



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Bc. Pavel Zdeněk

Testy nezávislosti dvou časových řad

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Zbyněk Pawlas, Ph.D.

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2023

Prohlašuji, že jsem tuto diplomovou práci vypracoval 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 Praze dne 19. července 2023

Pavel Zdeněk

Rád bych poděkoval svému vedoucímu, doc. RNDr. Zbyňku Pawlasovi, Ph.D., za všechny konzultace, poznámky, kontroly a veškerý čas, který mi při psaní diplomové práce věnoval.

Dále bych chtěl moc poděkovat své rodině, která mě během celého studia podporovala.

Název práce: Testy nezávislosti dvou časových řad

Autor: Pavel Zdeněk

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Zbyněk Pawlas, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Cílem této diplomové práce je představit několik testů nezávislosti pro časové řady řídicí se modelem ARMA a následně je mezi sebou v rámci simulační studie porovnat. Na začátku je připomenuta základní teorie nezávislosti spolu s kovariancí a korelací. Pro výběrový odhad křížové kovariance jsou formulovány a dokázány asymptotická nestrannost a konzistence, která je následně dokázána i pro korelaci. Po představení modelů ARMA jsou postupně popsány a diskutovány následující testy: Haughův test, využívající odhady bílých šumů a výběrové křížové korelace, modifikovaný t-test, pro který namísto náhodných výběrů předpokládáme slabě stacionární řady, a jako poslední test distanční kovariancí, využívající vlastností charakteristických funkcí. V simulační části jsou tyto testy porovnány spolu s testem nezávislosti pomocí Pearsonova korelačního koeficientu. Na závěr je prezentován ilustrativní příklad na finančních datech.

Klíčová slova: testy nezávislosti, časové řady, ARMA model, kovariance

Title: Tests of independence between two time series

Author: Pavel Zdeněk

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Zbyněk Pawlas, Ph.D., Department of Probability and Mathematical Statistics

Abstract: The goal of this diploma thesis is to introduce several tests of independence for time series following the ARMA model and then compare them within the simulation study. First, the basic theory of independence is reminded together with covariance and correlation. Asymptotic unbiasedness and consistency are derived for sample cross-covariance and also consistency for correlation. After the introduction of the ARMA model, each test is described and its advantages and disadvantages discussed. The following tests are included: Haugh test, using estimates of white noise and sample cross-correlation, modified t-test, for which we assume weakly stationary series instead of random samples, and lastly distance covariance test, which uses properties of characteristic functions. These tests are compared in the simulation study together with the standard independence test using Pearson correlation coefficient. At the end, an illustrative example with finance data is presented.

Keywords: tests of independence, time series, ARMA model, covariance

Obsah

Úvod	2
1 Základní definice a poznatky	3
1.1 Nezávislost	3
1.2 Kovariance a korelace	6
1.3 Výběrové křížové kovariance a korelace	8
1.4 ARMA model	12
1.5 Charakteristická funkce	16
2 Haughův test	18
2.1 Úvod	18
2.2 Konstrukce ARMA modelů	19
2.3 Test	20
2.4 Podobné testy	22
3 Modifikovaný t-test	24
3.1 Úvod	24
3.2 Test	25
4 Test distanční kovariancí	27
4.1 Úvod	27
4.2 Teorie distanční kovariance	27
4.3 Výběrová distanční kovariance	29
4.4 Asymptotické vlastnosti výběrové distanční kovariance	30
4.5 Test	32
5 Porovnání testů	34
5.1 Data	34
5.2 Porovnávané testy	41
5.3 Porovnání	41
6 Ilustrativní příklad na reálných datech	50
6.1 Popis dat	50
6.2 Použití testů nezávislosti	52
Závěr	54
Literatura	56
A Příloha	58

Úvod

V této práci se zaměříme na různé testy nezávislosti. Nezávislost je jedním ze základních vztahů mezi náhodnými veličinami. Dovoluje snadno získat sdružené rozdělení vektoru pomocí rozdělení marginálních, což nám dále umožňuje jednoduše počítat momenty funkcí náhodných veličin nebo například jejich sdružené charakteristické funkce.

Pro ověření nezávislosti mezi dvěma náhodnými výběry často stačí použít test nezávislosti pomocí Pearsonova korelačního koeficientu, který v této práci budeme nazývat standardním t-testem. V případě finančních časových řad se ale mnohdy nejedná o náhodné výběry, jelikož pozorování v daném čase často závisí na předchozích pozorováních. Musí se tím pádem použít komplexnější testy a těmi se budeme v této práci zabývat. Konkrétně se zaměříme na testy nezávislosti pro časové řady, které se speciálně řídí modelem ARMA. Testy nejen představíme, ale poté je i v simulační studii porovnáme. Podobné porovnání testů bylo provedeno například v článku Khan a Khan (2020).

V první kapitole nejdříve rozebereme pojem nezávislost v rámci časových řad. Poté zadefinujeme křížovou kovarianci a korelaci a jejich výběrové varianty. Pro tyto odhady odvodíme základní vlastnosti, jako je asymptotická nestranost a konzistence. Dále se zaměříme na definici modelu ARMA a tvrzení s ním související. Kapitulu ukončíme krátkým připomenutím charakteristických funkcí.

V druhé kapitole popíšeme test, který Larry D. Haugh představil v článku Haugh (1976). Tento test využívá výběrové křížové korelace, které jsou spočteny pro odhady bílých šumů, které získáme z pozorování díky odhadnutí modelu ARMA.

Modifikovanému t-testu z článku Alpargu a Buonaccorsi (2009) se budeme věnovat v rámci třetí kapitoly. Tento test je upravenou verzí standardního t-testu nezávislosti tak, aby stačilo předpokládat slabou stacionaritu namísto náhodného výběru.

Čtvrtá kapitola je věnovaná teorii distanční kovariance a testu na ní založeného z článku Székely a kol. (2007). Výběrová distanční korelace je obdobou Pearsonova korelačního koeficientu, pro kterou nemusíme předpokládat normální rozdělení pro použití k testování nezávislosti. Test distanční kovariancí je ve zmíněném článku navrhnut pro náhodné výběry, v této práci je přizpůsoben na ARMA modely.

V páté kapitole představíme simulační studii, ve které byly porovnány všechny testy popsané v této práci spolu se standardním t-testem.

Jako poslední v šesté kapitole vyzkoušíme použít testy na reálná finanční data, konkrétně na ceny indexů S&P 500 a FTSE 100.

1. Základní definice a poznatky

V této kapitole představíme všechnu základní teorii, kterou budeme v dalších částech práce potřebovat. Nejdříve upřesníme, co myslíme pod pojmem časová řada.

Definice 1 *Posloupnost náhodných veličin $\{X_t : t \in \mathbb{Z}\}$ nazýváme **časovou řadou** a index t nazýváme **časem**.*

Poznámka: V této práci budeme v některých případech značit časovou řadu $\{X_t : t \in \mathbb{Z}\}$ zkráceně symbolem X a řadu $\{Y_t : t \in \mathbb{Z}\}$ symbolem Y .

Pod pojmem časová řada si můžeme představit například ceny akcií, počty lidí, kteří denně navštíví obchodní centrum, nebo počty požárů v České republice za rok. Před tím než se zaměříme na nezávislost, tak se na chvíli zastavíme u měřitelnosti v prostoru $\mathbb{R}^{\mathbb{Z}}$. Příslušnou borelovskou σ -algebru zavedeme jako součinovou σ -algebru.

Poznámka: V této práci budeme symboly \times a \otimes značit kartézský součin množin.

Definice 2 *Označme $\mathcal{G}^{\mathbb{Z}} := \{\otimes_{t \in \mathbb{Z}} (-\infty, a_t], a_t \in \mathbb{R}\}$ systém množin v $\mathbb{R}^{\mathbb{Z}}$. Potom $\mathcal{B}^{\mathbb{Z}} := \mathcal{B}(\mathbb{R}^{\mathbb{Z}}) = \sigma\mathcal{G}^{\mathbb{Z}}$ je borelovská σ -algebra generovaná systémem $\mathcal{G}^{\mathbb{Z}}$. Řekneme, že množina $A \subseteq \mathbb{R}^{\mathbb{Z}}$ je **měřitelná**, pokud $A \in \mathcal{B}^{\mathbb{Z}}$.*

Nejdůležitějším pojmem v této práci je nezávislost, kterou budeme v případě časových řad testovat, a proto jí věnujeme následující podkapitolu.

1.1 Nezávislost

Definice 3 *Časové řady $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$ nazveme **nezávislé**, pokud platí:*

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

pro každé $A, B \in \mathcal{B}^{\mathbb{Z}}$.

Pro ověření definice 3 by bylo zapotřebí znát rozdělení obou časových řad jednotlivě i sdruženě, což je v praxi bez jakýchkoliv velmi silných předpokladů téměř nemožné. V této kapitole odvodíme větu, na jejíž myšlenke stojí dále popsané testy. Nejdříve však připomeňme definici distribuční funkce.

Definice 4 *Nechť $X = \{X_t : t \in \mathbb{Z}\}$ je časová řada. **Distribuční funkcí řady** X nazýváme funkci $F_X : \mathbb{R}^{\mathbb{Z}} \rightarrow [0, 1]$ splňující:*

$$F_X(\mathbf{x}) := P(X \leq \mathbf{x}) = P(X_t \leq x_t, \forall t \in \mathbb{Z})$$

pro každé $\mathbf{x} \in \mathbb{R}^{\mathbb{Z}}$, kde x_t je prvek vektoru \mathbf{x} na pozici t .

Poznámka: Sdruženou distribuční funkci časových řad $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$ značíme symbolem $F_{(X,Y)}$, čímž rozumíme:

$$F_{(X,Y)}(\mathbf{x}, \mathbf{y}) = P(X \leq \mathbf{x}, Y \leq \mathbf{y})$$

pro každé $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathbb{Z}}$.

Začneme větou, která ekvivalentně definuje nezávislost pomocí distribučních funkcí časových řad.

Věta 1 *Nechť $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$ jsou časové řady. Pak následující tvrzení jsou ekvivalentní:*

- (1) $\{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ jsou nezávislé.
- (2) Platí $F_{(X,Y)}(\mathbf{x}, \mathbf{y}) = F_X(\mathbf{x})F_Y(\mathbf{y})$ pro každé $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathbb{Z}}$.

Důkaz. (1) \implies (2): Označme x_t a y_s , kde $t, s \in \mathbb{Z}$, prvky vektoru $\mathbf{x} \in \mathbb{R}^{\mathbb{Z}}$, respektive $\mathbf{y} \in \mathbb{R}^{\mathbb{Z}}$. Nechť $A = \bigotimes_{t \in \mathbb{Z}} (-\infty, x_t] \in \mathcal{G}^{\mathbb{Z}}$ a $B = \bigotimes_{s \in \mathbb{Z}} (-\infty, y_s] \in \mathcal{G}^{\mathbb{Z}}$ jsou měřitelné množiny. Pak platí:

$$F_{(X,Y)}(\mathbf{x}, \mathbf{y}) = P(X \in A, Y \in B) = P(X \in A)P(Y \in B) = F_X(\mathbf{x})F_Y(\mathbf{y}).$$

(2) \implies (1): Nechť $X' = \{X'_t : t \in \mathbb{Z}\}$ a $Y' = \{Y'_t : t \in \mathbb{Z}\}$ jsou nezávislé náhodné časové řady s distribučními funkcemi $F_{X'}$ a $F_{Y'}$. Pak pro každé $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathbb{Z}}$ platí:

$$F_{(X',Y')}(\mathbf{x}, \mathbf{y}) \stackrel{(1) \implies (2)}{=} F_{X'}(\mathbf{x})F_{Y'}(\mathbf{y}) = F_X(\mathbf{x})F_Y(\mathbf{y}) \stackrel{(2)}{=} F_{(X,Y)}(\mathbf{x}, \mathbf{y}).$$

Jelikož distribuční funkce určuje pravděpodobnostní rozdělení jednoznačně, má (X, Y) stejné rozdělení jako (X', Y') . Tedy X a Y jsou taktéž nezávislé. \square

Věta 1 nám ověřování definice nijak neusnadňuje, jelikož je stále potřeba mít informace o celých časových řadách. Proto formulujeme dvě následující věty, ve kterých prozkoumáme jinou možnost.

Věta 2 *Nechť $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$ jsou časové řady. Pak následující tvrzení jsou ekvivalentní:*

- (1) $\{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ jsou nezávislé.
- (2) Platí:

$$\begin{aligned} F_{((X_{t_1}, X_{t_2}, \dots, X_{t_n}), (Y_{s_1}, Y_{s_2}, \dots, Y_{s_m}))}((x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_m)) &= \\ &= F_{(X_{t_1}, X_{t_2}, \dots, X_{t_n})}(x_1, x_2, \dots, x_n) F_{(Y_{s_1}, Y_{s_2}, \dots, Y_{s_m})}(y_1, y_2, \dots, y_m) \end{aligned}$$

pro každé $n, m \in \mathbb{N}$, různé $t_1, t_2, \dots, t_n \in \mathbb{Z}$, různé $s_1, s_2, \dots, s_m \in \mathbb{Z}$ a $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m \in \mathbb{R}$.

Důkaz. (1) \implies (2): Označme symbolem $\mathbf{x} \in \mathbb{R}^{\mathbb{Z}}$ vektor takový, že na pozicích $t \in \{t_1, t_2, \dots, t_n\}$ má prvky $x_1, x_2, \dots, x_n \in \mathbb{R}$, jinde ∞ . Analogicky vytvoříme vektor $\mathbf{y} \in \mathbb{R}^{\mathbb{Z}}$. Pak platí:

$$\begin{aligned} F_{((X_{t_1}, X_{t_2}, \dots, X_{t_n}), (Y_{s_1}, Y_{s_2}, \dots, Y_{s_m}))}((x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_m)) &= F_{(X,Y)}(\mathbf{x}, \mathbf{y}) = \\ &= F_X(\mathbf{x})F_Y(\mathbf{y}) = F_{(X_{t_1}, X_{t_2}, \dots, X_{t_n})}(x_1, x_2, \dots, x_n) F_{(Y_{s_1}, Y_{s_2}, \dots, Y_{s_m})}(y_1, y_2, \dots, y_m). \end{aligned}$$

(2) \implies (1): Necht $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathbb{Z}}$ a $N = 2n + 1 \in \mathbb{N}$. Označme $\mathbf{x}_N \in \mathbb{R}^{\mathbb{Z}}$ vektor, který bude mít na pozicích $-N, -(N-1), \dots, N$ stejné hodnoty jako vektor \mathbf{x} a na ostatních pozicích ∞ . Stejně tak vytvoříme vektor $\mathbf{y}_N \in \mathbb{R}^{\mathbb{Z}}$ z \mathbf{y} . Pro vektor \mathbf{x}_N platí, že po složkách konverguje k vektoru \mathbf{x} . Navíc po složkách platí $\mathbf{x} \leq \mathbf{x}_N$, takže konvergence je zprava. Stejně tak \mathbf{y}_N konverguje zprava k \mathbf{y} . Protože distribuční funkce je z definice spojitá zprava, platí také rovnost $\lim_{N \rightarrow \infty} F_{(X,Y)}(\mathbf{x}_N, \mathbf{y}_N) = F_{(X,Y)}(\mathbf{x}, \mathbf{y})$. Potom:

$$\begin{aligned} F_{(X,Y)}(\mathbf{x}, \mathbf{y}) &= \lim_{N \rightarrow \infty} F_{(X,Y)}(\mathbf{x}_N, \mathbf{y}_N) \\ &= \lim_{N \rightarrow \infty} F_{((X_{-N}, \dots, X_N), (Y_{-N}, \dots, Y_N))}((x_{-N}, \dots, x_N), (y_{-N}, \dots, y_N)) \\ &= \lim_{N \rightarrow \infty} F_{(X_{-N}, \dots, X_N)}(x_{-N}, \dots, x_N) F_{(Y_{-N}, \dots, Y_N)}(y_{-N}, \dots, y_N) \\ &= \lim_{N \rightarrow \infty} F_X(\mathbf{x}_N) F_Y(\mathbf{y}_N) \\ &= F_X(\mathbf{x}) F_Y(\mathbf{y}). \end{aligned}$$

Podle věty 1 jsou $\{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ nezávislé. □

Věta 3 Necht $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$ jsou časové řady. Pak následující tvrzení jsou ekvivalentní:

- (1) $\{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ jsou nezávislé.
(2) Platí:

$$\begin{aligned} P\left((X_{t_1}, X_{t_2}, \dots, X_{t_n}) \in A, (Y_{s_1}, Y_{s_2}, \dots, Y_{s_m}) \in B\right) &= \\ &= P\left((X_{t_1}, X_{t_2}, \dots, X_{t_n}) \in A\right) P\left((Y_{s_1}, Y_{s_2}, \dots, Y_{s_m}) \in B\right) \end{aligned}$$

pro každé $n, m \in \mathbb{N}$, různé $t_1, t_2, \dots, t_n \in \mathbb{Z}$, různé $s_1, s_2, \dots, s_m \in \mathbb{Z}$ a pro $A \subseteq \mathbb{R}^n, B \subseteq \mathbb{R}^m$ měřitelné.

Důkaz. (1) \implies (2): Označme množiny $A_t = \mathbb{R}, \forall t \notin T = \{t_1, t_2, \dots, t_n\}$ a $B_s = \mathbb{R}, \forall s \notin S = \{s_1, s_2, \dots, s_m\}$. Dále $A' = A \times \bigotimes_{t \notin T} A_t$ a $B' = B \times \bigotimes_{s \notin S} B_s$. Jelikož A je měřitelná množina v \mathbb{R}^n a \mathbb{R} je měřitelná množina v \mathbb{R} , bude množina A' měřitelná v součinové σ -algebře $\mathcal{B}^{\mathbb{Z}}$. Stejný argument platí i pro množinu B' . Celkem lze odvodit:

$$\begin{aligned} P\left((X_{t_1}, X_{t_2}, \dots, X_{t_n}) \in A, (Y_{s_1}, Y_{s_2}, \dots, Y_{s_m}) \in B\right) &= \\ &= P(X \in A', Y \in B') \\ &= P(X \in A') P(Y \in B') \\ &= P\left((X_{t_1}, X_{t_2}, \dots, X_{t_n}) \in A\right) P\left((Y_{s_1}, Y_{s_2}, \dots, Y_{s_m}) \in B\right). \end{aligned}$$

(2) \implies (1): Jelikož bod (2) platí pro libovolné množiny A a B , platí speciálně pro množiny tvaru $A = \bigotimes_{i=1}^n (-\infty, x_i]$ pro libovolné $x_1, x_2, \dots, x_n \in \mathbb{R}$ a B tvaru obdobného. Tím pádem je splněna podmínka (2) věty 2 a řady X a Y jsou nezávislé. □

Věty 2 a 3 nám říkají, že nezávislost lze ověřit pomocí informací o všech konečných marginálních a konečných sdružených rozděleních časových řad. Ve skutečnosti budeme mít vždy k dispozici pouze konečný počet pozorování, který budeme

značit $N \in \mathbb{N}$, a tím pádem můžeme otestovat nezávislost pouze pro těchto N členů. Idea testů je taková, že otestujeme za určitých předpokladů nezávislost na konečném vzorku dat, a pokud data nebudou v rozporu s testovanou vlastností, rozhodneme se nezávislost nezamítnout. Jedním ze způsobů jak testovat nezávislost časových řad jsou testy založené na výpočtu kovariance, respektive korelace, o kterých pojednává následující podkapitola.

1.2 Kovariance a korelace

V této části definujeme pojmy související s kovariancí a korelací a vysvětlíme jejich význam.

Definice 5 *Nechť $X = \{X_t : t \in \mathbb{Z}\}$ je časová řada, pro kterou platí $\mathbb{E}[X_t^2] < \infty$ pro každé $t \in \mathbb{Z}$. **Autokovarianční funkci** časové řady rozumíme funkci:*

$$R_X(t, s) := \text{cov}(X_t, X_s)$$

kde $t, s \in \mathbb{Z}$. Pokud navíc platí $\text{var } X_t > 0$ pro každé $t \in \mathbb{Z}$, můžeme definovat **autokorelační funkci** jako:

$$r_X(t, s) := \frac{R_X(t, s)}{\sqrt{R_X(t, t)}\sqrt{R_X(s, s)}}.$$

Poznámka: Platí: $R_X(t, t) = \text{var } X_t$, a tedy $r_X(t, s) = \text{corr}(X_t, X_s)$ pro každé $t, s \in \mathbb{Z}$. Navíc autokovarianční i autokorelační funkce jsou symetrické, pro každé $t, s \in \mathbb{Z}$ platí:

$$R_X(t, s) = R_X(s, t), \quad r_X(t, s) = r_X(s, t).$$

Autokorelační funkce nám dává představu o lineární závislosti uvnitř časové řady. Pokud je autokorelační funkce nulová ve všech bodech kromě $t = s$, pak řekneme, že časová řada je **nekorelovaná**. To však neznamená, že časová řada je posloupností nezávislých veličin. Platí pouze opačná implikace, a to že posloupnost nezávislých veličin s konečnými a kladnými rozptyly je nekorelovaná. Nemůžeme tedy podle korelace určit, jestli má časová řada nezávislé prvky, ale můžeme se na to podívat z jiné strany. Pokud autokorelační funkce je nenulová v nějakém bodě $t \neq s$, znamená to, že časová řada není posloupností nezávislých náhodných veličin. Pokud má časová řada sdruženě normální rozdělení, požadovaná ekvivalence platí. Nezávislost časových řad můžeme tedy přímo testovat pomocí korelace například za předpokladu normálního rozdělení.

Definice 6 *Řekneme, že časová řada $\{X_t : t \in \mathbb{Z}\}$ je **slabě stacionární**, pokud pro ni platí následující podmínky:*

- (1) *Konstantní střední hodnota: $\mathbb{E}[X_t] = c$ pro každé $t \in \mathbb{Z}$, kde $c \in \mathbb{R}$.*
- (2) *Konečné druhé momenty: $\mathbb{E}[X_t^2] < \infty$ pro každé $t \in \mathbb{Z}$.*
- (3) *Autokovarianční funkce $R_X(t, s)$ se dá zapsat jako funkce jedné proměnné argumentu $s - t$ pro každé $t, s \in \mathbb{Z}$.*

Co přesněji znamená podmínka (3), vysvětlíme v následující poznámce.

Poznámka: Časovou řadu, která splňuje podmínku (3), nazveme **kovariančně stacionární**. Pro takovou řadu zavedeme značení autokovarianční funkce jako:

$$R_X(s - t) := R_X(t, s),$$

a pokud $R_X(0) > 0$, tak značení autokorelační funkce jako:

$$r_X(s - t) := \frac{R_X(s - t)}{R_X(0)}$$

pro každé $t, s \in \mathbb{Z}$, jelikož $R_X(0) = R_X(t - t) = R_X(t, t) = \text{var } X_t$. Navíc díky symetrii pro každé $t \in \mathbb{Z}$ platí:

$$R_X(t) = R_X(-t), \quad r_X(t) = r_X(-t).$$

Slabá stacionarita znamená, že časová řada má konstantní první i druhé momenty a je kovariančně stacionární. Kovariančně stacionární časová řada má tu vlastnost, že autokovarianční a i případně autokorelační funkce nezávisí na časech, v kterých chceme funkce vyhodnotit, ale závisí pouze na tom, jak moc jsou časové okamžiky vzdálené. Neboli platí:

$$\text{cov}(X_t, X_{t+k}) = \text{cov}(X_s, X_{s+k}), \forall t, s, k \in \mathbb{Z}.$$

Definice 7 Necht $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$ jsou časové řady. **Křížovou kovarianční funkcí** rozumíme funkci:

$$R_{X,Y}(t, s) := \text{cov}(X_t, Y_s)$$

pro libovolné $t, s \in \mathbb{Z}$. Pokud navíc platí $\text{var } X_t > 0$ a $\text{var } Y_t > 0$ pro každé $t \in \mathbb{Z}$, můžeme definovat **křížovou korelační funkci** jako:

$$r_{X,Y}(t, s) := \frac{R_{X,Y}(t, s)}{\sqrt{R_X(t, t)}\sqrt{R_Y(s, s)}}.$$

Poznámka: Křížová kovarianční i křížová korelační funkce **nejsou symetrické**, tudíž pro každé $t, s \in \mathbb{Z}$ obecně neplatí rovnosti:

$$R_{X,Y}(t, s) = R_{X,Y}(s, t), \quad r_{X,Y}(t, s) = r_{X,Y}(s, t).$$

Křížová kovarianční a korelační funkce jsou analogií autokovarianční a autokorelační funkce. Nyní však vyjadřují míru lineární závislosti napříč dvěma časovými řadami. Podobně jako v případě autokovarianční a autokorelační funkce platí, že dvě navzájem nezávislé časové řady mají křížové funkce nulové ve všech bodech, opačná implikace ale platí pouze za předpokladu silnějších podmínek, například že jsou časové řady sdruženě normálně rozdělené.

Definice 8 Řekneme, že časové řady $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$ jsou **sdruženě slabě stacionární**, pokud jsou obě slabě stacionární a navíc platí, že křížová kovarianční funkce $R_{X,Y}(t, s)$ se dá zapsat jako funkce jedné proměnné argumentu $s - t$ pro každé $t, s \in \mathbb{Z}$.

