



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

BAKALÁRSKA PRÁCA

Timotej Mičuda

Vybrané aplikácie rekurentnej metódy najmenších štvorcov

Katedra pravděpodobnosti a matematické statistiky

Vedúci bakalárskej práce: RNDr. Radek Hendrych, Ph.D.

Študijný program: Finanční matematika

Praha 2023

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

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

V dne

Podpis autora

Rád by som podakoval vedúcemu tejto práce RNDr. Radekovi Hendrychovi, Ph.D. za ochotu, pomoc a cenné rady pri písaní bakalárskej práce. Taktiež za spríjemenie štúdia a pomoc pri ňom ďakujem Bc. Adamovi Plavčanovi, Bc. Antonovi Matisovi a Bc. Erikovi Mendrošovi.

Názov práce: Vybrané aplikácie rekurentnej metódy najmenších štvorcov

Autor: Timotej Mičuda

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedúci bakalárskej práce: RNDr. Radek Hendrych, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Táto bakalárska práca obsahuje predstavenie, odvodenie a implementáciu rekurentnej metódy najmenších štvorcov. Metóda je analógiou klasickej metódy najmenších štvorcov pre lineárnu regresnú úlohu, pričom používa rekurzívny postup na aktualizáciu koeficientov. Cieľom tejto práce je predstavenie algoritmu rekurentnej metódy a oboznámenie sa s jej výhodami. Metódu využijeme na štyri rôzne simulácie, v ktorých ukážeme jej vlastnosti a modifikácie. Taktiež metódu aplikujeme na odhady pri použití na reálne dáta v spojitosti s CAPM modelom.

Kľúčové slová: CAPM model, forgetting factor, metóda najmenších štvorcov, rekurentná metóda najmenších štvorcov

Title: Recursive least squares method: Selected applications

Author: Timotej Mičuda

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Radek Hendrych, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This bachelor's thesis introduces, derives and implements a recursive least squares method. The recursive least squares method is an analogy to the ordinary least squares for linear regression exercise. The method uses recursive procedure to update coefficients. The goal of this work is to introduce the recursive method algorithm and to become familiar with its benefits. The technique is employed in four simulation studies to demonstrate its statistical features. The method is also practically applied in the context of the CAPM model.

Keywords: CAPM model, forgetting factor, least squares method, recursive least squares method

Obsah

Úvod	2
1 Rekapitulácia	3
1.1 Koncept regresie	3
1.2 Lineárna regresia	4
1.3 Odhad metódou najmenších štvorcov	6
2 Rekurentná metóda najmenších štvorcov	10
2.1 Motivácia	10
2.2 Rekurzívny identifikačný algoritmus	11
2.3 Odvodenie metódy	12
2.4 Počiatočné podmienky algoritmu a vybrané vlastnosti odhadov . .	16
2.4.1 Počiatočné podmienky	16
2.4.2 Asymptotické vlastnosti	17
2.5 Modifikácie	18
3 Simulačné štúdie	22
3.1 Simulácia: Vývoj RLS odhadu	22
3.2 Simulácia: Preloženie dát polynómom	26
3.3 Simulácia: Skúmanie konvergenzie RLS odhadu k OLS odhadu . .	29
3.4 Simulácia: Dopad forgetting factoru na odhad	33
4 Praktická časť	35
4.1 Predstavenie	35
4.2 Vzorec CAPM modelu	35
4.3 Aplikácia RLS na CAPM model	35
Záver	43
Zoznam použitej literatúry	44
Zoznam obrázkov	45
Zoznam tabuliek	46

Úvod

V tejto bakalárskej práci sa budeme venovať rekurentnej metóde najmenších štvorcov. Jedná sa o široko používaný algoritmus, ktorý je významnou súčasťou adaptívnych systémov. Uvedieme si najskôr koncept regresie a klasický model lineárnej regresie, keďže rekurentná metóda najmenších štvorcov je založená práve na týchto princípoch.

Hlavnou zložkou teoretickej časti bude odvodenie algoritmu pomocou odhadu metódy najmenších štvorcov ako aj vysvetlenie potreby a výhod rekurentných metód a menovite rekurentnej metódy najmenších štvorcov. Taktiež sa pozrieme na voľbu počiatočných podmienok pre inicializáciu algoritmu ako aj vybrané vlastnosti odhadov. Simulácie uvedené v práci nám poslúžia pre lepšie pochopenie metódy a na ukázanie rôznych modifikácií a vlastností. Odhad rekurentnou metódou najmenších štvorcov budeme v simuláciách porovnávať s odhadom získaným klasickou metódou najmenších štvorcov ako numericky, tak aj na grafoch.

V praktickej časti sa pozrieme na využitie odvodeného algoritmu vo finančnej sfére v teórii portfólia. Konkrétne algoritmus použijeme pre odhadu faktoru β v CAPM modele, ktorý slúži na výpočet očakávaného zisku na investícii. Budeme odhadovať faktory β pre tri akcie z akciového indexu *S&P* 500 na základe reálnych historických dát.

1. Rekapitulácia

Časti tejto kapitoly vychádzajú z knihy Ljung (1999) a taktiež z knihy Anděl (2007).

1.1 Koncept regresie

Koncept regresie sa v teórii štatistiky používa k modelovaniu vzťahov medzi nezávislými premennými a závislou premennou. Závislá veličina je v tomto prípade skalárna (náhodná) veličina Y , ktorú chceme, čo najlepšie vysvetliť pomocou údajov obsiahnutých v nezávislých (náhodných) premenných $x(1), \dots, x(d)$. Cieľom regresie je nájsť matematický model, ktorý vhodne popisuje ich vzťah a umožňuje nám vytvoriť predikcie hodnôt závislých premenných Y založených na hodnotách nezávislých premenných. Regresia je používaná v rôznych odvetviach vedy i spoločnosti a to konkrétne napríklad v štatistike, ekonómii, sociológii, ale aj v psychológii a iných. Jednoduchým príkladom môže byť napríklad predikcia výšky platu osôb na základe ich vzdelania, veku, dĺžky praxe a lokácie. Príslušný model vytvoríme pomocou regresnej analýzy z pozorovaní, ktoré je možné získať priamo z firiem alebo z rôznych štatistických webov. Dynamické systémy sú jednou zo zložiek spadajúcou do regresného konceptu, kde Y je výstup systému v danom čase a $x(i)$ obsahujú informácie o náhodných veličinách Y . Uvedme si model teda v matematickom značení. Ako prvé si zavedieme označenie vektoru \mathbf{x} ako:

$$\mathbf{x} = \begin{pmatrix} x(1) \\ x(2) \\ \vdots \\ x(d) \end{pmatrix}.$$

Regresný problém vieme teda interpretovať ako hľadanie funkcie regresorov $g(\mathbf{x})$ tak, aby rozdiel

$$Y - g(\mathbf{x})$$

bol čo najmenší (tzn., že $\hat{Y} := g(\mathbf{x})$ je dobrým odhadom Y). Ak Y a \mathbf{x} sú náhodné veličiny môžeme sa napríklad pokúsiť o minimalizáciu

$$\mathbb{E} [Y - g(\mathbf{x})]^2. \tag{1.1}$$

Strednú hodnotu (1.1) môžeme prepísať ako:

$$\begin{aligned} \mathbb{E} [Y - g(\mathbf{x})]^2 &= \mathbb{E} \{ [Y - \mathbb{E}(Y|\mathbf{x})] - [g(\mathbf{x}) - \mathbb{E}(Y|\mathbf{x})] \}^2 \\ &= \mathbb{E} [Y - \mathbb{E}(Y|\mathbf{x})]^2 - 2 \cdot \mathbb{E} \{ [Y - \mathbb{E}(Y|\mathbf{x})][g(\mathbf{x}) - \mathbb{E}(Y|\mathbf{x})] \} \\ &\quad + \mathbb{E} [g(\mathbf{x}) - \mathbb{E}(Y|\mathbf{x})]^2 = \mathbb{E} [Y - \mathbb{E}(Y|\mathbf{x})]^2 + \mathbb{E} [g(\mathbf{x}) - \mathbb{E}(Y|\mathbf{x})]^2, \end{aligned}$$

čo je minimálne práve pre:

$$g(\mathbf{x}) = \mathbb{E}(Y|\mathbf{x}). \tag{1.2}$$

Je to teda podmienená stredná hodnota Y za podmienok $\mathbf{x}(1), \dots, \mathbf{x}(d)$. Túto funkciu (1.2) nazveme regresná funkcia alebo regresia Y na \mathbf{x} . Regresná funkcia zabezpečí aby rozdiel $Y - g(\mathbf{x})$ bol čo najmenší (minimálny).

Iný spôsob ako rozdiel minimalizovať je založený na hľadaní funkcie $\hat{g}(\mathbf{x})$, ktorá má maximálnu koreláciu s Y . Výsledkom pri použití tohto prístupu je ale znovu regresná funkcia (1.2), čo vyplýva z Ljung a Söderström (1983) zo strany 544. Regresná analýza je v podstate teda štatistická metóda, ktorá skúma vzťahy medzi závislou premennou a jednou alebo viacerými nezávislými premennými.

1.2 Lineárna regresia

Lineárna regresia je jednou zo základných a široko používaných regresných analýz. Podstatou tejto metódy je preloženie získaných, či nameraných dát vhodnou krivkou, ktorá najlepšie vystihuje vzťahy medzi premennými. Táto krivka v určitých aplikáciách býva jednoduchá priamka, ale obecné to môže byť ľubovoľný polynóm $n - tého$ rádu, ktorý lepšie dané vzťahy vysvetľuje. V regresii, kde predpokladáme, že tento existujúci vzťah je priamy lineárny a teda je parametrizácia omedzená na lineárnu sa nazýva lineárna regresia.

Vlastnosti veličín Y a \mathbf{x} nemusia byť vždy známe, čo je často aj prípadom v regresných analýzách, čo spôsobuje, že nie je možné vopred stanoviť regresnú funkciu $g(\mathbf{x})$. Regresná funkcia musí byť teda odhadnutá priamo z dostupných dát. Chceme teda namodelovať Y najvhodnejším spôsobom pomocou lineárnej kombinácie $x(i)$ ako:

$$g(\mathbf{x}) = \beta(1)x(1) + \beta(2)x(2) + \dots + \beta(d)x(d), \quad (1.3)$$

kde vektor $(\beta(1), \dots, \beta(d))^T$ je vektor neznámych parametrov. Práve tieto parametre potrebujeme odhadnúť, čo je cieľom lineárnej regresie. Ak pridáme zavedenie označenia vektoru

$$\boldsymbol{\beta} = \begin{pmatrix} \beta(1) \\ \beta(2) \\ \vdots \\ \beta(d) \end{pmatrix}$$

vieme vyjadrenie (1.3) prepísať ako:

$$g(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}. \quad (1.4)$$

Uvažujme, že pozorované dáta pochádzajú z N nezávislých pozorovaní vektora $(Y(t), \mathbf{x}(t))$, kde $t = 1, \dots, N$ a pre každé $\mathbf{x}(t)$ platí, že má $d < N$ zložiek $(x(t,1), \dots, x(t,d))^T$. Náhodnú veličinu $Y(t)$ v tomto kontexte nazývame odozva (príp. závislá premenná) a zložky náhodného vektora $\mathbf{x}(t)$ sú nazývané regresory (príp. nezávislé premenné). Formálne vieme potom model lineárnej regresie zapísať takto:

$$Y(t) = \beta(1)x(t,1) + \beta(2)x(t,2) + \dots + \beta(d)x(t,d) + \epsilon(t). \quad (1.5)$$

Povieme teda, že dáta $(Y(t), \mathbf{x}(t))$, $t = 1, \dots, N$, spĺňajú lineárny regresný model, ak sa riadia práve podľa vyjadrenia (1.5). Ako už vieme tak $\boldsymbol{\beta} = (\beta(1), \dots, \beta(d))^T$ je vektor neznámych parametrov a $\epsilon(1), \dots, \epsilon(N)$ sú nezávislé rovnako rozdelené náhodné veličiny, ktoré sú nazývané chybové členy a musia spĺňať tieto predpoklady $E \epsilon(t) = 0$ a $\text{var} \epsilon(t) = \sigma^2$ pre všetky t . Taktiež požadujeme nekorelovanosť

regresorov a chybového člena. Zložky vektora β nazývame regresné koeficienty pre dané regresory.

Model lineárnej regresie vieme zapísať aj iným spôsobom a to napríklad cez využitie podmienených momentov a v tom prípade model nadobúda tento tvar:

$$\begin{aligned} E(Y(t)|\mathbf{x}(t)) &= \beta(1)x(t,1) + \beta(2)x(t,2) + \dots + \beta(d)x(t,d), \\ \text{var}(Y(t)|\mathbf{x}(t)) &= \sigma^2. \end{aligned}$$

Ak máme dané $\mathbf{x}(t)$ tento zápis zdôrazňuje, že lineárny regresný model vyjadruje strednú hodnotu $Y(t)$ pomocou lineárneho vzťahu a taktiež predpokladá, že rozptyl $Y(t)$ je konštantný a nezávislý na $\mathbf{x}(t)$.

Obvykle sa v regresnej analýze sa volí $x(t,1) = 1$ pre všetky t . Parameter $\beta(1)$ potom nazývame absolútny člen. Takto zapísaný model lineárnej regresie, ale ešte nie je jej klasickým zápisom. Model je užitočné a možné prepísať nasledujúcim maticovým spôsobom, ktorý je nám určite známejší. Budeme uvažovať \mathbf{X} ako vektor, ktorý môže pozostávať z jednej alebo viacerých spojitých či diskretných veličín. \mathbf{X} teda bude maticou, ktorá je zložená z vektorov $\mathbf{x}(t)$. Prvý stĺpec tejto matice \mathbf{X} pozostáva zo samých jednotiek, keďže sa prvý člen volí rovný 1. Budeme teda pozeráť na to, ktoré zložky \mathbf{X} ovplyvňujú $E \mathbf{Y}$ a akým spôsobom. Uvedme si viac prehľadný a používaný maticový zápis, ktorý model lineárnej regresie uvádza v tomto tvare:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \quad (1.6)$$

kde

$$\begin{aligned} \mathbf{X} &= \begin{pmatrix} x(1,1) & x(1,2) & \dots & x(1,d) \\ x(2,1) & x(2,2) & \dots & x(2,d) \\ \vdots & \vdots & \dots & \vdots \\ x(N,1) & x(N,2) & \dots & x(N,d) \end{pmatrix} = \begin{pmatrix} 1 & x(1,2) & \dots & x(1,d) \\ 1 & x(2,2) & \dots & x(2,d) \\ \vdots & \vdots & \dots & \vdots \\ 1 & x(N,2) & \dots & x(N,d) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{x}^T(1) \\ \mathbf{x}^T(2) \\ \vdots \\ \mathbf{x}^T(N) \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} Y(1) \\ Y(2) \\ \vdots \\ Y(N) \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta(1) \\ \beta(2) \\ \vdots \\ \beta(d) \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon(1) \\ \epsilon(2) \\ \vdots \\ \epsilon(N) \end{pmatrix}, \end{aligned}$$

pričom \mathbf{X} je matica ($N \times d$), β je vektor ($d \times 1$), ϵ a \mathbf{Y} sú vektory ($N \times 1$). Matica \mathbf{X} sa nazýva regresná matica a jedným z predpokladov modelu lineárnej regresie je lineárna nezávislosť jej stĺpcov. Všetky predpoklady si uvedieme na konci tejto podkapitoly. Ako vidíme tak model (1.6) je vlastne len vektorovo prepísaný model (1.5).

Niektoré predpoklady, ktoré charakterizujú tzv. klasický model lineárnej regresie zapisovaný ako (1.5), či (1.6) sme si už spomenuli popri zavedení tejto metódy. Uvedme si ešte všetky tieto predpoklady na jednom mieste pre lepšiu prehľadnosť. Tieto predpoklady sa často uvádzajú v nasledujúcom tvare ako uvádza (Cipra (2008),str. 40):

- $E \epsilon(t) = 0$,
t.j. stredná hodnota reziduálnej zložky je nulová pre všetky t ;
- $\text{var} \epsilon(t) = \sigma^2 < \infty$,
t.j. rozptyl reziduálnej zložky je konštantný a konečný pre všetky t ;

- $\text{cov}(\epsilon(s), \epsilon(t)) = 0$ pre $s \neq t$,
t.j. reziduálne zložky sú navzájom nekorelované pre všetky $s \neq t$;
- $\text{cov}(x(t,i), \epsilon(t))$,
t.j. regresory sú rovnakom čase nekorelované s chybovou zložkou pre všetky i a t (slabšia podoba tohoto prepokladu je uvedená v nasledujúcom bode);
- $r(\mathbf{X})=d$,
t.j. nenáhodná matica \mathbf{X} má lineárne nezávislé stĺpce (uplatní sa iba pri nenáhodných regresoroch).

