

ÚVOD DO ČASOVÝCH ŘAD¹

VÍTEZSLAV VESELÝ

Katedra aplikované matematiky a informatiky

Ekonomicko-správní fakulta MU v Brně

Email: vesely@econ.muni.cz

OBSAH

- 1 Úvod
- 2 Základní pojmy
 - 2.1 Definice časové řady
 - 2.2 Systém distribučních funkcí
 - 2.3 Momentové charakteristiky (střední hodnota a autokovarianční funkce)
 - 2.4 Striktní a slabá stacionarita
 - 2.5 Lineární transformace časových řad
 - 2.6 Odhady momentových funkcí
3. Predikce v časových řadách
 - 3.1 Princip ortogonální projekce
 - 3.2 Durbin-Levinsonův a inovační algoritmus.
- 4 Modelování a odhad parametrů
 - 4.1 Předzpracování
 - Ošetření specifických efektů
 - Stabilizace rozptylu
 - Identifikace periodických komponent
 - 4.2 Dekompoziční model
 - Užití lineárního regresního modelu

Date: 24.listopad 2003, korekce 9.2.2009.

¹sborník semináře ANALÝZA DAT, listopad 2003 (TriloByte s r.o.)

- Filtrační techniky

4.3 Box-Jenkinsovy modely

- ARMA modely pro stacionární časové řady
- ARIMA modely pro kovariančně stacionární časové řady s trendem
- SARIMA modely pro kovariančně stacionární časové řady s trendem a sezónní komponentou.

5 Demonstrační ukázky pomocí knihovny TSA-M

Časová řada: = soubor pozorování náhodných veličin $\{x_t, t \in T\}$, kde t je zpravidla čas a $T \subseteq \mathbb{R}$.

1.1. UKÁZKY ČASOVÝCH ŘAD.

Obr. 1.1:

Měření proudu procházejícího odporem r v obvodu se střídavým napětím,
 $v(t) = a \cos(vt + \Theta)$, tj. $x_t = \frac{a}{r} \cos(vt + \Theta)$, kde a a Θ jsou zatíženy náhodnou chybou.

Obr. 1.2:

Růst populace v USA v letech 1790–1980 sledovaný v deseti-letých intervalech.

Obr. 1.3:

Počet stávek v USA v letech 1951–1980.

Obr. 1.4:

Měsíční počty tisíců cestujících v mezinárodní letecké dopravě v letech 1949–1960.

Obr. 1.5:

Počty slunečních skvrn v letech 1770–1869.

Obr. 1.6:

Počet úmrtí při nehodách v USA v letech 1973–1978.

1.2. OBLASTI UPLATNĚNÍ METOD ANALÝZY ČASOVÝCH ŘAD.

fyzika, technika:

- seismický záznam v geofyzice.
- řada nejvyšších denních teplot v meteorologii.
- průběh výstupního signálu určitého elektrického přístroje.
- tenzometrické měření povrchového napětí v provozu namáhané strojní součástky.

biologie, ekologie: sledování různých parametrů znečištění ovzduší.

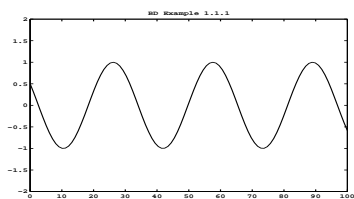
medicína: záznam EKG nebo EEG.

společenské vědy: změny v počtu a složení obyvatelstva.

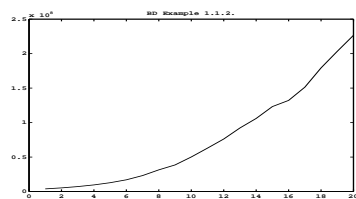
sociologie: vývoj rozvodovosti.

ekonomie: teorie časových řad = jedna z nejdůležitějších kvantitativních metod pro analýzu ekonomických dat, např.:

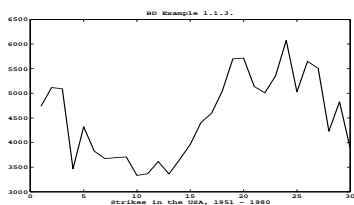
- analýza poptávky po určitém výrobku
- analýza objemu zemědělské produkce
- analýza počtu cestujících v letecké dopravě
- analýza vývoje kurzu akcií na burze



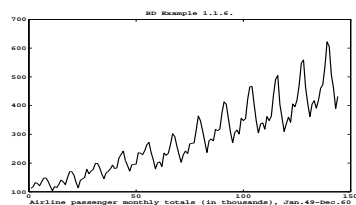
Obr. 1.1



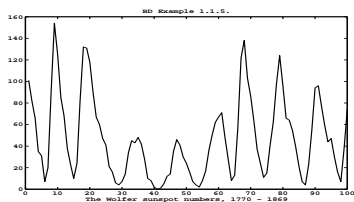
Obr. 1.2



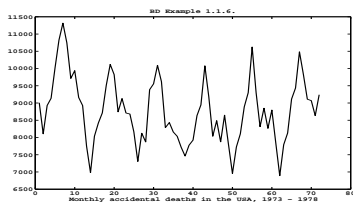
Obr. 1.3



Obr. 1.4



Obr. 1.5



Obr. 1.6

1.3. CÍL ANALÝZY.

Porozumění mechanismu, jímž se generují sledované údaje.
 Znalost modelu tohoto mechanismu \Rightarrow znalost algoritmu, jímž můžeme chování tohoto mechanismu simulovat na počítači \Rightarrow schopnost popsat s jistou přesností jeho chování:

- mezi časovými okamžiky měření (**interpolace**)
- v budoucnosti (**extrapolace, prognóza**)
- s cílem řídit a optimalizovat činnost určitého systému vhodnou volbou vstupních a počátečních podmínek (**regulace**), např. regulace složitých technologických procesů.

2.1. DEFINICE ČASOVÉ ŘADY.

Definice 2.1. Náhodný (stochastický) proces X je neprázdný systém ($T \neq \emptyset$) náhodných veličin definovaných na stejném pravděpodobnostním prostoru (Ω, \mathcal{A}, P) . Píšeme $X := \{X_t \mid t \in T\}$ nebo stručně $\{X_t\}$. Speciální případy:

$T \subseteq \mathbb{R} \dots$ proces se spojitým časem nebo náhodná funkce.

$T \subseteq \mathbb{Z} \dots$ proces s diskrétním časem, náhodná posloupnost nebo časová řada.

Poznámka 2.2. Indexová množina T je obvykle uspořádaná a interpretuje se jako spojitý nebo diskrétní časový interval. Může ale být také neuspořádaná, například souřadnice bodů v rovině (v meteorologii) nebo v 3-D prostoru (geofyzika).

Definice 2.3. Pro každý náhodný případ $\omega \in \Omega$ dostáváme funkci $x : T \rightarrow \mathbb{R}$ jako výsledek náhodného experimentu: $x(t) := X_t(\omega)$. Tato funkce se nazývá pozorování (trajektorie, realizace) procesu X .

Poznámka 2.4. Obrázky 1.1-1.6 ilustrují trajektorie různých náhodných procesů (časových řad).

Příklad 2.5 (Příklady časových řad).

- (1) Sinusovka s náhodnou amplitudou a fází (Obr. 1.1).
- (2) Binární proces házení mincí.
- (3) Náhodná procházka.
- (4) Proces větvení.

2.2. SYSTÉM DISTRIBUČNÍCH FUNKCÍ.

Definice 2.6 (Konzistentní systém distribučních funkcí procesu X). Označme $\mathcal{T} := \{t \mid t = [t_1, t_2, \dots, t_n] \in T^n, t_i \neq t_j, \text{ pro } i \neq j, n \in \mathbb{N}\}$. Pro každé $t \in \mathcal{T}$ libovolné délky $n \in \mathbb{N}$ nechť $F_t(x)$ značí adruženou distribuční funkci marginálního náhodného vektoru \mathbb{X}_t vybraného ze stochastického procesu $X = \{X_t\}_{t \in T}$ v časových okamžicích t_1, t_2, \dots, t_n . Systém $\{F_t\}_{t \in \mathcal{T}}$ úplně popisuje stochastické chování

procesu X a nazývá se **konzistentním systémem distribučních funkcí procesu X** (viz následující tvrzení).

Věta 2.7. *Systém distribučních funkcí $\{F_t\}_{t \in \mathcal{T}}$ z definice 2.6 se nazývá konzistentní protože má pro každé $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^n$ a $n \in \mathbb{N}$ následující dvě vlastnosti konzistence:*

- (i) $F_{\mathbf{t}_p}(\mathbf{x}_p) = F_{\mathbf{t}}(\mathbf{x})$ pro každou permutaci \mathbf{p} indexů $\{1, 2, \dots, n\}$. Zde značí \mathbf{t}_p , resp. \mathbf{x}_p vektory \mathbf{t} , resp. \mathbf{x} se složkami permutovanými podle permutace \mathbf{p} .
- (ii) $\lim_{x_i \rightarrow \infty} F_{\mathbf{t}}(\mathbf{x}) = F_{\mathbf{t}}(x_1, \dots, x_{i-1}, \infty, x_{i+1}, \dots, x_n) =: F_{\mathbf{t}(i)}(\mathbf{x}(i))$ pro každé $i \in \{1, 2, \dots, n\}$, kde $\mathbf{t}(i)$, resp. $\mathbf{x}(i)$ značí vektor \mathbf{t} , resp. \mathbf{x} s vynechanou i -tou složkou.

Definice 2.8. Stochastický proces se nazývá **normální** nebo **gaussovský**, když každá distribuční funkce $F_{\mathbf{t}}$ jeho konzistentního systému ($\mathbf{t} \in \mathcal{T}$) je sdruženou distribuční funkcí normálně rozloženého marginálního náhodného vektoru $\mathbb{X}_{\mathbf{t}}$.

Definice 2.9. Jestliže $X = \{X_t\}$ je časová řada, kde všechny náhodné veličiny X_t , $t \in T$ jsou **vzájemně nezávislé a stejně rozložené** se střední hodnotou μ a rozptylem σ^2 , budeme psát

$$\boxed{X \sim IID(\mu, \sigma^2)}$$

2.3. MOMENTOVÉ CHARAKTERISTIKY.

Nyní zavedeme **momentové funkce** jako analogie střední hodnoty a kovarianční matice náhodného vektoru, který můžeme pokládat za speciální případ konečné časové řady ($T = \{1, 2, \dots, n\}$).

Definice 2.10. Jsou-li dány stochastické procesy $X = \{X_t\}_{t \in T}$ a $Y = \{Y_t\}_{t \in T}$, oba definované na témž pravděpodobnostním prostoru, pak definujeme momentové funkce 1. a 2. řádu následovně:

- (1) **střední hodnota** X : $\mu_X : T \rightarrow \mathbb{R}$, kde $\mu_X(t) := EX_t$, pokud střední hodnoty existují pro všechna $t \in T$.

- (2) **autokovarianční funkce** $X: \gamma_X : T \times T \rightarrow \mathbb{R}$, kde
 $\gamma_X(r, s) := \text{cov}(X_r, X_s)$, pokud kovariance existují pro všechna
 $r, s \in T$.
- (3) **rozptyl** $X: \sigma_X^2 : T \rightarrow \mathbb{R}^+$, kde
 $\sigma_X^2(t) := \text{var}(X_t) = \text{cov}(X_t, X_t) = \gamma_X(t, t)$, pokud rozptyly
existují pro všechna $t \in T$.
- (4) **autokorelační funkce** $X: \rho_X : T \times T \rightarrow [-1, 1]$, kde

$$\rho_X(r, s) := \begin{cases} \frac{\gamma_X(r, s)}{\sqrt{\gamma_X(r, r)}\sqrt{\gamma_X(s, s)}} & \text{pro } \sqrt{\gamma_X(r, r)}\sqrt{\gamma_X(s, s)} \neq 0 \\ 0 & \text{jinak,} \end{cases}$$
pokud korelace existují pro všechna $r, s \in T$.
- (5) **vzájemná (křížová) kovarianční funkce** X a Y :
 $\gamma_{XY} : T \times T \rightarrow \mathbb{R}$, kde $\gamma_{XY}(r, s) := \text{cov}(X_r, Y_s)$, pokud
kovariance existují pro všechna $r, s \in T$.
- (6) **vzájemná (křížová) korelační funkce** X a Y :
 $\rho_{XY} : T \times T \rightarrow [-1, 1]$, kde

$$\rho_{XY}(r, s) := \begin{cases} \frac{\gamma_{XY}(r, s)}{\sqrt{\gamma_X(r, r)}\sqrt{\gamma_Y(s, s)}} & \text{pro } \sqrt{\gamma_X(r, r)}\sqrt{\gamma_Y(s, s)} \neq 0 \\ 0 & \text{jinak,} \end{cases}$$
pokud korelace existují pro všechna $r, s \in T$.

