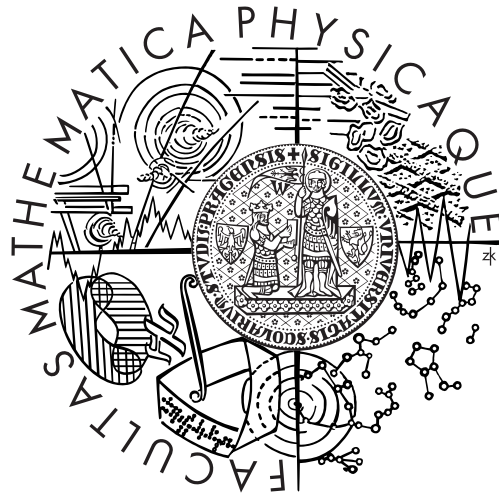


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Ján Krnáč

Metoda bootstrap ve finančních časových řadách

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: **doc. RNDr. Zuzana Prášková, CSc.**

Studijní program: **Matematika**

Studijní obor: **Finanční a pojistná matematika**

Praha 2011

Na tomto mieste by som rád poďakoval vedúcej diplomovej práce doc. RNDr. Zuzane Práškovej, CSc., za veľmi cenné pripomienky, rady, celkový čas, ktorý venovala mne a tejto práci a trpezlivosť pri mnohých konzultáciách.

Prehlasujem, že som svoju diplomovú prácu napísal samostatne a výhradne s použitím citovaných prameňov, literatúry a ďalších odborných zdrojov.

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

V Prahe dňa 3. augusta 2011

Ján Krnáč

.....

Názov práce: Metóda bootstrap vo finančných časových radoch

Autor: Ján Krnáč

Katedra: Katedra pravdepodobnosti a matematickej štatistiky

Vedúci diplomovej práce: doc. RNDr. Zuzana Prášková, CSc.

Abstrakt: Práca sa zaoberá základnými princípmi a vlastnosťami bootstrapových metód so zameraním na ich použitie pri modelovaní a štatistickom spracovaní lineárnych a nelineárnych finančných časových radov. Čitateľ sa zoznámí s podstatou bootstrapových metód vysvetlenou na nezávislých pozorovaniach, následne budú v práci predstavené metódy, ktoré sa môžu použiť v prípade časových radov. V práci sú predstavené parametrické bootstrapové metódy, ktoré môžeme použiť, ak poznáme parametrický model, z ktorého dáta pochádzajú, aj neparametrické bootstrapové metódy, ktoré sa používajú v prípade, ak predpokladáme všeobecný model neparametrickej autoregresie. Cieľ práce je porovnať jednotlivé metódy a následne ukázať použitie bootstrapových metód na príklade reálnych dát. V práci je uvedená aj základná teória o finančných časových radoch.

Kľúčové slová: bootstrap, podmienená heteroskedasticita, časové rady

Title: Bootstrap in financial time series

Author: Ján Krnáč

Department: Department of probability and mathematical statistics

Supervisor: doc. RNDr. Zuzana Prášková, CSc.

Abstract: In this diploma thesis we explain the main principles and properties of bootstrap methods, that can be used to conduct the statistical inference in linear and nonlinear financial time series. We will introduce basic ideas of bootstrap methods for the case when observations can be considered as independent random variables, and afterwards we will describe more advanced methods, that can be successfully used when we are dealing with time series. Thesis deals with both parametric bootstrap methods, that we can use when the underlying parametric model of observations is available, as well as with nonparametric bootstrap methods that are used when more general nonparametric model of time series data is considered. The main objective is to compare particular bootstrap methods and show the usage of these methods on real world data. There is also a basic time series theory included in the work.

Keywords: bootstrap, conditional heteroskedasticity, time series

Obsah

Úvod	8
1 Základné princípy metódy bootstrap	9
1.1 Predstavenie metódy bootstrap	9
1.2 Použitie metódy bootstrap	10
1.3 Bootstrap pre nezávislé pozorovania	10
1.4 Intervalové odhady pomocou metódy bootstrap	13
1.5 Voľba počtu bootstrapových výberov	16
2 Finančné časové rady	18
2.1 Modely s podmienenou homoskedasticitou	19
2.1.1 Autoregresný model	20
2.1.2 ARMA model	21
2.1.3 ARIMA model	21
2.2 Modely s podmienenou heteroskedasticitou	22
2.2.1 ARCH modely	22
2.2.2 GARCH modely	24
2.2.3 Odhady parametrov v ARCH a GARCH modeloch	25
2.2.4 Predpovede volatility	27
2.2.5 Modifikácie GARCH modelu	27
2.3 Odhady v neparametrickej autoregresii	29
3 Parametrické metódy bootstrap	32
3.1 Reziduálny bootstrap	32
3.2 Wild bootstrap	33
3.3 Párový bootstrap	34
4 Neparametrické metódy bootstrap	35
4.1 Autoregresný bootstrap	35
4.2 Neparametrický wild bootstrap	36
5 Bootstrap metódy a finančné časové rady	38
5.1 Reziduálny bootstrap	38
5.1.1 Ako nepresný je reziduálny bootstrap pri heteroskedastických chybách?	38
5.1.2 Asymptotické vlastnosti reziduálneho bootstrapu	44

5.2	Wild bootstrap	45
5.2.1	Rekurzívny wild bootstrap	45
5.2.2	Fixný wild bootstrap	46
5.3	Párový bootstrap	47
5.4	Autoregresný bootstrap	47
5.5	Wild bootstrap	51
5.6	Výber prvotných odhadov \tilde{m} , $\tilde{\sigma}$ a \tilde{P}_ε	52
6	Porovnanie metód	54
6.1	Simulované dáta	54
7	Použitie metódy bootstrap	71
7.1	Implementácia	77
	Záver	78
	Literatúra	79
	Použité značenie	82
	Zoznam tabuliek	83

Úvod

„One of main incentives for development of bootstrap is the desire to substitute computer power for theoretical calculations involving special distributions.“

Bradley Efron

Metóda bootstrap je intenzívna počítačová metóda, ktorá slúži na štatistickú analýzu dát. S rozvojom výpočtovej techniky sa jej využitie stáva jednoduchšie, a tým sa zvyšuje aj jej použiteľnosť v praxi. Metódu po prvýkrát zverejnil a zviditeľnil Bradley Efron (1979) a odvtedy sa teší veľkému záujmu, čo za posledné tri desaťročia prispelo k veľkému nárastu jej modifikácií a momentálne má veľmi veľký rozsah použiteľnosti pre široký okruh štatistických problémov. Efron zjednodušene označil metódu bootstrap ako typ Monte-Carlo metódy aplikovaných na pozorované dáta.

Výraz bootstrap si doslova môžeme preložiť ako „šnúrka na topánkach“. Tento výraz vznikol na základe príbehu o topiacom sa barónovi, ktorý nemal žiadnu inú možnosť na záchranu, ako zatiahnuť za šnúrku svojich topánok. Rovnako ako tento spôsob záchrany topiaceho sa človeka aj metóda bootstrap sa na prvý pohľad javí dosť nedôveryhodná. Avšak ako sa nakoniec ukázalo v príbehu, aj v štatistike, metóda je veľmi účinná.

Existuje veľké množstvo bootstrapových metód, a zďaleka nie každá metóda je vhodná pre všetky typy dát. Kritickým faktorom pre výber vhodnej bootstrapovej metódy je hlavne závislosť, resp. nezávislosť dát, a homoskedasticita, resp. heteroskedasticita dát. Naivným aplikovaním nevhodnej bootstrapovej metódy môžeme dostať veľmi nepresné výsledky. V práci sa budeme zaoberať hlavne časovými radmi, a v krátkosti ukážeme zlyhanie jednotlivých bootstrapových metód pri nesplnení požiadaviek, ktoré na dáta kladieme. Ďalšou prekážkou pri aplikovaní bootstrapových metód je znalosť, resp. predpoklad o modeli, z ktorého dáta pochádzajú. V práci predstavíme metódy, ktoré vyžadujú apriórnu znalosť určitého parametrického modelu, takisto ako metódy, ktoré k svojej implementácii vyžadujú veľmi všeobecný tvar podkladového modelu.

Práca je členená do siedmich kapitol. V 1. kapitole predstavíme najzákladnejšie princípy bootstrapovej metódy, spektrum jej použitia, teoretické pozadie metódy spolu s jednoduchým príkladom použitia metódy. Opíšeme aj princíp aplikovania metódy na konštrukciu intervalových odhadov. V kapitole 2 sa čitateľ zoznámí so základnými parametrickými a neparametrickými modelmi časových radov, ktoré sa používajú vo financiách. V kapitole 3 a 4 uvidíme stručný prehľad a algoritmy jednotlivých bootstrapových metód. V 5., resp 6 kapitole, už sú podrobnejšie študované vlastnosti metód pre parametrické modely, resp. neparametrické modely časových radov. Kapitola 6. obsahuje simulácie a porovnanie metód a nakoniec v kapitole 7. nájdeme príklad použitia bootstrapových metód v praxi, spolu s aplikáciou na reálne finančné dáta.

Všetky postupy a algoritmy, ktoré sú v práci uvedené, boli implementované v jazyku R. Algoritmy spolu s reálnymi dátami, ktoré boli použité pri príkladoch, resp. aplikáciách, sú obsahom priloženého CD.

Kapitola 1

Základné princípy metódy bootstrap

V tejto kapitole sa zameriame na základné teoretické vlastnosti metódy bootstrap a vysvetlíme si jej použitie v prípade nezávislých náhodných veličín.

1.1 Predstavenie metódy bootstrap

Pokiaľ by sme niečo o tejto metóde hľadali a nemali by sme žiadne informácie, v čom metóda spočíva, pravdepodobne by sme narazili na nasledujúce základné vysvetlenie. Metóda bootstrap simuluje, čo by sa stalo, ak by sme pozorovali opakujúce sa vzorky zo základného súboru tým spôsobom, že z dostupných dát vytvorí nové náhodné výbery (*prevzorkovania, replikácie*). Tieto náhodné výbery môžu mať menší rozmer, ako je rozmer pôvodného náhodného výberu dát a môžu byť vytvorené s opakovaním alebo bez opakovania. Najlepšie výsledky metóda väčšinou dosahuje, ak sú bootstrapové náhodné výbery konštruované s opakovaním a rovnakým rozsahom ako pôvodný výber. Obrázok 1.1 predstavuje najzákladnejší princíp metódy bootstrap.

\mathbf{X}	(originál)	→	(replikácie)	\mathbf{X}_1^*	\mathbf{X}_2^*	\mathbf{X}_3^*	...	\mathbf{X}_B^*
X_1				X_3	X_2	X_3	...	X_1
X_2				X_1	X_2	X_6	...	X_6
X_3			náhodný výber s vracaním	X_6	X_3	X_5	...	X_2
X_4		→		X_4	X_4	X_2	...	X_1
X_5				X_1	X_1	X_3	...	X_4
X_6				X_6	X_2	X_1	...	X_4

Obrázok 1.1. Základný princíp metódy bootstrap predstavuje náhodný výber s vracaním z pôvodnej vzorky dát s rozsahom n (v tomto prípade 6) a konštrukcia B replikácií.

1.2 Použitie metódy bootstrap

Predpokladajme, že máme k dispozícii náhodný výber X_1, \dots, X_n pozorovaných hodnôt. Keďže X_1, \dots, X_n sú náhodné veličiny, majú distribučnú funkciu F a nech $\theta = \theta(F)$ je neznámy parameter, ktorý chceme na základe realizácie tohto náhodného výberu X_1, \dots, X_n odhadnúť pomocou štatistiky $T_n = T_n(X_1, \dots, X_n)$. Inými slovami, chceme nájsť bodový odhad parametra θ , ktorý označíme $\hat{\theta} = T_n = T_n(X_1, \dots, X_n)$. Budeme predpokladať, že štatistiku (odhad) máme a môžeme pristúpiť k hľadaniu vlastností tohto odhadu. Najčastejšie štatistické otázky dopadajú na vychýlenie odhadu, rozptyl, resp. smerodajnú odchýlku odhadu, intervaly spoľahlivosti, prípadne kritické hodnoty pre testovanie hypotéz.

V tomto momente nastáva otázka, akú metódu zvoliť pre odhadovanie vlastností daného odhadu. Pokiaľ rozsah pôvodných dát je dostatočne veľký, môžeme sa oprieť o centrálnu limitnú vetu a asymptotické vlastnosti. V praxi ale veľmi často nastane situácia, keď tento predpoklad splnený nie je. Vtedy prichádza metóda bootstrap so svojím riešením.

Z informácií z predošlej a tejto sekcie si môžeme sformulovať základný postup, ako sa pomocou metódy bootstrap vlastnosti štatistík hľadajú. V nasledujúcej sekcii si ukážeme základný algoritmus aplikovateľný v prípade, že náš pôvodný výber, ktorý máme k dispozícii, tvoria nezávislé dáta.

1.3 Bootstrap pre nezávislé pozorovania

Majme X_1, \dots, X_n nezávislé a rovnako rozdelené náhodné veličiny s distribučnou funkciou F , a nech $\theta = \theta(F)$ je neznámy parameter určený rozdelením F , ktorý chceme na základe realizácie tohto náhodného výberu X_1, \dots, X_n odhadnúť pomocou štatistiky $T_n = T_n(X_1, \dots, X_n)$. Ďalej budeme predpokladať, že o distribučnej funkcii F máme minimum informácií, preto odhadovať vlastnosti štatistiky T_n na základe funkcie F nie je možné.

Metóda bootstrap je založená na takzvanom *plug-in* princípe. Ide o veľmi jednoduchý postup, kde neznámu distribučnú funkciu F nahradíme nejakým jej odhadom, najčastejšie *empirickou distribučnou funkciou* F_n definovanú vzťahom

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{[X_i \leq x]}.$$

V tomto jednoduchom princípe spočíva sila metódy bootstrap, ktorý zároveň predstavuje základný predpoklad metódy. Použitím empirickej distribučnej funkcie totiž nestrácame žiadnu informáciu o pôvodných dátach ani o pôvodnej distribučnej funkcii F . O vlastnostiach empirickej distribučnej funkcie spojených s bootstrapom sa môžeme dočítať v knihe van Es a Putter (2009, str. 5-6).

Poznámka 1.1. Metóda bootstrap tak, ako ju uviedol B. Efron vo svojej prvej publikácii, je takzvaný neparametrický bootstrap. Navyše existuje ešte parametrický bootstrap, ktorý je vhodný v prípade, že poznáme rozdelenie výberu $\mathbf{X} = X_1, \dots, X_n$,

čiže poznáme distribučnú funkciu $F(\beta)$ až na neznáme parametre. V tomto prípade sa postupuje tak, že odhadneme neznáme parametre a ďalej pracujeme s distribučnou funkciou $F(\hat{\beta})$. Vidíme, že parametrický bootstrap vyžaduje aspoň znalosť tvaru distribučnej funkcie, kým neparametrickú metódu bootstrap môžeme použiť na získanie vlastností odhadu bez akejkoľvek znalosti pôvodného rozdelenia.

Na to, aby sme mohli skúmať niektoré štatistické vlastnosti T_n (napr. intervaly spoľahlivosti), musíme poznať rozdelenie jej štandardizovanej verzie, označme ju $R_n = R_n(X_1, \dots, X_n; F)$. Napr. $R_n = \sqrt{n}(T_n - \theta)$. Ďalej si označme

$$H_F(x) = P[R_n(X_1, \dots, X_n; F) \leq x],$$

distribučnú funkciu štatistiky R_n . Veličiny X_1^*, \dots, X_n^* , ktoré dostaneme spomínaným prevzorkovaním, čiže náhodným výberom s vracaním z pôvodných dát X_1, \dots, X_n , tvoria tzv. *bootstrapový výber* so známym rozdelením daným distribučnou funkciou F_n (bootstrapový výber je takisto nezávislý). Poznamenajme, že z vlastností empirickej distribučnej funkcie plynú nasledujúce rovnosti

$$P^*(X_j^* = X_i) = \frac{1}{n}, \quad i, j = 1, \dots, n,$$

kde P^* značí pravdepodobnostnú mieru indukovanú bootstrapom, čiže podmienenú pravdepodobnosť pri daných X_1, \dots, X_n , a

$$E^* X_1^* = \sum_{i=1}^n X_i P^*(X_1^* = X_i) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n, \quad (1.1)$$

$$\text{var}^* X_1^* = \sum_{i=1}^n (X_i - E^* X_1^*)^2 P^*(X_1^* = X_i) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (1.2)$$

$E^*(\cdot)$ respektíve $\text{var}^*(\cdot)$ znamená, že ide o podmienené charakteristiky, čiže $E^* X_1^* = E(X_1^* | X_1, \dots, X_n)$ a $\text{var}^* X_1^* = \text{var}(X_1^* | X_1, \dots, X_n)$.

Podobne ako pre prvé dva bootstrapové momenty, pomocou funkcie F_n si skonštruujeme aj bootstrapové verzie štatistiky T_n a jej distribučnej funkcie. Dostávame $\theta^* = \theta(F_n)$, $T_n^* = T_n(X_1^*, \dots, X_n^*)$, $R_n^* = R_n(X_1^*, \dots, X_n^*, F_n)$. Vlastnosti a rozdelenie štatistiky T_n budeme skúmať na základe jej bootstrapovej verzie T_n^* . Ďalej budeme potrebovať distribučnú funkciu R_n^*

$$H_{F_n}^*(x) = P^*[R_n^* \leq x].$$

Stanoviť presné hodnoty týchto *teoretických charakteristík* získaných metódou bootstrap by si vyžadovalo veľké množstvo operácií, čo v praxi aj pri relatívne malom n nie je možné. Preto sa postupuje metódou *Monte-Carlo*, kde sa B -krát vygeneruje náhodný výber s rozdelením F_n (získame spomínané prevzorkovania), pre každý takto vygenerovaný výber sa spočítajú hodnoty štatistík T_n^* a R_n^* , čím získame $T_{n,1}^*, \dots, T_{n,B}^*$ a $R_{n,1}^*, \dots, R_{n,B}^*$, na základe ktorých si spočítame vlastnosti ako vychýlenie, rozptyl alebo

rozdelenie, ktoré potom vzťahujeme na charakteristiky pôvodných štatistík. Napr. pre vychýlenie platí

$$\widehat{\text{Bias}}^* T_n^* = \frac{1}{B} \sum_{b=1}^B T_{n,b}^* - T_n,$$

rozptyl dostaneme zo vzťahu

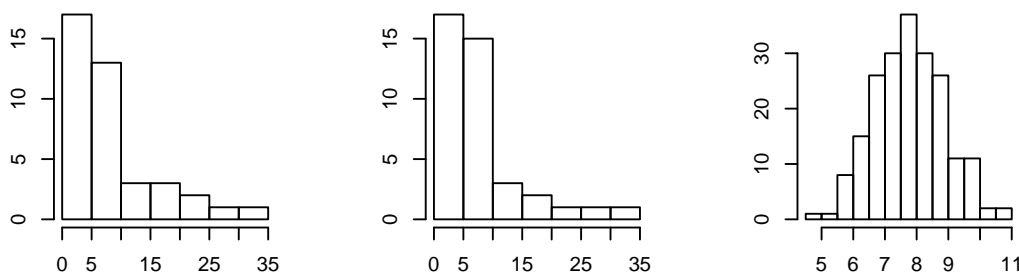
$$\widehat{\text{var}}^* T_n^* = \frac{1}{B} \sum_{b=1}^B (T_{n,b}^* - \frac{1}{B} \sum_{i=1}^B T_{n,i}^*)^2$$

a konečne rozdelenie R_n^*

$$\widehat{H}_{F_n}^*(x) = \frac{1}{B} \sum_{b=1}^B I_{[R_{n,b}^* \leq x]}.$$

Vidíme, že metóda bootstrap je akási kombinácia plug-in princípu a metódy Monte-Carlo.

Príklad 1.1. Skúmame časy medzi príchodom 41 áut v pevne zvolenom mieste na ceste do Bedfordshire, Anglicko. Pre dané dáta (prebrané z Hand et al. (1994)) sme vypočítali, že výberový priemer časov medzi jednotlivými príchodmi je $\bar{x} = 7.8$ a smerodajná odchýlka $s = 7.87$. Z histogramu na obrázku 1.2 vidíme, že rozdelenie dát je možné modelovať exponenciálnym rozdelením $\text{Exp}(\frac{1}{7.8})$, z čoho dostávame vzorce napr. pre smerodajnú odchýlku \bar{x} v tvare $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}} = 1.24$. Druhý spôsob, ako určiť smerodajnú odchýlku \bar{x} , je pomocou metódy bootstrap. Budeme postupovať tak, že si 200-krát ($B = 200$) napočítame \bar{x} pre 200 rôznych bootstrapových výberov z pôvodných dát, dostaneme $\bar{x}_1^*, \dots, \bar{x}_{200}^*$ a z týchto vypočítame $\sigma_{\bar{x}}^* = 1.23$ ako výberovú smerodajnú odchýlku. Vidíme, že smerodajná odchýlka vypočítaná pomocou bootstrapu je hodnotou veľmi blízko $\sigma_{\bar{x}}$. Na obrázku 1.2 môžeme vidieť vľavo histogram pôvodných pozorovaní doby medzi príchodom dvoch po sebe idúcich áut, v strede histogram jednej bootstrapovej vzorky a vpravo histogram $\bar{x}_1^*, \dots, \bar{x}_{200}^*$.



Obrázok 1.2. Postupne zľava: histogram doby medzi príchodom dvoch po sebe idúcich áut, histogram jednej bootstrapovej vzorky a nakoniec histogram $\bar{x}_1^*, \dots, \bar{x}_{200}^*$.

1.4 Intervalové odhady pomocou metódy bootstrap

V štatistike sa drvivá väčšina problémov konštrukcie intervalových odhadov opiera o aproximáciu pomocou centrálnej limitnej vety. V tejto sekcii si ukážeme, že intervalové odhady zostrojené pomocou metódy bootstrap môžu byť presnejšie ako odhady získané aproximáciou pomocou normálneho, resp. Studentovho rozdelenia. Pripomeňme si v skratke, ako vyzerá klasický interval spoľahlivosti

$$(T_n - u_{1-\alpha/2}S_n, T_n - u_{\alpha/2}S_n), \quad (1.3)$$

kde S_n je odhad smerodajnej odchýlky štatistiky T_n a $u_{\alpha/2}$ je príslušný α -kvantil, ktorý odvodíme z predpokladu, že

$$R_n = \frac{T_n - \theta}{S_n} \sim N(0, 1). \quad (1.4)$$

Tento predpoklad je na mieste iba pre $n \rightarrow \infty$. Pre dáta s konečnou početnosťou ide len o aproximáciu, ktorá sa navyše stáva nepresnejšou s klesajúcim rozsahom dát. V roku 1908 W. S. Gosset odvodil lepšiu aproximáciu pre prípad, že $T_n = \bar{X}$ a to

$$R_n = \frac{T_n - \theta}{S_n} \sim t_{n-1}, \quad (1.5)$$

kde t_{n-1} značí Studentovo rozdelenie o $n - 1$ stupňoch voľnosti. Táto aproximácia je presnou v prípade, že $T_n = \bar{X}$ a dáta pochádzajú z normálneho rozdelenia. Navyše interval spoľahlivosti zostrojený pomocou (1.5) býva dlhší ako interval zostrojený pomocou (1.4) (z dôvodu neznámej smerodajnej odchýlky). Ako vidíme, použitím (1.5) nepočítame so šikmou dát a takisto nekorigujeme výsledky o chyby plynúce z porušenia predpokladu $T_n = \bar{X}$. Nasledujúci odstavec v krátkosti predstaví metódu, ktorá rieši spomínané problémy.

Bootstrap-t intervaly

Prvou možnosťou, ako nahradiť klasický interval spoľahlivosti, je použiť bootstrap-t intervaly. Táto metóda odhaduje rozdelenie R_n priamo z dát. Opíšeme celý algoritmus. Najprv generujeme B výberov $\mathbf{X}_1^*, \dots, \mathbf{X}_B^*$, z pôvodného výberu $\mathbf{X} = X_1, \dots, X_n$. Pre každý takýto výber nadefinujeme

$$R_n^*(b) = \frac{T_n^*(b) - T_n}{S_n^*(b)}, \quad b = 1, \dots, B \quad (1.6)$$

kde $T_n^*(b)$ je hodnota štatistiky T_n pre bootstrapový výber \mathbf{X}_b^* a $S_n^*(b)$ je odhadnutá príslušná smerodajná odchýlka výberu \mathbf{X}_b^* . Ďalej budeme potrebovať získať kvantily rozdelenia veličiny $R_n^*(b)$. Tie spočítame ako $u_\alpha^* = X_{(\lfloor B\alpha \rfloor)}$, čiže ako výberové kvantily z usporiadaného výberu $R_n^*(1), \dots, R_n^*(B)$. Tým získame $100(1-\alpha)$ percentný bootstrap t-interval spoľahlivosti v tvare

$$(T_n - u_{1-\alpha/2}^* \cdot S_n, T_n - u_{\alpha/2}^* \cdot S_n). \quad (1.7)$$

Vidíme, že v algoritme na získanie bootstrap-t intervalu pracujeme s odhadom smerodajnej odchýlky $S_n^*(b)$ výberu \mathbf{X}_b^* . Ak tento odhad nemáme k dispozícii, zvyčajne ho hľadáme opäť metódou bootstrap. Tým sa však metóda stáva dosť výpočtovo náročnou, pretože potrebujeme B_1 výberov na získanie kvantilov a v každom tomto výbere B_2 výberov na získanie smerodajnej odchýlky, celkovo teda $B_1 \cdot B_2$ výberov.

Poznámka 1.2. Bootstrap-t metóda je popri výpočtovej náročnosti aj značne nepresná už pri malých počtoch odľahlých pozorovaní. Nasledujúca sekcia popisuje v skratke viac robustnejšiu metódu.

Percentilové intervaly

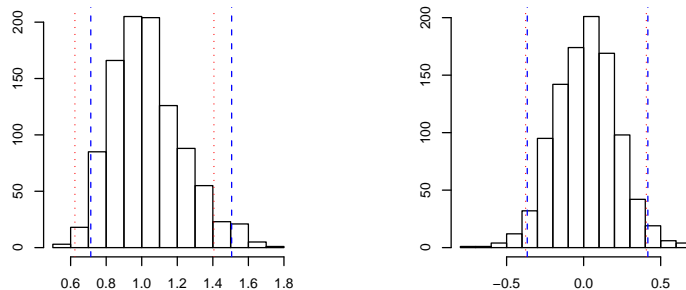
Predstavme si, že chceme vypočítať 95% interval spoľahlivosti pre odhad strednej hodnoty základného súboru. Percentilová metóda pracuje tak, že odhadne 0,025 a 0,975 percentné kvantily odhadu pomocou metódy bootstrap a tieto odhady potom použije ako koncové body 95% intervalu spoľahlivosti. Vo všeobecnosti to znamená, že metóda počíta intervalový odhad parametra θ pomocou kvantilov distribučnej funkcie *neštandardizovanej* štatistiky $T_n^*(b)$. Keď označíme $G_n^*(x)$ ako distribučnú funkciu $T_n^*(b)$, potom percentilový interval spoľahlivosti má tvar

$$\left(\hat{G}_n^{*-1}\left(\frac{\alpha}{2}\right), \hat{G}_n^{*-1}\left(1 - \frac{\alpha}{2}\right)\right),$$

kde $\hat{G}_n^{*-1}(\alpha)$ je α -percentný kvantil distribučnej funkcie $G_n^*(x)$.

Poznámka 1.3. Princíp získavania kvantilov z bootstrapových výberov je podobný ako pri bootstrap-t metóde. Nagenerujeme $\mathbf{X}_1^*, \dots, \mathbf{X}_B^*$ bootstrapových výberov, vypočítame hodnotu štatistiky $T_n^*(b)$ pre každý bootstrapový výber, zoradíme ich podľa veľkosti tak, aby $T_n^*(1) \leq \dots \leq T_n^*(B)$ a z nich získame dolnú a hornú hranicu intervalu podľa požadovaného koeficientu spoľahlivosti tohto intervalu v tvare $B \cdot \alpha/2$ a $B \cdot (1 - \alpha/2)$. Napr. pre $B = 1000$ a $\alpha = 5\%$ sú to 25-ta a 975-ta hodnota z usporiadaného súboru.

Príklad 1.2. Skúmame otázku korektnosti intervalového odhadu. Porovnáme dve metódy, a to interval zostrojený pomocou CLV a percentilový interval. Budeme mať k dispozícii náhodný výber z normovaného normálneho rozdelenia X_1, \dots, X_{20} . Parameter θ , ktorý chceme odhadnúť, si zvolíme e^μ , kde $\mu = EX_1$, v našom prípade je správna hodnota $e^0 = 1$. Ako odhad parametra θ použijeme štandardne $\hat{\theta} = T_n = e^{\bar{X}_n}$. Pre odhad smerodajnej odchýlky v intervale spoľahlivosti pomocou CLV použijeme bootstrap takisto ako na zostrojenie histogramu T_n^* (Keďže rozdelenie náhodného výberu poznáme, mohli by sme nagenerovať niekoľko výberov priamo, ale predpoklad o známom rozdelení nebudeme brať do úvahy a ďalšie výbery nagenerujeme metódou bootstrap).



Obrázok 1.3. Vľavo: Histogram rozdelenia štatistiky $T_n = \exp(\bar{X}_n)$ zostrojený pomocou 1000 bootstrapových replikácií. Červená bodkovaná čiara značí interval spoľahlivosti zostrojený pomocou centrálnej limitnej vety, modrá čiarkovaná čiara zasa percentilový interval. Vpravo: Rovnako ako vľavo, s tým rozdielom, že ide o histogram rozdelenia štatistiky $\log(T_n) = \bar{X}_n$.

95% percentilový interval spoľahlivosti sme počítali v tvare

$$(\hat{G}_n^{*-1}(\frac{\alpha}{2}), \hat{G}_n^{*-1}(1 - \frac{\alpha}{2})) = (0.71, 1.50),$$

a 95% interval spoľahlivosti na základe CLV v tvare

$$(e^{\bar{X}_n} + u_{\alpha/2} \cdot S_n^*, e^{\bar{X}_n} + u_{1-\alpha/2} \cdot S_n^*) = (0.62, 1.40).$$

Na obrázku 1.3 vidíme, že v prvom prípade, kde histogram vykazuje výrazne šikmé rozdelenie, nedáva klasický interval príliš dobré výsledky na rozdiel od percentilového intervalu, avšak po logaritmickej transformácii, ktorá znormalizuje šikmé rozdelenie, už obidve metódy vykazujú prakticky totožné výsledky.

Podme sa bližšie pozrieť na presnosť intervalových odhadov v číslach. Za týmto účelom porovnajme presnosť klasického intervalu spoľahlivosti a percentilového odhadu. Pod pojmom presnosť intervalového odhadu rozumieme schopnosť odhadu dodržať požadovaný koeficient spoľahlivosti. Ilustrujme si situáciu na príklade 1.2 v nasledujúcej tabuľke. Odhad smerodajnej odchýlky v intervale spoľahlivosti aproximáciou normálnym rozdelením sme opäť zostrojili bootstrapom.

Metóda zostrojenia intervalu	Chyba zľava (%)	Chyba sprava (%)
Aproximácia normálnym rozdelením	1.3	6.7
Percentilová	4.3	3.7

Tabuľka 1.1. Výsledky sú zostavené na základe 300 simulácií intervalov spoľahlivosti pre $\theta = \exp(\mu)$ pre náhodný vyber z normovaného normálneho rozdelenia o veľkosti 20. Čísla predstavujú percentuálne vyjadrenie situácie, kedy interval „minul“ správnu hodnotu 1.0 zľava, resp. sprava. Napr. chyba zľava znamená, že ľavý koncový bod intervalu je väčší ako 1.0. Ak budeme predpokladať, že zostrojené intervaly boli 95%, potom ideálne hodnoty chyba zľava aj chyba sprava sú rovné 2.5 %.

