

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Kateřina Pechmanová

Testování jednotkového kořene s aplikací na finanční časové řady

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Jitka Zichová, Dr.

Studijní program: Matematika

Studijní obor: Finanční matematika

Praha 2014

Ráda bych poděkovala RNDr. Jitce Zichové, Dr. za odborné vedení, cenné rady a celkovou spolupráci při psaní bakalářské práce.

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 21. května 2014

Kateřina Pechmanová

Název práce: Testování jednotkového kořene s aplikací na finanční časové řady

Autor: Kateřina Pechmanová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Jitka Zichová, Dr., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Tato práce se zabývá lineárními ARMA procesy, které jsou určeny pro popis chování časových řad, a též samotnou analýzou vybraných časových řad. Nejprve jsou zavedeny základní pojmy spolu s popisem daných ARMA modelů. Poté je představen Dickeyův-Fullerův test na jednotkový kořen, jakožto přístup k ověření nestacionarity časových řad. Důležitou částí práce jsou praktické aplikace těchto modelů a testů na simulovaná a reálná data. Reálná analyzovaná data zachycují vývoj směnného kurzu koruny vůči euru. Veškeré výpočty byly prováděny v softwaru Mathematica.

Klíčová slova: ARMA, stacionarita, Dickeyův-Fullerův test

Title: Unit root testing with applications to financial time series

Author: Kateřina Pechmanová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Jitka Zichová, Dr., Department of Probability and Mathematical Statistics

Abstract: This work deals with linear ARMA processes, which are intended to describe the behavior of time series, and also with analysis of selected time series. First, the basic concepts are introduced together with the descriptions of the ARMA models. Further, the Dickey-Fuller test for a unit root, as an approach to the verification of nonstationary time series, is introduced. An important part is the practical application of these models and tests on simulated and real data. Real analyzed data capture developments in the exchange rate of Czech crown against Euro. All calculations were performed in the Mathematica software.

Keywords: ARMA, stationarity, Dickey-Fuller test

Obsah

1	Zavedení pojmů, model ARMA	3
1.1	Základní definice	3
1.2	Stacionarita a invertibilita	3
1.3	Vyjádření ve tvaru lineárního procesu	4
1.4	Autokovarianční, autokorelační a parciální autokorelační funkce	4
1.5	Model ARMA	6
1.6	Autoregresní proces AR	6
1.7	Proces klouzavých součtů MA	8
1.8	Identifikace modelu	9
2	Dickeyův-Fullerův test na jednotkový kořen	10
2.1	τ_μ -test	10
2.2	τ_τ -test	12
2.3	Interpretace výsledků	14
3	Praktická aplikace na simulovaných a reálných datech	15
3.1	Simulovaná data	15
3.1.1	Generování časové řady podle nulové hypotézy	15
3.1.2	Generování časové řady podle 1. alternativy	16
3.1.3	Generování časové řady podle 2. alternativy	17
3.2	Reálná data	18
3.2.1	Logaritmická transformace dat	18
3.2.2	Netransformovaná data	21
	Literatura	25
	Seznam obrázků	26
	Seznam tabulek	27

Úvod

Modelování časových řad se hojně využívá v oblasti ekonomie a financí, přičemž široce používaným nástrojem jsou lineární ARMA procesy. Těmi je ovšem možné modelovat pouze stacionární časové řady. Vlastnost stacionarity však značné množství reálných dat postrádá, a proto bývají využívány různé způsoby, jak nestacionaritu ověřit. Jedním z nich jsou testy na jednotkový kořen.

V kapitole 1 jsou definovány základní pojmy a podrobně popsány modely, z nichž některé zmiňujeme v dalším textu. Veškeré informace jsou převzaty z knihy Cipra (2008). V kapitole 2 je popsán Dickeyův-Fullerův test, a to ve dvou variantách, jak byl publikován v článku Dickey a Fuller (1981). Oproti článku je podrobně provedeno odvození parametrů a rozepsána konstrukce věrohodnostního poměru. Kapitola 3 se zabývá praktickou aplikací testů na konkrétní časové řady, to jest na simulovaná a reálná data.

Kapitola 1

Zavedení pojmů, model ARMA

1.1 Základní definice

Definice 1 (náhodný proces). Náhodný proces $y_t, t \in T$ je množina náhodných veličin na stejném pravděpodobnostním prostoru (Ω, \mathcal{A}, P) indexovaná pomocí hodnot t z množiny T ($T \subset \mathbb{R}$) interpretovaných jako čas. Nadále budeme uvažovat celočíselný čas t . V takovém případě lze posloupnost $y_t, t \in \mathbb{Z}$ nazvat časovou řadou.

Definice 2 (bílý šum). Bílý šum je posloupnost ε_t nekorelovaných náhodných veličin s nulovou střední hodnotou a konstantním (konečným) rozptylem $\sigma^2 > 0$:

$$E(\varepsilon_t) = 0, \quad \text{var}(\varepsilon_t) = \sigma^2 > 0, \quad \text{cov}(\varepsilon_s, \varepsilon_t) = 0 \quad \text{pro } s \neq t.$$

Definice 3 (lineární proces). Lineární proces je definován jako

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = (1 + \psi_1 B + \psi_2 B^2 + \dots) \varepsilon_t = \psi(B) \varepsilon_t, \quad (1.1)$$

kde ε_t je bílý šum a B je operátor časového posunu.

1.2 Stacionarita a invertibilita

Stacionarita časové řady y_t naznačuje, že chování dané řady je stochasticky stálé. Rozlišujeme dva druhy stacionarity:

- *striktní stacionarita*: Pravděpodobnostní chování daného modelu je invariantní k posunům v čase. Tedy pravděpodobnostní rozdělení náhodného vektoru $(y_{t_1}, \dots, y_{t_k})$ je totožné jako rozdělení vektoru $(y_{t_1+h}, \dots, y_{t_k+h})$ pro libovolné h a libovolné časy t_1, \dots, t_k .
- *(slabá) stacionarita*: Postačující je, aby byl daný proces invariantní k posunům v čase jen v rámci jednotlivých momentů do druhého řádu, tedy pro každé s a t požadujeme

$$E(y_t) = \mu = \text{konst};$$

$$\text{cov}(y_s, y_t) = E(y_s - \mu)(y_t - \mu) = \text{cov}(y_{s+h}, y_{t+h}) \quad \text{pro libovolné } h.$$

Speciálně tedy platí i

$$\text{var}(y_t) = \sigma_y^2 = \text{konst}.$$

Lineární proces (1.1) nazveme invertibilní, když jej lze zapsat ve tvaru

$$y_t = \pi_1 y_{t-1} + \pi_2 y_{t-2} + \dots + \varepsilon_t,$$

tedy $\varepsilon_t = y_t - \pi_1 y_{t-1} - \pi_2 y_{t-2} - \dots = (1 - \pi_1 B - \pi_2 B^2 - \dots)y_t = \pi(B)y_t$.

(1.2)

Pro invertibilitu lineárního procesu je postačující podmínkou konvergence mocninné řady $\pi(z)$ pro $|z| \leq 1$, tedy konvergence na jednotkovém kruhu v komplexní rovině.

1.3 Vyjádření ve tvaru lineárního procesu

Modely založené na principu lineárního procesu (1.1) jsou vhodné k popisu chování časových řad pozorovaných v praxi. Uvažujme například stacionární proces y_t , jehož střední hodnota je nulová. Pro tento proces lze na základě minulých hodnot $Y_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$ předpovědět hodnotu y_t . Optimální předpověď ve smyslu minimalizace střední čtvercové chyby (MSE) je podmíněná střední hodnota $E(y_t|Y_{t-1})$. Chyba této předpovědi je

$$e_t = y_t - E(y_t|Y_{t-1})$$

a má vlastnosti bílého šumu. Chybu označujeme jako *inovaci*. Pokud platí, že proces y_t pochází z normálního rozdělení, je podmíněnou střední hodnotou $E(y_t|Y_{t-1})$ lineární kombinace hodnot y_{t-1}, y_{t-2}, \dots . Chybu e_t lze pak napsat ve tvaru

$$e_t = y_t - \pi_1 y_{t-1} - \pi_2 y_{t-2} - \dots$$

1.4 Autokovarianční, autokorelační a parciální autokorelační funkce

Autokovarianční a autokorelační funkce jsou běžné nástroje využívané pro kvantitativní popis silné korelovanosti, která zpravidla nastává u časových řad. Budeme předpokládat slabě stacionární časovou řadu y_t .

Autokovarianční funkce pro zpoždění k je definována jako

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

Autokorelační funkce (ACF) pro zpoždění k je definována jako

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_y^2} \quad k = \dots, -1, 0, 1, \dots$$

Autokovarianční i autokorelační funkce jsou funkce sudé, proto je stačí uvádět pouze pro $k \geq 0$. Také $\rho_0 = 1$ a $|\rho_k| \leq 1 \forall k$.