1.3 Odhad metódou najmenších štvorcov

Ako sme si už v predchádzajúcej podkapitole spomenuli väčšinou vopred nevieme, aké presné vzťahy medzi $Y(t)$ a $\mathbf{x}(t)$ existujú. Spôsob ako regresnú funkciu získame je pomocou historických dát. Použijeme ku konštrukcii regresnej funkcie pozorovania, ktoré sme namerali a sú pre nás už dostupné. Táto kolekcia predchádzajúcich pozorovaní (dát) nám určitú informáciu o \mathbf{Y} a \mathbf{X} dáva, čo je pre nás veľmi užitočné a potrebné. Je bežné tieto namerané hodnoty očíslovať postupne takýmto spôsobom:

$$Y(t), \mathbf{x}(t) \quad t = 1, \dots, N \quad (1.7)$$

Po očíslovaní historických pozorovaní a taktiež odchýliek jednotlivých pozorovaní vieme (1.1) prepísať do tvaru:

$$\frac{1}{N} \sum_{t=1}^N [Y(t) - g(\mathbf{x}(t))]^2.$$

Tento prepis jednoducho vyplýva z definície strednej hodnoty. So zadaným vektorom $\boldsymbol{\beta}$ a za pomoci vyjadrenia (1.4) v lineárnom prípade dostávame (1.1) ako:

$$V_n(\boldsymbol{\beta}) = \frac{1}{N} \sum_{t=1}^N [Y(t) - \mathbf{x}^T(t)\boldsymbol{\beta}]^2. \quad (1.8)$$

Vhodnou voľbou a výberom vektora $\boldsymbol{\beta}$ je teda ten, ktorý bude minimalizovať argument (1.8):

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^d} V_N(\boldsymbol{\beta}),$$

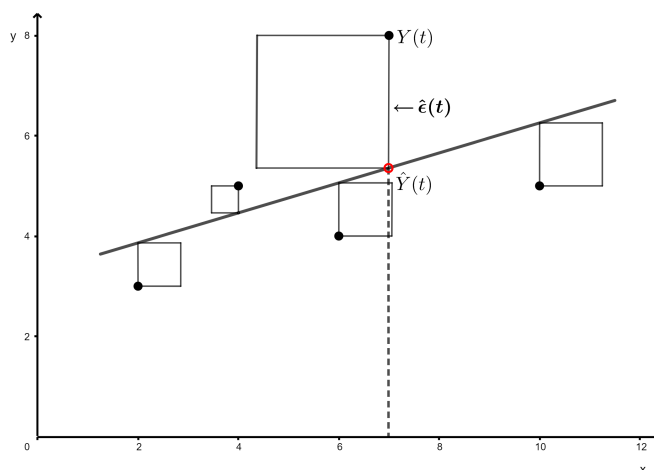
Ak sa na vyjadrenie pozrieme po častiach hľadáme jednoducho vektor $\boldsymbol{\beta}$, ktorý minimalizuje celkové rozdiely $(Y(t) - \mathbf{x}^T(t)\boldsymbol{\beta})^2$ pre každé t .

Definícia 1. *Odhad $\hat{\boldsymbol{\beta}}$, ktorý splňa $\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^d} V_N(\boldsymbol{\beta})$, nazývame odhadom metódou najmenších štvorcov.*

Na základe predchádzajúcich pozorovaní sme sa postupne dostali k odhadu vektora neznámych parametrov $\hat{\boldsymbol{\beta}}$, ktorý by sme v spoločne s vektorom \mathbf{x} použili ako predikčnú funkciu v takomto tvare:

$$\mathbf{x}^T \hat{\boldsymbol{\beta}}.$$

Táto metóda vyberania $\boldsymbol{\beta}$ má zmysel s, ale aj bez, zavedenia stochastickej štruktúry pre daný problém, teda výber vhodného odhadu funguje aj bez nutnosti



Obr. 1.1: Metóda najmenších štvorcov

toho, že Y a \mathbf{x} sú náhodné veličiny. Parameter $\hat{\beta}$ je jednoducho tá hodnota, ktorá podáva najlepšie výsledky pri použití na historické dáta.

Funkciu 1.8, ktorú minimalizujeme, vieme prepísať ako

$$\frac{1}{N} \sum_{t=1}^N (Y(t) - \beta^T \mathbf{x}(t))^2 = \frac{1}{N} (\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta).$$

Vyjadrenie pre odhad $\hat{\beta}$ následne z prepísaného tvaru dostaneme jednoducho pomocou maticových derivácií ako:

$$\begin{aligned} \frac{\partial}{\partial \beta} \frac{1}{N} (\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta) &= \frac{\partial}{\partial \beta} \frac{1}{N} (\mathbf{Y}^T \mathbf{Y} - 2\beta^T \mathbf{X}^T \mathbf{Y} + \beta^T \mathbf{X}^T \mathbf{X} \beta) \\ &= -\frac{2}{N} \mathbf{X}^T \mathbf{Y} + \frac{2}{N} \mathbf{X}^T \mathbf{X} \beta = -2 \left(\frac{1}{N} \mathbf{X}^T \mathbf{Y} - \frac{1}{N} \mathbf{X}^T \mathbf{X} \beta \right). \end{aligned}$$

Pri položení derivácie rovno nule, zistíme, že $\hat{\beta}$ rieši sústavu d lineárnych rovníc rovníc o d neznámych

$$\left(\frac{1}{N} \mathbf{X}^T \mathbf{X} \right) \hat{\beta} = \frac{1}{N} \mathbf{X}^T \mathbf{Y}.$$

Riešenie musí byť globálnym minimom, pretože minimalizovaná funkcia je konvexná v β . Keďže predpokladáme, že matica \mathbf{X} má plnú hodnotnosť d , matica $\mathbf{X}^T \mathbf{X}$ (štvorcová $d \times d$) má taktiež plnú hodnotnosť d . Teda existuje práve jedno riešenie danej sústavy

$$\hat{\beta} = \left(\frac{1}{N} \mathbf{X}^T \mathbf{X} \right)^{-1} \frac{1}{N} \mathbf{X}^T \mathbf{Y} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$

Ak sa vrátíme späť ku vektorovému vyjadreniu dostaneme, že všetky $\hat{\beta}$ spĺňajúce

$$\left[\frac{1}{N} \sum_{t=1}^N \mathbf{x}(t) \mathbf{x}^T(t) \right] \hat{\beta} = \frac{1}{N} \sum_{t=1}^N \mathbf{x}(t) Y(t) \quad (1.9)$$

sú globálnym minimom $V_N(\beta)$. Táto sústava lineárnych rovníc je všeobecne známa ako normálne rovnice. Ak je teda matica na ľavej strane invertovateľná, odhad pomocou metódy najmenších štvorcov vektorovo je vo tvare:

$$\hat{\beta} = \left[\frac{1}{N} \sum_{t=1}^N \mathbf{x}(t) \mathbf{x}^T(t) \right]^{-1} \frac{1}{N} \sum_{t=1}^N \mathbf{x}(t) Y(t). \quad (1.10)$$

Značenie je v súlade so značením zavedeným v (1.6). Odhad $\hat{\beta}$ vyberieme tým spôsobom, aby bol vektor $\hat{\mathbf{Y}}$ čo najbližšie vektoru \mathbf{Y} v euklidovskej vzdialenosti.

V (1.8) sú rôznym pozorovaniam priradené rovnaké váhy, ale občas je vhodné uvažovať, že každé pozorovanie má inú váhu. Priradenie rôznych váh daným nameraným pozorovaniam môže mať signifikantný vplyv na výsledné odhady, čo môže byť na jednej strane veľmi užitočné a je v praxi často používané, ale na druhej strane je preto voľbu váh potrebné robiť veľmi opatrne. Vážené najmenšie štvorce sú len doplnkom (1.8) o postupnosť čísel α_t a nadobúda takýto tvar:

$$V_N(\beta) = \sum_{t=1}^N \alpha_t [Y(t) - \mathbf{x}^T(t)\beta]^2.$$

Reálne váhy α_t môžu byť ako pozitívne tak negatívne, ale aj nulové. V praxi zvyčajne uvažujeme nezáporné váhy a to z toho dôvodu, aby negatívne váhy nerušili efekt tých pozitívnych. Ďalšou nevýhodou negatívnych váh je to, že môžu viesť k zaujatým výsledkom a nestálym odhadom. Váhy pozorovaní je kvôli ich výraznému dopadu na výsledky potrebné voliť opatrne a s ohľadom na charakter dát. Najčastejšie sa pokladanie rôznych váh pre pozorovania v praxi uvažuje v týchto dvoch prípadoch:

- Niektoré pozorovania $Y(t)$ môžu byť viac zafaržené chybovou zložkou ako iné, čo môže viesť na rôzne spoľahlivé pozorovania. V takomto prípade je potrebné menej spoľahlivým pozorovaniam udeliť menšiu váhu ako ostatným.
- Ak lineárny model nepoužíva celú škálu $\mathbf{x}(t)$ tak sa týmto viac diskutabilným a teda menej zásadným oblastiam priraďuje menšia váha. Je to v princípe rôzna relevantnosť daných častí $\mathbf{x}(t)$.

Ku koncu tejto podkapitoly si ešte stručne uvedme niektoré vlastnosti odhadu $\hat{\beta}$ metódy najmenších štvorcov, ktoré sú z kapitoly 5 v knihe Anděl (2007). Dané vybrané vlastnosti si spíšeme do nasledujúcej vety.

Veta 1. (Vlastnosti odhadov metódou najmenších štvorcov)

1. $E(\hat{\beta}|\mathbf{X}) = \beta$, $\hat{\beta}$ je pri danom \mathbf{X} nestranný odhad
2. $var(\hat{\beta}|\mathbf{X}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$
3. platí, že $\hat{\beta}|\mathbf{X} \sim N(\beta, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$, ak sú splnené predpoklady normality a to, že $\epsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ a taktiež, že $\mathbf{Y}|\mathbf{X} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$.

Dôkaz.

1. aplikujeme strednú hodnotu na odhad, ktorý následne rozpíšeme a upravíme

$$\begin{aligned} E(\hat{\beta}|\mathbf{X}) &= E(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{Y}|\mathbf{X}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E(\mathbf{X}\beta + \epsilon) \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\beta = \beta \end{aligned}$$

2. postupujeme podobne ako v prvom bode a použijeme pravidlá rozkladu rozptylu a upravíme

$$\begin{aligned}\text{var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) &= \text{var}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{Y}|\mathbf{X}) \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\text{var}(\mathbf{Y}|\mathbf{X})) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 I \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}\end{aligned}$$

3. z predpokladov vyplýva, že \mathbf{Y} má normálne rozdelenie a keďže $\boldsymbol{\beta}$ vzniká z \mathbf{Y} lineárnou transformáciou má rovnako normálne rozdelenie, ktorého parametre sme si už spočítali v bodoch 1. a 2. tejto vety.

□

2. Rekurentná metóda najmenších štvorcov

Časti tejto kapitoly vychádzajú z Ljung (1999) a taktiež z Ljung a Söderström (1983).

2.1 Motivácia

Najskôr sa v tejto kapitole pozrieme na motiváciu rekurzívnej metódy najmenších štvorcov, jej výhody a rozdiel oproti klasickej metóde najmenších štvorcov, keďže z nej práve rekurzívna metóda vychádza.

Predstavme si najskôr výrazy *system* a *model k danému systému*, ktoré sa nám zídu k lepšiemu pochopeniu metódy. Pod pomenovaním *system* rozumieme širokú škálu viacmenej komplexných objektov, o ktorých správanie sa zaujímate. Môžeme sa snažiť o kontrolu daného systému alebo jednoducho o pochopenie jeho štruktúr a fungovania. Príkladmi z praxe môže byť napríklad kontrola kvality produktu firmami, ktoré produkt vyrábajú na základe materiálov, teploty, hmotnosti a podobne pre optimálny výrobný proces. Ďalším príkladom sú banky, ktoré sa snažia dobre pochopiť ako funguje zlyhávanie subjektov pri splácaní na základe veku, príjmu, finančnej situácie a iných relevantných indikátorov.

Je potrebná znalosť vlastností systému pre správne pochopenie jeho správania. Určité chápanie a znalosť vlastností systému sa všeobecne nazývajú modelom. Je prakticky nevyhnutné poznať model systému, aby sme vedeli riešiť problémy, ktoré naväzujú na systém. Ako sa dostať k modelu systému je zvyčajne možné dvoma spôsobmi a to buď modelovaním, či identifikáciou. Modelovanie funguje na princípe, že sa priamo pozrieme dovnútra systému na vzťahy, ktoré v ňom figurujú, čo ale v spomenutých praktických prípadoch úplne nie je možné. Druhým spôsobom je identifikácia, ktorej základom je tvorba modelu pomocou výsledkov daného systému. Teda vytvoríme model z pozorovaní, ktoré z daného systému získame. V prípade bánk to budú jednoducho dáta o klientoch, ktorý mali poskytnutý úver. Identifikácia sa ešte delí na on-line a off-line identifikáciu. Off-line identifikácia modelu funguje tak, že zozbierame určitý počet dát a následne vytvoríme príslušný model. Naopak on-line identifikácia s každým novým získaným pozorovaním model aktualizuje a vezme do úvahy aj každé nové pozorovanie, čo vedie k adaptívnejšiemu a presnejšiemu modelu.

Nevýhodou klasickej metódy najmenších štvorcov pri jej použití pri on-line identifikácii je to, že spočítanie každého aktualizovaného odhadu s príchodom nového pozorovania spôsobuje, že výpočty sú pomerne časovo náročné a taktiež zaberajú veľké množstvo pamäte, keďže je potrebné mať uložené všetky doposiaľ namerané pozorovania. Rekurzívna metóda ako uvidíme funguje na princípe rekurzívnej a preto je možné staré pozorovania postupne vypúšťať a nezaťažovať s ním proces. Taktiež je metóda prakticky úplne oslobodená od väčších maticových násobení, či maticovej inverzie, čo je veľmi užitočné ak potrebujeme, aby sa model rýchlo aktualizoval pomocou nových pozorovaní a reagoval na dané prípadné zmeny. Rekurzívna identifikácia so sebou prináša aj dve nevýhody. V prvom rade je to, že pri on-line identifikácii je potrebné pred začatím pro-

cedúry predpokladať a zvoliť modelu určitú štruktúru, čo môže byť v niektorých prípadoch pomerne zložitú. Druhou nevýhodou sa ukázalo zhoršenie presnosti odhadov modelu pri rekurzívnych metódach v porovnaní s off-line metódami.

2.2 Rekurzívny identifikačný algoritmus

V tejto časti si predstavíme ako funguje rekurzívny identifikačný algoritmus, ktorým je aj rekurzívna metóda najmenších štvorcov. Základom pre všetky algoritmy, ktoré spadajú do tejto skupiny sú ako už vieme namerané dáta. Budeme označovať ako $z(t)$. Teda $z(t)$ predstavuje dáta alebo skupinu dát (môže byť vektor pozostávajúci z viac pozorovaní), ktoré sme obdržali v čase t . Budeme uvažovať, že akvizícia dát prebieha v diskretnom čase (prebieha v špecifických a fixne daných intervaloch) a nie v spojitom čase. To je kvôli tomu, že chceme, aby sa model aktualizoval iba s príchodom nového pozorovania, či skupiny pozorovaní. V čase t máme teda k dispozícii postupnosť pozorovaní $z(1), z(2), \dots, z(t)$. Zvyčajne sa uvažuje pre očíslovanie množina celých kladných čísel. Podstatou je teda odvodenie modelu, pre daný systém zo získaných dát z^t . Budeme používať označenie z^t pre celkovú dostupnú kolekciu pozorovaní do času t . Uvažujme teda, že: $z^t = \{z(t), z(t-1), \dots, z(1)\}$. Odvodenie modelu posluží pre stanovenie vektoru β , keďže to je práve vektor cez ktorý sme uviedli parametrizáciu modelov v prvej kapitole. Identifikačný problém teda prakticky spočíva v určení zobrazenia dát z^t k parametrom modelu. Matematicky sa tento vzťah značí ako:

$$z^t \rightarrow \hat{\beta}(t; z^t), \quad (2.1)$$

kde hodnota $\hat{\beta}(t; z^t)$ je odhad parametrov vektora β vytvorený na základe dostupných pozorovaní z^t . Pri on-line a rekurzívnej identifikácii je nevyhnutné spočítať odhad $\hat{\beta}$ pre každé t . Od tohoto momentu si, ale musíme dať pozor na rozlišovanie označenia $\beta(t)$ a $\beta(t)$. Ako sme si v (1.6) uviedli, tak vektor odhadovaných parametrov je d prvkový vektor, ktoré prvky značíme ako: $\beta(1), \dots, \beta(d)$. Povedali sme si, že pri rekurzívnom identifikačnom algoritme musíme pre každé nové pozorovania spočítať odhad. Pomocou označenia $\beta(t)$ máme teda na mysli t -tý odhad vektoru β . Teda odhad, ktorý bol vytvorený zo z^t . Okrem spočítania postupnosti odhadov taktiež požadujeme, aby algoritmus nebol početne náročný a nezaberal veľa pamäte, preto sú na neho kladené určité požiadavky (reštrikcie). Obvykle sa v off-line algoritmoch zozbiera N dát a následne sa vytvorí odhad $\hat{\beta} = \hat{\beta}(N; z^N)$. V našom prípade sa práve potrebujeme vyhnúť tomu, aby sme museli mať uloženú množinu pozorovaní z^N . Pre tieto účely budeme uvažovať samostatnú premennú $S(t)$, ktorá má fixné dimenzie a bude slúžiť ako funkcia obsahujúca všetky potrebné informácie o prechádzajúcich pozorovaniach a odhadoch. Tento parameter $S(t)$ a $\hat{\beta}(t)$ bude postupne aktualizovaný s príchodom nových pozorovaní. Bude v čase t postačujúce mať uložené len hodnoty $\{S(t), \hat{\beta}(t)\}$, ktorých hodnoty sa budú aktualizovať vzhľadom k tomu, ako bude algoritmus vyzerať. Pri odvodení každého algoritmu, ktorý spĺňa tieto nároky potrebujeme nájsť vhodné a fixné funkcie $H(\cdot, \cdot, \cdot)$ a $F(\cdot, \cdot, \cdot)$ tak aby platilo nasledovné:

$$\hat{\beta}(t) = F(\hat{\beta}(t-1), S(t), z(t)), \quad (2.2a)$$

$$S(t) = H(S(t-1), \hat{\beta}(t-1), z(t)). \quad (2.2b)$$

Určenie vhodných funkcií $H(\cdot, \cdot, \cdot)$ a $F(\cdot, \cdot, \cdot)$ je ďalšou podstatnou časťou tvorby vhodného rekurzívneho algoritmu. V praxi existuje viacero prístupov ako sa k odvodeniu používaných metód dostať. Pomôže nám k tomu aj predošlé vysvetlenie, čo znamenajú pojmy systém a aj príslušný model k danému systému. V tejto práci použijeme prístup, ktorý sa nazýva: Modifikácia off-line identifikačnej metódy. Ako už z názvu vyplýva prístup funguje na báze, že zoberieme ľubovoľný off-line identifikačný algoritmus, ktorý modifikujeme a upravíme tak, aby spĺňal obmedzenia pre rekurzívny identifikačný algoritmus. Použijeme teda klasickú metódu najmenších štvorcov, ktorú modifikujeme tak, aby spĺňala spomenuté reštrikcie a bolo ju možné považovať za rekurzívny identifikačný algoritmus. Konkrétne obmedzenia, ktoré je potrebné splniť sme si zadefinovali v (2.2).

