

Univerzita Karlova v Praze Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Monika Steinhübelová

Robustní odhady v modelu CAPM

Katedra pravděpodobnosti a matematické statistiky

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

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2012

Dovoľujem si vysloviť poďakovanie vedúcej mojej diplomovej práce doc. RNDr. Zuzane Práškovej, CSc. za cenné pripomienky a návrhy, za poskytnutie potrebnej literatúry, za jej čas strávený pri čítaní práce a pri konzultáciách. Moje poďakovanie patrí aj všetkým, ktorí mi pomáhali s korektúrami textu, najmä priateľovi a rodičom.

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

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 v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V dne

Název práce: Robustní odhady v modelu CAPM

Autor: Monika Steinhübelová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Zuzana Prášková, CSc., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Diplomová práce popisuje teorii modelu oceňování kapitálových aktív (CAPM) a problematiku robustních odhadů. Robustní odhady sú účinným nástrojom, ako dosiahnuť lepší odhad vzhľadom ku klasickému odhadovaniu metódou najmenších štvorcov pri nesplnení predpokladu normálneho rozdelenia reziduí alebo pri výskyte odľahlých pozorovaní v dátach. Teória M -odhadov, ktorá je následne aplikovaná aj v praktickej časti práce na mnohorozmernom CAPM modeli, je v práci podrobne spracovaná. Menej podrobne je uvedená aj teória R - a L -odhadov. Simulačná štúdia práce porovnáva jednotlivé konštruované odhady a dáva do kontrastu simultánne odhady v mnohorozmernom modeli a odhady konštruované po zložkách. Odhadovanie po zložkách je aplikované na daný model za predpokladu vzájomnej nezávislosti jednotlivých rovníc.

Klíčová slova: robustné odhady, mnohorozmerný lineárny model, CAPM model, simultánne odhady

Title: Robust estimators for CAPM

Author: Monika Steinhübelová

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Zuzana Prášková, CSc., Department of Probability and Mathematical Statistics

Abstract: The thesis describes the theory of capital asset pricing model (CAPM) and the issue of robust estimates. Robust methods are an effective tool to achieve better estimation relative to the classical least squares method when there is a failure to assume a normal distribution of errors or in the presence of outlying observations in the data. Theory of M -estimates, which is then applied in the practical part of the thesis to the multidimensional CAPM model is treated in detail. Theory of R - and L -estimates is explained in less detail. A simulation study compares simultaneous estimates in multivariate model and estimates designed individually when applied to the model assuming the mutual independence of equations.

Keywords: robust estimates, multivariate linear model, CAPM model, simultaneous estimates

Obsah

Úvod	3
1 CAPM model	5
1.1 CML priamka	6
1.2 SML priamka	7
1.2.1 Koeficient beta	10
1.2.2 Miera nerovnováhy alfa	12
2 Lineárny regresný model	13
2.1 Jednorozmerná lineárna regresia	13
2.1.1 OLS-odhad	14
2.2 Mnohorozmerná lineárna regresia	15
2.2.1 OLS-odhad	16
3 Teoretický základ štatistických odhadov	20
3.1 Parametrický štatistický model	20
3.2 Vlastnosti štatistických odhadov	21
4 Robustné štatistické odhady	22
4.1 Detekcia odľahlých pozorovaní klasickými nástrojmi	22
4.1.1 Detekcia odľahlých pozorovaní v jednorozmernom modeli	23
4.1.2 Detekcia odľahlých pozorovaní v mnohorozmernom modeli	24
4.2 Charakteristiky robustnosti	25
4.2.1 Teoretický základ	25
4.2.2 Influenčná funkcia	26
4.2.3 Rozptyl asymptoticky normálneho rozdelenia	28
4.2.4 Bod zlyhania	29
4.2.5 Miera chvostov	30
4.2.6 Príklad	31
4.3 M-odhady	32
4.3.1 Influenčná funkcia a bod zlyhania	34
4.3.2 Studentizované M-odhady	35
4.3.3 Asymptotické vlastnosti a minimaximálne robustné odhady	35
4.3.4 Najpoužívanejšie M-odhady	37
4.3.5 M-odhady v lineárnom regresnom modeli	37
4.4 L-odhady	40
4.4.1 Influenčná funkcia a bod zlyhania	41
4.4.2 Najpoužívanejšie L-odhady	41
4.4.3 L-odhady v lineárnom regresnom modeli	42

4.5	R-odhady	42
4.5.1	Influenčná funkcia	43
4.5.2	Najpoužívanejšie R-odhady	43
5	Robustné odhady v mnohorozmernom CAPM modeli	45
5.1	Odhadovanie parametrov po zložkách	46
5.2	Odhadovanie parametrov simultánne	46
5.3	Simulačná štúdia	47
5.4	Aplikácia odhadov na reálne dáta	55
5.4.1	Základné portfólio	55
5.4.2	Upravené portfólio	68
5.4.3	Zjednodušené portfólio	72
	Záver	80
	Zoznam použitej literatúry	82
	Zoznam obrázkov	84
	Zoznam tabuliek	85
	Zoznam použitých skratiek	86
	Prílohy	87
	Príloha č. 1	87
	Príloha č. 2	89

Úvod

V tejto diplomovej práci je spracovaná teória robustných odhadov. V práci sú zahrnuté M -, R - a L -odhady. V praktickej časti sú M -odhady následne aplikované na mnohorozmerný CAPM model. V jednotlivých kapitolách práce je uvedené:

Kapitola 1 obsahuje popis CAPM modelu. Sú tu uvedené vzťahy pre CML- a SML-priamku. Následne je model zapísaný v tvare lineárneho regresného modelu, na ktorý možno aplikovať mnoho štatistických analýz.

Kapitola 2 popisuje problematiku lineárneho regresného modelu, a to jednak jednorozmerného, jednak mnohorozmerného. Obsahom kapitoly je aj odvodenie OLS-odhadov pre jednotlivé modely.

Kapitola 3 uvádza niekoľko poznatkov z teórie pravdepodobnosti a matematickej štatistiky, ktoré sú potrebné k objasneniu vzťahov v nasledujúcej kapitole 4.

Kapitola 4 popisuje problematiku odľahlých pozorovaní. Tieto pozorovania spolu s porušením predpokladov OLS-odhadov často spôsobujú, že OLS-odhad nie je najlepším odhadom parametrov lineárneho regresného modelu. Kapitola sa preto venuje rôznym štatistickým metódam, ktorými sa možno vysporiadať s odľahlými pozorovaniami. V prvom rade je to klasický prístup pomocou detekcie týchto pozorovaní, a následne je podrobne popísaná teória robustných odhadov. Robustné odhady sa s odľahlými pozorovaniami dokážu vysporiadať bez ich predchádzajúcej detekcie. V práci je uvedená teória k M -, R - a L -odhadom potrebná k riešeniu problematiky, príklady týchto odhadov a ich využitie v lineárnom regresnom modeli. Najrozsiahlšie je spracovaná teória M -odhadov, ktorá je následne v kapitole 5 aplikovaná na mnohorozmerný CAPM model.

Kapitola 5 zahŕňa popis mnohorozmerného CAPM modelu a stručnú teóriu mnohorozmerných M -odhadov potrebnú k simulačnému experimentu. Aplikácia na reálne dáta je taktiež uvedená v tejto kapitole. Cieľom simulačnej štúdie bolo ukázať, že OLS-odhady často nie sú najlepším riešením odhadu parametrov mnohorozmerného lineárneho modelu. Na základe simulačnej analýzy je ukázané, že simultánne odhady mnohorozmerného CAPM modelu sú často efektívnejšie než odhady konštruované po zložkách za predpokladu nezávislosti jednotlivých rovníc mnohorozmerného modelu. Podľa záverov simulačnej analýzy boli v rámci tejto kapitoly vybrané odhadové funkcie M -odhadov. Tieto funkcie boli následne aplikované na reálne pozorovania (denné a mesačné výnosy vybraných cenných papierov). V kapitole sú uvažované aj dátové vzorky s ďalšími úpravami a následná analýza týchto dát.

Prílohy obsahujú popis výstupov jednotlivých simulácií a odkazy na jednotlivé dokumenty na priloženom CD v priečinku „Prílohy“.

Kapitola 1

CAPM model

Model oceňovania kapitálových aktív označovaný skratkou CAPM (Capital Assets Pricing Model) bol navrhnutý na začiatku šesťdesiatych rokov minulého storočia. Pre vznik modelu CAPM bola smerodajná teória portfólia a diverzifikácia portfólia, ktorou sa v 50. rokoch minulého storočia zaoberal H. Markowitz. CAPM model je špeciálnym prípadom Markowitzovho modelu portfólia, v ktorom má práve jedno aktívum nulové riziko a kladný výnos.

Vývojom modelu CAPM sa nezávisle na sebe zaoberali J. Treynor, W. Sharpe, J. Lintner a J. Mossin (pozri [24]). Tento model je jedným zo základných nástrojov súčasnej finančnej analýzy aj napriek tomu, že je založený na nasledujúcich zjednodušujúcich predpokladoch:

- (1) investori majú k dispozícii iba informácie o očakávaných výnosoch (stredných mierach zisku) a rizikách (kovariančnej štruktúre výnosov), t. j. na základe týchto charakteristík ohodnocujú svoje portfóliá,
- (2) investori majú racionálne chovanie (preferujú vyššie výnosy), t. j. medzi dvoma inak zhodnými portfóliami si vyberú to, ktoré ma vyšší zisk,
- (3) investori majú averziu k riziku (preferujú menšie riziko), t. j. z dvoch portfólií so zhodným ziskom si vyberú portfólio s menším rizikom,
- (4) jednotlivé aktíva sú obchodovateľné a nekonečne deliteľné, t. j. investor môže kúpiť zlomok akcie,
- (5) všetci investori majú rovnaký investičný horizont jedného obdobia,
- (6) na kapitálovom trhu existuje bezriziková sadzba, ktorá je rovnaká pre všetkých dlžníkov aj veriteľov,
- (7) dane a transakčné náklady sú zanedbateľné,
- (8) informácie sú voľne a okamžite dostupné všetkým investorom za rovnakých podmienok.

Základné predpoklady, ktoré sú často veľmi vzdialené od skutočnosti, nebránia tomu, aby sa model CAPM používal k znázorneniu chovania kapitálových aktív vo vzťahu k chovaniu celého trhu.

Keďže CAPM model dáva do súvislosti strednú mieru zisku uvažovaného portfólia (cenného papiera) s jeho mierou rizika beta (ozn. β), umožňuje vyšetrovať príspevky jednotlivých aktív k strednému výnosu a riziku celého portfólia. Objasňuje teda vzťah medzi očakávaným výnosom aktíva a jeho rizikom za podmienok trhovej rovnováhy, kde všetci investori volia optimálne trhové portfólio zložené zo všetkých investícií na trhu. Určenie (odhad) koeficientu beta je v praxi základným problémom pri výpočte výnosov aktív podľa modelu CAPM.

Teoreticky model CAPM umožňuje odhadnúť očakávaný výnos každého aktíva, najmä akcií, na kapitálovom trhu v okamihu jeho rovnováhy, pričom vychádza z lineárneho vzťahu medzi výnosnosťou jednotlivých akcií a výnosnosťou kombinácie akcií na trhu. Model CAPM dáva do súvislosti analyzované portfólio s trhovým portfóliom (indexom cenných papierov) a bezrizikovou investíciou. Pri praktickej aplikácii je možné trhové portfólio aproximovať pomocou rôznych indexov, napr. Standard&Poor's 500 Stock Price Index, Wilshire 5000 Stock Index, Dow Jones Index a i. (pozri [6]). Týmto nahradením sa hľadanie koeficientu beta stáva štatistickým problémom lineárnej regresie, pričom koeficient beta je smernicou vzniknutej priamky.

V modeli CAPM sa uvádzajú dva typy priamok:

- *priamku kapitálového trhu CML* (Capital Market Line) a
- *priamku trhu cenných papierov SML* (Security Market Line).

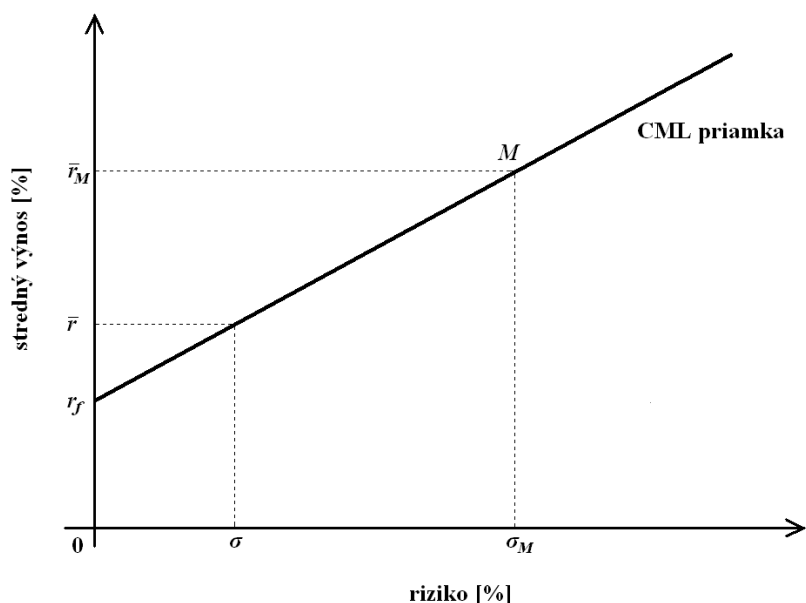
1.1 CML priamka

V rámci kapitálového trhu sa pre vysvetlenie vzťahu očakávaného výnosu (strednej miery zisku \bar{r}) a rizika (smerodajnej odchýlky σ) eficientného portfólia používa priamka kapitálového trhu (CML). V teórii portfólia má prípustná množina portfólií tzv. dáždnikovitý tvar a eficientná množina portfólií je daná jej „severozápadnou“ hranicou. Pokiaľ portfólio obsahuje aj bezrizikovú investíciu, je eficientná množina v tvare polpriamky (pozri [6], [7] a [8]). Túto skutočnosť zobrazuje práve CML priamka, ktorá je v tvare:

$$\bar{r} = r_f + \frac{\bar{r}_M - r_f}{\sigma_M} \sigma \quad (1.1)$$

- \bar{r} ... stredná miera zisku (výnos) eficientného portfólia,
 $\bar{r} = E(r) = \sum_{i=1}^n w_i \bar{r}_i$ pre n zložiek portfólia so strednými mierami zisku $\bar{r}_1, \dots, \bar{r}_n$ a s príslušnými váhami $w_i \geq 0$,
 $\sum_{i=1}^n w_i = 1$, $i = 1, \dots, n$
- σ ... riziko eficientného portfólia, $\sigma = \left(\sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij} \right)^{1/2}$,
kde $\sigma_{ij} = cov(r_i, r_j)$, $i, j = 1, \dots, n$
- \bar{r}_M, σ_M ... stredná miera zisku a riziko trhového portfólia (indexu cenného papiera)
- r_f ... miera zisku bezrizikovej investície

CML priamka teda prechádza bodmi $(0, r_f)$ a (σ_M, \bar{r}_M) , čo je zobrazené na nasledujúcom obrázku 1.1 prevzatom z [7].



Obr. 1.1: CML priamka.

1.2 SML priamka

Priamka trhu cenných papierov (SML) je obvykle používaná pre stanovenie stredného výnosu a rizika jednotlivých aktív. Na rozdiel od CML priamky, ktorá sa používa pre eficientné portfólio, SML priamka oceňuje aktíva na základe pohybu trhového indexu. Tento fakt vyplýva z toho, že SML priamka rozlišuje systematické a individuálne riziko, pričom systematické riziko analyzovaného aktíva vstupuje do modelu prostredníctvom vzťahu aktíva k trhovému portfóliu. SML priamku možno odvodiť pre ľubovoľné prípustné portfólio, ktoré leží v nezápornom kvadrante (σ, \bar{r}) pod CML priamkou (pozri [6], [7] a [8]).

Poznámka: Nasledujúca veta je aj s dôkazom prevzatá z [6].

Veta 1. *Nech trhové portfólio M možno vyjadriť ako kombináciu n prípustných portfólií s váhami w_{iM} , $i = 1, \dots, n$. Označme stredné miery zisku týchto prípustných portfólií \bar{r}_i a ich riziká (kovariancie) σ_{ij} , $i, j = 1, \dots, n$. Potom pre ľubovoľné $i = 1, \dots, n$ platí:*

$$\bar{r}_i = r_f + \frac{\bar{r}_M - r_f}{\sigma_M^2} \sigma_{iM}, \quad (1.2)$$

kde $\sigma_{iM} = \text{cov}(r_i, r_M) = \sum_{j=1}^n w_{jM} \sigma_{ij}$ je kovariancia mier zisku i -teho prípustného portfólia a trhového portfólia M .

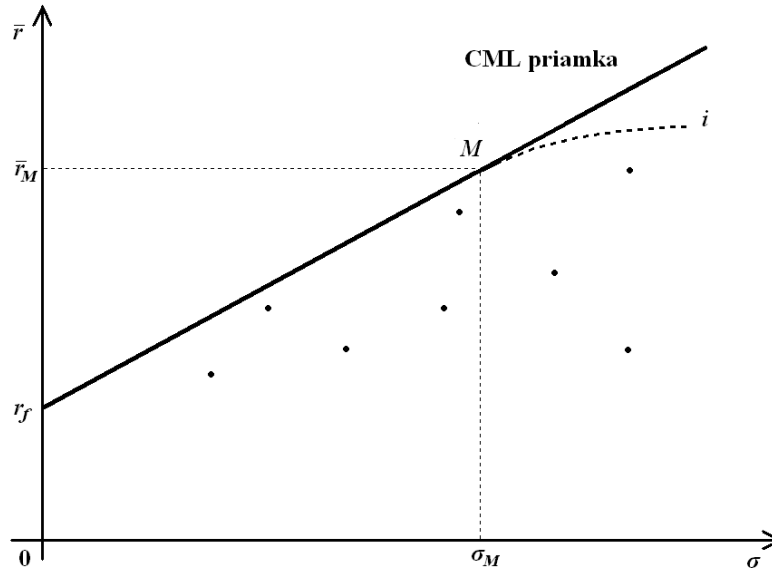
Dôkaz. Nech $\sigma_{iM} = cov(r_i, r_M) = \sum_{j=1}^n w_{jM} \sigma_{ij}$ je kovariancia mier zisku i -tého prípustného portfólia a trhového portfólia M . Potom smerodajnú odchýlku trhového portfólia možno vyjadriť v tvare:

$$\sigma_M = \left(\sum_{i=1}^n \sum_{j=1}^n w_{iM} w_{jM} \sigma_{ij} \right)^{\frac{1}{2}} = \left(\sum_{i=1}^n w_{iM} \sigma_{iM} \right)^{\frac{1}{2}}.$$

Nech P je ľubovoľné portfólio dané kombináciou portfólií i a M s váhami p_i a $1 - p_i$. Potom pre jeho strednú mieru zisku a smerodajnú odchýlku platí:

$$\begin{aligned} \bar{r}_P &= p_i \bar{r}_i + (1 - p_i) \bar{r}_M \\ \sigma_P &= [p_i^2 \sigma_i^2 + 2p_i(1 - p_i) \sigma_{iM} + (1 - p_i)^2 \sigma_M^2]^{\frac{1}{2}} \end{aligned}$$

Keďže portfólio P vzniklo kombináciou portfólií i a M , tak leží na krivke iM (pozri obr. 1.2).



