



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Bc. Jakub Hledík

Binomický autoregresní model

Katedra pravděpodobnosti a matematické statistiky

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

Studijní program: Matematika

Studijní obor: Pravděpodobnost, matematická
statistika a ekonometrie

Praha 2020

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

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

V dne

Podpis autora

Tímto bych chtěl poděkovat mé vedoucí RNDr. Šárce Hudecové, Ph.D. za velké množství cenných rad a připomínek a za veškerý čas, který mi při psaní práce ochotně věnovala.

Název práce: Binomický autoregresní model

Autor: Bc. Jakub Hledík

Katedra: Katedra pravděpodobnosti a matematické statistiky

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

Abstrakt: Binomický AR(1) proces je model pro celočíselné časové řady s konečným oborem hodnot a diskrétním časem, který má binomické marginální rozdělení a autokorelační strukturu stejnou jako standardní AR(1) proces. Tato práce se zabývá odvozením základních vlastností procesu, metodami odhadu parametrů a testy dobré shody. Uvádí se zde tři metody odhadu parametrů: Yuleova-Walkerova, podmíněných nejmenších čtverců a maximální věrohodnosti, u všech postupů se ukazují jejich asymptotické vlastnosti. Následují testy dobré shody, kde jsou nejprve shrnuty dvě známé metody založené na marginálním rozdělení a autokorelační funkci procesu. Ty jsou doplněny novou vlastní metodou založenou na vytvořující funkci. Vlastnosti všech testů jsou ukázány na simulacích, aplikace modelu je na závěr předvedena na reálných datech.

Klíčová slova: binomický AR(1) proces, odhad parametrů, test dobré shody

Title: Binomial autoregressive model

Author: Bc. Jakub Hledík

Department: Department of Probability and Mathematical Statistics

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

Abstract:

Binomial AR(1) process is a model for integer-valued time series with a finite range and discrete time. It has the binomial marginal distribution and the AR(1)-like autocorrelation structure. This thesis deals with deriving some basic properties of this process, methods of parameter estimation and goodness of fit testing. Three methods of parameter estimation are presented: Yule-Walker, the conditional least squares and the maximum likelihood method together with proofs of their asymptotical properties. Next, the goodness of fit testing is presented. At first, two known methods based on the marginal distribution and the autocorrelation function are summarized. Then our own method is added, based on the probability generating function. Several simulations are provided to show the properties of all tests. The application of this model is illustrated on a real dataset.

Keywords: binomial AR(1) process, parameter estimation, goodness of fit test

Obsah

Úvod	2
1 Základní vlastnosti	3
1.1 Definice binomického AR(1) procesu	4
1.2 Vlastnosti binomického AR(1) procesu	6
2 Odhady parametrů	12
2.1 Yuleova-Walkerova metoda	12
2.2 Metoda podmíněných nejmenších čtverců	14
2.3 Metoda maximální věrohodnosti	20
3 Testy dobré shody	24
3.1 Test založený na marginálním rozdělení	24
3.2 Test založený na autokorelační struktuře	26
3.3 Test založený na vytvořující funkci	29
4 Simulace	33
4.1 Podíl odhadů v přípustné množině	33
4.2 Empirická hladina testů dobré shody	34
4.3 Síla testů dobré shody	34
4.3.1 Beta-binomický AR(1) model	34
4.3.2 Binomický AR(2) model	36
4.3.3 Beta-binomický AR(2) model	37
4.4 Nesprávná velikost populace n	38
5 Aplikace na reálná data	52
Závěr	61
Seznam použité literatury	62

Úvod

Zatímco pro časové řady s reálnými hodnotami existuje nepřeberné množství modelů, tak časové řady s diskrétním oborem hodnot nejsou příliš zkoumány. Často je totiž obor možných hodnot natolik velký, že není problém uvažovat jej spojitě. Například při hledání modelu pro vývoj ceny akcie společnosti ČEZ zřejmě nemusíme brát v potaz, že se její cena může pohybovat pouze po celých desetihaléřích, pokud se cena nachází v řádech stokorun. Naproti tomu při modelování počtu společností na pražské burze, s jejichž akciemi proběhne alespoň jeden obchod v rámci zadaného časového intervalu, je oborem hodnot pouze množina $\{0, 1, \dots, 11\}$.

Zmíněnou časovou řadu s konečným oborem hodnot bychom mohli popsat například pomocí Markovových řetězců. Takové řešení by ale vyžadovalo popsat matici přechodu mezi jednotlivými stavy, což by ovšem vedlo k velkému množství parametrů a bylo by tak zapotřebí i velké množství pozorování k zajištění spolehlivých odhadů. Pro složitější autokorelační strukturu bychom navíc museli uvažovat řetězce vyšších řádů, čímž by se ještě více zvýšil počet parametrů. (McKenzie, 2003, kap. 2) sice zmiňuje některá zjednodušení vzhledem k redukci parametrů, i tak ale nenašly tyto modely široké uplatnění. V porovnání s ARMA modely by takto navíc nešlo jednoduše konstruovat typologicky podobné modely.

Tato práce se tak zabývá modelem popsaným pouhými dvěma parametry. Model má binomické marginální rozdělení a autokorelační strukturou stejnou jako v případě klasického AR(1) procesu. Tento model se v literatuře označuje jako *binomický AR(1) model* a byl navržen již v (McKenzie, 1985). Podrobně se mu ale věnuje až Christian Weiß, poprvé v (Weiß, 2009b) a v následujících letech i v dalších článcích, ze kterých budeme vycházet. V první kapitole uvádíme základní definice a podrobně rozepisujeme některé vlastnosti binomického AR(1) procesu. Dále ukážeme, že i takto definovaný proces splňuje markovskou podmínku, což nám umožní používat výsledky z této oblasti.

Druhá kapitola je věnována odhadům parametrů v představeném modelu. Shrnujeme v ní tři metody z (Weiß a Kim, 2013a) spolu s důkazy asymptotických vlastností. Tyto důkazy a veškerá odvození zde v porovnání s citovanou literaturou rozvádíme detailněji.

Ve třetí kapitole se zabýváme otázkou ověření, jestli je binomický AR(1) model vhodným kandidátem pro zadaná data. Nejprve jsou shrnuty dva testy dobré shody z (Weiß a Kim, 2015), každý z nich cílí na jinou charakteristiku binomického AR(1) modelu. U obou testů rozepisujeme podrobně po krocích jejich použití. Dále uvádíme vlastní test dobré shody, který je založený na vytvořující funkci. Všechny testy následně srovnáváme v simulační studii. Tyto testy jsme sami implementovali ve statistickém softwaru (R Core Team, 2020) a příslušné zdrojové kódy jsou k dispozici na <https://labk10.karlin.mff.cuni.cz/~hledikja/funkce.R>.

V poslední kapitole pak ukazujeme aplikaci binomického AR(1) modelu na reálná data. Jedná se právě o modelování počtu společností na pražské burze, s jejichž akciemi proběhne alespoň jeden obchod v rámci zadaného časového intervalu, kterým jsme v úvodu motivovali zkoumání celočíselných náhodných procesů.

1. Základní vlastnosti

Na rozdíl od náhodných procesů s neomezeným reálným oborem hodnot nemůžeme zapsat celočíselný autoregresní proces prvního řádu výrazem

$$X_t = \alpha \cdot X_{t-1} + \epsilon_t,$$

kde $|\alpha| < 1$, $\{\epsilon_t\}$ je bílý šum a ϵ_t jsou nezávislé na předchozích hodnotách procesu. Problém je v tom, že $\alpha \cdot X_{t-1}$ obecně není celočíselné a tím pádem by ϵ_t muselo být závislé na X_{t-1} , aby byla zachována jednak celočíselnost X_t , ale také aby hodnoty ležely v oboru hodnot pro X_t . Řešením proto bude uvažovat odlišnou strukturu vytváření nových hodnot než je přičítání nezávislých inovací $\{\epsilon_t\}$, ale také uvažovat jinou operaci než násobení, která jednak zajistí celočíselnost s vhodným oborem hodnot a zároveň v sobě bude obsahovat náhodnost. Představíme nyní operaci označovanou v anglické literatuře jako „thinning“, která tyto požadavky splňuje.

Definice 1 (Operace \circ). *Bud' X diskrétní náhodná veličina s hodnotami v \mathbb{N}_0 a $\alpha \in [0, 1]$, potom definujeme náhodnou veličinu $\alpha \circ X$ jako*

$$\alpha \circ X := \sum_{i=1}^X Z_i,$$

kde Z_i jsou iid náhodné veličiny s alternativním rozdělením se střední hodnotou α , nezávislé na X .

Takto definovaná operace má obě požadované vlastnosti, z definice bude náhodná veličina $\alpha \circ X$ opět celočíselná a se stejným oborem hodnot jako X . Podmíněně náhodnou veličinou X dokonce přesně známe její rozdělení, což ukážeme v následujícím tvrzení.

Tvrzení 1 (Podmíněné rozdělení $\alpha \circ X$). *Bud' X diskrétní náhodná veličina s hodnotami v \mathbb{N}_0 a $\alpha \in [0, 1]$, potom podmíněně X je $\alpha \circ X$ náhodná veličina s binomickým rozdělením s parametry X a α .*

Důkaz. Z definice je $\alpha \circ X = \sum_{i=1}^X Z_i$ pro Z_i nezávislé s rozdělením $\text{Alt}(\alpha)$. Podmíněně hodnotou $X = x$ je potom na pravé straně náhodná veličina s binomickým rozdělením $\text{Bi}(x, \alpha)$. □

Operace \circ je navíc v jistém smyslu podobná běžnému násobení. Jak totiž ukážeme v následujícím tvrzení, ve střední hodnotě jsou obě operace stejné. To již nebude platit pro rozptyl, ukážeme, že $\text{var}(\alpha \circ X) \geq \text{var}(\alpha \cdot X)$.

Tvrzení 2 (Momenty $\alpha \circ X$). *Pro $\alpha \in [0, 1]$ platí*

$$\begin{aligned} E[\alpha \circ X] &= \alpha E[X], \\ \text{var}[\alpha \circ X] &= \alpha^2 \text{var}[X] + \alpha(1 - \alpha) E[X]. \end{aligned}$$

Důkaz. Z vlastností podmíněné střední hodnoty a tvrzení 1 máme

$$\mathbf{E}[\alpha \circ X] = \mathbf{E}(\mathbf{E}[\alpha \circ X|X]) = \mathbf{E}[\alpha X] = \alpha \mathbf{E}[X].$$

Rozptyl dostaneme z téhož tvrzení a vlastností podmíněného rozptylu:

$$\begin{aligned} \text{var}[\alpha \circ X] &= \text{var}(\mathbf{E}[\alpha \circ X|X]) + \mathbf{E}(\text{var}[\alpha \circ X|X]) = \\ &= \text{var}[\alpha X] + \mathbf{E}[\alpha(1 - \alpha)X] \\ &= \alpha^2 \text{var}[X] + \alpha(1 - \alpha) \mathbf{E}[X]. \end{aligned}$$

□

Tím jsme shrnuli potřebné vlastnosti operace \circ a můžeme ji nyní použít k definici binomického AR(1) procesu.

1.1 Definice binomického AR(1) procesu

Definice 2 (Binomický AR(1) proces podle (Weiß, 2009b)). *Buď $n \in \mathbb{N}$, $\pi \in (0, 1)$, $\rho \in (\max\{\frac{-\pi}{1-\pi}, \frac{1-\pi}{-\pi}\}, 1)$ a označme $\beta := \pi(1 - \rho)$ a $\alpha := \beta + \rho$. Potom proces $(X_t, t \in \mathbb{Z})$ definovaný předpisem*

$$X_t = \alpha \circ X_{t-1} + \beta \circ (n - X_{t-1}),$$

kde obě operace \circ jsou navzájem nezávislé a navíc nezávislé na $(X_s, s < t)$, nazveme binomický AR(1) proces.

Na obrázku 1.1 vidíme přípustnou množinu parametrů (π, ρ) podle definice 2. Omezení hodnoty ρ zajistí, aby parametry (α, β) ležely v $(0, 1)^2$ a obě dvě operace \circ byly správně definovány.

Tento proces je původně zmiňován již v (McKenzie, 1985), autor zde ale definuje proces pomocí parametrů $\pi \in (0, 1/2)$ a $\alpha \in (0, 1)$, parametry β, ρ jsou z nich dopočítány. Zmiňme tedy nyní také tuto původní definici a ukažme, že jsou obě definice ekvivalentní.

Definice 3 (Binomický AR(1) proces podle (McKenzie, 1985)). *Buď $n \in \mathbb{N}$, $\pi \in (0, 1/2)$, $\alpha \in (0, 1)$ a označme $\beta = (1 - \alpha)\pi/(1 - \pi)$ a $\rho = (\alpha - \pi)/(1 - \pi)$. Potom proces $(X_t, t \in \mathbb{Z})$ definovaný předpisem*

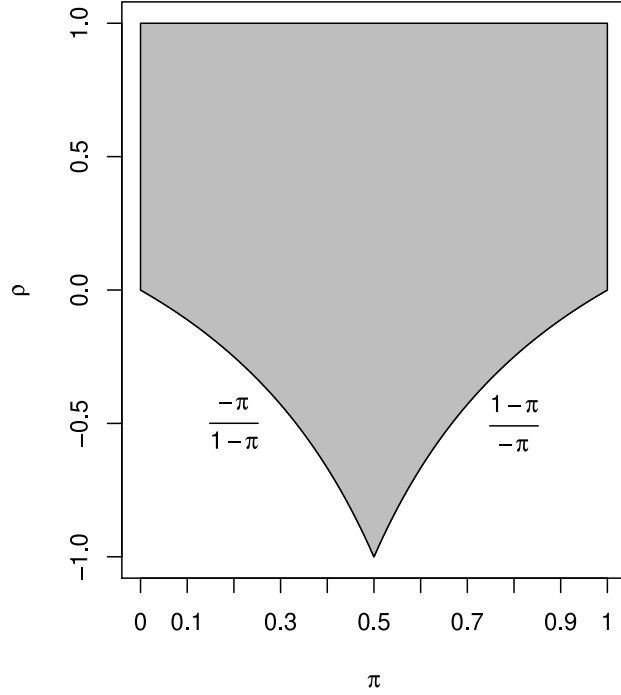
$$X_t = \alpha \circ X_{t-1} + \beta \circ (n - X_{t-1}),$$

kde obě operace \circ jsou navzájem nezávislé a navíc nezávislé na $(X_s, s < t)$, nazveme binomický AR(1) proces.

Tvrzení 3. *Definice 2 a 3 binomického AR(1) procesu jsou ekvivalentní.*

Důkaz. Buď $\pi \in (0, 1/2)$, $\alpha \in (0, 1)$, označme $\beta_W = \pi(1 - \rho_W)$ a $\rho_W = \alpha - \beta_W$ podle definice 2 a dále $\beta_{MK} = \frac{(1-\alpha)\pi}{1-\pi}$ a $\rho_{MK} = \frac{\alpha-\pi}{1-\pi}$ podle definice 3. Po dosazení dostáváme

$$\begin{aligned} \rho_W &= \alpha - \beta_W = \alpha - \pi(1 - \rho_W) = \alpha - \pi + \pi\rho_W, \\ (1 - \pi)\rho_W &= \alpha - \pi = (1 - \pi)\rho_{MK}. \end{aligned}$$



Obrázek 1.1: Přípustné hodnoty parametrů π a ρ .

Tímto se tedy shodují parametry $\rho_{MK} = \rho_W \equiv \rho$. Následně vyjádříme

$$\beta_W = \alpha - \rho_W = \alpha - \rho_{MK} = \alpha - \frac{\alpha - \pi}{1 - \pi} = \frac{\alpha - \alpha\pi - \alpha + \pi}{1 - \pi} = \frac{(1 - \alpha)\pi}{1 - \pi} = \beta_{MK}$$

a zbývá ověřit, že výše definované ρ leží v intervalu $(\max\{\frac{-\pi}{1-\pi}, \frac{1-\pi}{-\pi}\}, 1)$. Z předpokladu $\pi \in (0, 1/2)$ platí $\frac{-\pi}{1-\pi} \geq \frac{1-\pi}{-\pi}$ a navíc si můžeme všimnout, že ρ lze z McKenzieho definice také zapsat jako

$$\alpha + (1 - \alpha) \frac{-\pi}{1 - \pi} = \frac{\alpha - \pi\alpha - \pi + \alpha\pi}{1 - \pi} = \frac{\alpha - \pi}{1 - \pi} = \rho,$$

kde se na levé straně nachází konvexní kombinace čísel $\frac{-\pi}{1-\pi}$ a 1, takže skutečně $\rho \in (\frac{-\pi}{1-\pi}, 1)$.

Pro situaci $\pi \in (1/2, 1)$ budeme uvažovat proces Y_t s parametry $(\alpha, \tilde{\pi})$, kde $\tilde{\pi} := 1 - \pi \in (0, 1/2)$. V tomto procesu bude

$$\tilde{\beta} = \frac{(1 - \alpha)\tilde{\pi}}{1 - \tilde{\pi}} = \frac{(1 - \alpha)(1 - \pi)}{\pi}, \quad \tilde{\rho} = \frac{\alpha - \tilde{\pi}}{1 - \tilde{\pi}} = \frac{\alpha + \pi - 1}{\pi}$$

a jeho parametrizace bude díky předchozímu shodná s definicí 2. Nyní stačí uvažovat $X_t = n - Y_t$, který už bude taktéž parametrizován shodně s definicí 2. □

S přihlédnutím k této složitější konstrukci pro $\pi \in (1/2, 1)$ dáme přednost používání parametrizace procesu od Weiße podle definice 2. Parametrizace pomocí (ρ, π) se také ukáže výhodnější vzhledem k interpretaci těchto parametrů, k čemuž se ovšem dostaneme později.

Interpretaci binomického AR(1) procesu můžeme z pohledu parametrů (α, β) ilustrovat následujícím příkladem: V hotelu se nachází n pokojů. Označme X_t počet obsazených pokojů ve dni t , $n - X_t$ potom bude počet volných pokojů v témže dni. Během dne t se hosté v X_t obsazených pokojích rozhodnou, jestli zůstanou ubytovaní i v dalším dni — každý obsazený pokoj zůstane obsazený nezávisle na ostatních s pravděpodobností α . S pravděpodobností β potom bude každý neobsazený pokoj v dalším dni obsazen nezávisle na ostatních. Všechna rozhodování probíhají také nezávisle na počtu obsazených pokojů, a to jak aktuálním, tak i všech minulých. Dalším příkladem může být populace n jedinců, kde X_t je počet přítomných jedinců v čase t nebo například n telefonních linek v call centru, ze kterých je X_t používaných v čase t .

1.2 Vlastnosti binomického AR(1) procesu

Hned na začátku ukážeme, že binomický AR(1) proces splňuje markovskou vlastnost, což nám umožní používat výsledky z teorie Markovových řetězců. Následující značení a terminologii budeme používat v souladu se skripty (Prášková a Lachout, 2012), odkud také implicitně používáme tvrzení o základních vlastnostech Markovových řetězců. Netriviální vztahy budeme používat s odkazy na příslušná tvrzení.

Tvrzení 4. *Binomický AR(1) proces je Markovův řetězec.*

Důkaz. Přímo z předpisu pro X_t v definici 2 plyne, že rozdělení X_t podmíněně X_{t-1} již nezávisí na (X_{t-2}, \dots, X_1) a markovská podmínka je tak splněna. \square

Poznámka. V souladu s běžným značením v teorii Markovových řetězců budeme dále indexovat proces přirozenými čísly včetně nuly: $(X_t, t \in \mathbb{N}_0)$. O rozdělení náhodné veličiny X_0 budeme hovořit jako o *počátečním rozdělení*.

Tvrzení 5 (Pravděpodobnost přechodu). *Binomický AR(1) proces je homogenní Markovův řetězec a pro $i, j \in \{0, \dots, n\}$, $t \in \mathbb{Z}$ platí*

$$P(X_t = j | X_{t-1} = i) = \sum_{m=\max(0, i+j-n)}^{\min(i, j)} \binom{i}{m} \binom{n-i}{j-m} \alpha^m (1-\alpha)^{i-m} \beta^{j-m} (1-\beta)^{n-i+m-j}.$$

Důkaz. Z i přítomných jedinců v čase $t-1$ se na $X_t = j$ přítomných v čase t dostaneme následovně: m jedinců z i musí zůstat, tento jev má pravděpodobnost $\binom{i}{m} \alpha^m (1-\alpha)^{i-m}$. Zbýlých $j-m$ jedinců musí přibýt z $n-i$ jedinců nepřítomných v čase $t-1$. Tento jev má pravděpodobnost $\binom{n-i}{j-m} \beta^{j-m} (1-\beta)^{n-i+m-j}$. Z nezávislosti obou jevů je sdružená pravděpodobnost obou jevů součinem jejich dílčích pravděpodobností. Výslednou pravděpodobnost přechodu získáme součtem přes možná m . Zřejmě $m \leq i$ (nemůže zůstat víc jedinců než je jich přítomných) a také $m \leq j$, protože počet nově příchozích musí být nezáporný. Naopak $m \geq 0$, protože počet zůstávajících jedinců je nezáporný a také $m \geq j - (n-i) = i + j - n$

protože do požadovaných j jedinců v čase t může přibýt maximálně $n - i$ jedinců, tudíž jich musí dostatečný počet zůstat. Dohromady již dostáváme požadovanou rovnost, která nezávisí na t , což dokazuje homogenitu procesu. \square

Poznámka. V dalším budeme značit pravděpodobnost přechodu výrazem $p_{i,j}$ nebo pro lepší čitelnost také výrazem $p(j|i; \rho, \pi)$. Tímto zápisem také zdůrazňujeme, že tuto pravděpodobnost uvažujeme s parametry (ρ, π) .

Tvrzení 6. *Binomický AR(1) proces je ergodický Markovův řetězec.*

Důkaz. Z vyjádření pravděpodobností přechodu vidíme, že jsou všechny stavy navzájem přímo dosažitelné, a řetězec je tak nerozložitelný s neperiodickými stavy. Dále má konečnou množinu stavů, a tedy z (Prášková a Lachout, 2012, věta 2.19) jeho stavy trvalé nenulové. Z definice 2.39 je pak binomický AR(1) proces ergodický Markovův řetězec. \square

Z ergodicity řetězce plyne z (Prášková a Lachout, 2012, věta 2.26) existence stacionárního rozdělení, které je zároveň jediné. Nyní následují dvě zásadní tvrzení, která popíší marginální rozdělení a autokorelační strukturu binomického AR(1) procesu.

Tvrzení 7 (Stacionární rozdělení binomického AR(1) procesu). *Stacionární rozdělení binomického AR(1) procesu je $Bi(n, \pi)$.*

Důkaz. Ukážeme, že pokud má pro libovolné $t \in \mathbb{N}_0$ náhodná veličina X_t rozdělení $Bi(n, \pi)$, potom má X_{t+1} také rozdělení $Bi(n, \pi)$. Označme $P_X(s) := \mathbf{E} t^X$ vytvářející funkci náhodné veličiny X . Pro náhodnou veličinu X_t s binomickým rozdělením s parametry n a π má její vytvářející funkce tvar

$$\begin{aligned} P_{X_t}(s) &= \sum_{k=0}^n s^k \mathbf{P}(X_t = k) = \sum_{k=0}^n s^k \binom{n}{k} \pi^k (1 - \pi)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (s\pi)^k (1 - \pi)^{n-k} = (s\pi + 1 - \pi)^n. \end{aligned}$$

Pro náhodnou veličinu $X_{t+1} = \alpha \circ X_t + \beta \circ (n - X_t)$ potom platí

$$\begin{aligned} P_{X_{t+1}}(s) &= \mathbf{E} s^{\alpha \circ X_t + \beta \circ (n - X_t)} \\ &= \mathbf{E} \left[\mathbf{E} s^{\alpha \circ X_t + \beta \circ (n - X_t)} | X_t \right] \\ &= \mathbf{E} \left[\mathbf{E} s^{\alpha \circ X_t} | X_t \right] \mathbf{E} \left[\mathbf{E} s^{\beta \circ (n - X_t)} | X_t \right], \end{aligned}$$

kde poslední rovnost platí díky podmíněné nezávislosti $\alpha \circ X_t$ a $\beta \circ (n - X_t)$ podmíněně X_t . Tyto náhodné veličiny mají navíc podmíněně X_t binomické rozdělení, viz tvrzení 1. Máme tedy

$$\begin{aligned} \mathbf{E} \left[\mathbf{E} s^{\alpha \circ X_t} | X_t \right] &= \mathbf{E} \left[P_{\alpha \circ X_t | X_t}(s) \right] = \mathbf{E} (s\alpha + 1 - \alpha)^{X_t}, \\ \mathbf{E} \left[\mathbf{E} s^{\beta \circ (n - X_t)} | X_t \right] &= \mathbf{E} \left[P_{\beta \circ (n - X_t) | X_t}(s) \right] = \mathbf{E} (s\beta + 1 - \beta)^{n - X_t}. \end{aligned}$$

Dohromady potom

$$P_{X_{t+1}}(s) = (s\beta + 1 - \beta)^n \mathbb{E} \left[\frac{s\alpha + 1 - \alpha}{s\beta + 1 - \beta} \right]^{X_t},$$

kde druhý člen je vytvořující funkce náhodné veličiny X_t v bodě $\frac{s\alpha+1-\alpha}{s\beta+1-\beta}$. Platí tedy

$$\begin{aligned} P_{X_{t+1}}(s) &= (s\beta + 1 - \beta)^n \cdot P_{X_t} \left(\frac{s\alpha + 1 - \alpha}{s\beta + 1 - \beta} \right) & (1.1) \\ &= (s\beta + 1 - \beta)^n \cdot \left[\frac{s\alpha + 1 - \alpha}{s\beta + 1 - \beta} \pi + 1 - \pi \right]^n \\ &= [(1 - \pi)(s\beta + 1 - \beta) + \pi(s\alpha + 1 - \alpha)]^n \\ &= [s\beta + 1 - \beta - s\beta\pi - \pi + \beta\pi + s\pi\alpha + \pi - \alpha\pi]^n \\ &= [s\beta + 1 - \beta - s\beta\pi + \beta\pi + s\alpha\pi - \alpha\pi]^n \\ &= [s(\beta - \beta\pi + \alpha\pi) + 1 - \beta + \beta\pi - \alpha\pi]^n. \end{aligned}$$

Jelikož nyní $\beta - \beta\pi + \alpha\pi = \beta - \rho\pi = \rho - \rho\pi - \rho\pi = \pi$, dostáváme

$$P_{X_{t+1}}(s) = (s\pi + 1 - \pi)^n,$$

což je vytvořující funkce rozdělení $\text{Bi}(n, \pi)$. □

Tvrzení 8 (Autokorelační funkce binomického AR(1) procesu). *Nechť má binomický AR(1) proces počáteční rozdělení $\text{Bi}(n, \pi)$, potom je striktně stacionární a jeho autokorelační funkce je pro $k \in \mathbb{N}_0$ tvaru*

$$\rho(k) = \rho^k.$$

Důkaz. Buď $k, l \in \mathbb{N}$, potom s využitím homogenity a stacionárního rozdělení z tvrzení 7 platí

$$\begin{aligned} \mathbb{P}(X_l = i_0, X_{l+1} = i_1, \dots, X_{l+k} = i_k) &= \\ &= \mathbb{P}(X_l = i_0) \mathbb{P}(X_{l+1} = i_1 | X_l = i_0) \dots \mathbb{P}(X_{l+k} = i_k | X_{l+k-1} = i_{k-1}) \\ &= \mathbb{P}(X_l = i_0) p(i_1 | i_0) \dots p(i_k | i_{k-1}) \\ &= \mathbb{P}(X_0 = i_0) p(i_1 | i_0) \dots p(i_k | i_{k-1}) \\ &= \mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k), \end{aligned}$$

což dokazuje striktní stacionaritu procesu.