Ešte predtým ako sa pustíme na odvodenie rekurentnej metódy najmenších štvorcov, by v tejto časti práce bolo vhodné spomenúť takzvaný Kalmanov filter, ktorý je odvodený zo stavového modelu. Na rekurentnú metódu najmenších štvorcov je možné pozeráť ako na špeciálny prípad Kalmanovho filtra za určitých podmienok. Hlbšie sa, ale týmto prípadom Kalmanovho filtra a ani Kalmanovým filtrom nebudeme zaoberať kvôli pomerne dlhému odvodeniu a potrebnému vysvetlovaniu. Viac o Kalmanovom filtri je možné nájsť v knihe Cipra (2008) v podkapitole 12.7.

2.3 Odvodenie metódy

Uvažujeme znovu model lineárnej regresie (1.5), ktorý sme predstavili v prvej kapitole ako:

$$Y(t) = \boldsymbol{\beta}^T \mathbf{x}(t) + \epsilon(t).$$

Už vieme, že vektor $\boldsymbol{\beta}$ je možné odhadnúť z nameraných pozorovaní $Y(t)$ a $\mathbf{x}(t)$, kde $t = 1, \dots, N$. K získaniu tohoto odhadu využijeme minimalizáciu chybových členov $\epsilon(t)$ v tomto tvare:

$$V_N(\boldsymbol{\beta}) = \frac{1}{N} \sum_{t=1}^N \alpha_t [Y(t) - \boldsymbol{\beta}^T \mathbf{x}(t)]^2, \quad (2.3)$$

ktorú ako už vieme následne minimalizujeme vzhľadom k $\boldsymbol{\beta}$. Vo vzorci taktiež vystupuje deterministická reálna postupnosť čísel α_t , ktorá ako sme si už spomenuli slúži na to, aby sme mohli dávať rôzne váhy rôznym pozorovaniam. Vo väčšine použití tejto metódy je ale zvykom voliť koeficient α_t rovný 1. V tomto prípade minimalizáciou $V_N(\boldsymbol{\beta})$ sa dostávame k odhadu pomocou metódy najmenších v tomto tvare:

$$\hat{\boldsymbol{\beta}}(N) = \left[\sum_{t=1}^N \alpha_t \mathbf{x}(t) \mathbf{x}^T(t) \right]^{-1} \sum_{t=1}^N \alpha_t \mathbf{x}(t) Y(t), \quad (2.4)$$

ak príslušný inverz existuje. Podme sa teda pozrieť na to, ako prepísať tento algoritmus na rekurzívny algoritmus spĺňajúci reštrikcie (2.2). Najprv si pre zjednodušenie zavedieme nasledujúce označenie

$$\bar{R}(t) = \sum_{k=1}^t \alpha_k \mathbf{x}(k) \mathbf{x}^T(k). \quad (2.5)$$

Potom pomocou vyjadrenia (2.4) pre $\hat{\boldsymbol{\beta}}(t-1)$, malou úpravou a použitím práve zadaného označenia dostaneme:

$$\sum_{k=1}^{t-1} \alpha_k \mathbf{x}(k) Y(k) = \sum_{k=1}^{t-1} \alpha_k \mathbf{x}(k) \mathbf{x}^T(k) \hat{\boldsymbol{\beta}}(t-1).$$

a následne

$$\sum_{k=1}^{t-1} \alpha_k \mathbf{x}(k) Y(k) = \bar{R}(t-1) \hat{\boldsymbol{\beta}}(t-1). \quad (2.6)$$

Z definície $R(t)$ taktiež vyplýva, že:

$$\bar{R}(t-1) = \bar{R}(t) - \alpha_t \mathbf{x}(t) \mathbf{x}^T(t). \quad (2.7)$$

To môžeme jednoducho vidieť z nasledujúcej úpravy:

$$\bar{R}(t-1) - \bar{R}(t) = \sum_{k=1}^{t-1} \alpha_k \mathbf{x}(k) \mathbf{x}^T(k) - \sum_{k=1}^t \alpha_k \mathbf{x}(k) \mathbf{x}^T(k) = -\alpha_t \mathbf{x}(t) \mathbf{x}^T(t).$$

Prepíšme si ešte (2.6) na sumu do t a vyjadrenie prenásobme $\bar{R}^{-1}(t)$ zľava. Z čoho následne úpravami dostaneme:

$$\begin{aligned} \hat{\boldsymbol{\beta}}(t) &= \bar{R}^{-1}(t) \sum_{k=1}^t \alpha_k \mathbf{x}(k) Y(k) \\ &= \bar{R}^{-1}(t) \left[\sum_{k=1}^{t-1} \alpha_k \mathbf{x}(k) Y(k) + \alpha_t \mathbf{x}(t) Y(t) \right] \\ &= \bar{R}^{-1}(t) [\bar{R}(t-1) \hat{\boldsymbol{\beta}}(t-1) + \alpha_t \mathbf{x}(t) Y(t)] \\ &= \bar{R}^{-1}(t) \{ \bar{R}(t) \hat{\boldsymbol{\beta}}(t-1) + \alpha_t \mathbf{x}(t) [-\mathbf{x}^T(t) \hat{\boldsymbol{\beta}}(t-1) + Y(t)] \} \\ &= \hat{\boldsymbol{\beta}}(t-1) + \bar{R}^{-1}(t) \mathbf{x}(t) \alpha_t [Y(t) - \hat{\boldsymbol{\beta}}^T(t-1) \mathbf{x}(t)], \end{aligned} \quad (2.8)$$

kde druhá rovnosť je len rozpísanie sumy, tretia rovnosť je dosadenie vyjadrenia (2.6), v štvrtej rovnosti sme použili rozpísanie pomocou (2.7) a posledný krok spočíva v zjednodušení a miernom preusporiadaní výrazu. Prehodením strán vyjadrenia (2.7) dostaneme nasledujúci ekvivalentný výraz:

$$\bar{R}(t) = \bar{R}(t-1) + \alpha_t \mathbf{x}(t) \mathbf{x}^T(t). \quad (2.9)$$

Ak si zdefinujeme ešte označenie $R(t)$ ako

$$R(t) = \frac{1}{t} \bar{R}(t).$$

Pomocou tohoto vyjadrenia pre $R(t)$ aplikovaného na (2.9) a jednoduchou úpravou a dosadením získame:

$$\begin{aligned} R(t) &= \frac{1}{t} \bar{R}(t) = \frac{1}{t} [\bar{R}(t-1) + \alpha_t \mathbf{x}(t) \mathbf{x}^T(t)] \\ &= R(t-1) + \frac{1}{t} [\alpha_t \mathbf{x}(t) \mathbf{x}^T(t) - R(t-1)] \end{aligned} \quad (2.10)$$

Ak sa pozrieme na koncové vyjadrenia v (2.8) a (2.10) môžeme si všimnúť, že obe vyjadrenia v čase t potrebujú len hodnoty, ktoré sú z prechádzajúceho a súčasného kroku. Tieto výrazy vieme zhrnúť takýmto spôsobom:

$$\hat{\beta}(t) = \hat{\beta}(t-1) + \frac{1}{t}R^{-1}(t)\mathbf{x}(t)\alpha_t[Y(t) - \beta^T(t-1)\mathbf{x}(t)] \quad (2.11a)$$

$$R(t) = R(t-1) + \frac{1}{t}[\alpha_t\mathbf{x}(t)\mathbf{x}^T(t) - R(t-1)], \quad (2.11b)$$

kde vidíme, že v čase t potrebujeme mať uložené len hodnoty $\hat{\beta}(t), R(t), Y(t)$ a $\mathbf{x}(t)$, čím tento tvar rovníc naberá podobu rekurzívneho algoritmu. Pri akvizícii ďalších dát a teda nových pozorovaní nám postačia tieto hodnoty pre aktualizáciu odhadu. Vyjadrenia spĺňajú dané reštrikcie a potrebné dáta sú uložené, preto všetky prechádzajúce dáta môžeme zabudnúť.

Problém naďalej spočíva v tom, že rovnice (2.11), ktoré už máme v tvare rekurzívneho algoritmu sú pomerne náročné na počítanie, keďže v každom kroku je potrebné spočítať inverz matice R , čomu sa chceme, ak je to možné určite vyhnúť. Hlavným dôvodom je časová náročnosť invertovania matíc pri každom príchode nových pozorovaní. Existuje možnosť, ako sa danému problému vyhnúť a to pomocou zavedenia matice P , ktorú bude možné aktualizovať priamo a bez potreby maticového násobenia, čo by výpočty pomerne zjednodušilo. Definujme si maticu P takýmto spôsobom:

$$P(t) = \bar{R}^{-1}(t) = \frac{1}{t}R^{-1}(t).$$

To ako nám zadefinovanie tejto matice pomôže a ako to bude fungovať si hneď ukážeme, ale k tomu je potreba si najprv zaviesť lemma o maticovej inverzii.

Lemma 2 (Lemma o maticovej inverzii). *Nech A, B, C a D sú matice, ktoré majú vhodné a kompatibilné dimenzie. To znamená, že násobok BCD a sčítanec $A + BCD$ existujú a taktiež nech existujú príslušné inverzy A, B, C a D . Potom platí nasledujúce:*

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[DA^{-1}B + C^{-1}]^{-1}DA^{-1} \quad (2.12)$$

Dôkaz. Dôkaz si ukážeme jednoducho a to tak, že prenásobíme údajný inverz matice samotnou maticou, čím overíme, že je lemma pravdivé, ak bude výsledkom jednotková matica. To vychádza jednoducho z toho, že súčin matice a jej inverznej matice je rovné jednotkovej matici. Prenásobením pravej strany (2.12) sumou $A + BCD$ sprava dostávame výraz:

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

čím sme dokázali (2.12). □

So zavedením lemma o maticovej inverzii (2.12) a jeho aplikáciou v treťom riadku nasledujúceho odvodenia s týmto preznačením

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

a následnou úpravou dostávame

$$\begin{aligned} P(t) &= \bar{R}^{-1}(t) = (\bar{R}(t-1) + \alpha_t \mathbf{x}(t) \mathbf{x}^T(t))^{-1} & (2.13) \\ &= [P^{-1}(t-1) + \mathbf{x}(t) \alpha_t \mathbf{x}^T(t)]^{-1} \\ &= P(t-1) - P(t-1) \mathbf{x}(t) \left[\mathbf{x}^T(t) P(t-1) \mathbf{x}(t) + \frac{1}{\alpha_t} \right]^{-1} \mathbf{x}^T(t) P(t-1) \\ &= P(t-1) - \frac{P(t-1) \mathbf{x}(t) \mathbf{x}^T(t) P(t-1)}{1/\alpha_t + \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)}. \end{aligned}$$

Ako už vieme tak výhodou konečného(2.13) je to, že nie je potrebné v každom kroku invertovať maticu veľkosti $\dim \beta$, ale postačí nám invertovanie skaláru. Označenie \dim predstavuje dimenziu nejakého vektorového priestoru, ktorá je rovná počtu prvkov báze daného vektorového priestoru. V tomto prípade je vektorový priestor matica R , ktorej inverzii v každom kroku rekurzcie sa snažíme vyhnúť.

Je to v podstate iba určitá aktualizácia P . Taktiež sa pozrieme na to ako vyzerá $\alpha_t P(t) \mathbf{x}(t)$, kde za pomoci (2.13) dostávame, že:

$$\begin{aligned} \alpha_t P(t) \mathbf{x}(t) &= \alpha_t \left[P(t-1) - \frac{P(t-1) \mathbf{x}(t) \mathbf{x}^T(t) P(t-1)}{1/\alpha_t + \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)} \right] \mathbf{x}(t) & (2.14) \\ &= \alpha_t P(t-1) \mathbf{x}(t) - \frac{\alpha_t P(t-1) \mathbf{x}(t) \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)}{1/\alpha_t + \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)} \\ &= \frac{\frac{1}{\alpha_t} \alpha_t P(t-1) \mathbf{x}(t)}{1/\alpha_t + \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)} + \frac{\alpha_t P(t-1) \mathbf{x}(t) \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)}{1/\alpha_t + \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)} \\ &\quad - \frac{\alpha_t P(t-1) \mathbf{x}(t) \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)}{1/\alpha_t + \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)} \\ &= \frac{P(t-1) \mathbf{x}(t)}{1/\alpha_t + \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)}. \end{aligned}$$

Tento člen sa nám zíde pre lepší zápis (2.11), keďže práve člen (2.14) v ňom vystupuje. Odhad metódou najmenších štvorcov $\hat{\beta}(t)$, ktorý sme definovali v (2.4) môže byť teda rekurzívne spočítaný. Ak využijeme odvodenia, ktoré sme si teraz odvodili, je možné odhad rekurzívne spočítať prostredníctvom týchto vzorcov:

$$\hat{\beta}(t) = \hat{\beta}(t-1) + L(t)[Y(t) - \hat{\beta}^T(t-1) \mathbf{x}(t)], \quad (2.15a)$$

$$L(t) = \frac{P(t-1) \mathbf{x}(t)}{1/\alpha_t + \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)}, \quad (2.15b)$$

$$P(t) = P(t-1) - \frac{P(t-1) \mathbf{x}(t) \mathbf{x}^T(t) P(t-1)}{1/\alpha_t + \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)}. \quad (2.15c)$$

Tieto vzorce definujú algoritmus, ktorý je známy pod názvom rekurentná metóda najmenších štvorcov. Z (2.8) a (2.13) môžeme vidieť odvodenie pre $\hat{\beta}(t)$ a zároveň pre $P(t)$, $L(t)$ sme si odvodili v (2.14). Kvôli jej výhodám je tento algoritmus jednou z najviac používaných rekurzívnych identifikačných metód. Ďalej v práci budeme často používať skratku RLS odhad, ktorá označuje odhad rekurentnou metódou najmenších štvorcov a taktiež OLS odhad teda odhad klasickou metódou najmenších štvorcov. Musíme sa ešte povenovať tomu ako je algoritmus a rekurgia spustená. Potrebujeme mať určené počiatkové podmienky pre začatie rekurzcie. Problematikou inicializácie algoritmu sa budeme zaoberať konkrétne v nasledujúcej sekcii 2.4.1, kde sa pozrieme aj na niektoré z vlastností rekurentného odhadu.

2.4 Počiatkové podmienky algoritmu a vybrané vlastnosti odhadov

2.4.1 Počiatkové podmienky

Ako každý rekurzívny algoritmus aj rekurzívna metóda najmenších štvorcov, ktorú sme predstavili v podkapitole 2.3, vyžaduje typicky jeden alebo viac počiatkových hodnôt. Pri rekurzívnej metóde najmenších štvorcov definovaných ako algoritmus (2.15) potrebujeme hodnoty $\hat{\beta}(0)$ a $P(0)$. Tieto hodnoty sú nevyhnutné pre definovanie počiatkového stavu systému, či na určenie počiatkových podmienok pre rekurzívny proces. V algoritme (2.11) sú pre spustenie nevyhnutné taktiež dve hodnoty a to $\hat{\beta}(0)$ a $R(0)$. Ak sa znovu pozrieme na vyjadrenia z prechádzajúcej kapitoly, vidíme, že sme (2.15) odvodili z (2.4) za predpokladu, že $\bar{R}(t) = \sum_{k=1}^t \alpha_k \mathbf{x}(k) \mathbf{x}^T(k)$ je možné invertovať. Z toho jednoducho vyplýva to, že spustenie algoritmu (2.15) je možné len vtedy, ak je započatý v čase t_0 , kde t_0 je hodnota taká, že $\bar{R}(t_0)$ je invertovateľná. Typickým prípadom je, že $R(t)$ možné invertovať až od času $t_0 = \dim \mathbf{x}(t) = \dim \beta$. Vhodné spúšťacie hodnoty pre algoritmus (2.15) sú také, že začneme rekurziu v čase t_0 , pre hodnoty

$$P(t_0) = \left[\sum_{k=1}^{t_0} \alpha_k \mathbf{x}(k) \mathbf{x}^T(k) \right]^{-1}, \quad (2.16a)$$

$$\hat{\beta}(t_0) = P(t_0) \sum_{k=1}^{t_0} \alpha_k \mathbf{x}(k) Y(k). \quad (2.16b)$$

Nevýhodou tohoto prístupu je práve to, že je algoritmus možné spustiť až v čase t_0 . Vo veľa prípadoch, ale je potrebné a výhodné začať rekurziu už od prvého získaného pozorovania. Bežnejšie je preto začať rekurziu v čase $t = 0$ s nejakou invertovateľnou maticou $P(0)$ a vektorom $\hat{\beta}(0)$. Odhady vychádzajúce z (2.15) pomocou tohoto prístupu potom vyzerajú takto:

$$\hat{\beta}(t) = \left[P^{-1}(0) + \sum_{k=1}^t \alpha_k \mathbf{x}(k) \mathbf{x}^T(k) \right]^{-1} \left[P^{-1}(0) \hat{\beta}(0) + \sum_{k=1}^t \alpha_k \mathbf{x}(k) Y(k) \right], \quad (2.17)$$

čo dostaneme z toho, ak overíme, že (2.17) spĺňa rekurziu (2.15) s vyššie spomenutými počiatkovými podmienkami. Ak porovnáme (2.17) s (2.4) môžeme si

všimnúť, že dôležitosť počiatočných hodnôt s časom upadá s tým ako narastajú dané sumy. Táto možnosť nám pomôže hlavne v tom, že spustenie algoritmu bude možné skôr ako v čase t_0 , bez potreby mať najprv zozbieraných t_0 pozorovaní pre vytvorenie (2.16). Taktiež ako $P^{-1}(0) \rightarrow \mathbf{0}$ tak sa rovnako aj rekurzívne odhady blížia k off-line odhadom. To môžeme chápať tak, že pokiaľ sa jednotlivé prvky matice $P^{-1}(0)$ limitne približujú k nule tak sa RLS odhady približujú OLS odhadom. Pri alternatívnom spôsobe s voľbou invertovateľnej matice a vhodného vektora sa využíva voľba pre štart algoritmu tak, že volíme tieto hodnoty ako $P(0) = c \cdot I$, kde $c > 0$ je nejaká veľká konštanta a I je jednotková matica vhodnej dimenzie. Odhad vektoru β pre spustenie algoritmu sa volí ako nulový vektor, teda $\hat{\beta}(0) = 0$.