Poznámka: Pro dvojici sdruženě slabě stacionárních časových řad zavedeme značení křížové kovarianční funkce jako:

$$R_{X,Y}(s-t) := R_{X,Y}(t, s)$$

a v případě $R_X, R_Y > 0$ značení křížové korelační funkce jako:

$$r_{X,Y}(s-t) := \frac{R_{X,Y}(s-t)}{\sqrt{R_X(0)}\sqrt{R_Y(0)}}$$

pro každé $t, s \in \mathbb{Z}$. Stejně jako v nestacionárním případě platí, že křížová kovarianční a křížová korelační funkce obecně **nejsou symetrické**, takže pro každé $t \in \mathbb{Z}$ obecně neplatí rovnosti:

$$R_{X,Y}(t) = R_{X,Y}(-t), \quad r_{X,Y}(t) = r_{X,Y}(-t).$$

Sdružená slabá stacionarita je analogií ke kovarianční stacionaritě jedné řady. Pro určení nekorelovanosti dvou řad je potřeba vyšetřit funkci $R_{X,Y}(t)$ pro každé $t \in \mathbb{Z}$, na rozdíl od případu nekorelovanosti v rámci jedné řady, kde stačilo funkci $R_X(t)$ vyšetřit jenom pro t nezáporná. Stejně jako v jednorozměrném případě ale platí:

$$\text{cov}(X_t, Y_{t+k}) = \text{cov}(X_s, Y_{s+k}), \quad \forall t, s, k \in \mathbb{Z}.$$

Tato vlastnost nám umožňuje konstruovat smysluplné odhady křížových funkcí pro časové řady, jak se přesvědčíme v následující části.

1.3 Výběrové křížové kovariance a korelace

V této části uvedeme způsob výpočtu výběrové křížové kovarianční a výběrové korelační funkce a odvodíme některé jejich vlastnosti.

Nechť x_1, x_2, \dots, x_N a y_1, y_2, \dots, y_N jsou pozorování dvou sdruženě slabě stacionárních časových řad. Standardně budeme značit průměry symboly \bar{x} a \bar{y} a výběrové rozptyly symboly $\widehat{\sigma}_x^2$ a $\widehat{\sigma}_y^2$. Pro taková pozorování definujeme výběrovou křížovou kovarianční funkci jako:

$$\widehat{R}_{X,Y}(k) := \begin{cases} \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(y_{t+k} - \bar{y}), & \text{pro } 0 \leq k \leq N-1, \\ \frac{1}{N} \sum_{t=1-k}^N (x_t - \bar{x})(y_{t+k} - \bar{y}), & \text{pro } 1-N \leq k < 0. \end{cases}$$

Pokud jsou $\widehat{\sigma}_x^2$ a $\widehat{\sigma}_y^2$ nenulové, definujeme výběrovou křížovou korelační funkci jako

$$\widehat{r}_{X,Y}(k) := \frac{\widehat{R}_{X,Y}(k)}{\sqrt{\widehat{\sigma}_x^2}\sqrt{\widehat{\sigma}_y^2}}.$$

Výběrové odhady budeme používat hlavně v případě dvou normálně rozdělených bílých šumů, a proto definici bílého šumu připomeneme.

Definice 9 *Posloupnost náhodných veličin $\{X_t : t \in \mathbb{Z}\}$ nazveme **bílým šumem**, pokud jsou veličiny navzájem nekorelované, jejich střední hodnota je nulová a rozptyl je konstantní, kladný a konečný. Pokud jsou veličiny navíc nezávislé, hovoříme o **striktním bílém šumu**.*

Poznámka: Z definice bílého šumu vyplývá, že bílý šum je slabě stacionární časová řada.

V případě dvou striktních bílých šumů bude průměr konvergovat skoro jistě k 0 ze silného zákona velkých čísel, a tudíž ho můžeme ve vzorcích vynechat, čímž dostaneme nové verze vzorců:

$$\begin{aligned}\tilde{R}_{X,Y}(k) &:= \begin{cases} \frac{1}{N} \sum_{t=1}^{N-k} x_t y_{t+k}, & \text{pro } 0 \leq k \leq N-1, \\ \frac{1}{N} \sum_{t=1-k}^N x_t y_{t+k}, & \text{pro } 1-N \leq k \leq 0, \end{cases} \quad (1.1) \\ \tilde{r}_{X,Y}(k) &:= \frac{\tilde{R}_{X,Y}(k)}{\sqrt{\tilde{\sigma}_x^2} \sqrt{\tilde{\sigma}_y^2}},\end{aligned}$$

kde $\tilde{\sigma}_x^2$ a $\tilde{\sigma}_y^2$ jsou výběrové rozptyly spočtené bez odečítání průměru.

Nyní se budeme zabývat vlastnostmi výběrové křížové kovariance. Cílem bude odvodit předpoklady, které jsou potřeba k dokázání konzistence a dále komentovat jejich praktičnost. Nejdříve dokážeme asymptotickou nestrannost pro jednodušší tvar vzorce (1.1).

Tvrzení 4 *Nechť $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$ jsou sdruženě slabě stacionární časové řady, obě s nulovou střední hodnotou a necht' $k \in \mathbb{Z}$. Poté je $\tilde{R}_{X,Y}(k)$ asymptoticky nestranným odhadem $R_{X,Y}(k)$.*

Důkaz. Asymptotickou nestrannost odvodíme pro $k \geq 0$, pro záporné k je výpočet analogický. Pro střední hodnotu odhadu dostaneme:

$$\mathbb{E} \tilde{R}_{X,Y}(k) = \mathbb{E} \frac{1}{N} \sum_{t=1}^{N-k} X_t Y_{t+k} = \frac{1}{N} \sum_{t=1}^{N-k} R_{X,Y}(k) = \frac{N-k}{N} R_{X,Y}(k) \xrightarrow{N \rightarrow \infty} R_{X,Y}(k).$$

□

Pro dokázání tvrzení 4 nebylo potřeba žádných předpokladů kromě nulové střední hodnoty. V obecném případě, který následuje, důkaz komplikují odhady středních hodnot.

Tvrzení 5 *Nechť $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$ jsou sdruženě slabě stacionární časové řady a $k \in \mathbb{Z}$. Necht' navíc existuje $M \in \mathbb{N}$ takové, že pro každé $t > M$ platí $R_{X,Y}(t) = 0$ a $R_{X,Y}(-t) = 0$. Poté je $\hat{R}_{X,Y}(k)$ asymptoticky nestranným odhadem $R_{X,Y}(k)$.*

Důkaz. Označme μ_X střední hodnotu časové řady X a μ_Y střední hodnotu časové řady Y . Asymptotickou nestrannost odvodíme pro $k \geq 0$, pro záporné k je výpočet analogický. Začneme s výpočtem střední hodnoty odhadu:

$$\begin{aligned}\mathbb{E} N \hat{R}_{X,Y}(k) &= \mathbb{E} \sum_{t=1}^{N-k} (X_t - \bar{X})(Y_{t+k} - \bar{Y}) \\ &= \mathbb{E} \sum_{t=1}^{N-k} (X_t - \mu_X + \mu_X - \bar{X})(Y_{t+k} - \mu_Y + \mu_Y - \bar{Y}) \\ &= \mathbb{E} \sum_{t=1}^{N-k} \left[(X_t - \mu_X)(Y_{t+k} - \mu_Y) + (X_t - \mu_X)(\mu_Y - \bar{Y}) + \right. \\ &\quad \left. + (\mu_X - \bar{X})(Y_{t+k} - \mu_Y) + (\mu_X - \bar{X})(\mu_Y - \bar{Y}) \right].\end{aligned}$$

Nyní můžeme střední hodnotu rozdělit na součet čtyř součinů a každý tento součin řešit zvlášť.

1. člen:

$$\mathbb{E} \sum_{t=1}^{N-k} (X_t - \mu_X)(Y_{t+k} - \mu_Y) = \sum_{t=1}^{N-k} R_{X,Y}(t+k-t) = (N-k)R_{X,Y}(k).$$

Konvergence pro tento člen:

$$\frac{N-k}{N} R_{X,Y}(k) \xrightarrow{N \rightarrow \infty} R_{X,Y}(k).$$

2. člen:

$$\begin{aligned} \mathbb{E} \sum_{t=1}^{N-k} (X_t - \mu_X)(\mu_Y - \bar{Y}) &= \sum_{t=1}^{N-k} -\text{cov}(X_t, \bar{Y}) = \frac{1}{N} \sum_{t=1}^{N-k} \sum_{s=1}^N -\text{cov}(X_t, Y_s) \\ &= -\frac{1}{N} \sum_{t=1}^{N-k} \sum_{s=1}^N R_{X,Y}(s-t) \\ &= -\frac{1}{N} \left[(N-k) \sum_{t=0}^k R_{X,Y}(t) + \sum_{t=k+1}^{N-1} (N-t) R_{X,Y}(t) + \sum_{t=1-N+k}^{-1} (t+N-k) R_{X,Y}(t) \right]. \end{aligned}$$

Po vynásobení celého výrazu $\frac{1}{N}$ je první člen řádu N^{-1} a jde do 0. Díky předpokladu nekorelovanosti od určitého M jsou zbývající členy řádu M/N , a tudíž konvergují také do 0.

3. člen je analogický k 2. členu, a tedy také konverguje k 0 díky předpokladům.

4. člen:

$$\begin{aligned} \mathbb{E} \sum_{t=1}^{N-k} (\mu_X - \bar{X})(\mu_Y - \bar{Y}) &= (N-k) \text{cov}(\bar{X}, \bar{Y}) = \frac{N-k}{N^2} \sum_{t=1}^N \sum_{s=1}^N R_{X,Y}(s-t) \\ &= \frac{N-k}{N^2} \sum_{t=1-N}^{N-1} (N-|t|) R_{X,Y}(t). \end{aligned}$$

Tento člen je po přenásobení $\frac{1}{N}$ díky předpokladům řádu M/N a konverguje do 0.

Celkem jsme odvodili, že všechny členy kromě prvního konvergují k 0, a dokázali jsme tím, že střední hodnota výběrové křížové kovariance konverguje ke skutečné křížové kovarianci. Výsledkem je, že výběrová křížová kovariance je asymptoticky nestranný odhad. □

Na dokázání asymptotické nestrannosti v obecnějším případě stačí předpoklad o nekorelovanosti dvou pozorování, pokud jsou od sebe dostatečně časově daleko. Tento předpoklad je rozumný, jelikož M může být jakkoliv velké a v praxi se dá očekávat, že dvě velmi vzdálená pozorování už nebudou spolu nějak korelovaná. Jako další odvodíme konzistenci odhadu kovariance v případě dvou bílých šumů.

Tvrzení 6 *Nechť $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$ jsou časové řady a $k \in \mathbb{Z}$. Nechť jsou dále splněny následující předpoklady:*

- (1) $\{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ jsou sdruženě slabě stacionární bílé šumy.
- (2) Existuje $M' \in \mathbb{N}$ takové, že jsou veličiny $Z_t = X_t Y_{t+k}$ a $Z_s = X_s Y_{s+k}$ nekorelované, pokud $|t-s| > M'$.

(3) Existuje $C \in \mathbb{R}$ takové, že platí $|\mathbb{E} X_t Y_{t+k} X_s Y_{s+k} - \mathbb{E} X_t Y_{t+k} \mathbb{E} X_s Y_{s+k}| \leq C$ pro každé $t, s \in \mathbb{Z}$.

Poté je $\tilde{R}_{X,Y}(k)$ konzistentním odhadem $R_{X,Y}(k)$.

Důkaz. Ukážeme, že za splnění předpokladů je výběrová křížová kovariance asymptoticky nestranný odhad a navíc, že rozptyl odhadu konverguje k 0. Asymptotickou nestrannost jsme dokázali ve tvrzení 4. Stačí ukázat, že rozptyl konverguje k 0. Konzistenci stejně jako v důkazu tvrzení 4 dokážeme pro $k \geq 0$, pro záporné k by důkaz vypadal analogicky. Pro rozptyl odhadu máme:

$$\begin{aligned} \text{var } \tilde{R}_{X,Y}(k) &= \text{var} \left[\frac{1}{N} \sum_{t=1}^{N-k} X_t Y_{t+k} \right] = \frac{1}{N^2} \text{var} \left[\sum_{t=1}^{N-k} X_t Y_{t+k} \right] \\ &= \frac{1}{N^2} \sum_{t=1}^{N-k} \sum_{s=1}^{N-k} \text{cov}(X_t Y_{t+k}, X_s Y_{s+k}) \\ &= \frac{1}{N^2} \sum_{t=1}^{N-k} \sum_{s=1}^{N-k} (\mathbb{E} X_t Y_{t+k} X_s Y_{s+k} - \mathbb{E} X_t Y_{t+k} \mathbb{E} X_s Y_{s+k}). \end{aligned}$$

Díky předpokladu (2) platí $\mathbb{E} X_t Y_{t+k} X_s Y_{s+k} - \mathbb{E} X_t Y_{t+k} \mathbb{E} X_s Y_{s+k} = 0$ pro každé $|t - s| > M'$. Celý výraz navíc díky předpokladu (3) můžeme odhadnout následovně:

$$\begin{aligned} \frac{1}{N^2} \sum_{t=1}^{N-k} \sum_{s=1}^{N-k} (\mathbb{E} X_t Y_{t+k} X_s Y_{s+k} - \mathbb{E} X_t Y_{t+k} \mathbb{E} X_s Y_{s+k}) &\leq \frac{1}{N^2} \sum_{t=1}^{N-k} 2C(M' + 1) \\ &\leq \frac{2(N - k)C(M' + 1)}{N^2}. \end{aligned}$$

Tento výraz konverguje k 0 pro $N \rightarrow \infty$, a tedy i rozptyl konverguje k 0. Celkem jsme dokázali, že odhad výběrové křížové kovarianční funkce je za splnění předpokladů konzistentním odhadem. \square

V důkazu věty 6 jsme použili předpoklad, že členy $X_t Y_{t+k}$ a $X_s Y_{s+k}$ jsou nekorelované, pokud jsou od sebe dostatečně časově vzdáleny. Tento předpoklad je složitější na interpretaci, jelikož v něm nehraje roli jenom vazba mezi členy jedné a druhé posloupnosti, ale je ovlivněn i závislostí členů uvnitř obou časových řad. Pokud ale uvažujeme dvě posloupnosti bílých šumů, tak víme, že uvnitř jsou řady nekorelované a jde tedy o podobný předpoklad jako v předchozím případě. Jako poslední předpoklad jsme potřebovali konečnost kovariance smíšených členů, což je v praxi rozumný a splnitelný předpoklad. Z konzistence výběrové křížové kovariance jednoduše dostaneme konzistenci výběrové křížové korelace.

Důsledek 1 Za splnění všech předpokladů z tvrzení 6 a předpokladů $\text{var } X_t > 0$ a $\text{var } Y_t > 0$ pro každé $t \in \mathbb{Z}$, je $\tilde{r}_{X,Y}(k)$ konzistentním odhadem $r_{X,Y}(k)$.

Důkaz. Z věty 6 víme, že $\tilde{R}_{X,Y}(k)$ je konzistentním odhadem $R_{X,Y}(k)$. Stejně tak jsou i $\widetilde{\sigma}_X^2$ a $\widetilde{\sigma}_Y^2$ konzistentními odhady σ_X^2 a σ_Y^2 . Dostáváme:

$$\left(\tilde{R}_{X,Y}(k), \widetilde{\sigma}_X^2, \widetilde{\sigma}_Y^2 \right) \xrightarrow{N \rightarrow \infty} \left(R_{X,Y}(k), \sigma_X^2, \sigma_Y^2 \right).$$

Označme $g(R, X, Y) = \frac{R}{\sqrt{X}\sqrt{Y}}$. Z věty o spojitě transformaci dostáváme:

$$g \left(\tilde{R}_{X,Y}(k), \widetilde{\sigma}_X^2, \widetilde{\sigma}_Y^2 \right) \xrightarrow{N \rightarrow \infty} g \left(R_{X,Y}(k), \sigma_X^2, \sigma_Y^2 \right).$$

Dostáváme tak konzistenci odhadu $\tilde{r}_{X,Y}(k)$.

□

1.4 ARMA model

Jedním z nejpoužívanějších modelů pro časové řady je model ARMA, kterým se budou řídit generovaná data v pozdější simulační části 5.

Definice 10 Řekneme, že časová řada $\{X_t : t \in \mathbb{Z}\}$ se řídí **ARMA(m, n) modelem**, $m, n \in \mathbb{N}_0$, pokud:

$$X_t + a_1 X_{t-1} + \dots + a_m X_{t-m} = E_t + b_1 E_{t-1} + \dots + b_n E_{t-n}, \quad t \in \mathbb{Z}, \quad (1.2)$$

kde $a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_n \in \mathbb{R}$ jsou konstanty, dále platí $a_m \neq 0$ a $b_n \neq 0$ a $\{E_t : t \in \mathbb{Z}\}$ je bílý šum.

Dále se zaměříme na speciální druh časových řad, který budeme potřebovat v následujících testech.

Definice 11 Řekneme, že časová řada $\{X_t : t \in \mathbb{Z}\}$ je **(kauzální) lineární proces**, pokud existuje posloupnost konstant $\{g_j, j \in \mathbb{N}_0\}$ splňující $\sum_{j=0}^{\infty} |g_j| < \infty$ taková, že pro každé $t \in \mathbb{Z}$ platí:

$$X_t = \sum_{j=0}^{\infty} g_j E_{t-j},$$

kde $\{E_t : t \in \mathbb{Z}\}$ je bílý šum.

Poznámka: Bílý šum je lineární proces s $g_0 = 1$ a $g_j = 0$ pro každé $j \geq 1$.

V dalším tvrzení se přesvědčíme o tom, že každý lineární proces je slabě stacionární časovou řadou.

Tvrzení 7 Lineární proces je slabě stacionární časová řada.

Důkaz. Ověříme definici slabé stacionarity:

(1) Konstantní střední hodnota:

$$\mathbb{E} X_t = \mathbb{E} \sum_{j=0}^{\infty} g_j E_{t-j} = \sum_{j=0}^{\infty} g_j \mathbb{E} E_{t-j} = 0, \quad \forall t \in \mathbb{Z},$$

jelikož z definice 9 víme, že bílý šum má nulovou střední hodnotu.

(2) Konečné druhé momenty:

$$\mathbb{E} X_t^2 = \mathbb{E} \left[\sum_{j=0}^{\infty} g_j E_{t-j} \right]^2 = \mathbb{E} \left[\sum_{j=0}^{\infty} g_j^2 E_{t-j}^2 + \sum_{j < k} g_j g_k E_{t-j} E_{t-k} \right] = \sigma^2 \sum_{j=0}^{\infty} g_j^2 < \infty, \quad \forall t \in \mathbb{Z},$$

jelikož suma přes j a k je nulová díky $\mathbb{E} E_{t-j} E_{t-k} = 0$, což plyne z nekorelovanosti bílého šumu, a $\sum_{j=0}^{\infty} |g_j| < \infty \implies \sum_{j=0}^{\infty} g_j^2 < \infty$.

(3) Kovarianční stacionarita:

$$R_X(t, s) = \mathbb{E}[X_t X_s] = \mathbb{E} \left[\left(\sum_{j=0}^{\infty} g_j E_{t-j} \right) \left(\sum_{k=0}^{\infty} g_k E_{s-k} \right) \right] = \mathbb{E} \left[\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} g_j E_{t-j} g_k E_{s-k} \right].$$

Nenulové členy v sumě jsou všechny ty, pro které platí $t - j = s - k$, tedy členy $k = s - t + j$:

$$\mathbb{E}\left[\sum_{j=0}^{\infty} g_j E_{t-j}^2 g_{s-t+j}\right] = \sigma^2 \sum_{j=0}^{\infty} g_j g_{s-t+j} = R_X(s - t).$$

Dokázali jsme, že lineární proces splňuje všechny podmínky definice 6 a je slabě stacionární časovou řadou. □

Následující tvrzení nám udává postačující podmínky, za jejichž splnění je ARMA model lineárním procesem.

Tvrzení 8 *Nechť se časová řada $\{X_t : t \in \mathbb{Z}\}$ řídí modelem ARMA(m, n) s předpisem (1.2). Označme*

$$a(z) = 1 + a_1 z + a_2 z^2 + \dots + a_m z^m, \quad b(z) = 1 + b_1 z + b_2 z^2 + \dots + b_n z^n.$$

Nechť polynomy $a(z)$ a $b(z)$ nemají žádné společné kořeny a všechny kořeny polynomu $a(z)$ leží vně jednotkového kruhu. Pak je X lineárním procesem a zároveň slabě stacionární řadou. Navíc lze X zapsat jako:

$$X_t = \sum_{j=0}^{\infty} g_j E_{t-j}, \quad \forall t \in \mathbb{Z},$$

kde $\{g_j, j \in \mathbb{N}_0\}$ jsou koeficienty řady

$$g(z) = \sum_{j=0}^{\infty} g_j z^j = \frac{b(z)}{a(z)}, \quad |z| \leq 1. \quad (1.3)$$

Důkaz. Prášková (2016), Věta 5.6. □

Pokud budeme v testech požadovat, aby ARMA(m, n) byla lineárním procesem, využijeme tvrzení 8 k jejímu ověření. Navíc nám udává způsob, jak získat předpis lineárního procesu pro modely ARMA. Nyní formulujeme tvrzení, na jehož myšlenke jsou postaveny některé představené testy.

Tvrzení 9 *Nechť $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$ jsou lineární procesy generované bílými šumy $\{E_{X,t} : t \in \mathbb{Z}\}$ a $\{E_{Y,t} : t \in \mathbb{Z}\}$. Pokud jsou bílé šumy navzájem nezávislé, tak jsou i lineární procesy navzájem nezávislé¹.*

Důkaz. Z definice lineárního procesu existují posloupnosti konstant $\{g_{X,j}, j \in \mathbb{N}_0\}$ a $\{g_{Y,j}, j \in \mathbb{N}_0\}$ takové, že pro každé $t \in \mathbb{Z}$ platí:

$$X_t = \sum_{j=0}^{\infty} g_{X,j} E_{X,t-j}, \quad Y_t = \sum_{j=0}^{\infty} g_{Y,j} E_{Y,t-j}.$$

Jelikož jsou X a Y funkce navzájem nezávislých veličin, musí být také navzájem nezávislé. □

¹Kdybychom navíc předpokládali invertibilitu obou řad, tak místo implikace platí ekvivalence. Pro invertibilní řadu X existují konstanty $\{h_j, j \in \mathbb{N}_0\}$ splňující $\sum_{j=0}^{\infty} |h_j| < \infty$ takové, že pro každé $t \in \mathbb{Z}$ platí $E_t = \sum_{j=0}^{\infty} h_j X_{t-j}$, kde $\{E_t : t \in \mathbb{Z}\}$ je bílý šum.

Podle tvrzení 9 se tak nemusí testovat pouze přímo pozorované časové řady ARMA, ale dá se využít příslušných bílých šumů, které mají hezčí vlastnosti. Někdy může být na první pohled složité rozpoznat závislost mezi dvěma řadami ARMA, jak ukážeme v následujícím příkladu.

Příklad: Necht X a Y jsou časové řady řídicí se modely ARMA s následujícími předpisy:

$$X_t - 0,95X_{t-1} = E_{X,t} - 0,1E_{X,t-1} + 0,6E_{X,t-2}, \quad (1.4)$$

$$Y_t + 0,3Y_{t-1} - 0,2Y_{t-2} = E_{Y,t} - 0,7E_{Y,t-1}. \quad (1.5)$$

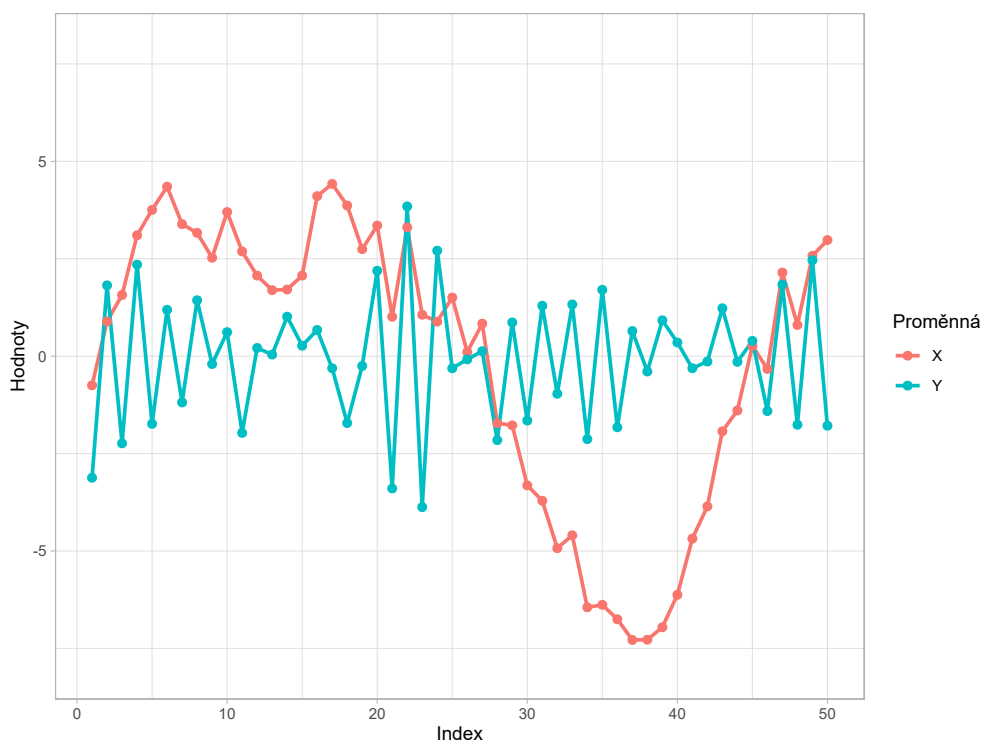
Pomocí výpočetního softwaru R (R Core Team (2022)) můžeme vygenerovat jednu možnou trajektorii obou řad. Postup generování dat je popsán v části 5.1. Z grafu 1.1 bychom mohli usoudit, že obě řady jsou navzájem nezávislé. Ve skutečnosti nejsou nezávislé, jelikož příslušné bílé šumy ($E_{X,t}, E_{Y,t}$) se řídí dvourozměrným normálním rozdělením s korelací rovnou 0,8, viz graf 1.2.

