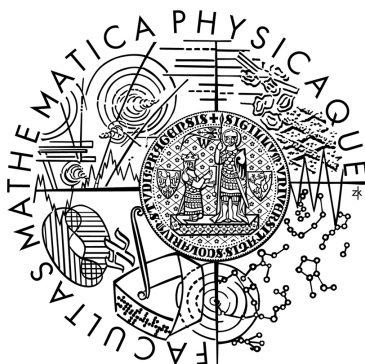


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Jakub Šindelář

M-odhady

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Prof. RNDr. Jaromír Antoch, CSc.

Studijní program: Matematika, Finanční matematika

2010

Děkuji Prof. RNDr. Jaromíru Antochovi, CSc. za vedení, konzultace a pomoc při vytváření bakalářské práce.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 28.5.2010

Jakub Šindelář

Obsah

1 Úvod	6
1.1 Robustní odhady	7
1.2 M-odhady	9
2 Modely	10
2.1 Lineární regresní model	10
2.2 Autoregresní (AR) model	10
2.3 ARMA model	11
3 Odlehlá pozorování	12
3.1 Odlehlá pozorování způsobená inovacemi	12
3.2 Odlehlá pozorování zapříčiněná aditivními efekty	13
4 Off-line robustní odhady v lineárních regresních modelech a ARMA modelech	14
4.1 M-odhad založený na Huberově návrhu	14
4.2 GM-odhad založený na Kraskerově a Welschově návrhu	16
5 Rekurzivní a adaptivní odhady	18
5.1 Rekurzivní metoda nejmenších čtverců (RLS)	18
5.2 Modifikovaný RLS algoritmus	21
5.3 Rekurzivní minimalizace Huberovy funkce (RHU)	22
5.3.1 Odvození RHU	23
5.4 Rekurzivní algoritmus založený na Kraskerově a Welschově odhadu (RKW)	25
5.4.1 Odvození RKW	25
5.5 Rekurzivní odhad měřítka	28

6 Aplikace algoritmů	30
6.1 Simulace dat	30
6.2 Aplikace algoritmů na nasimulovaná data	31
6.3 Hodnocení algoritmů	32
6.4 Diskuze výsledků	33
7 Závěr	43
Literatura	44

Název práce: M-odhady

Autor: Jakub Šindelář

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Prof. RNDr. Jaromír Antoch, CSc.

e-mail vedoucího: jaromir.antoch@karlin.mff.cuni.cz

Abstrakt: M-odhady tvoří důležitou alternativu k metodě nejmenších čtverců. Jedná se o velice širokou třídu odhadů. Proto jsem se v této práci zaměřil na rekurzivní robustní M-odhady parametrů v autoregresních modelech. V úvodu uvedu něco málo o robustních odhadech, myšlenky jejich vzniku a M-odhadech, jakožto třídě robustních odhadů. Dále se věnuji obecnému zavedení modelů, na které budeme naše algoritmy odhadu aplikovat, a fenoménu odlehlých pozorování, která se v datech často vyskytují. Odvodíme dva rekurzivní robustní algoritmy odhadu, které vycházejí z off-line odhadů, a porovnáme je na simulovaných datech společně s rekurzivní verzí metody nejmenších čtverců a její modifikací, kdy dochází k zamítnutí odlehlých pozorování.

Klíčová slova: M-odhady, robustní odhady, rekurzivní odhady, odlehlá pozorování, lineární regrese

Title: M-estimators

Author: Jakub Šindelář

Department: Department of Probability and Mathematical Statistics

Supervisor: Prof. RNDr. Jaromír Antoch, CSc.

Supervisor's e-mail address: jaromir.antoch@karlin.mff.cuni.cz

Abstract: M-estimators form an important alternative to the least squares method. It is a wide class of estimators. Therefore, I focused my attention to the methods for recursive robust estimation of autoregressive parameters. In the introduction, I will summarize some basic facts about robust estimates in general, the idea of their introduction, and M-estimates as a class of robust estimators. Further, I specify the models which will be used and describe the phenomenon of outliers, which often occur in the data. We derive two recursive algorithms, that are based on the off-line estimates, and compare them with recursively computed least squares and a modification, in which we reject the outliers from the estimation.

Keywords: M-estimators, robust estimation, recursive estimation, outliers, linear regression

Kapitola 1

Úvod

V této práci se budu věnovat metodám rekurzivních robustních odhadů parametrů v autoregresních modelech. V mnoha případech jako automatizace nebo adaptivní řízení systémů je potřeba použít on-line odhady parametrických modelů, aby parametry modelu sledovaly změny systému v čase. Metoda odhadu proto musí být adaptivní a předpokládat použití rekurzivního algoritmu odhadu. Kvůli možnosti výskytu odlehlých pozorování navíc požadujeme, aby použitý algoritmus odhadu minimalizoval vliv jakéhokoli případného odlehlého pozorování. Je důležité, ale většinou velmi obtížné, rozlišit chyby modelu způsobené hrubými chybami v datech od chyb modelu plynoucích z náhlých změn v systému. Žádoucí vlastností algoritmu by tedy bylo, aby byl při odhadu parametrů schopen chránit odhad před destruktivním vlivem hrubých chyb a zároveň zachoval možnost sledování změn v systému. V práci odvodíme dva rekurzivní algoritmy pro odhady parametrů AR modelů, které vycházejí z robustních off-line odhadů. Jeden pracuje s rekurzivní minimalizací kritériální funkce, kde rezidua vstupují skrze váhovou funkci navrženou Huberem v [4]. Druhý je rekurzivní verzí GM-odhadu navrženého Kraskerem a Welschem v [7]. Tento odhad je rozšířením Huberova odhadu, kde ke stanovení horní hranice vlivu jednotlivých pozorování užijeme míru informace v regresním vektoru. Těmito odvozeními ukážeme obecný postup pro získání rekurzivních algoritmů. V poslední kapitole na simulovaných datech srovnáme obě rekurzivní metody společně s rekurzivní metodou nejmenších čtverců a její modifikací, kdy odlehlá pozorování vynecháme z odhadu.

1.1 Robustní odhady

Při zpracování dat klasickými statistickými postupy, které jsou založeny na parametrických modelech (model je plně určen až na hodnoty několika parametrů, které nabývají reálných nebo vektorových hodnot), spoléháme na platnost určitých předpokladů jako jsou náhodnost, nezávislost pozorování, rozdělení chyb, homoskedasticita a další. Tyto postupy sice odvozují optimální postupy pro přesné parametrické modely, ale neříkají nic o jejich chování, pokud jsou modely platné pouze přibližně. *Robustní statistické metody* jsou modifikací klasických postupů, která neselže při malých odchylkách od základních předpokladů. Jsou optimální v okolí daného rozdělení, vzhledem k určité vzdálenosti a určitému kritériu optimality.

Pro naše účely bude robustnost značit necitlivost k odchylkám od předpokladů. Podle toho, vůči jakým předpokladům je metoda stabilní, mluvíme například o robustnosti vzhledem k rozdělení nebo o robustnosti vzhledem k modelu.

Technický termín „robustní“ byl zaveden roku 1953 G.E.P. Boxem, ale k jeho legitimnímu uznání došlo až o dekádu později. Přesněji je pojem robustnost definován např. v práci [3]. Nejednalo se však o revoluční koncept. Mezi vědci konce 19. a počátku 20. století bylo několik praktických statistiků, kteří naprosto jasně a funkčně rozuměli problému, znali nebezpečí rozdělení chyb s těžkými chvosty, navrhli pravděpodobnostní modely pro hrubé chyby, a dokonce vymysleli výborné robustní alternativy ke standardním odhadům.

Při práci s parametrickým modelem, u kterého většinou doufáme, že je dobrou aproximací reálné situace, ale nemůžeme předpokládat a nepředpokládáme, že je absolutně přesný, klademe na všechny statistické postupy následující požadavky:

1. Měly by mít dostatečně dobrou (optimální nebo skoro optimální) efektivitu na předpokládaném modelu, tedy by se měly lišit pouze minimálně od klasických optimálních metod při přesném splnění podmínek jejich optimality.
2. Měly by být robustní v tom smyslu, že při malých odchylkách od předpokladů modelu zůstávají vysoce efficientní na rozdíl od klasických optimálních metod.
3. Větší odchylky od modelu by neměly mít katastrofální následky.

Jednou z častých příčin nepřesnosti napozorovaných dat je výskyt hrubých chyb. Většinou se tyto chyby projeví jako odlehlá pozorování, která se vyskytují „daleko“ od hlavní skupiny dat. Často vznikají při kopírování nebo nepřesnostech při zápisu, ale mohou být způsobeny také volbou špatného modelu či rozděleními s těžkými chvosty. Problém s odlehlými pozorováními se ve statistice objevuje již od jejich počátků a způsoby, jak se s nimi vypořádat, jako subjektivní nebo objektivní zamítnutí, patří v širším smyslu k robustním metodám.

Dalšími faktory, které způsobují odchylky od ideálních předpokladů modelu jsou empirický charakter mnoha modelů a aproximační charakter mnoha teoretických modelů. Mnozí tvrdí, že pouze nepatrná chyba v matematickém modelu způsobí také jen nepatrnou chybu v našich závěrech. Ukázalo se však, že většina běžných statistických postupů (většinou ty, které pracují s předpokladem normálního rozdělení) jsou citlivé na zdánlivě nevýrazné odchylky od předpokladů. Mohli bychom si položit otázku, zda je nutné robustní metody vůbec používat, a zda by nestačilo aplikovat následující dvoukrokový přístup:

1. Vyčistit data aplikací nějakého pravidla pro zamítnutí odlehlých pozorování.
2. Použít klasický odhad a testovací metody na zbytek.

Tento přístup by bohužel nefungoval, a to hned z několika důvodů. Jen málokdy je možné oddělit oba kroky čistě. Například ve vícerozměrné regresi je obtížné rozpoznat problémová odlehlá pozorování, pokud nemáme spolehlivé robustní odhady a dobrý model. I když se původní skupina pozorování sestává z normálních pozorování s několika hrubými chybami, očištěná data nemusí být nutně normální. Obecně je prokazatelné, že nejlepší metody zamítnutí nedosahují úrovně nejlepších robustních metod, které se považují za nadřazenější, jelikož umožňují hladký přechod mezi úplným zamítnutím a úplným přijmutím pozorování. Navíc se ukázalo, že řada klasických zamítacích pravidel není schopna se vypořádat s více odlehlými pozorováními (může se stát, že druhé odlehlé pozorování maskuje to první, takže nezamítneme žádné).

1.2 M-odhady

M-odhady jsou jednou ze tříd robustních odhadů a byly zavedeny Huberem v práci [4]. Pracujeme-li například s lineárním regresním modelem

$$y_i = \mathbf{x}_i^T \boldsymbol{\theta} + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

pak nejznámějším odhadem $\boldsymbol{\theta}$ je klasický odhad *metodou nejmenších čtverců*, kdy minimalizujeme součet čtverců reziduí, tedy hledáme neznáme parametry takové, aby součet čtverců rozdílů mezi pozorovanými a napočítanými hodnotami byl minimální, tj.

$$\sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\theta})^2 := \min \quad \text{vzhledem k } \boldsymbol{\theta}.$$

Huber si položil otázku, zda není možné dosáhnout větší robustnosti minimalizováním jiné než kvadratické ztrátové funkce. Dostaneme se tak k definici M-odhadu, která je v podstatě zobecněním metody nejmenších čtverců. M-odhad parametru $\boldsymbol{\theta}$ v modelu (1.1) je definován jako řešení minimalizace

$$\sum_{i=1}^n \rho(y_i - \mathbf{x}_i^T \boldsymbol{\theta}) := \min \quad \text{vzhledem k } \boldsymbol{\theta},$$

kde ρ je vhodně zvolená ztrátová funkce, která by měla omezit vliv velkých reziduí a měla by tudíž být funkcí, jež konverguje do nekonečna pomaleji než kvadratická funkce. *Maximálně věrohodný odhad* je speciálním případem, kdy $\rho(t) = \log f(t)$ a $f(t)$ je předpokládaná hustota rozdělení chyby.

Odhad

$$\hat{\boldsymbol{\theta}}_n = \operatorname{argmin}_{\boldsymbol{\theta}} \sum_{i=1}^n \rho(y_i - \mathbf{x}_i^T \boldsymbol{\theta}) \quad (1.2)$$

je ekvivariantní vzhledem k posunutí, ale není obecně ekvivariantní vzhledem k změně měřítka. Tento nedostatek můžeme vyřešit *studentizací*, tj. odhad definujeme jako řešení minimalizace

$$\sum_{i=1}^n \rho \left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\theta}}{\hat{\sigma}} \right) := \min,$$

kde $\hat{\sigma}$ je robustní odhad měřítka souboru pozorování. Je také možné společně s odhadem parametrů modelu odhadovat i parametr měřítka jak navrhuje např. Huber v [4].