2.4. STRIKTNÍ A SLABÁ STACIONARITA.

Definice 2.11. Časová řada $X := \{X_t | t \in \mathbb{Z}\}$ se nazývá **striktně stacionární**, jestliže každá distribuční funkce jeho konzistentního systému $\{F_t\}_{t \in \mathcal{T}}$, nezávisí na posunutí: $F_t(\cdot) \equiv F_{t+h}(\cdot)$ pro každé $t \in \mathcal{T}$ a $h \in \mathbb{Z}$.

Definice 2.12. Časová řada $X := \{X_t | t \in \mathbb{Z}\}$ se nazývá **(slabě) stacionární**, jestliže jsou splněny následující podmínky:

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

Jestliže jsou splněny pouze podmínky (1) a (2), pak X se nazývá **kovariančně stacionární**.

Poznámka 2.13.

- (1) Zřejmě ze (2) vyplývá při $r = s$, že rozptyl stacionární časové řady je rovněž konstantní funkce: $\sigma_x^2(\cdot) \equiv \sigma_X^2$.
- (2) Jestliže platí (3), pak z $\gamma_X(r, s) = EX_r X_s - (EX_r)(EX_s) = EX_r X_s - \mu_x^2$ plyne, že (2) je ekvivalentní (a může tak být nahrazena) podmínkou:
 $EX_r X_s = EX_{r+h} X_{s+h}$ pro každé $r, s, h \in \mathbb{Z}$. Celkem vidíme, že všechny první a druhé momenty stacionární časové řady jsou invariantní k posunutí. Proto je slabá stacionarita také někdy označována jako **stacionarita druhého řádu**.

Poznámka 2.14. Zřejmě podmínka (2) z definice 2.12 může být také ekvivalentně nahrazena modifikovanou podmínkou:

(2') $\gamma_x(r, s)$ závisí pouze na rozdílu svých argumentů $r - s$.

Můžeme proto autokovarianční i autokorelační funkci stacionární časové řady zavést jako funkci jen jedné proměnné:

$$\begin{aligned}\gamma_X(h) &:= \gamma_X(t+h, t) \\ \rho_X(h) &:= \rho_X(t+h, t) = \frac{\gamma_X(t+h, t)}{\sigma_X \sigma_X} = \frac{\gamma_X(h)}{\gamma_X(0)} \\ \sigma_X^2 &= \gamma_X(t, t) = \gamma_X(0),\end{aligned}\tag{2.1}$$

kde $t, h \in \mathbb{Z}$ jsou libovolné.

Věta 2.15. Každá striktně stacionární časová řada s konečnými druhými momenty je stacionární.

Poznámka 2.16.

Obecně stacionarita nemá za následek striktní stacionaritu.

Věta 2.17. Každá gaussovská stacionární časová řada je striktně stacionární.

Definice 2.18. Časová řada $X = \{X_t\}$ se nazývá **bílý šum** se střední hodnotou μ a rozptylem σ^2 , jestliže $\mu_X(t) \equiv \mu$ a

$$\gamma_X(r, s) = \begin{cases} \sigma^2 & \text{pro } r = s \\ 0 & \text{jinak} \end{cases}.$$

Píšeme

$$X \sim WN(\mu, \sigma^2).$$

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

Poznámka 2.19. Je snadné ověřit následující implikace:

$$X \sim IID(\mu, \sigma^2) \Rightarrow X \sim WN(\mu, \sigma^2) \Rightarrow X \text{ je stacionární.}$$

Poznamenejme také, že žádná z opačných implikací neplatí. (viz též 2.16).

2.5. LINEÁRNÍ TRANSFORMACE ČASOVÝCH ŘAD.

Definice 2.20. Řekneme, že časová řada $X := \{X_t\}$ má **ohraničené druhé momenty**, jestliže existuje konstanta $C \in \mathbb{R}$ s vlastností $E|X_t|^2 \leq C < \infty$ pro každé t .

Věta 2.21. Časová řada $X := \{X_t\}$ má ohraničené druhé momenty právě když existují reálné konstanty $\mu_X < \infty$ a $\sigma_X < \infty$ takové, že $|\mu_X(t)| \leq \mu_X$ a $|\sigma_X(t)| \leq \sigma_X$ pro každé t , tj. právě když X má ohraničenou střední hodnotu i rozptyl.

Věta 2.22 (Hlavní věta o konvergenci). Jestliže $\Phi := \{\Phi_k\} \in \ell_1$ (tj. $\sum_k |\Phi_k| < \infty$) a $U := \{U_t\}$ je časová řada s ohraničenými druhými momenty, pak nekonečná řada $X_t := \sum_k \Phi_k U_{t-k}$ konverguje podle kvadratického středu¹ pro každé t a lineárně transformovaná časová řada $X := \{X_t\}$ má rovněž ohraničené druhé momenty.

¹tj. částečné součty konvergují dle kvadratického středu:
 $E|X_t - \sum_{k=-N}^N \Phi_k U_{t-k}|^2 \rightarrow 0$. Součet v tomto smyslu označujeme symbolem $\stackrel{2}{=}$

Poznámka 2.23. Lineární transformace s koeficienty $\Phi := \{\Phi_k\}$, která každou časovou řadu s ohraničenými druhými momenty transformuje opět na řadu s ohraničenými druhými momenty se nazývá **stabilní**. Z předchozí věty tedy vyplývá, že postačující podmínkou pro stabilitu lineární transformace je absolutní sumovatelnost posloupnosti Φ jejích koeficientů. Proto se tato podmínka také někdy nazývá **podmínkou stability**.

Důsledek 2.24. *Jestliže $\Phi \in \ell_1$ a $U = \{U_t\}$ je stacionární se střední hodnotou μ_U a autokovarianční funkcí γ_U , pak platí*

- (1) $X_t \stackrel{2}{=} \sum_k \Phi_k U_{t-k}$ konverguje pro každé t
- (2) $X := \{X_t\}$ je stacionární se střední hodnotou a autokovarianční funkcí:

$$\mu_X = \mu_U \sum_j \Phi_j \quad (2.2a)$$

$$\gamma_X(h) = \sum_j \sum_k \gamma_U(h - j + k), \quad h \geq 0. \quad (2.2b)$$

Důsledek 2.25.

$$\sigma_X^2 = \gamma_X(0) = \sum_j \sum_k \gamma_U(k - j) \quad (2.3a)$$

$$\sigma_X^2 \leq \sigma_U^2 \left(\sum_j |\Phi_j| \right)^2 \quad (2.3b)$$

Definice 2.26. Časová řada X , která je lineární transformací řady U jako ve větě 2.22 se také někdy nazývá časovou řadou **(lineárně) generovanou řadou** U . Jestliže je $\Phi_k = 0$ pro každé $k < 0$, pak X se nazývá **kauzální (kauzálně generovanou)**, neboť její hodnoty X_t nezávisí na budoucích hodnotách $U_\tau, \tau > t$ generující řady (jsou podmíněny jen hodnotami $U_\tau, \tau \leq t$). V takovém případě částečné součty při sumaci dle kvadratického středu nabudou tvaru $\sum_{k=0}^N \Phi_k U_{t-k}$ a samotnou řadu tak můžeme psát ve tvaru $X_t \stackrel{2}{=} \sum_{k=0}^{\infty} \Phi_k U_{t-k}$.

Poznámka 2.27 (Operátor zpětného posuvu).

Značíme pro $k \in \mathbb{N}$:

$$BU_t := U_{t-1}, \text{ atd. rekurentně } B^k U_t = B(B^{k-1} U_t) := U_{t-k}$$

$$B^{-1} U_t := U_{t+1}, \text{ atd. rekurentně } B^{-k} U_t := B^{-1}(B^{-(k-1)} U_t) := U_{t+k}$$

$B^0 U_t := U_t$. Pak formálně místo $X_t \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \Phi_k U_{t-k}$ píšeme stručně

$$X_t = \Phi(B) U_t, \text{ kde } \Phi(B) := \sum_{k=0}^{\infty} \Phi_k B^k, \text{ neboť}$$

$\sum_{k=0}^{\infty} \Phi_k U_{t-k} = \sum_{k=0}^{\infty} \Phi_k B^k U_t = (\sum_{k=0}^{\infty} \Phi_k B^k) U_t$ je operátor získaný formálním dosazením operátoru B za proměnnou v řadě (polynomu) $\Phi(z) = \sum_{k=0}^{\infty} \Phi_k z^k$, která(ý) hraje roli přenosové charakteristiky lineární transformace (ta je totiž konvolučního typu).

2.6. ODHADY MOMENTOVÝCH FUNKCÍ.

Definice 2.28. Necht $\mathbf{x} = [x_1, \dots, x_n]$ je n pozorování ($x_t = X_t(\omega)$ pro $t = 1, \dots, n$) stacionární časové řady se střední hodnotou μ , rozptylem σ^2 , autokovarianční funkcí $\gamma(\cdot)$ a autokorelační funkcí $\rho(\cdot)$.

Pak jejich odhady spočteme takto:

$$\hat{\mu} := \frac{1}{n} \sum_{j=1}^n x_j \dots \text{ odhad } \mu;$$

$$\hat{\gamma}(h) := \frac{1}{n} \sum_{j=1}^{n-h} (x_{j+h} - \hat{\mu})(x_j - \hat{\mu}), \quad 0 \leq h \leq n-1,$$

$$\hat{\gamma}(h) := \hat{\gamma}(-h), \quad -(n-1) \leq h < 0 \dots \text{ odhad } \gamma(h);$$

$$\hat{\sigma}^2 := \hat{\gamma}(0) \dots \text{ odhad rozptylu};$$

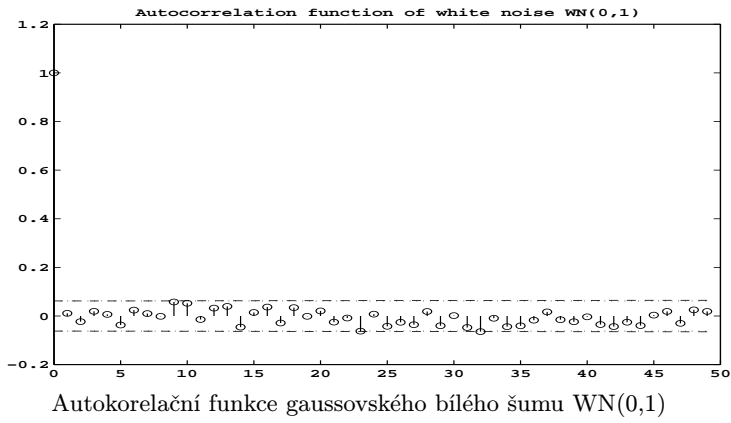
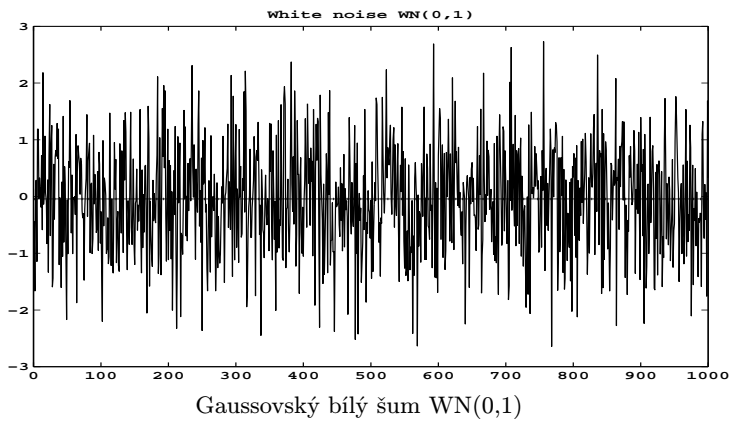
$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -(n-1) \leq h \leq n-1 \text{ (viz rovnici (2.1))}$$

... odhad autokorelační funkce v případě $\hat{\gamma}(0) \neq 0$, jinak $\hat{\rho}(h) := 0$.