Obr. 1.2: Odvodenie SML vzťahu.

Z obrázku je zrejmé, že smernica dotyčnice v bode M krivky iM sa rovná smernici CML priamky v bode M . Určíme preto všeobecne smernicu dotyčnice $\frac{d\bar{r}_P}{d\sigma_P}$ krivky iM v bode P :

$$\begin{aligned} \frac{d\bar{r}_P}{dp_i} &= \bar{r}_i - \bar{r}_M \\ \frac{d\sigma_P}{dp_i} &= \frac{p_i \sigma_i^2 + \sigma_{iM} - 2p_i \sigma_{iM} - \sigma_M^2 + p_i \sigma_M^2}{[p_i^2 \sigma_i^2 + 2p_i(1 - p_i) \sigma_{iM} + (1 - p_i)^2 \sigma_M^2]^{\frac{1}{2}}} \\ \Rightarrow \frac{d\bar{r}_P}{d\sigma_P} &= \frac{d\bar{r}_P}{dp_i} = \frac{(\bar{r}_i - \bar{r}_M) [p_i^2 \sigma_i^2 + 2p_i(1 - p_i) \sigma_{iM} + (1 - p_i)^2 \sigma_M^2]^{\frac{1}{2}}}{p_i \sigma_i^2 + \sigma_{iM} - 2p_i \sigma_{iM} - \sigma_M^2 + p_i \sigma_M^2}. \end{aligned} \quad (1.3)$$

V bode M je $p_i = 0$, preto z (1.3) platí:

$$\frac{d\bar{r}_M}{d\sigma_M} = \frac{(\bar{r}_i - \bar{r}_M) \sigma_M}{\sigma_{iM} - \sigma_M^2}. \quad (1.4)$$

Smernica CML priamky má tvar:

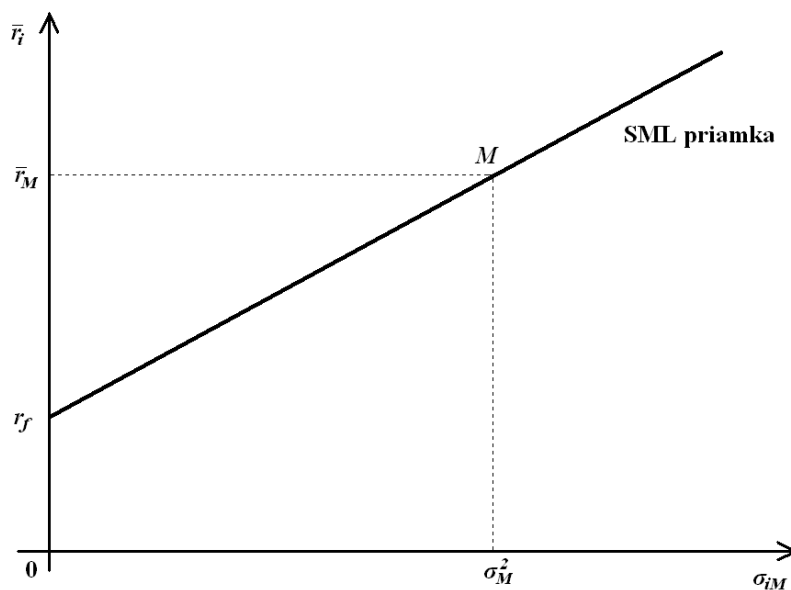
$$\frac{\bar{r}_M - r_f}{\sigma_M}. \quad (1.5)$$

Z rovnosti (1.4) a (1.5) dostávame:

$$\begin{aligned} \frac{(\bar{r}_i - \bar{r}_M)\sigma_M}{\sigma_{iM} - \sigma_M^2} &= \frac{\bar{r}_M - r_f}{\sigma_M} \\ (\bar{r}_i - \bar{r}_M)\sigma_M^2 &= (\bar{r}_M - r_f)(\sigma_{iM} - \sigma_M^2) \\ \bar{r}_i &= r_f + \frac{\bar{r}_M - r_f}{\sigma_M^2}\sigma_{iM}. \end{aligned}$$

□

Z predpisu (1.2) priamky SML je jasné, že riziko portfólia i je merané jeho kovarianciou s trhovým portfóliom a nie jeho smerodajnou odchýlkou ako u CML priamky vo vzťahu (1.1). Keďže je u SML priamky vzťah medzi mierou rizika a stredným výnosom portfólia i lineárny s kladnou smernicou, rizikovejšie portfóliá musia mať vyšší stredný výnos, čo ilustruje obrázok 1.3 prevzatý z [6].

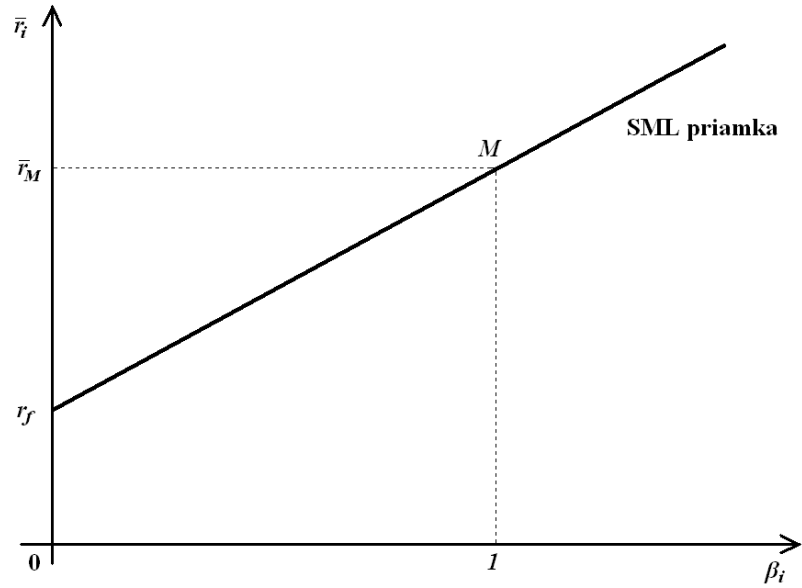


Obr. 1.3: SML priamka - kovariancia σ_{iM} .

Pri cenných papieroch, najmä pri akciách, sa často používa miera rizika beta, $\beta_i = \frac{\sigma_{iM}}{\sigma_M^2}$, pomocou ktorej možno upraviť vzťah (1.2) do tvaru:

$$\bar{r}_i = r_f + (\bar{r}_M - r_f) \beta_i, \quad i = 1, \dots, n, \quad (1.6)$$

čo znamená, že SML priamka predstavuje lineárny vzťah medzi rizikom investície do daného cenného papiera a zodpovedajúcou strednou mierou zisku. Inými slovami SML popisuje lineárnu závislosť stredného výnosu ľubovoľného aktíva na systematickom (trhovom) riziku meranom pomocou miery beta. Predpis SML priamky v tvare (1.6) je znázornený na obrázku 1.4 prevzatom z [6].



Obr. 1.4: SML priamka - miera beta.

Deterministický vzťah (1.6) je možné upraviť na stochastický, tzv. charakteristickú priamku portfólia, ktorý vysvetľuje mieru zisku i -teho portfólia ako náhodnú veličinu:

$$r_i = r_f + (r_M - r_f) \beta_i + \epsilon_i, \quad i = 1, \dots, n. \quad (1.7)$$

Reziduálna zložka ϵ_i reprezentuje náhodné odchýlky od deterministického lineárneho priebehu. Po jednoduchej úprave je predpis (1.7) v tvare:

$$r_i - r_f = (r_M - r_f) \beta_i + \epsilon_i, \quad i = 1, \dots, n, \quad (1.8)$$

ktorý je pre účely tejto práce použiteľnejší, keďže dáva do súvislosti odchýlky stredných výnosov portfólia i a trhového portfólia M od bezrizikovej miery zisku r_f . Z hľadiska štatistických odhadov sú k dispozícii pozorované miery zisku r_i , r_f , r_M a je možné odhadnúť mieru beta daného cenného papiera.

1.2.1 Koeficient beta

Koeficient beta je štandardizovaná miera rizika aktíva vzťahnutá k riziku celého trhu (t. j. miera systematického rizika). Predstavuje teda mieru rizika cenných papierov vo finančnej praxi. Po prepise (1.6) do tvaru:

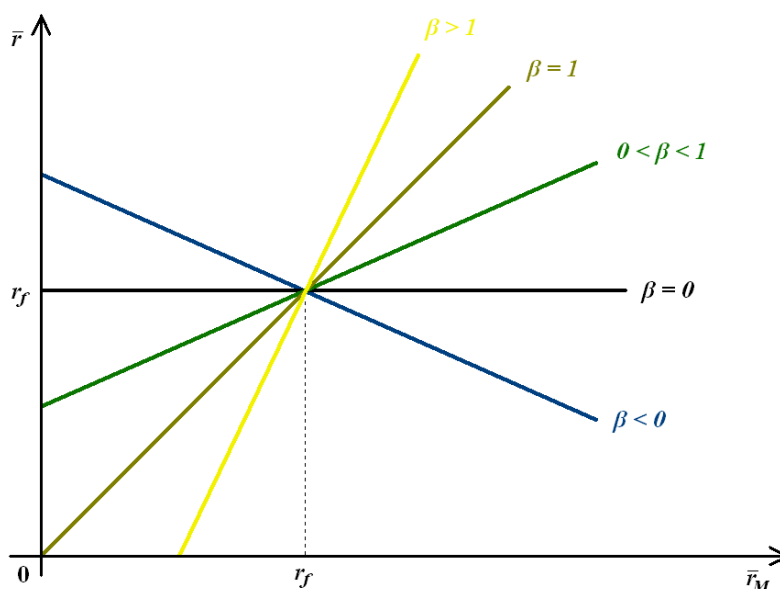
$$r_i - r_f = (r_M - r_f) \beta_i, \quad i = 1, \dots, n,$$

je zrejmé, že pre jednotkovú zmenu strednej miery zisku trhového portfólia (indexu cenných papierov) dochádza k väčšej zmene stredného výnosu uvažovaného cenného papiera. Je zrejmé, že čím väčší je koeficient β_i pre daný cenný papier, tým je toto aktívum nestabilnejšie. Rizikový faktor β_i preto vyjadruje citlivosť výnosu analyzovaného aktíva na zmeny trhového indexu. Je potrebné pripomenúť, že veľkosť koeficientu beta závisí aj na druhu podniku, ktorého akcie sa obchodujú (napr. energetika má $\beta < 1$ a poisťovníctvo $\beta > 1$).

Výnos analyzovaného aktíva (AA) vo vzťahu k veľkosti miery rizika sa vyvíja nasledovne (pozri tab. 1.1). Táto interpretácia miery beta je znázornená aj na obrázku 1.5 (pozri [8]).

$\beta < 0$	výnos AA sa pohybuje opačným smerom ako výnos trhového indexu
$\beta = 0$	výnos AA zodpovedá bezrizikóvemu výnosu
$0 < \beta < 1$	výnos AA sa pohybuje rovnakým smerom ako výnos trhového indexu, avšak pomalšie
$\beta = 1$	výnos AA zodpovedá výnosu trhového indexu, aktívum má priemerné systematické riziko
$\beta > 1$	výnos AA sa pohybuje rovnakým smerom ako výnos trhového indexu, avšak rýchlejšie

Tabuľka 1.1: Vývoj výnosu analyzovaného aktíva na základe koeficientu beta.



Obr. 1.5: Miera beta.

Taktiež pre koeficient beta platí:

$$\beta = \sum_{i=1}^n w_i \beta_i,$$

čo znamená, že pri kombinácii portfólií je miera beta kombináciou mier jednotlivých zložiek tohto portfólia so zodpovedajúcimi váhami.

Miera beta sa v praxi používa buď:

- *ex post* (t. j. napozorovaná z minulých dát) alebo
- *ex ante* (t. j. tak, že sa robia jej odhady do budúcnosti).

Stanovenie miery beta *ex post* je uvedené v praktickej časti tejto práce.

1.2.2 Miera nerovnováhy alfa

Miera alfa popisuje rozdiel medzi skutočnou strednou mierou zisku \bar{r}_i , $i = 1, \dots, n$ daného portfólia a tzv. vyváženou mierou zisku zo vzťahu (1.6):

$$\alpha_i = \bar{r}_i - (r_f + (\bar{r}_M - r_f)\beta_i) \quad (1.9)$$

Miera alfa teda umožňuje vyhľadávať cenné papiere vykazujúce rozdielne zisky v porovnaní s vyváženým priemerom pre uvažovaný typ finančného aktíva. Po úprave vzťahu (1.9) a zahrnutí reziduálnej zložky podobne ako v (1.8) vznikne:

$$r_i - r_f = \alpha_i + (r_M - r_f)\beta_i + \epsilon_i, \quad (1.10)$$

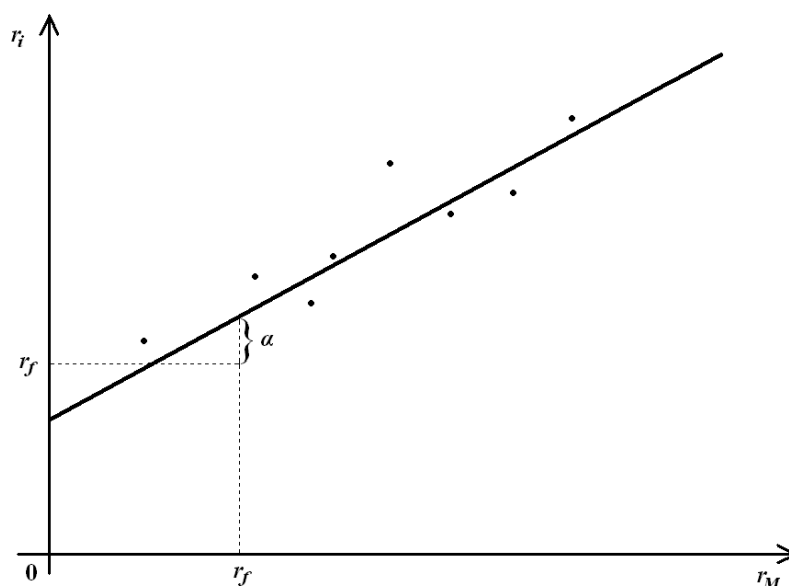
čo je opäť štatistický model lineárnej regresie. Rovnako ako u (1.8) sú dané historické výnosy r_i , r_f , r_M , avšak už je možné odhadnúť obe miery alfa i beta.

Podobne ako pre koeficient beta sa nižšie nachádza prehľadná tabuľka vzťahu výnosu analyzovaného aktíva (AA) s mierou nerovnováhy (pozri tab. 1.2).

$\alpha = 0$	AA je na trhu ocenené správne, t. j. medzi jeho výnosom a rizikom platia rovnovážne vzťahy dané SML priamkou
$\alpha > 0$	AA je na trhu podhodnotený, t. j. vykazuje v porovnaní s rovnovážnou hodnotou danou SML priamkou nadpriemerný výnos
$\alpha < 0$	AA je na trhu nadhodnotený, t. j. v porovnaní s očakávanou hodnotou vykazuje podpriemerný výnos

Tabuľka 1.2: Vývoj výnosu analyzovaného aktíva na základe koeficientu alfa.

Obrázok (1.6) znázorňuje mieru nerovnováhy alfa, pričom body v grafe zobrazujú pozorované hodnoty dvojíc (r_i, r_M) . Obrázok je prevzatý z knihy [8].



Obr. 1.6: Miera alfa.

Kapitola 2

Lineárny regresný model

Základnou myšlienkou teórie *lineárnej regresie* je určenie vzťahu medzi závislou premennou Y (vysvetľovaná premenná, regresand, efekt, odozva) a veličinami X_1, \dots, X_k (nezávislé premenné, vysvetľujúce premenné, regresory, príčiny, prediktory) pomocou lineárnej funkcie. Pre podrobnejšie informácie k tejto problematike možno siahnuť po knihách [1], [3] a [5].

Premennú Y je možné v regresnej analýze vyjadriť ako funkciu X_1, \dots, X_k v tvare:

$$Y = f(X_1, \dots, X_k) + \epsilon, \quad (2.1)$$

kde ϵ je náhodná chyba, ktorá môže vzniknúť napríklad nepresnosťami merania veličiny Y . Treba si však uvedomiť, že jednotlivé chyby nemožno pozorovať.

Regresný model (2.1) môže byť pre lineárnu funkciu f zapísaný nasledovne:

$$Y = \beta_1 X_1 + \dots + \beta_k X_k + \epsilon,$$

kde koeficienty β_1, \dots, β_k sú neznáme regresné parametre, ktoré sa na základe štatistických metód snažíme odhadnúť. Hľadáme teda najlepšiu lineárnu aproximáciu náhodnej veličiny Y pomocou X_1, \dots, X_k .

Cieľom v regresnej analýze je riešenie nasledujúcich problémov:

- (1) otestovať, na ktorých X_1, \dots, X_k závisí Y ,
- (2) zistiť ako na nich závisí,
- (3) predpovedať veličinu Y , ak sú k dispozícii pozorovania veličín X_1, \dots, X_k .

