



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Bc. Miroslava Gažová

**Vážené portmanteau testy pro analýzu
časových řad**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Šárka Hudecová, Ph.D.

Studijní program: Pravděpodobnost, matematická
statistika a ekonometrie

Studijní obor: Pravděpodobnost, matematická
statistika a ekonometrie

Praha 2024

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

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 má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

Touto cestou by som sa chcela poďakovať hlavne RNDr. Šárke Hudecovej, Ph.D., za vedenie, cenné rady a podporu pri písaní práce. Ďakujem za priateľský prístup a ochotu kedykoľvek pomôcť. Ďalej by som chcela poďakovať Dominikovi za ochotu a nekonečnú trpezlivosť. V neposlednej rade ďakujem svojim rodičom za podporu pri celom štúdiu.

Název práce: Vážené portmanteau testy pro analýzu časových řad

Autor: Miroslava Gažová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Šárka Hudecová, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Táto diplomová práca predstavuje vážené a nevážené portmanteau testy, ktoré sa používajú na testovanie adekvátnosti odhadnutého modelu ARMA. Modely ARMA sa hojne využívajú pri analýze časových rád. Čitateľ je oboznámený s váženými portmanteau testami s váhami založenými na jadrových funkciách a s geometricky klesajúcimi váhami. Ďalej detailne rozoberáme asymptotické rozdelenie testových štatistík. Na záver uvádzame rozsiahlu simulačnú štúdiu v programovacom jazyku R, ktorá skúma hladinu a silu uvažovaných testov. Cieľom simulačnej štúdie je najmä porovnať vážené varianty portmanteau testov s Boxovým-Piercovým a Ljungovým-Boxovým testom.

Klíčová slova: časové rady, modely ARMA, Boxov-Piercov test, Ljung-Boxov test, vážené portmanteau testy

Title: Weighted portmanteau tests for time series analysis

Author: Miroslava Gažová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Šárka Hudecová, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This thesis introduces weighted and unweighted portmanteau tests, which are used for testing goodness-of-fit of a fitted ARMA model. ARMA models are widely used for time series analysis. The reader is presented with weighted portmanteau tests with weights based on kernel functions and with geometrically decaying weights. Next, we deal with the asymptotic distribution of test statistics in greater detail. Lastly, we present an extensive simulation study in the R programming language, which computes the statistical size and power of the considered tests. The main goal of the simulation study is a comparison of weighted variants of portmanteau tests with Box-Pierce and Ljung-Box tests.

Keywords: time series, ARMA models, Box-Pierce test, Ljung-box test, weighted portmanteau tests

Obsah

Úvod	2
1 Základné pojmy	4
1.1 Časová rada	4
1.2 Autokovariančná a autokorelačná funkcia	4
1.3 ARMA proces	6
2 Základné teoretické výsledky	9
2.1 Odhad modelu ARMA(p, q)	9
2.1.1 Odhad modelu AR(1)	9
2.1.2 Odhad modelu ARMA(p, q)	11
2.2 Asymptotické rozdelenie výberových autokorelácií	12
2.3 Autokorelogram	15
3 Nevážené Q testy	17
3.1 Boxov-Piercov a Ljung-Boxov test	17
3.2 Asymptotické rozdelenie testových štatistík	18
3.2.1 Skutočné hodnoty parametrov	18
3.2.2 Odhadnuté hodnoty parametrov	20
4 Vážené Q testy	27
4.1 Asymptotické rozdelenie váženej štatistiky	28
4.2 Jadrové váhy	30
4.3 Geometricky klesajúce váhy	32
4.3.1 Aproximácia rozdelenia	33
5 Simulačné štúdie	35
5.1 Hladina testu	35
5.1.1 Hladina testov pre rôzne hodnoty n	36
5.1.2 Hladina testov pre rôzne hodnoty k	37
5.1.3 Hladina testov podľa korekcie	38
5.2 Sila testu	38
5.2.1 Sila testov podľa voľby ARMA parametrov	39
5.2.2 Sila testov pre rôzne hodnoty n	41
5.2.3 Sila testov pre rôzne hodnoty k	42
5.2.4 Sila testov podľa korekcie	42
5.2.5 Zhrnutie výsledkov	42
Záver	46
Zoznam použitej literatúry	47
Zoznam obrázkov	48
Zoznam tabuliek	50
A Appendix	51

Úvod

S časovými radami sa stretávame v mnohých oblastiach, ako sú napríklad ekonómia, poisťovníctvo a bankovníctvo alebo medicínsky výskum. Popis časových rád pomocou matematických modelov umožňuje získať mnoho cenných informácií, ako napríklad predikciu vývoja danej časovej rady v budúcnosti. V tejto práci sa budeme zaoberať modelmi typu ARMA, ktoré patria medzi modely Boxovej-Jenkinsonovej metodológie. Tieto modely popísali autori George Box a Gwilym Jenkins v 70. rokoch minulého storočia a dodnes sa tešia veľkej obľube pri analýze časových rád.

V našej práci sa budeme zaoberať štatistickými testami, pomocou ktorých je možné testovať adekvátnosť odhadnutého modelu ARMA pre danú časovú radu. Ako prvé uvedieme Boxov-Piercov a Ljung-Boxov test, ktoré sa typicky používajú v tejto situácii. Na základe článku Gallagher a Fisher (2015) sa budeme ďalej zaoberať akýmsi zobecnením týchto testov, a to tzv. *váženými portmanteau testami*. Testové štatistiky týchto testov sú taktiež založené na súčte druhých mocnín výberových autokorelácií, no s rastúcim oneskorením sa výberovým autokoreláciám pridáva iná (zpravidla menšia) váha. Práca je rozdelená na teoretickú a praktickú časť. Teoretická časť sa zaoberá najmä definíciou testovanej hypotézy, popisom testových štatistík a ich asymptotickým rozdelením. Praktická časť obsahuje simulačnú štúdiu, ktorá skúma vlastnosti jednotlivých testov na simulovaných dátach a porovnáva testy medzi sebou. Cieľom tejto práce je popísať nevážené portmanteau testy spolu s ich váženými variantami, určiť asymptotické rozdelenie testových štatistík a pomocou simulačných štúdií vyhodnotiť, či by niektoré varianty vážených testov mohli predstavovať vhodnú alternatívu k známemu Boxovmu-Piercovmu a Ljung-Boxovmu testu.

Práca je rozdelená do piatich kapitol. Prvá kapitola obsahuje základné definície, ktoré ďalej v práci používame. V druhej kapitole uvádzame základné teoretické výsledky, akými sú odhad parametrov ARMA modelu alebo určenie asymptotického rozdelenia výberových autokorelácií.

Tretia kapitola sa zaoberá neváženými portmanteau testami, a to konkrétne Boxovým-Piercovým a Ljung-Boxovým testom. Popisujeme asymptotické rozdelenie testových štatistík, ktoré je rozdelené na dva prípady. Hlavným výsledkom tejto kapitoly je veta 5, pomocou ktorej je možné určiť kritický obor oboch testov. Túto vetu názorne ilustrujeme na niekoľkých naväzujúcich príkladoch.

V štvrtej kapitole hovoríme o vážených portmanteau testoch, ktoré predstavujú hlavné zameranie tejto práce. Definujeme obecnú testovú štatistiku a s ňou dva typy váh, s ktorými sa v práci zaoberáme: váhy založené na jadrovej funkcii a geometricky klesajúce váhy. Uvádzame niekoľko typov jadrových funkcií, ktorých vhodnosť budeme testovať v simulačných štúdiách. Nakoniec uvádzame niekoľko možností, ako určiť asymptotické rozdelenie váženej testovej štatistiky.

Piata kapitola obsahuje komplexnú simulačnú štúdiu, ktorá skúma najmä hla-

dinu a silu testov v závislosti na rôznych parametroch, akými sú napríklad dĺžka simulovanej rady alebo voľba maximálneho oneskorenia pre výberové autokorelácie, z ktorých sú spočítané jednotlivé testové štatistiky. Výsledky simulácií sú popísané pomocou tabuliek a názorných grafov. Všetky simulácie boli spracované pomocou štatistického softvéru R (R Core Team (2021)).

Vlastným prínosom tejto práce je doplnenie teoretickej časti o mnohé príklady a obrázky, najmä príklady ilustrujúce vetu 5. Ďalej sa jedná o dôkaz lemmatu 2, doplnenie niektorých krokov v dôkaze lemmatu 3 a dôkaz vety 4. Vlastným prínosom je tiež presná formulácia vety 7 a následne jej dôkaz. Najmä v kapitole 4 sme zjednotili značenie a doplnili niektoré chýbajúce predpoklady príslušných viet. Typy váh pre vážené portmanteau testy sme čerpali z článku Gallagher a Fisher (2015), no pre váhy založené na jadrovej funkcii sme okrem Daniellovej jadrovej funkcie uvažovali aj iné jadrové funkcie, viz tabuľka 4.1. Práca je zakončená simulačnou štúdiou, ktorá skúma a porovnáva jednotlivé testy pre rôzne modelové situácie.

1. Základné pojmy

V prvej časti práce uvedieme základné definície pojmov, ktoré budeme ďalej používať. Definície boli čerpané z literatúry (Cipra, 2003; Brockwell a Davis, 1991).

1.1 Časová rada

Definícia 1 (časová rada). *Nech (Ω, \mathcal{F}, P) je pravdepodobnostný priestor, $T \subset \mathbb{R}$ je indexová množina. Ako náhodný proces označíme množinu náhodných veličín $\{Y_t, t \in T\}$ na rovnakom pravdepodobnostnom priestore (Ω, \mathcal{F}, P) . Časovou radou označíme náhodný proces $\{Y_t, t \in T\}$, kde $T \subset \mathbb{Z}$.*

Hodnoty $t \in T$ interpretujeme ako čas. V praktických aplikáciách môžeme pozorovať iba *trajektóriu*, resp. *realizáciu* náhodného procesu. Časovú radu $\{Y_t, t \in \mathbb{Z}\}$ budeme označovať aj ako $\{Y_t\}$. V tejto práci budeme uvažovať iba procesy tvorené reálnymi náhodnými veličinami, teda *reálne* náhodné procesy.

Definícia 2 (slabá stacionarita). *Časová rada $\{Y_t, t \in T\}$ s konečnými druhými momentami je slabo stacionárna, ak pre všetky $s, t \in T$ a pre ľubovoľné $h \in \mathbb{Z}$ také, že $s + h, t + h \in T$, platí:*

$$\begin{aligned}\mathbb{E}(Y_t) &= \mathbb{E}(Y_{t+h}) = \mu, \quad \mu \in \mathbb{R}, \\ \text{cov}(Y_s, Y_t) &= \mathbb{E}(Y_s - \mu)(Y_t - \mu) = \text{cov}(Y_{s+h}, Y_{t+h}).\end{aligned}$$

Z definície slabej stacionarity vyplýva, že aj rozptyl rady $\text{var}(Y_t) = \sigma_Y^2$ je konštantný. Stacionarita časovej rady vyjadruje, že priebeh rady je stabilný (teda stochasticky ustálený). Ak je rada slabo stacionárna, tak príslušné prvé dva momenty sú invariantné v čase. Implicitne predpokladáme, že ak je časová rada stacionárna, tak má konečné druhé momenty. Ďalej v práci budeme slabú stacionaritu označovať iba ako *stacionaritu*. Pod pojmom *stacionárna časová rada* budeme teda rozumieť stacionárnu radu z definície 2.

1.2 Autokovariančná a autokorelačná funkcia

Jedna z dôležitých vlastností časovej rady je korelovanosť jednotlivých členov. Túto vlastnosť zachytáva *autokovariančná funkcia*. V práci sa budeme ďalej zaoberať iba stacionárnymi časovými radami, preto budeme ako autokovariančnú funkciu označovať funkciu jednej premennej z definície 4. *Autokorelačná funkcia* predstavuje znormovanú autokovariančnú funkciu tak, aby jej hodnoty spadali do intervalu $[-1, 1]$.

Definícia 3 (autokovariančná funkcia). *Nech $\{Y_t, t \in T\}$ je náhodný proces s konečnými druhými momentami a strednou hodnotou $\mathbb{E}Y_t = \mu_t$. Autokovariančnú funkciu náhodného procesu definujeme ako funkciu dvoch premenných na $T \times T$ predpisom*

$$R(s, t) = \text{cov}(Y_s, Y_t) = \mathbb{E}(Y_s - \mu_s)(Y_t - \mu_t), \quad s, t \in T.$$

Rozptyl procesu v čase t je potom $R(t, t)$.

V nasledujúcich dvoch definíciách uvedieme autokovariančnú a autokorelačnú funkciu stacionárnej časovej rady. Keďže sa jedná o definície pre časovú radu, tak budeme uvažovať $t \in \mathbb{Z}$. Pre stacionárnu časovú radu platí, že jej autokovariančná funkcia je funkciou jednej premennej.

Definícia 4 (autokovariančná funkcia stacionárnej časovej rady). *Uvažujme stacionárnu časovú radu $\{Y_t, t \in \mathbb{Z}\}$ so strednou hodnotou $\mathbb{E}(Y_t) = \mu$. Autokovariančná funkcia tejto časovej rady pre oneskorenie k je definovaná ako*

$$\gamma_k = \text{cov}(Y_t, Y_{t-k}) = \mathbb{E}(Y_t - \mu)(Y_{t-k} - \mu), \quad k \in \mathbb{Z}.$$

Môžeme si všimnúť, že autokovariančná funkcia pre oneskorenie $k = 0$ je rovná rozptylu danej časovej rady, teda $\gamma_0 = \sigma_Y^2$. Ak je $\{Y_t\}$ stacionárna časová rada s autokovariančnou funkciou γ_k , tak variančná matica vektora (Y_1, \dots, Y_n) je rovná

$$\mathbf{\Gamma}_n = [\gamma_{|i-j|}]_{i,j=1}^n = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_0 \end{pmatrix}.$$

Definícia 5 (autokorelačná funkcia). *Autokorelačná funkcia (ACF) stacionárnej časovej rady $\{Y_t, t \in \mathbb{Z}\}$ s kladným rozptylom je definovaná pre oneskorenie k ako*

$$\rho_k = \frac{\text{cov}(Y_t, Y_{t-k})}{\sqrt{\text{var}(Y_t)}\sqrt{\text{var}(Y_{t-k})}} = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_Y^2}, \quad k \in \mathbb{Z}. \quad (1.1)$$

Autokovariančná aj autokorelačná funkcia sú symetrické, teda pre všetky oneskorenia $k \in \mathbb{Z}$ platí $\gamma_k = \gamma_{-k}$ a $\rho_k = \rho_{-k}$. Autokovariančná, resp. autokorelačná funkcia poskytuje dôležité informácie o závislosti jednotlivých náhodných veličín, ktoré tvoria časovú radu. V praxi máme iba konečné množstvo pozorovaní, preto potrebujeme obe funkcie odhadnúť. Budeme predpokladať, že máme k dispozícii pozorovania $Y_1, \dots, Y_n, n \in \mathbb{N}$, z časovej rady $\{Y_t\}$, z ktorých budeme odhad zostrojovať.

Definícia 6 (odhad autokovariančnej a autokorelačnej funkcie). *Odhad autokovariančnej funkcie stacionárnej časovej rady pre oneskorenie k (tj. odhad γ_k) z pozorovaní $Y_1, \dots, Y_n, n \in \mathbb{N}$, definujeme ako*

$$c_k = \frac{1}{n} \sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y}), \quad k = 0, 1, \dots, n-1,$$

kde $\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$ je odhad strednej hodnoty. *Odhad autokorelačnej funkcie pre oneskorenie k (tj. odhad ρ_k) definujeme ako*

$$r_k = \frac{c_k}{c_0}, \quad k = 0, 1, \dots, n-1. \quad (1.2)$$

Odhad autokovariančnej funkcie c_k budeme ďalej označovať ako *výberovú autokovarianciu* a odhad autokorelačnej funkcie r_k budeme označovať ako *výberovú autokoreláciu*.

Poznámka. Podľa (Box a kol., 2016, str. 31) je na výpočet odhadu autokovariančnej funkcie potrebných aspoň 50 pozorovaní. Výpočet odhadu autokovariančnej funkcie pre oneskorenie k do rádu $n - 1$ je teoreticky možné, no v praxi autori odporúčajú uvažovať hodnoty k do rádu $\frac{n}{4}$. Ak odhadujeme autokovariančnú funkciu pre menšie oneskorenia k , tak máme dostupných viac dvojíc pozorovaní, z ktorých môžeme odhad spočítať.

1.3 ARMA proces

Medzi základné metódy analýzy časových rád patria modely Boxovej-Jenkinsonovej metodológie, ktoré prvýkrát popísali autori George Box a Gwilym Jenkins. Tieto modely sú vhodné najmä na popis stacionárnych časových rád. Majú široké uplatnenie pri modelovaní a predikcii časových rád v rôznych oblastiach, ako sú ekonómia, epidemiológia alebo poisťovníctvo. V tejto časti sme čerpali najmä z literatúry (Cipra, 2003).

Definícia 7 (biely šum). *Postupnosť náhodných veličín $\{\varepsilon_t\}$ nazveme bielym šumom, ak sú jednotlivé náhodné veličiny nekorelované, majú nulovú strednú hodnotu a konštantný kladný rozptyl $\sigma^2 > 0$, teda:*

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

Zapisujeme aj ako

$$\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2).$$

Ak sú náhodné veličiny ε_t navyše nezávislé a rovnako rozdelené, tak zavedieme značenie

$$\{\varepsilon_t\} \sim \text{IID}(0, \sigma^2).$$

Poznámka. Biely šum (anglicky white noise) má konštantnú strednú hodnotu a rozptyl, ktoré sa nemenia v čase. Ide teda o stacionárnu časovú radu.

V tejto sekcii uvedieme definíciu zmiešaného procesu ARMA, ktorým sa budeme naďalej zaoberať v celej práci. Tento proces predstavuje dôležitý a často používaný nástroj na modelovanie časových rád. ARMA proces vznikol spojením autoregresného procesu (AR – autoregressive process) a procesu kľzavých súčtov (MA – moving averages). Nakoniec uvedieme definíciu kauzálneho a invertibilného ARMA procesu.

Definícia 8 (proces ARMA(p, q)). *Časová rada $\{Y_t, t \in T\}$ so strednou hodnotou $\mathbb{E}Y_t = 0$ sa riadi zmiešaným procesom ARMA rádu p a q , $p \geq 0, q \geq 0$, ak pre každé $t \in T$ platí*

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + v_1 \varepsilon_{t-1} + \dots + v_q \varepsilon_{t-q}, \quad (1.3)$$

kde $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$, $\varphi_1, \dots, \varphi_p, v_1, \dots, v_q \in \mathbb{R}$, $\varphi_p \neq 0, v_q \neq 0$.

Definícia 9 (kauzálny proces ARMA(p, q)). *Nech $\{Y_t, t \in T\}$ je časová rada riadiaca sa procesom ARMA(p, q) s bielym šumom $\{\varepsilon_t\}$. Ak existuje postupnosť konštant $\{c_j, j \in \mathbb{N}_0\}$ taká, že $\sum_{j=0}^{\infty} |c_j| < \infty$ a ak pre $\forall t \in T$ platí*

$$Y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j},$$

tak je $\{Y_t, t \in T\}$ kauzálny proces.

Definícia 10 (invertibilný proces ARMA(p, q)). *Nech $\{Y_t, t \in T\}$ je časová rada riadiaca sa procesom ARMA(p, q) s bielym šumom $\{\varepsilon_t\}$. Ak existuje postupnosť konštánt $\{d_j, j \in \mathbb{N}_0\}$ taká, že $\sum_{j=0}^{\infty} |d_j| < \infty$ a ak pre $\forall t \in T$ platí*

$$\varepsilon_t = \sum_{j=0}^{\infty} c_j Y_{t-j},$$

tak je $\{Y_t, t \in T\}$ invertibilný proces.

Poznámka. ARMA proces môžeme zjednodušene zapísať pomocou operátora posunutia B , ktorý definujeme nasledovne

$$\begin{aligned} BY_t &= Y_{t-1}, \\ B^k Y_t &= B^{k-1}(BY_t) = B^{k-1}Y_{t-1} = \dots = Y_{t-k}, \quad k \geq 1. \end{aligned}$$

ARMA proces z definície 8 môžeme teda zapísať ako

$$\varphi(B)Y_t = v(B)\varepsilon_t,$$

kde autoregresný operátor má tvar

$$\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p \tag{1.4}$$

a operátor kľzavých súčtov je v tvare

$$v(B) = 1 + v_1 B + \dots + v_q B^q. \tag{1.5}$$

Poznámka. Ak korene polynómu

$$\varphi(z) = 1 - \varphi_1 z - \dots - \varphi_p z^p$$

ležia zvonku jednotkového kruhu v \mathbb{C} , tak tvorí časová rada riadiaca sa modelom ARMA(p, q) kauzálny proces, z čoho plynie aj jej stacionarita.