Z tabuľky 1.1 vidíme, že interval zostrojený pomocou CLV má až príliš vysokú kryciu schopnosť zľava, avšak sprava prekračuje požadovanú chybu až o 4.2 %. Percentilová metóda dáva viac vyrovnané hodnoty zľava a sprava, ale celkovo nekryje odhad parametra príliš spoľahlivo. Toto je spôsobené neparametrickým prístupom, presnejšie tým, že percentilový interval nepracuje s rozdelením pôvodného výberu, ale používa empirické (bootstrapové) rozdelenia.

Ak zhrnieme doterajšie poznatky, môžeme konštatovať, že bootstrap-t intervaly majú veľmi dobré teoretické vlastnosti, ale ukázalo sa, že v praxi je často nepoužiteľná z dôvodu malej robustnosti voči odľahlým pozorováním a je výpočetne náročná. Percentilové intervaly sú univerzálnejšie, menej výpočtovo náročné, ale nemajú uspokojivé požadované pokrytie (viď tabuľka 1.1). Presnosť percentilových odhadov je iba $O_P(n^{-1/2})$, čo je spôsobené tým, že percentilové intervaly nezahŕňajú informáciu o pôvodnom odhade T_n a jeho smerodajnej odchýlke (viac v Hall (1992, str. 83-84)). Preto vznikli „vylepšené“ verzie percentilových intervalov, a to tzv. BC_a (*bias-corrected and accelerated*) intervaly, ktoré zlepšujú pokrytie intervalov a ABC (*approximate bootstrap confidence*) intervaly, ktoré navyše znižujú obrovskú výpočetnú náročnosť BC_a intervalov.

BC_a intervaly

Už vieme, že táto metóda zlepšuje presnosť percentilových intervalov spoľahlivosti. Zlepšenie spočíva v tom, že výber kvantilu závisí na dvoch konštantách \hat{a} a \hat{z}_0 nazývaných *akcelerácia* a *redukcia vychýlenia*. Veličina \hat{z}_0 zjednodušené meria a odstraňuje nesúlad mediánu odhadu T_n a T_n^* , a veličina \hat{a} predstavuje mieru zmeny smerodajnej odchýlky T_n . Tvar BC_a intervalového odhadu je teda podobný ako pri percentilovej metóde,

$$(\hat{G}_n^{*-1}(\alpha_1), \hat{G}_n^{*-1}(\alpha_2)),$$

kde $\hat{G}_n^{*-1}(\alpha)$ je opäť α -percentný kvantil distribučnej funkcie neštandardizovanej štatistiky a

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + u_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + u_{\alpha/2})}\right), \quad \alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + u_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + u_{1-\alpha/2})}\right),$$

kde $\Phi(\cdot)$ je distribučná funkcia normovaného normálneho rozdelenia a u_α je kvantil tohto rozdelenia.

1.5 Voľba počtu bootstrapových výberov

Pri používaní metódy bootstrap zvyčajne platí, že chyba bootstrapových odhadov je zložená z chyby aproximácie skutočných charakteristík metódou Monte Carlo (*resampling variability*) a takisto z variability danej pôvodným výberom (*sample variability*). Preto by počet B bootstrapových výberov (MC simulácií) mal byť taký veľký, aby chyba

plynúca z Monte Carlo aproximácií bola zanedbateľná. V skratke si opíšeme teoretické odvodenie optimálneho počtu B , napr. pre odhad rozptylu.

Predpokladajme X_1, \dots, X_n *iid* náhodné veľičiny. Už vieme, že odhad rozptylu štatistiky T_n na základe B bootstrapových výberov má tvar

$$\widehat{\text{var}}_B^* T_n^* = \frac{1}{B} \sum_{b=1}^B (T_{n,b}^* - \bar{T}_n^*)^2 = \frac{1}{B} \sum_{b=1}^B (T_{n,b}^* - \frac{1}{B} \sum_{i=1}^B T_{n,i}^*)^2.$$

Ak si označíme $\widehat{\text{var}}_\infty^* T_n^*$ ako ideálny rozptyl štatistiky, tak platí, že $\widehat{\text{var}}_B^* T_n^*$ je väčší ako $\widehat{\text{var}}_\infty^* T_n^*$ a takisto $\widehat{\text{var}}_B^* T_n^* \xrightarrow{P} \widehat{\text{var}}_\infty^* T_n^*$. Ak odhad rozptylu na základe B výberov je oveľa väčší ako ideálny odhad, tak chyba plynúca z MC simulácií je príliš veľká. Dá sa ukázať, že aby sme dosiahli to, že odhadnutý rozptyl je presne 5 % zo skutočného (ideálneho) rozptylu, môžeme sa riadiť nasledujúcou nerovnosťou,

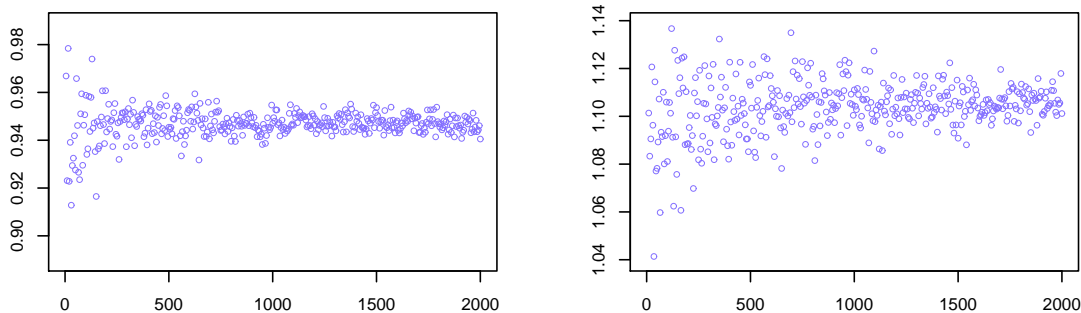
$$B \geq n \cdot \frac{40\text{var}^2(X_1)}{\text{E}(X_1 - \text{E}X_1)^4 - \text{var}^2(X_1)}.$$

Dôkaz nájdeme v knihe van. Es a Putter (2009, str. 17-20).

Pokiaľ by sme chceli odhadnúť kvantily, resp. distribučnú funkciu, môžeme sa riadiť nasledujúcim:

$$\sup_{x \in \mathbb{R}} |\widehat{H}_{F_n}^* - H_{F_n}^*| = \varepsilon_n + \sqrt{B^{-1} \log \log B},$$

kde $\varepsilon_n = \sup_{x \in \mathbb{R}} |H_{F_n}^* - H_F|$ (viď Prášková (2004a), str. 11). Pokiaľ by sme znova chceli dosiahnuť malý pomer medzi chybou MC aproximácie a ε_n , musí platiť $B^{-1} \log \log B = o(\varepsilon_n^2)$ a ak $\varepsilon_n = O_p(n^{-1})$, potom $B = n^2 \log \log n$. Z obrázku 1.4 vidíme, že $B = 200$ sa javí ako minimálna hranica počtu výberov pre odhad bodových charakteristík a $B = 1000$ ako minimálna hranica pre odhad kvantilov, resp. distribučnej funkcie.



Obrázok 1.4. Vľavo: bootstrapové odhady strednej hodnoty štatistiky $\exp(\bar{X}_n)$ v závislosti na počte bootstrapových výberov. Vpravo: bootstrapové odhady 90% kvantilu štatistiky $\exp(\bar{X}_n)$ v závislosti na počte bootstrapových výberov. Pozorovania X_1, \dots, X_{20} sú *iid* s $X_1 \sim N(0, 1)$.

V nasledujúcej kapitole opíšeme teóriu o časových radoch, ktorú budeme potrebovať pri aplikácii metódy bootstrap na časové rady s podmienenou heteroskedasticitou.

Kapitola 2

Finančné časové rady

Pod pojmom časový rad si môžeme predstaviť akúkoľvek postupnosť dát X_1, \dots, X_n , chronologicky usporiadanú v čase. Na túto postupnosť sa zjednodušene môžeme pozeráť ako na súbor čísel s časovým usporiadaním, z ktorých si napr. môžeme spočítať priemer, najnižšiu hodnotu a pod., tzn. študovať dynamiku časového radu (historický, deterministický pohľad). Druhým, veľmi dôležitým rysom takýchto postupností je ich náhodnosť. Pokiaľ zoberieme na zreteľ aj tento aspekt, časový rad pre nás už nepredstavuje len jednoduchú postupnosť čísel, ale akýsi *náhodný proces* s presne definovanými vlastnosťami. Na základe tohto faktu môžeme náhodný proces (v našom prípade špeciálne časový rad) konštruovať generátorom náhodných čísel s takými vlastnosťami, aby nám v zmysle náhodnosti vygeneroval postupnosť dát, veľmi podobnú časovému radu, ktorý chceme analyzovať.

Asi väčšina finančných štúdií pracuje nejakým spôsobom s výnosmi investície. Možno sa ponúka otázka, prečo sa namiesto výnosov nepracuje priamo s cenami príslušných aktív. Campbell, Lo a MacKinlay (1997) toto zdôvodnili tak, že po prvé pre priemerného investora sú výnosy kompletným a ľahko porovnateľným ukazovateľom investície. Po druhé, časový rad výnosov má z hľadiska jednoduchosti štatistickej analýzy oveľa príjemnejšie vlastnosti ako časový rad ceny aktíva.

Predpokladajme teda diskretný časový rad, popisujúci správanie logaritmických mier zisku (*log returns*)

$$r_t = \log \left(1 + \frac{P_t - P_{t-1}}{P_{t-1}} \right), \quad t = 1, 2, \dots,$$

kde $\{P_t, t = 0, 1, 2, \dots\}$ je cena finančného aktíva (cenného papieru, ceny akcie atď.) v čase t . Taylorov rozvoj veličiny r_t ukazuje, že r_t je hodnotami veľmi blízko argumentu $(P_t - P_{t-1})/P_{t-1}$, ktorý popisuje relatívnu zmenu procesu $\{P_t\}$ v čase. Existuje veľké množstvo rôznych modelov časových radov, takisto ako veľké množstvo teoretických, resp. empirických dôkazov o vhodnosti jednotlivých modelov pre rôzne typy časových radov z reálneho sveta. Niektoré základné, ktoré sa najčastejšie používajú na modelovanie finančných časových radov, teraz opíšeme.

Všeobecný model, ktorý si zavedieme a od ktorého sa budú odvíjať niektoré špecifickejšie modely, budeme uvažovať v tvare

$$X_t = m(X_{t-1}, \dots, X_{t-p}) + \sigma(X_{t-1}, \dots, X_{t-p})\varepsilon_t, \quad t \in \mathbb{Z}, \quad (2.1)$$

kde $m(\cdot)$ a $\sigma(\cdot)$ sú neznáme hladké funkcie (σ je nezáporná) a ε_t sú *iid* náhodné veličiny s nulovou strednou hodnotou a jednotkovým rozptylom. Zvyčajne máme k dispozícii pozorované hodnoty X_1, \dots, X_n , čiže budeme predpokladať, že ich hodnoty poznáme.

Model (2.1) vychádza zo všeobecného stochastického zápisu časového radu, kde X_t závisí na aktuálnej a všetkých minulých hodnotách inovácií, resp. šokov e_t , tj.

$$X_t = f(e_t, e_{t-1}, \dots),$$

kde f je neznáma funkcia. Funkcia f všeobecne môže byť nelineárna, potom hovoríme o nelineárnych modeloch. Avšak tento zápis nie je v praxi použiteľný, pretože obsahuje neobmedzene veľa parametrov. Častejší zápis tohto modelu je pomocou prvých dvoch podmienených momentov. Označme Ω_{t-1} ako σ -algebru generovanú akoukoľvek dostupnou informáciou do času $t-1$ vrátane. Typicky je táto informácia tvorená kombináciou prvkov z $\{X_{t-1}, X_{t-2}, \dots\}$ a $\{e_{t-1}, e_{t-2}, \dots\}$. Zápis pomocou prvých dvoch podmienených momentov je v tvare

$$E(X_t | \Omega_{t-1}) = \mu_t = m(\Omega_{t-1}), \quad \text{var}(X_t | \Omega_{t-1}) = \sigma_t^2 = \sigma^2(\Omega_{t-1}), \quad (2.2)$$

kde m a σ sú vhodné funkcie, $\sigma(x) > 0$, a spolu so zápisom procesu v tvare $X_t = \mu_t + e_t$, kde $e_t = \sigma_t \cdot \varepsilon_t$, nám dáva najčastejší tvar zápisu nelineárneho procesu

$$X_t = \mu_t + \sigma_t \cdot \varepsilon_t = m(\Omega_{t-1}) + \sigma(\Omega_{t-1}) \cdot \varepsilon_t, \quad (2.3)$$

kde znova ε_t sú *iid* náhodné veličiny s nulovou strednou hodnotou a jednotkovým rozptylom.

Kľúčovú rolu pri analýze časových radov zohráva schopnosť odhadnúť funkcie $m(\cdot)$ a $\sigma(\cdot)$, kde postup odhadovania sa výrazne líši v závislosti na tvare týchto funkcií. Avšak ani tvar týchto funkcií nemusí byť v čase analýzy radu známy, v tomto prípade sa musia použiť neparametrické metódy uvedené v sekcii 2.3.

Poznámka 2.1. Pre nelineárne $m(\cdot)$ a $\sigma(\cdot)$ môže byť model $X_t = m(X_{t-1}) + \sigma(X_{t-1})\varepsilon_t$ (model (2.1) pre $p = 1$) interpretovaný ako

$$dS_t = m(S_{t-1}) + \sigma(S_t)dW_t,$$

ktorý sa používa ako základný kameň Black-Scholesovho prístupu k modelovaniu cien opcií, kde W_t je štandardný Wienerov proces.

2.1 Modely s podmienenou homoskedasticitou

V tejto časti práce si predstavíme niekoľko základných modelov s podmienenou homoskedasticitou. Všeobecný model v tvare (2.1) sa v prípade podmienenej homoskedasticity zredukuje na tvar

$$X_t = m(X_{t-1}, \dots, X_{t-p}) + \sigma\varepsilon_t, \quad (2.4)$$

kde $\sigma > 0$ je konštanta a ε_t sú *iid* náhodné veličiny s nulovou strednou hodnotou a jednotkovým rozptylom. Označme ďalej $e_t = \sigma\varepsilon_t$. Všeobecne sa predpokladá, že náhodné veličiny e_t tvoria tzv. *biely šum*, tzn. postupnosť nekorelovaných a rovnako rozdelených náhodných veličín s

$$Ee_t = 0, \quad \text{var}(e_t) = \sigma^2, \quad \text{a} \quad \text{cov}(e_i, e_j) = 0, \quad \text{pre všetky } i \neq j.$$

Značíme $\{e_t\} \sim WN(0, \sigma^2)$. Postupnosť nezávislých a rovnako rozdelených náhodných veličín s nulovou strednou hodnotou a konečným rozptylom σ^2 je špeciálny prípad bieleho šumu, tzv. striktný biely šum. Značíme $\{\varepsilon_t\} \sim IID(0, \sigma^2)$.

2.1.1 Autoregresný model

Začnime s modelom, kde funkciu m môžeme zapísať v tvare $m(X_{t-1}, \dots, X_{t-p}) = \sum_{i=1}^p \varphi_i X_{t-i}$. Autoregresný model rádu p (*autoregressive model*) má tvar

$$\text{AR}(p): \quad X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + e_t, \quad (2.5)$$

kde $\{e_t\} \sim WN(0, \sigma^2)$.

Zápis $\text{AR}(p)$ modelu sa často uvádza pomocou tzv. *autoregresného operátora* $\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$, kde B je *operátor časového posunu* s vlastnosťami

$$BX_t = X_{t-1}, \quad (2.6)$$

$$B^j X_t = B^{j-1}(BX_t) = B^{j-1}X_{t-1} = \dots = X_{t-j}. \quad (2.7)$$

Pomocou autoregresného operátora môžeme model (2.5) zapísať ako

$$\varphi(B)X_t = e_t.$$

Stacionarita časových radov

Stacionarita časových radov je dôležitá vlastnosť, ktorá zaručuje určitým spôsobom časovú invariantnosť. Rozlišujeme dva typy stacionarity, ktoré si teraz zdefinujeme.

Definícia 2.1 (Slabá stacionarita) Časový rad $\{X_t, t \in \mathbb{Z}\}$ je slabo stacionárny, ak $E(X_t^2) < \infty$ pre všetky t a

1. $E(X_t)$ je konštantná, nezávislá na t , a
2. $\text{cov}(X_t, X_{t+k})$ je nezávislá na t pre všetky k

Definícia 2.2 (Striktná stacionarita) Časový rad $\{X_t, t \in \mathbb{Z}\}$ je striktnie stacionárny, ak pravdepodobnostné rozdelenie náhodného vektoru $(X_{t_1}, \dots, X_{t_n})$ je rovnaké ako rozdelenie vektoru $(X_{t_1+h}, \dots, X_{t_n+h})$ pre ľubovoľné $h \in \mathbb{Z}$, $n \in \mathbb{N}$ a $t_1, \dots, t_n \in \mathbb{Z}$.

Proces $\text{AR}(p)$ je stacionárny, ak všetky korene polynómu $\varphi(z)$ ležia mimo jednotkového kruhu v komplexnej rovine (Prášková (2004b)).

2.1.2 ARMA model

Zmiešané ARMA procesy sú veľmi obľúbeným a zároveň dôležitým nástrojom na analýzu časových radov. Z predpisu (2.8) procesu vidíme, že nezodpovedá všeobecnému modelu (2.4), ale ARMA procesy úzko súvisia s GARCH procesmi, ktorými sa budeme zaoberať nižšie, preto si aspoň v krátkosti uvedieme ich popis. Tomuto procesu sa hovorí zmiešaný, pretože je kombináciou procesov $AR(p)$ a $MA(q)$ (*moving average*):

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q}, \quad (2.8)$$

Zápis ARMA modelu (2.8) pomocou operátorov má tvar

$$\varphi(B)X_t = \theta(B)e_t, \quad (2.9)$$

kde $\theta(B)$ sa analogicky ako u autoregresného procesu, nazýva *operátor kľzavých súčtov* v tvare $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$.

Proces $ARMA(p, q)$ je (slabo) stacionárny, ak všetky korene polynómu $\varphi(z)$ ležia mimo jednotkového kruhu v komplexnej rovine. Čiže ak platí, že $|z_1|, \dots, |z_p| > 1$ (viac v Prášková (2004b)).

Poznámka 2.2. Proces $ARMA(1, 1)$: $X_t = \varphi_1 X_{t-1} + e_t + \theta_1 e_{t-1}$ je stacionárny pre $|\varphi_1| < 1$.

2.1.3 ARIMA model

Procesy $ARIMA(p, d, q)$ (*autoregressive integrated moving average*) sa využívajú k modelovaniu časových radov, kde dáta vykazujú určitú nestacionaritu, ale môžeme ich stacionarizovať vhodným diferencovaním. Všeobecný zápis modelu $ARIMA(p, d, q)$ uvedieme v tvare pomocou operátorov

$$\varphi(B)w_t = \alpha + \theta(B)e_t, \quad (2.10)$$

kde

$$w_t = \Delta^d X_t. \quad (2.11)$$

Pokiaľ budeme predpokladať, že veličina w_t predstavuje d -tú diferenciu modelovaného (nestacionárneho) časového radu X_t , a diferencovaním tohto radu vznikne rad stacionárny, môžeme výsledný rad (2.10) modelovať pomocou zmiešaného procesu ARMA.

Poznámka 2.3. Príkladom ARIMA procesu je náhodná prechádzka $X_t = X_{t-1} + e_t$, čiže model $ARIMA(0, 1, 0)$.

2.2 Modely s podmienenou heteroskedasticitou

Modely, ktoré sme si predstavili v predchádzajúcich sekciách, boli lineárne modely časových radov. Vo financiách avšak takýto predpoklad na dáta klásť nemôžeme, pretože ako sa ukázalo, veľa vzťahov v makroekonomických a finančných časových radoch je nelineárnych. Tieto (vnútorne) nelineárne vzťahy majú za následok fakt, že lineárne modely nedokážu zachytiť niektoré typické vlastnosti finančných radov. Môžeme spomenúť niekoľko takýchto vlastností:

- *leptokurtické rozdelenie* je rozdelenie, ktoré sa vyznačuje tzv „užším pásom“ a „ťažšími koncami“. Grafická podoba týchto zaužívaných výrazov znamená, že rozdelenie mier zisku je viac špicaté okolo stredu ako normálne rozdelenie (čiže významným faktorom týchto rozdelení je výrazne väčší koeficient špicatosti v porovnaní s normálnym rozdelením, kde koeficient špicatosti definujeme ako $\gamma_2 = E(X - EX)^4 / [var(X)]^2$), pričom na koncoch je jeho hustota väčšia;
- *zhlukovanie volatility* je jav, kde volatilita finančných trhov má tendenciu združovať sa do zhlukov malých a veľkých volatilit. Inak povedané, vysoké hodnoty volatility sú zvyčajne nasledované vysokými a nízke hodnoty nízkymi;
- *pákový efekt* je veľmi zaujímavá vlastnosť, ktorá súvisí s dynamikou volatility v čase. Ide o tendenciu volatility zväčšovať sa viac po poklese sledovanej finančnej veličiny, ako po náraste rovnakej veľkosti tejto veličiny.

Budeme predpokladať všeobecný zápis tohto modelu v tvare bez podmienenej strednej hodnoty,

$$X_t = \sigma(X_{t-1}, \dots, X_{t-p})\varepsilon_t,$$

ε_t sú *iid* náhodné veličiny s nulovou strednou hodnotou a jednotkovým rozptylom.

2.2.1 ARCH modely

Prvý stacionárny model, ktorý sčasti zachycoval vyššie spomínané vlastnosti finančných časových radov, použil Engle (1982) na modelovanie inflácie v UK. Bol to práve Robert F. Engle, ktorý pomenoval tento model ARCH (*Auto-Regressive Conditionally Heteroscedastic*). T. Bollerslev (bývalý PhD študent pod vedením práve Roberta F. Engela) a S. J. Taylor ho v roku 1986 zovšeobecnil za účelom lepšieho popisovania a správania finančných veličín v reálnom svete. Tento zovšeobecný model nazvali GARCH (*Generalized Auto-Regressive Conditionally Heteroscedastic*). GARCH je pravdepodobne najpoužívanejším finančným modelom časových radov a je ním inšpirované veľké množstvo ďalších sofistikovaných modelov.

Všeobecný model ARCH(m) má tvar

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_m X_{t-m}^2, \quad (2.12)$$

kde ε_t sú *iid* náhodné veličiny s nulovou strednou hodnotou a jednotkovým rozptylom (často sa predpokladá, že majú normálne rozdelenie $N(0, 1)$). Zo zápisu (2.12) vidíme, že smer zmeny ceny X_t je daný znamienkom ε_t a veľkosť tejto zmeny volatilitou σ_t^2 ,

ktorá je nezávislá na ε_t .

Stredná hodnota ARCH procesu je rovná 0, pretože

$$E(X_t) = E[E(X_t | \Omega_{t-1})] = E[E(\sigma_t \varepsilon_t | \Omega_{t-1})] = E[\sigma_t E(\varepsilon_t)] = E[\sigma_t \cdot 0] = 0.$$

kde $\Omega_{t-1} = \sigma(X_{t-1}, X_{t-2}, \dots)$ je σ -algebra generovaná informáciou obsiahnutou v procese $\{X_s, s \leq t-1\}$. Pre podmienený rozptyl dostávame

$$\text{var}(X_t | \Omega_{t-1}) = E[(X_t - E[X_t | \Omega_{t-1}])^2 | \Omega_{t-1}] = E[X_t^2 | \Omega_{t-1}] = \sigma_t^2 E(\varepsilon_t^2) = \sigma_t^2.$$

Pre nepodmienený rozptyl za predpokladu slabej stacionarity platí

$$\sigma_X^2 = \text{var}(X_t) = E(\sigma_t^2) E(\varepsilon_t^2) = \alpha_0 + \sum_{i=1}^m \alpha_i E(X_{t-i}^2) = \alpha_0 + \sigma_X^2 \sum_{i=1}^m \alpha_i,$$

z čoho dostávame

$$\sigma_X^2 = \frac{\alpha_0}{1 - \sum_{i=1}^m \alpha_i}. \quad (2.13)$$

Z výrazu (2.13) vidíme, že musí platiť $\sum_{i=1}^m \alpha_i < 1$, čo je zároveň nutná a postačujúca podmienka pre slabú stacionaritu procesu (Vaněček (2003), veta 4).

Poznámka 2.4. Z definície ARCH procesu vidíme, že sa dokáže vysporiadať s leptokurtickým rozdelením (pokiaľ je $\alpha_1 > 0$, viď Cipra (2008, str. 387)) aj so zhlukovaním volatility (viď Tsay (2003, str. 103)), ale už sa nedokáže vysporiadať s pákovým efektom. Tento nedostatok eliminujú niektoré modifikácie základného GARCH modelu.

Test na prítomnosť heteroskedasticity

Všimnime si, že ARCH(m) model môžeme zapísať ako AR(m) model v tvare

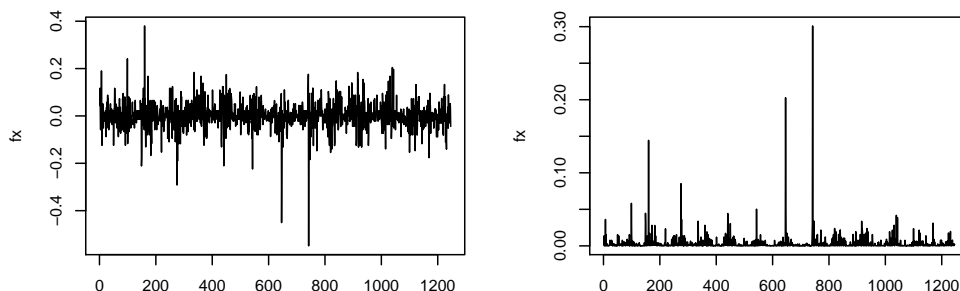
$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_m X_{t-m}^2 + Y_t,$$

kde $Y_t = X_t^2 - \sigma_t^2$, čiže chyby v tomto modeli nie sú iid. Na základe tejto skutočnosti, potom analýza prebieha rovnako ako v AR modeloch pomocou autokorelogramov (ACF), resp. parciálnych autokorelogramov (PACF). Najmä PACF pre X_t^2 môže poslúžiť ako vhodný nástroj na odhad rádu m . Testovanie na samotnú prítomnosť podmienenej heteroskedasticity, tzv. *ARCH efekt*, sa robí takisto na základe X_t^2 , a to pomocou *Ljung-Boxovho testu* založeného na tzv. portmanteau štatistike s kritickým oborom (na hladine významnosti α)

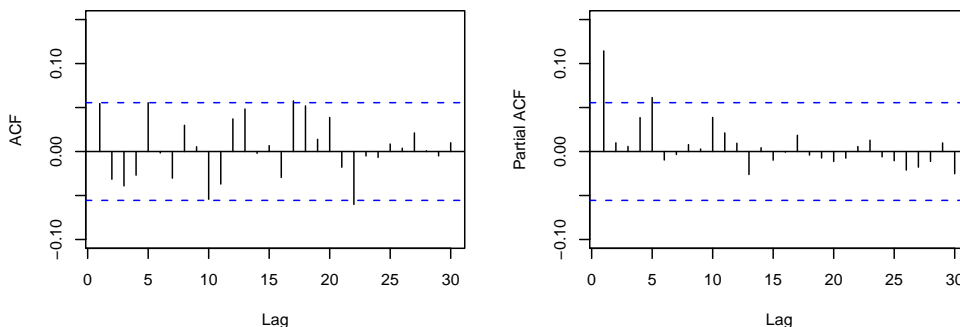
$$Q = n(n+2) \sum_{k=1}^m \frac{r_k^2}{n-k} \geq \chi_{1-\alpha}^2(m),$$

ktorý sa používa na testovanie nekorelovanosti, to znamená, že za platnosti nulovej hypotézy nekorelovanosti časového radu, má štatistika Q asymptoticky χ^2 rozdelenie o m stupňoch voľnosti. Poznamenajme, že r_k značí k -tý výberový autokorelačný koeficient.

Príklad 2.1. Ukážme, že ARCH efekt sa naozaj vyskytuje vo finančných časových radoch. Na obrázku 2.1 je znázornený priebeh časového radu percentuálnych zmien devízového kurzu EUR/USD meraného v 15-minútových intervaloch, v období od 25. 01. 2011 do 11. 02. 2011. Dĺžka časového radu je 1247. Hodnoty napočítané z $r_t = 100.(P_{t+1} - P_t)/P_t$, kde P_t je daný devízový kurz v čase t . Dáta boli získané pomocou programu XTB Trader určeného na internetové obchodovanie s cudzími menami.



Obrázok 2.1. Vľavo: priebeh časového radu r_t percentuálnych zmien devízového kurzu EUR/USD. Vpravo: priebeh časového radu r_t^2 druhých mocnín hodnôt percentuálnych zmien devízového kurzu EUR/USD.



Obrázok 2.2. Vľavo: ACF časového radu r_t . Vpravo: PACF časového radu r_t^2 .

Z obrázku 2.2 vidíme, že ACF pre časový rad r_t percentuálnych zmien vykazuje vlastnosti bieleho šumu, pretože všetky uvažované hodnoty autokorelačných koeficientov sú nevýznamné. Rad r_t^2 druhých mocnín percentuálnych zmien už ale vykazuje silnú autokoreláciu v bode 1, čo svedčí o podmienenej heteroskedasticite.

2.2.2 GARCH modely

Všeobecný model GARCH(m, s) má tvar

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i X_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2, \quad (2.14)$$

kde opäť ε_t sú *iid* náhodné veličiny s nulovou strednou hodnotou a jednotkovým rozptylom nezávislé na $\{X_{t-k}, k \geq 1\}$ pre všetky t a $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$.

Predpokladajme, že $\Omega_{t-1} = \sigma(X_{t-1}, X_{t-2}, \dots)$. Podobne ako pre ARCH proces, aj tu platí

$$\begin{aligned} E(X_t) &= E[E(X_t | \Omega_{t-1})] = E[E(\sigma_t \varepsilon_t | \Omega_{t-1})] = E[\sigma_t E(\varepsilon_t)] = E[\sigma_t \cdot 0] = 0, \\ \text{var}(X_t | \Omega_{t-1}) &= E(X_t^2 | \Omega_{t-1}) = E[\sigma_t^2 \varepsilon_t^2 | \Omega_{t-1}] = \sigma_t^2 E(\varepsilon_t^2 | \Omega_{t-1}) = \sigma_t^2, \\ \sigma_X^2 = \text{var}(X_t) &= E(\sigma_t^2) E(\varepsilon_t^2) = \alpha_0 + \sum_{i=1}^m \alpha_i E(X_{t-i}^2) + \sum_{j=1}^s \beta_j E(\sigma_{t-j}^2) \\ &= \alpha_0 + \sigma_X^2 \sum_{i=1}^m \alpha_i + \sigma_X^2 \sum_{j=1}^s \beta_j. \end{aligned}$$

Teda pre nepodmienený rozptyl σ_X^2 GARCH(m, s) procesu $\{X_t\}$ dostávame

$$\sigma_X^2 = \frac{\alpha_0}{1 - \sum_{i=1}^m \alpha_i - \sum_{j=1}^s \beta_j}. \quad (2.15)$$

Postačujúca podmienka pre existenciu rozptylu a zároveň pre slabú stacionaritu je

$$\sum_{i=1}^m \alpha_i + \sum_{j=1}^s \beta_j < 1. \quad (2.16)$$

Vidíme, že zovšeobecnenie modelu ARCH spočíva v pridaní ďalšieho regresoru do rovnice volatility, kde volatility procesu GARCH závisí na oneskorených hodnotách X_{t-j} a σ_{t-j}^2 , $j = 1, 2, \dots$ a rád tohto modelu je daný počtom týchto oneskorených hodnôt. Špeciálne model GARCH(1, 1) je najpoužívanejším modelom finančných časových radov a napriek svojej jednoduchosti je schopný veľmi dobre popísať všeobecné volatility štruktúry.

Tvar koeficientu špicatosti γ_2 pre GARCH modely nájdeme v knihe Tsay (2005, str. 114). Z predpokladov GARCH modelu o jeho parametroch vyplýva, že koeficient špicatosti $\gamma_2 > 0$, čo je v zhode s požiadavkou modelovať leptokurtické rozdelenie (a samozrejme rovnako ako ARCH model aj zhlukovanie volatility).