2.1 Jednorozmerná lineárna regresia

Pre potreby odhadu parametrov *jednorozmerného lineárneho regresného modelu* predpokladajme, že máme k dispozícii n pozorovaní Y_1, \dots, Y_n vysvetľovanej premennej \mathbf{Y} a n pozorovaných hodnôt vysvetľujúcich premenných $\mathbf{X}_1, \dots, \mathbf{X}_k$, t. j. hodnoty X_{ij} , $i = 1, \dots, n$, $j = 1, \dots, k$, čo možno zapísať ako vektor $\mathbf{Y} = (Y_1, \dots, Y_n)'$ a maticu $\mathbf{X}_{n \times k} = (x_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, k$, $k < n$.

Prvý stĺpec matice \mathbf{X} spravidla tvorí jednotkový vektor, aby nadrovina vyjadrujúca závislosť premennej \mathbf{Y} na premenných X_1, \dots, X_k neprechádzala vždy iba počiatkom.

Keďže sa v regresných modeloch nezavádzajú nadbytočné regresory, požadujeme lineárnu nezávislosť stĺpcov matice \mathbf{X} . Lineárny regresný model možno zapísať dvoma spôsobmi:

$$Y_i = \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \epsilon_i, \quad i = 1, \dots, n, \quad (2.2)$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.3)$$

kde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ je vektor neznámych parametrov a $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)'$ je vektor náhodných chýb (reziduálna zložka). V tomto modeli predpokladáme, že platí:

- vektor chýb $\boldsymbol{\epsilon}$ je vektor náhodných veličín a spĺňa podmienky:

$$E[\boldsymbol{\epsilon}] = \mathbf{0} \quad \text{a} \quad \text{var}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}, \quad (2.4)$$

kde \mathbf{I} je jednotková matica a $\sigma^2 > 0$ je neznámy parameter.

- \mathbf{X} je nenáhodná matica, teda aj vektor $\mathbf{X}\boldsymbol{\beta}$ je nenáhodný a preto platí:

$$E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta} \quad \text{a} \quad \text{var}[\mathbf{Y}] = \sigma^2 \mathbf{I}.$$

Z uvedeného je zrejmé, že zložky vektora $\boldsymbol{\epsilon}$ a teda aj \mathbf{Y} sú vzájomne nekorelované [26].

Model (2.3) možno rozpísať po zložkách:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

2.1.1 OLS-odhad

Nech existuje model v tvare (2.3) s predpokladmi na vektor reziduí (2.4). Za predpokladu plnej hodnosti matice \mathbf{X} sa k odhadu parametrov β_j , $j = 1, \dots, k$ najčastejšie používa *metóda najmenších štvorcov* (OLS). Metóda najmenších štvorcov spočíva v riešení úlohy:

$$\min_{\beta_1, \dots, \beta_k} \sum_{i=1}^n (Y_i - \sum_{j=1}^k x_{ij} \beta_j)^2 = \min_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \quad (2.5)$$

Riešením tejto minimalizačnej úlohy dostaneme vektor odhadov $\mathbf{b} = (b_1, \dots, b_k)'$, ktorý možno zapísať v tvare:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}.$$

Dôkaz tohto vzťahu možno nájsť napríklad v [1], strana 81.

Veta 2. Nech v modeli (2.3) platia predpoklady (2.4) a matica \mathbf{X} má plnú hodnotu. Potom pre odhad metódou najmenších štvorcov \mathbf{b} platí:

$$E[\mathbf{b}] = \boldsymbol{\beta} \quad \text{a} \quad \text{var}[\mathbf{b}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Dôkaz.

$$\begin{aligned} E[\mathbf{b}] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{Y}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} \\ \text{var}[\mathbf{b}] &= \text{var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{var}[\mathbf{Y}]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \end{aligned}$$

□

Výraz $S_e = (\mathbf{Y} - \widehat{\mathbf{Y}})'(\mathbf{Y} - \widehat{\mathbf{Y}})$, kde $\widehat{\mathbf{Y}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$ nazývame *reziduálny súčet štvorcov* a matica \mathbf{H} sa nazýva *projekčná matica*. Veličinu S_e vieme po jednoduchej úprave vyjadriť v tvare:

$$S_e = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}. \quad (2.6)$$

Keďže parameter σ^2 z rozptylu $\boldsymbol{\epsilon}$, resp. \mathbf{Y} nie je známy, je možné ho odhadnúť pomocou vzťahu:

$$s^2 = \frac{S_e}{n - k}. \quad (2.7)$$

Poznámka: Ak má vektor reziduí $\boldsymbol{\epsilon}$ normálne rozdelenie, tak OLS-odhad je zároveň tzv. *maximálne vierohodným odhadom*. Za predpokladu normálneho rozdelenia možno vykonať i štatistické testy pre regresné koeficienty (t-test, F-test) uvedené napríklad v [3].

2.2 Mnohorozmerná lineárna regresia

Mnohorozmerná lineárna regresia popisuje závislosť vektorových premenných $\mathbf{Y}_{\cdot 1}, \dots, \mathbf{Y}_{\cdot l}$ na vektorových premenných $\mathbf{X}_{\cdot 1}, \dots, \mathbf{X}_{\cdot k}$. Pre každú závislú premennú $\mathbf{Y}_{\cdot 1}, \dots, \mathbf{Y}_{\cdot l}$ je k dispozícii n pozorovaní, teda Y_{i1}, \dots, Y_{il} , $i = 1, \dots, n$. Mnohorozmerný lineárny regresný model môžno zapísať po zložkách:

$$\begin{pmatrix} Y_{11} & \dots & Y_{1l} \\ \vdots & \ddots & \vdots \\ Y_{n1} & \dots & Y_{nl} \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_{11} & \dots & \beta_{1l} \\ \vdots & \ddots & \vdots \\ \beta_{k1} & \dots & \beta_{kl} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} & \dots & \epsilon_{1l} \\ \vdots & \ddots & \vdots \\ \epsilon_{n1} & \dots & \epsilon_{nl} \end{pmatrix},$$

čo znamená, že vektory $\mathbf{Y}_{\cdot h}$, $h = 1, \dots, l$ predstavujú stĺpce matice \mathbf{Y} a podobne $\mathbf{X}_{\cdot j}$, $j = 1, \dots, k$ stĺpce matice $\mathbf{X} = (x_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, k$ (pozri [12]). Pre potreby OLS-odhadu je nutný predpoklad, že \mathbf{X} je regulárna.

Nech $\mathbf{Y}_{\cdot h} = (Y_{1h}, \dots, Y_{nh})'$, $h = 1, \dots, l$ je n -rozmerný vektor zodpovedajúci závislej premennej. Na každý takýto vektor možno aplikovať nasledujúci jednorozmerný lineárny regresný model

$$\mathbf{Y}_{\cdot h} = \mathbf{X}\boldsymbol{\beta}_{\cdot h} + \boldsymbol{\epsilon}_{\cdot h}, \quad h = 1, \dots, l, \quad (2.8)$$

kde $\mathbf{X}_{n \times k} = (\mathbf{X}_{\cdot 1}, \dots, \mathbf{X}_{\cdot k})$ je matica, ktorá má spravidla v prvom stĺpci vektor jednotiek a $\boldsymbol{\epsilon}_{\cdot h}$ sú reziduá spĺňajúce:

$$E\boldsymbol{\epsilon}_{\cdot h} = \mathbf{0} \quad \text{a} \quad \text{var}\boldsymbol{\epsilon}_{\cdot h} = \sigma_{hh}\mathbf{I}_n,$$

pričom $\sigma_{hh} > 0$, $h = 1, \dots, l$.

Je dôležité podotknúť, že matica \mathbf{X} je pre každý vektor $\mathbf{Y}_{\cdot h}$, $h = 1, \dots, l$ rovnaká, ale vektor reziduí $\boldsymbol{\epsilon}_{\cdot h} = (\epsilon_{1h}, \dots, \epsilon_{nh})'$ a vektor regresných parametrov $\boldsymbol{\beta}_{\cdot h} = (\beta_{1h}, \dots, \beta_{kh})'$ môžu byť rôzne pre rôzne $h = 1, \dots, l$.

Mnohorozmerný lineárny model môže byť uvedený v maticovom tvare:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.9)$$

kde $\mathbf{Y}_{n \times l} = (\mathbf{Y}_{\cdot 1}, \dots, \mathbf{Y}_{\cdot l})$, $\mathbf{X}_{n \times k} = (\mathbf{X}_{\cdot 1}, \dots, \mathbf{X}_{\cdot k})$, $\boldsymbol{\beta}_{k \times l} = (\boldsymbol{\beta}_{\cdot 1}, \dots, \boldsymbol{\beta}_{\cdot l})$ a $\boldsymbol{\epsilon}_{n \times l} = (\boldsymbol{\epsilon}_{\cdot 1}, \dots, \boldsymbol{\epsilon}_{\cdot l})$.

2.2.1 OLS-odhad

Zápis (2.8) nepredpokladá žiadne vzájomné vzťahy medzi premennými $\mathbf{Y}_{\cdot h}$, $h = 1, \dots, l$ (resp. medzi vektormi chýb $\boldsymbol{\epsilon}_{\cdot h}$, $h = 1, \dots, l$), t. j. jednotlivé rovnice je možné považovať za navzájom nezávislé. Pre tento model preto platí analogicky ako v kap. 2.1.1, že odhad metódou najmenších štvorcov je v tvare:

$$\mathbf{b}_{\cdot h} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_{\cdot h}, \quad h = 1, \dots, l. \quad (2.10)$$

Stredná hodnota odhadu $\mathbf{b}_{\cdot h}$ je $E\mathbf{b}_{\cdot h} = \boldsymbol{\beta}_{\cdot h}$ a rozptyl je $\text{var}\mathbf{b}_{\cdot h} = \sigma_{hh}(\mathbf{X}'\mathbf{X})^{-1}$, $h = 1, \dots, l$.

Medzi jednotlivými rovnicami sa dá zohľadniť možná závislosť. V ďalšom texte treba predpokladať, že $\text{cov}(\boldsymbol{\epsilon}_{\cdot i}, \boldsymbol{\epsilon}_{\cdot j}) = E[\boldsymbol{\epsilon}_{\cdot i}\boldsymbol{\epsilon}_{\cdot j}'] = \sigma_{ij}\mathbf{I}$, t. j.:

$$\begin{aligned} E[\epsilon_{\mu i}\epsilon_{\nu j}] &= \sigma_{ij}, \quad \mu = \nu, \\ &= 0, \quad \mu \neq \nu. \end{aligned}$$

Predpokladom teda je, že zodpovedajúce zložky vektorov $\boldsymbol{\epsilon}_{\cdot i}$, $\boldsymbol{\epsilon}_{\cdot j}$ sú vzájomne korelované. Pre potreby ďalších výpočtov (najmä pre účely odhadu parametrov $\boldsymbol{\beta}$) je vhodné použiť model (2.9) v tvare:

$$\begin{pmatrix} \mathbf{Y}_{\cdot 1} \\ \mathbf{Y}_{\cdot 2} \\ \vdots \\ \mathbf{Y}_{\cdot l} \end{pmatrix} = \begin{pmatrix} \mathbf{X} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{\cdot 1} \\ \boldsymbol{\beta}_{\cdot 2} \\ \vdots \\ \boldsymbol{\beta}_{\cdot l} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_{\cdot 1} \\ \boldsymbol{\epsilon}_{\cdot 2} \\ \vdots \\ \boldsymbol{\epsilon}_{\cdot l} \end{pmatrix}, \quad (2.11)$$

kde vektor chýb má strednú hodnotu $\mathbf{0}$ a variančnú maticu

$$\begin{pmatrix} \sigma_{11}\mathbf{I}_n & \sigma_{12}\mathbf{I}_n & \dots & \sigma_{1l}\mathbf{I}_n \\ \sigma_{12}\mathbf{I}_n & \sigma_{22}\mathbf{I}_n & \dots & \sigma_{2l}\mathbf{I}_n \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1l}\mathbf{I}_n & \sigma_{2l}\mathbf{I}_n & \dots & \sigma_{ll}\mathbf{I}_n \end{pmatrix}.$$

Tento model je možné interpretovať ako jednorozmerný, lebo ľavá strana (2.11) je nl -rozmerný vektor a pravá strana tohto modelu je súčin matice typu $nl \times kl$ a kl -rozmerného vektoru sčítaný s nl -rozmerným vektorom reziduí (pozri [5] a [12]).

Definícia. (Operátor Vec) *Nech $\mathbf{Y} = (Y_{ij})$ je matica typu $n \times l$. Potom $\text{Vec}(\mathbf{Y})$ je definovaný ako nl -rozmerný stĺpcový vektor, pre ktorý platí*

$$\text{Vec}(\mathbf{Y}) = (Y_{11}, \dots, Y_{n1}, Y_{12}, \dots, Y_{n2}, \dots, Y_{1l}, \dots, Y_{nl})'.$$

Poznámka: Pri zápise matice \mathbf{Y} typu $n \times l$ v tvare $\mathbf{Y} = (\mathbf{Y}_{\cdot 1}, \dots, \mathbf{Y}_{\cdot l})$ je $\text{Vec}(\mathbf{Y})$ v tvare:

$$\text{Vec}(\mathbf{Y}) = \begin{pmatrix} \mathbf{Y}_{\cdot 1} \\ \vdots \\ \mathbf{Y}_{\cdot l} \end{pmatrix} = (\mathbf{Y}'_{\cdot 1}, \dots, \mathbf{Y}'_{\cdot l})'.$$

Definícia. (Kroneckerov súčin matíc) *Nech \mathbf{A} je matica typu $m \times l$ a \mathbf{B} je matica typu $n \times k$. Potom Kroneckerov súčin matíc \mathbf{A} a \mathbf{B} je definovaný ako*

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1l}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \dots & a_{ml}\mathbf{B} \end{pmatrix}.$$

Lemma 1. *Pre Kroneckerov súčin matíc platí*

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD},$$

ak je násobenie matíc definované.

Dôkaz. Rozpísaním ľavej strany po zložkách vznikne:

$$\begin{aligned} (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) &= \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1l}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \dots & a_{ml}\mathbf{B} \end{pmatrix} \begin{pmatrix} c_{11}\mathbf{D} & \dots & c_{1n}\mathbf{D} \\ \vdots & \ddots & \vdots \\ c_{l1}\mathbf{D} & \dots & c_{ln}\mathbf{D} \end{pmatrix} = \\ &= \begin{pmatrix} \sum_{j=1}^l a_{1j}c_{j1}\mathbf{BD} & \dots & \sum_{j=1}^l a_{1j}c_{jn}\mathbf{BD} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^l a_{mj}c_{j1}\mathbf{BD} & \dots & \sum_{j=1}^l a_{mj}c_{jn}\mathbf{BD} \end{pmatrix}. \end{aligned}$$

Rozpísaním pravej strany po zložkách vznikne rovnaká matrica:

$$\begin{aligned} \mathbf{AC} \otimes \mathbf{BD} &= \begin{pmatrix} \sum_{j=1}^l a_{1j}c_{j1} & \dots & \sum_{j=1}^l a_{1j}c_{jn} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^l a_{mj}c_{j1} & \dots & \sum_{j=1}^l a_{mj}c_{jn} \end{pmatrix} \otimes \mathbf{BD} = \\ &= \begin{pmatrix} \sum_{j=1}^l a_{1j}c_{j1}\mathbf{BD} & \dots & \sum_{j=1}^l a_{1j}c_{jn}\mathbf{BD} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^l a_{mj}c_{j1}\mathbf{BD} & \dots & \sum_{j=1}^l a_{mj}c_{jn}\mathbf{BD} \end{pmatrix}. \end{aligned}$$

□

Lemma 2. Pre operátor Vec platí

$$Vec(\mathbf{AXB}) = (\mathbf{B}' \otimes \mathbf{A})Vec(\mathbf{X}). \quad (2.12)$$

Dôkaz. Nech $\mathbf{X}_{n \times k} = (\mathbf{X}_{\cdot 1}, \dots, \mathbf{X}_{\cdot k})$ a $\mathbf{B}_{n \times k} = (b_{ij}) = (\mathbf{B}_{\cdot 1}, \dots, \mathbf{B}_{\cdot k})$. Potom s -tý stĺpec matice \mathbf{AXB} , $s = 1, \dots, l$ je:

$$\begin{aligned} (\mathbf{AXB})_{\cdot s} &= \mathbf{AXB}_{\cdot s} = \mathbf{A} \sum_{i=1}^k \mathbf{X}_{\cdot i} b_{is} = (b_{1s}\mathbf{A}, \dots, b_{ks}\mathbf{A})(\mathbf{X}'_{\cdot 1}, \dots, \mathbf{X}'_{\cdot k})' = \\ &= [(b_{1s}, \dots, b_{ks})\mathbf{A}](\mathbf{X}'_{\cdot 1}, \dots, \mathbf{X}'_{\cdot k})' = (\mathbf{B}'_{\cdot s} \otimes \mathbf{A})Vec(\mathbf{X}). \end{aligned}$$

Odtiaľ plynie dokazovaný vzťah:

$$Vec(\mathbf{AXB}) = \begin{pmatrix} (\mathbf{AXB})_{\cdot 1} \\ \vdots \\ (\mathbf{AXB})_{\cdot l} \end{pmatrix} = \begin{pmatrix} \mathbf{B}'_{\cdot 1} \otimes \mathbf{A} \\ \vdots \\ \mathbf{B}'_{\cdot l} \otimes \mathbf{A} \end{pmatrix} Vec(\mathbf{X}) = (\mathbf{B}' \otimes \mathbf{A})Vec(\mathbf{X}).$$

□

Lemma 3. Pre operátor Vec platí

$$Vec(\mathbf{A} + \mathbf{B}) = Vec(\mathbf{A}) + Vec(\mathbf{B}). \quad (2.13)$$

Dôkaz. Dôkaz plynie z definície rozpísaním uvedeného vzťahu po zložkách. □

Ďalšie vzťahy pre Kroneckerov súčin a operátor Vec možno nájsť v článku [4]. Analógiou vzťahu (2.12) pre $Vec(\mathbf{XB})$ dostávame:

$$Vec(\mathbf{X}_{n \times k} \mathbf{B}_{k \times l}) = Vec(\mathbf{X}_{n \times k} \mathbf{B}_{k \times l} \mathbf{I}_l) = (\mathbf{I}_l \otimes \mathbf{X})Vec(\mathbf{B}).$$

Na základe vyššie uvedených vzťahov (2.12) a (2.13) je teda možné prepísať model (2.11) do tvaru:

$$Vec(\mathbf{Y}) = (\mathbf{I}_l \otimes \mathbf{X})Vec(\boldsymbol{\beta}) + Vec(\boldsymbol{\epsilon}), \quad (2.14)$$

kde pre $Vec(\boldsymbol{\epsilon})$ platí:

$$E[Vec(\boldsymbol{\epsilon})] = \mathbf{0} \quad \text{a} \quad var[Vec(\boldsymbol{\epsilon})] = \boldsymbol{\Sigma} \otimes \mathbf{I}_n,$$

pričom $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1}^l$, za predpokladu, že $\boldsymbol{\Sigma}$ je pozitívne definitná matica.