Poznámka. Pre prípad nenulovej strednej hodnoty $\mathbb{E}Y_t = \mu$ môžeme ARMA model zapísať ako

$$Y_t - \mu = \varphi_1(Y_{t-1} - \mu) + \dots + \varphi_p(Y_{t-p} - \mu) + \varepsilon_t + v_1 \varepsilon_{t-1} + \dots + v_q \varepsilon_{t-q},$$

teda

$$\begin{aligned} Y_t &= \alpha + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + v_1 \varepsilon_{t-1} + \dots + v_q \varepsilon_{t-q}, \\ \alpha &= (1 - \varphi_1 - \dots - \varphi_p)\mu. \end{aligned}$$

Posunutím získame časovú radu $\{Y_t - \mu\}$ s nulovou strednou hodnotou, a teda operátorovo môžeme zapísať túto časovú radu ako

$$\varphi(B)(Y_t - \mu) = v(B)\varepsilon_t, \quad \text{resp.} \quad \varphi(B)Y_t = \alpha + v(B)\varepsilon_t.$$

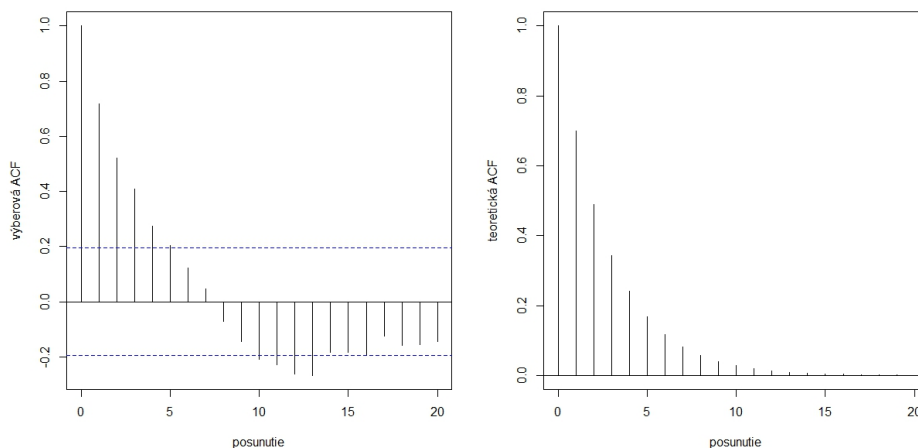
Príklad 1. Uvažujme časovú radu $\{Y_t\}$ riadiacu sa modelom AR(1) v tvare

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t, \quad |\varphi_1| < 1, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2). \tag{1.6}$$

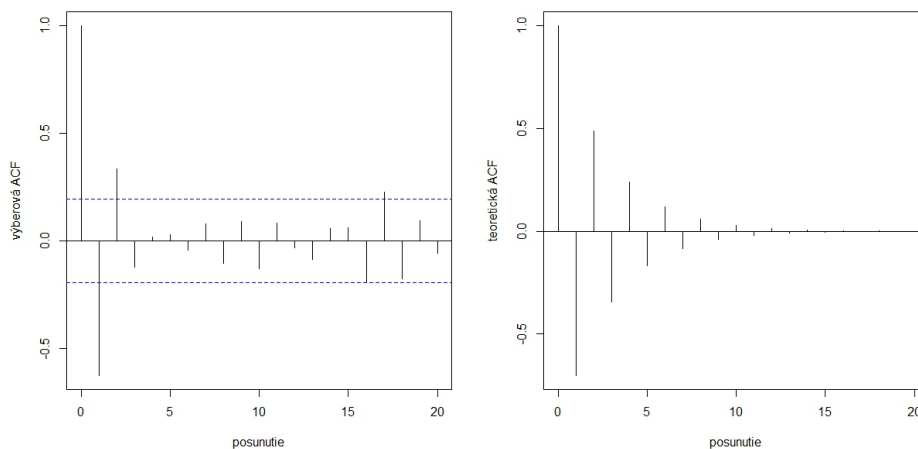
Pozrieme sa, ako vyzerá výberová autokorelačná funkcia tejto časovej rady pre voľbu parametrov $\varphi_1 = 0,7$ a $\varphi_1 = -0,7$ a porovnáme ju s teoretickou autokorelačnou funkciou ρ_k . Tvar ρ_k pre časovú radu splňujúcu (1.6) je

$$\rho_k = (\varphi_1)^k.$$

Presný výpočet môžeme nájsť napríklad v (Prášková a Lachout, 2001). Absolútne hodnoty autokorelačnej funkcie rady s predpisom (1.6) teda klesajú exponenciálne podľa koeficientu φ_1 . Na obrázku 2 vidíme výberovú a teoretickú autokorelačnú funkciu pre časovú radu (1.6) s hodnotou parametru $\varphi_1 = 0,7$. Obrázok 3 zobrazuje analogickú situáciu pre prípad $\varphi_1 = -0,7$. Výberové autokorelačné funkcie r_k sú odhady teoretickej autokorelačnej funkcie γ_k a z obrázkov môžeme vidieť, že nie sú úplne zhodné.



Obrázek 1.1: Výberová a teoretická autokorelačná funkcia časovej rady z modelu AR(1) a parametrom $\varphi_1 = 0,7$.



Obrázek 1.2: Výberová a teoretická autokorelačná funkcia časovej rady z modelu AR(1) a parametrom $\varphi_1 = -0,7$.

2. Základné teoretické výsledky

2.1 Odhad modelu ARMA(p, q)

Dôležitou súčasťou modelovania časových rád pomocou modelu ARMA(p, q) je odhad parametrov $\varphi_1, \dots, \varphi_p, \nu_1, \dots, \nu_q$. V tejto sekcii sa budeme zaoberať odhadom neznámych parametrov pomocou metódy maximálnej vierohodnosti. Táto metóda predpokladá znalosť rozdelenia náhodných veličín, z ktorých odhad zostrojujeme. Budeme teda predpokladať, že $\{\varepsilon_t\}$ tvorí postupnosť nezávislých náhodných veličín s rozdelením $N(0, \sigma^2)$, kde σ^2 je neznáme. Z toho budeme môcť odvodiť podmienené rozdelenie náhodných veličín danej časovej rady a zostrojiť maximálne vierohodný odhad (MLE). V tejto sekcii sme čerpali z literatúry (Brockwell a Davis, 1991, str. 256).

2.1.1 Odhad modelu AR(1)

Ako prvý ukážeme maximálne vierohodný odhad parametrov modelu AR(1), ktorý neskôr zobecníme pre odhad modelu ARMA(p, q). V tejto sekcii budeme používať označenie σ -algebry \mathcal{F}_{t-1} , ktorú definujeme ako

$$\mathcal{F}_{t-1} = \sigma\{Y_{t-1}, Y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}.$$

Máme k dispozícii dáta $\{Y_1, \dots, Y_n\}$, $n \in \mathbb{N}$, ktoré pochádzajú zo stacionárnej časovej rady $\{Y_t\}$ s nulovou strednou hodnotou, ktorá sa riadi AR(1) modelom

$$Y_t = \varphi Y_{t-1} + \varepsilon_t, \quad |\varphi| < 1. \quad (2.1)$$

Ďalej predpokladajme, že ε_t sú nezávislé náhodné veličiny s rozdelením $N(0, \sigma^2)$. Vieme, že ε_t je nezávislé s \mathcal{F}_{t-1} pre $\forall t$, pretože vďaka predpokladu $|\varphi| < 1$ môžeme časovú radu $\{Y_t\}$ vyjadriť ako kauzálny lineárny proces. Označme vektor neznámych parametrov ako $\boldsymbol{\theta} = (\varphi, \sigma^2)^\top$. Na základe predpokladu $\varepsilon_t \sim N(0, \sigma^2)$ dostávame

$$Y_t - \varphi Y_{t-1} \sim N(0, \sigma^2), \quad t = 2, \dots, n,$$

Podmienená hustota Y_t pri podmieňovaní celou σ -algebrou \mathcal{F}_{t-1} je ekvivalentná podmienenej hustote Y_t podmienenej znalosťou Y_{t-1} . Z toho dostaneme podmienené rozdelenie $Y_t | \mathcal{F}_{t-1}$ ako

$$Y_t | \mathcal{F}_{t-1} \sim N(\varphi Y_{t-1}, \sigma^2), \quad t = 2, \dots, n.$$

Na zostrojenie maximálne vierohodného odhadu potrebujeme poznať združené rozdelenie veličín Y_1, \dots, Y_n . Jeden z možných prístupov je vyjadrenie združenej hustoty pomocou podmienených hustôt. Z vyjadrenia (2.1) vieme, že

$$f_{Y_i | Y_{i-1}, \dots, Y_1}(y_i | y_{i-1}, \dots, y_1; \boldsymbol{\theta}) = f_{Y_i | Y_{i-1}}(y_i | y_{i-1}; \boldsymbol{\theta}). \quad (2.2)$$

Podmienená hustota $f_{Y_i | Y_{i-1}}(y_i | y_{i-1}; \boldsymbol{\theta})$ má explicitný tvar

$$f_{Y_i | Y_{i-1}}(y_i | y_{i-1}; \boldsymbol{\theta}) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(y_i - \varphi y_{i-1})^2\right). \quad (2.3)$$

Združená hustota má podľa vety o postupnom podmieňovaní tvar

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n; \boldsymbol{\theta}) = f_{Y_1}(y_1; \boldsymbol{\theta}) \prod_{i=2}^n f_{Y_i|Y_{i-1}}(y_i|y_{i-1}; \boldsymbol{\theta}). \quad (2.4)$$

Podmienená hustota veličín Y_1, \dots, Y_n pri danej počiatkovej hodnote $Y_1 = y_1$ má potom tvar

$$f_{Y_2, \dots, Y_n|Y_1}(y_2, \dots, y_n|y_1; \boldsymbol{\theta}) = \prod_{t=2}^n f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \boldsymbol{\theta}).$$

Z (2.2) a (2.3) dostávame vyjadrenie podmienenej vierohodnostnej funkcie ako

$$\mathcal{L}_n(\boldsymbol{\theta}) = \prod_{t=2}^n (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(Y_t - \varphi Y_{t-1})^2\right). \quad (2.5)$$

Maximálne vierohodný odhad vektora parametrov $\boldsymbol{\theta}$ maximalizuje (2.5). Ekvivalentne môžeme maximalizovať logaritmickú vierohodnosť, ktorá má tvar

$$\ell_n(\boldsymbol{\theta}) = \log(\mathcal{L}_n(\boldsymbol{\theta})) = \sum_{t=2}^n \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y_t - \varphi Y_{t-1})^2\right). \quad (2.6)$$

Maximálne vierohodné odhady (podmienené) dostaneme teda ako

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \ell_n(\boldsymbol{\theta}).$$

Poznámka. Podmienovanie vo výraze (2.4) (a teda prechod k podmienenému maximálne vierohodnému odhadu) nemá asymptoticky vplyv na výsledky a tento postup sa používa štandardne z toho dôvodu, že výsledný tvar logaritmickkej vierohodnosti uľahčí výpočet výsledného odhadu.

Poznámka. Maximálne vierohodný odhad modelu AR(1) môžeme explicitne vyjadriť. Uvažujme logaritmickú vierohodnosť (2.6). Maximálne vierohodný odhad parametrov φ a σ^2 dostaneme nasledovne. Ako prvé zderivujeme logaritmickú vierohodnosť podľa φ a σ^2 budeme považovať za konštantu.

$$\begin{aligned} \frac{\delta \ell_n(\boldsymbol{\theta})}{\delta \varphi} &= \sum_{t=2}^n -\frac{1}{2\sigma^2} \cdot 2(Y_t - \varphi Y_{t-1})(-Y_{t-1}) \\ &= \frac{1}{\sigma^2} \sum_{t=2}^n (Y_t - \varphi Y_{t-1})(Y_{t-1}) \\ &= \frac{1}{\sigma^2} \sum_{t=2}^n Y_t Y_{t-1} - \frac{\varphi}{\sigma^2} \sum_{t=2}^n Y_{t-1}^2. \end{aligned}$$

Ak položíme tento výraz rovno nule, tak získame maximálne vierohodný odhad

$$\hat{\varphi}_{MLE} = \frac{\sum_{t=2}^n Y_t Y_{t-1}}{\sum_{t=2}^n Y_{t-1}^2}.$$

Analogicky zderivovaním logaritmickkej vierohodnosti podľa parametru σ^2 dostávame

$$\begin{aligned} \frac{\delta \ell_n(\boldsymbol{\theta})}{\delta \sigma^2} &= \sum_{t=2}^n -\frac{1}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (Y_t - \varphi Y_{t-1})^2 \\ &= -\frac{n-1}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{t=2}^n (Y_t - \varphi Y_{t-1})^2. \end{aligned}$$

Položením tohoto výrazu rovno nule a úpravou získame maximálne vierohodný odhad

$$\hat{\sigma}_{MLE}^2 = \frac{\sum_{t=2}^n (Y_t - \hat{\varphi}_{MLE} Y_{t-1})^2}{n-1}.$$

Z výrazu (2.6) môžeme vidieť, že maximalizácia logaritmickej vierohodnosti podľa φ je ekvivalentná minimalizácii súčtu štvorcov

$$\sum_{t=2}^n (Y_t - \varphi Y_{t-1})^2.$$

2.1.2 Odhad modelu ARMA(p, q)

Maximálne vierohodný odhad parametrov modelu ARMA(p, q) ukážeme pomocou modifikácie postupu pre odhad parametrov AR(1). Máme opäť k dispozícii dáta $\{Y_1, \dots, Y_n\}$, $n \in \mathbb{N}$, pochádzajúce zo stacionárnej časovej rady s nulovou strednou hodnotou, ktorá sa riadi modelom ARMA(p, q).

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + v_1 \varepsilon_{t-1} + \dots + v_q \varepsilon_{t-q}, \quad (2.7)$$

kde ε_t sú nezávislé náhodné veličiny s rozdelením $N(0, \sigma^2)$ a ε_t je nezávislé na \mathcal{F}_{t-1} pre $\forall t$. Chceme odhadnúť parametre $\varphi_1, \dots, \varphi_p, v_1, \dots, v_q$ a neznámy rozptyl σ^2 , ktoré usporiadame do vektora

$$\boldsymbol{\psi} = (\varphi_1, \dots, \varphi_p, v_1, \dots, v_q, \sigma^2)^\top.$$

Z vyjadrenia (2.7) môžeme odvodiť podmienené rozdelenie $Y_t | \mathcal{F}_{t-1}$ ako

$$Y_t | \mathcal{F}_{t-1} \sim N\left(\sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{j=1}^q v_j \varepsilon_{t-j}, \sigma^2\right), \quad t = p + q + 1, \dots, n.$$

Pre odhad modelu AR(1) sme predpokladali znalosť jednej počiatkovej hodnoty $Y_1 = y_1$. Pre odhad modelu ARMA(p, q) budeme analogicky predpokladať znalosť $p + q$ počiatkových hodnôt veličín $Y_1, \dots, Y_p, \varepsilon_1, \dots, \varepsilon_q$. Vierohodnostná funkcia má v tomto prípade tvar

$$\mathcal{L}_n(\boldsymbol{\psi}) = \prod_{t=p+q}^n (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2} \left(Y_t - \sum_{i=1}^p \varphi_i Y_{t-i} - \sum_{j=1}^q v_j \varepsilon_{t-j}\right)^2\right).$$

Podmienené maximálne vierohodné odhady parametrov $\varphi_1, \dots, \varphi_p, v_1, \dots, v_q$ dostaneme ekvivalentne maximalizáciou logaritmickej vierohodnosti, ktorá má tvar

$$\ell_n(\boldsymbol{\psi}) = \sum_{t=p+q}^n \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left(Y_t - \sum_{i=1}^p \varphi_i Y_{t-i} - \sum_{j=1}^q v_j \varepsilon_{t-j}\right)^2\right).$$

Hľadané odhady môžeme vyjadriť ako

$$\hat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi}}{\operatorname{argmax}} \ell_n(\boldsymbol{\psi}).$$

Poznámka. Pri odhadovaní modelov AR(1) a ARMA(p, q) sme pre jednoduchosť predpokladali, že dáta pochádzajú z časovej rady s nulovou strednou hodnotou. Postup môžeme zobecniť aj pre prípad s nenulovou strednou hodnotou.

Majme teda dáta $\{Y_1, \dots, Y_n\}$, $n \in \mathbb{N}$, pochádzajúce zo stacionárnej časovej rady s neznámou strednou hodnotou μ , riadiacej sa modelom ARMA(p, q). Podľa druhej poznámky za definíciou 8 máme

$$Y_t = \alpha + \sum_{i=1}^p \varphi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^q v_j \varepsilon_{t-j}, \quad \alpha = (1 - \varphi_1 - \dots - \varphi_p)\mu,$$

kde ε_t sú nezávislé náhodné veličiny s rozdelením $N(0, \sigma^2)$ a ε_t je nezávislé na \mathcal{F}_{t-1} pre $\forall t$. Vektor neznámych parametrov bude v tomto prípade

$$\boldsymbol{\psi}_\mu = (\alpha, \varphi_1, \dots, \varphi_p, v_1, \dots, v_q, \sigma^2)^\top.$$

Analogicky môžeme odvodiť

$$Y_t | \mathcal{F}_{t-1} \sim N\left(\alpha + \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{j=1}^q v_j \varepsilon_{t-j}, \sigma^2\right), \quad t = p + q + 1, \dots, n.$$

Maximálne vierohodný odhad $\boldsymbol{\psi}_\mu$ dostaneme maximalizáciou logaritmickej vierohodnosti v tvare

$$\ell_n(\boldsymbol{\psi}_\mu) = \sum_{t=p+q}^n \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left(Y_t - \alpha - \sum_{i=1}^p \varphi_i Y_{t-i} - \sum_{j=1}^q v_j \varepsilon_{t-j} \right)^2 \right),$$

čo môžeme zjednodušene vyjadriť ako

$$\hat{\boldsymbol{\psi}}_\mu = \operatorname{argmax}_{\boldsymbol{\psi}_\mu} \ell_n(\boldsymbol{\psi}_\mu).$$

2.2 Asymptotické rozdelenie výberových autokorelácií

V tejto časti sa budeme zaoberať výberovými autokoreláciami, ktoré zohrávajú kľúčovú úlohu pri definícii testových štatistík, ktorými sa budeme zaoberať v nasledujúcich kapitolách. Hlavný výsledok tejto časti je veta 1, ktorá sa zaoberá asymptotickým rozdelením výberových autokorelácií. Túto vetu ďalej aplikujeme na výberové autokorelácie bieleho šumu v lemmatu 3. V tejto časti sme čerpali z knihy (Brockwell a Davis, 1991) a z článku (Box a Pierce, 1970). Medzi vlastné prínosy tejto sekcie patrí dôkaz lemmatu 2 a tiež podrobné rozpísanie niektorých krokov v dôkaze lemmatu 3.

Začneme vetou 1, ktorá popisuje asymptotické rozdelenie výberových autokorelácií. Opäť budeme predpokladať, že máme k dispozícii n pozorovaní časovej rady Y_1, \dots, Y_n .

Veta 1. *Nech $\{Y_t\}$ je stacionárna časová rada, pre ktorú platí*

$$Y_t - \mu = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim \text{IID}(0, \sigma^2).$$

Ďalej nech platí aspoň jeden z predpokladov

1. $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, $\mathbb{E}Z_t^4 < \infty$,
2. $\sum_{j=-\infty}^{\infty} \psi_j |j| < \infty$.

Nech ρ_k označuje autokorelačnú funkciu rady $\{Y_t\}$ pre oneskorenie k z definície 5 a r_k označuje výberovú autokoreláciu tejto rady z definície 6. Potom pre každé $k \in \mathbb{N}$ platí

$$\sqrt{n} \left[\begin{pmatrix} r_1 \\ \vdots \\ r_k \end{pmatrix} - \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_k \end{pmatrix} \right] \xrightarrow[n \rightarrow \infty]{D} N_k(\mathbf{0}, \mathbf{W}),$$

kde $\mathbf{0}$ je nulový vektor dimenzie k a \mathbf{W} je variančná matica typu $k \times k$, ktorej (i, j) -ty prvok je daný ako

$$w_{i,j} = \sum_{m=-\infty}^{\infty} \{ \rho(m+i)\rho(m+j) + \rho(m-i)\rho(m+j) + 2\rho(i)\rho(j)\rho^2(m) - 2\rho(i)\rho(m)\rho(m+j) - 2\rho(j)\rho(m)\rho(m+i) \}. \quad (2.8)$$

Dôkaz. Viz (Brockwell a Davis, 1991, kapitola 7.3, str. 225–226). □

V literatúre často nájdeme označenie pre 2.8 ako Bartletov vzorec. Vyjadrenie hodnôt $w_{i,j}$ je možné značne zjednodušiť, ako uvádzame v nasledujúcom lemmatu.

Lemma 2. *Prvky variančnej matice \mathbf{W} z vety 1 môžeme zjednodušené zapísať nasledujúcim spôsobom*

$$w_{i,j} = \sum_{m=1}^{\infty} \{ \rho(m+i) + \rho(m-i) - 2\rho(i)\rho(m) \} \times \{ \rho(m+j) + \rho(m-j) - 2\rho(j)\rho(m) \}. \quad (2.9)$$

Dôkaz. Princíp dôkazu ukážeme na nasledujúcom príklade

$$\begin{aligned} \sum_{m=-\infty}^{\infty} \rho(m+i) &= \rho(i) + \sum_{m=-\infty}^{-1} \rho(m+i) + \sum_{m=1}^{\infty} \rho(m+i) \\ &= \rho(i) + \sum_{m=1}^{\infty} \rho(-m+i) + \sum_{m=1}^{\infty} \rho(m+i) \\ &= \rho(i) + \sum_{m=1}^{\infty} \rho(m-i) + \sum_{m=1}^{\infty} \rho(m+i), \end{aligned}$$

pričom v poslednej rovnosti sme využili symetriu autokorelačnej funkcie, teda rovnosť $\rho(-m+i) = \rho(m-i)$. Rovnaký princíp použijeme aj pre výraz (2.8).

Pre $m = 0$ platí

$$\begin{aligned} &\rho(i)\rho(j) + \rho(-i)\rho(j) + 2\rho(i)\rho(j)\rho(0) - 2\rho(i)\rho(0)\rho(j) - 2\rho(j)\rho(0)\rho(i) \\ &= \rho(i)\rho(j) + \rho(i)\rho(j) + 2\rho(i)\rho(j) - 2\rho(i)\rho(j) - 2\rho(j)\rho(i) = 0, \end{aligned}$$

pričom sme využili rovnosť $\rho(0) = 1$. Výraz (2.8) môžeme teda prepísať nasledovne

$$\begin{aligned} w_{i,j} &= \sum_{m=1}^{\infty} \{ \rho(m+i)\rho(m+j) + \rho(m-i)\rho(m+j) + 2\rho(i)\rho(j)\rho^2(m) \\ &\quad - 2\rho(i)\rho(m)\rho(m+j) - 2\rho(j)\rho(m)\rho(m+i) \} \\ &\quad + \sum_{m=1}^{\infty} \{ \rho(m-i)\rho(m-j) + \rho(m+i)\rho(m-j) + 2\rho(i)\rho(j)\rho^2(m) \\ &\quad - 2\rho(i)\rho(m)\rho(m-j) - 2\rho(j)\rho(m)\rho(m-i) \}. \end{aligned}$$

Po preusporiadaní členov sumy môžeme vidieť platnosť dokazovanej rovnosti

$$\begin{aligned} w_{i,j} &= \sum_{m=1}^{\infty} \{ \rho(m+i)\rho(m+j) + \rho(m+i)\rho(m-j) - 2\rho(j)\rho(m)\rho(m+i) \\ &\quad + \rho(m-i)\rho(m+j) + \rho(m-i)\rho(m-j) - 2\rho(j)\rho(m)\rho(m-i) \\ &\quad - 2\rho(i)\rho(m)\rho(m+j) - 2\rho(i)\rho(m)\rho(m-j) + 4\rho(i)\rho(j)\rho^2(m) \} \\ &= (2.9). \end{aligned}$$

□

Po aplikovaní vety 1 na postupnosť *iid* bieleho šumu dostaneme, že výberové autokorelácie tejto postupnosti majú asymptoticky štandardné mnohorozmerné normálne rozdelenie. Tento dôsledok sformulujeme do nasledujúceho lemmatu.