Pro $k \in \mathbb{N}$ můžeme potom přímo spočítat autokovarianční funkci, která bude díky stacionaritě funkcí jedné proměnné:

$$\begin{aligned} R(k) &= \text{cov}(X_{t+k}, X_t) = \text{cov}[\alpha \circ X_{t+k-1} + \beta \circ (n - X_{t+k-1}), X_t] \\ &= \text{cov}[\alpha \circ X_{t+k-1}, X_t] + \text{cov}[\beta \circ (n - X_{t+k-1}), X_t], \end{aligned}$$

kde nyní pro přehlednost vyhodnotíme oba členy zvlášť. Označme dále $\mathcal{F}_t = \sigma(\bigcup_{0 \leq s \leq t} X_s)$ sigma algebru obsahující veškerou informaci o procesu do času t včetně. Potom zřejmě X_s je \mathcal{F}_t -měřitelná pro všechna $s \leq t$. Platí tedy

$$\begin{aligned} \text{cov}[\alpha \circ X_{t+k-1}, X_t] &= \mathbf{E}(\alpha \circ X_{t+k-1})X_t - \mathbf{E}(\alpha \circ X_{t+k-1})\mathbf{E} X_t \\ &= \mathbf{E} \mathbf{E}[(\alpha \circ X_{t+k-1})X_t | \mathcal{F}_{t+k-1}] - \mathbf{E}(\alpha X_{t+k-1})\mathbf{E} X_t \\ &= \mathbf{E} X_t \mathbf{E}[\alpha \circ X_{t+k-1} | \mathcal{F}_{t+k-1}] - \alpha \mathbf{E} X_{t+k-1} \mathbf{E} X_t \\ &= \mathbf{E} X_t (\alpha X_{t+k-1}) - \alpha \mathbf{E} X_{t+k-1} \mathbf{E} X_t \\ &= \alpha [\mathbf{E}(X_{t+k-1}X_t) - \mathbf{E} X_{t+k-1} \mathbf{E} X_t] \\ &= \alpha \text{cov}(X_{t+k-1}, X_t) = \alpha R(k-1) \end{aligned}$$

a obdobným způsobem dostaneme

$$\begin{aligned} \text{cov}[\beta \circ (n - X_{t+k-1}), X_t] &= \mathbf{E}[\beta \circ (n - X_{t+k-1})]X_t - \mathbf{E}[\beta \circ (n - X_{t+k-1})]\mathbf{E} X_t \\ &= \mathbf{E} \beta(n - X_{t+k-1})X_t - \mathbf{E}[\beta(n - X_{t+k-1})]\mathbf{E} X_t \\ &= \beta[n \mathbf{E} X_t - \mathbf{E}(X_{t+k-1}X_t) - n \mathbf{E} X_t + \mathbf{E} X_{t+k-1} \mathbf{E} X_t] \\ &= -\beta \text{cov}(X_{t+k-1}, X_t) = -\beta R(k-1). \end{aligned}$$

Dohromady potom

$$R(k) = \alpha R(k-1) - \beta R(k-1) = \rho R(k-1)$$

a tento rekurentní vztah lze již pro $k \in \mathbb{N}_0$ přímo přepsat do tvaru

$$R(k) = \rho^k R(0),$$

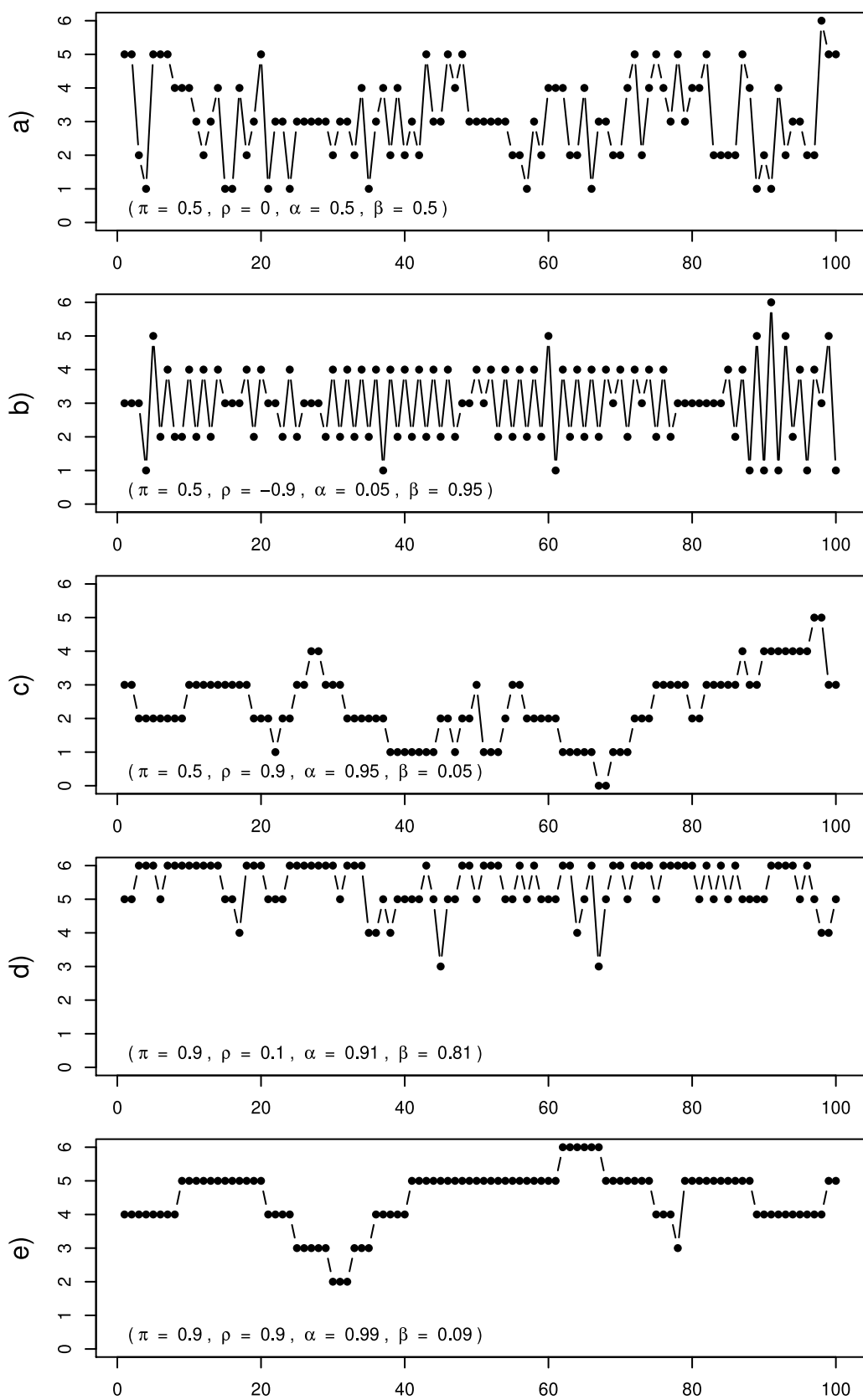
ze kterého přejdeme na autokorelační funkci vztahem

$$\rho(k) = R(k)/R(0) = \rho^k,$$

což jsme chtěli dokázat. □

Poznámka. Všimněme si, že autokorelační funkce tohoto celočíselného procesu je stejná jako v případě klasického AR(1) modelu.

Z předchozích tvrzení také plyne význam parametrů (π, ρ) . Tyto dva parametry reprezentují polohu a autokorelaci procesu, konkrétně parametr ρ je korelace procesu mezi sousedními časy a $n\pi$ je jeho střední hodnota. Můžeme si nyní interpretovat některé volby parametrů, na obrázku 1.2 pak nalezneme příklady jejich realizací. Na obrázku 1.2a vidíme nekorelovaný proces, pro něj platí $\alpha = \beta = \pi$, tedy jedinci odchází i přichází se stejnou pravděpodobností. Proces je v tomto případě ekvivalentní posloupnosti nezávislých stejně rozdělených veličin s binomickým rozdělením. Naopak na obrázcích 1.2b a 1.2c jsou realizace procesů s vysokou korelací (v absolutní hodnotě) sousedních pozorování. Pro zápornou korelaci to znamená, že jen zřídka zůstanou jedinci přítomní v čase t do času $t+1$ a zároveň téměř všichni nepřítomní se v dalším čase přidají. Pro kladnou korelaci tomu je přesně naopak, zde téměř všichni zůstávají a jen zřídka někdo nový přibude. Na obrázcích 1.2d a 1.2e můžeme sledovat ukázkou chování procesu



Obrázek 1.2: Příklady realizací procesů pro $n = 6, t = 0, \dots, 100$ a různé hodnoty parametrů π a ρ .

se střední hodnotou ležící blízko celkovému počtu všech jedinců n , v prvním případě s malou korelovaností mezi sousedními členy a ve druhém případě naopak s vysokou.

Mezi základními vlastnostmi uvedeme ještě tvrzení o momentech procesu podmíněně hodnotou z předchozího času, které využijeme později při odhadech parametrů.

Tvrzení 9 (Podmíněné jednokrokové momenty X_t). *Pro binomický $AR(1)$ proces platí*

$$\begin{aligned} E[X_t|X_{t-1}] &= \rho X_{t-1} + n\beta, \\ \text{var}[X_t|X_{t-1}] &= \rho(1 - \rho)(1 - 2\pi)X_{t-1} + n\beta(1 - \beta). \end{aligned}$$

Důkaz.

$$\begin{aligned} E[X_t|X_{t-1}] &= E[\alpha \circ X_{t-1} + \beta \circ (n - X_{t-1})|X_{t-1}] = \\ &= E[\alpha \circ X_{t-1}|X_{t-1}] + E[\beta \circ (n - X_{t-1})|X_{t-1}] = \\ &= \alpha X_{t-1} + \beta(n - X_{t-1}) = \\ &= (\alpha - \beta)X_{t-1} + \beta n = \rho X_{t-1} + n\beta. \end{aligned}$$

Obdobně postupujeme při výpočtu rozptylu:

$$\begin{aligned} \text{var}[X_t|X_{t-1}] &= \text{var}[\alpha \circ X_{t-1} + \beta \circ (n - X_{t-1})|X_{t-1}] = \\ &= \text{var}[\alpha \circ X_{t-1}|X_{t-1}] + \text{var}[\beta \circ (n - X_{t-1})|X_{t-1}] = \\ &= \alpha(1 - \alpha)X_{t-1} + \beta(1 - \beta)(n - X_{t-1}) = \\ &= [\alpha(1 - \alpha) - \beta(1 - \beta)]X_{t-1} + n\beta(1 - \beta) = \\ &= [(\pi(1 - \rho) + \rho)(1 - \pi(1 - \rho) - \rho) - \pi(1 - \rho)(1 - \pi(1 - \rho))]X_{t-1} \\ &\quad + n\beta(1 - \beta) = \\ &= [(\pi - \pi\rho + \rho)(1 - \pi + \pi\rho - \rho) - (\pi - \pi\rho)(1 - \pi + \pi\rho)]X_{t-1} \\ &\quad + n\beta(1 - \beta) = \\ &= [(\pi - \pi\rho)(1 - \pi + \pi\rho - \rho) - (\pi - \pi\rho)(1 - \pi + \pi\rho)]X_{t-1} \\ &\quad + [\rho(1 - \pi + \pi\rho - \rho)]X_{t-1} + n\beta(1 - \beta) = \\ &= [(\pi - \pi\rho)(1 - \pi + \pi\rho) - (\pi - \pi\rho)(1 - \pi + \pi\rho)]X_{t-1} \\ &\quad + [\rho(1 - \pi + \pi\rho - \rho) - \rho(\pi - \pi\rho)]X_{t-1} + n\beta(1 - \beta) = \\ &= [\rho(1 - \pi + \pi\rho - \rho) - \rho(\pi - \pi\rho)]X_{t-1} + n\beta(1 - \beta) = \\ &= [\rho(1 - 2\pi + 2\pi\rho - \rho)]X_{t-1} + n\beta(1 - \beta) = \\ &= \rho(1 - \rho)(1 - 2\pi)X_{t-1} + n\beta(1 - \beta), \end{aligned}$$

kde druhá rovnost platí díky nezávislosti obou operací \circ .

□

2. Odhady parametrů

V této sekci ukážeme několik způsobů odhadu parametrů (ρ, π) . Všechny zmíněné odhady jsou obsaženy v (Weiß a Kim, 2013a) a (Weiß a Kim, 2013b), odkud přebíráme výsledky. Zmíněné články obsahují také náznaky a myšlenky důkazů a odvození, které zde rozvádíme podrobněji. Ukážeme ale nejprve, že binomický AR(1) proces splňuje zákon velkých čísel.

Tvrzení 10. *Bud' $(X_t, t = 1, 2, \dots)$ stacionární binomický AR(1) proces. Potom pro $f : \{1, \dots, n\}^2 \rightarrow \mathbb{R}$ měřitelnou splňující $E |f(X_t, X_{t+1})| < \infty$ platí*

$$\frac{1}{T} \sum_{t=1}^T f(X_t, X_{t+1}) \xrightarrow[T \rightarrow \infty]{s.j.} E f(X_t, X_{t+1}).$$

Důkaz. Ve tvrzeních 5 a 6 jsme ukázali, že stacionární binomický AR(1) proces je homogenní a ergodický Markovův řetězec. Tím jsou splněny předpoklady Tvrzení 1.3 v (Billingsley, 1961), odkud již plyne tvrzení. □

Z výše uvedeného tvrzení plyne kromě konvergence výběrového průměru také konvergence výrazů

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T X_t X_{t+1} &\xrightarrow[T \rightarrow \infty]{s.j.} E X_t X_{t+1}, \\ \frac{1}{T} \sum_{t=1}^T s^{X_t} &\xrightarrow[T \rightarrow \infty]{s.j.} E s^{X_t}, \end{aligned}$$

které budeme později využívat.

2.1 Yuleova-Walkerova metoda

První metoda vychází z interpretace parametrů (ρ, π) , které mají význam střední hodnoty (po vynásobení konstantou n) a autokorelace prvního řádu. Můžeme je tak přímo odhadnout jako výběrový průměr respektive výběrovou autokorelaci prvního řádu:

$$\begin{aligned} \hat{\pi}_{YW} &:= \frac{1}{nT} \sum_{t=1}^T X_t = \frac{1}{n} \bar{X}_T, \\ \hat{\rho}_{YW} &:= \frac{\sum_{t=2}^T [(X_t - \bar{X}_T)(X_{t-1} - \bar{X}_T)]}{\sum_{t=1}^T (X_t - \bar{X}_T)^2}. \end{aligned}$$

Odhad $\hat{\pi}_{YW}$ je zřejmě nestranný, díky tvrzení 10 je konzistentní a můžeme navíc snadno spočítat jeho rozptyl.

Tvrzení 11 (Rozptyl $\hat{\pi}_{YW}$). *Bud' $n \in \mathbb{N}$ a necht' binomický AR(1) je stacionární. Potom rozptyl $\hat{\pi}_{YW}$ je*

$$\text{var}(\hat{\pi}_{YW}) = \frac{\pi(1-\pi)}{nT} \frac{1+\rho}{1-\rho} \left(1 + \frac{2\rho}{1-\rho^2} \frac{(\rho^T - 1)}{T} \right).$$

Důkaz.

$$\begin{aligned}
\text{var}(\hat{\pi}_{YW}) &= \text{var}\left(\frac{1}{nT} \sum_{t=1}^T X_t\right) = \frac{1}{(nT)^2} \text{var}\left(\sum_{t=1}^T X_t\right) \\
&= \frac{1}{(nT)^2} \left(\sum_{t=1}^T \text{var}(X_t) + \sum_{i \neq j} \text{cov}(X_i, X_j) \right) \\
&= \frac{1}{(nT)^2} \left(Tn\pi(1-\pi) + 2 \sum_{k=1}^{T-1} \sum_{l=k+1}^T \text{cov}(X_l, X_{l-k}) \right).
\end{aligned}$$

Kovarianci $\text{cov}(X_l, X_{l-k})$ můžeme vyjádřit jako $\rho(k) \text{var}(X_0) = \rho^k n\pi(1-\pi)$ a druhý výraz je tedy roven $2n\pi(1-\pi) \sum_{k=1}^{T-1} \rho^k (T-k)$. Sečteme nyní tuto sumu.

$$\begin{aligned}
\sum_{k=1}^{T-1} \rho^k (T-k) &= T \sum_{k=1}^{T-1} \rho^k - \sum_{k=1}^{T-1} k \rho^k \\
&= T \sum_{k=0}^{T-1} \rho^k - T - \sum_{k=0}^{T-1} k \rho^k \\
&= T \frac{1-\rho^T}{1-\rho} - T - \rho \sum_{k=0}^{T-1} k \rho^{k-1} \\
&= T \frac{1-\rho^T}{1-\rho} - T - \rho \left(\sum_{k=0}^{T-1} \rho^k \right)' \\
&= T \frac{1-\rho^T}{1-\rho} - T - \rho \left(\frac{1-\rho^T}{1-\rho} \right)' \\
&= T \frac{1-\rho^T}{1-\rho} - T - \rho \frac{-T\rho^{T-1}(1-\rho) + 1-\rho^T}{(1-\rho)^2} \\
&= \frac{T(1-\rho^T)(1-\rho) - T(1-\rho)^2 + T\rho^T(1-\rho) - \rho(1-\rho^T)}{(1-\rho)^2} \\
&= \frac{-T\rho + 2T\rho - T\rho^{T+1} - \rho + \rho^{T+1}}{(1-\rho)^2} = \frac{T\rho(1-\rho) + \rho(\rho^T - 1)}{(1-\rho)^2}.
\end{aligned}$$

Po dosazení máme

$$\begin{aligned}
\text{var}(\hat{\pi}_{YW}) &= \frac{1}{(nT)^2} \left(Tn\pi(1-\pi) + 2 \sum_{k=1}^{T-1} \sum_{l=k+1}^T \text{cov}(X_l, X_{l-k}) \right) \\
&= \frac{\pi(1-\pi)}{nT^2} \left(T + \frac{2T\rho(1-\rho) + 2\rho(\rho^T - 1)}{(1-\rho)^2} \right) \\
&= \frac{\pi(1-\pi)}{nT} \left(1 + \frac{2\rho}{1-\rho} + \frac{2\rho(\rho^T - 1)}{T(1-\rho)^2} \right) \\
&= \frac{\pi(1-\pi)}{nT} \left(\frac{1+\rho}{1-\rho} + \frac{2\rho}{(1-\rho)^2} \frac{(\rho^T - 1)}{T} \right) \\
&= \frac{\pi(1-\pi)}{nT} \frac{1+\rho}{1-\rho} \left(1 + \frac{2\rho}{1-\rho^2} \frac{(\rho^T - 1)}{T} \right).
\end{aligned}$$

□

Odhad $\hat{\rho}$ již není nestranný, ale díky tvrzení 10 je konzistentní. Přímé vyjádření jeho rozptylu je mírně složitější a nebudeme ho zde ukazovat. Jeho asymptotické vlastnosti dostaneme později, definujeme nejprve odhady metodou podmíněných nejmenších čtverců a ve tvrzení 13 ukážeme, že obě metody jsou asymptoticky ekvivalentní.

2.2 Metoda podmíněných nejmenších čtverců

Druhá metoda vychází z tvrzení 9, konkrétně z rovnosti

$$\mathbb{E}[X_t|X_{t-1}] = \rho X_{t-1} + n\beta.$$

Myšlenkou je porovnávat pozorované hodnoty X_t s jejich očekávanou hodnotou na základě pozorování z času $t-1$ za platnosti binomického AR(1) modelu s parametry (ρ, π) . Jako odhad $(\hat{\rho}_{CLS}, \hat{\pi}_{CLS})$ zvolíme (ρ, π) z přípustné množiny parametrů, které minimalizují ztrátovou funkci

$$M(\rho, \pi) := \sum_{t=2}^T (X_t - \mathbb{E}[X_t|X_{t-1}])^2 = \sum_{t=2}^T (X_t - \rho X_{t-1} - n\pi + n\pi\rho)^2.$$

Pro tuto úlohu dokážeme najít explicitní řešení. Minimum budeme hledat položením parciálních derivací ztrátové funkce rovno nule.

$$\begin{aligned} \frac{du}{d\pi} M(\rho, \pi) &= -2n(1-\rho) \sum_{t=2}^T (X_t - \rho X_{t-1} - n\pi + n\pi\rho) \\ \frac{du}{d\rho} M(\rho, \pi) &= 2 \sum_{t=2}^T [(X_t - \rho X_{t-1} - n\pi + n\pi\rho)(-X_{t-1} + n\pi)] \\ &= -2 \sum_{t=2}^T [(X_t - \rho X_{t-1} - n\pi + n\pi\rho)X_{t-1}] - \frac{\pi}{1-\rho} \frac{du}{d\pi} M(\rho, \pi) \end{aligned}$$

Položme tedy $\frac{du}{d\pi} M(\rho, \pi) = \frac{du}{d\rho} M(\rho, \pi) = 0$ a dostaneme soustavu

$$\begin{aligned} 0 &= -2n(1-\rho) \sum_{t=2}^T (X_t - \rho X_{t-1} - n\pi + n\pi\rho) \\ 0 &= -2 \sum_{t=2}^T [(X_t - \rho X_{t-1} - n\pi + n\pi\rho)X_{t-1}], \end{aligned}$$

kterou můžeme dále upravit do tvaru

$$0 = \sum_{t=2}^T (X_t - \rho X_{t-1} - n\pi + n\pi\rho) = \sum_{t=2}^T X_t - \rho \sum_{t=2}^T X_{t-1} - (T-1)n\pi(1-\rho) \quad (2.1)$$

$$\begin{aligned} 0 &= \sum_{t=2}^T [(X_t - \rho X_{t-1} - n\pi + n\pi\rho)X_{t-1}] \\ &= \sum_{t=2}^T X_t X_{t-1} - \rho \sum_{t=2}^T X_{t-1}^2 - n\pi(1-\rho) \sum_{t=2}^T X_{t-1}. \end{aligned} \quad (2.2)$$

Nyní odečtením $(\sum_{t=2}^T X_{t-1})$ -násobku rovnosti (2.1) od $(T-1)$ -násobku rovnosti (2.2) dostaneme rovnost, ve které se vyskytuje jediný parametr ρ :

$$0 = (T-1) \sum_{t=2}^T X_t X_{t-1} - \rho (T-1) \sum_{t=2}^T X_{t-1}^2 - \left(\sum_{t=2}^T X_t \right) \left(\sum_{t=2}^T X_{t-1} \right) + \rho \left(\sum_{t=2}^T X_{t-1} \right)^2,$$

ze které vyjádříme odhad parametru ρ jako

$$\hat{\rho}_{CLS} = \frac{(T-1) \sum_{t=2}^T X_t X_{t-1} - \left(\sum_{t=2}^T X_t \right) \left(\sum_{t=2}^T X_{t-1} \right)}{(T-1) \sum_{t=2}^T X_{t-1}^2 - \left(\sum_{t=2}^T X_{t-1} \right)^2}.$$

Z rovnosti (2.1) pak můžeme vyjádřit odhad parametru π pomocí $\hat{\rho}_{CLS}$ jako

$$\hat{\pi}_{CLS} = \frac{\sum_{t=2}^T X_t - \hat{\rho}_{CLS} \sum_{t=2}^T X_{t-1}}{n(T-1)(1 - \hat{\rho}_{CLS})}.$$

Tvrzení 12. *Ve stacionárním binomickém AR(1) modelu je odhad $(\hat{\rho}_{CLS}, \hat{\pi}_{CLS})$ metodou nejmenších čtverců konzistentní a platí*

$$\sqrt{T} \begin{pmatrix} \hat{\rho}_{CLS} - \rho \\ \hat{\pi}_{CLS} - \pi \end{pmatrix} \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}(\rho, \pi)),$$

kde

$$\mathbf{\Sigma}(\rho, \pi) = n^{-1} \begin{pmatrix} \frac{(1-2\pi)^2}{\pi(1-\pi)} \rho(1-\rho) + n(1-\rho^2) & (1-2\pi)\rho \\ (1-2\pi)\rho & \pi(1-\pi) \frac{1+\rho}{1-\rho} \end{pmatrix} \quad (2.3)$$

Důkaz. Důkaz je naznačen v (Weiß a Kim, 2013b) s využitím obecných asymptotických výsledků o stacionárních ergodických procesech ze sekce 3 v (Klimko a Nelson, 1978), zde jej provedeme podrobněji s rozepsáním jednotlivých kroků. Označme v souladu se značením použitým v citované literatuře

$$g_{t-1}(\rho, \pi) := E[X_t | X_{t-1}], \\ u_t(\rho, \pi) := X_t - g_{t-1} = X_t - E[X_t | X_{t-1}].$$

Parciální derivace funkce $g_{t-1}(\rho, \pi)$

$$\frac{\partial g_{t-1}}{\partial \rho}(\rho, \pi) = X_{t-1} - n\pi = X_{t-1} - E X_{t-1}, \\ \frac{\partial g_{t-1}}{\partial \pi}(\rho, \pi) = n(1-\rho)$$

jsou polynomy (navíc vždy jen v jedné proměnné), tudíž g_{t-1} má zřejmě spojitě parciální derivace až do řádu 3. Vzhledem k tomu, že X_t může nabývat jen konečný počet $n+1$ různých hodnot, je splněna podmínka pro kvadratickou integrovatelnost jednak X_t , ale také parciálních derivací g_{t-1} až do řádu 3. Tímto jsme ověřili, že g_{t-1} splňuje uvedené podmínky regularity a podle tvrzení 3.1 v (Klimko a Nelson, 1978) je $(\hat{\rho}_{CLS}, \hat{\pi}_{CLS})$ konzistentním odhadem (ρ, π) .