Model (2.14) prepíšeme do nasledujúceho tvaru:

$$\mathbf{y} = \mathbf{x}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\epsilon}},$$

kde $\mathbf{y} = Vec(\mathbf{Y})$, $\mathbf{x} = (\mathbf{I}_l \otimes \mathbf{X})$, $\tilde{\boldsymbol{\beta}} = Vec(\boldsymbol{\beta})$, $\tilde{\boldsymbol{\epsilon}} = Vec(\boldsymbol{\epsilon})$. Potom odhad parametrov v tomto modeli je:

$$\tilde{\mathbf{b}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}. \quad (2.15)$$

V modeli (2.9) je teda matica odhadov parametrov $\boldsymbol{\beta}$ daná vzťahom:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

pričom platí, že $Vec(\mathbf{b}) = \tilde{\mathbf{b}}$. Dôkaz tohto vzťahu pre odhad parametrov mnohorozmerného lineárneho modelu vychádza z predchádzajúcich definícií a dokázateľných lemat, pretože platí:

$$\begin{aligned}\mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{Y}_{.1}, \dots, \mathbf{Y}_{.l}] = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_{.1}, \dots, (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_{.l}] = \\ &= [\mathbf{b}_{.1}, \dots, \mathbf{b}_{.l}].\end{aligned}$$

Z vyššie ukázaného vzťahu vyplýva, že OLS-odhad odhadnutý po zložkách, je rovnaký, ako simultánny odhad. Je ale dôležité podotknúť, že variančné matice týchto odhadov nie sú totožné a iba simultánny odhad berie do úvahy všetky korelácie medzi náhodnými veličinami matice \mathbf{Y} (modely (2.8) a (2.11) nie sú všeobecne totožné, pokiaľ sa nepredpokladá, že $var[Vec(\boldsymbol{\epsilon})] = \boldsymbol{\Sigma} \otimes \mathbf{I}$).

Veta 3. (1) *Nech platia všetky predpoklady pre model (2.8) pre OLS-odhad $\mathbf{b}_{.h}$, $h = 1, \dots, l$ po zložkách v tvare (2.10). Potom platí $var[\mathbf{b}_{.h}] = \sigma_{hh}(\mathbf{X}'\mathbf{X})^{-1}$, $h = 1, \dots, l$, t. j. $[var\mathbf{b}_{.1}, \dots, var\mathbf{b}_{.l}] = diag(\boldsymbol{\Sigma}) \otimes (\mathbf{X}'\mathbf{X})^{-1}$.*

(2) *Nech platia všetky predpoklady pre model (2.11) pre simultánny OLS-odhad \mathbf{b} v tvare (2.15). Potom platí $var[\tilde{\mathbf{b}}] = \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}$.*

Dôkaz. (1) Pre odhad po zložkách $\mathbf{b}_{.h}$, $h = 1, \dots, l$ platí analogicky, ako bolo dokázané v kap. 2.1.1, že $var[\mathbf{b}_{.h}] = \sigma_{hh}(\mathbf{X}'\mathbf{X})^{-1}$, $h = 1, \dots, l$. Odtiaľ priamo plynie, že $[var\mathbf{b}_{.1}, \dots, var\mathbf{b}_{.l}] = diag(\boldsymbol{\Sigma}) \otimes (\mathbf{X}'\mathbf{X})^{-1}$.

(2) Pre simultánny odhad \mathbf{b} platí, že $var[Vec(\mathbf{b})] = var[\tilde{\mathbf{b}}]$. Potom analogicky ako v kap. 2.1.1 dostávame:

$$var[\tilde{\mathbf{b}}] = var[Vec(\mathbf{b})] = var[(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}] = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x} var[\mathbf{y}]\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}.$$

Platí:

$$\begin{aligned}(\mathbf{x}'\mathbf{x})^{-1} &= [(\mathbf{I}_l \otimes \mathbf{X}')(\mathbf{I}_l \otimes \mathbf{X})]^{-1} = [\mathbf{I}_l \otimes (\mathbf{X}'\mathbf{X})]^{-1} = \mathbf{I}_l \otimes (\mathbf{X}'\mathbf{X})^{-1}, \\ var[\mathbf{y}] &= var[\tilde{\boldsymbol{\epsilon}}].\end{aligned}$$

Odtiaľ plynie:

$$\begin{aligned}var [Vec(\mathbf{b})] &= \\ &= (\mathbf{I}_l \otimes (\mathbf{X}'\mathbf{X})^{-1})(\mathbf{I}_l \otimes \mathbf{X})(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)(\mathbf{I}_l \otimes \mathbf{X}')(\mathbf{I}_l \otimes (\mathbf{X}'\mathbf{X})^{-1}) = \\ &= (\mathbf{I}_l \otimes (\mathbf{X}'\mathbf{X})^{-1})(\boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X}))(\mathbf{I}_l \otimes (\mathbf{X}'\mathbf{X})^{-1}) = \\ &= \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} = \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}.\end{aligned}$$

□

Poznámka: Pre štatistické testy je potrebné predpokladať normálne rozdelenie reziduí ϵ_{ih} , teda treba predpokladať mnohorozmerné normálne rozdelenie pre $\boldsymbol{\epsilon}_{.h}$.

Kapitola 3

Teoretický základ štatistických odhadov

V nasledujúcich podkapitolách je uvedený teoretický základ pre prácu so štatistickými dátami. Teória pravdepodobnosti študuje matematické modely náhodných dejov. Preto táto kapitola uvádza niekoľko základných definícií, ktoré sú potrebné pre pochopenie ďalšieho textu.

3.1 Parametrický štatistický model

Definícia. (Náhodný výber) *Nech X_1, X_2, \dots je postupnosť nezávislých rovnako rozdelených náhodných veličín s rozdelením pravdepodobností P . Potom postupnosť X_1, X_2, \dots sa nazýva náhodný výber.*

Definícia. (Distribučná funkcia) *Nech X je náhodná veličina s rozdelením pravdepodobností P . Reálna funkcia:*

$$F(x) = P(X \leq x), \quad -\infty < x < \infty$$

sa nazýva distribučná funkcia náhodnej veličiny X .

Definícia. (Parameter rozdelenia) *Nech X je náhodná veličina s rozdelením pravdepodobností P , ktoré patrí do triedy $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ indexovanej parametrom $\theta \in \Theta \subseteq \mathbb{R}^n, n \in \mathbb{N}$. Potom parametrom rozdelenia P je akákoľvek hodnota $\theta = T(P)$, ktorá závisí na P .*

Nech F je distribučná funkcia príslušná P . Potom sa $\theta = T(F)$.

T nazýva štatistický funkcionál.

Definícia. (Parametrický štatistický model) *Nech X je náhodná veličina s rozdelením pravdepodobností $P, P \in \mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^n, n \in \mathbb{N}\}$. Potom sa trojica $\{\mathcal{X}, \mathcal{B}, P\}$ nazýva parametrický štatistický model.*

Obvykle sa za \mathcal{X} volí:

- $\mathcal{X} = \mathbb{R} \Rightarrow X(\omega) \in \mathcal{X}$ je reálna náhodná veličina,
- $\mathcal{X} = \mathbb{R}^n \Rightarrow X(\omega) \in \mathcal{X}$ je n -rozmerný náhodný vektor,
- $\mathcal{X} = \mathbb{R}^{n \times k} \Rightarrow$ náhodný pokus vedie ku k nezávislým n -rozmerným pozorovaniam.

Prirodzeným odhadom θ založeným na pozorovaniach X_1, \dots, X_n je $\hat{\theta}_n = T(P_n)$, kde P_n je empirické rozdelenie pravdepodobností vektora (X_1, \dots, X_n) :

$$P_n = \frac{1}{n} \sum_{i=1}^n I_{[X_i \in A]}, \quad A \in \mathcal{B}.$$

Distribučná funkcia príslušná P_n je empirická distribučná funkcia:

$$F_n(x) = P_n(X \leq x) = \frac{1}{n} \sum_{i=1}^n I_{[X_i \leq x]}, \quad x \in \mathbb{R}.$$

3.2 Vlastnosti štatistických odhadov

Definícia. (Nestrannosť) Nech X_1, \dots, X_n je náhodný výber s rozdelením pravdepodobností $P \in \mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^n\}$ a P_n je empirické rozdelenie pravdepodobností vektora (X_1, \dots, X_n) . Potom sa funkcionál $T(P_n)$ nazýva nestranným odhadom parametra θ , ak platí:

$$ET(P_n) = \theta, \quad \forall \theta \in \Theta.$$

Funkcionál $T^*(P_n)$ sa nazýva najlepším nestranným odhadom parametra θ , ak platí:

$$ET^*(P_n) = \theta \quad \text{a} \quad \text{var}T^*(P_n) \leq \text{var}T(P_n)$$

pre ľubovoľný iný funkcionál $T(P_n)$ spĺňajúci $ET(P_n) = \theta, \forall \theta \in \Theta$.

Definícia. (Slabá konzistencia) Nech X_1, \dots, X_n je náhodný výber s rozdelením pravdepodobností $P \in \mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^n\}$ a P_n je empirické rozdelenie pravdepodobností vektora (X_1, \dots, X_n) . Potom sa funkcionál $T(P_n)$ nazýva slabo konzistentným odhadom parametra θ , ak $T(P_n)$ konverguje v pravdepodobnosti k parametra θ :

$$T(P_n) \xrightarrow{P} \theta, \quad n \rightarrow \infty, \quad \forall \theta \in \Theta.$$

Poznámka: Konvergenciou postupnosti náhodných veličín X_1, X_2, \dots v pravdepodobnosti k náhodnej veličine C sa rozumie:

$$P(|X_n - C| \geq \epsilon) \rightarrow 0, \quad n \rightarrow \infty, \quad \epsilon > 0,$$

v skrátanom značení $X_n \xrightarrow{P} C, \quad n \rightarrow \infty$.

Definícia. (Vychýlenie) Nech X_1, \dots, X_n je náhodný výber s rozdelením pravdepodobností $P \in \mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^n\}$ a P_n je empirické rozdelenie pravdepodobností vektora (X_1, \dots, X_n) . Potom

$$E[T(P_n) - \theta]$$

sa nazýva vychýlenie odhadu $T(P_n)$.

Definícia. (Fisherovská konzistencia) Nech X_1, \dots, X_n je náhodný výber s rozdelením pravdepodobností $P \in \mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^n\}$ a P_n je empirické rozdelenie pravdepodobností vektora (X_1, \dots, X_n) . Potom odhad $\hat{\theta}_n$ nazývame fisherovsky konzistentným odhadom parametra θ , ak pre jeho funkcionál $\hat{\theta}_n = T(P_n)$ platí $T(P) = \theta$.

Fisherovská konzistencia je prirodzená vlastnosť odhadov a z hľadiska robustnosti je dôležitejšia než nevychýlenosť (nestrannosť) odhadov.

Kapitola 4

Robustné štatistické odhady

Klasické odhady, ako napríklad priemer, výberová smerodajná odchýlka alebo odhad metódou najmenších štvorcov v modeli lineárnej regresie, môžu byť veľmi ovplyvnené odľahlými pozorovaniami (stačí jedno odľahlé pozorovanie mimo hlavnú dátovú vzorku a odhad bude veľmi nepresný). Tieto klasické metódy často zlyhávajú a neposkytujú dobrý odhad hlavnej dátovej vzorky. Vplyv týchto odľahlých pozorovaní je vhodné minimalizovať. K tomu slúžia rôzne štatistické metódy.

Jednou z možností je odľahlé pozorovania odhaliť a nejakým spôsobom vplyv týchto pozorovaní v dátovom vzorke eliminovať, čo popisuje kap. 4.1. Druhou, alternatívnou možnosťou, je použitie pokročilých štatistických metód, napr. robustných štatistických odhadov.

Robustné štatistické odhady sa používajú za účelom lepšieho odhadu určitého parametra, ak sú v dátových vzorkách nejaké odľahlé pozorovania. Tieto pozorovania sú separované od hlavnej dátovej vzorky alebo sa iným spôsobom odlišujú od ostatných dát. Robustné metódy sa s týmito pozorovaniami dokážu vysporiadať bez toho, aby bolo nutné jednotlivé odľahlé pozorovania vopred detekovať.

Na rozdiel od klasických metód, robustné štatistické metódy poskytujú dobrý odhad v prípade, keď sa vo vzorke vyskytujú odľahlé pozorovania, ale aj keď vzorka neobsahuje žiadne odľahlé pozorovania.

4.1 Detekcia odľahlých pozorovaní klasickými nástrojmi

V nasledujúcej časti sú uvedené klasické prístupy k odľahlým pozorovaniam, a to samostatne pre jednorozmerný model lineárnej regresie a samostatne pre mnohorozmerný model lineárnej regresie. Pre podrobnosti k teórii modelov pozri kap. 2.1 a 2.2.

4.1.1 Detekcia odľahlých pozorovaní v jednorozmernom modeli

Odľahlé pozorovania su také pozorovania, ktoré sú separované od hlavnej dátovej vzorky, alebo sa nejakým spôsobom odlišujú od ostatných dát. Poznatky uvedené v tejto kapitole možno nájsť v [18]. Problematiku bližšie študujú aj knihy [17] a [20].

Klasickým štatistickým nástrojom pre detekciu odľahlých pozorovaní X_i z náhodného výberu X_1, \dots, X_n je:

$$t_i = \frac{X_i - \bar{X}}{s},$$

kde \bar{X} je priemer náhodného výberu a s udáva výberovú smerodajnú odchýlku. K detekcii odľahlých pozorovaní sa použije tzv. pravidlo 3σ , t. j. pozorovania, ktoré majú $|t_i| > 3$ sú považované za podozrivé. Toto pravidlo je založené na predpoklade, že tieto podozrivé pozorovania sú pravdepodobne z iného, než normálneho rozdelenia. Pre náhodnú veličinu z normovaného normálneho rozdelenia totiž platí:

$$P(|x| > 3) \doteq 0,003.$$

Bežne používanými metódami, ako sa s odľahlými pozorovaniami vysporiadať (s cieľom udržať normalitu pozorovaní) sú:

- odstránenie pozorovaní, pre ktoré platí $|t_i| > 3$, alebo
- nahradenie pozorovania, ktoré spôsobuje nenormalitu pozorovaním $\bar{X} \pm 3s$ (je možné nahradiť tým, ktoré je bližšie).

Pravidlo 3σ ale nefunguje rovnako v každom prípade, ako napr.:

- ak je v náhodnom výbere niekoľko odľahlých pozorovaní, ich efekty sa môžu navzájom rušiť a tieto odľahlé pozorovania nemusia byť rozpoznateľné,
- vo veľmi veľkom náhodnom výbere „dobrých“ dát môžu byť niektoré pozorovania vyhodnotené ako podozrivé (v normálne rozdelenom náhodnom výbere s počtom pozorovaní 1000 budú 3 pozorovania, ktoré budú vykazovať $|t_i| > 3$). V tomto prípade je lepšie, ak sa použije tzv. normálny QQ -graf,
- vo veľmi malom náhodnom výbere je pravidlo neefektívne, lebo pre všetky možné hodnoty platí:

$$|t_i| < \frac{n-1}{\sqrt{n}}. \quad (4.1)$$

Ak je teda $n \leq 10$, tak vždy platí $|t_i| < 3$.

Ukážeme teraz, že platí (4.1). Pre náhodný výber X_1, \dots, X_n platí:

$$\sum_{i=1}^n (X_i - \bar{X}) = 0.$$

Odtiaľ jednoduchým prepisom dostávame:

$$X_i - \bar{X} = - \sum_{\substack{j=1 \\ j \neq i}}^n (X_j - \bar{X})$$

$$(X_i - \bar{X})^2 = \left(\sum_{\substack{j=1 \\ j \neq i}}^n (X_j - \bar{X}) \right)^2.$$

Potom, s použitím Schwarzovej nerovnosti $(\sum a_j b_j)^2 \leq (\sum a_j^2) (\sum b_j^2)$, platí:

$$\begin{aligned} (X_i - \bar{X})^2 &\leq (n-1) \sum_{\substack{j=1 \\ j \neq i}}^n (X_j - \bar{X})^2 = \\ &= (n-1) \left[\sum_{j=1}^n (X_j - \bar{X})^2 - (X_i - \bar{X})^2 \right] \\ n (X_i - \bar{X})^2 &\leq (n-1) \sum_{j=1}^n (X_j - \bar{X})^2 \\ (X_i - \bar{X})^2 &\leq \frac{n-1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 = \\ &= \frac{(n-1)^2}{n} \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2 = \frac{(n-1)^2}{n} s^2 \\ \implies \frac{(X_i - \bar{X})^2}{s^2} &\leq \frac{(n-1)^2}{n} \\ t_i^2 &\leq \frac{(n-1)^2}{n} \\ |t_i| &\leq \frac{n-1}{\sqrt{n}}. \end{aligned}$$

4.1.2 Detekcia odľahlých pozorovaní v mnohorozmernom modeli

V mnohorozmernej regresii nie je detekcia tak jednoduchá a intuitívna, ako v jednorozmernej regresii. V mnohých prípadoch mnohorozmerné pozorovania nemôžu byť považované za odľahlé pozorovania, ak je každá premenná posudzovaná nezávisle. Testovanie odľahlých pozorovaní musí vziať do úvahy aj vzťah medzi premennými.

Štatistické metódy pre detekciu odľahlých pozorovaní v mnohorozmernom prípade väčšinou indikujú tie pozorovania, ktoré sa nachádzajú relatívne „ďaleko“ od stredu rozdelenia. Často sa používa tzv. *Mahalanobisova vzdialenosť* založená na odhadnutých parametroch mnohorozmerného rozdelenia.