Jako další vygenerujeme pozorování řad X a Y řídicí se předpisy:

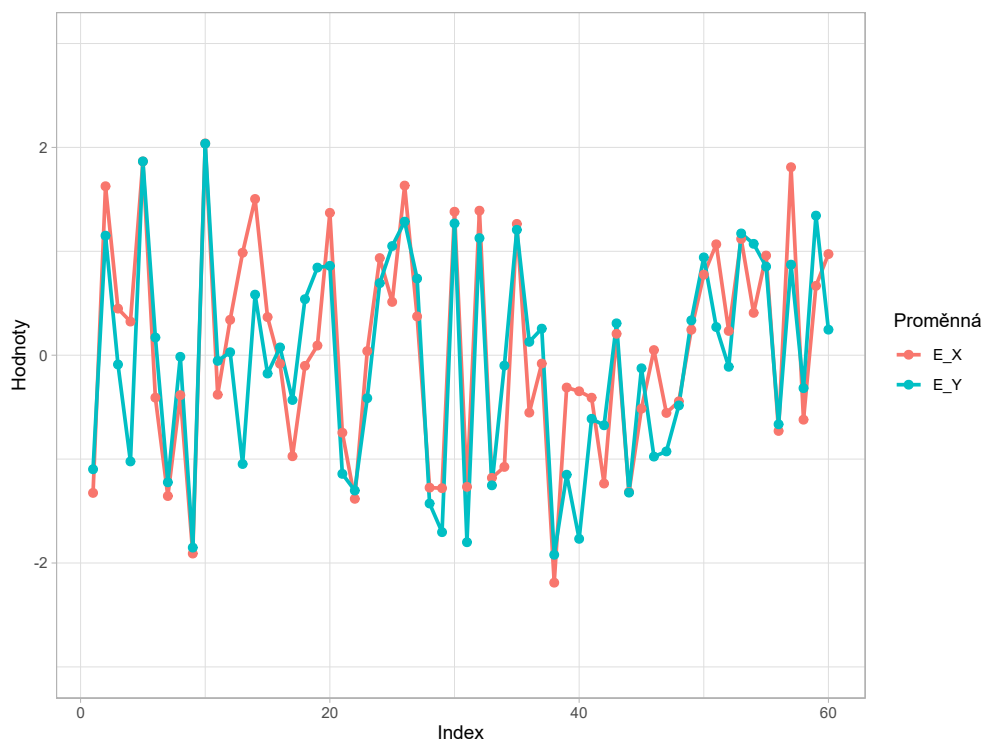
$$X_t + 0,3X_{t-1} = E_{X,t} - 0,7E_{X,t-1} + 0,6E_{X,t-2}, \quad (1.6)$$

$$Y_t + 0,3Y_{t-1} - 0,2Y_{t-2} = E_{Y,t} - 0,7E_{Y,t-1}. \quad (1.7)$$

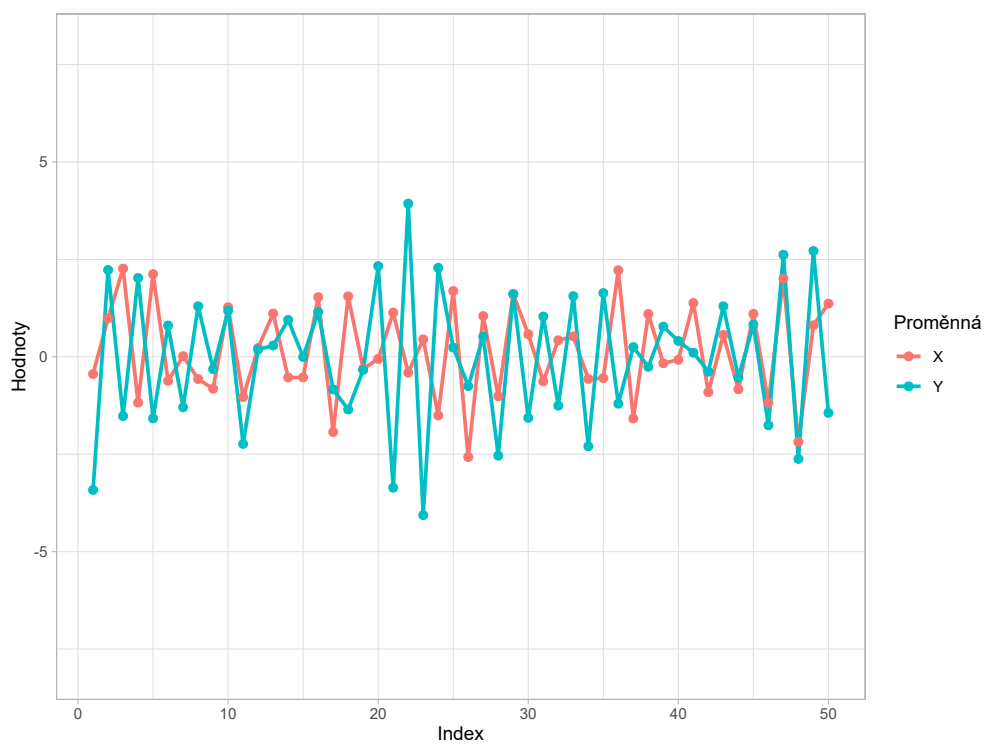
Z grafu 1.3 by se mohlo zdát, že jsou pozorované řady na sobě závislé. To je však pouze náhoda, jelikož jejich příslušné bílé šumy (viz graf 1.4) jsou nezávislé.



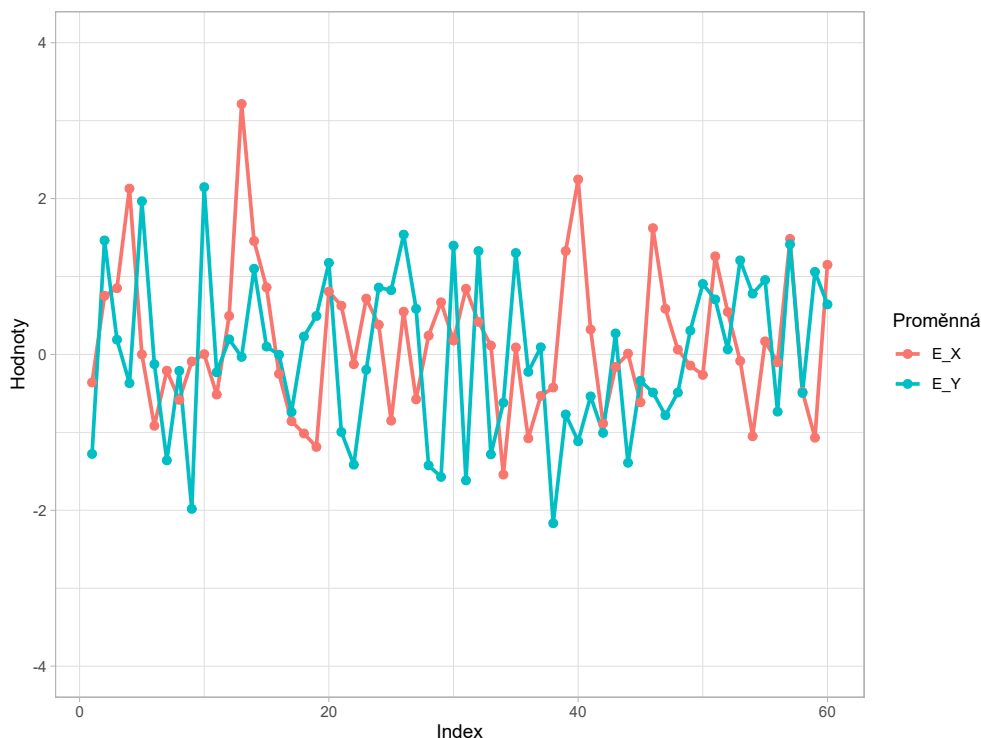
Obrázek 1.1: Graf znázorňující dvě časové řady X , řídicí se předpisem (1.4), a Y , řídicí se předpisem (1.5), které nejsou nezávislé.



Obrázek 1.2: Graf znázorňující bílé šумы příslušné řadám z grafu 1.1, které se řídí dvourozměrným normálním rozdělením s korelací ve stejném okamžiku rovnou 0,8.



Obrázek 1.3: Graf znázorňující dvě nezávislé časové řady X , řídicí se předpisem (1.6), a Y , řídicí se modelem (1.7).



Obrázek 1.4: Graf znázorňující bílé šумы příslušné řadám z grafu 1.3. Tyto bílé šумы jsou navzájem nezávislé.

1.5 Charakteristická funkce

Jedna z možností, jak popsat rozdělení náhodné veličiny či vektoru jednoznačně, je pomocí její charakteristické funkce.

Definice 12 *Nechť $\mathbf{X} = (X_1, X_2, \dots, X_p)$ je reálný náhodný vektor a $p \in \mathbb{N}$. Charakteristická funkce náhodného vektoru \mathbf{X} je definována jako $\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}[e^{i\langle \mathbf{t}, \mathbf{X} \rangle}]$, kde i je komplexní jednotka, $\mathbf{t} \in \mathbb{R}^p$ a $\langle \cdot, \cdot \rangle$ značí standardní skalární součin.*

Charakteristická funkce je účinný nástroj, pomocí kterého lze ověřit nezávislost dvou náhodných vektorů.

Tvrzení 10 *Nechť $\mathbf{X} = (X_1, X_2, \dots, X_p)$ a $\mathbf{Y} = (Y_1, Y_2, \dots, Y_q)$ jsou reálné náhodné vektory a $\varphi_{\mathbf{X}}, \varphi_{\mathbf{Y}}$ jsou jejich charakteristické funkce. Potom platí:*

$$\mathbf{X} \text{ a } \mathbf{Y} \text{ jsou nezávislé} \iff \varphi_{\mathbf{X}, \mathbf{Y}}(\mathbf{t}, \mathbf{s}) = \varphi_{\mathbf{X}}(\mathbf{t})\varphi_{\mathbf{Y}}(\mathbf{s}), \forall (\mathbf{t}, \mathbf{s}) \in \mathbb{R}^{p+q}.$$

Důkaz. \implies :

$$\begin{aligned} \varphi_{\mathbf{X}, \mathbf{Y}}(\mathbf{t}, \mathbf{s}) &= \mathbb{E}[e^{i\langle (\mathbf{t}, \mathbf{s}), (\mathbf{X}, \mathbf{Y}) \rangle}] = \mathbb{E}[e^{i(\langle \mathbf{t}, \mathbf{X} \rangle + \langle \mathbf{s}, \mathbf{Y} \rangle)}] \stackrel{\text{nezávislost}}{=} \mathbb{E}[e^{i\langle \mathbf{t}, \mathbf{X} \rangle}] \mathbb{E}[e^{i\langle \mathbf{s}, \mathbf{Y} \rangle}] \\ &= \varphi_{\mathbf{X}}(\mathbf{t})\varphi_{\mathbf{Y}}(\mathbf{s}) \end{aligned}$$

\impliedby : Nechť \mathbf{X}' a \mathbf{Y}' jsou nezávislé náhodné vektory, které mají stejné charakteristické funkce jako \mathbf{X} a \mathbf{Y} . Potom $\forall (\mathbf{t}, \mathbf{s}) \in \mathbb{R}^{p+q}$ platí:

$$\varphi_{\mathbf{X}', \mathbf{Y}'}(\mathbf{t}, \mathbf{s}) \stackrel{\text{stejně}}{=} \varphi_{\mathbf{X}'}(\mathbf{t})\varphi_{\mathbf{Y}'}(\mathbf{s}) = \varphi_{\mathbf{X}}(\mathbf{t})\varphi_{\mathbf{Y}}(\mathbf{s}) = \varphi_{\mathbf{X}, \mathbf{Y}}(\mathbf{t}, \mathbf{s}).$$

Jelikož charakteristická funkce určuje rozdělení náhodného vektoru jednoznačně, (\mathbf{X}, \mathbf{Y}) má stejné rozdělení jako $(\mathbf{X}', \mathbf{Y}')$, a tím pádem musí být i \mathbf{X} a \mathbf{Y} nezávislé.

□

V našem případě nás bude zajímat jednorozměrný případ pro $p = 1$, pro který platí $\varphi_X(t) = \mathbb{E}[e^{itX}]$, $t \in \mathbb{R}$.

2. Haughův test

První test, na který se zaměříme, vytvořil statistik Larry D. Haugh ve svém článku Haugh (1976). Rozhodl se svůj test zaměřit na sdruženě slabě stacionární řady, konkrétně ve svém článku konstruuje test pro dvě časové řady, které se řídí ARMA modely. Pro testování nezávislosti využívá křížovou korelační funkci.

2.1 Úvod

V podkapitole 1.3 jsme za platnosti určitých podmínek popsali některé z vlastností výběrové křížové kovarianční a korelační funkce. Haugh ve svém článku Haugh (1976) zkoumá vlastnosti výběrové křížové korelační funkce, které poté použije ke zkonstruování svého testu nezávislosti. Nejdůležitější vlastností odhadu ke zkonstruování použitelného testu je znát jeho asymptotické rozdělení za platnosti nulové hypotézy. Nulovou hypotézou Haughova testu je nezávislost dvou pozorovaných časových řad. Tu netestuje přímo na pozorovaných datech, ale testuje ji za pomoci příslušných bílých šumů. Nejdříve pro takové dva bílé šumy, které jsou navíc vzájemně nezávislé, spočteme rozptyl výběrové kovariance a kovarianci mezi samotnými odhady v jiných časových okamžicích.

Tvrzení 11 *Nechť máme k dispozici N pozorování dvou nezávislých bílých šumů $X = \{X_t : t \in \mathbb{Z}\}$ a $Y = \{Y_t : t \in \mathbb{Z}\}$. Pak pro $k, l \in \mathbb{Z}$, $|k|, |l| \leq N - 1$ a $k \neq l$ platí:*

$$\begin{aligned} \text{var } \tilde{R}_{X,Y}(k) &= \frac{N - |k|}{N^2} \text{var } X \text{ var } Y, \\ \text{cov}(\tilde{R}_{X,Y}(k), \tilde{R}_{X,Y}(l)) &= 0. \end{aligned}$$

Důkaz. Nechť $k \geq 0$. Pokud ve výpočtu rozptylu odhadu navážeme na důkaz věty 6, dostaneme dále:

$$\begin{aligned} \text{var } \tilde{R}_{X,Y}(k) &= \frac{1}{N^2} \sum_{t=1}^{N-k} \sum_{s=1}^{N-k} (\mathbb{E} X_t Y_{t+k} X_s Y_{s+k} - \mathbb{E} X_t Y_{t+k} \mathbb{E} X_s Y_{s+k}) \\ &= \frac{1}{N^2} \sum_{t=1}^{N-k} (\mathbb{E} X_t^2 \mathbb{E} Y_{t+k}^2 - \mathbb{E} X_t \mathbb{E} Y_{t+k} \mathbb{E} X_t \mathbb{E} Y_{t+k}) \\ &= \frac{N - k}{N^2} \text{var } X \text{ var } Y. \end{aligned}$$

Obecně pro $|k| \leq N - 1$ dostáváme $\text{var } \tilde{R}_{X,Y}(k) = \frac{N - |k|}{N^2} \text{var } X \text{ var } Y$.

Kovarianci odhadů spočteme pro nezáporná $k \neq l$ následovně:

$$\begin{aligned}
\text{cov}(\tilde{R}_{X,Y}(k), \tilde{R}_{X,Y}(l)) &= \mathbb{E}[\tilde{R}_{X,Y}(k)\tilde{R}_{X,Y}(l)] = \mathbb{E}\left[\left(\frac{1}{N} \sum_{t=1}^{N-k} X_t Y_{t+k}\right)\left(\frac{1}{N} \sum_{t=1}^{N-l} X_t Y_{t+l}\right)\right] \\
&= \frac{1}{N^2} \mathbb{E}\left[\sum_{t=1}^{N-k} \sum_{s=1}^{N-l} X_t X_s Y_{t+k} Y_{s+l}\right] \\
&= \frac{1}{N^2} \sum_{t=1}^{N-k} \sum_{s=1}^{N-l} \mathbb{E}(X_t X_s) \mathbb{E}(Y_{t+k} Y_{s+l}) \\
&= \frac{1}{N^2} \sum_{t=1}^{N-k} \sum_{s=1}^{N-l} R_X(s-t) R_Y(s+l-t-k) \\
&= 0.
\end{aligned}$$

Poslední rovnost plyne z toho, že jsou bílé šumy nekorelované a nemůže nastat $s = t$ a zároveň $s = t + k - l$. Dostáváme výsledek, že odhady výběrové kovariance v různých časových diferencích jsou za předpokladu nezávislosti navzájem nekorelované. □

2.2 Konstrukce ARMA modelů

V této části popíšeme obě řady, jejichž nezávislost chceme zkoumat, a jak odhadnout vhodný model. Necht' máme pozorování dvou sdruženě slabě stacionárních časových řad x_1, x_2, \dots, x_N a y_1, y_2, \dots, y_N , které se dají modelovat pomocí ARMA modelů:

$$\begin{aligned}
X_t + a_1 X_{t-1} + \dots + a_m X_{t-m} &= E_{X,t} + b_1 E_{X,t-1} + \dots + b_n E_{X,t-n}, \quad t \in \mathbb{Z}, \\
Y_t + c_1 Y_{t-1} + \dots + c_p Y_{t-p} &= E_{Y,t} + d_1 E_{Y,t-1} + \dots + d_q E_{Y,t-q}, \quad t \in \mathbb{Z},
\end{aligned}$$

kde hodnoty $a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_n$ a $c_1, c_2, \dots, c_p, d_1, d_2, \dots, d_q$ jsou neznámé parametry a je potřeba je odhadnout. Tedy platí, že $\{X_t : t \in \mathbb{Z}\}$ se řídí modelem ARMA(m, n) a $\{Y_t : t \in \mathbb{Z}\}$ se řídí modelem ARMA(p, q). Dále také předpokládáme, že jsou obě řady lineárním procesem.

Postup odhadnutí vhodného modelu ilustrujeme na řadě $\{X_t : t \in \mathbb{Z}\}$. K odhadnutí parametrů modelu ARMA(m, n) můžeme použít metodu podmíněných nejmenších čtverců. Zde máme k dispozici dva způsoby podmínění, které jsou převzaty z knihy Box a Jenkins (1970). Jeden spočívá v nahrazení prvních m hodnot bílého šumu nulou a minimalizování sumy:

$$\sum_{t=1+m}^N \hat{e}_{x,t}^2(\mathbf{a}, \mathbf{b}) = \sum_{t=1+m}^N (x_t + a_1 x_{t-1} + \dots + a_m x_{t-m} - b_1 \hat{e}_{x,t-1} - \dots - b_n \hat{e}_{x,t-n})^2,$$

kde $\hat{e}_{x,t} = 0$ pro $t = 1-n, 1-n+1, \dots, m-1, m$. Další způsob podmínění je založen na předpokladu, že všechny nepozorované hodnoty jsou nulové. Minimalizujeme tedy sumu:

$$\sum_{t=1}^N \hat{e}_{x,t}^2(\mathbf{a}, \mathbf{b}) = \sum_{t=1}^N (\hat{x}_t + a_1 \hat{x}_{t-1} + \dots + a_m \hat{x}_{t-m} - b_1 \hat{e}_{x,t-1} - \dots - b_n \hat{e}_{x,t-n})^2,$$

kde $\hat{e}_{x,t} = 0$ pro $t = 1 - n, 1 - n + 1, \dots, 0$, $\hat{x}_t = 0$ pro $t = 1 - m, 1 - m + 1, \dots, 0$ a $\hat{x}_t = x_t$ pro $t = 1, 2, \dots, N$. Asymptotické rozdělení testové statistiky nezávisí na volbě podmínění (viz Haugh (1976)).

Dalším úkolem je zvolit správně parametry m a n . Na to můžeme zvolit metodu porovnání hodnot výběrové autokorelační funkce s teoretickými hodnotami autokorelační funkce modelu ARMA, kde vybereme m a n takové, že si budou obě funkce nejvíce podobné tvarem. Druhá, více automatická metoda, spočívá ve výběru modelu pomocí informačních kritérií, kde vybereme model s nejnižší hodnotou daného informačního kritéria. Informační kritérium se skládá ze dvou částí. První částí je funkce, která určuje, jak moc sedí data na model. Může se například jednat o funkci výběrového rozptylu nebo věrohodnosti. Druhá část je penalizace za počet zvolených parametrů, jelikož platí, že čím více parametrů se pro odhad modelu použije, tím víc budou data odpovídat modelu. Bez penalizace by byl touto metodou vybrán vždy model s nejvíce parametry. Jedním z nejpoužívanějších kritérií je Akaikeho informační kritérium, které má následující předpis¹:

$$\text{AIC}(m, n) = \ln \hat{\sigma}_{m,n}^2 + \frac{2(m + n + 1)}{N},$$

kde $\hat{\sigma}_{m,n}^2$ je odhad rozptylu bílého šumu, pokud by se pro řadu předpokládal model ARMA(m, n). Po zvolení vhodného m a n a analogického odhadu p a q pro řadu Y a následného odhadnutí všech parametrů tak dostáváme finální odhadnuté předpisy:

$$\begin{aligned} X_t + \hat{a}_1 X_{t-1} + \dots + \hat{a}_m X_{t-m} &= E_{X,t} + \hat{b}_1 E_{X,t-1} + \dots + \hat{b}_n E_{X,t-n}, \quad t \in \mathbb{Z}, \\ Y_t + \hat{c}_1 Y_{t-1} + \dots + \hat{c}_p Y_{t-p} &= E_{Y,t} + \hat{d}_1 E_{Y,t-1} + \dots + \hat{d}_q E_{Y,t-q}, \quad t \in \mathbb{Z}. \end{aligned}$$

Nyní když máme odhadnuté modely pro obě řady, ověříme pomocí věty 8, zda jsou oba procesy lineární. Pokud ano, tak máme vše připravené k dalšímu kroku.

2.3 Test

Haugh se namísto přímo pozorovaných hodnot rozhodl využít odhadů parametrů, dopočítat odhady bílých šumů a testovat nezávislost na nich. Z tvrzení 9 víme, že pro lineární procesy platí, že pokud budou jejich příslušné bílé šumy nezávislé, tak poté budou i oba lineární procesy nezávislé.

Haugh volí postup výpočtu, který odpovídá druhému typu podmínění zmíněného výše a vypadá následovně:

$$\begin{aligned} \hat{e}_{x,1} &= x_1, \\ \hat{e}_{x,2} &= x_2 + \hat{a}_1 x_1 - \hat{b}_1 \hat{e}_{x,1}, \\ &\vdots \\ \hat{e}_{x,N} &= x_N + \hat{a}_1 x_{N-1} + \dots + \hat{a}_m x_{N-m} - \hat{b}_1 \hat{e}_{x,N-1} - \dots - \hat{b}_n \hat{e}_{x,N-n}. \end{aligned}$$

Analogicky se spočte odhad $\{\hat{e}_{y,t} : t = 1, 2, \dots, N\}$ bílého šumu pro druhou časovou řadu. Tento postup má výhodu v tom, že test bude mít k dispozici větší

¹Akaikeho informační kritérium můžeme v literatuře najít ve více tvarech, zmíněný vzorec je pouze jedna z možností.

počet dat než při výpočtu residuí za použití prvního způsobu podmínění, ale nevýhodu má v tom, že prvních m členů zanedbává AR strukturu a odhady tím pádem mohou být odchýlené. Proto by se v některých případech vyplatilo spočítat bílý šum i prvním způsobem podmínění, a to následovně:

$$\begin{aligned}\widehat{e}_{x,1+m} &= x_{1+m} + \widehat{a}_1 x_m + \cdots + \widehat{a}_m x_1, \\ \widehat{e}_{x,2+m} &= x_{2+m} + \widehat{a}_1 x_{1+m} + \cdots + \widehat{a}_m x_2 - \widehat{b}_1 \widehat{e}_{x,1+m}, \\ &\vdots \\ \widehat{e}_{x,N} &= x_N + \widehat{a}_1 x_{N-1} + \cdots + \widehat{a}_m x_{N-m} - \widehat{b}_1 \widehat{e}_{x,N-1} - \cdots - \widehat{b}_n \widehat{e}_{x,N-n},\end{aligned}$$

analogicky pro $\{\widehat{e}_{y,t} : t = 1 + p, 2 + p, \dots, N\}$. Dále budeme uvažovat vždy Haughův způsob výpočtu residuí, jelikož v simulaci budeme volit $m \leq 3$ a zanedbání prvních tří pozorování by mělo na výsledky pouze malý vliv.

Pro dva očištěné bílé šumy spočte výběrovou křížovou korelační funkci následovně:

$$\widetilde{r}_{\widehat{e}_x, \widehat{e}_y}(k) = \begin{cases} \frac{\sum_{t=1}^{N-k} \widehat{e}_{x,t} \widehat{e}_{y,t+k}}{\sqrt{\sum_{t=1}^N (\widehat{e}_{x,t})^2} \sqrt{\sum_{t=1}^N (\widehat{e}_{y,t})^2}}, & \text{pro } 0 \leq k \leq N-1, \\ \frac{\sum_{t=1-k}^N \widehat{e}_{x,t} \widehat{e}_{y,t+k}}{\sqrt{\sum_{t=1}^N (\widehat{e}_{x,t})^2} \sqrt{\sum_{t=1}^N (\widehat{e}_{y,t})^2}}, & \text{pro } 1-N \leq k < 0. \end{cases}$$

Základem celého Haughova článku je následující věta, která nám říká, že asymptotické rozdělení takto spočtených odhadů křížové korelace má stejné rozdělení jako odhady křížové korelace, které by se spočetly přímo z pozorování bílých šumů.

