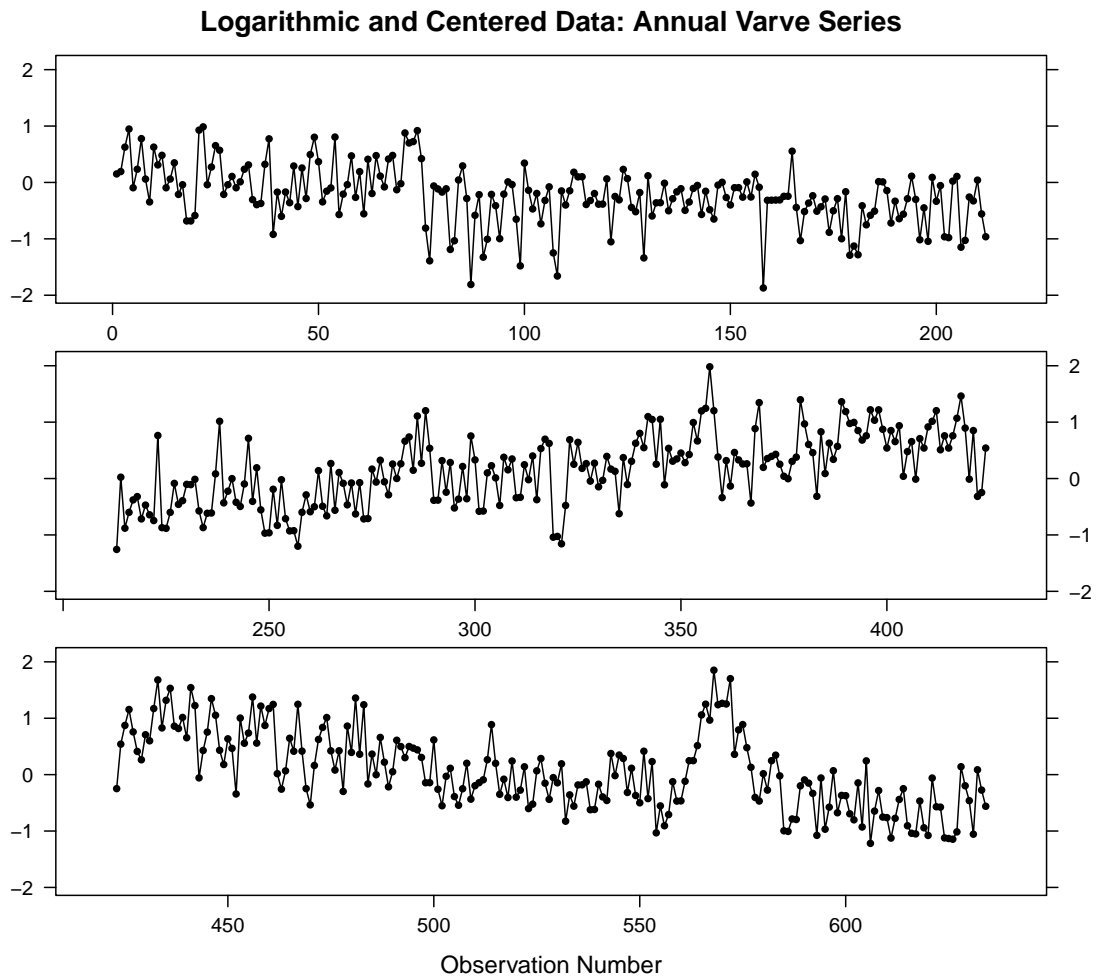
>data(varve)
>TXT<-"Annual Varve Series"
>pdf("PlotVarve1.pdf",width=8,height=7.25)
>TimeSeriesPlot(varve,SubLength=212,as.table=T,main=TXT,
                 col="black",type="o",pch=20,cex=0.75)
>dev.off()
```





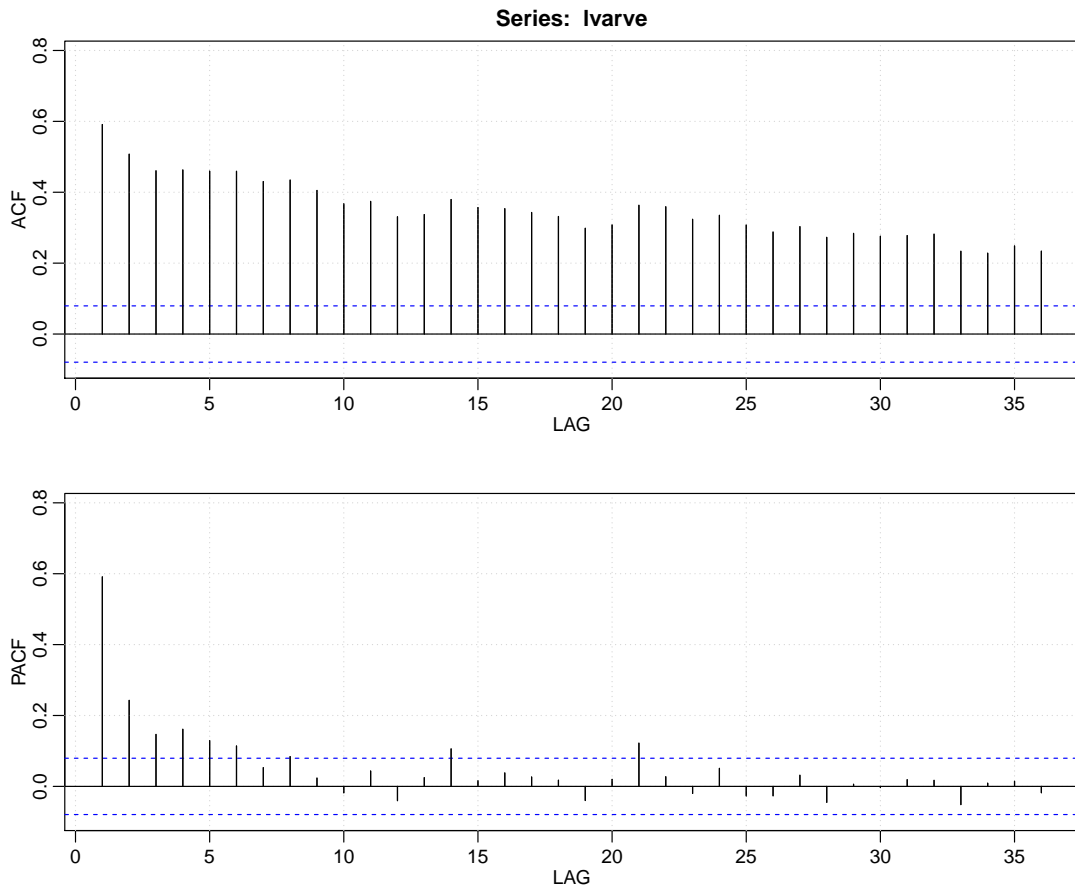
Z grafu je vidět zřejmá nestacionarita v rozptylu, proto budeme dále pracovat s logaritmovanými daty, která nejprve zobrazíme do grafu.

```
>TXT<-"Logarithmic and Centered Data: Annual Varve Series"
>lvarve = log(varve) - mean(log(varve))
>pdf("PlotLogVarve1.pdf",width=8,height=7.25)
>TimeSeriesPlot(lvarve,SubLength=212,as.table=T,main=TXT,
                 col="black",type="o",pch=20,cex=0.75)
>dev.off()
```



Dále vykreslíme průběhy autokorelační funkce (ACF) a parciální autokorelační funkce (PACF) pomocí funkce `acf2()` z balíčku `astsa`.

```
>pdf("LogVarve_acf2.pdf",width=10,height=8.25)
>pom<-acf2(lvarve)
>dev.off()
```



Ihned vidíme, že ACF klesá velmi pomalu, takže tato časová řada by mohla být vhodná pro analýzu dlouhé paměti.

### 3.2 Polynomický trend s ARMA reziduální částí

Nejprve se pokusíme modelovat dlouhodobé cyklické chování pomocí polynomu 3. řádu. Vzhledem k tomu, že jednotlivé hodnoty jsou v čase korelované, není vhodné použít klasický lineární regresní model, tj. funkci `lm()` z balíčku `stats`. Místo toho použijeme z balíčku `forecast` funkci `auto.arima()`, která dokáže pro reziduální část najít nejvhodnější řád ARMA procesu a přitom trend popsat pomocí zadané matice plánu, kterou nejprve vytvoříme.

```
>n <- length(varve)
>nn <- n
>tt <- 1:length(varve)
>tt <- ts(tt)
>ct <- tt - mean(tt)
>ct2 <- ct^2
```

```
>ct3 <- ct^3
>Xreg <- cbind(ct, ct2, ct3)
>fit_polynom3<-auto.arima(lvarve,trace=TRUE,xreg=Xreg)
```

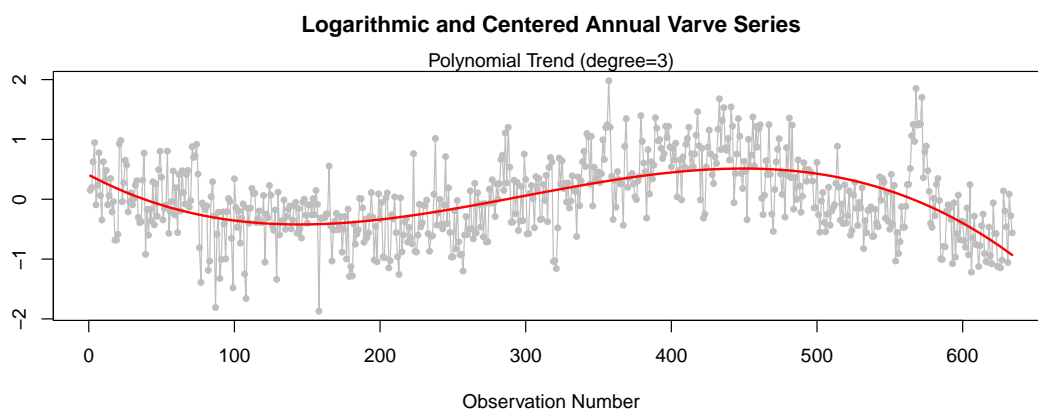
```
ARIMA(2,0,2) with non-zero mean : 863.2215
ARIMA(0,0,0) with non-zero mean : 981.5001
ARIMA(1,0,0) with non-zero mean : 880.6215
ARIMA(0,0,1) with non-zero mean : 906.9557
ARIMA(1,0,2) with non-zero mean : 860.7353
ARIMA(1,0,1) with non-zero mean : 864.0234
ARIMA(1,0,3) with non-zero mean : 861.8481
ARIMA(2,0,3) with non-zero mean : 864.2456
ARIMA(1,0,2) with zero mean      : 861.2278
ARIMA(0,0,2) with non-zero mean : 884.5786
```

Best model: ARIMA(1,0,2) with non-zero mean

Jako optimální byl navržen model ARMA(1,2).

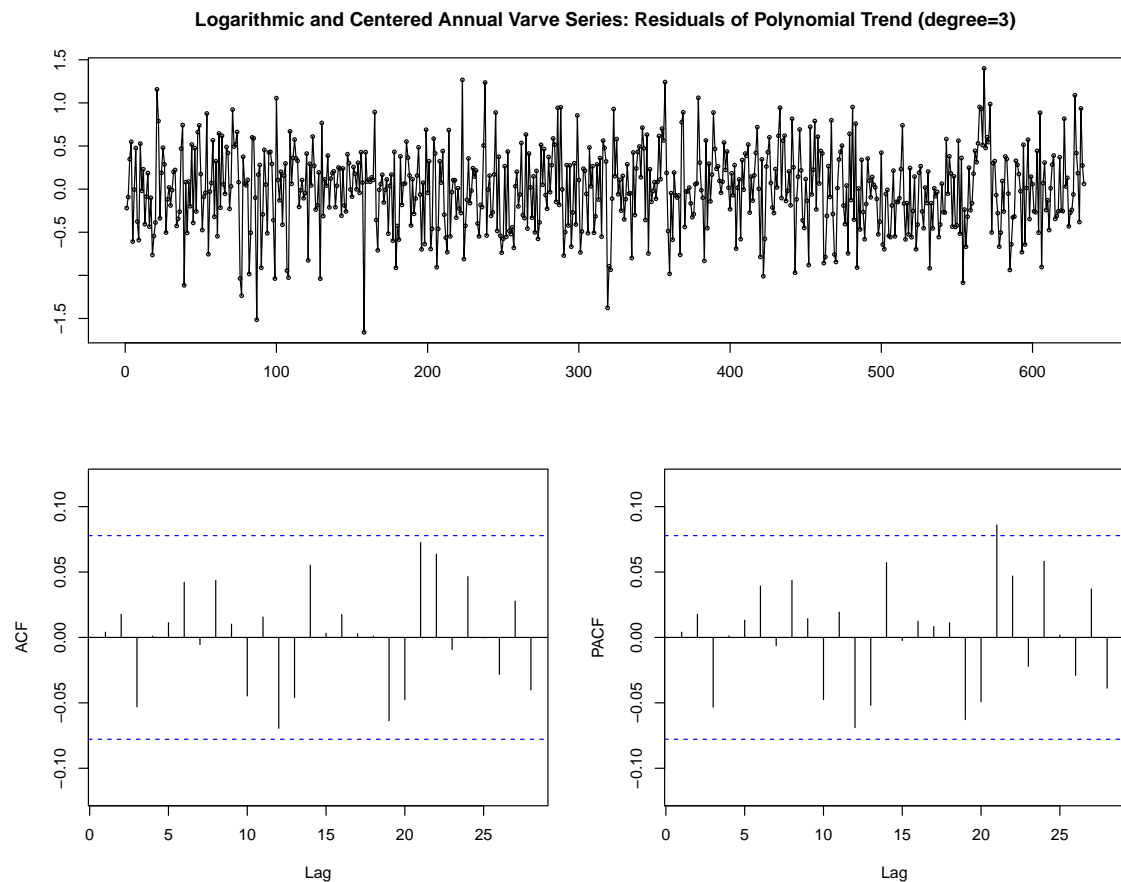
Abychom měli vizuální představu o odhadnutém modelu, polynomický trend spolu s daty nejprve zobrazíme.