Nech existuje n pozorovaní k -rozmerných náhodných výberov $\mathbf{X}_i, i = 1, \dots, n$. Výberový priemer sa označuje $\bar{\mathbf{X}}$ a výberová variančná matica \mathbf{S} . Platí:

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$$

Potom Mahalanobisova vzdialenosť pre výber $i, i = 1, \dots, n$ je daná vzťahom:

$$M_i = \left[\sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) \right]^{\frac{1}{2}}.$$

Veľká hodnota štatistiky M_i indikuje odľahlé pozorovanie. Viac informácií k tejto problematike možno nájsť v [17].

4.2 Charakteristiky robustnosti

V nasledujúcich podkapitolách sú vysvetlené charakteristiky robustnosti. Pre účely tejto kapitoly je potrebné zadefinovať niekoľko dôležitých vzťahov z teórie miery, funkcionálnej analýzy a teórie pravdepodobnosti. Tieto vzťahy uvedieme v nasledujúcej podkapitole. Viac informácií možno nájsť v knihách [9], [10], [13], [14] a [18].

4.2.1 Teoretický základ

Nech pre účely ďalšieho textu platí, že výberový priestor \mathcal{X} je metrický priestor s metrikou d a je úplný a separabilný. Nech \mathcal{B} je σ -algebra borelovských podmnožín \mathcal{X} (pozri [13]).

Definícia. (Prohorovova vzdialenosť) *Nech \mathcal{P} je systém všetkých pravdepodobnostných mier na $(\mathcal{X}, \mathcal{B})$. Potom \mathcal{P} je konvexná množina, na ktorej možno definovať Prohorovovu vzdialenosť dvoch prvkov $P, Q \in \mathcal{P}$:*

$$d_P(P, Q) = \inf\{\epsilon > 0 : P(A) \leq Q(A^\epsilon) + \epsilon, \forall A \in \mathcal{B}, A \neq \emptyset\},$$

kde $A^\epsilon = \{x \in \mathcal{X} : \inf_{y \in A} d(x, y) \leq \epsilon\}$ je uzavreté ϵ -okolie neprázdnej množiny A .

Množina \mathcal{P} všetkých pravdepodobnostných mier na $(\mathcal{X}, \mathcal{B})$ je metrický priestor (\mathcal{P}, δ) všetkých rozdelení pravdepodobností vzhľadom k metrike δ (napr. $\delta = d_P$). Preto na \mathcal{P} sa dá študovať spojitosť a ďalšie vlastnosti štatistického funkcionálu $T(P), P \in \mathcal{P}$.

Pre úplnosť teoretického základu uvedieme ešte 3 definície, ktoré budeme potrebovať v ďalšom texte – kontamináciu rozdelenia iným rozdelením a Gâteauxovu a Fréchetovu deriváciu.

Definícia. (Kontaminácia rozdelenia) *Nech \mathcal{P} je množina všetkých rozdelení pravdepodobností na $(\mathcal{X}, \mathcal{B})$ s metrikou δ . Nech $P, Q \in \mathcal{P}$ a $t \in [0, 1]$. Potom sa rozdelenie pravdepodobností*

$$P_t(Q) = (1 - t)P + tQ = P + t(Q - P)$$

nazýva kontaminácia P rozdelením Q v pomere t .

Definícia. (Gâteauxova derivácia) *Nech \mathcal{P} je množina všetkých rozdelení pravdepodobností na $(\mathcal{X}, \mathcal{B})$ s metrikou δ a T je štatistický funkcionál. Nech $P, Q \in \mathcal{P}$ a $t \in [0, 1]$. Potom funkcionál T je diferencovateľný v Gâteauxovom zmysle podľa P v smere Q , ak existuje limita*

$$T'_Q(P) = \lim_{t \rightarrow 0_+} \frac{T(P + t(Q - P)) - T(P)}{t}.$$

$T'_Q(P)$ sa nazýva Gâteauxova derivácia T podľa P v smere Q .

Po označení $\varphi(t) = T((1 - t)P + tQ)$, $t \in [0, 1]$ treba predpokladať, že $\varphi(t)$ má konečnú n -tú deriváciu $\varphi^{(n)}$, spojité derivácie $\varphi^{(k)}$ na intervale $(0, 1)$ a sprava spojité derivácie $\varphi_+^{(k)}$ v bode $t = 0$ pre $k = 1, \dots, n - 1$. Potom Gâteauxova derivácia $T'_Q(P)$ funkcionálu T je rovná obyčajnej derivácii sprava funkcie φ v bode 0:

$$T'_Q(P) = \varphi'(0_+).$$

Podobne sa dá definovať Gâteauxova derivácia rádu k :

$$T_Q^{(k)}(P) = \left[\frac{d^k}{dt^k} T(P + t(Q - P)) \right]_{t=0_+} = \varphi^{(k)}(0_+).$$

Definícia. (Fréchetova derivácia) *Nech \mathcal{P} je množina všetkých rozdelení pravdepodobností na $(\mathcal{X}, \mathcal{B})$ s metrikou δ . Nech T je štatistický funkcionál a $P \in \mathcal{P}$. Potom hovoríme, že funkcionál T je diferencovateľný vo Fréchetovom zmysle podľa P v smere Q , ak existuje lineárny funkcionál $L_P(Q - P)$ tak, že limita:*

$$\lim_{t \rightarrow 0} \frac{T(P + t(Q - P)) - T(P)}{t} = L_P(Q - P)$$

konverguje rovnomerne pre $Q \in \mathcal{P}$, $\delta(P, Q) \leq K$ pre ľubovoľné $K \in (0, \infty)$. $L_P(Q - P)$ sa nazýva Fréchetova derivácia T podľa P v smere Q .

Ak je funkcionál T diferencovateľný vo Fréchetovom zmysle, je diferencovateľný aj v Gâteauxovom zmysle. Existuje teda $T'_Q(P)$, $\forall Q \in \mathcal{P}$, že platí:

$$T'_Q(P) = L_P(Q - P)$$

4.2.2 Influenčná funkcia

Influenčná funkcia je jednou z najdôležitejších charakteristík odhadu. Definícia influenčnej funkcie štatistického funkcionálu T je veľmi úzko spojená s pojmom Gâteauxovej derivácie (pozri [13]).

Definícia. (Influenčná funkcia) *Nech \mathcal{X} je metrický priestor s metrikou d , ktorý je úplný a separabilný. Nech \mathcal{B} je σ -algebra borelovských podmnožín \mathcal{X} , \mathcal{P} je systém všetkých pravdepodobnostných mier na $(\mathcal{X}, \mathcal{B})$. Nech T je štatistický funkcionál v rozdelení pravdepodobností P a $Q = \delta_x$, $x \in \mathcal{X}$ je Diracova pravdepodobnosť v bode x (rozdelenie generované v bode x). Označme $T'_{\delta_x}(P) = T'_x(P)$ Gâteauxovu deriváciu T podľa P v smere δ_x . Potom definujeme influenčnú funkciu funkcionálu T predpisom:*

$$IF(x; T, P) = T'_x(P) = \lim_{t \rightarrow 0^+} \frac{T(P_t(\delta_x)) - T(P)}{t},$$

kde $P_t(\delta_x) = (1 - t)P + t\delta_x$.

Hodnota influenčnej funkcie $IF(x; T, P)$ meria vplyv kontaminácie funkcionálu T hodnotou x . Stredná hodnota influenčnej funkcie funkcionálu T udáva priemerný vplyv na chybu odhadovania cez všetky $x \in \mathcal{X}$. Jej hodnota je rovná nule:

$$E_P(IF(x; T, P)) = \int_{\mathcal{X}} T'_x(P) dP = 0.$$

Ak má byť funkcionál robustný, mal by mať ohraničenú influenčnú funkciu. Definícia robustnosti ale nie je úplne jednoznačná, pretože historicky sa tento pojem vyvíjal mnoho rokov a problémy citlivosti štatistických postupov k odchýlkam od daných podmienok uvažovalo mnoho štatistikov z rôznych hľadísk.

Definícia. (Spojitosť funkcionálu) *Nech $T_n = T_n(X_1, \dots, X_n)$ je postupnosť empirických štatistických funkcionálov, $T_n : (\mathcal{X}, \mathcal{B})^{\otimes n} \mapsto (\mathcal{T}_n, \mathcal{A}_n)$ a \mathcal{P} je systém všetkých pravdepodobnostných mier na $(\mathcal{X}, \mathcal{B})$ s Prohorovovou metrikou d_P . Potom postupnosť štatistik $\{T_n\}$ je (kvalitatívne) robustná pre rozdelenie pravdepodobností $P \in \mathcal{P}$, ak pre ľubovoľné $\epsilon > 0$ existuje $\delta > 0$ a $n_0 \in \mathbb{N}$ také, že pre všetky $Q \in \mathcal{P}$ a $n \geq n_0$ platí:*

$$d_P(P, Q) < \delta \quad \Rightarrow \quad d_P(\mathcal{L}_P(T_n), \mathcal{L}_Q(T_n)) < \epsilon,$$

kde $\mathcal{L}_P(T_n)$ je rozdelenie T_n pri P a $\mathcal{L}_Q(T_n)$ je rozdelenie T_n pri Q .

V definícii možno uvažovať aj iné vzdialenosti uvedené napr. v [13]. Je jasné, že spojitosť funkcionálu T nijako robustnosť nemeria (iba vypovedá o tom, či daný funkcionál je alebo nie je robustný). Preto patrí medzi kvalitatívne charakteristiky robustnosti.

Robustnosť funkcionálu je možné merať pomocou tzv. citlivosti funkcionálu T . Citlivosť funkcionálu rozlišujeme lokálnu a globálnu. Lokálna a globálna citlivosť funkcionálu T sú najpoužívanějšími kvantitatívnymi charakteristikami robustnosti. Tieto charakteristiky sú založené na influenčnej funkcii, a na rozdiel od spojitosti funkcionálu dokážu merať ohraničenosť influenčnej funkcie.

Definícia. (Citlivosť funkcionálu) *Nech $T_n = T_n(X_1, \dots, X_n)$ je empirický funkcionál zodpovedajúci vektora (X_1, \dots, X_n) . Pridajme k pozorovaniam X_1, \dots, X_n ďalšie pozorovanie Y . Potom vplyv Y na T_n je charakterizovaný rozdielom:*

$$T_{n+1}(X_1, \dots, X_n, Y) - T_n(X_1, \dots, X_n) := I(T_n, Y).$$

- (1) Citlivosťou funkcionálu T_n k pridaniu ďalšieho pozorovania Y pri daných X_1, \dots, X_n sa nazýva číslo:

$$S(T_n) = \sup_Y |I(T_n(X_1, \dots, X_n), Y)|.$$

- (2) Lokálnou citlivosťou funkcionálu T v rozdelení pravdepodobností P sa nazýva hodnota:

$$\lambda^* = \sup_{x, y \in \mathcal{X}, x \neq y} \left| \frac{IF(y; T, P) - IF(x; T, P)}{y - x} \right|,$$

ktorá meria vplyv nahradenia hodnoty x za hodnotu y na funkcionál T .

- (3) Globálnou citlivosťou funkcionálu T v rozdelení pravdepodobností P sa nazýva hodnota:

$$\gamma^* = \sup_{x \in \mathcal{X}} |IF(x; T, P)|,$$

ktorá udáva maximálnu hodnotu influenčnej funkcie.

4.2.3 Rozptyl asymptoticky normálneho rozdelenia

Veta 4. (CLV) Nech (\mathcal{P}, δ) je metrický priestor všetkých rozdelení pravdepodobností na $(\mathcal{X}, \mathcal{B})$ s metrikou δ . Nech P_n je empirické rozdelenie náhodného výberu (X_1, \dots, X_n) , $n = 1, 2, \dots$, ktoré splňa podmienku:

$$\sqrt{n}\delta(P_n, P) = O_p(1), \quad n \rightarrow \infty, \quad (4.2)$$

kde symbol $O_p(1)$ značí ohraničenosť v pravdepodobnosti¹. Potom za predpokladu, že je štatistický funkcionál T diferencovateľný vo Fréchetovom zmysle, $\text{var}_P(IF(x; T, P)) = E_P(IF(x; T, P))^2 > 0$ a $\text{var}_P(IF(x; T, P)) < \infty$, platí centrálna limitná veta, t. j. postupnosť $\sqrt{n}(T(P_n) - T(P))$ má asymptoticky normálne rozdelenie pre $n \rightarrow \infty$:

$$\mathcal{L}(\sqrt{n}(T(P_n) - T(P))) \rightarrow \mathcal{N}(0, \text{var}_P(IF(x; T, P))).$$

Dôkaz. Nech T je diferencovateľný vo Fréchetovom zmysle. Potom existuje $T'_Q(P)$, $\forall Q \in \mathcal{P}$, že platí:

$$\begin{aligned} T'_Q(P) &= \lim_{t \rightarrow 0^+} \frac{T(P + t(Q - P)) - T(P)}{t} - L_P(Q - P) + L_P(Q - P) = \\ &= L_P(Q - P), \end{aligned} \quad (4.3)$$

t. j. T je diferencovateľný aj v Gâteauxovom zmysle. Špeciálne platí $T'_x(P) = L_P(\delta_x - P) = IF(x; T, P)$.

Pre empirické rozdelenie P_n náhodného vektora (X_1, \dots, X_n) platí:

$$P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i},$$

¹ Ak $X_n \rightarrow_d X$, tak $X_n = O_p(1)$. Ak $\frac{X_n}{U_n} = O_p(1)$, tak $X_n = O_p(U_n)$. Ak $\frac{X_n}{U_n} \rightarrow_P 0$, tak $X_n = o_p(U_n)$.

kde δ_{X_i} je Diracova pravdepodobnosť pre $X_i \in \mathcal{X}$, $i = 1, \dots, n$. Potom je možné pre lineárny funkcionál L_P uviesť:

$$T'_{P_n}(P) = L_P(P_n - P) = L_P\left(\frac{1}{n} \sum_{i=1}^n \delta_{X_i} - P\right) = \frac{1}{n} \sum_{i=1}^n L_P(\delta_{X_i} - P) = \frac{1}{n} \sum_{i=1}^n T'_{X_i}(P).$$

Pre funkciu $\varphi(t)$ definovanú v kapitole 4.2.1 z Taylorovho rozvoja platí:

$$\varphi(t) = \varphi(u) + \sum_{k=1}^{n-1} \frac{\varphi^{(k)}(u)}{k!} (t-u)^k + \frac{\varphi^{(n)}(v)}{n!} (t-u)^n, \quad t \in [0, 1], u \in (0, t), v \in [u, t].$$

Pre $t = 1$, $u = 0$ potom platí:

$$T(Q) - T(P) = \sum_{k=1}^{n-1} \frac{T_Q^{(k)}(P)}{k!} + \frac{1}{n!} \left[\frac{d^n}{dt^n} T(P + t(Q - P)) \right]_{t=t^*}, \quad t^* \in [0, 1].$$

Za platnosti podmienky (4.2) potom platí:

$$\sqrt{n}(T(P_n) - T(P)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n T'_{X_i}(P) + R_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n IF(X_i; T, P) + o_p(1),$$

kde $o_p(1) = \sqrt{n}o(\delta(P_n - P))$.

Ak je rozptyl $\text{var}_P(IF(x; T, P))$ konečný, tak z klasickej centrálnej limitnej vety (pozri [13], [18]) platí:

$$\mathcal{L}(\sqrt{n}(T(P_n) - T(P))) \longrightarrow \mathcal{N}(\text{var}_P(IF(x; T, P))).$$

□

Označme $\sigma^2(T, P) = \text{var}_P(IF(x; T, P))$. Nech $\mathcal{P}_0 \subset \mathcal{P}$ je nejaké okolie. Potom vhodnou mierou robustnosti odhadu T_n je supremum rozptylu $\sigma^2(T, P)$ cez okolie \mathcal{P}_0 :

$$\sigma^2(T) = \sup_{P \in \mathcal{P}_0} \sigma^2(T, P). \quad (4.4)$$

Definícia. (Minimaximálne robustný odhad) *Odhad, ktorý minimalizuje (4.4) cez určitú triedu \mathcal{T} odhadov parametra θ sa nazýva minimaximálne robustným odhadom v triede \mathcal{T} .*

Viac informácií možno nájsť v [13] a [18].

4.2.4 Bod zlyhania

Nech existuje náhodný výber $\mathbf{x}^{(0)} = (x_1, \dots, x_n)$ a štatistický funkcionál $T_n(\mathbf{x}^{(0)})$ príslušný tomuto výberu. Vo výbere $\mathbf{x}^{(0)}$ možno nahradiť m ľubovoľných zložiek. Pre účely ďalšieho textu bude vybrané najviac nepriaznivé nahradenie s čo najnepriaznivejšími hodnotami. Tieto zložky môžu nadobúdať aj nekonečné hodnoty. Nový výber bude ozn. $\mathbf{x}^{(m)}$ a jemu príslušná hodnota odhadu bude $T_n(\mathbf{x}^{(m)})$.

Definícia. (Bod zlyhania odhadu) *Bodom zlyhania odhadu T_n vo výbere $\mathbf{x}^{(0)}$ sa nazýva číslo:*

$$\epsilon^*(T_n, \mathbf{x}^{(0)}) = \frac{m^*(\mathbf{x}^{(0)})}{n},$$

kde $m^*(\mathbf{x}^{(0)})$ je najmenšie celé číslo m , pre ktoré platí:

$$\sup_{\mathbf{x}^{(m)}} \|T_n(\mathbf{x}^{(m)}) - T_n(\mathbf{x}^{(0)})\| = \infty.$$

Ide o najmenší počet pozorovaní, ktorý po nahradení ľubovoľnými hodnotami vedie k nekonečným hodnotám funkcionálu T_n . Ak m^* nezávisí na počiatočnom výbere $\mathbf{x}^{(0)}$, tak je možné stanoviť limitu:

$$\epsilon^* = \lim_{n \rightarrow \infty} \epsilon^*(T_n, \mathbf{x}^{(0)}).$$

Čím väčšie je teda ϵ^* , tým väčší je odhad robustný (pozri [14], [18]).

Bod zlyhania je veľmi často používanou charakteristikou robustnosti odhadu. Modifikáciou môže byť napríklad situácia, kedy namiesto nahradenia m zložiek sa k pôvodnému výberu $\mathbf{x}^{(0)}$ pridá nových m nepriaznivých zložiek [13].