Pro odhad obou funkcí se využívá *odhad střední hodnoty*

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t.$$

Odhadem autokovarianční funkce je

$$c_k = \frac{1}{n} \sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y}), \quad k = 0, 1, \dots, n-1, \quad (1.3)$$

a odhadem autokorelační funkce je

$$r_k = \frac{c_k}{c_0}, \quad k = 0, 1, \dots, n-1.$$

Při zpracování časových řad je zpravidla požadováno $n > 50$ a $k < \frac{n}{4}$. Též je možné místo n dělit v (1.3) číslem $n - k$, čímž se zmenší vychýlení odhadu $E c_k - \gamma_k$, avšak současně se také zvětší střední čtvercová chyba odhadu definovaná jako $E(c_k - \gamma_k)^2$. Oba odhady jsou asymptoticky nestranné, jelikož při velkém n se $E(c_k)$ blíží skutečné hodnotě γ_k .

Obzvlášť autokorelační funkce je velmi důležitým ukazatelem, poněvadž je nápovědou k tomu, jaký typ modelu by bylo nejvhodnější použít k popisu pozorované časové řady. V rámci toho je důležité určit hodnotu $k = k_0$, za kterou začíná být autokorelační funkce nulová, případně zjistit, že daná hodnota neexistuje. Bod k_0 je označován jako *bod useknutí*.

V případě konkrétní časové řady je však teoretická autokorelační funkce ρ_k neznámá, a proto je zapotřebí získat představu o tom, jak blízko nule musí r_k být, abychom mohli s danou spolehlivostí prohlásit, že $\rho_k = 0$. K tomu slouží *Bartlettova aproximace*: pokud $\rho_k = 0$ pro $k > k_0$, pak při asymptotické normalitě časové řady platí

$$r_k \sim N \left(0, \sqrt{\frac{1}{n} \left(1 + 2 \sum_{j=1}^{k_0} r_j^2 \right)} \right) \quad \text{pro } k > k_0. \quad (1.4)$$

Kromě ACF je často využívána *parciální autokorelační funkce* (PACF). Funkci značíme ρ_{kk} a je definována jako parciální korelační koeficient mezi náhodnými veličinami y_t a y_{t-k} s pevnými hodnotami $y_{t-k+1}, \dots, y_{t-1}$. Platí $\rho_{00} = 1$ a $\rho_{11} = \rho_1$.

Odhad parciální autokorelační funkce ρ_{kk} značíme r_{kk} a jde o odhadnutý parametr $\hat{\varphi}$ z modelu

$$y_t = \delta + \varphi_{k1} y_{t-1} + \varphi_{k2} y_{t-2} + \dots + \varphi_{kk} y_{t-k} + \varepsilon_t.$$

V praxi je ovšem častější využití rekurentního výpočtu odhadu za využití vzorců

$$r_{11} = r_1, \quad r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} \cdot r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} \cdot r_j} \quad \text{pro } k > 1,$$

kde

$$r_{kj} = r_{k-1,j} - r_{kk} \cdot r_{k-1,k-j} \quad \text{pro } j = 1, \dots, k-1.$$

Za předpokladu $\rho_{kk} = 0$ pro $k > k_0$ platí při asymptotické normalitě tzv. *Que-nouilleova aproximace*

$$r_{kk} \sim N \left(0, \sqrt{\frac{1}{n}} \right) \quad \text{pro } k > k_0. \quad (1.5)$$

1.5 Model ARMA

Stochastický model ARMA je využíván k analýze časových řad, přičemž právě tento model umožňuje detailně rozpracovaný a softwarově dostupný přístup k časovým řadám pozorovaným v praxi.

Smíšený proces řádu p a q značíme jako $\text{ARMA}(p, q)$. Jeho tvar je

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad \text{tj. } \varphi(B)y_t = \theta(B)\varepsilon_t, \quad (1.6)$$

kde $\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ je *autoregresní operátor* a $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ je *operátor klouzavých součtů*. $\varphi_1, \dots, \varphi_p$ a $\theta_1, \dots, \theta_q$ jsou parametry modelu.

Střední hodnota stacionárního procesu $\text{ARMA}(p, q)$ je nulová. Autokorelační funkce procesu splňuje rovnici

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p} \quad \text{pro } k > q, \quad (1.7)$$

s řešením tvaru

$$\rho_k = \alpha_1 z_1^{-k} + \alpha_2 z_2^{-k} + \dots + \alpha_p z_p^{-k} \quad \text{pro } k \geq \max(0, q - p + 1), \quad (1.8)$$

kde z_1, \dots, z_p jsou navzájem různé kořeny polynomu $\varphi(z)$ ($|z_1|, \dots, |z_p| > 1$) a $\alpha_1, \dots, \alpha_p$ jsou pevné koeficienty.

Autokorelační funkce procesu $\text{ARMA}(p, q)$ je kombinace klesajících geometrických posloupností a sinusoid s různými frekvencemi a geometricky klesajícími amplitudami. Výjimkou jsou však počáteční hodnoty $\rho_0, \rho_1, \dots, \rho_{q-p}$, a to pokud $q \geq p$. Takováto lineární kombinace rovněž omezuje parciální autokorelační funkci procesu $\text{ARMA}(p, q)$. I zde jsou ale výjimkou počáteční hodnoty $\rho_{00}, \dots, \rho_{p-q, p-q}$ a to pokud $p \geq q$.

Model je možné zobecnit i pro případ nenulové střední hodnoty. $\text{ARMA}(p, q)$ se střední hodnotou μ bude mít tvar

$$y_t - \mu = \varphi_1 (y_{t-1} - \mu) + \dots + \varphi_p (y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (1.9)$$

tedy

$$y_t = \alpha + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad \text{kde } \alpha = (1 - \varphi_1 - \dots - \varphi_p)\mu \quad (1.10)$$

1.6 Autoregresní proces AR

Autoregresní proces řádu p značený jako $\text{AR}(p)$ je tvaru

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t, \quad \text{tj. } y_t - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p} = \varphi(B)y_t = \varepsilon_t. \quad (1.11)$$

Jedná se o proces $\text{ARMA}(p, 0)$, který vznikne, pokud invertovaný lineární proces (1.2) usekneme v bodě odpovídajícím zpoždění p .

Pokud každý z kořenů z_1, \dots, z_p polynomu $\varphi(z)$ leží vně jednotkového kruhu v komplexní rovině, pak je proces $\text{AR}(p)$ stacionární. V tomto případě má nulovou střední hodnotu, rozptyl

$$\sigma_y^2 = \frac{\sigma^2}{1 - \varphi_1 \rho_1 - \dots - \varphi_p \rho_p} \quad (1.12)$$

a autokorelační funkce procesu splňuje diferenční rovnici

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p} \quad \text{pro } k > 0. \quad (1.13)$$

Řešením této rovnice je

$$\rho_k = \alpha_1 z_1^{-k} + \alpha_2 z_2^{-k} + \dots + \alpha_p z_p^{-k} \quad \text{pro } k \geq 0, \quad (1.14)$$

kde z_1, \dots, z_p jsou navzájem různé kořeny polynomu $\varphi(z) (|z_1|, \dots, |z_p| > 1)$ a $\alpha_1, \dots, \alpha_p$ jsou pevné koeficienty. V případě, že z_i a z_j jsou komplexně sdružená čísla, je možné je nahradit pouze jedním členem a to ve tvaru $\alpha \cdot d^k \cdot \sin(\lambda k + \varphi)$, $0 < d < 1$. Pokud nejsou kořeny z_1, \dots, z_p navzájem různé, pak se ve vyjádření (1.14) vyskytne člen $(\beta_0 + \beta_1 k + \dots + \beta_r k^{r-1}) \cdot z_i^{-k}$. Tento člen je ovšem ve velké míře překrýván průběhem členu z_i^{-k} .

Autokorelační funkce procesu AR(p) je lineární kombinace klesajících geometrických posloupností a sinusoid s různými frekvencemi a geometricky klesajícími amplitudami. Parciální autokorelační funkce má bod useknutí $k_0 = p$.

Příklad (Proces AR(1)). Proces AR(1) tvaru

$$y_t = \varphi_1 y_{t-1} + \varepsilon_t \quad (1.15)$$

je stacionární pro $|\varphi_1| < 1$. Autokorelační funkce je

$$\rho_k = \varphi_1^k \quad \text{pro } k \geq 0.$$

První autokorelace procesu AR(1) je stejná jako jeho autoregresní parametr. Podstatnou roli hraje v modelu znaménko. Kladné φ_1 znamená setrvačnost ve znaménku sousedních hodnot dané časové řady, naopak záporné φ_1 předpovídá pro znaménko sousedních hodnot častější změny.

Parciální autokorelační funkce procesu AR(1) je tvaru

$$\rho_{11} = \varphi_1, \quad \rho_{kk} = 0 \quad \text{pro } k > 1.$$

Bod useknutí PACF je tedy $k_0 = 1$.

Příklad (Proces AR(2)). Proces

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$$

je stacionární pro $\varphi_1 + \varphi_2 < 1$, $\varphi_2 - \varphi_1 < 1$, $-1 < \varphi_2 < 1$.

Oblast, kde je proces stacionární, vyplní společně s osami pro hodnoty φ_1 a φ_2 vnitřek trojúhelníku s vrcholy $(-2, -1)$, $(0, 1)$ a $(2, 1)$. Autokorelační funkce je

$$\rho_k = \frac{z_1^{-1}(1 - z_2^{-2})z_1^{-k} - z_2^{-1}(1 - z_1^{-2})z_2^{-k}}{(z_1^{-1} - z_2^{-1})(1 + z_1^{-1}z_2^{-1})} \quad \text{pro } k \geq 0,$$

kde z_1 a z_2 jsou navzájem různé kořeny polynomu $\varphi(z)$. ρ_k nemá bod useknutí a je tvaru lineární kombinace dvou geometricky klesajících posloupností, nebo sinusoidy s geometricky klesající amplitudou.

