



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

BAKALÁRSKA PRÁCA

Ján Kostka

Lineárny regresný model s autokorelovanými reziduami

Katedra pravděpodobnosti a matematické statistiky

Vedúci bakalárskej práce: RNDr. Jitka Zichová, Dr.

Študijný program: Matematika

Študijný obor: Finanční matematika

Praha 2019

Prehlasujem, že som túto bakalársku prácu vypracoval(a) samostatne a výhradne s použitím citovaných prameňov, literatúry a ďalších odborných zdrojov.

Beriem na vedomie, že sa na moju prácu vzťahujú práva a povinnosti vyplývajúce zo zákona č. 121/2000 Sb., autorského zákona v platnom znení, najmä skutočnosť, že Univerzita Karlova má právo na uzavretie licenčnej zmluvy o použití tejto práce ako školského diela podľa §60 odst. 1 autorského zákona.

V dňa

Podpis autora

Na tomto mieste by som sa rád podakoval RNDr. Jitke Zichovej, Dr. za jej čas, rady a množstvo pripomienok k textu práce. Ďakujem rodičom za ich dôveru a za podporu počas celej doby štúdia.

Názov práce: Lineárny regresný model s autokorelovanými reziduami

Autor: Ján Kostka

Katedra: Katedra pravdepodobnosti a matematickej statistiky

Vedúci bakalárskej práce: RNDr. Jitka Zichová, Dr., Katedra pravdepodobnosti a matematickej statistiky

Abstrakt: Cieľom tejto práce je predstaviť postup pre analýzu lineárneho regresného modelu s autokorelovanými reziduami, ktorý je vhodný v prípade, keď sú pozorovania získané sledovaním veličín v čase. Pre reziduá predpokladáme lineárny model ARMA, prípadne ARIMA, čo rozširuje možnosti využitia. Analýza takýchto regresných modelov zahŕňa detekciu autokorelovanosti a s tým súvisiace testy nekorelovanosti reziduí, detekciu stacionarity s testom jednotkového koreňa, ďalej identifikáciu modelu pre reziduá a odhad všetkých parametrov regresného modelu metódou maximálnej vierohodnosti.

Kľúčové slová: autokorelovanosť reziduí, Durbin-Watsonov test, Ljung-Boxov test, Breusch-Godfreyho test, Dickey-Fullerov test, jednotkový koreň

Title: Linear regression model with autocorrelated residuals

Author: Ján Kostka

Department: Department of Probability and Mathematical Statistics

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

Abstract: The aim of this bachelor thesis is to introduce the algorithm for analysis of the linear regression model with autocorrelated residuals, which is applicable to time series data. For residuals, we assume the ARMA model, eventually ARIMA model, which enlarges the possibilities of application. The analysis of such regression models includes detection of autocorrelation and related tests, detection of stationarity and related unit root test, followed by model identification for residuals and maximum likelihood estimation of identified regression model.

Keywords: autocorrelated residuals, Durbin-Watson test, Ljung-Box test, Breusch-Godfrey test, Dickey-Fuller test, unit root nonstationarity

Obsah

Úvod	2
1 Úvod do časových radov	3
1.1 Stacionarita	3
1.2 Autokorelačná a parciálna autokorelačná funkcia	4
1.3 Modely lineárnych časových radov	5
1.4 Identifikácia modelu	7
1.5 Odhad modelu	8
2 Lineárny regresný model s autokorelovanými reziduami	9
2.1 Základy lineárnej regresie	9
2.2 Autokorelovanosť reziduí	10
2.3 Odhady podmienenou metódou maximálnej vierohodnosti	12
3 Testovanie reziduí	14
3.1 Testovanie autokorelovanosti	14
3.2 Testovanie stacionarity	16
4 Simulačná štúdia	18
4.1 Simulovaný regresný model	18
4.2 Analýza štatistických testov	19
5 Príklad	28
Záver	33
Zoznam použitej literatúry	34
Zoznam obrázkov	35
Zoznam tabuliek	36

Úvod

Lineárna regresia je užitočným nástrojom pri skúmaní vzťahu medzi veličinami náhodného vektoru

$$\begin{pmatrix} X_1 \\ \vdots \\ X_k \\ Y \end{pmatrix}.$$

Cieľ je nájsť lineárnu kombináciu X_1, \dots, X_k , ktorá by čo najlepšie vysvetľovala veličinu Y , teda koeficienty $\beta_0, \beta_1, \dots, \beta_k$, pre ktoré

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon,$$

v zmysle minimálneho rozptylu náhodnej chyby ε , resp. pre ktoré sú ε a X_i nekorelované. Aby sme koeficienty mohli odhadnúť, pozorujeme v praxi realizáciu vektorov, predpokladáme pre ne lineárny regresný model, pripomenutý v kapitole 2.1.

Vo finančnej praxi chceme najčastejšie pomocou lineárneho regresného modelu vysvetliť hodnotu nejakej veličiny Y_t v čase t pomocou jednej veličiny X_t . Ak pozorujeme X_t , máme lepšiu predstavu o Y_t v zmysle

$$\text{Var}[Y_t|X_t] = \text{Var}[\varepsilon_t] \leq \text{Var}[Y_t].$$

Pozorovania modelu, o ktorý sa zaujímate, sú teda vektory $\{(X_t, Y_t)^\top\}$, kde $\{Y_t\}$ a $\{X_t\}$ sú časové rady. Lineárny regresný model je však v praxi neadekvátny, nakoľko závislosť veličín v týchto radoch na ich minulých hodnotách často znamená, že závislosť existuje aj v časovom rade reziduí $\{\varepsilon_t\}$. Tú by sme mali zohľadniť, aby sme pri daných minulých pozorovaniach naozaj vysvetlili Y_t až na nepredikovateľné reziduá $\{u_t\}$.

Pre $\{\varepsilon_t\}$ budeme predpokladať nejaký stacionárny model, alebo stacionaritu ich vhodnej transformácie. Súhrnu potrebnej teórie časových radov je venovaná kapitola 1. To vedie k definícii lineárneho regresného modelu s autokorelovanými reziduami v kapitole 2.2. Za daných predpokladov pre časový rad reziduí vieme sformulovať obecný postup pre jeho analýzu. Transformovanie reziduí vyriešime transformáciou celého regresného modelu a za predpoladu stacionarity vieme parametre modelu odhadnúť metódou maximálnej vierohodnosti v kapitole 2.3.

Kľúčovým bodom analýzy je posúdiť, či sú rezidua nezávislé (za normality nekorelované). Budeme testovať hypotézu o nekorelovanosti reziduí, predstavíme Durbin-Watsonov, Ljung-Boxov a Breusch-Godfreyho test. V prípade korelovanosti posudzujeme či sú stacionárne alebo ich treba transformovať, v súvislosti s tým uvádzame Dickey-Fullerov test. Rozdiel v porovnaní s analýzou časových radov je, že reziduá nemôžeme pozorovať. Problematike sa venuje kapitola 3.

V praktickej časti práce skúmame uvedené štatistické testy. Simulovaním sa presvedčíme, že dodržiajú hladinu a odhadneme ich silu proti vybraným alternatívam. Na záver ilustrujeme postup z kapitoly 2.2 na reálnych dátach. Pomocou kurzu CZK/USD vysvetlíme kurz CZK/EUR.

1. Úvod do časových radov

V tejto kapitole zavedieme pojem časového radu, jeho základné vlastnosti a charakteristiky. Definície 1–10 sú prevzaté z prác Cipra (2008) a Tsay (2002).

Definícia 1. *Majme pevne daný pravdepodobnostný priestor (Ω, \mathcal{A}, P) . Nech $T \subseteq \mathbb{Z}$. Postupnosť náhodných veličín $\{X_t\}_{t \in T}$ nazývame časový rad. Index t je čas, veličiny X_t sú pozorovania.*

Časový rad s najjednoduchšou štruktúrou je postupnosť nezávislých rovnako rozdelených (i.i.d.) náhodných veličín.

Definícia 2. *Časový rad $\{u_t\}_{t \in T}$ nekorelovaných veličín s nulovou strednou hodnotou a spoločným rozptylom σ_u^2 sa nazýva slabý biely šum. Ak sú $\{u_t\}$ nezávislé rovnako rozdelené, hovoríme o striktnom bielom šume, ak $u_t \sim N(0, \sigma_u^2)$, hovoríme o Gaussovskom bielom šume.*

1.1 Stacionarita

Uvažujeme iba rady, ktoré sú v nejakom zmysle v čase invariantné. Táto vlastnosť sa nazýva stacionarita, rozlíšime striktnú a slabú. Neskôr budeme predpokladať normalitu pozorovaní, pre ktoré definície splývajú.

Postupnosť je slabo stacionárna keď má konštantnú strednú hodnotu a kovariancia pozorovaní sa v čase nemení, je funkciou vzdialenosti.

Definícia 3. *Časový rad $\{X_t\}_{t \in T}$ s konečnými druhými momentmi je slabo stacionárny, keď pre každé t, k je*

$$\begin{aligned} E[X_t] &= \mu, \\ \text{Cov}[X_t, X_{t+k}] &= \gamma_k. \end{aligned}$$

V striktno stacionárnej postupnosti sú rozdelenia všetkých podvektorov invariantné voči posunutiu.

Definícia 4. *Časový rad $\{X_t\}_{t \in T}$ je striktno stacionárny, keď pre každé $k, n, t_1 \dots t_n$*

$$\mathcal{L}(X_{t_1} \dots X_{t_n}) = \mathcal{L}(X_{t_1+k} \dots X_{t_n+k}).$$

Poznámka. Ak $\text{Var}(X_t) < \infty$ potom striktná stacionarita implikuje slabú. Slabý (resp. striktný) biely šum je slabo (resp. striktno) stacionárny a platí

$$\gamma_k = \begin{cases} \sigma_u^2 & \text{pre } k = 0, \\ 0 & \text{inak.} \end{cases}$$

Ďalej budeme pojmom stacionarita rozumieť slabú stacionaritu. Dôležitým príkladom stacionárneho časového radu je lineárny proces. Pozorovanie v čase t v lineárnom procese je (nekonečnou) lineárnou kombináciou bieleho šumu do času t . Neskôr ukážeme, že všetky postupnosti, ktoré budeme uvažovať, sa dajú za určitých podmienok previesť na tento tvar.

Definícia 5. Nech $\{u_t\}_{t \in \mathbb{Z}}$ je biely šum. Postupnosť $\{X_t\}_{t \in T}$ je kauzálny lineárny proces, keď pre každé $t \in T$ je

$$X_t = \sum_{i=0}^{\infty} c_i u_{t-i}$$

pre nejaké koeficienty $\{c_i\}_{i=0}^{\infty}$, $\sum |c_i| < \infty$.

Poznámka. Ak $\sum |c_i| < \infty$ potom $\sum_{i=0}^{\infty} c_i u_{t-i}$ konverguje P-skoro iste. Lineárny proces je teda dobre definovaný.

Odvodíme vlastnosti lineárnych procesov. Keďže je $\{u_t\}$ biely šum, platí

$$\mathbb{E}[X_t] = \mathbb{E}\left[\sum_{i=0}^{\infty} c_i u_{t-i}\right] = \sum_{i=0}^{\infty} \mathbb{E}[c_i u_{t-i}] = 0,$$

podobne pre rozpyl

$$\sigma_X^2 = \text{Var}[X_t] = \text{Var}\left[\sum_{i=0}^{\infty} c_i u_{t-i}\right] = \sum_{i=0}^{\infty} \text{Var}[c_i u_{t-i}] = \sum_{i=0}^{\infty} c_i^2 \text{Var}[u_{t-i}] = \sigma_u^2 \sum_{i=0}^{\infty} c_i^2.$$

Autokovariančná funkcia

$$\begin{aligned} \gamma_k &= \mathbb{E}\left[\left(\sum_{i=0}^{\infty} c_i u_{t-i}\right)\left(\sum_{i=0}^{\infty} c_i u_{t+k-i}\right)\right] = \mathbb{E}\left[\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_i c_j u_{t-i} u_{t+k-j}\right] \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_i c_j \mathbb{E}[u_{t-i} u_{t+k-j}] = \sum_{i=0}^{\infty} c_i c_{i+k} \mathbb{E}[u_{t-i}^2] = \sigma_u^2 \sum_{i=0}^{\infty} c_i c_{i+k} \end{aligned}$$

nezávisí na t , lineárny proces je teda stacionárny.

1.2 Autokorelačná a parciálna autokorelačná funkcia

Pre stacionárne postupnosti je možné zaviesť autokorelačnú a parciálnu autokorelačnú funkciu.

Definícia 6. Nech $\{X_t\}_{t \in T}$ je stacionárny časový rad. Potom definujeme autokorelačnú funkciu (ACF) pre meškanie k ,

$$\rho_k = \frac{\text{Cov}[X_t, X_{t+k}]}{\text{Var}[X_t]} = \frac{\gamma_k}{\gamma_0}.$$

V praxi pozorujeme úsek jednej realizácie $X_1(\omega) \dots X_n(\omega)$, $\omega \in \Omega$. To nás motivuje odhadnúť autokorelačnú funkciu výberovým korelačným koeficientom,

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)}{\sum_{t=1}^n (X_t - \bar{X}_n)^2}, \quad (1.1)$$

kde \bar{X}_n je výberový priemer a zároveň odhad strednej hodnoty μ stacionárneho radu.

Parciálna autokorelačná funkcia (PACF) je definovaná ako parciálny korelačný koeficient medzi veličinami X_t a X_{t-k} pri daných $X_{t-1} \dots X_{t-k+1}$. Parciálny korelačný koeficient je zavedený v knihe Anděl (2007, str. 39-43). PACF pre meškanie k budeme značiť ϕ_{kk} .