```
>beta<-matrix(coef(fit_polynom3)[4:7],ncol=1)
>FittedTr<-cbind(rep(1,n),Xreg) %*% beta
>Txt<-"Logarithmic and Centered Annual Varve Series"
>pdf("LogVarve-PolTr.pdf",width=10,height=4)
>TimeSeriesPlot(lvarve,main=Txt,col="gray",type="o",pch=20)
>lines(ts(FittedTr),col="red",lwd=2)
>mtext("Polynomial Trend (degree=3)")
>dev.off()
```



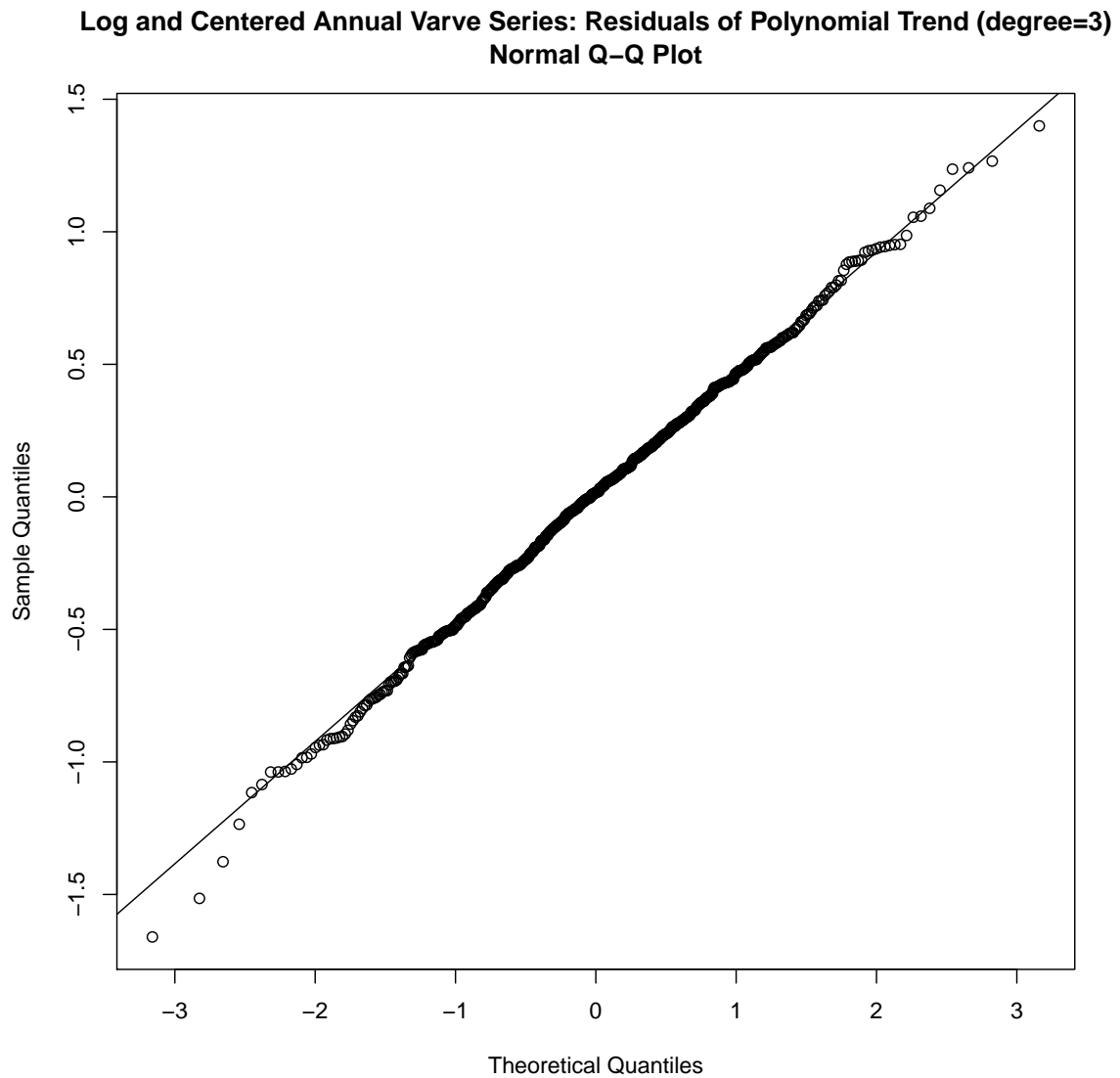
Získaná rezidua by měla být přibližně gaussovským bílým šumem, což ověříme graficky. Nejprve vykreslíme rezidua, ACF a PACF pomocí funkce `tsdisplay()` z balíčku `forecast`.

```
>ResidF <- resid(fit_polynom3)
>Txt<-"Logarithmic and Centered Annual Varve Series:
      Residuals of Polynomial Trend (degree=3)"
>pdf("LogVarve-PolTrRES.pdf",width=10,height=8)
>forecast::tsdisplay(ResidF,main=Txt)
>dev.off()
```



Normální rozdělení reziduí ověříme pomocí Q-Q grafu.

```
>Txt<-"Log and Centered Annual Varve Series:
      Residuals of Polynomial Trend (degree=3)\nNormal Q-Q Plot"
>pdf("LogVarve-PolTrRESqq.pdf",width=8.25,height=8.25)
>qqnorm(ResidF,main=Txt)
>qqline(ResidF)
>dev.off()
```



Zatím se rezidua přibližně jeví jako bílý šum. Poslední zkoušku reziduí provedeme pomocí funkce `auto.arima()` z balíčku `forecast`.

```
>pom <- auto.arima(ResidF,trace=TRUE)
```

```
ARIMA(2,0,2) with non-zero mean : 850.9519
ARIMA(0,0,0) with non-zero mean : 847.1898
ARIMA(1,0,0) with non-zero mean : 849.98
ARIMA(0,0,1) with non-zero mean : 849.1989
ARIMA(1,0,1) with non-zero mean : 851.9937
ARIMA(0,0,0) with zero mean      : 845.1773
ARIMA(1,0,0) with zero mean     : 847.962
ARIMA(0,0,1) with zero mean     : 847.18
```

```
ARIMA(1,0,1) with zero mean      : 849.9673
```

```
Best model: ARIMA(0,0,0) with zero mean
```

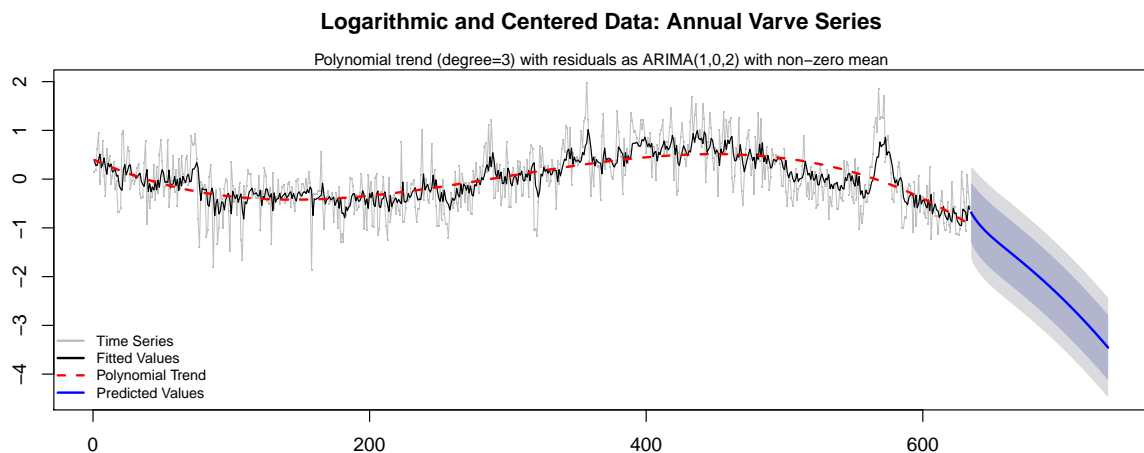
Vzhledem k tomu, že funkce `auto.arima()` z balíčku `forecast` pro rezidua navrhuje jako optimální model `ARMA(0,0)`, což je bílý šum, i tato poslední zkouška dopadla pozitivně.

Na závěr této části provedeme předpověď (například o 100 kroků dopředu) pomocí funkce `forecast()` z balíčku `forecast`. Nesmíme zapomenout vytvořit matici plánu pro predikci v čase.

```
>fitted <- lvarve - fit_polynom3$residuals
>m<-100
>CT<-ct[length(ct)]+c(1:m)
>CT2<-CT^2
>CT3<-CT^3
>XREG <- cbind(CT,CT2,CT3)
>pred.PolTr<-forecast(fit_polynom3,h=m,xreg=XREG)
```

Výslednou predikci graficky znázorníme pomocí funkce `plot.forecast()` z balíčku `forecast`.

```
>pdf("LogVarve-PolTr-final.pdf",width=10,height=4.25)
>op<-par(mar=c(2,2,4,0) + 0.25)
>plot(pred.PolTr,type="o",main=TXT,col="gray",pch=".",
      cex=0.5,lwd=0.75)
>lines(fitted,col="black")
>lines(ts(FittedTr),col="red",lwd=2,lty=2)
>legend("bottomleft",cex=0.75,
      legend = c("Time Series","Fitted Values","Polynomial Trend",
        "Predicted Values"), bty="n",lty = c(1,1,2,1), xjust = 1,
      yjust = 1,lwd=2, col=c("gray","black","red","blue"),
      pch=c(" ",NA,NA,NA))
>mtext(paste("Polynomial trend (degree=3) with residuals as",
      pred.PolTr$method),cex=0.85)
>par(op)
>dev.off()
```



### 3.3 Periodický trend s ARMA reziduální částí

Vzhledem k tomu, že polynomičtý trend je nevhodný k predikci budoucích hodnot, zvolíme raději popis trendu pomocí funkcí sinus a kosinus. Abychom to mohli provést, potřebujeme nejprve zjistit, které významné periody trend vytvářejí. Vyjdeme proto z periodogramu a pomocí Fisherova či Siegelova testu tyto významné periody určíme. Než tak učiníme, musíme z časové řady nejprve odstranit případný lineární trend.

```
>y <- as.vector(lvarve)
>x <- as.vector(time(lvarve))
>data<-data.frame(x,y)
>LinTrend <- lm( y ~ x,data=data)
>print(summary(LinTrend))
```

Call:

```
lm(formula = y ~ x, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.79214	-0.42844	-0.03282	0.39807	1.96235

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.1568956	0.0502189	-3.124	0.001864 **
x	0.0004942	0.0001370	3.606	0.000335 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6315 on 632 degrees of freedom

Multiple R-squared: 0.02016, Adjusted R-squared: 0.01861



F-statistic: 13 on 1 and 632 DF, p-value: 0.0003353

```
>res<-resid(LinTrend)
>TrigoTrend<-trigofit(res)
>str(TrigoTrend)
```

List of 10

```
$ periodogram: num [1:316] 3.3249 2.7248 0.1782 0.0366 0.4937 ...
$ freq       : num [1:316] 0.00993 0.01985 0.02978 0.0397 0.04963 ...
$ signA      : num [1:6] -0.19523 0.00471 -0.09925
                0.12834 -0.08717 ...
$ signB      : num [1:6] -0.3064 0.3289 0.1335 0.0726 0.1096 ...
$ signFreq   : num [1:6] 0.00993 0.01985 0.05956 0.08933 0.04963 ...
$ amplituda  : num [1:316] 0.02042 0.01849 0.00473 0.00214 0.00787 ...
$ faze       : num [1:316] 843.067 0.822 6.988 -0.884 -3.079 ...
$ signL      : logi [1:316] TRUE TRUE FALSE FALSE TRUE TRUE ...
$ A          : num [1:316] -0.19523 0.00471 0.04823
                -0.03088 -0.08717 ...
$ B          : num [1:316] -0.3064 0.3289 0.0689 0.0224 0.1096 ...
```

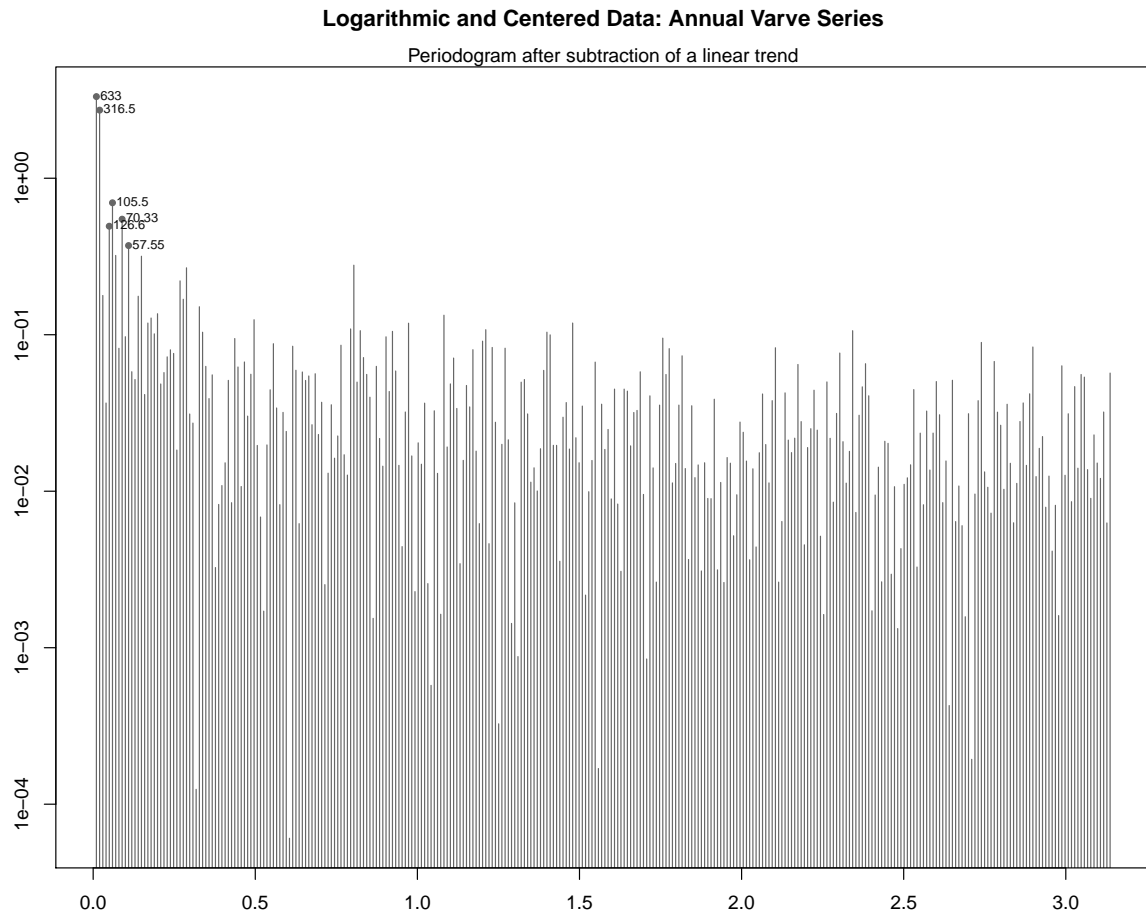
Jako statisticky významné byly určeny tyto periody

```
>round(FisherPeriods<-2*pi/TrigoTrend$signFreq,2)
```

```
[1] 633.00 316.50 105.50 70.33 126.60 57.55
```

Periodogram pro data s odstraněným lineárním trendem vykreslíme.