Kapitola 2

Modely

2.1 Lineární regresní model

Základním modelem, se kterým budeme pracovat, je klasický *lineární regresní model*, tedy

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.1)$$

kde

$\mathbf{y} = (y_1, \dots, y_n)^T$ je vektor pozorování (závislých proměnných),
 $\mathbf{X} = (x_{ij})_{i=1, \dots, n}^{j=1, \dots, p}$ je matice regresorů,
 $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ značí i -tý řádek matice \mathbf{X} ,
 $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ je vektor parametrů modelu,
 $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ je neznámý vektor chyb.

Později uvedené rekurzivní algoritmy odhadu vyžadují, aby model, který použijeme k popisu dat, byl lineární v odhadovaných parametrech. Model lze tedy zapsat právě ve tvaru lineárního regresního modelu. Proto se v celé práci soustředíme na modely typu (2.1).

2.2 Autoregresní (AR) model

Jedná se o druh náhodného procesu, který se typicky používá k modelování časových řad a k předpovědím. K vyjádření pozorování X_t v čase

t je použito několik předchozích pozorování. Počet p pozorování z minulosti, která jsou přímo využita k vyjádření pozorování v čase t , určuje řád AR modelu (značíme $AR(p)$). Nepřímo je dané pozorování ovlivněno všemi pozorováními z minulosti, ale jejich vliv exponenciálně klesá.

$$X_t = \sum_{i=1}^p \theta_i X_{t-i} + \varepsilon_t,$$

kde

$\theta_1, \dots, \theta_p$ jsou parametry modelu,

ε_t je bílý šum (proces s nulovou střední hodnotou a rozptylem σ^2).

2.3 ARMA model

Tento model pro časové řady se skládá ze dvou částí, a to z autoregresní části, kterou jsme již popsali výše (*autoregressive (AR)*), a části klouzavých průměrů (*moving average (MA)*). Je často označován jako $ARMA(p, q)$ model, kde p značí řád AR části a q řád MA části. Model lze zapsat následovně:

$$X_t = \varepsilon_t + \sum_{i=1}^p \theta_i X_{t-i} + \sum_{i=1}^q \varphi_i \varepsilon_{t-i},$$

kde

$\theta_1, \dots, \theta_p$ a $\varphi_1, \dots, \varphi_q$ jsou parametry modelu,

ε_t a $\varepsilon_{t-i}, i = 1, \dots, q$, je bílý šum.

Samotný *MA model* by vypadal takto:

$$X_t = \varepsilon_t + \sum_{i=1}^q \varphi_i \varepsilon_{t-i}$$

Kapitola 3

Odlehlá pozorování

Odlehlá pozorování v časových řadách mohou nepříznivě ovlivnit odhady metodou nejmenších čtverců a obyčejné M-odhady autoregresních parametrů. Obvykle uvažujeme dva různé druhy modelů pro odlehlá pozorování. V prvním modelu vliv osamocené odlehlé pozorování přetrvává způsobem, který je konzistentní s autoregresní strukturou modelu. V druhém případě je odlehlé pozorování přechodného charakteru, což znamená, že se neváže na autoregresní strukturu. Pokud se budeme držet terminologie zavedené Foxem v [2], označíme první druh odlehlých pozorování jako „odlehlá pozorování způsobená inovacemi“ (*innovations outliers*) a druhý jako „odlehlá pozorování zapříčiněná aditivními efekty“ (*additive effects outliers*). Modely, které zde uvedeme jsou dosti zjednodušeným popisem odlehlých pozorování, která se objevují v reálných systémech. Jednotlivá odlehlá pozorování nelze vždy jednoznačně klasifikovat jako IO nebo AO. Často se odlehlá pozorování objevují ve skupinkách a jsou určitým způsobem korelována. Aditivní odlehlá pozorování způsobují řadu problémů při odhadech parametrů modelu a pro tradiční metody, jako je metoda nejmenších čtverců, získáváme výsledky se značným zkreslením. Proto je potřeba aplikovat iterativní metody, které zahrnují detekci a odfiltrování takových pozorování.

3.1 Odlehlá pozorování způsobená inovacemi

Pracujeme s modelem

$$\begin{aligned}x_k &= \theta x_{k-1} + \epsilon_k \\y_k &= x_k,\end{aligned}$$

kde $|\theta| < 1$ je parametr autoregresní časové řady prvního řádu, tedy AR(1), $y_k, k = 1, \dots, n$, jsou pozorování a inovace ϵ_k jsou nezávislé a stejně rozdělené náhodné veličiny. Předpokládáme, že hustota inovací f je symetrická kolem počátku. Odlehlá pozorování se objeví, pokud f je hustota rozdělení s těžkými chvosty. Například, pokud f je hustota kontaminovaného normálního rozdělení, tedy směsí dvou normálních rozdělení, a to $N(0, 1)$ a $N(0, \sigma^2)$. Rozdělení inovací bychom pak mohli zapsat schématicky jako

$$\epsilon_k \sim (1 - \gamma) \times N(0, 1) + \gamma \times N(0, \sigma^2),$$

kde $\gamma > 0$ a $\sigma^2 > 1$, a budeme jej nazývat smíšený normální model. Potom ϵ_k , a tedy i x_k , budou obsahovat odlehlá pozorování. γ je většinou poměrně malé, takže s velkou pravděpodobností jsou inovace z rozdělení $N(0, 1)$ a nějaká malá část jich bude z rozdělení $N(0, \sigma^2)$. Pokud je f hustota rozdělení $N(0, 1)$, pak se taková odlehlá pozorování neobjeví, ale pořád mluvíme o IO modelu.

3.2 Odlehlá pozorování zapříčiněná aditivními efekty

Zde je model vyjádřen

$$\begin{aligned} x_k &= \theta x_{k-1} + \epsilon_k \\ y_k &= x_k + v_k, \end{aligned}$$

kde inovace ϵ_k jsou normálně rozdělené s nulovou střední hodnotou a rozptylem σ_ϵ^2 , tedy $\epsilon_k \sim N(0, \sigma_\epsilon^2)$. Předpokládáme, že aditivní efekty v_k jsou nezávislé stejně rozdělené se smíšeným normálním rozdělením a mohli bychom jejich rozdělení schématicky vyjádřit jako

$$v_k \sim (1 - \gamma) \times 0 + \gamma \times N(0, \sigma^2),$$

kde $\gamma > 0$ a $\sigma^2 > 1$. Znamená to, že většinou jsou aditivní efekty rovny nule a v malém množství případů jsou z rozdělení $N(0, \sigma^2)$, tedy $P(v_k = 0) = 1 - \gamma$ a $P(v_k \neq 0) = \gamma$.

V tomto modelu se pozorování skládají z normálního „jádra“ x_k a aditivních „efektů“ v_k , kterým v praxi říkáme efekty spíše než chyby, jelikož jsou způsobeny různými příčinami (překlepy při zadávání dat, posuny ekonomických časových řad z důvodu např. stávky). Předpokládáme, že γ je malé (obvykle $\gamma \leq 0.2$), protože se ukázalo, že odlehlá pozorování se v časových řadách objevují pouze občas.

Kapitola 4

Off-line robustní odhady v lineárních regresních modelech a ARMA modelech

V práci změníme značení u lineárního regresního modelu (2.1). Od tohoto okamžiku budeme pracovat s modelem zapsaným ve tvaru

$$y(t) = \mathbf{x}^T(t)\boldsymbol{\theta} + \varepsilon(t) \quad (4.1)$$

pro hodnotu pozorování v čase t .

4.1 M-odhad založený na Huberově návrhu

Hledáme odhad $\hat{\boldsymbol{\theta}}$, který minimalizuje kriteriální funkci

$$V(n, \boldsymbol{\theta}) = \sum_{i=1}^n \rho_c \left(\frac{\varepsilon(i, \boldsymbol{\theta})}{\sigma} \right), \quad (4.2)$$

kde $\varepsilon(i, \boldsymbol{\theta}) = y(i) - \mathbf{x}^T(i)\boldsymbol{\theta}$, a σ je parametr měřítka. $\rho_c(u)$ je váhová funkce, která může být zvolena například dle Huberova návrhu v [4]

$$\rho_c(u) = \begin{cases} \frac{1}{2}u^2, & |u| \leq c, \\ c|u| - \frac{1}{2}c^2, & |u| > c. \end{cases}$$

	$\rho(x)$	$\psi(x)$	
Fair	$A^2 \left(\frac{ x }{A} - \ln \left(1 + \frac{ x }{A} \right) \right)$	$\frac{x}{1 + \frac{ x }{A}}$	$x \in \mathcal{R}_1$
Huber	$\frac{x^2}{2}$ $B x - \frac{B^2}{2}$	x $B \text{sign}(x)$	$ x \leq B$ $ x > B$
Tukey	$\frac{C^2}{6} \left(1 - \left(1 - \left(\frac{x}{C} \right)^2 \right)^3 \right)$ $\frac{C^2}{6}$	$x \left(1 - \left(\frac{x}{C} \right)^2 \right)^2$ 0	$ x \leq C$ $ x > C$
Welsch	$\frac{D^2}{2} \left(1 - \exp \left\{ - \left(\frac{x}{D} \right)^2 \right\} \right)$	$x \exp \left\{ - \left(\frac{x}{D} \right)^2 \right\}$	$x \in \mathcal{R}_1$

Tabulka 4.1: Obvyklé volby váhových funkcí ρ používaných v M-odhadech

c je parametr, který omezuje vliv velkých reziduí. Lze zvolit i jiné váhové funkce. Například nějakou z Tabulky 4.1., kterou jsem převzal z [1]. Všimněme si, že pokud A, B, C a $D \rightarrow \infty$, dostaneme ve všech čtyřech případech $\psi(x) = x$, kde $\psi(x) = \frac{d}{dx} \rho(x)$, tedy odhad klasickou metodou nejmenších čtverců.

Pro simultánní odhad parametru měřítka σ Huber v [4] navrhuje využití rovnice (*Proposal 2*)

$$0 = \sum_{i=1}^n \chi_c \left(\frac{\epsilon(i, \hat{\theta})}{\hat{\sigma}} \right), \quad (4.3)$$

kde

$$\chi_c(u) = \psi_c^2(u) - b$$

a

$$\psi_c(u) = \frac{d}{du} \rho_c(u) = \begin{cases} u, & |u| \leq c, \\ c \text{sign}(u), & |u| > c; \end{cases}$$

$$b = E\{\psi_c^2(z)\}, \quad z \sim N(0, 1).$$

Taková volba korekčního parametru b zajišťuje, že pokud u (reziduum normalizované parametrem měřítka) je $N(0, 1)$, pak je odhad měřítka konzistentní, tj. jestliže roste rozsah výběru, pak je malá pravděpodobnost velké chyby v odhadu parametru. Lze také psát $\psi_c(u) = \max(\min(c, u), -c)$.

4.2 GM-odhad založený na Kraskerově a Welschově návrhu

Jelikož je influenční funkce M-odhadu (1.2) neohraničená vzhledem k pozorováním $\mathbf{x}(i)$, M-odhad je citlivý k případným vlivným bodům v regresní matici \mathbf{X} , tj. bodům které mají významný vliv na hodnotu odhadu. Vlivné body lze určit například z diagonálních prvků projekční matice. Tuto skutečnost nelze ovlivnit vhodnou volbou ψ , jelikož ta nám umožňuje eliminovat odlehlá pozorování pouze z hlediska pozorování $y(i)$. Je proto nutné zmínit statistiku robustnosti označovanou jako *bod selhání*. Ten značí část „špatných“ pozorování, která je odhad schopen „vstřebat“, aniž by byl příliš ovlivněn. Například výběrový průměr, pokud se soustředíme na odhad polohy rozdělení, má bod selhání 0, protože i jedno špatné pozorování může libovolně ovlivnit výběrový průměr. Na druhou stranu třeba medián má bod selhání 50 procent.

Právě z důvodu vytvoření odhadů s nenulovým bodem selhání byly odvozeny *GM odhady*, kde je definice M-odhadu doplněna vhodnými vahami, které redukuje vliv vybočujících hodnot $\mathbf{x}(i)$.