Věta 2.29. Necht $\mathbb{X} := [X_1, \dots, X_n]$ marginální vektor v X korepondující s vektorem pozorování \mathbf{x} . Potom jak matice

$$\hat{\Gamma}_n := [\hat{\gamma}(i-j)]_{i,j} = \begin{bmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \dots & \hat{\gamma}(n-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \dots & \hat{\gamma}(n-2) \\ \dots & \dots & \dots & \dots \\ \hat{\gamma}(n-1) & \hat{\gamma}(n-2) & \dots & \hat{\gamma}(0) \end{bmatrix},$$

kteřá je odhadem varianční matice $\text{var}\mathbb{X}$, tak i matice $\hat{R}_n := \frac{\hat{\Gamma}_n}{\hat{\gamma}(0)}$, kteřá je odhadem korelační matice $\rho(\mathbb{X})$, jsou symetrické a pozitivně semidefinitní.



Poznámka 2.30.

- (1) Odhad je spolehlivý pouze pro $n > 50$ a $h < \frac{n}{4}$.

- (2) Odhad $\hat{\gamma}(h)$ je vychýlený ($E\hat{\gamma}(h) \neq \gamma(h)$), protože součet pozorování je dělen n a nikoliv počtem stupňů volnosti $n-1-h$. Poznamenejme, že věta 2.29 neplatí pro nevychýlený odhad (matice $\hat{\Gamma}_n$ ztratí očekávanou vlastnost pozitivní semidefinitnosti). V každém případě je ale odhad $\hat{\gamma}(h)$ **asymptoticky nevychýlený** v tom smyslu, že $E\hat{\gamma}(h) \rightarrow \gamma(h)$ pro $n \rightarrow \infty$. Navíc je také **konzistentní podle kvadratického středu** v tom smyslu, že $E|\hat{\gamma}(h) - \gamma(h)|^2 \rightarrow 0$ pro $n \rightarrow \infty$, kde konvergence je dokonce rychlejší než v případě nevychýleného odhadu.
- (3) Z algebraického hlediska $\hat{\gamma}(h)$ lze psát ve tvaru skalárního součinu $\hat{\gamma}(h) = \frac{1}{n} \langle \mathbf{x}_0, \mathbf{x}_h \rangle$, kde $\mathbf{x}_h := [0, \dots, 0, \underbrace{x_1 - \hat{\mu}, \dots, x_n - \hat{\mu}}_{n-1-h}, 0, \dots, 0]$. Vektor \mathbf{x}_0 tak reprezentuje the původní vektor pozorování (doplněný na konci $n-1$ nulami), zatímco \mathbf{x}_h je jeho kopie posunutá (zpožděná) o h .
- Zřejmě $\|\mathbf{x}_0\|^2 = \|\mathbf{x}_h\|^2 = \sum_{j=1}^n |x_j - \hat{\mu}|^2$. Ze Schwarzovy nerovnosti dostáváme $|\langle \mathbf{x}_0, \mathbf{x}_h \rangle| \leq \|\mathbf{x}_0\|^2$, což vede na $|\hat{\gamma}(h)| \leq \frac{1}{n} \|\mathbf{x}_0\|^2 = \frac{1}{n} \langle \mathbf{x}_0, \mathbf{x}_0 \rangle = \hat{\gamma}(0)$. Odtud vidíme, že odhad autokorelační funkce zachovává její přirozenou vlastnost $|\hat{\rho}(h)| \leq 1$.
- (4) Vzhledem ke (3) je možno odhad $\hat{\rho}(h)$ interpretovat geometricky jako kosinus úhlu mezi původním a posunutým vektorem pozorování \mathbf{x}_0 a \mathbf{x}_h , což lze interpretovat jako míru jejich lineární závislosti (podobnosti): nula znamená ortogonalitu (úplnou lineární nezávislost=žádná korelace mezi oběma vektory), ± 1 odpovídá lineární závislosti (úplná korelace: jeden z nich je skalárním násobkem druhého).
- (5) Trend je charakterizován korelacemi na velkou vzdálenost, což se projevuje pomalým poklesem $\gamma(h)$ při $h \rightarrow \infty$. Periodická složka se projeví oscilatorickým chováním $\hat{\gamma}(h)$ s dominantní periodou této složky, případně směsí takových složek.

3.1. PRINCIP ORTOGONÁLNÍ PROJEKCE.

Definice 3.1 (Prostor $L_2(\Omega, \mathcal{A}, P)$).

$L_2 := L_2(\Omega, \mathcal{A}, P)$ definujeme jako množinu všech (komplexních) náhodných veličin nad týmž pravděpodobnostním prostorem (Ω, \mathcal{A}, P) , které mají konečné druhé momenty (resp. rozptyly - viz dále 3.5), tj. $L_2(\Omega, \mathcal{A}, P) := \{X \mid X \text{ náhodná veličina nad } (\Omega, \mathcal{A}, P), \mathbb{E}|X|^2 < \infty\}$.

Poznamenejme, že do tohoto prostoru zahrnujeme také všechny konstanty z \mathbb{C} , které považujeme za náhodné veličiny s nulovým rozptylem.

Věta 3.2. $L_2(\Omega, \mathcal{A}, P)$ je **Hilbertův prostor** se skalárním součinem $\langle X, Y \rangle = \mathbb{E}X\bar{Y}$ a normou $\|X\|_2 = \sqrt{\langle X, X \rangle} = \sqrt{\mathbb{E}|X|^2}$.

Důkaz. $L_2(\Omega, \mathcal{A}, P)$ je obdobou funkcionálního prostoru $L_2(\mathcal{J})$ tvořeného funkcemi absolutně integrovatelnými v kvadrátu na intervalu $\mathcal{J} \subseteq \mathbb{R}$. Totiž $\mathbb{E}|X|^2 = \int_{\Omega} |X(\omega)|^2 dP(\omega)$, takže namísto s Lebesgueovým integrálem pracujeme s obecnějším pojetím integrálu, kde Lebesgueova míra je nahrazena pravděpodobnostní mírou P :

• **Skalární součin $\langle X, Y \rangle$ existuje a je konečný** pro každé $X, Y \in L_2(\Omega, \mathcal{A}, P)$, jak snadno nahlédneme z nerovnosti

$$\begin{aligned} 4|X\bar{Y}| &= (|X| + |\bar{Y}|)^2 - (|X| - |\bar{Y}|)^2 \leq (|X| + |\bar{Y}|)^2 + (|X| - |\bar{Y}|)^2 = \\ &= 2(|X|^2 + |\bar{Y}|^2), \end{aligned}$$

odkud užitím $|\bar{Y}| = |Y|$ dostáváme

$$|X\bar{Y}| \leq \frac{1}{2}(|X|^2 + |Y|^2),$$

takže

$$\begin{aligned} |\mathbb{E}X\bar{Y}| &\leq \int_{\Omega} |X(\omega)\bar{Y}(\omega)| dP(\omega) \leq \\ &\leq \frac{1}{2} \left[\int_{\Omega} |X(\omega)|^2 dP(\omega) + \int_{\Omega} |Y(\omega)|^2 dP(\omega) \right] < \infty. \end{aligned}$$

$L_2(\Omega, \mathcal{A}, P)$ je vektorovým prostorem. Je uzavřený na násobení skaláry $c \in \mathbb{C}$, neboť $E|cX|^2 = |c|^2 E|X|^2 < \infty$. Uzavřenost vzhledem ke sčítání plyne z:

$$\begin{aligned} |X + Y|^2 &\leq (|X| + |Y|)^2 = |X|^2 + 2|XY| + |Y|^2 \Rightarrow \\ E|X + Y|^2 &\leq E|X|^2 + 2E|XY| + E|Y|^2 < \infty. \end{aligned}$$

• **Ověření, že $L_2(\Omega, \mathcal{A}, P)$ je úplný, neboli Hilbertův prostor**, je složitější, ale provádí se opět zcela analogicky jako v případě funkcionálního prostoru $L_2(\mathcal{J})$. Podrobnosti lze nalézt například v monografii [BD93, §2.10]. \square

Důsledek 3.3 (Schwarzova nerovnost).

$$|EXY|, |EX\bar{Y}| \leq \|X\|_2 \|Y\|_2 = \sqrt{E|X|^2} \sqrt{E|Y|^2}, \quad X, Y \in L_2(\Omega, \mathcal{A}, P).$$

Důsledek 3.4. $X \in L_2(\Omega, \mathcal{A}, P) \Rightarrow EX$ existuje.

Důkaz.

$$|EX| = |E(1 \cdot X)| \leq \underbrace{\sqrt{E|1|^2}}_1 \sqrt{E|X|^2} = \sqrt{E|X|^2} < \infty.$$

\square

Důsledek 3.5.

$$X, Y \in L_2(\Omega, \mathcal{A}, P) \Rightarrow X - EX, Y - EY \in L_2(\Omega, \mathcal{A}, P)$$

a

$$\langle X - EX, Y - EY \rangle = E(X - EX)(\overline{Y - EY}) = \text{cov}(X, Y)$$

existuje a splňuje Schwarzovu nerovnost

$$|\text{cov}(X, Y)| \leq \sqrt{E|X - EX|^2} \sqrt{E|Y - EY|^2} = \sigma_X \sigma_Y.$$

Důsledek 3.6.

$$\rho(X, Y) = \begin{cases} \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} & \text{pro } \sigma_x \sigma_y \neq 0 \\ 0 & \text{pro } \sigma_x \sigma_y = 0 \end{cases}$$

je tzv. **korelační koeficient náhodných veličin X a Y** , pro nějž platí $|\rho(X, Y)| \leq 1$ a speciálně $-1 \leq \rho(X, Y) \leq 1$ v případě reálných náhodných veličin X a Y .

Věta 3.7 (Věta o ortogonální projekci: [BD93], §2.3).

Je-li L uzavřený lineární podprostor Hilbertova prostoru H (například $H = L_2$) a $x \in H$, pak

(i) existuje jediný prvek $\hat{x} \in L$ takový, že

$$\|x - \hat{x}\| = \inf_{y \in L} \|x - y\| \quad (*)$$

a

(ii) nějaký prvek má minimální vlastnost (*) právě když $x - \hat{x} \perp L$, tj. $x - \hat{x} \perp y$ (neboli $\langle x - \hat{x}, y \rangle = 0$) pro každé $y \in L$. Píšeme $\hat{x} = P_L x$, kde P_L je tzv. **operátor ortogonální projekce na podprostor L** .

Definice 3.8.

Nechť $X_0 := 1, X_1, \dots, X_n \in L_2$ a $L := \mathcal{L}(X_0, X_1, \dots, X_n)$ je lineární podprostor v L_2 generovaný náhodnými veličinami X_0, X_1, \dots, X_n . Je-li $X \in L_2$ nějaká další náhodná veličina, pak $\hat{X} = P_L X$ se nazývá **nejlepší² lineární predikce** náhodné veličiny X pomocí náhodných veličin X_0, X_1, \dots, X_n .

Poznámka 3.9. Operátor P_L je korektně definován, neboť L má konečnou dimenzi ($\dim L \leq n + 1$) a je tak automaticky uzavřený. Protože $\hat{X} \in \mathcal{L}(X_0, X_1, \dots, X_n)$ existuje vektor $\beta_0 := [\beta_0, \beta_1, \dots, \beta_n]^T \in \mathbb{R}^{n+1}$ takový, že

$$\hat{X} = \beta_0 \underbrace{X_0}_1 + \beta_1 + \dots + \beta_n X_n. \quad (3.1a)$$

²rozumí se ve smyslu minimální střední kvadratické chyby.

I když \widehat{X} je jediný, nemusí být β_0 obecně jednoznačně určen. Je tomu tak jen když X_0, X_1, \dots, X_n jsou lineárně nezávislé, tj. tvoří bázi prostoru L .