Označme dále matici

$$\begin{aligned} \mathbf{V} &= \begin{pmatrix} \mathbb{E} \left[\frac{\partial g_{t-1}}{\partial \rho}(\rho, \pi) \cdot \frac{\partial g_{t-1}}{\partial \rho}(\rho, \pi) \right] & \mathbb{E} \left[\frac{\partial g_{t-1}}{\partial \rho}(\rho, \pi) \cdot \frac{\partial g_{t-1}}{\partial \pi}(\rho, \pi) \right] \\ \mathbb{E} \left[\frac{\partial g_{t-1}}{\partial \pi}(\rho, \pi) \cdot \frac{\partial g_{t-1}}{\partial \rho}(\rho, \pi) \right] & \mathbb{E} \left[\frac{\partial g_{t-1}}{\partial \pi}(\rho, \pi) \cdot \frac{\partial g_{t-1}}{\partial \pi}(\rho, \pi) \right] \end{pmatrix} \\ &= \begin{pmatrix} \mathbb{E} (X_{t-1} - \mathbb{E} X_{t-1})^2 & n(1-\rho) \mathbb{E} [X_{t-1} - \mathbb{E} X_{t-1}] \\ n(1-\rho) \mathbb{E} [X_{t-1} - \mathbb{E} X_{t-1}] & n^2(1-\rho)^2 \end{pmatrix} \\ &= \begin{pmatrix} n\pi(1-\pi) & 0 \\ 0 & n^2(1-\rho)^2 \end{pmatrix}, \end{aligned}$$

kde jsme využili vztahů $\mathbb{E} [X_{t-1} - \mathbb{E} X_{t-1}] = 0$ a

$$\mathbb{E} (X_{t-1} - \mathbb{E} X_{t-1})^2 = \text{var} X_{t-1} = n\pi(1-\pi).$$

Matice \mathbf{V} je diagonální s kladnými prvky na diagonále, takže je pozitivně definitní. Označme dále matici

$$\mathbf{W} = \begin{pmatrix} \mathbb{E} \left(\frac{\partial g_{t-1}}{\partial \rho} \cdot \frac{\partial g_{t-1}}{\partial \rho} \cdot u_t^2 \right) & \mathbb{E} \left(\frac{\partial g_{t-1}}{\partial \rho} \cdot \frac{\partial g_{t-1}}{\partial \pi} \cdot u_t^2 \right) \\ \mathbb{E} \left(\frac{\partial g_{t-1}}{\partial \pi} \cdot \frac{\partial g_{t-1}}{\partial \rho} \cdot u_t^2 \right) & \mathbb{E} \left(\frac{\partial g_{t-1}}{\partial \pi} \cdot \frac{\partial g_{t-1}}{\partial \pi} \cdot u_t^2 \right) \end{pmatrix}$$

s prvky $\mathbf{W} = (w_{ij})$, $i, j = 1, 2$, které vypočítáme postupně. Platí

$$\begin{aligned} w_{22} &= n^2(1-\rho)^2 \cdot \mathbb{E} (X_t - E[X_t|X_{t-1}])^2 \\ &= n^2(1-\rho)^2 \cdot \mathbb{E} \mathbb{E} [(X_t - E[X_t|X_{t-1}])^2 | X_{t-1}] \\ &= n^2(1-\rho)^2 \cdot \mathbb{E} \text{var}(X_t|X_{t-1}) \\ &= n^2(1-\rho)^2 \cdot \mathbb{E} [\rho(1-\rho)(1-2\pi)X_{t-1} + n\beta(1-\beta)] \\ &= n^2(1-\rho)^2 \cdot [\rho(1-\rho)(1-2\pi)n\pi + n\beta(1-\beta)] \\ &= n^3(1-\rho)^2\pi(1-\pi)(1-\rho^2), \end{aligned}$$

kde jsme pro vyjádření podmíněného rozptylu využili tvrzení 9. Dále platí

$$\begin{aligned} w_{12} &= n(1-\rho) \cdot \mathbb{E} \left\{ (X_{t-1} - \mathbb{E} X_{t-1}) \cdot (X_t - E[X_t|X_{t-1}])^2 \right\} \\ &= n(1-\rho) \cdot \mathbb{E} \left\{ \mathbb{E} [(X_{t-1} - \mathbb{E} X_{t-1}) \cdot (X_t - E[X_t|X_{t-1}])^2 | X_{t-1}] \right\} \\ &= n(1-\rho) \cdot \mathbb{E} \left\{ (X_{t-1} - \mathbb{E} X_{t-1}) \mathbb{E} [(X_t - E[X_t|X_{t-1}])^2 | X_{t-1}] \right\} \\ &= n(1-\rho) \cdot \mathbb{E} \left\{ (X_{t-1} - \mathbb{E} X_{t-1}) \cdot \text{var}(X_t|X_{t-1}) \right\} \\ &= n(1-\rho) \cdot \mathbb{E} \left\{ (X_{t-1} - \mathbb{E} X_{t-1}) \cdot [\rho(1-\rho)(1-2\pi)X_{t-1} + n\beta(1-\beta)] \right\} \\ &= n(1-\rho) \cdot \mathbb{E} \left\{ (X_{t-1} - \mathbb{E} X_{t-1}) \cdot \rho(1-\rho)(1-2\pi)X_{t-1} \right\} \\ &\quad + n(1-\rho) \cdot \mathbb{E} \left\{ (X_{t-1} - \mathbb{E} X_{t-1}) \cdot n\beta(1-\beta) \right\}. \end{aligned}$$

Druhý člen tohoto výrazu je nulový a první rozepíšeme přičtením a odečtením výrazu $\mathbb{E} X_{t-1}$ opět na dva členy, z nichž jeden bude nulový a druhý dokážeme snadno upravit do výsledného tvaru:

$$\begin{aligned} w_{12} &= n(1-\rho) \cdot \mathbb{E} \left\{ (X_{t-1} - \mathbb{E} X_{t-1}) \cdot \rho(1-\rho)(1-2\pi)(X_{t-1} - \mathbb{E} X_{t-1}) \right\} \\ &\quad + n(1-\rho) \cdot \mathbb{E} \left\{ (X_{t-1} - \mathbb{E} X_{t-1}) \cdot \rho(1-\rho)(1-2\pi) \mathbb{E} X_{t-1} \right\} \\ &= n(1-\rho) \cdot \mathbb{E} \left\{ (X_{t-1} - \mathbb{E} X_{t-1})^2 \cdot \rho(1-\rho)(1-2\pi) \right\} + 0 \\ &= n(1-\rho) \cdot \mathbb{E} \left\{ n\pi(1-\pi) \cdot \rho(1-\rho)(1-2\pi) \right\} \\ &= n(1-\rho) \cdot \mathbb{E} \left\{ \text{var}(X_{t-1}) \cdot \rho(1-\rho)(1-2\pi) \right\} \\ &= n^2(1-\rho)^2\rho(1-2\pi)\pi(1-\pi). \end{aligned}$$

Konečně zbývající prvek w_{11} spočítáme obdobně:

$$\begin{aligned}
w_{11} &= \mathbf{E} \left\{ (X_{t-1} - \mathbf{E} X_{t-1})^2 \cdot (X_t - E[X_t|X_{t-1}])^2 \right\} \\
&= \mathbf{E} \left\{ \mathbf{E} \left[(X_{t-1} - \mathbf{E} X_{t-1})^2 \cdot (X_t - E[X_t|X_{t-1}])^2 \mid X_{t-1} \right] \right\} \\
&= \mathbf{E} \left\{ (X_{t-1} - \mathbf{E} X_{t-1})^2 \cdot \mathbf{E} \left[(X_t - E[X_t|X_{t-1}])^2 \mid X_{t-1} \right] \right\} \\
&= \mathbf{E} \left\{ (X_{t-1} - \mathbf{E} X_{t-1})^2 \cdot \text{var}(X_t|X_{t-1}) \right\}.
\end{aligned}$$

Podmíněný rozptyl $\text{var}(X_t|X_{t-1})$ nyní rozepíšeme obdobně jako při výpočtu w_{12} pomocí přičtení a odečtení výrazu $\mathbf{E} X_{t-1}$ a dostáváme

$$\begin{aligned}
w_{11} &= \mathbf{E} \left\{ (X_{t-1} - \mathbf{E} X_{t-1})^2 \cdot \rho(1 - \rho)(1 - 2\pi)X_{t-1} \right\} \\
&\quad + \mathbf{E} \left\{ (X_{t-1} - \mathbf{E} X_{t-1})^2 \cdot n\beta(1 - \beta) \right\} \\
&= \mathbf{E} \left\{ (X_{t-1} - \mathbf{E} X_{t-1})^3 \cdot \rho(1 - \rho)(1 - 2\pi) \right\} \\
&\quad + \mathbf{E} \left\{ (X_{t-1} - \mathbf{E} X_{t-1})^2 \cdot \rho(1 - \rho)(1 - 2\pi) \mathbf{E} X_{t-1} \right\} \\
&\quad + \mathbf{E} \left\{ (X_{t-1} - \mathbf{E} X_{t-1})^2 \cdot n\beta(1 - \beta) \right\} \\
&= n\pi(1 - \pi)(1 - 2\pi) \cdot \rho(1 - \rho)(1 - 2\pi) \\
&\quad + n\pi(1 - \pi) \cdot \rho(1 - \rho)(1 - 2\pi)n\pi \\
&\quad + n\pi(1 - \pi) \cdot n\pi(1 - \rho)(1 - \pi + \pi\rho) \\
&= n\pi(1 - \pi)(1 - \rho) \left[(1 - 2\pi)^2\rho + \rho(1 - 2\pi)n\pi + n\pi(1 - \pi + \pi\rho) \right] \\
&= n\pi(1 - \pi)(1 - \rho) \left[(1 - 2\pi)^2\rho + n\pi(1 - \pi)(1 + \rho) \right].
\end{aligned}$$

Náhodné veličiny X_t a g_{t-1} mají konečné čtvrté momenty, a tím už jsou splněny všechny podmínky ve větě 3.2 v (Klimko a Nelson, 1978), ze které platí

$$\sqrt{T} \begin{pmatrix} \hat{\rho}_{CLS} - \rho \\ \hat{\pi}_{CLS} - \pi \end{pmatrix} \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \mathbf{V}^{-1} \mathbf{W} \mathbf{V}^{-1}).$$

Matice $\mathbf{V} = (v_{ij})$, $i, j = 1, 2$ je diagonální a snadno tak určíme její inverzi. Matice ve vyjádření asymptotického rozptylu je tak rovna

$$\begin{aligned}
\mathbf{V}^{-1} \mathbf{W} \mathbf{V}^{-1} &= \begin{pmatrix} (v_{11}^{-1})^2 w_{11} & v_{11}^{-1} v_{22}^{-1} w_{12} \\ v_{11}^{-1} v_{22}^{-1} w_{12} & (v_{22}^{-1})^2 w_{22} \end{pmatrix} \\
&= n^{-1} \begin{pmatrix} \frac{(1-2\pi)^2}{\pi(1-\pi)} \rho(1-\rho) + n(1-\rho^2) & (1-2\pi)\rho \\ (1-2\pi)\rho & \pi(1-\pi) \frac{1+\rho}{1-\rho} \end{pmatrix},
\end{aligned}$$

což jsme chtěli dokázat. □

V následujícím tvrzení ukážeme, že Yuleovy-Walkerovy odhady a odhady metodou nejmenších čtverců mají stejné asymptotické rozdělení.

Tvrzení 13. *Pro odhady parametrů ve stacionárním binomickém AR(1) modelu platí*

$$\sqrt{T} \begin{pmatrix} \hat{\rho}_{CLS} - \hat{\rho}_{YW} \\ \hat{\pi}_{CLS} - \hat{\pi}_{YW} \end{pmatrix} \xrightarrow[T \rightarrow \infty]{p} \mathbf{0}.$$

Důkaz. Začneme s odhady parametru π . Rozepíšeme si pro přehlednost nejprve oba odhady do vhodného tvaru, se kterým budeme dále pracovat:

$$\begin{aligned}\sqrt{T}\hat{\pi}_{CLS} &= \frac{1}{n(1-\hat{\rho}_{CLS})} \left[\frac{\sqrt{T}}{T-1} \sum_{t=2}^T X_t - \frac{\sqrt{T}}{T-1} \hat{\rho}_{CLS} \sum_{t=1}^{T-1} X_t \right] \\ &= \frac{1}{n(1-\hat{\rho}_{CLS})} \left[\frac{\sqrt{T}}{T-1} \sum_{t=1}^T X_t - \frac{\sqrt{T}}{T-1} \hat{\rho}_{CLS} \sum_{t=1}^T X_t \right. \\ &\quad \left. - \frac{\sqrt{T}}{T-1} X_1 + \frac{\sqrt{T}}{T-1} \hat{\rho}_{CLS} X_T \right],\end{aligned}$$

$$\begin{aligned}\sqrt{T}\hat{\pi}_{YW} &= \sqrt{T} \frac{1}{nT} \sum_{t=1}^T X_t \\ &= \frac{1}{n(1-\hat{\rho}_{CLS})} \left[\frac{1}{\sqrt{T}} (1-\hat{\rho}_{CLS}) \sum_{t=1}^T X_t \right] \\ &= \frac{1}{n(1-\hat{\rho}_{CLS})} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T X_t - \frac{1}{\sqrt{T}} \hat{\rho}_{CLS} \sum_{t=1}^T X_t \right].\end{aligned}$$

Odečtením obou výrazů od sebe dostáváme

$$\begin{aligned}\sqrt{T}(\hat{\pi}_{CLS} - \hat{\pi}_{YW}) &= \\ &= \frac{1}{n(1-\hat{\rho}_{CLS})} \left[\sum_{t=1}^T X_t \left(\frac{\sqrt{T}}{T-1} - \frac{1}{\sqrt{T}} \right) - \hat{\rho}_{CLS} \sum_{t=1}^T X_t \left(\frac{\sqrt{T}}{T-1} - \frac{1}{\sqrt{T}} \right) \right. \\ &\quad \left. - \frac{\sqrt{T}}{T-1} X_1 + \frac{\sqrt{T}}{T-1} \hat{\rho}_{CLS} X_T \right] \\ &= \frac{1}{n(1-\hat{\rho}_{CLS})} \left[\sum_{t=1}^T X_t \left(\frac{1}{\sqrt{T}(T-1)} \right) - \hat{\rho}_{CLS} \sum_{t=1}^T X_t \left(\frac{1}{\sqrt{T}(T-1)} \right) \right. \\ &\quad \left. - \frac{\sqrt{T}}{T-1} X_1 + \frac{\sqrt{T}}{T-1} \hat{\rho}_{CLS} X_T \right] \\ &= \frac{1}{\sqrt{T}} \frac{1}{n(1-\hat{\rho}_{CLS})} \left[\frac{\sum_{t=1}^T X_t}{T-1} - \hat{\rho}_{CLS} \frac{\sum_{t=1}^T X_t}{T-1} - X_1 \frac{T}{T-1} + \hat{\rho}_{CLS} X_T \frac{T}{T-1} \right].\end{aligned}$$

Tento výraz konverguje v pravděpodobnosti k nule, neboť $\hat{\rho}_{CLS}$ je konzistentní odhad, ostatní členy v hranaté závorce jsou omezené v pravděpodobnosti a výraz $1/\sqrt{T} \rightarrow 0$. Obdobně vyšetříme i odhady parametru ρ , které rovněž nejprve rozepíšeme do vhodného tvaru.

$$\begin{aligned}
\sqrt{T}\hat{\rho}_{CLS} &= \sqrt{T} \frac{(T-1) \sum_{t=2}^T X_t X_{t-1} - \left(\sum_{t=2}^T X_t\right) \left(\sum_{t=2}^T X_{t-1}\right)}{(T-1) \sum_{t=2}^T X_{t-1}^2 - \left(\sum_{t=2}^T X_{t-1}\right)^2} \\
&= \frac{1}{\sqrt{T}} \frac{\sum_{t=2}^T X_t X_{t-1} - \frac{1}{T-1} \left(\sum_{t=2}^T X_t\right) \left(\sum_{t=2}^T X_{t-1}\right)}{\frac{1}{T} \left[\sum_{t=2}^T X_{t-1}^2 - \frac{1}{T-1} \left(\sum_{t=2}^T X_{t-1}\right)^2\right]} \\
&=: \frac{1}{\sqrt{T} D_{CLS}} \left[\sum_{t=2}^T X_t X_{t-1} - \frac{\left(\sum_{t=2}^T X_t\right) \left(\sum_{t=2}^T X_{t-1}\right)}{T-1} \right] \\
&= \frac{1}{\sqrt{T} D_{CLS}} \left[\sum_{t=2}^T X_t X_{t-1} - \frac{\left(\sum_{t=1}^T X_t\right) \left(\sum_{t=2}^T X_{t-1}\right)}{T-1} + \frac{X_1 \left(\sum_{t=2}^T X_{t-1}\right)}{T-1} \right] \\
&= \frac{1}{\sqrt{T} D_{CLS}} \left[\sum_{t=2}^T X_t X_{t-1} - \frac{\left(\sum_{t=1}^T X_t\right) \left(\sum_{t=1}^T X_t\right)}{T-1} \right. \\
&\quad \left. + \frac{X_1 \left(\sum_{t=2}^T X_{t-1}\right)}{T-1} + \frac{X_T \left(\sum_{t=1}^T X_t\right)}{T-1} \right] \\
&= \frac{1}{\sqrt{T} D_{CLS}} \left[\sum_{t=2}^T X_t X_{t-1} - \frac{T^2}{T-1} (\bar{X}_T)^2 \right] + o_p(1),
\end{aligned}$$

kde jsme označili jmenovatele odhadu symbolem D_{CLS} a symbolem $o_p(1)$ zkráceně zapisujeme náhodnou veličinu konvergující v pravděpodobnosti k nule. Platí, že D_{CLS} konverguje v pravděpodobnosti k rozptylu procesu $\text{var } X_0$. Obdobně rozepíšeme i odhad $\hat{\rho}_{YW}$ a analogicky označíme jmenovatele symbolem D_{YW} , který rovněž konverguje v pravděpodobnosti k rozptylu procesu $\text{var } X_0$.

$$\begin{aligned}
\sqrt{T}\hat{\rho}_{YW} &= \sqrt{T} \frac{\sum_{t=2}^T \left[(X_t - \bar{X}_T)(X_{t-1} - \bar{X}_T) \right]}{\sum_{t=1}^T (X_t - \bar{X}_T)^2} \\
&= \frac{1}{\sqrt{T}} \frac{\sum_{t=2}^T X_t X_{t-1} - \bar{X}_T \sum_{t=2}^T X_t - \bar{X}_T \sum_{t=2}^T X_{t-1} + (T-1)(\bar{X}_T)^2}{\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}_T)^2} \\
&=: \frac{1}{\sqrt{T} D_{YW}} \left[\sum_{t=2}^T X_t X_{t-1} - \bar{X}_T \sum_{t=2}^T X_t - \bar{X}_T \sum_{t=2}^T X_{t-1} + (T-1)(\bar{X}_T)^2 \right] \\
&= \frac{1}{\sqrt{T} D_{YW}} \left[\sum_{t=2}^T X_t X_{t-1} - \left[T(\bar{X}_T)^2 - \bar{X}_T X_1 \right] \right. \\
&\quad \left. - \left[T(\bar{X}_T)^2 - X_T \bar{X}_T \right] + (T-1)(\bar{X}_T)^2 \right] \\
&= \frac{1}{\sqrt{T} D_{YW}} \left[\sum_{t=2}^T X_t X_{t-1} - T(\bar{X}_T)^2 + \bar{X}_T X_1 + X_T \bar{X}_T - (\bar{X}_T)^2 \right] \\
&= \frac{1}{\sqrt{T} D_{YW}} \left[\sum_{t=2}^T X_t X_{t-1} - T(\bar{X}_T)^2 \right] + o_p(1).
\end{aligned}$$

Odečtením obou výrazů od sebe dostáváme

$$\begin{aligned} \sqrt{T}(\hat{\rho}_{CLS} - \hat{\rho}_{YW}) &= \\ &= \frac{D_{YW} - D_{CLS}}{D_{YW}D_{CLS}} \frac{\sum_{t=2}^T X_t X_{t-1}}{\sqrt{T}} - \frac{\frac{T}{T-1} D_{YW} - D_{CLS}}{D_{YW}D_{CLS}} \sqrt{T} \bar{X}_T + o_p(1) \\ &= o_p(1)O_p(1) - o_p(1)O_p(1) + o_p(1) = o_p(1), \end{aligned}$$

kde $O_p(1)$ značíme náhodnou veličinu omezenou v pravděpodobnosti, což plyne z věty 7 v (Jones, 2004), která je obdobou centrální limitní věty pro Markovovy řetězce a jejíž předpoklady jsou pro binomický AR(1) proces splněny. Dohromady potom

$$\sqrt{T} \begin{pmatrix} \hat{\rho}_{CLS} - \rho \\ \hat{\pi}_{CLS} - \pi \end{pmatrix} - \sqrt{T} \begin{pmatrix} \hat{\rho}_{YW} - \rho \\ \hat{\pi}_{YW} - \pi \end{pmatrix} = \sqrt{T} \begin{pmatrix} \hat{\rho}_{CLS} - \hat{\rho}_{YW} \\ \hat{\pi}_{CLS} - \hat{\pi}_{YW} \end{pmatrix} \xrightarrow[T \rightarrow \infty]{p} \mathbf{0},$$

z čehož už plyne, že oba odhady mají stejné asymptotické rozdělení. □

2.3 Metoda maximální věrohodnosti

Dalším z možných postupů odhadu parametrů je přes maximalizování věrohodnosti jako funkce parametrů. Ta je definována jako sdružená pravděpodobnost, že při daných parametrech napozorujeme zadaná data tj. hodnoty $X_1 = x_1, \dots, X_T = x_T$. Tuto pravděpodobnost můžeme rozložit na pravděpodobnost výskytu v počátečním stavu a následně součin pravděpodobností přechodu mezi jednotlivými stavy (využili jsme homogenity procesu). Označme tedy

$$\begin{aligned} L(\rho, \pi) &:= \mathbf{P}(\mathbf{X} = \mathbf{x}; \rho, \pi) = \mathbf{P}(X_1 = x_1; \rho, \pi) \prod_{t=2}^T p(x_t | x_{t-1}; \rho, \pi) \\ &= \binom{n}{x_1} \pi^{x_1} (1 - \pi)^{n-x_1} \prod_{t=2}^T \sum_{m=\max(0, x_{t-1}+x_t-n)}^{\min(x_{t-1}, x_t)} \binom{x_{t-1}}{m} \binom{n-x_{t-1}}{x_t-m} \\ &\quad \cdot \alpha^m (1 - \alpha)^{x_{t-1}-m} \beta^{x_t-m} (1 - \beta)^{n-x_{t-1}+m-x_t}, \end{aligned}$$

kde tučným symbolem \mathbf{x} značíme vektor všech pozorovaných hodnot náhodných veličin $\mathbf{X} = (X_1, \dots, X_T)$. Za odhad parametrů (ρ, π) potom zvolíme hodnoty maximalizující tuto věrohodnostní funkci s pozorovanými hodnotami procesu. Takový odhad již nedokážeme explicitně vyjádřit a musíme jej hledat numericky.