Věta 12 *Nechť máme N pozorování dvou sdruženě slabě stacionárních lineárních procesů $\{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$, které se řídí ARMA modely, kde $\{e_{x,t} : t = 1, 2, \dots, N\}$ a $\{e_{y,t} : t = 1, 2, \dots, N\}$ jsou reálné hodnoty příslušných bílých šumů. Pokud jsou časové řady nezávislé, pak vektor $\sqrt{N} \widetilde{\mathbf{r}}_{\widehat{e}_x, \widehat{e}_y} = \sqrt{N} (\widetilde{r}_{\widehat{e}_x, \widehat{e}_y}(k_1), \dots, \widetilde{r}_{\widehat{e}_x, \widehat{e}_y}(k_M))$ má stejné asymptotické rozdělení jako vektor $\sqrt{N} \widetilde{\mathbf{r}}_{e_x, e_y} = \sqrt{N} (\widetilde{r}_{e_x, e_y}(k_1), \dots, \widetilde{r}_{e_x, e_y}(k_M))$ a to $\mathcal{N}_M(\mathbf{0}, \mathbf{I}_M)$, $M \in \mathbb{N}$, kde $\mathbf{0}$ značí nulový vektor v \mathbb{R}^M a $k_1, k_2, \dots, k_M \in \{1-N, 2-N, \dots, N-1\}$ je M různých čísel.*

Důkaz. Haugh (1976), věta 1. □

Z věty 12 plyne, že za platnosti nezávislosti časových řad má $\sqrt{N} \widetilde{r}_{\widehat{e}_x, \widehat{e}_y}(k)$ asymptoticky $\mathcal{N}(0, 1)$ rozdělení a pro různá k, l jsou odhady $\sqrt{N} \widetilde{r}_{\widehat{e}_x, \widehat{e}_y}(k)$ a $\sqrt{N} \widetilde{r}_{\widehat{e}_x, \widehat{e}_y}(l)$ asymptoticky nekorelované, a protože jsou sdruženě normálně rozdělené, tak jsou i asymptoticky nezávislé. To je motivace k sestavení Haughova testu včetně testové statistiky.

Haughův test testuje následující hypotézu:

$$\begin{aligned}H_0 &: \{X_t : t \in \mathbb{Z}\} \text{ a } \{Y_t : t \in \mathbb{Z}\} \text{ jsou nezávislé,} \\ H_A &: \{X_t : t \in \mathbb{Z}\} \text{ a } \{Y_t : t \in \mathbb{Z}\} \text{ nejsou nezávislé.}\end{aligned}$$

Testové statistiky navrhuje dvě:

$$S_M = N \sum_{k=-M}^M \widetilde{r}_{\widehat{e}_x, \widehat{e}_y}(k)^2 \quad \text{nebo} \quad S_M^* = N^2 \sum_{k=-M}^M \frac{\widetilde{r}_{\widehat{e}_x, \widehat{e}_y}(k)^2}{N - |k|}.$$

Obě tyto statistiky mají za platnosti nulové hypotézy asymptoticky χ_{2M+1}^2 rozdělení, protože obě jsou sumy čtverců $2M + 1$ asymptoticky nezávislých $\mathcal{N}(0, 1)$ veličin. Jediný rozdíl je, že v prvním případě používáme aproximaci tvaru $\text{var } \tilde{r}_{\hat{e}_x, \hat{e}_y}(k) \approx \frac{1}{N}$ a ve druhém případě přesnější $\text{var } \tilde{r}_{\hat{e}_x, \hat{e}_y}(k) \approx \frac{N-|k|}{N^2}$, která vyplývá z tvrzení 11. Druhá aproximace je doporučeno používat pro velká $|k|$ vůči N a je doporučeno ji používat hlavně v případě, kdy zvolíme $M > N/10$. Obě aproximace použijeme v simulační části.

Shrnutí Haughova testu

Předpoklady: $\{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ jsou sdruženě slabě stacionární časové řady, řídí se modely ARMA a navíc jsou lineárními procesy.

Hypotéza:

$$H_0 : \{X_t : t \in \mathbb{Z}\} \text{ a } \{Y_t : t \in \mathbb{Z}\} \text{ jsou nezávislé,}$$

$$H_A : \{X_t : t \in \mathbb{Z}\} \text{ a } \{Y_t : t \in \mathbb{Z}\} \text{ nejsou nezávislé.}$$

Testová statistika:

$$S_M = N \sum_{k=-M}^M \tilde{r}_{\hat{e}_x, \hat{e}_y}(k)^2 \quad \text{nebo} \quad S_M^* = N^2 \sum_{k=-M}^M \frac{\tilde{r}_{\hat{e}_x, \hat{e}_y}(k)^2}{N - |k|},$$

kde $M < N$ je předem zvolené číslo.

Kritický obor: Zamítáme H_0 , pokud $s_M \geq \chi_{2M+1}^2(1 - \alpha)$, kde s_M je naměřená hodnota statistiky S_M nebo S_M^* a $\chi_{2M+1}^2(1 - \alpha)$ je $(1 - \alpha)$ -kvantil χ_{2M+1}^2 rozdělení.

P-hodnota: $p = 1 - G(s_M)$, kde G je distribuční funkce χ_{2M+1}^2 rozdělení.

2.4 Podobné testy

Na Haughův test navázalo spousta dalších matematiků a v této kapitole si jich pár představíme. Jedněmi z nich byli P. D. Koch a S.-S. Yang ve svém článku Koch a Yang (1986). Ti se za stejných předpokladů zaměřili na situaci, kdy se autokorelační funkce nechová jako normálně rozdělená náhodná veličina. Například když má autokorelační funkce více po sobě jdoucích kladných hodnot. Jejich testová statistika má tvar:

$$r_{M,i}^* = N \sum_{k=-M}^{M-i} \left[\sum_{l=0}^i \hat{r}_{\hat{e}_x, \hat{e}_y}(k+l) \right]^2$$

kde $M < N$ a $i \in \{0, 1, \dots, M - 1\}$, která má za platnosti nulové hypotézy nezávislosti přibližně $\beta_i \chi_{v_i}^2$ rozdělení, kde:

$$v_i = \frac{A^2}{B}, \quad \beta_i = \frac{B}{A},$$

$$A = 2(M + 1)(i + 1) - (i + 1)^2,$$

$$B = (2M + 1 - 2i)(i + 1)^2 + 2 \sum_{j=1}^i j^2 [2M + i - 3(j - 1)].$$

Pro $i = 0$ tato statistika odpovídá testové statistice S_M . Pro $i = 1$ dostaneme:

$$r_{M,1}^* = N \sum_{k=-M}^{M-1} [\widehat{r}_{\widehat{e}_x, \widehat{e}_y}(k) + \widehat{r}_{\widehat{e}_x, \widehat{e}_y}(k+1)]^2.$$

V tomto případě budou po sobě následující dva kladné (nebo dva záporné) odhady mít v testové statistice větší váhu než v Haughově testu. Na druhou stranu pokud nastane situace $\widehat{r}_{\widehat{e}_x, \widehat{e}_y}(k) = -\widehat{r}_{\widehat{e}_x, \widehat{e}_y}(k+1)$ pro každé k , test by nezávislost nezamítl, i když by časové řady korelované byly.

V článku Rehman a Malik (2014) autoři uvádějí nový způsob výpočtu reziduí pro dvě řady řídicí se autoregresním modelem ARMA(1,0). Nejdříve určí časový index t_1 a k odhadu parametru modelu použijí pouze prvních t_1 pozorování. Poté odhadnou modelem hodnotu v čase $t_1 + 1$ a následně spočítají pro tento čas také residuum. V dalším kroku využijí pro tvorbu modelu prvních $t_1 + 1$ členů, odhadnou člen $t_1 + 2$, pro něj spočítají residuum a pokračují analogicky až do N . Na konci dostanou $N - t_1$ hodnot reziduí pro obě časové řady a pro tato data spočtou Pearsonův korelační koeficient. Dvě časové řady, které jsou nekorelované, ale řídí se stejným předpisem, budou mít vysoké hodnoty klasických odhadů korelace. Zmíněný postup by měl tento problém řešit.

3. Modifikovaný t-test

Druhý test, na který se zaměříme, je modifikovaný t-test od autorů G. Alpargu a J. Buonaccorsi z jejich článku Alpargu a Buonaccorsi (2009). Tento test využívá klasický Pearsonův korelační koeficient a efektivní velikost výběru.

3.1 Úvod

Nejdříve se zaměříme na výběrový Pearsonův korelační koeficient. Ten má tvar:

$$\hat{r}_{X,Y} = \frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})(y_t - \bar{y}) \left(\frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^2 \right)^{-1/2} \left(\frac{1}{N} \sum_{t=1}^N (y_t - \bar{y})^2 \right)^{-1/2}.$$

Za předpokladu konstantních prvních a druhých momentů na pozorované řady $\{x_t, t = 1, 2, \dots, N\}$ a $\{y_t, t = 1, 2, \dots, N\}$ lze koeficient zapsat jako:

$$\hat{r}_{X,Y} = \frac{1}{N} \sum_{t=1}^N \hat{R}_{X,Y}(t, t) (\hat{\sigma}_x^2 \hat{\sigma}_y^2)^{-1/2} = \frac{1}{N} \sum_{t=1}^N \hat{r}_{X,Y}(t, t),$$

kde $\hat{R}_{X,Y}(t, t) = (x_t - \bar{x})(y_t - \bar{y})$ je odhad kovariance $R_{X,Y}(t, t)$ a $\hat{r}_{X,Y}(t, t) = \hat{R}_{X,Y}(t, t) (\hat{\sigma}_x^2 \hat{\sigma}_y^2)^{-1/2}$ je odhad korelace $r_{X,Y}(t, t)$. Jelikož nemáme k dispozici více pozorování v bodech t , bez jakýchkoliv dalších předpokladů bude $\hat{r}_{X,Y}(t, t)$ nejlepší pokus o odhad $r_{X,Y}(t, t)$. Pearsonův korelační koeficient je v tomto případě průměr korelací ve stejném čase. Samotný koeficient nám nedává žádnou informaci o korelacích v rozdílných časech a jeho hodnota nám ani nevypovídá o nekorelovanosti časových řad ve stejném čase, jelikož nulový koeficient neznamena nulovost korelací $\hat{r}_{X,Y}(t, t)$. Pokud budeme předpokládat, že pozorované řady jsou sdruženě slabě stacionární, Pearsonův korelační koeficient se stane výběrovou křížovou korelací v bodě 0, tedy $\hat{r}_{X,Y}(0)$. Stejně jako v předchozím případě nám toto číslo nenese žádnou informaci o korelacích v rozdílných časech, ale alespoň jeho nulovost značí nulovou korelaci ve stejných časech. Pearsonův korelační koeficient nabývá ještě většího významu, pokud budeme předpokládat, že máme k dispozici dva náhodné výběry. Poté je koeficient přímo odhadem korelace mezi dvěma náhodnými veličinami. A na závěr, pokud budeme předpokládat, že náhodné výběry mají sdruženě normální rozdělení, nekorelovanost implikuje nezávislost. Za těchto nejpřísnějších předpokladů je konstruován standardní t-test se statistikou:

$$T_N = \hat{r}_{X,Y} \sqrt{\frac{N-2}{1 - \hat{r}_{X,Y}^2}},$$

která má za platnosti nulové hypotézy, kterou je nezávislost, t_{N-2} rozdělení.

3.2 Test

Autoři článku Alpargu a Buonaccorsi (2009) se rozhodli tento test modifikovat na případ, že pozorované řady jsou slabě stacionární a normálně rozdělené. Testová statistika se tím pádem musí upravit o informaci, že pozorování v rámci jedné řady mohou být korelované. Proto se nejdřív provede odhad kovarianční matice

$$\Sigma_X = \begin{pmatrix} R_X(0) & R_X(1) & R_X(2) & \dots & R_X(N-1) \\ R_X(1) & R_X(0) & R_X(1) & \dots & R_X(N-2) \\ R_X(2) & R_X(1) & R_X(0) & \dots & R_X(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_X(N-1) & R_X(N-2) & R_X(N-3) & \dots & R_X(0) \end{pmatrix},$$

jako

$$\hat{\Sigma}_X = \begin{pmatrix} \hat{R}_X(0) & \hat{R}_X(1) & \hat{R}_X(2) & \dots & \hat{R}_X(N-1) \\ \hat{R}_X(1) & \hat{R}_X(0) & \hat{R}_X(1) & \dots & \hat{R}_X(N-2) \\ \hat{R}_X(2) & \hat{R}_X(1) & \hat{R}_X(0) & \dots & \hat{R}_X(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{R}_X(N-1) & \hat{R}_X(N-2) & \hat{R}_X(N-3) & \dots & \hat{R}_X(0) \end{pmatrix},$$

kde

$$\hat{R}_X(k) = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$$

je odhad autokovarianční funkce v bodě k . Analogicky se provede odhad $\hat{\Sigma}_Y$ kovarianční matice Σ_Y pro pozorování $\{y_t, t = 1, 2, \dots, N\}$. Za předpokladů klasického t-testu má Pearsonův korelační koeficient rozptyl přibližně $\frac{1}{N-1}$. Za předpokladů modifikovaného t-testu byl v článku Dutilleul a kol. (1993) odvozen rozptyl koeficientu jako

$$\widehat{\sigma^2} = \frac{\text{tr}(\mathbf{B}\hat{\Sigma}_X\mathbf{B}\hat{\Sigma}_Y)}{\text{tr}(\mathbf{B}\hat{\Sigma}_X)\text{tr}(\mathbf{B}\hat{\Sigma}_Y)}, \quad (3.1)$$

kde $\mathbf{B} = \mathbf{I}_N - \frac{1}{N}\mathbf{J}$, \mathbf{J} je čtvercová matice samých jedniček řádu N . Pomocí této hodnoty je upravena velikost výběru na efektivní velikost výběru srovnáním s rozptylem koeficientu za předpokladu náhodných výběrů jako $\tilde{N} = \frac{1}{\widehat{\sigma^2}} + 1$. Dostáváme upravenou testovou statistiku:

$$\tilde{T}_N = \hat{r}_{X,Y} \sqrt{\frac{\tilde{N} - 2}{1 - \hat{r}_{X,Y}^2}},$$

která má za platnosti nulové hypotézy $t_{\tilde{N}-2}$ rozdělení.

Tento test má jednu velkou výhodu oproti Haughově testu. Kromě normality nevyžaduje, aby se řady řídily nějakým konkrétním modelem, nazývá se proto model-free t-test. Na správný odhad modelu je často zapotřebí velké množství dat, které tento test nepožaduje. Funguje tedy lépe při malém počtu dat nebo pokud by se pro data zvolil špatný model, kterým se data neřídí. Jeho hlavní nevýhodou je malá síla. Ta se dá zvýšit bootstrap verzí testu, ale tato verze má zase nevýhodu v tom, že je méně robustní. Další velká nevýhoda tohoto testu je,

že neodhalí případy, kdy je křížová korelace sice nulová pro $k = 0$, ale je nenulová pro nějaké $k \neq 0$. V simulační části práce použijeme tento test jak na pozorování řad X a Y , tak i na odhady bílých šumů E_X a E_Y .

Shrnutí modifikovaného t-testu

Předpoklady: $\{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ jsou slabě stacionární časové řady, které mají mnohorozměrné normální rozdělení.

Hypotéza:

$H_0 : \{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ jsou nezávislé,

$H_A : \{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ nejsou nezávislé.

Testová statistika:

$$\tilde{T}_N = \hat{r}_{X,Y} \sqrt{\frac{\tilde{N} - 2}{1 - \hat{r}_{X,Y}^2}}.$$

Kritický obor: Zamítáme H_0 , pokud $|\tilde{t}_N| \geq t_{\tilde{N}-2}(1 - \alpha/2)$, kde \tilde{t}_N je naměřená hodnota statistiky \tilde{T}_N a $t_{\tilde{N}-2}(1 - \alpha/2)$ je $(1 - \alpha/2)$ -kvantil $t_{\tilde{N}-2}$ rozdělení.

P-hodnota: $p = 2(1 - G(|\tilde{t}_N|))$, kde G je distribuční funkce $t_{\tilde{N}-2}$ rozdělení.

4. Test distanční kovariancí

Nyní se zaměříme na teorii distanční kovariance z článku Székely a kol. (2007). V tomto článku autoři navrhují distanční kovarianci a korelaci jako lepší variantu Pearsonova korelačního koeficientu. Autoři představují test nezávislosti pro dva náhodné výběry. Tento test přizpůsobíme pro případ dvou časových řad.

4.1 Úvod

Pro Pearsonův korelační koeficient platí, že je nulový, pokud jsou dvě náhodné veličiny nezávislé. Opačná implikace obecně neplatí. Toto motivovalo autory článku vymyslet koeficient, pro který platí obě implikace nehledě na rozdělení veličin. Rozhodli se využít charakteristických funkcí, pro které platí ekvivalence z tvrzení 10. Navíc charakteristické funkce jsou definovány i pro náhodné vektory a to umožňuje korelovat vícerozměrné objekty. Nechť \mathbf{X} je p -rozměrný a \mathbf{Y} q -rozměrný náhodný vektor. Distanční korelaci definují takovým způsobem, aby odrážela rozdíl sdružené charakteristické funkce od součinu marginálních, tedy rozdíl $\|\varphi_{\mathbf{X},\mathbf{Y}} - \varphi_{\mathbf{X}}\varphi_{\mathbf{Y}}\|$, kde $\|\cdot\|$ je vhodná norma. Pro naše účely stačí místo \mathbf{X} a \mathbf{Y} uvažovat jednorozměrné X a Y , a proto se budeme zabývat pouze tímto případem. Souvislost s časovými řadami ujasníme v podkapitole 4.5.

4.2 Teorie distanční kovariance

V této části definujeme distanční kovarianci a korelaci a dále se budeme zabývat jejich základními vlastnostmi.

Definice 13 Nechť X a Y jsou reálné náhodné veličiny s konečnou střední hodnotou. **Distanční kovarianci** mezi veličinami X a Y definujeme jako nezáporné číslo $\mathcal{V}(X, Y)$, které splňuje:

$$\mathcal{V}^2(X, Y) := \|\varphi_{X,Y} - \varphi_X\varphi_Y\|^2 = \frac{1}{\pi^2} \int_{\mathbb{R}^2} \frac{|\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2}{t^2 s^2} dt ds.$$

Dále pro X a Y definujeme **distanční korelaci** jako nezáporné číslo $\mathcal{R}(X, Y)$ splňující:

$$\mathcal{R}^2(X, Y) := \begin{cases} \frac{\mathcal{V}^2(X, Y)}{\mathcal{V}(X)\mathcal{V}(Y)}, & \mathcal{V}(X)\mathcal{V}(Y) > 0, \\ 0, & \mathcal{V}(X)\mathcal{V}(Y) = 0, \end{cases}$$

kde nezáporné číslo $\mathcal{V}(X)$ splňující $\mathcal{V}^2(X) = \mathcal{V}^2(X, X) = \|\varphi_{X,X} - \varphi_X\varphi_X\|^2$ nazýváme **distančním rozptylem**.

Poznámka: Pro náhodnou veličinu X platí:

$$\varphi_{X,X}(t, s) = \mathbb{E}[e^{i(tX+sX)}] = \mathbb{E}[e^{iX(t+s)}] = \varphi_X(t+s).$$

Nyní uvedeme dva příklady výpočtu distanční kovariance a korelace. V prvním případě uvedeme výsledky pro dvě normálně rozdělené veličiny a poté následuje příklad s exponenciálním rozdělením. První příklad je převzatý z článku Székely a kol. (2007) a druhý příklad je zpracován autorem diplomové práce. Na výpočet integrálů byl využit software Mathematica (Wolfram Research, Inc. (2023)).

Příklad: (Dvourozměrné normální rozdělení) Necht X a Y mají rozdělení $\mathcal{N}(0, 1)$ a $\text{corr}(X, Y) = \rho$. Poté platí:

$$\begin{aligned}\mathcal{V}^2(X) &= \frac{4}{\pi} \left(\frac{\pi}{3} + 1 - \sqrt{3} \right), \\ \mathcal{V}^2(X, Y) &= \frac{4}{\pi} \left(\rho \arcsin \rho + \sqrt{1 - \rho^2} - \rho \arcsin \frac{\rho}{2} - \sqrt{4 - \rho^2} + 1 \right), \\ \mathcal{R}^2(X, Y) &= \frac{\rho \arcsin \rho + \sqrt{1 - \rho^2} - \rho \arcsin \frac{\rho}{2} - \sqrt{4 - \rho^2} + 1}{\frac{\pi}{3} + 1 - \sqrt{3}}.\end{aligned}$$

Příklad: (Exponenciální rozdělení) Necht X a Y jsou dvě nezávislé náhodné veličiny, které se řídí exponenciálním rozdělením obě se střední hodnotou rovné jedné, a necht $Z = X + Y$. Naším cílem bude spočítat $\mathcal{V}(X, Z)$ a $\mathcal{R}(X, Z)$. Z definice exponenciálního rozdělení a díky větě o transformaci hustot dostáváme následující hustoty veličin X , Z a vektoru (X, Z) :

$$\begin{aligned}f_X(x) &= \begin{cases} e^{-x}, & x \geq 0, \\ 0, & \text{jinak,} \end{cases} \\ f_Z(z) &= \begin{cases} ze^{-z}, & z \geq 0, \\ 0, & \text{jinak,} \end{cases} \\ f_{(X,Z)}(x, z) &= \begin{cases} e^{-z}, & z \geq x, x \geq 0, \\ 0, & \text{jinak.} \end{cases}\end{aligned}$$

Jako další krok spočítáme charakteristické funkce pro všechny tři případy:

$$\begin{aligned}\varphi_X(t) &= \frac{i}{t + i}, \\ \varphi_Z(s) &= -\frac{1}{(s + i)^2}, \\ \varphi_{X,Z}(t, s) &= -\frac{1}{(s + i)(s + t + i)},\end{aligned}$$

kde $t, s \in \mathbb{R}$. Dále spočítáme, jak moc se liší reálná sdružená charakteristická funkce vektoru (X, Z) od případu, kdy X a Z by byly nezávislé náhodné veličiny:

$$\begin{aligned}|\varphi_{X,Z}(t, s) - \varphi_X(t)\varphi_Z(s)|^2 &= \left| -\frac{1}{(s + i)(s + t + i)} + \frac{i}{t + i} \frac{1}{(s + i)^2} \right|^2 \\ &= \left| \frac{st}{(s + i)^2(s + t + i)(t + i)} \right|^2 \\ &= \frac{s^2 t^2}{(s^2 + 1)^2 (t^2 + 1) ((s + t)^2 + 1)}.\end{aligned}$$

Nyní už stačí pouze vyčíslit $\mathcal{V}^2(X, Z)$:

$$\mathcal{V}^2(X, Z) = \frac{1}{\pi^2} \int_{\mathbb{R}^2} \frac{1}{(s^2 + 1)^2 (t^2 + 1) ((s + t)^2 + 1)} dt ds = \frac{1}{\pi^2} \frac{2\pi^2}{9} = \frac{2}{9}.$$

Analogicky můžeme spočítat $\mathcal{V}^2(X)$ a $\mathcal{V}^2(Z)$:

$$\mathcal{V}^2(X) = \frac{1}{\pi^2} \int_{\mathbb{R}^2} \frac{1}{(s^2 + 1)(t^2 + 1)((s + t)^2 + 1)} dt ds = \frac{1}{\pi^2} \frac{\pi^2}{3} = \frac{1}{3},$$

$$\mathcal{V}^2(Z) = \frac{1}{\pi^2} \int_{\mathbb{R}^2} \frac{(st - 2)^2 + 4(s + t)^2}{(s^2 + 1)^2(t^2 + 1)^2((s + t)^2 + 1)^2} dt ds = \frac{1}{\pi^2} \frac{79\pi^2}{108} = \frac{79}{108}.$$

Celkem dostaneme:

$$\mathcal{R}^2(X, Z) = \frac{\mathcal{V}^2(X, Z)}{\mathcal{V}^2(X)\mathcal{V}^2(Z)} = \frac{\frac{2}{9}}{\sqrt{\frac{1}{3}}\sqrt{\frac{79}{108}}} = \frac{4}{\sqrt{79}}.$$

Výsledná distanční kovariance mezi veličinami X a Z je $\frac{\sqrt{2}}{3} \doteq 0,471$ a distanční korelace je rovna $\frac{2}{\sqrt{79}} \doteq 0,671$. Pro srovnání klasická korelace mezi X a Z vychází $\frac{1}{\sqrt{2}} \doteq 0,707$.

Jak bylo zmíněno v úvodu kapitoly, koeficient distanční korelace byl vymyšlen se záměrem, aby jeho nulovost byla ekvivalentní s nezávislostí. V následující větě se o tom přesvědčíme.

Věta 13 *Pro náhodné veličiny X a Y s konečnou střední hodnotou platí:*

$$X \text{ a } Y \text{ jsou nezávislé} \iff \mathcal{R}^2(X, Y) = 0.$$

Důkaz. \implies : Pro nezávislé náhodné veličiny podle věty 10 získáváme $|\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)| = 0$ pro každé $t, s \in \mathbb{R}$. Tedy platí $\mathcal{V}^2(X, Y) = 0$ a z toho vyplývá nulovost $\mathcal{R}^2(X, Y)$.

\impliedby : Necht $\mathcal{V}^2(X)\mathcal{V}^2(Y) > 0$. Pak $\mathcal{R}^2(X, Y) = 0 \iff \mathcal{V}^2(X, Y) = 0$. Toto nastane pouze v případě, pokud $|\varphi_{X,Y}(t, s) - \varphi_X(t)\varphi_Y(s)|^2 = 0 \forall t, s \in \mathbb{R}$. Musí tedy platit $\varphi_{X,Y}(t, s) = \varphi_X(t)\varphi_Y(s) \forall t, s \in \mathbb{R}$. Platnost tohoto tvrzení vyplývá z věty 10. Pokud $\mathcal{V}^2(X)\mathcal{V}^2(Y) = 0$, tak musí platit $\mathcal{V}^2(X) = 0$ nebo $\mathcal{V}^2(Y) = 0$. BÚNO: Necht $\mathcal{V}^2(X) = 0$. To nastane právě tehdy, když $\mathbb{E}[e^{iX(t+s)}] = \mathbb{E}[e^{iXt}] \mathbb{E}[e^{iXs}] \forall t, s \in \mathbb{R}$. Toto nastane pouze v případě, že X je konstanta skoro jistě. Tím pádem jsou X a Y nezávislé. □

Poslední věta v této kapitole nám udává meze pro hodnoty distanční korelace.