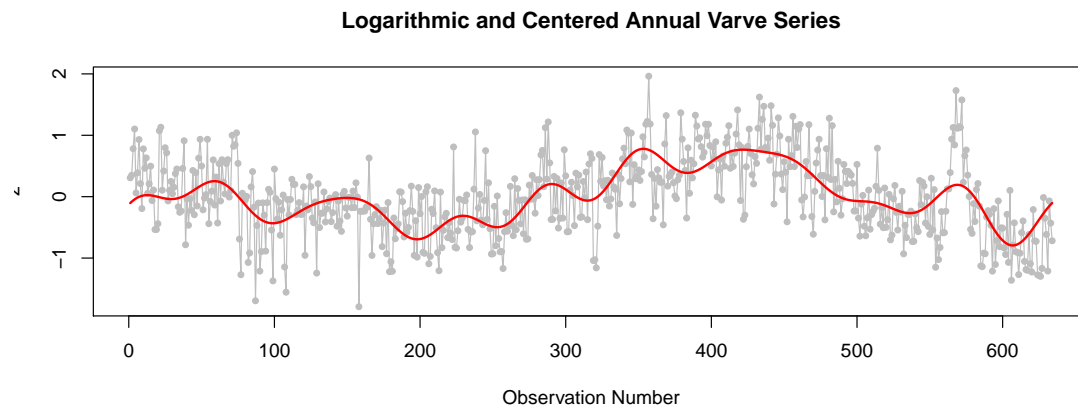
```
>col1<-"gray40"
>col2<-"gray5"
>pdf("LogVarve-Periodogram.pdf",width=10,height=8.25)
>op<-par(mar=c(2,2,4,0) + 0.25)
>with(TrigoTrend,plot(freq,periodogram,type="h",
                    log="y",col=col1,main=TXT))
>with(TrigoTrend,points(freq[signL],
                    periodogram[signL],pch=20,col=col1))
>with(TrigoTrend,text(freq[signL],periodogram[signL],col=col2,
                    round((2*pi)/freq[signL],2),pos=4,offset=0.175,cex=0.65))
>mtext("Periodogram after subtraction of a linear trend")
>par(op)
>dev.off()
```



Vidíme, že všechny významné frekvence jsou velmi blízko počátku, což opět svědčí o možné dlouhé paměti v časové řadě.

Nyní pro data bez lineárního trendu zobrazíme výsledný model tvořený šesti významnými periodami.

```
>TTIME<-seq(1,length(res),length=nn)
>predTrigoTrend<-trigoval(TTIME,A=TrigoTrend$signA,B=TrigoTrend$signB,
                           freq=TrigoTrend$signFreq)
>Txt<-"Logarithmic and Centered Annual Varve Series"
>pdf("LogVarve-TrigTr.pdf",width=10,height=4)
>TimeSeriesPlot(res,main=Txt,col="gray",type="o",pch=20)
>lines(predTrigoTrend,col="red",lwd=2)
>dev.off()
```



Protože jsme si však vědomi faktu, že nabízené periody jsou závislé na délce časové řady, raději pro vyhledání významných period použijeme algoritmus norských autorů popsany v části 1.2.5, který je založen na nelineární regresi.

```
>betaD<-damsleth(res,pstlevel=0.05)
```

Podívejme se, které významné periody byly vybrány.

```
>round(DamslethPeriods<-2*pi/betaD$freq,2)
```

```
freq  freq  freq  freq  freq
452.07 72.39 103.89 41.50 56.86
```

Vidíme, že tentokrát máme o jednu významnou periodu méně.

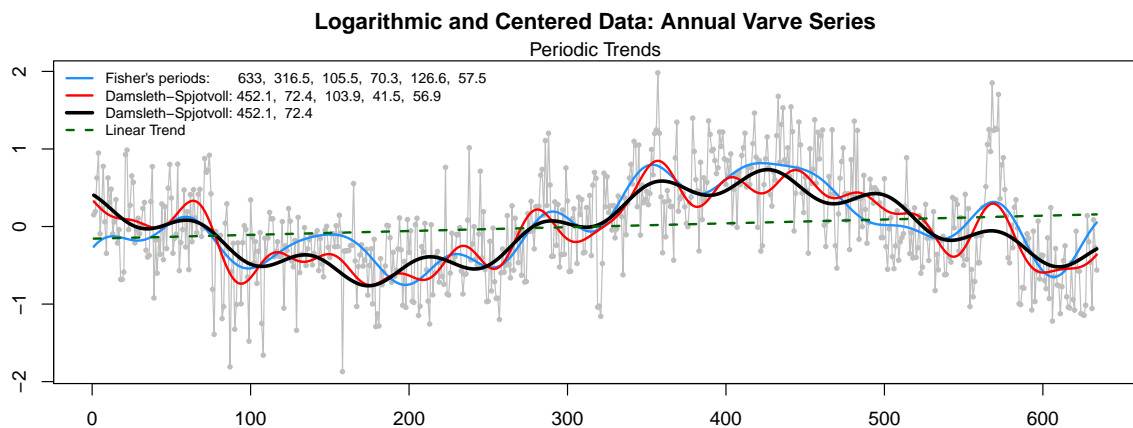
Oba dva modely graficky porovnejme. Pro konečné modelování použijeme jen první dvě významné periody, proto je tento model v grafu zachycen také.

```
>predTrigD<-trigoval(TTIME,A=betaD$A,B=betaD$B,freq=betaD$freq)
>predTrigD2<-trigoval(TTIME,A=betaD$A[1:2],
                      B=betaD$B[1:2],freq=betaD$freq[1:2])
>new <- data.frame(x = seq(x[1],x[n],length.out=nn))
>pred.lintrend<-predict(LinTrend, new)
>Txt<-"Logarithmic and Centered Annual Varve Series"
>pdf("LogVarve-TrigTrDall.pdf",width=10,height=4)
>op<-par(mar=c(2,2,3,0)+0.5)
>plot(y~x, type='o',main=Txt,cex.main=1.125,
      col="gray",pch=20,cex=0.75)
>lines(new$x,predTrigoTrend+pred.lintrend,col="dodgerblue",lwd=2)
>lines(new$x,predTrigD+pred.lintrend,lwd=2,col="red",lty=1)
>lines(new$x,predTrigD2+pred.lintrend,lwd=3,col="black",lty=1)
>lines(new$x,pred.lintrend,lwd=2,col="darkgreen",lty=2)
>TFP<-paste("Fisher's periods:      ",
```

```

      paste(round(FisherPeriods,1),collapse=","))
>TDS<-paste("Damsleth-Spjotvoll:",
           paste(round(DamslethPeriods,1),collapse=","))
>TDS2<-paste("Damsleth-Spjotvoll:",
            paste(round(DamslethPeriods[1:2],1),collapse=","))
>legend("topleft", legend = c(TFP,TDS,TDS2,"Linear Trend"),
      bty="n", lty = c(1,1,1,2), xjust = 1, yjust = 1,
      cex=0.75, lwd=c(2,2,3,2),
      col=c("dodgerblue","red","black","darkgreen"))
>mtext("Periodic Trends")
>par(op)
>dev.off()

```



Pro konečný model zvolíme první dvě významné periody a opět použijeme z balíčku `forecast` funkci `auto.arima()`, která dokáže pro reziduální část najít nejvhodnější řád ARMA procesu a přitom trend popsat pomocí zadané matice plánu, kterou nejprve vytvoříme.

```

>TIME<-time(varve)
>AA<-betaD$A[1:2]
>BB<-betaD$B[1:2]
>Freq=betaD$freq[1:2]
>xreg<-TIME
>for(k in 1:length(Freq))
>xreg<-cbind(xreg,trigpol2(TIME,AA[k],BB[k],Freq[k]))
>xreg<-as.matrix(xreg)
>colnames(xreg)<-c("time","cos1","sin1","cos2","sin2")
>fit_periodic2<-auto.arima(lvarve,trace=TRUE,xreg=xreg)

```

ARIMA(2,0,2) with non-zero mean : 851.9886

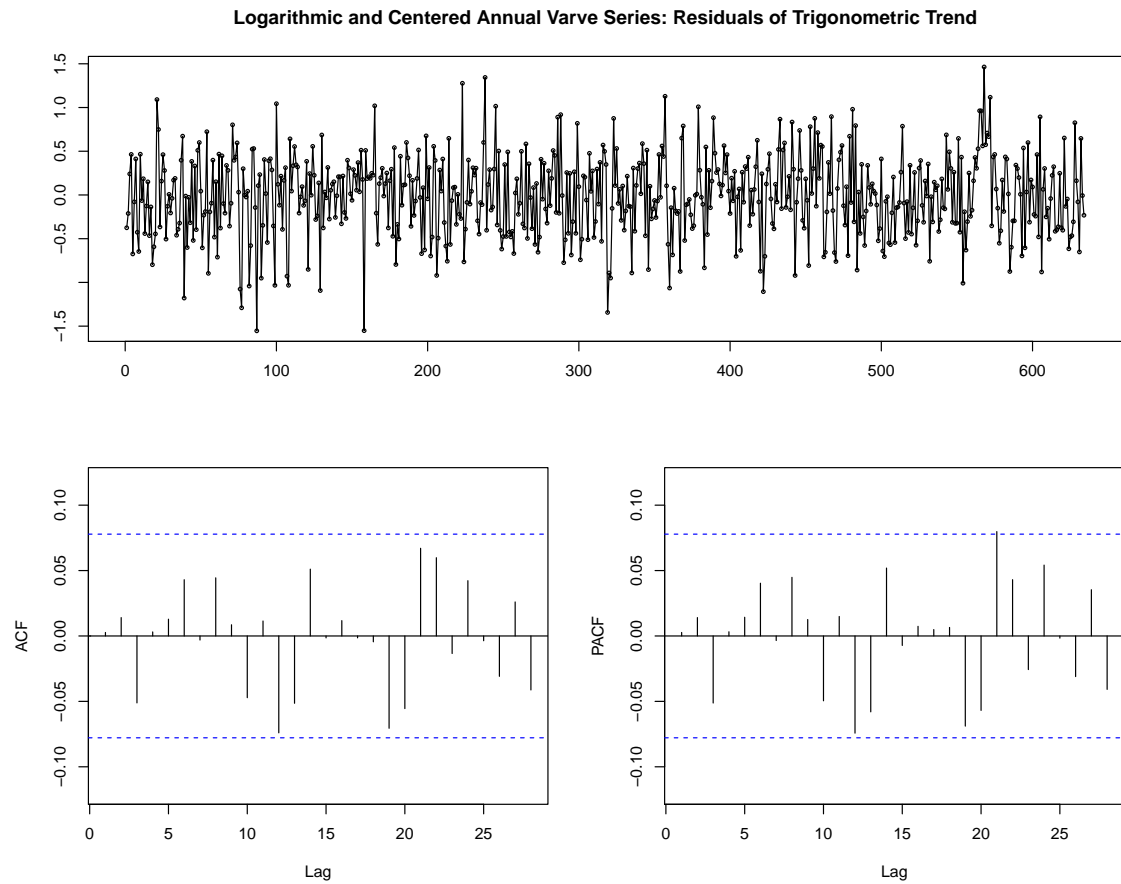
```
ARIMA(0,0,0) with non-zero mean : 923.3803
ARIMA(1,0,0) with non-zero mean : 854.9278
ARIMA(0,0,1) with non-zero mean : 869.0391
ARIMA(1,0,2) with non-zero mean : 849.2486
ARIMA(1,0,1) with non-zero mean : 849.7616
ARIMA(1,0,3) with non-zero mean : 850.5548
ARIMA(2,0,3) with non-zero mean : 853.049
ARIMA(1,0,2) with zero mean      : 848.1342
ARIMA(0,0,2) with zero mean      : 856.6082
ARIMA(2,0,2) with zero mean      : 850.8421
ARIMA(1,0,1) with zero mean      : 849.0481
ARIMA(1,0,3) with zero mean      : 849.3568
ARIMA(0,0,1) with zero mean      : 870.1664
ARIMA(2,0,3) with zero mean      : 851.8798
```

Best model: ARIMA(1,0,2) with zero mean

Obdobně jako u polynomického trendu, i v tomto případě jako optimální model pro rezidua byl zvolen ARMA(1,2).

Provedeme stejně jako u polynomického trendu obvyklou grafickou analýzu reziduí.