Ekvivalentne je možné zaviesť ϕ_{kk} ako koeficient pri X_{t-k} najlepšej lineárnej predikcie X_t pomocou $X_{t-1} \dots X_{t-k}$,

$$X_t = \phi_{k1}X_{t-1} + \dots + \phi_{kk}X_{t-k} + \varepsilon.$$

Odhady $\hat{\phi}_{kk}$, $k = 1, 2, \dots$ sa počítajú rekurentne. Vezmeme $\hat{\phi}_{11} = \hat{\rho}_1$ a

$$\hat{\phi}_{kk} = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \cdot \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \cdot \hat{\rho}_j} \quad \text{pre } k > 1,$$

kde

$$\hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk} \cdot \hat{\phi}_{k-1,k-j} \quad \text{pre } j = 1 \dots k-1.$$

Podrobne o PACF a jej odhade píše Wei (1990, str. 12-16).

1.3 Modely lineárnych časových radov

Autoregresný model

V kapitole 1.1 sme definovali lineárny proces s netriviálnou kovariančnou štruktúrou, ktorá je v čase invariantná. To nás vedie k myšlienke, že minulé pozorovania môžu byť užitočné v predikcii pozorovania v čase t a následne aj k definícii autoregresnej postupnosti.

Definícia 7. *Nech $\{u_t\}_{t \in \mathbb{Z}}$ je biely šum. $\{X_t\}_{t \in \mathbb{Z}}$ je autoregresný proces rádu p ($AR(p)$), keď pre každé $t \in \mathbb{Z}$ platí*

$$X_t = u_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p}. \quad (1.2)$$

Charakteristický polynóm procesu je $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$.

Poznámka. V autoregresnom modeli je X_t až na biely šum vysvetlená veličinami $X_{t-1} \dots X_{t-p}$. Platí

$$\begin{aligned} \mathbf{E} [X_t | X_{t-1} \dots X_{t-p}] &= \phi_1 X_{t-1} + \dots + \phi_p X_{t-p}, \\ \mathbf{Var} [X_t | X_{t-1} \dots X_{t-p}] &= \sigma_u^2. \end{aligned}$$

Nie každý autoregresný časový rad je stacionárny, existujú postupnosti, ktoré explodujú alebo sú náhodnou prechádzkou. Stacionarita úzko súvisí s charakteristickým polynómom $\Phi(z)$.

Veta 1. *Nech $\{X_t\}_{t \in \mathbb{Z}}$ je proces $AR(p)$. Ak sú všetky komplexné korene charakteristického polynómu väčšie ako 1 v absolútnej hodnote, potom je $\{X_t\}$ stacionárna postupnosť a kauzálny lineárny proces, t.j. pre každé t je $X_t = u_t + \sum_{i=1}^{\infty} a_i u_{t-i}$ pre nejaké $\{a_i\}$, $\sum |a_i| < \infty$.*

Dôkaz. Dôkaz je uvedený v práci Prášková (2007, str. 59-61).

Poznámka. V stacionárnom autoregresnom modeli je autokorelačná funkcia ρ_k obecné nenulová pre každé k . Dá sa ukázať, že platí rekurentný vzťah

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p}, \quad (1.3)$$

z toho plynie

$$\rho_k = \alpha_1 z_1^{-k} + \dots + \alpha_p z_p^{-k}$$

pre nejaké $\alpha_1 \dots \alpha_p$, ak sú $z_1 \dots z_p$ rôzne korene $\Phi(z)$. Pretože $|z_i| > 1$, autokorelačná funkcia konverguje k nule, asymptoticky exponenciálne.

Model kĺzavých súčtov

V modeli kĺzavých súčtov sú pozorovania konečnou lineárnou kombináciou minulého bieleho šumu.

Definícia 8. *Nech $\{u_t\}_{t \in \mathbb{Z}}$ je biely šum. $\{X_t\}_{t \in \mathbb{Z}}$ je proces kĺzavých súčtov rádu q ($MA(q)$), keď pre každé $t \in \mathbb{Z}$ platí*

$$X_t = u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q}. \quad (1.4)$$

Charakteristický polynóm procesu je $\Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$.

Postupnosť $MA(q)$ je z definície lineárny proces, z toho plynie stacionarita a nulová autokorelačná funkcia ρ_k pre $k > q$. Hovoríme, že q je bod useknutia ρ_k .

Už sme ukázali, že autoregresná postupnosť sa za určitých podmienok dá vyjadriť ako lineárna kombinácia bieleho šumu. Kvôli schopnosti predikovať je naopak vhodné vyjadriť časový rad ako nekonečnú lineárnu kombináciu minulých pozorovaní. Táto vlastnosť sa nazýva invertibilita, postupnosť $AR(p)$ je invertibilná z definície. Sformulujeme podmienku, kedy je invertibilná postupnosť $MA(q)$.

Veta 2. *Nech $\{X_t\}_{t \in \mathbb{Z}}$ je proces $MA(q)$. Ak sú všetky komplexné korene charakteristického polynómu $\Theta(z)$ väčšie ako 1 v absolútnej hodnote, potom je $\{X_t\}$ invertibilný proces, t.j. $u_t = X_t + \sum_{i=1}^{\infty} b_i X_{t-i}$ pre nejaké koeficienty $\{b_i\}_{i=1}^{\infty}$, $\sum |b_i| < \infty$.*

Dôkaz. Vid Prášková (2007, str. 71).

ARMA model

Časové rady AR a MA sú zobecnené v modeli $ARMA$.

Definícia 9. *Nech $\{u_t\}_{t \in \mathbb{Z}}$ je biely šum. $\{X_t\}_{t \in \mathbb{Z}}$ spĺňa model $ARMA(p,q)$, keď pre každé $t \in \mathbb{Z}$ platí*

$$X_t = u_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i u_{t-i}. \quad (1.5)$$

Poznámka. Ak použijeme operátor spätného posunutia B , pre ktorý platí $B^k X_t = X_{t-k}$, $k = 0, 1, 2, \dots$, dostaneme zápis postupnosti $ARMA(p, q)$ pomocou charakteristických polynómov:

$$\begin{aligned}\Phi(B)X_t &= \Theta(B)u_t && ARMA(p, q), \\ \Phi(B)X_t &= u_t && AR(p), \\ X_t &= \Theta(B)u_t && MA(q).\end{aligned}$$

Podmienka stacionarity pre $ARMA(p, q)$ je rovnaká ako u $AR(p)$, podmienka invertibility je za predpokladu stacionarity rovnaká ako u $MA(q)$, viac píše Prášková (2007, str. 67-72). Autokorelačná funkcia ρ_k aj parciálna autokorelačná funkcia ϕ_{kk} procesu $ARMA(p, q)$ nemajú bod useknutia. V stacionárnom a invertibilnom $ARMA$ procese konvergujú obe exponenciálne k nule a pre ρ_k platí od bodu q vzťah (1.3).

Poznámka. Definície 5, 7, 8 a 9 sa dajú rozšíriť aj pre nenulovú strednú hodnotu, my však chceme procesom $ARMA$ modelovať reziduá lineárneho regresného modelu (s interceptom), kde $\mu \neq 0$ stráca zmysel.

1.4 Identifikácia modelu

K identifikácii AR a MA modelov je možné použiť korelačné funkcie, ich odhady majú za predpokladu normality bieleho šumu približne normálne rozdelenie.

Pri identifikácii rádu autoregresných postupností je nástrojom PACF. Ak je $\{X_t\}_{t \in \mathbb{Z}}$ proces $AR(p)$, potom pre $k > p$ je $\phi_{kk} = 0$. Vychádza sa z aproximácie (Cipra, 2008, str. 331), že ak $\phi_{kk} = 0$ pre všetky $k > k_0$, potom pre veľké n má $\hat{\phi}_{kk}$ približne normálne rozdelenie s rozptylom $1/n$, teda

$$\sqrt{n}\hat{\phi}_{kk} \sim N(0, 1) \quad \text{pre } k > k_0.$$

Dôsledok. Hypotézu, že pozorovaný rad je realizáciou procesu $AR(p_0)$, zamietneme na hladine α , keď pre $k > p_0$

$$\left| \hat{\phi}_{kk} \right| > \frac{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}{\sqrt{n}}, \quad (1.6)$$

kde $\Phi^{-1}(\alpha)$ je kvantil normovaného normálneho rozdelenia.

Identifikáciu modelu kľzavých súčtov $MA(q)$ môžeme založiť na odhadnutej ACF, ktorá je pre $k > q$ dostatočne malá. Ak $\rho_k = 0$ pre všetky $k > k_0$, potom sa pre rozptyl $\hat{\rho}_k$ používa aproximácia $\text{Var}[\hat{\rho}_k] \approx n^{-1}(1 + 2\rho_1^2 + \dots + 2\rho_{k_0}^2)$. Cipra (2008, str. 331) uvádza

$$\sqrt{n}\hat{\rho}_k \sim N\left(0, 1 + 2 \sum_{i=1}^{k_0} \rho_i^2\right) \quad \text{pre } k > k_0.$$

Viac o vlastnostiach výberovej ACF je v knihe Box a kol. (1994, str. 32-34). V praxi $\rho_1 \dots \rho_{k_0}$ nepozorujeme a vezmeme ich odhady.

Dôsledok. Testujeme hypotézu, že pozorovaný rad je realizáciou procesu $MA(q_0)$. Model zamietame ako nevhodný, keď pre $k > q_0$ je

$$|\hat{\rho}_k| > \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\frac{1}{n} \left(1 + 2 \sum_{i=1}^{q_0} \hat{\rho}_i^2 \right)}. \quad (1.7)$$

Poznámka. Identifikáciu modelu pomocou ACF a PACF by sme chceli založiť na celom odhadnutom priebehu, pri rozhodovaní sa preto musíme zachovať do určitej miery subjektívne.

Informačné kritérium

Iným prístupom k identifikácii modelu je informačné kritérium, ktoré rozhodne o rádoch procesu $ARMA$ podľa pravidla

$$(p, q) = \arg \min_{k, l} A(k, l).$$

$A(k, l)$ je kritérium, v ktorom vystupuje odhadnutý rozptyl bieleho šumu a penalizačná funkcia, ktorá eliminuje príliš veľké rády.

Najznámejšie je kritérium AIC (Akaike information criterion), ktoré zavádza Cipra (2008, str. 342):

$$AIC(k, l) = \log \left(\hat{\sigma}_{u(k, l)}^2 \right) + \frac{2(k + l + 1)}{n},$$

kde $\hat{\sigma}_{u(k, l)}^2$ je rozptyl reziduí po odhade v modeli $ARMA(k, l)$. Problematike odhadu parametrov modelu sa venuje kapitola 1.5 a 2.3.

1.5 Odhad modelu

Odhad parametrov modelu $AR(p)$ môžeme založiť na regresii. Platí

$$X_t = u_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p},$$

pre $t = p + 1, \dots, n$ máme pozorované hodnoty časoveho radu a odhadneme $(\phi_1, \dots, \phi_p)^\top$ metódou najmenších štvorcov, ktorá je pripomenutá v kapitole 2.1. Iný prístup spočíva v momentovom odhade. Vzhľadom k (1.3) je $(\phi_1, \dots, \phi_p)^\top$ riešením sústavy

$$\left(\begin{array}{ccc|c} \rho_0 & \dots & \rho_{p-1} & \rho_1 \\ \vdots & \ddots & \vdots & \vdots \\ \rho_{p-1} & \dots & \rho_0 & \rho_p \end{array} \right).$$

Nahradením autokorelácií empirickými protajškami dostaneme momentový odhad koeficientov.

Obecne je možné odhadnúť parametre stacionárneho a invertibilného radu $ARMA(p, q)$ momentovo, v praxi sa však na odhadovanie používa software s vlastnými procedúrami. Najčastejšie sú implementované odhady, ako uvádza Cipra (2008, str. 345), založené na podmienenej metóde maximálnej vierohodnosti, ktorú formulujeme v kapitole 2.3 v kontexte odhadu parametrov lineárneho regresného modelu s autokorelovanými reziduami.

2. Lineárny regresný model s autokorelovanými reziduami

2.1 Základy lineárnej regresie

V lineárnej regresii, ako popisuje napr. Anděl (2007), vysvetľujeme náhodnú veličinu Y náhodným vektorom $\mathbf{X} = (1, X_1, \dots, X_k)^\top$. Y sa nazýva vysvetľovaná premenná, \mathbf{X} je vektor regresorov. Model má tvar

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon = \boldsymbol{\beta}^\top \mathbf{X} + \varepsilon,$$

kde ε je spojitá veličina s nulovou strednou hodnotou a konečným rozptylom σ_ε^2 , ε a \mathbf{X} sú nezávislé. ε nazývame reziduom.