Věta 14 *Pro náhodné veličiny X a Y s konečnou střední hodnotou platí:*

$$0 \leq \mathcal{R}(X, Y) \leq 1$$

Důkaz. Důkaz v obecnější formě lze najít v článku Lyons (2013), věta 2.3. □

4.3 Výběrová distanční kovariance

V této části zavedeme výběrovou distanční kovarianci a korelaci pro náhodné výběry jako odhad distanční kovariance a korelace a popíšeme jejich vztah k charakteristickým funkcím.

Nechť x_1, x_2, \dots, x_N jsou pozorování náhodného výběru. Pro tento výběr a indexy $k, l \in \{1, 2, \dots, N\}$ definujeme následující statistiky:

$$a_{kl} = |x_k - x_l|,$$

$$\bar{a}_{k\cdot} = \frac{1}{N} \sum_{l=1}^N a_{kl}, \quad \bar{a}_{\cdot l} = \frac{1}{N} \sum_{k=1}^N a_{kl}, \quad \bar{a}_{\cdot\cdot} = \frac{1}{N^2} \sum_{k,l=1}^N a_{kl},$$

$$A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot}.$$

Analogicky definujeme statistiky $b_{kl}, \bar{b}_{k\cdot}, \bar{b}_{\cdot l}, \bar{b}_{\cdot\cdot}, B_{kl}$ pro pozorování náhodného výběru y_1, y_2, \dots, y_N . Výběrovou distanční kovarianci $\hat{\mathcal{V}}_N(X, Y)$, distanční rozptyly $\hat{\mathcal{V}}_N(X)$ a $\hat{\mathcal{V}}_N(Y)$ a distanční korelaci $\hat{\mathcal{R}}_N(X, Y)$ definujeme jako nezáporná čísla splňující následující:

$$\hat{\mathcal{V}}_N^2(X, Y) = \frac{1}{N^2} \sum_{k,l=1}^N A_{kl} B_{kl}, \quad \hat{\mathcal{V}}_N^2(X) = \frac{1}{N^2} \sum_{k,l=1}^N A_{kl}^2, \quad \hat{\mathcal{V}}_N^2(Y) = \frac{1}{N^2} \sum_{k,l=1}^N B_{kl}^2,$$

$$\hat{\mathcal{R}}_N^2(X, Y) = \begin{cases} \frac{\hat{\mathcal{V}}_N^2(X, Y)}{\hat{\mathcal{V}}_N(X) \hat{\mathcal{V}}_N(Y)}, & \hat{\mathcal{V}}_N(X) \hat{\mathcal{V}}_N(Y) > 0, \\ 0, & \hat{\mathcal{V}}_N(X) \hat{\mathcal{V}}_N(Y) = 0. \end{cases}$$

Na první pohled není jasné, proč výběrový odhad distanční kovariance nevyužívá výběrové odhady charakteristických funkcí, které je přirozené definovat následovně:

$$\hat{\varphi}_{X,Y}(t, s) = \frac{1}{N} \sum_{k=1}^N \exp \{i(tx_k + sy_k)\},$$

$$\hat{\varphi}_X(t) = \frac{1}{N} \sum_{k=1}^N \exp \{i tx_k\}, \quad \hat{\varphi}_Y(s) = \frac{1}{N} \sum_{k=1}^N \exp \{i sy_k\}.$$

Následující věta nám tento princip osvětlí.

Věta 15 *Nechť X_1, X_2, \dots, X_N a Y_1, Y_2, \dots, Y_N jsou náhodné výběry. Poté platí:*

$$\hat{\mathcal{V}}_N^2(X, Y) = \|\hat{\varphi}_{X,Y} - \hat{\varphi}_X \hat{\varphi}_Y\|^2.$$

Důkaz. Důkaz v obecnější formě lze najít v článku Székely a kol. (2007), věta 1. □

Z věty 15 vyplývá, že odhad $\hat{\mathcal{V}}_N^2(X, Y)$ je přirozeným odhadem $\mathcal{V}^2(X, Y)$. Na rozdíl od odhadů charakteristických funkcí je ale velmi snadný na výpočet.

4.4 Asymptotické vlastnosti výběrové distanční kovariance

Nyní se podíváme na vybrané asymptotické vlastnosti odhadů. Jednou z nej důležitějších vlastností odhadů je konzistence.

Věta 16 *Nechť platí $\mathbb{E}|X| < \infty$ a $\mathbb{E}|Y| < \infty$. Poté platí:*

$$\widehat{\mathcal{V}}_N^2(X, Y) \xrightarrow[N \rightarrow \infty]{\text{s.j.}} \mathcal{V}^2(X, Y).$$

Důkaz. Székely a kol. (2007), věta 2. □

Pomocí věty 16 dostáváme následující důsledek.

Důsledek 2 *Nechť platí $\mathbb{E}|X| < \infty$, $\mathbb{E}|Y| < \infty$. Poté platí:*

$$\widehat{\mathcal{R}}_N^2(X, Y) \xrightarrow[N \rightarrow \infty]{\text{s.j.}} \mathcal{R}^2(X, Y).$$

Důkaz. Důkaz je obdobný jako v důsledku 1. □

Výběrová distanční kovariance i korelace jsou silně konzistentními odhady. Také můžeme pozorovat, že ke konzistenci stačí předpokládat pouze konečnost prvních absolutních momentů.

K sestrojení efektivního testu nás bude nejvíce zajímat asymptotické rozdělení odhadů. V další větě uvedeme, jak se odhady chovají asymptoticky, a to jak za předpokladu nezávislosti, tak také za jeho neplatnosti. Nejdříve ale zavedeme značení, které budeme dále používat:

$$S_N = \frac{1}{N^4} \sum_{k,l=1}^N a_{kl} \sum_{k,l=1}^N b_{kl},$$

$$Q = \sum_{j=1}^{\infty} \lambda_j Z_j^2,$$

kde Z_j jsou nezávislé $\mathcal{N}(0, 1)$ rozdělené náhodné veličiny a λ_j jsou konstanty závislé na rozdělení (X, Y) .

Věta 17 *Nechť platí $\mathbb{E}|X| + \mathbb{E}|Y| < \infty$. Potom:*

1) *Pokud X a Y jsou nezávislé, pak platí*

$$\frac{N\widehat{\mathcal{V}}_N^2(X, Y)}{S_N} \xrightarrow[N \rightarrow \infty]{\text{D}} Q$$

a navíc platí $\mathbb{E}Q = 1$.

2) *Pokud X a Y nejsou nezávislé, pak platí*

$$\frac{N\widehat{\mathcal{V}}_N^2(X, Y)}{S_N} \xrightarrow[N \rightarrow \infty]{\text{P}} \infty.$$

Důkaz. Székely a kol. (2007), důsledek 2. □

Pokud bychom k testování nulové hypotézy nezávislosti využili jako testovou statistiku výraz $\frac{N\widehat{\mathcal{V}}_N^2(X, Y)}{S_N}$, statistika by v případě její neplatnosti divergovala k nekonečnu, zatímco v případě její platnosti by se chovala jako náhodná veličina mající zobecněné χ^2 rozdělení se střední hodnotou rovné jedné. Stejně jako v případě konzistence stačí předpokládat pouze konečnost prvních absolutních momentů. Tento velmi slabý předpoklad můžeme v praxi ve většině případů považovat za splněný. Jediné, co ještě potřebujeme, je mít k dispozici dva náhodné výběry. Nyní když víme základní asymptotické chování odhadu, můžeme se podívat na popis testu distanční kovariancí.

4.5 Test

V článku Székely a kol. (2007) autoři konstruují test pomocí výběrové distanční kovariance následovně:

Předpoklady: $\{X_t : t = 1, 2, \dots, N\}$ a $\{Y_t : t = 1, 2, \dots, N\}$ jsou náhodné výběry.

Hypotéza:

$$H_0 : \{X_t : t = 1, 2, \dots, N\} \text{ a } \{Y_t : t = 1, 2, \dots, N\} \text{ jsou nezávislé,}$$

$$H_A : \{X_t : t = 1, 2, \dots, N\} \text{ a } \{Y_t : t = 1, 2, \dots, N\} \text{ nejsou nezávislé.}$$

Testová statistika:

$$V_N = \sqrt{\frac{N\hat{\mathcal{V}}_N^2(X, Y)}{S_N}}.$$

Kritický obor: Zamítáme H_0 , pokud $v_N \geq u(1 - \alpha/2)$, kde v_N je naměřená hodnota statistiky V_N a $u(1 - \alpha/2)$ je $(1 - \alpha/2)$ -kvantil $\mathcal{N}(0, 1)$ rozdělení.

P-hodnota: $p = 2(1 - \Phi(v_N))$, kde Φ je distribuční funkce $\mathcal{N}(0, 1)$ rozdělení.

V následující větě se podíváme na to, jaké hladiny tento test dosahuje.

Věta 18 Označme α_N dosaženou hladinu testu nezávislosti pomocí distanční kovariance na hladině $0 < \alpha \leq 0,215$. Pro tuto hodnotu platí:

$$\lim_{N \rightarrow \infty} \alpha_N \leq \alpha$$

Důkaz. Székely a kol. (2007), věta 6. □

Věta 18 nám říká, že reálná hladina provedeného testu bude asymptoticky nejvýše dosahovat požadované hladiny. Tím pádem test zamítá nulovou hypotézu častěji, než by bylo žádané. Takové testy nazýváme konzervativní. Tato nerovnost je výsledkem aproximace asymptotického rozdělení testové statistiky pomocí normálního rozdělení namísto rozdělení veličiny \sqrt{Q} , jejíž rozdělení by bylo velmi složité získat a navíc záleží na rozdělení (X, Y) .

V našem případě časových řad $\{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ nejsou náhodnými výběry, a tak tento test nemůžeme použít, aniž bychom udělali nějaké úpravy. Stejně jako v kapitole 2 budeme chtít uplatnit tvrzení 9. Navíc ale potřebujeme, abychom místo bílých šumů měli k dispozici náhodné výběry. Předpokládáme, že bílý šum je posloupnost stejně rozdělených náhodných veličin, je přirozené. Jediné, co ještě potřebujeme předpokládat, je nezávislost. Nezávislost můžeme také snadno získat pomocí předpokladu, že oba bílé šumy mají sdružené normální rozdělení.

Postup navrhovaného testu bude kombinace Haughova testu a testu distanční kovariancí. Máme k dispozici pozorování dvou lineárních procesů x_1, x_2, \dots, x_N a y_1, y_2, \dots, y_N . Nejdříve je potřeba odhadnout modely $\text{ARMA}(m, n)$ resp. $\text{ARMA}(p, q)$ pro $\{X_t : t \in \mathbb{Z}\}$ resp. $\{Y_t : t \in \mathbb{Z}\}$. Po odhadnutí modelů spočítáme hodnoty bílých šumů $\hat{e}_{x,t}$ a $\hat{e}_{y,t}$. Poté už stačí pouze spočítat statistiky $\hat{\mathcal{V}}_N^2(\hat{E}_X, \hat{E}_Y)$ a S_N a nakonec použít test distanční kovariancí.

Shrnutí testu pomocí distanční kovariance

Předpoklady: $\{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ se řídí modely ARMA, jsou oba lineárními procesy a příslušné bílé šумы jsou navíc posloupností stejně rozdělených nezávislých náhodných veličin.

Hypotéza:

$H_0 : \{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ jsou nezávislé,

$H_A : \{X_t : t \in \mathbb{Z}\}$ a $\{Y_t : t \in \mathbb{Z}\}$ nejsou nezávislé.

Testová statistika:

$$V_N = \sqrt{\frac{N\hat{\mathcal{V}}_N^2(\hat{E}_X, \hat{E}_Y)}{S_N}}.$$

Kritický obor: Zamítáme H_0 , pokud $v_N \geq u(1 - \alpha/2)$, kde v_N je naměřená hodnota statistiky V_N a $u(1 - \alpha/2)$ je $(1 - \alpha/2)$ -kvantil $\mathcal{N}(0, 1)$ rozdělení.

P-hodnota: $p = 2(1 - \Phi(v_N))$, kde Φ je distribuční funkce $\mathcal{N}(0, 1)$ rozdělení.

5. Porovnání testů

Cílem této kapitoly je porovnat výše zmíněné testy spolu se standardním t-testem v různých nastaveních na simulovaných datech. Na výpočty používáme programovací jazyk R (R Core Team (2022)), ve kterém jsme připravili funkce pro všechny tři výše popsané testy.

5.1 Data

Vstupní data budou generována následovně. Nejdříve je potřeba vygenerovat bílé šумы E_X a E_Y . Pro ty budeme předpokládat dvourozměrné normální rozdělení:

$$\begin{pmatrix} E_{X,t} \\ E_{Y,t} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad t \in \mathbb{Z},$$

kde ρ je mezi -1 a 1 . Právě pomocí parametru ρ budeme ovlivňovat, jak moc jsou časové řady závislé. Dvě pozorování bílých šumů, ať už v rámci jedné řady nebo napříč oběma řadami, budou vždy pro různé časy nezávislé. Nyní když máme k dispozici bílé šумы, tak se můžeme vrhnout na generování řad X , která se řídí modelem $\text{ARMA}(m, n)$, a Y , která se řídí modelem $\text{ARMA}(p, q)$. Pro prvních m (resp. p) pozorování však nemáme k dispozici všechny potřebné pozorování X (resp. Y) a pro prvních n (resp. q) pozorování nemáme k dispozici ani potřebné bílé šумы. Proto pro vygenerování N pozorování bude potřeba vygenerovat $N+Q$ pozorování, kde poté prvních Q pozorování zahodíme, jelikož jsou potřeba pouze ke startu modelu a mohly by zhoršit následné odhady parametrů.

Poznámka: Díky předpokladu korelovanosti bílých šumů ve stejném časovém okamžiku dáváme šanci t-testům, které nedokážou odhalit korelaci, která by se vyskytovala mezi časy t a $t+k$ pro $k \neq 0$. Tento případ v simulaci nebyl zkoumán.

Testy budeme porovnávat na dvou nastaveních předpisů ARMA . V prvním případě se bude X řídit modelem $\text{ARMA}(1, 0)$ a Y modelem $\text{ARMA}(0, 1)$ s předpisy:

$$X_t + 0,8X_{t-1} = E_{X,t}, \quad (5.1)$$

$$Y_t = E_{Y,t} - 0,4E_{Y,t-1}. \quad (5.2)$$

Toto nastavení budeme nazývat nastavením s **jednoduchými modely**. V druhém případě se bude X řídit modelem $\text{ARMA}(2, 2)$ a Y modelem $\text{ARMA}(1, 3)$ s předpisy:

$$X_t - 0,95X_{t-1} + 0,1X_{t-2} = E_{X,t} - 0,1E_{X,t-1} - 0,5E_{X,t-2}, \quad (5.3)$$

$$Y_t + 0,5Y_{t-1} = E_{Y,t} + 0,6E_{Y,t-1} + 0,2E_{Y,t-2} - 0,5E_{Y,t-3}. \quad (5.4)$$

Toto nastavení nazveme nastavením s **komplexními modely**. Můžeme tak sledovat, jestli se výsledky budou lišit na základě předpisů řad.

Poznámka: Takto generované řady splňují potřebné předpoklady pro všechny testy:

1) Lineární proces: Využijeme tvrzení 8. Pro (5.1) dostáváme polynomy:

$$a(z) = 1 + 0,8z, b(z) = 1.$$

Polynom $a(z)$ má jediný kořen $-1,25$ a $b(z)$ žádné kořeny nemá. Pro (5.2) získáváme:

$$a(z) = 1, b(z) = 1 - 0,4z.$$

Polynom $a(z)$ kořeny nemá a $b(z)$ má kořen $2,5$. Následují polynomy pro předpis (5.3):

$$a(z) = 1 - 0,95z + 0,1z^2, b(z) = 1 - 0,1z - 0,5z^2.$$

Kořeny polynomu $a(z)$ jsou přibližně $1,206$ a $8,294$ a polynom $b(z)$ má kořeny přibližně rovné $1,318$ a $-1,518$. Jako poslední polynomy pro (5.4):

$$a(z) = 1 + 0,5z, b(z) = 1 + 0,6z + 0,2z^2 - 0,5z^3.$$

Kořen pro $a(z)$ vychází -2 a kořeny pro $b(z)$ vychází přibližně $1,745$ a $-0,672 \pm 0,833i$. Ve všech čtyřech případech mají polynomy $a(z)$ a $b(z)$ různé kořeny a navíc kořeny polynomu $a(z)$ leží vně jednotkového kruhu. Tím pádem jsou vždy splněny předpoklady tvrzení 8 a ve všech čtyřech případech se jedná o lineární procesy.

2) Slabá stacionarita: Lineární procesy jsou automaticky slabě stacionárními řadami podle věty 7.

3) Normální rozdělení: Z definice lineárního procesu 11 víme, že existují konstanty $\{g_{X,j}, j \in \mathbb{N}_0\}$ splňující $\sum_{j=0}^{\infty} |g_{X,j}| < \infty$ tak, že pro každé $t \in \mathbb{Z}$ platí identita $X_t = \sum_{j=0}^{\infty} g_{X,j} E_{X,t-j}$. Jelikož je $E_{X,t}$ posloupnost nekorelovaných normálně rozdělených veličin, X_t bude mít rozdělení $\mathcal{N}(0, \sum_{j=0}^{\infty} g_{X,j}^2)$, kde konečnost $\sum_{j=0}^{\infty} g_{X,j}^2$ vyplývá z konečnosti $\sum_{j=0}^{\infty} |g_{X,j}| < \infty$. Analogicky by se dokázala normalita pro řadu Y .

4) Sdružená slabá stacionarita: Pro takto generované řady platí:

$$\begin{aligned} R_{X,Y}(t, s) &= \text{cov}(X_t, Y_s) = \mathbb{E}[X_t Y_s] = \mathbb{E}\left[\left(\sum_{j=0}^{\infty} g_{X,j} E_{X,t-j}\right)\left(\sum_{k=0}^{\infty} g_{Y,k} E_{Y,s-k}\right)\right] \\ &= \mathbb{E}\left[\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} g_{X,j} E_{X,t-j} g_{Y,k} E_{Y,s-k}\right] = \mathbb{E}\left[\sum_{j=0}^{\infty} g_{X,j} g_{Y,s-t+j} E_{X,t-j} E_{Y,t-j}\right] \\ &= \sum_{j=0}^{\infty} g_{X,j} g_{Y,s-t+j} \rho = R_{X,Y}(s-t). \end{aligned}$$

Pro $k \in \mathbb{Z}$ dostáváme rovnosti:

$$R_{X,Y}(k) = \rho \sum_{j=0}^{\infty} g_{X,j} g_{Y,j+k}, \quad r_{X,Y}(k) = \rho \frac{\sum_{j=0}^{\infty} g_{X,j} g_{Y,j+k}}{\sqrt{\sum_{j=0}^{\infty} g_{X,j}^2} \sqrt{\sum_{j=0}^{\infty} g_{Y,j}^2}}. \quad (5.5)$$

5) Bílé šумы jsou náhodné výběry: Pozorování bílých šumů jsou v čase generována nezávisle na sobě a řídí se stejným rozdělením, tudíž se jedná o náhodný výběr.

Testy chceme porovnat nejenom na základě zvolených předpisů pro časové řady, ale taky v závislosti na počtu dostupných pozorování. Proto byla simulace provedena s počtem pozorování N rovným 50, 100 a 200. Je potřeba mít na mysli, že tento počet ovlivňuje sílu testů jak přímo, tak i nepřímo přes kvalitu odhadnutých modelů ARMA, která poté může ovlivnit výsledky testů. Pro nastartování modelu ARMA bylo použito $Q = 30$ ve všech případech.

Dále je potřeba otestovat, jak testy reagují na sílu korelace mezi bílými šumy. Korelace ρ byla zvolena postupně od -1 do 1 s krokem $0,1$, kde $\rho = 0$ odpovídá platnosti nulové hypotézy nezávislosti časových řad.

Celkem dostáváme 2 možné kombinace předpisů řad ARMA, 3 varianty délky testovaných řad a 21 možností pro korelaci ρ , dohromady 126 testovaných nastavení.

Než se pustíme do porovnání, tak ukážeme dva příklady možných trajektorií testovaných řad.

Příklad: Jako první se podíváme na jednu trajektorii řad řídicích se jednoduchými modely (5.1) a (5.2), délku řad zvolíme $N = 50$ a korelaci bílých šumů $\rho = 0,5$. Na grafu 5.1 vidíme jeden z možných průběhů tohoto nastavení. Po odhadnutí modelů ARMA pro obě řady dostaneme odhady bílých šumů znázorněné v grafu 5.2.

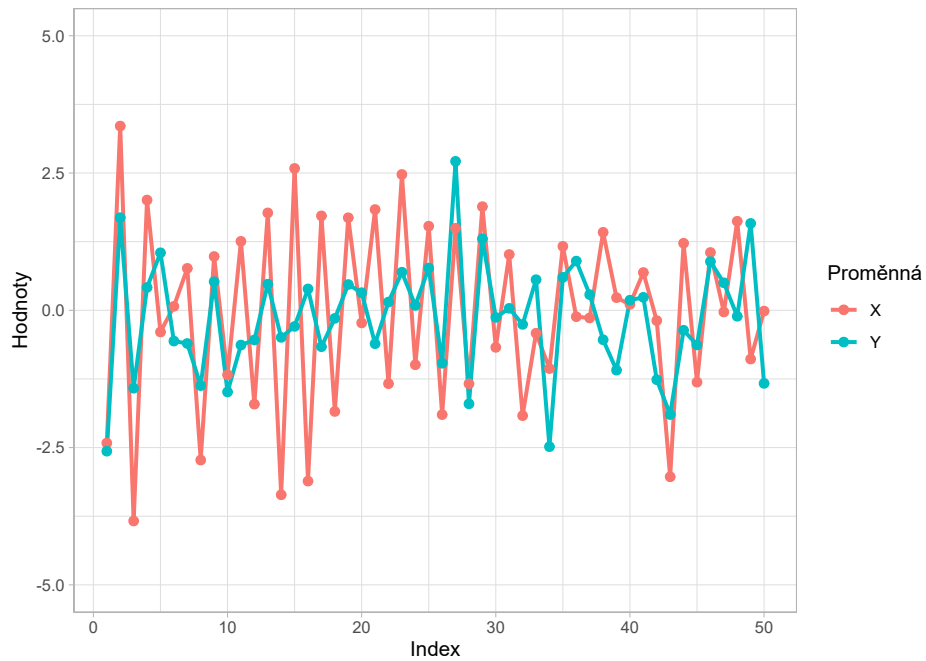
Na grafu 5.3 pozorujeme hodnoty výběrové křížové korelace mezi pozorováními X a Y spolu s teoretickými hodnotami, které byly spočteny díky koeficientům lineárních procesů získané pomocí (1.3) a vzorcům (5.5). Vidíme, že výběrová korelace pro malé $|k|$ docela dobře kopíruje tvar reálné korelace. Pro $k \geq 2$ je důsledkem předpisu řady Y reálná korelace nulová. Výběrová korelace však dosahuje relativně vysokých hodnot, což je způsobeno charakterem řad ARMA, kde i dvě nezávislé řady mohou mít vysoké hodnoty výběrové korelace, protože se řídí podobnými předpisy.

Výběrová křížová korelační funkce pro odhady bílých šumů je znázorněná na grafu 5.4. Odhad korelace mezi bílými šumy pro $k = 0$ vychází $0,456$.

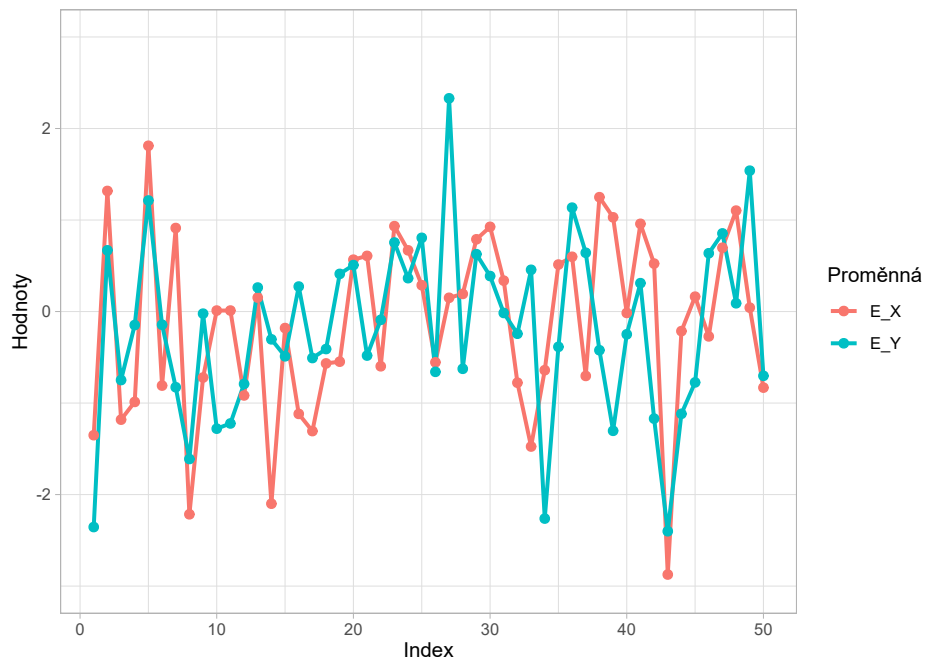
Příklad: V druhém případě se zaměříme na trajektorii vygenerovanou pro komplexní modely (5.3) a (5.4) délky $N = 50$ a korelaci mezi bílými šumy tentokrát opačnou $\rho = -0,5$. Vygenerované pozorování obou řad pozorujeme na grafu 5.5. Díky odhadům modelů ARMA můžeme zpětně dopočítat bílé šumy znázorněné na grafu 5.6.

Porovnání výběrové a reálné křížové korelace pozorujeme na grafu 5.7. V tomto případě se výběrová funkce liší od té reálné více než v příkladu s jednoduchými modely. I když byla zvolena korelace mezi bílými šumy $-0,5$, tak díky předpisům je hodnota reálné korelace v bodě $k = 0$ rovna hodnotě pouze $-0,224$.