Parciální autokorelační funkce procesu AR(2) má bod useknutí $k_0 = 2$.

1.7 Proces klouzavých součtů MA

Proces klouzavých součtů řádu q značíme $MA(q)$ a je tvaru

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} = \theta(B)\varepsilon_t.$$

Jedná se o proces $ARMA(0, q)$, který vzniká useknutím lineárního procesu (1.1) v bodě odpovídajícím zpoždění q .

Proces $MA(q)$ je vždy stacionární, střední hodnota je nulová, rozptyl je

$$\sigma_y^2 = (1 + \theta_1^2 + \dots + \theta_q^2)\sigma^2$$

a autokorelační funkce

$$\rho_k = \begin{cases} \frac{\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & \text{pro } k = 1, \dots, q \\ 0 & \text{pro } k > q. \end{cases}$$

Bod useknutí k_0 pro autokorelační funkci je tedy roven řádu modelu q . U parciální autokorelační funkce ρ_{kk} procesu $MA(q)$ bod useknutí není. Tuto funkci omezuje lineární kombinace geometricky klesajících posloupností a sinusoid s geometricky klesajícími amplitudami.

Pokud jsou všechny kořeny z_1, \dots, z_q polynomu $\theta(z)$ vně jednotkového kruhu v komplexní rovině, pak je proces $MA(q)$ invertibilní.

Příklad (Proces $MA(1)$). Bod useknutí ACF procesu $MA(1)$ je $k_0 = 1$. Parciální autokorelační funkce tohoto procesu je

$$\rho_{kk} = \frac{(-1)^{k-1} \theta_1^k (1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}} \quad \text{pro } k = 1, 2, \dots$$

Příklad (Proces $MA(2)$). Proces $MA(2)$ je tvaru

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2},$$

s autokorelační funkcí

$$\rho_k = \begin{cases} \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2} & \text{pro } k = 1 \\ \frac{\theta_2}{1+\theta_1^2+\theta_2^2} & \text{pro } k = 2 \\ 0 & \text{pro } k > 2, \end{cases}$$

s bodem useknutí $k_0 = 2$.

1.8 Identifikace modelu

K identifikaci modelu pomohou obecné poznatky o tvaru autokorelační a parciální autokorelační funkce procesů $AR(p)$, $MA(q)$ a $ARMA(p, q)$:

	$AR(p)$	$MA(q)$	$ARMA(p, q)$
ρ_k	neexistuje k_0 ; ρ_k ve tvaru křivky U	$k_0 = q$	neexistuje k_0 ; ρ_k ve tvaru křivky U po prvních $q - p$ hodnotách
ρ_{kk}	$k_0 = p$	neexistuje k_0 ; ρ_{kk} ve tvaru křivky U	neexistuje k_0 ; ρ_{kk} ve tvaru křivky U po prvních $p - q$ hodnotách

Tabulka 1.1: Vlastnosti autokorelační a parciální autokorelační funkce modelů $AR(p)$, $MA(q)$ a $ARMA(p, q)$. U značí křivku tvaru lin. kombinace klesajících geom. posloupností a sinusoid s geom. klesajícími amplitudami.

Identifikace spočívá v přiřazení nejvhodnějšího typu modelu pozorované časové řady pomocí charakterizací obsažených v tabulce 1.1 konfrontovaných s odhady ACF a PACF. Odpovídající typ modelu vybíráme na základě prohlídky grafů odhadnuté ACF a PACF zkoumané časové řady. Pokud není jasné, jaký typ modelu je vhodné zvolit, testuje se potenciální bod useknutí k_0 , a to za pomoci Bartlettovy aproximace (1.4). Přibližný (asymptotický) kritický obor na hladině významnosti 5% je

$$|r_k| \geq 2 \sqrt{\frac{1}{n} \left(1 + 2 \sum_{j=1}^{k_0} r_j^2 \right)} \quad \text{pro } k > k_0. \quad (1.16)$$

Využít lze též Quenouilleovu aproximaci (1.5) s kritickým oborem na hladině významnosti 5%

$$|r_{kk}| \geq 2 \sqrt{\frac{1}{n}} \quad \text{pro } k > k_0. \quad (1.17)$$

Pro ověření adekvátnosti identifikovaného modelu je vhodné provést diagnostiku vypočteného bílého šumu. Ta spočívá především v ověření nekorelovanosti s využitím (1.4) a normality pomocí histogramu, případně některého z testů implementovaných v použitém softwaru.

Poznámka (Odhad modelu). Konstrukce odhadů parametrů ve zmíněných modelech jsou záležitostí softwaru. Implementované procedury vycházejí z metody nejmenších čtverců nebo metody maximální věrohodnosti za předpokladu normality bílého šumu.

Kapitola 2

Dickeyův-Fullerův test na jednotkový kořen

Ekonomické a finanční časové řady v praxi velmi často nejsou stacionární, což je ovšem základní podmínkou pro jejich modelování pomocí ARMA procesů. Konkrétně AR(1) je stacionární pro $|\varphi_1| < 1$, viz kapitola 1.6. Pokud ale $\varphi_1 = 1$, model AR(1) přejde na model náhodné procházky $y_t = y_{t-1} + \varepsilon_t$. Přitom $z = \frac{1}{\varphi}$ je kořenem autoregresního polynomu $\varphi(z)$. Proto se testování hypotézy náhodné procházky proti alternativám obsahujícím proces AR(1) nazývá testy na jednotkový kořen. Metodologie pro testování hypotézy náhodné procházky proti alternativám souvisejícím s AR(1) byla navržena v článku Dickey a Fuller (1981).

V rámci Dickeyova-Fullerova test budeme uvažovat dvě varianty, značené jako τ -testy.

Označení testu	Hypotéza	Alternativa	Testová statistika
τ_μ -test	$y_t = y_{t-1} + \varepsilon_t$	$y_t = \alpha + \varphi y_{t-1} + \varepsilon_t$	Φ_1
τ_τ -test	$y_t = y_{t-1} + \varepsilon_t$	$y_t = \alpha + \beta t + \varphi y_{t-1} + \varepsilon_t$	Φ_2

Tabulka 2.1: Testy na jednotkový kořen.

2.1 τ_μ -test

Testujeme $H_0 : y_t = y_{t-1} + \varepsilon_t$ proti

$$H_{1,\mu} : y_t = \alpha + \varphi y_{t-1} + \varepsilon_t,$$

kde $t = 1, 2, 3, \dots, n$, y_0 je pevně dané a $\varepsilon_t \sim N(0, \sigma^2)$ je bílý šum.

K odhadu parametrů α a φ využijeme regresní přístup (viz Anděl, 2005). Nechť \mathbb{X} je matice $n \times 2$,

$$\mathbb{X} = \begin{pmatrix} 1 & y_0 \\ 1 & y_1 \\ \vdots & \vdots \\ 1 & y_{n-1} \end{pmatrix}$$

a $\mathbf{y}' = (y_1, y_2, \dots, y_n)$.

Odhad vektoru parametrů $\boldsymbol{\theta}_\mu = (\alpha, \varphi)'$ metodou nejmenších čtverců je

$$\hat{\boldsymbol{\theta}}_\mu = \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbf{y}. \quad (2.1)$$

Jeho složky lze psát ve tvaru

$$\hat{\varphi}_\mu = \left[\sum_{t=1}^n (y_{t-1} - \bar{y}_{(-1)})^2 \right]^{-1} \sum_{t=1}^n (y_t - \bar{y}_{(0)})(y_{t-1} - \bar{y}_{(-1)})$$

$$\hat{\alpha}_\mu = \bar{y}_{(0)} - \hat{\varphi}_\mu \bar{y}_{(-1)},$$

kde $\bar{y}_{(i)} = n^{-1} \sum_{t=1}^n y_{t+i}$ pro $i = -1, 0$.

Označíme

$$S_{\varepsilon\mu}^2 = (n-2)^{-1} \sum_{t=1}^n (y_t - \hat{\alpha}_\mu - \hat{\varphi}_\mu y_{t-1})^2.$$

Odhad $\hat{\boldsymbol{\theta}}_\mu$ pomocí metody nejmenších čtverců je totožný s odhadem metodou maximální věrohodnosti. Věrohodnostní funkce za platnosti $H_{1,\mu}$ je

$$L = \left(\frac{1}{\sqrt{2\pi}\sigma_1} \right)^n \prod_{t=1}^n e^{-\frac{\varepsilon_t^2}{2\sigma_1^2}},$$

kde $\varepsilon_t = y_t - \alpha - \varphi y_{t-1}$. Logaritmická věrohodnostní funkce je

$$l = \log L = -\frac{1}{2}n \log(2\pi) - n \log \sigma_1 - (2\sigma_1^2)^{-1} \sum_{t=1}^n (y_t - \alpha - \varphi y_{t-1})^2.$$

Pro logaritmickou věrohodnostní funkci platí

$$(2\sigma_1^2)^{-1} \sum_{t=1}^n (y_t - \alpha - \varphi y_{t-1})^2 > 0,$$

a proto maximalizace l vzhledem k parametrům α, φ je ekvivalentní minimalizaci $\sum_{t=1}^n \varepsilon_t^2$.