V lineárnom regresnom modeli (LRM) máme pozorovania

$$\begin{pmatrix} \mathbf{X}_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} \mathbf{X}_2 \\ Y_2 \end{pmatrix}, \dots, \begin{pmatrix} \mathbf{X}_n \\ Y_n \end{pmatrix}, \quad \mathbf{X}_t = (1, X_{t1}, \dots, X_{tk})^\top,$$

ktoré spĺňajú

$$Y_t = \beta_0 + \beta_1 X_{t1} + \dots + \beta_k X_{tk} + \varepsilon_t, \quad t = 1, 2, \dots, n. \quad (2.1)$$

Predpokladá sa, že $\boldsymbol{\varepsilon} = \{\varepsilon_t\}_{t=1}^n$ sú i.i.d. náhodné veličiny, $\mathbf{E}[\varepsilon_t] = 0$, $\text{Var}[\varepsilon_t] = \sigma_\varepsilon^2$. Označme

$$\mathbb{X} = \begin{pmatrix} 1 & X_{11} & \dots & X_{1k} \\ 1 & X_{21} & \dots & X_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \dots & X_{nk} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1^\top \\ \mathbf{X}_2^\top \\ \vdots \\ \mathbf{X}_n^\top \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}.$$

Maticový zápis lineárneho regresného modelu je potom

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Pre odhadovanie parametrov $\boldsymbol{\beta}$ sa používa odhad metódou najmenších štvorcov, ktorý je konzistentný a za predpokladu $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ je aj odhadom metódou maximálnej vierohodnosti. Je známe, že $\hat{\boldsymbol{\beta}}$ sa spočíta ako

$$\hat{\boldsymbol{\beta}} = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \mathbf{Y},$$

ak má \mathbb{X} plnú hodnotu $k + 1$.

- $\hat{\mathbf{Y}} = \mathbb{X}\hat{\boldsymbol{\beta}}$ je odhad $\mathbf{E}[\mathbf{Y}|\mathbb{X}]$ a nazýva sa vektor vyrovnaných hodnôt odozvy.
- $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}}$ je vektor odhadnutých rezidií.
- $SS_e = \hat{\boldsymbol{\varepsilon}}^\top \hat{\boldsymbol{\varepsilon}}$ je reziduálny súčet štvorcov.

- $SS_c = \sum_{t=1}^n [Y_t - \bar{Y}_n]^2$ nazývame celkový súčet štvorcov.
- $R^2 = 1 - \frac{SS_e}{SS_c}$ je koeficient determinácie.

Rozptylová matica odhadu je $\text{Var}[\hat{\beta}] = \sigma_\varepsilon^2 (\mathbb{X}^\top \mathbb{X})^{-1}$. Odhad smerodatnej odchylky $\hat{\beta}_i$ je teda $\sqrt{\hat{\sigma}_\varepsilon^2 x_{ii}}$ kde x_{ii} je i -ty diagonálny prvok $(\mathbb{X}^\top \mathbb{X})^{-1}$.

Za normality reziduí je navyše

$$\frac{SS_e}{\sigma_\varepsilon^2} \sim \chi_{n-k-1}^2,$$

z toho plynie, že $\hat{\sigma}_\varepsilon^2 = \frac{SS_e}{n-k-1}$ je nestranný odhad σ_ε^2 . Ďalej t -štatistika

$$T = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}_\varepsilon^2 x_{ii}}}$$

má t -rozdelenie s $n-k-1$ stupňami voľnosti, čo sa používa na testovanie hypotézy $\beta_i = c_0$, najčastejšie pre $c_0 = 0$.

2.2 Autokorelovanosť reziduí

Vo finančnej ekonometrii je typické, že pozorovania regresného modelu nemajú charakter náhodného výberu, ale sú sledované ako vývoj finančných veličín v čase. Ak sú v modeli časové rady

$$\{Y_t\}_{t=1}^n, \quad \{X_{ti}\}_{t=1}^n, \quad i = 1 \dots k,$$

autokorelované, model v praxi často vykazuje aj porušenie predpokladu nekorelovanosti reziduí $\{\varepsilon_t\}_{t=1}^n$.

Cipra (2008, str. 94) uvádza možné príčiny autokorelovanosti reziduí, napríklad:

1. Medzi regresormi chýbajú niektoré veličiny, ktorých časové rady vykazujú autokorelovanosť, ktorá sa potom presunie do reziduí.
2. Medzi regresormi chýbajú minulé hodnoty vysvetľovanej premennej alebo niektorých regresorov.
3. Regresný vzťah je nelineárny.

V prípade autokorelovanosti reziduí je vhodným modelom pre chyby $\{\varepsilon_t\}_{t=1}^n$ stacionárny a invertibilný $ARMA$ proces. Dostaneme lineárny regresný model s autokorelovanými reziduami.

Definícia 10. Časový rad regresorov a odoziev

$$\begin{pmatrix} \mathbf{X}_1 \\ Y_1 \end{pmatrix}, \begin{pmatrix} \mathbf{X}_2 \\ Y_2 \end{pmatrix}, \dots, \begin{pmatrix} \mathbf{X}_n \\ Y_n \end{pmatrix}$$

splňa lineárny regresný model s autokorelovanými reziduami (LRM s ACR), keď platí

$$\mathbf{Y} = \mathbb{X}\beta + \varepsilon,$$

$\varepsilon \sim ARMA(p,q)$, ε je nezávislá s \mathbb{X} .

Model je teda zobecnením (2.1) s použitím (1.5). Po zložkách má tvar

$$\begin{aligned} Y_t = & \beta_0 + \beta_1 X_{t1} + \cdots + \beta_k X_{tk} \\ & + \phi_1 \varepsilon_{t-1} + \cdots + \phi_p \varepsilon_{t-p} \\ & + u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q}, \end{aligned} \quad (2.2)$$

kde $\{u_t\}_{t \in \mathbb{Z}}$ je biely šum, $\varepsilon_t = Y_t - \boldsymbol{\beta}^\top \mathbf{X}_t$.

Poznámka. V LRM s ACR budeme vždy predpokladať normalitu bieleho šumu $\{u_t\}$.

Naším postupom pri analýze LRM s autokorelovanými reziduami bude algoritmus naznačený v knihe Tsay (2002, str. 71).

1. Lineárny regresný model

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

odhadneme metódou najmeších štvorcov. Odhadnuté chyby $\hat{\boldsymbol{\varepsilon}}$ sú predmetom záujmu, nakoľko nám pomôžu detekovať možné autokorelácie v $\boldsymbol{\varepsilon}$. Chceme testovať hypotézu $H_0 : \boldsymbol{\varepsilon}$ je biely šum. K tomu budeme používať Ljung-Boxov, Durbin-Watsonov alebo Breusch-Godfreyho test, ktoré sú popísané v kapitole 3.1. Testy sú navrhnuté na testovanie hypotézy H_0 o nekorelovanosti reziduí, testová štatistika a kritický obor sú typicky funkciami $\hat{\boldsymbol{\varepsilon}}$ a počtu regresorov v modeli. Ak hypotézu nemôžeme zamietnuť, prijmeme klasický LRM (2.1), v ktorom je $\hat{\boldsymbol{\beta}}$ najlepší nestranný odhad. Ak hypotézu zamietame, sú $\varepsilon_1, \dots, \varepsilon_n$ autokorelované a pokračujeme krokom 2.

2. Aby sme mohli pre reziduá identifikovať a odhadnúť vhodný *ARMA* model, potrebujeme posúdiť ich stacionaritu. Okrem prehliadky grafu $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ môžeme aplikovať test na jednotkový koreň. Koreň $z \approx 1$ v autoregresnom polynóme reziduí spôsobuje nestacionaritu, jeho prítomnosť budeme testovať Dickey-Fullerovým testom, viac o problematike je v kapitole 3.2. Ak hypotézu o jednotkovom koreni nemôžeme zamietnuť, vezmeme prvé diferencie regresorov a odozvy, dostaneme nový model

$$\Delta Y_t = \beta_0 + \beta_1 \Delta X_{t1} + \cdots + \beta_k \Delta X_{tk} + \varepsilon_t \quad t = 2, 3, \dots, n$$

a vrátíme sa na 1. krok. Ak hypotézu o jednotkovom koreni zamietame, možno považovať $\boldsymbol{\varepsilon}$ za stacionárnu postupnosť. Na reziduá $\hat{\boldsymbol{\varepsilon}}$ aplikujeme metódy z kapitoly 1.4 a identifikujeme model *ARMA*(p, q). Budeme teda odhadovať model v tvare (2.2). Prejdeme na krok 3.

3. Parametre modelu $\beta_0 \dots \beta_k, \phi_1 \dots \phi_p, \theta_1 \dots \theta_q$ odhadneme metódou maximálnej vierohodnosti z kapitoly 2.3. Adekvátnosť modelu overíme testovaním nekorelovanosti reziduí $\mathbf{u} = (u_1, \dots, u_n)^\top$.

Ak bolo potrebných viac iterácií na stacionarizovanie reziduálnej zložky $\boldsymbol{\varepsilon}$, svedčí to o prítomnosti viacnásobného jednotkového koreňa v autoregresnom polynóme. Výsledkom analýzy je odhadnutý model

$$\Delta^d Y_t = \hat{\boldsymbol{\beta}}^\top \Delta^d \mathbf{X}_t + \sum_{i=1}^p \hat{\phi}_i \hat{\varepsilon}_{t-i} + \hat{u}_t + \sum_{i=1}^q \hat{\theta}_i \hat{u}_{t-i}, \quad t = d+1, \dots, n. \quad (2.3)$$

Zo vzťahu

$$\Delta^d Y_t = Y_t - \binom{d}{1} Y_{t-1} + \binom{d}{2} Y_{t-2} - \dots + (-1)^d Y_{t-d}$$

dostaneme regresný vzťah pre pôvodné Y_t .

2.3 Odhady podmienenou metódou maximálnej vierohodnosti

O odhade parametrov v modeli *ARMA* píše Wei (1990, str. 137). Predpokladajme, že pozorujeme časový rad $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$ generovaný modelom *ARMA*(p, q),

$$u_t = \varepsilon_t - \phi_1 \varepsilon_{t-1} - \dots - \phi_p \varepsilon_{t-p} - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q}.$$

Ak sú dané počiatkové hodnoty $\mathbf{u}^* = (u_0 \dots u_{1-q})^\top$, $\boldsymbol{\varepsilon}^* = (\varepsilon_0 \dots \varepsilon_{1-p})^\top$, môžeme pre dané $\boldsymbol{\phi} = (\phi_1 \dots \phi_p)^\top$, $\boldsymbol{\theta} = (\theta_1 \dots \theta_q)^\top$ spočítať odhad reziduí

$$u_t(\boldsymbol{\phi}, \boldsymbol{\theta} \mid \mathbf{u}^*, \boldsymbol{\varepsilon}^*) = u_t(\boldsymbol{\phi}, \boldsymbol{\theta}), \quad t = 1 \dots n$$

rekurentne:

$$u_t(\boldsymbol{\phi}, \boldsymbol{\theta}) = \varepsilon_t - \phi_1 \varepsilon_{t-1} - \dots - \phi_p \varepsilon_{t-p} - \theta_1 u_{t-1}(\boldsymbol{\phi}, \boldsymbol{\theta}) - \dots - \theta_q u_{t-q}(\boldsymbol{\phi}, \boldsymbol{\theta}).$$

Ak predpokladáme $\mathbf{E}[u_t] = \mathbf{E}[\varepsilon_t] = 0 \forall t$, môžeme vziať napr. $u_0 = \dots = u_{1-q} = 0$, $\varepsilon_1 = \dots = \varepsilon_{1-p} = 0$.

Predpokladáme, že \mathbf{u} je normálny biely šum, ktorého hustota je

$$f_{\mathbf{u}}(x_1 \dots x_n) = \frac{1}{(2\pi)^{n/2} \sigma_u^n} \exp \left\{ -\frac{1}{2\sigma_u^2} \sum_{t=1}^n x_t^2 \right\}.$$

Vierohodnostná funkcia modelu je potom

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma_u^2 \mid \mathbf{u}^*, \boldsymbol{\varepsilon}^*) = \frac{1}{\sigma_u^n} \exp \left\{ -\frac{1}{2\sigma_u^2} \sum_{t=1}^n [u_t(\boldsymbol{\phi}, \boldsymbol{\theta})]^2 \right\}.$$

Logaritmická vierohodnosť je

$$\ell(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma_u^2 \mid \mathbf{u}^*, \boldsymbol{\varepsilon}^*) = -n \log \sigma_u - \frac{1}{2\sigma_u^2} S(\boldsymbol{\phi}, \boldsymbol{\theta} \mid \mathbf{u}^*, \boldsymbol{\varepsilon}^*).$$

kde

$$S(\boldsymbol{\phi}, \boldsymbol{\theta} \mid \mathbf{u}^*, \boldsymbol{\varepsilon}^*) = S(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{t=1}^n [u_t(\boldsymbol{\phi}, \boldsymbol{\theta})]^2.$$

Z toho vidíme, že ak ℓ je maximálna, potom $\boldsymbol{\phi}, \boldsymbol{\theta}$ minimalizujú súčet reziduálnych štvorcov $S(\boldsymbol{\phi}, \boldsymbol{\theta})$.

Definícia 11. Nech $\boldsymbol{\varepsilon} \sim \text{ARMA}(p, q)$. $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})^\top$ je odhad podmienenou vierohodnosťou pri daných $\mathbf{u}^*, \boldsymbol{\varepsilon}^*$, keď

$$(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})^\top = \arg \min_{\boldsymbol{\phi}, \boldsymbol{\theta}} S(\boldsymbol{\phi}, \boldsymbol{\theta} \mid \mathbf{u}^*, \boldsymbol{\varepsilon}^*).$$

$\hat{\mathbf{u}} = \mathbf{u}(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ je odhadnutý biely šum a $SS_e = S(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})$ je reziduálny súčet štvorcov. Odhad rozptylu bieleho šumu je založený na SS_e , rovnako ako v LRM, to znamená

$$\hat{\sigma}_u^2 = \frac{SS_e}{n - p - q}.$$

Problematika hľadania odhadu je softwarová záležitosť, používajú sa iteračné algoritmy a ako počiatočné riešenie sa vezmú momentové odhady parametrov. Dá sa uvažovať varianta, v ktorej zvolíme počiatočnú podmienku $u_p = u_{p-1} = \dots = u_{p+1-q} = 0$ a počítame $u_t(\boldsymbol{\phi}, \boldsymbol{\theta})$ pre $t = p + 1, \dots, n$. Potom je

$$S(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{t=p+1}^n [u_t(\boldsymbol{\phi}, \boldsymbol{\theta})]^2$$

a vo výpočte $\hat{\sigma}_u^2$ by sa upravil menovateľ, vzhľadom k menšiemu počtu odhadnutých reziduí. V tejto forme je odhad implementovaný najčastejšie.