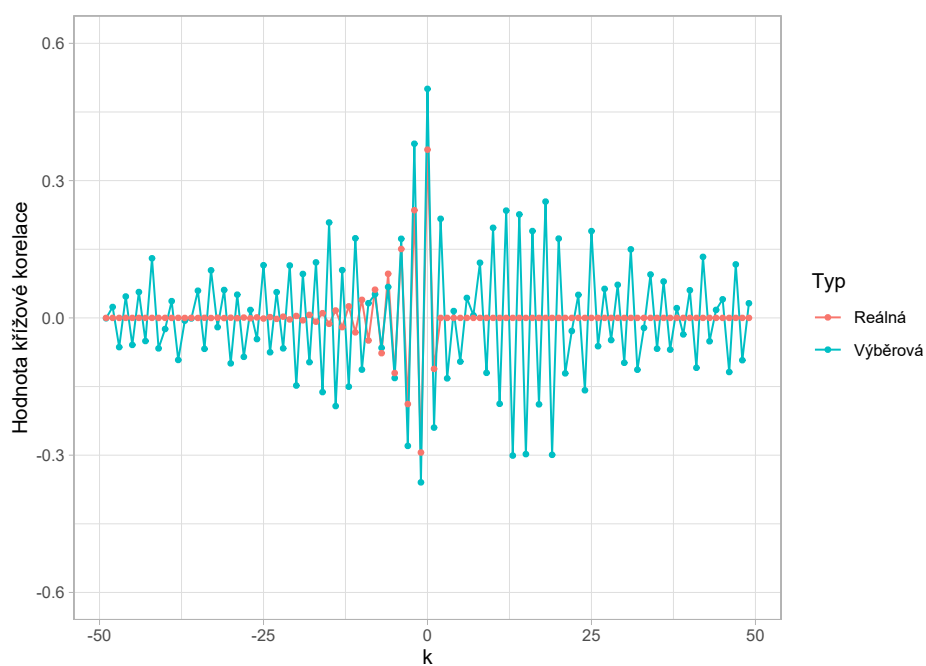
Jako poslední se můžeme podívat na graf 5.8 s hodnotami výběrové křížové korelace mezi odhady bílých šumů. Korelace pro $k = 0$ je rovná $-0,328$. Odhad je tedy horší, než v prvním případě. To může být způsobené komplexnějšími předpisy a s tím souvisejícím odhadem modelů ARMA, ale také pouhou náhodou. V obou případech jsme vygenerovali pouze 50 pozorování, což je relativně malý počet pozorování pro přesné odhady.



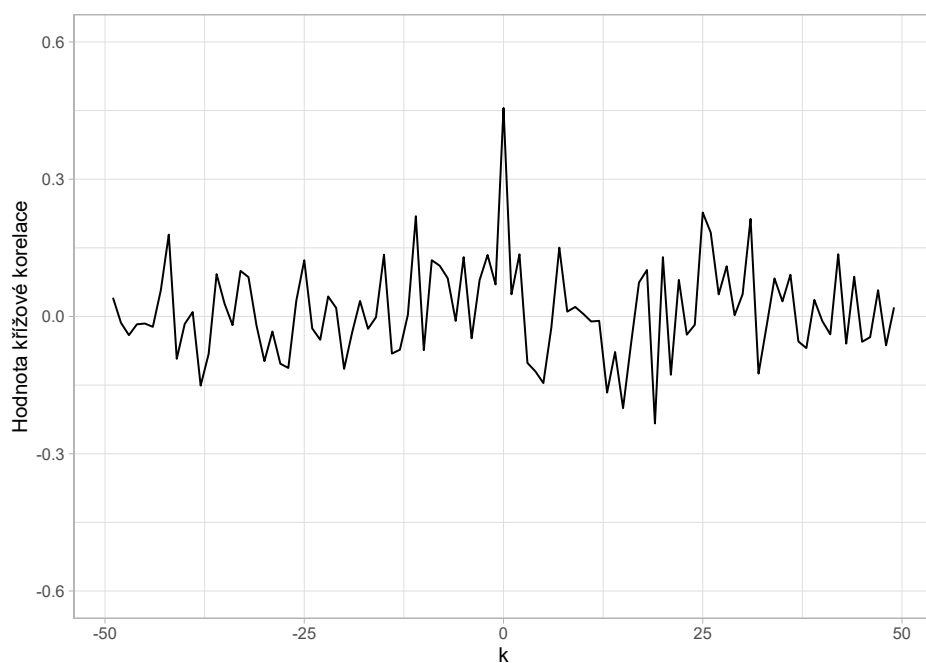
Obrázek 5.1: Graf znázorňující vygenerované pozorování řad X a Y řídicí se jednoduchými modely s korelací ρ mezi bílými šумы ve stejném časovém okamžiku rovnou 0,5.



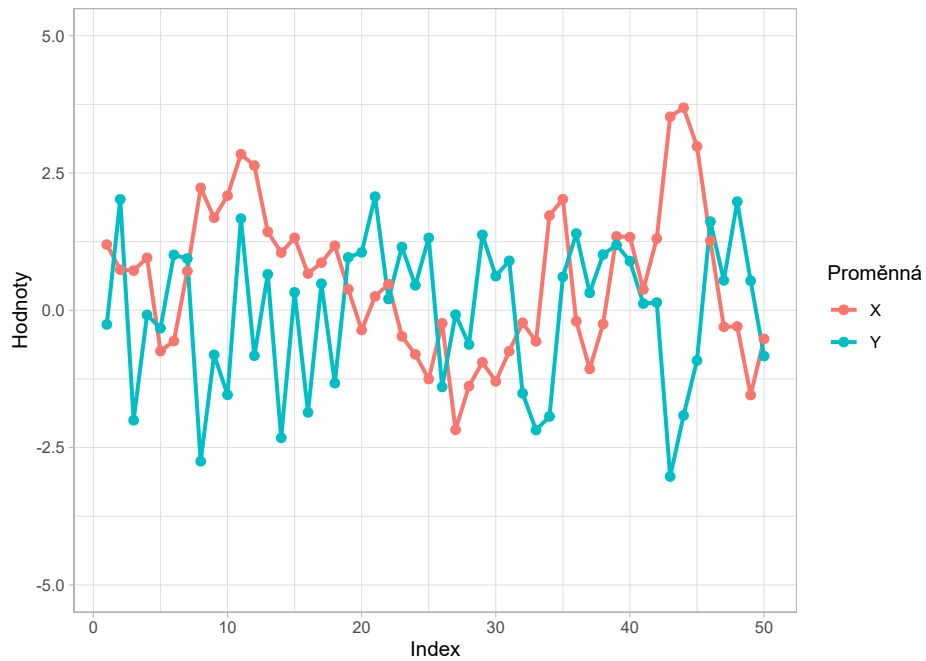
Obrázek 5.2: Graf znázorňující odhadnuté bílé šумы příslušné řadám X a Y , které se řídí jednoduchými modely s korelací ρ mezi bílými šумы ve stejném časovém okamžiku rovnou 0,5.



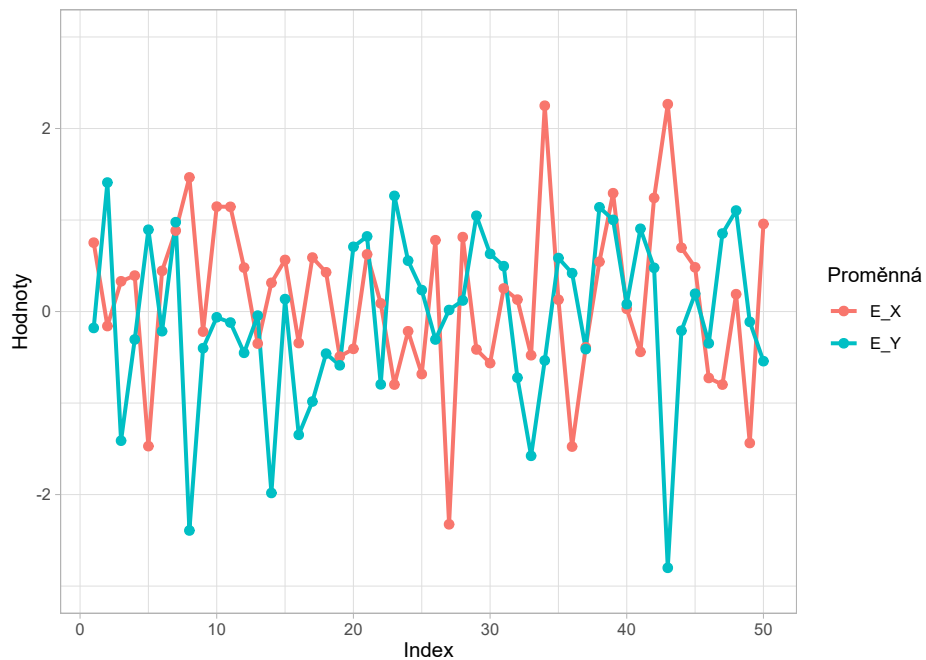
Obrázek 5.3: Graf znázorňující hodnoty výběrové křížové korelace $\hat{r}_{X,Y}(k)$ mezi pozorováními řad X a Y z grafu 5.1 spolu s jejími teoretickými hodnotami $r_{X,Y}(k)$.



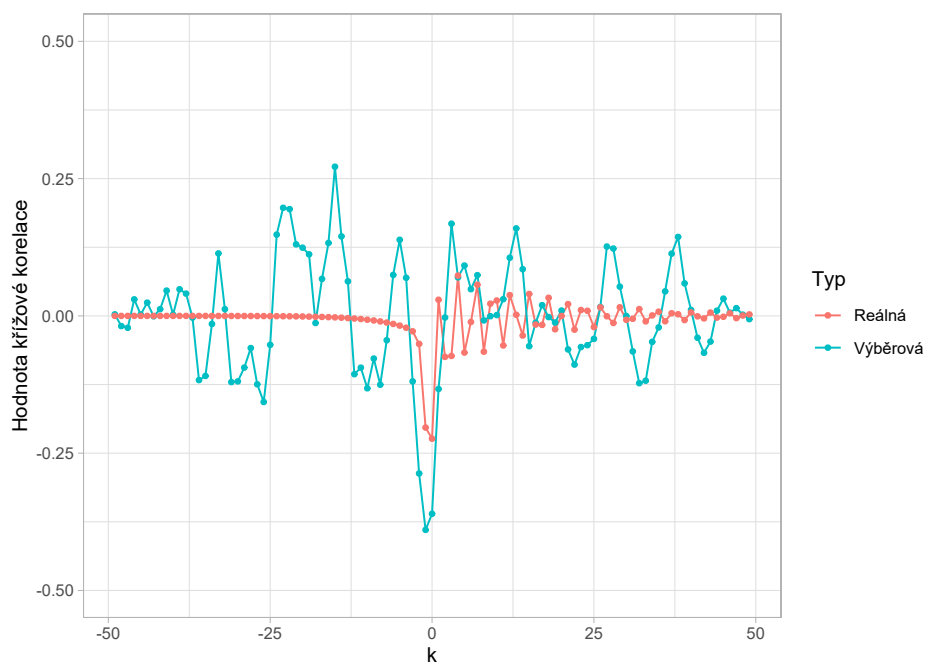
Obrázek 5.4: Graf znázorňující hodnoty výběrové křížové korelace $\tilde{r}_{\hat{E}_X, \hat{E}_Y}(k)$ mezi odhady bílých šumů příslušných řadám X a Y z grafu 5.2.



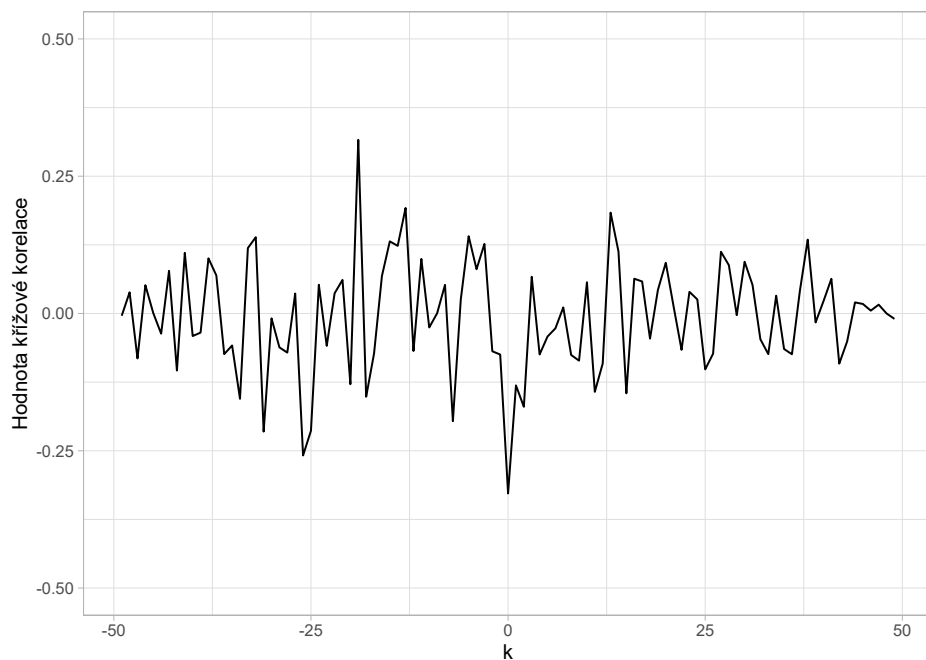
Obrázek 5.5: Graf znázorňující vygenerované pozorování řad X a Y řídicí se komplexními modely s korelací ρ mezi bílými šумы ve stejném časovém okamžiku rovnou $-0,5$.



Obrázek 5.6: Graf znázorňující odhadnuté bílé šумы příslušné řadám X a Y , které se řídí komplexními modely s korelací ρ mezi bílými šумы ve stejném časovém okamžiku rovnou $-0,5$.



Obrázek 5.7: Graf znázorňující hodnoty výběrové křížové korelace $\hat{r}_{X,Y}(k)$ mezi pozorováními řad X a Y z grafu 5.5 spolu s jejími teoretickými hodnotami $r_{X,Y}(k)$.



Obrázek 5.8: Graf znázorňující hodnoty výběrové křížové korelace $\tilde{r}_{\hat{E}_X, \hat{E}_Y}(k)$ mezi odhady bílých šumů příslušných řadám X a Y z grafu 5.6.

5.2 Porovnávání testy

Porovnávat budeme následující sadu sedmi testů:

- Haughův test s využitím první verze testové statistiky S_M – H1,
- Haughův test s využitím druhé verze testové statistiky S_M^* – H2,
- Modifikovaný t-test použitý na pozorování řad ARMA – Mt-ARMA,
- Modifikovaný t-test použitý na odhadnuté bílé šумы – Mt-WN,
- Test distanční kovariancí použitý na odhadnuté bílé šумы – DCT,
- Standardní t-test použitý na pozorování řad ARMA – St-ARMA,
- Standardní t-test použitý na odhadnuté bílé šумы – St-WN.

Standardní t-test použitý přímo na pozorované řady je zahrnutý jakožto kontrolní test, který by neměl v případě modelů ARMA fungovat. Pokud ho ale použijeme na odhadnuté bílé šумы, tak za našich předpokladů jsou splněny předpoklady tohoto standardního testu nezávislosti a můžeme tak pozorovat, jak se s tímto klasickým testem poměřují testy ostatní. Tento test se bude pouze mírně lišit od modifikovaného t-testu, a to pouze kvůli rozdílu použití N a \tilde{N} , kde \tilde{N} může být upraveno o možnou náhodnou kovarianci uvnitř odhadů bílých šumů. Pro oba Haughovy testy bylo zvoleno $M = N/10$, tedy 5, 10 a 20. Hladina byla zvolena $\alpha = 0,05$ pro všechny testy.

Poznámka: Nemáme zaručeno, že odhady bílých šumů mají stejné vlastnosti, jako které předpokládáme pro reálné bílé šумы. Pokud se ale pozorované řady opravdu řídí modelem ARMA a máme dostatek pozorování ke kvalitnímu odhadu těchto modelů, budou se odhady bílých šumů blížit těm reálným a i jejich vlastnosti budou podobné.

5.3 Porovnání

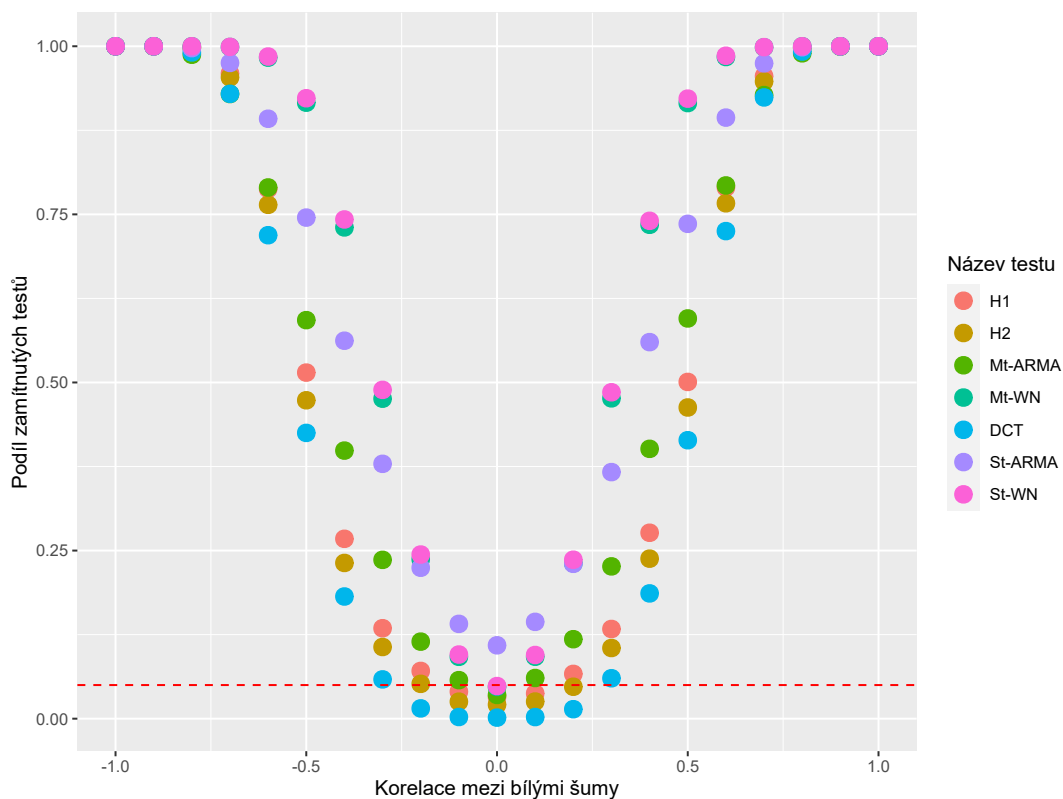
Každé ze 126 možných nastavení bylo otestováno 10000krát. V rámci jednoho testování byl postup následující:

- 1) Vygenerovat dvojrozměrný bílý šum s korelací ρ .
- 2) Vygenerovat příslušné pozorování řad X a Y pomocí daných předpisů.
- 3) Odhadnout modely ARMA, kde nejvhodnější model byl vybrán z množiny modelů $\text{ARMA}(i, j)$, kde $i, j \in \{0, 1, 2, 3\}$, $i + j \neq 0$. Výběr byl proveden na základě Akaikeho informačního kritéria.
- 4) Na základě odhadnutých modelů ARMA spočíst odhady bílých šumů.
- 5) Použít postupně všechny zkoušené testy.

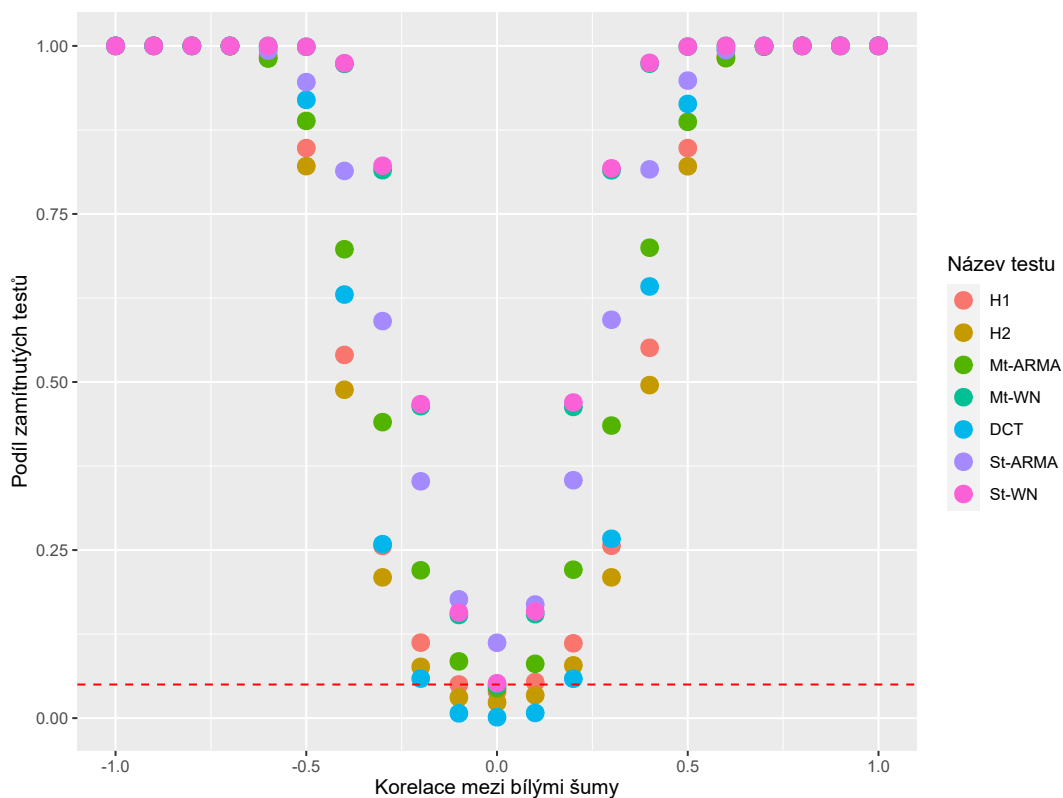
Pozorovanou hodnotou je podíl zamítnutých nulových hypotéz. V případě platnosti nulové hypotézy, tedy pro případ $\rho = 0$, by podíl zamítnutých testů měl být co nejblíže rovný požadované hladině 0,05. V opačném případě bychom požadovali, aby test měl poměr zamítnutých testů co největší. Výsledky simulací uvádíme jak graficky, tak číselně v tabulkách v příloze A:

- Jednoduché modely s $N = 50$, tabulka A.1, graf 5.9.
- Jednoduché modely s $N = 100$, tabulka A.2, graf 5.10.
- Jednoduché modely s $N = 200$, tabulka A.3, graf 5.11.
- Komplexní modely s $N = 50$, tabulka A.4, graf 5.12.
- Komplexní modely s $N = 100$, tabulka A.5, graf 5.13.
- Komplexní modely s $N = 200$, tabulka A.6, graf 5.14.

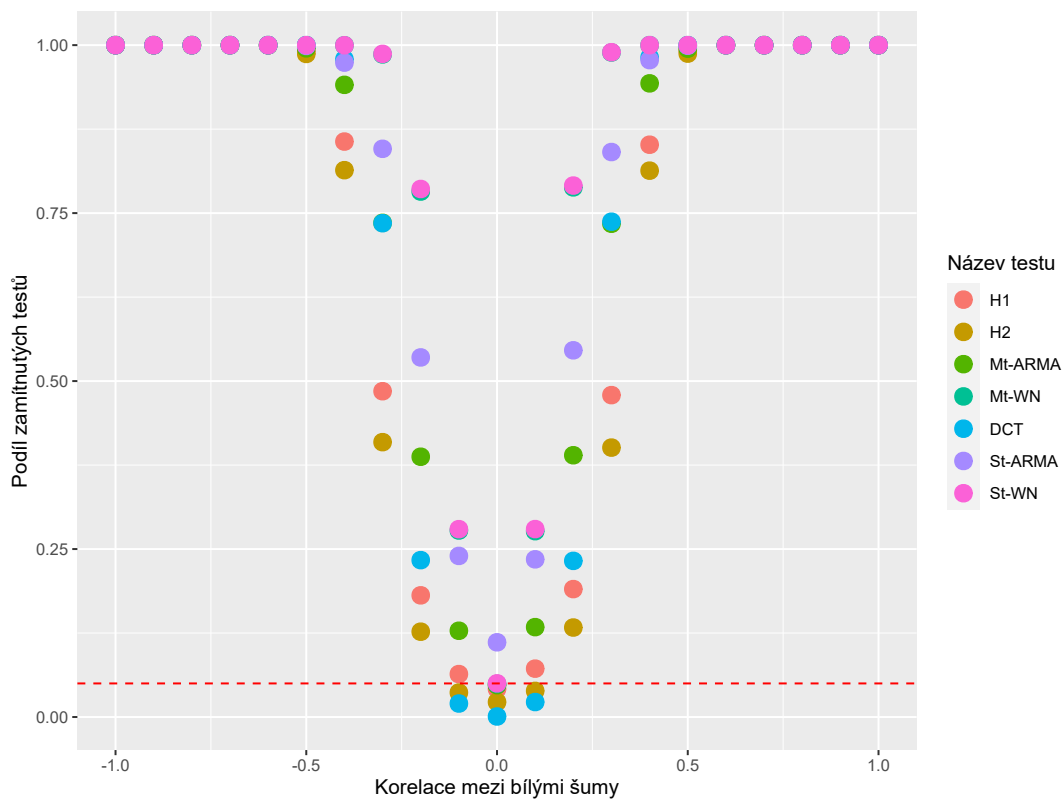
Navíc uvádíme grafy 5.15 a 5.16, kde můžeme pozorovat, jak se podíl zamítnutých testů mění na základě počtu dostupných pozorování pro každý test zvlášť. Nejdříve budeme diskutovat každý test jednotlivě a poté bude následovat shrnutí.