Lemma 3. *Uvažujme postupnosť $\{\varepsilon_t\} \sim \text{IID}(0, \sigma^2)$. Potom pre výberové autokorelácie tejto postupnosti definované ako (1.2) platí*

$$\sqrt{n} \begin{pmatrix} r_1 \\ \vdots \\ r_k \end{pmatrix} \xrightarrow[n \rightarrow \infty]{D} \text{N}_k(\mathbf{0}, \mathbf{I}_k), \quad (2.10)$$

kde $\mathbf{0}$ je nulový vektor dimenzie k a \mathbf{I}_k je jednotková matica.

Dôkaz. Rada $\{\varepsilon_t\}$ tvorí postupnosť nezávislých a rovnako rozdelených náhodných veličín. Nezávislé náhodné veličiny sú zároveň nekorelované, preto platí nasledujúca implikácia

$$\{\varepsilon_t\} \sim \text{IID}(0, \sigma^2) \implies \rho_k = 0 \text{ pre } \forall k \neq 0. \quad (2.11)$$

Z vety 1 a vzťahov (2.9) a (2.11) dostávame vyjadrenie pre (i, j) -ty prvok kovariačnej matice \mathbf{W} ako

$$w_{i,j} = \begin{cases} 1 & i = j, \\ 0 & \text{inak.} \end{cases} \quad (2.12)$$

Platnosť tohoto vyjadrenia môžeme nahliadnuť nasledovne. Pre $i = j$ pletí

$$w_{i,j} = \sum_{m=1}^{\infty} \left(\rho(m+i) + \rho(m-i) - 2\rho(i)\rho(m) \right)^2 = 1.$$

Všimnime si, že pre $m = i$ je člen $\rho(m - i) = \rho(0) = 1$, teda i -ty sčítanec sumy je rovný jednej. Pre $m \neq i$ sú jednotlivé členy sumy nulové, nakoľko $\rho(k) = 0$ pre všetky $k \neq 0$.

Pre $i \neq j$ platí vzťah (2.9). Pre $m = i$ je člen $\rho(m - i) = \rho(0) = 1$ a ostatné členy v prvom činiteli sú nulové. Druhý činitel je ale pre $m = i$ nulový, pretože pre $m = i$ je $\rho(m - j) \neq \rho(0)$. Pre $m \neq i$ sú sčítance sumy taktiež nulové. \square

Poznámka. V lemmatu 3 musíme predpokladať $\{\varepsilon_t\} \sim \text{IID}(0, \sigma^2)$ z toho dôvodu, aby bol splnený predpoklad vety 1 o rozdelení časovej rady $\{Z_t\}$.

2.3 Autokorelogram

Autokorelogram (označovaný aj ako korelogram) je graf výberových autokorelácií, ktoré sú zobrazené pre jednotlivé oneskorenia k . Najčastejšie sa korelogram používa k posúdeniu či daná časová rada $\{\varepsilon_t\}$ tvorí biely šum. Nulová hypotéza má tvar

$$H_0 : \{\varepsilon_t\} \sim \text{IID}(0, \sigma^2). \quad (2.13)$$

Ak platí H_0 , tak $\{\varepsilon_t\}$ tvorí postupnosť nekorelovaných náhodných veličín, a teda vektor výberových autokorelácií má podľa lemmatu 3 asymptotické rozdelenie $N_k(\mathbf{0}, \mathbf{I}_k)$. Odtiaľ potom platí pre všetky $i = 1, \dots, k$,

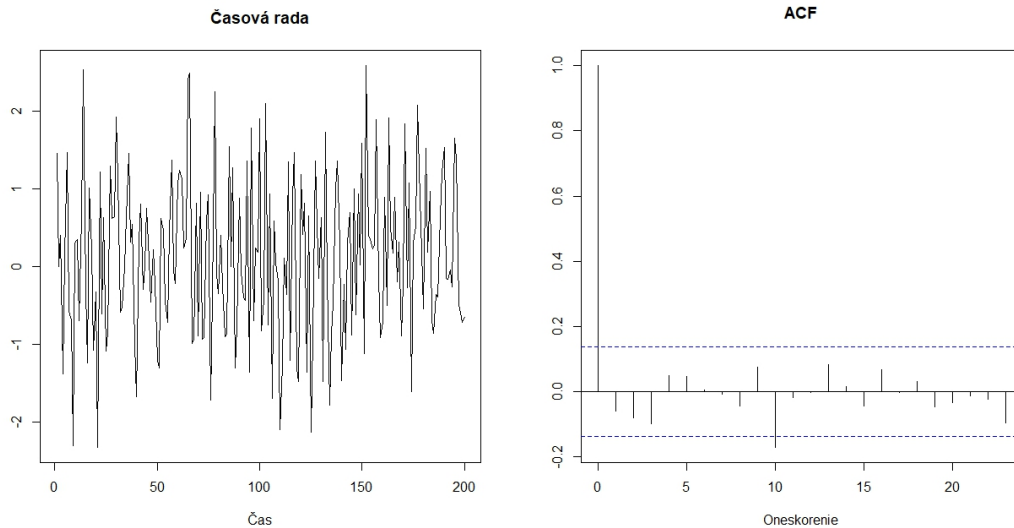
$$\sqrt{n} \cdot r_i \xrightarrow[n \rightarrow \infty]{D} N(0, 1).$$

Naviac r_i a r_j sú asymptoticky nekorelované pre $i \neq j$. Označme α -tý kvantil štandardného normálneho rozdelenia ako u_α . Z predošlého vyjadrenia dostávame

$$\begin{aligned} P\left(\sqrt{nr_i} \in (-u_{1-\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}})\right) &\xrightarrow[n \rightarrow \infty]{} 1 - \alpha, \quad \forall i = 1, \dots, k, \\ P\left(r_i \in \left(-\frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right)\right) &\xrightarrow[n \rightarrow \infty]{} 1 - \alpha, \quad \forall i = 1, \dots, k. \end{aligned} \quad (2.14)$$

Autokorelogram zobrazuje výberové autokorelácie r_i , $i = 1, \dots, k$ pre pevne zvolené k . Vodorovné prerušované čiary na korelograme (viz obrázok 2.1) vyznačujú hornú a dolnú mez intervalu z (2.14), do ktorého náleží r_i s pravdepodobnosťou blížiacou sa k $(1 - \alpha)$. V praxi sa používa najčastejšie hodnota $\alpha = 0,05$. Ak viac ako 5 % výberových autokorelácií vyčnieva z daného intervalu (alebo ak niektoré výberové autokorelácie vyčnievajú výrazne), tak korelogram naznačuje, že hypotéza H_0 neplatí.

Na obrázku 2.1 vidíme postupnosť bieleho šumu $\text{IID}(0, 1)$ spolu s jej korelogramom. Z 20 prvých výberových autokorelácií sa jedna pohybuje za hranicami vyznačeného intervalu, teda aspoň 95 % výberových autokorelácií do intervalu spadá (autokorelačnú funkciu pre oneskorenie $k = 0$ nepočítame, pretože má vždy hodnotu 1).



Obrázek 2.1: Postupnosť bieleho šumu IID(0, 1) a jej korelogram.

Pomocou korelogramu posudzujeme zvlášť pre každú hodnotu výberovej autokorelácie či sa nachádza vo vyznačenom intervale. Našou motiváciou je testovať nekorelovanosť časovej rady pomocou jednej testovej štatistiky, ktorá by mohla predstavovať potenciálne silnejší test. S týmto zámerom boli vytvorené Boxov-Piercov a Ljung-Boxov test, ktorými sa budeme zaoberať v nasledujúcej kapitole. Testová štatistika týchto testov je založená na súčte druhých mocnín výberových autokorelácií.

3. Nevážené Q testy

V tejto kapitole sa budeme zaoberať Boxovým-Piercovým a Ljung-Boxovým testom, ktoré patria medzi nevážené Q testy. V literatúre ich môžeme nájsť tiež pod názvom *portmanteau* testy. Výraz *portmanteau* pochádza z francúzštiny a v preklade znamená lodný kufor, ktorý môže obsahovať veľa menších predmetov. Táto analógia súvisí s typom alternatív, ktoré tieto testy uvažujú. Portmanteau testy sú štatistické testy, u ktorých je nulová hypotéza jasne definovaná, no alternatíva môže zahŕňať veľké množstvo rôznych scenárov. V nasledujúcej časti popíšeme základnú myšlienku, ktorá stojí za vznikom testových štatistík Boxova-Piercova a Ljung-Boxova testu. Autormi Boxova-Piercova testu sú matematici George E. P. Box a David A. Pierce, Ljung-Boxov test je pomenovaný podľa autorov Greta M. Ljung a Georga E. P. Boxa. Obe testové štatistiky sú založené na súčte druhých mocnín výberových autokorelácií testovanej časovej rady.

V tejto kapitole sa budeme zaoberať definíciou testových štatistík a následne ich asymptotickými rozdeleniami. Hlavným výsledkom je veta 5, ktorá umožňuje určiť kritický obor oboch testov. Vlastným prínosom tejto kapitoly je dôkaz vety 4 a ilustrácia vety 5 pomocou niekoľkých naväzujúcich príkladov a obrázkov. V celej kapitole sme čerpali najmä z literatúry (Brockwell a Davis, 1991) a (Box a kol., 2016).

3.1 Boxov-Piercov a Ljung-Boxov test

Predstavme si situáciu, kedy pozorovanú časovú radu $\{Y_t\}$ modelujeme pomocou procesu ARMA(p, q), teda

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + v_1 \varepsilon_{t-1} + \dots + v_q \varepsilon_{t-q}, \quad (3.1)$$

$$\varepsilon_t \sim \text{WN}(0, \sigma^2).$$

Parametre $\varphi_1, \dots, \varphi_p, v_1, \dots, v_q$ môžeme odhadnúť pomocou metódy maximálnej vierohodnosti ako $\hat{\varphi}_1, \dots, \hat{\varphi}_p, \hat{v}_1, \dots, \hat{v}_q$. Našou úlohou je overiť či je odhadnutý model adekvátny pre pozorovanú postupnosť. Ak $\{\varepsilon_t\}$ skutočne tvorí postupnosť nekorelovaných náhodných veličín $\text{WN}(0, \sigma^2)$, tak môžeme usúdiť, že všetka systematická závislosť v dátach je už vyjadrená pomocou odhadnutého modelu. Postupnosť $\{\varepsilon_t\}$ v praxi väčšinou nepozorujeme, preto budeme používať tzv. *reziduá* e_t , ktoré sú definované nasledovne:

$$e_t = Y_t - \hat{\varphi}_1 Y_{t-1} - \dots - \hat{\varphi}_p Y_{t-p} - \hat{v}_1 e_{t-1} - \dots - \hat{v}_q e_{t-q}, \quad (3.2)$$

pričom prvých q hodnôt reziduí položíme rovno nule. Kvôli splneniu predpokladov pre asymptotické rozdelenie testových štatistík Boxova-Piercova a Ljung-Boxova testu budeme testovať špecifickejšiu hypotézu

$$H_0 : \varepsilon_t \sim \text{IID}(0, \sigma^2). \quad (3.3)$$

Oba uvažované testy sú založené na testovaní nulovosti prvých k korelácií testovanej časovej rady, kde k je pevne dané. Označme ρ_i autokorelačnú funkciu

postupnosti $\{\varepsilon_t\}$ pre oneskorenie i . Platnosť hypotézy H_0 potom implikuje platnosť hypotézy

$$\tilde{H}_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0.$$

Testové štatistiky Boxova-Piercova a Ljung-Boxova testu sú založené na súčte druhých mocnín výberových autokorelácií testovanej postupnosti. V praxi budeme pri testovaní hypotézy (3.3) používať výberové autokorelácie spočítané z reziduí e_t .

Testové štatistiky

Majme pozorovania Z_1, \dots, Z_n , pre ktoré chceme testovať hypotézu H_0 definovanú ako (3.3). Nech $k \in \mathbb{N}$ je pevne dané, $k < n$ a nech r_i , $i = 1, \dots, k$ označujú výberové autokorelácie spočítané z pozorovaní Z_1, \dots, Z_n . Testová štatistika Boxova-Piercova testu má tvar

$$Q = n \sum_{i=1}^k r_i^2$$

a testová štatistika Ljung-Boxova testu má tvar

$$Q^* = n(n+2) \sum_{i=1}^k \frac{1}{n-i} r_i^2.$$

3.2 Asymptotické rozdelenie testových štatistík

Odvodenie asymptotického rozdelenia testových štatistík Q a Q^* rozdělíme na dva prípady podľa toho či poznáme skutočné hodnoty parametrov modelu $\text{ARMA}(p, q)$ alebo ich budeme odhadovať. Asymptotické rozdelenie testových štatistík Q a Q^* sa totiž pre tieto dva prípady líši.

3.2.1 Skutočné hodnoty parametrov

V prvom prípade budeme predpokladať, že poznáme skutočné hodnoty parametrov $\varphi_1, \dots, \varphi_p, \nu_1, \dots, \nu_q$. Máme teda k dispozícii postupnosť $\{\varepsilon_t\}$ a žiadne hodnoty nepotrebujeme odhadovať. Pod túto situáciu spadá aj prípad, kedy priamo testujeme či pozorovaná časová rada tvorí postupnosť bieleho šumu $\text{IID}(0, \sigma^2)$. Budeme testovať nulovú hypotézu (3.3). Za platnosti (3.3) konvergujú testové štatistiky Boxova-Piercova a Ljung-Boxova testu k chí-kvadrát rozdeleniu s k stupňami voľnosti, ako ukážeme vo vete 4.

Veta 4. *Nech $\{\varepsilon_t\} \sim \text{IID}(0, \sigma^2)$ a r_i , $i = 1, \dots, k$, sú výberové autokorelácie tejto postupnosti pre pevné $k \in \mathbb{N}$. Potom platí*

$$Q = n \sum_{i=1}^k r_i^2 \xrightarrow[n \rightarrow \infty]{D} \chi_k^2, \quad Q^* = n(n+2) \sum_{i=1}^k \frac{1}{n-i} r_i^2 \xrightarrow[n \rightarrow \infty]{D} \chi_k^2.$$

Dôkaz. Pomocou vety 1 a vzťahov (2.11) a (2.12) dostávame asymptotické rozdelenie výberových autokorelácií bieleho šumu $\{\varepsilon_t\} \sim \text{IID}(0, \sigma^2)$ ako

$$\sqrt{n} \begin{pmatrix} r_1 \\ \vdots \\ r_k \end{pmatrix} \xrightarrow[n \rightarrow \infty]{D} \mathbf{Z}, \quad \mathbf{Z} \sim N_k(\mathbf{0}, \mathbf{I}_k), \quad (3.4)$$

kde \mathbf{Z} je náhodný vektor

$$\mathbf{Z} = (Z_1, \dots, Z_k)^\top.$$

Definujme funkciu h , ktorá je spojitá na \mathbb{R}^k , predpisom

$$h: \mathbb{R}^k \rightarrow \mathbb{R}, \quad h((x_1, \dots, x_k)^\top) = \sum_{i=1}^k x_i^2. \quad (3.5)$$

Z vety o spojitaj transformácii dostávame

$$n \sum_{i=1}^k r_i^2 = h(\sqrt{n}(r_1, \dots, r_k)^\top) \xrightarrow[n \rightarrow \infty]{D} h(\mathbf{Z}) = \sum_{i=1}^k Z_i^2.$$

Vieme, že súčet k druhých mocnín nezávislých náhodných veličín so štandardným normálnym rozdelením $N(0, 1)$ má χ_k^2 rozdelenie, preto platí

$$Q = n \sum_{i=1}^k r_i^2 \xrightarrow[n \rightarrow \infty]{D} \chi_k^2.$$

Asymptotické rozdelenie testovej štatistiky Q^* odvodíme nasledovne. Zo vzťahu

$$\frac{n+2}{n-i} \xrightarrow[n \rightarrow \infty]{} 1, \quad i, n \in \mathbb{N}, i < n, i \text{ pevné},$$

dostávame pre pevné k nasledujúcu konvergenciu

$$\begin{pmatrix} \sqrt{\frac{n+2}{n-1}} & 0 & \dots & 0 \\ 0 & \sqrt{\frac{n+2}{n-2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{\frac{n+2}{n-k}} \end{pmatrix} \xrightarrow[n \rightarrow \infty]{P} \mathbf{I}_k, \quad (3.6)$$

pričom matica z (3.6) je typu $k \times k$. Pomocou (3.4) a Cramérovej-Sluckého vety dostávame

$$\sqrt{n} \begin{pmatrix} \sqrt{\frac{n+2}{n-1}} & 0 & \dots & 0 \\ 0 & \sqrt{\frac{n+2}{n-2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{\frac{n+2}{n-k}} \end{pmatrix} \begin{pmatrix} r_1 \\ \vdots \\ r_k \end{pmatrix} = \sqrt{n} \begin{pmatrix} \sqrt{\frac{n+2}{n-1}} r_1 \\ \vdots \\ \sqrt{\frac{n+2}{n-k}} r_k \end{pmatrix} \xrightarrow[n \rightarrow \infty]{D} \mathbf{Z},$$

pričom pre náhodnú veličinu \mathbf{Z} platí $\mathbf{Z} \sim N_k(\mathbf{0}, \mathbf{I}_k)$. Z vety o spojitaj transformácii pre funkciu (3.5) dostávame

$$n \sum_{i=1}^k \frac{n+2}{n-i} r_i^2 = h \left(\sqrt{n} \left(\sqrt{\frac{n+2}{n-1}} r_1, \dots, \sqrt{\frac{n+2}{n-k}} r_k \right)^\top \right) \xrightarrow[n \rightarrow \infty]{D} h(\mathbf{Z}) = \sum_{i=1}^k Z_i^2.$$

Pomocou rovnakého argumentu ako pre testovú štatistiku Q dostávame asymptotické rozdelenie

$$Q^* = n \sum_{i=1}^k \frac{n+2}{n-i} r_i^2 \xrightarrow[n \rightarrow \infty]{D} \chi_k^2.$$

□

Poznámka. Motiváciou za vytvorením testovej štatistiky Ljung-Boxova testu bol fakt, že rozptyl výberových autokorelácií r_k môžeme presnejšie vyjadriť ako $\frac{(n-k)}{n(n+2)}$, viz (Box a kol., 2016, str. 289).

3.2.2 Odhadnuté hodnoty parametrov

V reálnych situáciách väčšinou nepoznáme skutočné hodnoty parametrov modelu ARMA a preto ich musíme odhadnúť. Zaujíma nás znovu nulová hypotéza (3.3), no v tomto prípade už nepozorujeme priamo postupnosť $\{\varepsilon_t\}$. Testové štatistiky Boxova-Piercova a Ljung-Boxova testu spočítame z reziduí e_t , čo zároveň zmení ich asymptotické rozdelenia. Reziduí e_t sú funkciou maximálne vierohodných odhadov $\hat{\varphi}$, \hat{v} , a teda netvorí postupnosť *iid* náhodných veličín. Predpoklad vety 4 nie je splnený, preto ju už nemôžeme použiť. Hľadané asymptotické rozdelenie popíšeme v nasledujúcej časti, v ktorej sme priamo čerpali z (Brockwell a Davis, 1991, str. 308).

Veta 5 popisuje asymptotické rozdelenie vektora výberových autokorelácií reziduí. Tento výsledok, spolu s pomocným lemmatom 2, použijeme pri odvodzovaní hľadaného asymptotického rozdelenia testových štatistík.

Veta 5. *Uvažujme stacionárnu časovú radu $\{Y_t\}$, ktorá sa riadi kauzálnym a invertibilným procesom ARMA(p, q), definovanom ako (1.3). Definujme reziduí e_t ako (3.2) a označme ako r_i výberové autokorelácie postupnosti $\{e_t\}$ pre oneskorenie $i = 1, \dots, k$, $k \in \mathbb{N}$ pevne dané. Označíme*

$$\mathbf{r} = (r_1, \dots, r_k)^\top. \quad (3.7)$$

Nech $\varphi(z), v(z)$ sú polynómy definované ako (1.4) a (1.5). Definujme polynóm

$$c(z) = \varphi(z)v(z)$$

a mocninnú radu

$$a(z) = (c(z))^{-1} = \sum_{j=0}^{\infty} a_j z^j.$$

Pre $j < 0$ položíme $a_j = 0$. Predpokladajme, že $k \geq p + q$. Definujme matice

$$\mathbf{T}_k = [a_{i-j}]_{i,j=1}^{k,p+q}, \quad \tilde{\mathbf{\Gamma}}_{p+q} = \left[\sum_{h=0}^{\infty} a_h a_{h+|i-j|} \right]_{i,j=1}^{p+q},$$

$$\mathbf{Q} = \mathbf{T}_k \tilde{\mathbf{\Gamma}}_{p+q}^{-1} \mathbf{T}_k^\top = [q_{ij}]_{i,j=1}^k. \quad (3.8)$$

Potom platí

$$\sqrt{n} \cdot \mathbf{r} \xrightarrow[n \rightarrow \infty]{D} N_k(\mathbf{0}, (\mathbf{I}_k - \mathbf{Q})). \quad (3.9)$$

Dôkaz. Viz (Box a Pierce, 1970). □

Poznámka. Podľa (Brockwell a Davis, 1991, str. 308) je matica $\tilde{\mathbf{\Gamma}}_{p+q}$ variančná matica vektora $(X_1, \dots, X_{p+q})^\top$, kde $\{X_t\}$ je časová rada riadiaca sa modelom $\text{AR}(p+q)$ s autoregresným polynómom $c(z)$, pre ktorú platí $\varepsilon_t \sim \text{WN}(0, 1)$.