2.4.2 Asymptotické vlastnosti

V tejto podsekcii sa pozrieme na niektoré z vlastností odhadu vytvoreného pomocou rekurentnej metódy najmenších štvorcov. Predpokladajme teda v tejto časti, že dáta boli generované pomocou modelu

$$Y(t) = \beta_0^T \mathbf{x}(t) + \epsilon(t), \quad (2.18)$$

kde $\epsilon(t)$ sú nezávislé rovnako rozdelené náhodné veličiny s nulovou strednou hodnotou a konečným rozptylom. Budeme skúmať ako sa odhad (2.4) správa so zvyšujúcim sa N (počtom pozorovaní). Použitím vyjadrenia (2.18) pre $Y(t)$ v (2.4) dostávame celkovo:

$$\begin{aligned} \hat{\beta}(N) &= \left[\sum_{t=1}^N \alpha_t \mathbf{x}(t) \mathbf{x}^T(t) \right]^{-1} \sum_{t=1}^N \alpha_t \mathbf{x}(t) Y(t) \\ &= \left[\sum_{t=1}^N \alpha_t \mathbf{x}(t) \mathbf{x}^T(t) \right]^{-1} \sum_{t=1}^N \alpha_t \mathbf{x}(t) [\mathbf{x}^T(t) \beta_0 + \epsilon(t)] \\ &= \left[\sum_{t=1}^N \alpha_t \mathbf{x}(t) \mathbf{x}^T(t) \right]^{-1} \left[\sum_{t=1}^N \alpha_t [\mathbf{x}(t) \mathbf{x}^T(t) \beta_0 + \mathbf{x}(t) \epsilon(t)] \right] \\ &= \beta_0 + \left[\sum_{t=1}^N \alpha_t \mathbf{x}(t) \mathbf{x}^T(t) \right]^{-1} \sum_{t=1}^N \alpha_t \mathbf{x}(t) \epsilon(t) \\ &= \beta_0 + \left[\frac{1}{N} \sum_{t=1}^N \alpha_t \mathbf{x}(t) \mathbf{x}^T(t) \right]^{-1} \frac{1}{N} \sum_{t=1}^N \alpha_t \mathbf{x}(t) \epsilon(t). \end{aligned} \quad (2.19)$$

Vo vyjadrení sme použili jednoduché prepísanie a to konkrétne, že $\beta_0^T \mathbf{x}(t) = \mathbf{x}^T(t) \beta_0$, keďže oba vektory sú rovnakej veľkosti. Z konečného vyjadrenia (2.19) môžeme pozorovať niektoré vlastnosti odhadu $\hat{\beta}(N)$. Konkrétne sú teda takéto:

- odhad $\hat{\beta}(N)$ je dostatočne blízko β_0
- odhad konverguje k β_0 s N idúcim do nekonečna.

Ak je chybový člen $\epsilon(t)$ v (2.18) malý v porovnaní s $\mathbf{x}(t)$ môžeme vidieť, že $\hat{\beta}(N)$ bude blízko β_0 . Podľa zákona veľkých čísel bude suma $(1/N) \sum_{t=1}^N \alpha_t \mathbf{x}(t) \epsilon(t)$ konvergovať k jej strednej hodnote s N idúcim do nekonečna. Spomenutá stredná hodnota závisí na korelácii medzi chybovým členom $\epsilon(t)$ a vektorom dát $\mathbf{x}(t)$.

Stredná hodnota je nulová jedine v tom prípade, ak $\epsilon(t)$ a $\mathbf{x}(t)$ sú nekorelované. To nastáva napríklad v tom prípade, ak je postupnosť $\{\epsilon(t)\}$ postupnosťou nekorelovaných náhodných veličín s nulovou strednou hodnotou a konštantným kladným rozptylom. Takto definovaná postupnosť je taktiež nazývaná biely šum. Potom $\epsilon(t)$ nie je závislá s ničím, čo sa stalo do času $t - 1$ a preto platí, že $E \epsilon(t)\mathbf{x}(t)$. Nekorelovanosť $\mathbf{x}(t)$ a $\epsilon(t)$ ako vieme je jedným z predpokladov klasickej regresie. Pre N idúce do nekonečna teda rozdiel OLS odhadu a RLS odhadu ide k nule, čo si ukážeme neskôr v tejto práci na simulačnej štúdii. RLS odhad sa od OLS odhadu líši hlavne v tom, že ho je možné spočítať rekurzívne a vyžaduje počiatočné podmienky. Okrem toho sa asymptotické vlastnosti týchto dvoch odhadov zhodujú, práve kvôli tomu, že RLS algoritmus je odvodený z klasickej metódy najmenších štvorcov. Ostatné tieto asymptotické vlastnosti sú ukázané v kapitolách 8 a 9 v knihe Ljung (1999).

2.5 Modifikácie

V prvej kapitole sme si uviedli koncept regresie od ktorého sme sa postupne prepracovali k lineárnej regresii a následne odhadu pomocou metódy najmenších štvorcov. Neskôr v motivácii pri rekurentnej metóde najmenších štvorcov sme si ukázali ako dôležité je použitie rekurzívnej identifikácie v praxi a ako podstatné je mať modely, ktorých dynamika sa mení v čase. My sme si ukázali ako z off-line modelu, ktorý je konštantný (nemenný) sa vieme dostať k adaptívnemu modelu. V tejto časti skúsime adaptívny a dynamický systém, ktorým rekurzívna metóda najmenších štvorcov je, modifikovať tak, aby model ešte lepšie reagoval na zmeny, ktoré sa s príchodom nových pozorovaní dejú.

Jeden spôsob ako to je možné dosiahnuť je pomocou takzvanej diskontácie pozorovaní. Konkrétne sa táto modifikácia nazýva rekurzívna metóda najmenších štvorcov s diskontovanými meraniami. Budeme sa v podstate snažiť zachytiť trend vývoja modelu cez postupné vypúšťanie starších pozorovaní, ktoré sú menej relevantné. Táto postupnosť bude v niečom podobná ako postupnosť váh $\alpha(t)$, ktorú sme si zaviedli už na začiatku práce. Bude to v princípe len vylepšenie tejto postupnosti pre naše potreby. Pri klasickej metóde najmenších štvorcov vieme, že je potrebné minimalizovať (2.3), teda sumu

$$V_N(\boldsymbol{\beta}) = \frac{1}{N} \sum_{t=1}^N \alpha_t [Y(t) - \boldsymbol{\beta}^T \mathbf{x}(t)]^2. \quad (2.20)$$

Túto sumu je následne minimalizovaná cez vektor $\boldsymbol{\beta}$, aby sme sa dostali k vyjadreniu odhadu daného vektora. Ak sa na vyjadrenie kritéria (2.20) dobre pozrieme môžeme vidieť, že je to v podstate len priemerné správanie odhadu počas času $1 \leq t \leq N$. V niektorých aplikáciach, kde je potrebné do modelu zahrnúť všetky pozorovania rovnomerne je táto vlastnosť využiteľná. Ak, ale máme pomerne rýchlo meniace sa dáta, ktorých trend chceme modelom zachytiť mali by sme uvažovať nad vytvorením odhadu, ktorý viac reprezentuje momentálne vlastnosti systému v reálnom čase, kedy príslušný odhad vytvárame. V tomto prípade by bolo pre nás logické uvažovať diskontovanie starších pozorovaní. To dosiahneme

takýmto spôsobom:

$$\tilde{V}_N(\boldsymbol{\beta}) = \sum_{t=1}^N \bar{\gamma}(N,t)[Y(t) - \boldsymbol{\beta}^T \mathbf{x}(t)]^2, \quad (2.21)$$

kde $\bar{\gamma}(N,t)$ sa typicky zväčšuje s t pre dané N . Pozorovania bližšie k času N majú väčší koeficient ako pozorovania, ktoré sme namerali na začiatku spustenia algoritmu. Keďže $\tilde{V}_N(\boldsymbol{\beta})$ je stále kvadratická v $\boldsymbol{\beta}$ postup pri minimalizácii tohoto výrazu je totožný s minimalizáciou (2.20) ešte z podkapitoly 1.3. Tým pádom vieme odhad $\hat{\boldsymbol{\beta}}$ spočítať nasledovne:

$$\hat{\boldsymbol{\beta}}(N) = \left[\sum_{t=1}^N \bar{\gamma}(N,t) \mathbf{x}(t) \mathbf{x}^T(t) \right]^{-1} \left[\sum_{t=1}^N \bar{\gamma}(N,t) \mathbf{x}(t) Y(t) \right]. \quad (2.22)$$

Toto vyjadrenie predstavuje východisko k výpočtu off-line odhadu daného vektora s použitím $\bar{\gamma}(N,t)$. Ak chceme postupnosť odhadov počítat rekurzívne za pomoci vyššie uvedeného vyjadrenia pre off-line odhady musíme uvažovať, že aj $\bar{\gamma}(N,t)$ je v určitom tvare. Odhady je možné spočítat rekurzívne, ak zavedieme $\bar{\gamma}(t,k)$ v špeciálnom tvare, aby platilo, že tento koeficient je závislý na poradí daného pozorovania v čase. Predpokladajme preto, že tento koeficient je možné vyjadriť v nasledujúcom tvare:

$$\bar{\gamma}(t,k) = \lambda(t) \bar{\gamma}(t-1,k), \quad 1 \leq k \leq t-1. \quad (2.23a)$$

Čo vieme prepísať ako:

$$\begin{aligned} \bar{\gamma}(t,k) &= \lambda(t) \bar{\gamma}(t-1,k) = \lambda(t) \lambda(t-1) \bar{\gamma}(t-2,k) \\ &= \left[\prod_{j=k+1}^t \lambda(j) \right] \bar{\gamma}(k,k) = \left[\prod_{j=k+1}^t \lambda(j) \right] \alpha_k, \end{aligned} \quad (2.23b)$$

kde vidíme, že platí rovnosť $\bar{\gamma}(k,k) = \alpha_k$ a to pri zohľadnení konvencie, že $\prod_{j=k+1}^t \lambda(j) = 1$. Zvyčajne je taktiež dané, že $\lambda(k) \leq 1$. Ak $\lambda(k) = \lambda$ pre všetky k , tak to znamená jedine to, že tento koeficient je konštantný pre všetky t a preto z (2.23b) dostaneme:

$$\bar{\gamma}(t,k) = \lambda^{t-k} \alpha_k, \quad (2.24)$$

čo pri použití v (2.21) dáva danému vyjadreniu takzvaný exponenciálne zabúdajúci tvar. Takže v tomto tvare sú počiatkové pozorovania postupne vypúšťané a to tak, že k nim je priradený menší koeficient. To sme chceli touto modifikáciou dosiahnuť. V takom prípade, kde platí (2.24) nazývame λ ako *forgetting factor*. Zavedieme si znovu rovnaké označenie ako (2.5), ale v tomto prípade pre (2.22) ako:

$$\bar{R}(t) = \sum_{k=1}^t \bar{\gamma}(t,k) \mathbf{x}(k) \mathbf{x}^T(k). \quad (2.25)$$

Potom platí

$$\begin{aligned}
\bar{R}(t) &= \sum_{k=1}^t \bar{\gamma}(t,k) \mathbf{x}(k) \mathbf{x}^T(k) \\
&= \sum_{k=1}^{t-1} \bar{\gamma}(t,k) \mathbf{x}(k) \mathbf{x}^T(k) + \bar{\gamma}(t,t) \mathbf{x}(t) \mathbf{x}^T(t) \\
&= \lambda(t) \sum_{k=1}^{t-1} \bar{\gamma}(t-1,k) \mathbf{x}(k) \mathbf{x}^T(k) + \alpha_t \mathbf{x}(t) \mathbf{x}^T(t) \\
&= \lambda(t) \bar{R}(t-1) + \alpha_t \mathbf{x}(t) \mathbf{x}^T(t).
\end{aligned} \tag{2.26}$$

Teraz budeme postupovať rovnako ako v (2.8), ale použijeme nové vyjadrenie (2.26) pre $\bar{R}(t)$, z čoho následne dostaneme:

$$\begin{aligned}
\hat{\beta}(t) &= \bar{R}^{-1}(t) \left[\sum_{k=1}^t \bar{\gamma}(t,k) \mathbf{x}(k) Y(k) \right] \\
&= \bar{R}^{-1}(t) \left[\sum_{k=1}^{t-1} \bar{\gamma}(t,k) \mathbf{x}(k) Y(k) + \bar{\gamma}(t,t) \mathbf{x}(t) Y(t) \right] \\
&= \bar{R}^{-1}(t) \left[\lambda(t) \sum_{k=1}^{t-1} \bar{\gamma}(t-1,k) \mathbf{x}(k) Y(k) + \alpha_t \mathbf{x}(t) Y(t) \right] \\
&= \bar{R}^{-1}(t) [\lambda(t) \bar{R}(t-1) \hat{\beta}(t-1) + \alpha_t \mathbf{x}(t) Y(t)] \\
&= \bar{R}^{-1}(t) \{ [\bar{R}(t) - \alpha_t \mathbf{x}(t) \mathbf{x}^T(t)] \hat{\beta}(t-1) + \alpha_t \mathbf{x}(t) Y(t) \} \\
&= \bar{R}^{-1}(t) \{ \bar{R}(t) \hat{\beta}(t-1) + \alpha_t \mathbf{x}(t) [Y(t) - \mathbf{x}^T(t) \hat{\beta}(t-1)] \} \\
&= \hat{\beta}(t-1) + \bar{R}^{-1}(t) \mathbf{x}(t) \alpha_t [Y(t) - \mathbf{x}^T(t) \hat{\beta}(t-1)].
\end{aligned}$$

Dostali sme sa presne k tomu istému finálnemu vyjadreniu ako v postupe (2.8). Takže odhady $\hat{\beta}(t)$ s, ale aj bez zavedenia forgetting factoru majú zhodné aktualizáčnè vyjadrenia. Aj v tomto prípade sa ale potrebujeme vyhnúť invertovaniu matice $\bar{R}(t)$ v každom kroku rekurzcie. Použijeme k tomu znovu označenie

$$P(t) = \bar{R}^{-1}(t)$$

a budeme postupovať rovnako ako. Aplikujeme teda lemma 2 na (2.25). Z čoho dostávame

$$\begin{aligned}
P(t) &= \bar{R}^{-1}(t) = [\lambda(t) \bar{R}(t-1) + \alpha_t \mathbf{x}(t) \mathbf{x}^T(t)]^{-1} \\
&= \frac{1}{\lambda(t)} P(t-1) - \frac{1}{\lambda(t)} P(t-1) \mathbf{x}(t) \left[\mathbf{x}^T(t) \frac{1}{\lambda(t)} P(t-1) \mathbf{x}(t) + \frac{1}{\alpha_t} \right]^{-1} \\
&\quad \times \mathbf{x}^T(t) P(t-1) \frac{1}{\lambda(t)} \\
&= \frac{1}{\lambda(t)} P(t-1) - \frac{\frac{1}{\lambda(t)} P(t-1) \mathbf{x}(t) \mathbf{x}^T(t) P(t-1) \frac{1}{\lambda(t)}}{\mathbf{x}^T(t) \frac{1}{\lambda(t)} P(t-1) \mathbf{x}(t) + \frac{1}{\alpha_t}} \\
&= \frac{1}{\lambda(t)} P(t-1) - \frac{P(t-1) \mathbf{x}(t) \mathbf{x}^T(t) P(t-1)}{\lambda(t) \mathbf{x}^T(t) P(t-1) \mathbf{x}(t) + [\lambda^2(t)/\alpha_t]} \\
&= \frac{1}{\lambda(t)} \left[P(t-1) - \frac{P(t-1) \mathbf{x}(t) \mathbf{x}^T(t) P(t-1)}{[\lambda(t)/\alpha_t] + \mathbf{x}^T(t) P(t-1) \mathbf{x}(t)} \right].
\end{aligned}$$

V prvom riadku sme si len prepísali vyjadrenie pomocou (2.26), na ktorý sme následne aplikovali lemma 2 a potom už len jednoducho upravili. Týmto postupom sme sa teda dostali k tomuto algoritmu:

$$\hat{\beta}(t) = \hat{\beta}(t-1) + L(t)[Y(t) - \mathbf{x}^T(t)\hat{\beta}(t-1)], \quad (2.27a)$$

$$L(t) = \frac{P(t-1)\mathbf{x}(t)}{[\lambda(t)/\alpha_t] + \mathbf{x}^T(t)P(t-1)\mathbf{x}(t)} = \alpha_t P(t)\mathbf{x}(t), \quad (2.27b)$$

$$P(t) = \frac{1}{\lambda(t)} \left[P(t-1) - \frac{P(t-1)\mathbf{x}(t)\mathbf{x}^T(t)P(t-1)}{[\lambda(t)/\alpha_t] + \mathbf{x}^T(t)P(t-1)\mathbf{x}(t)} \right], \quad (2.27c)$$

kde $\hat{\beta}(t)$ a $P(t)$ sme si práve odvodili. Ukážme si ešte ako sme sa dostali k vyjadreniu $L(t)$:

$$\begin{aligned} L(t) &= \alpha_t P(t)\mathbf{x}(t) = \alpha_t \left[\frac{1}{\lambda(t)} \left[P(t-1) - \frac{P(t-1)\mathbf{x}(t)\mathbf{x}^T(t)P(t-1)}{[\lambda(t)/\alpha_t] + \mathbf{x}^T(t)P(t-1)\mathbf{x}(t)} \right] \right] \mathbf{x}(t) \\ &= \alpha_t \left[\frac{1}{\lambda(t)} P(t-1) - \frac{P(t-1)\mathbf{x}(t)\mathbf{x}^T(t)P(t-1)}{\lambda(t)\mathbf{x}^T(t)P(t-1)\mathbf{x}(t) + [\lambda^2(t)/\alpha_t]} \right] \mathbf{x}(t) \\ &= \alpha_t \left[\frac{P(t-1)\mathbf{x}^T(t)\mathbf{x}(t)P(t-1) + \frac{\lambda(t)P(t-1)}{\alpha_t}}{\lambda(t)\mathbf{x}^T(t)P(t-1)\mathbf{x}(t) + [\lambda^2(t)/\alpha_t]} \right. \\ &\quad \left. - \frac{P(t-1)\mathbf{x}(t)\mathbf{x}^T(t)P(t-1)}{\lambda(t)\mathbf{x}^T(t)P(t-1)\mathbf{x}(t) + [\lambda^2(t)/\alpha_t]} \right] \mathbf{x}(t) \\ &= \alpha_t \left[\frac{\frac{\lambda(t)P(t-1)}{\alpha_t}}{\lambda(t)\mathbf{x}^T(t)P(t-1)\mathbf{x}(t) + [\lambda^2(t)/\alpha_t]} \right] \mathbf{x}(t) \\ &= \frac{P(t-1)\mathbf{x}(t)}{[\lambda(t)/\alpha_t] + \mathbf{x}^T(t)P(t-1)\mathbf{x}(t)} \end{aligned}$$

Ukázali sme sa si ako sa postupne dostať, ku všetkým potrebným častiam rekurzívneho algoritmu so zavedením forgetting factoru, čoho dôsledkom dostávame, že (2.21) môže byť rekurzívne minimalizované pomocou algoritmu (2.27), ktorý sme si v tejto podkapitole odvodili. Algoritmus musí spĺňať požiadavky stanovené pre forgetting factor a teda musí podliehať dvojici vyjadrení (2.23). Ak sa znovu pozrieme na obidva tvary rekurzívnych identifikačných algoritmov, môžeme si všimnúť, že obyčajný algoritmus rekurentnej metódy najmenších štvorcov (2.15) je špeciálnym prípadom modifikovaného (2.27), kde $\lambda(t) = 1$. Podstata a efekt forgetting factoru $\lambda(t)$ v (2.27c) je zjavne to, aby sme člen $P(t)$ a teda aj nárast $L(t)$ vedeli určitým spôsobom korigovať pre naše potreby vzhľadom ku charakteru dát. Ako už vieme tak zvyčajne platí, že $\lambda(k) \leq 1$ pre všetky k . Potom z (2.27) vidíme, že koeficient $\lambda(t)$ prirodzene zväčšuje $P(t)$ a rovnako aj $L(t)$. Ak $\lambda(t) \leq 1 - \delta$, $\delta > 0$, potom $P(t)$ nebude mať tendenciu ísť do nuly a algoritmus bude preto vždy pripravený zaznamenávať meniacu sa dynamiku systému a zmeny, ktoré s ňou prichádzajú.