V LRM s ACR uvažujeme obecnjšie aj reziduá s jednotkovým koreňom, model

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

teda vznikol z pôvodného modelu diferencovaním a má $N = n - d$ pozorovaní, $\boldsymbol{\varepsilon}$ je stacionárny a invertibilný $ARMA(p, q)$ proces. Metóda odhadu je založená na podmienenej vierohodnosti, ktorú aplikujeme na odhadnuté reziduá v LRM, ktoré sú dané metódou najmenších štvorcov. Získame odhad

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})^\top = \arg \min_{\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}} S(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta} \mid \mathbf{u}^*, \boldsymbol{\varepsilon}^*)$$

kde

$$u_t(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) = Y_t - \boldsymbol{\beta}^\top \mathbf{X}_t - \phi_1 \hat{\varepsilon}_{t-1} - \dots - \phi_p \hat{\varepsilon}_{t-p} - \theta_1 u_{t-1}(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}) - \dots - \theta_q u_{t-q}(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}).$$

Rozptyl bieleho šumu odhadneme ako

$$\hat{\sigma}_u^2 = \frac{SS_e}{N - (k + 1) - p - q} = \frac{S(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})}{n - d - (k + 1) - p - q}.$$

3. Testovanie reziduí

3.1 Testovanie autokorelovanosti

V 1. kroku algoritmu z kapitoly 2.2 sme uviedli potrebu testovať hypotézu

$$H_0 : \varepsilon \text{ je biely šum,}$$

kde $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$ sú reziduá v regresnom modeli (2.1). V tejto kapitole uvedieme niektoré štatistické testy, ktoré môžeme k tomu využiť.

Durbin-Watsonov test

O Durbin-Watsonovom teste píše Cipra (2008, str. 96). Test vychádza z predpokladu $\varepsilon \sim AR(1)$, teda po zložkách

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t, \quad t = 1, \dots, n,$$

ktorým sa autokorelovanosť reziduí modeluje najčastejšie. Parameter ρ je vzhľadom k (1.3) korelačný koeficient s meškáním 1, ε je biely šum práve keď $\rho = 0$. Hypotézu H_0 teda môžeme napísať v tvare

$$H_0 : \rho = 0.$$

Testujeme ju proti alternatíve $H_1 : \rho \neq 0$. Testová štatistika spočítaná z odhadnutých reziduí je

$$DW = \frac{\sum_{t=2}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2} \approx 2(1 - \hat{\rho}), \quad (3.1)$$

kde

$$\hat{\rho} = \frac{\sum_{t=2}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=1}^n \hat{\varepsilon}_t^2}$$

je odhad ρ podľa (1.1). Zrejme máme

- $\rho = 0$ (platí H_0): $DW \approx 2$
- $\rho = 1$ (susedné reziduá sú silno pozitívne korelované): $DW \approx 0$
- $\rho = -1$ (susedné reziduá sú silno negatívne korelované): $DW \approx 4$

Štatistika DW nemá žiadne z bežne používaných rozdelení. Jej kritické hodnoty $0 < d_L < d_U < 2$ závisia počte regresorov v modeli (2.1) a sú tabelované, napr. v štatistických tabuľkách Likeš a Laga (1978, str. 240-242) pre test proti jednostrannej alternatíve $H_1 : \rho > 0$. Tabuľka 3.1 udáva kritické hodnoty pre model s parametrami $(\beta_0, \beta_1)^\top$ a $n = 100$. Kritické hodnoty dávajú oblasti, viď Cipra (2008):

- $DW \in \mathcal{C} = (0, d_L) \cup (4 - d_L, 4) \implies$ zamietneme H_0 , reziduá sú autokorelované.

- $DW \in (d_U, 4 - d_U) \implies H_0$ nemôžeme zamietnuť.
- $DW \in (d_L, d_U) \cup (4 - d_U, 4 - d_L) \implies$ neučíňujeme žiadny záver.

Nevýhodou testu je, že detekuje len autokorelovanosť susedných reziduí, pričom v praxi sa často prejavajú ako významné až autokorelácie s väčším meškaním. Poznamenajme ešte, že pre skúmanie autokorelovanosti susedných reziduí možno ako orientačný nástroj použiť bodový graf závislosti ε_t na ε_{t-1} .

Hladina	d_L	d_U
1%	1,52	1,56
5%	1,65	1,69

Pozn: Test proti jednostrannej alternatíve $H_1 : \rho > 0$.

Tabuľka 3.1: Kritické hodnoty Durbin-Watsonovho testu.

Ljung-Boxov test

Test, ktorý detekuje aj autokorelácie s väčším meškaním, pochádza z článku Ljunga a Boxa (Ljung a Box, 1978) a je určený pôvodne na testovanie nekorelovanosti bieleho šumu po odhade modelu $ARMA(p, q)$. Test je založený na odhadoch ρ_1, \dots, ρ_m , ktoré sa spočítajú z $\hat{\varepsilon}$ podľa (1.1). Testujeme hypotézu

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0,$$

proti $H_1 :$ niektoré $\rho_i \neq 0$. Testová štatistika

$$LB = n(n+2) \sum_{i=1}^m \frac{\hat{\rho}_i^2}{n-i} \quad (3.2)$$

má za platnosti H_0 asymptoticky χ^2 rozdelenie s m stupňami voľnosti. Za platnosti H_1 sú niektoré $\rho_i \neq 0$, v prospech alternatívy teda hovoria väčšie hodnoty testovej štatistiky. Kritický obor má preto tvar

$$\mathcal{C} = (\chi_m^2(1-\alpha), \infty),$$

kde $\chi_m^2(\alpha)$ značí kvantil χ^2 rozdelenia. Voľba čísla m môže ovplyvňovať vlastnosti testu, Tsay (2002, str. 25) odporúča vziať $m = \log(n)$.

Breusch-Godfreyho test

Tento test predpokladá model $\varepsilon \sim AR(p)$, $p \geq 1$,

$$\varepsilon_t = u_t + \phi_1 \varepsilon_{t-1} + \dots + \phi_p \varepsilon_{t-p}.$$

Testujeme hypotézu

$$H_0 : \phi_1 = \phi_2 = \dots = \phi_p = 0,$$

proti alternatíve $H_1 :$ niektoré $\phi_i \neq 0$. Pre odhadnuté reziduá $\hat{\varepsilon}$ uvažujeme pomocný regresný model

$$\hat{\varepsilon}_t = \gamma_0 + \gamma_1 X_{t1} + \dots + \gamma_k X_{tk} + \phi_1 \hat{\varepsilon}_{t-1} + \dots + \phi_p \hat{\varepsilon}_{t-p} + u_t. \quad (3.3)$$

Za testovú štatistiku (podľa Cipra, 2008, str. 98) vezmeme

$$BG = (n - p)R^2 \quad (3.4)$$

s kritickým oborom $\mathcal{C} = (\chi_p^2(1 - \alpha), \infty)$. R^2 je koeficient determinácie v modeli (3.3), zavedený v kapitole 2.1. Číslo p sa volí vopred, pre mesačné alebo štvrtročné dáta Cipra (2008, str. 98) odporúča brať frekvenciu dát za rok.

3.2 Testovanie stacionarity

V chovaní stacionárnych reziduí $\varepsilon \sim ARMA(p, q)$ a nestacionárnych s autoregresným polynómom v tvare $\Phi(z) = (1 - z)^d(1 - \phi_1 z - \dots - \phi_{p-d} z^{p-d})$ sú výrazné rozdiely. Vo vyjadrení $\varepsilon_t = \sum_{i=0}^{\infty} c_i u_{t-i}$ koeficienty nekonvergujú k nule a rozptyl pozorovaní (podmieneny počiatkom radu) lineárne rastie, je teda porušený aj predpoklad homoskedasticity reziduí v LRM. Nestacionaritu spôsobenú jednotkovým koreňom vieme vyriešiť prechodom k diferenciám

$$\Delta \varepsilon = (\Delta \varepsilon_2, \dots, \Delta \varepsilon_n)^\top,$$

ktoré modelujeme ako stacionárny a invertibilný *ARMA* proces. V takomto prípade hovoríme o *ARIMA* procese. Obecné môže byť na stacionarizáciu potrebné diferencovať viac krát, hovoríme $\varepsilon \sim ARIMA(p, d, q)$, keď $\Delta^d \varepsilon \sim ARMA(p, q)$.

Obecnejšie je možné uvažovať model, v ktorom $\Phi(z) = u(z)\phi(z)$, polynóm $u(z)$ má korene na jednotkovej kružnici. Podrobne o problematike reziduí s jednotkovým koreňom, ako aj o postupe v obecnom prípade, sa píše v článku Tsay (1984). My sa však zaoberáme iba prípadom keď $u(z) = (1 - z)^d$.

Jednotkový koreň môžeme rozpoznať veľmi pomalým poklesom odhadnutej ACF. Aby sme rozlíšili hraničné prípady, budeme používať štatistický test.

Dickey-Fullerov test

Test, ktorý budeme používať na hypotézu o jednotkovom koreni, navrhli Dickey a Fuller v knihe Fuller (1976, str. 372). Test je pôvodne určený pre časové rady, my však reziduá nepozorujeme a testovú štatistiku musíme spočítať z ich odhadu metódou najmenších štvorcov. Test vychádza z modelu $\varepsilon \sim AR(1)$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t,$$

ktorého autoregresný polynóm má jediný koreň $1/\rho$. Testujeme hypotézu

$$H_0 : \rho = 1,$$

proti alternatíve $H_1 : \rho < 1$. Parameter ρ odhadneme metódou najmenších štvorcov a za štatistiku vezmeme t-štatistiku pre regresný koeficient ρ . Počítame ju z odhadnutých reziduí $\hat{\varepsilon}$ a má tvar

$$DF = \frac{\hat{\rho} - 1}{\sqrt{\frac{SS_e}{(n-2)} (\sum_{t=1}^{n-1} \hat{\varepsilon}_t^2)^{-1}}} = \frac{\sum_{t=2}^n \hat{\varepsilon}_{t-1} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})}{\hat{\sigma}_u \sqrt{\sum_{t=1}^{n-1} \hat{\varepsilon}_t^2}} \quad (3.5)$$

kde

$$\hat{\rho} = \frac{\sum_{t=2}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=1}^{n-1} \hat{\varepsilon}_t^2}, \quad SS_e = (n-2)\hat{\sigma}_u^2 = \sum_{t=2}^n (\hat{\varepsilon}_t - \hat{\rho}\hat{\varepsilon}_{t-1})^2.$$

Záporné hodnoty testovej štatistiky hovoria v neprospech hypotézy, kritický obor má tvar $(-\infty, u(\alpha))$. DF sa za platnosti hypotézy neriadi štandardným rozdelením, kvantily $u(\alpha)$ sa pre vybrané rozsahy a hladiny musia spočítať simulačne. Pre test v analýze časových rád autori uvádzajú hodnoty uvedené v tabuľke 3.2. Kvantily pre LRM s ACR s parametrami $(\beta_0, \beta_1)^\top$ sme stanovili simulačne v kapitole 4.2.

Hladina	Kvantil
1%	-2,58
5%	-1,95
10%	-1,62

Tabuľka 3.2: Kvantily Dickey-Fullerovej štatistiky pre časové rady.

4. Simulačná štúdia

Cieľom tejto kapitoly je presvedčiť sa pomocou simulovania LRM s ACR o vlastnostiach testov popísaných v kapitole 3. Budeme skúmať ich hladinu a silu za rôznych predpokladov pre ε . Všetky výpočty realizujeme pomocou programu Mathematica (verzia 11.3).

4.1 Simulovaný regresný model

Uvažujeme lineárny regresný model s jednou vysvetľujúcou premennou a $n = 500$ pozorovaniami,

$$Y_t = 2.65 + 1.59X_t + \varepsilon_t, \quad t = 1, \dots, 500. \quad (4.1)$$

Regresory generujeme ako náhodný výber z rozdelenia $N(2, 9/4)$, index t interpretujeme ako čas. Ďalej generujeme reziduá

$$\varepsilon = \{\varepsilon_1, \dots, \varepsilon_{500}\}$$

v tvare časovej rady, uvažujeme modely, ktoré vychádzajú z bieleho šumu

$$u_t \sim N(0, 1/4).$$

1. $\{\varepsilon_t\}$ je biely šum, teda $\varepsilon_t = u_t$. Značíme $\varepsilon \sim \text{BŠ}$.
2. $\{\varepsilon_t\}$ je rad kĺzavých súčtov $MA(1)$, $\varepsilon_t = \theta u_{t-1} + u_t$.
3. $\{\varepsilon_t\}$ je náhodná prechádzka $ARIMA(0, 1, 0)$,
 $\varepsilon_t = \varepsilon_{t-1} + u_t$. Značíme $\varepsilon \sim RW$.
4. $\{\varepsilon_t\}$ je autoregresný rad $AR(1)$, $\varepsilon_t = \phi \varepsilon_{t-1} + u_t$.
5. $\{\varepsilon_t\}$ je rad $ARMA(1,1)$, $\varepsilon_t = \phi \varepsilon_{t-1} + \theta u_{t-1} + u_t$.