Príklad 2. Vetu 5 aplikujeme na časovú radu $\{Y_t\}$ riadiacu sa procesom $\text{AR}(1)$, ktorú môžeme vyjadriť ako

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t, \quad |\varphi_1| < 1, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2).$$

Polynómy $\varphi(z)$ a $v(z)$ majú tvar $\varphi(z) = 1 - \varphi_1 z$, $v(z) = 1$, a teda

$$c(z) = 1 - \varphi_1 z.$$

Mocninnú radu $a(z)$ dostaneme pomocou vzorca pre súčet geometrickej rady nasledovne

$$a(z) = (c(z))^{-1} = \frac{1}{1 - \varphi_1 z} = \sum_{k=0}^{\infty} \varphi_1^k z^k$$

a položíme

$$a_k = \begin{cases} \varphi_1^k, & k \geq 0, \\ 0, & k < 0. \end{cases}$$

Preskúmame tvar matice \mathbf{Q} pre $k = 4$. Pre prehľadnosť budeme φ_1 označovať len ako φ . Matica \mathbf{T}_k je v tomto prípade obdĺžniková matica typu 4×1

$$\mathbf{T}_4 = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 1 \\ \varphi \\ \varphi^2 \\ \varphi^3 \end{pmatrix}.$$

Matica $\tilde{\mathbf{\Gamma}}_1$ je typu 1×1 a má tvar

$$\tilde{\mathbf{\Gamma}}_1 = \left(\sum_{h=0}^{\infty} a_h^2 \right) = \left(\sum_{h=0}^{\infty} (\varphi^h)^2 \right) = \left(\frac{1}{1 - \varphi^2} \right).$$

Inverznú maticu k matici $\tilde{\mathbf{\Gamma}}_1$ dostaneme jednoducho ako $\tilde{\mathbf{\Gamma}}_1^{-1} = (1 - \varphi^2)$. Hľadaná matica \mathbf{Q} má teda tvar

$$\mathbf{Q} = \begin{pmatrix} 1 \\ \varphi \\ \varphi^2 \\ \varphi^3 \end{pmatrix} (1 - \varphi^2) (1, \varphi, \varphi^2, \varphi^3) = (1 - \varphi^2) \begin{pmatrix} 1 & \varphi & \varphi^2 & \varphi^3 \\ \varphi & \varphi^2 & \varphi^3 & \varphi^4 \\ \varphi^2 & \varphi^3 & \varphi^4 & \varphi^5 \\ \varphi^3 & \varphi^4 & \varphi^5 & \varphi^6 \end{pmatrix}.$$

Pre zdôraznenie závislosti na φ budeme maticu \mathbf{Q} označovať ako \mathbf{Q}_φ . Asymptotické rozdelenie vektora $(r_1, r_2, r_3, r_4)^\top$ dostaneme nasledovne

$$\sqrt{n} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{pmatrix} \xrightarrow[n \rightarrow \infty]{D} N_4(\mathbf{0}, (\mathbf{I}_4 - \mathbf{Q}_\varphi)). \quad (3.10)$$

Z tvaru matice \mathbf{Q}_φ nie je ťažké určiť, aký tvar má matica \mathbf{Q} pre obecné k . Môžeme si všimnúť, že pre $\varphi \neq 0$ má matica $(\mathbf{I}_k - \mathbf{Q})$ mimo diagonálu nenulové čísla, teda výberové autokorelácie $r_i, i = 1, \dots, k$ nie sú asymptoticky nezávislé.

Veta 6. *Nech je matica \mathbf{Q} definovaná ako (3.8). Potom pre testové štatistiky Q a Q^* platí*

$$Q \xrightarrow[n \rightarrow \infty]{D} \mathbf{Z}, \quad Q^* \xrightarrow[n \rightarrow \infty]{D} \mathbf{Z}, \quad \mathbf{Z} = \sum_{i=1}^k \lambda_i Y_i^2, \quad (3.11)$$

kde $\lambda_1, \dots, \lambda_k$ sú vlastné čísla matice $(\mathbf{I}_k - \mathbf{Q})$ a Y_1, \dots, Y_k sú nezávislé náhodné veličiny s rozdelením $N(0, 1)$.

Dôkaz. Veta plynie pre testovú štatistiku Q s využitím (3.9) a lemmatu 9. Pre testovú štatistiku Q^* ukážeme platnosť vety nasledovne. Keďže platí (3.9), tak analogicky ako v dôkaze vety 4 odvodíme

$$\sqrt{n} \begin{pmatrix} \sqrt{\frac{n+2}{n-1}} & 0 & \dots & 0 \\ 0 & \sqrt{\frac{n+2}{n-2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{\frac{n+2}{n-k}} \end{pmatrix} \mathbf{r} = \sqrt{n} \begin{pmatrix} \sqrt{\frac{n+2}{n-1}} r_1 \\ \vdots \\ \sqrt{\frac{n+2}{n-k}} r_k \end{pmatrix} \xrightarrow[n \rightarrow \infty]{D} N_k(\mathbf{0}, (\mathbf{I}_k - \mathbf{Q})).$$

S využitím tohoto vzťahu a lemmatu 9 dostávame platnosť vety (3.11) pre testovú štatistiku Q^* . □

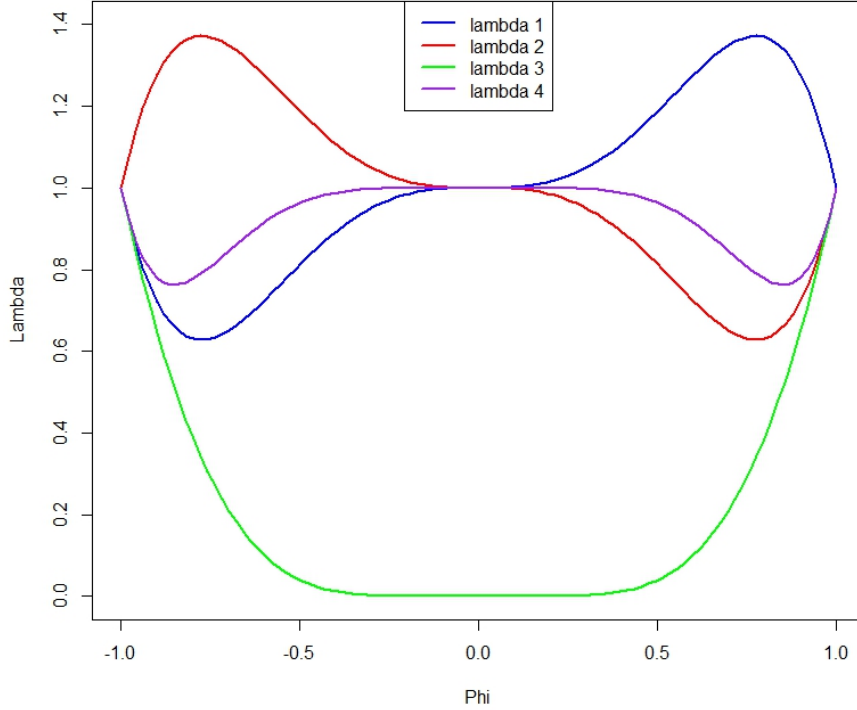
Príklad 3. Uvažujme znovu situáciu z príkladu 2. Chceme odvodiť asymptotické rozdelenie testových štatistík Q a Q^* pre časovú radu $\{Y_t\}$. Na prvý pohľad môžeme vidieť, že matica $(\mathbf{I}_4 - \mathbf{Q}_\varphi)$ nie je idempotentná, preto nemôžeme použiť druhú časť z lemmatu 9. Ak by sme chceli použiť vetu 6 pre maticu \mathbf{Q}_φ , tak by sme potrebovali nájsť vlastné čísla matice $(\mathbf{I}_4 - \mathbf{Q}_\varphi)$, ktorá má tvar

$$(1 - \varphi^2) \begin{pmatrix} \frac{\varphi^2}{1-\varphi^2} & \varphi & \varphi^2 & \varphi^3 \\ \varphi & \frac{1-\varphi^2+\varphi^4}{1-\varphi^2} & \varphi^3 & \varphi^4 \\ \varphi^2 & \varphi^3 & \frac{1-\varphi^4+\varphi^6}{1-\varphi^2} & \varphi^5 \\ \varphi^3 & \varphi^4 & \varphi^5 & \frac{1-\varphi^6+\varphi^8}{1-\varphi^2} \end{pmatrix}. \quad (3.12)$$

S použitím výpočetného softvéru sme našli tieto štyri vlastné čísla matice (3.12)

$$\begin{aligned} \lambda_1 &= -2\varphi^5 + 2\varphi^3 + 1, & \lambda_3 &= \frac{1}{2} \left(\varphi^8 - \sqrt{\varphi^{16} - 16\varphi^{10} + 30\varphi^8 - 16\varphi^6 + 1 + 1} \right), \\ \lambda_2 &= 2\varphi^5 - 2\varphi^3 + 1, & \lambda_4 &= \frac{1}{2} \left(\varphi^8 + \sqrt{\varphi^{16} - 16\varphi^{10} + 30\varphi^8 - 16\varphi^6 + 1 + 1} \right). \end{aligned}$$

Na obrázku 3.1 môžeme vidieť graf vlastných čísel $\lambda_1, \dots, \lambda_4$ v závislosti na parametre $\varphi \in (-1, 1)$. Môžeme si všimnúť, že pre φ blízke krajným hodnotám intervalu $(-1, 1)$ sú všetky vlastné čísla blízko 1, a teda rozdelenie náhodnej veličiny (3.11) je blízko rozdeleniu χ_4^2 . Naopak, pre hodnoty φ blízke 0 sú tri vlastné čísla blízko 1 a jedno vlastné číslo je blízko 0, teda rozdelenie náhodnej veličiny (3.11) je blízko χ_3^2 rozdeleniu. Ďalej si môžeme všimnúť, že pre parameter φ aj



Obrázek 3.1: Graf vlastných čísel $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ v závislosti na parametre φ .

$-\varphi$ dostaneme rovnaké hodnoty vlastných čísel. Vetu 6 by sme síce mohli použiť, no rozdelenie náhodnej veličiny (3.11) pre obecné φ nevieme explicitne vyjadriť pomocou nejakého známeho rozdelenia. Pre ilustráciu sme hľadané rozdelenie odhadli empiricky pomocou simulácie v softvéri R (R Core Team (2021)). Empirickú distribučnú funkciu pre parametre φ z množiny $\{0,80; 0,87; 0,91; 0,96; 0,99\}$ nájdeme na obrázku 3.2. Modrá prerušovaná čiara označuje 95%-ný kvantil príslušných rozdelení.

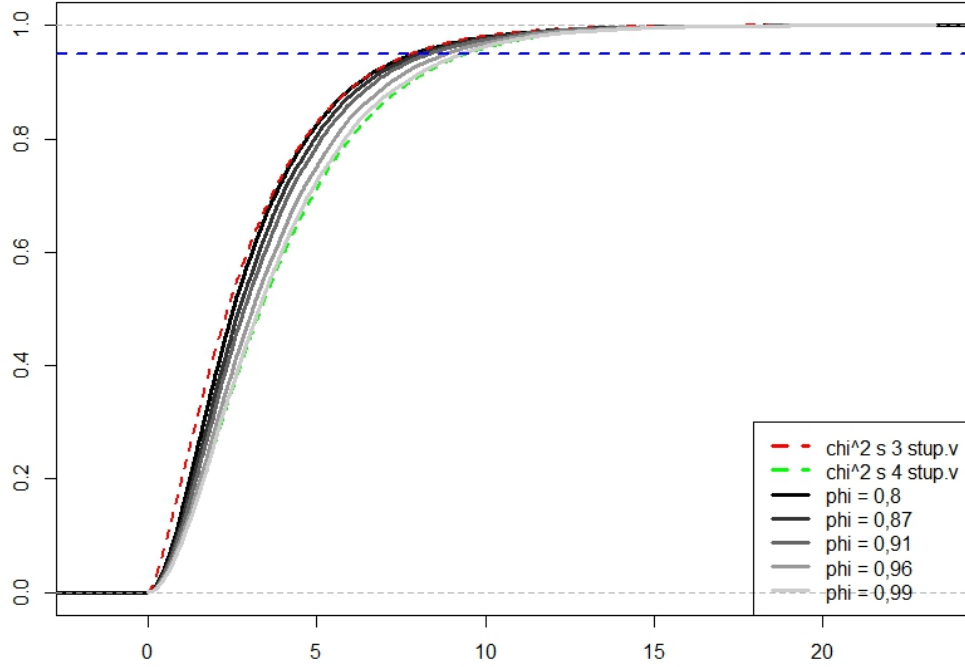
Veta 6 nám síce poskytuje predstavu o rozdelení testových štatistík Q a Q^* , no presné rozdelenie nie je jednoduché vyjadriť, keďže matica $(\mathbf{I}_k - \mathbf{Q})$ nie je obecné idempotentná a vlastné čísla závisia na neznámych parametroch. Rozdelenie náhodnej veličiny (3.11) budeme teda aproximovať, z čoho získame približné rozdelenie Q , resp. Q^* .

Označme ako $\{Y_t\}$ časovú radu, ktorá splňuje predpoklady vety 5. Predpokladajme, že k závisí na n tak, že $k \rightarrow \infty$ ak $n \rightarrow \infty$. Z predpokladu kauzality vety 5 môžeme vyjadriť všetky členy rady $\{Y_t\}$ ako

$$Y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} |c_j| < \infty.$$

Nech podľa (Box a Pierce, 1970) platia nasledujúce dva predpoklady:

- $k = O(n^{1/2})$,
- $c_j = O(n^{-1/2})$ pre $j \geq k$.



Obrázek 3.2: Empirické distribučné funkcie náhodnej veličiny $\sum_{i=1}^4 \lambda_i Y_i^2$ pre hodnoty parametru φ z množiny $\{0,80; 0,87; 0,91; 0,96; 0,99\}$, pričom λ_i , $i = 1, \dots, 4$ sú vlastné čísla matice (3.12) a Y_i , $i = 1, \dots, 4$ sú nezávislé náhodné veličiny s rozdelením $N(0, 1)$. Modrá prerušovaná čiara označuje 95%-ný kvantil príslušných rozdelení.

Za týchto podmienok môžeme maticu $\tilde{\mathbf{\Gamma}}_{p+q}$ aproximovať pomocou matice $\mathbf{T}_k^\top \mathbf{T}_k$, pričom matica \mathbf{T}_k má hodnotu $p+q$, viz (Brockwell a Davis, 1991, str. 310). Túto aproximáciu môžeme chápať ako

$$\tilde{\mathbf{\Gamma}}_{p+q} - \mathbf{T}_k^\top \mathbf{T}_k \xrightarrow{n \rightarrow \infty} \mathbf{0}_{p+q},$$

pričom $\mathbf{0}$ označuje nulovú maticu. Maticu \mathbf{Q} môžeme potom aproximovať maticou

$$\mathbf{T}_k (\mathbf{T}_k^\top \mathbf{T}_k)^{-1} \mathbf{T}_k^\top, \quad (3.13)$$

s hodnotou

$$\text{rank}(\mathbf{T}_k (\mathbf{T}_k^\top \mathbf{T}_k)^{-1} \mathbf{T}_k^\top) = \text{rank}(\mathbf{T}_k) = p + q.$$

Matica (3.13) je projekčná matica, teda je idempotentná. Matica

$$\mathbf{I}_k - \mathbf{T}_k (\mathbf{T}_k^\top \mathbf{T}_k)^{-1} \mathbf{T}_k^\top \quad (3.14)$$

predstavuje maticu projekcie do kolmého priestoru, teda je taktiež idempotentná. Túto skutočnosť môžeme overiť aj pomocou jednoduchého výpočtu:

$$\begin{aligned} & (\mathbf{I}_k - \mathbf{T}_k (\mathbf{T}_k^\top \mathbf{T}_k)^{-1} \mathbf{T}_k^\top) (\mathbf{I}_k - \mathbf{T}_k (\mathbf{T}_k^\top \mathbf{T}_k)^{-1} \mathbf{T}_k^\top) = \\ & = \mathbf{I}_k - \mathbf{T}_k (\mathbf{T}_k^\top \mathbf{T}_k)^{-1} \mathbf{T}_k^\top - \mathbf{T}_k (\mathbf{T}_k^\top \mathbf{T}_k)^{-1} \mathbf{T}_k^\top + \mathbf{T}_k (\mathbf{T}_k^\top \mathbf{T}_k)^{-1} \mathbf{T}_k^\top \mathbf{T}_k (\mathbf{T}_k^\top \mathbf{T}_k)^{-1} \mathbf{T}_k^\top \\ & = \mathbf{I}_k - \mathbf{T}_k (\mathbf{T}_k^\top \mathbf{T}_k)^{-1} \mathbf{T}_k^\top. \end{aligned}$$

Matica (3.14) má hodnotu

$$\text{rank}(\mathbf{I}_k - \mathbf{T}_k(\mathbf{T}_k^\top \mathbf{T}_k)^{-1} \mathbf{T}_k^\top) = \text{rank}(\mathbf{I}_k) - \text{rank}(\mathbf{T}_k) = k - p - q. \quad (3.15)$$

Maticu $(\mathbf{I}_k - \mathbf{Q})$ môžeme teda aproximovať idempotentnou maticou (3.14). S využitím lemmatu 9, vety 6 a vzťahu (3.15) dostávame, že testová štatistika Q (resp. Q^*) má pre k a n dost veľké približne rozdelenie χ_{k-p-q}^2 . Nulovú hypotézu (3.3) teda zamietame na hladine α , ak

$$Q > \chi_{1-\alpha}^2(k - p - q), \quad \text{resp.} \quad Q^* > \chi_{1-\alpha}^2(k - p - q), \quad (3.16)$$

kde $\chi_{1-\alpha}^2(k - p - q)$ označuje $(1 - \alpha)$ -tý kvantil rozdelenia χ_{k-p-q}^2 .

Príklad 4. Uvažujme znovu situáciu z príkladu 2. Matica $\tilde{\Gamma}_1$ má jeden prvok, ktorý predstavuje nekonečný súčet geometrickej rady

$$\tilde{\Gamma}_1 = \left(\sum_{h=0}^{\infty} (\varphi^2)^h \right).$$

Maticu $\tilde{\Gamma}_1$ aproximujeme pomocou konečného súčtu geometrickej rady ako

$$\mathbf{T}_4^\top \mathbf{T}_4 = (1 + \varphi^2 + \varphi^4 + \varphi^6) = \left(\sum_{h=0}^3 (\varphi^2)^h \right).$$

Maticu \mathbf{Q}_φ teda aproximujeme ako

$$\frac{1}{1 + \varphi^2 + \varphi^4 + \varphi^6} \begin{pmatrix} 1 & \varphi & \varphi^2 & \varphi^3 \\ \varphi & \varphi^2 & \varphi^3 & \varphi^4 \\ \varphi^2 & \varphi^3 & \varphi^4 & \varphi^5 \\ \varphi^3 & \varphi^4 & \varphi^5 & \varphi^6 \end{pmatrix}.$$

Poznámka. Pri odvodzovaní približného rozdelenia testových štatistík Q a Q^* je dôležitý predpoklad $k \rightarrow \infty$. Pre malé k už táto aproximácia rozdelenia fungovať nemusí.

Príklad 5. Príklad 2 môžeme modifikovať aj pre časovú radu $\{Y_t\}$ riadiacu sa modelom MA(1), teda

$$Y_t = \varepsilon_t + v_1 \varepsilon_{t-1}, \quad |v_1| < 1, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2).$$

Pre prehľadnosť budeme znovu označovať v_1 ako v . Matica \mathbf{Q} (ktorú v tomto prípade označíme ako \mathbf{Q}_v) má tvar ako matica \mathbf{Q}_φ pre $\varphi = -v$, teda

$$\mathbf{Q}_v = (1 - v^2) \begin{pmatrix} 1 & -v & v^2 & -v^3 \\ -v & v^2 & -v^3 & v^4 \\ v^2 & -v^3 & v^4 & -v^5 \\ -v^3 & v^4 & -v^5 & v^6 \end{pmatrix}.$$

Ostatné výsledky dostaneme analogicky ako v predošlom príklade. Vráťme sa opäť do kontextu sekcie 3.2.1 a uvažujme modely AR(1) a MA(1). Ak poznáme

skutočnú hodnotu parametru φ_1 , resp. v_1 , tak podľa (3.4) je asymptotický rozptyl výberových autokorelácií bieleho šumu $\{\varepsilon_t\}$ rovný

$$\frac{1}{n}.$$

Zo vzťahu (3.9) a z tvaru matice (3.12) môžeme ale vidieť, že asymptotický rozptyl výberovej autokorelácie r_1 spočítanej z reziduí e_t definovaných ako (3.2) pre model AR(1) je rovný

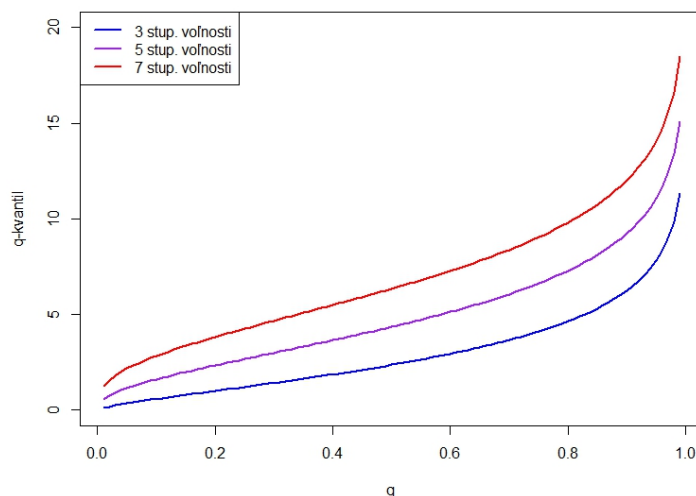
$$\frac{\varphi^2}{n}$$

a pre model MA(1) analogicky

$$\frac{v^2}{n}.$$

Hodnoty $\frac{\varphi^2}{n}$, resp. $\frac{v^2}{n}$ môžu byť pre $|\varphi_1| < 1, |v_1| < 1$ významne menšie než hodnota $\frac{1}{n}$. Všimnime si ale, že hodnoty na diagonále matíc \mathbf{Q}_φ a \mathbf{Q}_v sa postupne znižujú. Asymptotický rozptyl výberových autokorelácií spočítaných z reziduí e_t pre modely AR(1) a MA(1) sa teda približujú hodnote $\frac{1}{n}$ so zvyšujúcim sa oneskorením k . Môžeme teda vidieť, že odhad asymptotického rozptylu ako $\frac{1}{n}$ pre tieto výberové autokorelácie môže značne nadhodnocovať skutočný rozptyl pre malé hodnoty oneskorení k , no pre veľké hodnoty k môže byť vhodnou aproximáciou.