3. Simulačné štúdie

Táto časť práce by mala poslúžiť k lepšiemu uchopeniu rekurentnej metódy najmenších štvorcov na zopár simuláciách z náhodne generovaných dát. Budeme sa pozerat' aj na odhady pomocou klasickej metódy najmenších štvorcov, keďže práve to je metóda z ktorej sme on-line adaptívny algoritmus odvodili. Simulácie taktiež využijeme na ukázanie niektorých vlastností, ktoré boli v práci prestavené a taktiež ako sa odhady adaptujú, aktualizujú a vyvíjajú s príchodom nových pozorovaní.

3.1 Simulácia: Vývoj RLS odhadu

V tejto simulácii sa pozrieme ako sa odhad pomocou rekurentnej metódy najmenších štvorcov postupne vyvíja s príchodom nových dát. Uvidíme ako sa odhady menia v závislosti na tom, ktoré pozorovania sú pri vytvorení odhadu dostupné. Za naše vstupné hodnoty budeme uvažovať 70 náhodných dát, ktoré budeme generovať z jednoduchého modelu $Y(t) = 2 + 5 \cdot t + \epsilon(t)$, kde t sú hodnoty $1, \dots, 70$ a ϵ budeme brať ako chybové členy generované normálnym rozdelením so strednou hodnotou 0, čo je dané z predpokladov chybového člena a smerodajnú odchýlku nastavíme na 25. Seed nastavíme na hodnotu 42. Pozrieme sa konkrétne na to ako sa odhady postupne menia po 15, 25, 45, 70 prijatých pozorovaniach, ako prekrývajú dostupné dáta a ako po príchode nových sa odhady adaptujú a vyvinú. Všetky vybrané odhady si vykreslíme do grafu a vypíšeme ich hodnoty aj do tabuľky spolu s odhadom vytvoreným pomocou klasickej metódy najmenších štvorcov aplikovanej na všetkých 70 pozorovaní.

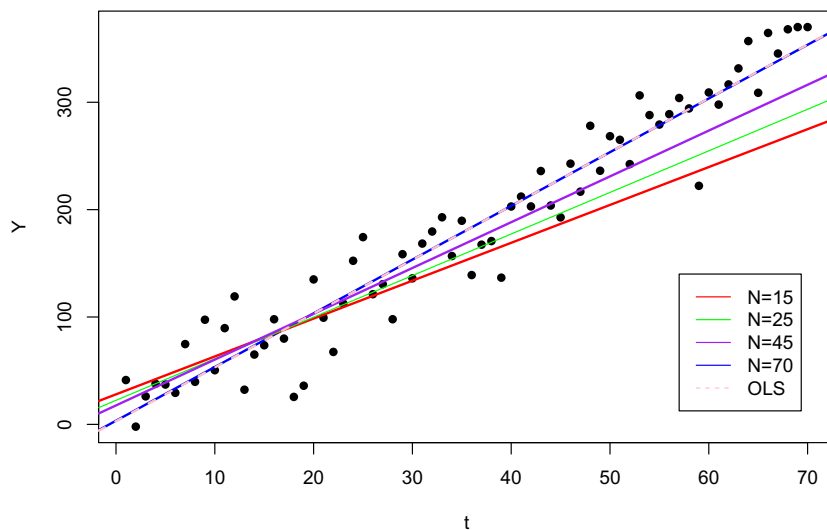
Vykreslíme si tiež grafy vývojov hodnôt a a b pri RLS odhadoch pre každé pozorovanie, na ktorých budeme môcť vidieť ich vývoj v každom rekurzívnom kroku. Ako inicializáciu $P(0)$ v simulačných štúdiách používame $10^7 \cdot I$, kde 10^7 je v tomto prípade konštanta c a I je jednotková matica 2×2 , čoho dôvod vyplýva zo sekcie 2.4.1. Chceme v tomto prípade začať rekurziu už po prvom pozorovaní, ale keďže potrebujeme, aby maticu R bolo možné invertovať, voľba ako v (2.16a) nedáva v tomto prípade zmysel. Tým pádom sa prikláňame k alternatívnej voľbe $P(0)$ ako $10^7 \cdot I$.

	RLS(N=15)	RLS(N=25)	RLS(N=45)	RLS(N=70)	OLS(N=70)
a	27.98288	22.38913	17.65541	3.403752	2.923
b	3.527962	3.873058	4.266276	5.00198	5.012

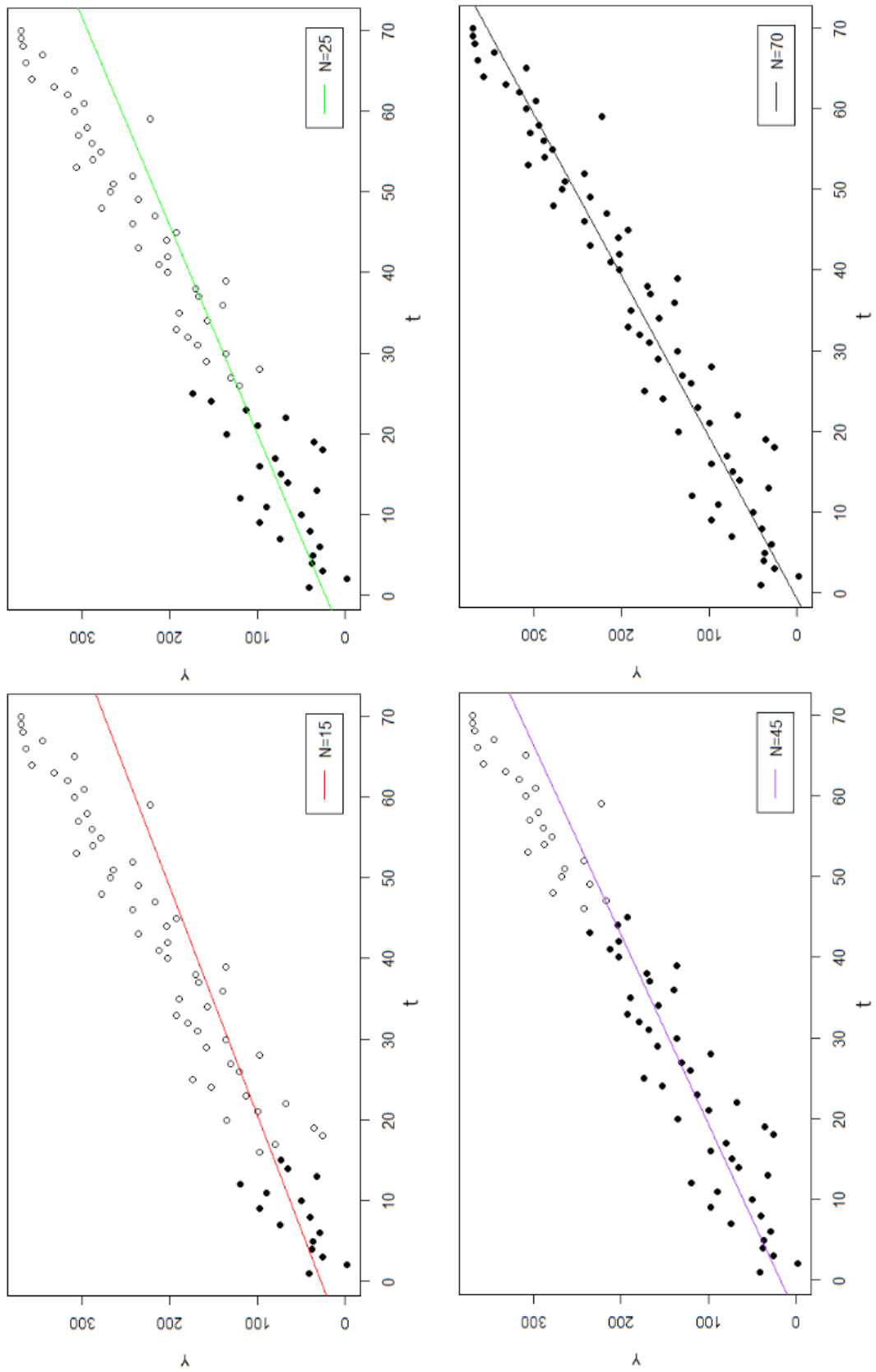
Tabuľka 3.1: Vývoj RLS odhadu s príchodom nových dát a porovnanie s OLS odhadom

V tabuľke 3.1 sme si zapísali hodnoty odhadov a vidíme ako sa rekurzívne odhady menia s príchodom nových dát a adaptujú sa na zmeny. V tabuľke sme tiež pridali stĺpec s OLS odhadom spočítaným zo všetkých 70 dát ku ktorému RLS odhady pomaly smerujú. Na obrázku 3.1 vidíme na grafe vykreslených všetkých 70 pozorovaní a aj ich preloženie pomocou odhadov vytvorených pre práve príslušné dostupné dáta a taktiež OLS odhad. Vidíme, že OLS odhad na obrázku 3.1 prekrýva RLS odhad pre $N = 70$, keďže tieto odhady sú už pomerne

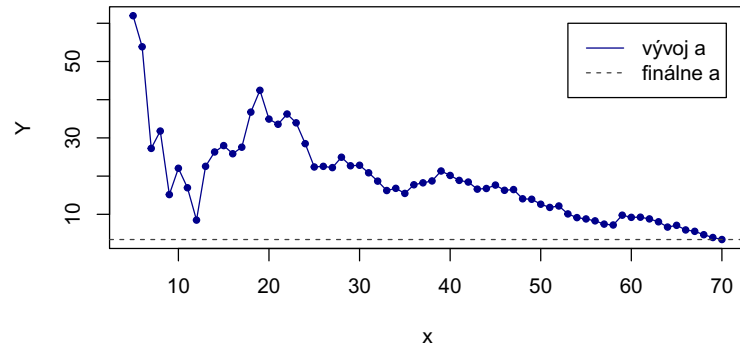
blízko pri sebe. Na obrázku 3.2 sme si ešte postupné odhady pre rôzne N vykreslili do samostatných grafov, kde sú odhady vytvorené z plne vyplnených bodov a zvyšné nevyplnené body ešte nie sú dostupné. Simulácia hlavne mierila na to, ukázať ako rozlične funguje výpočet pomocou klasickej metódy najmenších štvorcov a pomocou rekurzívneho algoritmu. Na obrázkoch 3.3 a 3.4 vidíme vývoje odhadov pre koeficienty a a b , kde vidíme peknú adaptáciu algoritmu na zmeny spôsobené prichádzajúcimi pozorovaniami a postupný prechod ku konečnému odhadu.



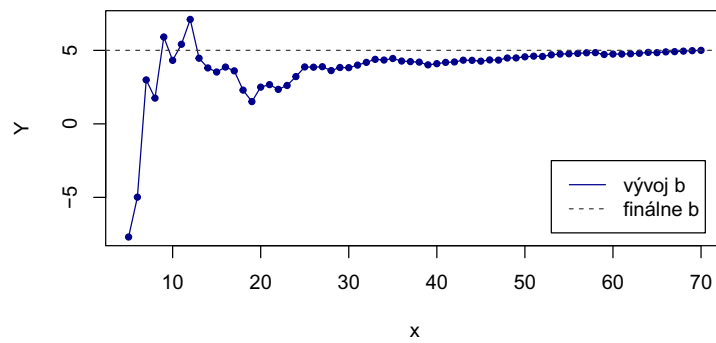
Obr. 3.1: Vykreslenie odhadov pre vybrané N a ich postupná aktualizácia



Obr. 3.2: Postupné odhady pre rôzne N , kde plné body znázorňujú aktuálne dostupné pozorovania pre RLS odhady



Obr. 3.3: Vývoj koeficientu a



Obr. 3.4: Vývoj koeficientu b

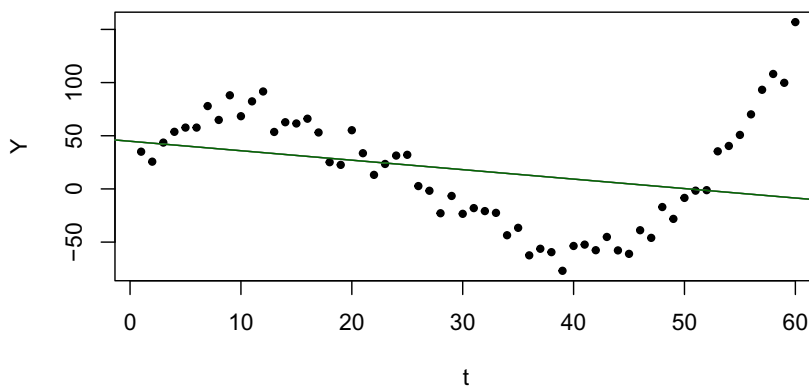
3.2 Simulácia: Preloženie dát polynómom

Rovnako ako klasickú metódu najmenších štvorcov tak aj rekurentnú metódu je možné použiť na prekladanie získaných dát nie len jednoduchou priamkou, ale aj ľubovoľným polynómom nejakého stupňa. V tejto časti si teda ukážeme ako je možné rekurentný algoritmus použiť pre odhad koeficientov pre prekrytie dát nejakým polynómom. Doteraz sme v každej simulácii kvôli jednoduchosti a v závislosti na dátach odhadovali len tvar jednoduchej priamky. Teraz sa od tohoto odkloníme a budeme generovať dáta pomocou zložitejšieho modelu, s pomocou ktorého sa dostaneme k pozorovaniam, pri ktorých už preloženie priamkou bude nepostačujúce. Budeme teda odhadovať koeficienty daného n – tého polynómu, ktorý sa bude najlepšie hodiť pre dané dáta. Taktiež sa rovnako ako v prvej simulačnej štúdii pozrieme na vývoj rekurzívneho odhadu v čase a jeho smerovanie k OLS odhadu, ale pre všetky 4 odhadované koeficienty.

Vygenerujeme si teda náhodné dáta a vytvoríme odhady príslušného polynómu, ktorý postupne dostupné dáta čo najlepšie vystihuje a prekladá. Dáta v tomto prípade budeme generovať z modelu:

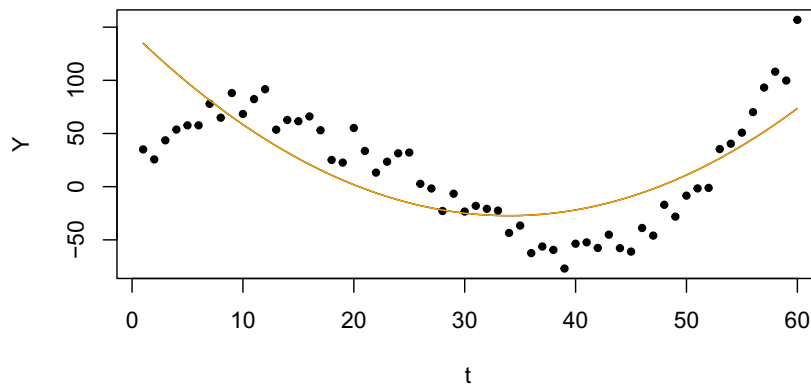
$$Y(t) = 10 + 12 \cdot t - 0,7 \cdot t^2 + 0,009 \cdot t^3 + \epsilon(t).$$

ϵ v tomto prípade znovu predstavuje chybový člen s nulovou strednou hodnotou a smerodajnou odchýlkou 10. V tomto prípade budeme generovať 60 náhodných pozorovaní. Ako prvé sa pozrime na to, ako vygenerované dáta vyzerajú a ako by ich prekrytie pomocou jednoduchej priamky vyzeralo.