Pro odhad parametru σ_1^2 položíme

$$\frac{\partial \log L}{\partial \sigma_1^2} = 0$$

$$-\frac{n}{2\hat{\sigma}_1^2} + \frac{1}{2\hat{\sigma}_1^4} \sum_{t=1}^n (y_t - \alpha - \varphi y_{t-1})^2 = 0$$

$$\hat{\sigma}_1^2 = \frac{1}{n} \sum_{t=1}^n (y_t - \alpha - \varphi y_{t-1})^2$$

Vzhledem k tomu, že

$$\hat{\sigma}_1^2 = \frac{n-2}{n} \frac{1}{n-2} \sum_{t=1}^n (y_t - \alpha - \varphi y_{t-1})^2, \quad (2.2)$$

dostáváme po dosazení odhadů $\hat{\alpha}_\mu, \hat{\varphi}_\mu$ do (2.2)

$$\hat{\sigma}_1^2 = (n-2)n^{-1}S_{\varepsilon\mu}^2.$$

Za platnosti $H_{1,\mu}$ tedy nastává maximum funkce l v $(\hat{\sigma}_1^2, \hat{\theta}'_\mu)$, kde $\hat{\theta}_\mu$ již bylo definováno v (2.1).

Za platnosti H_0 je $(\alpha, \varphi) = (0, 1)$. Pak $\hat{\sigma}_0^2 = n^{-1} \sum_{t=1}^n (y_t - y_{t-1})^2$.

Díky získaným odhadům za platnosti nulové a alternativní hypotézy lze zkonstruovat věrohodnostní poměr

$$\begin{aligned} \frac{L(\hat{\sigma}_1^2)}{L(\hat{\sigma}_0^2)} &= \frac{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_1}\right)^n \exp\left\{-\frac{1}{2\hat{\sigma}_1^2} \sum_{t=1}^n (y_t - \hat{\alpha}_\mu - \hat{\varphi}_\mu y_{t-1})^2\right\}}{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_0}\right)^n \exp\left\{-\frac{1}{2\hat{\sigma}_0^2} \sum_{t=1}^n (y_t - y_{t-1})^2\right\}} = \\ &= \frac{\hat{\sigma}_0^n}{\hat{\sigma}_1^n} \exp\left\{\frac{-\sum_{t=1}^n (y_t - \hat{\alpha}_\mu - \hat{\varphi}_\mu y_{t-1})^2}{2\frac{1}{n} \sum_{t=1}^n (y_t - \hat{\alpha}_\mu - \hat{\varphi}_\mu y_{t-1})^2} - \frac{-\sum_{t=1}^n (y_t - y_{t-1})^2}{2\frac{1}{n} \sum_{t=1}^n (y_t - y_{t-1})^2}\right\} = \\ &= \frac{\hat{\sigma}_0^n}{\hat{\sigma}_1^n} \exp\left\{-\frac{n}{2} + \frac{n}{2}\right\} = \left(\frac{\hat{\sigma}_0}{\hat{\sigma}_1}\right)^n. \end{aligned}$$

Dále lze psát

$$\begin{aligned} \frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2} &= \frac{\hat{\sigma}_0^2}{\frac{n-2}{n} S_{\varepsilon\mu}^2} = \frac{n}{n-2} \left(\frac{\hat{\sigma}_0^2}{S_{\varepsilon\mu}^2} - 1 + 1\right) = \\ &= \frac{n}{n-2} \left(\frac{\hat{\sigma}_0^2}{S_{\varepsilon\mu}^2} - 1\right) + \frac{n}{n-2} = \frac{n}{n-2} \frac{\hat{\sigma}_0^2}{S_{\varepsilon\mu}^2} - \frac{n}{n-2} + \frac{2}{n-2} + 1 = \\ &= \left(\frac{n}{2} \frac{\hat{\sigma}_0^2}{S_{\varepsilon\mu}^2} - \frac{n}{2}\right) \frac{2}{n-2} + \frac{2}{n-2} + 1 = \left(\frac{n}{2} \frac{\hat{\sigma}_0^2}{S_{\varepsilon\mu}^2} - \frac{n}{2} + 1\right) \frac{2}{n-2} + 1 = \\ &= \frac{1}{2S_{\varepsilon\mu}^2} [n\hat{\sigma}_0^2 - (n-2)S_{\varepsilon\mu}^2] \frac{2}{n-2} + 1. \end{aligned}$$

Po dosazení do věrohodnostního poměru dostaneme

$$\begin{aligned} \frac{L(\hat{\sigma}_1^2)}{L(\hat{\sigma}_0^2)} &= \frac{\hat{\sigma}_0^n}{\hat{\sigma}_1^n} = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2}\right)^{\frac{n}{2}} = \left[1 + \frac{1}{2S_{\varepsilon\mu}^2} [n\hat{\sigma}_0^2 - (n-2)S_{\varepsilon\mu}^2] \frac{2}{n-2}\right]^{\frac{n}{2}} = \\ &= [1 + 2(n-2)^{-1}\Phi_1]^{\frac{n}{2}}, \end{aligned}$$

kde $\Phi_1 = (2S_{\varepsilon\mu}^2)^{-1} [n\hat{\sigma}_0^2 - (n-2)S_{\varepsilon\mu}^2]$.

2.2 τ_τ -test

Stejným způsobem lze postupovat i v případě testu $H_0 : y_t = y_{t-1} + \varepsilon_t$ proti

$$H_{1,\tau} : y_t = \alpha + \beta t + \varphi y_{t-1} + \varepsilon_t,$$

kde $t = 1, 2, 3, \dots, n$, y_0 je pevně dané a $\varepsilon_t \sim N(0, \sigma^2)$. Nechť \mathbb{X} je matice $n \times 3$,

$$\mathbb{X} = \begin{pmatrix} 1 & 1 & y_0 \\ 1 & 2 & y_1 \\ \vdots & \vdots & \vdots \\ 1 & n & y_{n-1} \end{pmatrix}$$

a necht' $\mathbf{y}' = (y_1, y_2, \dots, y_n)$. Pak pro odhad parametrů opětovně využijeme regresi. Odhad vektoru parametrů $\boldsymbol{\theta}_\tau = (\alpha, \beta, \varphi)'$ metodou nejmenších čtverců je

$$\hat{\boldsymbol{\theta}}_\tau = (\hat{\alpha}_\tau, \hat{\beta}_\tau, \hat{\varphi}_\tau)' = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbf{y}. \quad (2.3)$$

Označíme

$$S_{\varepsilon_\tau}^2 = (n-3)^{-1} \sum_{t=1}^n (y_t - \hat{\alpha}_\tau - \hat{\beta}_\tau t - \hat{\varphi}_\tau y_{t-1})^2 = (n-3)^{-1} \mathbf{y}' [\mathbb{I} - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'] \mathbf{y}.$$

Odhad $\hat{\boldsymbol{\theta}}_\tau$ pomocí metody nejmenších čtverců je i v tomto případě stejný jako odhad metodou maximální věrohodnosti. Věrohodnostní funkce za platnosti $H_{1,\tau}$ je

$$L = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \prod_{t=1}^n e^{-\frac{\varepsilon_t^2}{2\sigma^2}},$$

kde $\varepsilon_t = y_t - \alpha - \beta t - \varphi y_{t-1}$. Logaritmická věrohodnostní funkce je pak

$$l = \log L = -\frac{1}{2}n \log(2\pi) - n \log \sigma - (2\sigma^2)^{-1} \sum_{t=1}^n (y_t - \alpha - \beta t - \varphi y_{t-1})^2.$$

I zde pro logaritmickou věrohodnostní funkci platí

$$(2\sigma^2)^{-1} \sum_{t=1}^n (y_t - \alpha - \beta t - \varphi y_{t-1})^2 > 0,$$

a proto je maximalizace l vzhledem k α, β a φ ekvivalentní minimalizaci $\sum_{t=1}^n \varepsilon_t^2$.

Za platnosti $H_{1,\tau}$ nastává maximum v $(\hat{\sigma}_1^2, \hat{\boldsymbol{\theta}}_\tau')$, kde $\hat{\boldsymbol{\theta}}_\tau$ již bylo definováno v (2.3) a $\hat{\sigma}_1^2 = (n-3)n^{-1}S_{\varepsilon_\tau}^2$.

Za platnosti H_0 je $(\alpha, \beta, \varphi) = (0, 0, 1)$. Pak $\hat{\sigma}_0^2 = n^{-1} \sum_{t=1}^n (y_t - y_{t-1})^2$.