Vektor odoziev potom dostaneme z regresného vzťahu. Vygenerujme príklad modelu (4.1) s reziduami v tvare bieleho šumu. Pozorovania

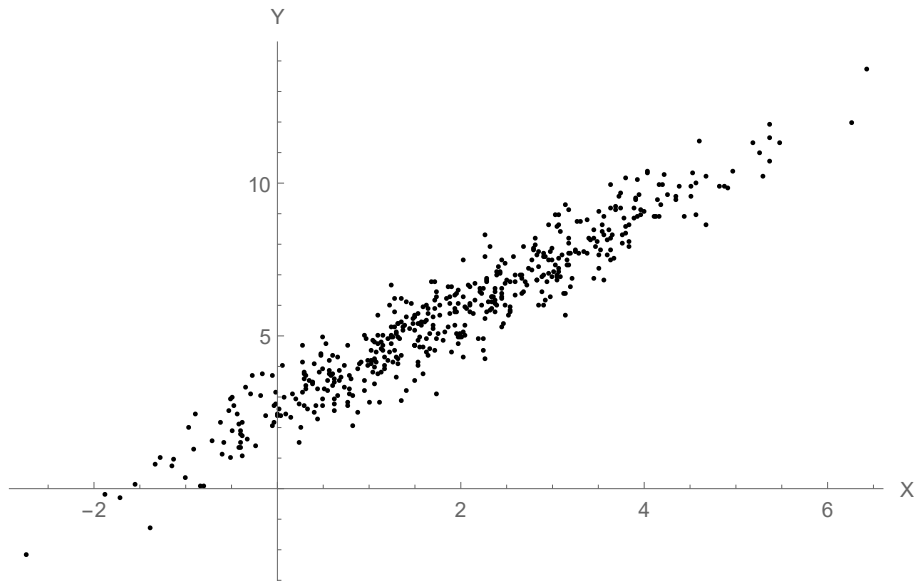
$$X_1, \dots, X_{500}, \quad Y_1, \dots, Y_{500}$$

si najlepšie prehladneme pomocou grafu dvojíc $\{(X_t, Y_t)^\top\}_{t=1}^{500}$, ktorý sa nazýva korelogram a obsahuje informáciu o regresnom vzťahu, viď obr. 4.1.

Príklad modelu (4.1) s reziduami v tvare náhodnej prechádzky je na obrázku 4.2. Regresný vzťah však vidíme, až keď vezmeme časové diferencie pozorovaní,

$$\{\Delta X_t\}_{t=1}^{500}, \quad \{\Delta Y_t\}_{t=1}^{500}.$$

ako ukazujú korelogramy 4.2 a 4.3.



Obr. 4.1: Korelogram (X_t, Y_t) , $\varepsilon \sim \text{BŠ}$.

4.2 Analýza štatistických testov

Lineárny regresný model

$$Y_t = 2.65 + 1.59X_t + \varepsilon_t, \quad t = 1, \dots, 500.$$

odhadneme metódou najmenších štvorcov a z odhadnutých reziduí $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_{500}$ spočítame hodnoty štatistík DW, LB, BG a DF . Budeme generovať 1000 nezávislých LRM a na ich základe si prezrieme empirické rozdelenie testových štatistík pri uvedených predpokladoch 1.–5. pre reziduá. Pre zamietanie hypotéz používame hladinu $\alpha = 0.05$.

Ljung-Boxova štatistika sa počíta zo vzorca (3.2). Za najväčšie meškanie zvolíme $m = \lceil \log 500 \rceil = 7$, kritická hodnota testu je potom $\chi_7^2(0.95) = 14.07$.

V Breusch-Godfrey teste volíme parameter p — rád autoregného procesu. Voľba podľa frekvencie dát za nejaké významnejšie časové obdobie stráca pri simulovaných dátach zmysel, ak však platí hypotéza, mala by byť dodržaná hladina testu nech volíme p akékoľvek. My berieme $p = 4$ a za kritickú hodnotu teda $\chi_4^2(0.95) = 9.49$. Testová štatistika sa konštruuje podľa (3.4).

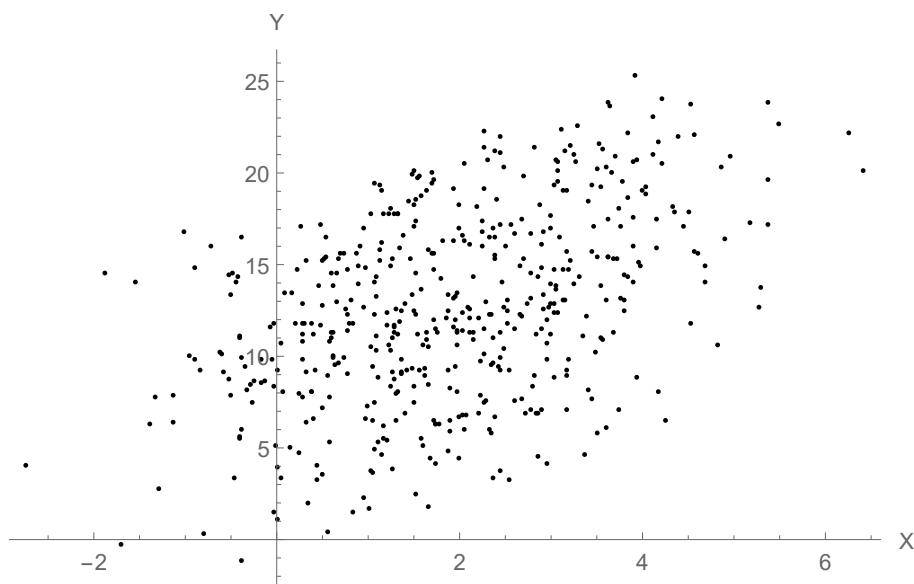
Biely šum

Reziduá generujeme ako proces bieleho šumu,

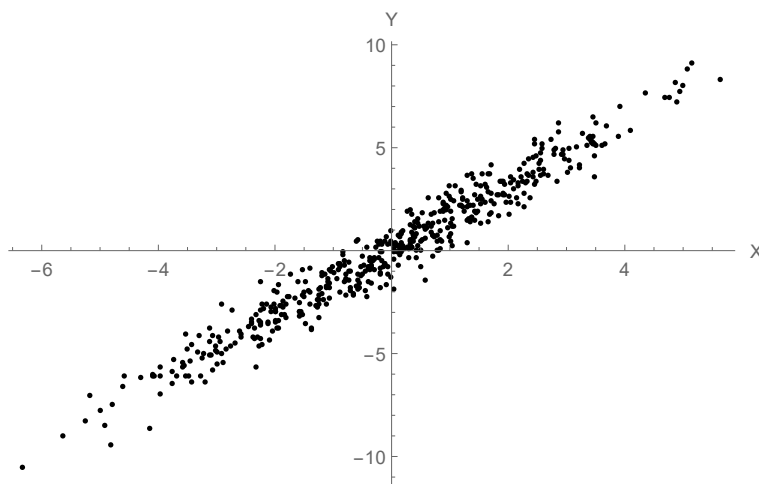
$$\varepsilon_t = u_t, \quad t = 1, \dots, 500.$$

Platí klasický LRM a hypotéza $H_0 : \varepsilon$ je biely šum, testov DW, LB a BG . Očakávame, že približne v 50 z 1000 pozorovaní týchto testov bude ležať v ich kritických oboroch.

Histogram Ljung-Boxovej štatistiky (obr. 4.4) je v súlade s jej teoretickým rozdelením. Výberový priemer je 6.87 a výberový 95% kvantil je 13.72, čo napovedá, že empirické rozdelenie je v dobrej zhode s χ_7^2 . Test zamietá hypotézu v 45 prípadoch z 1000.



Obr. 4.2: Korelogram (X_t, Y_t) , $\varepsilon \sim RW$.



Obr. 4.3: Korelogram $(\Delta X_t, \Delta Y_t)$, $\varepsilon \sim RW$.

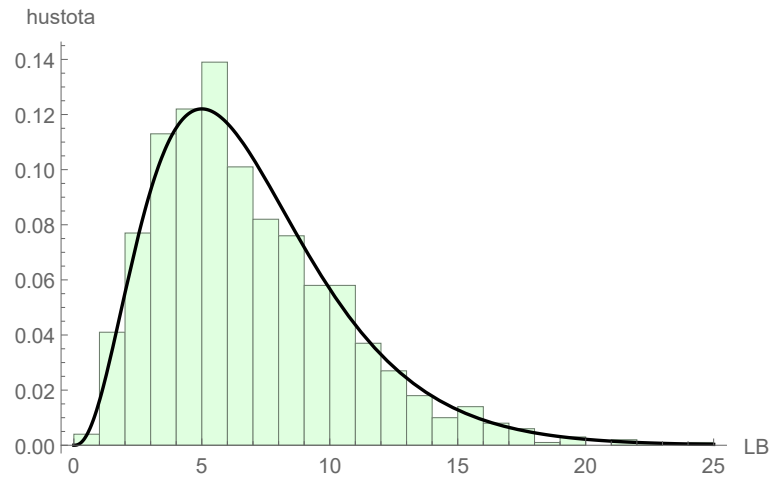
Breusch-Godfreyho štatistika má histogram spolu s hustotou χ_4^2 na obrázku 4.5. Empirické rozdelenie odpovedá skutočnému, výberový priemer je 3.98, čo opäť dobre odpovedá $E[BG] = 4$. Výberový 95% kvantil je 9.06. Štatistika prekračuje kritickú hodnotu v 42 prípadoch, čo sa na prvý pohľad pomerne líši od očakávaných 50. Môžeme použiť binomický test, ktorý hypotézu

$$P[\text{Zam.}] = 0.05$$

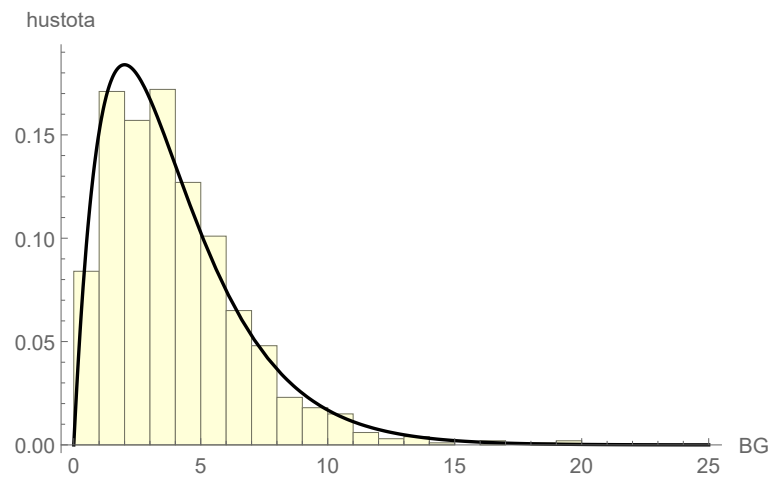
nemôže zamietnuť v prospech jednostrannej alternatívy, s p -hodnotou 0.14.

Rozdelenie Durbin-Watsonovej štatistiky nie je známe, môžeme si však prehliaadnuť jeho histogram (obr. 4.6). Test proti jednostrannej alternatíve $H_1 : \rho > 0$ dáva kritický obor $(0, d_L)$ a región nezamietania $(d_U, 4)$. Kritické hodnoty pre $n = 500$ nemáme k dispozícii. Výberový 5% kvantil je rovný 1.86.

Rozdelenie Dickey-Fullerovej štatistiky sa za platnosti alternatívy $\varepsilon \sim B\check{S}$ výrazne líši od rozdelenia za $\varepsilon \sim RW$, test zamietajú hypotézu o jednotkovom koreni



Obr. 4.4: Hustota L-B štatistiky za $\varepsilon \sim \text{BŠ}$.



Obr. 4.5: Hustota B-G štatistiky za $\varepsilon \sim \text{BŠ}$.

vždy s rezervou a sila testu proti tejto alternatíve je veľmi blízka 1. Empirické rozdelenie štatistiky je na obrázku 4.7.