4.2.5 Miera chvostov

Nech existuje náhodný výber (X_1, \dots, X_n) z rozdelenia pravdepodobností so spojitou distribučnou funkciou $F(x - \theta)$, $\theta \in \mathbb{R}$. Miera chvostov sa uplatňuje najmä pri posudzovaní odhadov parametrov posunutia a regresie.

Definícia. (Chvosty rozdelenia pravdepodobností) *Pravdepodobnosť $P_\theta(T_n - \theta > a)$, resp. $P_\theta(\theta - T_n > a)$, pri veľkých $a > 0$ sa nazýva pravým, resp. ľavým chvostom rozdelenia pravdepodobností štatistického funkcionálu T_n .*

Pre symetrické rozdelenia sú chvosty rozdelenia pravdepodobností charakterizované vzťahom $P_\theta(|T_n - \theta| > a) = P_0(|T_n| > a)$.

Obmedzíme sa na odhady T_n *ekvivariantné vzhľadom k posunutiu:*

$$T_n(X_1 + c, \dots, X_n + c) = T_n(X_1, \dots, X_n) + c, \quad \forall c \in \mathbb{R}, \quad \forall X_1, \dots, X_n.$$

Chovanie odhadu T_n parametra θ možno charakterizovať pomocou priebehu pravdepodobnosti $P_\theta(|T_n - \theta| > a)$ pre:

(1) pevné $a > 0$ a $n \rightarrow \infty$

- ak je T_n konzistentným odhadom θ , tak pre ľubovoľné $a > 0$ platí:

$$\lim_{n \rightarrow \infty} P_\theta(|T_n - \theta| > a) = 0,$$

(2) pevné n a $a \rightarrow \infty$

- od dobrého odhadu T_n očakávame, že pri pevnom n je:

$$\lim_{a \rightarrow \infty} P_\theta(|T_n - \theta| > a) = 0,$$

a že táto konvergencia je čo najrýchlejšia vzhľadom k $a \rightarrow \infty$.

V knihe [14] sa uvádza nasledujúca štatistika popisujúca mieru chovania chvostov za predpokladu, že existuje ekvivariantný odhad T_n .

Definícia. (Miera chovania chvostov ekvivariantného odhadu) *Nech* (X_1, \dots, X_n) *je náhodný výber z rozdelenia pravdepodobností so spojitou symetrickou distribučnou funkciou* $F(x - \theta)$, $0 < F(x) < 1$, $F(-x) = (1 - F(x))$, $x, \theta \in \mathbb{R}$. *Potom mierou chovania chvostov ekvivariantného odhadu* T_n *rozumieme:*

$$B(T_n; a) = \frac{-\log P_\theta(|T_n - \theta| > a)}{-\log(1 - F(a))} = \frac{-\log P_0(|T_n| > a)}{-\log(1 - F(a))}, \quad a > 0.$$

Čím väčšia je hodnota $1 \leq \lim_{a \rightarrow \infty} B(T_n; a) \leq n$, tým lepší je odhad. Hodnota $B(T_n; a)$ totiž udáva, koľkokrát rýchlejšie konverguje pravdepodobnosť $P_0(|T_n| > a)$ k nule pre $a \rightarrow \infty$ než horná hranica rýchlosti chvostov $1 - F(a)$.

Hovoríme, že distribučná funkcia F má:

- *exponenciálne chvosty*, ak $\lim_{a \rightarrow \infty} \frac{-\log(1-F(a))}{ba^r} = 1$ pre nejaké $b > 0$, $r \leq 1$,
- *ťažké chvosty*, ak $\lim_{a \rightarrow \infty} \frac{-\log(1-F(a))}{m \log a} = 1$ pre nejaké $m > 0$.

Poznámka: Distribučné funkcie s exponenciálnymi chvostami sa označujú ako typ I. Medzi tieto rozdelenia patrí napríklad normálne rozdelenie s $r = 2$. Distribučné funkcie s ťažkými chvostami sa označujú ako typ II. Medzi ne patrí napríklad t -rozdelenie s m stupňami voľnosti, $m > 1$.

4.2.6 Príklad

Na záver tejto kapitoly uvedieme ako príklad parametra rozdelenia pravdepodobností P strednú hodnotu (príklad vychádza z [13], [14] a [18]). Stanovíme influenčnú funkciu tohto štatistického funkcionálu a jej lokálnu a globálnu citlivosť:

- štatistický funkcionál a empirický funkcionál:

$$\begin{aligned} E_P X &= T(P) = \int_{\mathcal{X}} x dP \\ T(P_n) &= \int_{\mathcal{X}} x dP_n = \bar{X} \end{aligned}$$

- Gâteauxova derivácia:

$$\begin{aligned} \varphi(t) &= \int_{\mathcal{X}} x d((1-t)P + tQ) = (1-t)E_P X + tE_Q X \\ \varphi'(t) &= E_Q X - E_P X \end{aligned}$$

$$\Rightarrow T'_Q(P) = \varphi'(0_+) = E_Q X - E_P X$$

- influenčná funkcia:

$$\begin{aligned} IF(x; T, P) &= T'_x(P) = x - E_P X \\ E_P(IF(x; T, P)) &= 0 \\ \text{var}_P(IF(x; T, P)) &= E_P(x - E_P X)^2 = \text{var}_P X = \sigma_P^2 \end{aligned}$$

$$\Rightarrow \mathcal{L}(\sqrt{n}(\bar{X}_n - E_P X)) \rightarrow \mathcal{N}(0, \sigma_P^2)$$

- lokálna a globálna citlivosť:

$$\gamma^* = \sup_{x \in \mathcal{X}} |x - E_P X| = \infty$$

$$\lambda^* = \sup_{x, y \in \mathcal{X}, x \neq y} \left| \frac{y - E_P X - x + E_P X}{y - x} \right| = 1$$

\Rightarrow priemer nie je robustný, ale nie je citlivý k lokálnemu nahradzovaniu hodnôt [13].

Pre empirické funkcionály (priemer a medián) určíme bod zlyhania pre ľubovoľný počítačový výber $\mathbf{x}^{(0)}$ a mieru chvostov štatistického odhadu:

- $T_n = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$:

$$\epsilon^*(T_n, \mathbf{x}^{(0)}) = \frac{1}{n} \longrightarrow 0, \quad n \longrightarrow \infty$$

ak má distribučná funkcia F exponenciálne chvosty, tak

$$\lim_{a \rightarrow \infty} B(T_n; a) = n,$$

ak má distribučná funkcia F ťažké chvosty, tak

$$\lim_{a \rightarrow \infty} B(T_n; a) = 1,$$

- $T_n = \tilde{X} = X_{\frac{n+1}{2}}$ pre nepárne n :

$$\epsilon^*(T_n, \mathbf{x}^{(0)}) = \frac{n+1}{2n} \longrightarrow \frac{1}{2}, \quad n \longrightarrow \infty,$$

ak má distribučná funkcia F exponenciálne alebo ťažké chvosty, tak

$$\lim_{a \rightarrow \infty} B(T_n; a) = \frac{n+1}{2}.$$

4.3 M-odhady

Triedu M -odhadov zaviedol P. J. Huber a ich vlastnosti študovalo mnoho rôznych autorov. Teória M -odhadov je podrobne rozoberaná práve v knihe [10]. Pre viac informácií možno siahnuť po [13] alebo [18].

Nech X_1, \dots, X_n je náhodný výber s rozdelením pravdepodobností P a F je neznáma distribučná funkcia z triedy distribučných funkcií \mathcal{F} príslušná P . Hľadáme odhad parametra θ . Tento parameter je možné vyjadriť ako funkcionál $T(P)$ rozdelenia P .

Funkcionál $T(P)$ je pre M -odhad vyjadrený ako minimalizačná úloha. Odhad parametra θ sa získa nahradením P vo funkcionáli T empirickým rozdelením

prislúchajúcim vektora X_1, \dots, X_n . M -odhad T_n je teda definovaný ako riešenie úlohy pre $\theta \in \Theta$:

$$\min_{\theta \in \Theta} \sum_{i=1}^n \rho(X_i, \theta) = \min_{\theta \in \Theta} E_{P_n}[\rho(X, \theta)],$$

kde $\rho(\cdot, \cdot)$ je vhodne zvolená funkcia.

Poznámka: Špeciálne, ak existuje parameter rozdelenia P_θ s hustotou $f(x, \theta)$, tak maximálne vierohodný odhad je riešením úlohy:

$$\min_{\theta \in \Theta} \sum_{i=1}^n -\log f(X_i, \theta).$$

Ak je navyše funkcia ρ spojitou diferencovateľná vzhľadom k θ s deriváciou $\psi(\cdot, \theta) = \frac{\partial}{\partial \theta} \rho(\cdot, \theta)$, tak T_n je riešením rovnice:

$$\begin{aligned} \sum_{i=1}^n \psi(X_i, \theta) &= 0, \quad \theta \in \Theta \\ \frac{1}{n} \sum_{i=1}^n \psi(X_i, T_n) &= E_{P_n}[\psi(X, T_n)] = 0, \quad T_n \in \Theta. \end{aligned} \quad (4.5)$$

Z (4.5) vyplýva, že štatistický funkcionál $T(P)$ príslušný T_n je definovaný ako riešenie úlohy, resp. rovnice:

$$\begin{aligned} \min \int_{\mathcal{X}} \rho(x, T(P)) dP(x) &= \min E_P[\rho(X, T(P))], \quad T(P) \in \Theta \\ \int_{\mathcal{X}} \psi(x, T(P)) dP(x) &= E_P[\psi(X, T(P))] = 0, \quad T(P) \in \Theta. \end{aligned}$$

Za predpokladu, že má táto úloha práve jedno riešenie, je funkcionál $T(P)$ fisherovsky konzistentný [13].

Je možné použiť náhodný výber X_1, \dots, X_n nezávislých pozorovaní s rovnakou neznámou distribučnou funkciou $F(x - \theta)$, $\theta \in \mathbb{R}$. Potom možno zaviesť model s parametrom posunutia θ . V tomto modeli je M -odhad T_n definovaný ako riešenie úlohy:

$$\min_{\theta \in \mathbb{R}} \sum_{i=1}^n \rho(X_i - \theta). \quad (4.6)$$

Za predpokladu, že $\rho(\cdot)$ je absolútne spojitou diferencovateľná s deriváciou $\psi(\cdot)$, sa dá úloha (4.6) prepísať do tvaru:

$$\sum_{i=1}^n \psi(X_i - \theta) = 0, \quad \theta \in \mathbb{R}. \quad (4.7)$$

Opäť za predpokladu, že úloha $\min \int_{\mathcal{X}} \rho(x - \theta) dP(x)$ má jediné riešenie ($\theta = 0$), je funkcionál $T(P)$ fisherovsky konzistentný.

Ako uvádza [13], v modeli s parametrom posunutia θ platí:

(1) T_n je ekvivariantný vzhľadom k posunutiu, t. j. platí:

$$T_n(X_1 + c, \dots, X_n + c) = T_n(X_1, \dots, X_n) + c, \quad \forall c \in \mathbb{R}$$

(2) T_n nie je všeobecne ekvivariantný vzhľadom k merítku, t. j. neplatí:

$$T_n(cX_1, \dots, cX_n) = cT_n(X_1, \dots, X_n), \quad c > 0$$

Ak je funkcia ρ v (4.6) rýdzo konvexná funkcia (a teda ψ je rastúca), tak je M -odhad určený jednoznačne. Ak je ρ v niektorom úseku lineárna funkcia, je ψ v tomto úseku konštantná a rovnica (4.7) môže mať viac koreňov. Z týchto koreňov obvykle volíme jeden podľa pravidla:

$$T_n = \frac{1}{2}(T_n^+ + T_n^-) = \frac{1}{2} \left(\inf_t \left\{ \sum_{i=1}^n \psi(X_i - t) < 0 \right\} + \sup_t \left\{ \sum_{i=1}^n \psi(X_i - t) > 0 \right\} \right).$$

4.3.1 Influenčná funkcia a bod zlyhania

Nech funkcia $\rho(\cdot, \theta)$ je absolútne spojito diferencovateľná vzhľadom k θ s deriváciou $\psi(\cdot, \theta)$. Nech ďalej úloha (4.6) má jediné riešenie $T(P)$. Za predpokladu, že $P_t = (1-t)P + t\delta_x$, kde δ_x je Diracova pravdepodobnosť v bode x platí, že $T(P_t)$ je riešením rovnice:

$$\begin{aligned} \int_{\mathcal{X}} \psi(y, T(P_t)) dP((1-t)P + t\delta_x) &= 0 \\ (1-t) \int_{\mathcal{X}} \psi(y, T(P_t)) dP(y) + t\psi(x, T(P_t)) &= 0 \end{aligned}$$

Derivovaním tohto vzťahu vzhľadom k t vznikne:

$$\begin{aligned} - \int_{\mathcal{X}} \psi(y, T(P_t)) dP(y) + (1-t) \frac{dT(P_t)}{dt} \int_{\mathcal{X}} \left[\frac{\partial}{\partial \theta} \psi(y, \theta) \right]_{\theta=T(P_t)} dP(y) \\ + \psi(x, T(P_t)) + t \frac{dT(P_t)}{dt} \left[\frac{\partial}{\partial \theta} \psi(x, \theta) \right]_{\theta=T(P_t)} = 0 \end{aligned}$$

a prechodom k limite pre $t \rightarrow 0_+$ je výsledkom influenčná funkcia M -odhadu v tvare:

$$IF(x; T, P) = \frac{\psi(x, T(P))}{-\int_{\mathcal{X}} \left[\frac{\partial}{\partial \theta} \psi(y, \theta) \right]_{\theta=T(P)} dP(y)}.$$

V prípade, že existuje model s parametrom posunutia, je influenčná funkcia v tvare:

$$IF(x; T, P) = \frac{\psi(x, T(P))}{-\int_{\mathcal{X}} \psi'(y) dP(y)}.$$

Pre model s parametrom posunutia je možné použiť neohraničenú funkciu $\psi(\cdot)$. V [13] sa uvádza, že za týchto predpokladov je bod zlyhania M -odhadu parametra posunutia:

$$\epsilon^* = \lim_{n \rightarrow \infty} \epsilon^*(T_n, \mathbf{x}^{(0)}) = 0.$$

V tomto prípade je M -odhad nerobustný. Ak je ale funkcia ψ ohraničená a nepárna a ρ je stredom symetrie rozdelenia pravdepodobností, tak platí:

$$\epsilon^* = \lim_{n \rightarrow \infty} \epsilon^*(T_n, \mathbf{x}^{(0)}) = \frac{1}{2}$$

a M -odhad je teda robustný.

4.3.2 Studentizované M-odhady

Ako už bolo uvedené, M -odhad je ekvivariantný vzhľadom k posunutiu, ale nie je všeobecne ekvivariantný vzhľadom k merítku. Tento nedostatok možno odstrániť pomocou studentizácie M -odhadu vhodnou škálovacou štatistikou $S_n = S_n(X_1, \dots, X_n)$, ktorá pre $(X_1, \dots, X_n) \in \mathbb{R}^n$ spĺňa:

- (1) $S_n > 0$ skoro všade,
- (2) S_n je invariantná vzhľadom k posunutiu:

$$S_n(X_1 + c, \dots, X_n + c) = S_n(X_1, \dots, X_n), \quad c \in \mathbb{R},$$

- (3) S_n je ekvivariantná vzhľadom k merítku:

$$S_n(cX_1, \dots, cX_n) = cS_n(X_1, \dots, X_n), \quad c > 0.$$

Naviac možno predpokladať, že:

$$\sqrt{n}(S_n - S(P)) = O_p(1), \quad n \rightarrow \infty,$$

kde $S(P)$ je štatistický funkcionál príslušný S_n .

Za týchto predpokladov je možné odvodiť studentizovaný M -odhad ekvivariantný vzhľadom k posunutiu aj k merítku. Studentizovaný M -odhad je riešením úlohy pre $\theta \in \Theta$:

$$\min_{\theta \in \Theta} \sum_{i=1}^n \rho \left(\frac{X_i - \theta}{S_n} \right).$$

Ak má ρ spojitú deriváciu ψ , možno M -odhad odvodiť ako riešenie rovnice:

$$\sum_{i=1}^n \psi \left(\frac{X_i - \theta}{S_n} \right) = 0, \quad \theta \in \Theta.$$

Poznámka: Za škálovaciu štatistiku S_n je možné zvoliť napr. výberovú smerodajnú odchýlku, medzikvartilové rozpätie (rozdiel horného a dolného kvartilu) alebo mediánovú absolútnu odchýlku (MAD):

$$S_n = \operatorname{med}_{1 \leq i \leq n} |X_i - \tilde{X}|.$$

4.3.3 Asymptotické vlastnosti a minimaximálne robustné odhady

Už bolo v práci uvedené, že postupnosť $\sqrt{n}(T(P_n) - T(P))$ má asymptoticky normálne rozdelenie za určitých predpokladov. Taktiež bolo uvedené, že funkcionály $T(P)$ a $T(F)$ sú zámenné. Preto je možné za týchto predpokladov písať:

$$\mathcal{L}(\sqrt{n}(T_n - T(F))) \rightarrow \mathcal{N}(0, \sigma^2(T, F)), \quad (4.8)$$

kde F je distribučná funkcia vzhľadom k rozdeleniu pravdepodobností P a $\sigma^2(T, F) = \operatorname{var}_F(IF(x; T, F)) = \int_{\mathcal{X}} IF^2(x; T, F) dF(x)$.

Centrálnu limitnú vetu možno uviesť aj pre M -odhad. Nech existuje postupnosť nezávislých pozorovaní X_1, X_2, \dots s rovnakou distribučnou funkciou $F(x, \theta)$, $\theta \in \Theta$, kde Θ je otvorený interval v \mathbb{R} . Za predpokladu, že $\rho(\cdot, \theta)$ je absolútne spojitá v θ s deriváciou $\psi(\cdot, \theta) = \frac{\partial}{\partial \theta} \rho(\cdot, \theta)$ hľadáme minimum $E_\theta \rho(X, t)$. Ak je $\psi(x, \theta)$ spojitá v θ , tak úloha hľadania minima prechádza na hľadanie M -odhadu T_n medzi koreňmi rovnice:

$$\sum_{i=1}^n \psi(X_i, \theta) = 0, \quad \theta \in \Theta. \quad (4.9)$$

Ak $E_\theta \rho(X, t)$ má jediné minimum v bode $t = \theta$ a je splnená buď podmienka na hladkosť $\psi(x, \theta)$ alebo $F(x, \theta)$, tak ako uvádza [13] existuje postupnosť $\{T_n\}$ koreňov rovnice (4.9), že:

$$\sqrt{n}(T_n - \theta) = O_p(1) = \frac{1}{\sqrt{n}\gamma(\theta)} \sum_{i=1}^n \psi(X_i, \theta) + O_p\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow \infty$$

kde $\gamma(\theta) = E_\theta \left[\frac{\partial}{\partial \theta} \psi(X, \theta) \right]$. Platí:

$$\mathcal{L}(\sqrt{n}(T_n - \theta)) \rightarrow \mathcal{N}(0, \sigma^2(\psi, F)),$$

kde $\sigma^2(\psi, F) = \frac{E_\theta[\psi^2(X, \theta)]}{\gamma^2(\theta)}$.