Z odhadů za platnosti nulové a alternativní hypotézy analogickým postupem jako v případě τ_μ -testu získáme věrohodnostní poměr

$$\left[\frac{\hat{\sigma}_1}{\hat{\sigma}_0} \right]^n = [1 + 3(n-3)^{-1}\Phi_2]^{\frac{n}{2}},$$

kde

$$\Phi_2 = (3S_{\varepsilon_\tau}^2)^{-1} [n\hat{\sigma}_0^2 - (n-3)S_{\varepsilon_\tau}^2].$$

2.3 Interpretace výsledků

Tabulky 2.2 a 2.3 obsahují potřebné $(1 - \alpha)$ -kvantily pro zvolené hladiny významnosti α rovné postupně 10%, 5%, 2.5% a 1%. Alternativa i nulová hypotéza jednotlivých modelů jsou uvedeny u příslušných tabulek. Nulovou hypotézu v obou případech zamítáme pro velké hodnoty Φ_1 , respektive Φ_2 . Odpovídá to metodologii testu založeném na věrohodnostním poměru, která je popsána v knize Anděl (2005)

Kvantily rozdělení statistik Φ_1, Φ_2 byly převzaty z článku Dickey a Fuller (1981). Hodnoty pro výběry konečného rozsahu n byly získány z napočítaných statistik pro výběry generované modelem s $y_1 = 0$ a $y_t = y_{t-1} + \varepsilon_t$, $t = 2, 3, \dots, n$, pro $n = 25, 50, 100, 250$ a 500 . Pro $n = 25$ byla generována tři opakování 50000 výběrů, pro $n = 50, 100$ a 250 dvě opakování a jedno opakování pro $n = 500$. Simulace pro limitní případ $n = \infty$ byla zkonstruována na základě procesu popsaného v práci Dickey (1976).

n	Pravděpodobnost $(1-\alpha)$			
	0.90	0.95	0.975	0.99
25	4.12	5.18	6.30	7.88
50	3.94	4.86	5.80	7.06
100	3.86	4.71	5.57	6.70
250	3.81	4.63	5.45	6.52
500	3.79	4.61	5.41	6.47
∞	3.78	4.59	5.38	6.43

Tabulka 2.2: Kvantily rozdělení Φ_1 za platnosti $(\alpha, \varphi) = (0, 1)$ v případě modelu $y_t = \alpha + \varphi y_{t-1} + \varepsilon_t$.

n	Pravděpodobnost $(1-\alpha)$			
	0.90	0.95	0.975	0.99
25	4.67	5.68	6.75	8.21
50	4.31	5.13	5.94	7.02
100	4.16	4.88	5.59	6.50
250	4.07	4.75	5.40	6.22
500	4.05	4.71	5.35	6.15
∞	4.03	4.68	5.31	6.09

Tabulka 2.3: Kvantily rozdělení Φ_2 za platnosti $(\alpha, \beta, \varphi) = (0, 0, 1)$ v případě modelu $y_t = \alpha + \beta t + \varphi y_{t-1} + \varepsilon_t$.

Kapitola 3

Praktická aplikace na simulovaných a reálných datech

3.1 Simulovaná data

Časové řady, na kterých budeme demonstrovat aplikaci testů na jednotkový kořen, budeme postupně generovat podle nulové hypotézy a obou alternativ. Takto generované řady následně podrobíme testům za účelem ověření jejich funkčnosti a spolehlivosti.

3.1.1 Generování časové řady podle nulové hypotézy

Řadu generujeme podle modelu odpovídajícího nulové hypotéze, tedy $y_t = y_{t-1} + \varepsilon_t$, kde $t = 1, 2, \dots, n$ pro $n = 250$. Počáteční hodnotu y_0 stanovíme zprůměrováním předvýběru deseti členů generovaných podle nulové hypotézy, kde $y_0 = 0$. Hodnoty bílého šumu generujeme z $N(0,1)$.

Uvedeme výpis výpočtů z programu Mathematica.

```
> n = 250;
> epsilon = RandomVariate[NormalDistribution[0, 1], n];
> yt0[0] := 0;
> yt0[t_] := yt0[t-1] + epsilon[[t]];
> y0 = 1/10*Sum[yt0[i], {i, 1, 10}]
> yt[0] := y0;
> yt[t_] := yt[t-1] + epsilon[[t]];
> radayt = Table[yt[i], {i, 1, n}];
```

Nejprve provedeme τ_μ -test, kdy hypotézu testujeme proti $H_1 : \alpha + \varphi y_{t-1} + \varepsilon_t$.

```
> PrumerY0 = n^-1 Sum[yt[i], {i, 1, n}];
> PrumerYMinus1 = n^-1 Sum[yt[i-1], {i, 1, n}]
> OdhadPhi = (Sum[(yt[i-1] - PrumerYMinus1)^2, {i, 1, n}])^-1*
  Sum[(yt[i] - PrumerY0)*(yt[i-1] - PrumerYMinus1), {i, 1, n}];
> OdhadAlpha = PrumerY0 - OdhadPhi*PrumerYMinus1;
> Smi = (n-2)^-1*
  Sum[(yt[i] - OdhadAlpha - OdhadPhi*yt[i-1])^2, {i, 1, n}];
> Odhad1Sigma1 = (n-2)*n^-1*Smi;
> Odhad1Sigma0 = n^-1*Sum[(yt[i] - yt[i-1])^2, {i, 1, n}];
```

```
> F1 = (2*Smi)^-1*(n*Odhad1Sigma0 - (n-2)*Smi)
[1] 2.8593
```

Pro $n = 250$ na hladině významnosti 5% je odpovídající kvantil 4.63 (viz tabulka 2.2). Vzhledem k $\Phi_1 = 2.8593$ tedy nulovou hypotézu nezamítáme.

Následně generovanou časovou řadu podrobíme τ_τ -testu, kdy hypotézu testujeme proti $H_1 : \alpha + \beta t + \varphi y_{t-1} + \varepsilon_t$.

```
> X = Table[{1, i, yt[i-1]}, {i, 1, n}];
> Xtran = Transpose[X];
> Theta = Inverse[Xtran.X].Xtran.radayt ;
> Stau = (n-3)^-1*radayt.(IdentityMatrix[n]
      - X.Inverse[Xtran.X].Xtran).radayt;
> Odhad2Sigma1 = (n-3) n^-1 Stau;
> Odhad2Sigma0 = n^-1*Sum[(yt[i] - yt[i-1])^2 , {i, 1, n}];
> F2 = ((3 Stau)^-1)*(n*Odhad2Sigma0 - (n-3)*Stau)
[1] 1.90892
```

Pro $n = 250$ na hladině významnosti $\alpha = 5\%$ je odpovídající kvantil 4.75 (viz tabulka 2.3). Vzhledem k $\Phi_2 = 1.90892$ tedy nulovou hypotézu nezamítáme.

V obou aplikovaných testech nulová hypotéza $y_t = y_{t-1} + \varepsilon_t$ nebyla zamítnuta a výsledky tedy souhlasí s faktem, že jsme danou časovou řadu generovali právě podle této hypotézy.

3.1.2 Generování časové řady podle 1. alternativy

Řadu generujeme podle modelu odpovídajícího alternativě $y_t = \alpha + \varphi y_{t-1} + \varepsilon_t$, kde $t = 1, 2, \dots, n$ pro $n = 250$, $\alpha = 0.3$ a $\varphi = 0.5$. Počáteční hodnotu y_0 i hodnoty bílého šumu ε_t generujeme analogickým způsobem jako v předchozím případě.

Odhady parametrů alternativy jsou

$$\hat{\alpha}_\mu = 0.292237 \quad \text{a} \quad \hat{\varphi}_\mu = 0.454199.$$

Odlišnost od skutečných α a φ souvisí s relativně malým rozsahem výběru n . Testová statistika τ_μ -testu je

$$\Phi_1 = 46.4102.$$

Na hladině významnosti $\alpha = 5\%$ je odpovídající kvantil 4.63, nulovou hypotézu zamítáme. Jde přitom o očekávaný závěr, jelikož časová řada byla generována podle alternativy.

V případě τ_τ -testu jsou odhady parametrů

$$\hat{\alpha}_\tau = 0.155831, \quad \hat{\beta}_\tau = 0.0011189 \quad \text{a} \quad \hat{\varphi}_\tau = 0.446636.$$

Projevilo se, že generovaná řada neobsahuje lineární trend βt .

Testová statistika je

$$\Phi_2 = 31.4777$$

a jelikož je na hladině významnosti $\alpha = 5\%$ odpovídající kvantil 4.75, nulovou hypotézu zamítáme.

3.1.3 Generování časové řady podle 2. alternativy

Řadu generujeme podle modelu odpovídajícího alternativě $y_t = \alpha + \beta t + \varphi y_{t-1} + \varepsilon_t$, kde $t = 1, 2, \dots, n$ pro $n = 250$, $\alpha = 0.3$, $\beta = 0.7$ a $\varphi = 0.5$. Počáteční hodnotu y_0 i hodnoty bílého šumu ε_t generujeme analogickým způsobem jako v předchozích případech.

Odhady volených parametrů v rámci τ_μ -testu jsou

$$\hat{\alpha}_\mu = 1.28114 \quad \text{a} \quad \hat{\varphi}_\mu = 1.00052.$$

Odchyvky od skutečné hodnoty parametrů jsou podstatně větší než v předchozích případech, a to především z důvodu volby n a přítomnosti lineárního trendu v generované řadě. Odchyvky však nemají na konečný výsledek testu zásadní vliv.

Testová statistika je

$$\Phi_1 = 129.703.$$

Jelikož je na hladině významnosti $\alpha = 5\%$ odpovídající kvantil 4.63, nulovou hypotézu zamítáme.