Věta 3.10. Vektor β_0 ze vztahu (3.1a) se spočte jako řešení následujícího systému $n + 1$ lineárních rovnic (tzv. **normální systém rovnic**), kde $i = 0, 1, \dots, n$:

$$\langle X_0, X_i \rangle \beta_0 + \langle X_1, X_i \rangle \beta_1 + \dots + \langle X_n, X_i \rangle \beta_n = \langle X, X_i \rangle \quad (3.1b)$$

neboli dle věty 3.2

$$E(X_0 X_i) \beta_0 + E(X_1 X_i) \beta_1 + \dots + E(X_n, X_i) \beta_n = E(X X_i). \quad (3.1c)$$

Důkaz. Dle 3.7:

$\widehat{X} = P_L X \Leftrightarrow X - \widehat{X} \perp Y$ pro každé $Y \in L \Leftrightarrow X - \widehat{X} \perp X_i$ pro každé $i = 0, 1, \dots, n \Leftrightarrow \langle X - \widehat{X}, X_i \rangle = 0$ pro každé $i = 0, 1, \dots, n \Leftrightarrow \langle \widehat{X}, X_i \rangle = \langle X, X_i \rangle$ pro každé $i = 0, 1, \dots, n \Leftrightarrow \langle \sum_{j=0}^n \beta_j X_j, X_i \rangle = \langle X, X_i \rangle$ pro každé $i = 0, 1, \dots, n \Leftrightarrow \sum_{j=0}^n \beta_j \langle X_j, X_i \rangle = \langle X, X_i \rangle$ pro každé $i = 0, 1, \dots, n$, což je právě (3.1b). \square

Příklad 3.11 ($n = 0$). $L = \mathcal{L}(X_0) = \mathcal{L}(1) = \mathbb{R}$ je lineární podprostor všech (reálných) konstant v L_2 . Pak (3.1b) se redukuje jen na jednu rovnici:

$$E(\underbrace{X_0 X_0}_1) \beta_0 = E(X \underbrace{X_0}_1).$$

Odtud máme $\beta_0 = EX$ a tedy $\widehat{X} = \beta_0 X_0 = EX$.

Tedy EX je nejlepší lineární predikce náhodné veličiny X pomocí konstanty. Dosažená minimální střední kvadratická chyba je $E(X - EX)^2 = \text{var} X$.

Důsledek 3.12. Jestliže $EX = EX_1 = \dots = EX_n = 0$, pak $\widehat{X} = P_L X = P_{\mathcal{L}(X_1, \dots, X_n)} X$, $\beta_0 = 0$ a (3.1c) přejde v systém n rovnic, kde $i = 1, \dots, n$:

$$E(X_1 X_i) \beta_1 + \dots + E(X_n, X_i) \beta_n = E(X X_i). \quad (3.1d)$$

Položíme-li $\mathbb{X} := [X_1, \dots, X_n]^T$, můžeme jej přepsat do maticového tvaru

$$(\text{var}\mathbb{X})\boldsymbol{\beta} = \text{cov}(X, \mathbb{X})^T, \quad (3.1e)$$

kde $\boldsymbol{\beta} := [\beta_1, \dots, \beta_n]^T \in \mathbb{R}^n$.

Důkaz. První rovnice ($i = 0$) v (3.1c) bude tvaru:

$$\underbrace{\text{E}(1.1)}_1 \beta_0 + \underbrace{\text{E}(X_1.1)}_0 \beta_1 + \dots + \underbrace{\text{E}(X_n.1)}_0 \beta_n = \underbrace{\text{E}(X.1)}_0.$$

Odtud dostáváme $\beta_0 = 0$, takže ve zbývajících rovnicích pro $i = 1, \dots, n$ bude $\text{E}(X_0 X_1) \beta_0 = 0$, což je právě systém normálních rovnic pro $P_{\mathcal{L}(X_1, \dots, X_n)} X$. \square

Důsledek 3.13 (Nejlepší lin. predikce pro stacionární časové řady).

*Nechť $X := \{X_t\}$ je stacionární časová řada s nulovou střední hodnotou $\mu_X = 0$ a autokovarianční funkcí γ_X . Pro dané $k \in \mathbb{N}$ nechtě $\hat{X}_{t+k}^{(k)}$ značí nejlepší (k -krokovou) lineární predikci náhodné veličiny X_{t+k} pomocí n předchozích veličin $X_t, X_{t-1}, \dots, X_{t+1-n}$ ve tvaru $\hat{X}_{t+k}^{(k)} = \sum_{j=1}^n \Phi_{n,j}^{(k)} X_{t+1-j}$. Pak vektor $\boldsymbol{\Phi}_n^{(k)} := [\Phi_{n1}^{(k)}, \dots, \Phi_{nn}^{(k)}]^T$ nalezneme řešením tzv. **Yule-Walkerova systému lineárních rovnic**:*

$$\underbrace{\begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(n-2) \\ \vdots & \vdots & \dots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(0) \end{bmatrix}}_{\Gamma_n} \underbrace{\begin{bmatrix} \Phi_{n1}^{(k)} \\ \Phi_{n2}^{(k)} \\ \vdots \\ \Phi_{nn}^{(k)} \end{bmatrix}}_{\boldsymbol{\Phi}_n^{(k)}} = \underbrace{\begin{bmatrix} \gamma(k) \\ \gamma(k+1) \\ \vdots \\ \gamma(k+n-1) \end{bmatrix}}_{\boldsymbol{\gamma}_n^{(k)}} \quad (3.2)$$

Důkaz. Plyne z 3.12 po záměnách:

$X \rightsquigarrow X_{t+k}, X_j \rightsquigarrow X_{t+1-j}, \beta_j \rightsquigarrow \Phi_{nj}^{(k)}$ a tedy pak

$$\text{var}\mathbb{X} = [\text{cov}(X_i X_j)] \rightsquigarrow [\text{cov}(X_{t+1-i}, X_{t+1-j})] = [\gamma(t+1-i - (t+1-j))] = [\gamma(j-i)] = \Gamma_n \text{ a podobně}$$

$$\text{cov}(X, \mathbb{X})^T \rightsquigarrow [\mathbb{E}(X_{t+k}, X_{t+1-j})] = \dots = [\gamma(t+k - (t+1-j))] = [\gamma(k+j-1)] = \gamma_n^{(k)}. \quad \square$$

Věta 3.14 (Střední kvadratická chyba).

Střední kvadratická chyba $v_n^{(k)} := \mathbb{E}(X_{t+k} - X_{t+k}^{(k)})^2$ *nejlepší* k -*krokové lineární predikce se spočte dle vztahu:*

$$v_n^{(k)} = \gamma(0) - \sum_{j=1}^n \Phi_{nj}^{(k)} \gamma(k+j-1) = \gamma(0) - \Phi_n^T \gamma_n^{(k)}. \quad (3.3)$$

Důkaz. $v_n^{(k)} = \mathbb{E}(X_{t+k} - X_{t+k}^{(k)})^2 = \|X_{t+k} - X_{t+k}^{(k)}\|^2 =$
 $\langle X_{t+k} - X_{t+k}^{(k)}, X_{t+k} - X_{t+k}^{(k)} \rangle = \langle X_{t+k} - X_{t+k}^{(k)}, X_{t+k} \rangle -$
 $\underbrace{\langle X_{t+k} - X_{t+k}^{(k)}, X_{t+k}^{(k)} \rangle}_0 = \langle X_{t+k}, X_{t+k} \rangle - \langle X_{t+k}^{(k)}, X_{t+k} \rangle =$
 $\|X_{t+k}\|^2 - \langle \sum_{j=1}^n \Phi_{nj}^{(k)} X_{t+1-j}, X_{t+k} \rangle =$
 $\|X_{t+k}\|^2 - \sum_{j=1}^n \Phi_{nj}^{(k)} \langle X_{t+1-j}, X_{t+k} \rangle =$
 $\mathbb{E}(X_{t+k}^2) - \sum_{j=1}^n \Phi_{nj}^{(k)} \mathbb{E}(X_{t+1-j} X_{t+k}) =$
 $\gamma(0) - \sum_{j=1}^n \Phi_{nj}^{(k)} \gamma(t+1-j - (t+k)) = \gamma(0) - \sum_{j=1}^n \Phi_{nj}^{(k)} \gamma(1-j-k) =$
 $\gamma(0) - \sum_{j=1}^n \Phi_{nj}^{(k)} \gamma(k+j-1). \quad \square$

Definice 3.15. Necht $X, Y, X_1, \dots, X_n \in L_2$ a \hat{X} a \hat{Y} jsou nejlepší lineární predikce náhodných veličin X a Y pomocí X_1, \dots, X_n . Pak $\rho(X, Y | X_1, \dots, X_n) := \rho(X - \hat{X}, Y - \hat{Y})$ se nazývá **parciální korelační koeficient náhodných veličin X a Y při daných X_1, \dots, X_n .**

Interpretace: parciální korelace mezi X a Y při daných X_1, \dots, X_n je korelace mezi X a Y očištěná o tu svoji část, která je zprostředkována náhodnými veličinami X_1, \dots, X_n .

Definice 3.16. Necht $X = \{X_t | t \in \mathbb{Z}\}$ je stacionární časová řada s autokorelační funkcí $\rho_X(\cdot)$. Pak **parciální autokorelační funkci**

(pacf) $\alpha_X(\cdot)$ časové řady X definujeme následovně:

$$\alpha_X(0) = \rho_X(0) = 1,$$

$$\alpha_X(1) = \rho_X(1) = \rho(X_{t+1}, X_t)$$

$$\alpha_X(h) = \rho(X_{t+h}, X_t | X_{t+1}, \dots, X_{t+h-1}) \text{ for } h \geq 2.$$

Věta 3.17. Jestliže $X = \{X_t | t \in \mathbb{Z}\}$ je stacionární časová řada se střední hodnotou $\mu_X = 0$ a parciální autokorelační funkcí $\alpha_X(\cdot)$, pak $\alpha_X(n) = \Phi_{n,n}$ pro $n \geq 1$, kde $\Phi_n = [\Phi_{n,1}, \dots, \Phi_{n,n}]^T$ je řešením problému nejlepší lineární predikce.

Důkaz. See [BD93, §5.2]. □

$\alpha_X(h)$ lze počítat rekurzivně užitím mezivýsledných vztahů (3.4b) dále uvedeného Durbin-Levinsonova algoritmu 3.19.

Jiný postup počítá $\Phi_{n,n}^{(k)}$ z (3.2) pomocí Cramerova pravidla podle následujícího důsledku. Bohužel tato metoda není výpočetně příliš efektivní pro velká n .

Důsledek 3.18. Je-li matice $\Gamma_n := [\gamma_x(i-j)]_{i,j=1}^n$ regulární, pak

$$\alpha_X(n) = \Phi_{n,n} = \frac{\det \Gamma_n^*}{\det \Gamma_n},$$

kde $\Gamma_n^* := [\Gamma(:,1), \dots, \Gamma(:,n-1), \gamma_n]$ a $\gamma_n = [\gamma_X(1), \dots, \gamma_X(n)]^T$.

3.2. DURBIN-LEVINSONŮV A INOVAČNÍ ALGORITMUS.

Věta 3.19 (Durbin-Levinsonův algoritmus pro rekur. výpočet Φ_n).
 Nechť $X = \{X_t \mid t \in \mathbb{Z}\}$ je stacionární časová řada s nulovou střední hodnotou $\mu_X = 0$ a autokovarianční funkcí $\gamma_X(h) \rightarrow 0$ pro $h \rightarrow \infty$.
 Jestliže $\hat{X}_{n+1} = \Phi_{n,1}X_n + \dots + \Phi_{n,n}X_1$ je nejlepší lineární predikce, pak pro koeficienty $\Phi_{n,j}$ a střední kvadratickou chybu $v_n = E|X_{n+1} - \hat{X}_{n+1}|^2$ platí následující rekurentní vztahy.

Počáteční krok pro $n = 0$:

$$v_0 = \gamma_X(0). \quad (3.4a)$$

Rekurentní krok pro $n > 0$:

$$\Phi_{n,n} = \left[\gamma_X(n) - \overbrace{\sum_{j=1}^{n-1} \Phi_{n-1,j} \gamma_X(n-j)}{=0 \text{ pro } n=1} \right] v_{n-1}^{-1}, \quad (3.4b)$$

$$v_n = v_{n-1}(1 - |\Phi_{n,n}|^2), \quad (3.4c)$$

$$\begin{bmatrix} \Phi_{n,1} \\ \Phi_{n,2} \\ \vdots \\ \Phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \Phi_{n-1,1} \\ \Phi_{n-1,2} \\ \vdots \\ \Phi_{n-1,n-1} \end{bmatrix} - \Phi_{n,n} \begin{bmatrix} \bar{\Phi}_{n-1,n-1} \\ \bar{\Phi}_{n-1,n-2} \\ \vdots \\ \bar{\Phi}_{n-1,1} \end{bmatrix} \quad \text{pro } n > 1. \quad (3.4d)$$

Důkaz. Viz [BD93, str. 162]. \square

Představme ještě jiný algoritmus pro výpočet jednokrokové nejlepší lineární predikce. Jedná se o tzv **inovační algoritmus (IA)**, který může být na rozdíl od Durbin-Levinsonova algoritmu (DLA) aplikován na libovolnou i nestacionární časovou řadu s nulovou střední hodnotou $\mu_X = 0$. Tento algoritmus není nic jiného než souřadnicový přepis Gram-Schmidtova ortogonalizačního procesu.