Obr. 3.5: Preloženie dát jednoduchou priamkou

Ako na obrázku 3.5 vidíme, tak charakter dát v tejto simulácii už je na prvý pohľad podstatne zložitejší ako v prechádzajúcich simuláciách. Po preložení dát jednoduchou priamkou vidíme, že tento odhad je už dosť nedostačujúci a teda nevie preložiť vygenerované dáta tak, aby dobre vystihol ich charakter a vlastnosti. Pozrime sa ešte nato, ako ako by vyzeralo preloženie týchto dát pomocou polynómu druhého stupňa.

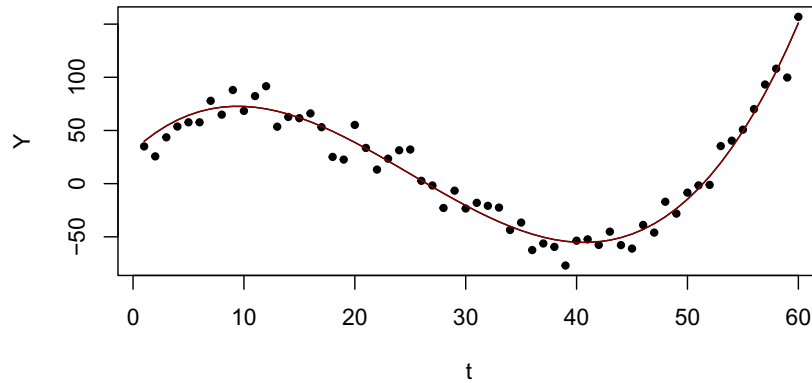


Obr. 3.6: Preloženie dát pomocou kvadratickej funkcie

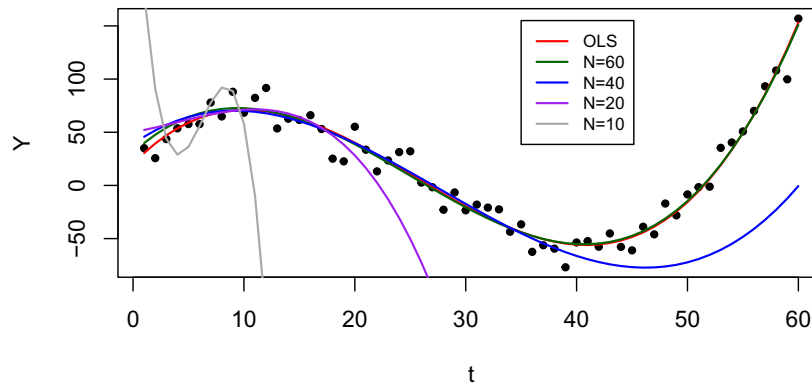
Z obrázku 3.6 vidíme, že tento odhad sa už pokúša lepšie adaptovať na zmeny a vystihnúť trend dát, ale odhadnutý tvar kvadratickej krivky stále nie je pre preloženie vygenerovaných dát dostačujúce. Kubický polynóm bude v tomto prípade vzhľadom na model, z ktorého dáta generujeme najvhodnejšiu voľbu. Odhadovaný tvar kubickej krivky nadobúda nasledujúci tvar $y = a + b \cdot x + c \cdot x^2 + d \cdot x^3$ respektíve $y = a + b \cdot x + c \cdot x^2$ pre kvadratickú funkciu a rovnakým spôsobom pre vyššie polynómy. V princípe funguje algoritmus na úplne rovnakej báze, ale jediným rozdielom je to, že $P(0)$ pri inicializácii algoritmu bude matica veľkosti 4×4 a respektíve pri kvadratickej funkcii 3×3 a teda aj dimenzia odhadu bude vektor príslušnej veľkosti. $P(0)$ sme pri každom odhade znovu volili ako $10^7 \cdot I$, pre vhodnú veľkosť jednotkovej matice. V poslednom rade sa pozrieme na preloženie dát kubickou funkciou a ako daný odhad vystihuje pozorovania. V tabuľke 3.2 vidíme odhadnuté koeficienty a , b , c a d v jednej tabuľke pre vybrané počty pozorovaní ako aj odhad vytvorený pomocou klasickej metódy najmenších štvorcov. Z tabuľky je možné jasne pozorovať vývoj a smerovanie RLS odhadov k OLS odhadu rovnako ako predchádzajúcej simulácii. Kvôli zložitosti dát sú, ale rozdiely RLS a OLS odhadov oveľa väčšie ako pri prekladaní dát jednoduchou priamkou. Obrázok 3.7 znázorňuje preloženie dát konečným RLS odhadom, takže odhadom, ktorý má minimálnu štvorcovú odchýlku. Môžeme vidieť, že tento odhad už pekne vystihuje vygenerované dáta. Pozrieme sa ešte na to ako odhady vytvorené z 10, 20, 40 a 60 dostupných pozorovaní prekladajú dáta, kde taktiež budeme vidieť ich postupný vývoj. Preloženie dát týmito odhadmi môžeme pozorovať na obrázku 3.8. Odhad vytvorený zo všetkých 60 pozorovaní pri vykreslení už veľmi dobre odpovedá OLS odhadu. Zvyšné odhady z 10, 20 a 40 dát už sa výslednému odhadu pomerne vzdialujú a to je hlavne spôsobené komplexitou systému, z ktorých sú dáta generované. Tieto odhady nemajú šancu vystihnúť meniaci sa trend generovaných dát. Hlavnou podstatou tejto podkapitoly bolo ukávanie toho, že je rekurentnú metódu najmenších štvorcov rovnako ako klasickú metódu najmenších štvorcov možné použiť na prekladanie dát ako jednoduchou priamkou tak aj ľubovoľným polynómom vhodného rádu,. Taktiež sme pozorovali postupný vývoj RLS odhadov pre vybrané N a ako odhad postupne smeruje k odhadu, ktorý najlepšie vystihuje namerané dáta.

	RLS $N = 10$	$N = 20$	$N = 40$	$N = 60$	OLS $N = 60$
a	320.753	50.207	38.925	30.800	20.181
b	-170.262	1.755	7.354	9.641	10.905
c	30.948	0.211	-0.481	-0.629	-0.670
d	-1.654	-0.017	0.005	0.008	0.008

Tabulka 3.2: Vývoj RLS odhadu koeficientov kubického polynómu s príchodom nových dát a porovnanie s OLS odhadom



Obr. 3.7: Preloženie dát pomocou kubickej funkcie



Obr. 3.8: Vývoj odhadu pomocou kubického polynómu pre rôzne N

3.3 Simulácia: Skúmanie konvergenzie RLS odhadu k OLS odhadu

Podstatou tejto simulácie bude ukázať ako pre zvyšujúci sa počet pozorovaní konverguje rozdiel odhadu vytvoreného pomocou klasickej metódy najmenších štvorcov a odhadu vytvoreného rekurentnou metódou najmenších štvorcov k nule. Táto vlastnosť vyplýva z podkapitoly 2.4.2, kde sme si ukázali, že odhad musí byť dostatočne blízko β_0 . Spomenutú konvergenciu si predvedieme pomocou postupného zväčšovania počtu dát generovaných z jednoduchého modelu v tvare $Y(t) = 2 + 5 \cdot t + \epsilon(t)$. V tomto prípade ϵ bude znovu predstavovať chybové členy dané normálnym rozdelením so strednou hodnotou 0, ale so smerodajnou odchýlkou 35.

Chceme teda ukázať, že pre rastúce N sa prakticky RLS odhad približuje bližšie a bližšie OLS odhadu. Ako počty pozorovaní pre túto simuláciu budeme uvažovať hodnoty $N = 25, 50, 100, 250, 500, 1000, 2500, 5000$. Hodnoty t budeme brať v každom prípade ako $1, 2, 3, \dots, N$ z ktorých následne vygenerujeme hodnoty Y ako: $Y(t) = 2 + 5 \cdot t + \epsilon(t)$. Na vygenerované dáta aplikujeme rekurzívny algoritmus (2.15), ktorý nám vytvorí postupnosť odhadov, pre každé prístupné pozorovania. Ako inicializáciu $P(0)$ v simulácii znovu použijeme $10^7 \cdot I$, čoho dôvod sme si vysvetlili už v predchádzajúcej simulácii. Pre každý rozsah dát N spustíme cyklus 1000 krát s rôzne generovanými chybovými členmi a v každom prípade pre vygenerované hodnoty Y spočítame príslušné OLS a RLS odhady. Následne si spočítame rozdiel v konečného RLS odhadu a OLS odhadu a týchto 1000 hodnôt spriemerujeme pre všetky rozsahy N , čím sa dostaneme k priemerných odchýlkám týchto dvoch odhadov. Tieto rozdiely budeme separátne počítat pre odhady koeficientov a a b .

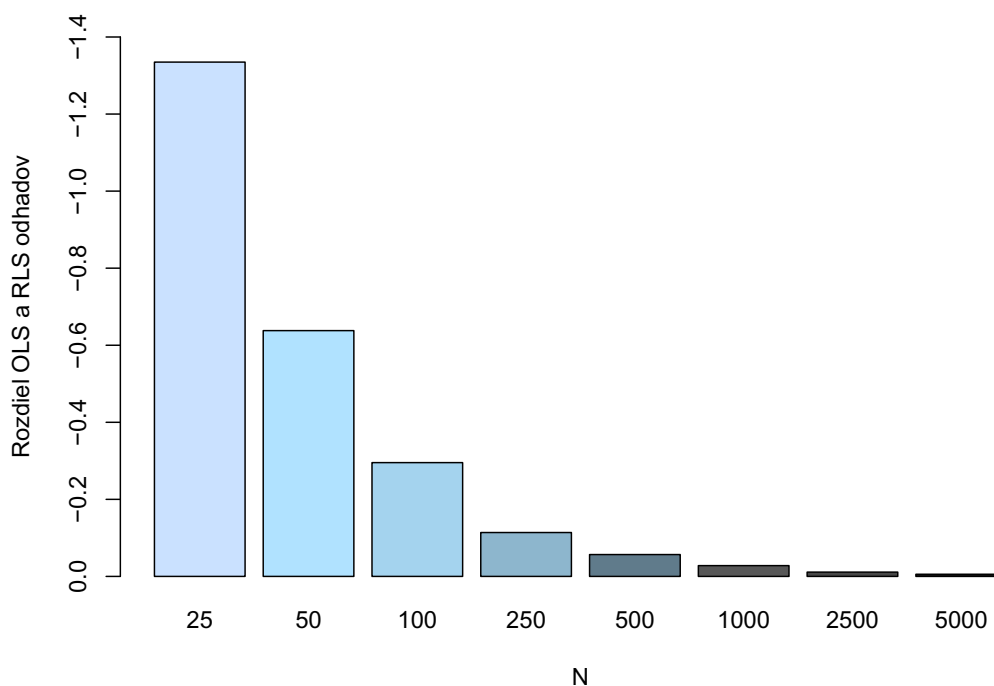
Okrem pozorovania konvergenzie rozdielov daných odhadov k nule si vykreslíme boxplot (krabicový graf) pre všetky konečné odhady z cyklu pre každé N , kde budeme pozorovať postupne sa znižujúce rozptyly RLS odhadov s pribúdajúcim počtom pozorovaní, čo je ďalšia z vlastností rekurentnej metódy najmenších štvorcov. Ako budeme z grafov vidieť, je podstatné mať čo najväčšie množstvo pozorovaní, ak chceme aby odhady boli čo najpresnejšie. Ako už vieme tak odhadujeme tvar priamky, ktorá čo najlepšie vystihuje namerané hodnoty. Táto priamka je v tvare $a + b \cdot x$, kde a teda označuje prienik s osou y a b označuje sklon odhadnutej priamky. Vypočítané priemerné rozdiely hodnôt odhadov si zaznamenáme do tabuľky a taktiež si ich hodnoty vykreslíme do stĺpcového grafu.

V tabuľke 3.3 máme vypísané priemerné rozdiely v oboch odhadovaných koeficientoch a a b a je jasné, že počet pozorovaní má na tieto spriemerované rozdiely signifikantný vplyv. Hodnoty pomerne rýchlo smerujú k nule, čo sme si chceli aj ukázať. Na obrázkoch 3.9 a 3.10 môžeme vidieť jasný trend znižujúcich sa rozdielov v oboch hodnotách odhadov a teda aj peknú konvergenciu rozdielu RLS odhadu a OLS odhadu k nule so zvyšujúcim sa počtom N . RLS odhady vytvorené z menej pozorovaní sú OLS odhadu ešte pozorovateľne vzdialené, ale pre 1000, 2500 a 5000 pozorovaní dávajú už veľmi podobné hodnoty. Taktiež môžeme pozorovať, že rozdiely v hodnote koeficientu a sú patrične väčšie ako v hodnote b , čo ale nie je vzhľadom ku tvaru priamky prekvapivé. Obrázky 3.11 a 3.12 ešte postupne zobrazujú rozptyly vytvorených odhadov pre rôzne počty pozorovaní ako pre

koeficient a tak aj koeficient b . V oboch prípadoch je viditeľné, že sa rozptyly odhadov postupne znižujú s rastom počtu pozorovaní. Pre čo najlepšiu presnosť odhadov je teda podstatné získať a použiť, čo najviac pozorovaní.

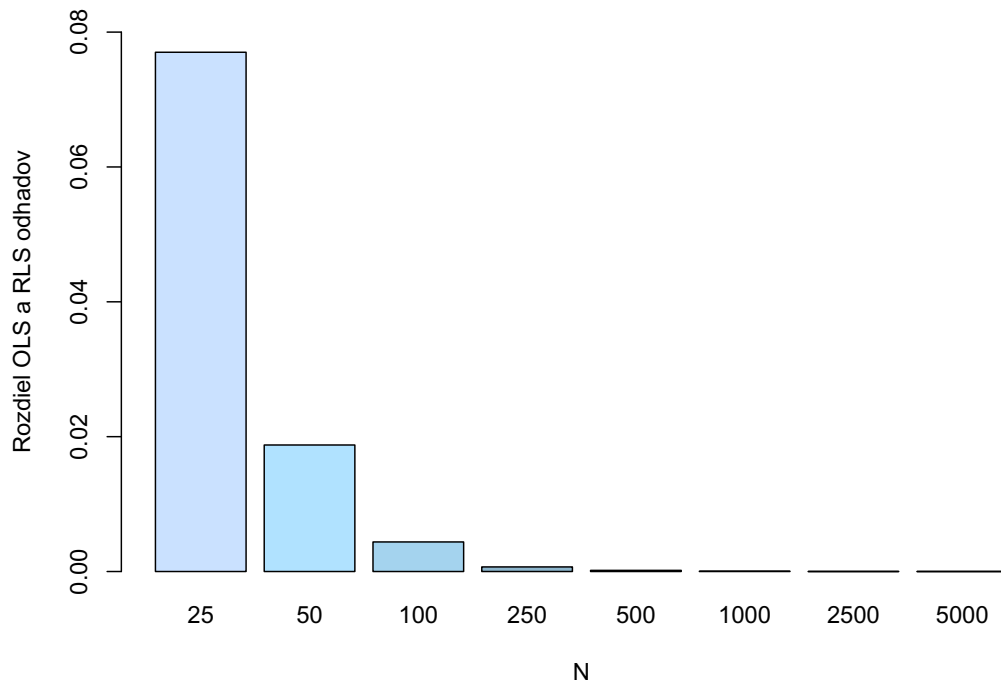
	a	b
N=25	-1.334856358	0.077010943754
N=50	-0.637945492	0.018763102724
N=100	-0.295337145	0.004386195227
N=250	-0.113897083	0.000680659857
N=500	-0.056726696	0.000169840408
N=1000	-0.027998579	0.000041955913
N=2500	-0.011149909	0.000006687271
N=5000	-0.005589795	0.000001676603

Tabuľka 3.3: Priemerné rozdiely OLS a RLS odhadov pre rôzne N

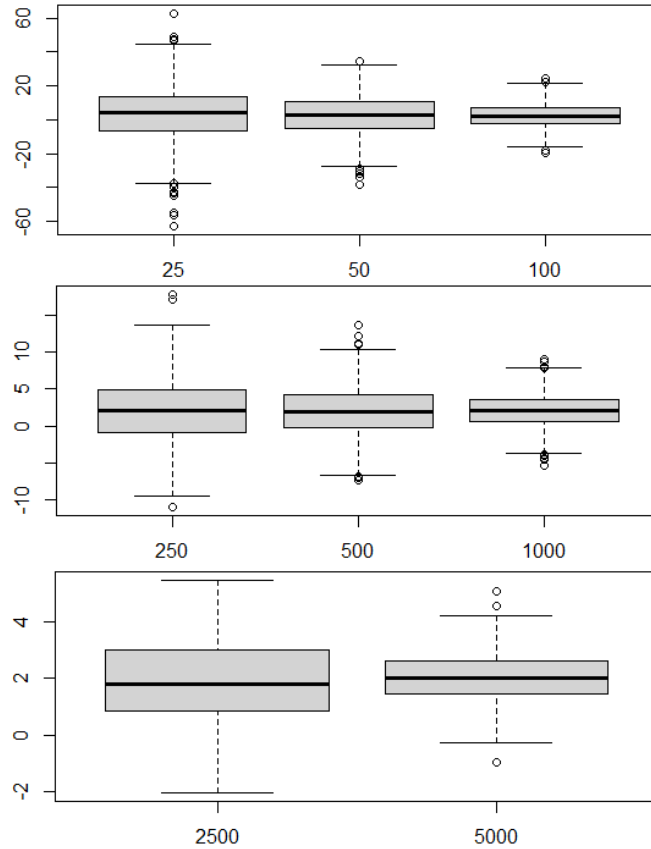


Obr. 3.9: Rozdiely v hodnote a medzi RLS a OLS odhadmi pre rôzne N

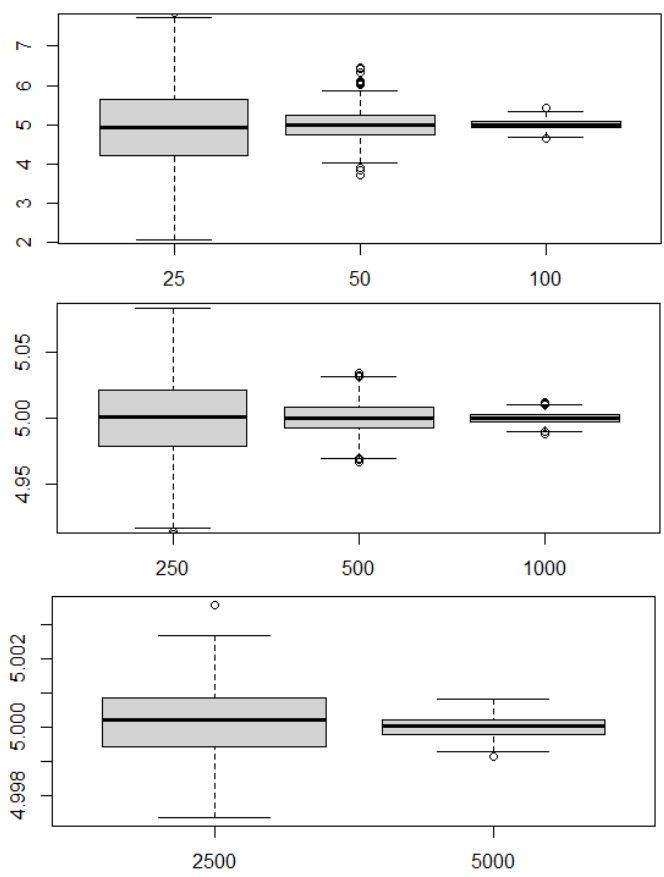
Na jednoduchšej simulácii sme si teda predviedli konvergenciu rozdielu odhadu vytvoreného rekurentnou metódou najmenších štvorcov a odhadu vytvoreného všeobecnou metódou najmenších štvorcov k nule pre rastúci počet pozorovaní. Taktiež sme mohli pozorovať ako sa postupne znižuje aj rozptyl vytvorených RLS odhadov v závislosti na počte pozorovaní z boxplotov, ktoré sme z odhadov vytvorili.