Pro zjednodušení můžeme dále použít podmíněnou věrohodnost, tedy vyjádřit si pravděpodobnost napozorování hodnot procesu podmíněně počáteční hodnotou

procesu. Označme tuto podmíněnou věrohodnost následovně:

$$L^{(C)}(\rho, \pi) := \mathbf{P}(\mathbf{X} = \mathbf{x} | X_1 = x_1; \rho, \pi) = \prod_{t=2}^T p(x_t | x_{t-1}; \rho, \pi)$$

$$l^{(C)}(\rho, \pi) := \log L^{(C)}(\rho, \pi) = \sum_{t=2}^T \log p(x_t | x_{t-1}; \rho, \pi) =$$

$$= \sum_{t=2}^T \log \left[\sum_{m=\max(0, x_{t-1}+x_t-n)}^{\min(x_{t-1}, x_t)} \binom{x_{t-1}}{m} \binom{n-x_{t-1}}{x_t-m} \cdot \alpha^m (1-\alpha)^{x_{t-1}-m} \beta^{x_t-m} (1-\beta)^{n-x_{t-1}+m-x_t} \right].$$

Jako maximálně věrohodný odhad parametrů (ρ, π) budeme uvažovat hodnoty maximalizující podmíněnou logaritmickou věrohodnost, označme

$$(\hat{\rho}_{ML}, \hat{\pi}_{ML}) := \arg \max_{(\rho, \pi) \in P} l^{(C)}(\rho, \pi),$$

kde P značí přípustnou oblast pro parametry (ρ, π) podle definice 2.

Ukáže se, že i v případě korelovaných pozorování platí vlastnosti a asymptotika maximálně věrohodných odhadů při splnění jistých podmínek regularity obdobně jako při nezávislých pozorováních. Zformulujeme nyní obecnější tvrzení pro vlastnosti maximálně věrohodných odhadů pro Markovovy řetězce s diskretním časem a konečně mnoha stavy.

Tvrzení 14. *Bud' $(X_t, t = 1, 2, \dots)$ Markovův řetězec s konečně mnoha stavy, jehož pravděpodobnosti přechodu závisejí na K -rozměrném parametru $\boldsymbol{\theta} \in \Theta$. Označme $D(\boldsymbol{\theta})$ množinu všech dvojic (i, j) takových, že $p(j|i; \boldsymbol{\theta}) > 0$. Jestliže je splněno*

1. $D(\boldsymbol{\theta}) \equiv D$ nezávisí na parametru $\boldsymbol{\theta}$,
2. $p(j|i; \boldsymbol{\theta})$ má v Θ spojité parciální derivace třetího řádu,
3. $|D| \times K$ matice s prvky $\frac{\partial}{\partial \theta_k} p(j|i; \boldsymbol{\theta})$ pro $(i, j) \in D, k = 1, \dots, K$, má v Θ hodnost K ,
4. $(X_t, t = 1, 2, \dots)$ je ergodický,

potom platí, že existuje posloupnost odhadů $\{\hat{\boldsymbol{\theta}}_T; T = 1, 2, \dots\}$, kde $\hat{\boldsymbol{\theta}}_T$ je funkcí prvních T hodnot řetězce, konvergující v pravděpodobnosti pro $T \rightarrow \infty$ ke skutečné hodnotě parametru $\boldsymbol{\theta}$ a zároveň každé $\hat{\boldsymbol{\theta}}_T$ je s pravděpodobností jdoucí k jedné pro $T \rightarrow \infty$ řešením soustavy věrohodnostních rovnic

$$\frac{\partial}{\partial \theta_k} L(\boldsymbol{\theta}) = 0, \quad k = 1, \dots, K.$$

Navíc platí

$$\sqrt{T} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta})),$$

kde $\mathbf{I}(\boldsymbol{\theta}) = -E \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log p(X_t | X_{t-1}; \boldsymbol{\theta})$.

Důkaz. Důkaz je uveden v (Billingsley, 1961, kap. 5). □

Uvedené tvrzení můžeme nyní aplikovat konkrétně pro binomický AR(1) proces.

Tvrzení 15. *Ve stacionárním binomickém AR(1) modelu existuje konzistentní odhad parametrů $(\hat{\rho}_{ML}, \hat{\pi}_{ML})$ metodou maximální podmíněné věrohodnosti a platí*

$$\sqrt{T} \begin{pmatrix} \hat{\rho}_{ML} - \rho \\ \hat{\pi}_{ML} - \pi \end{pmatrix} \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \mathbf{I}^{-1}(\rho, \pi)),$$

kde $\mathbf{I}(\rho, \pi) = -E \frac{\partial^2}{\partial(\rho, \pi)^T \partial(\rho, \pi)} \log p(X_t | X_{t-1}; \rho, \pi)$.

Důkaz. Nejprve ověříme podmínky 1 až 4 v tvrzení 14. V binomickém AR(1) modelu jsou všechny stavy navzájem dosažitelné s kladnou pravděpodobností nezávisle na parametrech (ρ, π) z přípustné množiny. Pravděpodobnosti přechodu $p(j|i; \rho, \pi)$ jsou podle tvrzení 5 polynomy dvou proměnných (ρ, π) , což znamená, že mají spojitě derivace třetího řádu. Ergodicitu procesu jsme již zmiňovali dříve a zbývá ukázat, že matice v podmínce 3 má hodnotu 2. K tomu stačí ukázat lineární nezávislost dvou řádků například pro $(i, j) \in \{(0, 0), (1, 0)\}$. Vyjádříme si

$$\begin{aligned} p(0|0; \rho, \pi) &= (1 - \beta)^n = (1 - \rho + \rho\pi)^n, \\ p(0|1; \rho, \pi) &= (1 - \alpha)(1 - \beta)^{n-1} = (1 - \pi)(1 - \rho)(1 - \rho + \rho\pi)^{n-1} \end{aligned}$$

a rozepíšeme parciální derivace obou výrazů:

$$\begin{aligned} \frac{\partial}{\partial \rho} p(0|0; \rho, \pi) &= n(\pi - 1)(1 - \rho + \rho\pi)^{n-1}, \\ \frac{\partial}{\partial \pi} p(0|0; \rho, \pi) &= n\rho(1 - \rho + \rho\pi)^{n-1}, \\ \frac{\partial}{\partial \rho} p(0|1; \rho, \pi) &= (n - 1)(1 - \pi)(1 - \rho)(\pi - 1)(1 - \rho + \rho\pi)^{n-2} \\ &\quad - (1 - \pi)(1 - \rho + \rho\pi)^{n-1} \\ &= (1 - \pi)(1 - \rho + \rho\pi)^{n-2} [(1 - \rho)(n - 1)(\pi - 1) - (1 - \rho - \rho\pi)], \\ \frac{\partial}{\partial \pi} p(0|1; \rho, \pi) &= (1 - \rho)(1 - \rho + \rho\pi)^{n-2} [\rho(1 - \rho)(n - 1) - (1 - \rho - \rho\pi)]. \end{aligned}$$

Jak nyní můžeme vidět, tyto dva řádkové vektory $\left(\frac{\partial}{\partial \rho} p(0|0; \rho, \pi) \quad \frac{\partial}{\partial \pi} p(0|0; \rho, \pi) \right)$ a $\left(\frac{\partial}{\partial \rho} p(0|1; \rho, \pi) \quad \frac{\partial}{\partial \pi} p(0|1; \rho, \pi) \right)$ jsou pro všechna (ρ, π) z přípustné množiny lineárně nezávislé a matice z podmínky 3 tak má hodnotu 2.

Zbývá ještě ospravedlnit použití podmíněné věrohodnosti namísto úplné. Jelikož platí $l(\rho, \pi) = l^{(C)}(\rho, \pi) + \log \mathbf{P}(X_1 = x_1; \rho, \pi)$, liší se obě věrohodnosti pouze ve členu, který závisí pouze na prvním pozorování a s rostoucím počtem pozorování T je tak za ověřených předpokladů regularity zanedbatelný vzhledem k součtu logaritmu pravděpodobností přechodu ve výrazu $l^{(C)}(\rho, \pi)$ (viz Billingsley, 1961, str. 4). □

Poznámka. Použití podmíněné maximální věrohodnosti není nutné a vede pouze k mírnému zjednodušení výrazů. Pro konečný rozsah výběru jsou rozdíly mezi podmíněnou a nepodmíněnou věrohodností zanedbatelné, viz (Weiß a Kim, 2013a, kap. 3.3). Při výpočtech budeme v této práci používat podmíněnou věrohodnost.

3. Testy dobré shody

V této kapitole se budeme zabývat úlohou testování, jestli náhodné veličiny X_1, X_2, \dots , ze kterých pozorujeme data, pochází z binomického AR(1) procesu. Souhrnně můžeme zformulovat tuto hypotézu jako

H_0 : Existují (ρ, π) z přípustné množiny podle definice 2 takové, že se $(X_t, t = 1, 2, \dots)$ řídí stacionárním binomickým AR(1) modelem s těmito parametry.

Nejprve budeme vycházet z článku (Weiß a Kim, 2015), odkud shrnujeme první dva přístupy. Následně navrhneme test dobré shody inspirovaný článkem (Hudecová a kol., 2015) založený na vytvořující funkci procesu. Jednotlivé metody poté budou porovnány v následující kapitole věnované simulacím.

3.1 Test založený na marginálním rozdělení

První přístup se omezuje pouze na ověření marginálního rozdělení procesu, empirické četnosti výskytů jednotlivých stavů jsou porovnány s teoretickými pomocí běžné Pearsonovy statistiky. Na rozdíl od nezávislých stejně rozdělených náhodných veličin ale tato statistika u korelovaného procesu nekonverguje v distribuci k χ^2 rozdělení ale k váženému součtu čtverců normálních náhodných veličin s různými váhami, jak se ukáže v následujícím tvrzení.

Tvrzení 16 (Pearsonova statistika). *Nechť $(X_t, t = 1, 2, \dots)$ je stacionární binomický AR(1) proces. Označme pro $i = 1, \dots, n$ počet výskytů ve stavu i jako $N_i := \sum_{t=1}^T \mathbb{1}[X_t = i]$ a jednotlivé pravděpodobnosti výskytů $p_i := P(X_t = i)$. Potom platí*

$$I_1 := \sum_{i=0}^n \frac{(N_i - Tp_i)^2}{Tp_i} \xrightarrow{T \rightarrow \infty} \sum_{j=1}^n \frac{1 + \rho^j}{1 - \rho^j} \cdot Z_j^2,$$

kde Z_1, \dots, Z_n jsou nezávislé náhodné veličiny s rozdělením $\mathcal{N}(0, 1)$.

Důkaz. Viz tvrzení 1 v (Weiß, 2009a). □

Poznámka. Ve speciálním případě $\rho = 0$ má I_1 asymptoticky χ^2 rozdělení o $(n + 1) - 1 = n$ stupních volnosti a tato situace je tak shodná s testováním dobré shody náhodného výběru X_1, \dots, X_T s rozdělením $\text{Bi}(n, \pi)$ pro známé π , viz (Anděl, 2005, s. 271).

Pro další postup se nám bude hodit vyjádření Pearsonovy statistiky ve vektorové formě, platí

$$\begin{aligned} I_1 &= \sum_{i=0}^n \frac{(N_i - Tp_i)^2}{Tp_i} = T \sum_{i=0}^n \frac{(N_i/T - p_i)^2}{p_i} = \\ &= T(\hat{\mathbf{p}} - \mathbf{p})^T D^{-1}(\hat{\mathbf{p}} - \mathbf{p}), \end{aligned}$$

kde jsme označili $\mathbf{p} = (p_0, \dots, p_n)^T$, dále $\hat{\mathbf{p}} = (N_0/T, \dots, N_n/T)^T$ jako empirický odhad vektoru \mathbf{p} a matici $D = \text{diag}(\mathbf{p})$.

Pearsonova statistika pro iid data s sebou nese dobře známý problém, že pokud je pro nějaký stav i očekávaný počet výskytů Tp_i malý, aproximace jejího rozdělení pomocí χ^2 rozdělení není příliš přesná. Namísto kontrolování všech stavů $0, \dots, n$ tak můžeme sloučením některých stavů vytvořit menší počet kategorií, ve kterých budou očekávané počty výskytů dostatečně velké. Vytvoříme $n - k - l + 1$ kategorií sloučením krajních stavů od 0 do k v jednu kategorii, obdobně stavů $n - l$ až n do další kategorie a stavy uprostřed ponecháme samostatně. Za k, l přitom zvolíme co nejmenší čísla $k, l \in N_0$, taková, že $T \cdot P(X = k + 1) \geq 5$ a $T \cdot P(X = n - l - 1) \geq 5$. Ve výsledku tedy porovnááme empirické četnosti kategorií

$$\{0, \dots, k\}, \{k + 1\}, \dots, \{n - l - 1\}, \{n - l, \dots, n\},$$

s teoretickými za platnosti binomického AR(1) modelu opět pomocí Pearsonovy statistiky.

Vyjádríme si také maticově transformaci stavů na výše definované kategorie, tu můžeme blokově zapsat ve tvaru

$$\mathbf{A}^{(k,l)} := (A_k | I_{n-l-k+1} | A_l),$$

kde matice A_k typu $(n - k - l + 1) \times k$ má pouze v prvním řádku vektor jedniček, matice A_l typu $(n - k - l + 1) \times l$ má pouze v posledním řádku vektor jedniček a $I_{n-l-k+1}$ je čtvercová jednotková matice. Z pravděpodobností výskytu v jednotlivých stavech \mathbf{p} nyní určíme pravděpodobnost výskytu v jednotlivých kategoriích jako

$$\mathbf{A}^{(k,l)} \mathbf{p} = (p_1 + \dots + p_k, p_{k+1}, \dots, p_{n-l-1}, p_{n-l} + \dots + p_n)^T.$$

Označme potom pro kategorie stavů kombinovanou Pearsonovu statistiku

$$I_1^{(k,l)} := T \left[\mathbf{A}^{(k,l)} (\hat{\mathbf{p}} - \mathbf{p}) \right]^T \left[\text{diag}(\mathbf{A}^{(k,l)} \mathbf{p}) \right]^{-1} \left[\mathbf{A}^{(k,l)} (\hat{\mathbf{p}} - \mathbf{p}) \right],$$

pro kterou nalezneme v následujícím tvrzení asymptotické vlastnosti. Tím bude opět vážený součet nezávislých náhodných veličin s normovaným normálním rozdělením, tentokrát ale se složitějšími váhami.

Tvrzení 17 (Kombinovaná Pearsonova statistika). *Nechť $(X_t, t = 1, 2, \dots)$ je stacionární binomický AR(1) proces. Označme pro $k, l \in \mathbb{N}$ matici*

$$\Sigma^{k,l} := \mathbf{A}^{(k,l)} \Sigma \left[\mathbf{A}^{(k,l)} \right]^T,$$

kde $\Sigma = 2(\mathbf{I} - \mathbf{P}^T + \mathbf{p}\mathbf{1}^T)^{-1} \mathbf{D} - \mathbf{D} - \mathbf{p}\mathbf{p}^T$ a symbolem $\mathbf{1}$ značíme vektor jedniček délky $n + 1$.

Nechť $r \in \mathbb{N}$ je takové, že $\lambda_1^{(k,l)}, \dots, \lambda_r^{(k,l)}$ jsou nenulová vlastní čísla matice

$$\left[\text{diag}(\mathbf{A}^{(k,l)} \mathbf{p}) \right]^{-1} \Sigma^{k,l},$$

potom platí

$$I_1^{(k,l)} \xrightarrow[T \rightarrow \infty]{d} \sum_{j=1}^r \lambda_j^{(k,l)} \cdot Z_j^2,$$

kde Z_1, \dots, Z_r jsou nezávislé náhodné veličiny s rozdělením $\mathcal{N}(0, 1)$.

Důkaz. Viz tvrzení 2.1 v (Weiß a Kim, 2015). □

Problémem obou výše uvedených verzí Pearsonovy statistiky je, že se v nich uvažuje plně specifikovaný model. Pokud chceme ověřit, jestli data mohou pocházet z nespécifikovaného binomického AR(1) modelu a jsou v souladu s hypotézou

$$H_0: X_t \sim \text{Bi}(n, \cdot),$$

musíme nejprve odhadnout parametry (ρ, π) jako by platil binomický AR(1) model a poté pomocí Pearsonovy statistiky testovat, jestli je marginální rozdělení procesu $X_t \sim \text{Bi}(n, \hat{\pi})$. Celý postup je tak následující:

- 1) Odhadneme z dat parametry (ρ, π) za předpokladu stacionárního binomického AR(1) modelu a na jejich základě určíme vektor očekávaných pravděpodobností stavů $\tilde{\mathbf{p}} = (\tilde{p}_1, \dots, \tilde{p}_n)$.
- 2) Zvolme $k, l \in N_0$ nejmenší splňující $T\tilde{p}_{k+1} \geq 5$ a $T\tilde{p}_{n-l-1} \geq 5$.
- 3) Nalezneme $\tilde{\lambda}_1^{(k,l)}, \dots, \tilde{\lambda}_r^{(k,l)}$ vlastní čísla matice $[\text{diag}(\mathbf{A}^{(k,l)}\tilde{\mathbf{p}})]^{-1} \tilde{\Sigma}^{k,l}$, pro matici $\tilde{\Sigma}$ s odhadnutým vektorem $\tilde{\mathbf{p}}$.
- 4) Zamítneme H_0 na hladině α pokud $I_1^{k,l} > q_{1-\alpha}$, kde q_α je $(1 - \alpha)$ kvantil rozdělení $\sum_{j=1}^r \tilde{\lambda}_j^{(k,l)} \cdot Z_j^2$, kde Z_1, \dots, Z_r jsou nezávislé náhodné veličiny s rozdělením $\mathcal{N}(0, 1)$.

Poznámka. (Weiß a Kim, 2015) uvažují tento test pouze k inferenci o parametru π za předpokladu známé hodnoty parametru ρ . V takovém případě ověřujeme hypotézu $H_0: X_t \sim \text{Bi}(n, \pi_0)$. Použití celého testu se liší pouze v prvním kroku, kde namísto odhadu vektoru pravděpodobností $\tilde{\mathbf{p}}$ použijeme vektor očekávaných pravděpodobností \mathbf{p}_0 za platnosti hypotézy.

V případě použití Pearsonovy statistiky pro iid data se asymptotickému χ^2 rozdělení odebírá stupeň volnosti za každý odhadnutý parametr, viz (Anděl, 2005, s. 272). U binomického AR(1) procesu není popsána analogická korekce, bez ní očekáváme výsledný test konzervativní, jak ověříme v kapitole věnované simulacím. V duchu odebírání stupňů volnosti za každý odhadnutý parametr také z vlastní iniciativy vyzkoušíme porovnávat testovou statistiku $I_1^{k,l}$ také s kvantily rozdělení $\sum_{j=1}^{r-2} \tilde{\lambda}_j^{(k,l)} \cdot Z_j^2$, kde Z_1, \dots, Z_{r-2} jsou nezávislé náhodné veličiny s rozdělením $\mathcal{N}(0, 1)$. Použití Pearsonovy statistiky pro test dobré shody s neznámými parametry zmiňuje (Weiß, 2018), autor tam pro tuto situaci odvozuje asymptotické rozdělení testové statistiky, které je ve stejném tvaru jako ve tvrzení 17 a liší se pouze váhami $\lambda_j^{(k,l)}$, jejichž výpočet je komplikovanější. V této práci tak zůstaneme u výše prezentované verze testu a ověříme její použitelnost v situaci s neznámými parametry v simulační studii. Takové použití v (Weiß a Kim, 2015) nebylo ověřeno.

3.2 Test založený na autokorelační struktuře

Druhý přístup k testování ověřuje správnost autokorelační struktury procesu za použití testové statistiky podobné Q-testům nekorelovanosti náhodných pro-

cesů, kterou přizpůsobíme binomickému AR(1) procesu. Testová statistika bude porovnávat empirické hodnoty parciální autokorelační funkce s jejich teoretickými hodnotami za platnosti modelu.

Definice 4 (Parciální autokorelační funkce). *Bud' $(X_t, t \in \mathbb{Z})$ stacionární náhodný proces a $\rho(k)$ pro $k \in \mathbb{N}_0$ jeho autokorelační funkce. Parciální autokorelační funkci $\rho_{part}(k)$ definujeme předpisem $\rho_{part}(1) = \rho(1)$ a pro $k \in \mathbb{N}, k \geq 2$ jako*

$$\rho_{part}(k) := \frac{\text{cov}(X_t, X_{t+k} | X_{t+1}, \dots, X_{t+k-1})}{\sqrt{\text{var}(X_t | X_{t+1}, \dots, X_{t+k-1}) \text{var}(X_{t+k} | X_{t+1}, \dots, X_{t+k-1})}}$$

Hodnoty $\rho_{part}(k)$ se prakticky určují pomocí Durbinova-Levinsonova algoritmu, položením $\rho_{part}(k) = r_{k,k}$, kde hodnoty $r_{k,k}$ spočítáme rekurentně z rovností

$$\begin{aligned} r_{1,1} &= \rho(1), \\ r_{k,j} &= r_{k-1,j} - r_{k,k} \cdot r_{k-1,k-j}, & \text{pro } j = 1, \dots, k-1, \\ r_{k,k} &= \frac{\rho(k) - \sum_{j=1}^{k-1} r_{k-1,j} \cdot \rho(k-j)}{1 - \sum_{j=1}^{k-1} r_{k-1,j} \cdot \rho(j)}, & \text{pro } k = 2, 3, \dots \end{aligned}$$

Z tohoto rekurentního předpisu získáme také odhad parciální autokorelační funkce $\hat{\rho}_{part}$, kde pouze zaměníme autokorelační funkci $\rho(k)$ jejím empirickým odhadem $\hat{\rho}(k)$.

Tvrzení 18. *Nechť $(X_t, t = 1, 2, \dots)$ je stacionární binomický AR(1) proces, potom pro všechna $i \in \mathbb{N}, i \geq 2$ platí $\rho_{part}(i) = 0$.*

Důkaz. Pro libovolné $t \in \mathbb{N}$ a $i = 2, 3, \dots$ platí

$$\text{cov}(X_t, X_{t+i} | X_{t+1}, \dots, X_{t+i-1}) = 0,$$

neboť z markovské vlastnosti závisí X_{t+k} pouze na X_{t+k-1} skrze operátor \circ , který je z definice nezávislý na minulých hodnotách procesu. Tedy X_t a X_{t+i} jsou podmíněně X_{t+i-1} nekorelované a $\rho_{part}(i) = 0$. □

Pokud se tedy proces $(X_t, t \in \mathbb{Z})$ řídí binomickým AR(1) procesem, měly by jeho výběrové parciální autokorelace být blízké nule od řádu dva. V následujícím tvrzení nalezneme jejich asymptotické rozdělení.

Tvrzení 19 (Quenouilleova aproximace pro binomický AR(1) proces). *Nechť $(X_t, t = 1, 2, \dots)$ je stacionární binomický AR(1) proces, $K \in \mathbb{N}$ pevné a označme $\hat{\boldsymbol{\rho}}_{part} = (\hat{\rho}_{part}(2), \dots, \hat{\rho}_{part}(K))^T$. Potom platí*

$$\sqrt{T} \hat{\boldsymbol{\rho}}_{part} \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

kde $\boldsymbol{\Sigma}$ je diagonální matice s prvky σ_i na diagonále tvaru

$$\sigma_i = 1 + \frac{(1 - 2\pi)^2}{n\pi(1 - \pi)} \cdot \frac{\rho^i}{(1 + \rho)^2},$$

pro $i = 2, \dots, K$.

Důkaz. Viz tvrzení 4.4 v (Weiß a Kim, 2015). □

Tvrzení nám již dává vše potřebné k inferenci o parciální autokorelační funkci procesu. Test dobré shody s binomickým AR(1) modelem realizujeme tak, že pro předem zvolené $m \in \mathbb{N}, m \geq 2$, budeme testovat hypotézu

$$H_0: \rho_{part}(i) = 0, \text{ pro všechna } i = 2, 3, \dots, m.$$

Testová statistika bude podobná Boxově-Pierceově Q-testu vyšetřujícímu nekorelovanost náhodného procesu. Vyjdeme z toho, že v binomickém AR(1) procesu jsou teoretické parciální autokorelace $\rho_{part}(i) = 0$ pro $i = 2, 3, \dots$ a jejich empirické protějšky $\hat{\rho}_{part}(i)$ jsou asymptoticky nezávislé a normálně rozdělené

$$\sqrt{T} \frac{\hat{\rho}_{part}(i)}{\sigma_i} \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(0, 1). \quad (3.1)$$

Součtem čtverců těchto výrazů tak dostaneme testovou statistiku

$$Q_{BP}(m) = T \sum_{i=2}^m \frac{\hat{\rho}_{part}^2(i)}{\sigma_i^2},$$

která má asymptoticky χ^2 rozdělení o $(m - 1)$ stupních volnosti. Pro aplikaci na testování dobré shody s binomickým AR(1) modelem s neznámými hodnotami parametrů (ρ, π) ležícími v přípustné množině odhadneme konzistentně výrazy σ_i dosazením konzistentních odhadů pro (ρ, π) předvedených v kapitole 2. Pro potřeby reference na citovaný článek označíme testovou statistiku s odhadnutým rozptylem hvězdičkou:

$$Q_{BP}^*(m) = T \sum_{i=2}^m \frac{\hat{\rho}_{part}^2(i)}{\hat{\sigma}_i^2},$$

Při standardních Q-testech nekorelovanosti náhodného procesu se spíše uvažuje modifikace Boxovy-Piercovy testové statistiky označovaná jako Boxova-Ljungova statistika, která je pro konečný rozsah výběru blíže asymptotickému χ^2 rozdělení, viz například (Cipra, 2008, s. 348). Také (Weiß a Kim, 2015) ukazují v několika simulacích, že i v případě binomického AR(1) procesu mají testové statistiky

$$Q_{BL}(m) = T(T + 2) \sum_{i=2}^m \frac{\hat{\rho}_{part}^2(i)}{(T - i)\sigma_i^2},$$

$$Q_{BL}^*(m) = T(T + 2) \sum_{i=2}^m \frac{\hat{\rho}_{part}^2(i)}{(T - i)\hat{\sigma}_i^2}$$

lepší vlastnosti než základní Boxova-Pierceova verze. Hypotézu H_0 zamítneme na hladině α , pokud pro předem zvolené $m \in \mathbb{N}$ hodnota testové statistiky $Q_{BL}^*(m) \geq \chi_{m-1}^2(1 - \alpha)$.