2.2.3 Odhady parametrov v ARCH a GARCH modeloch

Najpoužívanejšia metodika na odhad parametrov v ARCH a GARCH modeloch je metóda maximálnej vierohodnosti (MLE). Metóda najmenších štvorcov nie je vhodná napríklad z toho dôvodu, že za prítomnosti heteroskedasticity odhady pomocou OLS metódy nie sú nestranné. Nepoužívajú sa ale klasické MLE odhady z dôvodu zložitosti a takisto z toho dôvodu, že ak nesprávne použijeme pre model podmienené normálne rozdelenie (tj. $\varepsilon_t \sim N(0, 1)$), klasické MLE odhady budú väčšinou konzistentne odhadovať

len parametre modelu, ale už nie ich rozptyľové matice. Preto sa vo väčšine prípadov používajú tzv. QMLE odhady (*quasi-maximum likelihood*), ktoré počítajú odhady parametrov modelu podľa vzorcov, ktoré by sme dostali s maximálne vierohodným odhadom za predpokladu normality (preto sa tento postup označuje niekedy ako robustný voči normálnemu rozdeleniu). Tento postup navrhli T.Bollerslev a J.M.Wooldridge v roku 1992.

Ukážme si najprv postup odhadu parametrov pre ARCH model. Označme $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_m)$ ako parametre ARCH(m) modelu. Podmienehá hustota X_t za danej Ω_{t-1} (informácia obsiahnutá v procese do času $t - 1$) je v tvare

$$f_1(x_t) = \frac{1}{\sqrt{2\pi\sigma_t^2(\boldsymbol{\alpha})}} \exp\left(-\frac{x_t^2}{2\sigma_t^2(\boldsymbol{\alpha})}\right),$$

kde $\sigma_t(\boldsymbol{\alpha}) = \sigma_t$ z (2.12). Keďže ε_t sú nezávislé na $\Omega_{t-1} = \sigma(X_{t-1}, X_{t-2}, \dots)$, združená hustota (X_1, \dots, X_n) má tvar

$$f_n(x_1, \dots, x_n) = \prod_{t=m+1}^n \frac{1}{\sqrt{2\pi\sigma_t^2(\boldsymbol{\alpha})}} \exp\left(-\frac{x_t^2}{2\sigma_t^2(\boldsymbol{\alpha})}\right) \times f(\boldsymbol{\alpha}, x_1, \dots, x_m),$$

kde sa ďalej predpokladá, že x_1, \dots, x_m sú známe. Preto pre podmienenú (podmieňujeme x_1, \dots, x_m) logaritmickeú vierohodnostnú funkciu píšeme

$$l(\boldsymbol{\alpha}) = \sum_{t=m+1}^n \left(-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \sigma_t^2(\boldsymbol{\alpha}) - \frac{1}{2} \frac{x_t^2}{\sigma_t^2(\boldsymbol{\alpha})} \right) \quad (2.17)$$

a volatility vyjadrujeme pomocou $\alpha_0, \dots, \alpha_m$ rekurentne na základe vzťahu

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_m \varepsilon_{t-m}^2, \quad t = m + 1, \dots, n.$$

QMLE odhad parametra $\boldsymbol{\alpha}$ je potom definovaný ako vektor $\hat{\boldsymbol{\alpha}}$, ktorý maximalizuje $l(x_1, \dots, x_m | \boldsymbol{\alpha})$, čiže

$$\hat{\boldsymbol{\alpha}} = \operatorname{argmax}_{\boldsymbol{y} \in \Theta} l(x_1, \dots, x_m | \boldsymbol{y}),$$

kde Θ je nejaká prípustná množina.

Odhady v modeli GARCH sa takisto konštruujú QMLE metódou, s tým rozdielom, že v GARCH(m, s) modeli nemôžeme σ_t^2 zapísať ako súčet konečného počtu minulých pozorovaní X_{t-i} , a keďže v praxi máme len konečné množstvo pozorovaných hodnôt X_t , musíme σ_t^2 aproximovať určitou useknutou časťou procesu. Zvolme preto ν ako parameter, pre ktorý $\nu > m$, $\nu = \nu(n) \rightarrow \infty$ a $\nu/n \rightarrow 0$ pre $n \rightarrow \infty$. Potom podmienená logaritmickeá vierohodnostná funkcia pre odhad parametrov GARCH modelu má tvar

$$l(\boldsymbol{\theta}) = \sum_{t=\nu}^n \left(-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \sigma_t^2(\boldsymbol{\theta}) - \frac{1}{2} \frac{x_t^2}{\sigma_t^2(\boldsymbol{\theta})} \right),$$

kde $\boldsymbol{\theta} = (\alpha_0, \dots, \alpha_m, \beta_1, \dots, \beta_s)$ je vektor parametrov modelu GARCH(m, s).

2.2.4 Predpovede volatility

Už vieme, že podmienenému rozptylu σ_t^2 hovoríme volatilita v čase t . Zaoberajme sa znova iba modelom ARCH. Volatilita ARCH(m) modelu má tvar

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_m \varepsilon_{t-m}^2.$$

Volatilitu na základe tejto rovnice môžeme predpovedať rovnako ako v klasickom AR modeli. Predpokladajme teda, že poznáme hodnoty volatility až do času t . Hodnotu v čase $t + 1$, čiže predpoveď o jeden krok dopredu, vypočítame zo vzťahu

$$\hat{\sigma}_{t+1}^2(t) = \hat{\sigma}_{t+1}^2 = \alpha_0 + \alpha_1 \varepsilon_t^2 + \alpha_2 \varepsilon_{t-1}^2 + \dots + \alpha_m \varepsilon_{t+1-m}^2.$$

Výraz $\hat{\sigma}_{t+1}^2(t)$ znamená, že ide o predpoveď volatility σ_{t+1}^2 konštruovanú v čase t . Často sa uvádza alternatívny zápis v tvare $\sigma_{t+1}^2 = \hat{\sigma}_t^2(1)$, kde $\hat{\sigma}_t^2(1)$ značí predpoveď volatility známej do času t o jeden krok dopredu. Predpoveď o dva kroky dopredu získame z

$$\hat{\sigma}_{t+2}^2(t) = \alpha_0 + \alpha_1 \hat{\sigma}_{t+1}^2(t) + \alpha_2 \varepsilon_t^2 + \dots + \alpha_m \varepsilon_{t+2-m}^2.$$

Všeobecne predpoveď o k krokov dopredu vypočítame pomocou

$$\hat{\sigma}_{t+k}^2(t) = \alpha_0 + \sum_{i=1}^m \alpha_i \hat{\sigma}_{t+k-i}^2(t),$$

kde $\hat{\sigma}_{t+k-i}^2(t) = \varepsilon_{t+k-i}^2$ ak $k - i \leq 0$.

2.2.5 Modifikácie GARCH modelu

Veľkým nedostatkom pôvodného modelu GARCH je jeho neschopnosť modelovať pákový efekt (viď začiatok tejto kapitoly). Portfólio analýzy nelineárnych časových radov, ale obsahuje veľké množstvo modifikácií pôvodného modelu, ktoré si so spomínaným nedostatkom, a samozrejme s ďalšími inými, vedia poradiť. Spomenieme si aspoň niekoľko základných.

1. IGARCH

V knihe Cipra (2008) sa dočítame, že v základnom modeli GARCH, predpovede volatility $\hat{\sigma}_{t+\tau}^2(t)$ konvergujú pre $\tau > 0$, $\tau \rightarrow \infty$ k nepodmienenému rozptylu procesu. Už vieme, že model GARCH pri splnení podmienok (2.16) je stacionárny vo volatilita. V modeli IGARCH (všimnime si analógiu s modelom ARIMA) však informácia o volatilita v čase t pretrváva vo všetkých predpovedných horizontoch.

Tvar IGARCH(m, s) má rovnaké vyjadrenie ako model GARCH(m, s) s tým rozdielom, že pre model IGARCH platí $\sum_{i=1}^{\max\{m,s\}} (\alpha_i + \beta_i) = 1$, z čoho vyplýva, že nepodmienený rozptyl tohto procesu neexistuje.

2. GJR GARCH

Tento model sa na rozdiel od pôvodného modelu GARCH úspešne vysporiadáva s pákovým efektom. Úspešnú modifikáciu základného modelu GARCH navrhli Glosten, Jagannathan, a Runkle (1993), podľa ktorých sa model pomenoval ako GJR GARCH. Zápis GJR GARCH (1, 1) sa väčšinou uvádza v tvare,

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma_1 X_{t-1}^2 I_{t-1}^-, \quad (2.18)$$

kde $I_t^- = 1$ pre $X_t < 0$ a $I_t^- = 0$ pre $X_t \geq 0$. Význam poslednej sumy v modeli môžeme interpretovať tak, že „dobré správy“ ($X_{t-1} \geq 0$) zväčšia volatilitu menej, prostredníctvom α_i , ako „zlé správy“ ($X_{t-1} < 0$) prostredníctvom $\alpha_1 + \gamma_1$. Toto tvrdenie platí, ak $\gamma_1 > 0$ a ide skutočne o spomínaný pákový efekt.

3. GARCH-M

Vo financiách často závisí výnos aktíva od jeho volatility. Pokiaľ investor investuje do vysoko rizikového aktíva (zvýšené riziko sa prejaví zvýšenou volatilitou), očakáva samozrejme vyšší výnos. Z tohto dôvodu Engle, Lilien a Robins (1987) navrhli modifikáciu ARCH modelu za účelom zohľadnenia spomínaného javu a pridali volatilitu do rovnice strednej hodnoty. Rozšírením na GARCH model získame GARCH-M model. Napr. GARCH(1, 1)-M model má tvar

$$X_t = \gamma_1 \sigma_t^2 + e_t, \quad e_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

Parameter γ_1 je tzv. „risk premium parameter“ a pokiaľ je $\gamma_1 > 0$, zväčšená volatilita sa prejaví zvýšenou úrovňou radu.

4. EGARCH

Inú formu asymetrie, ako ponúka model GJR-GARCH, navrhol Nelson (1991). Tento model sa v literatúre uvádza vo viacerých formách, tu sa obmedzíme na EGARCH(1,1) v tvare:

$$X_t = \sigma_t \varepsilon_t, \quad \ln \sigma_t^2 = \alpha_0 + \beta_1 \ln \sigma_{t-1}^2 + g(\varepsilon_{t-1}),$$

kde

$$g(x) = \theta x + \gamma(|x| - E|x|), \quad (2.19)$$

kde $\alpha_0, \beta_1, \theta$ a γ sú reálne parametre. Asymetria funkcie $g(x)$ vyplýva z jej zápisu vo forme

$$g(\varepsilon_t) = \begin{cases} (\theta + \gamma)\varepsilon_t - \gamma E|\varepsilon_t| & \text{ak } \varepsilon_t \geq 0, \\ (\theta - \gamma)\varepsilon_t - \gamma E|\varepsilon_t| & \text{ak } \varepsilon_t < 0. \end{cases}$$

Poznámka 2.5. Pre chyby ε_t so štandardným normálnym rozdelením je $E|\varepsilon_t| = \sqrt{2/\pi}$.

Model EGARCH(1,1) môžeme zapísať v ekvivalentnom tvare

$$X_t = \sigma_t \varepsilon_t, \quad (1 - \beta_1 B) \ln \sigma_t^2 = \alpha_0 + g(\varepsilon_{t-1})$$

a pre $\ln \sigma_t^2$ dostávame

$$(1 - \beta_1 B) \ln \sigma_t^2 = \begin{cases} \alpha_*(\gamma + \theta) \varepsilon_{t-1} & \text{ak } \varepsilon_{t-1} \geq 0, \\ \alpha_*(\gamma - \theta) (-\varepsilon_{t-1}) & \text{ak } \varepsilon_{t-1} < 0, \end{cases}$$

kde $\alpha_* = \alpha_0 - \gamma \sqrt{2/\pi}$. Platí, že v tomto modeli sa priebeh volatility riadi nelineárne na základe znamienka X_{t-1} . Presnejšie platí

$$\sigma_t^2 = \sigma_{t-1}^{2\alpha} \exp(\alpha_*) \cdot \begin{cases} \exp(\gamma + \theta) \frac{X_{t-1}}{\sigma_{t-1}} & \text{ak } X_{t-1} \geq 0, \\ \exp(\gamma - \theta) \frac{|X_{t-1}|}{\sigma_{t-1}} & \text{ak } X_{t-1} < 0. \end{cases}$$

Vidíme, že priebeh volatility je rozdielne ovplyvnený znamienkom X_{t-1} pomocou koeficientov $(\gamma + \theta)$ a $(\gamma - \theta)$. Pákový efekt nastáva pokiaľ $\theta < 0$.

2.3 Odhady v neparametrickej autoregresii

Neparametrický prístup k analýze časových radov predstavuje iný spôsob, ako popisovať priebeh (finančných) časových radov. Pripomeňme si všeobecný zápis neparametrického modelu:

$$X_t = m(X_{t-1}, \dots, X_{t-p}) + \sigma(X_{t-1}, \dots, X_{t-p}) \varepsilon_t, \quad t = 1, \dots, n, \quad (2.20)$$

kde $m(\cdot)$ a $\sigma(\cdot)$ sú neznáme hladké funkcie (σ je nezáporná) a ε_t sú *iid* náhodné veličiny s nulovou strednou hodnotou a jednotkovým rozptylom. V parametrických metódach, kde predpokladáme, že funkcie $m(\cdot)$ a $\sigma(\cdot)$ majú určitú známu formu a závisia na parametri $\theta \in \mathbb{R}^k$, sa analýza zvyčajne uberať smerom k odhadu $\hat{\theta}$ parametra θ , ktorý je v istom zmysle optimálny (napr. metódou najmenších štvorcov alebo maximálnej vierohodnosti). Avšak pri použití neparametrických metód nekladíme na funkcie $m(\cdot)$ a $\sigma(\cdot)$ žiadne iné požiadavky ako hladkosť funkcií. Model (2.20) sa nazýva *nonparametric autoregressive conditional heteroscedastic* (NARCH) model alebo v prípade konštantnej $\sigma(\cdot)$ *nonparametric autoregressive* (NAR) model.

Avšak model (2.20) je užitočný iba v prípade, ak $p = 1$ alebo 2 . V prípade vyšších hodnôt p sa odhad neparametrickej formy funkcií $m(\cdot)$ a $\sigma(\cdot)$ stáva veľmi zložitý, pokiaľ nepredpokladáme obrovský počet pozorovaní. Ďalej budeme predpokladať, že $p = 1$.

Neparametrické modely sa nepoužívajú len v autoregresii. Základný princíp si môžeme predstaviť na klasickej regresii. Majme najjednoduchší model s dvoma (finančnými) premennými x_t a y_t , medzi ktorými je nejaký vzťah reprezentovaný *hladkou* funkciou m , čiže

$$Y_t = m(X_t) + \varepsilon_t, \quad t = 1, \dots, n, \quad (2.21)$$

kde $m(\cdot)$ je neznáma a $\{\varepsilon_t\} \sim IID(0, 1)$. Najčastejšou úlohou neparametrického prístupu je odhadnúť funkciu $m(\cdot)$ na základe pozorovaných dát X_1, \dots, X_n a Y_1, \dots, Y_n . Najjednoduchší odhad, ktorý vychádza zo zákona veľkých čísel ($\bar{\varepsilon}_n = \frac{1}{n} \sum \varepsilon_t \rightarrow 0$), predpokladá, že odhadom funkcie m je aritmetický priemer

$$\hat{m}(x) = \frac{1}{n} \sum_{t=1}^n Y_t. \quad (2.22)$$

Tento odhad už ale na prvý pohľad nie je veľmi dobrý, pretože v ňom nie sú zahrnuté hodnoty X_t , ktoré sa od rôznych hodnôt argumentu x môžu značne líšiť. Preto sa namiesto aritmetického priemeru použije vážený priemer

$$\hat{m}(x) = \frac{1}{n} \sum_{t=1}^n w_t(x) \cdot Y_t, \quad (2.23)$$

kde váhy $w_t(x)$ sa riadia podľa toho, ako blízko (ďaleko) je hodnota X_t od hodnoty x .

Z (2.23) vychádzajú tzv. *jadrové odhady*, kde jadro definujeme nasledovne.

Definícia 2.3. Nech pre ľubovoľnú symetrickú funkciu $K : \mathbb{R} \rightarrow \mathbb{R}$ platí

$$\int K(y) dy = 1$$

a

$$\int y^j K(y) dy = 0, \quad j = 1, \dots, l.$$

Potom K nazývame jadro rádu l .

Poznámka 2.6. Požiadavka na symetriu jadra nie je povinná. Väčšinou sa ale pracuje so symetrickými jadrami, preto túto požiadavku ponecháme.

My budeme pracovať iba s jadrami rádu 1. Zoberme si teda znova model (2.21) a skúmajme ho bližšie. Za účelom odvodenia odhadu funkcie $m(x)$ si zapíšeme túto funkciu v tvare:

$$m(x) = E(Y_t | X_t = x) = \int y f(y | x) dy = \frac{\int y f(x, y) dy}{\int f(x, y) dy},$$

kde $f(x, y)$ je združená hustota náhodných veličín X a Y . Vidíme, že k odhadu funkcie $m(x)$ nám postačí odhadnúť združenú hustotu $f(x, y)$. Budeme pracovať s neparametrickým odhadom združenej hustoty v tvare

$$\hat{f}(x, y) = \frac{1}{nh_x h_y} \sum_{t=1}^n K\left(\frac{x - X_t}{h_x}\right) K\left(\frac{y - Y_t}{h_y}\right) = \frac{1}{n} \sum_{t=1}^n K_{h_x}(x - X_t) K_{h_y}(y - Y_t),$$

kde $K(\cdot)$ je *jadro* a h_x , resp. h_y , je tzv. *šírka okna*, resp. *vyhladzovací parameter*. Všimnime si, že sme označili $K_h(\cdot) = K_h(\cdot/h)/h$. Tento zápis je veľmi častý v teórii neparametrických odhadov. Pomocou tohto odhadu dostávame

$$\int y \hat{f}(x, y) dy = \frac{1}{n} \int y \sum_{t=1}^n K_{h_x}(x - X_t) K_{h_y}(y - Y_t) dy.$$

Na základe vlastností z definície jadra $\int K(y)dy = 1$ a $\int yK(y)dy = 0$, potom takisto platí $\int yK_{h_y}(y - Y_t) dy = Y_t$ a čitateľ sa zjednoduší na

$$\int y \hat{f}(x, y) dy = \frac{1}{n} \sum_{t=1}^n K_{h_x}(x - X_t) Y_t.$$

Pre menovateľ podobným postupom získame

$$\int \hat{f}(x, y) dy = \frac{1}{n} \sum_{t=1}^n K_{h_x}(x - X_t) \int K_{h_y}(y - Y_t) dy = \frac{1}{n} \sum_{t=1}^n K_{h_x}(x - X_t) = \hat{f}(x).$$

Teda pre konečný neparametrický (jadrový) odhad $\hat{m}(x)$ platí:

$$\hat{m}(x) = \frac{\sum_{t=1}^n K_{h_x}(x - X_t) Y_t}{\sum_{t=1}^n K_{h_x}(x - X_t)}.$$

Tomuto odhadu sa hovorí *Nadaraya-Watsonov* odhad pre podmienenú strednú hodnotu. Pre autoregresný model položíme $Y_t = X_{t+1}$ a dostávame

$$\hat{m}(x) = \frac{\sum_{t=1}^{n-1} K_{h_x}(x - X_t) X_{t+1}}{\sum_{t=1}^{n-1} K_{h_x}(x - X_t)}. \quad (2.24)$$

Pre odhad podmieneného rozptylu sa ponúka použiť známy vzťah $\sigma^2(x) = E(X_{t+1}^2 | X_t = x) - m^2(x)$, ktorý vedie na odhad $\hat{\sigma}^2(x) = \hat{v}(x) - \hat{m}^2(x)$, kde $\hat{v}(x)$ a $\hat{m}(x)$ sú príslušné jednotlivé odhady pre $v(x) = E(X_{t+1}^2 | X_t = x)$ a $m(x)$. Avšak tento priamy postup vedie k tomu, že odhad $\hat{\sigma}^2(x)$ môže byť značne vychýlený (viac v Fan a Yao (1998, str. 8-9)). Preto Fan a Yao (1998) navrhli nasledujúce tvary odhadov funkcie $\sigma^2(x)$

$$\hat{\sigma}^2(x) = \frac{\sum_{t=1}^{n-1} K_{h_x}(x - X_t) (X_{t+1} - \hat{m}(X_t))^2}{\sum_{t=1}^{n-1} K_{h_x}(x - X_t)},$$

v prípade, že uvažujeme model s nenulovou podmienenou strednou hodnotou alebo

$$\hat{\sigma}^2(x) = \frac{\sum_{t=1}^{n-1} K_{h_x}(x - X_t) (X_{t+1} - \bar{X}_n)^2}{\sum_{t=1}^{n-1} K_{h_x}(x - X_t)},$$

v prípade modelu bez podmienenej strednej hodnoty $m(\cdot)$. V druhom výraze je $\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t$.

Kapitola 3

Parametrické metódy bootstrap

Nezávislý bootstrap, ktorý Efron (1979) predstavil vo svojej práci, má neustále veľké množstvo aplikácií na rôzne štatistické problémy. Avšak vtedajšia všeobecná mienka o bootstrape ako univerzálnom nástroji, ktorý dáva presné výsledky automaticky pre všetky problémy, bola značne zavádzajúca. Prvý dôkaz, ktorý vyvracal tento predpoklad, uviedol Singh (1981, str. 1187 - 1195), ktorý tvrdil, že použitie *iid* bootstrapu na závislé dáta môže viesť k nekonzistencii výsledkov (viď tiež Lahiri (2003, str. 21 - 22)). Odvtedy bolo navrhnutých veľa metód, ktoré závislosť dát zohľadňujú. Ukážeme si základné vlastnosti a princípy týchto metód, ktoré budeme v ďalších kapitolách využívať pri aplikáciách na finančné časové rady.

3.1 Reziduálny bootstrap

Ide o základnú a najpriamočiarejšiu metódu v prípade, že nemôžeme predpokladať nezávislé dáta. Spadá do skupiny parametrického bootstrapu, čiže je vhodný za predpokladu, že poznáme model, ktorým sa dáta riadia. Aby sme popísali postup metódy, budeme pracovať s modelom $AR(p)$

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (3.1)$$

kde $\{\varepsilon_t\}$ je striktný biely šum (tj. postupnosť nezávislých a rovnako rozdelených náhodných veličín s nulovou strednou hodnotou a konštantným konečným rozptylom $\sigma^2 > 0$), rád modelu p je známy, proces $\{X_t\}$ je stacionárny (viď kapitola 2). $T_n = T_n(X_1, \dots, X_n)$ bude znova štatistika, ktorej vlastnosti budeme chcieť odhadnúť. Základ metódy je založený na fakte, že pokiaľ použijeme konzistentné odhady pre neznáme parametre $\varphi_1, \dots, \varphi_p$, odhadnuté chyby (rezíduá) sa správajú ako „približne nezávislé“ náhodné veličiny a môžeme použiť metódu popísanú v 1. kapitole. Popíšeme detailne celý algoritmus pre dáta X_1, \dots, X_n , o ktorých vieme, že sa riadia modelom (3.1).

1. Odhadneme parametre modelu (3.1) niektorým z konzistentných odhadov a dostaneme odhad rezíduí v tvare

$$\hat{\varepsilon}_t = X_t - \hat{\varphi}_1 X_{t-1} - \dots - \hat{\varphi}_p X_{t-p}, \quad t = 1, \dots, n.$$

2. Pokiaľ pracujeme s modelom bez interceptu, skonštruujeme centrovane reziduá $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \bar{\varepsilon}$, kde $\bar{\varepsilon}$ je aritmetický priemer rezíduí $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$. Spočítame empirickú distribučnú funkciu F_n centrovaných rezíduí $\tilde{\varepsilon}_t$. (Centrovat reziduá môžeme, aj keď pracujeme s modelom s interceptom, ale nie je to nutné.)
3. Generujeme bootstrapové verzie $\varepsilon_1^*, \dots, \varepsilon_n^*$ rezíduí $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n$ pomocou funkcie F_n .
4. Skonštruujeme bootstrapové výbery podľa predpisu

$$X_t^* = \hat{\varphi}_1 X_{t-1}^* + \dots + \hat{\varphi}_p X_{t-p}^* + \varepsilon_t^*, \quad t = 1, \dots, n.$$

Počiatočné hodnoty môžeme zvoliť napríklad $X_{1-p}^* = \dots = X_0^* = 0$, resp. $X_{1-p}^* = X_{1-p}, \dots, X_0^* = X_0$.

V kroku 2 konštruujeme centrovane reziduá. Len spomenieme, že centrovanie je naozaj nutné v prípade predpokladu, že pracujeme s modelom bez interceptu, pretože inak by výsledná bootstrapová aproximácia vykazovala náhodnú chybu, a tým by sa celý postup stal nepoužiteľným.

3.2 Wild bootstrap

Reziduálny bootstrap sa na prvý pohľad môže javiť ako ideálny nástroj na analýzu časových radov. Avšak ukázalo sa, že v modeloch s (podmienenou) heteroskedasticitou reziduálny bootstrap dobre nekopíruje pôvodnú závislostnú štruktúru a navyše nekonzistentne odhaduje rozdelenie odhadov parametrov v modeli.

Wild bootstrap po prvý-krát zaviedol Wu (1986), ktorý metódu použil na regresný model s heterogénnymi chybami. Na časové rady, kde chyby nemôžeme považovať za nezávislé rovnako rozdelené náhodné veličiny, ju neskôr použila Liu (1988). Metóda sa používa v dvoch variantoch, ktoré sa líšia v predpise, podľa ktorého sa generuje bootstrapový výber. V oboch variantoch budeme pracovať s $AR(p)$ modelom, v ktorom tentokrát chyby e_t nebudeme považovať za nezávislé a rovnako rozdelené. Definujme teda model:

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + e_t, \quad t \in \mathbb{Z} \quad (3.2)$$

kde e_t vykazujú určitý druh (podmienenej) heteroskedasticity, napr. sa môžu riadiť modelom GARCH.

Poznámka 3.1. Model (3.2) s GARCH(m, s) chybami sa nazýva $AR(p)$ -GARCH(m, s) model.

Klasický wild bootstrap (regresná metóda)

Táto metóda generuje bootstrapové výbery regresne na základe rovnice

$$X_t^* = \hat{\varphi}_1 X_{t-1}^* + \dots + \hat{\varphi}_p X_{t-p}^* + e_t^*, \quad t = 1, \dots, n, \quad (3.3)$$

kde

$$e_t^* = \hat{e}_t \cdot \varepsilon_t,$$

a $\hat{e}_t = X_t - \hat{\varphi}_1 X_{t-1} - \dots - \hat{\varphi}_p X_{t-p}$ sú vypočítané rezíduá, pričom $\hat{\varphi}_1, \dots, \hat{\varphi}_p$ sú odhady parametrov modelu pomocou metódy najmenších štvorcov a ε_t je *iid* postupnosť náhodných veličín s nulovou strednou hodnotou a jednotkovým rozptylom, nezávislých na X_0, \dots, X_n .

Vidíme teda, že regresory X_{t-1}, \dots, X_{t-p} v bootstrapovom modeli (3.3) sú konštantné. Preto sa tomuto modelu hovorí *fixed-design wild bootstrap*.

Rekurzívny wild bootstrap (autoregresná metóda)

Na rozdiel od prvej metódy sa bootstrapové výbery negenerujú na základe regresie, ale na základe autoregresie, čiže

$$X_t^* = \hat{\varphi}_1 X_{t-1}^* + \dots + \hat{\varphi}_p X_{t-p}^* + e_t^*, \quad t = 1, \dots, n, \quad (3.4)$$

kde sa nový proces chýb znova riadi výrazom $e_t^* = \hat{e}_t \cdot \varepsilon_t$. Pre túto metódu potrebujeme ešte definovať počiatočné podmienky - napr v tvare $X_{1-p}^* = \dots = X_0^* = 0$. Tomuto modelu sa hovorí *recursive-design wild bootstrap*.

Výsledky štúdií ukazujú, že rekurzívny wild bootstrap vykazuje veľmi dobré výsledky pri aplikácii na veľké množstvo rôznych modelov s podmienenou heteroskedasticitou, čo si ukážeme aj v tejto práci. Obvykle je rekurzívny wild bootstrap viac presný ako klasický wild bootstrap pri malých počtoch pozorovaní a tiež sa ukázalo, že presnosť rekurzívneho wild bootstrapu je porovnateľná s klasickým reziduálnym bootstrapom, za predpokladu *iid* náhodných chýb. Navyše je metóda väčšinou presnejšia aj ako aproximácia normálnym rozdelením.

3.3 Párový bootstrap

Posledný parametrický prístup, ako odhadovať vlastnosti časových radov s podmienenou heteroskedasticitou, ktorý si predstavíme, je párový bootstrap (*Pairwise bootstrap*, *Pair(s) bootstrap*). Metóda spočíva v „bootstrapovaní párov“ vysvetľovanej a vysvetľujúcich premenných v autoregresii a je akýmsi rozšírením metódy do kontextu autoregresie, ktorú použil Freedman (1981) na odhad korelácie.

Pracujme s modelom (3.2) a označme

$$Z = \{(X_t, \mathbf{X}_{t-1}), t = 1, \dots, n\},$$

množinu všetkých párov pozorovaných dát X_1, \dots, X_n , kde $\mathbf{X}_{t-1} = (X_{t-1}, \dots, X_{t-p})$. Bootstrapové výbery generujeme z množiny Z klasickým prevzorkovaním s povolením opakovania prvkov. Získame množinu

$$Z^* = \{(X_t^*, \mathbf{X}_{t-1}^*), t = 1, \dots, n\},$$

kde X_t^* , $t = 1, \dots, n$ tvorí vektor vysvetľovanej premennej a \mathbf{X}_{t-1}^* , $t = 1, \dots, n$ sú vektory vysvetľujúcich premenných. Hodnotu štatistiky T_n^* potom počítame regresiou X_t^* na \mathbf{X}_{t-1}^* .

Kapitola 4

Neparametrické metódy bootstrap

4.1 Autoregresný bootstrap

Táto metóda na rozdiel od všetkých ostatných, ktoré sme spomínali v kapitole 3, pracuje s neparametrickým vyjadrením časového radu. V kapitole 2 sme uviedli, že v prípade neparametrických metód je teoretické či praktické odvodzovanie vlastností časového radu zložité pre veľké p . Preto sa obmedzíme na model

$$X_t = m(X_{t-1}) + \sigma(X_{t-1})\varepsilon_t, \quad t = 1, \dots, n, \quad (4.1)$$

kde $m(\cdot)$ a $\sigma(\cdot)$ sú neznáme hladké funkcie (σ je nezáporná) a ε_t sú *iid* náhodné veličiny s nulovou strednou hodnotou a jednotkovým rozptylom a neznámou hustotou p_ε .

Podobne ako parametrické metódy aj táto metóda potrebuje k svojej implementácii odhad podmienených momentov $m(\cdot)$ a $\sigma(\cdot)$. Keďže v parametrických metódach sme predpokladali, že funkcie $m(\cdot)$ a $\sigma(\cdot)$ majú určitý známy tvar, kde neznáme boli parametre, postačilo nám odhadnúť tieto parametre k implementácii prevzorkovacieho algoritmu, v neparametrických metódach ale tvar funkcií $m(\cdot)$ a $\sigma(\cdot)$ známy nie je. Preto musíme napr. metódami, ktoré sme uviedli v kapitole 2, odhadnúť tvar týchto funkcií.