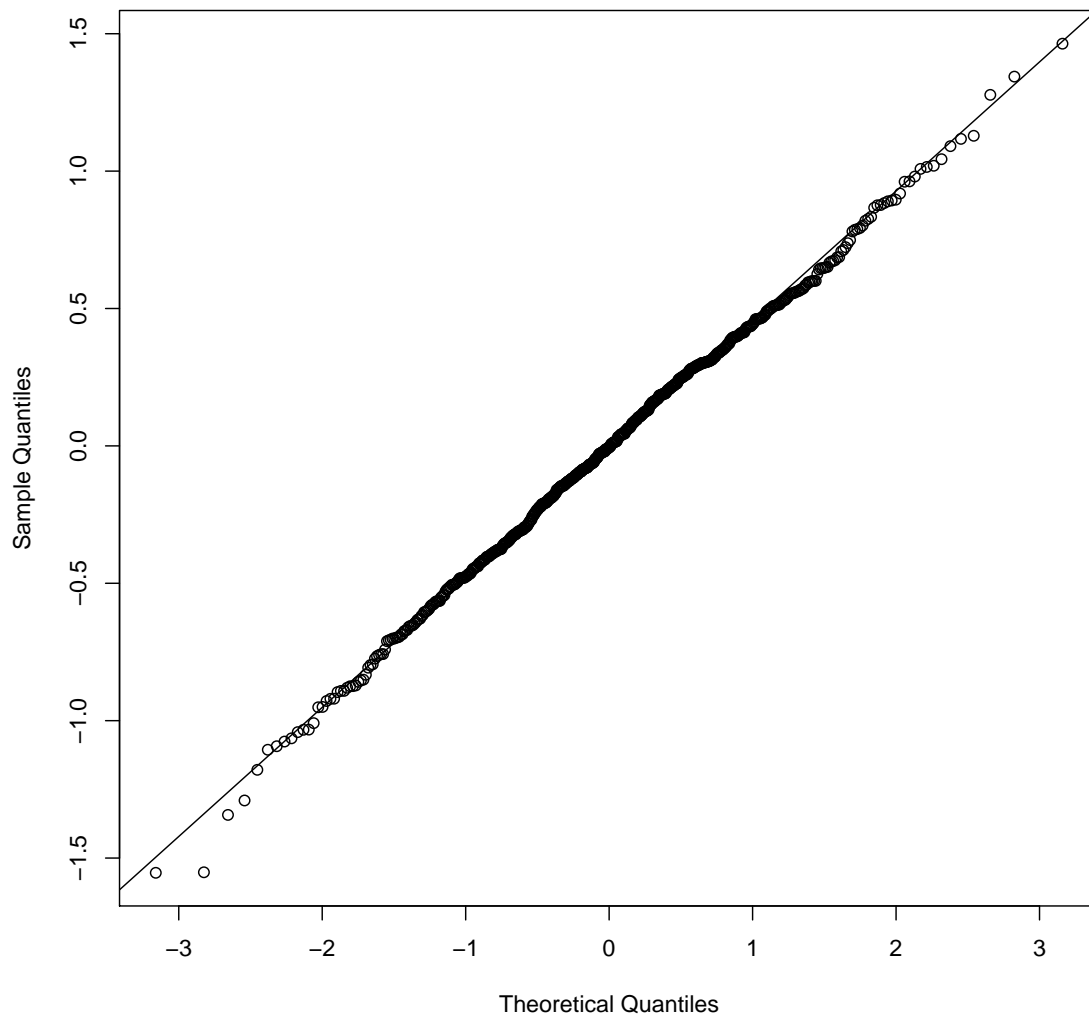
```
>RESIDF<-residuals(fit_periodic2)
>Txt<-"Logarithmic and Centered Annual Varve Series:
      Residuals of Trigonometric Trend"
>pdf("LogVarve-TrigTrRES.pdf",width=10,height=8)
>forecast::tsdisplay(RESIDF,main=Txt)
>dev.off()
```



```

>Txt<-"Log and Centered Annual Varve Series:
      Residuals of Trigonometric Trend\nNormal Q-Q Plot"
>pdf("LogVarve-TrigTrRESqq.pdf",width=8.25,height=8.25)
>qqnorm(RESIDF,main=Txt)
>qqline(RESIDF)
>dev.off()
    
```

**Log and Centered Annual Varve Series: Residuals of Trigonometric Trend  
Normal Q-Q Plot**



```
>pom <- auto.arima(RESIDF,trace=TRUE)
```

```
ARIMA(2,0,2) with non-zero mean : Inf
ARIMA(0,0,0) with non-zero mean : 832.9442
ARIMA(1,0,0) with non-zero mean : 835.3411
ARIMA(0,0,1) with non-zero mean : 834.9588
ARIMA(1,0,1) with non-zero mean : 837.2979
ARIMA(0,0,0) with zero mean      : 831.2098
ARIMA(1,0,0) with zero mean      : 833.5667
ARIMA(0,0,1) with zero mean      : 833.2166
ARIMA(1,0,1) with zero mean      : 835.5076
```

```
Best model: ARIMA(0,0,0) with zero mean
```

Vidíme, že rezidua se jeví jako bílý šum.

Na závěr této části opět provedeme predikci budoucích 100 hodnot pomocí funkce `forecast()` z balíčku `forecast`. Nesmíme zapomenout vytvořit matici plánu pro predikci v čase.

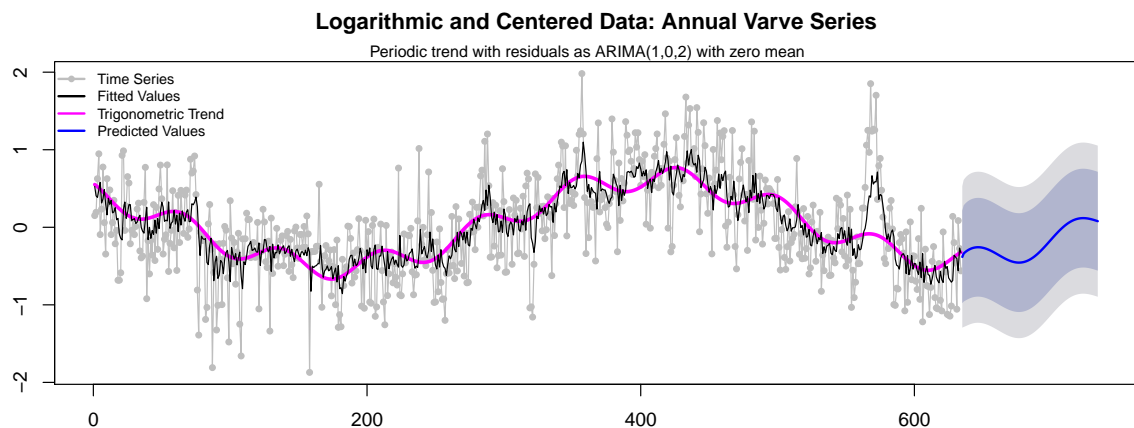
```
>Fitted <- lvarve - fit_periodic2$residuals
>m<-100
>TTIME<-TIME[length(TIME)]+c(1:m)
>Xreg<-TTIME
>for(k in 1:length(Freq))
>Xreg<-cbind(Xreg, trigpol2(TTIME, AA[k], BB[k], Freq[k]))
>Xreg<-as.matrix(Xreg)
>colnames(Xreg)<-c("time", "cos1", "sin1", "cos2", "sin2")

>pred.TrigTr<-forecast(fit_periodic2, h=m, xreg=Xreg)
```

Výslednou predikci graficky znázorníme pomocí funkce `plot.forecast()` z balíčku `forecast`.

```
>Beta<-matrix(coef(fit_periodic2)[4:8], ncol=1)
>FittedPTr<-xreg %*% Beta
>Txt<-"Logarithmic and Centered Annual Varve Series"
>pdf("LogVarve-TrigTrD-final.pdf", width=10, height=4)
>op<-par(mar=c(2, 2, 3, 0)+0.5)
>plot(pred.TrigTr, type="o", main=Txt, col="gray", pch=20, cex=0.75)
>lines(ts(FittedPTr), lwd=3, col="magenta", lty=1)
>lines(Fitted, col="black")
>legend("topleft", cex=0.75, legend = c("Time Series", "Fitted Values",
    "Trigonometric Trend", "Predicted Values"),
    bty="n", lty = c(1, 1, 1, 1), xjust = 1, yjust = 1, lwd=2,
    col=c("gray", "black", "magenta", "blue"), pch=c(20, NA, NA, NA))
>mtext(paste("Periodic trend with residuals as", pred.TrigTr$method),
    cex=0.85)
>par(op)
>dev.off()
```





### 3.4 Balíčky forecast a fracdiff

V předchozích modelech jsme zatím s faktem, že data mohou mít dlouhou paměť, neuvažovali. Pro práci s ARFIMA modely nejprve použijeme funkce z balíčku `forecast` [17]. Půjde především o funkci `arfima()`, která dokáže sama vybrat vhodný model s dlouhou pamětí. Kombinuje funkci `fracdiff()` z balíčku `fracdiff` [11] a funkci `auto.arima()` z balíčku `forecast`, a to následujícím způsobem:

1. Parametr  $d$  je nejprve odhadnut pomocí funkce `fracdiff()` pro model  $ARFIMA(2, d, 0)$ .
2. Na datech se dále provede frakcionální diference řádu  $d$ , pak se pomocí funkce `auto.arima()` určí parametry  $p$  a  $q$ .
3. Nakonec se pomocí funkce `fracdiff()` pro model  $ARFIMA(p, d, q)$  provedou konečné odhady parametrů.

```
>fit_forecast<-forecast::arfima(lvarve,estim="mle",trace=T)
ARIMA(2,0,2) with non-zero mean : 876.2168
ARIMA(0,0,0) with non-zero mean : 871.9092
ARIMA(1,0,0) with non-zero mean : 871.4302
ARIMA(0,0,1) with non-zero mean : 870.2264
ARIMA(1,0,1) with non-zero mean : 871.2864
ARIMA(0,0,2) with non-zero mean : 871.114
ARIMA(1,0,2) with non-zero mean : 873.2604
ARIMA(0,0,1) with zero mean : 868.2688
ARIMA(1,0,1) with zero mean : 869.3584
ARIMA(0,0,0) with zero mean : 869.9485
ARIMA(0,0,2) with zero mean : 869.1567
ARIMA(1,0,2) with zero mean : 871.3229
```

Best model: ARIMA(0,0,1) with zero mean

Jako nejvhodnější je doporučen model  $ARFIMA(0, d, 1)$  s parametry

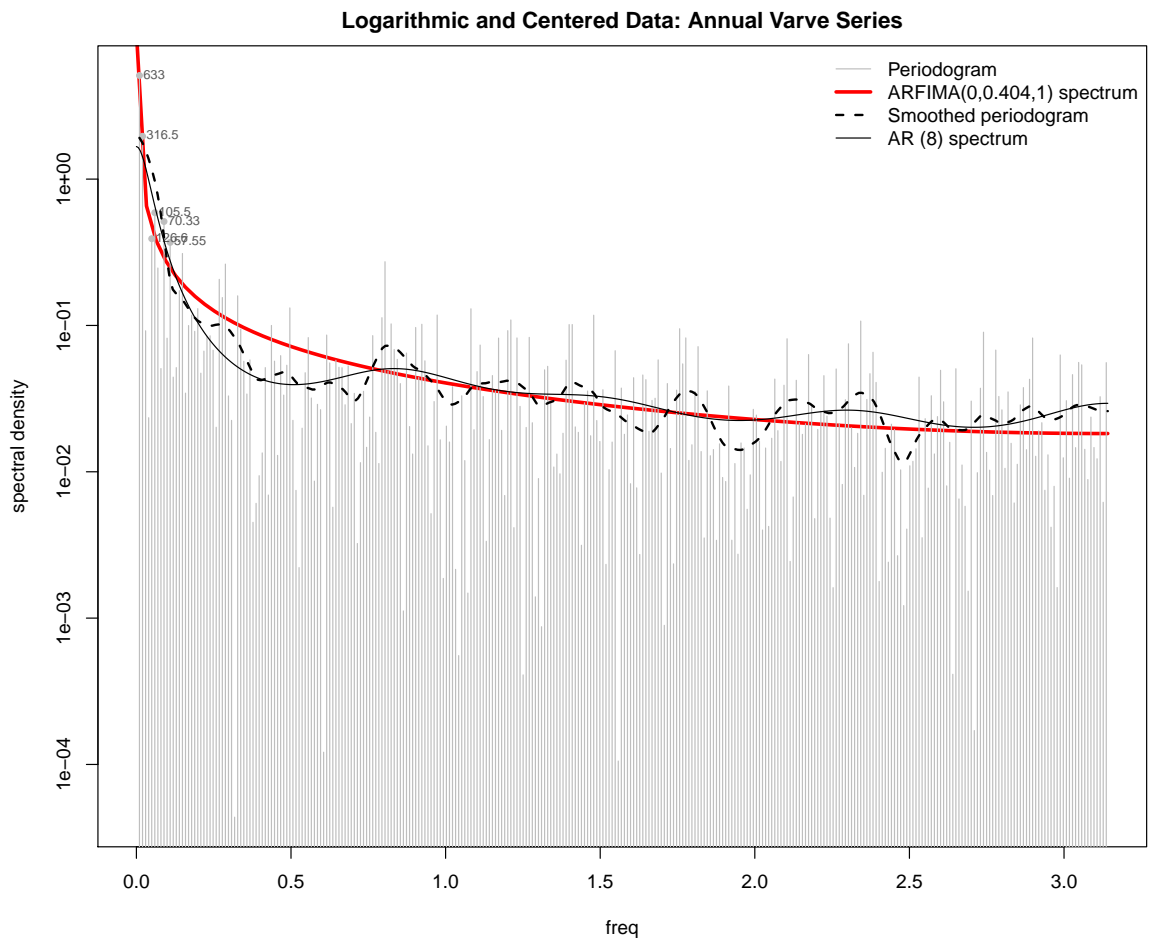
```
>fit_forecast$d
[1] 0.404032
>fit_forecast$ma
ma1
 0.06373206
```

Nyní porovnáme různé odhady spektrálních hustot s hustotou modelu  $ARFIMA(0, d, 1)$  danou vzorcem (2.16) pro odhadnuté parametry.

```
>fit<-fit_forecast
>NameModel<-paste("ARFIMA(",length(fit$ar),"",round(fit$d,3),"",
                  length(fit$ma), "")",sep="")
>TrigoTrend<-trigofit(lvarve)

>spARFIMA<-spectrum.arfima(d = FinalLong$d,ar=FinalLong$ar,
                          ma=FinalLong$ma,sd.innov=FinalLong$sigma)
>KERNEL<-kernel("modified.daniell", c(5,5))
>spec<-spectrum(lvarve,KERNEL,method="pgram",plot=FALSE,
                fast=TRUE,detrend=FALSE,log="y")
>spAR<-spec.ar(lvarve,plot=FALSE)

>pdf("LogVarveSP_arfima.pdf",width=10,height=8.25)
>op<-par(mar=c(4,4,2,0) + 0.25)
>Ylim<-range(c(TrigoTrend$periodogram,spec$spec/(2*pi)))
>plot(spARFIMA,xlim=c(0.00125, pi),log="y",col="red",lwd=3,ylim=Ylim,
      xlab="freq",ylab="spectral density",main=TEXT)
>with(TrigoTrend,lines(freq,periodogram,type="h",log="y",col="gray"))
>with(TrigoTrend,points(freq[signL],periodogram[signL],
                       pch=20,col="gray"))
>with(TrigoTrend,text(freq[signL],periodogram[signL],col="gray35",
                     round((2*pi)/freq[signL],2),pos=4,offset=0.175,cex=0.65))
>lines(2*pi*spec$freq,spec$spec/(2*pi),col="black",lwd=2,lty=2)
>lines(2*pi*spAR$freq,spAR$spec/(2*pi),col="black",lwd=1,lty=1)
>legend("topright",cex=1,
       legend = c("Periodogram",paste(NameModel,"spectrum"),
                 "Smoothed periodogram",spAR$method),
       bty="n",lty = c(1,1,2,1), xjust = 1, yjust = 1,lwd=c(1,3,2,1),
       col=c("gray","red","black","black"))
>par(op)
>dev.off()
```