V případě τ_τ -testu jsou odhady parametrů

$$\hat{\alpha}_\tau = -0.212744, \quad \hat{\beta}_\tau = 0.880806 \quad \text{a} \quad \hat{\varphi}_\tau = 0.371888.$$

Přesnějších odhadů bychom opět docílili volbou většího n .

Testová statistika je

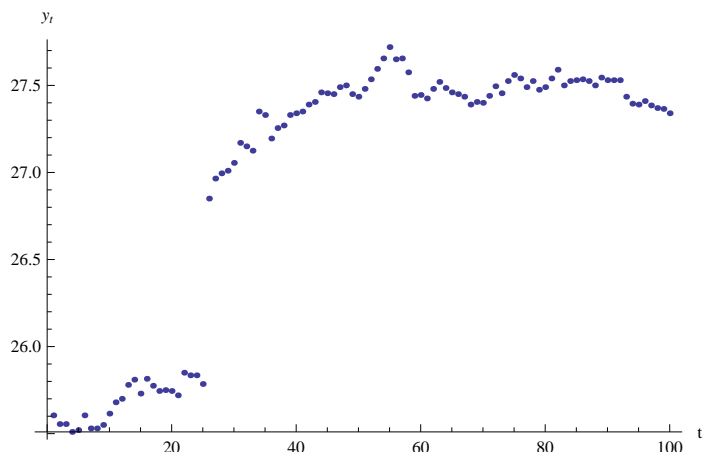
$$\Phi_2 = 181.8.$$

Na hladině významnosti $\alpha = 5\%$ je odpovídající kvantil 4.75, nulovou hypotézu tedy zamítáme. Jde o očekávaný závěr, jelikož časová řada byla generována podle dané alternativy.

Jak je z výše popsaných výsledků patrné, oba testy na jednotkový kořen lze považovat za funkční. Ve všech případech byly výsledky očekávané a nulová hypotéza se nezamítla pouze v případě, kdy jsme časovou řadu podle této hypotézy generovali.

3.2 Reálná data

Pro identifikaci modelů ARMA a aplikaci testů na jednotkový kořen využijeme data České národní banky, a to denní vývoj směnného kurzu koruny k euru v pracovních dnech od 1.10.2013 do 25.2.2014, to jest řadu délky $n = 100$.

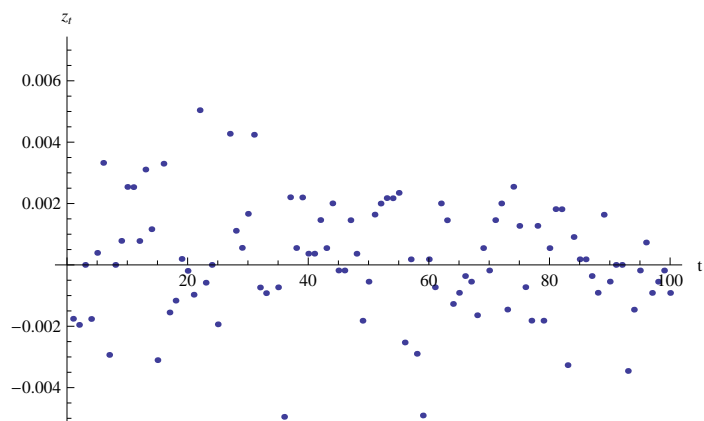


Obrázek 3.1: Vývoj kurzu Eura.

3.2.1 Logaritmická transformace dat

Již z grafu 3.1 lze vidět, že zkoumaná časová řada není stacionární. Proto provedeme obvyklou transformaci pro finanční časové řady

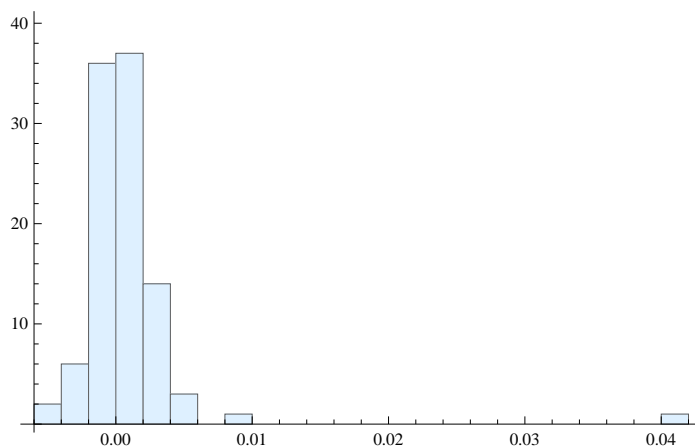
$$z_t = \log y_t - \log y_{t-1} \quad t = 1, \dots, n.$$



Obrázek 3.2: Transformovaná data z_t .

V případě identifikace modelu ověříme vzhledem k použití aproximací (1.4) a (1.5) normalitu dat testem normality a konstrukcí histogramu.

Z histogramu 3.3 lze pozorovat, že transformovaná data nepocházejí z normálního rozdělení. To potvrzuje i test provedený v softwaru Mathematica. Jeho výstupem je p-hodnota, tedy nejmenší hladina, na které zamítáme normalitu dat.

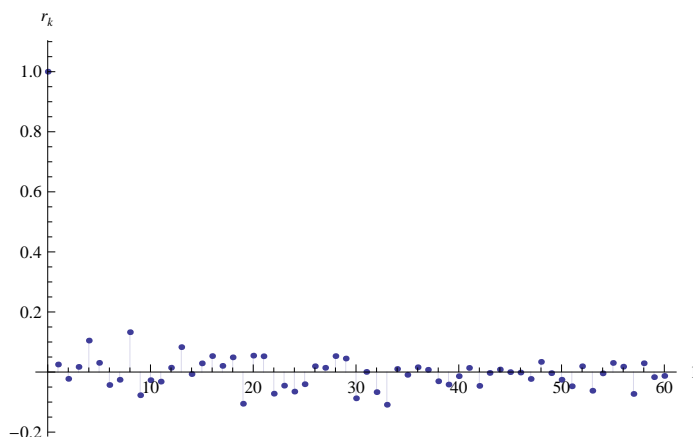


Obrázek 3.3: Histogram časové řady z_t .

```
> DistributionFitTest[LogData, NormalDistribution[]]
[1] 1.77636*10^-15
```

Získaná p-hodnota je rovna 1.77636×10^{-15} , vzhledem k čemuž zamítáme nulovou hypotézu, že zkoumaná data pocházejí z normálního rozdělení.

Předpoklad normality tedy není splněn, a to zřejmě z důvodu několika málo extrémně velkých hodnot a značné špičatosti histogramu. Navzdory tomu se o identifikaci modelu pokusíme. Nejprve využijeme grafy funkce ACF a PACF.



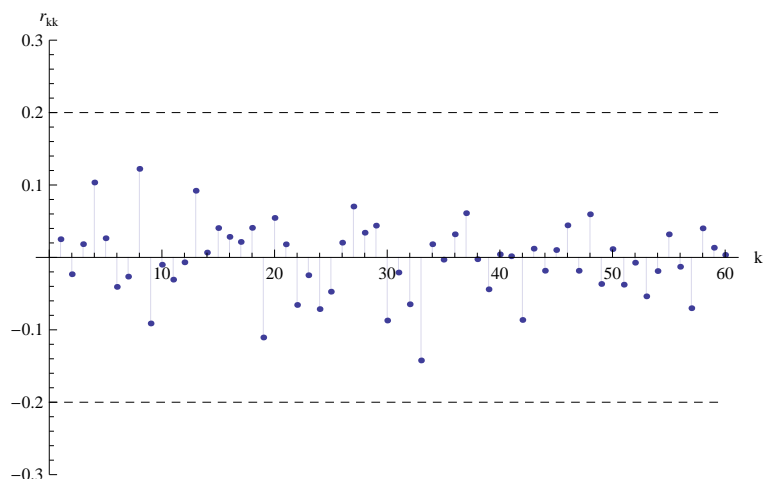
Obrázek 3.4: Graf ACF pro z_t .

Z korelogramu 3.4 lze vidět, že nelze identifikovat bod useknutí. Stejně je tomu i v případě parciálního korelogramu 3.5. Oba grafy poukazují k možnosti, že by se mohlo jednat o bílý šum. Pro ověření, zda je tomu tak, využijeme testy založené na Bartlettově (1.16) a Quenouilleově (1.17) aproximaci. Kritický obor (1.16) má při nekorelovanosti veličin v časové řadě dolní mez $2\sqrt{\frac{1}{n}}$. Platí

$$|r_k| < 2\sqrt{\frac{1}{n}} = 0.2 \quad \text{pro } k \geq 1$$

a

$$|r_{kk}| < 2\sqrt{\frac{1}{n}} = 0.2 \quad \text{pro } k \geq 1,$$



Obrázek 3.5: Graf PACF pro z_t .

jak je vidět i v grafu 3.5. ACF i PACF lze tedy považovat za nulové a pro z_t lze identifikovat bílý šum. Provedeme testy na jednotkový kořen na řady z_t a $\log y_t$.