Je možné odvodiť ešte minimaximálny robustný odhad za predpokladu konvergencie (4.8). Ako už bolo uvedené, za mieru robustnosti funkcionálu T je možné považovať maximum asymptotického rozptylu $\sigma^2(T, P)$ cez určitú triedu rozdelení pravdepodobností \mathcal{P} . Taktiež platí, že maximum asymptotického rozptylu $\sigma^2(T, F)$ je miera robustnosti cez určitú triedu distribučných funkcií \mathcal{F} :

$$\sigma^2(T) = \sup_{F \in \mathcal{F}} \sigma^2(T, F).$$

V tomto prípade je možné použiť triedu M -funkcionálov \mathcal{T} a hľadať funkcionál T_0 taký, že:

$$\sigma^2(T_0) \leq \sigma^2(T), \quad \forall T \in \mathcal{T}. \quad (4.10)$$

Definícia. (Minimaximálne robustný funkcionál) *Ak existuje funkcionál T_0 , pre ktorý platí (4.10), tak sa nazýva minimaximálne robustný, lebo spĺňa:*

$$\sigma^2(T_0) = \inf_{T \in \mathcal{T}} \sup_{F \in \mathcal{F}} \sigma^2(T, F)$$

Nech F je neznámy prvok systému distribučných funkcií \mathcal{F} . Nech $F_0 \in \mathcal{F}$ je distribučná funkcia, ktorá minimalizuje Fisherovu informáciu na \mathcal{F} (t. j. najmenej priaznivé rozdelenie systému \mathcal{F}), že:

$$\mathcal{I}(F_0) = \int_{\mathbb{R}} \left(\frac{f'_0(x)}{f_0(x)} \right)^2 dF_0 = \min_{F \in \mathcal{F}} \mathcal{I}(F).$$

Nech $T_0 \in \mathcal{T}$ je asymptotickým odhadom θ pre distribučnú funkciu F_0 , t. j. $\sigma^2(T_0, F_0) = \frac{1}{\mathcal{I}(F_0)}$. Ak je $\sigma^2(T_0, F_0) \leq \sup_{F \in \mathcal{F}} \sigma^2(T_0, F)$, tak platí:

$$\inf_{T \in \mathcal{T}} \sup_{F \in \mathcal{F}} \sigma^2(T, F) = \frac{1}{\mathcal{I}(F_0)}, \quad \forall T \in \mathcal{T}, \quad \forall F \in \mathcal{F}.$$

Ak za triedu distribučných funkcií \mathcal{F} vezmeme symetrický kontaminačný model:

$$\mathcal{F}_G = \{F : F = (1 - \epsilon)G + \epsilon H, H \in \mathcal{F}\},$$

kde G je pevná distribučná funkcia, $0 \leq \epsilon < 1$ je pevné a H prebieha cez pevnú triedu \mathcal{F} distribučných funkcií, tak existuje minimaxálny robustný M -odhad.

4.3.4 Najpoužívanejšie M -odhady

Špeciálnou triedou M -odhadov sú tzv. L_p odhady, ktoré minimalizujú:

$$\sum_{i=1}^n |\epsilon_i|^p,$$

kde ϵ_i je pri predchádzajúcom značení argument funkcie $\rho(\epsilon_i)$ a $1 \leq p \leq 2$. Ak $p = 2$, tak ide o klasický OLS-odhad. Čím menšie sa volí p , tým robustnejší je daný odhad.

V nasledujúcej prehľadnej tabuľke (pozri tab. 4.1) sú uvedené možné voľby funkcií ρ a ψ pre M -odhady modelu s parametrom posunutia (pozri [9] a [25]).

Odhad	ρ	ψ	Medze pre x
Priemer	$\frac{1}{2}x^2$	x	$-\infty < x < \infty$
Useknutý priemer (trimming)	nekonvexná funkcia	x 0	$ x < c$ $ x \geq c$
Winsorizovaný priemer	nekonvexná funkcia	$-c$ x c	$x \leq -c$ $ x < c$ $x \geq c$
Medián	$ x $	$\text{sgn}(x)$	$-\infty < x < \infty$
Huberova funkcia (k), $k > 0$	$\frac{1}{2}x^2$ $k x - \frac{1}{2}k^2$	x $k \text{sgn}(x)$	$ x \leq k$ $ x > k$
Hampelova po častiach lineárna funkcia (a, b, c), $0 < a < b < c$	$\frac{1}{2}x^2$ $a x - \frac{1}{2}a^2$ $ab - \frac{1}{2}a^2 + (c-b)\frac{a}{2} \left[1 - \left(\frac{c- x }{c-b} \right)^2 \right]$ $ab - \frac{1}{2}a^2 + (c-b)\frac{a}{2}$	x $a \text{sgn}(x)$ $a \frac{c- x }{c-b} \text{sgn}(x)$ 0	$ x \leq a$ $a < x \leq b$ $b < x \leq c$ $ x > c$
Andrewsova sínusová funkcia	$\frac{1}{\pi^2}(1 - \cos \pi x)$ $\frac{2}{\pi^2}$	$\frac{1}{\pi} \sin \pi x$ 0	$ x \leq 1$ $ x > 1$
Tukeyho funkcia (biweight)	$\frac{k^2}{6} [1 - (1 - (\frac{x}{k})^2)^3]$ $\frac{k^2}{6}$	$x(1 - (\frac{x}{k})^2)^2$ 0	$ x \leq k$ $ x > k$

Tabuľka 4.1: Najpoužívanejšie M -odhady pre model s parametrom posunutia.

Poznámka: Ako uvádza [13], Kagan, Linnik a Rao dokázali, že odhad metódou najmenších štvorcov v lineárnom regresnom modeli je veľmi citlivý k odchýlkam od normálneho rozdelenia chýb. Tento odhad je prípustný vzhľadom ku kvadratickej stratovej funkcii práve vtedy, keď sú chyby normálne rozdelené.

4.3.5 M -odhady v lineárnom regresnom modeli

Pre účely tejto kapitoly bude použitý lineárny regresný model v tvare (2.3) uvedený v kapitole 2.1. V tomto modeli predstavuje $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ vektor vzájomne

nezávislých reziuí s rovnakou distribučnou funkciou F . Táto distribučná funkcia nie je všeobecne známa. Predpokladá sa ale, že F patrí do určitého systému distribučných funkcií \mathcal{F} .

Ako už bolo uvedené, k odhadu parametrov β_j , $j = 1, \dots, k$ sa najčastejšie používa *metóda najmenších štvorcov*. Tento odhad bol označený v predchádzajúcom texte ako \mathbf{b} (pozri kap. 2.1.1). Pre distribučnú funkciu F s konečným druhým momentom je odhad \mathbf{b} najlepším nestranným lineárnym odhadom $\boldsymbol{\beta}$ [13].

Ako uvádzajú viacerí autori (pozri [13], [18]), odhad metódou najmenších štvorcov je veľmi nerobustný a citlivý k:

- odľahlým pozorovaniam Y_i , $i = 1, \dots, n$,
- odľahlým pozorovaniam jednotlivých elementov matice \mathbf{X} ,
- odchýlkam od normálneho rozdelenia chýb ϵ_i , $i = 1, \dots, n$.

Tento odhad navyše zlyháva, ak má rozdelenie chýb ťažké konce. Preto sa ku klasickému odhadu metódou najmenších štvorcov hľadajú robustné alternatívy.

Pre lineárny regresný model (2.3) sa M -odhad parametra $\boldsymbol{\beta}$ definuje ako riešenie \mathbf{M}_n úlohy minimalizácie vzhľadom k $\boldsymbol{\beta} \in \mathbb{R}$:

$$\min_{\beta_1, \dots, \beta_k} \sum_{i=1}^n \rho \left(Y_i - \sum_{j=1}^k x_{ij} \beta_j \right) = \min_{\boldsymbol{\beta}} \sum_{i=1}^n \rho(Y_i - \mathbf{X}_i \boldsymbol{\beta}),$$

kde \mathbf{X}_i , $i = 1, \dots, n$ je i -ty riadok matice \mathbf{X} a $\rho: \mathbb{R} \rightarrow \mathbb{R}$ je absolútne spojitá, obvykle konvexná funkcia s deriváciou ψ . Odhad $\boldsymbol{\beta}$ v tomto prípade teda riešením rovníc:

$$\sum_{i=1}^n \psi(Y_i - \mathbf{X}_i \boldsymbol{\beta}) \mathbf{X}_i = \mathbf{0}.$$

Z podmienok uvedených v [13] možno ukázať, že M -odhad parametra $\boldsymbol{\beta}$ je asymptoticky normálny pre $n \rightarrow \infty$, t. j. platí:

$$\mathcal{L}(\sqrt{n}(\mathbf{M}_n - \boldsymbol{\beta})) \rightarrow \mathcal{N}_k(\mathbf{0}, \sigma^2(\psi, F)\mathbf{Q}^{-1}),$$

kde $\mathbf{Q} = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}' \mathbf{X}$ je podľa predpokladov pozitívne definitná matica a pre $\sigma^2(\psi, F)$ platí:

$$\sigma^2(\psi, F) = \frac{E[\psi^2(\epsilon_1)]}{(E[\psi'(\epsilon_1)])^2}.$$

Pre M -odhad \mathbf{M}_n platí:

- \mathbf{M}_n je ekvvariantný vzhľadom k regresii, platí:

$$\mathbf{M}_n(\mathbf{Y} - \mathbf{X}\mathbf{b}) = \mathbf{M}_n(\mathbf{Y}) + \mathbf{b}, \quad \forall \mathbf{b} \in \mathbb{R}^k,$$

- \mathbf{M}_n nie je všeobecne ekvvariantný vzhľadom k merítku, neplatí:

$$\mathbf{M}_n(c\mathbf{Y}) = c\mathbf{M}_n(\mathbf{Y}), \quad c > 0.$$

Ekvivariantný odhad vzhľadom k merítku je možné získať buď studentizáciou, alebo tak, že sa spolu s regresným parametrom β odhaduje aj merítko:

(1) Studentizovaný M -odhad je riešením úlohy:

$$\min_{\beta} \sum_{i=1}^n \rho \left(\frac{Y_i - \mathbf{X}_i \beta}{S_n} \right) \quad (4.11)$$

kde $S_n = S_n(\mathbf{Y}) \geq 0$ je vhodná škálovacia štatistika invariantná vzhľadom k regresii a ekvivariantná vzhľadom k merítku, t. j.:

$$S_n(c(\mathbf{Y} + \mathbf{X}\mathbf{b})) = cS_n(\mathbf{Y}), \quad \forall \mathbf{b} \in \mathbb{R}^k, \quad c > 0.$$

Ak je $\psi(x) = \frac{\partial}{\partial x} \rho(x)$ spojitá funkcia, tak \mathbf{M}_n je koreňom sústavy rovníc:

$$\sum_{i=1}^n \mathbf{X}_i \psi \left(\frac{Y_i - \mathbf{X}_i \beta}{S_n} \right) = \mathbf{0}. \quad (4.12)$$

Táto sústava môže mať viac koreňov, ale iba jeden z nich vedie ku globálnemu minimu úlohy (4.11). Za platnosti podmienok pre:

- škálovaciu štatistiku S_n uvedených v kapitole 4.3.2,
- funkcie² ρ a ψ a
- maticu³ \mathbf{X}

vždy existuje aspoň jeden koreň sústavy (4.12), ktorý je \sqrt{n} -konzistentným odhadom β . Znenie tohoto tvrdenia možno aj s dôkazom nájsť v [15].

(2) Merítko spolu s regresným parametrom (t. j. dvojicu (\mathbf{M}_n, s)) sa dá odhadovať napríklad ako riešenie úlohy:

$$\min_{\beta} \sum_{i=1}^n \sigma \rho \left(\frac{Y_i - \mathbf{X}_i \beta}{\sigma} \right) + a\sigma,$$

kde $\beta \in \mathbb{R}$, $\sigma > 0$ a $a > 0$ je vhodná konštanta. To vedie k sústave rovníc:

$$\begin{aligned} \sum_{i=1}^n \mathbf{X}_i \psi \left(\frac{Y_i - \mathbf{X}_i \beta}{\sigma} \right) &= \mathbf{0} \\ \sum_{i=1}^n \chi \left(\frac{Y_i - \mathbf{X}_i \beta}{\sigma} \right) &= a, \end{aligned}$$

kde $\chi(x) = x\psi(x) - \rho(x)$, $a = \int_{\mathbb{R}} \chi(x) d\Phi(x)$ a Φ je distribučná funkcia $\mathcal{N}(0, 1)$.

² $h(t) = \int \rho \left(\frac{z-t}{s(P)} \right) dP(z)$ má jednoznačné minimum v bode $t = 0$ a funkcia ψ je hladká.

³ $x_{i1} = 1, i = 1, \dots, n; \frac{1}{n} \sum_{i=1}^n \|\mathbf{X}_i\|^4 = O_p(1); \lim_{n \rightarrow \infty} \mathbf{Q}_n = \mathbf{Q}$, kde $\mathbf{Q}_n = \frac{1}{n} \mathbf{X}' \mathbf{X}$ a \mathbf{Q} je pozitívne definitná.

Otázkou zostáva, ako spomedzi všetkých koreňov sústavy (4.12) nájsť ten, ktorý je \sqrt{n} -konzistentným odhadom β . K tomu možno použiť aproximovaný eficientný koreň sústavy (4.12) pomocou tzv. *jednokrokovej verzie*. Tá reprezentuje prvý krok Newton - Raphsonovho algoritmu riešenia algebraických rovníc.

Jednokroková verzia M -odhadu $\mathbf{M}_n^{(1)}$ začína počiatočným \sqrt{n} -konzistentným odhadom $\mathbf{M}_n^{(0)}$ parametra β . Potom sa určí $\mathbf{M}_n^{(1)}$ pomocou vzťahu:

$$\mathbf{M}_n^{(1)} = \begin{cases} \mathbf{M}_n^{(0)} + \frac{1}{c} \mathbf{W}_n, & c \neq 0, \\ \mathbf{M}_n^{(0)}, & c = 0, \end{cases}$$

kde charakteristika \mathbf{W}_n je daná vzťahom:

$$\mathbf{W}_n = \mathbf{Q}_n^{-1} \sum_{i=1}^n \mathbf{X}_i \psi \left(\frac{Y_i - \mathbf{X}_i \mathbf{M}_n^{(0)}}{S_n} \right), \quad \mathbf{Q}_n = \frac{1}{n} \mathbf{X}' \mathbf{X},$$

a c je odhad funkcionálu γ :

$$\begin{aligned} \gamma &= \frac{1}{S(F)} \int_{\mathbb{R}} \psi' \left(\frac{x}{S(F)} \right) dF(x), & \text{ak } \psi \text{ je spojitá,} \\ \gamma &= \int_{\mathbb{R}} f(xS(F)) d\psi(x), & \text{ak } \psi \text{ je nespojitá.} \end{aligned}$$

Poznámka: Pre absolútne spojitú funkciu ψ je možné použiť odhad:

$$c = \frac{1}{nS_n} \sum_{i=1}^n \psi' \left(\frac{Y_i - \mathbf{X}_i \mathbf{M}_n^{(0)}}{S_n} \right).$$

Jednokroková verzia $\mathbf{M}_n^{(1)}$ je dobrou aproximáciou konzistentného odhadu \mathbf{M}_n , ak je ψ dostatočne hladká a platí:

- $\left\| \mathbf{M}_n - \mathbf{M}_n^{(1)} \right\| = O_p\left(\frac{1}{n}\right)$ všobecne, resp.
- $\left\| \mathbf{M}_n - \mathbf{M}_n^{(1)} \right\| = O_p\left(\frac{1}{\sqrt[4]{n^3}}\right)$ za prítomnosti skokov vo funkcii ψ .

4.4 L-odhady

L -odhad je označenie odhadu, ktorý je založený na usporiadaných pozorovaniach náhodného výberu X_1, \dots, X_n , ozn. $X_{[1]} \leq \dots \leq X_{[n]}$. Tento typ odhadu je riešením rovnice:

$$T_n = \sum_{i=1}^n c_{[i]} h(X_{[i]}) + \sum_{j=1}^k a_j h^*(X_{[np_j+1]}), \quad (4.13)$$

kde $0 < p_1 < \dots < p_k < 1$, $c_{[1]}, \dots, c_{[n]}$ a a_1, \dots, a_k sú dané koeficienty a $h(\cdot)$, $h^*(\cdot)$ sú dané funkcie. Navyiac pre koeficienty $C_{[i]}$, $i = 1, \dots, n$ platí:

$$c_{[i]} = \int_{\frac{i-1}{n}}^{\frac{i}{n}} J(s) ds \approx \frac{1}{n} J\left(\frac{i}{n+1}\right), \quad i = 1, \dots, n,$$

t. j. tieto koeficienty sú ohraňované váhovou funkciou $J : [0, 1] \rightarrow \mathbb{R}$. L -odhad je teda tvorený dvoma zložkami:

- prvá zložka zahŕňa všetky usporiadané pozorovania a
- druhá zložka je lineárnou kombináciou konečného počtu výberových kvantilov.

Preto odhady, ktoré obsahujú iba jednu z týchto dvoch zložiek je možné označiť ako L -odhady typu I, resp. typu II.

4.4.1 Influenčná funkcia a bod zlyhania

Pomocou empirickej kvantilovej funkcie $Q_n(t) = F_n^{-1}(t) = \inf_x \{F_n(x) \geq t\}$, $0 < t < 1$ sa dá L -odhad typu I s váhovou funkciou J spĺňajúcou $\int_0^1 J(u)du = 1$ prepísať do tvaru:

$$T_n = \int_0^1 J(s)h(Q_n(s))ds.$$

Influenčnú funkciu štatistického funkcionálu $T(F)$ príslušnému tomuto špeciálnemu L -odhadu je možné zapísať v tvare odvodenom v [13]:

$$IF(x; T, F) = \int_{-\infty}^{\infty} F(y)h'(y)J(F(y))dy - \int_x^{\infty} J(F(y))dy,$$

kde F a h sú absolútne spojité funkcie.