Věta 3.20 (Inovační algoritmus pro rekurentní výpočet Φ_n).

Nechť $X = \{X_t \mid t \in \mathbb{Z}\}$ je časová řada s nulovou střední hodnotou $\mu_X = 0$ s maticí smíšených momentů $[\kappa_{i,j}]_{i,j=1}^n$, $\kappa_{i,j} := \mathbb{E}(X_i X_j)$, která je regulární pro každé n . Pak jednokrokové nejlepší lineární predikce $\hat{X}_{n+1} := \hat{X}_{n+1}^{(1)}$, $n \geq 0$, a jejich střední kvadratické chyby $v_n := v_n^{(1)}$ spočteme rekurentně pomocí následujících vztahů:

$$\hat{X}_{n+1} = \begin{cases} 0 & \text{pro } n = 0 \\ \sum_{j=1}^n \Theta_{n,j} (X_{n+1-j} - \hat{X}_{n+1-j}) & \text{pro } n \geq 1 \end{cases}, \quad (3.5a)$$

užitím

$$v_0 = \kappa(1, 1) \quad (3.5b)$$

$$\Theta_{n,n-k} = v_k^{-1} (\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \Theta_{k,k-j} \Theta_{n,n-j} v_j) \quad (3.5c)$$

$$v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \Theta_{n,n-j}^2 v_j. \quad (3.5d)$$

postupně v pořadí: $v_0; \Theta_{1,1}, v_1; \Theta_{2,2}, \Theta_{2,1}, v_2; \dots$

4.1. PŘEDZPRACOVÁNÍ.

4.1.1. Ošetření specifických efektů.

a) Volba okamžiků pozorování:

- jsou přímo **diskrétní svou povahou**, např. úroda obilí za jednotlivé roky.
- vznikají **diskretizací spojitě časové řady**, např. teplota v danou denní dobu na daném místě.
- vznikají **akumulací (agregací)** hodnot za určité období, např. denní množství srážek, roční výroba závodu. Místo akumulace se někdy provádí průměrování.

Je-li dána možnost volby, je třeba jí věnovat pozornost:

- **málo bodů** \Rightarrow unikne charakteristický rys řady.
- **mnoho bodů** \Rightarrow zvýší se výpočetní náročnost.
- **ekvidistantní diskretizace** zpravidla usnadní numerické zpracování, ale neumožňuje adaptivně měnit hustotu diskretizace v závislosti na lokálním charakteru řady.
- **při agregaci** se mohou porušit vlastnosti původní řady.

b) Problémy s kalendářem:

- různá délka kalendářních měsíců.
- 4 nebo 5 víkendů v měsíci.
- různý počet pracovních dnů v měsíci.
- pohyblivé svátky: např. svátek na začátku měsíce sníží prodej potravin za tento měsíc, ale zvýší jej za předchozí v důsledku efektu předzásobení.

Příklad ‘očistění’ časové řady např. od proměnlivé délky měsíce:

zavedeme ‘standardní’ měsíc o délce 30 dnů a pak údaj třebas o produkci za leden přenásobíme korekčním faktorem $\frac{30}{31}$.

c) **Problémy s nekompatibilitou jednotlivých měření:**

Příklad: hodnota nějakého ukazatele se jeden rok týká např. 85 podniků, další rok jen 82 apod.

d) **Problémy s délkou časových řad:**

- zvětšení počtu měření (např. půlením časových intervalů mezi body) nemusí vždy znamenat zvětšení množství informace.
- někdy se mohou objevit protichůdné tendence: metoda zpracování vyžaduje delší řadu, ale na druhé straně řada vzniklá dlouhodobým sledováním může měnit charakteristiky svého modelu v čase \Rightarrow obtíže s konstrukcí modelu.

4.1.2. *Stabilizace rozptylu.*

Většina běžně užívaných modelů předpokládá alespoň kovariančně stacionární chování analyzované řady $X := \{X_t\}$, tj. zejména konstantní rozptyl σ_X^2 (tzv. **homoskedasticita**). Pokud je tato podmínka porušena (tzv. **heteroskedasticita**), je třeba buď užít vhodný model a nebo řadu transformovat na řadu s konstantním rozptylem. Takové transformaci říkáme **stabilizace rozptylu**.

Obvykle předpokládáme exponenciální model závislosti směrodatné odchylky na střední hodnotě: $\sigma_X(t) = \sigma_0 \mu_X(t)^\Theta$. V praxi se nejčastěji vyskytují 2 případy:

- $\Theta = 0$: řada je již homoskedastická a není třeba ji transformovat;
- $\Theta = 1$: směrodatná odchylka řady závisí lineárně na střední hodnotě.

Existují postupy, které umožňují odhadnout parametr Θ z pozorované řady. Heteroskedasticitu pak odstraníme jednou z níže uvedených transformací, kde zvolíme parametr $\lambda = 1 - \Theta$.

Box-Coxova transformace pro řadu $X_t > 0$:

$$Y_t := \begin{cases} \frac{X_t^\lambda - 1}{\lambda} & \text{pro } \lambda \neq 0 \\ \ln(X_t) & \text{pro } \lambda = 0 \text{ (lineární nárůst při } \Theta = 1) \end{cases}.$$

Mocninná transformace pro řadu $X_t > 0$:

$$Y_t := \begin{cases} X_t^\lambda & \text{pro } \lambda \neq 0 \\ \ln(X_t) & \text{pro } \lambda = 0 \text{ (lineární nárůst při } \Theta = 1) \end{cases}.$$

Poznámka 4.1.

- (1) Jestliže není splněna podmínka $X_t > 0$ nebo pozorované hodnoty jsou blízké nule, nelze uvedené transformace přímo použít. V takovém případě je možno řadu vhodně posunout do kladných hodnot. Tento posuv však představuje natolik umělý zásah do stochastického chování originální řady, že může vést k jejímu znehodnocení a nevěrohodnosti výsledného modelování.
- (2) Mocninná transformace není spojitá pro $\lambda \rightarrow 0$, takže je nutno se vyhýbat malým nenulovým hodnotám parametru λ .
- (3) Postup odhadování parametru Θ , resp. λ lze nalézt např. v [Cip86, s.144-145].

4.1.3. *Identifikace periodických komponent.*

Identifikace je založena na rozkladu T -periodické funkce $x(t)$ do její

Fourierovy řady:

$$x(t) = \sum_{k=-\infty}^{\infty} c_k e^{\frac{i2\pi kt}{T}} = \frac{a_0}{2} + \sum_{k=1}^{\infty} A_k \cos\left(\frac{2\pi kt}{T} - \varphi_k\right),$$

kde se pak snažíme identifikovat energeticky nejsilnější harmonické komponenty, tj. takové, které mají velkou hodnotu kvadrátu své amplitudy A_k^2 , což je veličina úměrná energii (nebo střednímu výkonu) této složky na intervalu délky T .

K tomu slouží tzv. **periodogram**:

Periodogram = odhad energetické spektrální hustoty reálné T -periodické

funkce $x(t)$, kde za c_k , $k = 1, \dots, m$, $m := \lfloor \frac{N-1}{2} \rfloor$, dosadíme odhad \hat{c}_k spočtený pomocí DFT_N^- (operátor Diskrétní Fourierovy transformace — viz dále) po ekvidistantní diskretizaci jedné periody $x(t)$:

$T = N\Delta t$, $x_n = x(n\Delta t)$, $n = 0, 1, \dots, N$, $\mathbf{x} := [x_0, \dots, x_{N-1}]$.

Periodogram je tedy posloupnost hodnot $\{I_k\}_{k=1}^m$, kde

$$I_k = I(\omega_k) = 2T|\hat{c}_k|^2 = 2N\Delta t \left| \frac{1}{N} X_k \right|^2 = \frac{2}{N} \Delta t |X_k|^2, \quad k = 1, \dots, m \quad (4.1)$$

a $\mathbf{X} := [X_1, \dots, X_N]$ je transformací vektoru \mathbf{x} pomocí DFT_N^- :

$$X_k = \sum_{t=0}^{N-1} x_t e^{-i \frac{2\pi kt}{N}} = \sum_{t=1}^N x_t e^{-i \frac{2\pi kt}{N}}.$$

Druhá rovnost je zde důsledkem N -periodicity $x_0 = x_N$.

Z hlediska kvalitativních závěrů nezáleží na multiplikační konstantě, proto bez újmy na obecnosti můžeme předpokládat $\Delta t = 1$ a dostáváme tak **výsledný vztah pro hodnoty I_k periodogramu** ve tvaru

$$I_k = I(\omega_k) = \frac{2}{N} |X_k|^2 = \quad (4.2)$$

$$= \frac{2}{N} \left\{ \left(\sum_{t=1}^N x_t \cos \frac{2\pi kt}{N} \right)^2 + \left(\sum_{t=1}^N x_t \sin \frac{2\pi kt}{N} \right)^2 \right\}, \quad k = 1, \dots, m. \quad (4.3)$$

Velká hodnota I_k indikuje velkou energii k -té harmonické komponenty, tj. silné zastoupení složky $x_k(t) = c_k e^{i\omega_k t} + c_{-k} e^{-i\omega_k t} = A_k \cos(\omega_k t - \varphi_k)$ v rozvoji $x(t)$ do Fourierovy řady ($\omega_k := \frac{2\pi k}{T}$).

Věta 4.2.

$$I_k = 2 \sum_{|h| < N} \hat{\gamma}(h) e^{-i \frac{2\pi h k}{N}} \text{ pro } k = 1, 2, \dots, \left\lfloor \frac{N-1}{2} \right\rfloor,$$

kde $\hat{\gamma}(h)$ je odhad autokovarianční funkce vektoru $\mathbf{x} = (x_1, \dots, x_N)^T$ spočtený dle 1.46.

Důkaz.

$$I_k = \frac{2}{N} |X_k|^2 = \frac{2}{N} X_k \bar{X}_k = \frac{2}{N} \left(\sum_{s=1}^N x_s e^{-i \frac{2\pi s k}{N}} \right) \left(\sum_{t=1}^N \bar{x}_t e^{i \frac{2\pi t k}{N}} \right).$$

Protože platí

$$\sum_{s=1}^N e^{-i \frac{2\pi s k}{N}} = \sum_{t=1}^N e^{i \frac{2\pi t k}{N}} = 0 \text{ pro } k \neq 0 \pmod{N},$$

lze psát

$$\begin{aligned} I_k &= \frac{2}{N} \left(\sum_{s=1}^N (x_s - \hat{x}) e^{-i \frac{2\pi s k}{N}} \right) \left(\sum_{t=1}^N \overline{(x_t - \hat{x})} e^{i \frac{2\pi t k}{N}} \right) = \\ &= \frac{2}{N} \sum_{s=1}^N \sum_{t=1}^N (x_s - \hat{x}) \overline{(x_t - \hat{x})} e^{-i \frac{2\pi (s-t) k}{N}} \quad h \equiv s-t \\ &= 2 \sum_{|h| < N} \frac{1}{N} \underbrace{\sum_{t=1}^{N-h} (x_{t+h} - \hat{x}) \overline{(x_t - \hat{x})}}_{\hat{\gamma}(h)} e^{-i \frac{2\pi h k}{N}}. \end{aligned}$$

□

4.1.4. Fisherův test periodicity.

Položme $m = \lfloor \frac{N-1}{2} \rfloor$ a definujme statistiku

$$W = \max_{k=1, \dots, m} Y_k, \quad \text{kde} \quad Y_k = \frac{I_k}{\sum_{j=1}^m I_j}.$$

Protože $I_j \geq 0$ pro $j = 1, 2, \dots, m$, je $0 \leq W \leq 1$.

Nechť $X_t = P_t + E_t$, $E_t \sim WN(0, \sigma^2)$ gaussovský.

Platí-li nulová hypotéza $H_0 : X_t = E_t$, pak testová statistika W má na intervalu $[0, 1]$ rozdělení, pro něž platí

$$P(W > x) = \sum_{\substack{j=1 \\ 1-jx > 0}}^m (-1)^{j-1} \binom{m}{j} (1-jx)^{m-1}, \quad 0 < x < 1,$$

přičemž $P(W > x) \approx m(1-x)^{m-1}$ je dobrá aproximace pro $m \leq 50$.

Nechť $g_F(1-\alpha)$ značí $(1-\alpha)$ -kvantil rozdělení statistiky W (tabelováno), tj.