Nejprve aplikujeme testy na časovou řadu z_t . V případě τ_μ testu je testová statistika

$$\Phi_1 = 141.217.$$

Pro zvolenou hladinu $\alpha = 5\%$ je odpovídající kvantil 4.71, proto hypotézu $H_0 : z_t = z_{t-1} + \varepsilon_t$ zamítáme ve prospěch alternativy $H_1 : z_t = \alpha + \varphi z_{t-1} + \varepsilon_t$. Odhady parametrů jsou

$$\hat{\alpha}_\mu = 0.0334586 \text{ a } \hat{\varphi}_\mu = 0.000637767.$$

Podobně postupujeme i v případě τ_τ -testu. Testová statistika je

$$\Phi_2 = 95.8675.$$

Na hladině $\alpha = 5\%$ je odpovídající kvantil roven 4.88, hypotézu $H_0 : z_t = z_{t-1} + \varepsilon_t$ tedy zamítáme ve prospěch alternativy $H_1 : z_t = \alpha + \beta t + \varphi z_{t-1} + \varepsilon_t$. Odhady parametrů jsou

$$\hat{\alpha}_\tau = 0.0017707, \hat{\beta}_\tau = -0.0000224352 \text{ a } \hat{\varphi}_\tau = 0.0387998.$$

Potvrdilo se, že řadu z_t lze považovat za stacionární.

Analogickým způsobem testujeme i časovou řadu $l_t = \log y_t$. V τ_μ -testu je testová statistika

$$\Phi_1 = 2.39496.$$

Pro zvolenou hladinu $\alpha = 5\%$ je odpovídající kvantil 4.71, proto hypotézu $H_0 : l_t = l_{t-1} + \varepsilon_t$ nezamítáme. Testová statistika v τ_τ -testu je

$$\Phi_2 = 1.58347.$$

Na hladině $\alpha = 5\%$ je odpovídající kvantil roven 4.88, nulovou hypotézu tedy nezamítáme.

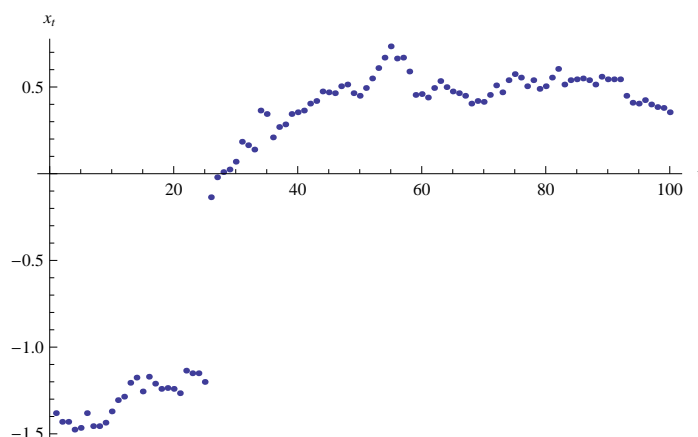
Testy na jednotkový kořen prokázaly, že $\log y_t$ lze identifikovat jako náhodnou procházku. Tento fakt odpovídá předchozímu zjištění, že časová řada z_t je bílý šum.

3.2.2 Netransformovaná data

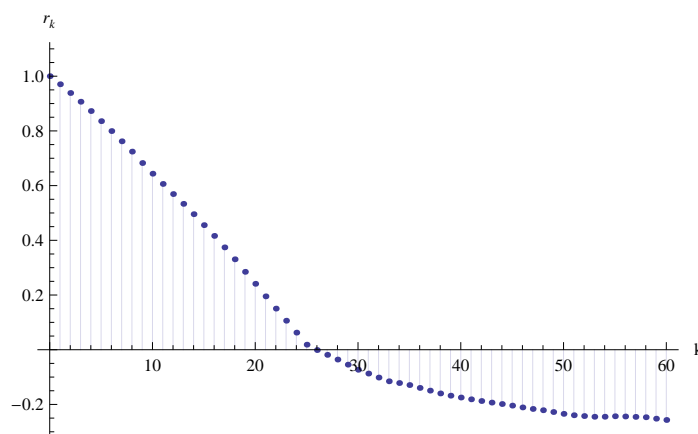
Vzhledem k tomu, že pro transformovaná data se nepodařilo identifikovat ARMA model, budeme se zabývat původní časovou řadou s odečteným průměrem

$$x_t = y_t - \frac{1}{n} \sum_{t=1}^n y_t.$$

Takto upravenou řadu znázorňuje obrázek 3.6. Je patrné, že prvních 25 pozorování se zásadně liší od dalších hodnot a lze nalézt i další výkyvy.



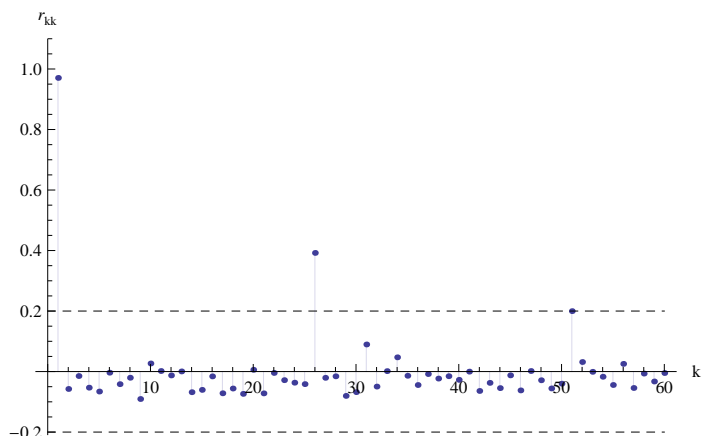
Obrázek 3.6: Graf časové řady po odečtení aritmetického průměru.



Obrázek 3.7: Graf ACF pro data x_t .

Graf funkce ACF 3.7 není příliš vypovídající a těžko z něj lze odhadnout, jakému modelu by data mohla odpovídat. Lineární a relativně pomalý pokles pro $k = 1, \dots, 27$ poukazuje na nestacionaritu dat. Napovědět může i graf PACF na obrázku 3.8. Zde je výrazné maximum v bodě $k = 1$. Pro ověření, že jde o model AR(1), lze využít test založený na Quenouilleově aproximaci (1.17). Aby bylo možné označit model za AR(1), je nutné splnění nerovnosti

$$|r_{kk}| < 2\sqrt{\frac{1}{n}} \quad \text{pro } k > 1.$$



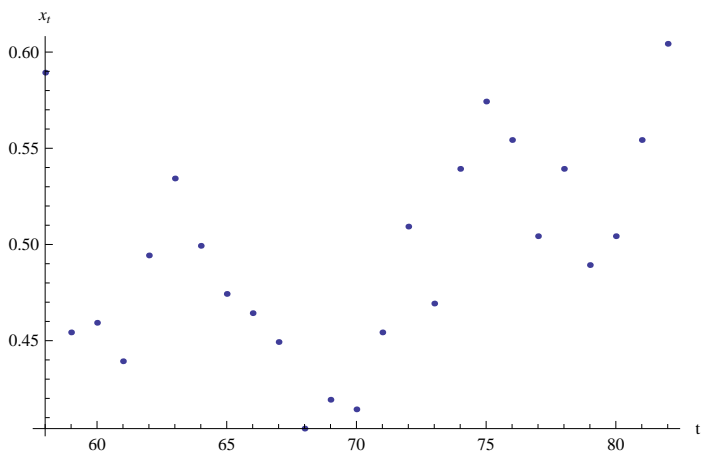
Obrázek 3.8: Graf PACF pro data x_t .

Jak je patrné z grafu, nerovnost není splněna pro $k = 26$ a 51 . Druhé porušení nerovnosti lze zanedbat, avšak vzhledem k výraznému rozdílu v případě $k = 26$ bychom mohli model AR(1) vyloučit. Graf ACF a PACF by bylo taktéž značně obtížné omezit křivkou U, proto je nepravděpodobné, že by šlo o model ARMA(p, q). Identifikace modelu se tedy nezdařila, což taktéž poukazuje k nestacionárním datům. Proto provedeme testy na jednotkový kořen.

τ_μ -test s $H_0 : x_t = x_{t-1} + \varepsilon_t$ proti $H_1 : x_t = \alpha + \varphi x_{t-1} + \varepsilon_t$ dává $\Phi_1 = 2.39762$. Na zvolené hladině významnosti $\alpha = 5\%$ je odpovídající kvantil 4.71, tedy H_0 nezamítáme.

Stejně tak se H_0 nezamítne v případě τ_τ -testu, kdy testujeme $H_0 : x_t = x_{t-1} + \varepsilon_t$ proti $H_1 : x_t = \alpha + \beta t + \varphi x_{t-1} + \varepsilon_t$. V tomto případě je totiž $\Phi_2 = 1.58626$ a odpovídající kvantil na hladině významnosti $\alpha = 5\%$ je roven 4.88.

Jelikož jsme v obou případech nulovou hypotézu nezamítli, testy prokázaly, že zkoumaná časová řada je nestacionární. Nestacionarita je zjevně způsobena odlehlými hodnotami x_1, \dots, x_{25} a rostoucím trendem u hodnot x_{26}, \dots, x_{55} . Proto se pokusíme o identifikaci modelu pro případ 25 pozorování, konkrétně pro data x_{58}, \dots, x_{82} na obrázku 3.9. Nejprve však ověříme předpokládanou stacionaritu.