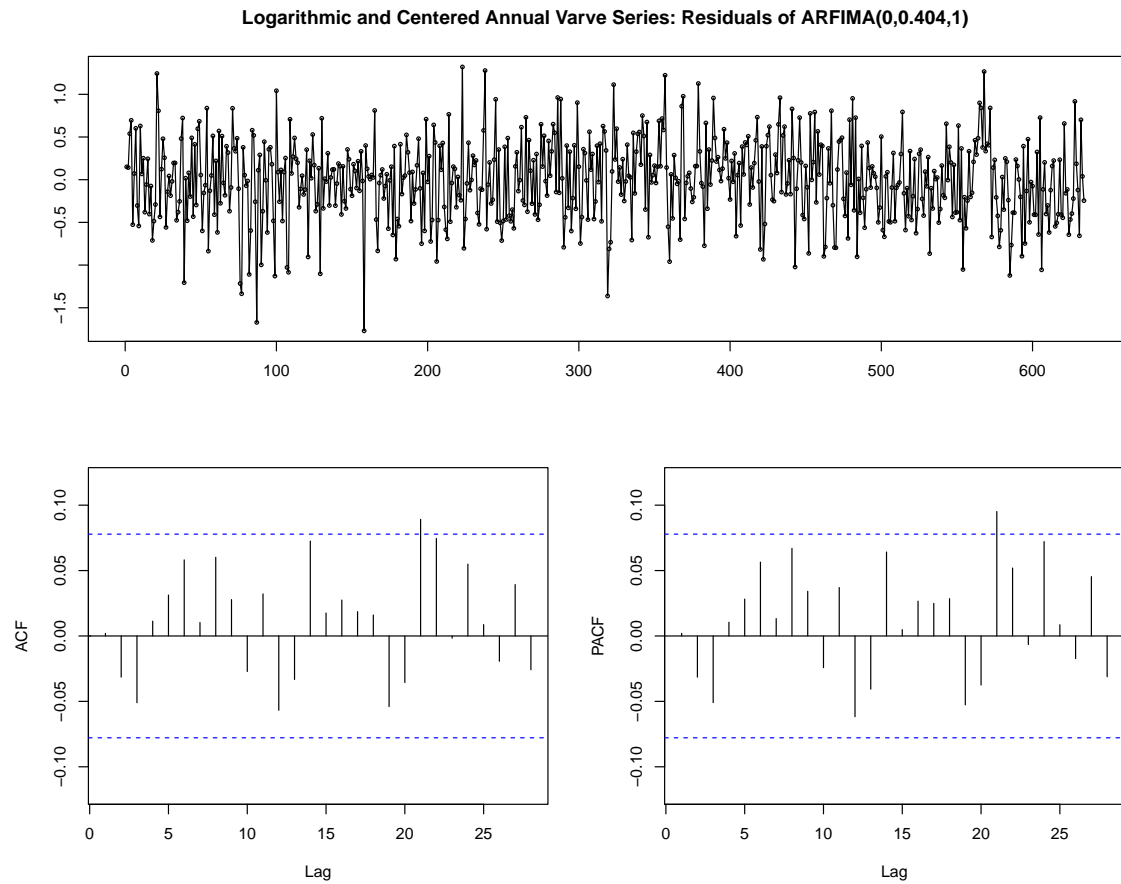


Z grafu je vidět, že u analyzovaných dat je výrazný fenomén dlouhé paměti, neboť různé metody odhadu spektrální hustoty se příliš neliší od teoretické hustoty ARFIMA modelu pro odhadnuté parametry. Jako v předchozích modelech i nyní na závěr provedeme grafickou analýzu reziduí ARFIMA modelu.

```

>res<-residuals(fit)

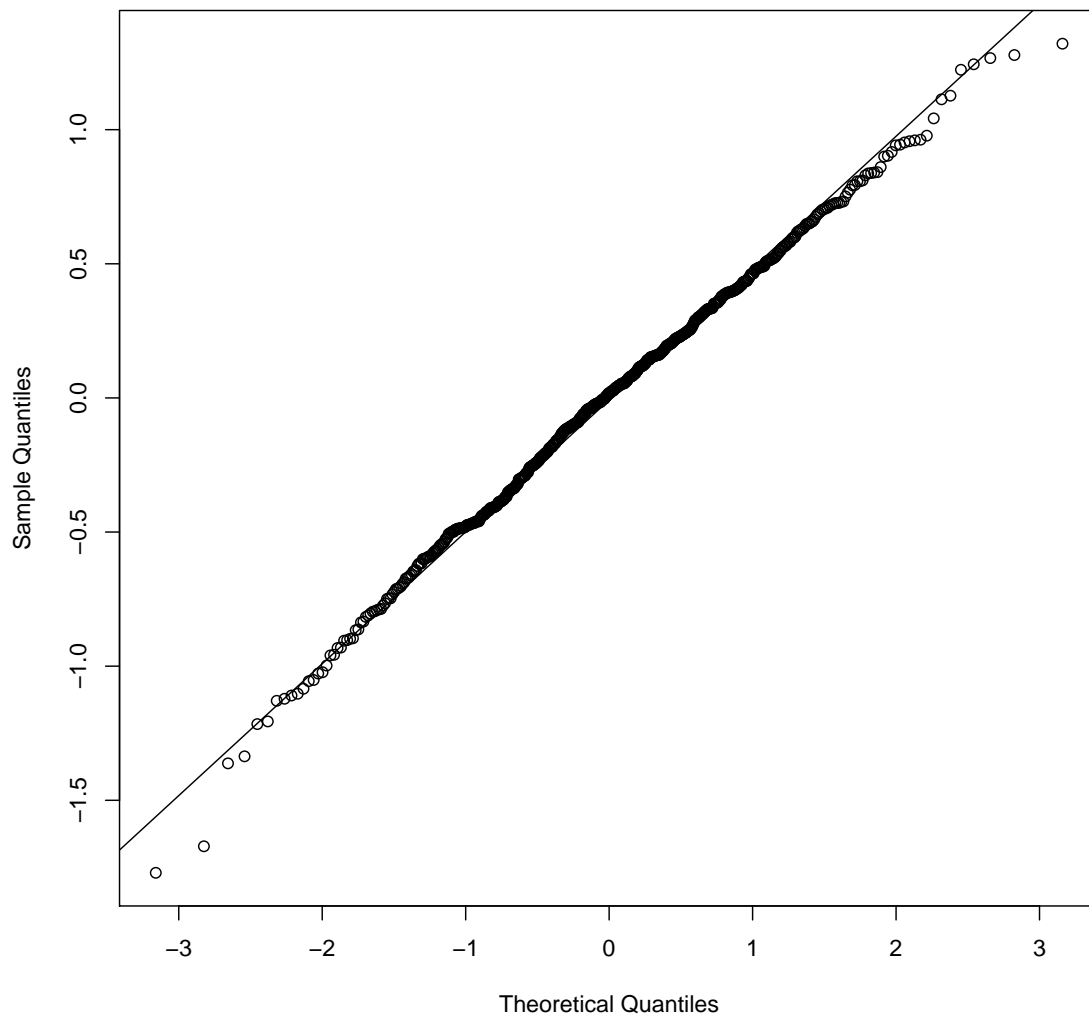
>Txt<-paste("Logarithmic and Centered Annual Varve Series:
            Residuals of", NameModel)
>pdf("LogVarve-ARFIMA-forecast.pdf",width=10,height=8)
>forecast::tsdisplay(res,main=Txt)
>dev.off()
    
```



```

>Txt<-paste("Log and Centered Annual Varve Series: Residuals of",
            NameModel,"\nNormal Q-Q Plot")
>pdf("LogVarve-ARFIMA-forecast-qq.pdf",width=8.25,height=8.25)
>qqnorm(res,main=Txt)
>qqline(res)
>dev.off()
    
```

**Log and Centered Annual Varve Series: Residuals of ARFIMA(0,0.404,1)  
Normal Q-Q Plot**



Na základě předchozích dvou grafů lze říci, že rezidua by mohla být bílým šumem. Pro kontrolu, zda jde skutečně o bílý šum, na tato rezidua použijeme ještě funkci `auto.arima()`.

```
>fit.res<-auto.arima(res,trace=TRUE)
```

```
ARIMA(2,0,2) with non-zero mean : 866.4524
ARIMA(0,0,0) with non-zero mean : 868.3551
ARIMA(1,0,0) with non-zero mean : 871.267
ARIMA(0,0,1) with non-zero mean : 870.3714
ARIMA(1,0,2) with non-zero mean : 874.0519
ARIMA(3,0,2) with non-zero mean : 875.5106
ARIMA(2,0,1) with non-zero mean : 874.6936
```

```
ARIMA(2,0,3) with non-zero mean : 873.9958
ARIMA(1,0,1) with non-zero mean : 873.2925
ARIMA(3,0,3) with non-zero mean : Inf
ARIMA(2,0,2) with zero mean      : Inf
```

Best model: ARIMA(2,0,2) with non-zero mean

Avšak výsledek není příliš potěšující, neboť místo očekávaného *ARIMA*(0,0,0) je doporučen model *ARIMA*(2,0,2).

Vyzkoušíme tedy ještě nejjednodušší model *ARFIMA*(0,*d*,0), což je *FDWN* model. K odhadu použijeme rovnou funkci `fracdiff()` z balíčku `fracdiff`.

```
>fit_fracdiff<-fracdiff(lvarve,nar=0,nma=0,M=30)
>summary(fit_fracdiff)
Call:
  fracdiff(x = lvarve, nar = 0, nma = 0, M = 30)
```

Coefficients:

```
Estimate Std. Error z value Pr(>|z|)
d 3.842e-01 4.590e-06 83706 <2e-16 ***
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
sigma[eps] = 0.479934
[d.tol = 0.0001221, M = 30, h = 4.591e-06]
Log likelihood: -434.7 ==> AIC = 873.491 [2 deg.freedom]
```

Zajímají nás především parametry *d* a  $\sigma^2$  :

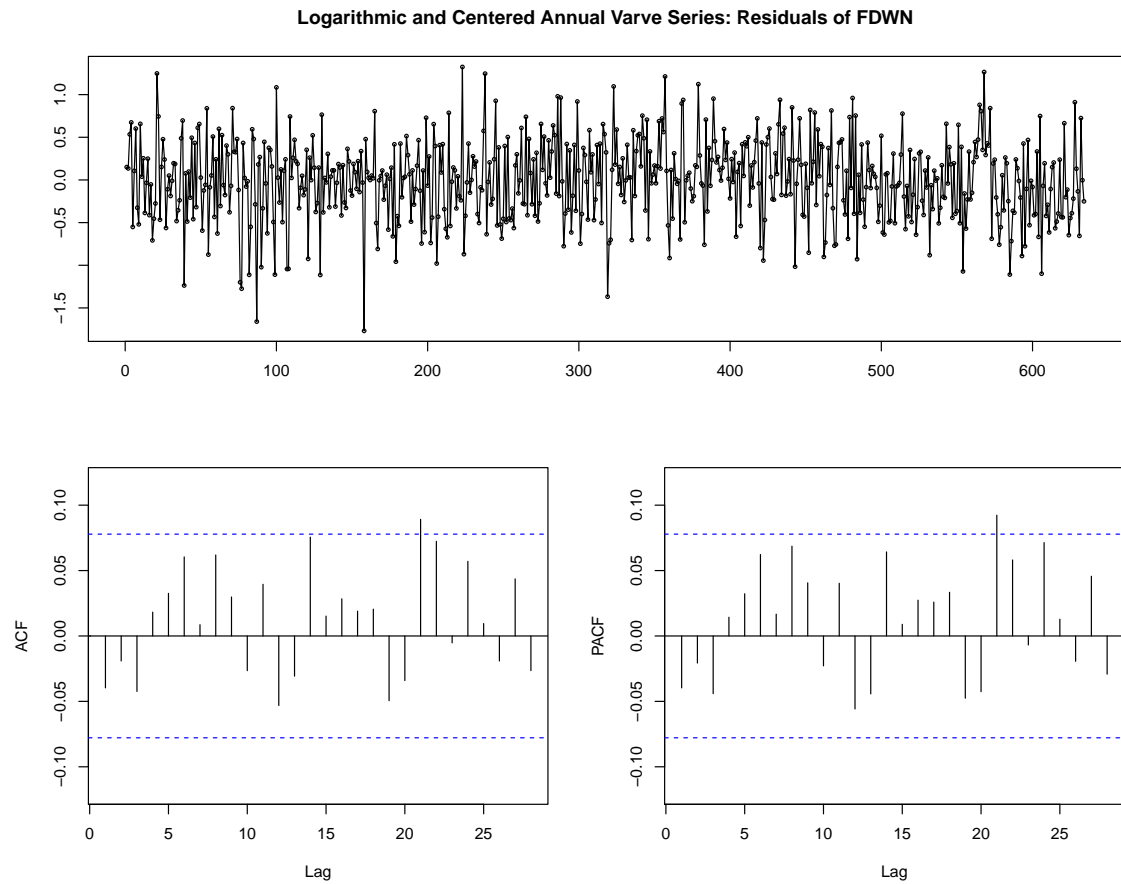
```
>fit_fracdiff$d
[1] 0.3841688
```

```
>fit_fracdiff$sigma^2
[1] 0.2303367
```