Majme teda pozorovania X_1, \dots, X_n . Základná myšlienka metódy je kopírovať štruktúru pozorovaných dát z modelu (4.1) tak, že prevzorkovania konštruujeme na základe vzťahu

$$X_t^* = \tilde{m}(X_{t-1}^*) + \tilde{\sigma}(X_{t-1}^*)\varepsilon_t^*, \quad t = 1, \dots, n, \quad (4.2)$$

kde ε_t^* sú *iid* s hustotou \tilde{p}_ε a \tilde{m} , $\tilde{\sigma}$ a \tilde{p}_ε sú počiatkové (a tzv. nadhladené, *oversmoothed*) odhady m , σ a p_ε . Tvar týchto odhadov nájdeme v sekcii 5.6. Pri aplikácii tejto metódy sa musíme vyhnúť pozorovaniám, kde hustota X_{t-1} je blízka 0. Navyše odhady funkcií $m(\cdot)$ a $\sigma(\cdot)$, ktoré budeme používať pri generovaní bootstrapových dát, musia byť konštruované na základe vyhladzovacieho parametra g , ktorý konverguje k 0 pomalšie ako parameter h , na základe ktorého budeme konštruovať bootstrapové odhady $m^*(\cdot)$ a $\sigma^*(\cdot)$. V knihe Härdle et al. (2004, veta 4.3) je dokázané, že optimálny vyhladzovací parameter pre odhad $m(\cdot)$ je rádu $h \sim c.n^{-1/5}$. Uvedme si podrobne celý algoritmus:

1. Určíme vyhladzovací parameter h (pre rôzne postupy ako určiť tento parameter, vid Härdle (1990, str. 150-189), prípadne Fan a Yao (2003, str. 199-200)).
2. Stanovíme počiatočný vyhladzovací parameter g tak, aby $g_n/h_n \rightarrow \infty$, pri $n \rightarrow \infty$ (keďže platí $h_n \sim c.n^{-1/5}$, kde $c > 0$ je nejaká konštanta, a teda h závisí na počte pozorovaní n , stačí zvoliť $g \gg h$).
3. Použijeme parameter g na odhad \hat{m} a $\hat{\sigma}$ (vid Nadaraya-Watsonove odhady definované v sekcii 2.3). Na základe týchto odhadov vypočítame rezíduá v tvare

$$\hat{\varepsilon}_t = \frac{X_t - \hat{m}_g(X_{t-1})}{\hat{\sigma}_g(X_{t-1})}, \quad t = 2, \dots, n.$$

4. Určíme množinu $C_n = \{t; X_{(n,\alpha)} < X_{t-1} < X_{(n(1-\alpha))}\}$, kde $X_{(n,\alpha)}$ je $n\alpha$ -té pozorovanie v usporiadanom výbere pozorovaní a n je počet pozorovaní. Konštantu α volíme, väčšinou $\alpha = 0.01$.
5. Z rezíduí vypočítaných v kroku 3 vyberieme len také $\hat{\varepsilon}_t$, kde $t \in C_n$. Získame množinu rezíduí $\{\hat{\varepsilon}_t, t \in C_n\}$, a z nich vypočítame centrované rezíduá

$$\tilde{\varepsilon}_t = \hat{\varepsilon}_t - \bar{\varepsilon}_M,$$

kde $\bar{\varepsilon}_M = \frac{1}{M} \sum_{i=1}^M \hat{\varepsilon}_i$, a M je počet prvkov množiny C_n .

6. Z množiny $\{\tilde{\varepsilon}_t, t \in C_n\}$ klasickým výberom s vracaním nagenerujeme bootstrapové rezíduá ε_t^* , $t = 2, \dots, n$. Položíme $X_1^* = X_1$ a konštruujeme bootstrapové vzorky pomocou vzťahu (4.2). Na základe X_1^*, \dots, X_n^* už potom počítame hodnotu štatistiky T_n^* .

Poznámka 4.1. Pokiaľ je v kroku 6 objekt analýzy neparametrický odhad, musíme od tohto momentu pri počítaní bootstrapovej verzie odhadu použiť vyhladzovací parameter h . Dôvod použitia dvoch vyhladzovacích parametrov g a h , takisto ako dôvod konštrukcie množiny C_n vysvetlíme v kapitole 5.

4.2 Neparametrický wild bootstrap

Metóda podobne ako autoregresný bootstrap pracuje s neparametrickým vyjadrením časového radu v tvare

$$X_t = m(X_{t-1}) + \sigma(X_{t-1})\varepsilon_t, \quad t = 1, \dots, n.$$

Celý postup implementácie tejto metódy je rovnaký ako u autoregresného bootstrapu, až na spôsob, akým sa generujú bootstrapové rezíduá. Neparametrický wild bootstrap pracuje s rezíduami v tvare $\hat{\varepsilon}_t = X_t - \hat{m}_g(X_{t-1})$, z ktorých sa bez akéhokoľvek obmedzovania na množinu C_n konštruujú bootstrapové rezíduá v tvare

$$\varepsilon_t^* = \hat{\varepsilon}_t \cdot \xi_t, \quad t = 2, \dots, n,$$

kde ξ_t je *iid* postupnosť náhodných veličín s nulovou strednou hodnotou a jednotkovým rozptylom. Bootstrapové vzorky sa počítajú na základe

$$X_t^* = \tilde{m}(X_{t-1}) + \varepsilon_t^*, \quad t = 2, \dots, n,$$

kde môžeme položiť $X_1^* = X_1$ a \tilde{m} je opäť počiatočný (prvotný) odhad autoregresnej funkcie (presný tvar viď sekcia 5.6). Vidíme, že postup získavania bootstrapových rezíduí je rovnaký ako pre parametrický fixný wild bootstrap.

Existuje veľké množstvo ďalších bootstrapových postupov, ktoré sa dajú použiť v kontexte časových radov. Napríklad *sitový bootstrap*, ktorý aproximuje proces generovania dát (DGP, *data generating process*) autoregresným modelom rádu p , kde p rastie so zvyšujúcim sa počtom pozorovaní. Viac o tejto metóde nájdeme v Bühlmann (1997). Ďalšia metóda, ktorú zaviedol Rajarshi (1990), pracuje s faktom, že model (4.1) je Markovský proces. Metóda sa intuitívne volá Markovský bootstrap a jej základný postup spočíva v neparametrickom odhade Markovskej prechodovej hustoty, pomocou ktorej potom konštruujeme bootstrapové výbery prevzorkovaním rezíduí. Iný postup generovania prevzorkovaní pre Markovský bootstrap navrhli Papanaroditis, Politis (2001).

Kapitola 5

Bootstrap metódy a finančné časové rady

V tejto kapitole si ukážeme niektoré teoretické vlastnosti spomínaných bootstrapových metód a postupy, ako ich aplikovať na finančné časové rady pri ich analýze. Na základe pozorovaných dát, v našom prípade logaritmických mier zisku, sa analýza týchto radov zameriava hlavne na konštrukciu bodových alebo intervalových odhadov podmienenej strednej hodnoty m , resp. volatility σ . Bolo navrhnutých niekoľko postupov, ako tieto odhady získať, cez parametrické metódy, ktoré požadujú splnenie určitých predpokladov vzťahnutých na tieto funkcie, až po neparametrické metódy založené na veľmi slabých predpokladoch a znalostiach, ktoré sme schopní o m , resp. σ , sformulovať. Postupy na získavanie týchto odhadov nie sú náplňou tejto práce, aj keď niektoré základné postupy sme spomenuli v kapitole 2. Zameriame sa teda na priradovanie vlastností týmto odhadom.

5.1 Reziduálny bootstrap

5.1.1 Ako nepresný je reziduálny bootstrap pri heteroskedastických chybách?

Už sme spomenuli, že reziduálny bootstrap nie je najlepšou metódou na analýzu vlastností odhadov v modeli s podmienenou heteroskedasticitou. Ilustrujme si to na nasledujúcom príklade, kde reziduálny bootstrap porovnáme s wild bootstrapom na základe porovnania presnosti intervalových odhadov pre odhad parametra modelu AR(1).

Príklad 5.1. Uvažujme model AR(1)-ARCH(1) v tvare $X_t = \varphi_1 X_{t-1} + e_t$, $e_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2$, $\varepsilon_t \sim N(0, 1)$. Budeme konštruovať 90% percentilové intervaly v tvare $(\hat{G}_n^{-1}(\frac{\alpha}{2}), \hat{G}_n^{-1}(1 - \frac{\alpha}{2}))$ pre odhad $\hat{\varphi}_1$ parametra φ_1 konštruovaný pomocou OLS metódy za predpokladu nulovosti interceptu φ_0 . Percentilovú metódu sme zvolili z dôvodu nižšej výpočtovej náročnosti a takisto z dôvodu možnosti porovnania s výsledkami bootstrap- t intervalov, ktoré budú konštruované v simuláciách na konci práce. Kvantily pre jednotlivé intervaly spoľahlivosti budeme získať pomocou reziduálneho a wild bo-

otstrapu. Výsledky kvality pokrytia obidvoch metód popisuje nasledujúca tabuľka.

DGP: $X_t = \varphi_1 X_{t-1} + e_t$, $e_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2$, $\varepsilon_t \sim N(0, 1)$					
n	φ_1	α_1	Reziduálny bootstrap	Rekurzívny wild bootstrap	
50	0	0	89.8	90.0	
	0	0.5	79.0	85.7	
	0.9	0	85.7	81.8	
	0.9	0.5	83.5	84.0	
100	0	0	90.0	90.0	
	0	0.5	75.0	85.9	
	0.9	0	87.8	84.1	
	0.9	0.5	85.4	84.5	
200	0	0	90.2	90.1	
	0	0.5	71.7	87.0	
	0.9	0	88.1	86.9	
	0.9	0.5	85.6	86.0	

Tabuľka 5.1. Percentuálne pokrytie percentilových intervalov spoľahlivosti pre parameter φ_1 za predpokladu nulovosti interceptu φ_0 . Výsledky získané na základe 5000 simulácií, pre každú simuláciu generované 1000 bootstrapových replikácií.

Z tabuľky 5.1 vidíme, že pri $\varphi_1 = 0$ a $\alpha_1 = 0$, kedy sa model zredukuje na biely šum, dávajú obidve metódy rovnako dobré výsledky. Pri $\varphi_1 = 0$ sa model zredukuje na čistý ARCH(1) model a reziduálny bootstrap začína výrazne strácať na wild bootstrap, navyše z tabuľky vyplýva, že v tomto prípade sa nepresnosť reziduálnej metódy zvyšuje s rastúcim počtom pozorovaní. Pri $\alpha_1 = 0$ ide o AR(1) model s homoskedastickými chybami a reziduálna metóda má mierne navrch. Pri plnom AR(1)-ARCH(1) modeli znova reziduálna metóda zaostáva za wild bootstrapom.

Aby sme lepšie pochopili, prečo nie je reziduálny bootstrap schopný presnejšie kopírovať podmienenú heteroskedasticitu, sformulujeme nasledujúce tvrdenia. Budeme pracovať so stacionárnym modelom AR(p) v tvare (3.2). Kľúčová podmienka oprávnenosti a správnosti použitia akejkoľvek metódy bootstrap je schopnosť metódy konzistentne odhadnúť rozdelenie odhadu $\hat{\varphi}$ parametra $\varphi = (\varphi_1, \dots, \varphi_p)'$ získaného OLS metódou. Všeobecne teda musí platiť:

$$\sup_{x \in \mathbb{R}^p} \left| P^*(\sqrt{n}(T_n^* - T_n) \leq x) - P(\sqrt{n}(T_n - \varphi) \leq x) \right| \xrightarrow{P} 0,$$

kde v našom prípade, keďže pracujeme s modelom AR(p) a odhadujeme parametre modelu, položíme $T_n = \hat{\varphi}$ a $T_n^* = \hat{\varphi}^*$.

Definícia 5.1. (Martingal) Nech $\{\mathcal{F}_t\}$ je rastúca postupnosť, ktorú tvoria σ -algebry ($\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$) a $\{X_t\}$ je postupnosť náhodných veličín merateľných na \mathcal{F}_t . Potom

$\{X_t\}$ sa nazýva *martingal* vzhľadom k $\{\mathcal{F}_t\}$, ak platí

$$E(X_t | \mathcal{F}_{t-1}) = X_{t-1}.$$

Postupnosť $X_1, X_2 - X_1, X_3 - X_2, \dots$ sa nazýva *postupnosť martingalových diferencií*.

Predpokladajme teda, že máme pozorované dáta $\{X_0, X_1, \dots, X_n\}$, o ktorých vieme, že pochádzajú z modelu

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + e_t, \quad t \in \mathbb{Z},$$

kde $\{e_t, t \in \mathbb{Z}\}$ je *postupnosť martingalových diferencií* (m.d.s.) a $\varphi = (\varphi_1, \dots, \varphi_p)'$ je parameter, ktorý chceme pomocou metódy najmenších štvorcov odhadnúť (všimnime si, že nekladíme silnejší predpoklad nezávislosti náhodných veličín e_t). Budeme uvažovať nasledujúce predpoklady o inováciách $\{e_t\}$:

Predpoklady A1

1. $E(e_t | \mathcal{F}_{t-1}) = 0$ skoro iste, kde $\mathcal{F}_{t-1} = \sigma(e_{t-1}, e_{t-2}, \dots)$, je σ -algebra generovaná $\{e_{t-1}, e_{t-2}, \dots\}$;
2. $E(e_t^2) = \sigma^2 < \infty$;
3. $\frac{1}{n} \sum_{t=1}^n E(e_t^2 | \mathcal{F}_{t-1}) \xrightarrow{P} \sigma^2 > 0$ pre $n \rightarrow \infty$;
4. $\tau_{r,s} \equiv \sigma^{-4} E(e_t^2 e_{t-r} e_{t-s}) \leq K$ pre všetky $t, r, s \geq 1$ a $K > 0$, navyše platí $\tau_{r,r} > 0$ pre všetky r ;
5. $\frac{1}{n} \sum_{t=1}^n e_{t-r} e_{t-s} E(e_t^2 | \mathcal{F}_{t-1}) \xrightarrow{P} \sigma^4 \tau_{r,s} > 0$ pre ľubovoľné $r \geq 1, s \geq 1$ a $n \rightarrow \infty$;
6. $E|e_t|^{4r} \leq K$ pre nejaké $r > 1$ a $K > 0$.

Predpoklady A1 nahradzujú obvyklý predpoklad nezávislosti a rovnakého rozdelenia chýb $\{e_t\}$. V tomto tvare ich uviedli Goncalves a Kilian (2004) a poznamenajme, že povoluju závislé, ale nekorelované chyby. Vidíme, že nepredpisujú priamo podmienenú heteroskedasticitu, ale pokrývajú veľké množstvo modelov s podmienenou heteroskedasticitou ako ARCH, GARCH a modely stochastickej volatility. Ukážme teraz, že ARCH a GARCH modely spĺňajú predpoklady A1. Dôkaz predvedieme na GARCH(1, 1) modeli. Najprv budeme potrebovať nasledujúce definície a tvrdenia.

Definícia 5.2. (Ergodický proces) Nech (Ω, \mathcal{A}, P) je pravdepodobnostný priestor. Merateľná transformácia $T : \Omega \rightarrow \Omega$ sa nazýva transformácia zachováajúca mieru, ak $P(T^{-1}(A)) = P(A)$ pre všetky $A \in \mathcal{A}$ a jav A sa nazýva invariantný voči T ak $T^{-1}(A) = A$. Nech je ďalej ε σ -algebra všetkých invariantných javov voči T . Potom transformácia zachováajúca mieru T je ergodická, ak $P(E) = 1$ alebo $P(E) = 0$ pre každý $E \in \varepsilon$. Striktne stacionárny proces $\{X_t, t \in \mathbb{Z}\}$ je ergodický, ak

$$X_t(\omega) = X_1(T^{t-1}(\omega)),$$

kde $\omega \in \Omega$ a T je ergodická a mieru zachovávajúca transformácia.

Veta 5.1. (Ergodická) Nech veličiny $\{Z_t\}$ tvoria striktné stacionárny a ergodický proces s $E|Z_t| < \infty$. Potom pre $n \rightarrow \infty$ platí

$$\frac{1}{n} \sum_{t=1}^n Z_t \xrightarrow{a.s.} EZ_1.$$

Dôkaz: Dôkaz nájdeme v Stout (1974, veta 3.5.7). □

Tvrdenie 5.1. Nech veličiny $\{e_t\}$ tvoria GARCH(1, 1) proces v tvare

$$e_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

a nech $\varepsilon_t \sim IID(0, 1)$ a $\alpha_0 > 0$. Ak $E \{\log(\alpha_1 \varepsilon_t^2 + \beta_1)\} < 0$, potom σ_t^2 je striktné stacionárny a ergodický.

Dôkaz: Dôkaz tvrdenia môžeme nájsť v Nelson (1990, veta 2). □

Tvrdenie 5.1 hovorí, že podmienka v tvare $E \{\log(\alpha_1 \varepsilon_t^2 + \beta_1)\} < 0$ zaručuje striktnú stacionaritu a ergodicitu procesu $\{\sigma_t^2\}$, čiže aj striktnú stacionaritu a ergodicitu procesu $\{e_t\}$ (viď tiež Bougerol a Picard (1992, veta 1.3)). Uvedme pre úplnosť, že keďže podmienka existencie konečného rozptylu modelu GARCH(1, 1) je $\alpha_1 + \beta_1 < 1$, tak vzhľadom k Jensenovej nerovnosti a striktnej konkávnosti funkcie $\log(\cdot)$ nerovnosť $E \{\log(\alpha_1 \varepsilon_t^2 + \beta_1)\} < \log E(\alpha_1 \varepsilon_t^2 + \beta_1) = \log(\alpha_1 E\varepsilon_t^2 + \beta_1) = \log(\alpha_1 + \beta_1) \leq 0$, platí aj pre $\alpha_1 + \beta_1 = 1$, čiže proces $\{e_t\}$ je striktné stacionárny aj v prípade, keď druhý moment procesu nie je konečný. Takýmto procesom sa hovorí integrované GARCH (IGARCH) procesy, ktoré sme spomenuli v sekcii 2.2.5. Pre dokázanie vety, ktorá zaručuje, že GARCH procesy spĺňujú predpoklady A1, ale musíme predpokladať ďalšiu vlastnosť, z ktorej plynie konečnosť momentov procesu.

Tvrdenie 5.2. Predpokladajme, že veličiny $\{e_t\}$ sa riadia GARCH(1, 1) procesom v tvare

$$e_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

a nech $\alpha_0 > 0$ a $p > 0$ a $E \{\log(\alpha_1 \varepsilon_t^2 + \beta_1)\} < 0$, potom $E \sigma_t^{2p} < \infty$ vtedy a len vtedy, ak $E \{(\alpha_1 \varepsilon_t^2 + \beta_1)^p\} < 1$.

Dôkaz: Tvrdenie je dokázané v Nelson (1990, str. 322). □

Na základe predošlých tvrdení dokážeme nasledujúcu vetu.

Veta 5.2. Predpokladajme, že sa náhodné veličiny $\{e_t\}$ riadia GARCH(1, 1) procesom v tvare

$$e_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

kde ε_t sú nezávislé a rovnako rozdelené náhodné veličiny so štandardným normálnym rozdelením. Nech ďalej $\alpha_0, \alpha_1 > 0, \beta_1 \geq 0$ a $E \{\log(\alpha_1 \varepsilon_t^2 + \beta_1)\} < 0$ a $E \{(\alpha_1 \varepsilon_t^2 + \beta_1)^4\} < \infty$.

1. Potom $\{e_t\}$ splňujú predpoklady A1.

Dôkaz: Bod (1) plynie veľmi ľahko, pretože

$$\mathbb{E}(e_t \mid e_{t-1}, e_{t-2}, \dots) = \mathbb{E}(\sigma_t \varepsilon_t \mid e_{t-1}, e_{t-2}, \dots) = \sigma_t \mathbb{E}(\varepsilon_t \mid e_{t-1}, e_{t-2}, \dots),$$

a keďže ε_t je nezávislé na e_{t-1}, e_{t-2}, \dots už z definície GARCH procesu, platí

$$\sigma_t \mathbb{E}(\varepsilon_t \mid e_{t-1}, e_{t-2}, \dots) = \sigma_t \mathbb{E}(\varepsilon_t) = 0,$$

a teda GARCH proces je m.d.s. Z definície GARCH procesu je zrejmé, že stacionaritu a vlastnosti momentov σ_t môžeme vziať na e_t . Preto bod (2) nám zaistuje nasledujúci vzťah:

$$\begin{aligned} \sigma^2 &= \text{var}(e_t) = \mathbb{E}e_t^2 = \mathbb{E}\sigma_t^2 \mathbb{E}\varepsilon_t^2 = \alpha_0 + \alpha_1 \mathbb{E}e_{t-1}^2 + \beta_1 \mathbb{E}\sigma_{t-1}^2 \\ &= \alpha_0 + \sigma^2 \alpha_1 + \sigma^2 \beta_1. \end{aligned}$$

Teda

$$\mathbb{E}e_t^2 \equiv \sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1},$$

kde $\sigma^2 < \infty$ plynie z predpokladu vety, že $\mathbb{E}\{(\alpha_1 \varepsilon_t^2 + \beta_1)^4\} < 1$ a tvrdenia 5.2. Bod (3) plynie z ergodickej vety - keďže pre striktné stacionárny a ergodický proces podľa ergodickej vety platí, že $\frac{1}{n} \sum_{t=1}^n Z_t \xrightarrow{\text{a.s.}} \mathbb{E}Z_t$ a

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E}(e_t^2 \mid e_{t-1}, e_{t-2}, \dots) = \frac{1}{n} \sum_{t=1}^n \mathbb{E}(\sigma_t^2 \varepsilon_t^2 \mid e_{t-1}, e_{t-2}, \dots) = \frac{1}{n} \sum_{t=1}^n \sigma_t^2,$$

dostávame pomocou tvrdenia 5.1, že

$$\frac{1}{n} \sum_{t=1}^n \sigma_t^2 \xrightarrow{\text{a.s.}} \mathbb{E}\sigma_1^2 = \mathbb{E}e_1^2 = \sigma^2,$$

z čoho (keďže je konvergencia takmer iste silnejší typ konvergenie ako konvergencia v pravdepodobnosti) takisto vyplýva konvergencia v pravdepodobnosti. Kladná hodnota σ^2 je zaručená podmienkami na parametre modelu. Zostáva teda dokázať bod 4,5 a 6. Dokážme najprv bod (6), ktorý plynie z tvrdenia 2 a predpokladu tejto vety, že $\mathbb{E}\{(\alpha_1 \varepsilon_t^2 + \beta_1)^4\} < 1$. Z tvrdenia 2 teda máme: ($p = 4$)

$$\mathbb{E}\{(\alpha_1 \varepsilon_t^2 + \beta_1)^4\} < 1 \Rightarrow \mathbb{E}\sigma_t^8 < \infty,$$

a vzhľadom ku striktnej stacionarite σ_t^2 platí $\mathbb{E}\sigma_t^8 = K_1 < \infty$. Pomocou Hölderovej nerovnosti dostaneme

$$\mathbb{E}|e_t|^{4r} = \mathbb{E}|\sigma_t \varepsilon_t|^{4r} = \mathbb{E}|\sigma_t^{4r} \varepsilon_t^{4r}| \leq \left[\mathbb{E}|\sigma_t|^8\right]^{\frac{r}{2}} \left[\mathbb{E}|\varepsilon_t|^{\frac{8r}{2-r}}\right]^{\frac{2-r}{2}},$$

kde prvý člen na pravej strane nerovnosti je menší ako $r/2$ -tá mocnina $K_1 < \infty$ a druhý menší ako $K_2 < \infty$. Ak položíme $K \equiv K_1^{r/2} \cdot K_2$, dostávame požadovanú nerovnosť.

Bod (4) dokážeme takto: keďže $\tau_{r,s} \equiv \sigma^{-4} \mathbb{E}(e_t^2 e_{t-r} e_{t-s})$, dvojitým aplikovaním *Cauchy-Schwarzovej nerovnosti* dostávame

$$\sigma^4 \tau_{r,s} \leq [\mathbb{E}(e_t^4)]^{\frac{1}{2}} [\mathbb{E}(e_{t-r}^2 e_{t-s}^2)]^{\frac{1}{2}} \leq [\mathbb{E}(e_t^4)]^{\frac{1}{2}} \left\{ [\mathbb{E}(e_{t-r}^4)]^{\frac{1}{2}} [\mathbb{E}(e_{t-s}^4)]^{\frac{1}{2}} \right\}^{\frac{1}{2}}. \quad (5.1)$$

Z Hölderovej nerovnosti takisto vyplýva, že

$$\mathbb{E} |e_t|^4 \leq (\mathbb{E} |e_t|^{4r})^{1/r}$$

a keďže z vlastnosti (6) platí $\mathbb{E} |e_t|^{4r} \leq K$ pre $r > 1$, tak pre $r = 1$ platí $\mathbb{E}(e_t^4) \leq K_1$ pre nejakú kladnú konštantu $K_1 < \infty$. Tým sme zhora obmedzili členy v hranatých zátvorkách na pravej strane nerovnosti v (5.1). Čiže $\tau_{r,s} \leq K$ pre $t, r, s \geq 1$ a $K > 0$. Pre $r = s$ máme

$$\tau_{r,r} = \frac{1}{\sigma^4} \mathbb{E}(e_t^2 e_{t-r}^2) = \frac{1}{\sigma^4} \mathbb{E}(e_{t-r}^2 \mathbb{E}(e_t^2 | \mathcal{F}_{t-r}))$$

a pre $r \geq 1$

$$\mathbb{E}(e_t^2 | \mathcal{F}_{t-r}) = \mathbb{E}(\mathbb{E}(e_t^2 | \mathcal{F}_{t-1}) | \mathcal{F}_{t-r}) = \mathbb{E}(\alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2 | \mathcal{F}_{t-r}) \geq \alpha_0 > 0,$$

keďže $\alpha_0 > 0$, $\alpha_1 \geq 0$ a $\beta_1 \geq 0$. Nakoniec bod (5) plynie z rovnosti

$$\frac{1}{n} \sum_{t=1}^n e_{t-r} e_{t-s} \mathbb{E}(e_t^2 | \mathcal{F}_{t-1}) = \frac{1}{n} \sum_{t=1}^n e_{t-r} e_{t-s} \sigma_t^2,$$

a z faktu, že σ_t^2 sa dá zapísať ako funkcia $\{e_{t-1}, e_{t-2}, \dots\}$. Keďže proces e_t je stacionárny a ergodický, potom aj $f(e_{t-1}, e_{t-2}, \dots)$ je stacionárny a ergodický proces podľa (Stout (1974, veta 3.5.8)). Z tej istej vety ale vyplýva, že aj proces $\{e_{t-r} e_{t-s} \sigma_t^2\}$ je striktnie stacionárny a ergodický pre ľubovoľné $r \geq 1$, $s \geq 1$ a teda môžeme aplikovať ergodickú vetu

$$\frac{1}{n} \sum_{t=1}^n e_{t-r} e_{t-s} \mathbb{E}(e_t^2 | \mathcal{F}_{t-1}) \xrightarrow{a.s.} \mathbb{E}(e_{t-r} e_{t-s} \mathbb{E}(e_t^2 | \mathcal{F}_{t-1})) = \mathbb{E}(e_{t-r} e_{t-s} e_t^2) = \sigma^4 \tau_{r,s} > 0,$$

pre $r \geq 1$, $s \geq 1$. □

Vráťme sa späť k modelu (3.2). Budeme chcieť ukázať, že reziduálny bootstrap nekonzistentne odhaduje rozdelenie odhadu $\hat{\varphi}$ pri podmienene heteroskedastických chybách. Aby sme ale tvrdenie o konzistencii, resp. nekonzistencii, bootstrapového odhadu mohli vysloviť, musíme najprv vedieť, ako vyzerá asymptotické rozdelenie odhadu $\hat{\varphi}$ za platnosti A1.1-A1.6. To nám osvetlí nasledujúca veta.

Asymptotické vlastnosti OLS odhadu

Zavedme si najprv označenie $b_j = (\theta_{j-1}, \dots, \theta_{j-p})'$, $\varphi^{-1}(B) = \sum_{j=0}^{\infty} \theta_j B^j$, kde $\varphi(B) \cdot \theta(B) = 1$ a $\theta_0 = 1$. OLS odhad pre parameter φ v modeli (3.2) má tvar

$$\hat{\varphi} = \left(\sum_{t=1}^n \mathbf{X}_{t-1} \mathbf{X}'_{t-1} \right)^{-1} \sum_{t=1}^n \mathbf{X}_{t-1} X_t.$$

Veta 5.3. Za platnosti predpokladov A1.1 - A1.6 platí:

$$\sqrt{n}(\hat{\varphi} - \varphi) \xrightarrow{D} N(0, \mathbf{C}),$$

kde

$$\begin{aligned} \mathbf{C} &= \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}, \\ \mathbf{A} &= \sigma^2 \sum_{j=1}^{\infty} b_j b_j', \quad \mathbf{B} = \sigma^4 \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} b_i b_j' \tau_{i,j}. \end{aligned}$$

Dôkaz: viď Goncalves a Kilian (2004, str. 108-110). □

Vidíme teda, že rozdelenie $\hat{\varphi}$ je asymptoticky normálne s kovariančnou maticou \mathbf{C} . Ako sme už písali o pár riadkov vyššie, podmienka správnosti použitia metódy bootstrap, je jej schopnosť konzistentne odhadnúť rozdelenie odhadu $\hat{\varphi}$. Túto skutočnosť môžeme zistiť práve podľa schopnosti správne určiť asymptotické rozdelenie $\hat{\varphi}$.

Poznámka 5.1. Keby chyby v modeli boli podmienené homoskedastické, platila by rovnosť $\mathbf{B} = \sigma^2 \mathbf{A}$, pretože by platilo

$$\tau_{i,i} = \sigma^{-4} \mathbf{E}(e_t^2 e_{t-i}^2) = \sigma^{-4} \mathbf{E}(e_{t-i}^2 \mathbf{E} e_t^2 | \mathcal{F}_{t-1}) = \sigma^{-4} \mathbf{E}(e_{t-i}^2 \sigma^2) = \sigma^{-4} \sigma^4 = 1$$

pre všetky $i = 1, 2, \dots$. Podobne sa dá ukázať, že $\tau_{i,j} = 0$ pre všetky $i \neq j$. Napr. stacionárny proces AR(1) môžeme zapísať v tvare lineárneho procesu, a platí $\varphi(B)X_t = e_t$, $X_t = \varphi(B)^{-1}e_t$ a keďže postupnou rekurziou $X_t = e_t + \varphi_1 e_{t-1} + \varphi_1^2 e_{t-2} + \dots$ dostávame $X_t = \sum_{j=0}^{\infty} \varphi_1^j e_{t-j} = \sum_{j=0}^{\infty} \theta_j e_{t-j}$, z čoho vyplýva, že asymptotická kovariančná matica sa zredukuje na tvar

$$\mathbf{C} = \left(\sigma^2 \sum_{j=0}^{\infty} \theta_j^2 \right)^{-1} \cdot \left(\sigma^4 \sum_{j=0}^{\infty} \theta_j^2 \right) \cdot \left(\sigma^2 \sum_{j=0}^{\infty} \theta_j^2 \right)^{-1} = 1 / \sum_{j=0}^{\infty} \theta_j^2 = 1 / \sum_{j=0}^{\infty} \varphi_1^{2j} = 1 - \varphi_1^2.$$

5.1.2 Asymptotické vlastnosti reziduálneho bootstrapu

Štandardný reziduálny bootstrap konštruje prevzorkovania na základe vzťahu

$$X_t^* = \hat{\varphi}_1 X_{t-1}^* + \dots + \hat{\varphi}_p X_{t-p}^* + e_t^*, \quad t = 1, \dots, n \quad (5.2)$$

pri určitých počiatkových podmienkach a podobne ako u pôvodného modelu aj pre model (5.2) si napíšeme matice

$$\mathbf{A}^* = \frac{1}{n} \sum_{t=1}^n \mathbf{E}^*(\mathbf{X}_{t-1}^* \mathbf{X}_{t-1}^{*'}),$$

$$\mathbf{B}^* = \text{var}^* \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbf{X}_{t-1}^* e_t^*).$$

Pretože na základe

$$\begin{aligned} \mathbf{E}^* e_1^* &= \frac{1}{n} \sum_{t=1}^n \tilde{e}_t = \frac{1}{n} \sum_{t=1}^n (\hat{e}_t - \bar{e}) = 0 \\ \text{var}^* e_1^* &= \frac{1}{n} \sum_{t=1}^n \tilde{e}_t^2 = \frac{1}{n} \sum_{t=1}^n \hat{e}_t^2 - \bar{e}^2 \equiv \hat{\sigma}^2, \end{aligned}$$

platí, že rezíduá ε_t^* sú nezávislé a rovnako rozdelené náhodné veličiny s rozdelením $\varepsilon_t^* \sim IID(0, \hat{\sigma}^2)$ a \mathbf{X}_{t-1}^* sú (podmienene) nezávislé, platí

$$\mathbf{B}^* = \frac{1}{n} \sum_{t=1}^n \mathbf{E}^*(\mathbf{X}_{t-1}^* \mathbf{X}_{t-1}^{*'} e_t^{*2}) = \frac{1}{n} \sum_{t=1}^n \mathbf{E}^*(\mathbf{X}_{t-1}^* \mathbf{X}_{t-1}^{*'}) \mathbf{E}^*(e_t^{*2}) = \hat{\sigma}^2 \mathbf{A}^*.$$

Z rovnosti vyššie vyplýva, že matica $\mathbf{C}^* = \mathbf{A}^{*-1} \mathbf{B}^* \mathbf{A}^{*-1} = \hat{\sigma}^2 \mathbf{A}^{*-1}$ a v pravdepodobnosti konverguje k $\sigma^2 \mathbf{A}^{-1}$ (viď Goncalves a Kilian 2004). Z toho vyplýva, že limitné rozdelenie $\sqrt{n}(\hat{\varphi}^* - \hat{\varphi})$ je $N(0, \sigma^2 \mathbf{A}^{-1})$, čo ale nie je to správne rozdelenie podľa vety 5.3.