Obr. 3.10: Rozdiely v hodnote b medzi RLS a OLS odhadmi pre rôzne N



Obr. 3.11: Rozptyly odhadov koeficientu a pre rôzne N



Obr. 3.12: Rozptyly odhadov koeficientu b pre rôzne N

3.4 Simulácia: Dopad forgetting factoru na odhad

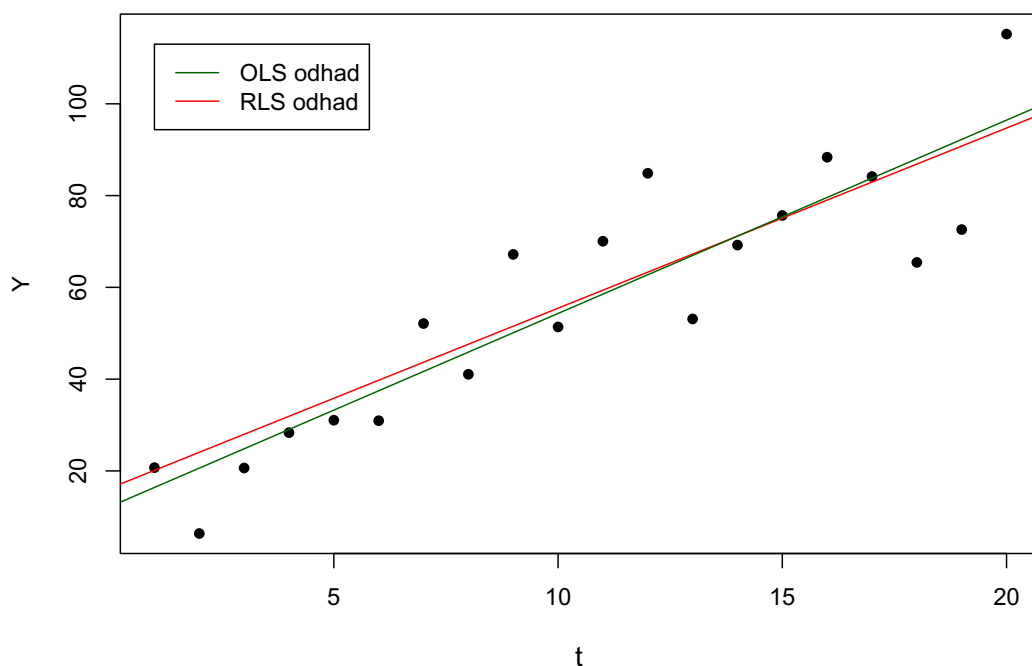
Táto simulačná štúdia sa pozrie na to ako sa odhad pomocou rekurentnej metódy najmenších štvorcov mení pri rôznych nastaveniach forgetting factoru. Simuláciu znovu budeme aplikovať na dáta generované jednoduchým modelom, ktorý sme už použili v prechádzajúcich simuláciách. Zmena bude len v počte generovaných pozorovaní. Kvôli pomerne rýchlemu približovaniu RLS odhadu k OLS odhadu, odhady s rôznymi forgetting faktormi λ sa adaptujú a tiež smerujú rýchlo k sebe. Tým pádom najlepšie bude možné pozorovať dopad rôznych λ pri menej pozorovaniach. Vykreslenie prekrytia dát odhadmi by z konvergenencie rozdielov OLS a RLS odhadov k nule nemalo veľkú výpovednú hodnotu. V tomto prípade sa pozrieme na odhady v krokoch rekurzívnej pre rôzne hodnoty λ a budeme taktiež sledovať ako sa odhady postupne začnú čoraz viac približovať a postupne zhodovať. Dáta ako sme si už spomenuli budeme generovať modelom $Y(t) = 2 + 5 \cdot t + \epsilon(t)$, kde t sú hodnoty $1, \dots, 20$ a ϵ budeme uvažovať ako chybové členy generované normálnym rozdelením s nulovou strednou hodnotou a smerodajnou odchýlkou 10. Seed nastavíme na hodnotu 42.

	$\lambda=1$	$\lambda=0.99$	$\lambda=0.95$	$\lambda=0.90$	$\lambda=0.75$
1.	207095845	205024886	196741052	186386260	155321884
	207095845	205024886	196741052	186386260	155321884
2.	82838342	82009958	78696425	74554508	62128757
	-41419161	-41004970	-39348203	-37277245	-31064369
3.	143.47295	143.47295	143.4729	143.4729	143.47288
	-47.85041	-47.85041	-47.8504	-47.8504	-47.85038
4.	54.506152	54.506152	54.506150	54.506149	54.506142
	-9.721772	-9.721772	-9.721771	-9.721771	-9.721769
5.	34.389845	34.389845	34.389844	34.389844	34.389841
	-2.178156	-2.178156	-2.178156	-2.178156	-2.178155
6.	28.619138	28.619138	28.6191372	28.6191370	28.6191359
	-0.254587	-0.254587	-0.2545869	-0.2545869	-0.2545866
7.	16.581905	16.581905	16.581905	16.581905	16.581905
	3.356583	3.356583	3.356583	3.356583	3.356583
...
19.	19.281098	19.281098	19.281098	19.281098	19.281097
	3.524461	3.524461	3.524461	3.524461	3.524461
20.	16.20263	16.20263	16.20263	16.20263	16.20263
	3.92600	3.92600	3.92600	3.92600	3.92600

Tabuľka 3.4: Vývoj odhadov v čase

V tabuľke 3.1 môžeme pozorovať odhady pre rôzne hodnoty λ pre vybrané hodnoty spočítané rekurzívnou metódou. Vidíme, že hodnoty počiatočných odhadov sa pomerne líšia, keďže na tieto pozorovania má forgetting factor najväčší vplyv. Je možné si všimnúť, že do 7 pozorovania sa už odhady pre všetky rôzne λ zhodujú a postupne pomaly konvergujú k odhadu pomocou OLS. Takže vplyv voľby forgetting factoru najviac ovplyvňuje začiatok rekurzívnej, ale postupne sa

odhady rýchlo formujú na požadovaný výsledok. Začiatocným pozorováním sa so znižujúcim λ pokladá menšia dôležitosť a chceme pokladať väčšiu dôležitosť na pozorovania, ktoré sme získali neskôr v čase, keďže sa dajú považovať za viac aktuálne a signifikantné. Ak na rovnaké dáta aplikujeme klasickú metódu najmenších štvorcov dostaneme sa ku výslednému odhadu 12.177, 4.214. Je vidno, že RLS odhady pomaly smerujú k OLS odhadu, ale kvôli rozsahu dát sú ešte viditeľne vzdialené, čo je možné pozorovať aj na obrázku 3.13, kde sme si vykreslili prekrytie dát ako RLS odhadom tak OLS odhadom pre všetkých 20 pozorovaní. RLS odhady pre rôzne λ sú vtedy už zhodné a preto nám stačí vykresliť jeden odhad, ktorý reprezentuje všetky.



Obr. 3.13: Rozdiel preloženia dát OLS a RLS odhadom pri $N = 20$

4. Praktická časť

4.1 Predstavenie

V tejto časti práce sa pozrieme na jednu z možností využitia rekurentnej metódy najmenších štvorcov pri jej aplikácii na reálne dáta. Potrebu využitia metódy a výhody jej rekuzie sme si už v predchádzajúcich častiach práce spomenuli a preto nie je prekvapivé, že sa s jej pomocou rieši škála problémov. V tejto časti sa pozrieme na jej praktické využitie vo financiách a to konkrétne pri CAPM modeli. CAPM model je v skratke Capital asset pricing model, čo je v preklade model oceňovania kapitálových aktív. Tento model tvorí dôležitú zložku novodobej teórie portfólia, ktorá sa zaoberá konštrukciou a analýzou investičného portfólia. Teória portfólia je v súčasnosti založená na diverzifikácii portfólia, čo vedie k zníženiu rizika spadajúceho na investora, keďže aktíva nie sú dostatočne vzájomne korelované. CAPM model poskytuje možnosť ako vieme zmerať rizikovosť určitého aktíva v kontexte väčšieho portfólia. Spočítaním očakávaného výnosu a rizika jednotlivých aktív použitím CAPM modelu, môžu investori zostrojiť optimálne portfólio s minimálnym rizikom a maximálnym očakávaným výnosom, alebo prípadne iné portfólio s inak nastaveným zámerom.

4.2 Vzorec CAPM modelu

CAPM je vlastne model opisujúci vzťah medzi očakávaným výnosom a rizikom investície do daného aktíva. Ako sa uvádza v Cipra (2005) v podkapitole 12.3 tak vzorec, ktorý sa pri výpočte oceňovania aktív používa je nasledujúci:

$$E R_i = R_f + \hat{\beta}_i(E R_m - R_f), \quad (4.1)$$

kde $E R_i$ označuje očakávanú výnosovú mieru kapitálového aktíva i , R_f je bezriziková výnosová miera, $E R_m$ označuje očakávanú výnosovú mieru trhu a $\hat{\beta}$ sa používa ako faktor, ktorý označuje senzitivitu aktíva. Bodku nad β použijeme jednoducho len pre rozlíšenie označenia od vektora neznámych používaného vo zvyšku práce. Všeobecne sa tento koeficient označuje ako β . Trh má podľa definície $\hat{\beta}$ danú rovnú 1. To znamená, že ak $\hat{\beta}$ daného aktíva je rovná 1, aktívum reaguje na zmeny na trhu presne ako trh. Ak má aktívum $\hat{\beta}$ väčšie ako 1 tak je aktívum viac senzitívne na zmeny a teda pri náraste, či poklese trhu sú viac nestále, ale správajú sa podľa trendu trhu. Z čoho opačne vyplýva, že aktíva s $\hat{\beta}$ menším ako 1 sú menej volatilné ako trh. Špeciálnym prípadom sú aktíva, ktorých faktor $\hat{\beta}$ je záporný, čo predstavuje negatívnu koreláciu s trhom. Inak povedané, ak trh rastie výnos daného aktíva klesá a opačne. Príkladom takéhoto aktíva je napríklad zlato, ktoré je pri nepriaznivých podmienkach na trhu kupované investormi pre udržanie hodnoty peňazí a ako páka proti inflácii.

4.3 Aplikácia RLS na CAPM model

Zápis (4.1) sa vhodne modifikuje pri aplikáciách spojených s regresnou analýzou. Ako sa uvádza v knihe Cipra (2008) na strane 32 je možné klasický zápis

pre naše potreby modifikovať takto:

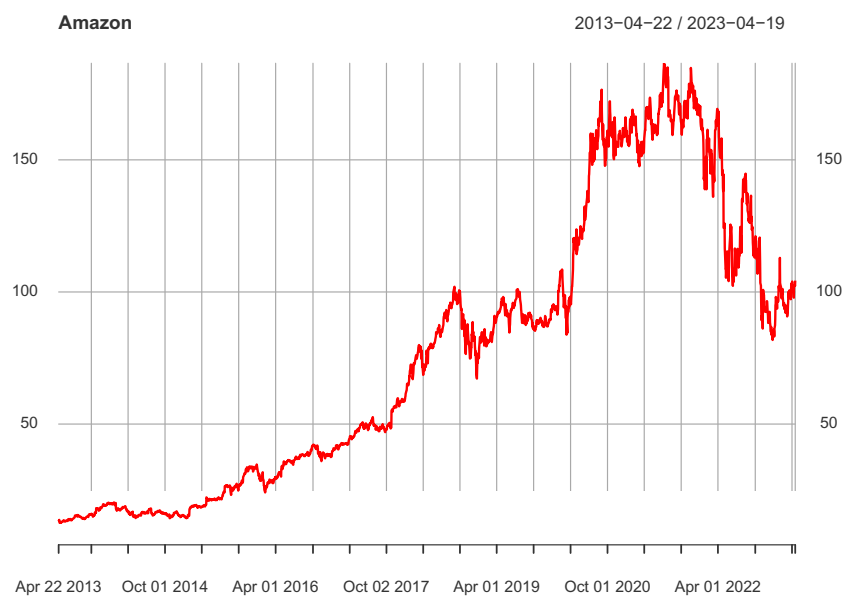
$$R_i(t) - R_f(t) = \alpha_i + \beta_i(R_m(t) - R_f(t)) + \epsilon(t), \quad t = 1, \dots, N \quad (4.2)$$

kde $R_i(t)$ je výnos na aktíve i v danom čase t (v našom prípade v danom dni), $R_f(t)$ je výnos bezrizikového aktíva v čase t , $R_m(t)$ je teda výnos indexu v danom čase t a α_i a β_i sú koeficienty, ktoré chceme odhadnúť pomocou rekurentnej metódy najmenších štvorcov a klasickej metódy najmenších štvorcov. Na chybovú zložku kladieme rovnaké predpoklady ako v prípade lineárneho regresného modelu. Uviedli sme si teda CAPM model v tvare klasickej regresnej rovnice.

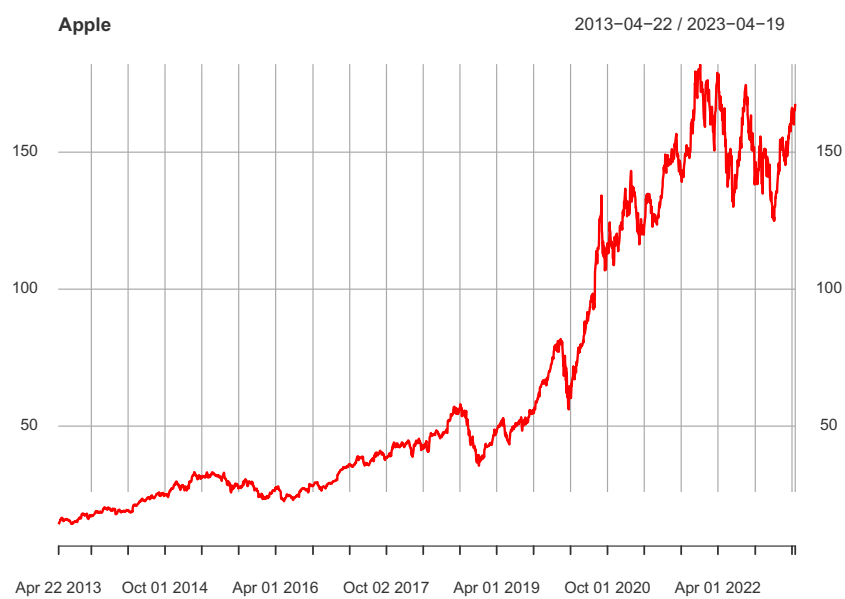
Budeme sa teda snažiť pomocou rekurzívneho algoritmu odhadovať β faktor pre 3 rôzne akcie z akciového indexu Standard and Poor's 500 (S&P 500 (^GSPC)). Tento koeficient je odhadovaný na základe historických dát akcii a takzvanej bezrizikovej investície a taktiež ich vzájomného správania na trhu. Konkrétne budeme β odhadovať pre akcie Amazon (AMZN), Apple (AAPL) a Tesla (TSLA). Ako bezrizikovú výnosovú mieru použijeme výnosy 10 ročných amerických štátnych dlhopisov (10 year US Treasury Bonds). Grafy cien jednotlivých aktív ako aj indexu S&P 500 za posledných 10 rokov si vykreslíme a taktiež k nim pridáme tabuľku so základnými štatistickými charakteristikami v danom období 10 rokov.

Všetky dáta sú prevzaté z Yahoo Finance. Pozorovania, ktoré budú vstupovať do algoritmu, si vytvoríme z každodenných zmien vybraných akcii, indexu a výnosov dlhopisov za posledné roky. Z týchto hodnôt si následne vytvoríme hodnoty závislej premennej ako aj nezávislých premenných, ktorých vzájomné vzťahy nás zaujímajú. Ako je zo zápisu (4.2) vidieť tak hodnoty závislej premennej Y spočítame ako rozdiel medzidenných percentuálnych zmien akcie i (R_i) a v danom dni určenú bezrizikovú výnosovú mieru (R_f). Na druhej strane nezávislú premennú vytvoríme pomocou rozdielu medzidenných percentuálnych zmien indexu S&P 500 (R_m) od ktorých znovu odpočítame v danom dni určenú bezrizikovú výnosovú mieru (R_f). Na získané hodnoty Y a x teda použijeme RLS algoritmus a aj klasickú metódu najmenších štvorcov. Pri rekurzívnom odhade volíme forgetting faktor λ rovný 1 rovnako ako aj váhy pozorovaní *alpha* a algoritmus inicializujeme s $P(0)$ v tvare $10^7 \cdot I$, tak ako v simuláciách a rovnako vektor $\beta(0)$ je pri spustení volený ako nulový vektor, kde znovu táto voľba vyplýva z 2.4.1. Mohli sme sa v tejto časti prikloniť ku klasickej voľbe počiatočných podmienok, keďže máme pomerne dost dát, ale zostaneme vo voľbe počiatočných podmienok konzistentný a preto sme sa znovu priklonili ku alternatívnej voľbe hodnôt pre spustenie algoritmu. Dáta budeme brať do dátumu 20.04.2023 a odhady spočítame postupne za posledný rok, posledné tri roky, posledných päť rokov ako aj posledných desať rokov od uvedeného dátumu.