$P(W \leq g_F(1-\alpha)) = 1-\alpha$. Pak H_0 zamítáme s rizikem α , pokud $W > g_F(1-\alpha)$. Je-li v takovém případě $I_{k_0} = \max_{k=1, \dots, m} I_k$, pak k_0 -tou harmonickou komponentu považujeme za významnou. Další významnou harmonickou komponentu zjistíme z periodogramu délky $m-1$, kde vynecháme I_{k_0} , a tak pokračujeme dále, dokud není hypotéza H_0 přijata.

4.1.5. Siegelův test periodicity.

Tento test je vhodnější než Fisherův v případě většího množství harmonických komponent. Siegelova statistika je tvaru

$$T_\lambda = \sum_{k=1}^m (Y_k - \lambda g_F(1-\alpha))^+, \quad 0 < \lambda < 1 \quad (\text{doporučená hodnota je } \lambda = 0,6).$$

H_0 zamítáme s rizikem α , jestliže $T_\lambda > t_\lambda(1 - \alpha)$, kde $t_\lambda(1 - \alpha)$ je $(1 - \alpha)$ -kvantil rozdělení T_λ (tabelováno).

Poznámka 4.1 (Praktická doporučení).

- (1) Vzhledem k povaze nulové hypotézy ve výše uvedených testech by v testovaných datech neměl být obsažen trend. Doporučuje se protom alespoň hrubé předběžné odstranění dlouhodobého trendu. Jeho přítomnost totiž snižuje citlivost testů.
- (2) Vzhledem k povaze rozvoje do Fourierovy řady musí být N dělitelné délkami period všech harmonických složek, které se snažíme detekovat (obvykle stačí, aby N bylo dělitelné délkou dominantní periody). V případě potřeby je proto třeba od začátku řady ubrat co nejmenší počet pozorování a snížit tak celkovou délku n na $N = n - r$, které vyhovuje. Činíme tak ovšem jen pro účel identifikace periodických komponent.
- (3) Požadavek (2) je obtížné splnit, jestliže délka dominantní periody není předem známa. V takovém případě postupně zkoušíme $r = 0, 1, \dots$ a vybereme $N = n - r$, pro něž jsou v periodogramu nejvyšší a nejostřejší špičky (lokální maxima), z nichž je co nejvíce vůči sobě navzájem harmonických (perioda jedné komponenty dělí periodu jiné komponenty).

4.2. DEKOMPOZIČNÍ MODEL.

$$\text{Aditivní model (AM): } X_t = \underbrace{Tr_t + \overbrace{Sz_t + C_t}^{P_t}}_{D_t} + E_t,$$

$$\text{Multiplikativní model (MM): } X_t = Tr_t \cdot \underbrace{\overbrace{Sz_t \cdot C_t}^{P_t}}_{D_t} \cdot E_t,$$

kde značí

$Tr_t \dots$ dlouhodobý trend

Sz_t ... sezónní komponenta (perioda je předem známa)
 C_t ... cyklická komponenta (perioda není předem známa)
 P_t ... celková periodická komponenta
 $D_t = \mu_X(t)$... deterministická komponenta (střední hodnota)
 E_t ... stacionární náhodná komponenta (obvykle bílý šum nebo IID):
 $\mu_E = 0$ pro AM, resp. $\mu_E = 1$ pro MM.
 AM lze použít jen pro homoskedastická data, neboť $\sigma_X(t) = \sigma(D_t + E_t) = \sigma(E_t) = \sigma_E$ je konstantní vzhledem ke stacionaritě E_t .
 Naopak MM lze použít jen pro lineárně heteroskedastická data, neboť $\sigma_X(t) = \sigma(D_t \cdot E_t) = D_t \sigma(E_t) = \mu_X(t) \sigma_E$ závisí lineárně na střední hodnotě $\mu_X(t)$. V takovém případě lze MM převést na AM logaritmickou transformací (srov. s 4.1.2).

4.2.1. Parametrizované modely.

Označení .

$\mathbf{t} = (t_1, \dots, t_n)^T$, $t_i \neq t_j$ pro $i \neq j$... vektor "času",
 $\mathbb{X} := \mathbb{X}_{\mathbf{t}} = (X_{t_1}, \dots, X_{t_n})^T$... náhodný vektor populace vybrané z časové řady,
 $\mathbf{x} := \mathbf{x}_{\mathbf{t}} = (x_1, \dots, x_n)^T$, $x_i = X_{t_i}(\omega)$... pozorování vektoru populace \mathbb{X} .

A. Parametrické metody (P).

Tyto metody používáme v případě, že je znám analytický tvar $D_t = D(t; \beta_1, \dots, \beta_p)$ v aditivním modelu $X_t = D_t + E_t$ (multiplikatívni model převedeme na aditivní logaritmováním) s neznámými parametry $\beta = (\beta_1, \dots, \beta_p)^T$, které odhadneme minimalizací výrazu $N(\beta_1, \dots, \beta_p) = \|D(\mathbf{t}; \beta_1, \dots, \beta_p) - \mathbf{x}\|^2$ ve vhodné normě $\|\cdot\|$. Zpravidla se používá euklidovská norma a klasická metoda nejmenších čtverců:

$$N(\beta_1, \dots, \beta_p) = \sum_{i=1}^n |D(t_i, \beta_1, \dots, \beta_p) - x_i|^2.$$

Lokální extrém hledáme řešením systému rovnic

$$\frac{\partial N(\beta_1, \dots, \beta_p)}{\partial \beta_j} = 0 \text{ pro } j = 1, \dots, p$$

s ověřením podmínky pro lokální minimum (Hessova matice je pozitivně definitní):

$$J = \left[\frac{\partial^2 N(\beta_1, \dots, \beta_p)}{\partial \beta_i \partial \beta_j} \right]_{i,j} > 0.$$

Pokud D závisí lineárně na β_1, \dots, β_p , pak tento systém je systémem p lineárních rovnic pro p neznámých β_1, \dots, β_p a dostáváme klasický **lineární regresní model**.

V případě nelineární regrese použijeme pro minimalizaci $N(\beta_1, \dots, \beta_p)$ **vhodnou optimalizační metodu**, například *Optimization Toolbox* v prostředí systému MATLAB.

Lineární regresní model: $D(t; \beta_1, \dots, \beta_p) = \varphi_1(t)\beta_1 + \dots + \varphi_p(t)\beta_p$, kde φ_i jsou dané.

$$\mathbb{X} = D\beta + \mathbb{E}, \quad \mathbb{E} = (E_{t_1}, \dots, E_{t_n})^T \sim \text{IID}(0, \sigma^2),$$

kde

$$D = [\varphi_1(\mathbf{t}), \dots, \varphi_p(\mathbf{t})], \quad \varphi_j(\mathbf{t}) = (\varphi_j(t_1), \dots, \varphi_j(t_n))^T$$

je matice s plnou sloupcovou hodnotí p .

Pro vektor pozorování přejde výše uvedený model v

$$\mathbf{x} = D\beta + \mathbf{e}, \quad \mathbf{e} = (e_1, \dots, e_n)^T, \text{ kde } e_i = E_{t_i}(\omega).$$

Klasické dekompoziční metody

- **Metoda malého trendu** předpokládá v každé sezónní periodě přibližně konstantní trend, který se odhadne jako průměr. Po jeho odečtení odhadneme sezónní složku sezónním průměrováním (viz [BD93, str. 21]).

- **Metoda sezónních klouzavých průměrů** využívá pro odhad trendu obyčejných klouzavých průměrů s délkou rovnou sezónní periodě. Dále postupujeme jako v předchozím případě (viz [BD93, str. 22]).
- **lineární regresní model (ozn. polfs)**, v němž trend je modelován polynomem (pol) a periodická složka zkrácenou Fourierovou řadou (fs=Fourier Series) v goniometrickém tvaru nazývaný též **metoda skrytých period**, do níž zařadíme pouze významné harmonické složky s periodou T_j dle periodogramu:

$$Tr_t = \beta_1 t^p + \beta_2 t^{p-1} + \dots + \beta_{p+1}, \quad 0 \leq p \quad (4.4a)$$

$$P_t = \sum_{j=1}^q a_j \cos\left(\frac{2\pi t}{T_j}\right) + b_j \sin\left(\frac{2\pi t}{T_j}\right). \quad (4.4b)$$

V modelu tedy odhadujeme $(p+1)+2q$ neznámých parametrů $\beta_i, i = 1, 2, \dots, p+1$ a $a_j, b_j, j = 1, 2, \dots, q$.

- **regresní model (ozn. polper)**, v němž trend je opět modelován polynomem (4.4a) a u periodické složky P_t (per) jsou její hodnoty v jedné její základní (sezónní) periodě přímo odhadovanými parametry: $s_k := P_k$ pro $k = 2, \dots, s$ a $s_1 := P_1$ určíme z dodatečné podmínky $\frac{1}{s} \sum_{k=1}^s s_k = 0$: Celkem tedy v modelu odhadujeme $p+1+s-1 = p+s$ parametrů a pak klademe $P_t = s_k, t = (j-1)s+k$ (tj., je-li P_t k -tou hodnotou v j -té základní periodě).

4.2.2. Filtrační techniky.

Za vhodných předpokladů konstruujeme lineární nebo nelineární operátor S (smoother = vyhlazovač) takový, že

$$S(\{X_t\}) =: \{\hat{D}_t\} \approx \{D_t\}$$

je dobrý odhad D_t v tom smyslu, že $\{\hat{D}_t\} \rightarrow \{D_t\}$ ve vhodné konvergenci při $n \rightarrow \infty$, tj. odhad je **asymptoticky přesný**.

- Metoda klouzavých průměrů (lineární konvoluční filtr):

$$Y_t = \sum_{\tau=-q}^q w_\tau X_{t+\tau} = \underbrace{\sum_{\tau=-q}^q w_\tau D_{t+\tau}}_{\hat{D}_t := S(\{D_t\})} + \underbrace{\sum_{\tau=-q}^q w_\tau E_{t+\tau}}_{\hat{E}_t := S(\{E_t\})},$$

kde požadujeme:

- (1) Co nejmenší zkreslení D_t , tj. $D_t \approx \sum_{\tau=-q}^q w_\tau D_{t+\tau}$.

Standardní požadavek je, aby bez zkreslení byly filtrem zachovány polynomy dostatečně vysokého stupně. Je-li D_t po částech polynomického tvaru, bude pak aproximace \hat{D}_t velmi dobrá. Odpověď dává následující věta:

Věta 4.2. Platí $P_t = \sum_{\tau=-q}^q P_{t+\tau}$ pro každý polynom $P_t = c_0 + c_1 t + \dots + c_k t^k$ nejvýše k -ho stupně právě když váhy splňují následující 2 podmínky

$$\begin{aligned} \sum_{\tau=-q}^q w_\tau &= 1 \\ \sum_{\tau=-q}^q \tau^r w_\tau &= 0 \text{ pro } r = 1, \dots, k \end{aligned} \tag{4.5}$$

- (2) Co největší redukci rozptylu náhodné složky E_t , tj.

$$0 \approx \sum_{\tau=-q}^q w_\tau E_{t+\tau}.$$

Mírou pro tento účel je tzv. **redukční intenzita klouzavých průměrů** definovaná jako podíl rozptylů σ_Y^2 / σ_X^2 za předpokladu, že $E_t \sim WN(0, \sigma^2)$.

Spočtíme nejprve rozptyl σ_Y^2 :

$$\begin{aligned} \sigma_Y^2 &= \text{var} \hat{E}_t = \text{var} \left(\sum_{\tau=-q}^q w_\tau E_{t+\tau} \right) = \\ &= \sum_{\tau=-q}^q w_\tau^2 \text{var}(E_{t+\tau}) = \sigma^2 \sum_{\tau=-q}^q w_\tau^2 = \sigma^2 \|\mathbf{w}\|^2. \end{aligned}$$

$$\text{Odtud } \sigma_Y^2 / \sigma_X^2 = \text{var} \hat{E}_t / \text{var} E_t = \sigma^2 \|\mathbf{w}\|^2 / \sigma^2 = \|\mathbf{w}\|^2.$$

Požadujeme tedy $\|\mathbf{w}\|^2 < 1$ co nejmenší. Dá se ukázat,

že z tohoto pohledu vyhovují nejlépe obyčejné klouzavé průměry $w_\tau = \frac{1}{2q+1}$, kde

$\|\mathbf{w}\|^2 = \sum_{\tau=-q}^q \frac{1}{(2q+1)^2} = \frac{2q+1}{(2q+1)^2} = \frac{1}{2q+1}$. Ty však zase nejsou příliš optimální z hlediska požadavku (1). Zachovávají totiž pouze polynomy nejvýše stupně 1, tj. (po částech) lineární průběh D_t a nehodí se tedy na řady s vyšší dynamikou ve střední hodnotě.