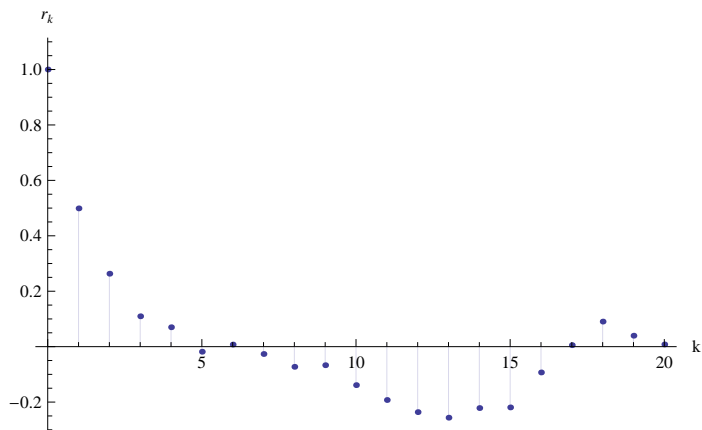


Obrázek 3.9: Graf časové řady x_{58}, \dots, x_{82} .

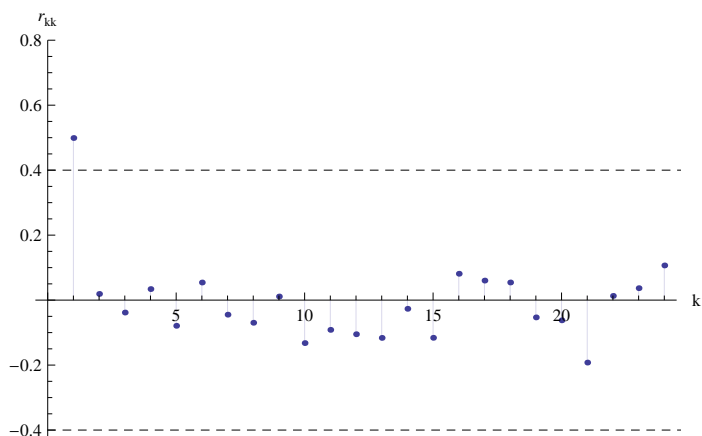
V případě τ_μ -testu je $\Phi_1 = 4.44665$ a odpovídající kvantil na hladině významnosti $\alpha = 5\%$ je 5.18, tedy nulovou hypotézu nezamítáme. Na hladině $\alpha = 10\%$

bychom vzhledem ke kvantilu 4.12 nulovou hypotézu zamítli. Nulovou hypotézu zamítáme v případě τ_τ -testu, jelikož $\Phi_2 = 6.01748$ a odpovídající kvantil na hladině $\alpha = 5\%$ je 5.68.

K identifikaci modelu opět využijeme grafy ACF a PACF.



Obrázek 3.10: Graf ACF pro data x_{58}, \dots, x_{82} .



Obrázek 3.11: Graf PACF pro data x_{58}, \dots, x_{82} .

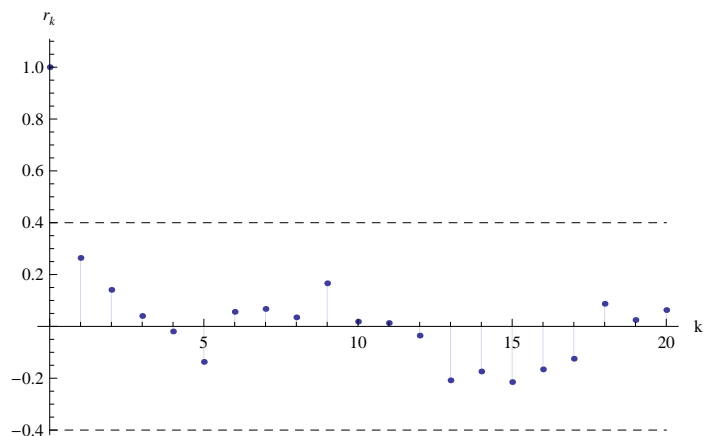
Z grafu PACF na obrázku 3.11 lze identifikovat bod useknutí $k_0 = 1$. V případě ACF bod useknutí patrně neexistuje a její tvar spolu s tvarem PACF vede k tomu, že model lze identifikovat jako AR(1). Odhady parametrů jsou

$$\begin{aligned}\hat{\varphi} &= 0.498936, \\ \hat{\sigma}^2 &= 0.00217286,\end{aligned}$$

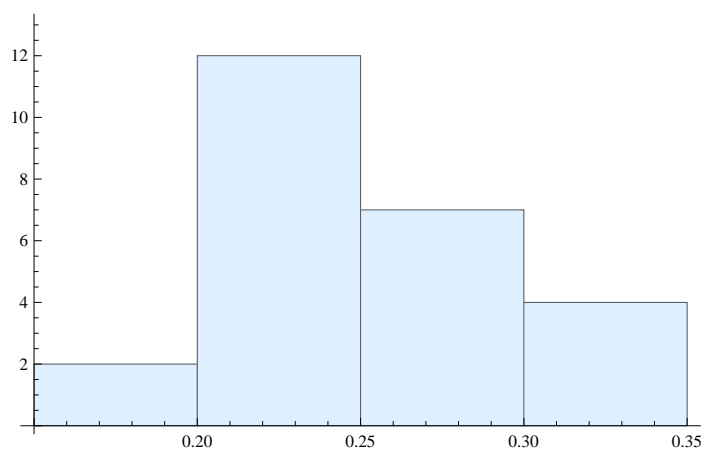
kde σ^2 je rozptyl bílého šumu. $|\hat{\varphi}| < 1$, což odpovídá stacionárnímu procesu.

Adekvátnost zvoleného modelu AR(1) ověříme ještě diagnostikou vypočteného bílého šumu podle vztahu $\hat{\varepsilon}_t = x_t - \hat{\varphi}x_{t-1}$.

Z ACF na obrázku 3.12 lze vidět, že pro vypočtený bílý šum platí $|r_k| < 2\sqrt{\frac{1}{n}}$ pro $k = 1, 2, \dots$ a je tedy nekorelovaný. Normalita vypočteného bílého šumu se nepotvrdila, jak naznačuje asymetrický histogram na obrázku 3.13.



Obrázek 3.12: Graf ACF pro vypočtený bílý šum.



Obrázek 3.13: Histogram vypočteného bílého šumu.

Závěrem lze říci, že jsme prokázali nestacionaritu zkoumané řady v rozsahu $n = 100$. Nestacionaritu je možné odstranit přechodem k diferencím logaritmů. Rozbor korelačních funkcí transformované řady prokázal, že ji lze považovat za bílý šum. Tomu odpovídající hypotéza náhodné procházky pro logaritmovanou řadu byla potvrzena testy na jednotkový kořen.

Nestacionaritu lze rovněž odstranit zkrácením původní řady a eliminací problematických úseků. Pro hodnoty směnného kurzu z konce prosince 2013 a ledna 2014 lze pak identifikovat stacionární AR(1) model, což potvrdily i testy na jednotkový kořen. Výsledky jsou smysluplné i přes porušení předpokladu normality bílého šumu.

Literatura

ANDĚL, J. (2005). *Základy matematické statistiky*. První vydání. Matfyzpress, Praha. ISBN 80-86732-40-1.

CIPRA, T. (2008). *Finanční ekonometrie*. První vydání. Ekopress, Praha. ISBN 978-80-86929-43-9.

DICKEY, D. A. (1976). *Estimation and hypothesis testing in nonstationary time series*. PhD thesis, Iowa State University.

DICKEY, D. A. a FULLER, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, **49**, 1057 – 1072.

Seznam obrázků

3.1	Vývoj kurzu Eura.	18
3.2	Transformovaná data z_t	18
3.3	Histogram časové řady z_t	19
3.4	Graf ACF pro z_t	19
3.5	Graf PACF pro z_t	20
3.6	Graf časové řady po odečtení aritmetického průměru.	21
3.7	Graf ACF pro data x_t	21
3.8	Graf PACF pro data x_t	22
3.9	Graf časové řady x_{58}, \dots, x_{82}	22
3.10	Graf ACF pro data x_{58}, \dots, x_{82}	23
3.11	Graf PACF pro data x_{58}, \dots, x_{82}	23
3.12	Graf ACF pro vypočtený bílý šum.	24
3.13	Histogram vypočteného bílého šumu.	24

Seznam tabulek

1.1	Vlastnosti autokorelační a parciální autokorelační funkce modelů $AR(p)$, $MA(q)$ a $ARMA(p,q)$. U značí křivku tvaru lin. kombinace klesajících geom. posloupností a sinusoid s geom. klesajícími amplitudami.	9
2.1	Testy na jednotkový kořen.	10
2.2	Kvantily rozdělení Φ_1 za platnosti $(\alpha, \varphi) = (0, 1)$ v případě modelu $y_t = \alpha + \varphi y_{t-1} + \varepsilon_t$	14
2.3	Kvantily rozdělení Φ_2 za platnosti $(\alpha, \beta, \varphi) = (0, 0, 1)$ v případě modelu $y_t = \alpha + \beta t + \varphi y_{t-1} + \varepsilon_t$	14