Poznámka 5.2. Rozdelenie $N(0, \sigma^2 \mathbf{A}^{-1})$ by bolo správne, keby sme navyše predpokladali podmienenú homoskedasticitu. Pre model AR(1) s podmienene homoskedastickými chybami platí $\sqrt{n}(\hat{\varphi} - \varphi) \xrightarrow{D} N(0, 1 - \varphi_1^2)$.

5.2 Wild bootstrap

5.2.1 Rekurzívny wild bootstrap

Teraz budeme chcieť ukázať, že metóda na rozdiel od reziduálnej metódy konzistentne odhaduje rozdelenie odhadov parametra $\varphi = (\varphi_1, \dots, \varphi_p)$ v AR(p) modeli (3.2) získaných metódou najmenších štvorcov aj za prítomnosti podmienene heteroskedastických chýb.

Pripomeňme si ešte raz spôsob, akým rekurzívny wild bootstrap generuje prevzorovania:

$$X_t^* = \hat{\varphi}_1 X_{t-1}^* + \dots + \hat{\varphi}_p X_{t-p}^* + e_t^* \quad t = 1, \dots, n,$$

kde $e_t^* = \hat{e}_t \varepsilon_t$, \hat{e}_t sú odhadnuté rezíduá a $\varepsilon_t \sim IID(0, 1)$. Odhad parametra $\hat{\varphi}$ na základe rekurzívneho wild bootstrapu má tvar

$$\hat{\varphi}^* = \left(\sum_{t=1}^n \mathbf{X}_{t-1}^* \mathbf{X}_{t-1}^{*'} \right)^{-1} \sum_{t=1}^n \mathbf{X}_{t-1}^* X_t^*.$$

Aby sme dokázali asymptotickú validitu (konzistentnosť) metódy, budeme musieť sprísniť požiadavky, ktoré kladieme na inovácie e_t . Spolu s predpokladmi A1 potom pre metódu dokážeme, že bootstrapový odhad asymptotickej kovariančnej matice C konverguje k správnej limite.

Predpoklady A2

1. $\mathbf{E}(e_t^2 e_{t-r} e_{t-s}) = 0$ pre všetky $r \neq s$ a všetky $t, r, s \geq 1$;

2. $E|e_t|^{4r} \leq K$ pre nejaké $r \geq 2$, $K > 0$ a všetky t .

Poznámka 5.3. Ako ukážeme nižšie, na dokázanie konzistentnosti fixného wild bootstrapu a párového bootstrapu stačia iba predpoklady A1, kým predpoklady A2 sú nevyhnutné pre odvodenie konzistencie rekurzívneho wild bootstrapu.

Predpoklad A2.1 požaduje, aby $\tau_{r,s} = 0$ pre všetky $r \neq s$, čo splňuje ARCH(p) model s inováciami, ktoré majú symetrické rozdelenie (viď Milhoj (1985)). Neskôr sa platnosť tohto tvrdenia rozšírila aj na prípad GARCH(p, q) modelov so symetrickým rozdelením inovácií. Z toho ale vyplýva, že A2.1 vylučuje niektoré asymetrické modely, napr. model EGARCH v tvare 2.19. Predpoklad A2.2 požaduje existenciu momentov aspoň do rádu 8 (predpoklad A1.6 požaduje existenciu len do rádu $4r$ pre $r > 1$).

Veta 5.4. (Konzistencia rekurzívneho wild bootstrapu) Za platnosti predpokladov A1 a A2 platí

$$\sup_{x \in \mathbb{R}^p} \left| P^*(\sqrt{n}(\hat{\varphi}_{rwb}^* - \hat{\varphi}) \leq x) - P(\sqrt{n}(\hat{\varphi} - \varphi) \leq x) \right| \xrightarrow{P} 0,$$

kde P^* značí pravdepodobnostnú mieru indukovanú (rekurzívnym wild) bootstrapom a $\hat{\varphi}_{rwb}^*$ je bootstrapový ekvivalent $\hat{\varphi}$.

Dôkaz: viď Goncalves a Kilian (2004, str. 110). □

5.2.2 Fixný wild bootstrap

Na rozdiel od rekurzívneho wild bootstrapu, na dokázanie schopnosti fixného wild bootstrapu konzistentne odhadovať rozptyl, a tým pádom aj limitné rozdelenie $\hat{\varphi}$, nepotrebujeme žiadne iné predpoklady ako predpoklady A1. Odhad parametra $\hat{\varphi}$ na základe fixného wild bootstrapu má tvar,

$$\hat{\varphi}^* = \left(\sum_{t=1}^n \mathbf{X}_{t-1} \mathbf{X}'_{t-1} \right)^{-1} \sum_{t=1}^n \mathbf{X}_{t-1} X_t^*.$$

Vidíme, že bootstrapová verzia odhadu parametru v AR(p) modeli sa počíta regresne, kde regresory sú pôvodné dáta.

Veta 5.5. (Konzistencia fixného wild bootstrapu) Za platnosti predpokladov A1 platí

$$\sup_{x \in \mathbb{R}^p} \left| P^* \sqrt{n}(\hat{\varphi}_{fwb}^* - \hat{\varphi}) \leq x) - P(\sqrt{n}(\hat{\varphi} - \varphi) \leq x) \right| \xrightarrow{P} 0,$$

kde P^* značí pravdepodobnostnú mieru indukovanú (fixným wild) bootstrapom a $\hat{\varphi}_{fwb}^*$ je bootstrapový ekvivalent $\hat{\varphi}$.

Dôkaz: viď Goncalves a Kilian (2004, str.110-111). □

5.3 Párový bootstrap

Veľmi dobré výsledky pri podmienene heteroskedastických chybách dosahuje aj párový bootstrap. Základný princíp sme popísali v kapitole 3 a takisto platí, že metóda dokáže konzistentne odhadnúť limitné rozdelenie OLS odhadu v autoregresnom modeli, kde pre chyby platia predpoklady A1. Výpočet odhadu parametra $\hat{\varphi}$ na základe párového bootstrapu má tvar

$$\hat{\varphi}^* = \left(\sum_{t=1}^n \mathbf{X}_{t-1}^* \mathbf{X}_{t-1}^{*'} \right)^{-1} \sum_{t=1}^n \mathbf{X}_{t-1}^* X_t^*,$$

kde ale vysvetľovaná premenná a regresory netvoria jeden časový rad, ako je to v prípade reziduálneho alebo rekurzívneho wild bootstrapu. Keďže jednotlivé prevzorkovania sa vytvárajú náhodným výberom s vracaním z „párov“ vysvetľovanej a vysvetľujúcich premenných, bootstrapové prevzorkovania vysvetľovanej premennej tvoria jeden vektor a bootstrapové prevzorkovania príslušných opozdených hodnôt tvoria samostatné vektory. Odhad $\hat{\varphi}^*$ teda počítame regresiou X_t^* na \mathbf{X}_{t-1}^* .

Rovnako ako v prípade fixného wild bootstrapu, na dokázanie konzistencie párového bootstrapu nám postačia predpoklady A1.

Veta 5.6. (Konzistencia párového bootstrapu) Za platnosti predpokladov A1 platí

$$\sup_{x \in \mathbb{R}^p} \left| P^*(\sqrt{n}(\hat{\varphi}_{pb}^* - \hat{\varphi}) \leq x) - P(\sqrt{n}(\hat{\varphi} - \varphi) \leq x) \right| \xrightarrow{P} 0,$$

kde P^* značí pravdepodobnostnú mieru indukovanú (párovým) bootstrapom a $\hat{\varphi}_{pb}^*$ je bootstrapový ekvivalent $\hat{\varphi}$.

Dôkaz: viď Goncalves a Kilian (2004, str. 111-112). □

5.4 Autoregresný bootstrap

V tejto sekcii si uvedieme niektoré vlastnosti neparametrického časového radu 5.3, ktoré budeme potrebovať k zaisteniu konzistentnosti autoregresného bootstrapu. Najprv si zavedieme tieto vlastnosti. Model, s ktorým budeme pracovať, je

$$X_t = m(X_{t-1}) + \sigma(X_{t-1})\varepsilon_t, \quad t = 1, \dots, n, \quad (5.3)$$

kde p_ε je hustota veličín ε_t , $\varepsilon_t \sim IID(0, 1)$ a tvar bootstrapového procesu generovania replikácií je v tvare

$$X_t^* = \tilde{m}(X_{t-1}^*) + \tilde{\sigma}(X_{t-1}^*)\varepsilon_t^*, \quad t = 1, \dots, n, \quad (5.4)$$

kde ε_t^* sú *iid* náhodné veličiny s hustotou \tilde{p}_ε a podmieneným rozdelením \tilde{P}_ε a \tilde{m} , $\tilde{\sigma}$ sú prvotné odhady m , σ . Tvar týchto prvotných odhadov spolu s tvarom \tilde{P}_ε si ukážeme v sekcii 5.6 venovanej ich konštrukcii. Najprv si zavedme všeobecný tvar odhadov funkcií m , σ .

Pre odhady funkcií m , σ budeme uvažovať klasické Nadaraya-Watson jadrové odhady v tvare

$$\hat{m}_h(x) = \frac{\sum_{t=1}^{n-1} K_h(x - X_t) X_{t+1}}{\sum_{t=1}^{n-1} K_h(x - X_t)},$$

$$\hat{\sigma}_{h'}^2(x) = \frac{\sum_{t=1}^n K_{h'}(x - X_t) (X_{t+1} - \hat{m}_{h'}(X_t))^2}{\sum_{t=1}^n K_{h'}(x - X_t)}.$$

Aby sme mohli vysloviť tvrdenia týkajúce sa správnosti bootstrapovej metódy, musíme podobne ako v parametrických metódach predpokladať určité „dobré“ vlastnosti o (5.3). V AR(p)-GARCH(m, s) modeloch sme predpokladali stacionaritu. V neparametrických modeloch budeme potrebovať, aby časový rad bol geometricky ergodický a absolútne regulárny (čiže tzv. β -mixing). Základný predpoklad, ktorý umožňuje dokázať tieto vlastnosti, je ten, že proces (5.3) je *markovský reťazec*. Stacionarita časového radu potom plynie z faktu, že ergodický Markovský reťazec je stacionárny (viď Fan a Yao (2003, veta 2.2)). Geometrickú ergodicitu a absolútnu regularitu dosiahneme určitými predpokladmi o m , σ a p_ε . Uvedieme si tieto predpoklady.

Predpoklady B1

1. $|m(x)| + \sigma(x)E|\varepsilon_1| \leq C_1 + C_2|x|$ pre všetky $x \in \mathbb{R}$ a nejaké $C_1 < \infty$, $C_2 < 1$;
2. $\sigma(x) \geq C_3 > 0$ pre všetky $x \in \mathbb{R}$ a nejaké $C_3 > 0$;
3. $p_\varepsilon(x) \geq C_4 > 0$ pre všetky

$$x \in \left[-C_5 - \sup_{|x| \leq C_6} \{(m(x) - x)/\sigma(x)\}, C_5 - \inf_{|x| \leq C_6} \{(m(x) - x)/\sigma(x)\} \right]$$

a nejaké $C_6 > C_1/(1 - C_2)$ a $C_4, C_5 > 0$.

Poznamenajme, že konštanty C_1, \dots, C_6 nezávisia na n . Pre špeciálny prípad, keď $m(x) = \varphi x$ a $\sigma(x) = \sigma$, sa model (5.3) zredukuje na AR(1) model, a predpoklad B1.1 sa takisto zredukuje na klasickú podmienku stacionarity AR procesu, $|\varphi| < 1$. Predpoklad B2.2 požaduje kladnú hodnotu $\sigma(x)$ pre všetky x a predpoklad B1.3 hovorí, že hustota inovácií p_ε je spojitá a kladná na \mathbb{R} . Nižšie uvedieme niekoľko viet, ktoré zaručujú geometrickú ergodicitu a absolútnu regularitu procesu (5.3) spolu s tvrdeniami o presnosti aproximácie stacionárneho rozdelenia pôvodného procesu (5.3) bootstrapovým procesom. Tvrdenia boli čerpané z článkov Franke et al. (2002a), Franke et al. (2002b) a Franke et.al (2004). Tvrdenia spolu s dôkazmi sú značne technické, preto uvedieme iba ich znenie a v záujme prehľadnosti práce sa v prípade dôkazu odkážeme na citovaný zdroj. Zavedme si najprv definíciu absolútnej regularity. Táto definícia bola prebraná z Fan, Yao (2003). Alternatívnu definíciu môžeme nájsť v článku Franke et al. (2002b, str. 560).

Definícia 5.3. (β -mixing) Striktne stacionárny proces $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ sa nazýva absolútne regulárny alebo β -mixing, ak pre tzv. mixingové koeficienty,

$$\beta(n) = E \left\{ \sup_{B \in \mathcal{F}_n^\infty} |P(B) - P(B | X_0, X_{-1}, X_{-2}, \dots)| \right\},$$

platí $\beta(n) \rightarrow 0$, pre $n \rightarrow \infty$. \mathcal{F}_i^j značí σ -algebru generovanú $\{X_t, i \leq t \leq j\}$.

Zjednodušene povedané, ide o vlastnosť časového radu, ktorá zaručuje, že daný časový rad môže byť vnímaný ako postupnosť náhodných veličín, pre ktorú sú minulé a budúce hodnoty asymptoticky nezávislé.

Veta 5.7. (Vlastnosti NARCH procesu) Predpokladajme, že proces (5.3) splňuje predpoklady B1, potom

1. $\{X_t\}$ je geometricky ergodický, čiže platí

$$\int \|P^n(x, \cdot) - \pi\| \mu(dx) \leq C_\mu \rho_\mu^{-n},$$

pre nejaké $\rho_\mu > 1$, kde $P^n(x, \cdot)$ značí podmienené rozdelenie X_n na $X_0 = x$, $\|\cdot\|$ značí totálnu variáciu a μ zodpovedá stacionárnemu rozdeleniu π , ak počiatočné rozdelenie $\{X_t\}$ je π alebo zodpovedá Diracovej miere δ_{x_0} , ak proces $\{X_t\}$ začína v nejakom nenáhodnom bode x_0 ;

2. $\{X_t\}$ je absolútne regulárny s geometricky klesajúcimi mixingovými koeficientmi $\beta(m)$, tzn. $\beta(m) \leq C_\mu \rho_\mu^{-m}$.

Dôkaz: viď Franke et.al (2004, veta 1). □

Vidíme teda, že za predpokladu, že proces (5.3) splňuje predpoklady B1, je tento proces geometricky ergodický a absolútne regulárny. Uvažujme teraz, že bootstrapový proces je konštruovaný na základe (5.4). Samozrejme, budeme chcieť, aby aj tento proces bol geometricky ergodický a absolútne regulárny. Splnenie týchto vlastností pre bootstrapový proces (5.4) plynie z podobných predpokladov ako pre pôvodný proces, s tým rozdielom, že kladieme nejaké požiadavky na \tilde{m} , $\tilde{\sigma}$ a \tilde{p}_ε namiesto m , σ a p_ε .

Predpoklady B2: Nech existuje vhodná postupnosť množín $\Omega_n \subseteq \mathbb{R}^{n+1}$, pre ktorú $P((X_0, \dots, X_n) \notin \Omega_n) = o(1)$ pre $n \rightarrow \infty$, taká, že pre $(X_0, \dots, X_n) \in \Omega_n$ sú splnené nasledujúce vlastnosti:

1. $|\tilde{m}(x)| + \tilde{\sigma}(x)E^*|\varepsilon_1^*| \leq C_1 + C_2|x|$ pre všetky $x \in \mathbb{R}$ a nejaké $C_1 < \infty$, $C_2 < 1$;
2. $\sup_{x \in \mathbb{R}} |\tilde{p}_\varepsilon(x) - p_\varepsilon(x)| = O(n^{-v})$ pre nejaké $v > 0$.

E^* ako vždy značí podmienenú strednú hodnotu (bootstrapovej náhodnej veličiny pri daných X_0, \dots, X_n).

Nasledujúca veta zaručuje, že bootstrapový proces (5.4) je geometricky ergodický a absolútne regulárny.

Veta 5.8. (Vlastnosti bootstrapového procesu) Predpokladajme, že proces (5.3) splňuje predpoklady B1, potom

1. $\{X_t^*\}$ je geometricky ergodický;
2. $\{X_t^*\}$ je absolútne regulárny s geometricky klesajúcimi mixingovými koeficientmi $\beta^*(m)$, tzn. $\beta^*(m) \leq C_\mu \rho_\mu^{-m}$, kde C_μ a ρ_μ nezávisia na hodnotách X_0, \dots, X_n .

Dôkaz: Dôkaz pre podmienené homoskedastické dáta ($\sigma(X_{t-1}) = \sigma = \text{konšt.}$) nájdeme v článku Franke et.al (2002b, veta 1 a veta 2). V článku Franke et.al (2004, veta 2) sú tieto vlastnosti dokázané aj pre podmienené heteroskedastický prípad. \square

Avšak tým, že dosiahneme, že proces $\{X_t^*\}$ je geometricky ergodický a absolútne regulárny, ešte nezaručíme, že bude kopírovať vlastnosti pôvodného procesu. Základný predpoklad, ktorý musí bootstrapový proces splňovať, je, že stacionárne rozdelenie $\{X_t^*\}$ je „blízke“ stacionárnemu rozdeleniu $\{X_t\}$. Toto zaručíme tým, že budeme pracovať s konzistentnými odhadmi \tilde{m} , $\tilde{\sigma}$ a \tilde{p}_ε funkcií m , σ a p_ε . V sekcii 5.6 definujeme odhady \tilde{m} , $\tilde{\sigma}$ a \tilde{p}_ε , ktoré splňujú podmienku konzistencie a tým zaručujú validitu bootstrapového algoritmu.

Pred tým, ako vyslovíme vetu o vzťahu medzi stacionárnym rozdelením π a π^* procesu $\{X_t\}$ resp. $\{X_t^*\}$, stanovme najprv predpoklady o prvotných odhadoch \tilde{m} , $\tilde{\sigma}$ a \tilde{p}_ε .

Predpoklady B3: Nech existuje postupnosť množín $\{M_n\}$ taká, že $\dots \subseteq M_n \subseteq M_{n+1} \subseteq \dots$ a $P(X_t \notin M_n) = O(n^{-\nu})$. Nech ďalej rovnako ako v B2 existuje vhodná postupnosť množín $\Omega_n \subseteq \mathbb{R}^{n+1}$, pre ktorú $P((X_0, \dots, X_n) \notin \Omega_n) = o(1)$ pre $n \rightarrow \infty$, taká, že pre $(X_0, \dots, X_n) \in \Omega_n$ sú splnené nasledujúce vlastnosti:

1. $\sup_{x \in M_n} |\tilde{m}(x) - m(x)| = O(n^{-\nu})$;
2. $\sup_{x \in M_n} |\tilde{\sigma}(x) - \sigma(x)| = O(n^{-\nu})$;
3. $\inf_{x \in \mathbb{R}} \tilde{\sigma}(x) \geq C > 0$;
4. $\int |\tilde{p}_\varepsilon(z) - p_\varepsilon(z)| dz = O(n^{-\nu})$;
5. $\int |p_\varepsilon(z) - p_\varepsilon(z+c)| dz = O(c)$ a $\int |p_\varepsilon(z) - p_\varepsilon(z/(1+c))| dz = O(c)$ pre $c \rightarrow 0$.

Veta 5.9. (Validita bootstrapového algoritmu) Predpokladajme, že model, z ktorého dáta X_0, \dots, X_n pochádzajú, sa riadi (5.3), bootstrapový proces sa riadi (5.4) a sú splnené predpoklady B1, B2 a B3. Potom pre $\nu > 0$ z predpokladov B3 a ľubovoľne veľké $a < \infty$, platí

$$\sup_B \left\{ (\lambda(B)n^{-\nu} + n^{-a})^{-1} |\pi(B) - \pi^*(B)| \right\} \leq C,$$

ak $(X_0, \dots, X_n) \in \Omega_n$, kde množiny B sú lebesgueovsky merateľné a $\lambda(\cdot)$ značí Lebesgueovu mieru.

Dôkaz: Dôkaz pre podmienené homoskedastický prípad nájdeme v článku Franke et.al (2002b, veta 3). V článku Franke et.al (2004, veta 3) je táto veta dokázaná pre podmienené heteroskedastické dáta. \square

Veta 5.9 nám hovorí, že stacionárne rozdelenie pôvodného procesu, môže byť aproximované stacionárnym rozdelením bootstrapového procesu s presnosťou, ktorá sa rovná nejakej mocnine n^{-1} , a to rovnomerne nad množinami s konečnou Lebesgueovou mierou. Tento fakt môže byť veľmi užitočný pri odvodzovaní validity bootstrapového algoritmu, aplikovaného na veľmi širokú skupinu štatistík časového radu (5.3). Poznamenajme, že bootstrapové odhady konštruované pomocou tejto metódy majú tvar

$$\hat{m}_h^*(x) = \frac{\sum_{t=1}^{n-1} K_h(x - X_t^*) X_{t+1}^*}{\sum_{t=1}^{n-1} K_h(x - X_t^*)},$$

$$\hat{\sigma}_{h'}^{*2}(x) = \frac{\sum_{t=1}^n K_{h'}(x - X_t^*) (X_{t+1}^* - \hat{m}_{h'}^*(X_t^*))^2}{\sum_{t=1}^n K_{h'}(x - X_t^*)}.$$

V článku Franke et.al (2002a, veta 1) je za platnosti všeobecných a zložitých predpokladov dokázané, že pre ľubovoľné $x_0 \in \mathbb{R}$ platí

$$\sup_{x \in \mathbb{R}} \left| P^* \left\{ \sqrt{nh}(\hat{m}_h^*(x_0) - \tilde{m}(x_0)) \right\} \leq x \right| - P \left\{ \sqrt{nh}(\hat{m}_h(x_0) - m(x_0)) \right\} \leq x \right| \xrightarrow{P} 0,$$

$$\sup_{x \in \mathbb{R}} \left| P^* \left\{ \sqrt{nh}(\hat{\sigma}_h^{*2}(x_0) - \tilde{\sigma}^2(x_0)) \right\} \leq x \right| - P \left\{ \sqrt{nh}(\hat{\sigma}_h^2(x_0) - \sigma^2(x_0)) \right\} \leq x \right| \xrightarrow{P} 0,$$

kde $P^*(\cdot \leq x)$ značí podmienené rozdelenie za daných X_1, \dots, X_n .

5.5 Wild bootstrap

Nadalej pracujeme s modelom (5.3). Pripomeňme, že neparametrický wild bootstrap konštruuje prevzorkovania na záklde vzťahu

$$X_t^* = \tilde{m}(X_{t-1}) + \varepsilon_t^*, \quad t = 2, \dots, n,$$

kde $\varepsilon_t^* = \hat{\varepsilon}_t \xi_t$, $\hat{\varepsilon}_t = X_t - \hat{m}_g(X_{t-1})$ a ξ_t je *iid* postupnosť náhodných veličín s nulovou strednou hodnotou a jednotkovým rozptylom. Vidíme, že táto metóda sa dá použiť iba na konštrukciu odhadu $\hat{m}_h^*(x)$ veľmi dobre poslúži v prípade, že funkcia $\sigma(x)$ nejakým spôsobom nespĺňa predpoklady pre úspešné použitie autoregresného bootstrapu. Uvedme, že odhad funkcie $m(x)$ konštruovaný na základe bootstrapových prevzorkovaní má tvar

$$\hat{m}_h^*(x) = \frac{\sum_{t=1}^{n-1} K_h(x - X_t) X_{t+1}^*}{\sum_{t=1}^{n-1} K_h(x - X_t)}.$$

V článku Franke et.al (2002a, veta 1) je podobne ako autoregresný bootstrap dokázané, že pre ľubovoľné $x_0 \in \mathbb{R}$ platí

$$\sup_{x \in \mathbb{R}} \left| P^* \left\{ \sqrt{nh}(\hat{m}_h^*(x_0) - \tilde{m}(x_0)) \right\} \leq x \right| - P \left\{ \sqrt{nh}(\hat{m}_h(x_0) - m(x_0)) \right\} \leq x \right| \xrightarrow{P} 0,$$

kde $P^*(\cdot \leq x)$ značí podmienené rozdelenie za daných X_1, \dots, X_n a \hat{m}_h^* je konštruovaná na základe algoritmu neparametrického wild bootstrapu. Pre znenie všetkých potrebných predpokladov potrebných k dokázaniu tejto konvergencie znova odkazujeme na citovaný zdroj.

5.6 Výber prvotných odhadov \tilde{m} , $\tilde{\sigma}$ a \tilde{P}_ε

Ako sme spomenuli vyššie, konzistentnosť bootstrapového algoritmu kriticky závisí na konzistencii odhadov \tilde{m} , $\tilde{\sigma}$ a \tilde{P}_ε . Pri konštrukcii týchto odhadov a ich použití na konštrukciu bootstrapového modelu, narážame na problém nízkej hustoty stacionárneho rozdelenia π . V týchto oblastiach (v našom prípade intervaloch) nízkej hustoty je počet hodnôt pôvodného časového radu veľmi nízky. Našťastie môžeme očakávať, že počet hodnôt v týchto intervaloch bude nízky aj v prípade bootstrapového procesu. Preto sa ukazuje, že konzistenciu odhadov \tilde{m} , $\tilde{\sigma}$ a \tilde{P}_ε stačí požadovať, iba na intervaloch, kde hustota stacionárneho rozdelenia je nad nejakou stanovenou hodnotou.

Franke et.al (2002a) navrhli za účelom konštrukcie prvotných odhadov \tilde{m} , $\tilde{\sigma}$, nasledujúcu modifikáciu (useknutie) Nadaraya-Watsonovho odhadu pre m a σ . Pre konštanty K_m , k_σ a K_σ označme $\bar{K}_m = K_m |\bar{X}|$, $\bar{k}_\sigma = k_\sigma \bar{\varepsilon}_{C_n}^2$ a $\bar{K}_\sigma = K_\sigma \bar{\varepsilon}_{C_n}^2$, kde \bar{X} je výberový priemer vzorky X_1, \dots, X_n a $\bar{\varepsilon}_{C_n}^2$ je výberový priemer druhých mocnín reziduí $\hat{\varepsilon}_i^2$. Definujme

$$\tilde{m}(x) = \begin{cases} \hat{m}_g(x) & \text{keď } |\hat{m}_g(x)| \leq \bar{K}_m, \\ \bar{K}_m & \text{keď } \hat{m}_g(x) > \bar{K}_m, \\ -\bar{K}_m & \text{keď } \hat{m}_g(x) < -\bar{K}_m, \end{cases}$$

$$\tilde{\sigma}(x) = \begin{cases} \hat{\sigma}_{g'}(x) & \text{keď } \bar{k}_\sigma \leq \hat{\sigma}_{g'}(x) \leq \bar{K}_\sigma, \\ \bar{K}_\sigma & \text{keď } \hat{\sigma}_{g'}(x) > \bar{K}_\sigma, \\ \bar{k}_\sigma & \text{keď } \hat{\sigma}_{g'}(x) < \bar{k}_\sigma, \end{cases}$$

kde $g = g(n)$ a $g' = g'(n)$ sú vyhladzovacie parametre, pre ktoré $g(n)/h(n) \rightarrow \infty$, $g'(n)/h'(n) \rightarrow \infty$, pri $n \rightarrow \infty$. Táto definícia prvotných odhadov funkcií m , σ hovorí, že tieto odhady nahradíme nejakými konštantami, ak sú ich hodnoty pod resp. nad stanovenými hodnotami, ktoré závisia na X_1, \dots, X_n . Táto procedúra sa uvažuje preto, pretože hodnoty odhadov $\tilde{m}(x)$ a $\tilde{\sigma}(x)$ závisia na absolútnej hodnote parametra x a pokiaľ je hodnota $|x|$ príliš veľká, klasický „neuseknutý“ Nadaraya-Watsonov odhad stáva značne nepresný. Ďalej si definujme \tilde{P}_ε ako empirickú distribučnú funkciu, na základe

ktorej budeme *iid* prevzorkovaním vyberať bootstrapové inovácie ε_t^* . Pre postupnosť $\gamma_n \rightarrow \infty$, položíme $C_n = \{1 \leq t \leq n : |\hat{\varepsilon}_t| \leq \gamma_n\}$, kde

$$\hat{\varepsilon}_t = \frac{X_t - \hat{m}_g(X_{t-1})}{\hat{\sigma}_{g'}(X_{t-1})}, \quad t = 2, \dots, n.$$

Potom \tilde{P}_ε je definovaná ako empirická distribučná funkcia náhodných veličín $\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}_{C_n}$, pre $t \in C_n$, kde $\bar{\hat{\varepsilon}}_{C_n}$ je výberový priemer $\hat{\varepsilon}_t$ pre $t \in C_n$.

V článku Franke et.al (2002a) je dokázané, že pre K_m a K_σ dostatočne veľké, k_σ dostatočne malé a vyššie definované \tilde{m} , $\tilde{\sigma}$ existuje taká postupnosť γ_n pre ktorú platí,

$$\sup_{|x| \leq \gamma_n} |\tilde{m}(x) - m(x)| = o_P(1)$$

a

$$\sup_{|x| \leq \gamma_n} |\tilde{\sigma}(x) - \sigma(x)| = o_P(\gamma_n^{-1}),$$

z čoho plynie rovnomerná konzistentnosť odhadov $\tilde{m}(x)$ a $\tilde{\sigma}(x)$ na intervaloch typu $I = [-\gamma_n, \gamma_n]$. Pre zoznam všetkých technických predpokladov, ktoré musia byť splnené, aby vyššie uvedené rovnosti platili viď Franke et.al (2002a).

Poznámka 5.4. Tento postup je potrebný pokiaľ pracujeme s Nadaraya-Watson odhadmi, prípadne s lokálne lineárnymi odhadmi. Iné tvary odhadov \tilde{m} , $\tilde{\sigma}$, ktoré nepotrebujú podobné useknutie časového radu na interval definovaný vyššie uvažujú napr. v článku Franke et.al (2004) a to tzv. odhady metódou N-najbližších susedov (N-nearest neighbour).