Poznámka. Řád m , do kterého kontrolujeme souhrnně parciální autokorelace, by měl být zvolen s ohledem na uvažovanou alternativu. Více se o tom zmíníme v rámci simulační studie.

Poznámka. V kauzálním AR(1) procesu se používá Quenouilleova aproximace $\sqrt{T}\hat{\rho}_{part}(i) \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(0, 1)$ pro $i = 2, 3, \dots$ (Cipra, 2008, s. 332). Asymptotický rozptyl v tvrzení 19 se tak v případě binomického AR(1) procesu liší pouze členem

$$\frac{(1 - 2\pi)^2}{n\pi(1 - \pi)} \cdot \frac{\rho^i}{(1 + \rho)^2},$$

který mizí s rostoucím n nebo i .

3.3 Test založený na vytvořující funkci

Předchozí metody cílily vždy jen na jednu z charakteristik binomického AR(1) procesu. Pokusíme se tedy nyní definovat testovou statistiku, která by zahrnovala jak autokorelační strukturu procesu tak i jeho marginální rozdělení. Inspirujeme se testovou statistikou uvedenou v (Hudecová a kol., 2015), která porovnává empirickou vytvořující funkci procesu s teoretickou za platnosti binomického AR(1) modelu.

Předpokládejme nejprve, že pro X_1, X_2, \dots platí stacionární binomický AR(1) model se známými parametry (ρ, π) . Potom z rovnosti (1.1) v důkazu tvrzení 7 můžeme vyjádřit vytvořující funkci náhodné veličiny X_t pro $s \in R$ jako

$$P_{X_t}(s) = (s\beta + 1 - \beta)^n \cdot P_{X_{t-1}}\left(\frac{s\alpha + 1 - \alpha}{s\beta + 1 - \beta}\right),$$

ze stacionarity navíc vyplývá, že $P_{X_{t-1}} \equiv P_{X_t}$ a celou rovnost tak můžeme implicitně zapsat jako

$$P_X(s) = (s\beta + 1 - \beta)^n \cdot P_X\left(\frac{s\alpha + 1 - \alpha}{s\beta + 1 - \beta}\right),$$

což můžeme přepsat do tvaru

$$P_X(s) - (s\beta + 1 - \beta)^n \cdot P_X\left(\frac{s\alpha + 1 - \alpha}{s\beta + 1 - \beta}\right) = 0. \quad (3.2)$$

Za platnosti stacionárního binomického AR(1) modelu je empirická vytvořující funkce

$$\hat{P}_X(s) = \frac{1}{T} \sum_{t=1}^T s^{X_t}$$

pro $s \in R$ konzistentním odhadem skutečné vytvořující funkce $P_X(s)$, což plyne

z tvrzení 10. Potom i rozdíl

$$\begin{aligned}
& \hat{P}_X(s) - (s\beta + 1 - \beta)^n \cdot \hat{P}_X\left(\frac{s\alpha + 1 - \alpha}{s\beta + 1 - \beta}\right) = \\
&= \frac{1}{T} \sum_{t=1}^T s^{X_t} - (s\beta + 1 - \beta)^n \cdot \hat{P}_X\left(\frac{s\alpha + 1 - \alpha}{s\beta + 1 - \beta}\right) \\
&= \frac{1}{T} \sum_{t=1}^T s^{X_t} - (s\beta + 1 - \beta)^n \frac{1}{T} \sum_{t=1}^T \left(\frac{s\alpha + 1 - \alpha}{s\beta + 1 - \beta}\right)^{X_t} \\
&= \frac{1}{T} \sum_{t=1}^T s^{X_t} - \frac{1}{T} \sum_{t=1}^T \left[(s\alpha + 1 - \alpha)^{X_t} (s\beta + 1 - \beta)^{n-X_t} \right] \\
&= \frac{1}{T} \sum_{t=1}^T \left[s^{X_t} - (s\alpha + 1 - \alpha)^{X_t} (s\beta + 1 - \beta)^{n-X_t} \right]
\end{aligned} \tag{3.3}$$

bude pro $s \in \mathbb{R}$ s rostoucím T blízký nule. Při testování dobré shody neznáme parametry (ρ, π) binomického AR(1) procesu, ze kterého by měla pozorování pocházet. Parametry (α, β) v rovnosti výše tak nahradíme jejich konzistentními odhady $(\hat{\alpha}, \hat{\beta})$ za platnosti binomického AR(1) modelu a budeme vyšetřovat, jestli je uvedený výraz blízký nule. Hypotézou tak bude

$$H_0: \exists(\alpha, \beta) \in (0, 1)^2 \text{ takové, že platí rovnost (3.2).}$$

Rovnost dvou funkcí můžeme obecně porovnat různými způsoby, například v jednom bodě, několika bodech nebo třeba jako supremum jejich rozdílů na nějaké množině. Všimněme si nejprve, že výraz (3.3) můžeme interpretovat jako rozdíl dvou odhadů vytvářející funkce náhodné veličiny X_t . První člen je neparametrický odhad bez platnosti jakéhokoliv modelu pro data. Druhý člen označíme jako semiparametrický odhad, neboť jsme za předpokladu platnosti binomického AR(1) modelu upravili vytvářející funkci do tvaru, který závisí na parametrech, ale zároveň část výrazu odhadujeme empirickou vytvářející funkcí. Lze snadno ukázat, že na základě znalostí derivací vytvářející funkce v nule už známe kompletně rozdělení daného procesu. Dále na základě derivací zleva v bodě jedna dokážeme určit momenty procesu. Rozhodli jsme se tak porovnávat rozdíly obou odhadů na intervalu $(0, 1)$, a to integrací čtverců rozdílů obou funkcí. Definujme tak testovou statistiku

$$S_T(\mathbf{X}) = \frac{1}{T} \int_0^1 \left\{ \sum_{t=1}^T \left[s^{X_t} - (s\hat{\alpha} + 1 - \hat{\alpha})^{X_t} (s\hat{\beta} + 1 - \hat{\beta})^{n-X_t} \right] \right\}^2 ds$$

a proti hypotéze H_0 budou svědčit vysoké hodnoty této statistiky. Jelikož by bylo obtížné vyjádřit rozdělení $S_T(\mathbf{X})$, budeme testovat hypotézu H_0 pomocí parametrického bootstrapového testu. Rozdělení testové statistiky tak nalezneme pomocí Monte-Carlo principu, generováním B nezávislých realizací stacionárního binomického AR(1) procesu s parametry $(\hat{\rho}, \hat{\pi})$, pomocí nichž odhadneme empiricky distribuční funkci S_T . Celý postup je následující:

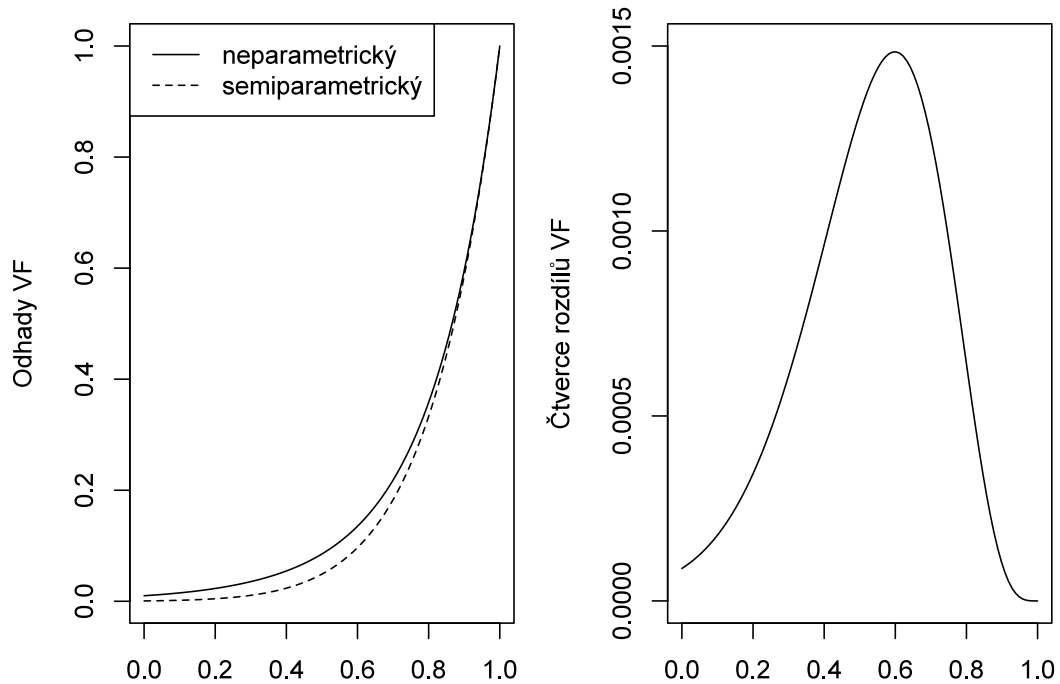
- 1) Odhadneme z dat parametry (ρ, π) za předpokladu stacionárního binomického AR(1) modelu a spočítáme hodnotu testové statistiky $S_T(\mathbf{x})$.

- 2) S odhadnutými parametry $(\hat{\rho}, \hat{\pi})$ generujeme B nezávislých realizací stacionárního binomického AR(1) procesu délky T . Označme generovaná data $\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_B^*$, kde $\mathbf{x}_i^* = (x_{i1}^*, x_{i2}^*, \dots, x_{iT}^*)$.
- 3) Na těchto generovaných datech poté spočítáme hodnoty testové statistiky $S_T(\mathbf{x}_1^*), S_T(\mathbf{x}_2^*), \dots, S_T(\mathbf{x}_B^*)$ a odhadneme p-hodnotu testu výrazem

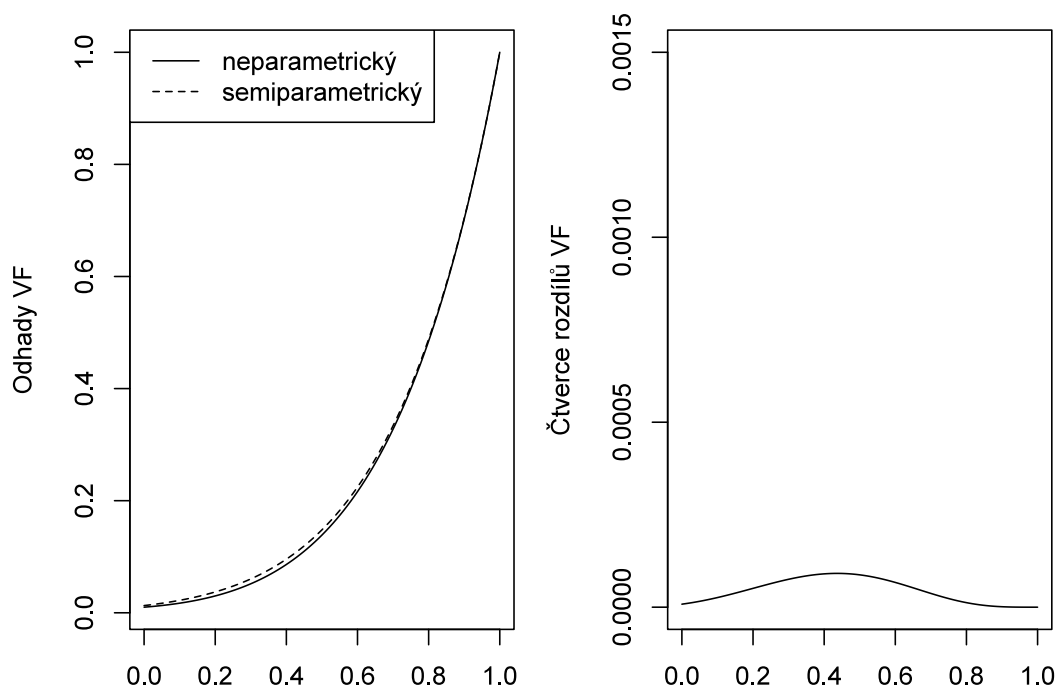
$$\hat{p}_B = \frac{1 + \sum_{b=1}^B [S_T(\mathbf{x}_b^*) \geq S_T(\mathbf{x})]}{B + 1}.$$

Hypotézu H_0 zamítneme na hladině α pokud $\hat{p}_B \leq \alpha$.

Na obrázku 3.1 ilustrujeme výpočet testové statistiky porovnáním neparametrického odhadu vytvářející funkce a semiparametrického za platnosti AR(1) modelu. Pro názornost rozdílů jsme použili realizaci náhodného procesu $X_t = \min\{Z_t, 10\}$, kde Z_t jsou iid z rozdělení $\text{Po}(5)$, o délce $T = 100$. Pro porovnání přidáváme také obrázek 3.2 za platnosti hypotézy pro binomický AR(1) proces s parametry $n = 10$, $\pi = 0,35$, $\rho = 0,5$, $T = 100$.



Obrázek 3.1: Ukázka výpočtu testové statistiky VF testu pro realizaci náhodného procesu $X_t = \min\{Z_t, 10\}$, kde Z_t jsou iid z rozdělení $\text{Po}(5)$, o délce $T = 100$.



Obrázek 3.2: Ukázka výpočtu testové statistiky VF testu pro binomický AR(1) proces s parametry $n = 10$, $\pi = 0,35$, $\rho = 0,5$, $T = 100$.

4. Simulace

V této kapitole postupně uvedeme výsledky několika různých simulací k porovnání představených testů dobré shody. Pro zkrácení zápisu budeme dále používat zkratku *MR* pro test dobré shody založený na marginálním rozdělení, *AF* pro test založený na autokorelační funkci a *VF* pro navržený test na základě vytvořující funkce. Uvažovanou vlastní korekci MR testu pomocí redukce pomyslných stupňů volnosti pro situaci s neznámými parametry zmíněnou v závěru sekce 3.1 budeme v tabulkách s výsledky značit zkratkou *MR adj*.

Malá část našich simulací replikuje některé výsledky uvedené již v (Weiß a Kim, 2015). Ty jednak doplňujeme o náš navržený VF test, dále ilustrujeme použití MR testu pro situaci s neznámým marginálním rozdělením a přidáváme některé další alternativy, které autoři ve zmíněném článku neuvažovali. Veškeré simulace jsou vyhodnoceny na základě 4000 opakování, velikost bootstrapových výběrů pro VF test jsme zvolili $B = 1001$. Výpočty byly provedeny ve statistickém software R ve verzi 4.0.3 (R Core Team, 2020) za použití vlastních funkcí, které dáváme k dispozici ke stažení z <https://labk10.karlin.mff.cuni.cz/~hledikja/funkce.R>. Nejprve se ale ještě krátce zaměříme na odhady parametrů.

4.1 Podíl odhadů v přípustné množině

Při praktických aplikacích můžeme narazit na dva problémy při odhadech parametrů, ty jednak nemusí být pro některé realizace procesu dobře definovány nebo nemusí nutně ležet uvnitř přípustné množiny znázorněné na obrázku 1.1.

První situace nastane u CLS nebo YW odhadu parametru ρ , pokud je jmenovatel příslušného zlomku nulový. Ten ovšem vyjde v obou případech nulový právě když je výběrový rozptyl pozorování nulový, neboli pozorujeme realizaci, která je v čase konstantní. Takový problém nastane typicky jen pro malý počet pozorování T a zároveň pro hodnoty π velmi blízké 0 nebo 1 spolu s velmi vysokou korelovaností procesu. Jak dokládá tabulka 4.1, i pro $T = 25$ a parametry $\pi = 0,1$, $\rho = 0,9$ je více než 95 % odhadů dobře definovaných a pro další použití se tak touto situací prakticky nemusíme zabývat.

Můžeme ovšem narazit na odhad (ρ, π) metodou CLS nebo YW, který leží mimo přípustnou množinu. Vzhledem k tomu, že odhady π neleží v otevřeném intervalu $(0, 1)$ jen v případě konstantní realizace, tak bude problémem pouze odhad autokorelace ρ , který může vyjít jednak menší nebo roven $\max\{\frac{-\pi}{1-\pi}, \frac{1-\pi}{-\pi}\}$, nebo naopak větší nebo roven 1 v případě CLS odhadů. S ohledem na konzistenci všech odhadů se pak opět bude jednat o problém způsobený nízkým počtem pozorování, pokud zároveň ρ leží blízko zmíněným hranicím. Například pro $\pi = 0,1$ musí ležet $\rho \in (-0,111, 1)$. Zkoušeli jsme simulace pro různá nastavení parametrů (ρ, π) a pro ilustraci také uvažovat velmi krátké realizace $T \in \{25, 50\}$. V tabulce 4.2 vidíme, že pro volbu $\pi = 0,1$, $\rho = -0,1$ je i pro proces délky $T = 500$ podíl přípustných odhadů pouze 60,2 % u YW odhadů a 59,9 % u CLS odhadů. S menším počtem pozorování pak klesá až 44,4 % respektive k 43,5 %. Obecně pozorujeme, že YW odhad vyjde v přípustné množině o něco častěji než CLS odhad, rozdíly jsou však velmi malé. Připomeňme, že oba odhady jsou asymptot-

ticky ekvivalentní. Pro $\pi = 0,5$ a $\rho \in [-0,9, 0,9]$ jsme potom už od délky $T \geq 100$ nezaznamenali žádný nepřipustný odhad.

Vzhledem k tomu, že přípustná množina je otevřená, nemusí ani věrohodnost nabývat na této množině svého maxima. Ve výpočetním softwaru se ale tento problém díky konečné přesnosti výpočtů neprojeví a dostáváme výsledky ležící v přípustné množině. I z tohoto důvodu jsme pro veškeré odhady v rámci následujících simulací zvolili právě ML odhady.

4.2 Empirická hladina testů dobré shody

V tabulce 4.3 uvádíme dosažené hladiny testů pro binomické AR(1) procesy s parametry $n = 10$, $\pi \in \{0,10, 0,50\}$ a $\rho \in \{-0,50, 0, 0,25, 0,50, 0,75\}$. Nastavení $\rho = -0,5$ bylo uvažována pouze pro $\pi = 0,5$, aby byl proces definovaný. Podle očekávání vychází MR test jako velmi konzervativní. Předepsané hladině 5 % se blíží jen v případě $\pi = 0,5$ a $\rho \in \{-0,50, 0\}$. Můžeme také vidět, že dosažená hladina u navržené korekce MR testu pro nekorelovaný nebo negativně korelovaný proces i pro $T = 1000$ výrazně překračuje stanovených 5 %. Kromě toho se navíc pro výrazněji pozitivně korelovaný proces ani nepodařilo výrazněji eliminovat konzervativnost původního MR testu, použití tohoto testu tak není vhodné. Empirické hladiny AF a VR testů jsou i pro $T = 100$ blízko požadovaným 5 %, a to pro všechna nastavení parametrů (ρ, π) .

Uvedeme zde ještě simulaci dosažené hladiny pro populaci o velikosti $n = 5$. Můžeme si všimnout, že AF test byl už pro $n = 10$ lehce konzervativní pro $\pi = 0,1$ a vyšší autokorelaci $\rho \in \{0,50, 0,75\}$. Pro $n = 5$ je konzervativnost o něco výraznější, viz tabulka 4.4. Obecně jsou ale výsledky pro obě uvažované hodnoty n velmi podobné.

Vzhledem k tomu, že budeme později analyzovat reálná data s populací o velikosti $n = 11$, budeme v dalších simulacích uvažovat už jen $n = 10$, jako nastavení relevantní vzhledem k těmto datům. Weiß a Kim (2015) uvažovali n v rozmezí mezi 5 a 30, v článku se pak uvádí výsledky pro $n = 10$. Autoři zde zmiňují, že se výsledky pro různá n příliš neliší.

4.3 Síla testů dobré shody

Sílu tří uvedených testů dobré shody budeme ověřovat proti třem různým alternativám a bude nás zajímat zejména použití navrženého testu na základě vytvořující funkce v porovnání s ostatními testy. První alternativa se od binomického AR(1) procesu bude lišit pouze marginálním rozdělením se shodnou autokorelační strukturou. Druhá alternativa se naopak bude za stejného marginálního rozdělení lišit svou autokorelační strukturou a konečně třetí alternativa se bude lišit jak marginálním rozdělením, tak autokorelační strukturou. Ve všech případech budeme shodně uvažovat $T \in \{100, 250, 500, 1000\}$, $\pi \in \{0,1, 0,5\}$ a $n = 10$.

4.3.1 Beta-binomický AR(1) model

V první alternativě budeme uvažovat beta-binomický AR(1) model uvedený v (Weiß a Kim, 2014). Ten se od binomického AR(1) procesu liší pouze tím, že

parametry (α, β) v něm neuvažujeme pevné ale jako náhodné veličiny s beta rozdělením (připomeňme, beta rozdělení $\text{Be}(a, b)$ má hustotu $f(x) \propto x^{a-1}(1-x)^{b-1}$ pro $x \in (0, 1)$). Tím umožníme modelovat proces s větší variabilitou než v případě binomického AR(1) procesu. Před uvedením definice tohoto procesu bude třeba rozšířit zavedenou operaci \circ . Definice 1 uvažovala tuto operaci pouze mezi konstantou a náhodnou veličinou, pokud je nyní A náhodná veličina s hodnotami v intervalu $(0, 1)$, budeme pro náhodnou veličinu X nezávislou s A rozumět zápisem $A \circ X$ náhodnou veličinou, která má podmíněně $X = x, A = \alpha$ binomické rozdělení s parametry (x, α) . Nyní již uvedeme definici beta-binomického AR(1) procesu.

Definice 5 (Beta-binomický AR(1) model). *Bud' $n \in \mathbb{N}$, $\phi \in (0, 1)$, $\pi \in (0, 1)$, $\rho \in \left(\max\left\{\frac{-\pi}{1-\pi}, \frac{1-\pi}{-\pi}\right\}, 1\right)$ a označme $\beta := \pi(1-\rho)$ a $\alpha := \beta + \rho$. Dále označme náhodné veličiny $\alpha_\phi \sim \text{Be}\left(\frac{1-\phi}{\phi} \cdot \alpha, \frac{1-\phi}{\phi} \cdot (1-\alpha)\right)$ a $\beta_\phi \sim \text{Be}\left(\frac{1-\phi}{\phi} \cdot \beta, \frac{1-\phi}{\phi} \cdot (1-\beta)\right)$. Potom proces $(X_t, t \in \mathbb{Z})$ definovaný předpisem*

$$X_t = \alpha_\phi \circ X_{t-1} + \beta_\phi \circ (n - X_{t-1}),$$

kde obě operace \circ a náhodné veličiny α_ϕ, β_ϕ jsou navzájem nezávislé a navíc nezávislé na $(X_s, s < t)$, nazveme beta-binomický AR(1) proces.

Míra přidané variability závisí v takto definovaném modelu na parametru ϕ , který autoři označují jako disperzní parametr. Tento parametr totiž ovlivňuje marginální rozptyl procesu, jak bude dále ukázáno. Pro $\phi \rightarrow 0$ platí $\alpha_\phi \xrightarrow{P} \alpha$ a $\beta_\phi \xrightarrow{P} \beta$ a model se tak limitně blíží binomickému AR(1) modelu. Pro stacionární beta-binomický AR(1) proces s autokorelační funkcí $\rho(k)$ platí pro $k \in \mathbb{N}$

$$\begin{aligned} \mathbb{E} X_t &= n\pi, \\ \text{var} X_t &= n\pi(1-\pi) \cdot \frac{(1-\phi)(1+\rho) + n\phi \cdot [1 - 2\pi(1-\pi)(1-\rho)]}{(1-\phi)(1+\rho) + \phi \cdot [1 - 2\pi(1-\pi)(1-\rho)]}, \\ \rho(k) &= \rho^k. \end{aligned}$$

Tyto vlastnosti zde nebudeme dokazovat, důkaz je uveden v (Weiß a Kim, 2014). Jak vidíme, autokorelační struktura beta-binomického AR(1) procesu je shodná s binomickým AR(1) procesem. Test dobré shody založený na autokorelační funkci tak nebude schopen tyto procesy od sebe rozlišit. Marginální rozptyl pak pro $\phi \rightarrow 0$ konverguje k $n\pi(1-\pi)$, jak již bylo konstatováno výše, a budeme tedy chtít vyhodnotit sílu MR a VF testů pro ϕ blízké nule.