Poznámka. Ak by sme v kritickom obore (3.16) použili rozdelenie χ_k^2 namiesto χ_{k-p-q}^2 (teda ak by sme použili chí-kvadrát rozdelenie s väčším počtom stupňov voľnosti), tak by sme dostali viac konzervatívny test. Odôvodnenie môžeme vidieť na obrázku 3.3, na ktorom sú zobrazené kvantilové funkcie rozdelenia χ^2 pre rôzne počty stupňov voľnosti. So zvyšujúcim sa počtom stupňov voľnosti sú kritické hodnoty (resp. kvantily) väčšie. Z tohoto dôvodu je kritický obor menší pre vyšší počet stupňov voľnosti (kritická hodnota je naopak väčšia) a teda príslušný test by bol konzervatívnejší.



Obrázek 3.3: Kvantilová funkcia χ^2 rozdelenia pre rôzne počty stupňov voľnosti.

4. Vážené Q testy

V tejto kapitole sa budeme zaoberať váženými portmanteau testami, ktoré predstavujú zobecnenie Boxova-Piercova a Ljung-Boxova testu, ktorými sme sa doposiaľ zaoberali. Uvedieme obecnú testovú štatistiku váženého portmanteau testu Q_W pre obecné váhy w_i a ďalej sa budeme venovať jej asymptotickému rozdeleniu. V celej kapitole sme čerpali najmä z článku (Gallagher a Fisher, 2015). Vlastným prínosom tejto kapitoly je presná formulácia vety 7 a jej dôkaz, zjednotenie značenia vo vetách 7 a 8 a tiež ilustrácia hlavných výsledkov pomocou niekoľkých príkladov a obrázkov.

Ljung-Boxov a Boxov-Piercov test patria medzi prvé portmanteau testy, pomocou ktorých bolo možné testovať adekvátnosť odhadnutého modelu ARMA. Neskôr sa objavovali rôzne modifikácie týchto testov, ktoré predstavovali potenciálne silnejšie testy. Medzi ne patrí napríklad Fisherov a Gallagherov test, ktorý predstavili autori v článku Fisher a Gallagher (2012). Testová štatistika Q_{FG} tohoto testu je založená na váženej sume výberových autokorelácií.

$$Q_{FG} = n(n+2) \sum_{i=1}^k \left(\frac{k-i+1}{k} \right) \frac{r_i^2}{n-i}.$$

Testové štatistiky Q , Q^* a Q_{FG} sa dajú zapísať jednotne v tvare

$$Q_W = n \sum_{i=1}^k w_i r_i^2 \quad (4.1)$$

pre vhodne zvolené váhy $w_i, i = 1, \dots, k$. Toto nás motivuje uvažovať obecný tvar testovej štatistiky (4.1) a preskúmať aj iné tvary váh w_i . Testy založené na testovej štatistike (4.1) budeme podľa článku (Gallagher a Fisher, 2015) nazývať *vážené portmanteau testy* alebo *vážené Q testy*. V práci budeme ďalej predpokladať, že $w_i \geq 0, \forall i = 1, \dots, k$, a tiež že váhy w_i závisia na parametroch n a k :

$$w_i = w_{i,n,k}.$$

Testová štatistika

Majme pozorovania Z_1, \dots, Z_n , pre ktoré by sme chceli testovať nulovú hypotézu (3.3). Nech $k \in \mathbb{N}$ je pevne dané, $k < n$ a nech $r_i, i = 1, \dots, k$ označujú výberové autokorelácie spočítané z pozorovaní Z_1, \dots, Z_n . Testová štatistika váženého portmanteau testu má tvar

$$Q_W = n \sum_{i=1}^k w_i r_i^2. \quad (4.2)$$

V tejto kapitole sa budeme zaoberať váženými portmanteau testami založenými na testovej štatistike (4.2). V prvej časti sa budeme zaoberať asymptotickým rozdelením tejto testovej štatistiky a v druhej časti uvedieme niektoré typy váh, ktoré je možné použiť.

Poznámka. Testové štatistiky Boxova-Piercova a Ljung-Boxova testu môžeme zapísať v tvare (4.2), no v kontexte našej práce ich pre jednoduchosť nazývame ako nevážené testy.

4.1 Asymptotické rozdelenie váženej štatistiky

V tejto časti sa budeme venovať rozdeleniu váženej testovej štatistiky (4.2). Vo vete 7 uvedieme asymptotické rozdelenie pre fixné k s váhami, ktoré konvergujú k nezápornému reálnemu číslu. Vo vete 8 budeme predpokladať, že k rastie spolu s n a uvedieme asymptotické rozdelenie pre váhy, ktoré nie sú sčítateľné pre $n \rightarrow \infty, k \rightarrow \infty, k/n \rightarrow 0$.

Veta 7. *Uvažujme časovú radu $\{Y_t\}$, ktorá splňuje (3.1). Definujme reziduá e_t ako (3.2) a označme r_i ako výberové autokorelácie postupnosti $\{e_t\}$ pre oneskorenie $i = 1, \dots, k$, $k \in \mathbb{N}$ pevne dané. Označíme $\mathbf{r} = (r_1, \dots, r_k)^\top$. Predpokladajme, že w_i sú nezáporné váhy, pre ktoré platí*

$$w_i \xrightarrow[n \rightarrow \infty]{} w_i^*, \quad \forall i = 1, \dots, k, \quad w_i^* \in [0, \infty). \quad (4.3)$$

Označme maticu \mathbf{Q} ako (3.8) a definujme matice

$$\begin{aligned} \mathbf{W}^{1/2} &= \text{diag}(\sqrt{w_1}, \dots, \sqrt{w_k}), \\ \mathbf{W}^* &= \text{diag}(w_1^*, \dots, w_k^*), \quad \mathbf{W}^{*1/2} = \text{diag}(\sqrt{w_1^*}, \dots, \sqrt{w_k^*}). \end{aligned}$$

Potom platí

$$\sqrt{n} \cdot \mathbf{W}^{1/2} \cdot \mathbf{r} \xrightarrow[n \rightarrow \infty]{D} N_k(\mathbf{0}, \mathbf{W}^{*1/2}(\mathbf{I}_k - \mathbf{Q})\mathbf{W}^{*1/2})$$

a pre testovú štatistiku Q_W platí

$$Q_W \xrightarrow[n \rightarrow \infty]{D} \mathbf{Z}, \quad \mathbf{Z} = \sum_{i=1}^k \lambda_i Y_i^2, \quad (4.4)$$

kde λ_i sú vlastné čísla matice $(\mathbf{I}_k - \mathbf{Q})\mathbf{W}^*$ a Y_1, \dots, Y_k sú nezávislé náhodné veličiny s rozdelením $N(0, 1)$.

Dôkaz. Matica $\mathbf{W}^{*1/2}$ je diagonálna, teda $(\mathbf{W}^{*1/2})^\top = \mathbf{W}^{*1/2}$. Zo vzťahov (4.3) a (3.9) následne dostaneme

$$\sqrt{n} \cdot \mathbf{W}^{1/2} \cdot \mathbf{r} = \sqrt{n} \begin{pmatrix} \sqrt{w_1} r_1 \\ \vdots \\ \sqrt{w_k} r_k \end{pmatrix} \xrightarrow[n \rightarrow \infty]{D} N_k(\mathbf{0}, \mathbf{W}^{*1/2}(\mathbf{I}_k - \mathbf{Q})\mathbf{W}^{*1/2}).$$

S využitím tohoto vzťahu a lemmatu 9 dostávame

$$Q_W \xrightarrow[n \rightarrow \infty]{D} \mathbf{Z}, \quad \mathbf{Z} = \sum_{i=1}^k \lambda_i Y_i^2, \quad (4.5)$$

pričom λ_i sú vlastné čísla matice

$$\mathbf{W}^{*1/2}(\mathbf{I}_k - \mathbf{Q})\mathbf{W}^{*1/2}.$$

Z lineárnej algebry vieme, že podobné matice majú rovnaké vlastné čísla. Matice A, B sú podobné, ak existuje invertovateľná matica P taká, že

$$B = P^{-1}AP.$$

Môžeme vidieť, že matica $\mathbf{W}^{*1/2}(\mathbf{I}_k - \mathbf{Q})\mathbf{W}^{*1/2}$ je podobná s maticou

$$(\mathbf{I}_k - \mathbf{Q})\mathbf{W}^{*1/2}\mathbf{W}^{*1/2} = (\mathbf{I}_k - \mathbf{Q})\mathbf{W}^*,$$

pretože

$$\begin{aligned}\mathbf{W}^{*1/2}(\mathbf{I}_k - \mathbf{Q})\mathbf{W}^{*1/2} &= (\mathbf{W}^{*1/2})^{-1}\mathbf{W}^{*1/2}(\mathbf{I}_k - \mathbf{Q})\mathbf{W}^{*1/2}\mathbf{W}^{*1/2} \\ &= (\mathbf{I}_k - \mathbf{Q})\mathbf{W}^*.\end{aligned}$$

Dostávame teda, že Q_W konverguje v distribúcii k náhodnej veličine \mathbf{Z} definovanou ako (4.5), kde λ_i sú vlastné čísla matice $(\mathbf{I}_k - \mathbf{Q})\mathbf{W}^*$. □

Poznámka. Veta 7 platí aj pre váhy w_i , ktoré nezávisia na n . V tomto prípade platia výsledky vety 7 pre $w_i^* = w_i, \forall i = 1, \dots, k$.

Veta 7 nám síce dá predstavu o rozdelení Q_W , no podobne ako v príklade 3 nemusíme byť schopí toto rozdelenie vyjadriť pomocou nejakého známeho rozdelenia. Veta 8 nám naopak poskytne presný tvar asymptotického rozdelenia pre modifikáciu testovej štatistiky Q_W .

Veta 8. *Uvažujme časovú radu $\{Y_t\}$, ktorá splňuje (3.1). Ďalej uvažujme testovú štatistiku (4.2). Ak platia podmienky*

- $\sum_{i=1}^k w_i^2 \rightarrow \infty$ pre $n \rightarrow \infty, k \rightarrow \infty, \frac{k}{n} \rightarrow 0$,
- $\mathbb{E}[\varepsilon_t^8] < \infty$,

tak pre testovú štatistiku Q_W platí vzťah

$$\frac{Q_W - \sum_{i=1}^k w_i}{\sqrt{2 \sum_{i=1}^k w_i^2}} \xrightarrow[n \rightarrow \infty]{D} N(0, 1). \quad (4.6)$$

Dôkaz. Viz (Gallagher a Fisher, 2015). □

Testové štatistiky Q a Q^* sú špeciálne prípady štatistiky Q_W , pričom váhy odpovedajúce štatistike Q majú tvar $w_i^Q = 1, \forall i = 1, \dots, k$ a váhy odpovedajúce štatistike Q^* majú tvar $w_i^{Q^*} = \frac{n+2}{n-i}, \forall i = 1, \dots, k$. Pozrieme sa, čo nám hovorí veta 8 pre testovú štatistiku Q . Prvú podmienku z vety 8 môžeme vyjadriť ako

$$\sum_{i=1}^k (w_i^Q)^2 = k \rightarrow \infty \quad \text{a zároveň } n \rightarrow \infty, k/n \rightarrow 0.$$

Pre váhy w_i^Q platí

$$\sum_{i=1}^k w_i^Q = \sum_{i=1}^k (w_i^Q)^2 = k, \quad (4.7)$$

teda z vety 8 dostávame

$$\frac{Q - k}{\sqrt{2k}} \xrightarrow[n \rightarrow \infty]{D} N(0, 1). \quad (4.8)$$

Z predchádzajúcej kapitoly vieme, že pre $n \rightarrow \infty, k \rightarrow \infty, k = O(n^{1/2})$ môžeme rozdelenie testovej štatistiky Q aproximovať pomocou χ_{k-p-q}^2 . Z lemmatu 10 potom dostávame, že

$$\frac{Q - (k - p - q)}{\sqrt{2(k - p - q)}} \quad (4.9)$$

má približne rozdelenie $N(0, 1)$. Pre $k \rightarrow \infty$ sú parametre p a q vo výraze (4.9) zanedbateľné, teda približné rozdelenie (4.9) odpovedá dôsledku (4.8) vety 8.

V ďalšej časti sa budeme zaoberať výberom váh w_i . Budeme uvažovať dve kategórie váh podľa článku (Gallagher a Fisher, 2015). Prvú kategóriu tvoria váhy, ktoré sú založené na jadrovej funkcii $\kappa(u)$. Ako druhé budeme uvažovať váhy, ktoré sa geometricky zmenšujú na základe zvoleného koeficientu.

4.2 Jadrové váhy

Tieto váhy sú založené na druhej mocnine jadrovej funkcie $\kappa(u)$. Jadrové funkcie majú značné využitie v štatistike, napríklad v teórii jadrových odhadov. Podrobnosti o jadrových funkciách je možné nájsť v knihách (Fan a Yao, 2003; Racine, 2008), z ktorých sme čerpali v tejto podkapitole. V tejto časti definujeme jadrovú funkciu jednej premennej a uvedieme niekoľko príkladov najznámejších jadrových funkcií. Na záver definujeme váhy w_i založené na jadrovej funkcii $\kappa(u)$.

Definícia 11 (jadrová funkcia). *Jadrová funkcia je integrovateľná funkcia $\kappa(u)$, ktorá splňuje*

$$\int_{-\infty}^{\infty} \kappa(u) du = 1.$$

V našej práci budeme uvažovať iba reálne a symetrické jadrové funkcie. Najpoužívanejšie jadrové funkcie uvádzame v tabuľke 4.1. Pre prehľadnosť budeme označovať jadrovú funkciu ako *jadro*. Jednotlivé jadrové funkcie definujeme pomocou indikátora, ktorý má nasledujúci tvar

$$\mathbb{I}\{|u| \leq 1\} = \begin{cases} 1, & |u| \leq 1, \\ 0, & |u| > 1. \end{cases}$$

Váhy založené na jadrovej funkcii $\kappa(u)$ majú tvar

$$w_i = \frac{n+2}{n-i} \left[\kappa\left(\frac{i}{k}\right) \right]^2. \quad (4.10)$$

Intuícia za vytvorením váh (4.10) je nasledujúca. Budeme vychádzať z váh Ljung-Boxova testu, ktoré môžeme vyjadriť v kontexte vážených testov ako

$$w_i^{Q^*} = \frac{n+2}{n-i} \cdot \mathbb{I}\left\{ \left| \frac{i}{k} \right| \leq 1 \right\}.$$

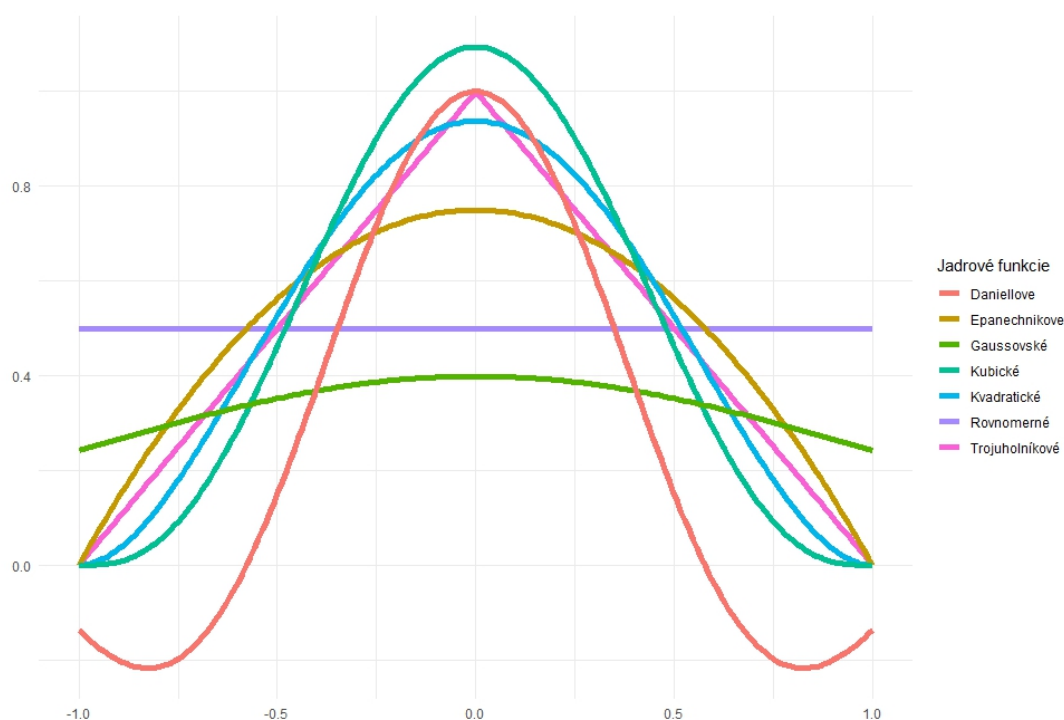
Tieto váhy by sme chceli rozšíriť aj pre ďalšie jadrové funkcie. Vzhľadom k tomu, že niektoré jadrové funkcie môžu nadobúdať záporné hodnoty (napr. Daniellovo jadro), vezmeme druhú mocninu jadrovej funkcie. Dostávame teda obecný tvar jadrových váh ako

$$w_i = \frac{n+2}{n-i} \left[\kappa\left(\frac{i}{k}\right) \right]^2.$$

Rovnomerné jadro	$\kappa(u) = \frac{1}{2}\mathbb{I}\{ u \leq 1\}$
Trojuholníkové jadro	$\kappa(u) = (1 - u)\mathbb{I}\{ u \leq 1\}$
Epanechnikovo jadro	$\kappa(u) = \frac{3}{4}(1 - u^2)\mathbb{I}\{ u \leq 1\}$
Kvadratické jadro	$\kappa(u) = \frac{15}{16}(1 - u^2)^2\mathbb{I}\{ u \leq 1\}$
Kubické jadro	$\kappa(u) = \frac{35}{32}(1 - u^2)^3\mathbb{I}\{ u \leq 1\}$
Gaussovské jadro	$\kappa(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$
Daniellovo jadro	$\kappa(u) = \frac{\sin(\sqrt{3}\pi u)}{\sqrt{3}\pi u}\mathbb{I}\{ u \leq 1\}$

Tabulka 4.1: Príklady jadrových funkcií.

Na obrázku 4.1 môžeme vidieť graf jadrových funkcií z tabuľky 4.1 na intervale $[-1, 1]$. Z obrázku môžeme vidieť, že váhy založené na jadrových funkciách zaisťujú, že dôležitosť jednotlivých výberových autokorelácií klesá so zvyšujúcim sa oneskorením k (s výnimkou rovnomerných váh a Daniellových váh, u ktorých môžeme vidieť mierny nárast pri krajoch intervalu).



Obrázok 4.1: Graf jadrových funkcií na intervale $[-1, 1]$.

Na určenie asymptotického rozdelenia váženej štatistiky s jadrovými váhami budeme v tejto práci používať vetu 8. V nasledujúcom príklade overíme prvý predpoklad tejto vety pre rovnomerné jadrové váhy.

Príklad 6. Pre rovnomerné jadrové váhy ukážeme platnosť prvého predpokladu vety 8. Váhy založené na rovnomernej jadrovej funkcii majú tvar

$$w_i = \frac{1}{4} \cdot \frac{n+2}{n-i}.$$

Pre tieto váhy platí

$$w_i^2 = \frac{1}{16} \left(\frac{n+2}{n-i} \right)^2 \xrightarrow{n \rightarrow \infty} \frac{1}{16},$$

a teda

$$\sum_{i=1}^k \frac{1}{16} \left(\frac{n+2}{n-i} \right)^2 \rightarrow \infty, \quad \text{pre } n \rightarrow \infty, k \rightarrow \infty, k/n \rightarrow 0.$$

Podobným spôsobom môžeme tento predpoklad overiť aj pre ostatné jadrové váhy. V simulačných štúdiách v nasledujúcej kapitole budeme používať kritický obor vážených testov s jadrovými váhami založený na vete 8 a vzťahu (4.6).

4.3 Geometricky klesajúce váhy

Geometricky klesajúce váhy zaistujú, že dôležitosť výberových autokorelácií v testovej štatistike (4.2) klesá exponenciálne so zvyšujúcim sa oneskorením k . Geometricky klesajúce váhy sa teda znižujú podľa zvoleného parametru a . Autori článku Gallagher a Fisher (2015) navrhujú uvažovať váhy v tvare

$$w_i = (p+q)a^{i-1},$$

pričom p, q sú rády uvažovaného modelu ARMA(p, q) a $0 < a < 1$ je zvolený parameter. Pre súčet geometricky klesajúcich váh a pre súčet ich druhých mocnín platí

$$\sum_{i=1}^k w_i = (p+q) \frac{1-a^k}{1-a}, \quad \sum_{i=1}^k w_i^2 = (p+q)^2 \frac{1-a^{2k}}{1-a^2}. \quad (4.11)$$

Poznámka. Geometricky klesajúce váhy nespĺňujú prvý predpoklad vety 8, keďže platí

$$\sum_{i=1}^k w_i^2 = \sum_{i=1}^k (p+q)^2 a^{2(i-1)} = (p+q)^2 \sum_{i=0}^{k-1} a^{2i} \xrightarrow{k \rightarrow \infty} \frac{(p+q)^2}{1-a^2} \neq \infty.$$

Vetu 8 teda nemôžeme použiť na určenie kritického oboru váženého testu s geometricky klesajúcimi váhami. Autori článku Gallagher a Fisher (2015) odporúčajú aproximovať rozdelenie štatistiky Q_W s geometrickými váhami pomocou vhodného gama rozdelenia, ktorému sa budeme venovať v nasledujúcej sekcii.