Vidíme, že pri konštrukcii prvotného odhadu používame vyhladzovací parameter g pre odhad podmienenej strednej hodnoty, prípadne g' pre odhad volatility. Tento postup vychádza zo všeobecne známeho problému pri aplikácii bootstrapových metód, a to ako sa vysporiadať z vychýlením pôvodných odhadov. Zoberme si napríklad situáciu, že sa chceme zaoberať analýzou vlastností odhadu $\hat{\sigma}_h(x)$ pomocou bootstrapu. Potom rozložením štandardizovanej štatistiky na stochastickú časť a časť, ktorá reprezentuje vychýlenie pôvodného odhadu,

$$\sqrt{nh}(\hat{\sigma}_h^2(x) - \sigma^2(x)) = \sqrt{nh} \left[\hat{\sigma}_h^2(x) - E(\hat{\sigma}_h^2(x)) \right] + \sqrt{nh} \left[E(\hat{\sigma}_h^2(x)) - \sigma^2(x) \right],$$

môžeme vidieť, že ak chceme aby bootstrap pracoval správne, mal by odhadnúť korektne obidva výrazy na pravej strane. V neparametrickom prípade sa toto docieli tzv. „nadhladením“ (oversmoothing), a to tým spôsobom, že na výpočet reziduí a generovanie bootstrapových replík sa použije parameter g tak, aby splňoval $g > h$ a $h/g \rightarrow 0$ pri $n \rightarrow \infty$. Vyhladzovací parameter h sa následne používa pri zostrojení $\hat{\sigma}_h^*(x)$.

Kapitola 6

Porovnanie metód

V tejto kapitole sa pokúsime potvrdiť teoretické vlastnosti metód, ktoré sme uviedli v predošlých kapitolách. Výpočty rozdelíme do dvoch častí. V prvej časti budeme dáta simulovať pomocou nami zvoleného procesu (DGP) tak, aby nagenované dáta zodpovedali požiadavkom, ktoré na ne v teoretickej časti kladíme. V druhej časti bootstrapové metódy použijeme pri práci s reálnymi dátami z finančného sveta.

6.1 Simulované dáta

Parametrické modely

V tejto časti budeme pracovať s modelom AR(1), ktorý môžeme považovať za reprezentatívny príklad rodiny autoregresných postupností rádu p . Všeobecný DGP je

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + e_t, \quad t = 1, \dots, n,$$

kde povolíme heteroskedasticitu dát spôsobenú tým, že chyby e_t sa budú riadiť GARCH procesom (hovoríme o tzv. AR-GARCH modeli). Ďalej budeme uvažovať rôzne hodnoty parametrov φ_0 a φ_1 , rôzne tvary rozdelenia, ktorými sa riadia chyby ε_t a takisto rôznu veľkosť n pozorovaného (v tomto prípade simulovaného) súboru dát. Presný tvar DGP bude vždy uvedený v tabuľke s príslušnými výsledkami.

Aby sme zhrnuli možné varianty, tak budeme uvažovať DGP s $\varphi_0 \in \{0, 1\}$, $\varphi_1 \in \{0, 0.5, 0.9\}$, a chyby budeme postupne generovať modelmi GARCH, EGARCH a GJR GARCH pre hodnoty parametrov α a β tak, aby $\alpha + \beta < 1$. Všetky modely GARCH budú maximálne rádu GARCH(1, 1), kde základný model GARCH(1, 1) budeme uvažovať v tvare

$$e_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2$$

a náhodné veličiny ε_t sa budú riadiť rozdelením $N(0, 1)$. Parameter α_0 bol volený tak, aby sa nepodmienený rozptyl $\text{var}(e_t) = \sigma^2 = 1$. Zo zápisu modelov vidíme, že pokiaľ $\varphi_0 = \varphi_1 = 0$, model sa zredukuje na GARCH(1, 1) a ak navyše $\beta = 0$, na ARCH(1) model a nakoniec v prípade, že aj $\alpha = 0$ sa model zredukuje na striktný biely šum.

Pri počítaní bootstrap-t intervalov, potrebujeme poznať rozptyl $\sqrt{n}(\hat{\varphi}_1 - \varphi_1)$. Pre wild bootstrap, a párový bootstrap budeme pracovať s odhadom kovariančnej matice, ktorý je robustný voči heteroskedasticite, v tvare

$$\hat{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}(\hat{\varepsilon}_t^2)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \quad (6.1)$$

kde \mathbf{X} značí maticu regresorov v AR modeli. Tento odhad sa označuje ako tzv. HC0 (*heteroskedasticity consistent*). Pre reziduálny bootstrap, ktorý predpokladá, že asymptotická kovariančná matica je v tvare $C = 1 - \varphi_1^2$ budeme používať klasický odhad $\tilde{C} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

Tabuľka 6.1 zobrazuje výsledky simulácií za účelom skúmania presnosti pokrytia bootstrap-t intervalov v tvare

$$(\hat{\varphi}_1 - u_{1-\alpha/2}^* \cdot \sqrt{\hat{C}_{22}}, \hat{\varphi}_1 - u_{\alpha/2}^* \cdot \sqrt{\hat{C}_{22}}),$$

kde u_α^* je α -kvantil rozdelenia Studentizovanej štatistiky $\sqrt{n}(\hat{\varphi}_1^* - \hat{\varphi}_1)/\hat{C}_{22}^*$ vypočítaný príslušnou bootstrapovou metódou. V prípade reziduálneho bootstrapu nahradíme \hat{C} za \tilde{C} . Poznamenajme, že \hat{C}^* resp. \tilde{C}^* je HC0 odhad kovariančnej matice, určený na základe bootstrapových dát, čiže

$$\hat{C}^* = (\mathbf{X}^*\mathbf{X}^*)^{-1}\mathbf{X}^*\text{diag}(\hat{\varepsilon}_t^{*2})\mathbf{X}^*(\mathbf{X}^*\mathbf{X}^*)^{-1},$$

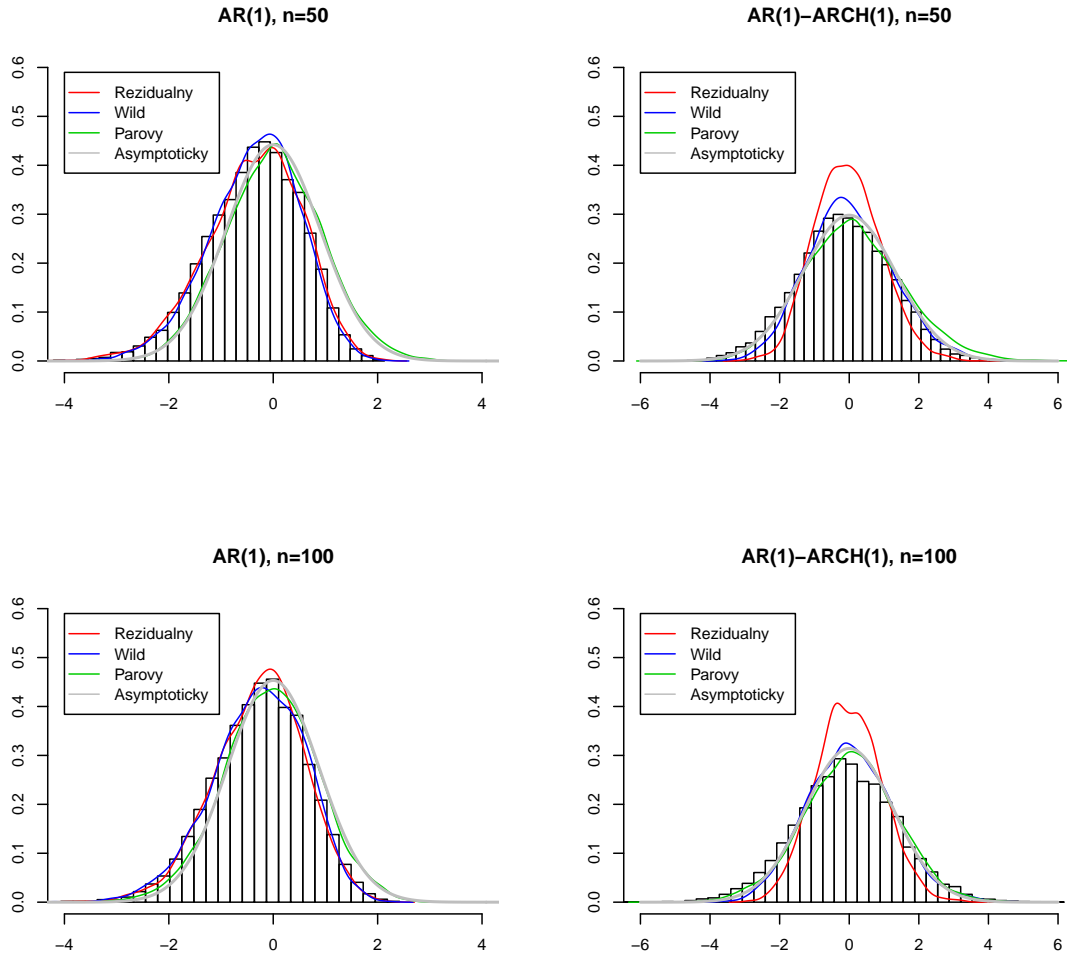
kde $\hat{\varepsilon}_t^* = X_t^* - \hat{\varphi}_0^* - \hat{\varphi}_1^*X_{t-1}^*$ sú odhadnuté rezíduá na základe bootstrapových dát.

Z tabuľky 6.1 jasne vyplýva, že reziduálny bootstrap poskytuje veľmi vysokú presnosť, resp. úspešnosť pokrytia v prípade, že chyby sú *iid*, ale v prípade podmienene heteroskedastických chýb sa presnosť rapídne znižuje. Navyše môžeme pozorovať ďalšie zhoršenie s rastúcim počtom pozorovaní. Ďalšia metóda v poradí - rekurzívny wild bootstrap, vykazuje veľmi vysokú presnosť v prípade *iid* aj ARCH/GARCH chýb. Jeho pokrytie je najbližšie nominálnej hodnote koeficientu spoľahlivosti s pomedzi všetkých uvažovaných metód. Navyše sa ukazuje, že nesplnenie prísnejších predpokladov A1, konkrétne porušenie predpokladu A2.2 o existencii ôsmeho momentu, nemá vplyv na presnosť metódy rekurzívny wild bootstrap pre uvažované dĺžky radov. Modely, ktoré nesplňujú túto podmienku, sú označené hviezdičkou. Fixný wild bootstrap je naopak vo väčšine prípadoch menej presný ako rekurzívna metóda, a dokonca aj ako párový bootstrap. S rastúcim n sa tieto rozdiely znižujú. Intervaly spoľahlivosti založené na aproximácii normálnym rozdelením sú presnejšie v prípade dát s ARCH/GARCH chybami, ale menej presné v prípade *iid* chýb.

DGP: $X_t = \varphi_1 X_{t-1} + e_t$, $e_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = \alpha_0 + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2$, $\varepsilon_t \sim N(0, 1)$								
n	φ_1	α	β	Reziduálny b.	Rek. wild b.	Fix. wild b.	Párový b.	Asymptotický
50	0	0	0	89.7	90.0	89.1	89.8	86.7
*	0	0.5	0	77.0	88.8	86.9	89.1	85.2
*	0	0.3	0.6	81.7	89.5	88.9	89.4	85.4
	0	0.3	0	81.9	89.6	88.9	89.5	86.2
	0	0.1	0.8	87.1	89.3	87.4	89.0	85.6
	0.9	0	0	83.8	83.4	79.5	80.1	76.7
*	0.9	0.5	0	80.0	85.0	80.1	81.3	77.1
*	0.9	0.3	0.6	80.7	84.5	80.8	81.9	77.2
	0.9	0.3	0	82.5	85.2	81.7	82.1	77.5
	0.9	0.1	0.8	83.6	83.4	79.8	80.1	75.9
100	0	0	0	88.9	88.9	89.3	89.7	88.3
*	0	0.5	0	72.9	88.9	87.9	89.3	86.9
*	0	0.3	0.6	77.5	89.7	88.5	89.5	87.0
	0	0.3	0	79.8	89.6	89.1	89.6	87.0
	0	0.1	0.8	84.6	90.2	89.3	89.8	88.3
	0.9	0	0	88.0	87.7	84.3	84.2	82.9
*	0.9	0.5	0	82.1	87.3	84.7	85.1	82.9
*	0.9	0.3	0.6	81.8	88.5	85.3	85.3	82.1
	0.9	0.3	0	82.6	87.5	85.4	85.8	83.2
	0.9	0.1	0.8	86.7	87.6	86.9	84.4	82.2

Tabuľka 6.1. Percentuálne pokrytie 90% bootstrap-t intervalov spoľahlivosti pre parameter φ_1 za predpokladu nulovosti interceptu φ_0 . Výsledky získané na základe 5000 simulácií, pre každú simuláciu nagenované 1000 bootstrapových replikácií. Hviezdičkou sú označené modely, v ktorých 8. moment GARCH procesu nie je konečný a nesplňuje podmienky A2.

Na obrázku 6.1 vidíme histogram skutočného rozdelenia štatistiky $\sqrt{n}(\hat{\varphi}_1 - \varphi_1)$ spolu s aproximáciou jej rozdelenia pomocou 3 bootstrapových metód. Šedou farbou je vyznačená hustota rozdelenia $N(0, C)$, kde C je HC0 odhad v tvare (6.1). Dĺžka radu bola $n = 50$ a $n = 100$. Na ľavom obrázku boli dáta generované z AR(1) modelu v tvare $X_t = 0.5X_{t-1} + \varepsilon_t$, kde $\varepsilon_t \sim N(0, 1)$. Na obrázkoch vpravo dáta pochádzajú z AR(1)–ARCH(1) modelu v tvare $X_t = 0.01X_{t-1} + e_t$, $e_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = 0.5 + 0.5e_{t-1}^2$, $\varepsilon_t \sim N(0, 1)$. Pre metódu wild bootstrap budeme uvažovať len rekurzívny variant.



Obrázok 6.1. Vľavo: Histogram rozdelenia štatistiky $R_n = \sqrt{n}(\hat{\varphi}_1 - \varphi_1)$, kde dáta pochádzajú z AR(1) modelu a aproximácia tohto rozdelenia bootstrapovými metódami. Vpravo: Histogram rozdelenia štatistiky R_n spolu s bootstrapovými aproximáciami tohto rozdelenia, kde dáta pochádzajú z modelu AR(1)-ARCH(1). Histogramy určené na základe 10000 MC simulácií. Bootstrapové aproximácie určené na základe 5000 bootstrapových výberov.

V práci Goncalves, Kilian (2004) bola vyšetrovaná úspešnosť pokrytia bootstrap-t intervalov pre AR-GARCH model spolu s viacerými modifikáciami základného GARCH procesu. Poďme sa pozrieť, či podobné výsledky získame aj pre hodnoty charakteristík bodových odhadov. Tabuľka 6.2 obsahuje hodnoty vychýlenia, smerodajnej odchýlky a MSE pre odhad rozptylu parametra φ_1 pomocou dvoch bootstrapových metód. Pre porovnanie uvádzame aj presnosť odhadu HC0. Reziduálny bootstrap vynecháme, keďže sme v tabuľke 6.1 ukázali, že pre podmienene heteroskedastické chyby nie je táto metóda vhodná.

Pohľad na tabuľku 6.2 prezradzuje, že podobne ako pri intervalových odhadoch, aj pri bodových odhadoch je najlepšou voľbou metóda wild bootstrap. Vidíme, že pre túto metódu je pri $\varphi_1 = 0$ vychýlenie vyššie ako pre ostatné metódy, ale smerodajná odchýlka (rozptyl) odhadu je najmenšia pre všetky uvažované modely, a to v takej miere, že podľa

MSE sa metóda wild bootstrap javí ako najlepšia. Všimnime si, že vychýlenie odhadu rozptylu pomocou párového bootstrapu je malé, ale smerodajná odchýlka je príliš veľká, čo robí z párového bootstrapu veľmi neeficientnú metódu. Odhad pomocou HC0 je rovnako kvalitný ako odhad pomocou párového bootstrapu vo väčšine simulácií. Takisto ako pri skúmaní úspešnosti pokrytia intervalov spoľahlivosti, aj tu platí, že nesplnenie teoretických predpokladov kladených na konzistenciu metódy wild bootstrap (označené hviezdíčkou), má malý vplyv na jej presnosť pri takýchto malých počtoch pozorovaní.

DGP: $X_t = \varphi_1 X_{t-1} + e_t$, $e_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = \alpha_0 + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2$, $\varepsilon_t \sim N(0, 1)$													
					Vychýlenie			Smer. odchýlka			MSE		
n	φ_1	α	β	$n\text{var}\hat{\varphi}_1$	HC0	Wild	Pair	HC0	Wild	Pair	HC0	Wild	Pair
50	0	0	0	0.99	-0.04	-0.12	0.02	0.25	0.19	0.26	0.07	0.05	0.07
*	0	0.5	0	1.73	-0.10	-0.48	0.01	0.74	0.40	0.87	0.56	0.39	0.76
*	0	0.3	0.6	1.47	-0.12	-0.34	-0.03	0.56	0.32	0.59	0.33	0.22	0.35
	0	0.3	0	1.45	-0.12	-0.37	-0.06	0.47	0.28	0.49	0.23	0.22	0.24
	0	0.1	0.8	1.09	-0.04	-0.19	0.03	0.30	0.21	0.32	0.09	0.08	0.10
	0.9	0	0	0.43	-0.13	-0.13	-0.10	0.17	0.16	0.18	0.05	0.04	0.04
*	0.9	0.5	0	0.58	-0.20	0.02	-0.16	0.30	0.26	0.32	0.13	0.07	0.13
*	0.9	0.3	0.6	0.56	-0.18	0.01	-0.14	0.25	0.24	0.32	0.10	0.06	0.12
	0.9	0.3	0	0.51	-0.16	0.03	-0.13	0.23	0.18	0.24	0.08	0.03	0.07
	0.9	0.1	0.8	0.46	-0.15	0.02	-0.11	0.19	0.15	0.20	0.06	0.02	0.05
100	0	0	0	1.00	-0.05	-0.07	0.00	0.19	0.16	0.19	0.04	0.03	0.04
*	0	0.5	0	2.12	-0.20	-0.59	-0.06	0.98	0.58	1.17	1.01	0.69	1.37
*	0	0.3	0.6	1.76	-0.10	-0.33	-0.07	0.72	0.50	0.73	0.53	0.36	0.54
	0	0.3	0	1.57	-0.12	-0.27	-0.01	0.49	0.35	0.52	0.26	0.20	0.27
	0	0.1	0.8	1.20	-0.07	-0.14	-0.03	0.27	0.21	0.29	0.08	0.06	0.09
	0.9	0	0	0.32	-0.07	0.03	-0.06	0.10	0.08	0.11	0.02	0.01	0.02
*	0.9	0.5	0	0.46	-0.10	0.04	-0.10	0.25	0.23	0.28	0.07	0.05	0.09
*	0.9	0.3	0.6	0.44	-0.08	0.05	-0.07	0.24	0.23	0.28	0.07	0.06	0.08
	0.9	0.3	0	0.39	-0.10	0.05	-0.09	0.16	0.16	0.16	0.03	0.03	0.03
	0.9	0.1	0.8	0.35	-0.07	0.05	-0.06	0.12	0.11	0.14	0.02	0.02	0.02

Tabuľka 6.2. Kvalita aproximácie rozptylu parametra φ_1 v AR(1)-GARCH(1,1). Hviezdíčkou sú označené modely, v ktorých 8. moment GARCH procesu nie je konečný, a tým pádom nesplňuje podmienky A2. Správna hodnota štatistiky $n\text{var}\hat{\varphi}_1$ bola získaná metódou MC na základe 5000 simulácií, výsledky boli získané prostredníctvom 2000 simulácií, pre každú simuláciu nagenerovaných 500 bootstrapových replikácií.

V predchádzajúcich kapitolách sme uviedli, že metóda wild bootstrap by teoreticky nemala vykazovať dobré výsledky pre asymetrické modely. Tabuľka 6.3 ale ukazuje, že aj napriek nepriaznivému teoretickému základu presnosť metódy wild bootstrap sa pre asymetrické dáta zásadne nezhoršila.

DGP: $X_t = \varphi_1 X_{t-1} + e_t$, $e_t = \sigma_t \varepsilon_t$, $\ln \sigma_t^2 = \alpha_0 + \beta \ln \sigma_{t-1}^2 + \theta \varepsilon_{t-1} + \gamma (\varepsilon_{t-1} - \sqrt{2/\pi})$, $\varepsilon_t \sim N(0, 1)$															
							Vychýlenie			Smer. odchýlka			MSE		
n	φ_1	α_0	β	θ	γ	$n\text{var}\hat{\varphi}_1$	HC0	Wild	Pair	HC0	Wild	Pair	HC0	Wild	Pair
50	0	0.01	0.70	-0.30	0.60	1.79	-0.12	-0.52	0.08	0.83	0.42	1.10	0.71	0.45	1.22
	0.9	0.01	0.70	-0.30	0.60	0.67	-0.17	0.00	-0.10	0.47	0.29	0.50	0.25	0.08	0.26
	0	0.01	0.10	0.30	0.90	1.76	-0.20	-0.50	0.09	0.70	0.35	1.25	0.52	0.38	1.58
	0.9	0.01	0.10	0.30	0.90	0.57	-0.17	-0.01	-0.12	0.34	0.22	0.37	0.14	0.05	0.15
**	0.09	-0.71	0.86	-0.08	0.26	1.13	-0.07	-0.18	0.01	0.35	0.23	0.36	0.13	0.09	0.13
100	0	0.01	0.70	-0.30	0.60	2.20	-0.12	-0.53	0.01	1.19	0.66	1.21	1.45	0.71	1.47
	0.9	0.01	0.70	-0.30	0.60	0.54	-0.09	0.04	-0.07	0.44	0.28	0.42	0.20	0.08	0.19
	0	0.01	0.10	0.30	0.90	1.99	-0.15	-0.41	0.06	0.73	0.40	0.95	0.55	0.34	0.90
	0.9	0.01	0.10	0.30	0.90	0.42	-0.10	0.04	-0.08	0.21	0.19	0.27	0.05	0.04	0.08
**	0.09	-0.71	0.86	-0.08	0.26	1.22	-0.03	-0.13	0.03	0.35	0.25	0.36	0.12	0.08	0.13
864**	0.09	-0.71	0.86	-0.08	0.26	1.36	0.00	-0.03	0.02	0.20	0.20	0.22	0.04	0.04	0.05

Tabuľka 6.3. Kvalita aproximácie rozptylu parametra φ_1 v AR(1)-EGARCH(1,1). Správna hodnota štatistiky $n\text{var}\hat{\varphi}_1$ bola získaná metódou MC na základe 5000 simulácií, výsledky boli získané prostredníctvom 2000 simulácií, pre každú simuláciu nagenovaných 500 bootstrapových replikácií. (**) Tento model bol použitý v knihe Tsay (2005) na modelovanie vývoja mesačných logaritmickejch výnosov akcií spoločnosti IBM od januára 1926 do decembra 1997 pre 864 hodnôt.

Pre GJR GARCH model dostávame takisto dobré výsledky pre metódu wild bootstrap, ako ukazuje tabuľka 6.4.

DGP: $X_t = \varphi_1 X_{t-1} + e_t$, $e_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma_1 e_{t-1}^2 I_{t-1}^-$, $\varepsilon_t \sim N(0, 1)$															
							Vychýlenie			Smer. odchýlka			MSE		
n	φ_1	α_0	α_1	β_1	γ_1	$n\text{var}\hat{\varphi}_1$	HC0	Wild	Pair	HC0	Wild	Pair	HC0	Wild	Pair
50	0	0.01	0.10	0.80	0.20	1.09	-0.03	-0.14	0.03	0.34	0.22	0.70	0.12	0.07	0.48
	0.9	0.01	0.10	0.80	0.20	0.47	-0.14	0.04	-0.03	0.18	0.16	0.40	0.05	0.03	0.16
	0	0.01	0.10	0.80	0.70	1.36	-0.10	-0.27	0.01	0.47	0.35	0.80	0.23	0.20	0.63
	0.9	0.01	0.10	0.80	0.70	0.52	-0.11	0.02	0.00	0.27	0.18	0.48	0.09	0.03	0.23
100	0	0.01	0.10	0.80	0.20	1.22	-0.03	-0.13	0.00	0.31	0.24	0.56	0.10	0.07	0.31
	0.9	0.01	0.10	0.80	0.20	0.36	-0.08	0.03	-0.03	0.13	0.12	0.22	0.02	0.02	0.05
	0	0.01	0.10	0.80	0.70	1.49	-0.01	-0.23	0.01	0.65	0.54	0.92	0.43	0.34	0.84
	0.9	0.01	0.10	0.80	0.70	0.42	-0.03	0.05	0.01	0.24	0.20	0.38	0.06	0.04	0.14

Tabuľka 6.4. Kvalita aproximácie rozptylu parametra φ_1 v AR(1)-GJR-GARCH(1,1). Správna hodnota štatistiky $n\text{var}\hat{\varphi}_1$ bola získaná metódou MC na základe 5000 simulácií, výsledky boli získané prostredníctvom 2000 simulácií, pre každú simuláciu nagenovaných 500 bootstrapových replikácií.

Vráťme sa na chvíľu k percentilovým intervalovým odhadom. V kapitole 5, príklade 5.1 sme konštruovali percentilové intervaly pre model AR(1)-ARCH(1) v prípade, že sme predpokladali nulovosť interceptu φ_0 . V skutočnosti ale tento predpoklad je veľmi

ťažko realizovateľný. Preto si teraz ukážeme, ako sa percentilové intervaly spoľahlivosti správajú v prípade, ak tento predpoklad nulovosti interceptu pri výpočte OLS odhadu zamietneme.

DGP: $X_t = \varphi_1 X_{t-1} + e_t$, $e_t = \sigma_t \varepsilon_t$, $\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2$, $\varepsilon_t \sim N(0, 1)$							
n	φ_1	α_1	Reziduálny b.	Rek. wild b.	Fix. wild b.	Párový b.	Asymptotický
50	0	0	90.2	90.0	88.1	89.8	86.7
	0	0.5	78.2	84.5	79.0	87.1	85.1
	0	0.9	69.6	81.7	73.7	87.1	82.8
	0.5	0	82.5	80.3	86.1	86.5	85.2
	0.9	0	46.2	44.3	77.2	79.4	76.7
	0.5	0.5	74.2	79.2	78.3	86.9	78.0
100	0	0	89.9	89.9	88.6	89.6	88.3
	0	0.5	74.5	87.1	78.3	88.4	86.1
	0	0.9	61.7	82.0	73.7	86.4	84.2
	0.5	0	85.8	85.0	87.0	88.2	87.5
	0.9	0	61.5	62.0	83.0	82.6	82.9
	0.9	0.5	75.3	88.5	80.0	87.2	86.2
500	0	0	89.7	89.1	88.5	88.9	88.5
	0	0.5	68.8	87.0	80.0	88.6	88.3
	0	0.9	43.8	82.8	75.2	86.6	86.0
	0.5	0	89.3	88.9	88.5	89.0	89.5
	0.9	0	81.7	81.7	87.5	88.0	87.7
	0.9	0.5	69.6	86.5	80.7	88.5	89.1

Tabuľka 6.5. Percentuálne pokrytie 90% percentilových intervalov spoľahlivosti pre parameter φ_1 za predpokladu nenulovosti interceptu φ_0 . Výsledky získané na základe 5000 simulácií, pre každú simuláciu nagenovaných 1000 bootstrapových replikácií.

Vidíme, že percentilové intervaly spoľahlivosti sú v prípade silnej závislosti veľmi nepresné. Tento fakt je spôsobený tým, že klasický OLS odhad, ktorý predpokladá nenulový intercept, nie je nestranný, pokiaľ regresory tvoria oneskorené hodnoty vysvetľovanej premennej. Na rozdiel od klasickej lineárnej regresie, kde OLS odhad parametra je nestranný, v autoregresii musíme byť opatrní pri používaní bootstrapových metód. Keďže základná podstata metódy bootstrap je, že $\hat{\varphi}^*$ zodpovedá $\hat{\varphi}$ tak, ako $\hat{\varphi}$ zodpovedá φ , tak pri neplatnosti druhého predpokladu, generovaním bootstrapových replík (ktoré sa „naivne“ spoliehajú na korektnosť $\hat{\varphi}$), sa stáva bootstrapový odhad ešte viac vychýlený. Všimnime si, že zlé výsledky dávajú metódy, ktoré sa naozaj pri konštrukcii $\hat{\varphi}^*$ spoliehajú na $\hat{\varphi}$, ako reziduálny a rekurzívny wild bootstrap. Párový bootstrap, ktorý bootstrapové odhady generuje priamo z dát (a nie na základe odhadnutých reziduí), je presnejší.

Tento nedostatok môžeme opraviť tým, že budeme pracovať s bootstrap-t intervalmi, ktoré pri konštrukcii bootstrapových replík zohľadňujú pôvodný odhad $\hat{\varphi}$ (prostredníctvom centrovania) a tým implicitne odstraňujú vychýlenie. Druhá možnosť je použiť

BC_a intervaly, ktoré nevyžadujú počítanie smerodajnej odchýlky. Tretia možnosť je odhad $\hat{\varphi}$ očistiť od vychýlenia predtým, ako aplikujeme bootstrapovú metódu.

Neparametrické modely

Pozrime sa teraz, aké výsledky dávajú bootstrapové metódy v praxi, ak nemáme k dispozícii parametrický model. Dáta budeme simulovať z nasledujúcich modelov:

$$\text{M1: } X_t = 4 \sin(X_{t-1}) + \varepsilon_t,$$

$$\text{M2: } X_t = 0.5X_{t-1} + \sqrt{0.5 + 0.25X_{t-1}^2}\varepsilon_t,$$

$\varepsilon_t \sim N(0, 1)$ v oboch prípadoch. Zoberme si najprv model M1.

Model M1

Ukážme si prostredníctvom krátkej simulácie, že neparametrické bootstrapové metódy správne aproximujú rozdelenie jadrového odhadu podmienenej strednej hodnoty modelu M1. Vidíme, že $m(x) = 4 \sin(x)$, $\sigma(x) \equiv 1$. Budeme predpokladať, že konštantnosť $\sigma(x)$ je dopredu známa, preto pre dáta zvolíme model $X_t = m(X_{t-1}) + \varepsilon_t$, čiže model s podmienene homoskedastickými chybami. Budeme skúmať rozdelenie štatistiky

$$\sqrt{nh}(\hat{m}_h(\pi/2) - m(\pi/2)) \quad (6.2)$$

a jej bootstrapovej aproximácie

$$\sqrt{nh}(\hat{m}_h^*(\pi/2) - \tilde{m}(\pi/2)), \quad (6.3)$$

kde $\tilde{m}(x) = \hat{m}_g(x)$. Vyhľadovací parameter bol zvolený $h = 0.9$, $g = 1.2$ pre $n = 50$, $h = 0.8$, $g = 1.1$ pre $n = 100$ a $h = 0.7$, $g = 1.0$ pre $n = 200$. Správne rozdelenie štatistiky (6.2) sme určili na základe 5000 Monte Carlo replikácií, rozdelenie bootstrapovej štatistiky na základe 1000 bootstrapových replikácií.