Zvolili jsme tři různé hodnoty $\phi \in \{0,025, 0,050, 0,100\}$ a ostatní parametry stejně jako v předchozím. V tabulkách 4.5, 4.6 a 4.7 nalezneme sílu všech testů proti těmto alternativám. Obecně si můžeme všimnout, že síla všech testů je výrazně nižší pro procesy s větší pozitivní korelací. Dále je VF test silnější než MR test ve většině případů, srovnatelné jsou pouze u procesů se zápornou autokorelací. Pro $\phi = 0,100$ je již pro 500 pozorování dostatečná síla k zamítnutí téměř všech testů, o něco menší síla je pouze u zmíněných procesů s velkou pozitivní autokorelací. Například pro $\phi = 0,050$ a $(\rho, \pi) = (0,1, 0,75)$ jsme pak i pro proces délky $T = 250$ zamítli VF testem hypotézu binomického AR(1) procesu jen u 16,7 % případů. (Weiß a Kim, 2015) zkoumali tuto alternativu pouze pro model se známým parametrem π (a tedy určeným marginálním rozdělením).

4.3.2 Binomický AR(2) model

Jako další alternativu použijeme binomický AR(2) model uvedený v (Weiß, 2009b). Ten vzniká obdobně jako binomický AR(1) proces s tím rozdílem, že nová hodnota v čase t závisí i na hodnotě procesu v čase $t - 2$ a tím se změní autokorelační struktura procesu.

Definice 6 (Binomický AR(2) model). *Bud' $n \in \mathbb{N}$, $\delta \in (0, 1)$, $\pi \in (0, 1)$, $\rho \in (\max\{\frac{-\pi}{1-\pi}, \frac{1-\pi}{-\pi}\}, 1)$ a označme $\beta := \pi(1-\rho)$ a $\alpha := \beta + \rho$. Dále označme náhodnou veličinu $D_t \sim \text{Alt}(\delta)$, potom proces $(X_t, t \in \mathbb{Z})$ definovaný předpisem*

$$X_t = D_t \cdot [\alpha \circ X_{t-1} + \beta \circ (n - X_{t-1})] + (1 - D_t) \cdot [\alpha \circ X_{t-2} + \beta \circ (n - X_{t-2})],$$

kde všechny operace \circ a náhodná veličina D_t jsou navzájem nezávislé a navíc nezávislé na $(X_s, s < t)$, nazveme binomický AR(2) proces.

Míru odlišnosti od binomického AR(1) procesu můžeme kontrolovat nastavením parametru δ , pro $\delta \rightarrow 1$ se model limitně blíží binomickému AR(1) procesu. Označme

$$r := \frac{\delta\rho}{1 - (1 - \delta)\rho},$$

potom pro stacionární binomický AR(2) proces s autokorelační funkcí $\rho(k)$ platí pro $k \in \mathbb{N}, k \geq 2$

$$\begin{aligned} \mathbf{E} X_t &= n\pi, \\ \text{var} X_t &= n\pi(1 - \pi), \\ \rho(1) &= r, \quad \rho(k) \neq r^k. \end{aligned}$$

Navíc také platí $X_t \sim \text{Bi}(n, \pi)$. Důkaz těchto vlastností je uveden v (Weiß, 2009b). Tuto alternativu tedy nedokážeme odlišit pomocí MR testů.

V simulacích budeme kontrolovat autokorelaci prvního řádu nastavením parametru $r \in \{-0,25, 0,25, 0,50, 0,75\}$. Nekorelovaný AR(2) proces ($r = 0$) nemá smysl uvažovat, neboť v tomto speciálním případě splňuje i AR(1) definici. Vzhledem k takto specifikované alternativě volíme pro AF test $m = 2$ a budeme tak tímto testem zkoumat pouze autokorelaci druhého řádu. Parametr delta volíme $\delta \in \{0,5, 0,7, 0,9\}$. V tabulkách 4.8, 4.9 a 4.10 nalezneme sílu všech testů proti těmto alternativám. Síla VF testů je sice v porovnání s MR testy vyšší, ale ani pro $\delta = 0,5$ a proces délky $T = 1000$ nezamítneme hypotézu AR(1) procesu ve více než 20 % případů. Naproti tomu AF testu stačí procesy délky $T = 250$ pro zamítnutí více než 80 % hypotéz. Nižší sílu testů jsme zjistili pro $\rho = 0,25$, což souvisí s poznámkou výše, že pro $\rho = 0$ máme opět binomický AR(1) proces. Pro $\delta = 0,7$ vypadají AF testy dostatečně silné pro $T \geq 500$ pozorování i pro méně korelované procesy. Naopak hodnota π nemá výrazný vliv na výslednou sílu testu.

(Weiß a Kim, 2015) uvádí pro tuto alternativu obdobné simulace pro AF testy. Je zde také blíže diskutována volba m pro výpočet testové statistiky, autoři ukazují, že pro tuto alternativu vychází nejlépe právě volba $m = 2$, což bylo důvodem pro její zvolení v našich simulacích. Obecně jsou naše pozorování ohledně AF testů ve shodě s tam uvedenými.

4.3.3 Beta-binomický AR(2) model

Třetí alternativou, kterou se budeme zabývat, je beta-binomický AR(2) model uvedený v (Weiß a Kim, 2014). Jedná se o kombinaci předchozích dvou modelů a bude se tak od binomického AR(1) procesu lišit marginálním rozdělením i autokorelační strukturou zároveň.

Definice 7 (Beta-binomický AR(2) model). *Bud' $n \in \mathbb{N}$, $\delta \in (0, 1)$, $\phi \in (0, 1)$, $\pi \in (0, 1)$, $\rho \in \left(\max\{\frac{-\pi}{1-\pi}, \frac{1-\pi}{-\pi}\}, 1\right)$ a označme $\beta := \pi(1 - \rho)$ a $\alpha := \beta + \rho$. Dále označme náhodné veličiny $D_t \sim \text{Alt}(\delta)$, $\alpha_\phi \sim \text{Be}\left(\frac{1-\phi}{\phi} \cdot \alpha, \frac{1-\phi}{\phi} \cdot (1 - \alpha)\right)$ a $\beta_\phi \sim \text{Be}\left(\frac{1-\phi}{\phi} \cdot \beta, \frac{1-\phi}{\phi} \cdot (1 - \beta)\right)$, potom proces $(X_t, t \in \mathbb{Z})$ definovaný předpisem*

$$X_t = D_t \cdot [\alpha_\phi \circ X_{t-1} + \beta_\phi \circ (n - X_{t-1})] + (1 - D_t) \cdot [\alpha_\phi \circ X_{t-2} + \beta_\phi \circ (n - X_{t-2})],$$

kde všechny operace \circ a náhodné veličiny $D_t, \alpha_\phi, \beta_\phi$ jsou navzájem nezávislé a navíc nezávislé na $(X_s, s < t)$, nazveme beta-binomický AR(2) proces.

Podobně jako v předchozím můžeme nahlédnout, že pro $\delta \rightarrow 1$ se proces limitně blíží beta-binomickému AR(1) procesu, pro $\phi \rightarrow 0$ binomickému AR(2) procesu a pro $\delta, \phi \rightarrow 0$ zároveň pak binomickému AR(1) procesu. Pro stacionární beta-binomický AR(2) proces s autokorelační funkcí $\rho(k)$ platí pro $k \in \mathbb{N}, k \geq 2$

$$\begin{aligned} \mathbb{E} X_t &= n\pi, \\ \text{var} X_t &= n\pi(1 - \pi) \cdot \frac{(1 - \phi)(1 + \rho) + n\phi \cdot [1 - 2\pi(1 - \pi)(1 - \rho)]}{(1 - \phi)(1 + \rho) + \phi \cdot [1 - 2\pi(1 - \pi)(1 - \rho)]}, \\ \rho(1) &= r, \quad \rho(k) \neq r^k, \end{aligned}$$

tyto vlastnosti jsou bez důkazu uvedeny v (Weiß a Kim, 2014). Tato alternativa již není pro simulace uvedené v (Weiß a Kim, 2015) vůbec uvažována.

V simulacích budeme opět kontrolovat autokorelaci prvního řádu nastavením parametru $r \in \{-0,25, 0,25, 0,50, 0,75\}$. Primárně se budeme zajímat, jestli bude síla VF testu vyšší proti alternativě kombinující změnu marginálního rozdělení a autokorelační struktury zároveň. U AF testů budeme kontrolovat opět pouze autokorelaci druhého řádu volbou $m = 2$. Zvolili jsme malé odchylky od binomického AR(1) procesu pomocí parametrů $\delta = 0,9$ a $\phi = 0,025$. Můžeme si všimnout, že dosažená síla AF testu v tabulce 4.12 odpovídá síle AF testu proti binomickému AR(2) procesu, změna marginálního rozdělení navíc nemá na tento test podle očekávání vliv. Totéž sledujeme i u MR testů ve srovnání se silou proti beta-binomickému AR(1) procesu v tabulce 4.5. Zde ovšem musíme brát v potaz, že neporovnáváme stejné alternativy, neboť v tabulce 4.5 jsme nastavovali parametr ρ a nyní kontrolujeme r . Pro $\delta = 0,9$ ale nejsou ve většině nastavení hodnoty r a ρ příliš vzdálené a můžeme tak obě tabulky rámcově srovnávat. Pokud se konečně zaměříme na VF testy, tak v porovnání s tabulkou 4.5 taktéž nepozorujeme žádné zlepšení. Test se tak nezdá citlivější proti alternativě přidávající navíc změnu autokorelační struktury oproti srovnatelné změně pouze marginálního rozdělení. Podobného výsledku jsme dosáhli i při vyšším vlivu autokorelace druhého řádu nastavením $\delta = 0,7$. V tabulce 4.12 vidíme výsledky simulací, dosažená síla MR a VF testů je opět téměř identická jako v tabulce 4.5 pro beta-binomický AR(1) proces.

4.4 Nesprávná velikost populace n

Na závěr budeme uvažovat ještě jednu alternativu odlišnou od předchozích. V literatuře se vždy předpokládá známá velikost populace n , jejíž počty v čase modelujeme binomickým AR(1) procesem. Zkoušeli jsme proto simulovat výsledky testů dobré shody pro binomický AR(1) proces se skutečnou hodnotou $n = 5$, o kterém ale mylně předpokládáme $n = 10$. Ostatní parametry byly voleny stejně jako při zkoumání empirické hladiny testů, přidali jsme navíc možnost $\pi = 0,9$. V tomto extrémním případě se plně projevilo, že proces ve skutečnosti nemůže nabýt hodnot $6, \dots, 10$. Střední hodnota takového procesu je $E X_t = n\pi = 4,5$ a případné binomické rozdělení s parametry $n = 10$ a $p = 0,45$ (tedy stejnou střední hodnotou) by přitom mělo nabývat těchto hodnot s pravděpodobností zhruba 26 %. MR test tak již pro $T = 100$ pozorování zamítá téměř všechny hypotézy binomického AR(1) procesu a podobně tomu je i pro VF testy. Korelační struktura procesu naproti tomu zůstává i při chybně specifikovaném n nezměněná a AF testy zamítají i ve zmíněném extrémním případě $\pi = 0,9$ jen lehce nad 5 % hypotéz binomického AR(1) procesu. Tabulky 4.13 a 4.14 shrnují počet zamítnutých testů pro všechna nastavení parametrů.

Podíl dobře definovaných odhadů								
Metoda	π	ρ	T					
			25	50	100	250	500	
CLS	0,1	-0,10	100,0	100,0	100,0	100,0	100,0	
		0	100,0	100,0	100,0	100,0	100,0	
		0,10	100,0	100,0	100,0	100,0	100,0	
		0,25	100,0	100,0	100,0	100,0	100,0	
		0,50	100,0	100,0	100,0	100,0	100,0	
		0,75	99,9	100,0	100,0	100,0	100,0	
		0,90	95,8	99,9	100,0	100,0	100,0	
		YW	0,1	-0,10	100,0	100,0	100,0	100,0
0	100,0	100,0		100,0	100,0	100,0		
0,10	100,0	100,0		100,0	100,0	100,0		
0,25	100,0	100,0		100,0	100,0	100,0		
0,50	100,0	100,0		100,0	100,0	100,0		
0,75	99,9	100,0		100,0	100,0	100,0		
0,90	96,2	99,9		100,0	100,0	100,0		

Tabulka 4.1: Procento dobře definovaných CLS a YW odhadů na základě 4000 realizací binomického AR(1) procesu s populací $n = 10$.

Podíl přípustných odhadů (z dobře definovaných)								
Metoda	π	ρ	T					
			25	50	100	250	500	
CLS	0,1	-0,10	43,5	49,3	51,2	56,0	59,9	
		0	63,1	73,0	85,0	96,0	99,3	
		0,10	77,6	89,7	97,4	99,8	100,0	
		0,25	90,8	98,3	100,0	100,0	100,0	
		0,50	98,8	100,0	100,0	100,0	100,0	
		0,75	99,1	100,0	100,0	100,0	100,0	
		0,90	95,1	98,7	100,0	100,0	100,0	
		YW	0,1	-0,10	44,4	49,8	51,9	56,2
0	64,1	73,7		85,3	96,1	99,3		
0,10	78,0	89,8		97,4	99,8	100,0		
0,25	91,2	98,3		100,0	100,0	100,0		
0,50	98,9	100,0		100,0	100,0	100,0		
0,75	99,6	100,0		100,0	100,0	100,0		
0,90	99,1	99,8		100,0	100,0	100,0		

Tabulka 4.2: Procento přípustných CLS a YW odhadů z dobře definovaných odhadů (viz tabulka 4.1) na základě 4000 realizací binomického AR(1) procesu s populací $n = 10$.

Binomický AR(1) model ($n = 10$)							
T	π	ρ	MR	MR adj	AF	VF	
100	0,1	0	1,9	9,7	5,3	5,0	
		0,25	1,0	5,8	4,6	4,9	
		0,50	0,7	1,9	3,7	5,5	
		0,75	0,9	1,8	3,5	5,7	
	0,5	-0,50	3,6	6,8	5,5	4,8	
		0	2,6	8,7	5,4	4,5	
		0,25	1,8	4,7	6,3	4,4	
		0,50	0,5	1,5	5,2	4,3	
		0,75	0,8	1,2	5,2	3,4	
	250	0,1	0	2,4	10,3	4,9	5,5
			0,25	1,3	4,4	5,0	4,5
0,50			0,7	1,7	3,9	5,5	
0,75			0,4	0,9	3,9	5,5	
0,5		-0,50	4,1	6,4	4,8	5,1	
		0	3,2	7,5	4,9	5,6	
		0,25	1,8	4,5	5,1	5,4	
		0,50	0,6	1,3	5,1	4,2	
		0,75	0,6	0,9	5,4	4,3	
500		0,1	0	2,2	8,7	5,1	4,8
			0,25	1,4	4,9	4,6	4,9
	0,50		1,1	2,1	3,7	5,3	
	0,75		0,5	1,0	3,6	5,3	
	0,5	-0,50	4,8	7,4	5,0	5,2	
		0	3,7	8,5	4,6	5,2	
		0,25	2,5	5,9	5,3	5,0	
		0,50	1,0	2,2	5,3	4,6	
		0,75	0,7	1,1	5,8	4,1	
	1000	0,1	0	3,1	9,7	4,8	5,6
			0,25	1,6	4,3	4,9	4,6
0,50			0,9	1,8	4,2	4,7	
0,75			0,8	1,2	4,1	5,4	
0,5		-0,50	5,1	7,2	4,6	5,4	
		0	4,4	9,0	4,3	5,1	
		0,25	2,6	5,1	5,0	5,4	
		0,50	1,8	3,1	5,1	5,2	
		0,75	1,2	1,4	5,2	5,2	

Tabulka 4.3: Hladina testů dobré shody – procento špatně zamítnutých hypotéz pro 4000 realizací binomického AR(1) modelu ($n = 10$).

Binomický AR(1) model ($n = 5$)							
T	π	ρ	MR	MR adj	AF	VF	
100	0,1	0	2,2	10,9	5,5	5,0	
		0,25	1,1	4,5	4,5	5,1	
		0,50	0,6	2,0	3,0	5,5	
		0,75	0,4	1,1	2,5	5,3	
	0,5	-0,50	3,9	6,6	5,1	5,6	
		0	2,4	7,7	5,0	4,9	
		0,25	1,2	4,3	5,4	4,3	
		0,50	0,7	2,2	5,4	5,0	
		0,75	1,0	1,6	5,3	3,8	
	250	0,1	0	2,1	12,1	4,5	4,9
			0,25	1,3	4,5	4,3	4,8
0,50			0,7	2,4	3,3	5,4	
0,75			0,5	1,4	2,6	4,8	
0,5		-0,50	4,4	7,0	5,4	5,1	
		0	2,3	8,8	5,3	5,5	
		0,25	1,5	4,6	5,2	5,4	
		0,50	0,7	1,9	5,1	5,6	
		0,75	0,5	0,9	4,7	4,0	
500		0,1	0	2,4	12,2	5,2	4,6
			0,25	1,4	5,5	4,5	5,6
	0,50		0,6	2,0	3,3	4,7	
	0,75		0,7	1,6	3,2	4,6	
	0,5	-0,50	4,5	7,3	4,9	4,5	
		0	3,1	10,5	4,7	5,3	
		0,25	1,3	4,3	5,6	5,7	
		0,50	0,6	1,3	5,1	4,7	
		0,75	0,5	0,9	5,0	4,7	
	1000	0,1	0	3,2	10,6	5,0	5,2
			0,25	1,7	4,6	4,7	4,4
0,50			1,4	2,4	3,5	4,8	
0,75			1,0	1,8	2,9	5,1	
0,5		-0,50	4,5	7,3	4,9	5,2	
		0	2,7	9,8	4,7	4,3	
		0,25	1,4	4,6	5,1	5,0	
		0,50	0,6	1,4	4,3	5,0	
		0,75	0,3	0,5	5,3	5,1	

Tabulka 4.4: Hladina testů dobré shody – procento špatně zamítnutých hypotéz ze 4000 realizací binomického AR(1) modelu ($n = 5$).

Beta-binomický AR(1) model ($n = 10$)								
ϕ	T	π	ρ	MR	MR adj	AF	VF	
0,025	100	0,1	0,25	5,3	13,4	4,8	14,6	
			0,50	1,3	3,5	4,0	7,8	
			0,75	0,9	1,6	3,8	4,0	
	0,5	-0,50	0,25	7,9	10,9	5,8	11,1	
			0,50	4,0	9,0	5,0	12,3	
			0,75	1,8	3,5	5,5	8,7	
			0,25	0,7	1,1	4,6	5,7	
			0,50	14,2	28,1	5,0	30,4	
			0,75	5,3	9,2	4,3	16,8	
	250	0,1	0,25	14,2	28,1	5,0	30,4	
			0,50	5,3	9,2	4,3	16,8	
			0,75	1,5	2,4	4,3	5,1	
0,5			-0,50	0,25	15,3	19,5	5,4	16,5
				0,50	9,4	15,8	5,4	20,3
				0,75	5,5	7,9	5,2	15,3
500	0,1	0,25	2,1	2,7	5,0	9,3		
		0,50	34,9	50,2	5,1	52,6		
		0,75	15,3	21,4	4,2	29,7		
		0,5	-0,50	0,25	4,4	6,0	4,8	10,0
				0,50	26,6	31,2	4,8	26,3
				0,75	17,4	27,3	5,4	32,2
1000	0,1	0,25	12,4	16,5	4,9	23,5		
		0,50	5,0	5,9	4,8	12,5		
		0,75	66,5	79,0	5,1	82,8		
		0,5	-0,50	0,25	34,1	44,2	4,9	54,3
				0,50	10,5	13,2	4,4	22,4
				0,75	43,0	48,2	5,3	41,1
0,5	-0,50	0,25	35,5	46,5	5,1	54,8		
		0,50	24,6	30,8	5,3	40,1		
		0,75	10,9	12,5	4,7	20,8		
		0,25	10,9	12,5	4,7	20,8		

Tabulka 4.5: Síla testů dobré shody – procento správně zamítnutých hypotéz ze 4000 realizací beta-binomického AR(1) modelu s disperzním parametrem $\phi = 0,025$.

Beta-binomický AR(1) model ($n = 10$)							
ϕ	T	π	ρ	MR	MR adj	AF	VF
0,050	100	0,1	0,25	18,9	34,3	5,5	39,7
			0,50	6,2	12,0	4,4	20,5
			0,75	1,5	3,1	4,1	6,0
		0,5	-0,50	18,9	24,0	5,1	21,9
			0,25	11,4	20,0	5,4	27,9
			0,50	6,4	9,7	5,1	19,4
	250	0,1	0,25	56,9	72,7	5,4	77,4
			0,50	26,7	36,6	4,2	48,1
			0,75	7,6	10,9	4,8	16,7
		0,5	-0,50	42,8	48,4	4,7	41,3
			0,25	36,2	47,9	4,7	54,7
			0,50	23,6	31,0	5,3	39,1
500	0,1	0,25	91,3	96,1	5,1	96,1	
		0,50	61,0	69,8	5,1	76,4	
		0,75	21,9	26,3	4,5	34,5	
	0,5	-0,50	68,8	73,5	4,7	67,2	
		0,25	67,5	77,0	5,9	81,0	
		0,50	50,0	57,7	5,6	65,2	
1000	0,1	0,25	99,8	99,9	5,1	100,0	
		0,50	93,6	96,1	5,0	97,3	
		0,75	52,2	57,8	4,7	65,2	
	0,5	-0,50	93,0	94,6	4,8	92,2	
		0,25	94,6	97,3	4,6	97,9	
		0,50	84,6	88,7	4,7	91,7	
		0,75	51,0	54,0	4,9	60,1	

Tabulka 4.6: Síla testů dobré shody – procento správně zamítnutých hypotéz ze 4000 realizací beta-binomického AR(1) modelu s disperzním parametrem $\phi = 0,050$.

Beta-binomický AR(1) model ($n = 10$)							
ϕ	T	π	ρ	MR	MR adj	AF	VF
0,100	100	0,1	0,25	60,6	77,2	5,0	82,8
			0,50	30,4	43,4	4,9	56,8
			0,75	8,3	12,5	4,7	16,9
		0,5	-0,50	52,6	59,9	5,9	57,3
			0,25	48,7	61,7	5,3	68,8
			0,50	32,1	40,5	5,2	53,6
	250	0,1	0,25	98,0	99,2	4,9	99,6
			0,50	82,4	88,3	5,1	92,4
			0,75	38,9	46,6	4,8	52,2
		0,5	-0,50	91,4	93,5	4,8	90,3
			0,25	92,2	95,5	5,3	96,8
			0,50	82,4	86,9	5,4	88,4
500	0,1	0,25	100,0	100,0	4,7	100,0	
		0,50	99,2	99,6	4,9	99,6	
		0,75	77,6	82,1	5,5	84,2	
	0,5	-0,50	99,8	99,9	5,4	99,6	
		0,25	99,9	100,0	5,4	100,0	
		0,50	99,0	99,4	5,1	99,6	
1000	0,1	0,25	100,0	100,0	4,7	100,0	
		0,50	100,0	100,0	5,3	100,0	
		0,75	98,7	99,1	5,2	98,9	
	0,5	-0,50	100,0	100,0	4,7	100,0	
		0,25	100,0	100,0	5,4	100,0	
		0,50	100,0	100,0	5,4	100,0	
		0,75	99,3	99,4	4,8	98,8	

Tabulka 4.7: Síla testů dobré shody – procento správně zamítnutých hypotéz ze 4000 realizací beta-binomického AR(1) modelu s disperzním parametrem $\phi = 0,100$.

Binomický AR(2) model ($n = 10$)							
δ	T	π	r	MR	MR adj	AF	VF
0,5	100	0,1	0,25	1,5	7,4	37,5	7,0
			0,50	3,6	8,1	76,0	12,5
			0,75	10,1	14,4	85,6	20,5
		0,5	-0,25	10,3	17,3	91,6	13,5
			0,25	2,1	5,8	41,0	4,9
			0,50	3,5	6,8	80,8	9,1
	250	0,1	0,25	2,2	6,9	80,3	7,8
			0,50	2,9	6,1	99,3	12,4
			0,75	8,1	11,1	99,9	20,1
		0,5	-0,25	11,2	16,9	100,0	15,3
			0,25	2,4	5,6	83,8	6,4
			0,50	4,2	6,7	99,8	9,9
500	1000	0,1	0,25	2,3	5,6	97,7	7,3
			0,50	3,9	6,6	100,0	12,5
			0,75	7,6	10,2	100,0	19,7
		0,5	-0,25	10,8	16,8	100,0	15,7
			0,25	3,5	6,8	99,1	6,7
			0,50	4,9	7,8	100,0	12,1
	1000	0,1	0,25	2,4	6,5	100,0	6,7
			0,50	4,7	7,4	100,0	13,7
			0,75	8,7	11,1	100,0	20,1
		0,5	-0,25	11,0	16,0	100,0	16,6
			0,25	3,5	7,1	100,0	6,8
			0,50	5,6	8,6	100,0	11,5
			0,75	12,1	14,0	100,0	17,7

Tabulka 4.8: Síla testů dobré shody – procento správně zamítnutých hypotéz ze 4000 realizací binomického AR(2) modelu s parametrem $\delta = 0,5$.

Binomický AR(2) model ($n = 10$)							
δ	T	π	r	MR	MR adj	AF	VF
0,7	100	0,1	0,25	1,3	6,2	11,7	5,6
			0,50	1,4	3,7	27,7	8,0
			0,75	4,5	6,8	39,7	12,2
		0,5	-0,50	50,4	57,0	78,9	34,3
			0,25	2,0	6,1	11,8	6,1
			0,50	1,3	2,8	31,4	5,5
	250	0,1	0,25	1,7	5,8	27,3	6,9
			0,50	1,6	3,5	66,2	8,2
			0,75	2,9	4,3	85,5	10,9
		0,5	-0,50	53,6	58,8	99,2	40,0
			0,25	2,2	5,2	28,2	5,6
			0,50	1,6	3,2	71,6	7,0
500	1000	0,1	0,25	1,7	5,1	51,3	5,8
			0,50	1,6	3,2	93,5	8,0
			0,75	2,7	3,8	99,1	11,0
		0,5	-0,50	56,0	61,1	100,0	40,5
			0,25	2,7	5,7	54,6	5,9
			0,50	2,1	3,6	95,2	7,4
	1000	0,1	0,25	2,0	6,1	81,1	6,1
			0,50	1,7	3,1	99,8	7,9
			0,75	2,6	3,9	100,0	11,8
		0,5	-0,50	57,6	61,5	100,0	44,0
			0,25	3,0	6,1	83,9	5,9
			0,50	3,2	4,5	99,9	8,0
		0,75	4,9	5,8	100,0	10,5	

Tabulka 4.9: Síla testů dobré shody – procento správně zamítnutých hypotéz ze 4000 realizací binomického AR(2) modelu s parametrem $\delta = 0,7$.