Reziduá ako kĺzavé súčty

Reziduá generujeme ako proces kĺzavých súčtov $MA(1)$ s parametrom θ ,

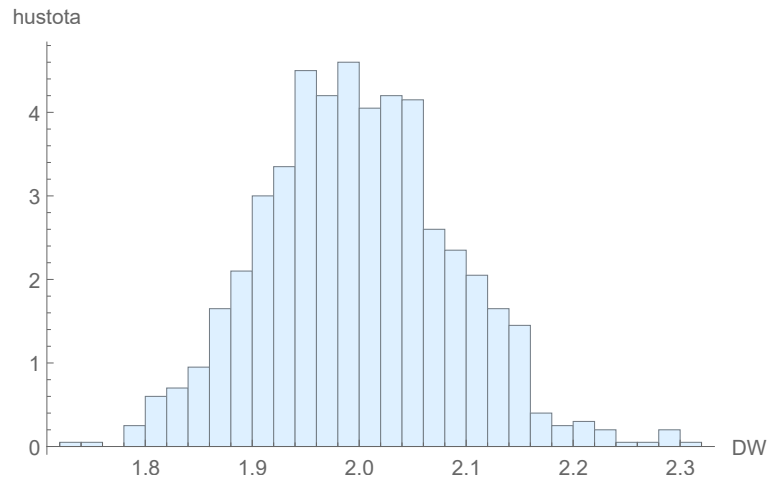
$$\varepsilon_t = \theta u_{t-1} + u_t, \quad t = 1, \dots, 500.$$

Pre $\theta \neq 0$ neplatí hypotéza bieleho šumu, ani hypotéza o jednotkovom koreni. Pre $|\theta| < 1$ je ε invertibilná postupnosť. Splnený je model pre Ljung-Boxov test. Zameriame sa na odhad sily

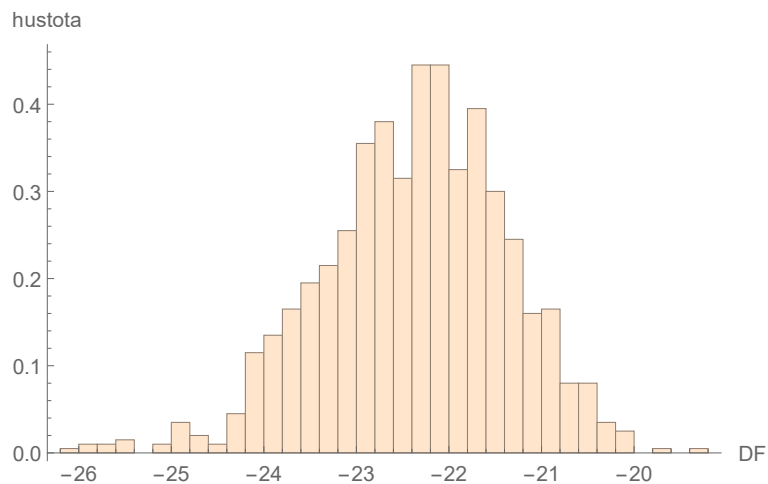
$$\beta = P[\text{Zam.}]$$

testov bieleho šumu L-B a B-G proti alternatívam blízkyh hypotéze, ktorá platí práve keď $\theta = 0$. Budeme voliť θ malé a pozorovať četnosti zamietania n_z z celkového počtu $n = 1000$, potom odhadneme silu ako

$$\hat{\beta} = \frac{n_z}{n}.$$



Obr. 4.6: Hustota D-W štatistiky za $\varepsilon \sim \text{BŠ}$.



Obr. 4.7: Hustota D-F štatistiky za alternatívy $\varepsilon \sim \text{BŠ}$.

- $\theta = 0.05$

Ljung-Boxov test zamieta hypotézu v 84 modeloch z 1000, z toho spočítame odhad sily.

$$\hat{\beta}_{LB} = 0.08$$

Breusch-Godfrey test zamieta v 99 prípadoch.

$$\hat{\beta}_{BG} = 0.1$$

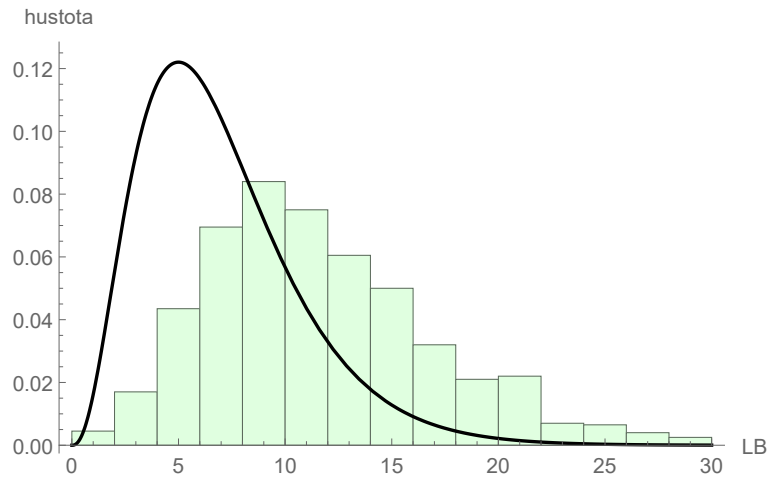
- $\theta = 0.1$

Ljung-Boxov test zamieta v 290 prípadoch a Breusch-Godfrey test v 378 prípadoch.

$$\hat{\beta}_{LB} = 0.29$$

$$\hat{\beta}_{BG} = 0.38$$

Porovnanie empirického rozdelenia Ljung-Boxovej štatistiky s hustotou χ_7^2 za platnosti platnosti hypotézy je na obrázku 4.8.



Obr. 4.8: Hustota χ_7^2 a histogram L-B za alternatívy.

- $\theta = 0.15$

Ljung-Boxov test zamietá v 654 prípadoch a Breusch-Godfrey test v 771 prípadoch.

$$\hat{\beta}_{LB} = 0.65$$

$$\hat{\beta}_{BG} = 0.77$$

Odhadnutá sila B-G testu vyšla opäť väčšia. Otestujme teda, či je rozdiel četností zamietania významný. Máme 1000 pozorovaní výsledkov testov, z ktorých si môžeme vytvoriť kontingenčnú tabuľku.

	LB Zam.	LB Nez.
BG Zam.	630	141
BG Nez.	24	205

Tabuľka 4.1: Kontingenčná tabuľka zamietania L-B a B-G testu.

Chceme otestovať hypotézu

$$H_0 : \beta_{LB} = \beta_{BG},$$

ktorá platí práve keď $P[LB \text{ Zam.} \ \& \ BG \text{ Nez.}] = P[LB \text{ Nez.} \ \& \ BG \text{ Zam.}]$, z čoho už lepšie vidíme, že hypotéza neplatí. Úloha vedie na test lineárnej kombinácie pravdepodobností multinomického vektoru, hypotézu H_0 zamietame s p-hodnotou blízko 0. Platí

$$\beta_{BG} > \beta_{LB}.$$

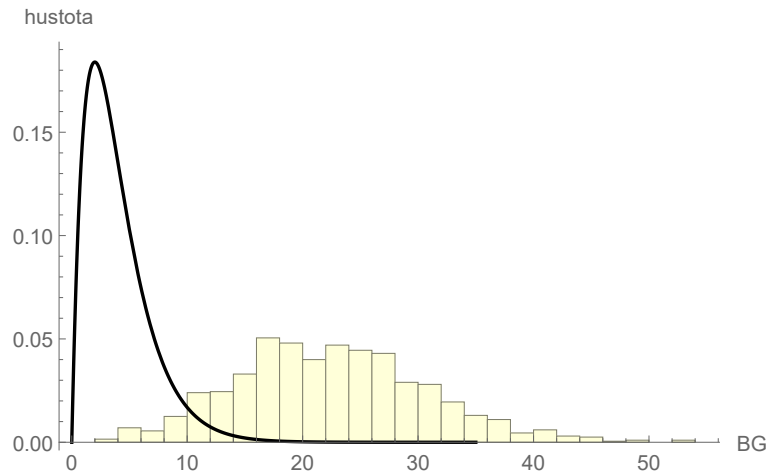
- $\theta = 0.2$

Ljung-Boxov test zamietá v 915 prípadoch a Breusch-Godfrey test v 955 prípadoch.

$$\hat{\beta}_{LB} = 0.92$$

$$\hat{\beta}_{BG} = 0.96$$

To už napovedá, že sila Breusch-Godfrey testu sa pri $\theta > 0.2$ rýchlo priblíži k 1. Histogram BG štatistiky spolu s hustotou χ_4^2 je na obrázku 4.9.



Obr. 4.9: Hustota χ_4^2 a histogram B-G štatistiky za alternatív.

Poznámka. Silu testov sme skúmali aj pre model $AR(1)$, $\varepsilon_t = \phi\varepsilon_{t-1} + u_t$. Pre rovnaké numerické hodnoty parametru ϕ vyšli sily testov podobne ako v modeli $MA(1)$. Sila sa blíži 1 pri $\phi > 0.2$ a Breusch-Godfrey test sa tiež ukázal ako lepší.

Reziduá s jednotkovým koreňom

Reziduá generujeme ako náhodnú prechádzku (nestacionárny autoregresný rad),

$$\varepsilon_t = \varepsilon_{t-1} + u_t, \quad t = 1, \dots, 500.$$

Platí hypotéza o jednotkovom koreni Dickey-Fullerovho testu. Štatistika DF sa počíta podľa vzorca (3.5), hypotézu $H_0 : \varepsilon \sim RW$ zamietame keď $DF < u(0.05)$, kde $u(\alpha)$, $\alpha \in (0, 1)$ je kvantil rozdelenia DF za platnosti hypotézy. Rozdelenie ani jeho kvantily nemáme k dispozícii, pre model s jedným regresorom

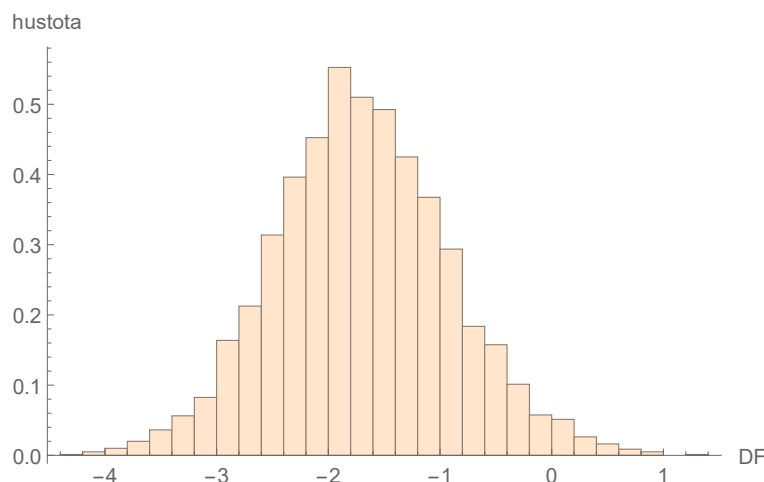
$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

stanovíme kvantil simulačne ako výberový kvantil na základe 4000 regresných modelov (4.1).

$$u(0.05) = -2.95$$

Empirické rozdelenie DF štatistiky zo 4000 pozorovaní je na obrázku 4.10.

Poznámka. Pre korektnosť dodajme, že rozdelenie DF nezávisí na parametroch β_0, β_1 a σ_u^2 . To sa nahliadne z toho, že $\hat{\varepsilon} = (\mathbb{I} - \mathbb{H}) \mathbf{Y} = (\mathbb{I} - \mathbb{H}) \varepsilon$ a ďalej keď sa vezme ε/σ_u miesto ε .



Obr. 4.10: Hustota D-F štatistiky za hypotézy $\varepsilon \sim RW$.

Autoregresné rezidá

Reziduá generujeme ako autoregresný rad $AR(1)$ s parametrom ϕ ,

$$\varepsilon_t = \phi\varepsilon_{t-1} + u_t, \quad t = 1, \dots, 500.$$

Pre $\phi \neq 0$, $|\phi| < 1$ je ε stacionárna postupnosť, neplatí hypotéza bieleho šumu ani hypotéza o jednotkovom koreni. Pre $\phi = 0$ platí hypotéza bieleho šumu, pre $\phi = 1$ platí hypotéza o jednotkovom koreni. Je splnený model pre všetky testy – DW, LB, BG aj DF. Zamerajme sa tentokrát na alternatívy blízke hypotéze o jednotkovom koreni. Budeme uvažovať ϕ blízke 1 a odhadovať silu Dickey-Fullerovho testu. Ako kritickú hodnotu budeme používať 5% kvantil -2.95 , ktorý sa získal simulačne.

- **$\phi = 0.98$**

Dickey-Fullerov test zamieta hypotézu v 344 prípadoch z 1000 keď platí alternatíva $\phi = 0.98$.

$$\hat{\beta}_{DF} = 0.34$$

- **$\phi = 0.96$**

Dickey-Fullerov test zamieta hypotézu v 866 prípadoch.

$$\hat{\beta}_{DF} = 0.87$$

Empirické rozdelenie testovej štatistiky spolu s empirickým rozdelením za platnosti $\varepsilon \sim RW$ si môžeme prezrieť na obrázku 4.11.

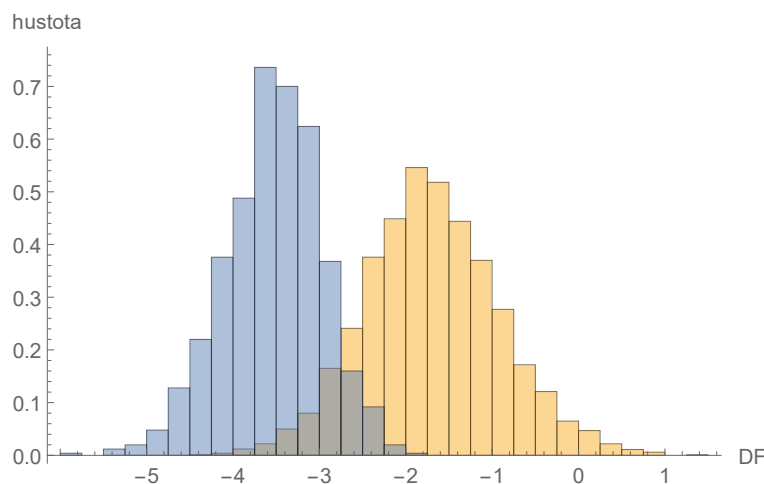
- **$\phi = 0.94$**

Dickey-Fullerov test nemôže zamietnuť hypotézu o jednotkovom koreni iba v 3 prípadoch. Odhad sily po zaokrúhlení je

$$\hat{\beta}_{DF} = 1.$$

Tým sme sa presvedčili, že sila testu sa pri $\phi < 0.94$ blíži 1.

Poznámka. Už pri $\phi = 0.99$ sa rozdelenie DF štatistiky líši tak, že test zamieta s pravdepodobnosťou okolo 0.13, čo je podstatne viac ako hladina $\alpha = 0.05$.



Obr. 4.11: Histogram D-F štatistiky za hypotézy a alternatívy.