Krasker a Welsch v [7] zavedli odhad $(\boldsymbol{\theta}_n, \sigma_n, \mathbf{A}_n)$, kde jednotlivá pozorování $\mathbf{x}(i)$ mají omezený vliv na odhad, a kde \mathbf{A}_n nese informaci o rozptýlenosti regresorů. Vezmeme v podstatě Huberův odhad pro lineární regresní model (viz např. [5]), ve kterém míru informace v regresním vektoru, $\sqrt{\mathbf{x}^T(i)\hat{\mathbf{A}}_n^{-1}\mathbf{x}(i)}$, uijeme ke stanovení horní hranice vlivu jednotlivých pozorování. Tento odhad lze užít pro lineární regresní model (4.1), kde data $(y(i), \mathbf{x}(i))$ jsou nezávislá a inovace jsou normální veličiny s nulovou střední hodnotou. Použijeme následující off-line odhad

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \psi_c \left(\frac{\sqrt{\mathbf{x}^T(i)\hat{\mathbf{A}}_n^{-1}\mathbf{x}(i)} \epsilon(i, \hat{\boldsymbol{\theta}})}{\hat{\sigma}} \right) \frac{\mathbf{x}(i)}{\sqrt{\mathbf{x}^T(i)\hat{\mathbf{A}}_n^{-1}\mathbf{x}(i)}} \hat{\sigma} \\ 0 &= \frac{1}{n} \sum_{i=1}^n \chi_c \left(\frac{\epsilon(i, \hat{\boldsymbol{\theta}})}{\hat{\sigma}} \right) \\ \hat{\mathbf{A}}_n &= \frac{1}{n} \sum_{i=1}^n g_1 \left(\frac{a}{\sqrt{\mathbf{x}^T(i)\hat{\mathbf{A}}_n^{-1}\mathbf{x}(i)}} \right) \mathbf{x}(i)\mathbf{x}^T(i), \end{aligned}$$

kde

$$g_1(u) = E_z\{\min(z^2, u^2)\} = \min\left\{1, \frac{a}{\left|\frac{\epsilon(i, \hat{\boldsymbol{\theta}})}{\hat{\sigma}}\right| \sqrt{\mathbf{x}^T(i) \hat{\mathbf{A}}_n^{-1} \mathbf{x}(i)}}}\right\}^2 \left(\frac{\epsilon(i, \hat{\boldsymbol{\theta}})}{\hat{\sigma}}\right)^2$$

za předpokladu, že z je náhodná veličina s rozdělením $N(0, 1)$ nezávislá na $\mathbf{x}(i)$. $\hat{\sigma}$ je odhad parametru σ , např. pomocí (4.3). Místo toho, abychom měli pouze jeden ladící parametr pro všechny tři části odhadu, použijeme parametr c v části, kde pracujeme s chybou modelu, a jiný parametr a pro odhad kovariance. V druhé rovnici jsme se rozhodli k odhadu parametru měřítka pomocí Huberova návrhu (4.3). Poslední rovnice definuje robustní odhad kovariance regresorů pomocí vah použitých na vybočující hodnoty regresorů prostřednictvím funkce $g_1(u)$.

Krasker a Welsch v [7] rozebírají vhodnou volbu omezujícího parametru a vzhledem k asymptotické vydatnosti a porovnávají s odhadem MNČ pro normálně rozdělené chyby (relativní vydatnost). Abychom získali relativní vydatnost např. 0.95 nebo 0.99, navrhují volbu parametru a jako 1.596 nebo 2.093 krát $\sqrt{\dim \mathbf{x}(i)}$.

Tyto rovnice nelze přímo převést na situaci závislých posloupností pozorování a rekurzivního odhadu.

Kapitola 5

Rekurzivní a adaptivní odhady

Algoritmy pro rekurzivní odhady parametrů, kterými se budeme dále zabývat, vyžadují, aby model, ve kterém odhadujeme hodnoty parametrů, byl v těchto parametrech lineární. Budeme se tedy zabývat modelem typu

$$y(t) = \mathbf{x}^T(t)\boldsymbol{\theta} + e(t).$$

Navíc předpokládáme, že $\mathbf{x}(t)$ a $e(t)$ jsou vzájemně nezávislé.

Budeme porovnávat čtyři algoritmy pro rekurzivní odhady parametrů tohoto modelu. Uvedeme tradiční rekurzivní metodu nejmenších čtverců. Její modifikaci, kdy pozorování, pro která máme velké hodnoty reziduí, označíme jako odlehlá pozorování a při výpočtech s nimi nebudeme pracovat. Dále algoritmus založený na rekurzivní minimalizaci Huberovy kriteriální funkce společně s rekurzivním řešením Huberova návrhu na odhad měřítka. Čtvrtý algoritmus je založen na rekurzivním řešení odhadu navrženého Kraskerem a Welschem. U všech algoritmů budeme pracovat s faktorem pro diskontování starších pozorování, takže bude docházet k exponenciálnímu zapomínání informací z pozorování více odlehlých v čase.

5.1 Rekurzivní metoda nejmenších čtverců (RLS)

V off-line případě metody nejmenších čtverců získáme odhad parametru $\boldsymbol{\theta}$ jako argument minimalizující příslušnou kriteriální funkci, tedy

$$\hat{\boldsymbol{\theta}}(t) = \operatorname{argmin}_{\boldsymbol{\theta}} \frac{1}{2} \sum_{i=1}^t \lambda^{t-i} \epsilon^2(i, \boldsymbol{\theta}).$$

Typicky volíme parametr pro zapomínání $\lambda \leq 1$ a $\epsilon(i, \boldsymbol{\theta}) = y(i) - \mathbf{x}^T(i)\boldsymbol{\theta}$ je reziduum v čase i .

Odhad parametru lze nalézt analyticky jako řešení rovnice

$$\begin{aligned} 0 &= \frac{d}{d\boldsymbol{\theta}} \frac{1}{2} \sum_{i=1}^t \lambda^{t-i} (y(i) - \mathbf{x}^T(i)\boldsymbol{\theta})^2_{|\hat{\boldsymbol{\theta}}(t)} \\ &= \sum_{i=1}^t \lambda^{t-i} \mathbf{x}(i)\mathbf{x}^T(i)\hat{\boldsymbol{\theta}}(t) - \sum_{i=1}^t \lambda^{t-i} \mathbf{x}(i)y(i) \end{aligned}$$

Za předpokladu, že existuje inverze k $\sum_{i=1}^t \lambda^{t-i} \mathbf{x}(i)\mathbf{x}^T(i)$, a díky tomu, že kriteriální funkce je v proměnné $\boldsymbol{\theta}$ konvexní, vyjádříme off-line odhad MNČ

$$\hat{\boldsymbol{\theta}}(t) = \left[\sum_{i=1}^t \lambda^{t-i} \mathbf{x}(i)\mathbf{x}^T(i) \right]^{-1} \left[\sum_{i=1}^t \lambda^{t-i} \mathbf{x}(i)y(i) \right]. \quad (5.1)$$

Toto vyjádření odhadu parametru $\boldsymbol{\theta}$ lze přepsat do rekurzivní podoby. Nejprve označíme

$$\bar{\mathbf{R}}(t) = \sum_{i=1}^t \lambda^{t-i} \mathbf{x}(i)\mathbf{x}^T(i). \quad (5.2)$$

Poté z (5.1) dostaneme, že

$$\sum_{i=1}^{t-1} \lambda^{t-i} \mathbf{x}(i)y(i) = \bar{\mathbf{R}}(t-1)\hat{\boldsymbol{\theta}}(t-1)$$

a z vyjádření (5.2) vyplývá

$$\bar{\mathbf{R}}(t) = \lambda \bar{\mathbf{R}}(t-1) + \mathbf{x}(t)\mathbf{x}^T(t). \quad (5.3)$$

Odtud máme

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t) &= \bar{\mathbf{R}}^{-1}(t) \left[\sum_{i=1}^{t-1} \lambda^{t-i} \mathbf{x}(i)y(i) + \mathbf{x}(t)y(t) \right] \\ &= \bar{\mathbf{R}}^{-1}(t) \left[\lambda \sum_{i=1}^{t-1} \lambda^{t-1-i} \mathbf{x}(i)y(i) + \mathbf{x}(t)y(t) \right] \\ &= \bar{\mathbf{R}}^{-1}(t) \left[\lambda \bar{\mathbf{R}}(t-1)\hat{\boldsymbol{\theta}}(t-1) + \mathbf{x}(t)y(t) \right] \\ &= \bar{\mathbf{R}}^{-1}(t) \left\{ \bar{\mathbf{R}}(t)\hat{\boldsymbol{\theta}}(t-1) + \mathbf{x}(t) \left[y(t) - \mathbf{x}^T(t)\hat{\boldsymbol{\theta}}(t-1) \right] \right\} \\ &= \hat{\boldsymbol{\theta}}(t-1) + \bar{\mathbf{R}}^{-1}(t)\mathbf{x}(t) \left[y(t) - \mathbf{x}^T(t)\hat{\boldsymbol{\theta}}(t-1) \right]. \end{aligned}$$

Nyní jsme již získali rekurzivní algoritmus, avšak toto vyjádření není z početního hlediska nejšťastnější, jelikož je potřeba v každém kroku invertovat matici. Proto zavedeme označení

$$\mathbf{P}_\lambda(t) = \overline{\mathbf{R}}^{-1}(t),$$

a použijeme následující tvrzení o inverzní matici.

Tvrzení 5.1. (*O inverzní matici*) *Nechť \mathbf{A} , \mathbf{B} , \mathbf{C} a \mathbf{D} jsou matice slučitelné dimenze takové, že součin \mathbf{BCD} a součet $\mathbf{A} + \mathbf{BCD}$ existují. Potom $[\mathbf{A} + \mathbf{BCD}]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}[\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1}]^{-1}\mathbf{DA}^{-1}$.*

Důkaz. Vynásobením rovnosti zprava výrazem $\mathbf{A} + \mathbf{BCD}$ dostaneme

$$\begin{aligned} PS &= \mathbf{I} + \mathbf{A}^{-1}\mathbf{BCD} - \mathbf{A}^{-1}\mathbf{B}[\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1}]^{-1}\mathbf{D} \\ &\quad - \mathbf{A}^{-1}\mathbf{B}[\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1}]^{-1}\mathbf{DA}^{-1}\mathbf{BCD} \\ &= \mathbf{I} + \mathbf{A}^{-1}\mathbf{B}[\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1}]^{-1} \\ &\quad \times \{[\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1}]\mathbf{CD} - \mathbf{D} - \mathbf{DA}^{-1}\mathbf{BCD}\} \\ &= \mathbf{I} + \mathbf{A}^{-1}\mathbf{B}[\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1}]^{-1} \times 0 \\ &= \mathbf{I} \end{aligned}$$

Čímž je tvrzení dokázáno. □

Použitím předchozího tvrzení na (5.3) s

$$\mathbf{A} = \lambda\mathbf{P}_\lambda^{-1}(t-1), \quad \mathbf{B} = \mathbf{x}(t), \quad \mathbf{C} = 1, \quad \mathbf{D} = \mathbf{x}^T(t)$$

dostáváme

$$\begin{aligned} \mathbf{P}_\lambda(t) &= [\lambda\mathbf{P}_\lambda^{-1}(t-1) + \mathbf{x}(t)\mathbf{x}^T(t)]^{-1} \\ &= \frac{1}{\lambda}\mathbf{P}_\lambda(t-1) - \frac{1}{\lambda}\mathbf{P}_\lambda(t-1)\mathbf{x}(t) \left[\mathbf{x}^T(t)\frac{1}{\lambda}\mathbf{P}_\lambda(t-1)\mathbf{x}(t) + 1 \right]^{-1} \\ &\quad \times \mathbf{x}^T(t)\mathbf{P}_\lambda(t-1)\frac{1}{\lambda} \\ &= \frac{1}{\lambda} \left[\mathbf{P}_\lambda(t-1) - \frac{\mathbf{P}_\lambda(t-1)\mathbf{x}(t)\mathbf{x}^T(t)\mathbf{P}_\lambda(t-1)}{\lambda + \mathbf{x}^T(t)\mathbf{P}_\lambda(t-1)\mathbf{x}(t)} \right]. \end{aligned}$$

Dostáváme tedy algoritmus

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{P}_\lambda(t)\mathbf{x}(t)\left[y(t) - \mathbf{x}^T(t)\hat{\boldsymbol{\theta}}(t-1)\right],$$

$$\mathbf{P}_\lambda(t) = \frac{1}{\lambda} \left[\mathbf{P}_\lambda(t-1) - \frac{\mathbf{P}_\lambda(t-1)\mathbf{x}(t)\mathbf{x}^T(t)\mathbf{P}_\lambda(t-1)}{\lambda + \mathbf{x}^T(t)\mathbf{P}_\lambda(t-1)\mathbf{x}(t)} \right].$$

Matice $\mathbf{P}_\lambda(t)$ obsahuje informace o kovariancích regresorů pozorovaných před časem t . Inverzní matici k $\mathbf{P}_\lambda(t)$ získáme následovně:

$$\begin{aligned} \mathbf{P}_\lambda^{-1}(t) &= \overline{\mathbf{R}}(t) = \lambda\mathbf{P}_\lambda^{-1}(t-1) + \mathbf{x}(t)\mathbf{x}^T(t) = \dots = \\ &= \lambda^t\mathbf{P}_\lambda^{-1}(0) + \sum_{i=1}^t \lambda^{t-i}\mathbf{x}(i)\mathbf{x}^T(i), \end{aligned}$$

kde $\mathbf{P}_\lambda(0)$ je počáteční hodnota.