Binomický AR(2) model ($n = 10$)							
δ	T	π	r	MR	MR adj	AF	VF
0,9	100	0,1	0,25	1,1	5,7	4,9	5,5
			0,50	0,7	2,2	5,8	5,3
			0,75	1,8	2,8	5,7	7,1
		0,5	-0,50	4,7	7,3	11,9	5,8
			0,25	1,9	5,5	5,0	4,8
			0,50	0,9	2,4	6,4	4,7
	250	0,1	0,25	1,6	4,9	6,5	5,4
			0,50	0,4	1,5	8,8	5,8
			0,75	1,0	1,7	14,2	7,4
		0,5	-0,50	5,9	8,6	18,7	6,6
			0,25	2,3	5,1	5,8	5,5
			0,50	0,8	2,0	11,1	5,7
500	0,1	0,25	1,8	5,3	8,3	5,3	
		0,50	1,0	2,2	15,5	5,1	
		0,75	1,1	1,5	28,6	6,5	
		0,5	-0,50	6,4	9,2	29,5	6,5
			0,25	3,0	5,6	8,1	5,2
			0,50	1,4	2,7	19,7	5,5
	1000	0,1	0,25	1,6	4,8	11,9	5,0
			0,50	1,1	2,3	30,6	5,8
			0,75	1,1	1,5	53,4	6,4
		0,5	-0,50	7,5	9,8	49,6	6,8
			0,25	2,6	6,0	12,2	5,6
			0,50	1,7	2,7	35,7	6,2
		0,75	1,7	2,1	63,0	6,6	

Tabulka 4.10: Síla testů dobré shody – procento správně zamítnutých hypotéz ze 4000 realizací binomického AR(2) modelu s parametrem $\delta = 0,9$.

Beta-binomický AR(2) model ($n = 10$)											
ϕ	δ	T	π	ρ	MR	MR adj	AF	VF			
0,025	0,70	100	0,1	0,25	5,7	13,1	12,1	14,4			
				0,50	2,5	6,1	26,4	10,6			
				0,75	3,7	6,4	41,1	9,6			
			0,5	-0,50	53,8	59,8	80,0	35,0			
				0,25	5,2	10,0	10,6	12,5			
				0,50	3,8	5,9	32,1	11,1			
			0,75	0,25	4,2	6,0	52,9	9,7			
				250	0,1	0,25	13,7	26,1	26,4	29,0	
						0,50	6,9	11,6	65,7	17,8	
			0,75			4,3	6,5	82,9	11,3		
			0,5	-0,50	59,5	64,2	99,1	41,4			
					0,25	10,8	18,0	28,3	20,2		
		0,50			7,8	11,1	73,5	16,3			
		0,75	0,25	6,6	7,8	92,4	12,3				
			500	0,1	0,25	31,6	46,7	50,2	49,9		
					0,50	16,8	22,9	92,3	29,1		
		0,75			9,1	11,6	99,0	16,2			
		0,5	-0,50	64,2	68,2	100,0	46,9				
				0,25	19,5	29,5	53,5	34,2			
				0,50	15,1	20,0	95,4	25,1			
		0,75	0,25	10,8	12,6	99,9	16,5				
			1000	0,1	0,25	61,9	75,1	81,9	78,3		
					0,50	34,5	42,8	99,9	51,2		
		0,75			16,6	19,9	100,0	25,5			
0,5	-0,50	69,0	71,9	100,0	51,6						
		0,25	36,7	47,6	84,2	54,2					
		0,50	27,5	34,6	99,9	39,9					
0,75	0,25	19,1	21,8	100,0	24,1						

Tabulka 4.11: Síla testů dobré shody – procento správně zamítnutých hypotéz ze 4000 realizací beta-binomického AR(2) modelu s disperzním parametrem $\phi = 0,025$ a parametrem $\delta = 0,7$.

Beta-binomický AR(2) model ($n = 10$)											
ϕ	δ	T	π	ρ	MR	MR adj	AF	VF			
0,025	0,90	100	0,1	0,25	5,4	13,7	4,7	14,6			
				0,50	1,7	3,8	5,2	8,4			
				0,75	1,5	3,0	6,1	5,7			
			0,5	-0,50	10,9	14,4	13,3	12,5			
				0,25	4,4	9,2	5,2	13,0			
				0,50	2,5	4,2	6,3	9,9			
			0,75	1,5	2,1	9,1	6,1				
				250	0,1	0,25	14,7	27,1	6,7	30,2	
						0,50	6,0	10,2	9,2	17,3	
			0,75			2,4	3,6	15,3	6,9		
			0,5	-0,50	18,5	22,4	18,2	17,5			
				0,25	9,2	16,7	6,7	19,7			
		0,50		5,3	8,0	11,0	13,6				
		0,75	2,9	3,6	19,2	9,7					
			500	0,1	0,25	34,7	50,6	7,1	54,8		
					0,50	14,8	20,8	16,8	29,7		
		0,75			5,8	7,7	28,6	11,2			
		0,5	-0,50	27,0	31,7	28,8	27,8				
			0,25	19,4	28,7	9,0	34,4				
			0,50	11,9	16,6	19,8	23,0				
		0,75	6,8	8,0	36,8	13,9					
			1000	0,1	0,25	63,9	77,4	13,3	80,8		
					0,50	34,3	43,5	32,6	53,4		
		0,75			13,5	16,7	54,5	23,1			
0,5	-0,50	44,1	48,9	48,5	42,8						
	0,25	36,5	48,1	12,2	54,8						
	0,50	25,9	32,1	35,8	40,5						
0,75	13,0	14,7	65,0	21,3							

Tabulka 4.12: Síla testů dobré shody – procento správně zamítnutých hypotéz ze 4000 realizací beta-binomického AR(2) modelu s disperzním parametrem $\phi = 0,025$ a parametrem $\delta = 0,9$.

Binomický AR(1) model (nesprávné n)							
T	π	ρ	MR	MR adj	AF	VF	
100	0,1	0	1,8	10,6	5,5	5,7	
		0,25	0,7	5,0	4,1	6,3	
		0,50	0,7	2,3	2,4	7,3	
		0,75	0,7	1,9	1,8	7,0	
	0,5	-0,50	29,4	39,6	4,6	25,6	
		0	27,9	53,4	4,9	47,0	
		0,25	21,8	41,3	5,1	41,9	
		0,50	14,3	23,2	5,0	31,5	
		0,75	9,3	12,6	4,5	16,7	
	0,9	0	100,0	100,0	5,1	100,0	
		0,25	100,0	100,0	6,4	100,0	
		0,50	100,0	100,0	7,5	98,9	
		0,75	97,6	98,2	7,5	63,6	
	250	0,1	0	2,0	15,1	4,6	8,9
			0,25	1,3	6,4	4,0	8,6
			0,50	0,7	2,9	2,9	9,4
0,75			0,6	1,6	2,1	7,7	
0,5		-0,50	86,3	91,2	4,6	58,7	
		0	92,2	98,6	5,2	89,0	
		0,25	83,0	93,0	4,9	83,7	
		0,50	55,5	66,6	4,7	65,0	
		0,75	20,9	25,5	4,3	37,7	
0,9		0	100,0	100,0	5,2	100,0	
		0,25	100,0	100,0	5,7	100,0	
		0,50	100,0	100,0	7,2	100,0	
		0,75	100,0	100,0	7,9	99,8	

Tabulka 4.13: Procento zamítnutých hypotéz ze 4000 realizací binomického AR(1) modelu se skutečným parametrem $n = 5$ a odhadovaným jako by $n = 10$ pro $T \in \{100, 250\}$.

Binomický AR(1) model (nesprávné n)							
T	π	ρ	MR	MR adj	AF	VF	
500	0,1	0	2,1	13,5	5,2	11,5	
		0,25	1,9	7,9	4,1	12,3	
		0,50	0,7	2,6	2,7	11,3	
		0,75	0,6	1,4	2,4	9,1	
	0,5	-0,50	99,9	100,0	4,1	87,7	
		0	100,0	100,0	4,7	99,6	
		0,25	99,9	100,0	5,3	98,7	
		0,50	95,2	97,6	4,8	92,2	
		0,75	48,8	53,9	4,5	61,0	
	0,9	0	100,0	100,0	4,9	100,0	
		0,25	100,0	100,0	6,1	100,0	
		0,50	100,0	100,0	7,1	100,0	
		0,75	100,0	100,0	7,1	100,0	
	1000	0,1	0	4,8	17,9	5,0	21,9
			0,25	2,7	9,0	4,3	20,1
			0,50	1,3	3,2	2,7	15,9
0,75			0,7	1,5	2,4	11,9	
0,5		-0,50	100,0	100,0	3,9	99,3	
		0	100,0	100,0	4,7	100,0	
		0,25	100,0	100,0	4,8	100,0	
		0,50	100,0	100,0	3,9	99,8	
		0,75	92,9	94,6	4,5	87,0	
0,9		0	100,0	100,0	5,0	100,0	
		0,25	100,0	100,0	6,1	100,0	
		0,50	100,0	100,0	6,5	100,0	
		0,75	100,0	100,0	8,3	100,0	

Tabulka 4.14: Procento zamítnutých hypotéz ze 4000 realizací binomického AR(1) modelu se skutečným parametrem $n = 5$ a odhadovaným jako by $n = 10$ pro $T \in \{500, 1000\}$.

5. Aplikace na reálná data

V článku (Weiß a Kim, 2013a) analyzují autoři počty obchodovaných titulů na korejské burze cenných papírů v pětiminutových intervalech, které modelují binomickým AR(1) procesem ($n = 22, T = 70$). V podobném duchu se tak pokusíme aplikovat tento model i na pražskou burzu.

Index PX se v týdnu od 12. 10. do 16. 10. 2020 skládal z 11 společností obchodovaných na Burze cenných papírů Praha. Kontinuální obchodování na této burze probíhá od 9:00 do 16:20, toto období jsme rozdělili do intervalů po 5 minutách (uvažovali jsme zprava uzavřené intervaly s přesností časových údajů na celé sekundy). Z analýzy jsme se rozhodli vynechat první a poslední interval dne, které se viditelně odlišovaly větším množstvím obchodů oproti zbytku dne. Dostáváme tak 86 intervalů, pro které označíme X_t počet společností z indexu PX, u jejichž akcií proběhl v t -tém intervalu alespoň jeden obchod na burze. Připravená data jsou k dispozici ke stažení na <https://labk10.karlin.mff.cuni.cz/~hledikja/bcpp.RData>. Počty obchodů jsme získali za období jednoho týdne, zkusíme tak použít binomický AR(1) model v rámci jednotlivých dnů ($T = 86$) a také souhrnně na celý týden ($T = 430$). Uvažujeme tak celkem šest hypotéz:

$H_{i,0}$: Data z i -tého dne jsou realizací stacionárního binomického AR(1) procesu s neznámými parametry (ρ_i, π_i) pro $i = 1, \dots, 5$ a

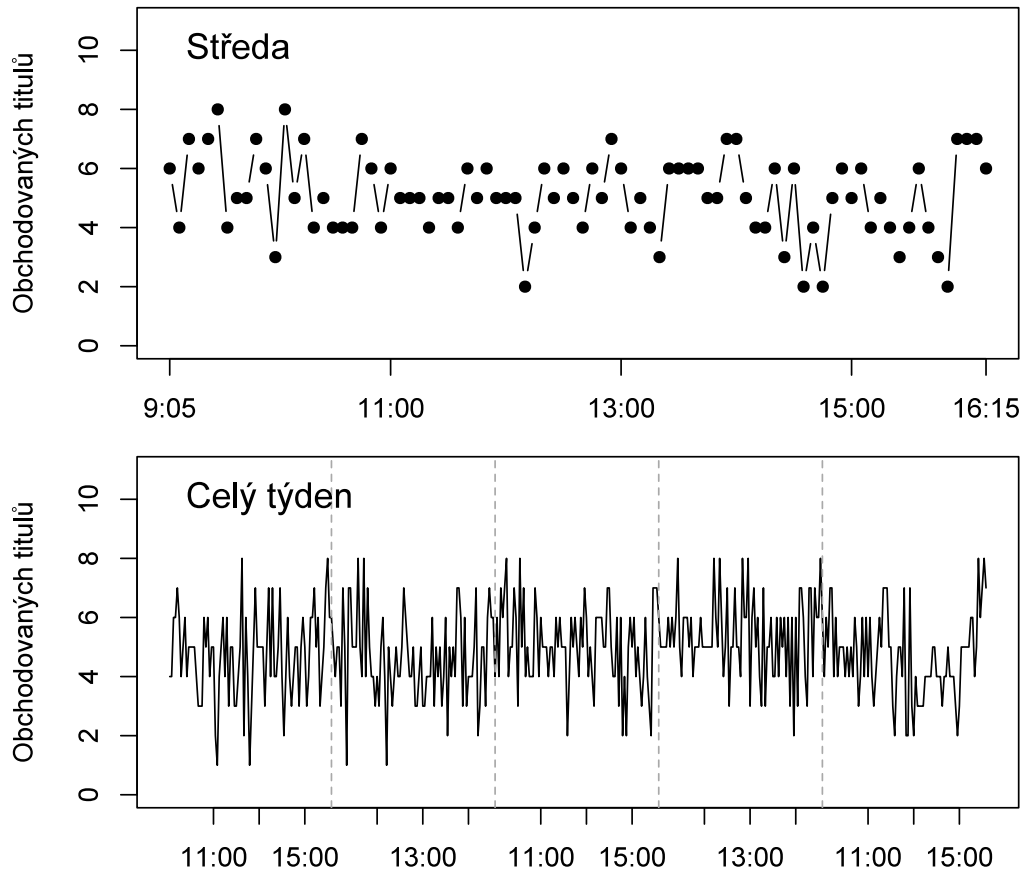
H_0 : Data za celý týden jsou realizací stacionárního binomického AR(1) procesu s neznámými parametry (ρ, π) .

V tabulce 5.1 uvádíme popisné charakteristiky všech šesti uvažovaných časových řad. V každém pětiminutovém intervalu proběhl alespoň 1 obchod a maximálně se obchodovalo s akciemi 8 společností. Nejčastěji se obchodovalo 4 nebo 5 titulů, ve středu a zejména ve čtvrtek je průměrný počet v rámci jednoho intervalu vyšší než ve zbývajících dnech. Na obrázku 5.1 můžeme vidět průběh počtu obchodovaných společností v čase pro středu a celý týden.

Období	Min	Q1	Medián	Q3	Max	Modus	Průměr	Sm.odch.
Pondělí	1	4	5,0	6	8	5	4,72	1,46
Úterý	1	4	4,5	5	8	4	4,60	1,46
Středa	2	4	5,0	6	8	5	5,08	1,37
Čtvrtek	2	5	5,0	6	8	5	5,38	1,33
Pátek	2	4	4,5	5	8	4	4,65	1,41
Týden	1	4	5,0	6	8	5	4,89	1,43

Tabulka 5.1: Výběrové charakteristiky počtů obchodů v jednotlivých obdobích.

Nejprve budeme zkoumat marginální rozdělení procesů. Na obrázku 5.2 jsme vykreslili empirické četnosti jednotlivých stavů (počtů obchodovaných společností) v porovnání s teoretickými četnostmi pro střeďeční data. Jako teoretické rozdělení jsme použili binomické rozdělení se střední hodnotou rovnající se výběrovému průměru z tabulky 5.1. Jak si můžeme všimnout, empirické marginální



Obrázek 5.1: Počty obchodovaných společností z indexu PX v čase.

rozdělení je o něco špičatější než teoretické binomické. Tyto rozdíly ovšem nejsou podle MR testu dobré shody statisticky významné. Tento test s p-hodnotou 0,345 nezamítá hypotézu $H_{3,0}$ binomického AR(1) procesu, p-hodnoty pro ostatní dny uvádíme v tabulce 5.2. V každém z jednotlivých dnů nezamítáme na základě MR testu příslušnou hypotézu $H_{i,0}$ binomického AR(1) procesu. Podobné rozdíly vidíme i na obrázku 5.3 za celý týden. Vzhledem k většímu počtu pozorování už ale vychází p-hodnota MR testu 0,057, tedy na hraně zamítnutí. S ohledem na zjištění, jak moc se ukázal MR test konzervativní v provedených simulacích, tak máme pochybnosti o vhodnosti binomického rozdělení pro marginální rozdělení počtu obchodovaných společností v celém týdnu.

Druhý pohled bude směřovat na autokorelační strukturu procesů. Na obrázku 5.4 můžeme vidět výběrovou parciální autokorelační funkci pro střední data. Modrými přerušovanými čarami je pro každý řád autokorelace k znázorněn očekávaný interval s 95% pravděpodobností podle rovnosti (3.1). Můžeme si všimnout, že v naší situaci je $\sigma_k \approx 1$ pro $k = 2, 3, \dots$ a intervaly tak mají přibližně konstantní šířku $2 \cdot \frac{u_{1-\alpha/2}}{\sqrt{T}}$, kde $u_{1-\alpha/2}$ je $1 - \alpha/2$ kvantil normovaného normálního rozdělení. Vidíme také, že střední pozorování vykazují jen slabou autokorelovanost řádu 1. Pro všechna $k = 2, 3, \dots, 20$ pak leží výběrová autokorelační funkce $\hat{\rho}_{part}$ ve vyznačeném rozmezí a nemáme tedy důvod uvažovat o autokorelaci až do řádu 20. Podobně tomu je i pro ostatní dny, v tabulce 5.2 uvádíme p-hodnoty pro AF testy s třemi volbami maximálního testovaného řádu autokorelace m . Ani jeden test by nevedl k zamítnutí příslušné hypotézy $H_{i,0}$ binomického AR(1) pro-

cesu. Pro období za celý týden se zdají být významné autokorelace řádu 7 a 8, kontrolování autokorelace do řádu 8 ($m = 8$) by vedlo k zamítnutí hypotézy H_0 binomického AR(1) procesu s p-hodnotou 0,011.

Konečně třetí pohled bude ověřovat vytvořující funkci procesu navrženým VF testem. Opět si nejprve znázorníme situaci graficky, u tohoto testu ale nebudou tyto grafy tolik vypovídající jako u předchozích testů. Uvádíme je tak spíše pro ilustraci fungování VF testu. Na obrázku 5.6 vidíme porovnání neparametrické empirické vytvořující funkce se semiparametrickým odhadem za platnosti modelu podle rovnosti (3.3) a čtverec jejich rozdílů, ze kterého je počítána testová statistika VF testu. Statistickou významnost testové statistiky pro středěční data vidíme na obrázku 5.8, s p-hodnotou 0,096 nezamítáme hypotézu $H_{3,0}$ binomického AR(1) modelu. Stejně tomu je i pro ostatní dny, dosažené p-hodnoty VF testů pro ostatní dny shrnuje tabulka 5.2. Pro čtvrtěční data ovšem vyšla p-hodnota 0,065, tedy jen těsně nad hranicí zamítnutí. Naproti tomu pro celotýdenní data s p-hodnotou 0,007 už zamítáme hypotézu H_0 binomického AR(1) modelu, jak ilustruje obrázek 5.9 a rozdíly odhadů vytvořujících funkcí na obrázku 5.7, které jsou sice menší než v případě středěčních dat, ale s ohledem na počet pozorování jsou významnější.

Období	MR	MR.adj	AF ($m = 2$)	AF ($m = 5$)	AF ($m = 8$)	VF
Pondělí	0,842	0,647	0,309	0,692	0,197	0,512
Úterý	0,350	0,166	0,698	0,327	0,359	0,235
Středa	0,345	0,157	0,360	0,320	0,611	0,096
Čtvrtek	0,132	0,058	0,952	0,942	0,976	0,065
Pátek	0,468	0,293	0,267	0,095	0,127	0,168
Týden	0,057	0,023	0,125	0,171	0,011	0,007

Tabulka 5.2: P-hodnoty testů dobré shody pro jednotlivá období.

Jelikož denní data nejsou v rozporu s hypotézami $H_{i,0}$ binomického AR(1) modelu, můžeme pro ně odhadnout parametry (ρ, π) . Na obrázku 5.10 vidíme graficky znázorněné bodové odhady pomocí tří představených metod spolu s konfidenčními množinami. Bodové odhady parametrů pak pro všechny dny uvádíme v tabulce 5.3, marginální intervaly spolehlivosti poté v tabulce 5.4. Můžeme si všimnout, že zatímco odhady parametru π jsou nejen pro všechny metody, ale také přes jednotlivé dny velmi podobné, odhady parametru ρ se liší u některých dnů i v rámci jednotlivých metod. To by mohlo být způsobeno tím, že binomický AR(1) model v těchto dnech tolik neodpovídá datům. Yule-Walkerův odhad totiž nevyžaduje žádnou strukturu modelu a uvažuje pouze empirickou autokorelaci na rozdíl od odhadu metodou maximální věrohodností. Dále vidíme, že marginální intervaly spolehlivosti pro ρ jsou výrazně širší než pro π . To plyne z asymptotické varianční matice pro CLS (a tedy i YW) odhad, konkrétně z rovnosti (2.3). Pro ρ malé tak například platí $\text{var}(\hat{\rho}_{CLS}) \approx 1/T$, zatímco pro $\pi = 0,5$ je marginální rozptyl odhadu $\hat{\pi}_{CLS}$ čtvrtinový. Všimněme si také, že hodnota $\rho = 0$ leží v intervalech spolehlivosti pro první čtyři dny, pro které tak nelze vyloučit ani nekorelovanost jednotlivých pozorování v čase.

Shrňme nyní výsledky z této kapitoly. Z celotýdenního pohledu se binomický AR(1) proces nezdá být vhodným modelem, a to jak z pohledu autokorelační

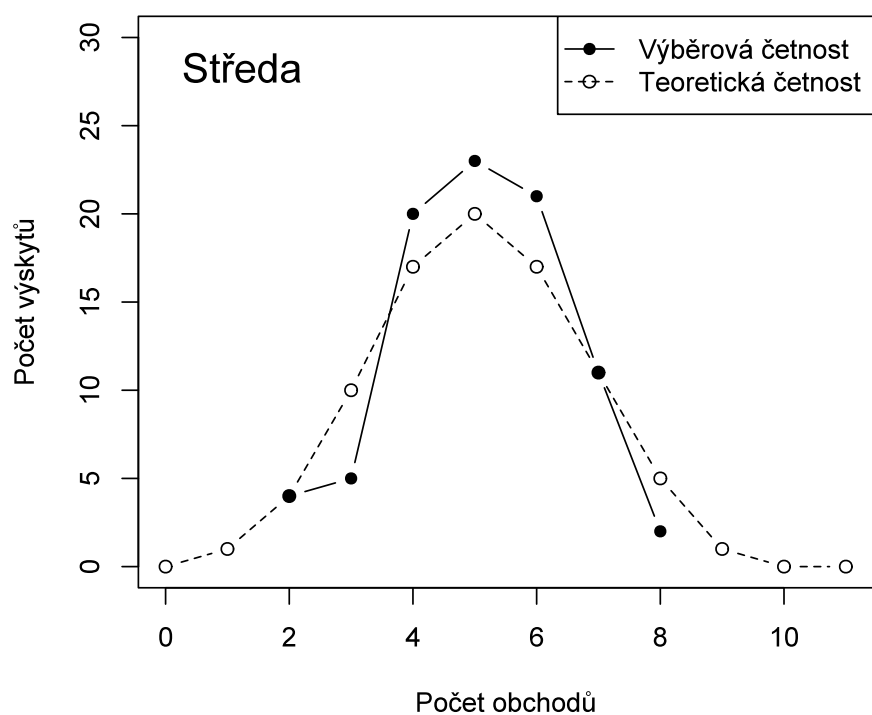
struktury, tak i marginálního rozdělení, které je špičatější než očekávané binomické. Z obrázku 5.1 se také můžeme domnívat, že by nemusel být splněn předpoklad stacionarity, počty obchodovaných společností se zdají být například o něco vyšší ke konci obchodních dnů. Chování procesu by se také nemohlo za platnosti hypotézy H_0 lišit pro jednotlivé dny v týdnu. Z pohledu na jednotlivé dny jsme nenalezli dostatečné důkazy k zamítnutí hypotézy binomického AR(1) procesu uvažovaného pro každý den zvlášť. Pokud bychom věřili předpokladu stacionarity uvnitř jednoho dne, mohl by tak být tento model vhodný. Při pohledu na odhad parametrů ρ by pak dokonce pro některé dny mohlo být možné zjednodušit model na nekorelovaná pozorování v čase.