Reziduá ARMA

Reziduá generujeme ako $ARMA(1,1)$ proces s parametrami $\rho = 0.5$ a $\theta = 0.5$,

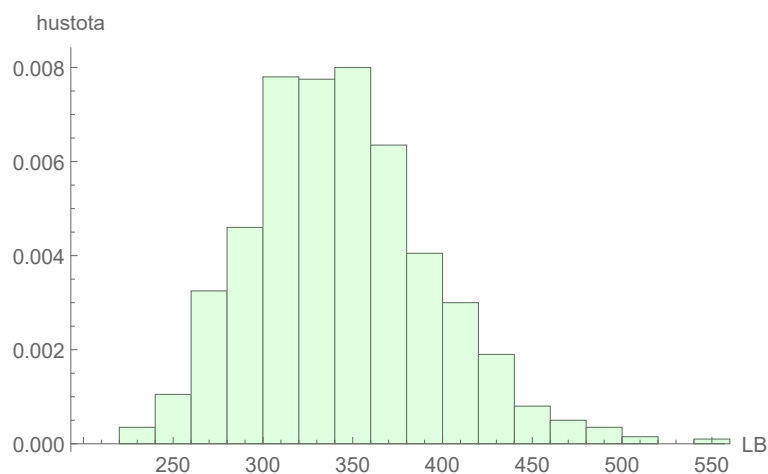
$$\varepsilon_t = 0.5\varepsilon_{t-1} + 0.5u_{t-1} + u_t, \quad t = 1, \dots, 500.$$

Neplatí hypotéza o nekorelovanosti reziduí ani hypotéza o jednotkovom koreni. Je splnený model pre Ljung-Boxov test.

Ljung-Boxov test zamietá hypotézu bieleho šumu vo všetkých 1000 prípadoch, histogram testovej štatistiky (obr. 4.12) potvrdzuje, že jej rozdelenie je hlboko v kritickom obore a pre silu testu platí

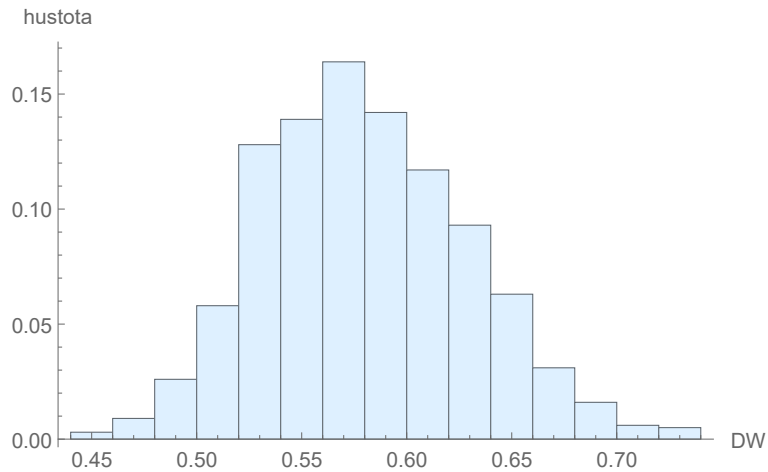
$$\beta_{LB} = P[\text{Zam.}] \approx 1.$$

Breusch-Godfreyho test má proti tejto alternatíve podobné vlastnosti ako Ljung-Boxov. Testová štatistika je veľká a sila testu je blízka 1.



Obr. 4.12: Hustota L-B štatistiky za alternatívy $\varepsilon \sim ARMA$.

Na obrázku 4.13 uvádzame aj histogram Durbin-Watsonovej štatistiky. Potvrdzuje sa, že jej rozdelenie sa líši od rozdelenia za platnosti hypotézy bieleho



Obr. 4.13: Hustota D-W štatistiky za alternatívy $\varepsilon \sim ARMA$.

šumu, všetky simulované hodnoty štatistiky ležia s rezervou pod kvantilom, ktorý sme odhadli empiricky.

Dickey-Fullerov test zamietá vo všetkých prípadoch hypotézu o jednotkovom koreni. S prihliadnutím na fakt, že výberový 95% kvantil -8.49 je s rezervou v kritickom obore $(-\infty, -2.95)$, môžeme tvrdiť

$$\beta_{DF} \approx 1.$$

Veľká sila testov je spôsobená tým, že zvolený rad je vzdialenou alternatívu hypotézy bieleho šumu aj jednotkového koreňa, ako aj dostatočným počtom pozorovaní.

5. Príklad

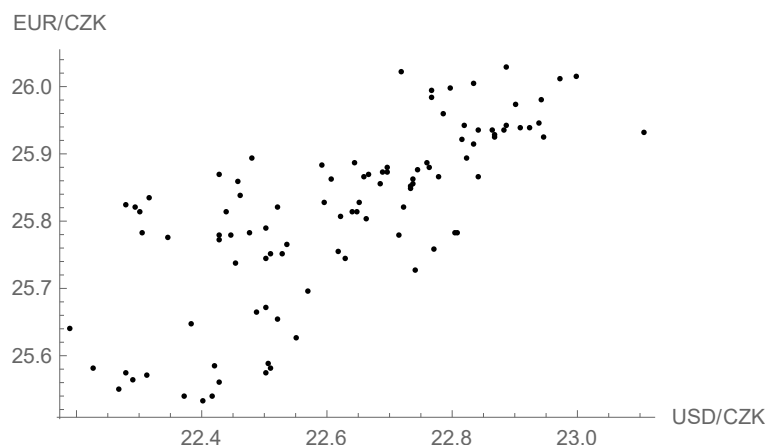
V tejto časti práce nadviažeme na kapitolu 2.2, kde sme popísali obecný postup pri analýze LRM s ACR. Aplikujeme ho na reálne finančné dáta. Spracovanie je inšpirované spracovaním príkladu v knihe Tsay (2002, str. 66–72).

Sledujeme časové rady kurzov CZK/EUR a CZK/USD za 100 obchodných dní na prelome roku 2018/2019.

t	Dátum	USD/CZK	EUR/CZK
1	8.10.2018	22,347	25,776
2	9.10.2018	22,427	25,779
3	10.10.2018	22,438	25,814
⋮	⋮	⋮	⋮
98	29.1.2019	22,529	25,752
99	30.1.2019	22,534	25,767
100	31.1.2019	22,445	25,779

Tabuľka 5.1: Ukážka časových rád kurzov.

Či sú kurzy CZK/EUR a CZK/USD korelované, sa dá posúdiť z obrázku 5.1.



Obr. 5.1: Korelogram EUR/CZK a USD/CZK.

Náš cieľ je vysvetliť kurz CZK/EUR pomocou CZK/USD. Korelogram z pozorovaní 5.1 naznačuje, že medzi kurzami môže byť dlhodobá rovnováha daná lineárnym vzťahom. Pre jednoduchosť označme

$$\begin{aligned}X_t &= \text{CZK/USD}_t, \\Y_t &= \text{CZK/EUR}_t.\end{aligned}$$

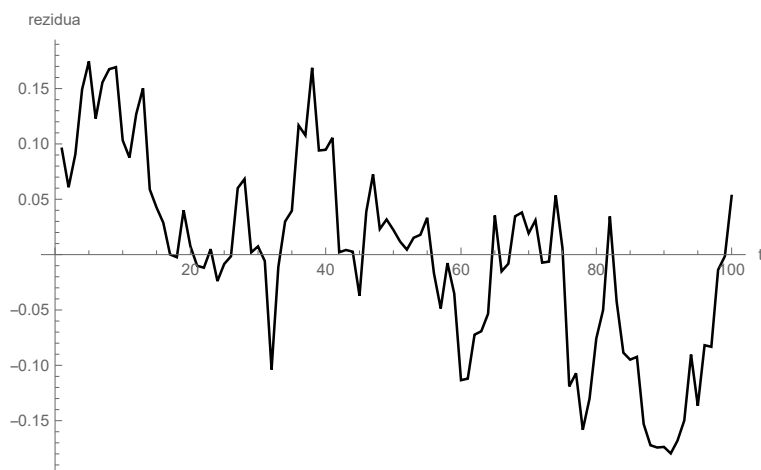
To nás vedie k tomu, že budeme modelovať vzťah medzi kurzami lineárnym regresným modelom s autokorelovanými reziduami,

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t = 1, \dots, 100.$$

Metódou najmenších štvorcov dostávame odhadnutý model

$$Y_t = 15.21 + 0.47X_t + \hat{\varepsilon}_t, \quad t = 1, \dots, 100.$$

Koeficient $\hat{\beta}_1$ potvrdzuje, že medzi X a Y je pomerne silný lineárny vzťah, avšak model zrejme nie je adekvátny ako ukazuje graf $\hat{\varepsilon}$ na obrázku 5.2, z ktorého vidieť, že reziduá ε sa dajú len ťažko považovať za nekorelované, skôr pripomínajú náhodnú prechádzku.

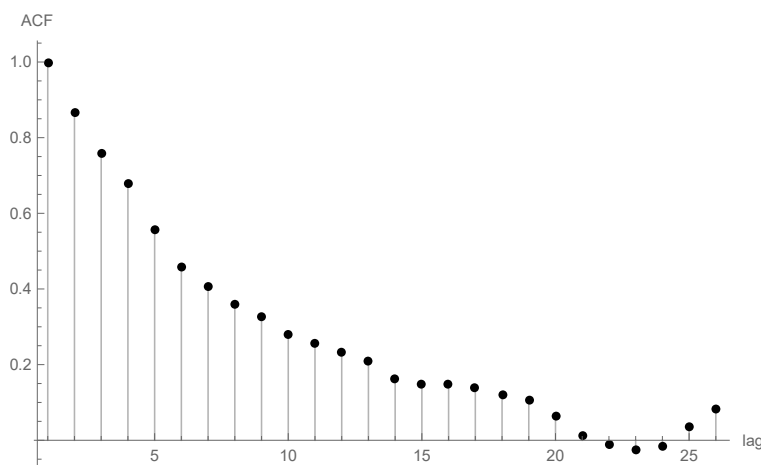


Obr. 5.2: Odhadnuté reziduá v LRM pre kurzy.

Viac o chovaní reziduí napovie odhad ich autokorelácií. Keby boli reziduá nekorelované, potom podľa (1.7) by malo pre väčšinu $\hat{\rho}_k$, $k > 0$ platiť

$$|\hat{\rho}_k| < \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\frac{1}{n}} \approx 0.2,$$

ak používame hladinu významnosti $\alpha = 0.05$.



Obr. 5.3: Odhadnutá ACF reziduí v LRM pre kurzy.

Testy bieleho šumu podľa očakávania zamietajú hypotézu s rezervou, Durbin-Watsonov test dáva

$$DW = 0.25,$$

čo naznačuje veľkú koreláciu susedných reziduí. V Ljung-Boxovom teste berieme $m = \lceil \log 100 \rceil = 4$ a v Breusch-Godfreyho teste vezmeme $p = 7$. Testové štatistiky sú veľké.

$$LB = 218.41 > \chi_4^2(0.95)$$

$$BG = 70.10 > \chi_7^2(0.95)$$

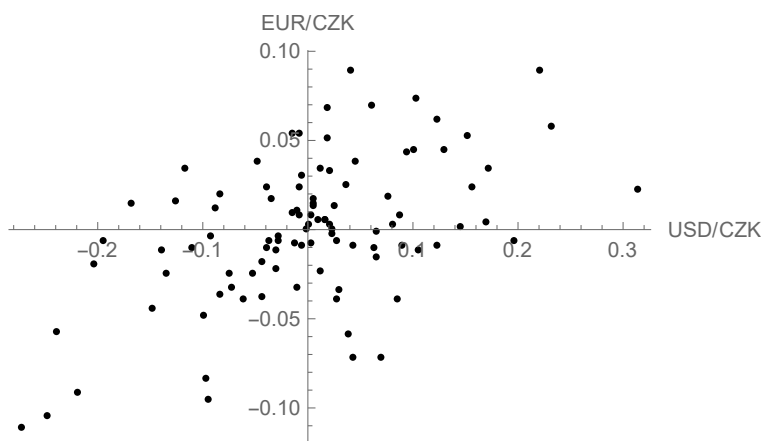
ACF každého stacionárneho *ARMA* procesu klesá exponenciálne k nule, my však pozorujeme veľmi pozvoľný pokles, čo napovedá, že autoregresný polynóm reziduí má koreň blízky 1. Predvedme teda test na jednotkový koreň. Hodnota Dickey-Fullerovej štatistiky je

$$DF = -2.68 \notin \mathcal{C} = (-\infty, -2.95).$$

Dickey-Fullerov test nemôže zamietnuť hypotézu o jednotkovom koreni. Reziduá nemožno považovať za stacionárne a identifikovať pre ne nejaký stacionárny model. Vezmeme teda časové diferencie a dostávame LRM s ACR v tvare

$$\Delta Y_t = \gamma_0 + \gamma_1 \Delta X_t + \Delta \varepsilon_t, \quad t = 2, \dots, 100.$$

Po diferencovaní je regresný vzťah zreteľnejší, korelogram 5.4 podstatne viac pripomína obrázok 4.1.



Obr. 5.4: Korelogram $\Delta \text{EUR}/\text{CZK}$ a $\Delta \text{USD}/\text{CZK}$.