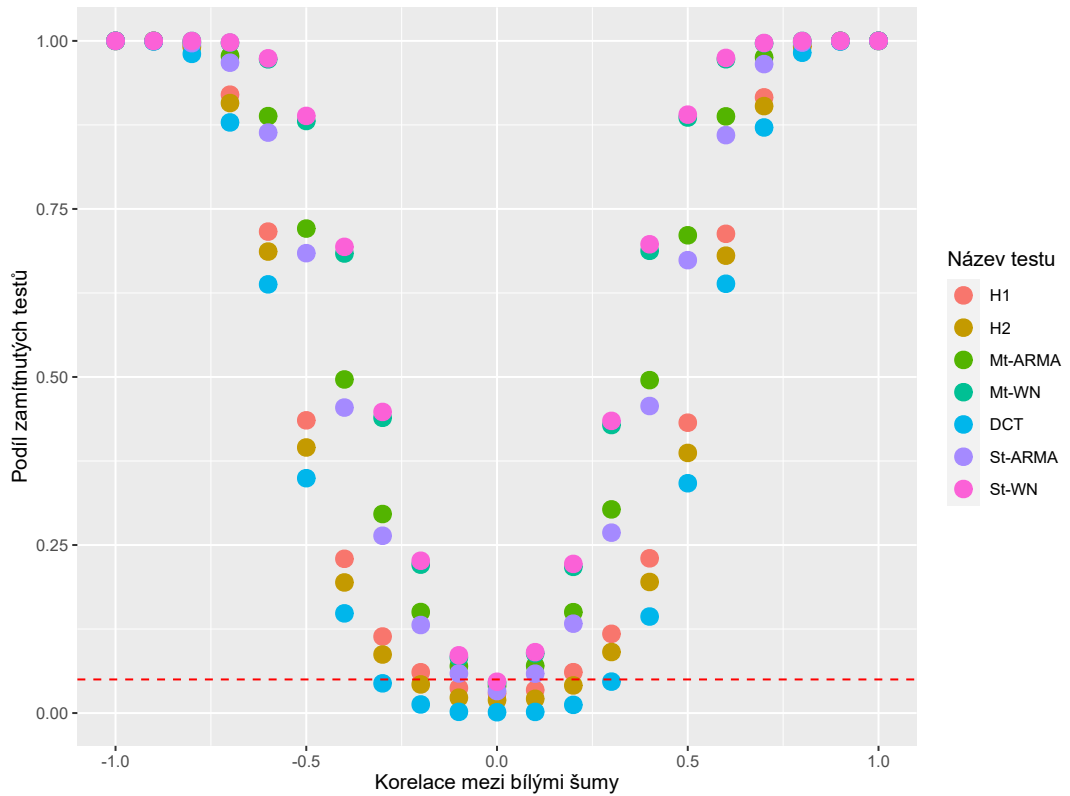
Obrázek 5.9: Grafické porovnání testů, které byly použity v případě jednoduchého modelu pro $N = 50$. Přerušovaná červená čára značí hladinu 0,05.



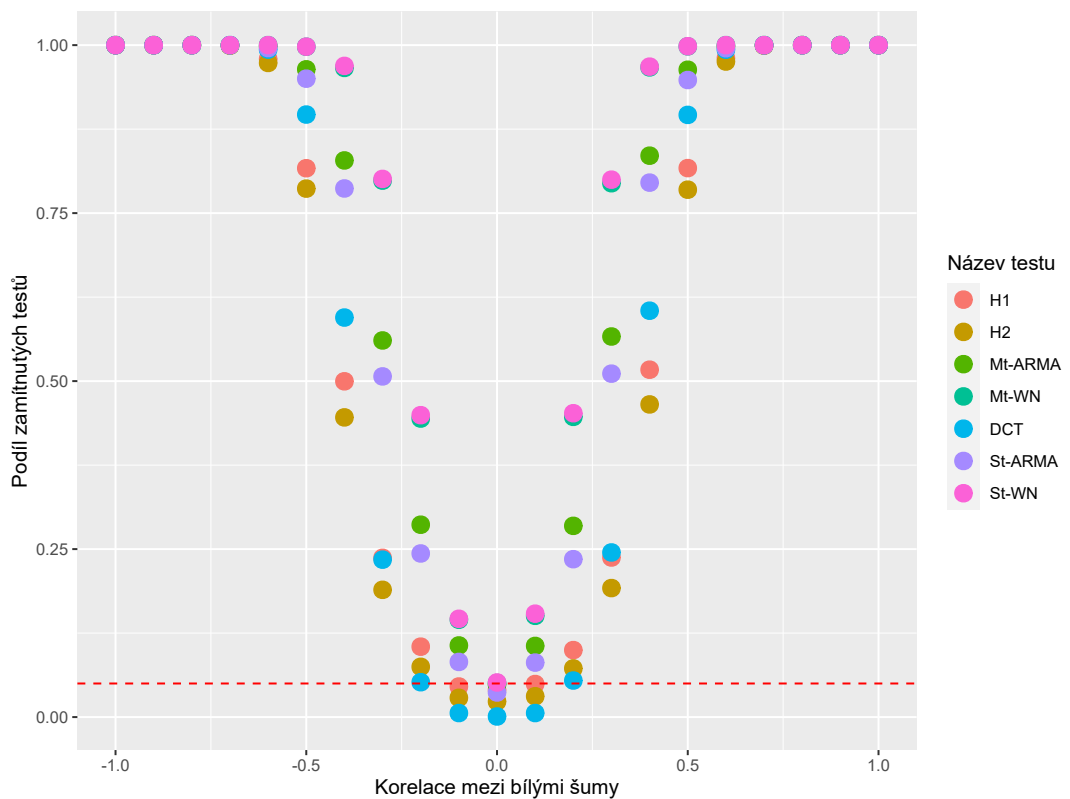
Obrázek 5.10: Grafické porovnání testů, které byly použity v případě jednoduchého modelu pro $N = 100$. Přerušovaná červená čára značí hladinu 0,05.



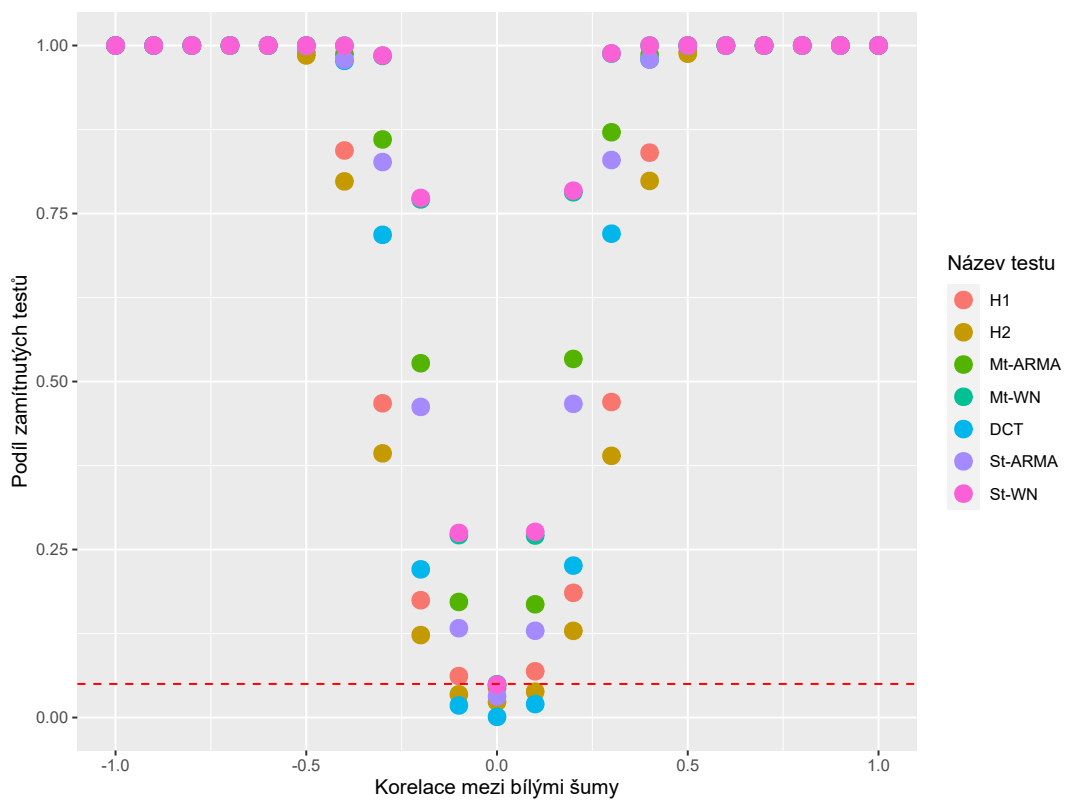
Obrázek 5.11: Grafické porovnání testů, které byly použity v případě jednoduchého modelu pro $N = 200$. Přerušovaná červená čára značí hladinu 0,05.



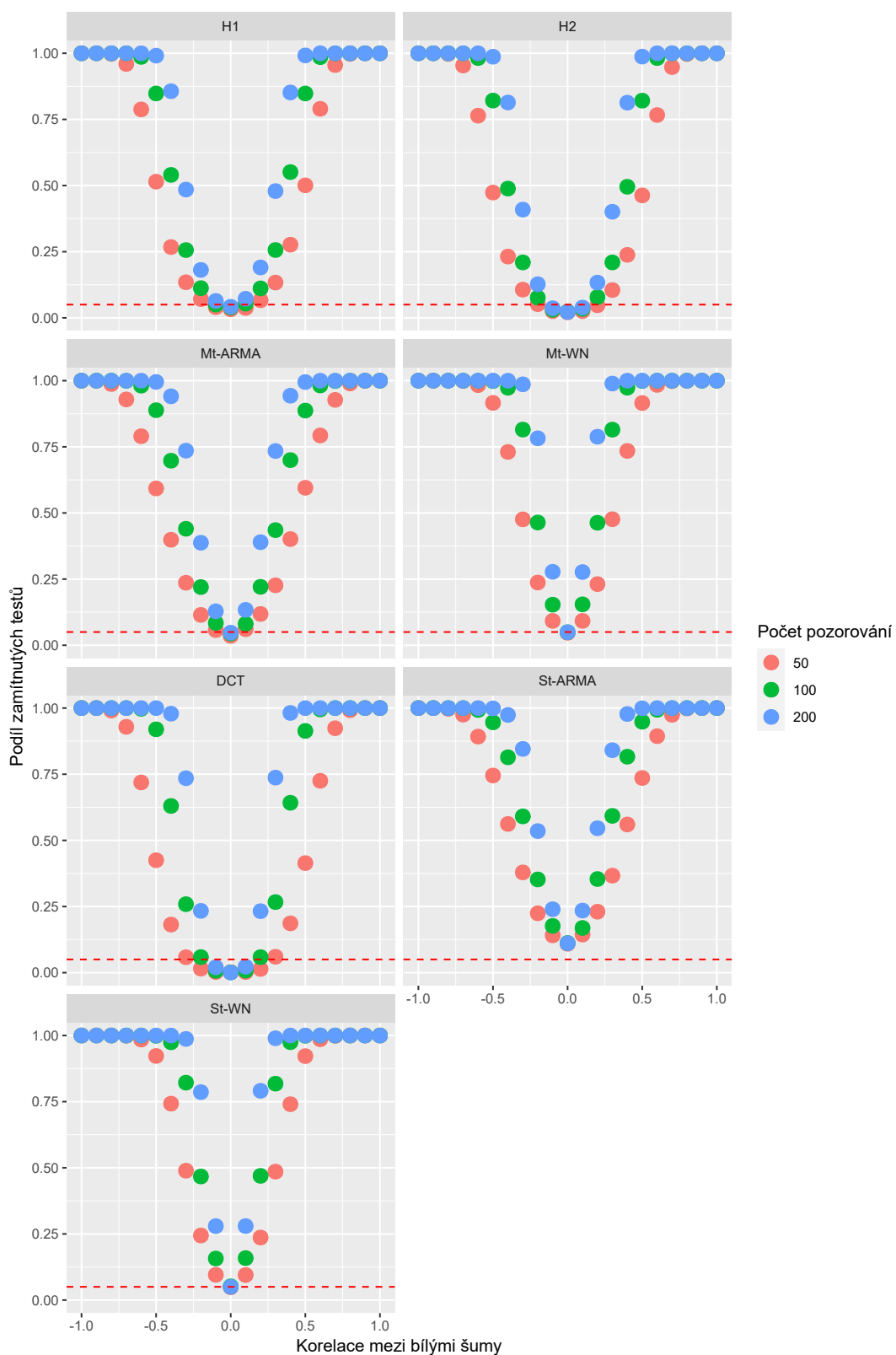
Obrázek 5.12: Grafické porovnání testů, které byly použity v případě komplexního modelu pro $N = 50$. Přerušovaná červená čára značí hladinu 0,05.



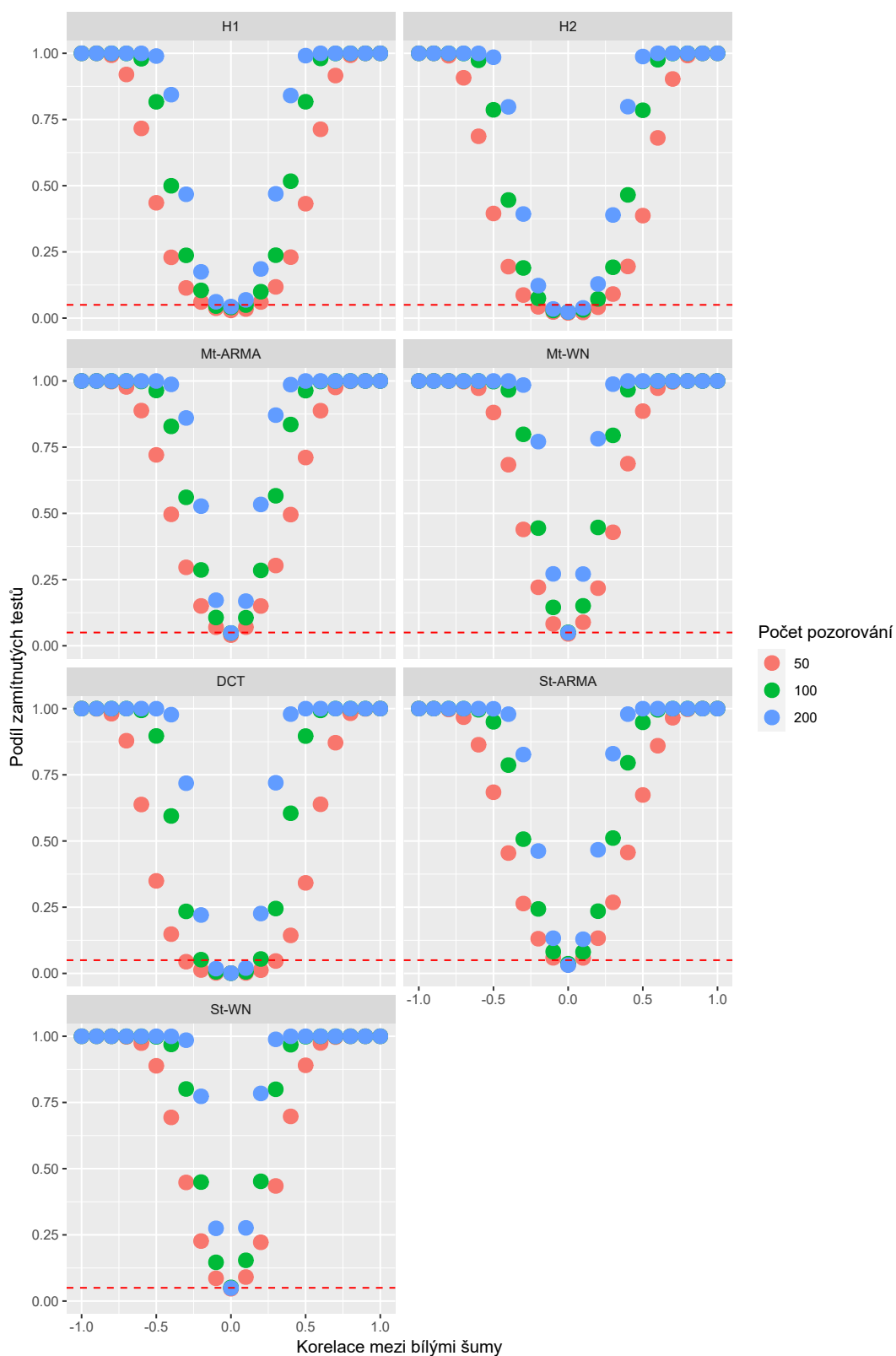
Obrázek 5.13: Grafické porovnání testů, které byly použity v případě komplexního modelu pro $N = 100$. Přerušovaná červená čára značí hladinu 0,05.



Obrázek 5.14: Grafické porovnání testů, které byly použity v případě komplexního modelu pro $N = 200$. Přerušovaná červená čára značí hladinu 0,05.



Obrázek 5.15: Grafy znázorňující vztah podílu zamítnutých testů na počtu pozorování pro každý test zvlášť v případě použití na jednoduché modely. Přerušovaná červená čára značí hladinu 0,05.



Obrázek 5.16: Grafy znázorňující vztah podílu zamítnutých testů na počtu pozorování pro každý test zvlášť v případě použití na komplexní modely. Přerušovaná červená čára značí hladinu 0,05.

Haughovy testy

Ani jedna z verzí Haughova testu nedosahuje v případě platnosti nulové hypotézy požadované hladiny 0,05. Zatímco první verze se blíží k hladině s rostoucím počtem dat, druhá verze stagnuje na hodnotách mezi 0,019 a 0,023 ve všech šesti nastaveních a je tudíž o něco konzervativnější než první verze testu. Obě verze testu vykazují při slabé korelovanosti menší sílu oproti ostatním testům a síla testu se zvyšuje nejpomaleji s rostoucím počtem dat. Tento test závisí na volbě hodnoty parametru M , jehož úpravou by mohlo dojít k vylepšení výsledků této simulace. V našem případě je generovaná korelace pouze pro $k = 0$, zbylá korelace je zcela náhodná. Proto by se možná vyplatilo použít menší M než $M = N/10$. Nalezení optimální hodnoty M však nebylo předmětem tohoto porovnání. Nevýhodou tohoto testu je, že závisí na správném odhadu modelu ARMA.

Modifikované t-testy

V případě jednoduchých modelů ARMA verze tohoto testu nedosahuje požadované hladiny při malém počtu dat (0,035). S rostoucím počtem dat se ale reálná hladina blíží té požadované. Síla testu je o něco lepší než u Haughových testů, pro počet dat $N = 200$ dosahuje dobrých výsledků už od korelace $|\rho| \geq 0,3$. U tohoto testu můžeme pozorovat největší rozdíl mezi jednoduchými a komplexními modely. V tomto případě hraje roli složitost autokovarianční matice, která následně ovlivňuje hodnotu \tilde{N} . Zásadní výhodou této verze testu je, že nezávisí na správně odhadnutém modelu ARMA. Na druhou stranu nedokáže odhalit korelaci posunutou v čase.

Verze modifikovaného t-testu využívající odhady bílých šumů lépe dosahuje požadované hladiny než verze ARMA. Síla tohoto testu je v případě jednoduchého modelu významně lepší než ARMA verze a i v případě komplexního modelu dosahuje lepších výsledků. Síla v případě velkého počtu dat je velmi vysoká už pro korelace $|\rho| = 0,3$, a to přibližně 0,985. Jeho hlavní nevýhodou je, že stejně jako Haughův test závisí na správném odhadu modelu ARMA a stejně jako první verze testu neodhalí korelaci posunutou v čase.

Test distanční kovariancí

Testování potvrdilo větu 18, hladina tohoto testu se pohybuje ve všech případech kolem velmi nízké hodnoty 0,001. Dále si můžeme všimnout, že pro počet dat $N = 50$ se jedná o test s nejmenší silou. Zároveň je to ale test, který, co se týče síly testu, dělá největší pokroky s přibývajícími daty. Pro $N = 200$ je v případě komplexního modelu síla pro korelace $0,2 \leq |\rho| \leq 0,5$ pro tento test větší, než pro Haughovy testy. V případě jednoduchého modelu se silou blíží dokonce i modifikovanému t-testu použitého na pozorování ARMA. Jeho nevýhodou je potřeba odhadu modelu ARMA, jehož nesprávný odhad by mohl vést ke špatným výsledkům testu. Při malém počtu dat je až moc konzervativní, a tak se nevyplatí ho používat. S narůstajícím N by se ale vylepšovala jak kvalita odhadnutého modelu, tak dále síla tohoto testu. Výhodou tohoto testu je, že na rozdíl od modifikovaných t-testů nevyžaduje normální rozdělení bílých šumů. Na druhou stranu se musí jednat o striktní bílý šum.

Standardní t-testy

Nejdříve se zaměříme na verzi testu použitou přímo na pozorování řad ARMA. Tento test v případě komplexních modelů nedosahuje požadované hladiny 0,05 nehledě na počet využitých dat. Naopak v případě jednoduchých modelů hladinu více než dvojnásobně překračuje. Výsledky tohoto testu velmi závisí na předpisu obou řad, tudíž je velmi nespolehlivý a tím pádem i nepoužitelný.

Klasická verze t-testu, která využívá odhady bílých šumů, dodržuje hladinu nejlépe ze všech testů a jeho síla je nejvyšší ze všech použitých testů. V případě více dat si dokáže poradit s nízkou korelací $0,2 \leq |\rho| \leq 0,4$ a i pro $|\rho| = 0,1$, velmi nízkou korelací, kterou je těžké odhalit, má sílu průměrně 0,274. Jako všechny ostatní testy využívající odhady bílých šumů tak i tento test závisí na správných odhadech modelů ARMA a stejně jako modifikovaný t-test nedokáže odhalit korelaci posunutou v čase

Shrnutí

Z testování vyplynulo, že na komplexnosti předpisů modelů ARMA závisí pouze modifikovaný t-test použitý na pozorování ARMA. V ostatních případech ovlivňuje složitost modelu pouze odhad modelu ARMA, což pak ale může ovlivnit výsledky testů. Dalším poznatkem je, že síla testů nezávisí na znaménku korelace bílých šumů, což můžeme snadno vypožorovat ze symetričnosti grafů 5.15 a 5.16.

Jak se dalo očekávat, tak standardní t-test aplikovaný přímo na pozorování, je v případě řad ARMA nepoužitelný. Haughův test v našem případě se zvoleným parametrem $M = N/10$ zaostává za ostatními testy a musela by se buďto vyzkoušet jiná volba M , nebo zkusit některý z popisovaných testů z podkapitoly 2.4. Nejlépe z našeho testování vychází standardní i modifikovaný t-test použitý na odhady bílých šumů. Je ale dobré znovu poznamenat, že takto velmi dobré výsledky t-testů byly umožněny způsobem generování bílých šumů. Kdyby ρ byla korelace mezi E_X v čase t a E_Y v čase $t + k$ pro nějaké $k \neq 0$, t-testy by nebyly schopné tuto korelaci odhalit. Pokud jsme si jisti odhadem modelu ARMA, ale bílý šum není normálně rozdělený, lze v případě jeho nezávislosti a většího počtu dat použít test distanční kovariancí. Pokud si nejsme jisti ani odhadem modelu ARMA, tak nám nic jiného nezbyvá, než zkusit použít modifikovaný t-test přímo na pozorování.

6. Ilustrativní příklad na reálných datech

V této kapitole vyzkoušíme testy nezávislosti na reálná data. Jako pozorování nám poslouží ceny indexů **Standard and Poor's 500** (S&P 500) a **Financial Times Stock Exchange 100 Index** (FTSE 100) v období červen 2021 až květen 2023. Data byla stažena ze stránky The Wall Street Journal (2023).

6.1 Popis dat

S&P 500 je burzovní index skládající se z 503 největších firem na americkém burzovním trhu. Čím větší podíl má firma na akciovém trhu, tím větší je jeho váha v rámci indexu. K 5. 7. 2023 mezi 5 firem s nejvyšší váhou patří Apple, Microsoft, Amazon, Nvidia a Alphabet Inc (vlastníci společnost Google). Ceny tohoto indexu uvádíme v amerických dolarech.

FTSE 100 je obdoba indexu S&P 500, ale tentokrát se jedná o 100 největších firem na burze ve Velké Británii. K 5. 7. 2023 patří mezi 5 největších firem Astra a Zeneca, Shell, HSBC, Unilever a BP. Ceny tohoto indexu uvádíme v britských librách.

Oba indexy se často používají jako indikátory ekonomické situace v daných zemích. Pro testování byly vybrány proto, abychom zjistili, jestli se ceny na burzovních trzích USA a Velké Británie chovají nezávisle na sobě. Data pochází z doby, kdy už největší ekonomický vliv pandemie covidu-19 opadal, a tak by tato událost neměla zkreslovat vztah mezi těmito indexy.