Poznámka. Autori článku Gallagher a Fisher (2015) navrhujú okrem predošlých dvoch typov váh aj tzv. *adaptívne váhy*. Tieto váhy sú založené na výberových parciálnych autokoreláciách $\{\pi_{ii}\}$. Testová štatistika (4.2) s adaptívnymi váhami má tvar

$$Q_{W_A} = n \sum_{i=1}^{k_0} \frac{n+2}{n-i} r_i^2 + n \sum_{i=k_0+1}^k w_i r_i^2,$$

kde $k_0 = \min(\log(n), M)$, M je zvolená konečná horná hranica a w_i sú váhy definované ako

$$w_i = -\log(1 - |\pi_{ii}|).$$

Takto definované váhy sú z princípu náhodné, a preto nezapadajú do konceptu vety 7 ani vety 8. V našej práci ich teda nebudeme ďalej uvažovať.

4.3.1 Aproximácia rozdelenia

Podľa autorov článku (Gallagher a Fisher, 2015) je vhodné aproximovať rozdelenie váženej testovej štatistiky (4.2) pomocou vhodného gama rozdelenia. Autori odporúčajú rozdelenie $\Gamma(\alpha, \beta)$, $\alpha, \beta > 0$ s parametrami

$$\alpha = \frac{\left(\sum_{i=1}^k w_i\right)^2}{2\left(\sum_{i=1}^k w_i^2 - p - q\right)}, \quad \beta = \frac{2\left(\sum_{i=1}^k w_i^2 - p - q\right)}{\sum_{i=1}^k w_i}. \quad (4.12)$$

Pripomeňme, že hustota náhodnej veličiny $X \sim \Gamma(\alpha, \beta)$, $\alpha, \beta > 0$ má tvar

$$f(x; \alpha, \beta) = \begin{cases} \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

kde Γ je gama funkcia.

Príklad 7. Ak dosadíme vyjadrenia (4.11) pre geometricky klesajúce váhy do predpisov pre parametre α a β definované ako (4.12), tak dostaneme

$$\alpha = \frac{(p+q)(1-a^k)^2(1+a)}{2\left[(p+q)(1-a^{2k}) - (1-a^2)\right](1-a)},$$

$$\beta = \frac{2\left[(p+q)(1-a^{2k}) - (1-a^2)\right]}{(1+a)(1-a^k)}.$$

Kedže platí $k \in \mathbb{N}$, $a \in (0, 1)$, $p, q \geq 0$, $p+q \neq 0$, tak parametre α a β sú kladné, ak je splnená podmienka

$$p+q > \frac{1-a^2}{1-a^{2k}}.$$

Príklad 8. Pozrieme sa, ako vyzerá aproximácia rozdelenia testovej štatistiky Q podľa (4.12). Vieme, že χ_k^2 rozdelenie je špeciálny prípad gama rozdelenia. Presnejšie, pre náhodnú veličinu X platí

$$X \sim \chi_{k-p-q}^2 \iff X \sim \Gamma(\alpha, \beta),$$

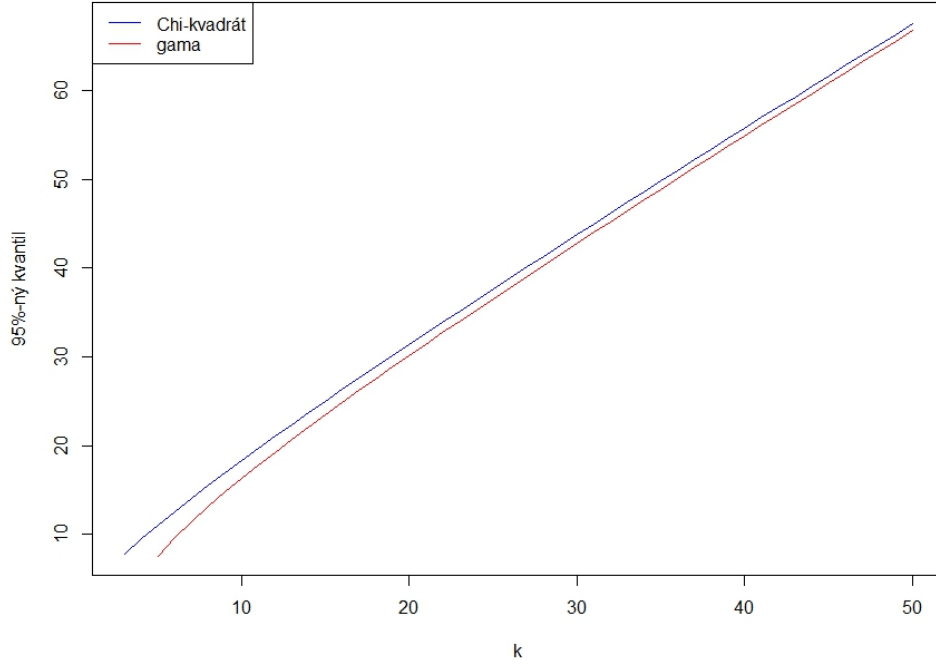
$$\alpha = \frac{k-p-q}{2}, \quad \beta = 2.$$

S využitím (4.7) dostávame parametre gama rozdelenia pre váhy w_i^Q ako

$$\alpha_Q = \frac{k^2}{2(k-p-q)}, \quad \beta_Q = \frac{2(k-p-q)}{k}. \quad (4.13)$$

Pre $k \rightarrow \infty$ sa α_Q správa ako α , analogicky β_Q sa pre $k \rightarrow \infty$ správa ako β . Vidíme teda, že $\Gamma(\alpha_Q, \beta_Q)$ je dobrou aproximáciou rozdelenia χ_{k-p-q}^2 . Na obrázku 4.2 je zobrazený graf 95%-ného kvantilu pre χ_k^2 rozdelenie a gama rozdelenie s parametrami (4.13) a $p = q = 2$. Môžeme vidieť, že pre zväčšujúce sa hodnoty k sú hodnoty týchto kvantilov približne rovnaké.

Aproximáciu rozdelenia váženej štatistiky Q_W pomocou gama rozdelenia budeme v práci používať len pre geometricky klesajúce váhy. Pre váhy založené na jadrových funkciách táto aproximácia nie je obecné vhodná, keďže pre niektoré kombinácie parametrov môžeme dostať záporné hodnoty parametrov α a β , viz nasledujúci príklad.



Obrázek 4.2: Graf 95%-ného kvantilu pre χ_k^2 rozdelenie a gama rozdelenie s parametrami (4.13) a $p = q = 2$ v závislosti na k .

Príklad 9. Nech $\{Y_t\}$ je časová rada riadiaca sa procesom ARMA(2, 1) s dĺžkou $n = 100$. Zvoľme $k = 20$. Nech Q_{W_R} je testová štatistika definovaná ako (4.2) s rovnomernými jadrovými váhami, pričom výberové autokorelačné funkcie tejto štatistiky sú spočítané z reziduí $\{e_t\}$. Ak by sme chceli aproximovať rozdelenie štatistiky Q_{W_R} pomocou gama rozdelenia, tak by sme získali parameter α definovaný ako (4.12) nasledovne:

$$\begin{aligned} \alpha &= \frac{\left(\sum_{i=1}^k w_i\right)^2}{2\left(\sum_{i=1}^k w_i^2 - p - q\right)} = \frac{\left(\sum_{i=1}^{20} \frac{100+2}{100-i} \cdot \frac{1}{4}\right)^2}{2\left(\sum_{i=1}^{20} \left(\frac{100+2}{100-i}\right)^2 \cdot \left(\frac{1}{4}\right)^2 - 2 - 1\right)} \\ &= \frac{\left(\frac{102}{4} \sum_{i=1}^{20} \frac{1}{100-i}\right)^2}{2\left(\frac{10404}{16} \sum_{i=1}^{20} \left(\frac{1}{100-i}\right)^2 - 2 - 1\right)}. \end{aligned}$$

Pomocou výpočetného softvéru dostaneme približnú hodnotu $\sum_{i=1}^{20} \frac{1}{100-i} \doteq 0,2244$ a $\sum_{i=1}^{20} \left(\frac{1}{100-i}\right)^2 \doteq 0,00253$, a teda $\alpha \doteq -12,0837$. Keďže parametre gama rozdelenia α, β musia byť kladné, tak v tomto prípade nemôžeme použiť aproximáciu zo sekcie 4.3.1. Podobný problém môže nastať aj pri použití ostatných jadrových funkcií z tabuľky 4.1.

5. Simulačné štúdie

V tejto kapitole budeme skúmať jednotlivé vážené testy pomocou simulačných štúdií. Pre vážené testy založené na geometricky klesajúcich váhach budeme uvažovať voľbu parametru a z množiny $\{0,4; 0,6; 0,8\}$. Prvá simulačná štúdia bude zameraná na *hladinu testu* a druhá na *silu testu*. V oboch simulačných štúdiách sú náhodne generované časové rady z príslušných ARMA modelov a biely šum je generovaný z t -rozdelenia s 10 stupňami voľnosti. Všetky simulácie boli prevedené pre 1000 iterácií. Výpočty boli prevedené pomocou programovacieho jazyka R (R Core Team (2021)). Zdrojový kód k simulačným štúdiám je v prílohe práce pod názvom **simulation_study.R**.

Kritický obor Boxova-Piercova a Ljung-Boxova testu je založený na (3.16). Kritické obory vážených testov v našej simulácii sme prispôbili jednotlivým kategóriám váh. Pre testy založené na geometricky klesajúcich váhach sme použili gama rozdelenie s parametrami (4.12). Pre väčšinu vážených testov s jadrovými váhami nebolo možné použiť gama aproximáciu, keďže pre niektoré kombinácie parametrov boli parametre α a β z (4.12) záporné (viz príklad 9). Pre testy s jadrovými váhami sme teda použili modifikovanú štatistiku z (4.6) a kritický obor založený na štandardnom normálnom rozdelení $N(0, 1)$. Všetky testy sú realizované na hladine významnosti $\alpha = 0,05$.

5.1 Hladina testu

V prvej simulačnej štúdii sa budeme zaoberať otázkou či je za platnosti nulovej hypotézy dodržaná hladina testov $\alpha = 0,05$. Hladinu testu môžeme zjednodušene chápať ako pravdepodobnosť chyby prvého druhu, teda pravdepodobnosť zamietnutia platnej hypotézy. Túto pravdepodobnosť odhadneme tak, že spočítame pomer zamietnutí platnej nulovej hypotézy. V tejto simulácii budeme náhodne generovať časové rady $\{Y_t\}$ zo stacionárneho AR(1) modelu v tvare

$$Y_t = \varphi_1 Y_{t-1} + \varepsilon_t,$$

kde $\{\varepsilon_t\}$ tvorí *iid* postupnosť so Studentovým t -rozdelením s 10 stupňami voľnosti. Pre tieto rady budeme odhadovať AR(1) model a jednotlivé testy použijeme na reziduá (3.2). Odhadnutú pravdepodobnosť chyby prvého druhu (resp. empirickú hladinu) budeme počítat pre rôzne prípady, konkrétne budeme meniť nasledujúce hodnoty:

- **Parameter φ_1 :** Budeme uvažovať parameter φ_1 z $\{0,1; 0,5; 0,9\}$.
- **Dĺžka simulovanej rady:** Simulácie prevedieme pre časové rady s dĺžkou $n \in [100, 1500]$.
- **Korekcia pre jadrové váhy:** Váhy založené na jadrových funkciách budeme uvažovať s tzv. *korekciou* $\frac{n+2}{n-i}$ a bez tejto korekcie. Budeme teda porovnávať empirickú hladinu testov s jadrovými váhami pre váhy

$$w_{i_1} = \frac{n+2}{n-i} \left[\kappa \left(\frac{i}{k} \right) \right]^2, \quad w_{i_2} = \left[\kappa \left(\frac{i}{k} \right) \right]^2. \quad (5.1)$$

- **Oneskorenie** k : Testy prevedieme pre rôzne hodnoty $k \in [1, 50]$.

5.1.1 Hladina testov pre rôzne hodnoty n

V prvej časti simulácie budeme počítat empirickú hladinu testov pre rôzne hodnoty dĺžky časovej rady n . Všetky testy boli realizované s oneskorením $k = 20$ a pre vážené testy s jadrovými váhami bola použitá korekcia $\frac{n+2}{n-i}$. Empirické hladiny testov uvádzame v tabuľkách 5.1, 5.2, 5.3, 5.4 a 5.5.

φ_1	Box-Pierce	Ljung-Box	Rovnomer.	Trojuhol.	Epanech.	Kvadr.
0,1	0,027	0,059	0,035	0,013	0,025	0,015
0,5	0,027	0,062	0,039	0,017	0,025	0,019
0,9	0,035	0,067	0,039	0,022	0,027	0,023
φ_1	Kubické	Gauss.	Daniellove	$a = 0,4$	$a = 0,6$	$a = 0,8$
0,1	0,016	0,036	0,013	0,010	0,016	0,022
0,5	0,018	0,040	0,018	0,025	0,020	0,026
0,9	0,020	0,043	0,023	0,103	0,061	0,039

Tabuľka 5.1: Empirické hladiny testov pre časové rady generované z AR(1) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 100$.

φ_1	Box-Pierce	Ljung-Box	Rovnomer.	Trojuhol.	Epanech.	Kvadr.
0,1	0,040	0,064	0,034	0,017	0,026	0,021
0,5	0,038	0,062	0,031	0,017	0,026	0,024
0,9	0,049	0,070	0,042	0,027	0,028	0,027
φ_1	Kubické	Gauss.	Daniellove	$a = 0,4$	$a = 0,6$	$a = 0,8$
0,1	0,020	0,031	0,014	0,015	0,028	0,027
0,5	0,020	0,023	0,016	0,030	0,030	0,030
0,9	0,029	0,033	0,033	0,116	0,086	0,058

Tabuľka 5.2: Empirické hladiny testov pre časové rady generované z AR(1) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 200$.

Ako prvé môžeme porovnať empirické hladiny Boxova-Piercova a Ljung-Boxova testu. Pre malý počet pozorovaní n je hladina Boxova-Piercova testu výrazne menšia a hladina Ljung-Boxova testu sa viac približuje teoretickej hladine $\alpha = 0,05$. So zvyšujúcim sa počtom pozorovaní n sa ale rozdiel v hladinách znižuje a oba testy majú prijateľnú empirickú hladinu. Túto skutočnosť môžeme pozorovať aj na obrázku 5.1. Z vážených testov s geometricky klesajúcimi váhami dodržiavajú hladinu najpresnejšie testy s vyššou hodnotou parametru a . Menšia voľba parametru sa neukázala ako vhodná z hľadiska dodržiavania hladiny. Napríklad empirická hladina v tabuľkách pre $\varphi_1 = 0,9$ a $a = 0,4$ je výrazne vyššia s porovnaním s empirickou hladinou ostatných testov. Z tohoto dôvodu budeme v simulačných štúdiách na hladinu testu skúmať len geometricky klesajúce váhy s $a = 0,8$. Z príslušných tabuliek si tiež môžeme všimnúť, že empirická hladina testov sa

φ_1	Box-Pierce	Ljung-Box	Rovnomer.	Trojuhol.	Epanech.	Kvadr.
0,1	0,044	0,048	0,040	0,014	0,023	0,015
0,5	0,045	0,047	0,039	0,020	0,026	0,022
0,9	0,048	0,055	0,044	0,023	0,027	0,024
φ_1	Kubické	Gauss.	Daniellove	$a = 0,4$	$a = 0,6$	$a = 0,8$
0,1	0,015	0,036	0,016	0,011	0,019	0,029
0,5	0,019	0,032	0,019	0,026	0,025	0,030
0,9	0,021	0,034	0,020	0,114	0,076	0,048

Tabulka 5.3: Empirické hladiny testov pre časové rady generované z AR(1) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 500$

φ_1	Box-Pierce	Ljung-Box	Rovnomer.	Trojuhol.	Epanech.	Kvadr.
0,1	0,034	0,038	0,031	0,014	0,021	0,021
0,5	0,034	0,036	0,026	0,016	0,023	0,020
0,9	0,035	0,037	0,035	0,023	0,031	0,026
φ_1	Kubické	Gauss.	Daniellove	$a = 0,4$	$a = 0,6$	$a = 0,8$
0,1	0,020	0,023	0,016	0,015	0,017	0,022
0,5	0,018	0,024	0,016	0,028	0,027	0,025
0,9	0,023	0,033	0,020	0,108	0,067	0,040

Tabulka 5.4: Empirické hladiny testov pre časové rady generované z AR(1) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 1000$.

zvyšuje, keď sa zvyšuje parameter φ_1 .

Pre prehľadnosť sme okrem tabuliek s presnými hodnotami použili aj grafickú vizualizáciu, viz obrázok 5.1. Tento graf obsahuje empirické hladiny testov pre jednu hodnotu $\varphi_1 = 0,1$ a pre počet pozorovaní n z intervalu $[100, 1500]$. Do grafu sme už zakomponovali iba jednu verziu váženého testu s geometricky klesajúcimi váhami s hodnotou parametru $a = 0,8$. Prerušovaná čiara na obrázku 5.1 vyznačuje hladinu $\alpha = 0,05$. Zo všetkých uvažovaných testov dosahoval hladinu najlepšie Ljung-Boxov test a Boxov-Piercov test. Z vážených testov dosahoval hladinu najlepšie test s rovnomernými jadrovými váhami (pre konkrétny model zvolený v simulácii). Prijateľne hladinu dosahoval aj vážený test s Gaussovskými váhami, geometricky klesajúcimi váhami s hodnotou parametru $a = 0,8$ a Epanechnikovými váhami. Testy s trojuholníkovými váhami a Daniellovými váhami boli v kontexte našej simulácie skôr konzervatívne.

5.1.2 Hladina testov pre rôzne hodnoty k

V druhej časti simulácie sa budeme zaoberať vplyvom zvoleného maximálneho oneskorenia k na hladinu testov. Znovu budeme uvažovať iba jednu verziu váženého testu s geometricky klesajúcimi váhami s $a = 0,8$. V tejto simulácii budeme generovať rady z AR(1) modelu s parametrom $\varphi_1 = 0,1$ a dĺžkou $n = 500$. Pre vážené testy s jadrovými váhami bola použitá korekcia $\frac{n+2}{n-i}$. Hodnoty oneskorení

φ_1	Box-Pierce	Ljung-Box	Rovnomer.	Trojuhol.	Epanech.	Kvadr.
0,1	0,042	0,043	0,033	0,021	0,028	0,025
0,5	0,042	0,045	0,034	0,023	0,025	0,027
0,9	0,048	0,049	0,036	0,028	0,036	0,032
φ_1	Kubické	Gauss.	Daniellove	$a = 0,4$	$a = 0,6$	$a = 0,8$
0,1	0,028	0,038	0,022	0,010	0,017	0,030
0,5	0,025	0,035	0,020	0,027	0,029	0,033
0,9	0,036	0,035	0,030	0,117	0,086	0,057

Tabulka 5.5: Empirické hladiny testov pre časové rady generované z AR(1) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 2000$.

k budeme uvažovať z intervalu $[1, 50]$.

Výsledky tejto simulácie sme zobrazili graficky na obrázku 5.2. Môžeme vidieť, že pre väčšinu testov sa empirická hladina zvyšuje s rastúcim k . Empirická hladina váženého testu s geometricky klesajúcimi váhami bola ako jediná skoro nezmenená. Rýchly nárast empirickej hladiny môžeme pozorovať pre vážený test s Gaussovskými váhami a rovnomernými váhami. V nasledujúcej simulačnej štúdiu zaoberajúcej sa silou testov sme tieto dva testy vyhodnotili ako najslabšie (vzhľadom ku konkrétnej voľbe parametrov v simulácii).

5.1.3 Hladina testov podľa korekcie

V tretej časti simulácie sme testovali hladinu vážených jadrových testov bez korekcie, teda používali sme tvar váh w_{i_2} z (5.1). Keď sme previedli simulácie z časti 5.1.1 bez použitia korekcie, tak výsledné hodnoty empirických hladín boli skoro identické. Túto skutočnosť ilustruje obrázok 5.3. Usudzujeme teda, že voľba korekcie nemá zásadný vplyv na empirickú hladinu testov.

5.2 Sila testu

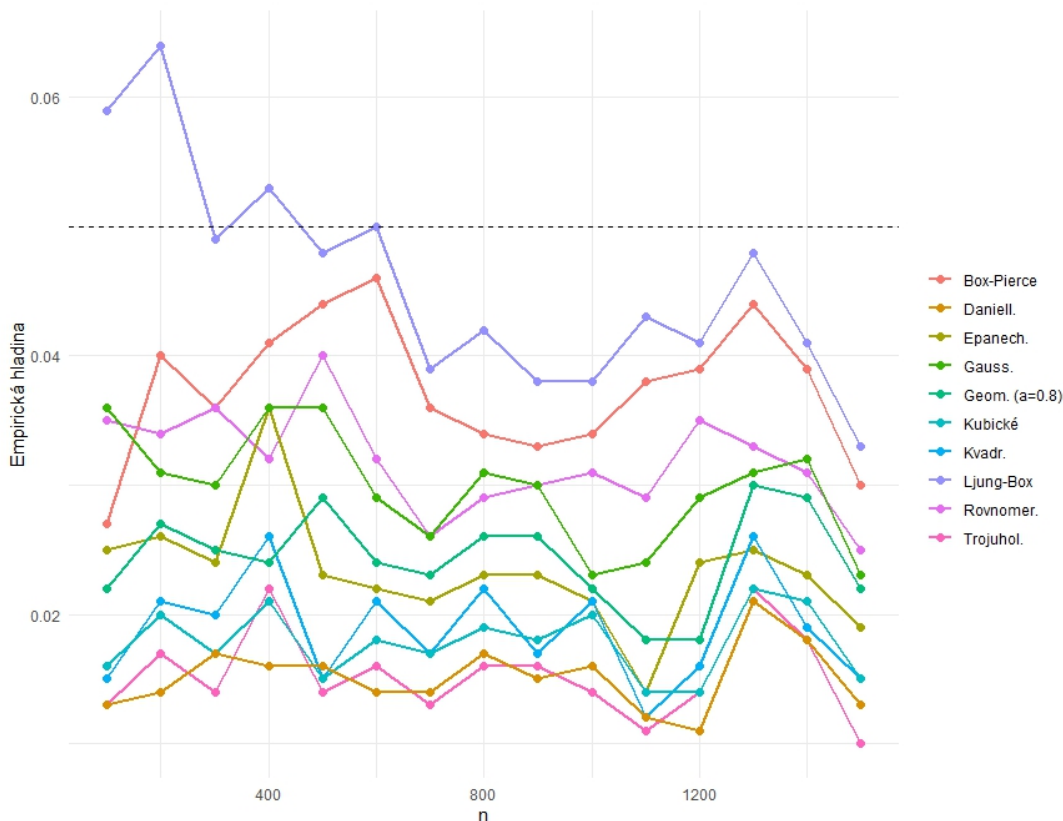
V druhej simulácii budeme skúmať silu jednotlivých testov. Sila testu predstavuje pravdepodobnosť zamietnutia neplatnej hypotézy pri konkrétnej alternatíve. Posudzovanie sily testov je pomerne komplexná úloha, keďže sila závisí na danej alternatíve, ktorú vyhodnocujeme. V tejto simulácii sa zameriame na alternatívu, v ktorej je nevhodne identifikovaný rád modelu. Budeme teda generovať časové rady z modelov AR(2) a ARMA(1, 1) a odhadovať ich nevhodným modelom AR(1).