5.2 Modifikovaný RLS algoritmus

Jedná se o modifikaci RLS algoritmu, kdy s velkými chybami predikce, klasifikovanými jako odlehlá pozorování, tedy s příslušnými pozorováními, pracujeme jako s chybějícími. Každé reziduum pro odhad parametru srovnáme s rekurzivní hodnotou odhadu měřítka. Pokud reziduum přesahuje c -násobek odhadu měřítka, pak neupravíme odhad parametrů, odhad matice \mathbf{P}_λ ani residuální odhad rozptylu. Algoritmus tedy vypadá následovně:

$$k_\sigma(t) = \max\left(\frac{1}{t}, 1 - \lambda\right),$$

$$\hat{\sigma}^2(t) = \hat{\sigma}^2(t-1) + k_\sigma(t)(d_c\epsilon^2(t, \hat{\boldsymbol{\theta}}(t-1)) - \hat{\sigma}^2(t-1))I_{\{|\epsilon(t, \hat{\boldsymbol{\theta}}(t-1))| < c\hat{\sigma}(t-1)\}},$$

$$\mathbf{P}_\lambda(t) = \frac{\mathbf{P}_\lambda(t-1)}{\lambda} - \frac{1}{\lambda} \frac{\mathbf{P}_\lambda(t-1)\mathbf{x}(t)\mathbf{x}^T(t)\mathbf{P}_\lambda(t-1)}{\lambda + \mathbf{x}^T(t)\mathbf{P}_\lambda(t-1)\mathbf{x}(t)} I_{\{|\epsilon(t, \hat{\boldsymbol{\theta}}(t-1))| < c\hat{\sigma}(t-1)\}},$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{P}_\lambda(t)\mathbf{x}(t)\left[y(t) - \mathbf{x}^T(t)\hat{\boldsymbol{\theta}}(t-1)\right] I_{\{|\epsilon(t, \hat{\boldsymbol{\theta}}(t-1))| < c\hat{\sigma}(t-1)\}}.$$

Přírůstek $k_\sigma(t)$ zavádíme, aby umožnil sledování pomalu se měnícího odhadu měřítka. Podobně faktor měřítka d_c je použit k získání konzistentního odhadu měřítka v případě, že jsou rezidua náhodné proměnné s rozdělením $N(0, \sigma^2)$. Toho dosáhneme, pokud

$$d_c^{-1} = E\{\phi_c^2(z)\}, \quad z \sim N(0, 1),$$

$$\phi_c(u) = \begin{cases} u, & |u| \leq c, \\ 0, & |u| > c. \end{cases}$$

Tato metoda má zcela jasnou nevýhodu. Pokud jsou odhady příliš daleko od skutečných hodnot parametrů, pak budou chyby v modelu vždy příliš velké a ani parametry, ani matice \mathbf{P}_λ nebudou upraveny.

5.3 Rekurzivní minimalizace Huberovy funkce (RHU)

Jedná se o algoritmus založený na rekurzivní minimalizaci Huberovy kritériální funkce (4.2) v kombinaci s rekurzivním řešením Huberova návrhu pro odhad měřítka (4.3).

Na RLS lze nahlížet buď jako na rekurzivní formulaci off-line odhadu nebo jako na rekurzivní řešení minimalizace kritériální funkce MNČ. Oba přístupy jsou ekvivalentní. Pokud bychom však uvažovali nekonvexní kritériální funkce, pak mohou oba přístupy vést k rozdílným algoritmům v závislosti na povaze aproximací, které provedeme. Při rekurzivní minimalizaci Huberovy funkce (pracujeme navíc s parametrem exponenciálního zapomínání λ)

$$V_{Hu}^\lambda(t, \boldsymbol{\theta}) = \sum_{i=1}^t \lambda^{t-i} \rho_c \left(\frac{\epsilon(i, \boldsymbol{\theta})}{\sigma} \right) \sigma^2,$$

kde

$$\rho_c(u) = \begin{cases} \frac{1}{2}u^2, & |u| \leq c, \\ c|u| - \frac{1}{2}c^2, & |u| > c, \end{cases}$$

se snažíme minimalizovat $V_{Hu}^\lambda(t, \boldsymbol{\theta})$ za použití aproximací, které jsou ospravedlnitelné, pokud se odhady parametrů nacházejí blízko jejich skutečné hodnotě. Tyto aproximace a změnu kriteriální funkce mezi po sobě jdoucími kroky použijeme k získání vztahu mezi odhady parametrů po sobě jdoucích kroků.

5.3.1 Odvození RHU

Předpokládáme, že parametr měřítka σ je známý (o jeho rekurzivním odhadu budeme mluvit v kapitole 5.5). Nechť $\hat{\boldsymbol{\theta}}(t-1)$ je odhad v čase $t-1$, pak $\hat{\boldsymbol{\theta}}(t)$ by měl být odhad v čase t získaný použitím algoritmu na příslušná pozorování. Odhad $\hat{\boldsymbol{\theta}}(t)$ by měl minimalizovat kriteriální funkci $V_{Hu}^\lambda(t, \boldsymbol{\theta})$. Z Taylorova rozvoje kriteriální funkce v bodě $\hat{\boldsymbol{\theta}}(t-1)$ dostaneme

$$\begin{aligned} V_{Hu}^\lambda(t, \boldsymbol{\theta}) &= V_{Hu}^\lambda(t, \hat{\boldsymbol{\theta}}(t-1)) + \left[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(t-1) \right]^T \nabla_{\boldsymbol{\theta}} V_{Hu}^\lambda(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} \\ &\quad + \frac{1}{2} \left[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(t-1) \right]^T \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{Hu}^\lambda(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} \left[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(t-1) \right] \\ &\quad + o\left(\left| \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(t-1) \right|^2 \right), \end{aligned}$$

kde $\nabla_{\boldsymbol{\theta}}$ značí gradient vzhledem k $\boldsymbol{\theta}$ a $o(x)$ je funkce pro kterou platí $\frac{o(x)}{|x|} \rightarrow 0$ pro $|x| \rightarrow 0$. Vezmeme derivaci výše uvedeného Taylorova rozvoje vzhledem k $\boldsymbol{\theta}$. Předpokládáme, že gradient kriteriální funkce je roven 0 pro rekurzivně získaný odhad parametru, tedy

$$\forall t : \nabla_{\boldsymbol{\theta}} V_{Hu}^\lambda(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t)} = 0, \quad (5.4)$$

Získáme následující rovnici pro vylepšení odhadu parametru

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) - \left[\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{Hu}^\lambda(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} \right]^{-1} \nabla_{\boldsymbol{\theta}} V_{Hu}^\lambda(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} \\ &\quad + o\left(\left| \hat{\boldsymbol{\theta}}(t) - \hat{\boldsymbol{\theta}}(t-1) \right| \right). \end{aligned} \quad (5.5)$$

Kriteriální funkci $V_{Hu}^\lambda(t, \boldsymbol{\theta})$ lze zapsat rekurzivně ve tvaru

$$V_{Hu}^\lambda(t, \boldsymbol{\theta}) = \lambda V_{Hu}^\lambda(t-1, \boldsymbol{\theta}) + \rho_c \left(\frac{\epsilon(t, \boldsymbol{\theta})}{\sigma} \right)^2. \quad (5.6)$$

Zderivováním (5.6) podle $\boldsymbol{\theta}$ a za použití předpokladu (5.4), pokud získanou rovnicí vyjádříme pro $\hat{\boldsymbol{\theta}}(t-1)$, dostáváme

$$\nabla_{\boldsymbol{\theta}} V_{Hu}^{\lambda}(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} = -\mathbf{x}(t)\psi_c \left(\frac{\epsilon(t, \hat{\boldsymbol{\theta}}(t-1))}{\sigma} \right) \sigma. \quad (5.7)$$

Dále vyjádříme druhou derivaci rekurzivního vzorce (5.6) v bodě $\hat{\boldsymbol{\theta}}(t-1)$ a dostaneme

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{Hu}^{\lambda}(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} &= \lambda \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{Hu}^{\lambda}(t-1, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} \\ &\quad + \mathbf{x}(t)\mathbf{x}^T(t)\psi'_c \left(\frac{\epsilon(t, \hat{\boldsymbol{\theta}}(t-1))}{\sigma} \right), \end{aligned} \quad (5.8)$$

kde $\psi'_c(u) = \frac{d}{du}\psi_c(u)$.

Pokud předpokládáme, že $\hat{\boldsymbol{\theta}}(t)$ je blízko $\hat{\boldsymbol{\theta}}(t-1)$, pak lze ospravedlnit následující dvě aproximace:

- Výraz $o(|\hat{\boldsymbol{\theta}}(t) - \hat{\boldsymbol{\theta}}(t-1)|)$ v (5.5) lze zanedbat.
- $\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{Hu}^{\lambda}(t-1, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)}$ v (5.8) můžeme nahradit výrazem $\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{Hu}^{\lambda}(t-1, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-2)}$.

Nakonec označíme inverzi k druhé derivaci kritériální funkce jako

$$\mathbf{P}_{\lambda}(t) \equiv \left[\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{Hu}^{\lambda}(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} \right]^{-1},$$

a použitím Tvzení 5.1, kde dosadíme

$$\mathbf{A} = \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{Hu}^{\lambda}(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-2)}, \quad \mathbf{B} = \mathbf{x}(t), \quad \mathbf{C} = \frac{1}{\lambda}, \quad \mathbf{D} = \mathbf{x}^T(t),$$

dostáváme následující algoritmus:

$$\mathbf{P}_{\lambda}(t) = \frac{\mathbf{P}_{\lambda}(t-1)}{\lambda} - \frac{1}{\lambda} \psi'_c \left(\frac{\epsilon(t, \hat{\boldsymbol{\theta}}(t-1))}{\sigma} \right) \frac{\mathbf{P}_{\lambda}(t-1)\mathbf{x}(t)\mathbf{x}^T(t)\mathbf{P}_{\lambda}(t-1)}{\lambda + \mathbf{x}^T(t)\mathbf{P}_{\lambda}(t-1)\mathbf{x}(t)}, \quad (5.9)$$

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \mathbf{P}_{\lambda}(t)\mathbf{x}(t)\psi_c \left(\frac{\epsilon(t, \hat{\boldsymbol{\theta}}(t-1))}{\sigma} \right) \sigma.$$

Výraz (5.9) jsme zjednodušili tím, že $\psi'_c(u)$ je buď jedna nebo nula pro Huberovu funkci $\psi_c(u)$ a zanedbáme fakt, že v některých bodech derivace neexistuje.

5.4 Rekurzivní algoritmus založený na Kraskerově a Welschově odhadu (RKW)

Tento algoritmus je založený na rekurzivním řešení off-line odhadu navrženého Kraskerem a Welschem, který jsme uvedli v kapitole 4.2. Rekurzivní algoritmus získáme obdobně jakou u RHU, a to rekurzivní minimalizací kriteriální funkce

$$V_{KW}^\lambda(t, \boldsymbol{\theta}) = \sum_{i=1}^t \lambda^{t-i} \frac{\sigma^2}{\mathbf{x}^T(i) \hat{\mathbf{A}}_t^{-1} \mathbf{x}(i)} \rho_c \left(\sqrt{\mathbf{x}^T(i) \hat{\mathbf{A}}_t^{-1} \mathbf{x}(i)} \frac{\epsilon(i, \boldsymbol{\theta})}{\sigma} \right).$$

Tento vzorec je důležité doplnit rekurzivním odhadem kovarianční matice regresorů.