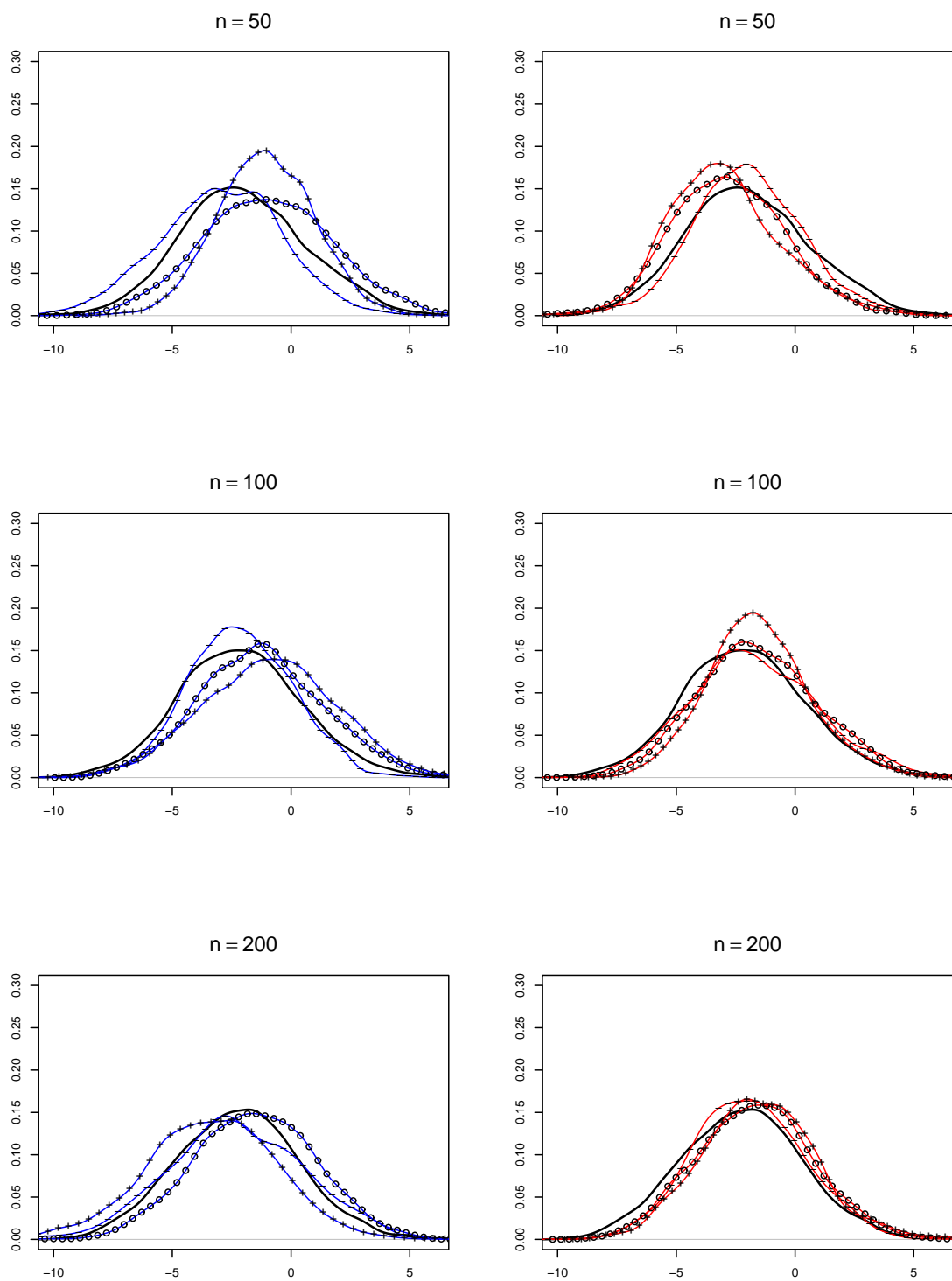
Aby sme číselne vyjadrili presnosť aproximácie rozdelenia štatistiky (6.2) bootstrapovými metódami, počítajme tzv. odmocninovú priemernú štvorcovú chybu (root average squared error) v tvare

$$\text{RASE}_j = \left[\frac{1}{m} \sum_{i=1}^m \{f_j^*(x_i) - f(x_i)\}^2 \right]^{1/2},$$

kde RASE_j značí hodnotu odmocninovej priemernej štvorcovej chyby pre j -tú simuláciu, pričom v každej takejto simulácii vygenerujeme nový podkladový časový rad z modelu M1 a pre tento časový rad vypočítame bootstrapovou metódou 1000 hodnôt štatistiky (6.3). Z týchto hodnôt potom odhadneme rozdelenie f_j^* tejto štatistiky pre j -tu simuláciu klasickým jadrovým odhadom. Hodnoty RASE sme počítali v 100 ekvidistantných bodoch z intervalu $[-10, 5]$. Skutočnú hustotu určenú na základe MC simulácií sme označili $f(x_i)$. Hodnoty RASE nám budú slúžiť aj na určenie reprezentatívnej vzorky

bootstrapových aproximácií rozdelenia štatistiky (6.2), a to tým spôsobom, že zo 100 simulácií vyberieme len tie, ktoré (zhruba) zodpovedajú mediánu, hornému a dolnému kvartilu z hodnôt $RASE_1, \dots, RASE_{100}$. Takto určíme 3 prípady, ktoré zakreslíme do grafov.

Na obrázku 6.2 vidíme správne rozdelenie štatistiky (6.2) spolu s 3 bootstrapovými aproximáciami pre 3 rôzne pôvodné časové rady. Vidíme, že presnosť autoregresného bootstrapu je relatívne veľká už pri $n = 50$, čo je pre neparametrické metódy značne nízky počet pozorovaní. Pre $n = 100$ pozorujeme ďalšie spresnenie metódy. Z grafov vidíme, že wild bootstrap sa zdá byť menej presný ako autoregresný bootstrap pre všetky dĺžky časového radu, čo potvrdzuje aj tabuľka 6.6. Pre $n = 200$ už obidve metódy vykazujú výborné výsledky.

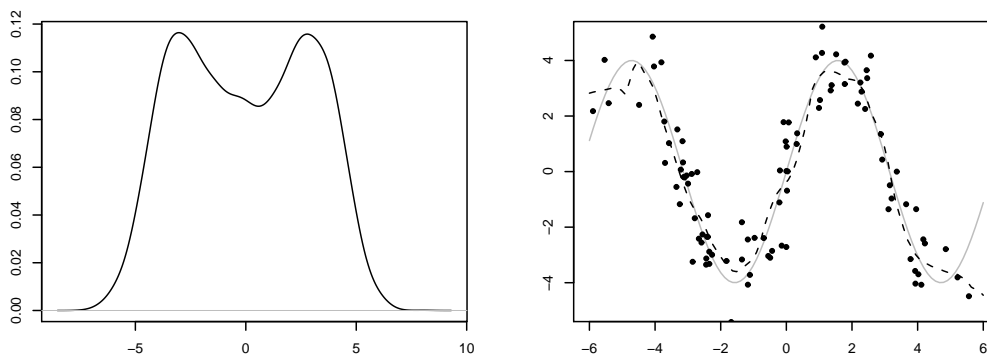


Obrázok 6.2. Hustota štatistiky (6.2) pre $n = 50, 100, 200$ vypočítaná pomocou 5000 MC simulácií (—) spolu s 3 aproximáciami metódou neparametrický wild bootstrap (ľavý panel) a autoregresný bootstrap (pravý panel). Pre každú metódu zakreslené 3 prípady zodpovedajúce 3 hodnotám RASE, dolný kvartil (- - -), medián (o o o) a horný kvartil (+ + +).

n	Autoregresný bootstrap			Wild bootstrap		
	dolný kvartil	medián	horný kvartil	dolný kvartil	medián	horný kvartil
50	0.0123	0.0191	0.0261	0.0170	0.0237	0.0343
100	0.0098	0.0149	0.0212	0.0132	0.0120	0.0299
200	0.0082	0.0115	0.0175	0.0108	0.0185	0.0245

Tabuľka 6.6. Hodnoty dolného kvartilu, mediánu a horného kvartilu RASE odhadnutej hustoty štatistiky (6.2) metódou autoregresný a neparametrický wild bootstrap voči odhadnutej hustote skutočného rozdelenia štatistiky spočítanej na základe 5000 MC simulácií. Medián spočítaný na základe simulácií 100 rôznych časových radov $\{X_t\}$.

Na obrázku 6.3 je pre znázornenie zobrazená stacionárna hustota X_t simulovaná na základe realizácie procesu z modelu M1 s $n = 10000$ a ďalej tvar funkcie $m(x)$ pre hodnoty $x \in [-6, 6]$ spolu s odhadom $\hat{m}_h(x)$ pre $n = 50$, $h = 0.9$. Vidíme, že odhad $\hat{m}_h(x)$ je značne nepresný na intervaloch nízkej stacionárnej hustoty, čiže približne pre $|x| > 5$.



Obrázok 6.3. Vľavo hustota stacionárneho rozdelenia procesu M1. Vpravo: znázornený tvar funkcie $m(x)$ (—) pre $x \in [-5, 5]$ spolu s jej odhadom $\hat{m}_h(x)$ (- - -), $n = 50$, $h = 0.9$ pre jednu realizáciu. Počet bodov z tohto intervalu, v ktorých boli funkcie konštruované, je 100.

Model M2

V modeli M1 sme predviedli, že bootstrapová aproximácia rozdelenia podmienenej strednej hodnoty je relatívne presná pri podmienene homoskedastických chybách. V tomto prípade vidíme, že v modeli M2 sú chyby generované ako ARCH(1) proces. Budeme znova skúmať rozdelenie Nadaraya-Watsonovho odhadu $\hat{m}(x)$ a $\hat{\sigma}(x)$ v bode $x = 1$. Najprv nás bude zaujímať presnosť aproximácie rozdelenia štatistiky

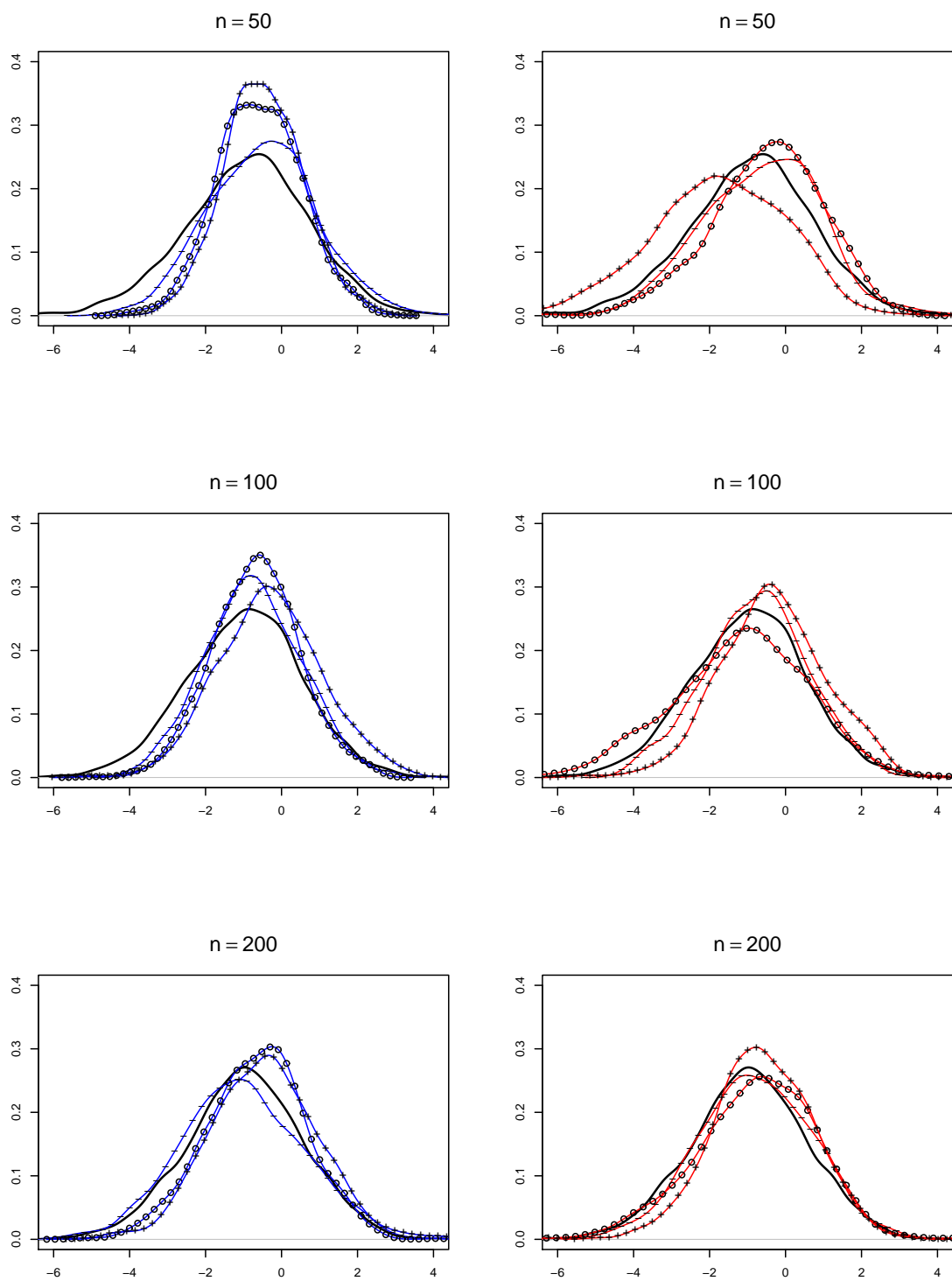
$$\sqrt{nh}(\hat{m}_h(1) - m(1)) \quad (6.4)$$

pomocou neparametrického autoregresného a neparametrického wild bootstrapu. Správne rozdelenie štatistiky (6.4) určíme metódou Monte Carlo pomocou 5000 simulácií pre 3

rôzne dĺžky časového radu, $n = 50$, $n = 100$ a $n = 200$ generovaného pomocou M2. Bootstrapová štatistika má tvar

$$\sqrt{nh}(\hat{m}_h^*(1) - \tilde{m}(1)). \quad (6.5)$$

Na obrázku 6.4 vidíme skutočné rozdelenie štatistiky (6.4) a bootstrapovú aproximáciu pomocou dvoch bootstrapových metód. Znovu zobrazujeme iba 3 prípady zo 100, ktoré zodpovedajú 3 hodnotám ASE_j , $j = 1, \dots, 100$, rovnako ako v modeli M1. Hodnoty ASE sme tentokrát počítali v 100 bodoch z intervalu $[-8, 6]$. Vyhľadovací parameter bol zvolený $h = 1.0$, $g = 1.2$ pre $n = 50$, $h = 0.9$, $g = 1.1$ pre $n = 100$ a $h = 0.8$, $g = 1.0$ pre $n = 200$. Z obrázku je zrejmé, že wild bootstrap nezachycuje správne rozptyl štatistiky. Autoregresný bootstrap vykazuje znova lepšie výsledky, a to pre všetky uvažované dĺžky simulovaných časových radov.



Obrázok 6.4. Hustota štatistiky (6.4) pre $n = 50, 100, 200$ vypočítaná pomocou 5000 MC simulácií (—) spolu s 3 aproximáciami metódou neparametrický wild bootstrap (ľavý panel) a autoregresný bootstrap (pravý panel). Pre každú metódu zakreslené 3 prípady zodpovedajúce 3 hodnotám RASE, dolný kvartil (- - -), medián (o o o) a horný kvartil (+ + +).

n	Autoregresný bootstrap			Wild bootstrap		
	dolný kvartil	medián	horný kvartil	dolný kvartil	medián	horný kvartil
50	0.0207	0.0298	0.0414	0.0260	0.0358	0.0514
100	0.0132	0.0208	0.0334	0.0209	0.0301	0.0391
200	0.0123	0.0195	0.0292	0.0152	0.0224	0.0332

Tabuľka 6.7. Hodnoty dolného kvartilu, mediánu a horného kvartilu RASE odhadnutej hustoty štatistiky (6.4) metódou autoregresný a neparametrický wild bootstrap voči odhadnutej hustote skutočného rozdelenia štatistiky spočítanej na základe 5000 MC simulácií. Medián spočítaný na základe simulácií 100 rôznych časových radov $\{X_t\}$.

Pre štatistickú analýzu volatility môžeme použiť autoregresný bootstrap. Pretože neparametrické odhady v prípade skúmania volatility sú príliš nepresné pri malých počtoch pozorovaní, budeme pracovať s mierne dlhším radom ako v prípade podmienenej strednej hodnoty, konkrétne budeme uvažovať $n = 100, 200, 400, 800$. Budeme teda skúmať kvalitu aproximácie rozdelenia štatistiky

$$\sqrt{nh}(\hat{\sigma}_h(1) - \sigma(1)) \quad (6.6)$$

pomocou neparametrického autoregresného bootstrapu. Správne rozdelenie štatistiky (6.6) určíme metódou Monte Carlo pomocou 5000 simulácií pre 4 rôzne dĺžky časového radu, $n = 100, 200, 400, 800$ generovaného pomocou M2. Bootstrapová štatistika má tvar

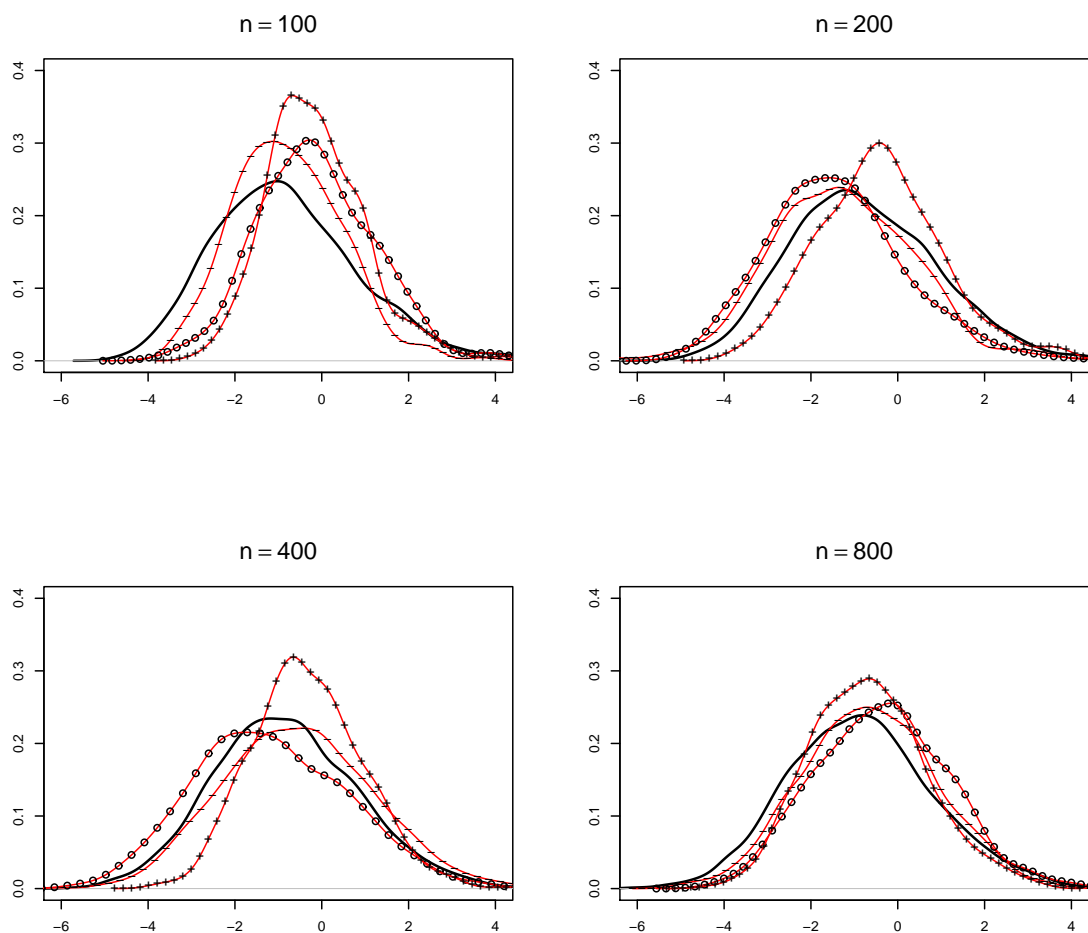
$$\sqrt{nh}(\hat{\sigma}_h^*(1) - \tilde{\sigma}(1)). \quad (6.7)$$

Tentokrát zo simulácií vyplynulo, že najlepšie výsledky sme dostali použitím rôznych vyhladzovacích parametrov pre m a σ . Hodnoty vyhladzovacích parametrov pre jednotlivé počty pozorovaní sú uvedené v tabuľke 6.8.

	h_m	g_m	h	g
100	0.9	1.1	1.0	1.2
200	0.8	1.0	0.9	1.1
400	0.7	0.9	0.8	1.0
800	0.6	0.8	0.7	0.9

Tabuľka 6.8. Hodnoty vyhladzovacích parametrov pre jednotlivé počty pozorovaní použitých pri aproximácii rozdelenia štatistiky (6.6).

Z obrázku 6.5 vidíme, že pri $n = 100$ je aproximácia rozdelenia štatistiky (6.6) značne nepresná. V porovnaní s aproximáciou rozdelenia štatistiky pre podmienenú strednú hodnotu je odhad rozdelenia štatistiky pre volatility menej presný pri rovnakom počte pozorovaní. So zvyšujúcim sa počtom pozorovaní sa stáva aproximácia presnejšia, čo vyplýva aj z tabuľky 6.9. Pre $n = 800$ už je presnosť aproximácie porovnateľná s presnosťou aproximácie pre podmienenú strednú hodnotu pri $n = 200$.



Obrázok 6.5. Hustota štatistiky (6.6) pre $n = 100, 200, 400, 800$ vypočítaná pomocou 5000 MC simulácií (—) spolu s 3 aproximáciami metódou autoregresný bootstrap. Zakreslené 3 prípady zodpovedajúce 3 hodnotám RASE, dolný kvartil (- - -), medián (o o o) a horný kvartil (+ + +).

Autoregresný bootstrap			
n	dolný kvartil	medián	horný kvartil
100	0.0261	0.0488	0.0685
200	0.0225	0.0311	0.0482
400	0.0163	0.0251	0.0403
800	0.0134	0.0221	0.0361

Tabuľka 6.9. Hodnoty dolného kvartilu, mediánu a horného kvartilu RASE odhadnutej hustoty štatistiky (6.6) metódou autoregresný bootstrap voči odhadnutej hustote skutočného rozdelenia štatistiky spočítanej na základe 5000 MC simulácií. Medián spočítaný na základe simulácií 100 rôznych časových radov $\{X_t\}$.

Nakoniec prostredníctvom krátkej simulácie porovnáme presnosť parametrických a neparametrických bootstrapových metód. Budeme študovať presnosť aproximácie rozptylu OLS odhadu $n\text{var}\hat{\varphi}_1$ pre 3 rôzne jednoduché časové rady. Prvý model, ktorý budeme uvažovať, je klasický autoregresný model $X_t = \varphi_1 X_{t-1} + \varepsilon_t$. Tabuľka 6.10 obsahuje hodnoty vychýlenia, smerodajnej odchýlky a MSE pre 4 bootstrapové metódy a $\varphi_1 = 0.5$. Tabuľka 6.11 obsahuje rovnaké charakteristiky pre $\varphi_1 = 0$. Vidíme, že pre dáta bez podmienenej heteroskedasticity sú z hľadiska vychýlenia parametrické metódy presnejšie. Smerodajná odchýlka vychádza najmenšia pre parametrický rekurzívny wild bootstrap (RW) a hodnoty MSE naznačujú, že parametrické metódy sú lepšou voľbou ako neparametrický wild bootstrap (NW) pre malé počty pozorovaní. Autoregresný bootstrap (AR) ale nezaostáva a vykazuje podobné výsledky ako parametrický fixný wild bootstrap (FW). Vyhľadovací parameter pre neparametrické metódy bol zvolený $h = 1.2$ pre $n = 50$, $h = 1.1$ pre $n = 100$ a $h = 1.0$ pre $n = 200$.

		Vychýlenie				Smer. odchýlka				MSE			
	$n\text{var}\hat{\varphi}_1$	RW	FW	NW	AR	RW	FW	NW	AR	RW	FW	NW	AR
50	0.85	-0.07	-0.05	-0.19	0.09	0.15	0.23	0.21	0.19	0.03	0.05	0.08	0.05
100	0.79	-0.03	-0.03	-0.12	0.09	0.13	0.18	0.16	0.15	0.02	0.03	0.04	0.03
200	0.77	-0.03	-0.01	-0.09	0.07	0.10	0.13	0.11	0.11	0.01	0.02	0.02	0.02

Tabuľka 6.10. Kvalita aproximácie rozptylu parametra φ_1 v modeli $X_t = 0.5X_{t-1} + \varepsilon_t$. Správna hodnota štatistiky $n\text{var}\hat{\varphi}_1$ bola získaná metódou MC na základe 10000 simulácií, výsledky boli získané prostredníctvom 2000 simulácií, pre každú simuláciu nagenovaných 500 bootstrapových replikácií.

		Vychýlenie				Smer. odchýlka				MSE			
	$n\text{var}\hat{\varphi}_1$	RW	FW	NW	AR	RW	FW	NW	AR	RW	FW	NW	AR
50	0.98	-0.09	-0.01	-0.21	-0.09	0.18	0.26	0.24	0.24	0.04	0.07	0.10	0.06
100	1.00	-0.08	-0.02	-0.15	-0.06	0.16	0.20	0.19	0.19	0.03	0.04	0.06	0.04
200	1.00	-0.06	-0.02	-0.10	-0.03	0.13	0.15	0.16	0.17	0.02	0.02	0.03	0.03

Tabuľka 6.11. Kvalita aproximácie rozptylu parametra φ_1 v modeli $X_t = \varepsilon_t$. Správna hodnota štatistiky $n\text{var}\hat{\varphi}_1$ bola získaná metódou MC na základe 10000 simulácií, výsledky boli získané prostredníctvom 2000 simulácií, pre každú simuláciu nagenovaných 500 bootstrapových replikácií.

Tabuľka 6.12 popisuje kvalitu aproximácie, pokiaľ do modelu pridáme podmienenú heteroskedasticitu. Okrem evidentného zhoršenia všetkých ukazovateľov vidíme, že vychýlenie parametrického fixného wild bootstrapu je najväčšie zo všetkých uvažovaných metód, naopak najlepšie si vedie rekurzívny wild bootstrap a autoregresný bootstrap. Smerodajná odchýlka odhadov rozptylu je najmenšia v prípade autoregresnej metódy. Nakoniec z hľadiska MSE sú dokonca všetky ostatné (tzn. aj obidve neparametrické

metódy) presnejšie ako parametrická metóda fixný wild bootstrap. Parametrická metóda rekurzívny wild bootstrap je v tomto prípade presnejšia ako neparametrický wild bootstrap, ale menej presná ako autoregresný bootstrap. Vyhľadovací parameter pre neparametrické metódy bol zvolený $h = 1$ pre $n = 50$, $h = 0.9$ pre $n = 100$ a $h = 0.8$ pre $n = 200$.

		Vychýlenie				Smer. odchýlka				MSE			
	$n\text{var}\hat{\varphi}_1$	RW	FW	NW	AR	RW	FW	NW	AR	RW	FW	NW	AR
50	1.35	-0.27	-0.36	-0.30	-0.24	0.27	0.33	0.34	0.20	0.15	0.24	0.21	0.10
100	1.44	-0.22	-0.38	-0.25	-0.22	0.27	0.27	0.29	0.17	0.12	0.22	0.15	0.08
200	1.50	-0.16	-0.39	-0.23	-0.19	0.30	0.25	0.27	0.15	0.12	0.21	0.13	0.06

Tabuľka 6.12. Kvalita aproximácie rozptylu parametra φ_1 v modeli $X_t = \sqrt{0.5 + 0.25X_{t-1}^2}\varepsilon_t$. Správna hodnota štatistiky $n\text{var}\hat{\varphi}_1$ bola získaná metódou MC na základe 10000 simulácií, výsledky boli získané prostredníctvom 2000 simulácií, pre každú simuláciu nagenovaných 500 bootstrapových replikácií.

Kapitola 7

Použitie metódy bootstrap

V tejto časti si ukážeme, ako môžeme metódu bootstrap použiť v praxi. Medzi dôležité otázky štatistického skúmania môže napríklad patriť aj testovanie na parametrické štruktúry, čiže hľadanie odpovede na otázku, či dané dáta môžeme modelovať určitým parametrickým modelom. Vhodnou cestou k nájdeniu tejto odpovede sú neparametrické metódy. Zjednodušene teda hľadáme odpoveď na otázku, ako ďaleko je neparametrický model od parametrického. Hjellvik a Tjøstheim (1995) navrhli testy založené na L_2 -vzdialenosti medzi neparametrickým a parametrickým odhadom pre strednú hodnotu a podmienený rozptyl. Avšak ako sa ukázalo, v tomto prípade je testovanie hypotéz pomocou asymptotických aproximácií veľmi nepresné, pokiaľ nemáme k dispozícii obrovské množstvo dát. Ako už vieme, bootstrapové metódy často poskytujú lepšie výsledky ako asymptotické metódy pri malom počte pozorovaní, avšak nikde nebola teoreticky dokázaná validita bootstrapového algoritmu v tomto kontexte. Kreiss, Neumann a Yao (2008) prišli so svojím článkom, kde túto medzeru zaplňujú. Nimi navrhnutý postup si teraz popíšeme.

Uvažujme jednoduchý model

$$Y_t = m(\mathbf{X}_t) + \varepsilon_t, \quad t \geq 1, \quad (7.1)$$

kde $m(\mathbf{x}) = E(Y_t | \mathbf{X}_t = \mathbf{x})$ a $E(\varepsilon_t | Y_t) = 0$ pre všetky t . Neuvažujeme, že ε_t , $t \geq 1$ sú nezávislé, čo nám dovoľí zahrnúť do tohto modelu podmienenú heteroskedasticitu. Poznamenajme, že model (7.1) obsahuje aj autoregresný model, kde \mathbf{X}_t obsahuje oneskorené hodnoty Y_t . Nulová hypotéza, ktorú chceme testovať, má tvar

$$H_0 : m(\cdot) \in \{m_\theta | \theta \in \Theta\}$$

proti alternatíve

$$H_1 : m(\cdot) \notin \{m_\theta | \theta \in \Theta\}$$

čiže testujeme, či autoregresnú funkciu (podmienenú strednú hodnotu) môžeme vyjadriť v určitej parametrickej forme. Ako príklad si môžeme uviesť testovanie hypotézy na lineárne regresný tvar autoregresnej funkcie, t.j. $m_\theta(x_1, \dots, x_p) = \sum_{i=1}^p \theta_i x_i$, respektíve zaujímavý príklad s aplikáciou v ekonometrii - testovanie na ARCH štruktúru. Na prvý pohľad sa môže zdať, keďže ARCH štruktúra nie je obsiahnutá v podmienenej strednej hodnote ale vo volatilitate, že takto formulovaný test je irelevantný. Ale ako sme spomenuli

v sekcii 2.2.1, ARCH model sa dá ekvivalentne zapísať ako AR model druhých mocnín, čiže napr. nasledujúce tvary ARCH(1) modelu sú zameniteľné:

$$X_t = \sqrt{\alpha_0 + \alpha_1 X_{t-1}^2} \varepsilon_t,$$

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + (\alpha_0 + \alpha_1 X_{t-1}^2)(\varepsilon_t^2 - 1),$$

a teda tento problém je špeciálny prípad testovania hypotézy H_0 pre $Y_t = X_t^2$.

Základná myšlienka spočíva v použití L_2 -vzdialenosti medzi neparametrickým (budeme uvažovať Nadaraya-Watsonov odhad v tvare (2.24)) a parametrickým odhadom autoregresnej funkcie ako testovej štatistiky, čiže

$$T_n = \int_{R^p} \left(\frac{1}{n} \sum_{t=1}^n K_h(\mathbf{x} - \mathbf{X}_t) \{Y_t - m_{\hat{\theta}}(\mathbf{X}_t)\} \right)^2 w(\mathbf{x}) d\mathbf{x}, \quad (7.2)$$

kde $K_h(\cdot) = K_h(\cdot/h)/h$ je jadro definované v kapitole 3, h je vyhladzovací parameter a $w(\cdot)$ je váhová funkcia. Za odhad $\hat{\theta}$ parametra θ funkcie $m(\cdot)$ zoberieme akýkoľvek \sqrt{n} -konzistentný odhad.

K implementácii akejkoľvek hypotézy budeme potrebovať kritické obory testovej štatistiky, ktoré získame pomocou bootstrapovej metódy. V sekcii 5.5 sme uviedli vlastnosti neparametrického wild bootstrapu a túto metódu teraz použijeme na zostrojenie kritických hodnôt. V práci Kreiss, Neumann a Yao (2008, str. 3) je uvedené, že asymptotické rozdelenie štatistiky $nh^{p/2}T_n$ je rovnaké ako asymptotické rozdelenie štatistiky $nh^{p/2}S_n$, kde

$$S_n = \int_{R^p} \left(\frac{1}{n} \sum_{t=1}^n K_h(\mathbf{x} - \mathbf{X}_t) \varepsilon_t \right)^2 w(\mathbf{x}) d\mathbf{x}.$$

Tento fakt nám dovoľuje aproximovať rozdelenie T_n , aplikovaním bootstrapovej metódy „iba“ na štatistiku S_n . Bootstrapovú štatistiku definujeme ako

$$S_n^* = \int_{R^p} \left(\frac{1}{n} \sum_{t=1}^n K_h(\mathbf{x} - \mathbf{X}_t) \varepsilon_t^* \right)^2 w(\mathbf{x}) d\mathbf{x},$$

kde, ako sme uviedli v sekcii 5.5, $\varepsilon_t^* = \hat{\varepsilon}_t \xi_t$, $\hat{\varepsilon}_t = Y_t - \hat{m}_h(X_t)$ a $\xi_t \sim IID(0, 1)$. Hypotézu H_0 zamietame, ak $T_n > u_{1-\alpha}^*$, kde $u_{1-\alpha}^*$ je α -percentný kvantil (podmieneného) rozdelenia S_n^* . Všetky potrebné teoretické tvrdenia zaručujúce asymptotickú validitu tohto bootstrapového procesu nájdeme v článku Kreiss, Neumann a Yao (2008).