Období	ρ			π		
	MLE	CLS	YW	MLE	CLS	YW
Pondělí	0,26	0,20	0,19	0,43	0,43	0,43
Úterý	0,01	0,01	0,01	0,42	0,42	0,42
Středa	0,18	0,10	0,10	0,46	0,46	0,46
Čtvrtek	-0,32	-0,18	-0,18	0,49	0,49	0,49
Pátek	0,52	0,38	0,37	0,43	0,43	0,42

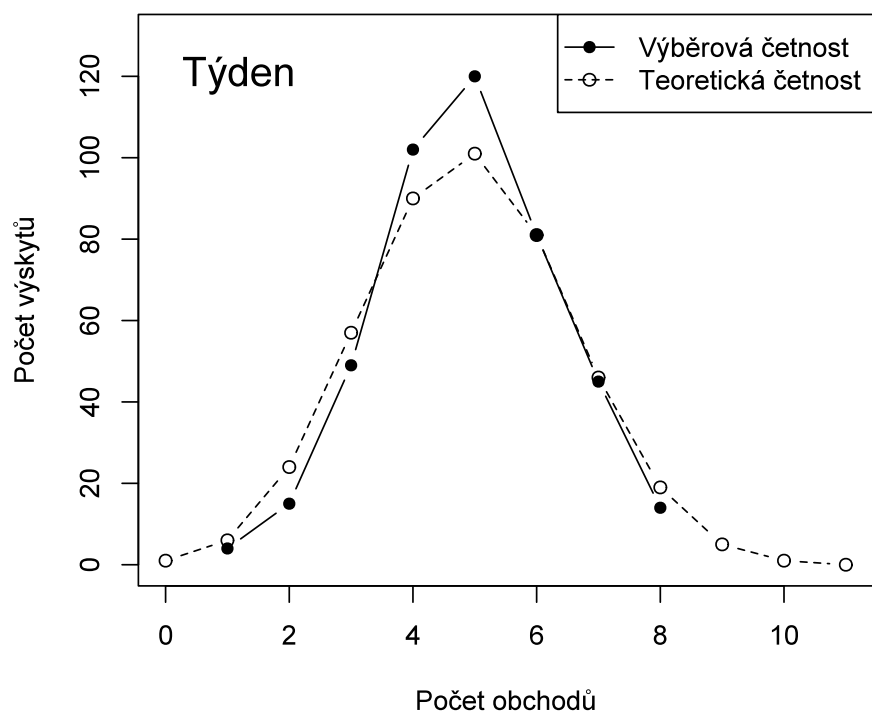
Tabulka 5.3: Bodové odhady parametrů po dnech.

Parametr	Období	MLE		CLS		YW	
		D	H	D	H	D	H
ρ	Pondělí	-0,04	0,43	-0,01	0,40	-0,01	0,40
	Úterý	-0,26	0,28	-0,20	0,22	-0,20	0,22
	Středa	-0,22	0,42	-0,11	0,31	-0,11	0,31
	Čtvrtek	-0,44	0,09	-0,39	0,03	-0,39	0,03
	Pátek	0,22	0,55	0,19	0,58	0,18	0,57
π	Pondělí	0,39	0,47	0,39	0,47	0,39	0,47
	Úterý	0,39	0,45	0,39	0,45	0,39	0,45
	Středa	0,42	0,50	0,43	0,50	0,43	0,50
	Čtvrtek	0,47	0,51	0,46	0,52	0,46	0,52
	Pátek	0,37	0,48	0,38	0,47	0,38	0,47

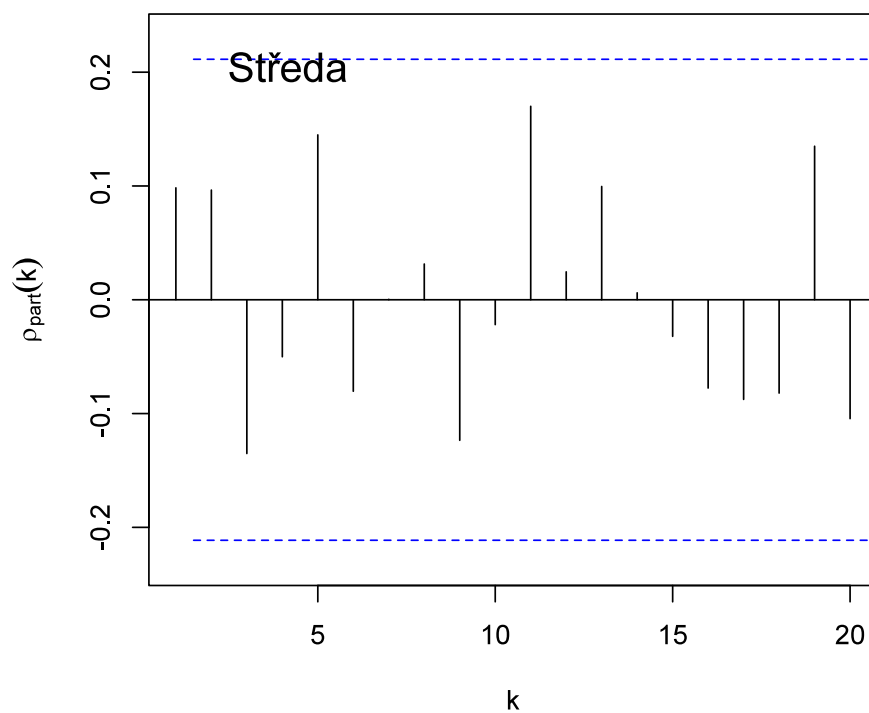
Tabulka 5.4: Dolní (D) a horní (H) hranice marginálních 95% intervalů spolehlivosti pro (ρ, π) po jednotlivých dnech.



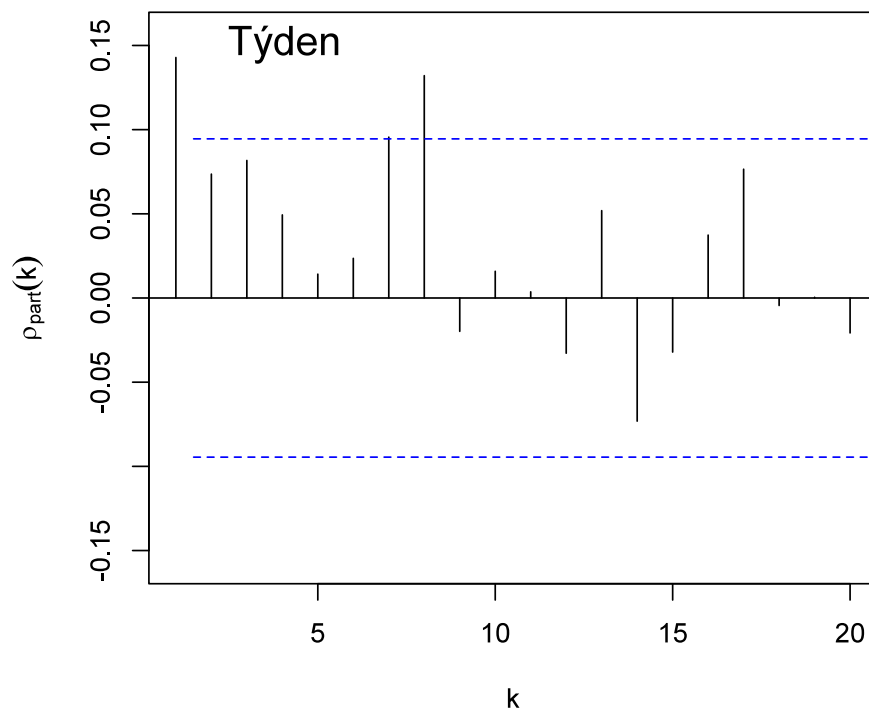
Obrázek 5.2: Empirické a teoretické rozdělení počtu obchodovaných společností ve středu.



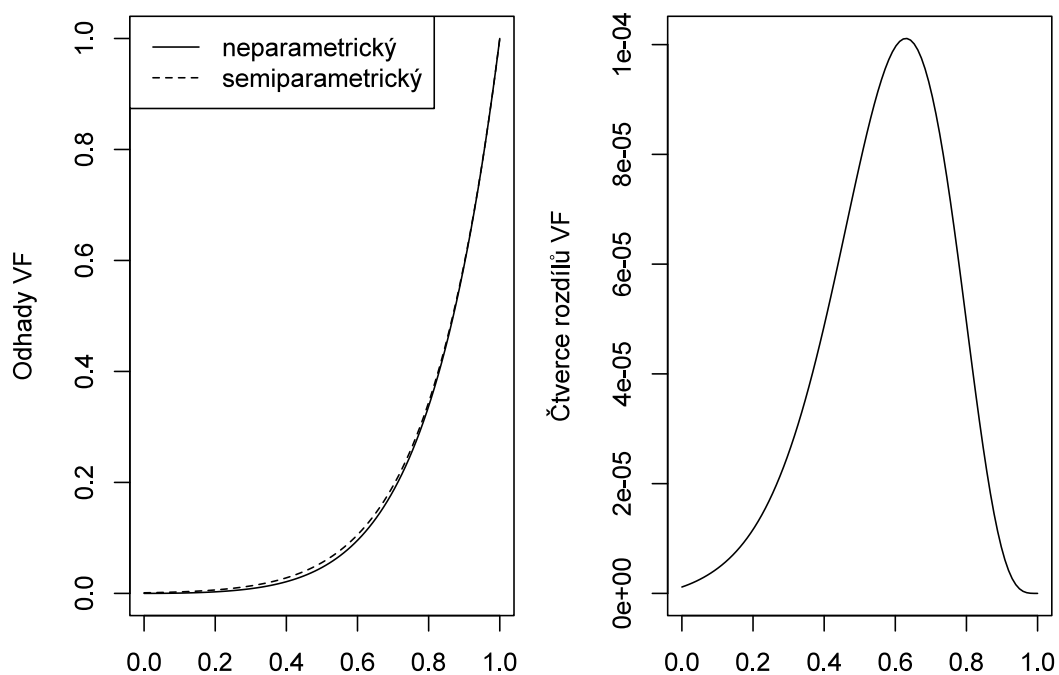
Obrázek 5.3: Empirické a teoretické rozdělení počtu obchodovaných společností za celý týden.



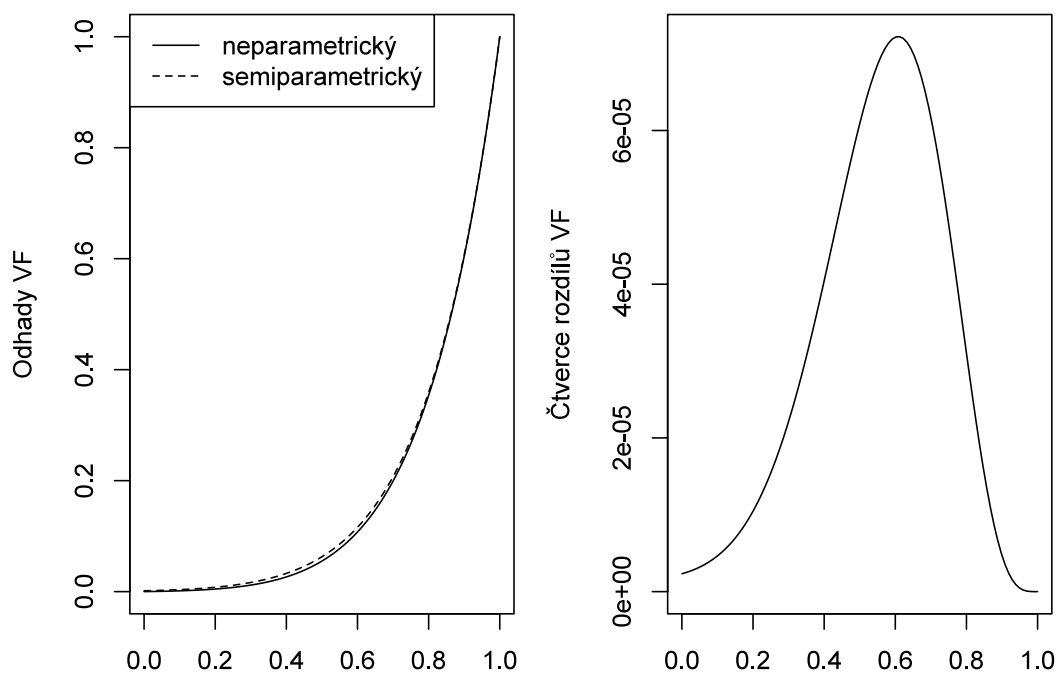
Obrázek 5.4: Výběrová parciální autokorelační funkce pro středeční data.



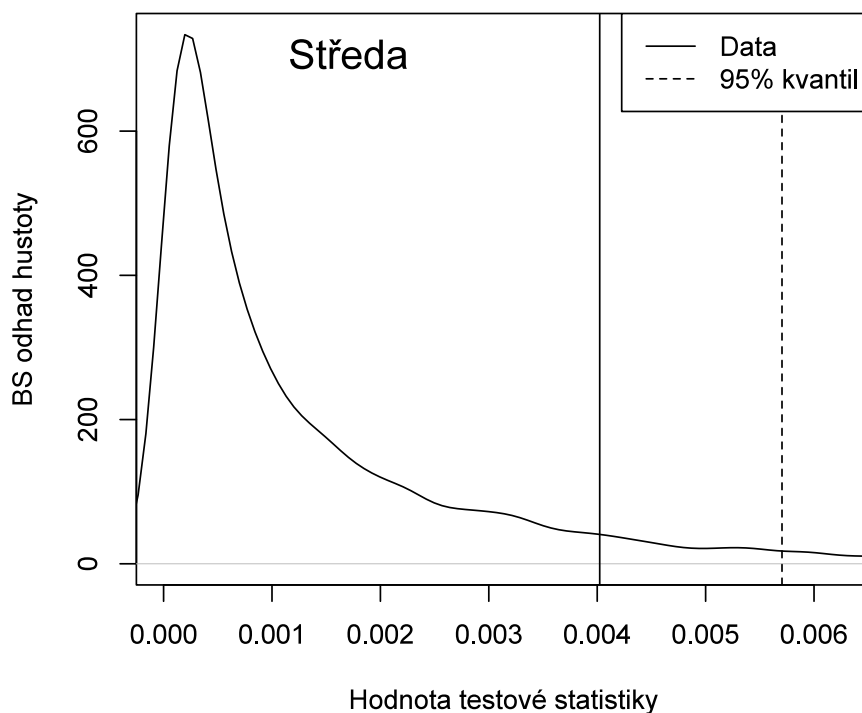
Obrázek 5.5: Výběrová parciální autokorelační funkce pro celotýdenní data.



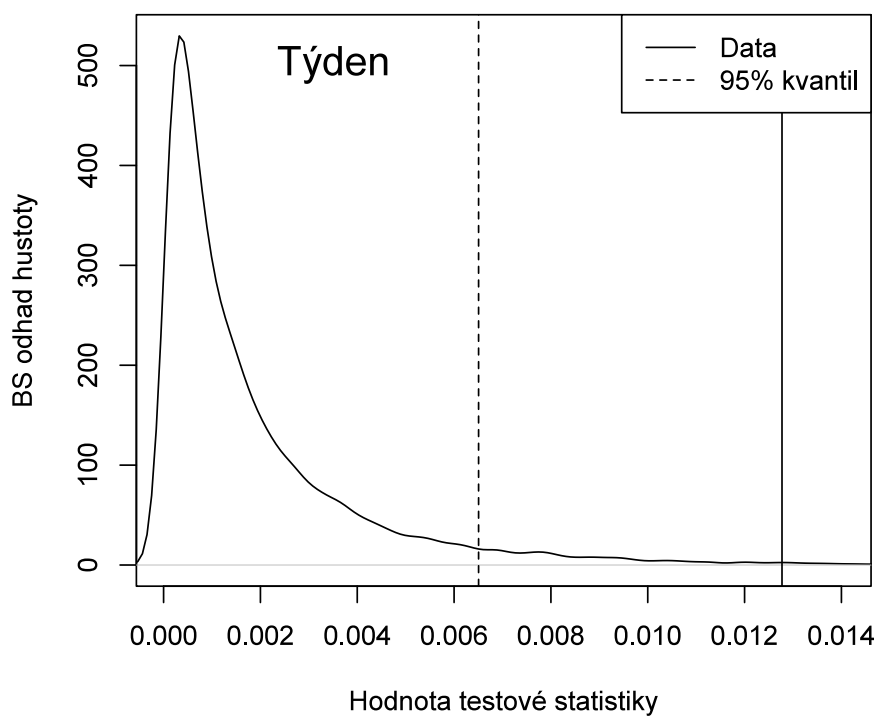
Obrázek 5.6: Rozdíly mezi neparametrickým odhadem vytvořující funkce pro střední data a semiparametrickým odhadem za platnosti binomického AR(1) modelu.



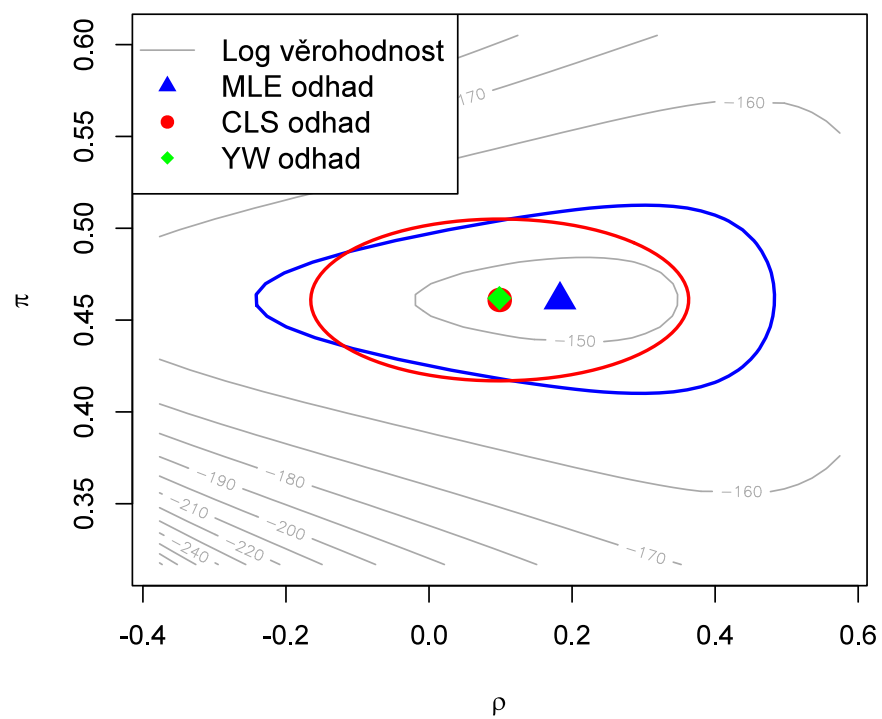
Obrázek 5.7: Rozdíly mezi neparametrickým odhadem vytvořující funkce pro týdenní data a semiparametrickým odhadem za platnosti binomického AR(1) modelu.



Obrázek 5.8: Bootstrapový odhad ($B = 10001$) rozdělení testové statistiky VF testu pro středeční data, svislou čarou vyznačena hodnota naměřená hodnota a přerušovaně vyznačen 95% kvantil.



Obrázek 5.9: Bootstrapový odhad ($B = 10001$) rozdělení testové statistiky VF testu pro celotýdenní data, svislou čarou vyznačena hodnota naměřená hodnota a přerušovaně vyznačen 95% kvantil.



Obrázek 5.10: Odhady parametrů pro střední data spolu s 95% konfidenčními oblastmi (Waldova typu pro CLS a na základě poměru věrohodnosti pro MLE odhad).

Závěr

V této práci jsme se zabývali jedním z možných modelů pro celočíselné časové řady s diskrétním časem a konečnou množinou stavů. Jedná se o model s binomickým marginálním rozdělením a autokorelační strukturou podobnou standardnímu AR(1) procesu, který by mohl být vhodným kandidátem pro modelování korelovaných pozorování nabývajících pouze omezeného množství hodnot na množině $\{0, 1, \dots, n\}$.

Nejprve jsme ukázali podrobná odvození některých základních vlastností tohoto procesu. Dále jsme z literatury shrnuli tři možné metody odhadů parametrů a rozepsali podrobněji některá odvození a důkazy. Nakonec jsme se věnovali otázce identifikace, jestli je binomický AR(1) proces vhodným modelem pro data, pomocí tří testů dobré shody. První dva přístupy (MR, AF) jsme převzali z (Weiß a Kim, 2015) a následně navrhli vlastní test založený na vytvářící funkci procesu (VF).

Tyto testy dobré shody jsme následně zkoumali v simulační studii. Nejprve jsme ověřovali, jestli testy dodržují předepsanou hladinu. Ukázalo se, že dosažená hladina u VF testu je blízko 5 % pro různé rozsahy výběrů. AF test byl také dostatečně blízko 5 %, ovšem pro některé hodnoty parametrů vycházel lehce konzervativní, výrazněji to bylo vidět pro $n = 5$. MR test byl podle očekávání velmi konzervativní a jeho korekce, kterou jsme navrhli a vyzkoušeli, naopak vůbec nedodržovala předepsanou hladinu a její použití bychom nedoporučili.

Dále jsme zjišťovali sílu jednotlivých testů proti různým alternativám. Žádný z uvažovaných testů dobré shody se neukázal být univerzální proti všem alternativám. Námí navržený VF test se ukázal podobný MR testu a byl citlivý zejména na porušení modelu z pohledu marginálního rozdělení. Pro většinu nastavení parametrů se ukázal silnější než MR test napříč různými rozsahy výběru. Na druhou stranu byl ovšem slabší proti výrazněji záporně korelovaným procesům.

V závěrečné kapitole jsme pak ukázali možnost aplikace binomického AR(1) modelu na finanční data. Analyzovali jsme časovou řadu počtu obchodovaných společností na pražské burze po 5 minutách, která nabývá hodnot v množině $\{0, 1, \dots, 11\}$. Z pohledu denních dat jsme pomocí testů dobré shody nenarazili na významný problém a použití binomického AR(1) modelu by tak mohlo být pro daná data vhodné. Předvedli jsme také bodové a intervalové odhady odvozené v kapitole 2 spolu s jejich grafickým znázorněním.

Zkoumaná problematika nabízí prostor pro další výzkum. Testovou statistiku námí navrženého VF testu by bylo možné dále rozšířit uvažováním vhodné váhové funkce $w(s) : (0, 1) \rightarrow \mathbb{R}$ do tvaru

$$S_T(\mathbf{X}) = \frac{1}{T} \int_0^1 w(s) \left\{ \sum_{t=1}^T \left[s^{X_t} - (s \hat{\alpha} + 1 - \hat{\alpha})^{X_t} (s \hat{\beta} + 1 - \hat{\beta})^{n-X_t} \right] \right\}^2 ds$$

a pokusit se nalézt vhodný tvar váhové funkce pro různé alternativy. Dále by se dala zkoumat situace s neznámou velikostí populace n a navrhnout vhodný odhad tohoto parametru s vyhodnocením efektu na testy dobré shody. To vše už by bylo nad rámec této práce.

Seznam použité literatury

- ANDĚL, J. (2005). *Základy matematické statistiky*. Matfyzpress, Praha. ISBN 80-86732-40-1.
- BILLINGSLEY, P. (1961). *Statistical Inference for Markov Processes*. The University of Chicago Press, Chicago.
- CIPRA, T. (2008). *Finanční ekonometrie*. Ekopress, Praha. ISBN 978-80-86929-43-9.
- HUDECOVÁ, Š., HUŠKOVÁ, M. a MEINTANIS, S. G. (2015). Tests for time series of counts based on the probability-generating function. *Statistics*, **49**(2), 316–337. doi: 10.1080/02331888.2014.979826.
- JONES, G. L. (2004). On the markov chain central limit theorem. *Probab. Surveys*, **1**, 299–320. doi: 10.1214/154957804100000051.
- KLIMKO, L. A. a NELSON, P. I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics*, **6**(3), 629–642.
- MCKENZIE, E. (1985). Some simple models for discrete variate time series. *JAWRA Journal of the American Water Resources Association*, **21**(4), 645–650. doi: 10.1111/j.1752-1688.1985.tb05379.x.
- MCKENZIE, E. (2003). Discrete variate time series. *Handbook of Statistics*, **21**, 573–606. doi: 10.1016/S0169-7161(03)21018-X.
- PRÁŠKOVÁ, Z. a LACHOUT, P. (2012). *Základy náhodných procesů I*. Matfyzpress, Praha. ISBN 978-80-7378-210-8.
- R CORE TEAM (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- WEISS, C. H. (2009a). Monitoring correlated processes with binomial marginals. *Journal of Applied Statistics*, **36**(4), 399–414. doi: 10.1080/02664760802468803.
- WEISS, C. H. (2009b). A new class of autoregressive models for time series of binomial counts. *Communications in Statistics - Theory and Methods*, **38**(4), 447–460. doi: 10.1080/03610920802233937.
- WEISS, C. H. (2018). Goodness-of-fit testing of a count time series' marginal distribution. *Metrika*, **81**. doi: 10.1007/s00184-018-0674-z.
- WEISS, C. H. a KIM, H.-Y. (2013a). Parameter estimation for binomial AR(1) models with applications in finance and industry. *Statistical Papers*, **54**. doi: 10.1007/s00362-012-0449-y.
- WEISS, C. H. a KIM, H.-Y. (2013b). Binomial AR(1) processes: moments, cumulants, and estimation. *Statistics*, **47**(3), 494–510. doi: 10.1080/02331888.2011.605893.

- WEISS, C. H. a KIM, H.-Y. (2014). Diagnosing and modeling extra-binomial variation for time-dependent counts. *Applied Stochastic Models in Business and Industry*, **30**(5), 588–608. doi: 10.1002/asmb.2005.
- WEISS, C. H. a KIM, H.-Y. (2015). Goodness-of-fit tests for binomial AR(1) processes. *Statistics*, **49**(2), 291–315. doi: 10.1080/02331888.2014.974606.