Klasickou technikou hledání vah je **postupná polynomiální regrese**, kdy postupně zleva prokládáme segmenty dat délky $2q+1$ regresní polynom zvoleného stupně. Jsou-li data ekvidistantní, pak takto získané váhy nezávisí na poloze segmentu a jsou tedy korektně definovány.

Hledání **optimálních vah** zachovávajících polynomy do zadaného stupně při co nejlepší redukční intenzitě vedou na úlohu kvadratického programování (minimalizujeme $\|\mathbf{w}\|^2 = \sum_{\tau=-q}^q w_\tau^2$) při lineárních omezeních daných rovnicemi (4.5). Příkladem kvalitně navržených vah jsou tzv. **Spencerovy 15-ti bodové váhy**:

$[w_0, \dots, w_7] = \frac{1}{320} [74, 67, 46, 21, 3, -5, -6, -3]$, kde

$w_{-\tau} = w_\tau$ pro $|\tau| \leq 7$. Tyto váhy zachovávají polynomy až do stupně 3 při redukční intenzitě 0.1926, což je velmi dobrý výsledek ve srovnání s dolní mezí $\frac{1}{15} = 0.0667$ danou obyčejnými klouzavými průměry délky 15.

- Exponenciální vyrovnávání
- Jádrové vyhlazování
- Waveletové vyhlazování

4.3. BOX-JENKINSOVY MODELY.

4.3.1. ARMA modely pro stacionární časové řady.

Definice 4.3 (ARMA proces).

Stochastický proces $X = \{X_t \mid t \in \mathbb{Z}\}$ se nazývá **ARMA procesem** řádu p, q ($0 \leq p, q < \infty$), píšeme $X \sim ARMA(p, q)$, jestliže

$$X_t = \underbrace{\Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p}}_{\text{Autoregresní část (AR)}} + \underbrace{Z_t + \Theta_1 Z_{t-1} + \Theta_2 Z_{t-2} + \cdots + \Theta_q Z_{t-q}}_{\text{Část klouzavých součtů (MA=Moving Average)}}, \quad (4.6a)$$

kde $Z := \{Z_t \mid t \in \mathbb{Z}\} \sim WN(0, \sigma^2)$ a $\Phi_p \neq 0$, $\Theta_q \neq 0$, $\sigma \neq 0$.
Přepíšeme-li (4.6a) do ekvivalentního tvaru

$$X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \cdots - \Phi_p X_{t-p} = Z_t + \Theta_1 Z_{t-1} + \Theta_2 Z_{t-2} + \cdots + \Theta_q Z_{t-q}, \quad (4.6b)$$

lze použít též zkrácený zápis dle poznámky 2.27:

$$\Phi(B)X_t = \Theta(B)Z_t \quad (4.6c)$$

kde při $\Phi_0 = \Theta_0 = 1$ je

$$\Phi(z) = 1 - \Phi_1 z - \cdots - \Phi_p z^p, \quad \Theta(z) = 1 + \Theta_1 z + \cdots + \Theta_q z^q, \quad z \in \mathbb{C}.$$

Poznámka 4.4. V předchozí definici bez újmy na obecnosti předpokládáme $\Theta_0 = 1$, neboť v opačném případě stačí nahradit bílý šum $\{Z_t\} \sim WN(0, \sigma^2)$ šumem $\{\Theta_0 Z_t\} \sim WN(0, (\Theta_0 \sigma)^2)$.

a) Autoregresní proces (AR proces):

$X \sim AR(p) = ARMA(p, 0)$:
 $\Phi(B)X_t = Z_t$, neboť $\Theta(z) \equiv 1$.
Tedy

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + Z_t. \quad (4.6d)$$

V tomto případě připouštíme i $p = \infty$ za předpokladu, že $\Phi := \{\Phi_j\}_{j=1}^{\infty} \in \ell_1$.

b) Proces klouzavých součtů (MA proces):

$X \sim MA(q) = ARMA(0, q)$:
 $X_t = \Theta(B)Z_t$, neboť $\Phi(z) \equiv 1$.
Tedy

$$X_t = Z_t + \Theta_1 Z_{t-1} + \Theta_2 Z_{t-2} + \cdots + \Theta_q Z_{t-q}. \quad (4.6e)$$

V tomto případě připouštíme i $q = \infty$ za předpokladu, že $\Theta := \{\Theta_j\}_{j=1}^{\infty} \in \ell_1$.

c) Bílý šum (jediný proces, který je současně AR i MA):

$X \sim ARMA(0, 0) = AR(0) = MA(0) = WN(0, \sigma^2)$: $X_t = Z_t$.

d) Smíšený ARMA proces:

$X \sim ARMA(p, q)$, $0 < p, q < \infty$: Netriviální směs autoregrese a klouzavých součtů.

Definice 4.6.

Buď $X = \{X_t | t \in \mathbb{Z}\}$, $X \sim ARMA(p, q)$. X se nazývá **kauzální ARMA proces**, jestliže existuje $\psi = \{\psi_j\}_{j=0}^{\infty}$, $\sum_{j=0}^{\infty} |\psi_j| < \infty$ (tj.

$\psi \in \ell_1$) tak, že

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad (\text{zkráceně } X_t = \psi(B)Z_t), \quad t \in \mathbb{Z}. \quad (4.7a)$$

X se nazývá **invertibilní ARMA proces**, jestliže existuje

$\pi = \{\pi_j\}_{j=0}^{\infty}$, $\sum_{j=0}^{\infty} |\pi_j| < \infty$ (tj. $\pi \in \ell_1$) tak, že

$$\sum_{j=0}^{\infty} \pi_j X_{t-j} = Z_t \quad (\text{zkráceně } \pi(B)X_t = Z_t), \quad t \in \mathbb{Z}. \quad (4.7b)$$

Poznámka 4.7.

Tedy **kauzalita** znamená, že ARMA proces X je kauzálním lineárním procesem, který lze generovat bílým šumem $\{Z_t\}$, neboli $X \sim MA(\infty)$.

Podobně **invertibilita** znamená, že bílý šum $\{Z_t\}$ je kauzálním lineárním procesem, který je generován ARMA procesem X , což je ekvivalentní s $X \sim AR(\infty)$.

Věta 4.8. *Nechť $\{X_t\} \sim ARMA(p, q) : \Phi(B)X_t = \Theta(B)Z_t$ takový, že polynomy $\Phi(z)$ a $\Theta(z)$ nemají společné kořeny. Pak platí*

- (i) $\{X_t\}$ je kauzální právě když $\Phi(z) \neq 0$ pro všechna $z \in \mathbb{C}$, $|z| \leq 1$ (všechny kořeny $\Phi(z)$ musí být vně uzavřeného jednotkového kruhu). V takovém případě ψ_j jsou určeny Taylorovým rozvojem racionální funkce lomené $\psi(z) := \frac{\Theta(z)}{\Phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j$, přičemž posloupnost $\psi := \{\psi_j\} \in \ell_1$ a je určena jednoznačně.
- (ii) $\{X_t\}$ je invertibilní právě když $\Theta(z) \neq 0$ pro všechna $z \in \mathbb{C}$, $|z| \leq 1$ (všechny kořeny $\Theta(z)$ musí být vně uzavřeného jednotkového kruhu). V takovém případě π_j jsou určeny Taylorovým rozvojem racionální funkce lomené $\pi(z) := \frac{\Phi(z)}{\Theta(z)} = \sum_{j=0}^{\infty} \pi_j z^j$, přičemž posloupnost $\pi := \{\pi_j\} \in \ell_1$ a je určena jednoznačně.

Věta 4.9 (Autokovarianční funkce MA procesu).
 $\{X_t\} \sim MA(q)$, $q \leq \infty$ je stacionární náhodný proces s nulovou střední hodnotou $\mu_X = 0$ a autokovarianční funkcí

$$\gamma_X(h) = \sigma^2 \sum_{k=0}^q \Theta_{h+k} \Theta_k \text{ pro } h \geq 0. \quad (4.8a)$$

Speciálně pro $q < \infty$ je

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{k=0}^{q-h} \Theta_{h+k} \Theta_k & \text{pro } 0 \leq h \leq q \\ 0 & \text{pro } h > q \end{cases} \quad (4.8b)$$

a zejména $\gamma_X(q) = \sigma^2 \Theta_q \neq 0$, neboť $\Theta_0 = 1$.

Důkaz. Vztahy jsou důsledkem 2.24. □

Důsledek 4.10.

$$\sigma_X^2 = \gamma_X(0) = \sigma^2 \sum_{k=0}^q |\Theta_k|^2. \quad (4.9)$$

$$\sigma_X^2 = \sigma^2(1 + |\Theta_1|^2 + |\Theta_2|^2 + \dots + |\Theta_q|^2) \text{ pro } q < \infty.$$

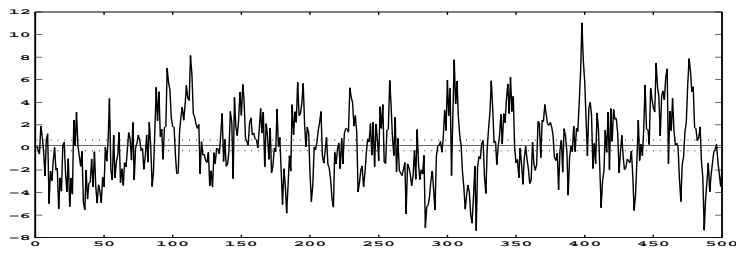
$$\rho_X(h) = \frac{\sum_{k=0}^{q-h} \Theta_{h+k} \Theta_k}{\sum_{k=0}^q |\Theta_k|^2} \text{ pro } h \geq 0. \quad (4.10)$$

Věta 4.11 (Parciální autokorelační funkce kauzálního AR procesu).
Nechť $\{X_t\} \sim AR(p)$, $p < \infty$ je kauzální AR proces. Pak $\{X_t\}$ je stacionární s nulovou střední hodnotou $\mu_X = 0$ a parciální autokorelační funkcí α_X , pro niž $\alpha_X(p) = \Phi_p \neq 0$ a $\alpha_X(h) = 0$ pro $h > p$.

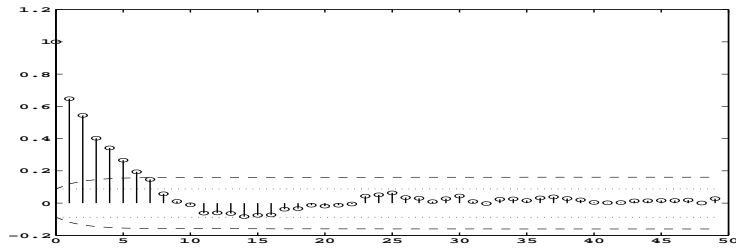
Na obrázcích 2, 3 a 4 jsou typické průběhy odhadnuté autokorelační a parciální autokorelační funkce simulovaných procesů po řadě $AR(2)$, $MA(2)$ a $ARMA(2,2)$. Čárkovaný pás udává interval spolehlivosti 95% pro nulovou hodnotu. Vidíme, že v případě procesu $AR(2)$ na obr. 2, resp. $MA(2)$ na obr. 3 skutečně $\alpha_X(h) \approx 0$, resp. $\rho_X(h) \approx 0$ pro $h > 2$ v souladu s větami 4.11, resp. 4.9. V ostatních případech obálka $\rho_X(h)$, resp. $\alpha_X(h)$ (v případě $ARMA(2,2)$ na obr. 4 dokonce

obou) vykazuje exponenciálně klesající, případně navíc i oscilatorický, průběh. Dá se totiž ukázat, že ρ_X i α_X jsou v těchto případech lineární kombinací klesajících geometrických posloupností a kosinusovek s geometricky klesající amplitudou.