Opět provedeme grafickou analýzu reziduí nově uvažovaného modelu *FDWN*

```
>res<-residuals(fit_fracdiff)

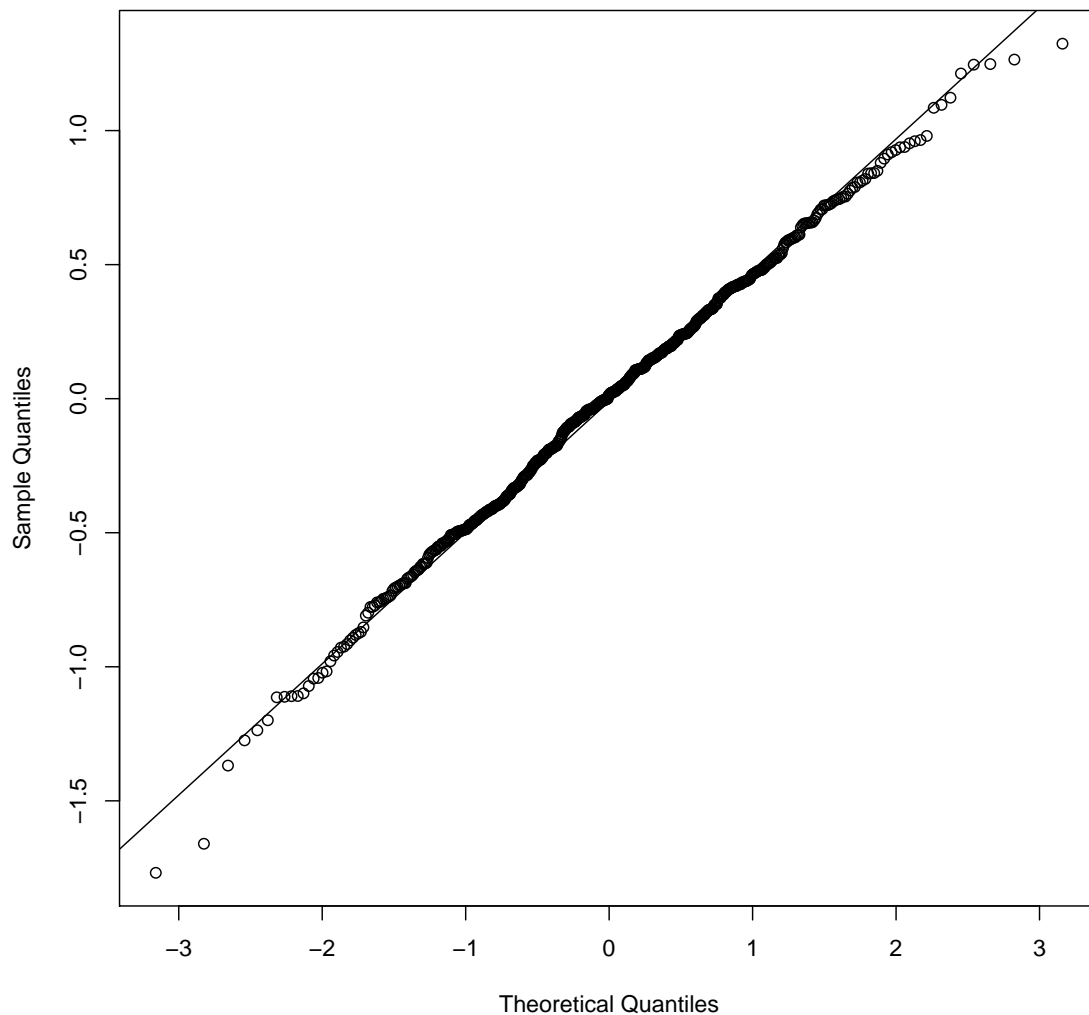
>Txt<-"Logarithmic and Centered Annual Varve Series:
  Residuals of FDWN"
>pdf("LogVarve-ARFIMA-fracdiff.pdf",width=10,height=8)
>forecast::tsdisplay(res,main=Txt)
>dev.off()
```



```

>Txt<-paste("Log and Centered Annual Varve Series: Residuals of FDWN",
            "\nNormal Q-Q Plot")
>pdf("LogVarve-ARFIMA-fractional-diff-qq.pdf",width=8.25,height=8.25)
>qqnorm(res,main=Txt)
>qqline(res)
>dev.off()
    
```

**Log and Centered Annual Varve Series: Residuals of FDWN  
Normal Q-Q Plot**



Na základě předchozích dvou grafů lze opět říci, že rezidua by mohla být bílým šumem. Pro kontrolu na tato rezidua použijeme ještě funkci `auto.arima()`.

```
>fit.res<-auto.arima(res,trace=TRUE)
```

```
ARIMA(2,0,2) with non-zero mean : 877.6048
ARIMA(0,0,0) with non-zero mean : 869.9346
ARIMA(1,0,0) with non-zero mean : 871.8495
ARIMA(0,0,1) with non-zero mean : 870.9089
ARIMA(1,0,1) with non-zero mean : 873.1227
ARIMA(0,0,0) with zero mean      : 867.978
ARIMA(1,0,0) with zero mean     : 869.8972
ARIMA(0,0,1) with zero mean     : 868.9509
```



ARIMA(1,0,1) with zero mean : 871.1856

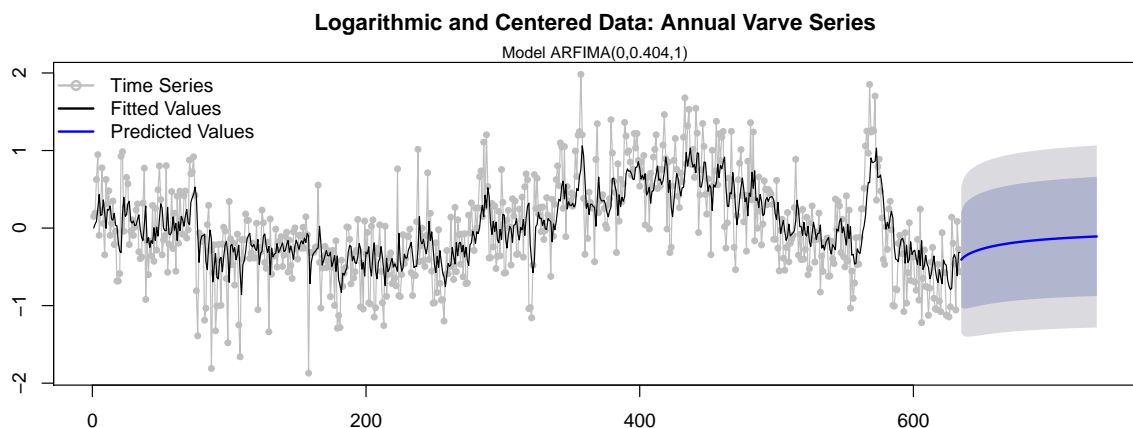
Best model: ARIMA(0,0,0) with zero mean

V tomto případě je již výsledek potěšující, neboť je doporučen očekávaný model *ARIMA(0,0,0)*.

Podívejme se ještě na predikce pomocí modelu *ARFIMA(0,0.404,1)* a FDWN modelu s parametrem  $d = 0,384$ .

```
>m<-100
>pred.forecast<-forecast(fit_forecast,h=m)
>pred.fracdiff<-forecast(fit_fracdiff,h=m)

>Txt<-"Logarithmic and Centered Annual Varve Series"
>pdf("LogVarve-ARFIMA-pred-forecast.pdf",width=10,height=4)
>op<-par(mar=c(2,2,3,0)+0.5)
>plot(pred.forecast,type="o", main=Txt,col="gray",pch=20,cex=0.75)
>lines(fit_forecast$fitted,col="black")
>mtext(paste("d =",round(FinalLong$d,3)),side=1,line=-1)
>legend("topleft",
       legend = c("Time Series","Fitted Values","Predicted Values"),
       bty="n",lty = c(1,1,1), xjust = 1, yjust = 1,lwd=2,
       col=c("gray","black","blue"),pch=c(1,NA,NA))
>mtext(paste("Model",NameModel),cex=0.85)
>par(op)
>dev.off()
```

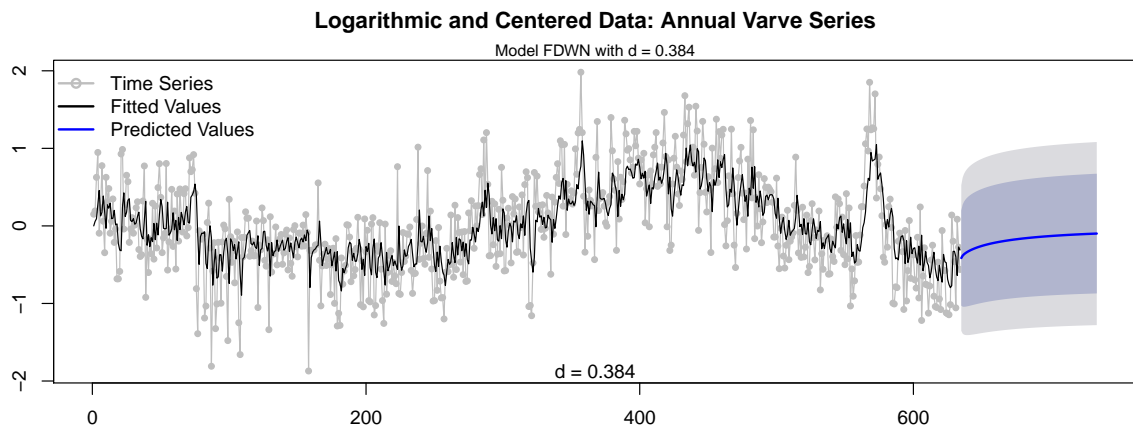


```
>Txt<-"Logarithmic and Centered Annual Varve Series"
>pdf("LogVarve-ARFIMA-pred-fracdiff.pdf",width=10,height=4)
>op<-par(mar=c(2,2,3,0)+0.5)
>plot(pred.fracdiff,type="o", main=Txt,col="gray",pch=20,cex=0.75)
```

```

>lines(fitted(fit_fracdiff),col="black")
>mtext(paste("d =",round(FinalLong$d,3)),side=1,line=-1)
>legend("topleft",
       legend = c("Time Series","Fitted Values","Predicted Values"),
       bty="n",lty = c(1,1,1), xjust = 1, yjust = 1,lwd=2,
       col=c("gray","black","blue"),pch=c(1,NA,NA))
>mtext(paste("Model FDWN with d =",round(fit_fracdiff$d,3)),cex=0.85)
>par(op)
>dev.off()

```



### 3.5 Srovnání s dalšími balíčky

Na závěr tohoto příkladu ještě uvedeme odhady ARFIMA modelů pomocí dalších balíčků, a to konkrétně

- `fracdiff` (autoři: Chris Fraley, Fritz Leisch, Valderio Reisen a Artur Lemonte) viz [11],
- `FGN` (autoři: A. I. McLeod a Justin Q. Veenstra) viz [18],
- `arfima` (autoři: Justin Q. Veenstra, A. I. McLeod) viz [24],
- `longmemo` (autoři: Jan Beran, Brandon Whitcher a Martin Maechler) viz [3],
- `afmtools` (autoři: Javier E. Contreras-Reyes, Georg M. Goerg, Wilfredo Palma) viz [4].

### 3.5.1 Balíček fracdiff

Velmi oblíbeným a nejčastěji používaným balíčkem pro modely s dlouhou pamětí je balíček `fracdiff`. Oficiálně se anglicky balíček nazývá *Fractionally differenced ARIMA aka ARFIMA(p,d,q) models*.

Odhady neznámých parametrů modelů jsou založeny především na metodě maximální věrohodnosti (MLE).

Snad právě díky oblíbenosti tohoto balíčku byla i do balíčku `forecast` zařazena řada funkcí, které s objekty třídy „arfima“ pracují.

Kromě funkce `fracdiff()` s MLE odhady neznámých parametrů (viz [13]) můžeme použít i funkci `fdGPH()` s Geweke a Porter-Hudakovými odhady založenými na regresních rovnicích využívajících periodogram.

```
>library(fracdiff)
>fit_fracdiff_fdGPH<-fdGPH(lvarve)
>(d_fdGPH<-fit_fracdiff_fdGPH$d)
[1] 0.4839232
```

Obdobný způsob odhadu nabízí další funkce, která na rozdíl od předchozí využívá vyhlazeného periodogramu.

```
>fit_fracdiff_fdSperio<-fdSperio(lvarve)
>(d_fdSperio<-fit_fracdiff_fdSperio$d)
[1] 0.5741439
```

### 3.5.2 Balíček FGN

Oficiální název balíčku je v angličtině *Fractional Gaussian Noise and power law decay time series model fitting*. Odhady parametrů modelů FGN (frakcionální gaussovský bílý šum) jsou založeny na exaktní MLE metodě a Whittleově MLE metodě.

```
>library(FGN)
>fit_FGN_FitFGN<-FitFGN(lvarve)
>summary(fit_FGN_FitFGN)
H = 0.815, R-sq = 34.2%
length of series = 634 , number of parameters = 2
mean = 1.93595071515089e-17 RMSE = 0.517085914542709
loglikelihood = 462.535 , AIC = -921.1 , BIC = -912.2
100 123.3 0.04953686
```

Parametr  $d$  vypočítáme ze vztahu  $d = H - 0.5$ , kde  $H$  je tzv. Hurstův koeficient.

```
>(d_FitFGN<-fit_FGN_FitFGN$H-0.5)
[1] 0.3147981

>(sigma2_FitFGN<-fit_FGN_FitFGN$sigmsqHat^2)
[1] 0.07149091
```

Opět se zaměříme na rezidua modelu.

```
>res<-residuals(fit_FGN_FitFGN)
>fit.res<-auto.arima(res,trace=TRUE)
ARIMA(2,1,2) with drift      : 1717.548
ARIMA(0,1,0) with drift     : 2191.179
ARIMA(1,1,0) with drift     : 1971.043
ARIMA(0,1,1) with drift     : 1708.968
ARIMA(1,1,1) with drift     : 1705.173
ARIMA(1,1,2) with drift     : 1706.5
ARIMA(1,1,1)                : 1703.572
ARIMA(0,1,1)                : 1707.406
ARIMA(2,1,1)                : 1719.87
ARIMA(1,1,0)                : 1969.025
ARIMA(1,1,2)                : 1704.87
ARIMA(0,1,0)                : 2189.166
ARIMA(2,1,2)                : 1716.074
```

Best model: ARIMA(1,1,1)

Tento model nedopadl nejlépe.

Nyní budeme uvažovat FDWN model, který se v balíčku FGN nazývá model typu FD.

Nejprve provedeme odhad parametrů modelu FDWN pomocí Whittleova přístupu.

```
>fit_FGN_warfima<-warfima(lvarve,order=c(0,0,0),lmodel="FD")
>(d_warfima<-fit_FGN_warfima$dHat)
[1] 0.3780139
```

Následně využijeme exaktní MLE odhad parametrů modelu FDWN.

```
>fit_FGN_earfima<-earfima(lvarve,order=c(0,0,0),lmodel="FD")
>(d_earfima<-fit_FGN_earfima$dHat)
[1] 0.3728782
```

### 3.5.3 Balíček arfima

Oficiální název balíčku je v angličtině *Fractional ARIMA Time Series Modeling*. Odhady parametrů modelů jsou založeny na exaktní MLE metodě.

Opět budeme uvažovat model FDWN, který se v tomto balíčku označuje jako typ „d“.

```
>library(arfima)
>fit_arfima<-arfima::arfima(lvarve,order=c(0,0,0),
                           lmodel="d",numeach=c(3,3))
>fit_arfima
Number of modes: 1
```

Call:

```
arfima::arfima(z = lvarve, order = c(0, 0, 0),
              numeach = c(3, 3), lmodel = "d")
```

Coefficients for fits:

	Coef.1:	SE.1:
d.f	0.372788	0.0273454
Fitted mean	-0.0365793	0.264639
logl	466.028	
sigma <sup>2</sup>	0.229718	

Starred fits are close to invertibility/stationarity boundaries

Opět se zaměříme na rezidua modelu.

```
>res<-residuals(fit_arfima)[[1]]
>fit.res<-auto.arima(res,trace=TRUE)
ARIMA(2,0,2) with non-zero mean : 878.2154
ARIMA(0,0,0) with non-zero mean : 869.6446
ARIMA(1,0,0) with non-zero mean : 872.0669
ARIMA(0,0,1) with non-zero mean : 871.1432
ARIMA(1,0,1) with non-zero mean : 873.6563
ARIMA(0,0,0) with zero mean      : 867.6675
ARIMA(1,0,0) with zero mean      : 870.09
ARIMA(0,0,1) with zero mean      : 869.1618
ARIMA(1,0,1) with zero mean      : 871.6852
```

Best model: ARIMA(0,0,0) with zero mean

Vzhledem k tomu, že odhad parametrů je velmi podobný odhadům získaným pomocí funkce `fracdiff()` z balíčku `fracdiff`, není překvapením, že se rezidua jeví obdobně jako model ARIMA(0,0,0).