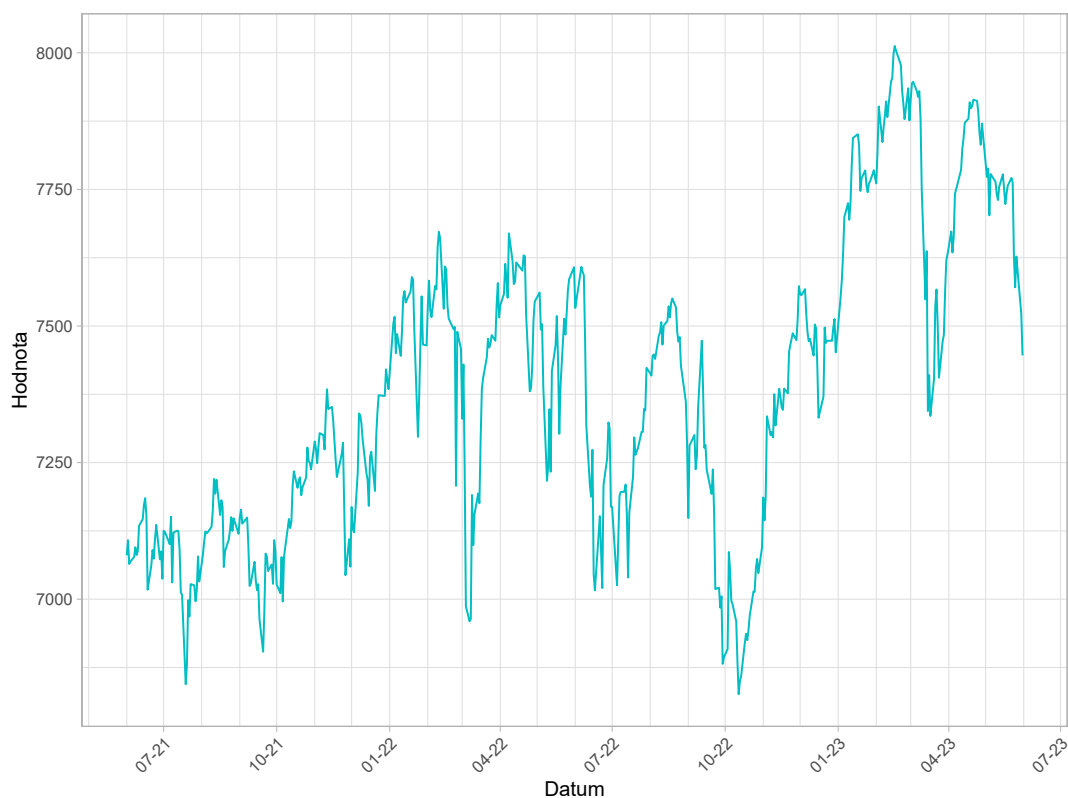
Pro správné určení závislosti je potřeba vybrat všechny dny, kdy bylo obchodováno s oběma indexy zároveň. Dostáváme tak průnik o velikosti $N = 493$. V tabulce 6.1 uvádíme základní výběrové statistiky. Na grafech 6.1 a 6.2 můžeme pozorovat grafické znázornění vývoje ceny indexů ve zkoumaném období. Můžeme si všimnout, že některé záporné výkyvy můžeme najít v cenách obou indexů, jakožto například v červenci 2022 nebo v říjnu 2022.

Výběrová statistika	S&P 500	FTSE 100
Minimum	3 577,03	6 826,15
1. kvartil	3 969,61	7 148,50
Medián	4 166,45	7 361,31
Průměr	4 199,84	7 375,65
3. kvartil	4 436,75	7 554,09
Maximum	4 793,54	8 012,53
Rozptyl	83 767,11	74 580,95

Tabulka 6.1: Výběrové statistiky ceny indexů S&P 500 v amerických dolarech a FTSE 100 v britských librách.



Obrázek 6.1: Vývoj ceny indexu S&P 500 od června 2021 do května 2023. Ceny jsou znázorněny v amerických dolarech.



Obrázek 6.2: Vývoj ceny indexu FTSE 100 od června 2021 do května 2023. Ceny jsou znázorněny v britských librách.

6.2 Použití testů nezávislosti

Prvním krokem je odhadnout modely ARMA. Jelikož se jedná o časové řady s nulovou střední hodnotou, je potřeba ceny vycentrovat odečtením průměrů, jinak probíhá odhad modelů stejně jako v simulační části 5. Pro oba indexy dostáváme na základě hodnoty Akaikeho informačního kritéria jako nejvhodnější model ARMA(1, 0). Pro centrované ceny S&P 500 dostáváme odhadnutý předpis:

$$X_t - 0,9828X_{t-1} = E_{X,t}, \quad (6.1)$$

a pro centrované FTSE 100 zase předpis:

$$Y_t - 0,9695Y_{t-1} = E_{Y,t}. \quad (6.2)$$

Jedná se tedy o autoregresní modely s velmi vysokými hodnotami autokorelačních funkcí. Nyní když máme odhadnuté modely, můžeme snadno dopočítat odhad bílého šumu a postupně vyzkoušet všech sedm testů, které byly použity v rámci simulace. Pro Haughovy testy použijeme $M = 40$. V tabulce 6.2 uvádíme p-hodnoty použitých testů.

H1	H2	Mt-ARMA	Mt-WN	DCT	St-ARMA	St-WN
$7,4246 \cdot 10^{-8}$	$1,8931 \cdot 10^{-7}$	0,8880	$< 10^{-21}$	$1,1521 \cdot 10^{-5}$	0,3070	$1,0787 \cdot 10^{-20}$

Tabulka 6.2: Tabulka p-hodnot testů nezávislosti pro ceny indexů S&P 500 a FTSE 100.

Z tabulky 6.2 vidíme, že jediné dva testy, které nezamítly nulovou hypotézu nezávislosti, jsou ARMA verze modifikovaného t-testu a standardního t-testu. Důvodem je, že Pearsonův korelační koeficient mezi cenami indexů vychází na nízkých $-0,046$. Na grafu 6.3, znázorňující výběrovou křížovou korelaci mezi cenami indexů, můžeme pozorovat, že pro malé $|k|$ dosahuje korelace hodnot mezi $-0,2$ a $-0,35$, zatímco pro velké $|k|$ dosahuje korelace hodnot opačných. Pearsonův korelační koeficient v tomto případě není schopný odhalit korelaci mezi oběma řadami, a je tím pádem při určování nezávislosti nedostačující statistikou.

Modifikovaný t-test použitý přímo na ceny indexů má velmi vysokou p-hodnotu oproti standardnímu t-testu. To je způsobené hodnotou \tilde{N} , která vychází pro tuto verzi testu 11,803. Vysoká úroveň autokorelace v rámci obou řad totiž zapříčiňuje vysokou hodnotu rozptylu (3.1), která pak má za následek velmi nízkou hodnotu \tilde{N} .

Všechny ostatní testy, které zároveň všechny využívají odhady bílých šumů, nulovou hypotézu nezávislosti zamítají. Graf 6.4 znázorňuje hodnoty výběrové křížové korelace mezi odhady bílých šumů. Vidíme, že korelace je nejvyšší v bodě $k = 0$ a poté osciluje mezi hodnotami $-0,1$ a $0,1$. Pro odhady bílých šumů vychází Pearsonův korelační koeficient rovný 0,403. Výběrová distanční korelace nabývá hodnoty 0,372.

Jako poslední ověříme předpoklady použitých testů. Zprvce nás zajímá normalita bílých šumů. K jejímu otestování použijeme Shapirův-Wilkův test. V obou případech je nulová hypotéza normality zamítnuta, pro residua S&P 500 s p-hodnotou 0,016 a pro FTSE 100 dokonce s p-hodnotou $4,934 \cdot 10^{-11}$.

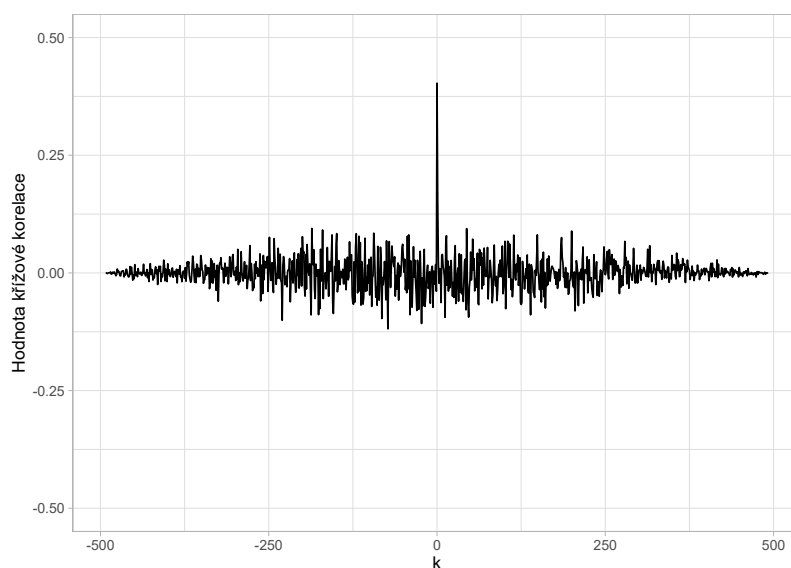
Oba odhadnuté modely ARMA s předpisy (6.1) a (6.2) mají kořeny polynomu $a(z)$ vně jednotkového kruhu (přibližně 1,018 a 1,031) a polynom $b(z)$ žádné kořeny nemá. Jedná se tedy v obou případech podle tvrzení 8 o lineární procesy.

Dále otestujeme nekorelovanost bílých šumů pomocí Ljungové-Boxova testu. Pokud budeme chtít testovat významnost prvních padesáti hodnot autokorelační funkce, tak pro oba bílé šумы nezamítáme nulovou hypotézu nekorelovanosti, pro S&P 500 s p-hodnotou 0,760 a pro FTSE 100 0,180.

Bílé šумы tudíž splňují nekorelovanost, ale nesplňují normalitu. V tomto případě bychom se řídili Haughovým testem a nulovou hypotézu zamítli. Takže ceny indexů S&P 500 a FTSE 100 nejsou nezávislé.



Obrázek 6.3: Hodnoty výběrové křížové korelace $\hat{r}_{X,Y}(k)$ pro ceny S&P 500 a FTSE 100.



Obrázek 6.4: Hodnoty výběrové křížové korelace $\tilde{r}_{\hat{E}_X, \hat{E}_Y}(k)$ mezi odhadnutými bílými šумы pro ceny S&P 500 a FTSE 100.

Závěr

Předmětem této práce byla kompilace informací o testech nezávislosti pro časové řady řídicí se modely ARMA a následující porovnání těchto testů na simulovaných datech.

V první kapitole jsme se věnovali definici nezávislosti a odvodili pro ni pár ekvivalentních tvrzení. Následovala podkapitola o kovarianci a korelaci v rámci jedné řady i mezi dvěma časovými řadami. Pro výběrový odhad křížové kovariance byly autorem formulovány a dokázány asymptotická nestrannost a konzistence, která byla následně dokázána i pro korelaci. Dále jsme se zabývali představením modelu ARMA a definicí lineárního procesu. Součástí této kapitoly bylo tvrzení 9 o vztahu nezávislosti bílých šumů s nezávislostí lineárních procesů, na jehož myšlenke jsou sestaveny některé z dále popsaných testů. V závěru kapitoly jsme si připomněli definici charakteristických funkcí.

Druhá kapitola byla věnována Haughovu testu nezávislosti. V úvodu kapitoly byly autorem odvozeny rozptyl a kovariance mezi odhady bílých šumů za předpokladu nezávislosti. V rámci této kapitoly jsme popsali dva možné postupy odhadnutí modelu ARMA. Následovalo představení Haguhova testu a popis dalších variací testu, které na Haughův test navazovaly.

V třetí kapitole jsme se zabývali modifikovaným t-testem nezávislosti. Na rozdíl od standardního t-testu požaduje, aby místo nezávislých pozorování byly řady pouze slabě stacionární. V závěru kapitoly jsme diskutovali o výhodách a nevýhodách tohoto testu.

Distanční kovarianci a korelaci byla věnována čtvrtá kapitola této práce. Jedná se o navrhovanou náhradu Pearsonova korelačního koeficientu, která na místo sledování lineární závislosti měří, jak moc se liší charakteristická funkce vektoru dvojice náhodných veličin od součinu marginálních charakteristických funkcí, který by se rovnal sdružené funkci za platnosti předpokladu nezávislosti. Byly představeny dva příklady výsledků distanční kovariance a korelace, z toho jeden spočtený autorem diplomové práce. V další podkapitole se zavedla výběrová verze distanční kovariance a korelace a vysvětlila se spojitost s výběrovými odhady charakteristických funkcí. Oba výběrové odhady distanční kovariance i korelace jsou silně konzistentními odhady a po představení aproximace rozdělení budoucí testové statistiky jsme tak mohli přejít ke konstrukci testu nezávislosti. Ten je v původním článku konstruován pro náhodné výběry. V této práci jsme ukázali, že pokud budeme předpokládat, že jsou příslušné bílé šumy oba posloupností nezávislých veličin, můžeme test distanční kovariancí použít na ně a využít tak stejný princip jako v Haughově testu.

Obsahem páté kapitoly byla simulační studie, ve které jsme ve více nastaveních porovnávali výše popsané testy a standardní t-test. Simulace byla zpracována ve výpočetním softwaru R, kde funkce pro všechny tři popsané testy byly vytvo-

řené autorem. Po představení postupu generování dat bylo provedeno ověření, že takto generovaná data splňují předpoklady všech testů. Výsledky testů, které jsou v práci jak graficky tak číselně uvedeny, jsme poté popsali pro každý test zvlášť, kde jsme navíc diskutovali o výhodách či nevýhodách každého testu. V našem případě měl největší sílu standardní t-test použitý na odhady bílých šumů. Námí zkonstruovaný test pomocí distanční kovariance je velmi konzervativní, ale s přibývajícím daty rostla jeho síla nejrychleji ze všech testů.

V poslední šesté kapitole jsme vyzkoušeli testy nezávislosti na finančních datech, konkrétně na cenách indexů S&P 500 a FTSE 100, kde byla ilustrována nevýhoda Pearsonova korelačního koeficientu. Po diskuzi výsledků všech testů a ověření předpokladů jsme usoudili, že nezávislost v tomto případě zamítáme na základě Haughova testu.

Literatura

- ALPARGU, G. a BUONACCORSI, J. (2009). A model-free test for independence between time series. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**(1), 115–132. doi: 10.1198/jabes.2009.0007.
- BOX, G. E. P. a JENKINS, G. M. (1970). *Time Series Analysis – Forecasting and Control*. Holden-Day, Inc., San Francisco, Calif.
- DUTILLEUL, P., CLIFFORD, P., RICHARDSON, S. a HEMON, D. (1993). Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, **49**(1), 305–314. doi: 10.2307/2532625.
- HAUGH, L. D. (1976). Checking the independence of two covariance-stationary time series: A univariate residual cross-correlation approach. *Journal of the American Statistical Association*, **71**(354), 378–385. doi: 10.1080/01621459.1976.10480353.
- KHAN, W. M. a KHAN, A. U. I. (2020). Most stringent test of independence for time series. *Communications in Statistics – Simulation and Computation*, **49**(11), 2808–2826. doi: 10.1080/03610918.2018.1527350.
- KOCH, P. D. a YANG, S.-S. (1986). A method for testing the independence of two time series that accounts for a potential pattern in the cross-correlation function. *Journal of the American Statistical Association*, **81**(394), 533–544. doi: 10.1080/01621459.1986.10478301.
- LYONS, R. (2013). Distance covariance in metric spaces. *The Annals of Probability*, **41**(5), 3284–3305. doi: 10.1214/12-aop803.
- PRÁŠKOVÁ, Z. (2016). *Základy náhodných procesů II*. Univerzita Karlova Nakladatelství Karolinum.
- R CORE TEAM (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- REHMAN, A.-U. a MALIK, M. I. (2014). The modified R a robust measure of association for time series. *Electronic Journal of Applied Statistical Analysis*, **7**(1), 1–13. doi: 10.1285/i20705948v7n1p1.
- SZÉKELY, G. J., RIZZO, M. L. a BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, **35**(6), 2769–2794. doi: 10.1214/009053607000000505.

THE WALL STREET JOURNAL (2023). Market data. <https://www.wsj.com/market-data>.

WOLFRAM RESEARCH, INC. (2023). Mathematica, Version 13.3. URL <https://www.wolfram.com/mathematica>. Champaign, IL.

A. Příloha

V příloze uvádíme číselné výsledky simulací popsané v kapitole 5.

ρ	H1	H2	Mt-ARMA	Mt-WN	DCT	St-ARMA	St-WN
-1,0	1,0000	1,0000	0,9999	1,0000	1,0000	1,0000	1,0000
-0,9	1,0000	1,0000	0,9998	1,0000	1,0000	0,9999	1,0000
-0,8	0,9981	0,9977	0,9874	1,0000	0,9908	0,9972	1,0000
-0,7	0,9595	0,9538	0,9290	0,9989	0,9292	0,9752	0,9991
-0,6	0,7878	0,7642	0,7900	0,9831	0,7190	0,8922	0,9847
-0,5	0,5147	0,4735	0,5927	0,9161	0,4250	0,7452	0,9227
-0,4	0,2675	0,2316	0,3987	0,7306	0,1817	0,5621	0,7424
-0,3	0,1345	0,1066	0,2362	0,4759	0,0584	0,3790	0,4890
-0,2	0,0710	0,0518	0,1146	0,2370	0,0153	0,2245	0,2441
-0,1	0,0402	0,0253	0,0573	0,0920	0,0025	0,1410	0,0954
0,0	0,0323	0,0209	0,0353	0,0472	0,0016	0,1091	0,0485
0,1	0,0375	0,0255	0,0605	0,0921	0,0024	0,1440	0,0947
0,2	0,0666	0,0475	0,1182	0,2311	0,0141	0,2303	0,2364
0,3	0,1334	0,1051	0,2265	0,4764	0,0601	0,3667	0,4854
0,4	0,2765	0,2380	0,4013	0,7345	0,1863	0,5600	0,7404
0,5	0,5008	0,4628	0,5952	0,9155	0,4141	0,7359	0,9221
0,6	0,7900	0,7665	0,7930	0,9840	0,7251	0,8939	0,9859
0,7	0,9555	0,9477	0,9274	0,9985	0,9239	0,9746	0,9987
0,8	0,9978	0,9974	0,9895	1,0000	0,9921	0,9985	1,0000
0,9	1,0000	1,0000	0,9994	1,0000	0,9999	0,9999	1,0000
1,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabulka A.1: Tabulka podílů zamítnutých testů v závislosti na korelaci ρ mezi bílými šumy pro jednoduché modely při použití $N = 50$.

ρ	H1	H2	Mt-ARMA	Mt-WN	DCT	St-ARMA	St-WN
-1,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,7	0,9999	0,9999	0,9994	1,0000	1,0000	0,9999	1,0000
-0,6	0,9864	0,9820	0,9812	1,0000	0,9972	0,9929	1,0000
-0,5	0,8480	0,8211	0,8884	0,9986	0,9197	0,9461	0,9988
-0,4	0,5403	0,4885	0,6975	0,9732	0,6301	0,8140	0,9743
-0,3	0,2561	0,2094	0,4403	0,8151	0,2588	0,5906	0,8217
-0,2	0,1123	0,0763	0,2199	0,4640	0,0588	0,3523	0,4671
-0,1	0,0501	0,0308	0,0844	0,1534	0,0071	0,1767	0,1571
0,0	0,0393	0,0234	0,0444	0,0492	0,0013	0,1121	0,0519
0,1	0,0536	0,0340	0,0808	0,1546	0,0077	0,1690	0,1586
0,2	0,1113	0,0785	0,2208	0,4629	0,0586	0,3539	0,4695
0,3	0,2564	0,2094	0,4352	0,8149	0,2668	0,5926	0,8179
0,4	0,5509	0,4954	0,6998	0,9735	0,6421	0,8163	0,9748
0,5	0,8480	0,8209	0,8870	0,9989	0,9137	0,9484	0,9990
0,6	0,9854	0,9819	0,9819	1,0000	0,9950	0,9936	1,0000
0,7	0,9996	0,9995	0,9987	1,0000	1,0000	0,9998	1,0000
0,8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
0,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
1,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabulka A.2: Tabulka podílů zamítnutých testů v závislosti na korelaci ρ mezi bílými šумы pro jednoduché modely při použití $N = 100$.

ρ	H1	H2	Mt-ARMA	Mt-WN	DCT	St-ARMA	St-WN
-1,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,6	0,9999	0,9999	1,0000	1,0000	1,0000	1,0000	1,0000
-0,5	0,9909	0,9869	0,9952	1,0000	0,9999	0,9988	1,0000
-0,4	0,8565	0,8139	0,9409	0,9998	0,9789	0,9740	0,9998
-0,3	0,4849	0,4091	0,7355	0,9862	0,7350	0,8457	0,9870
-0,2	0,1811	0,1270	0,3874	0,7821	0,2335	0,5352	0,7859
-0,1	0,0639	0,0364	0,1284	0,2776	0,0201	0,2398	0,2796
0,0	0,0418	0,0223	0,0474	0,0492	0,0008	0,1112	0,0501
0,1	0,0720	0,0389	0,1338	0,2766	0,0223	0,2347	0,2797
0,2	0,1904	0,1332	0,3896	0,7883	0,2324	0,5458	0,7910
0,3	0,4791	0,4010	0,7342	0,9891	0,7372	0,8409	0,9895
0,4	0,8518	0,8132	0,9431	1,0000	0,9817	0,9777	1,0000
0,5	0,9918	0,9874	0,9948	1,0000	0,9998	0,9991	1,0000
0,6	1,0000	1,0000	0,9999	1,0000	1,0000	1,0000	1,0000
0,7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
0,8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
0,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
1,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabulka A.3: Tabulka podílů zamítnutých testů v závislosti na korelaci ρ mezi bílými šумы pro jednoduché modely při použití $N = 200$.

ρ	H1	H2	Mt-ARMA	Mt-WN	DCT	St-ARMA	St-WN
-1,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,9	1,0000	0,9999	1,0000	1,0000	0,9991	1,0000	1,0000
-0,8	0,9926	0,9907	0,9974	0,9998	0,9807	0,9966	0,9999
-0,7	0,9200	0,9075	0,9774	0,9972	0,8787	0,9676	0,9977
-0,6	0,7163	0,6866	0,8881	0,9726	0,6378	0,8636	0,9741
-0,5	0,4355	0,3951	0,7208	0,8808	0,3494	0,6842	0,8884
-0,4	0,2295	0,1943	0,4963	0,6837	0,1483	0,4545	0,6936
-0,3	0,1139	0,0871	0,2961	0,4393	0,0442	0,2638	0,4480
-0,2	0,0607	0,0426	0,1503	0,2208	0,0128	0,1310	0,2266
-0,1	0,0373	0,0229	0,0698	0,0825	0,0018	0,0589	0,0859
0,0	0,0290	0,0194	0,0402	0,0448	0,0012	0,0327	0,0466
0,1	0,0343	0,0212	0,0702	0,0887	0,0017	0,0586	0,0906
0,2	0,0608	0,0411	0,1501	0,2175	0,0122	0,1330	0,2219
0,3	0,1179	0,0908	0,3030	0,4285	0,0468	0,2684	0,4347
0,4	0,2302	0,1950	0,4953	0,6877	0,1436	0,4568	0,6974
0,5	0,4320	0,3870	0,7108	0,8860	0,3419	0,6738	0,8903
0,6	0,7130	0,6805	0,8877	0,9727	0,6387	0,8598	0,9747
0,7	0,9160	0,9030	0,9755	0,9964	0,8712	0,9655	0,9969
0,8	0,9925	0,9915	0,9985	0,9999	0,9824	0,9974	0,9999
0,9	0,9999	0,9999	1,0000	1,0000	0,9990	1,0000	1,0000
1,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabulka A.4: Tabulka podílů zamítnutých testů v závislosti na korelaci ρ mezi bílými šумы pro komplexní modely při použití $N = 50$.

ρ	H1	H2	Mt-ARMA	Mt-WN	DCT	St-ARMA	St-WN
-1,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,7	0,9995	0,9993	0,9999	1,0000	1,0000	0,9999	1,0000
-0,6	0,9794	0,9735	0,9981	1,0000	0,9933	0,9956	1,0000
-0,5	0,8168	0,7865	0,9640	0,9976	0,8967	0,9501	0,9977
-0,4	0,4997	0,4460	0,8284	0,9661	0,5946	0,7867	0,9691
-0,3	0,2370	0,1894	0,5605	0,7984	0,2343	0,5070	0,8009
-0,2	0,1048	0,0748	0,2863	0,4443	0,0519	0,2435	0,4493
-0,1	0,0455	0,0287	0,1066	0,1449	0,0060	0,0821	0,1463
0,0	0,0405	0,0232	0,0483	0,0507	0,0009	0,0366	0,0515
0,1	0,0492	0,0310	0,1059	0,1507	0,0060	0,0811	0,1539
0,2	0,0997	0,0724	0,2846	0,4468	0,0544	0,2350	0,4522
0,3	0,2375	0,1920	0,5664	0,7945	0,2450	0,5111	0,7998
0,4	0,5170	0,4653	0,8355	0,9669	0,6048	0,7954	0,9680
0,5	0,8170	0,7849	0,9634	0,9982	0,8962	0,9481	0,9984
0,6	0,9806	0,9756	0,9980	1,0000	0,9932	0,9954	1,0000
0,7	0,9994	0,9994	1,0000	1,0000	1,0000	0,9999	1,0000
0,8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
0,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
1,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabulka A.5: Tabulka podílů zamítnutých testů v závislosti na korelaci ρ mezi bílými šумы pro komplexní modely při použití $N = 100$.

ρ	H1	H2	Mt-ARMA	Mt-WN	DCT	St-ARMA	St-WN
-1,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
-0,6	0,9999	0,9998	1,0000	1,0000	1,0000	1,0000	1,0000
-0,5	0,9896	0,9852	0,9999	1,0000	0,9996	0,9993	1,0000
-0,4	0,8438	0,7978	0,9871	0,9998	0,9772	0,9790	0,9998
-0,3	0,4676	0,3931	0,8602	0,9845	0,7183	0,8266	0,9852
-0,2	0,1746	0,1227	0,5272	0,7710	0,2205	0,4622	0,7732
-0,1	0,0617	0,0345	0,1720	0,2714	0,0180	0,1329	0,2747
0,0	0,0436	0,0234	0,0481	0,0496	0,0010	0,0314	0,0490
0,1	0,0688	0,0385	0,1685	0,2708	0,0200	0,1291	0,2763
0,2	0,1856	0,1290	0,5336	0,7816	0,2261	0,4668	0,7840
0,3	0,4695	0,3894	0,8709	0,9879	0,7200	0,8296	0,9884
0,4	0,8406	0,7985	0,9861	0,9999	0,9792	0,9792	0,9999
0,5	0,9911	0,9878	0,9999	1,0000	0,9998	0,9997	1,0000
0,6	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
0,7	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
0,8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
0,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
1,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Tabulka A.6: Tabulka podílů zamítnutých testů v závislosti na korelaci ρ mezi bílými šumy pro komplexní modely při použití $N = 200$.