5.4.1 Odvození RKW

Algoritmus získáme jako rekurzivní a aproximativní minimalizaci kriteriální funkce $V_{KW}^\lambda(t, \boldsymbol{\theta})$ společně s rekurzivním a aproximačním výrazem pro update robustního odhadu kovarianční matice $\hat{\mathbf{A}}_t$. Stejně jako u RHU je odhad měřítka oddělen od odhadu parametrů modelu. K odvození používáme stejné procedury jako u RHU. Nejdříve zapíšeme Taylorův rozvoj (druhého řádu) kriteriální funkce s využitím předpokladu, že derivace kriteriální funkce v čase t je nula, pokud je vyjádřena pro $\hat{\boldsymbol{\theta}}(t)$. Zavedeme modifikovanou kriteriální funkci

$$V_{KW}^{\lambda,r}(t, \boldsymbol{\theta}) = \sum_{i=1}^t \lambda^{t-i} \frac{\sigma^2}{\mathbf{x}^T(i) \hat{\mathbf{A}}^{-1}(i) \mathbf{x}(i)} \rho_c \left(\sqrt{\mathbf{x}^T(i) \hat{\mathbf{A}}^{-1}(i) \mathbf{x}(i)} \frac{\epsilon(i, \boldsymbol{\theta})}{\sigma} \right), \quad (5.10)$$

kde je odhad kovarianční matice, tedy $\hat{\mathbf{A}}^{-1}(t) (= \hat{\mathbf{A}}_t^{-1})$, v $V_{KW}^\lambda(t, \boldsymbol{\theta})$ nahrazen odhadem kovariance v čase, pro který je dostupný. Taylorův rozvoj druhého řádu použijeme na modifikovanou kriteriální funkci, což vede k následující rovnici pro update odhadu parametru

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) - \left[\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{KW}^{\lambda,r}(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} \right]^{-1} \nabla_{\boldsymbol{\theta}} V_{KW}^{\lambda,r}(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} \\ &\quad + o(|\hat{\boldsymbol{\theta}}(t) - \hat{\boldsymbol{\theta}}(t-1)|). \end{aligned} \quad (5.11)$$

Modifikovanou kriteriální funkci (5.10) lze zapsat rekurzivně ve tvaru

$$V_{KW}^{\lambda,r}(t, \boldsymbol{\theta}) = \lambda V_{KW}^{\lambda,r}(t-1, \boldsymbol{\theta}) + \frac{\sigma^2}{\mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t) \mathbf{x}(t)} \rho_c \left(\sqrt{\mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t) \mathbf{x}(t)} \frac{\epsilon(t, \boldsymbol{\theta})}{\sigma} \right). \quad (5.12)$$

Pokud zderivujeme obě strany rovnosti (5.12) podle $\boldsymbol{\theta}$ a pro odhad parametru $\hat{\boldsymbol{\theta}}(t-1)$ využijeme (5.4) a předpokládáme, že platí i pro kriteriální funkci $V_{KW}^{\lambda,r}(t, \boldsymbol{\theta})$, pak dostaneme

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} V_{KW}^{\lambda,r}(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} &= -\mathbf{x}(t) \frac{\sigma}{\sqrt{\mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t) \mathbf{x}(t)}} \\ &\quad \times \psi_c \left(\sqrt{\mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t) \mathbf{x}(t)} \frac{\epsilon(t, \hat{\boldsymbol{\theta}}(t-1))}{\sigma} \right). \end{aligned} \quad (5.13)$$

Derivace druhého řádu kriteriální funkce, tedy jejího rekurzivního zápisu v (5.12), vyjádřená pro odhad $\hat{\boldsymbol{\theta}}(t-1)$ dává

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{KW}^{\lambda,r}(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} &= \lambda \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{KW}^{\lambda,r}(t-1, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} \\ &\quad + \mathbf{x}(t) \mathbf{x}^T(t) \psi_c' \left(\sqrt{\mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t) \mathbf{x}(t)} \frac{\epsilon(t, \hat{\boldsymbol{\theta}}(t-1))}{\sigma} \right). \end{aligned} \quad (5.14)$$

Stejně jako u odvození RHU, i zde provedeme následující aproximace:

- Výraz $o(|\hat{\boldsymbol{\theta}}(t) - \hat{\boldsymbol{\theta}}(t-1)|)$ v (5.11) zanedbáme.
- $\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{KW}^{\lambda,r}(t-1, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)}$ nahradíme výrazem $\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{KW}^{\lambda,r}(t-1, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-2)}$.

Provedeme-li výše uvedené aproximace a použijeme Tvrzení 5.1 s

$$\mathbf{A} = \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{KW}^{\lambda,r}(t-1, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-2)}, \quad \mathbf{B} = \mathbf{x}(t), \quad \mathbf{C} = \frac{1}{\lambda}, \quad \mathbf{D} = \mathbf{x}^T(t),$$

na (5.14) a zavedeme označení

$$\mathbf{P}_{\lambda}(t) \equiv \left[\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} V_{KW}^{\lambda,r}(t, \boldsymbol{\theta})|_{\hat{\boldsymbol{\theta}}(t-1)} \right]^{-1},$$

pak dostaneme výraz (5.17). Pokud dosadíme výraz (5.13) do (5.11), využijeme aproximace o zanedbání členu $o(x)$ a dosadíme $\mathbf{P}_\lambda(t)$, pak dostaneme finální výraz pro update odhadu parametru (5.18).

Abychom získali rekurzivní vzorec pro odhad kovarianční matice $\hat{\mathbf{A}}(t)$, musíme odhad v i -tém členu součtu nahradit tím, co získáme v čase $i - 1$. Potom dostaneme vyjádření

$$\hat{\mathbf{A}}(t) = \frac{1}{t} \sum_{i=1}^t g_1 \left(\frac{a}{\sqrt{\mathbf{x}^T(i) \hat{\mathbf{A}}^{-1}(i-1) \mathbf{x}(i)}} \right) \mathbf{x}(i) \mathbf{x}^T(i),$$

což lze rekurzivně zapsat jako

$$\hat{\mathbf{A}}(t) = \hat{\mathbf{A}}(t-1) + \frac{1}{t} \left(g_1 \left(\frac{a}{\sqrt{\mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t-1) \mathbf{x}(t)}} \right) \mathbf{x}(t) \mathbf{x}^T(t) - \hat{\mathbf{A}}(t-1) \right). \quad (5.15)$$

Použitím Tvzení 5.1 lze (5.15) převést na (5.16), čímž získáme kompletní odvození RKW algoritmu.

$$\begin{aligned} \hat{\mathbf{A}}^{-1}(t) &= \frac{t}{t-1} \hat{\mathbf{A}}^{-1}(t-1) - \frac{t}{t-1} g_1 \left(\frac{a}{\sqrt{\mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t-1) \mathbf{x}(t)}} \right) \\ &\quad \times \frac{\hat{\mathbf{A}}^{-1}(t-1) \mathbf{x}(t) \mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t-1)}{t + \mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t-1) \mathbf{x}(t) g_1 \left(\frac{a}{\sqrt{\mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t-1) \mathbf{x}(t)}} \right)}, \quad (5.16) \end{aligned}$$

$$\begin{aligned} \mathbf{P}_\lambda(t) &= \frac{\mathbf{P}_\lambda(t-1)}{\lambda} - \frac{1}{\lambda} \psi'_c \left(\sqrt{\mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t) \mathbf{x}(t)} \frac{\epsilon(t, \hat{\boldsymbol{\theta}}(t-1))}{\sigma} \right) \\ &\quad \times \frac{\mathbf{P}_\lambda(t-1) \mathbf{x}(t) \mathbf{x}^T(t) \mathbf{P}_\lambda(t-1)}{\lambda + \mathbf{x}^T(t) \mathbf{P}_\lambda(t-1) \mathbf{x}(t)}, \quad (5.17) \end{aligned}$$

$$\begin{aligned} \hat{\boldsymbol{\theta}}(t) &= \hat{\boldsymbol{\theta}}(t-1) + \mathbf{P}_\lambda(t) \mathbf{x}(t) \frac{\sigma}{\sqrt{\mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t) \mathbf{x}(t)}} \\ &\quad \times \psi_c \left(\sqrt{\mathbf{x}^T(t) \hat{\mathbf{A}}^{-1}(t) \mathbf{x}(t)} \frac{\epsilon(t, \hat{\boldsymbol{\theta}}(t-1))}{\sigma} \right). \quad (5.18) \end{aligned}$$

Poznamenejme, že (5.17) bylo zjednodušeno předpokladem, že $\psi'_c(u)$ je nula nebo jedna pro Huberovu funkci $\psi_c(u)$.

5.5 Rekurzivní odhad měřítka

Algoritmus pro rekurzivní odhad měřítka je založen na Huberově návrhu (4.3). Abychom se dostali k rekurzivnímu odhadu měřítka, je nutné zavést funkci

$$X_\lambda(t, \sigma) = \sum_{i=1}^t \lambda^{t-i} \chi_c \left(\frac{\epsilon(t, \boldsymbol{\theta})}{\sigma} \right),$$

kde předpokládáme, že známe množinu parametrů $\boldsymbol{\theta}$.

Funkci X_λ lze zapsat rekurzivně jako

$$X_\lambda(t, \sigma) = \lambda X_\lambda(t-1, \sigma) + \chi_c \left(\frac{\epsilon(t, \boldsymbol{\theta})}{\sigma} \right). \quad (5.19)$$

Pokud použijeme Taylorův rozvoj prvního řádu funkce X_λ v bodě $\hat{\sigma}(t-1)$, a předpokládáme, že odhad měřítka v čase t splňuje (4.3), tedy $X_\lambda(t, \hat{\sigma}(t)) = 0$, pak dostaneme

$$0 = X_\lambda(t, \hat{\sigma}(t-1)) + [\hat{\sigma}(t) - \hat{\sigma}(t-1)] \nabla_\sigma X_\lambda(t, \sigma)|_{\hat{\sigma}(t-1)} + o(|\hat{\sigma}(t) - \hat{\sigma}(t-1)|). \quad (5.20)$$

První výraz napravo v (5.20) lze vyjádřit z (5.19), pokud za σ dosadíme $\hat{\sigma}(t-1)$ a využijeme předpoklad $X_\lambda(t-1, \hat{\sigma}(t-1)) = 0$. Dostáváme

$$X_\lambda(t, \hat{\sigma}(t-1)) = \chi_c \left(\frac{\epsilon(t, \boldsymbol{\theta})}{\hat{\sigma}(t-1)} \right).$$

Vyjádřením derivace výrazu (5.19) v $\hat{\sigma}(t-1)$ získáme

$$\nabla_\sigma X_\lambda(t, \sigma)|_{\hat{\sigma}(t-1)} = \lambda \nabla_\sigma X_\lambda(t-1, \sigma)|_{\hat{\sigma}(t-1)} - 2 \frac{\epsilon^2(t, \boldsymbol{\theta})}{\hat{\sigma}^3(t-1)} I_{\{|\epsilon(t, \boldsymbol{\theta})| \leq c\hat{\sigma}(t-1)\}}.$$

Nyní provedeme následující aproximace:

- Výraz $o(|\hat{\sigma}(t) - \hat{\sigma}(t-1)|)$ v (5.20) zanedbáme.
- $\nabla_\sigma X_\lambda(t-1, \sigma)|_{\hat{\sigma}(t-1)}$ nahradíme výrazem $\nabla_\sigma X_\lambda(t-1, \sigma)|_{\hat{\sigma}(t-2)}$.

Označíme-li $h_\lambda(t) \equiv -\nabla_\sigma X_\lambda(t, \sigma)|_{\hat{\sigma}(t-1)}$, pak můžeme psát

$$h_\lambda(t) = \lambda h_\lambda(t-1) + 2 \frac{\epsilon^2(t, \boldsymbol{\theta})}{\hat{\sigma}^3(t-1)} I_{\{|\epsilon(t, \boldsymbol{\theta})| \leq c\hat{\sigma}(t-1)\}},$$

a

$$\hat{\sigma}(t) = \hat{\sigma}(t-1) + \frac{\chi_c \left(\frac{\epsilon(t, \theta)}{\hat{\sigma}(t-1)} \right)}{h_\lambda(t)}.$$

Vzhledem k symetrii funkce χ_c může odhad měřítka konvergovat k pozitivní hodnotě i jejímu zápornému „protikladu“. Abychom se vyhnuli zápornému odhadu měřítka, omezíme množinu přípustných hodnot na kladnou reálnou polopřímku. V praxi má toto omezení význam pouze v počáteční fázi použití tohoto algoritmu.

Kapitola 6

Aplikace algoritmů

V této kapitole použijeme uvedené algoritmy odhadu na nasimulovaná data, která budou buď zcela bez odlehlých pozorování, nebo budou obsahovat odlehlá pozorování způsobená inovacemi a nebo odlehlá pozorování zapříčiněná aditivními efekty. Příslušný autoregresní model a tři druhy šumů působících na data použijeme obdobně jako v [9]. Pro ilustraci problému a zejména z důvodu průhlednosti získaných výsledků jsme zvolili nejjednodušší možný model. Uvažuji tedy lineární regresní model

$$y(t) = x(t)\theta + e(t)$$

a naším cílem je odhadnout parametr θ co nejpřesněji, přičemž uvažuji různé druhy chyb vyskytujících se v datech, viz dále. K simulacím, výpočtům i tvorbě grafů použijeme statistický software R. Funkce nutné k výpočtům jsem naprogramoval a jsou k nahlédnutí na příloženém CD.