### 3.5.4 Balíček `longmemo`

Oficiální název balíčku je v angličtině *Statistics for Long-Memory Processes (Jan Beran) – Data and Functions*.

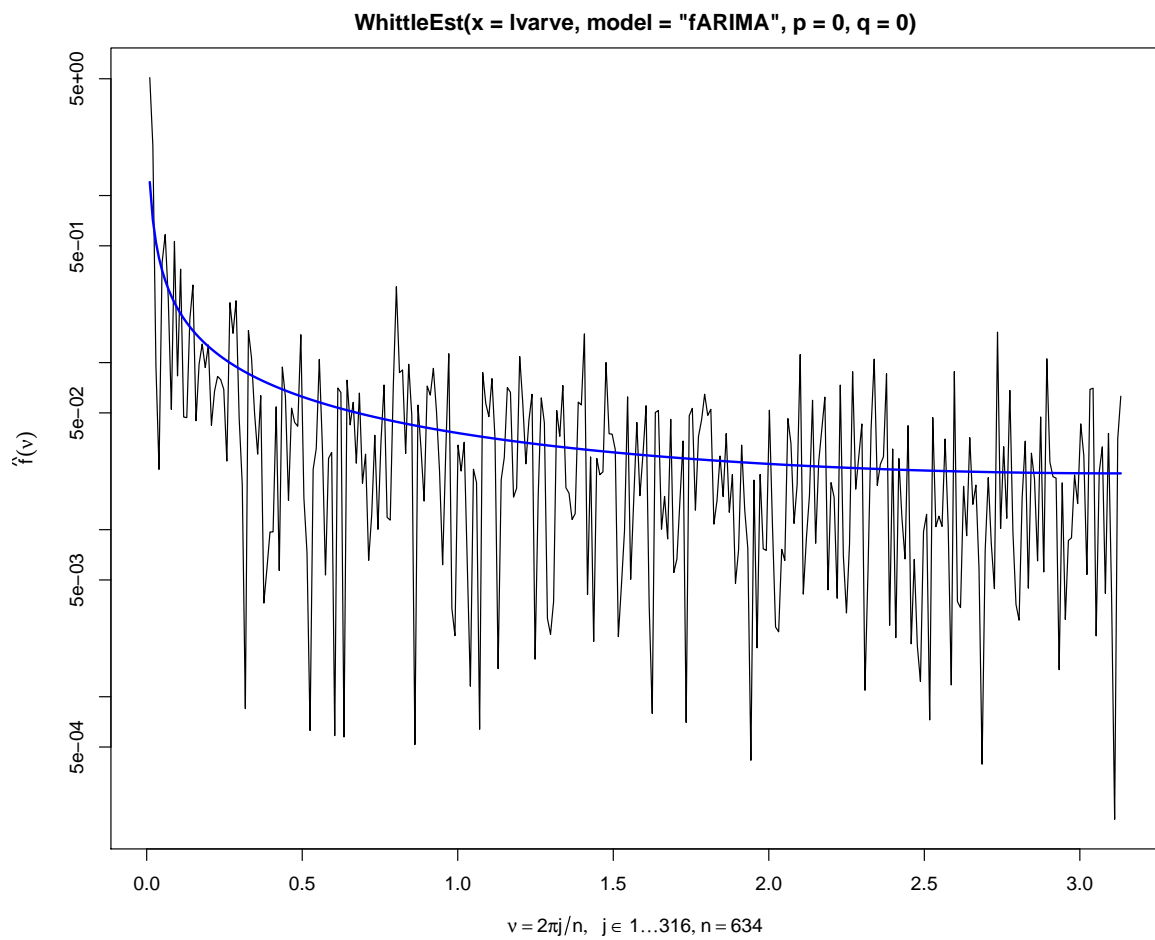
Odhady parametrů modelů jsou založeny na exaktní MLE metodě.

```
>library(longmemo)
>fit_longmemo<-WhittleEst(lvarve,model="fARIMA",p=0,q=0)
>fit_longmemo$coeff
      Estimate Std. Error  z value      Pr(>|z|)
H 0.8783005 0.03112879 28.21505 3.822219e-175

>(d_longmemo<-fit_longmemo$coeff[1]-0.5)
[1] 0.3783005
```

Protože tento balíček nabízí také grafické výstupy, které se týkají odhadů spektrální hustoty, využijeme jich.

```
>pdf("LogVarveSP_longmemo.pdf",width=10,height=8.25)
>op<-par(mar=c(4,4,2,0) + 0.25)
>plot(fit_longmemo,log="y")
>par(op)
>dev.off()
```



### 3.5.5 Balíček afmtools

Oficiální název balíčku je v angličtině *Estimation, Diagnostic and Forecasting Functions for ARFIMA models*. Odhad parametrů vychází z Whittleovy metody popsané v části 2.4.3.

```
>library(afmtools)
>fit_afmtools_arfima.whittle<-arfima.whittle(lvarve)
>print(fit_afmtools_arfima.whittle)

Call: arfima.whittle(series = lvarve)
Estimation method: Whittle

Parameter estimates:
      Estimate Std. Error  t value Pr(>|t|)
d  0.383836    0.065223    5.885 3.98e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

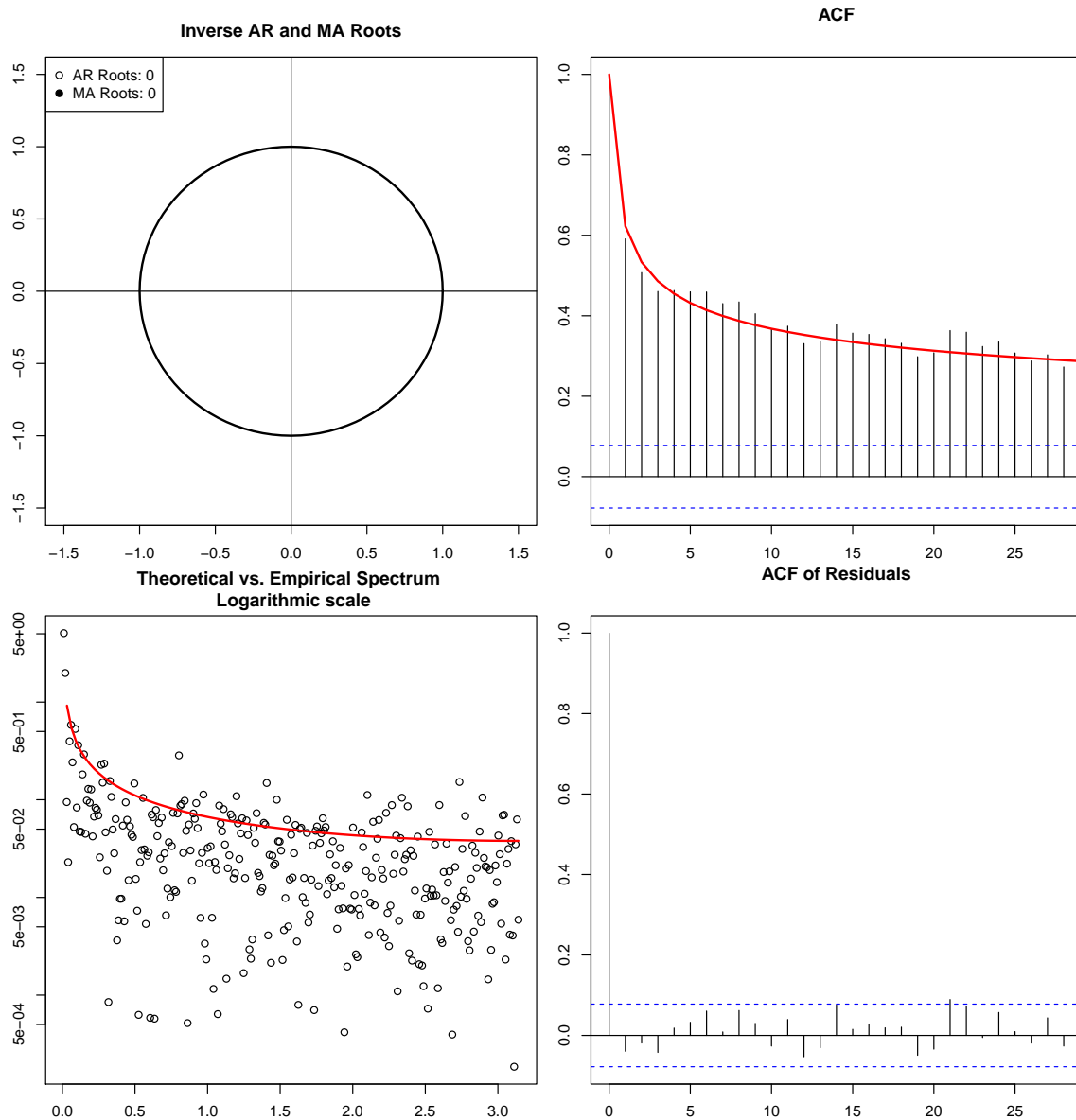
Estimated standard deviation of residuals
(ratio of periodogram and spectrum):
0.6369951

>(d_arfima.whittle<-fit_afmtools_arfima.whittle$d)
[1] 0.3838358

>(d_arfima.whittle<-fit_afmtools_arfima.whittle$sd.innov)

V balíčku je také velmi zajímavě zpracovaná funkce plot.arfima()

>pdf("LogVarveSP_afmtools.pdf",width=10,height=10.25)
>plot(fit_afmtools_arfima.whittle)
>dev.off()
```



Opět se zaměříme na rezidua modelu.

```
>res<-resid(fit_afmtools_arfima.whittle)
>fit.res<-auto.arima(res,trace=TRUE)
```

```
ARIMA(2,0,2) with non-zero mean : 875.3103
ARIMA(0,0,0) with non-zero mean : 869.9286
ARIMA(1,0,0) with non-zero mean : 871.8606
ARIMA(0,0,1) with non-zero mean : 870.9213
ARIMA(1,0,1) with non-zero mean : 873.1423
ARIMA(0,0,0) with zero mean      : 867.9721
ARIMA(1,0,0) with zero mean     : 869.9083
```



ARIMA(0,0,1) with zero mean : 868.9632

ARIMA(1,0,1) with zero mean : 871.2059

Best model: ARIMA(0,0,0) with zero mean

Také v tomto případě se rezidua modelu chovají jako bílý šum.

## Závěr

V první kapitole byla připomenuta teorie modelování jednorozměrných časových řad z [9]. V druhé kapitole byla uvedena definice dlouhé paměti, frakcionálně diferencovaného bílého šumu, jeho zobecnění na ARFIMA modely a jejich základní vlastnosti, tj. stacionarita, kauzalita, invertibilita a spektrální hustota. Na závěr kapitoly je popsán Whittleův odhad logaritmické věrohodnostní funkce. V třetí kapitole jsme na reálných zkoumaných datech ověřili, že předpovědi pomocí polynomickeho a periodického trendu pro časové řady vykazující dlouhou paměť se výrazně liší od modelů ARFIMA. V programovacím prostředí R jsme uvedli odhady parametrů  $p, d, q, \sigma^2$  pomocí funkcí z balíčků `forecast` a `fracdiff`. Na závěr kapitoly jsme provedli srovnání s balíčky `FGN`, `arfima`, `longmemo` a `afmtools`.

# Přílohy

Zde jsou uvedeny pomocné funkce užití v 3. kapitole. Pochází z výukových materiálů pro předmět *Praktika z náhodných procesů*, viz [10].

## Pomocná funkce tperiod

Tuto funkci vytvořila Iveta Hellebrandová jako součást své diplomové práce [14]. Funkce slouží k výpočtu periodogramu a určení významných period pomocí Fisherova a Siegelova testu.

```
tperiod <- function(x, figure=FALSE, log="y")
# Additive model: x(t)=y(t)+e(t) is assumed where
# y(t) ... deterministic component with periodicities to be detected
# e(t) ... error component, e: IID(0,sigma2)
# =====
# INPUT:
# x ... is an 1 x n vector of observations
# fig ... logical
#       figure handle for the periodogram with emphasized
#       significant harmonics
#       FALSE=default=no plot
# OUTPUT:
# list containing the following components:
# I ... periodogram: vector of length n2
#       n2=trunc((n-1)/2)
# kF ... indices of significant harmonics using Fisher's test
# kS ... indices of significant harmonics using Siegel's test

# Iveta Hellebrandova, 2011
# according to the MATLAB function "tperiod":
# (C) 1997-1999, V. Vesely, Masaryk University of Brno
# (Czech Republic) Rev. 1, Dec. 22, 1999

{
if (min(dim(as.matrix(x)))>1) stop('x is not a vector')
n<-length(x)
m<-trunc((n-1)/2)           # length of the periodogram
```

```
if (m<5) stop('x must be longer than 10')
m<-min(m,1000)
I<-fft(x)
I<-(2/n)*abs(I[2:(m+1)])^2 # Computing the periodogram
II<-I
lF<-numeric(); lS<-numeric(); k<-numeric()
# source('fisher.r')
while (m>4) { # loop to detect significant periodicities
  f <- fisher(m) # Fisher's and Siegel's 95# percentile
  gf <- f$pfisher
  tsi <- f$psiegel
  Y <- II/sum(II)
  W <- max(Y) # Fisher's statistic
  kmax <- which.max(Y)
  Ts<-sum(pmax(0, Y - 0.6*gf)) # Siegel's statistic
  if (length(lF)==0 & W<=gf) lF <- length(k)
  if (length(lS)==0 & Ts<=tsi) lS <- length(k)
  if (W>gf | Ts>tsi) {
    k <- c(k,kmax[1])
    II[kmax[1]] <- 0 # scratch component just detected
    m <- m-1
  } else {
    break
  }
}

if (figure) {
  plot(I,log=log)
  lines(I, type='h', lty='dashed')
  points(k,I[k],pch=8)
  text(k,I[k],round(n/k,1), pos=4)
  abline(h=0)
  if (lS>lF)
    title("PERIODOGRAM (Siegel's test)") else
    title("PERIODOGRAM (Fisher's test)")
}
kF <- k[1:lF]
kS <- k[1:lS]
return(list(I=I, kF=kF, kS=kS))
}
```