Klasická metóda najmenších štvorcov je jednou z možností ako spočítať koeficient β , ale v závislosti na dátach je možné samozrejme použiť na spočítanie koeficientu aj rôzne iné metódy ako odhad pomocou maximálnej vierohodnosti, analýza časového radu, či prevalenčná analýza (Cross-Sectional Analysis). S koeficientom β sa zvyčajne na finančných stránkach spomínajú historické hodnoty troch, či piatich rokov. Budeme si taktiež všimnúť zmeny odhadov v závislosti na voľbe obdobia, z ktorého sme dáta čerpali. Získané hodnoty odhadov si zapíšeme do tabuľky. Pozrime sa ešte pred spočítaním odhadov β na tabuľky so štatistickými charakteristikami cien a medzidenných výnosov, ktoré vo výpočtoch vystupujú. Taktiež vykreslíme vývoje cien skúmaných aktív v danom desaťročí.



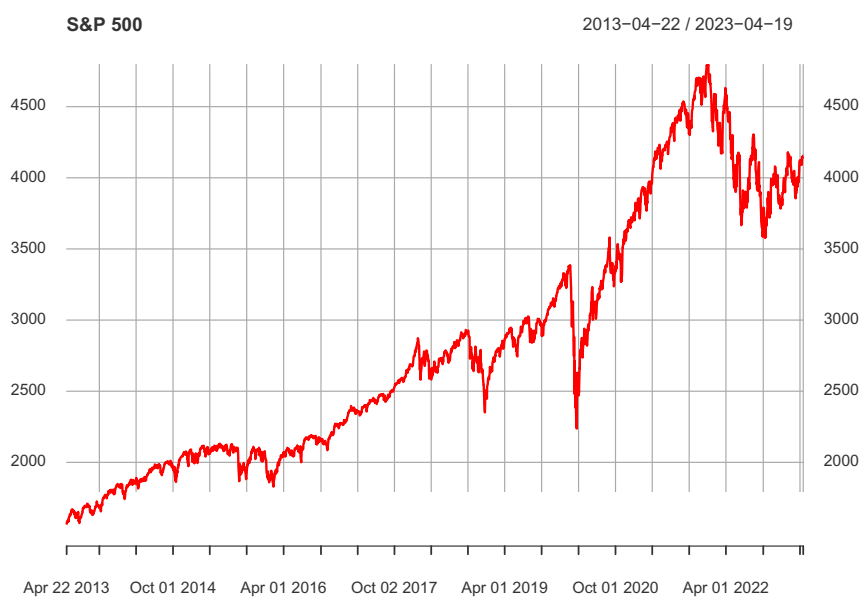
Obr. 4.1: Vývoj ceny akcie Amazon



Obr. 4.2: Vývoj ceny akcie Apple



Obr. 4.3: Vývoj ceny akcie Tesla



Obr. 4.4: Vývoj ceny indexu S&P 500

	Amazon	Apple	Tesla	S&P 500
Priemer	76.28	66.37	78.24	2816.55
Minimum	12.41	14.06	3.35	1562.50
Maximum	186.57	182.01	409.97	4796.56
Smerodajná odchýlka	52.37	50.25	102.79	872.10

Tabuľka 4.1: Štatistické charakteristiky cien aktív, indexu a dlhopisu za posledných 10 rokov

	Amazon	Apple	Tesla	S&P 500	Dlhopis
Priemer	0.00104	0.00114	0.00224	0.00045	0.02199
Minimum	-0.14049	-0.28647	-0.21063	-0.11984	0.00499
Maximum	0.14131	0.11981	0.24395	0.09382	0.04234
Smerodajná odchýlka	0.02072	0.01806	0.036324	0.01114	0.00737

Tabuľka 4.2: Štatistické charakteristiky medzidených výnosov aktív, indexu a dlhopisu za posledných 10 rokov

Ako je na grafoch možné pozorovať (obrázky 4.1, 4.2, 4.3, 4.4), každé aktívum zaznamenalo jasný nárast s miernymi prepismi v cene až približne do roku 2020, kde najskôr nastala krátka stagnácia a následne pomerný prepis v ich cenách. Hlavnou príčinou poklesu ceny aktív bude pandémie COVID-19 ako aj neskorší vojnový konflikt na Ukrajine. Pozrime sa teda na odhadnuté hodnoty koeficientu pomocou rekurzívnej a klasickej metódy najmenších štvorcov vo vybraných časových obdobiach:

	Amazon	Apple	Tesla
RLS(posledný 1 rok)	1.65629	1.30219	1.74202
OLS(posledný 1 rok)	1.64768	1.29438	1.73320
RLS(posledné 3 roky)	1.23242	1.17225	1.57059
OLS (posledné 3 roky)	1.22915	1.16917	1.56707
RLS (posledných 5 rokov)	1.09687	1.16984	1.40097
OLS (posledných 5 rokov)	1.09611	1.16908	1.40020
RLS(posledných 10 rokov)	1.11252	1.13276	1.36823
OLS(posledných 10 rokov)	1.11267	1.13289	1.36821

Tabuľka 4.3: Odhady koeficientu $\hat{\beta}$ pre CAPM model spočítaného k dátumu 20.04.2023

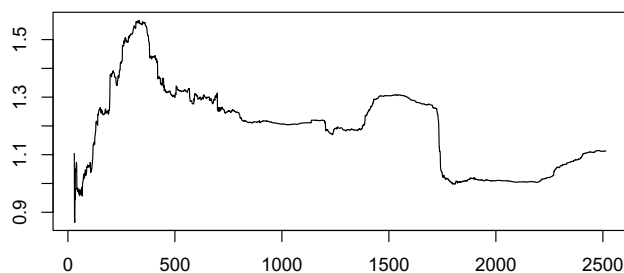
V tabuľke 4.3 môžeme vidieť spočítané RLS a OLS odhady, vytvorené z historických dát za posledných 1, 3, 5 a 10 rokov. Odhady koeficientu $\hat{\beta}$ sa pri akciách značne menia v závislosti na zvolenom období, z ktorého sme odhady vytvárali. Akcie Tesly mali v každom období najväčší odhadnutý koeficient, takže môžeme tvrdiť, že práve táto akcia najviac reagovala na zmeny na trhu. Pre Apple je tento koeficient pri odhadoch z dát v poslednom roku a troch rokoch oproti zvyšným dvom akciám viditeľne menší, takže ich akcie majú dosť podobný vývoj ako trh aj keď práve trh rastie alebo klesá. Akcie Amazonu a Apple mali v odhadoch za posledných 3, 5 a 10 rokov veľmi podobné koeficienty $\hat{\beta}$. Taktiež sú tieto odhady pomerne blízko 1, čo značí, že ich vývoj je veľmi podobný trendu trhu. To

môže byť výhodné pri nepriaznivom stave na trhu, keďže by ich pokles nemal byť tak závažný ako napríklad pri Tesle. Samozrejme, ale na druhej strane, ak sa trhu bude dariť Amazon a Apple by nemal poskytovať taký rapidný nárast ceny aktíva ako Tesla. Apple mal celkovo najstabilnejšie odhady $\hat{\beta}$ naprieč všetkými skúmanými obdobiami. Z konvergenencie rozdielu RLS a OLS odhadu k nule s rastúcim počtom pozorovaní si je možné v tabuľke všimnúť, že RLS odhady sa veľmi približujú OLS odhadom a so zväčšujúcim sa obdobím, v ktorom sme dáta uvažovali sa ich rozdiely rapidne znižujú. Túto vlastnosť sme si už ukázali na simulácii 3.3 v prechádzajúcej kapitole.

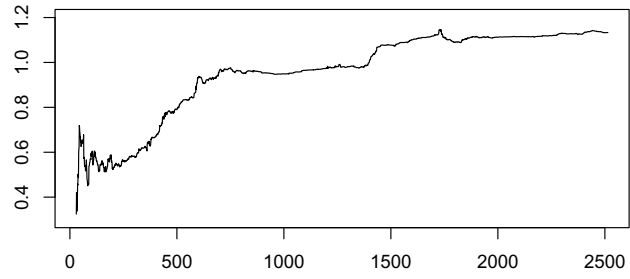
	Amazon	Apple	Tesla
$\lambda = 1$	1.11252898	1.13276775	1.36823410
$\lambda = 0.95$	1.11252897	1.13276774	1.36823409
$\lambda = 0.9$	1.11252895	1.13276773	1.36823407
$\lambda = 0.75$	1.11252890	1.13276767	1.36823401
$\lambda = 0.65$	1.11252885	1.13276762	1.36823394

Tabuľka 4.4: Ohady $\hat{\beta}$ za posledných 10 rokov pre rôzne forgetting factory

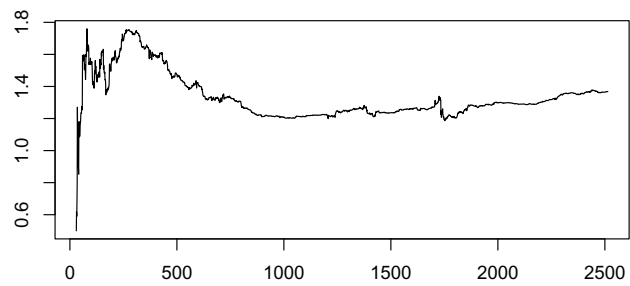
Tabuľka 4.4 ešte obsahuje výsledné hodnoty $\hat{\beta}$ spočítané za posledných 10 rokov pomocou rekurentnej metódy najmenších štvorcov pri rôznej voľbe forgetting factorov. Vidíme, že odhady sa pri takomto počte pozorovaní odlišujú minimálne. Môžeme si všimnúť, že so znižujúcim sa λ sa znižujú aj hodnoty faktoru $\hat{\beta}$, čo je spôsobené tým, ako algoritmus vyzerá. Grafy na obrázkoch 4.5, 4.6 a 4.7 nám poskytujú náhľad na to ako sa koeficient $\hat{\beta}$ pre každú z akcií aktualizoval pri rekurzii s príchodom nových pozorovaní. Grafy sú vytvorené z odhadov, v ktorých je $\lambda = 1$. Ako sme si už ukázali, tak pri takomto počte pozorovaní majú rôzne forgetting factory na výsledky už minimálny vplyv a preto by ani rôzne λ nebolo možné na vývojoch rozlíšiť. Vývoje odhadov $\hat{\beta}$ sú vytvorené len pre obdobie posledných desiatich rokov. Vývoj $\hat{\beta}$ pre Apple počas celého obdobia prakticky narastal. Koeficient Amazonu zaznamenal väčšie kmitanie, ale v menšom rozsahu. Môžeme taktiež pozorovať, že koeficient okolo pozorovania 400 dosiahol svoje maximum presahujúce hodnotu 1.5. Vývoj pri akcii Tesla najskôr dosiahol svoje maximum a následne prakticky osciloval v okolí hodnoty 1.3.



Obr. 4.5: Vývoj koeficientu $\hat{\beta}$ pre Amazon za posledné desaťročie



Obr. 4.6: Vývoj koeficientu $\dot{\beta}$ pre Apple za posledné desaťročie



Obr. 4.7: Vývoj koeficientu $\dot{\beta}$ pre Tesla za posledné desaťročie

Odhady $\hat{\beta}$ počítané pomocou rekurzívnej metódy najmenších štvorcov teda poskytujú solídne výsledky. Rekurzívny algoritmus poskytuje efektívny spôsob ako spočítať odhady $\hat{\beta}$, ktoré sú priamo vytvorené z historických dát a vzájomných vzťahov na trhu. Okrem toho sa oplatí koeficient $\hat{\beta}$ počítať rekurzívne aj z nasledujúcich dôvodov:

- On-line algoritmus poskytuje aktualizáciu odhadu s príchodom každého nového pozorovania. Po každom dni nám trh poskytne nové hodnoty, ktoré je možné implementovať a odhad vylepšovať
- Jednoduchosť algoritmu je tiež veľkou výhodou. Algoritmus je priamočiary a ľahko aplikovaný na dáta bez potreby špeciálnych znalostí či softveru
- Flexibilita a robustnosť metódy a jej možnosť aplikovania na širokú škálu dát s tým, že algoritmus nevyžaduje veľa predpokladov, ktoré dáta musia spĺňať.

Samozrejme, že aj odhad $\hat{\beta}$ pomocou RLS algoritmu má svoje nevýhody ako napríklad senzitivita modelu na odľahlé pozorovania, čo môže viesť k zaujatému odhadu alebo závislosť modelu na predpokladoch, či problematika numerickej stability pri zaokrúhľovaní. Aj napriek tomu, že odhad neberie do úvahy ostatné faktory trhu ovplyvňujúce výsledné $\hat{\beta}$ je tento odhad solídnym štartovacím bodom odhadovania a môže byť používaný kvôli jeho výhodám ako základ pre hodnotenie ostatných metód.

Záver

Bakalárska práca mala za úlohu predstavenie, odvodenie a ukázanie aplikácii rekurentnej metódy najmenších štvorcov. Vychádzali sme zo základného konceptu regresie, z ktorého sme postupne prešli na lineárnu regresiu a odhad pomocou metódy najmenších štvorcov. OLS odhad bol podstatným krokom k odvodeniu rekurentného algoritmu, ktoré tvorilo základ teoretickej časti spoločne s časťou ohľadom počiatočných podmienok algoritmu ako aj vybraných vlastnostiach. Vlastnosti a modifikácie metódy sme si priblížili a ilustrovali na simuláciách v samostatnej kapitole v práci.

Praktická časť sa venovala aplikácii rekurentnej metódy na reálne finančné dáta pri odhade faktoru β v CAPM modeli. Pri výpočte odhadu sme brali do úvahy historické dáta z rôzne dlhých období a výsledné odhady sme porovnali s OLS odhadmi. Dostali sme sa k odhadom koeficientu β , ktorých hodnoty sa líšili v závislosti na období z ktorého boli vytvorené.

Pri rekurentnej metóde najmenších štvorcov nie je prekvapivé jej použitie pri riešení vybraných problémov v rôznych oblastiach spoločnosti. Značnou výhodou je jej implementácia na aktualizáciu odhadu s príchodom každého nového pozorovania bez potreby náročných a zdĺhavých výpočtov a taktiež jej adaptácia na zmenu aj za prítomnosti rušivých členov.

Zoznam použitej literatúry

- ANDĚL, J. (2007). *Základy matematické statistiky*. Matfyzpress, Praha.
- CIPRA, T. (2005). *Praktický průvodce finanční a pojistnou matematikou*. Ekopress, Praha.
- CIPRA, T. (2008). *Finanční ekonometrie*. Ekopress, Praha.
- LJUNG, L. (1999). *System identification: Theory for the User, 2nd edition*. Prentice Hall.
- LJUNG, L. a SÖDERSTRÖM, T. (1983). *Theory and Practice of Recursive Identification*. The MIT Press, Cambridge, USA.

Zoznam obrázkov

1.1	Metóda najmenších štvorcov	7
3.1	Vykreslenie odhadov pre vybrané N a ich postupná aktualizácia .	23
3.2	Postupné odhady pre rôzne N , kde plné body znázorňujú aktuálne dostupné pozorovania pre RLS odhady	24
3.3	Vývoj koeficientu a	25
3.4	Vývoj koeficientu b	25
3.5	Preloženie dát jednoduchou priamkou	26
3.6	Preloženie dát pomocou kvadratickej funkcie	27
3.7	Preloženie dát pomocou kubickej funkcie	28
3.8	Vývoj odhadu pomocou kubického polynómu pre rôzne N	28
3.9	Rozdiely v hodnote a medzi RLS a OLS odhadmi pre rôzne N . .	30
3.10	Rozdiely v hodnote b medzi RLS a OLS odhadmi pre rôzne N . .	31
3.11	Rozptyly odhadov koeficientu a pre rôzne N	31
3.12	Rozptyly odhadov koeficientu b pre rôzne N	32
3.13	Rozdiel preloženia dát OLS a RLS odhadom pri $N = 20$	34
4.1	Vývoj ceny akcie Amazon	37
4.2	Vývoj ceny akcie Apple	37
4.3	Vývoj ceny akcie Tesla	38
4.4	Vývoj ceny indexu S&P 500	38
4.5	Vývoj koeficientu $\hat{\beta}$ pre Amazon za posledné desaťročie	40
4.6	Vývoj koeficientu $\hat{\beta}$ pre Apple za posledné desaťročie	41
4.7	Vývoj koeficientu $\hat{\beta}$ pre Tesla za posledné desaťročie	41

Zoznam tabuliek

3.1	Vývoj RLS odhadu s príchodom nových dát a porovnanie s OLS odhadom	22
3.2	Vývoj RLS odhadu koeficientov kubického polynómu s príchodom nových dát a porovnanie s OLS odhadom	28
3.3	Priemerné rozdiely OLS a RLS odhadov pre rôzne N	30
3.4	Vývoj odhadov v čase	33
4.1	Štatistické charakteristiky cien aktív, indexu a dlhopisu za posledných 10 rokov	39
4.2	Štatistické charakteristiky medzidených výnosov aktív, indexu a dlhopisu za posledných 10 rokov	39
4.3	Odhady koeficientu β pre CAPM model spočítaného k dátumu 20.04.2023	39
4.4	Odhady β za posledných 10 rokov pre rôzne forgetting factory . . .	40