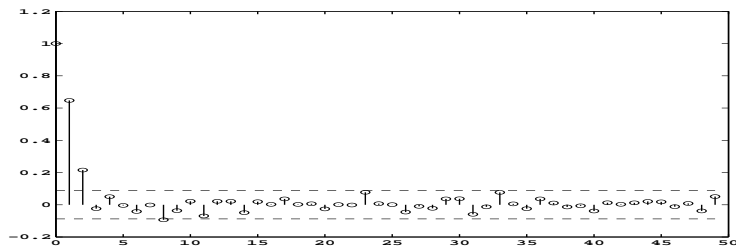
Pozorování procesu, $n = 500$



Autokorelační funkce

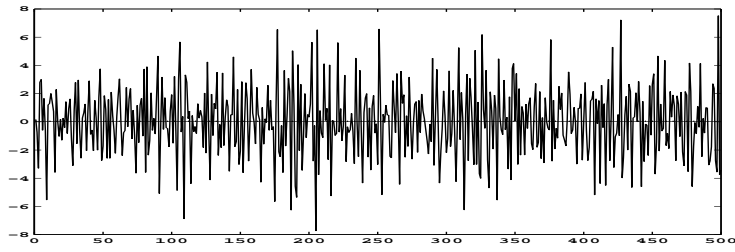


Parciální autokorelační funkce

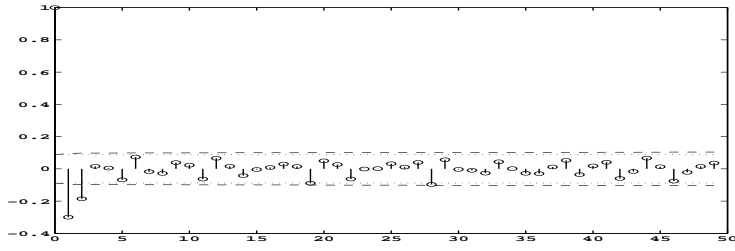


OBRÁZEK 2. $AR(2) : \Phi = [0.5, 0.2], \sigma^2 = 2.25$.

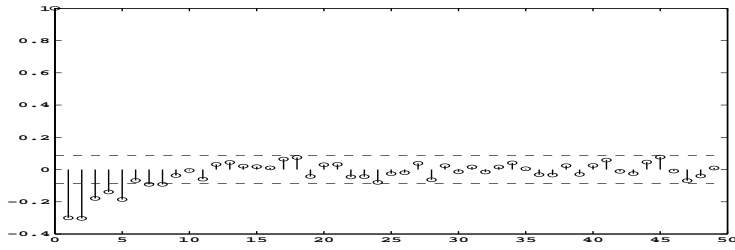
Pozorování procesu, $n = 500$



Autokorelační funkce

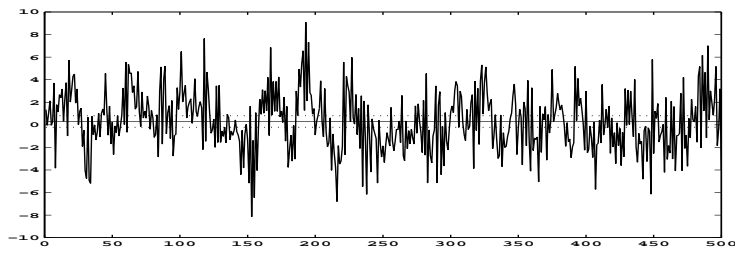


Parciální autokorelační funkce

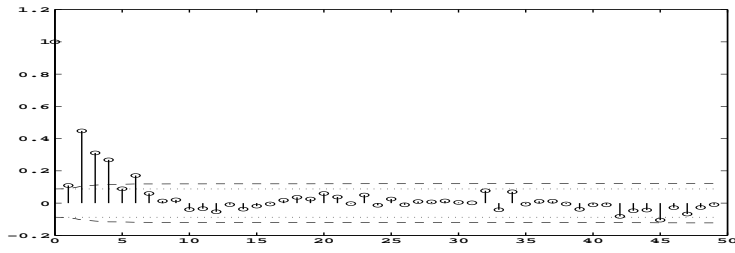


OBRÁZEK 3. $MA(2)$: $\Theta = [-0.5, -0.2]$, $\sigma^2 = 2.25$.

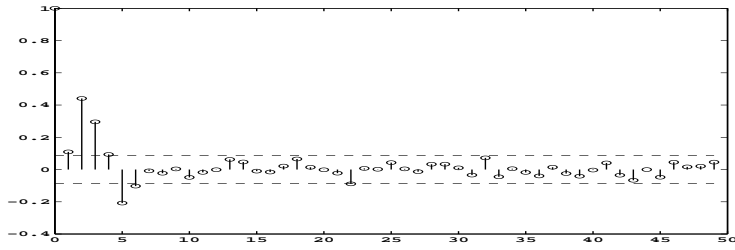
Pozorování procesu, $n = 500$



Autokorelační funkce



Parciální autokorelační funkce



OBRÁZEK 4. $ARMA(2, 2)$: $\Phi = [0.5, 0.2]$, $\Theta = [-0.6, 0.3]$, $\sigma^2 = 2.25$.

4.3.2. (S)ARIMA modely pro kovariančně stacionární časové řady.

Do popisu těchto modelů vstupuje operátor diferencování, který lze také ekvivalentně vyjádřit pomocí operátoru zpětného posuvu B popsaném v poznámce 2.27:

Diferencování časové řady na vzdálenost $s \in \mathbb{N}$:

$Y_t := \Delta_s X_t := X_t - X_{t-s} = X_t - B^s X_t = (1 - B^s)X_t$, atd. rekurentně pro libovolný řád diferencování $d \in \mathbb{N}$:

$Y_t := \Delta_s^d X_t := \Delta_s(\Delta_s^{d-1} X_t) = (1 - B^s)^d X_t$.

Pro $s > 1$ se $\Delta_s^d = (1 - B^s)^d$ také nazývá operátorem **sezónního diferencování řádu d** , neboť s hraje roli délky sezónní periody.

Pro $s = 1$ značíme $\Delta^d := \Delta_1^d = (1 - B)^d$ a operátor nazýváme operátorem **nesezónního diferencování řádu d** .

Konstrukce modelů:

$\{X_t\} \sim SARIMA(p, d, q, P, D, Q, s)$, jestliže pro $V_t := \Delta_s^D X_t$ platí:

$$V_t = \underbrace{\tilde{\Phi}_1 V_{t-s} + \tilde{\Phi}_2 V_{t-2s} + \cdots + \tilde{\Phi}_P V_{t-Ps}}_{\text{Sezónní autoregresní část}} + E_t + \underbrace{\tilde{\Theta}_1 E_{t-s} + \tilde{\Theta}_2 E_{t-2s} + \cdots + \tilde{\Theta}_Q E_{t-Qs}}_{\text{Sezónní část klouzavých součtů}}, \quad (4.11a)$$

kde $\{\Delta^d E_t\} \sim ARMA(p, q)$ s parametry jako v (4.11b), přičemž $P, Q, d, D \geq 0$ a $s > 1$ jsou opět celá čísla a $\tilde{\Phi}_P \neq 0, \tilde{\Theta}_Q \neq 0$.

$\{X_t\} \sim ARIMA(p, d, q) = SARIMA(p, d, q, 0, 0, 0, 1)$ neboli

$X_t = V_t = E_t$, tj. pro $W_t := \Delta^d X_t \sim ARMA(p, q)$ platí:

$$W_t = \Phi_1 W_{t-1} + \Phi_2 W_{t-2} + \cdots + \Phi_p W_{t-p} + Z_t + \Theta_1 Z_{t-1} + \Theta_2 Z_{t-2} + \cdots + \Theta_q Z_{t-q}. \quad (4.11b)$$

Ve všech výše uvedených modelech navíc předpokládáme, že popisují **kauzální** časovou řadu, neboli současně platí $\{X_t\} \sim MA(\infty)$,

přičemž posloupnost MA-koeficientů je absolutně sumovatelná a příslušná řada konverguje dle kvadratického středu.

Použití a interpretace modelu *ARIMA*:

Jestliže aplikujeme operátor Δ na (alespoň lokálně) polynomiální průběh, dojde ke snížení stupně polynomu (nejméně) o 1 (snadno se ověří). Tedy Δ^d sníží stupeň nejméně o d . Model *ARIMA*(p, d, q) se tedy hodí na stacionárně kovarianční časovou řadu $X := \{X\}_t$, kde stacionarita je porušena jen ve střední hodnotě $\mu_X(t)$, která není konstantní a má po částech polynomiální charakter stupně nejvýše d . Nesezónním diferencováním řádu d dosáhneme buď konstantní nebo dokonce nulové střední hodnoty u diferencované řady. Je-li takto získaná řada stacionární (v což obvykle doufáme), pak lze použít model *ARIMA*.

Použití a interpretace modelu *SARIMA*:

Zde předpokládáme, že všechny částečné mezisezónní řady $\{X_{k+\tau s}\}_\tau$ pro $k = 0, 1, \dots, s-1$ se chovají jako *ARIMA*(P, D, Q) s týmiž parametry $\tilde{\Phi}_i$ a $\tilde{\Theta}_j$ (tj. nezávisí na k).

Reziduální řada $\{E_t\}$ získaná z tohoto modelu je ovšem dekernelována jen mezisezónně a zatížena zbytkovým kolísáním ve střední hodnotě $\mu_E(t)$ a tudíž předpokládáme pro ni opět model *ARIMA*, avšak s jinými řády p, d, q a jinými parametry Φ_i a Θ_j .

Obvykle stačí volit $D = 1$ (odstraní mezisezónně konstantní sezónní složku). Výjimečně volíme $D = 2$, jestliže je tendence k mezisezónnímu lineárnímu nárůstu nebo poklesu. Pozorované mezisezónní řady bývají totiž většinou velmi krátké a tudíž takové chování je z nich obtížně ověřitelné.

5. DEMONSTRAČNÍ UKÁZKY POMOCÍ KNIHOVNY TSA-M

V práci [Ves00] lze nalézt podrobný popis knihovny funkcí TSA-M vytvořené autorem tohoto příspěvku jako *toolbox* pro numerické výpočetní prostředí MATLAB. Některé možnosti knihovny jsou zde ilustrovány na reálném datovém souboru zachycujícím měsíční počty úmrtí při automobilových nehodách v USA v letech 1973–1978 převzatém z [BD93, Example 1.1.6] (viz Obr. 1.6 z odst. 1.1).

V ukázkách je modelována predikce o jednu sezónu (12 měsíců) jednak pomocí klasického lineárního regresního modelu a jednak užitím modelu $SARIMA(0, 1, 1, 0, 1, 1, 12)$, který se po podrobnější analýze ukázal pro zvolená data jako optimální.

Ukázky současně názorně demonstrují posloupnost systematických kroků nutných při analýze časové řady:

Předzpracování → Volba modelu → Identifikace modelu → Předběžný odhad jeho parametrů → Finální odhad parametrů → Verifikace modelu.

LITERATURA

- [And76] Jiří Anděl, *Statistická analýza časových řad*, SNTL, Praha, 1976.
- [BD93] Peter J. Brockwell and Richard A. Davis, *Time series: Theory and methods*, 2-nd ed., Springer-Verlag, New York, 1991 (corrected 2-nd printing 1993).
- [BD94] ———, *ITSM for Windows. A user's guide to time series modelling and forecasting*, Springer-Verlag, New York, 1994, 2 diskettes included.
- [BD02] ———, *Introduction to time series and forecasting*, 2-nd ed., Springer-Verlag, New York, 2002.
- [Cip86] Tomáš Cipra, *Analýza časových řad s aplikacemi v ekonomii*, SNTL, Praha, 1986.
- [Ham94] James D. Hamilton, *Time series analysis*, Princeton University Press, Princeton, NJ 08540, 1994.
- [Pri89] M. B. Priestley, *Spectral analysis and time series*, vol. I–II, Academic Press, London, 1989.
- [Ves00] V. Veselý, *Knihovna programů TSA-M pro analýzu časových řad*, Zborník zo XIV. letnej školy biometrie: Biometrické metódy a modely v pôdohospodárskej vede, výskume a výuke, Račkova dolina, 2. - 6. októbra 2000 (P. Flák, ed.), VUZVN (Research Inst. of Animal Production Nitra, Slovakia), ASAPV (Agency of Slovak Academy for Agricultural Science), 2000, pp. 239–248.

DOC. RNDR. VÍTĚZSLAV VESELÝ, CSC., KATEDRA APLIKOVANÉ MATEMATIKY A INFORMATIKY, EKONOMICKO-SPRÁVNÍ FAKULTA MU V BRNĚ, LIPOVÁ 41A, 602 00 BRNO

TEL.: +420 5 49498330, FAX: +420 5 49491720

E-mail address: vesely@econ.muni.cz, *URL:* <http://www.math.muni.cz/~vesely>