6.1 Simulace dat

Simulovaná data, se kterými budeme pracovat, budou všechna z následujícího čistě autoregresního modelu prvního řádu

$$z(t) = 0.8z(t-1) + e(t). \tag{6.1}$$

Každá posloupnost dat se skládá z 3005 pozorování a je simulována buď bez odlehlých pozorování, nebo s odlehlými pozorováními způsobenými inovacemi a nebo s odlehlými pozorováními zapříčiněnými aditivními efekty. Pokud simulujeme data s odlehlými pozorováními, pak se neobjeví v prvních

pěti pozorováních. Pracujeme s následujícími třemi druhy šumů působícími na náš AR(1) model:

- *Bez odlehlých pozorování:* Pokud předpokládáme, že pozorování neobsahují odlehlá pozorování, pak jsou inovace $\{e(t)\}$ nezávislé stejně rozdělené náhodné veličiny s rozdělením $N(0, 1)$.
- *Odlehlá pozorování způsobená inovacemi:* Pokud jsou pozorování $\{y(t)\}$ simulována s IO, pak jsou stále rovna hodnotám procesu $\{z(t)\}$, ale inovace $\{e(t)\}$ jsou z kontaminovaného normálního rozdělení. Inovace jsou nezávislé, ale s nějakou malou pravděpodobností p jsou z rozdělení s větším rozptylem. V naší simulaci pracujeme s rozdělením inovací, které lze schematicky zapsat následujícím způsobem:

$$e(t) \sim 0.95 \times N(0, 1) + 0.05 \times N(0, 6.25).$$

- *Odlehlá pozorování zapříčiněná aditivními efekty:* Pokud jsou pozorování kontaminována aditivními efekty, pak se mohou lišit od procesu kvůli aditivnímu šumu $\{v(t)\}$, tedy

$$y(t) = z(t) + v(t).$$

$\{z(t)\}$ simulujeme jako v situaci bez odlehlých pozorování a aditivní šum $\{v(t)\}$ je nula s pravděpodobností 0.95 a $N(0, 6.25)$ s pravděpodobností 0.05, tedy

$$v(t) \sim 0.95 \times 0 + 0.05 \times N(0, 6.25).$$

Na Obrázku 6.1 je příklad pozorování vytvořených pomocí AR(1) modelu (6.1) s třemi druhy šumů uvedenými výše. V obou případech s odlehlými pozorováními je vidět několik pozorování, která vybočují. Je také vidět, že oba typy odlehlých pozorování rozdílně působí na simulovaná data. IO ovlivňují samotný proces a způsobují tak trvalé „podráždění“, zatímco AO se projevují jako aditivní hroty v době jejich výskytu.

6.2 Aplikace algoritmů na nasimulovaná data

Pro každou posloupnost simulovaných dat použijeme čtyři algoritmy odhadu, a to RLS , $RHU(c)$, $RKW(c, a)$ a $RMO(c)$ s parametry c a a , které jsme

v práci specifikovali. Rekurzivní odhady pro všechny metody začneme použitím RLS algoritmu na prvních 5 pozorování. V čase $t = 6$ tedy inicializujeme studovaný algoritmus parametry $\hat{\theta}(5)_{RLS}$ a kovarianční maticí $P_\lambda(5)_{RLS}$. *RLS* je inicializováno kovarianční maticí, která je 100 násobkem jednotkové matice, a také robustní odhad kovariance v *RKW*, tedy $\hat{A}^{-1}(5) = 100I$.

Počáteční hodnota $\hat{\sigma}^2(5)$ je známý rozptyl simulace inovací bez odlehlých pozorování a počáteční hodnota $h_\lambda(5)$ v rekurzivní verzi Huberova návrhu (4.3) je rovna 1, tedy $h_\lambda(5) = 1$. Pro *RHU*(c), *RKW*(c, a) a *RMO*(c) použijeme $c = 2, 3$ a pro každou hodnotu c *RKW*(c, a) použijeme $a = 2, 3, 4$. Jelikož data simulujeme s konstantní hodnotou parametru $\theta = 0.8$, použijeme pro všechny algoritmy zapomínací faktor $\lambda = 1$.

Obrázek 6.2 ukazuje průběh odhadů parametru θ pro náš autoregresní model prvního řádu (6.1), a to na datech bez odlehlých pozorování. Základní posloupnost inovací jsme zvolili stejnou ve všech odhadech. Vidíme, že není skoro žádný rozdíl mezi RLS a *RKW*(2,3) použitými na simulovaná data. Oba algoritmy dobře konvergují k hledanému parametru. Podobný průběh mají i *RMO*(2) a *RHU*(2) algoritmy, které zde graficky neznázorníme.

Na Obrázku 6.3 je zobrazen průběh odhadů pro data s odlehlými pozorováními způsobenými inovacemi získaných RLS a *RKW*(2,3) algoritmy. Vidíme, že oba algoritmy konvergují k hledanému parametru opět celkem dobře. Konvergence je sice rychlejší, ale více rozkolísaná, což je způsobeno tím, že odlehlá pozorování ovlivňují celý proces. *RHU*(2) a *RMO*(2) mají opět podobný průběh.

Pokud se podíváme na průběh odhadů ovlivněných odlehlými pozorováními zapříčiněnými aditivními efekty na Obrázku 6.4 a 6.5, vidíme již daleko větší rozdíl, a to zejména pro RLS algoritmus, kde odlehlá pozorování způsobují posun odhadu. I u *RHU*(2) je vidět pomalejší konvergence k hledané hodnotě parametru než u *RKW*(2,3), který ukazuje poměrně hladký průběh odhadů a zlepšení vůči *RHU*(2). *RMO*(2) dokonce konverguje k hledanému parametru nejpřesněji, jelikož klasifikuje některá odlehlá pozorování. U tohoto algoritmu ovšem neustále hrozí možnost, že se odhady stanou tak špatnými, že rezidua modelu budou příliš velká vzhledem k odhadům rozptylu, což znamená, že odhad nevylepšíme.

6.3 Hodnocení algoritmů

K hodnocení algoritmů jsme spočetli 6 různých charakteristik, a to pro 1000 rozdílných simulací a vždy se všemi třemi druhy odlehlých pozorování. Všechny

ny algoritmy použijeme na všechny posloupnosti simulovaných dat. Pro každou posloupnost spočteme součet čtverců odchylek odhadnutých hodnot parametrů od skutečné hodnoty parametru pro pozorování č. 2006 až č. 3005

$$SSD = \sum_{i=2006}^{3005} (\hat{\theta}(i) - \theta)^2.$$

Budeme pracovat s jejich průměrem \overline{SSD} pro našich 1000 simulací. Spočítáme také průměrnou hodnotu $\hat{\theta}(t)$, standartní odchylku $s_{\hat{\theta}(t)}$ a 5, 50 a 95 procentní kvantily $\hat{\theta}(t)_{0.05}$, $\hat{\theta}(t)_{0.50}$ a $\hat{\theta}(t)_{0.95}$ rekurzivně získaných odhadů pro $t_1 = 2005$ a $t_2 = 3005$.

6.4 Diskuze výsledků

V Tabulce 6.1 jsou průměrné hodnoty charakteristik pro 1000 simulací dat bez odlehlých pozorování, viz výše. Je vidět, že horší hodnoty vychází pro nižší hodnoty parametrů a a c , což je rozumný výsledek, jelikož v datech nemáme žádné chybné informace.

V druhé Tabulce 6.2 je vidět, že s výskytem odlehlých pozorování způsobených inovacemi v datech se RHU, RKW a RMO dokonce zlepšily. Pro RLS algoritmus je výsledek prakticky stejný a dochází jen k malému zhoršení. To se shoduje se závěry Martina a Yohaie v [8], kteří ukázali, že asymptotický rozptyl odhadů metodou nejmenších čtverců AR a MA parametrů je nezávislý na rozdělení inovací. Avšak pro rozdělení inovací s těžkými chvosty mohou být odhady metodou nejmenších čtverců neeficientní ve srovnání s maximálně věrohodnými odhady, protože odhady MNČ příliš nezískají ze zvýšené přesnosti dosažitelné díky rozdělení inovací s těžkými chvosty. Pokud neznáme rozdělení, pak lze M-odhad použít, abychom využili zvýšené podráždění.

Třetí Tabulka 6.3 ukazuje zvětšení odchylek pro data s AO. K největšímu vychýlení, tedy ovlivnění odhadu odlehlými pozorováními, dochází u RLS. Všechny robustní varianty dosahují lepších výsledků. Pro velikost kontaminace 5% jsou výsledky pro RMO, RKW a RHU podobné. Teprve s rostoucí kontaminací (například 20%) je vidět, že nejlepších výsledků dosahuje RKW. U RMO dochází ke zhoršení odhadu pouze pomalu s růstem kontaminace, takže se postupně srovnává s ostatními odhady a pro velké hodnoty kontaminace se pak stává nejlepší. Stejný trend platí pro hodnoty c a a a postupně

dávají menší hodnoty lepší výsledky. Porovnáním výsledků pro různá odlehlá pozorování zjišťujeme, že rozumnou volbou je $(c, a) = (2, 3)$. Poznamenejme, že velká část \overline{SSD} vychází z vychýlení odhadů.

Algoritmus	\overline{SSD}	$\hat{\theta}(t_1)$	$s_{\hat{\theta}(t_1)}$	$\hat{\theta}(t_1)_{0.05}$	$\hat{\theta}(t_1)_{0.50}$	$\hat{\theta}(t_1)_{0.95}$
RLS	0.155	0.799	0.013	0.775	0.799	0.820
RHU(2)	0.158	0.799	0.013	0.775	0.799	0.820
RHU(3)	0.155	0.799	0.013	0.775	0.799	0.820
RKW(2,2)	0.176	0.799	0.014	0.773	0.799	0.822
RKW(2,3)	0.173	0.799	0.014	0.774	0.799	0.821
RKW(2,4)	0.173	0.799	0.014	0.774	0.799	0.821
RKW(3,2)	0.163	0.799	0.014	0.775	0.799	0.821
RKW(3,3)	0.162	0.799	0.013	0.775	0.799	0.821
RKW(3,4)	0.161	0.799	0.013	0.775	0.799	0.821
RMO(2)	0.223	0.799	0.016	0.771	0.800	0.824
RMO(3)	0.160	0.799	0.013	0.775	0.800	0.820

Algoritmus	$\hat{\theta}(t_2)$	$s_{\hat{\theta}(t_2)}$	$\hat{\theta}(t_2)_{0.05}$	$\hat{\theta}(t_2)_{0.50}$	$\hat{\theta}(t_2)_{0.95}$
RLS	0.799	0.011	0.780	0.800	0.817
RHU(2)	0.799	0.011	0.780	0.799	0.817
RHU(3)	0.799	0.011	0.780	0.800	0.817
RKW(2,2)	0.799	0.012	0.779	0.800	0.817
RKW(2,3)	0.799	0.011	0.779	0.799	0.817
RKW(2,4)	0.799	0.011	0.779	0.799	0.817
RKW(3,2)	0.799	0.011	0.779	0.800	0.817
RKW(3,3)	0.799	0.011	0.779	0.800	0.817
RKW(3,4)	0.799	0.011	0.780	0.800	0.817
RMO(2)	0.799	0.013	0.776	0.800	0.821
RMO(3)	0.799	0.011	0.780	0.800	0.817

Tabulka 6.1: Výsledky pro AR(1) model a data bez OP, kde $t_1 = 2005$ a $t_2 = 3005$

Algoritmus	\overline{SSD}	$\hat{\theta}(t_1)$	$s_{\hat{\theta}(t_1)}$	$\hat{\theta}(t_1)_{0.05}$	$\hat{\theta}(t_1)_{0.50}$	$\hat{\theta}(t_1)_{0.95}$
RLS	0.161	0.798	0.014	0.774	0.800	0.821
RHU(2)	0.146	0.799	0.013	0.776	0.800	0.819
RHU(3)	0.151	0.799	0.013	0.775	0.800	0.820
RKW(2,2)	0.162	0.799	0.014	0.775	0.799	0.821
RKW(2,3)	0.161	0.799	0.013	0.775	0.799	0.821
RKW(2,4)	0.160	0.799	0.013	0.775	0.799	0.820
RKW(3,2)	0.155	0.799	0.013	0.774	0.799	0.820
RKW(3,3)	0.154	0.799	0.013	0.775	0.799	0.820
RKW(3,4)	0.155	0.799	0.013	0.774	0.799	0.820
RMO(2)	0.198	0.799	0.015	0.772	0.799	0.823
RMO(3)	0.149	0.799	0.013	0.777	0.800	0.820