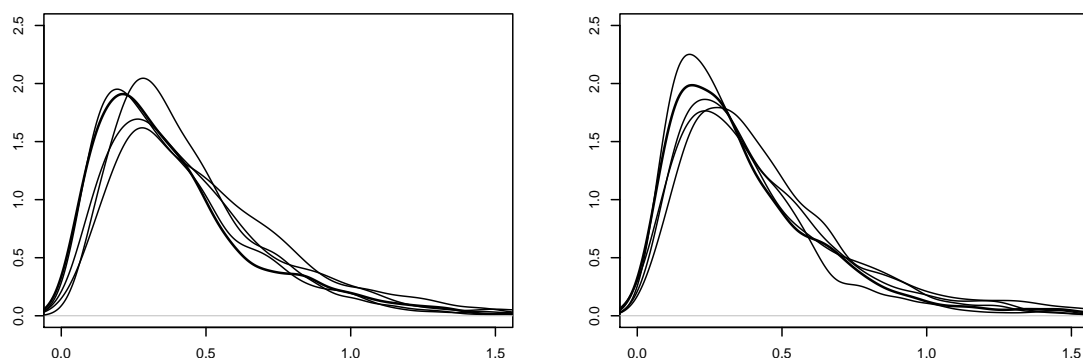
Predtým, ako tento postup použijeme na reálne dáta, poďme sa pozrieť, ako blízko je rozdelenie bootstrapovej štatistiky skutočnému rozdeleniu štatistiky S_n pomocou krátkej simulácie. Budeme uvažovať nasledujúce modely:

$$\text{GOF1: } X_t = 0.9 \cdot X_{t-1} + \varepsilon_t,$$

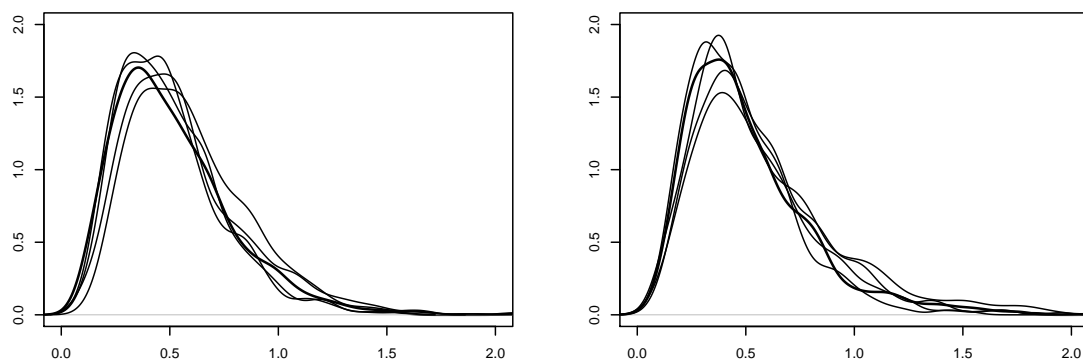
$$\text{GOF2: } X_t = 4 \cdot \sin(X_{t-1}) + \varepsilon_t,$$

kde budeme uvažovať $\varepsilon_t \sim N(0, 1)$ a $\varepsilon_t \sim Exp(1)$. Začnime s modelom GOF1. Máme podozrenie, že dáta by mohli pochádzať z autoregresného modelu $AR(1)$. Preto testujeme hypotézu $m(x) \in \{\theta_1 x\}$. Počet MC simulácií pre rôzne podkladové časové rady

je 500, počet bootstrapových replikácií 1000, správne rozdelenie štatistiky $nh^{1/2}T_n$ určené MC metódou pomocou 5000 simulácií. Na obrázku 7.1 vidíme správne rozdelenie štatistiky $nh^{1/2}T_n$ pre model GOF1 spolu so 4 aproximáciami metódou neparametrický wild bootstrap. Na obrázku 7.2 potom rozdelenie štatistiky $nh^{1/2}T_n$ spolu s aproximáciami pre model GOF2. Vidíme, že v oboch prípadoch bootstrapová metóda poskytuje veľmi dobrú aproximáciu rozdelenia spomínanej štatistiky. Tabuľka 7.1 ďalej obsahuje hodnoty empirických hladín testov vypočítaných pomocou bootstrapovej metódy v modeloch GOF1 a GOF2 pre dve rôzne rozdelenia chýb ε_t a $n = 100$. Nominálna hodnota hladiny testu bola $\alpha = 5\%$. Vidíme, že pre vyrovnávací parameter blízko optimálneho, sú empirické hodnoty testov blízko nominálnej hodnoty.



Obrázok 7.1. Hustota štatistiky $nh^{1/2}T_n$ pre model $X_t = 0.9 \cdot X_{t-1} + \varepsilon_t$ a $n = 100$ vypočítaná pomocou 5000 MC simulácií (hrubá čiara), spolu so 4 aproximáciami metódou neparametrický wild bootstrap. Vľavo model, kde $\varepsilon_t \sim N(0, 1)$, vpravo model s $\varepsilon_t \sim Exp(1)$.

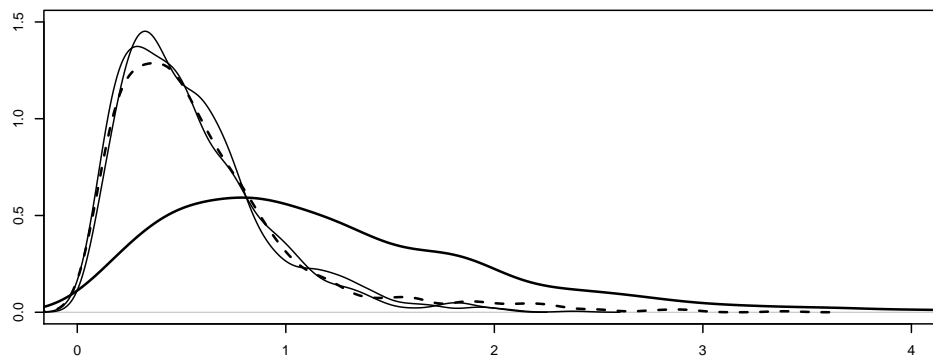


Obrázok 7.2. Hustota štatistiky $nh^{1/2}T_n$ pre model $X_t = 4 \cdot \sin(X_{t-1}) + \varepsilon_t$ a $n = 100$ vypočítaná pomocou 5000 MC simulácií (hrubá čiara), spolu so 4 aproximáciami metódou neparametrický wild bootstrap. Vľavo model, kde $\varepsilon_t \sim N(0, 1)$, vpravo model s $\varepsilon_t \sim Exp(1)$.

$\mathcal{L}(\varepsilon_t)$	GOF1			GOF2			
	$h = 1.2$	$h = 1.5$	$h = 1.8$	$h = 0.7$	$h = 1.0$	$h = 1.3$	
normal	0.092	0,040	0.025	normal	0.081	0.036	0.030
exp	0.088	0.044	0.021	exp	0.098	0.048	0.031

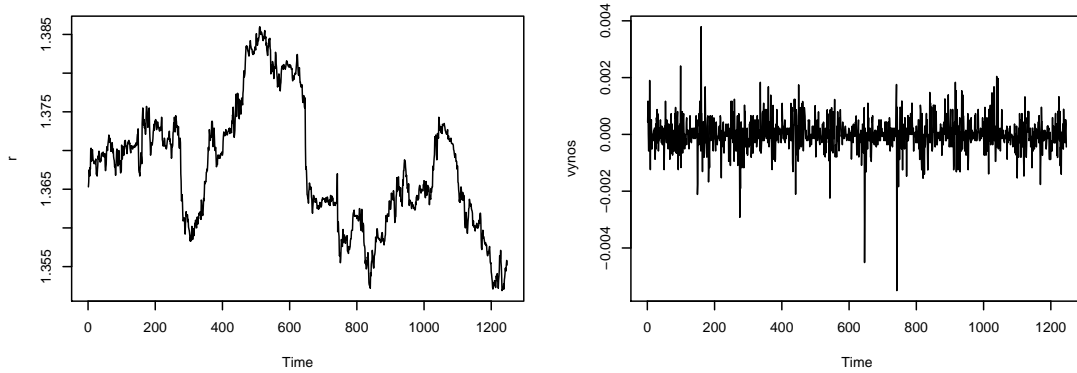
Tabuľka 7.1. Hodnoty empirických hladín testov vypočítaných pomocou bootstrapovej metódy v modeloch GOF1 a GOF2 pre $n = 100$. Nominálna hodnota hladiny testu $\alpha = 0.05$. Hodnoty vypočítané pomocou 500 MC simulácií, počet bootstrapových replikácií 1000.

Princíp tohto postupu spočíva v tom, že pokiaľ predpokladáme, že dáta pochádzajú z iného modelu ako v skutočnosti, rozdelenie štatistiky T_n bude iné ako rozdelenie štatistiky S_n^* . V skutočnosti platí, že rozdelenie S_n^* (resp S_n) vždy zodpovedá rozdeleniu za platnosti nulovej hypotézy. Ako príklad môžeme uviesť nasledujúcu situáciu. Predpokladajme, že máme pozorovania, ktoré v skutočnosti pochádzajú z modelu $X_t = 0.9 \cdot \sin(X_{t-1}) + \varepsilon_t$, ale my máme podozrenie, že môžu pochádzať z modelu $X_t = 0.9 \cdot X_{t-1} + \varepsilon_t$. Preto testujeme hypotézu o linearite $m(\cdot)$. Pokiaľ aplikujeme vyššie uvedený postup na dáta, ktoré pochádzajú z nelineárneho modelu a vypočítame rozdelenie T_n aj S_n^* , vidíme na obrázku 7.3, že rozdelenie oboch štatistík je značne odlišné. Podobná simulácia bola konštruovaná v práci Kreiss, Neumann a Yao (2008).



Obrázok 7.3. Hustota štatistiky $nh^{1/2}T_n$ (hrubá čiara), hustota štatistiky $nh^{1/2}S_n$ (prerušovaná čiara) a dve bootstrapové aproximácie pre $n = 100$. Rozdelenie štatistík $nh^{1/2}T_n$ a $nh^{1/2}S_n$ vypočítané pomocou 5000 MC simulácií, bootstrapové aproximácie určené na základe 1000 replikácií.

V tejto časti budeme aplikovať metódu bootstrap na analýzu reálnych finančných časových radov. Zoberme si dáta z príkladu 2.1, čiže časový rad relatívnych zmien devízového kurzu EUR/USD meraného v 15-minútových intervaloch v období od 25. 01. 2011 do 11. 02. 2011. Budeme opäť pracovať s časovým radom $\{r_t, t = 1, \dots, n\}$, $r_t = (P_{t+1} - P_t)/P_t$, kde P_t je daný devízový kurz v čase t . Dĺžka časového radu je $n = 1247$. Na obrázku 7.4 vidíme priebeh devízového kurzu spolu s percentuálnymi zmenami. V príklade 2.1 sme z autokorelogramov a parciálnych autokorelogramov vyčítali, že časový rad druhých mocnín vykazuje autokorelovanosť, z čoho plynie prítomnosť ARCH štruktúry v tomto časovom rade.



Obrázok 7.4. Priebeh časového radu $\{P_t\}$ percentuálnych zmien devízového kurzu EUR/USD meraného v 15-minútových intervaloch, v období od 25. 01. 2011 do 11. 02. 2011, spolu priebehom časového radu $\{r_t\}$.

Budeme teda chcieť testovať hypotézu, či model ARCH(1) je vhodný model pre časový rad $\{r_t\}$. Implikovaný model má tvar $r_t = \sqrt{\alpha_0 + \alpha_1 r_{t-1}^2} \varepsilon_t$ a ako sme uviedli vyššie, tento zápis je ekvivalentný s $r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + (\alpha_0 + \alpha_1 r_{t-1}^2)(\varepsilon_t^2 - 1)$. Chyby ε_t budeme predpokladať nezávislé a s rozdelením $N(0, 1)$. Pomocou štatistiky (7.2) budeme merať L_2 vzdialenosť medzi parametrickým odhadom $\hat{\alpha}_0 + \hat{\alpha}_1 x^2$ a neparametrickým odhadom veličiny $m_2(x) = E(r_t^2 \mid r_{t-1} = x)$ na intervale $[-0.001, 0.001]$. V tomto intervale sa nachádza 94 % pozorovaných dát.

Pre konštrukciu parametrického odhadu využijeme funkciu *garchFit* v štatistickom programe R, ktorá konštruje QMLE odhady. Na odhad volatility sme použili vlastnú funkciu založenú na Nadaraya-Watsonovom odhade v tvare

$$\hat{m}_2(x) = \frac{\sum_{t=1}^{n-1} K_h(x - r_t) r_{t+1}^2}{\sum_{t=1}^n K_h(x - r_t)}.$$

Na implementáciu tejto metódy budeme potrebovať optimálny vyhladzovací parameter h . Použijeme postup navrhnutý v práci Yao a Tong (1998), ktorý je založený na minimalizácii

$$\min_{h \in H_m} ECV_m(h) = \frac{1}{n - m} \sum_{t=m+1}^n [X_{t+1} - \hat{f}(X_t)]^2 w(X_t),$$

kde konštanta $m \in \{1, \dots, n\}$ nám rozdelí dáta X_1, \dots, X_n na dve časti, $\{X_t, 1 \leq t \leq m\}$ a $\{X_t, m + 1 \leq t \leq n\}$ a \hat{f} je odhad autoregresnej funkcie $m(\cdot)$ na základe prvých m pozorovaní. Množinu H_m v tomto článku uvažujú v tvare

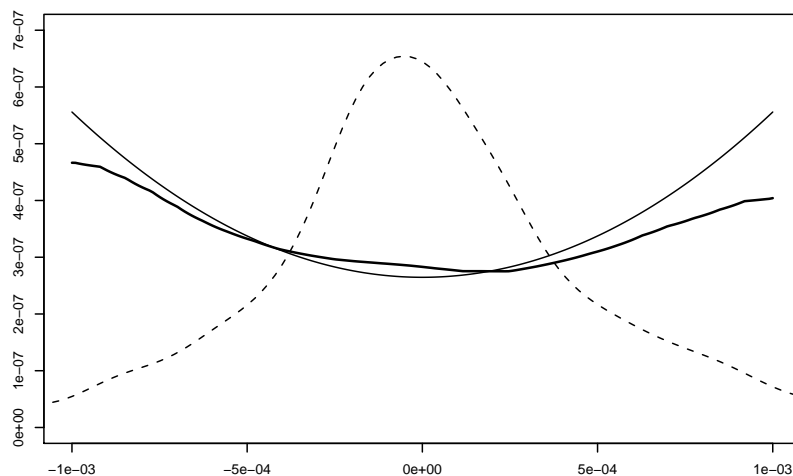
$$H_m = [am^{-\frac{1}{5} - \varepsilon_0}, bm^{-\frac{1}{5} + \varepsilon_0}],$$

kde $0 < a < b < \infty$ a $\varepsilon_0 \in (0, \frac{1}{150})$ sú konštanty. Optimálny vyhladzovací parameter h určíme zo vzťahu

$$h = h_1 \left(\frac{m}{n} \right)^{1/5},$$

kde h_1 je vyhladzovací parameter, pre ktorý je $ECV_m(h)$ minimálna.

Na obrázku 7.5 vidíme pre porovnanie parametrický a neparametrický odhad autoregresnej funkcie. V miestach, kde je hustota dostatočne veľká, rozdiely medzi parametrickým a neparametrickým odhadom je malý, čo by značilo, že ARCH model je vhodný kandidát na modelovanie našich dát. V intervaloch extrémnych hodnôt radu r_t už sa rozdiely medzi odhadmi zväčšujú. Optimálny vyhladzovací parameter bol zvolený $h = 7 \cdot 10^{-4}$.



Obrázok 7.5. Parametrický odhad $\hat{\alpha}_0 + \hat{\alpha}_1 x^2$ (hrubá čiara), neparametrický odhad veličiny $E(r_t^2 | r_{t-1} = x)$ (tenká čiara) a hustota X_1, \dots, X_n (prerušovaná čiara) vynásobená 10^{-9} .

Hypotéza H_0 teda testuje validitu modelu $X_t = \sigma_\alpha(X_{t-1})\varepsilon_t$, kde $\sigma_\alpha(x) = \sqrt{\alpha_0 + \alpha_1 x^2}$. Položme najprv v definícii (7.2) $w(x) \equiv 1$. Na hladine $\alpha = 10\%$ má testová štatistika hodnotu $nh^{1/2}T_n = 3.03 \cdot 10^{-12}$. Na základe 10000 bootstrapových simulácií sme vypočítali $u_{1-\alpha}^* = 3.36 \cdot 10^{-12}$, z toho vyplýva, že ARCH model pre naše dáta nemôžeme zamietnuť.

V práci Kreiss, Neumann a Yao (2008) použili túto hypotézu na potvrdenie asymetrie v časovom rade denných uzatváracích cien indexu DAX nemeckej burzy cenných papierov. Použijeme ich postup na otestovanie asymetrie v časovom rade percentuálnych zmien devízového kurzu EUR/USD. Budeme postupovať rovnako ako vyššie, s tým rozdielom, že použijeme dve rôzne váhy, $w_1 = 1_{[-0.001, -0.0001]}$ a $w_2 = 1_{[0.0001, 0.001]}$, čiže testujeme vhodnosť ARCH modelu jednotlivo na zápornej a kladnej osi. Na hladine $\alpha = 5\%$ je hodnota štatistiky rovná $nh^{1/2}T_n = 2.34 \cdot 10^{-12}$ a $u_{1-\alpha}^* = 1.99 \cdot 10^{-12}$ pre váhu w_1 a $nh^{1/2}T_n = 6.50 \cdot 10^{-13}$ a $u_{1-\alpha}^* = 1.94 \cdot 10^{-12}$ pre váhu w_2 . Vidíme, že test na hladine $\alpha = 5\%$ zamietol hypotézu o ARCH štruktúre na zápornej osi, ale nezamietol tú istú hypotézu na osi kladnej. Tento výsledok by sa dal vysvetliť asymetrickým tvarom funkcie volatility.

7.1 Implementácia

Značná časť výpočtov bola naprogramovaná pomocou štatistického softvéru R 2.11.0. Softvér obsahuje množstvo vnútorne implementovaných procedúr, ktoré uľahčili a zrýchlili priebeh simulácií. Metóda bootstrap, ktorá pracuje na princípe metód Monte Carlo, využíva hlavne opakované volania toho istého algoritmu pre rôzne dáta. Keďže software R, aj napriek tomu, že patrí medzi najrýchlejšie štatistické počítačové programy, nie je dobre optimalizovaný pre vnorené cykly, bolo potrebné zefektívniť časť výpočtovo náročných výpočtov. To sme dosiahli použitím jazyka Fortran, v ktorom sú neimplementované základné „rekurzívne“ procedúry, ako napr. generovanie autoregresných postupností, usporiadavanie výberov a pod. Ďalšia optimalizácia kódu sa dosiahne hlavne vektorizovaním, čiže nahradením cyklu za vektorové operácie. S použitím týchto praktík sa dosiahlo zrýchlenie niektorých výpočtov až o 600 %. Parametrické metódy sú jednoduchšie na naimplementovanie a obvykle aj rýchlejšie, neparametrické metódy, ktoré vyžadujú zložité generovanie bootstrapového radu pomocou neparametrických odhadov (a v prípade autoregresného bootstrapu dokonca rekurzívne, keďže nemôžeme vektorizovať) sú výpočtovo náročnejšie. Uvedme napríklad, že sada simulácií na určenie empirických hladín testov v „goodnes of fit“ trvala približne 3 hodiny pre $n = 100$ a 5 hodín pre $n = 200$. Výpočet kritických hodnôt v simulácii na reálnych dátach prebiehal približne 10 minút.

Záver

Cieľom práce bolo porovnať presnosť bootstrapových metód pre parametrické a neparametrické modely časových radov. Ukázali sme, že metódu bootstrap môžeme s úspechom využiť pri štatistickom skúmaní vlastností odhadov, odhade ich charakteristík, hľadani konfidenčných intervalov, resp. kritických hodnôt.

V prípade parametrických modelov časových radov sme skúmali vlastnosti štyroch rôznych bootstrapových metód. Reziduálny bootstrap môžeme použiť v prípade, že naše dáta nevykazujú podmienenú heteroskedasticitu, kedy sú výsledky získané touto metódou veľmi presné. V prípade podmienene heteroskedastických chýb už ale metóda zaostáva, pretože *iid* prevzorkovaním rezíduí sa stráca pôvodná závislostná štruktúra. Ďalšie metódy, ktoré sme predstavili, už túto závislostnú štruktúru zohľadňujú a ich presnosť je oveľa lepšia. Ako najpresnejšia metóda sa ukázal rekurzívny wild bootstrap, ktorý svojou presnosťou preyšuje aj jeho klasický (fixný) variant - fixný wild bootstrap, takisto ako aj metódu párový bootstrap. Metóda párový bootstrap zaostáva hlavne z dôvodu veľkej variability odhadov charakteristík.

V prípade neparametrických modelov časových radov majú metódy bootstrap takisto svoje uplatnenie. V práci sme porovnali dve metódy. Autoregresný a neparametrický wild bootstrap. Metóda autoregresný bootstrap, ktorá kompletne kopíruje závislostnú štruktúru dát pri prevzorkovaní, sa ukázala ako spoľahlivejšia metóda oproti neparametrickému wild bootstrapu. Presnosť metód sme porovnávali na základe schopnosti správne odhadnúť rozdelenie jadrových odhadov pre autoregresnú funkciu. Pre odhad rozdelenia jadrového odhadu volatility sme ukázali, že metóda autoregresný bootstrap je na tento účel vhodný nástroj. Pre volatilitu ale platí, že aproximácia rozdelenia tohto odhadu pri relatívne malom počte pozorovaní nie je presná. Tento nedostatok môžeme pripísať všeobecne známemu fakt, že neparametrické odhady volatility sú spoľahlivé až pri väčšom počte pozorovaní.

Pri porovnaní parametrických a neparametrických metód sa ukázalo, že neparametrické metódy nezaostávajú za parametrickými v prípade podmienene homoskedastických dát a dokonca v prípade podmienene heteroskedastických dát neparametrická autoregresná metóda preyšuje svojou presnosťou obidve parametrické metódy.

V aplikácii na reálne dáta sme ukázali, že metóda neparametrický wild bootstrap môže byť použitá pri konštruovaní testov zhody s modelom. Konkrétne sme skúmali otázku vhodnosti určitého parametrického modelu pre pozorované dáta. Pomocou bootstrapovej metódy sme konštruovali kritické hodnoty testovej štatistiky. Algoritmus sme následne použili pre časový rad percentuálnych zmien devízového kurzu EUR/USD meraného v 15-minútových intervaloch.

Literatúra

Bougerol, P. a Picard, N. (1992). *Strict stationarity of generalized autoregressive processes*. Journal of Econometrics, 52, 115-127.

Bühlmann, P. (1997). Sieve bootstrap for time series. Bernoulli 3, 123–148.

Campbell, J. Y., Lo, A. W. a MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Cipra, T. (2008). *Finanční ekonometrie*. Ekopress.

Efron, B. (1979). *Bootstrap Methods: Another Look at the Jackknife*. The Annals of Statistics, 7, 1-26.

Engle, R. F. (1982). *Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation*. Econometrica, 50, 987-1007.

Engle, R. F., Lilien, D.M. a Robins, R.P. (1987). *Estimating time varying risk premia in the term structure: the ARCH-M model*. Econometrica, 55, 391-407.

van Es, B. a Putter, H. (2009). *The Bootstrap*. Universiteit van Amsterdam, Korteweg-de Vries Institute for Mathematics.

Fan, J. a Yao, Q. (1998). *Efficient estimation of conditional variance functions in stochastic regression*. Biometrika, 85, 1(645)-25(660).

Fan, J. a Yao, Q. (2003): *Nonlinear Time Series, Nonparametric and Parametric Methods*. Springer, New York.

Franke, J., Kreiss, J.-P. a Mammen, E. (2002a). *Bootstrap of kernel smoothing in nonlinear time series*. Bernoulli, 8, 1–37.

Franke, J., Kreiss, J.-P., Mammen, E. a Neumann, M. H. (2002b). *Properties of the nonparametric autoregressive bootstrap*. Journal of Time Series Analysis, 23, 555–585.

- Franke, J., Neumann, M. H. a Stockis, J.-P. (2004). *Bootstrapping nonparametric estimators of the volatility function*. Journal of Econometrics, 118, 189–218.
- Freedman, D. A. (1981). *Bootstrapping Regression Models*. Annals of Statistics, 9, 1218–1228.
- Glosten, L. R., Jagannathan, R. a Runkle, D. E. (1993). *On the relation between the expected value and the volatility of nominal excess return on stocks*. Journal of Finance, 48, 1779–1801.
- Goncalves, S. a Kilian, L. (2004). *Bootstrapping autoregressions with conditional heteroskedasticity of unknown form*. Journal of Econometrics, 123, 89–120.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag.
- Hand, D., Daly, F., Lunn A., McConway, K. a Ostrowski, E. (1994). *A Handbook of Small Data Sets*. Chapman & Hall.
- Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, New York.
- Härdle, W., Müller, M., Sperlich, S. a Werwatz, A. (2004). *An introduction to Non- and Semiparametric Models*. Springer, Berlin.
- Hjellvik, V. a Tjøstheim, D. (1995). *Nonparametric tests of linearity for time series*. Biometrika, 82, 351–368.
- Kreiss, J.-P., Neumann, M. H. a Yao, Q. (2008). *Bootstrap tests for simple structures in nonparametric time series regression*. Statistics and its interface, Volume 1, 367–380.
- Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer.
- Liu, R. Y. (1988): *Bootstrap Procedure under Some Non-i.i.d. Models*, Annals of Statistics, 16, 1696–1708.
- Milhoj, A. (1985). *The Moment Structure of ARCH Processes*. Scandinavian Journal of Statistics, 12, 281–292.
- Nelson, D. B. (1990). *Stationarity and persistence in the GARCH(1, 1) model*. Econometric Theory, 6, 318–334.
- Nelson, D. B. (1991). *Conditional Heteroskedasticity in Asset Returns: A New Approach*. Econometrica, 59, 347–370.

- Paparoditis, E. a Politis, D. N. (2001). *The local bootstrap for Markov processes*. Journal of Statistical Planning and Inference, 108, 301-328.
- Prášková, Z. (2004a). *Metoda bootstrap*. Robust'2004.
- Prášková, Z. (2004b). *Základy náhodných procesů II*. Karolinum, Praha.
- Rajarshi, M. B. (1990): *Bootstrap in Markov-Sequences Based on Estimates of Transition Density*. Annals of the Institute of Statistical Mathematics, 42, 253-268.
- Singh, K. (1981). *On the asymptotic accuracy of efron's bootstrap*. Annals of Statistics, 9 (6), 1187-1195.
- Stout, W. F. (1974). *Almost Sure Convergence*. Academic Press, New York.
- Tsay, R. S. (2005). *Analysis of Financial Time Series*. Wiley, New York (druhé vydanie).
- Vaněček, P. (2003). *Diskrétní modely s měnící se volatilitou*. Diplomová práce, MFF.
- Wu, C. F. J. (1986). *Jackknife, bootstrap and other resampling methods in regression analysis*. Annals of Statistics, 14.
- Yao, Q. a Tong, H. (1998). *Cross-validators bandwidth selections for regression estimation based on dependent data*. Journal of statistical planning and inference, 68 (2), 387-415.

Použité značenie

$T_n = T_n(X_1, \dots, X_n)$	odhad parametra θ na základe výberu X_1, \dots, X_n
$T_n^* = T_n(X_1^*, \dots, X_n^*)$	odhad parametra θ na základe bootstrapového výberu X_1^*, \dots, X_n^*

OLS	odhad metódou najmenších štvorcov (Ordinárny Least Squares)
DGP	Proces generovania dát (Data Generating Process)
n	počet pozorovaní
$\text{diag}(a_t)$	diagonálna matica s prvkami $\{a_t, t = 1, \dots, n\}$ na diagonále
m.d.s.	postupnosť martingalových diferencií (Martingale Difference Sequence)
<i>iid</i>	nezávislé a rovnako rozdelené (náhodné veličiny)

\xrightarrow{P}	konvergencia v pravdepodobnosti
\xrightarrow{D}	konvergencia v distribúcii
$\xrightarrow{a.s.}$	konvergencia takmer iste

$x_n = o(n^k)$	$x_n/n^k \rightarrow 0$, pre $n \rightarrow \infty$, kde $\{x_n\}$ je reálna postupnosť
$x_n = O(n^k)$	$\limsup_{n \rightarrow \infty} x_n /n^k < \infty$, kde x_n je reálna postupnosť
$X_n = o_P(n^k)$	$x_n/n^k \xrightarrow{P} 0$, pre $n \rightarrow \infty$, kde $\{X_n\}$ je reálna postupnosť
$X_n = O_P(n^k)$	$(\forall \varepsilon > 0) (\exists M \in (0, \infty))$ také, že $\sup_{n \geq 1} P(x_n /n^k > M) < \varepsilon$, kde $\{X_n\}$ je postupnosť náhodných veličín

Zoznam tabuliek

1.1	Percentuálne vyjadrenie nepokrytia správnej hodnoty štatistiky konfidenčnými intervalmi.	15
5.1	Percentuálne pokrytie percentilových intervalov spoľahlivosti pre parameter φ_1 za predpokladu nulovosti interceptu φ_0 .	39
6.1	Percentuálne pokrytie 90% bootstrap-t intervalov spoľahlivosti pre parameter φ_1 za predpokladu nenulovosti interceptu φ_0 .	56
6.2	Kvalita aproximácie rozptylu parametru φ_1 v AR(1)-GARCH(1,1).	58
6.3	Kvalita aproximácie rozptylu parametru φ_1 v AR(1)-EGARCH(1,1).	59
6.4	Kvalita aproximácie rozptylu parametru φ_1 v AR(1)-GJR-GARCH(1,1).	59
6.5	Percentuálne pokrytie 90% percentilových intervalov spoľahlivosti pre parameter φ_1 za predpokladu nenulovosti interceptu φ_0 .	60
6.6	Hodnoty dolného kvartilu, mediánu a horného kvartilu RASE odhadnutej hustoty štatistiky (6.2) metódou autoregresný a neparametrický wild bootstrap.	64
6.7	Hodnoty dolného kvartilu, mediánu a horného kvartilu RASE odhadnutej hustoty štatistiky (6.4) metódou autoregresný a neparametrický wild bootstrap	67
6.8	Hodnoty vyhladzovacích parametrov pre jednotlivé počty pozorovaní použitých pri aproximácii rozdelenia štatistiky (6.6).	67
6.9	Hodnoty dolného kvartilu, mediánu a horného kvartilu RASE odhadnutej hustoty štatistiky (6.6) metódou autoregresný bootstrap	68
6.10	Kvalita aproximácie rozptylu parametru φ_1 v modeli $X_t = 0.5X_{t-1} + \varepsilon_t$	69
6.11	Kvalita aproximácie rozptylu parametru φ_1 v modeli $X_t = \varepsilon_t$	69
6.12	Kvalita aproximácie rozptylu parametru φ_1 v modeli $X_t = \sqrt{0.5 + 0.25X_{t-1}^2}\varepsilon_t$	70
7.1	Hodnoty empirických hladín testov vypočítaných pomocou bootstrapovej metódy v modeloch GOF1 a GOF2 pre $n = 100$.	74