V prvom prípade uvažujeme generované rady $\{Y_t\}$ zo stacionárneho modelu

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim t_{10},$$

pričom $\{\varepsilon_t\}$ tvorí postupnosť *iid* náhodných veličín so Studentovým rozdelením s 10 stupňami voľnosti. Budeme testovať nulovú hypotézu

$$H_0 : \{Y_t\} \text{ sa riadi modelom AR(1) s } iid \text{ inováciami.}$$



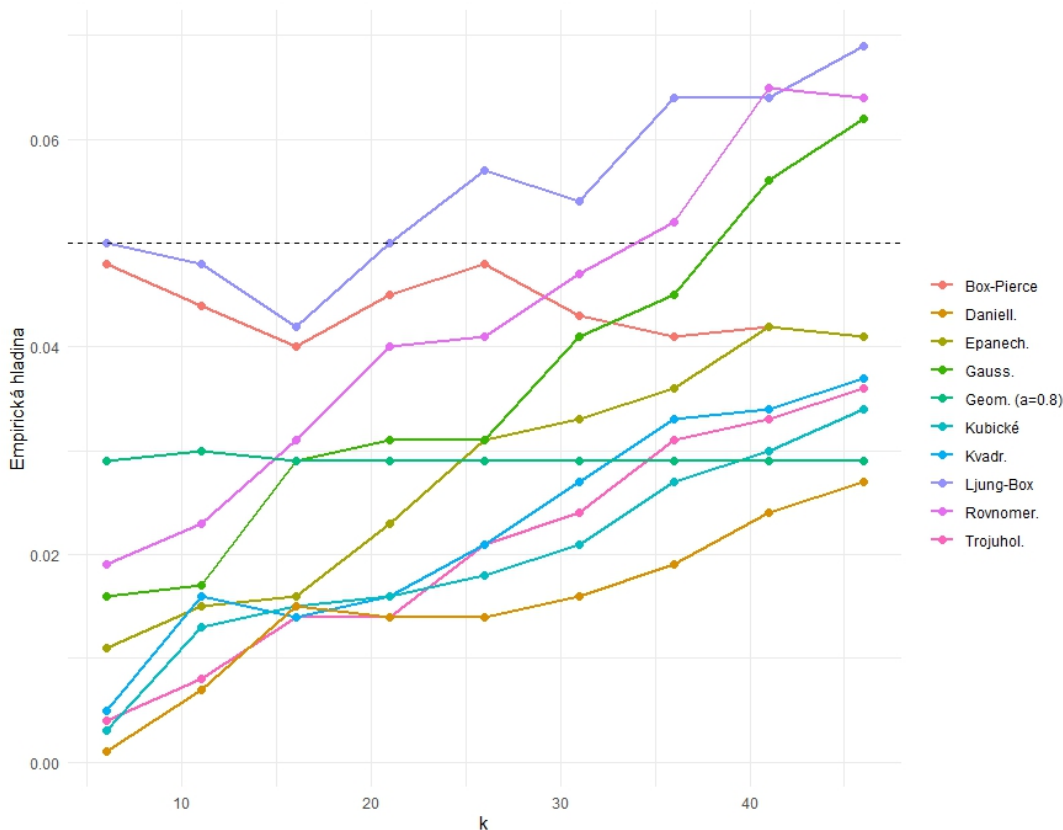
Obrázek 5.1: Empirické hladiny testov pre rôzne hodnoty n , časové rady generované z AR(1) modelu s $\varphi_1 = 0,1$ a odhadnuté ako AR(1), $k = 20$.

Pre rady generované zo stacionárneho modelu ARMA(1, 1) postupujeme analogicky. Silu testov budeme odhadovať pomocou pomeru zamietnutí neplatnej nulovej hypotézy. Pre vážené testy s jadrovými váhami bola v simuláciách použitá korekcia $\frac{n+2}{n-i}$. Budeme sa zaoberať najmä týmito faktormi:

- **Voľba parametru φ_2 :** Budeme uvažovať rôzne hodnoty parametru φ_2 pre rady generované z modelu AR(2).
- **Voľba modelu:** Budeme generovať rady z modelov AR(2) a ARMA(1, 1), aby sme mohli porovnať sily testov pre rôzne modely.
- **Dĺžka simulovanej rady:** Simulácie prevedieme pre časové rady s dĺžkou $n \in [100, 1000]$.
- **Oneskorenie k :** Simulácie prevedieme pre rôzne hodnoty $k \in [1, 50]$.

5.2.1 Sila testov podľa voľby ARMA parametrov

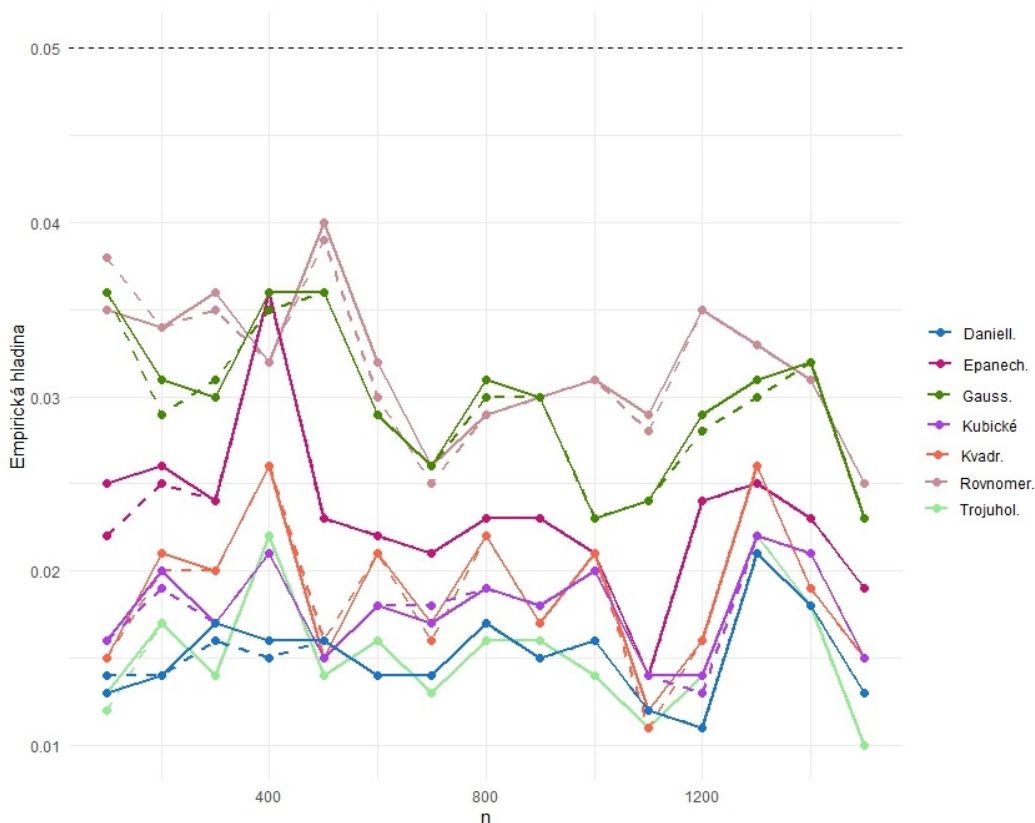
V prvej simulácii budeme generovať časové rady s dĺžkou $n = 100$ zo stacionárneho modelu AR(2). Budeme sa zaoberať voľbou parametrom φ_2 . V tejto simulácii budeme overovať či sa sila testov zväčšuje s rastúcou hodnotou φ_2 .



Obrázek 5.2: Empirické hladiny testov pre rôzne hodnoty k , časové rady generované z AR(1) modelu s $\varphi_1 = 0,1$ a odhadnuté ako AR(1), $n = 500$.

V tabuľke 5.6 nájdeme empirické sily testov pre konkrétne voľby parametrov φ_1 a φ_2 . V tejto tabuľke môžeme pozorovať rastúci trend so zvyšujúcou sa hodnotou φ_2 . Rastúci trend môžeme jasne vidieť aj na obrázku 5.4, ktorý predstavuje grafickú vizualizáciu simulácie z tabuľky 5.6. Podobnú simuláciu sme vytvorili aj pre záporné hodnoty parametrov φ_2 a výsledky môžeme nájsť na obrázku 5.5. Zo simulácií je zrejmé, že sila testov sa zväčšuje, ak sa zväčšuje $|\varphi_2|$. Neformálne môžeme tento výsledok chápať tak, že čím viac je testovaná rada „vzdialenejšia“ od nulovej hypotézy (v zmysle veľkosti parametru φ_2), tým väčšiu silu majú uvažované testy.

Zo simulácií môžeme vidieť, že pre voľbu parametru $\varphi_2 \in [-0,2; 0,2]$ je najsilnejší Ljung-Boxov test, no pre ostatné voľby parametru φ_2 je silnejší test s Daniellovými váhami a geometricky klesajúcimi váhami s parametrom $a = 0,8$. Pre voľbu $|\varphi_2| > 0,35$ sú oproti Ljung-Boxovu testu silnejšie aj testy s trojuholníkovými, kubickými alebo kvadratickými váhami. Môžeme teda vidieť, že pre niektoré voľby parametrov φ_1 a φ_2 sú mnohé vážené testy potenciálne silnejšie než často používaný Ljung-Boxov test. Zo simulácií sú tieto testy silnejšie najmä pre testované rady, ktoré majú väčšiu „vzdialenosť“ od nulovej hypotézy. Vo všetkých uvedených simuláciách v tejto sekcii môžeme vidieť, že Ljung-Boxov test má väčšiu silu než Boxov-Piercov test.



Obrázek 5.3: Empirické hladiny testov pre rôzne hodnoty n pre testy s korekciou (plné čiary) a testy bez korekcie (prerušované čiary), časové rady generované z AR(1) modelu s $\varphi_1 = 0,1$ a odhadnuté ako AR(1), $n = 500$.

5.2.2 Sila testov pre rôzne hodnoty n

V druhej simulácii sa budeme zaoberať otázkou či sa sila testov zväčšuje, ak sa zväčšuje dĺžka simulovanej časovej rady. V tejto simulácii budeme generovať rady z ARMA(1,1) modelu s parametrami $\varphi_1 = 0,5$, $v_2 = 0,2$, tieto rady budeme odhadovať nevhodným modelom AR(1) a postupne budeme vyhodnocovať empirické sily testov pre n z intervalu $[100, 1000]$. Výsledky simulácie vidíme na obrázku 5.6. Môžeme pozorovať jasný rastúci trend pre zväčšujúce sa n . Podobne ako v predošlej simulácii sú aj v tomto prípade najsilnejšie testy vážený test s Daniellovými váhami a s geometricky klesajúcimi váhami s parametrom $a = 0,8$.

Môžeme si všimnúť, že pre váženú testovú štatistiku s rovnomernými váhami Q_{WR} platí $Q_{WR} = \frac{1}{4}Q^*$. Vážený test s rovnomernými váhami je teda ekvivalentný s Ljung-Boxovým testom. V našich simuláciách sú ale kritické obory oboch testov založené na inom rozdelení a empirická sila Ljung-Boxova testu je vyššia ako sila váženého testu s rovnomernými váhami. Jedným z možných vysvetlení je skutočnosť, že asymptotické rozdelenie (4.6) platí pre $k \rightarrow \infty$, no v simuláciách sme pracovali s fixným k .

φ_1	φ_2	Box-Pierce	Ljung-Box	Rovnom.	Trojuhol.	Epan.	Kvadr.
0,3	0,10	0,054	0,102	0,045	0,033	0,037	0,032
0,3	0,20	0,120	0,178	0,084	0,128	0,117	0,130
0,3	0,30	0,282	0,396	0,189	0,393	0,320	0,375
0,3	0,40	0,592	0,696	0,456	0,763	0,684	0,745
0,3	0,50	0,853	0,895	0,779	0,949	0,921	0,941
0,3	0,60	0,974	0,983	0,948	0,993	0,987	0,990
φ_1	φ_2	Kubické	Gauss.	Daniell.	$a = 0,4$	$a = 0,6$	$a = 0,8$
0,3	0,10	0,033	0,045	0,035	0,018	0,037	0,041
0,3	0,20	0,138	0,096	0,142	0,117	0,166	0,163
0,3	0,30	0,411	0,259	0,450	0,472	0,544	0,502
0,3	0,40	0,775	0,598	0,799	0,837	0,861	0,829
0,3	0,50	0,952	0,868	0,960	0,973	0,979	0,970
0,3	0,60	0,992	0,978	0,995	0,997	0,998	0,997

Tabulka 5.6: Empirické sily testov pre časové rady generované z AR(2) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 100$.

5.2.3 Sila testov pre rôzne hodnoty k

V tejto simulácii budeme vyhodnocovať empirickú silu testov pre rôzne hodnoty oneskorenia k . Budeme generovať časové rady z ARMA(1, 1) modelu s parametrami $\varphi_1 = 0,6$ a $v_1 = 0,3$, dĺžkou $n = 200$ a oneskorenie k budeme uvažovať do hodnoty 50. Tieto rady sme odhadovali modelom AR(1). Výsledky simulácie nájdeme na obrázku 5.7. Môžeme pozorovať klesajúci trend empirických síl testov so zväčšujúcim sa k . Na základe výsledkov tejto simulácie usudzujeme, že menšia hodnota k môže zabezpečiť potenciálne silnejší test. Empirická sila testu s geometricky klesajúcimi váhami je v tejto simulácii skoro konštantná pre všetky hodnoty k . Táto skutočnosť závisí aj na voľbe parametru a , čím sa detailnejšie zaoberáme v Appendixe na obrázkoch A.1 a A.2.

5.2.4 Sila testov podľa korekcie

V poslednej simulácii sa budeme zaoberať vplyvom korekcie $\frac{n+2}{n-i}$ na silu vážených testov s jadrovými váhami. Simulácia je prevedená pre časové rady generované z ARMA(1, 1) modelu s parametrami $\varphi_1 = 0,6$, $v_1 = 0,3$ a odhadnuté pomocou AR(1) modelu. V tabuľke 5.7 nájdeme výsledky tejto simulácie. Môžeme si všimnúť, že empirické sily testov bez použitia korekcie sú o niečo vyššie, no tento rozdiel je veľmi malý. Z tejto simulácie teda nemôžeme usudzovať, že by mali vážené jadrové testy vyššiu silu s použitím korekcie $\frac{n+2}{n-i}$.

5.2.5 Zhrnutie výsledkov

V tejto kapitole sme skúmali hladinu a silu vážených testov v závislosti na rôznych faktoroch, ktoré by ich mohli ovplyvňovať. Empirická hladina testov sa síce zvyšovala a približovala požadovanej hladine $\alpha = 0,05$ pre rastúce k , no sila testov sa pre rastúce k pri fixnom n znižovala. Môžeme teda usúdiť, že je vhodnejšie pou-

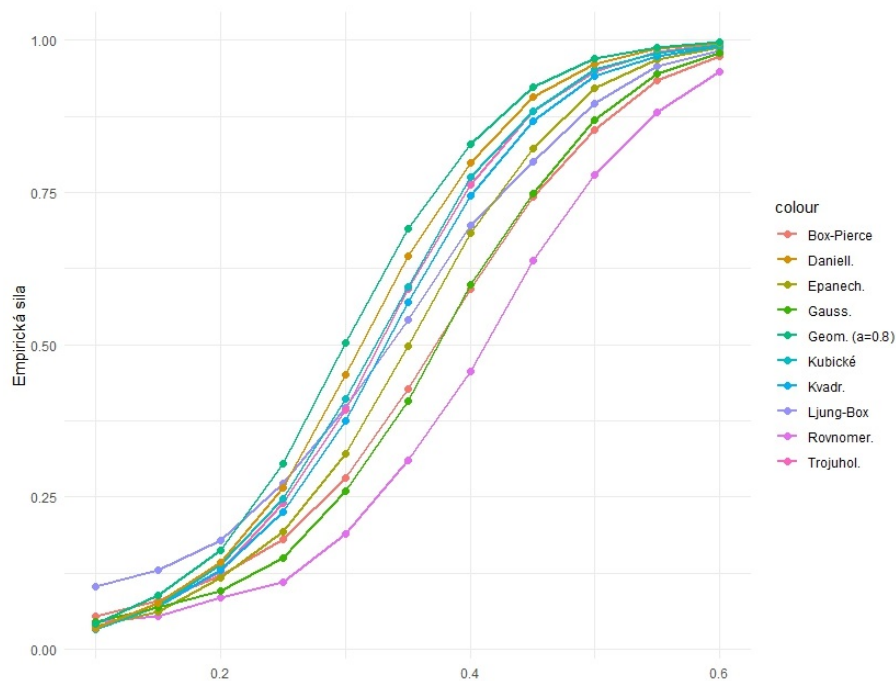
Bez korekcie							
k	Rovnom.	Trojuhol.	Epan.	Kvadr.	Kubické	Gauss.	Daniell.
10	0,429	0,657	0,592	0,641	0,669	0,521	0,698
20	0,307	0,567	0,473	0,523	0,562	0,421	0,601
30	0,230	0,498	0,395	0,460	0,490	0,309	0,528
40	0,196	0,450	0,329	0,396	0,443	0,252	0,481
50	0,151	0,399	0,282	0,345	0,386	0,217	0,447

S korekciou							
k	Rovnom.	Trojuhol.	Epan.	Kvadr.	Kubické	Gauss.	Daniell.
10	0,421	0,654	0,589	0,642	0,667	0,516	0,699
20	0,295	0,562	0,471	0,519	0,559	0,395	0,596
30	0,211	0,487	0,383	0,457	0,483	0,290	0,525
40	0,177	0,442	0,308	0,382	0,436	0,239	0,472
50	0,131	0,377	0,264	0,330	0,371	0,198	0,434

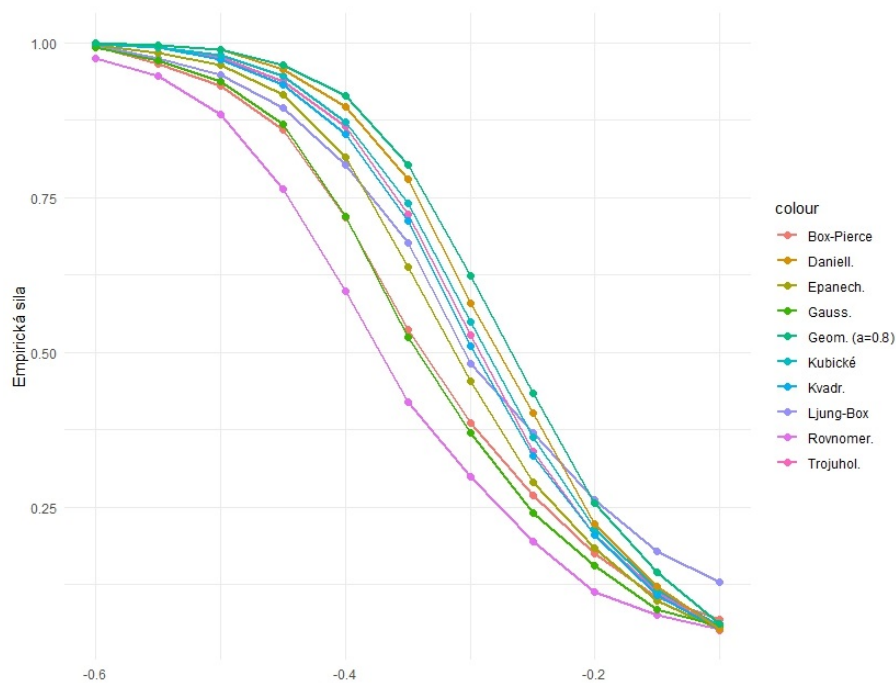
Tabulka 5.7: Empirické sila testov pre časové rady generované z ARMA(1,1) modelu s parametrami $\varphi_1 = 0,6$, $v_1 = 0,3$ a odhadnuté ako AR(1) model pomocou vážených jadrových testov s korekciou a bez korekcie, $n = 200$.

žiť menšie maximálne oneskorenie k . V praxi sa odporúča voliť k v závislosti na n , ako napríklad $k = \frac{n}{3}$ alebo $k = \frac{n}{4}$, viz (Box a kol., 2016, str. 31). Rastúca hodnota n síce nemala zásadný vplyv na hladinu testov, no sila testov značne stúpala. Môžeme teda vyvodit, že jednotlivé testy budú silnejšie, ak ich budeme používať pre časové rady s dostatočne veľkým počtom pozorovaní. Použitie korekcie pre vážené testy s jadrovými váhami nemalo signifikantný efekt na hladinu alebo silu testov. Z jednotlivých simulácií môžeme pozorovať, že testy s najväčšou empirickou silou sú vážený test s Daniellovými a geometricky klesajúcimi váhami. Túto skutočnosť môžeme vidieť pre rady generované z modelov AR(2) aj AR(1,1). V simuláciách boli najsilnejšie vážené testy s geometricky klesajúcimi váhami s vyššou hodnotou parametru a , preto môžeme odporučiť voľbu $a = 0,8$ alebo $a = 0,9$. Empirická hladina testu s geometricky klesajúcimi váhami bola prijateľná, no vážený test s Daniellovými váham bol skôr konzervatívny. V mnohých situáciách bola empirická sila týchto testov vyššia než sila známeho Ljung-Boxova testu. Musíme ale podotknúť, že Ljung-Boxov test dodržiaval v našich simuláciách hladinu o niečo presnejšie než test s Daniellovými a geometrickými váhami. Aj napriek tomu sa domnievame, že tieto dva vážené testy by mohli predstavovať potenciálne silnejšie testy než Ljung-Boxov test a mohli by sa stať vhodnou alternatívou v praktických aplikáciách.