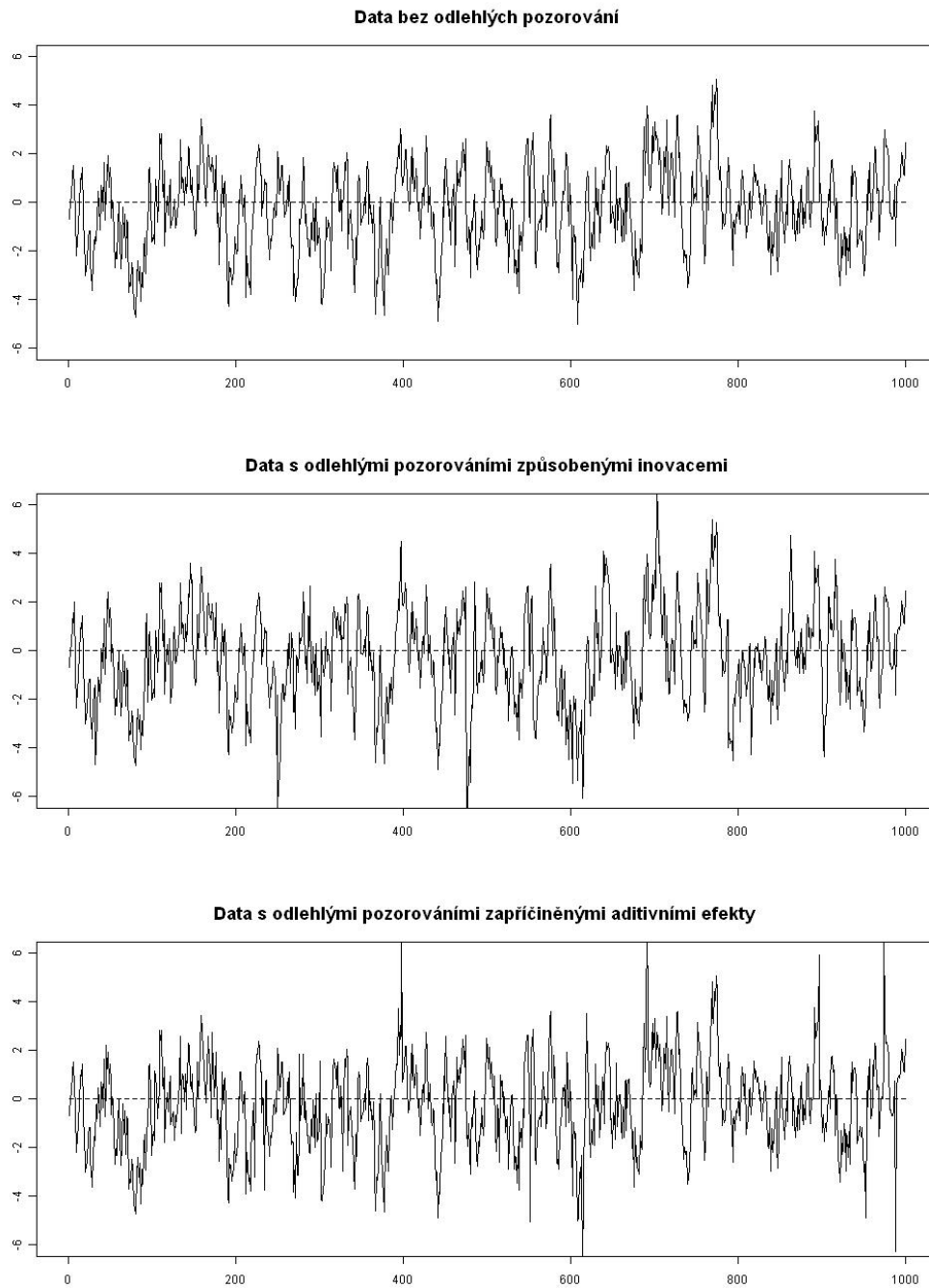
Algoritmus	$\hat{\theta}(t_2)$	$s_{\hat{\theta}(t_2)}$	$\hat{\theta}(t_2)_{0.05}$	$\hat{\theta}(t_2)_{0.50}$	$\hat{\theta}(t_2)_{0.95}$
RLS	0.799	0.011	0.780	0.799	0.816
RHU(2)	0.799	0.010	0.780	0.799	0.815
RHU(3)	0.799	0.011	0.780	0.799	0.815
RKW(2,2)	0.799	0.011	0.779	0.799	0.817
RKW(2,3)	0.799	0.011	0.779	0.799	0.817
RKW(2,4)	0.799	0.011	0.779	0.799	0.817
RKW(3,2)	0.799	0.011	0.780	0.799	0.817
RKW(3,3)	0.799	0.011	0.780	0.799	0.816
RKW(3,4)	0.799	0.011	0.780	0.799	0.816
RMO(2)	0.799	0.012	0.778	0.800	0.820
RMO(3)	0.799	0.011	0.780	0.799	0.816

Tabulka 6.2: Výsledky pro AR(1) model a data s IO, kde $t_1 = 2005$ a $t_2 = 3005$

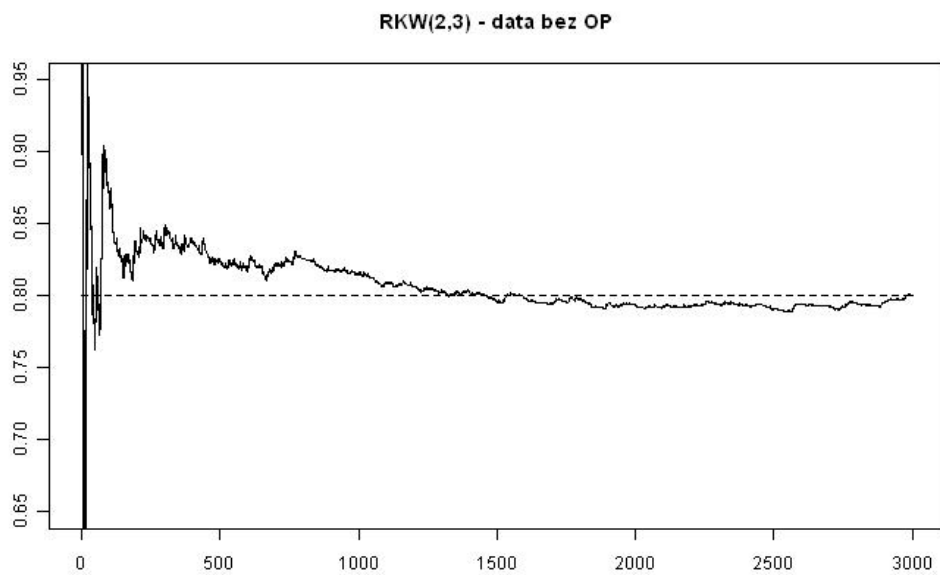
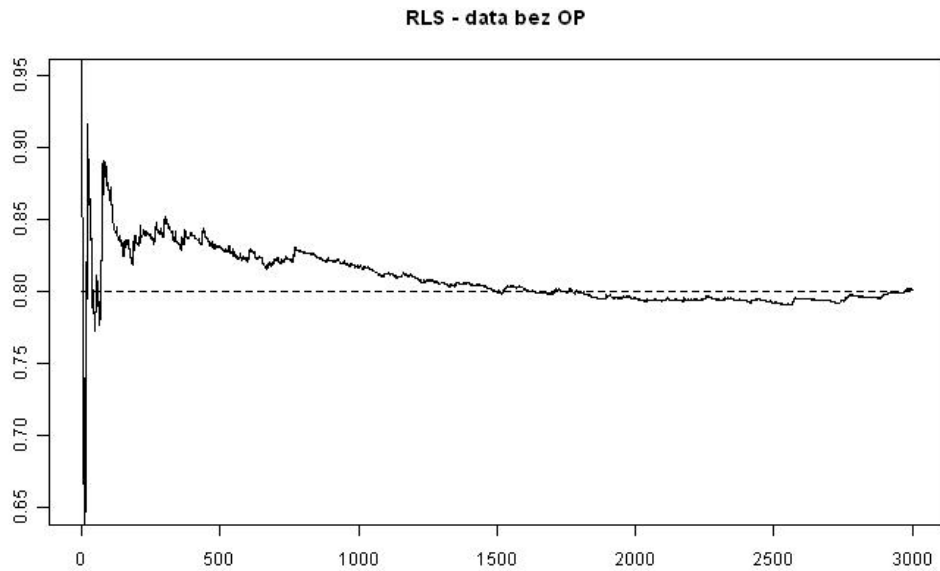
Algoritmus	\overline{SSD}	$\hat{\theta}(t_1)$	$s_{\hat{\theta}(t_1)}$	$\hat{\theta}(t_1)_{0.05}$	$\hat{\theta}(t_1)_{0.50}$	$\hat{\theta}(t_1)_{0.95}$
RLS	0.210	0.799	0.015	0.770	0.800	0.823
RHU(2)	0.187	0.799	0.015	0.774	0.800	0.822
RHU(3)	0.195	0.799	0.015	0.772	0.799	0.823
RKW(2,2)	0.199	0.799	0.015	0.772	0.799	0.824
RKW(2,3)	0.197	0.799	0.015	0.772	0.799	0.823
RKW(2,4)	0.197	0.799	0.015	0.772	0.799	0.823
RKW(3,2)	0.194	0.799	0.015	0.773	0.799	0.823
RKW(3,3)	0.194	0.799	0.015	0.773	0.799	0.823
RKW(3,4)	0.194	0.799	0.015	0.773	0.799	0.823
RMO(2)	0.246	0.799	0.017	0.769	0.800	0.827
RMO(3)	0.186	0.799	0.014	0.774	0.800	0.823

Algoritmus	$\hat{\theta}(t_2)$	$s_{\hat{\theta}(t_2)}$	$\hat{\theta}(t_2)_{0.05}$	$\hat{\theta}(t_2)_{0.50}$	$\hat{\theta}(t_2)_{0.95}$
RLS	0.799	0.012	0.778	0.800	0.818
RHU(2)	0.799	0.012	0.778	0.800	0.818
RHU(3)	0.799	0.012	0.778	0.800	0.818
RKW(2,2)	0.799	0.012	0.778	0.800	0.819
RKW(2,3)	0.799	0.012	0.778	0.800	0.819
RKW(2,4)	0.799	0.012	0.777	0.800	0.819
RKW(3,2)	0.799	0.012	0.778	0.800	0.819
RKW(3,3)	0.799	0.012	0.778	0.800	0.819
RKW(3,4)	0.799	0.012	0.778	0.800	0.819
RMO(2)	0.799	0.014	0.774	0.800	0.822
RMO(3)	0.799	0.012	0.778	0.799	0.819

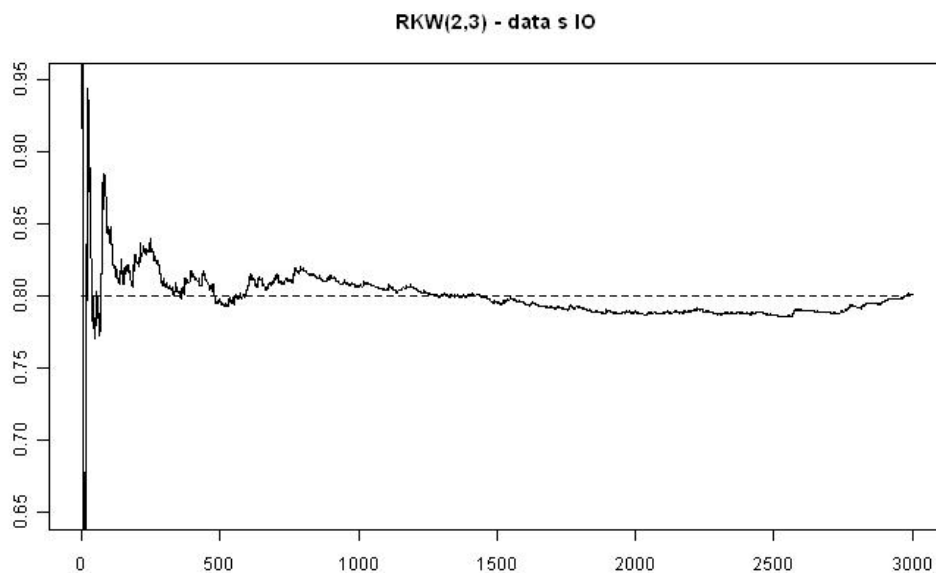
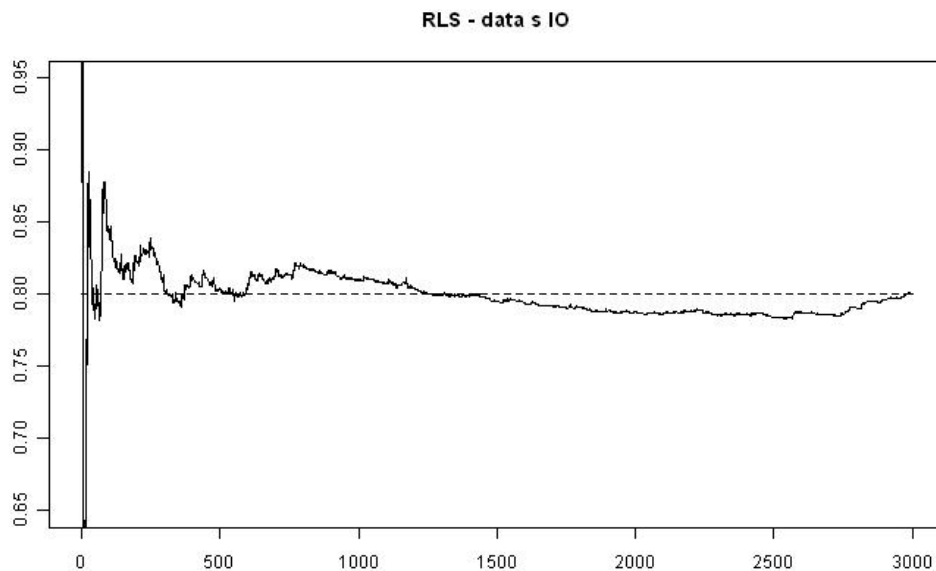
Tabulka 6.3: Výsledky pro AR(1) model a data s AO, kde $t_1 = 2005$ a $t_2 = 3005$