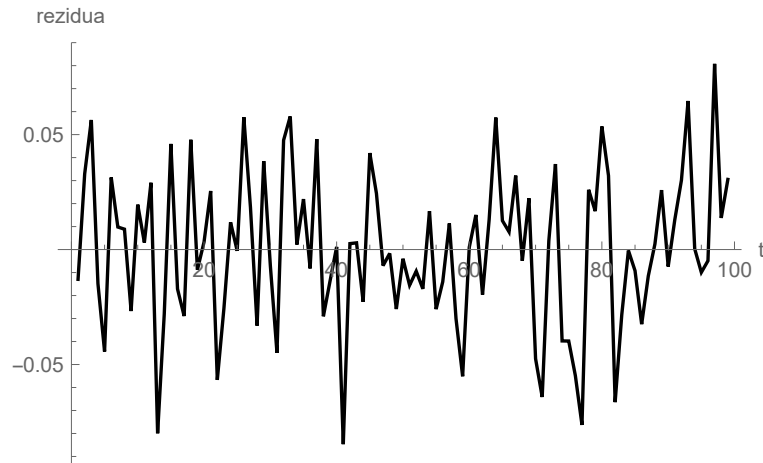
Metódou najmenších štvorcov dostávame odhad modelu

$$\Delta Y_t = -0.0002 + 0.198 \Delta X_t + \widehat{\Delta \varepsilon}_t, \quad t = 2, \dots, 100.$$

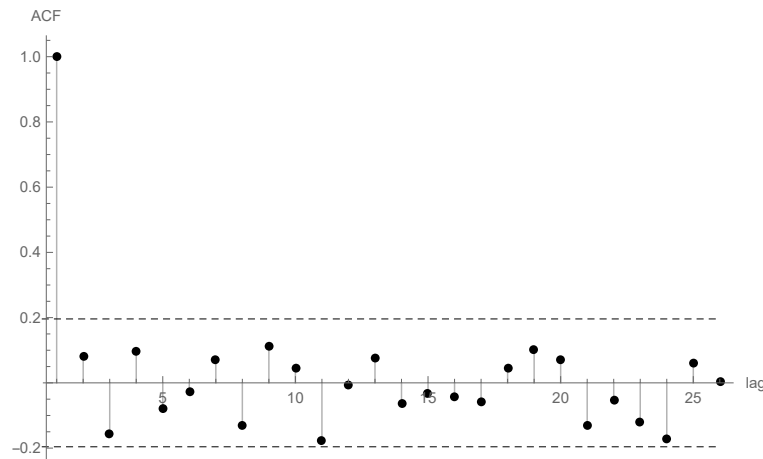
Na obrázku 5.5 si môžeme prezrieť odhadnuté reziduá $\Delta \varepsilon$. Podľa grafu ACF (obr. 5.6) hneď zamietame hypotézu o jednotkovom koreni a dokonca nemôžeme zamietnuť hypotézu o nekorelovanosti reziduí, identifikujeme pre ne $\Delta \varepsilon \sim \text{BŠ}$. Všetky $\hat{\rho}_k$, $k > 0$ sa ukázali ako nevýznamné. K rovnakému záveru dospejeme prehliadkou PACF.

Otestujme teda hypotézu $H_0 : \Delta \varepsilon \sim \text{BŠ}$ testami z kapitoly 3. Spočítame Durbin-Watsonovu štatistiku, jej kritické hodnoty sú uvedené v tabuľke 3.1. Dostávame

$$DW = 1.82 > d_U = 1.69,$$



Obr. 5.5: Odhadnuté reziduá po diferencovaní.



Obr. 5.6: Odhadnutá ACF reziduú po diferencovaní.

Durbin-Watsonov test teda nemôže zamietnuť hypotézu H_0 . Spočítame Ljung-Boxovu a Breusch-Godfreyho štatistiku.

$$LB = 4.90 < \chi_4^2(0.95) = 9.49$$

$$BG = 7.68 < \chi_7^2(0.95) = 14.07$$

Všetky testy hypotézy bieleho šumu už nemôžu zamietnuť H_0 . Prijímame teda klasický LRM,

$$\Delta Y_t = \gamma_0 + \gamma_1 \Delta X_t + u_t,$$

$\{u_t\} = \{\Delta \varepsilon_t\}$ sú nekorelované, v ktorom je odhad metódou najmenších štvorcov

$$\Delta Y_t = -0.0002 + 0.198 \Delta X_t + \hat{u}_t$$

najlepší nestranný odhad parametrov. Koeficient $\hat{\gamma}_0$ je malý. Ak je model platný, musí platiť $H_0 : \gamma_0 = 0$. Hodnota t -štatistiky pre H_0 je

$$T = -0.047 \notin \mathcal{C},$$

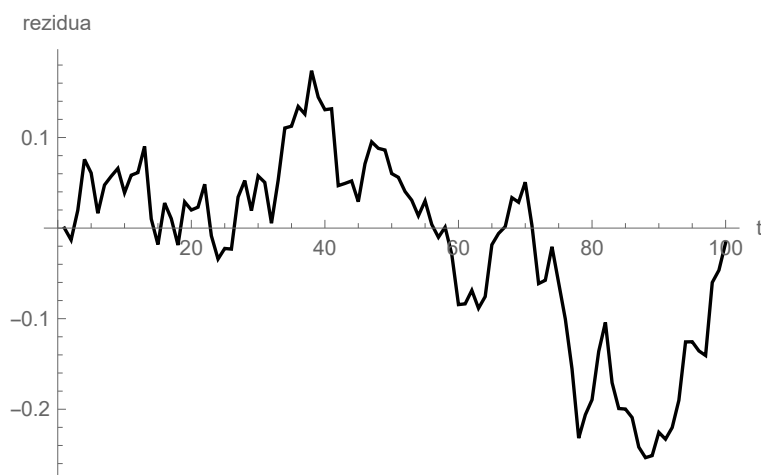
hypotézu zamietame vtedy keď $|T| > 1.98$, podľa príslušných kvantilov rozdelenia t_{97} . Konečne dostávame regresný vzťah pre Y_t :

$$Y_t = Y_{t-1} + 0.198 (X_t - X_{t-1}) + \hat{u}_t.$$

Odhadnutý rozptyl reziduí $\{u_t\}$ je

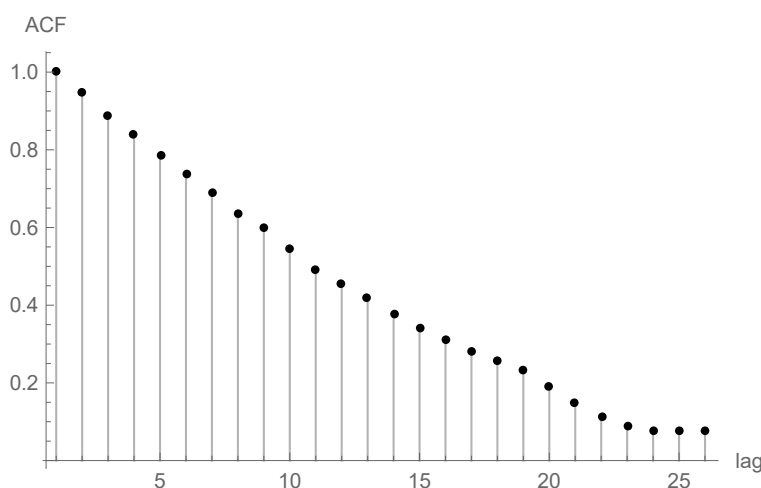
$$\hat{\sigma}_u^2 = \frac{\hat{\mathbf{u}}^\top \hat{\mathbf{u}}}{99 - 2} = 0.0035.$$

Všimnime si, že za platnosti modelu je $\beta_1 = \gamma_1$. Odhad $\hat{\beta}_1$ sa však výrazne líši od $\hat{\gamma}_1$, ktorý je konzistentný. To napovedá, že pre $\varepsilon \sim RW$ nemusí metóda najmenších štvorcov dávať konzistentné odhady parametrov $(\beta_0, \beta_1)^\top$. Vezmime teda v pôvodnom modeli $\hat{\beta}_1 = \hat{\gamma}_1$ a dostaneme realistickejší odhad reziduí ε . Intercept $\hat{\beta}_0$ zvolme napr. tak aby rad začínal v 0.



Obr. 5.7: Odhadnuté reziduá po konzistentnom odhade parametrov.

Graf nových $\hat{\varepsilon}$ zreteľnejšie pripomína proces $ARIMA(0,1,0)$, ktorý sme pre ε identifikovali. Podobne aj ACF na obrázku 5.8 viac podporuje túto hypotézu.



Obr. 5.8: ACF reziduí po konzistentnom odhade modelu.

Pre zaujímavosť uvedme ešte hodnotu Dickey-Fullerovej štatistiky.

$$DF = -1.60$$

Záver

V práci sme rozobrali niekoľko problematík, spojených s LRM s ACR, aby sme mohli formulovať obecný postup pre jeho analyzovanie. Predstavovaný model tak môžeme použiť, kedykoľvek máme daný regresný model s časovými pozorovaniami a pristupovať k jeho analýze algoritmicky.

Popísali sme testy hypotézy bieleho šumu a test jednotkového koreňa. V simulačnej štúdií sme odhadli hladinu testov a porovnali ju so skutočnou hladinou. Užitočné bolo odhadnúť silu testov proti vybraným alternatívam v okolí hypotézy. Pri skúmaní testov bieleho šumu sme dospeli k záveru

$$\beta_{BG} > \beta_{LB},$$

pri reziduách $MA(1)$ a $AR(1)$. Stanovili sme kritický obor pre Dickey-Fullerov test ako

$$\mathcal{C} = (-\infty, -2.95).$$

Príklad z kapitoly 5 umožnil ilustrovať všetky kroky algoritmu okrem identifikácie $ARMA$ modelu pomocou AIC a odhadu regresného modelu podmienenou vierohodnosťou. Ukázalo sa, aké užitočné je uvažovať aj reziduá s jednotkovým koreňom $ARIMA$ a neobmedziť sa len na stacionárne $ARMA$ rady. Vysvetlili sme $Y_t = CZK/EUR_t$ pomocou $X_t = CZK/USD_t$ ako

$$Y_t = Y_{t-1} + 0.198 (X_t - X_{t-1}) + \hat{u}_t.$$

Zistili sme, že pri existencii jednotkového koreňa dáva metóda najmenších štvorcov nekonzistentné odhady parametrov a následne aj odhady reziduí, ktoré nezodpovedajú skutočnosti. To zdôrazňuje užitočnosť kritického oboru \mathcal{C} pri zamietaní hypotézy o jednotkovom koreni reziduí. Ukázalo sa to ako dôležité práve v skúmanom modeli.

Zoznam použitej literatúry

- ANDĚL, J. (2007). *Základy matematické statistiky*. Druhé opravené vydání. Matfyzpress, Praha. ISBN 80-7378-001-1.
- BOX, G. E. P., JENKINS, G. M. a REINSEL, G. C. (1994). *Time Series Analysis*. Prentice-Hall, New Jersey. ISBN 0-13-060774-6.
- CIPRA, T. (2008). *Finanční ekonometrie*. Ekopress, Praha. ISBN 978-80-86929-43-9.
- FULLER, W. (1976). *Introduction to Statistical Time Series*. Wiley, New York. ISBN 0-471-28715-6.
- LIKEŠ, J. a LAGA, J. (1978). *Základní statistické tabulky*. SNTL, Praha.
- LJUNG, G. M. a BOX, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65**(2), 297–303.
- PRÁŠKOVÁ, Z. (2007). *Základy náhodných procesů II*. Karolinum, Praha. ISBN 978-80-246-0971-3.
- TSAY, R. S. (1984). Regression models with time series errors. *Journal of the American Statistical Association*, **79**(385), 118–124.
- TSAY, R. S. (2002). *Analysis of Financial Time Series*. Wiley, New York. ISBN 0-471-41544-8.
- WEI, W. W. S. (1990). *Time Series Analysis*. Addison-Wesley, New York. ISBN 0-201-15911-2.

Zoznam obrázkov

4.1	Korelogram (X_t, Y_t) , $\varepsilon \sim \text{BŠ}$	19
4.2	Korelogram (X_t, Y_t) , $\varepsilon \sim RW$	20
4.3	Korelogram $(\Delta X_t, \Delta Y_t)$, $\varepsilon \sim RW$	20
4.4	Hustota L-B štatistiky za $\varepsilon \sim \text{BŠ}$	21
4.5	Hustota B-G štatistiky za $\varepsilon \sim \text{BŠ}$	21
4.6	Hustota D-W štatistiky za $\varepsilon \sim \text{BŠ}$	22
4.7	Hustota D-F štatistiky za alternatívy $\varepsilon \sim \text{BŠ}$	22
4.8	Hustota χ_7^2 a histogram L-B za alternatívy.	23
4.9	Hustota χ_4^2 a histogram B-G štatistiky za alternatívy.	24
4.10	Hustota D-F štatistiky za hypotézy $\varepsilon \sim RW$	25
4.11	Histogram D-F štatistiky za hypotézy a alternatívy.	26
4.12	Hustota L-B štatistiky za alternatívy $\varepsilon \sim ARMA$	26
4.13	Hustota D-W štatistiky za alternatívy $\varepsilon \sim ARMA$	27
5.1	Korelogram EUR/CZK a USD/CZK.	28
5.2	Odhadnuté reziduá v LRM pre kurzy.	29
5.3	Odhadnutá ACF reziduí v LRM pre kurzy.	29
5.4	Korelogram $\Delta \text{EUR/CZK}$ a $\Delta \text{USD/CZK}$	30
5.5	Odhadnuté reziduá po diferencovaní.	31
5.6	Odhadnutá ACF reziduí po diferencovaní.	31
5.7	Odhadnuté reziduá po konzistentnom odhade parametrov.	32
5.8	ACF reziduí po konzistentnom odhade modelu.	32

Zoznam tabuliek

3.1	Kritické hodnoty Durbin-Watsonovho testu.	15
3.2	Kvantily Dickey-Fullerovej štatistiky pre časové rady.	17
4.1	Kontingenčná tabuľka zamietania L-B a B-G testu.	23
5.1	Ukážka časových rád kurzov.	28