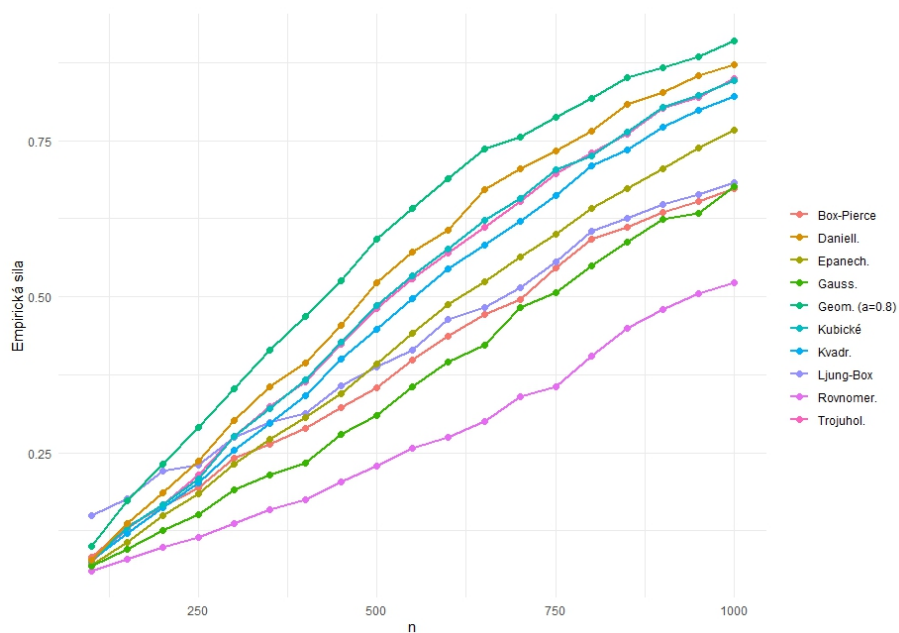
V simuláciách môžeme pozorovať, že empirická sila váženého testu s rovnomernými váhami je o niečo nižšia než empirická sila Ljung-Boxova testu aj napriek tomu, že sa jedná o ekvivalentné testy. Toto môže byť spôsobené voľbou kritického oboru pre vážené testy, ktorý sme zostrojili na základe vety 8. Domnievame sa, že v ďalšom výzkuve by bolo možné nájsť vhodnú aproximáciu rozdelenia veličiny z (4.5) vo vete 7, podobne ako sme aproximovali rozdelenie za vetou 6.



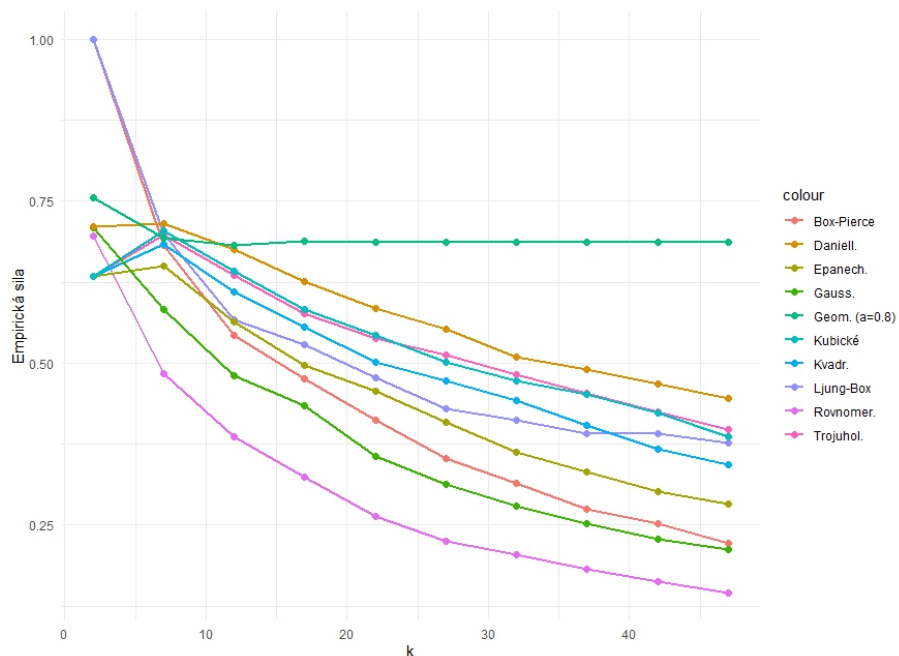
Obrázek 5.4: Empirické síla testov pre časové rady generované z AR(2) modelu s parametrami $\varphi_1 = 0,3$ a $\varphi_2 \in [0,1;0,6]$ (x -ová os), odhadnuté ako AR(1), $k = 20$, $n = 100$.



Obrázek 5.5: Empirické síla testov pre časové rady generované z AR(2) modelu s parametrami $\varphi_1 = 0,3$ a $\varphi_2 \in [-0,6; -0,1]$ (x -ová os), odhadnuté ako AR(1), $k = 20$, $n = 100$.



Obrázek 5.6: Empirické síly testů pro časové řady generované z ARMA(1,1) modelu s parametry $\varphi_1 = 0,5$ a $v_1 = 0,2$, odhadnuté jako AR(1), $k = 20$, $n \in [100, 1000]$.



Obrázek 5.7: Empirická síla testů pro časové řady generované z ARMA(1,1) modelu s parametry $\varphi_1 = 0,6$, $v_1 = 0,3$ a odhadnuté jako AR(1), $n = 200$, $k \in [1, 50]$.

Záver

Hlavným cieľom tejto práce bolo prehľadne popísať vážené portmanteau testy pre určenie adekvátnosti odhadnutého modelu ARMA a tieto testy porovnať s Boxovým-Piercovým a Ljung-Boxovým testom. V teoretickej časti sme sa zaoberali najmä podrobným popisom vážených a nevážených portmanteau testov a asymptotickým rozdelením ich testových štatistík. V praktickej časti sme testovali hladinu a silu jednotlivých testov pomocou simulačných štúdií.

Na začiatku práce sme sa zaoberali odhadovaním parametrov ARMA modelu pomocou metódy maximálnej vierohodnosti. Taktiež sme opísali asymptotické rozdelenie výberových autokorelácií, ktoré sme využili ďalej v práci. V kapitole o nevážených portmanteau testoch sme sa podrobne venovali Boxovmu-Piercovmu a Ljung-Boxovmu testu. Venovali sme sa aproximácii asymptotického rozdelenia testových štatistík, ktoré sme doplnili pomocou niekoľkých príkladov a obrázkov.

V kapitole o vážených portmanteau testoch sme zaviedli váženú testovú štatistiku Q_W a predstavili dve kategórie vážiacich schém. Uvideli sme aj niekoľko nových váh založených na jadrovej funkcii, ktoré sme neskôr testovali pomocou simulačných štúdií. Taktiež sme sa venovali asymptotickému rozdeleniu váženej testovej štatistiky, ktoré sme určovali (alebo aproximovali) iným spôsobom pre jadrové váhy a pre geometrické váhy.

V praktickej časti sme aplikovali jednotlivé testy na simulované dáta. Najväčšiu empirickú silu mali v simuláciách vážené testy s Daniellovými váhami a geometrickými váhami s parametrom $a = 0,8$. Tieto vážené testy síce nedodržiavali hladinu α s takou presnosťou ako Boxov-Piercov alebo Ljung-Boxov test, no ich sila bola v simuláciach vyššia než sila Boxova-Piercova a Ljung-Boxova testu. Aj napriek tomu považujeme vážené testy s Daniellovými a geometrickými váhami za možnú alternatívu k známemu Boxovmu-Piercovmu a Ljung-Boxovmu testu. Veríme, že problematika vážených portmanteau testov stále poskytuje priestor pre ďalší výzkum, najmä pri voľbe vhodných váh alebo určovaní rozdelenia testových štatistík pre konkrétne vážiace schémy.

Zoznam použitej literatúry

- BOX, G. E. P. a PIERCE, D. A. (1970). Distribution of residual autocorrelations in autoregressive- integrated moving average time series models. *Journal of the American Statistical Association*, Dec., 1970, **65**(332), 1509–1526.
- BOX, G. E. P., JENKINS, G. M., REINSEL, G. C. a LJUNG, G. M. (2016). *Time Series analysis*. by John Wiley and Sons, New Jersey, fifth edition edition. ISBN 978-1-118-67502-1.
- BROCKWELL, P. J. a DAVIS, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York, second edition edition. ISBN 0-387-97429-6.
- CIPRA, T. (2003). *Finanční ekonometrie*. Ekopress, Praha. ISBN 978-80-86929-93-4.
- FAN, J. a YAO, Q. (2003). *Nonlinear Time Series, Nonparametric and Parametric Methods*. Springer, New York. ISBN 0-387-95170-9.
- FISHER, T. J. a GALLAGHER, C. M. (2012). New weighted portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association*, **107**(498), 777–787.
- GALLAGHER, C. M. a FISHER, T. J. (2015). On weighted portmanteau tests for time-series goodness-of-fit. *Journal of Time Series Analysis*, **36**(1), 67–83.
- OMELKA, M. (2021). *NMSA331 Matematická statistika 1, Poznámky k přednášce*. Praha. URL https://www.karlin.mff.cuni.cz/~omelka/Vyuka_nmsa331_2324.php.
- PRÁŠKOVÁ, Z. a LACHOUT, P. (2001). *Základy náhodných procesů*. Nakladatelství Karolinum, Praha. ISBN 80-7184-688-0.
- R CORE TEAM (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- RACINE, J. S. (2008). *Nonparametric Econometrics: A Primer*. now Publishers Inc., Hanover, MA 02339, USA. ISBN 978-1-60198-110-3.

Zoznam obrázkov

1.1	Výberová a teoretická autokorelačná funkcia časovej rady z modelu AR(1) a parametrom $\varphi_1 = 0,7$	8
1.2	Výberová a teoretická autokorelačná funkcia časovej rady z modelu AR(1) a parametrom $\varphi_1 = -0,7$	8
2.1	Postupnosť bieleho šumu IID(0, 1) a jej korelogram.	16
3.1	Graf vlastných čísel $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ v závislosti na parametre φ	23
3.2	Empirické distribučné funkcie náhodnej veličiny $\sum_{i=1}^4 \lambda_i Y_i^2$ pre hodnoty parametre φ z množiny $\{0,80; 0,87; 0,91; 0,96; 0,99\}$, pričom $\lambda_i, i = 1, \dots, 4$ sú vlastné čísla matice (3.12) a $Y_i, i = 1, \dots, 4$ sú nezávislé náhodné veličiny s rozdelením $N(0, 1)$. Modrá prerušovaná čiara označuje 95%-ný kvantil príslušných rozdelení.	24
3.3	Kvantilová funkcia χ^2 rozdelenia pre rôzne počty stupňov voľnosti.	26
4.1	Graf jadrových funkcií na intervale $[-1, 1]$	31
4.2	Graf 95%-ného kvantilu pre χ_k^2 rozdelenie a gama rozdelenie s parametrami (4.13) a $p = q = 2$ v závislosti na k	34
5.1	Empirické hladiny testov pre rôzne hodnoty n , časové rady generované z AR(1) modelu s $\varphi_1 = 0,1$ a odhadnuté ako AR(1), $k = 20$	39
5.2	Empirické hladiny testov pre rôzne hodnoty k , časové rady generované z AR(1) modelu s $\varphi_1 = 0,1$ a odhadnuté ako AR(1), $n = 500$	40
5.3	Empirické hladiny testov pre rôzne hodnoty n pre testy s korekciou (plné čiary) a testy bez korekcie (prerušované čiary), časové rady generované z AR(1) modelu s $\varphi_1 = 0,1$ a odhadnuté ako AR(1), $n = 500$	41
5.4	Empirické sila testov pre časové rady generované z AR(2) modelu s parametrami $\varphi_1 = 0,3$ a $\varphi_2 \in [0,1; 0,6]$ (x -ová os), odhadnuté ako AR(1), $k = 20, n = 100$	44
5.5	Empirické sila testov pre časové rady generované z AR(2) modelu s parametrami $\varphi_1 = 0,3$ a $\varphi_2 \in [-0,6; -0,1]$ (x -ová os), odhadnuté ako AR(1), $k = 20, n = 100$	44
5.6	Empirické sila testov pre časové rady generované z ARMA(1, 1) modelu s parametrami $\varphi_1 = 0,5$ a $\nu_1 = 0,2$, odhadnuté ako AR(1), $k = 20, n \in [100, 1000]$	45
5.7	Empirická sila testov pre časové rady generované z ARMA(1, 1) modelu s parametrami $\varphi_1 = 0,6, \nu_1 = 0,3$ a odhadnuté ako AR(1), $n = 200, k \in [1, 50]$	45
A.1	Empirická sila testov pre časové rady generované z ARMA(1, 1) modelu s parametrami $\varphi_1 = 0,6, \nu_1 = 0,3$ a odhadnuté ako AR(1), $n = 200, k \in [1, 50]$	52

A.2	Empirická sila testov pre časové rady generované z AR(2) modelu s parametrami $\varphi_1 = 0,6$, $\varphi_2 = -0,3$ a odhadnuté ako AR(1), $n = 200$, $k \in [1, 50]$	52
-----	---	----

Zoznam tabuliek

4.1	Príklady jadrových funkcií.	31
5.1	Empirické hladiny testov pre časové rady generované z AR(1) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 100$	36
5.2	Empirické hladiny testov pre časové rady generované z AR(1) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 200$	36
5.3	Empirické hladiny testov pre časové rady generované z AR(1) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 500$	37
5.4	Empirické hladiny testov pre časové rady generované z AR(1) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 1000$	37
5.5	Empirické hladiny testov pre časové rady generované z AR(1) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 2000$	38
5.6	Empirické sila testov pre časové rady generované z AR(2) modelu a odhadnuté ako AR(1) model, $k = 20$, $n = 100$	42
5.7	Empirické sila testov pre časové rady generované z ARMA(1,1) modelu s parametrami $\varphi_1 = 0,6$, $v_1 = 0,3$ a odhadnuté ako AR(1) model pomocou vážených jadrových testov s korekciou a bez korekcie, $n = 200$	43

A. Appendix

Lemma 9. *Nech $\mathbf{X} = (X_1, \dots, X_k)^\top$ a $\mathbf{X} \sim N_k(\mathbf{0}, \Sigma)$. Potom náhodná veličina $\sum_{i=1}^k X_i^2$ má rovnaké rozdelenie ako náhodná veličina $\sum_{i=1}^k \lambda_i Y_i^2$, kde $\lambda_1, \dots, \lambda_k$ sú vlastné čísla matice Σ a Y_1, \dots, Y_k sú nezávislé náhodné veličiny s rozdelením $N(0, 1)$.*

Ak je matica Σ idempotentná, tak má náhodná veličina $\sum_{i=1}^k X_i^2$ rozdelenie chí-kvadrát, pričom počet stupňov voľnosti sa rovná hodnosti matice Σ .

Dôkaz. Viz (Omelka, 2021, Appendix). □

Lemma 10. *Nech pre náhodnú veličinu X_n platí $X_n \sim \chi_n^2, n \in (0, \infty)$. Potom*

$$\frac{X_n - n}{\sqrt{2n}} \xrightarrow[n \rightarrow \infty]{D} N(0, 1).$$

Dôkaz. Keďže pre X_n platí $X_n \sim \chi_n^2$, tak náhodnú veličinu X_n môžeme zapísať ako

$$X_n = \sum_{i=1}^n Z_i^2,$$

pričom Z_1, \dots, Z_n sú nezávislé a rovnako rozdelené náhodné veličiny s rozdelením $N(0, 1)$. Potom aj Z_1^2, \dots, Z_n^2 tvoria nezávislé a rovnako rozdelené náhodné veličiny a z vlastností normálneho rozdelenia dostávame

$$\mathbb{E}(Z_1^2) = 1, \quad \text{Var}(Z_1^2) = 2.$$

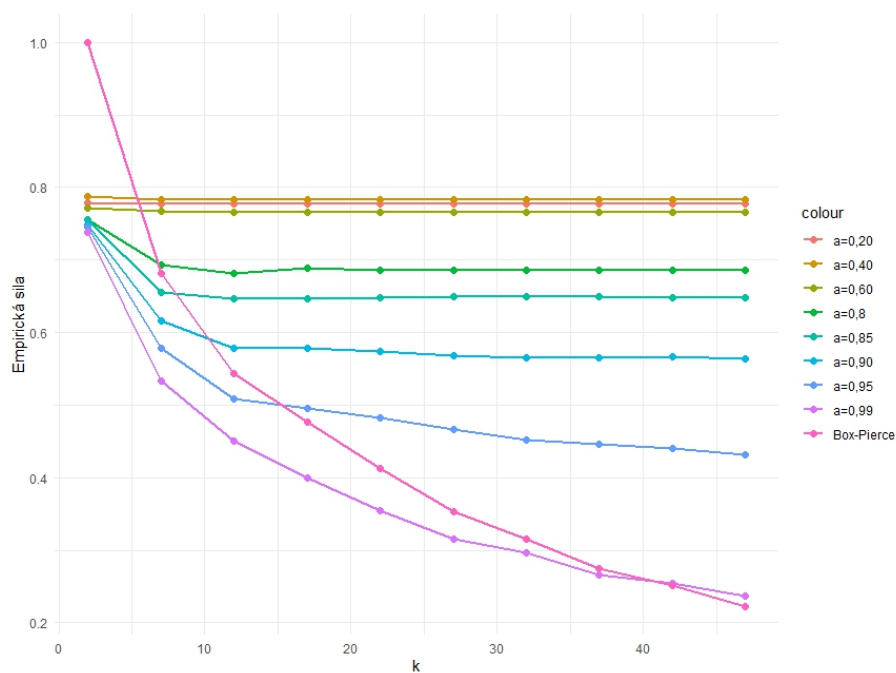
Aplikovaním Centrálnej limitnej vety na postupnosť Z_1^2, Z_2^2, \dots dostávame

$$\frac{\sum_{i=1}^n (Z_i^2 - 1)}{\sqrt{2n}} = \frac{\sum_{i=1}^n Z_i^2 - n}{\sqrt{2n}} = \frac{X_n - n}{\sqrt{2n}} \xrightarrow[n \rightarrow \infty]{D} N(0, 1).$$

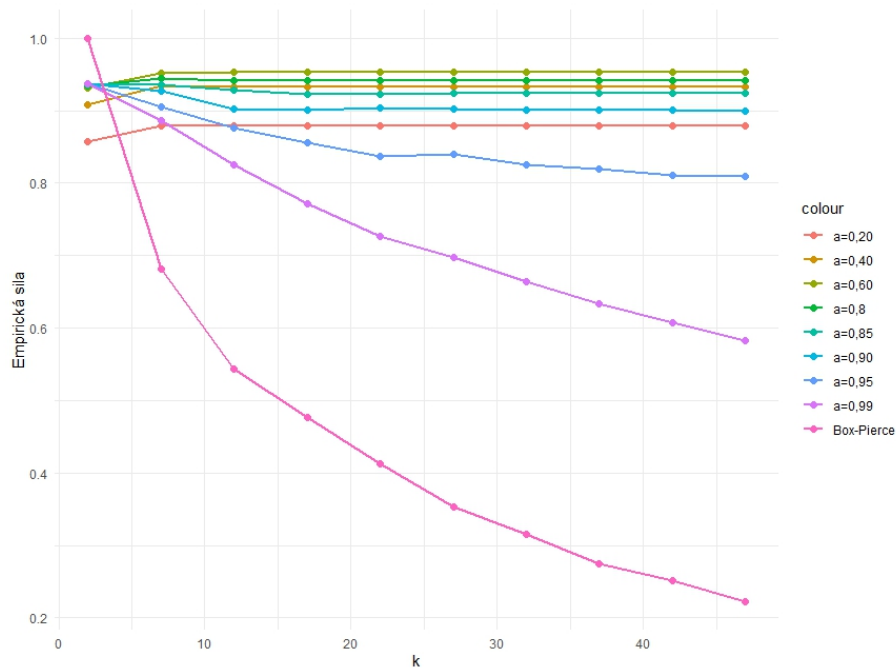
□

Sila vážených testov s geometricky klesajúcimi váhami v závislosti na k

Na obrázku 5.7 môžeme vidieť, že empirická sila váženého testu s geometricky klesajúcimi váhami s parametrom $a = 0,8$ sa skoro nemení s rastúcim k . Pozrieme sa či dostaneme podobný výsledok aj pre iné voľby parametru a . Na obrázku A.1 môžeme vidieť simuláciu s rovnakými parametrami ako v simulácii z obrázku 5.7, ale pre vážené testy s geometrickými váhami pre rôzne parametry a . Pridali sme aj empirickú silu Boxova-Piercova testu, keďže Boxov-Piercov test odpovedá váženému testu s geometrickými váhami s parametrom $a = 1$ a bez člena $(p + q)$ v definícii geometrických váh. Môžeme vidieť, že pre menšie hodnoty a je empirická sila skoro konštantná, no pre vyššie hodnoty a sa tvar empirickej sily v závislosti na k viac podobá tvaru pre Boxov-Piercov test. Sklon kriviek s empirickou hladinou závisí aj na voľbe modelu, z ktorého rady simulujeme, ako môžeme vidieť na obrázku A.2.



Obrázek A.1: Empirická síla testů pro časové řady generované z ARMA(1,1) modelu s parametry $\varphi_1 = 0,6$, $v_1 = 0,3$ a odhadnuté jako AR(1), $n = 200$, $k \in [1, 50]$.



Obrázek A.2: Empirická síla testů pro časové řady generované z AR(2) modelu s parametry $\varphi_1 = 0,6$, $\varphi_2 = -0,3$ a odhadnuté jako AR(1), $n = 200$, $k \in [1, 50]$.