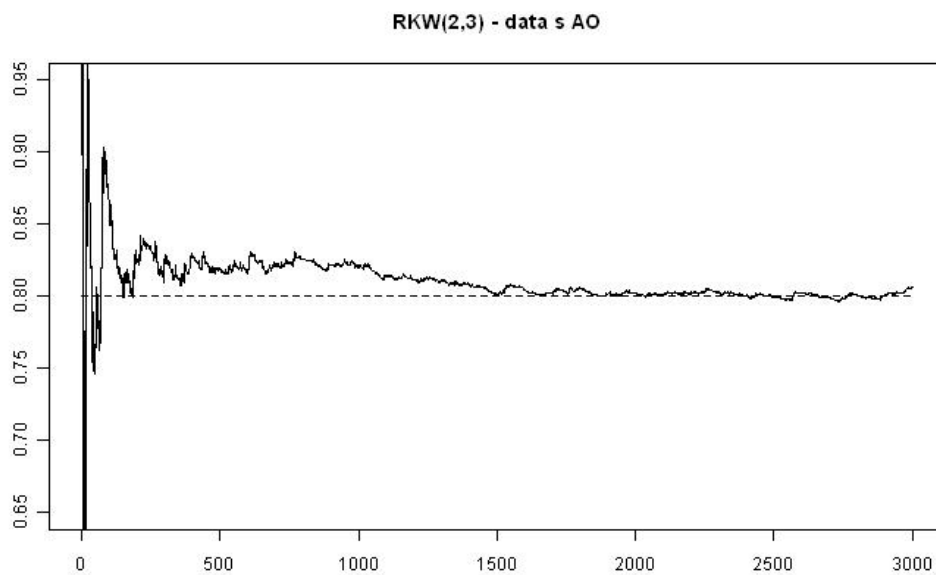
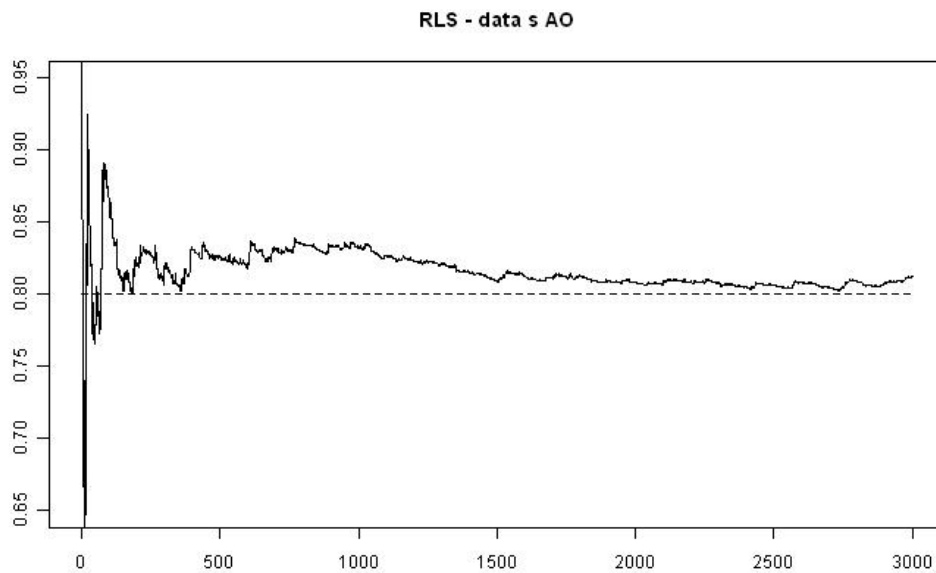
Obrázek 6.1: Data simulovaná s AR(1) modelem pro tři druhy odlehlých pozorování a se stejným základním šumem



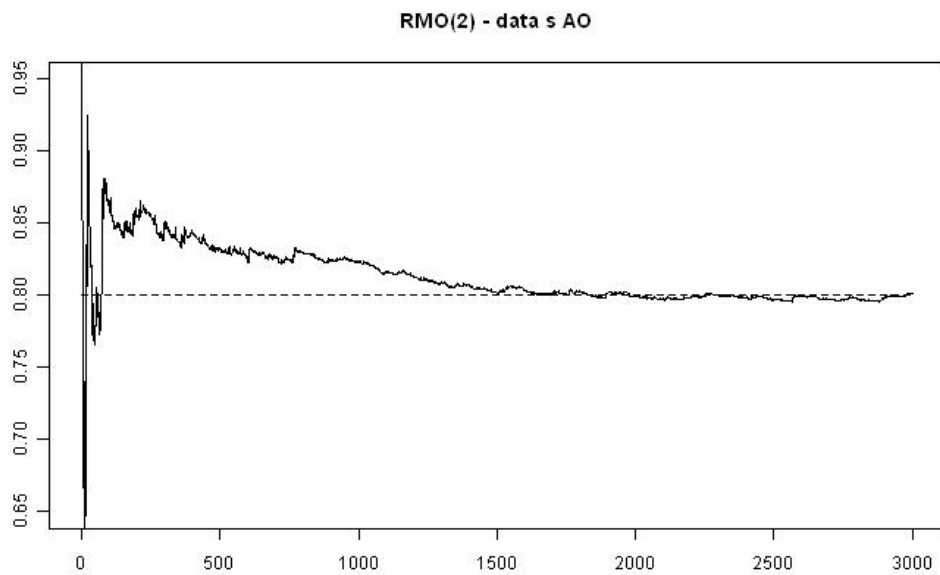
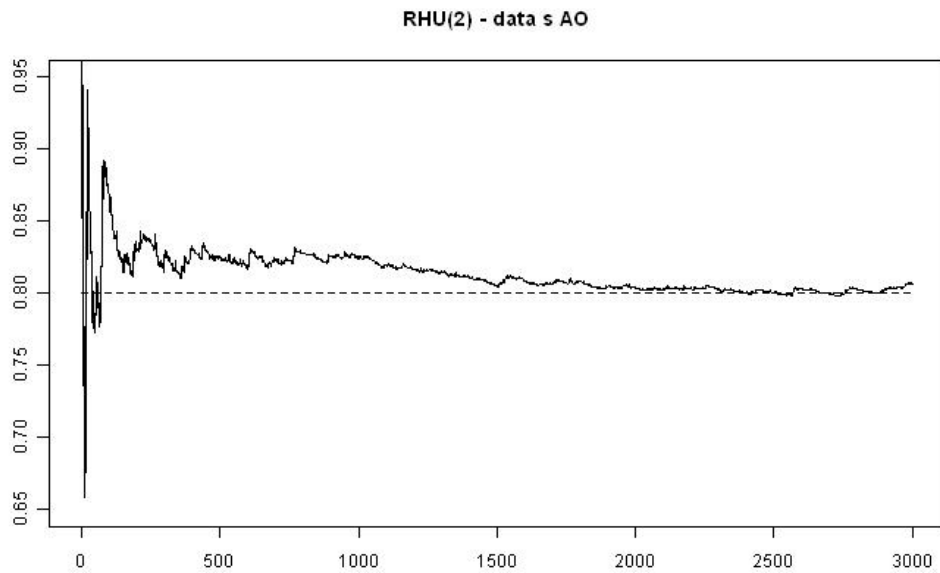
Obrázek 6.2: Průběh odhadů parametru v AR(1) modelu pro data bez odlehlých pozorování a RLS a RKW(2,3) algoritmy



Obrázek 6.3: Průběh odhadů parametru v AR(1) modelu pro data s odlehlými pozorováními způsobenými inovacemi a RLS a RKW(2,3) algoritmy



Obrázek 6.4: Průběh odhadů parametru v AR(1) modelu pro data s odlehlými pozorováními zapříčiněnými aditivními efekty a RLS a RKW(2,3) algoritmy



Obrázek 6.5: Průběh odhadů parametru v AR(1) modelu pro data s odlehlými pozorováními zapříčiněnými aditivními efekty a RHU(2) a RMO(2) algoritmy

Kapitola 7

Závěr

Úspěšné použití algoritmů pro odhady parametrů na reálná data vyžaduje, aby byly algoritmy robustní, tedy redukovaly vliv chyb pocházejících z odlehlých pozorování různých druhů. Tradiční metody robustních odhadů uvažují většinou off-line situace, avšak v mnoha praktických případech je nutno použít on-line odhady. V práci jsme nastínili metodu získávání rekurzivních robustních on-line algoritmů odhadu. Tuto metodu jsme použili k odvození dvou rekurzivních robustních algoritmů založených na minimalizaci robustifikované kritériální funkce, a to Huberovy a Kraskerovy a Welschovy.

Abychom vyšetřili vlastnosti odhadů, použili jsme je na simulovaná data. Jedním z výsledků našich simulací je, že RLS algoritmus není vhodný pro odhad parametrů, pokud počítáme s odlehlými pozorováními zapříčiněnými aditivními efekty. Pokud ho však rozšíříme o schopnost zacházet s detekovanými odlehlými pozorováními jako s chybějícími pozorováními, pak se s AO dokáže vypořádat. RHU se ukázal jako poměrně citlivý na AO, zatímco RKW je schopný se s odlehlými pozorováními celkem dobře vypořádat a způsobenou odchylku redukovat. Je nutno poznamenat, že RMO algoritmus, který odřízne část dat, se může zaseknout v situacích, kde všechna data považujeme za odlehlá pozorování a odhad se tak nezmění. To se ostatním algoritmům stát nemůže. Pokud pracujeme pouze s malou kontaminací odlehlými pozorováními, pak všechny robustní odhady slouží jako dobré vylepšení metody nejmenších čtverců. K většímu rozdílu mezi algoritmy dochází až v případě, kdy pracujeme s daty, která jsou hodně kontaminována.

Literatura

- [1] Antoch J., Eklblom H. (2002): *Selected Algorithms for Robust M- and L-Regression Estimators*, Developments in Robust Statistics, Springer, Berlin, 32–49.
- [2] Fox A.J. (1972): *Outliers in time series*, Journal of the Royal Statistical Society, Ser. B, 34, 350–363.
- [3] Hampel F.R. (1971): *A general qualitative definition of robustness*, Ann. Math. Statist. 42, 1887–1896.
- [4] Huber P.J. (1964): *Robust estimation of a location parameter*, Ann. Math. Statist. 36, 73–101.
- [5] Huber P.J. (1973): *Robust regression: Asymptotics, conjectures and Monte Carlo*, Anal. Statist., 5, 799–821.
- [6] Jurečková J. (2001): *Robustní statistické metody*, Karolinum, Praha.
- [7] Krasker W.S., Welsch R.E. (1982): *Efficient bounded-influence regression estimates*, J. Amer. Statist. Assoc., 77, 595–604.
- [8] Martin R.D., Yohai V.J. (1984): *Robustness in time series and estimating ARMA models*, Department of Statistics, University of Washington, Technical report No.50.
- [9] Sejling K., Madsen H., Holst J., Holst U., and Englund J.-E. (1994): *Methods for recursive robust estimation of AR parameters*, Computational Statistics and Data Analysis 17, 509–536.