## Pomocná funkce fisher

Tuto funkci vytvořila Iveta Hellebrandová jako součást své diplomové práce [14].

```
fisher <- function(m)
# 95# PERCENTILES FOR THE FISHER'S AND SIEGEL'S TEST OF PERIODICITY
# =====
# INPUT:
# m ... vector of lengths of the periodogram, 5<=m[j]<=1000
# OUTPUT:
# list containing the following components:
# pfisher... pfisher[j]=Fisher's 95% percentile for m[j]
# psiegel... psiegel[j]=Siegel's 95% percentile for m[j]

# Iveta Hellebrandova, 2011
# according to the MATLAB function "fisher":
# (C) 1997-1999, V. Vesely, Masaryk University of Brno
# (Czech Republic) Rev. 1, Dec. 22, 1999

{
if (any(m<5)) stop('Periodogram must have length at least 5')
if (any(m>1000)) stop('The periodogram length must not exceed 1000')
GF <- rbind(c(0.6837722340, 0.274),
            c(0.4449525606, 0.1812916218),
            c(0.3346119552, 0.1401751436),
            c(0.2704042202, 0.1159434232),
            c(0.2280456085, 0.0997325432),
            c(0.1978406246, 0.0880198043),
            c(0.1751294860, 0.0791060851),
            c(0.1573827449, 0.0720634801),
            c(0.1431035999, 0.0663391976),
            c(0.1313472721, 0.0615819252),
            c(0.1130887896, 0.0541010364),
            c(0.0995257984, 0.0484581309),
            c(0.0890258722, 0.0440297152),
            c(0.0806402063, 0.0404493552),
            c(0.0737781851, 0.0374866267),
            c(0.0522136336, 0.0279346469),
            c(0.0407367370, 0.0226517317),
            c(0.0335542444, 0.0192460210),
            c(0.0286124993, 0.0168450268),
            c(0.0249931178, 0.0150496264),
            c(0.0222217969, 0.0136497353),
            c(0.0200280827, 0.0125235715),
            c(0.0182461407, 0.0115953645),
            c(0.0155221997, 0.0101493846),
            c(0.0135334433, 0.0090692380),
            c(0.0120144724, 0.0082276189),
            c(0.0108145470, 0.0075508836),
```

```
c(0.0098415283, 0.0069932836))

mGF <- c(seq(5,50,5),seq(60,100,10),seq(150,500,50),seq(600,1000,100))
pfisher <- spline(mGF,GF[,1],xout=m)$y
psiegel <- spline(mGF,GF[,2],xout=m)$y
return(list(pfisher=pfisher, psiegel=psiegel))
}
```

## Pomocná funkce damsleth

V této funkci je naprogramována iterativní metoda nalezení skrytých period pomocí nelineární regrese (Gauss-Newtonova metoda). Počáteční hodnoty neznámých parametrů se určují pomocí Fisherova testu periodogramu.

```
damsleth<-function(X,pstlevel=0.05)
{
  if (min(dim(as.matrix(X)))>1) stop('X is not a vector')
  n<-length(X);
  TIME<-1:n
  m<-as.integer(n/2)
  Y<-X
  aa<-NULL;bb<-NULL;freq<-NULL
  konec<-FALSE
  krok<-1
  while (!konec & krok<=m)
  {
    # krok 1
    # pomocí Fisherova testu se určí prvotní přiblížení
    beta0<-fishper(Y,pstlevel)
    # krok 2
    # nelineární regrese
    if (is.null(beta0))
      konec<-TRUE
    else {
      modelnls<-nls(Y~trigpol1(TIME,A,B,freq),
                    start=list(A=beta0$A,B=beta0$B,freq=beta0$freq))
      aa<-c(aa,coef(modelnls)[1])
      bb<-c(bb,coef(modelnls)[2])
      freq<-c(freq,coef(modelnls)[3])
      Y<-resid(modelnls)
      krok<-krok+1
    }
  }
  return(list(A=aa,B=bb,freq=freq))
}
```

## Pomocná funkce fishper

funkce pro určení nejvíce signifikantní periody-Fisherův test, výpočet periodogramu

```

fishper<-function(XX,pstlevel=0.05)
{
# vytvoření omega(k)=(2*pi/n)*k k-1,...,n/2

if (min(dim(as.matrix(XX)))>1) stop('X is not a vector')
n<-length(XX)
if (n %% 2==0)
  {X<-XX[-1]
  n<-n-1
  } else
  X<-XX

m<-as.integer(n/2)

omega<-matrix((1:m),ncol=1)*2*pi/n
#
# a(k)=sqrt(2/n)*suma X(t)*cos(omega(k)*t)
#
a<-sqrt(2/n)*cos(kronecker(matrix((1:n),nrow=1),omega))%%X
#
# b(k)=sqrt(2/n)*suma X(t)*sin(omega(k)*t)
#
b<-sqrt(2/n)*sin(kronecker(matrix((1:n),nrow=1),omega))%%X
I<-1/(4*pi)*(a^2+b^2)
a<-a*sqrt(2/n)
b<-b*sqrt(2/n)
# Fisherův test pro maximální hodnotu periodogramu
Vo<-max(I)
ind<-which.max(I)
sumV<-sum(I)
pstG<-0
Wo<-Vo/sumV
# Pst(W>Wo)=m*(1-1*Wo)^(m-1)+(m nad 2)*(1-2*Wo)^(m-1)+....
# sečítá se tak dlouho, dokud (1-k*Wo)>0
k<-1
pstG<-0;
sign<-1
while (k*Wo<1)
  {pstG<-pstG+sign*comnum(m,k)*(1-k*Wo)^(m-1)
  k<-k+1
  sign<-sign*(-1)
  }
}

```

```

ifelse(pstG<=pstlevel,
      return(list(A=a[ind],B=b[ind],freq=omega[ind])),
      return(NULL))
}

```

## Pomocná funkce na výpočet kombinačního čísla

kombinační číslo  $\binom{n}{k} = (n/1) * ((n-1)/2) * \dots * ((n-k+1)/k)$

```

comnum<-function(n,k)
comb<-prod(seq(n,n-k+1,by=-1)/(1:k))

```

## Pomocná funkce trigpol1

funkce určená k výpočtu trigonometrického polynomu s jednou frekvencí

```

# beta0=list(A, B, freq)
# y(t)=A*cos(TIME*freq)+B*sin(TIME*freq)
trigpol1<-function(TIME,A,B,freq)
{
y<-A*cos(freq*TIME)+B*sin(freq*TIME)
}

```

## Pomocná funkce trigoval

trigonometrický polynom z významných period

```

trigoval<-function(TIME,A,B,freq)
#function Y=trigoval(T,aa,bb,freq)
#
#           p
# trig(t)= sum(aa(i)*cos((freq(i)*t)+bb(i)*sin((freq(i)*t)
#           i=1
{
if (min(dim(as.matrix(TIME)))>1) stop('TIME is not a vector')
Y<-cos(kronecker(matrix(TIME,ncol=1),
                    matrix(freq,nrow=1)))*%*%matrix(A,ncol=1)+
sin(kronecker(matrix(TIME,ncol=1),
                matrix(freq,nrow=1)))*%*%matrix(B,ncol=1)
}

```



## Pomocná funkce trigofit

funkce pro nalezení amplitud a fází skrytých period, výpočet periodogramu, vytvoření  $\omega(k) = (2 * \pi/n) * k, k = 1, \dots, n/2$ .

```
trigofit<-function(XX,pstlevel=0.05)
# function [I,omega,aa,bb,freq,amplituda,faze,L,a,b]=
#                                     =trigofit(XX,pstlevel)
{
if (min(dim(as.matrix(XX)))>1) stop('X is not a vector')
n<-length(XX)
if (n %% 2==0)
  {X<-XX[-1]
  n<-n-1
  } else
  X<-XX
m<-as.integer(n/2)
omega<-matrix((1:m),ncol=1)*2*pi/n
L<-rep(FALSE,m)
a<-sqrt(2/n)*cos(kronecker(matrix((1:n),nrow=1),omega))%%X
b<-sqrt(2/n)*sin(kronecker(matrix((1:n),nrow=1),omega))%%X
I<-1/(4*pi)*(a^2+b^2)
a<-a*sqrt(2/n)
b<-b*sqrt(2/n)
amplituda=sqrt(a^2+b^2)*sqrt(2/n)
faze=tan(b/a)
```

## Metoda skrytých period

```
o<-order(I)
V<-I[o]
sumV<-sum(V)
i<-m
pstG<-0
pocet<-0
aa<-NULL
bb<-NULL
freq<-NULL
while ((pstG<=pstlevel)&(i<=m))
  {
  Wo<-V[i]/sumV
  sumV<-sumV-V[i]
  # Pst(W>Wo)=m*(1-1*Wo)^(m-1)+(m nad 2)*(1-2*Wo)^(m-1)+...
  # sečítá se tak dlouho, dokud (1-k*Wo)>0
  k<-1
```

```
pstG<-0
sign<-1
while (k*Wo<1)
{
  pstG<-pstG+sign*comnum(m,k)*(1-k*Wo)^(m-1);
  k<-k+1;
  sign<-sign*(-1);
}
if (pstG<=pstlevel)
{
  L[o[i]]=TRUE
  aa<-c(aa,a[o[i]])
  bb<-c(bb,b[o[i]])
  freq=c(freq,omega[o[i]])
  pocet<-pocet+1
}
i<-i-1
}
return(list(periodogram=as.vector(I),freq=as.vector(omega),signA=aa,
  signB=bb,signFreq=freq,amplituda=as.vector(amplituda),
  faze=as.vector(faze),signL=L,A=as.vector(a),
  B=as.vector(b)))
}
```

## Pomocná funkce trigpol2

funkce pro vytvoření 2 sloupců trigonometrického polynomu s jednou frekvencí

```
trigpol2<-function(TIME,A,B,freq)
{
return(cbind(A*cos(freq*TIME),B*sin(freq*TIME)))
}
```

# Seznam použité literatury

- [1] J. ANDĚL, *Long memory time series models*, Kybernetika, Volume **22** (1986), č. 2, str. 105-123.
- [2] J. ANDĚL, *Statistische Analyse von Zeitreihen*, Akademie-Verlag, Berlin 1984.
- [3] J. BERAN, B. WHITCHER a M. MAECHLER, **longmemo**: *Statistics for Long-Memory Processes (Jan Beran) – Data and Functions*, 2011. Dostupný z URL: <http://cran.r-project.org/web/packages/longmemo/longmemo.pdf>.
- [4] J. E. CONTRERAS-REYES, M. GEORG a W. PALMA, **amftools**: *Estimation, Diagnostic and Forecasting Functions for ARFIMA models*, 2011. Dostupný z URL: <http://www.icesi.edu.co/CRAN/web/packages/amftools/amftools.pdf>.
- [5] J. E. CONTRERAS-REYES a W. PALMA, *Statistical analysis of autoregressive fractionally integrated moving average models in R*, Computational Statistics, Springer-Verlag Berlin Heidelberg, Volume **28** (2013), č. 5, str. 2309-2331.
- [6] E. DAMSLETH a E. SPJØTVOLL, *Estimation of Trigonometric Components in Time Series*, Journal of the American Statistical Association, Volume **77** (1982), č. 378, str. 381-387.
- [7] F. X. DIEBOLD a G. D. RUDEBUSCH, *Long memory and persistence in aggregate output*, Journal of Monetary Economics **24** (1989), č. 2, str. 189-209.
- [8] R. A. FISHER, *Tests of significance in harmonic analysis*, Proc. Roy. Soc. A, **125** (1929), str. 54-59.
- [9] M. FORBELSKÁ, *Stochastické modelování jednorozměrných časových řad*, 1. vyd. Brno: Masarykova univerzita, 2009. 4761/Př-3/09-17/31. ISBN 978-80-210-4812-6.
- [10] M. FORBELSKÁ, *Praktikum z náhodných procesů*, studijní materiály předmětu M0130, Masarykova univerzita, [cit. 4. 1. 2015]. Dostupný z URL: <https://www.math.muni.cz/~kolacek/docs/frvs/M0130/skripty/FunkceM0130.R>
- [11] C. FRALEY, F. LEISCH a M. MAECHLER, **fracdiff**: *Fractionally Differenced ARIMA aka ARFIMA(p,d,q) Models*, 2012. Dostupný z URL: <http://cran.r-project.org/web/packages/fracdiff/fracdiff.pdf>.

- [12] J. GEWEKE a S. PORTER-HUDAK, *The Estimation and Application of Long Memory Time Series Models*, Journal of Time Series Analysis, Volume 4 (1983), č. 4, str. 221-238.
- [13] J. HASLETT a A. E. RAFTERY, *Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource (with Discussion)*, Applied Statistics 38 (1989), str. 1-50.
- [14] I. HELLEBRANDOVÁ, *Stochastické modelování časových řad pomocí strukturálních modelů*, diplomová práce. Brno: Masarykova univerzita, 2011.
- [15] J. R. M. HOSKING, *Fractional differencing*, Biometrika 68 (1981), č. 1, str. 165-176.
- [16] H. HURST, *Long Term Storage Capacity of Reservoirs*, Transactions of the American Society of Civil Engineers 116 (1951), str. 770-799.
- [17] R. J. HYNDMAN, G. ATHANASOPOULOS, S. RAZBASH, D. SCHMIDT, Z. ZHOU, Y. KHAN, C. BERGMEIR a E. WANG, **forecast**: *Forecasting functions for time series and linear models*, 2014. Dostupný z URL: <http://cran.r-project.org/web/packages/forecast/forecast.pdf>.
- [18] A. I. McLEOD a J. Q. VEENSTRA, **FGN**: *Fractional Gaussian Noise and power law decay time series modeling*, 2014. Dostupný z URL: <http://cran.r-project.org/web/packages/FGN/FGN.pdf>.
- [19] W. PALMA, *Long-Memory Time Series - Theory and Methods*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2007. ISBN 978-0-470-1 1402-5.
- [20] J. PETRÁNEK, *Geologická encyklopedie*, Česká geologická služba, 2007. [cit. 4. 1. 2015]. Dostupný z URL: <http://www.geology.cz/aplikace/encyklopedie/term.pl?varva>
- [21] R Development Core Team. *R: A language and environment for statistical computing* [online]. R Foundation for Statistical Computing, Vídeň, 2012 [cit. 4. 1. 2015]. ISBN 3-900051-07-0. Dostupné z URL: <http://www.r-project.org/>.
- [22] H. W. RUST, *Detection of Long-Range Dependence – Applications in Climatology and Hydrology*, disertační práce. Univerzita Postupim, Postupim, 2007.
- [23] D. STOFFER **astsa**: *Applied Statistical Time Series Analysis*, 2014. Dostupný z URL: <http://cran.r-project.org/web/packages/astsa/astsa.pdf>.
- [24] J. Q. VEENSTRA a A. I. McLEOD, **arfima**: *Fractional ARIMA Time Series Modeling*, 2014. Dostupný z URL: <http://cran.r-project.org/web/packages/arfima/arfima.pdf>.