Nech pre váhovou funkciou J navyše platí, že $J(u) = 0$, $0 < u < \alpha$ a $1 - \alpha < u < 1$. Ak $\epsilon_n^* = \frac{m_n}{n}$ je bod zlyhania L -odhadu (4.13), tak platí:

$$\lim_{n \rightarrow \infty} \epsilon_n^* = \alpha.$$

4.4.2 Najpoužívanejšie L-odhady

V nasledujúcej prehľadnej tabuľke sú uvedené najznámejšie L -odhady (pozri [13]).

Odhad	Rovnica
Výberové rozpätie	$R_n = X_{[n]} - X_{[1]}$
Giniho priemerná diferenciacia	$G_n = \frac{2}{n(n-1)} \sum_{i=1}^n (2i - n - 1)X_{[i]}$
α -useknutý priemer, $0 < \alpha < \frac{1}{2}$	$\bar{X}_{n\alpha} = \frac{1}{n-2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{[i]}$
α -winsorizovaný priemer	$\bar{W}_{n\alpha} = \frac{1}{n} \{ [n\alpha]X_{[[n\alpha]+1]} + \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{[i]} + [n\alpha]X_{[n-[n\alpha]]} \}$
Senov vážený priemer, $0 < k < \frac{n}{2}$	$T_{n,k} = \binom{n}{2k+1}^{-1} \sum_{i=k+1}^{n-k} \binom{i}{k} \binom{n-i+1}{k} X_{[i+1]}$

Tabuľka 4.2: Najpoužívanejšie L-odhady.

Poznámka: $[n\alpha]$ označuje celú časť čísla.

4.4.3 L-odhady v lineárnom regresnom modeli

Rozšírenie L -odhadov na lineárny regresný model (2.3) nie je jednoduché. Odvodiť sa to podarilo až Koenkerovi a Bassettovi. Tí za predpokladu, že matica \mathbf{X} má v prvom stĺpci jednotky, definovali regresný α -kvantil $\widehat{\boldsymbol{\beta}}(\alpha)$, $\alpha \in (0, 1)$ ako riešenie úlohy:

$$\min \sum_{i=1}^n \rho_{\alpha}(Y_i - \mathbf{X}_i' \boldsymbol{\beta}), \quad \boldsymbol{\beta} \in \mathbb{R}^k,$$

kde $\rho_{\alpha}(x) = |x| \{ \alpha I_{[x>0]} + (1 - \alpha) I_{[x<0]} \}$, $x \in \mathbb{R}$.

Uvedenú úlohu možno riešiť simplexovou metódou, lebo $\rho_{\alpha}(x)$ je po častiach lineárna konvexná funkcia (pozri [13]). Výsledkom riešenia sú regresné kvantily. Pomocou regresných kvantilov je možné definovať L -odhady parametra $\boldsymbol{\beta}$ lineárneho regresného modelu (2.3). Najznámejším je regresný medián označovaný ako L_1 -odhad, ktorý reprezentuje α -kvantil pre $\alpha = \frac{1}{2}$. Pre praktické použitie je najzaujímavejší (α_1, α_2) -useknutý odhad metódou najmenších štvorcov:

$$T_n(\alpha_1, \alpha_2) = (\mathbf{X}' \mathbf{A}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{A} \mathbf{Y},$$

kde $\mathbf{A} = \text{diag}(a_i)$, $a_i = I[\mathbf{X}_i' \widehat{\boldsymbol{\beta}}(\alpha_1) < Y_i < \mathbf{X}_i' \widehat{\boldsymbol{\beta}}(\alpha_2)]$ sú váhy váženého odhadu metódou najmenších štvorcov a $0 < \alpha_1 < \alpha_2 < 1$. L -odhady regresného parametra sú ekvivariantné aj k regresii aj k merítku [13].

Riešením duálnej úlohy vznikne regresné skóre s poradím. Tieto skóre majú teda vlastnosti podobné vlastnostiam zoradených pozorovaní a dajú sa zapísať v tvare:

$$\widehat{\mathbf{a}}_n(\alpha) = (\widehat{a}_{n1}, \dots, \widehat{a}_{nn})', \quad 0 \leq \alpha \leq 1.$$

Z duality $\widehat{\boldsymbol{\beta}}(\alpha)$ a $\widehat{\mathbf{a}}_n(\alpha)$ plynie pre $i = 1, \dots, n$:

$$\widehat{a}_{ni} = \begin{cases} 1, & \text{ak } Y_i > \mathbf{X}_i' \widehat{\boldsymbol{\beta}}(\alpha), \\ \in (0, 1), & \text{ak } Y_i = \mathbf{X}_i' \widehat{\boldsymbol{\beta}}(\alpha), \\ 0, & \text{ak } Y_i < \mathbf{X}_i' \widehat{\boldsymbol{\beta}}(\alpha). \end{cases}$$

Regresné skóre s poradím sa často využívajú pre studentizáciu M -odhadov ako škálové štatistiky, keďže sú invariantné vzhľadom k regresii a ekvivariantné vzhľadom k merítku (pozri [13]).

4.5 R-odhady

Nech X_1, \dots, X_n je náhodný výber z rozdelenia so spojitou distribučnou funkciou a $R_i = \sum_{j=1}^n I_{[X_j \leq X_i]}$ je poradie X_i medzi X_1, \dots, X_n , $i = 1, \dots, n$. Pre empirickú distribučnú funkciu F_n výberu X_1, \dots, X_n je možné písať $R_i = nF_n(X_i)$, $i = 1, \dots, n$. Poradia sú invariantné k triede rýdzo monotónnych transformácií.

R -odhady boli navrhnuté ako inverzia testov založených na poradiach. Tieto testy majú tú vlastnosť, že za platnosti nulovej hypotézy rozdelenie testového kritéria nezávisí na distribučnej funkcii pozorovaní.

Pre náhodný výber X_1, \dots, X_n so spojitou distribučnou funkciou $F(x - \theta)$ so stredom symetrie θ sa hypotézu o strede symetrie ($H_0 : \theta = \theta_0$) testuje znamienkovým testom založeným na poradiach. Testová štatistika tohoto testu je v tvare:

$$S_n(\theta_0) = \sum_{i=1}^n \operatorname{sgn}(X_i - \theta_0) a_n(R_{|i|}^+(\theta_0)),$$

kde $R_{|i|}^+(\theta_0)$ je poradie $|X_i - \theta_0|$ medzi $|X_1 - \theta_0|, \dots, |X_n - \theta_0|$, $i = 1, \dots, n$ a $a_n(1) \leq \dots \leq a_n(n)$ sú skóre generované neklesajúcou skórovou funkciou $\varphi^+ : [0, 1) \rightarrow \mathbb{R}^+$, $\varphi^+(0) = 0$:

$$a_n(i) = \varphi^+ \left(\frac{i}{n+1} \right), \quad i = 1, \dots, n.$$

Za platnosti nulovej hypotézy a $F(x) + F(-x) = 1$, $x \in \mathbb{R}$ sú $\operatorname{sgn}(X_i - \theta_0)$ a $R_{|i|}^+(\theta_0)$ stochasticky nezávislé a $S_n(t)$ je nerastúca schodovitá funkcia. Preto za platnosti H_0 je $E_{\theta_0} S_n(\theta_0) = 0$ a rozdelenie $S_n(\theta_0)$ je symetrické okolo nuly.

Odhadom θ_0 je hodnota t , ktorá rieši rovnicu $S_n(t) = 0$. Keďže $S_n(t)$ nie je spojitá, nemusí riešenie tejto rovnice existovať. R -odhad sa preto definuje v tvare:

$$T_n = \frac{1}{2} \left(\sup_t \{S_n(t) > 0\} + \inf_t \{S_n(t) < 0\} \right).$$

R -odhady sú ekvivariantné vzhľadom k posunutiu aj vzhľadom k merítku [13]. Všeobecne možno R -odhad tvorený skórovou funkciou φ^+ vyjadriť ako riešenie rovnice:

$$\int_{-\infty}^{\infty} \varphi(F_n(y) - F_n(2T_n - y)) dF_n(y) = 0, \quad (4.14)$$

kde $\varphi(u) = \operatorname{sgn}(u - \frac{1}{2})\varphi^+(2u - 1)$, $0 < u < 1$.

4.5.1 Influenčná funkcia

Influenčná funkcia štatistického funkcionálu $T(F)$ príslušnému (4.14) je uvedená v [13] v tvare:

$$IF(x; T, F) = \frac{\varphi(F(x))}{\int_{\mathbb{R}} \varphi(F(y))(-f'(y))dy},$$

kde f je absolútne spojitá hustota symmetrickej distribučnej funkcie F .

4.5.2 Najpoužívanejšie R-odhady

Nech skóre $a_n(i)$ je generované skórovou funkciou φ v tvare:

$$a_n(i) = n \int_{\frac{i-1}{n}}^{\frac{i}{n}} \varphi(s) ds, \quad i = 1, \dots, n.$$

Za φ sa najčastejšie volí (pozri [10]):

- $\varphi(t) = t - \frac{1}{2} \quad \dots \quad$ Hodges-Lehmann

- $\varphi(t) = |t - \frac{1}{2}| - \frac{1}{4}$... Ansari-Bradley-Siegel-Tukey
- $\varphi(t) = (t - \frac{1}{2})^2 - \frac{1}{12}$... Mood
- $\varphi(t) = \Phi^{-1}(t)$... normálne skóre, kde Φ je distribučná funkcia normálneho rozdelenia
- $\varphi(t) = \Phi^{-1}(t)^2 - 1$... Klotz

Najznámejším R -odhadom je Hodges-Lehmannov odhad v tvare:

$$T_n = \operatorname{med}_{1 \leq i \leq j \leq n} \left\{ \frac{X_i - X_j}{2} \right\}.$$

Kapitola 5

Robustné odhady v mnohorozmernom CAPM modeli

Nech existuje CAPM model v tvare:

$$r_{ti} - r_{tf} = \alpha_i + \beta_i(r_{tm} - r_{tf}) + \epsilon_{ti}, \quad (5.1)$$

kde $i = 1, \dots, n$ označuje cenný papier a $t = 1, \dots, T$ predstavuje čas (pozri [2], [11] a [24]). Ide teda o mnohorozmerný lineárny model:

$$\begin{pmatrix} r_{11} - r_{1f} & \dots & r_{1n} - r_{1f} \\ \vdots & \ddots & \vdots \\ r_{T1} - r_{Tf} & \dots & r_{Tn} - r_{Tf} \end{pmatrix} = \\ = \begin{pmatrix} 1 & r_{1m} - r_{1f} \\ \vdots & \vdots \\ 1 & r_{Tm} - r_{Tf} \end{pmatrix} \begin{pmatrix} \alpha_1 & \dots & \alpha_n \\ \beta_1 & \dots & \beta_n \end{pmatrix} + \begin{pmatrix} \epsilon_{11} & \dots & \epsilon_{1n} \\ \vdots & \ddots & \vdots \\ \epsilon_{T1} & \dots & \epsilon_{Tn} \end{pmatrix},$$

kde r_{ti} , r_{tm} a r_{tf} sú miery zisku i -teho cenného papiera, trhového indexu a bezrizikového aktíva v čase t .

Za účelom ďalšieho popisu potrebnej teórie bude mnohorozmerný model CAPM (5.1) uvažovaný v tvare:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (5.2)$$

kde

- zložky matice \mathbf{Y} sú $y_{ti} = r_{ti} - r_{tf}$, $i = 1, \dots, n$ a $t = 1, \dots, T$,
- matica \mathbf{X} je typu $T \times 2$ a má v prvom stĺpci jednotky a v druhom stĺpci vektor so zložkami $x_{t2} = r_{tm} - r_{tf}$, $t = 1, \dots, T$,
- matica $\boldsymbol{\beta}$ je typu $2 \times n$ a má v prvom riadku vektor so zložkami α_i , $i = 1, \dots, n$ a v druhom riadku vektor so zložkami β_i , $i = 1, \dots, n$.

Ide teda o model s n rovnicami, pričom v každej rovnici je T pozorovaní.

5.1 Odhadovanie parametrov po zložkách

Pre model (5.2) je možné za platnosti predpokladov uvedených v kapitole 4.3 a za predpokladu, že jednotlivé cenné papiere $i = 1, \dots, n$ sú vzájomne nezávislé (t. j. ϵ_i sú nezávislé vektory), zapísať všeobecný M -odhad ako riešenie minimalizácie (pozri [16]):

$$M(\mathbf{b}_i) = \min_{\beta_i \in \mathbb{R}^2} \sum_{t=1}^T \rho(y_{ti} - \mathbf{X}_t \beta_i), \quad i = 1, \dots, n,$$

kde \mathbf{X}_t , $t = 1, \dots, T$ predstavuje riadkový vektor matice \mathbf{X} a β_i , $i = 1, \dots, n$ predstavuje stĺpcový vektor matice β . To znamená, že model (5.2) sa rozpadne na n nezávislých rovníc a v každej rovnici sa dá aplikovať M -odhad pre odhad parametrov β_i .

Túto úlohu je možné za určitých predpokladov uvedených v kapitole 4.3 prepísať do tvaru:

$$\sum_{t=1}^T \psi(y_{ti} - \mathbf{X}_t \beta_i) \mathbf{X}_t = 0, \quad i = 1, \dots, n,$$

pričom samozrejme musí platiť $\psi = \rho'$. Pomocou tohto prepisu sa teda úloha minimalizácie stáva úlohou riešenia sústavy rovníc. Existuje možnosť použiť aj studentizovaný odhad, ak sa v minimalizačnej úlohe nahradí funkcia ρ funkciou ρ_0 :

$$\rho(y_{ti} - \mathbf{X}_t \beta_i) = \rho_0 \left(\frac{y_{ti} - \mathbf{X}_t \beta_i}{s_i} \right),$$

kde s_i je škálovacia štatistika, t. j. skalárna veličina (pre podrobnosti pozri kap. 4.3.5). V tomto prípade je M -odhad riešením rovnice

$$\sum_{t=1}^T \psi \left(\frac{y_{ti} - \mathbf{X}_t \beta_i}{s_i} \right) \mathbf{X}_t = 0, \quad i = 1, \dots, n.$$

5.2 Odhadovanie parametrov simultánne

Cieľom tejto kapitoly je zohľadniť korelácie medzi jednotlivými rovnicami v modeli (5.2) podobne, ako bolo ukázané v kapitole 2.2.1 pre OLS-odhad.

Model v tvare (5.2) možno pomocou Vec operátoru a Kroneckerovho súčinu prepísať do tvaru:

$$Vec(\mathbf{Y}) = (\mathbf{I}_n \otimes \mathbf{X}) Vec(\beta) + Vec(\epsilon),$$

kde pre $Vec(\epsilon)$ platí:

$$E[Vec(\epsilon)] = 0 \quad \text{a} \quad var[Vec(\epsilon)] = \Sigma \otimes \mathbf{I}_T,$$

kde Σ je pozitívne definitná matica.

Je zrejmé, že zložky $Vec(\mathbf{Y})$, ktoré pre tento model predstavujú jeden vektor, a riadkové vektory matice $(\mathbf{I}_n \otimes \mathbf{X})$ je možné preindexovať pomocou indexu $k = 1, \dots, T \cdot n$. Tento model sa dá totiž podobne ako v kapitole 2.2.1 prepísať do tvaru:

$$\mathbf{y} = \mathbf{x}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\epsilon}},$$

kde $\mathbf{y} = Vec(\mathbf{Y})$, $\mathbf{x} = (\mathbf{I}_n \otimes \mathbf{X})$, $\tilde{\boldsymbol{\beta}} = Vec(\boldsymbol{\beta})$, $\tilde{\boldsymbol{\epsilon}} = Vec(\boldsymbol{\epsilon})$.

Pre tento model je všeobecný M -odhad riešením minimalizačnej úlohy:

$$M(Vec(\mathbf{b})) = \min_{Vec(\mathbf{b}) \in \mathbb{R}^{2 \cdot n}} \sum_{k=1}^{T \cdot n} \rho(y_k - \mathbf{x}_k \cdot Vec(\boldsymbol{\beta})).$$

Táto úloha sa dá za určitých predpokladov uvedených v kapitole 4.3 prepísať do tvaru:

$$\sum_{k=1}^{T \cdot n} \psi(y_k - \mathbf{x}_k \cdot Vec(\boldsymbol{\beta})) \mathbf{x}_k = 0,$$

pričom samozrejme musí platiť $\psi = \rho'$.

Asymptotická variančná matica odhadu je odvodená v článku [16].

Poznámka: Možno použiť aj studentizovaný odhad, ak sa v minimalizačnej úlohe nahradí funkcia ρ funkciou ρ_0 , ako bolo uvedené v predchádzajúcej kapitole.

5.3 Simulačná štúdia

Odhady po zložkách a simultánný M -odhad parametrov daného modelu v tvare (5.2) nie sú všeobecne rovnaké. Odhady sa líšia väčšinou až na štvrtom desatinnom mieste, viac sa líšia relatívne eficiencie jednotlivých konštruovaných odhadov. Ilustrujú to výsledky nasledujúcej simulačnej štúdie spracované v programe R-project¹. Výstupy v praktickej časti boli inšpirované najmä článkom [11] a skriptami k programu R-project [23] a [22].

Cieľom praktickej štúdie bolo vytvoriť niekoľko možných simulácií modelu (5.2) z rozdelení, ktoré by mohli reálne v praxi nastať, resp. z rozdelenia r_{tm} pozorovaného na dátach. Pre účely simulácií bol použitý model (5.2) v zjednodušenom tvare:

$$r_{ti} = \alpha_i + \beta_i r_{tm} + \epsilon_{ti}, \quad (5.3)$$

kde $t = 1, \dots, T$ a $i = 1, \dots, n$ boli zvolené vzhľadom k počtu pozorovaní v dátovom vzorku.

V dátovom vzorku boli k dispozícii pozorovania záverečných cien P_t z obdobia 01.11.1993 - 03.04.2003 pre 3 cenné papiere (Microsoft, GE a Ford), preto $n = 3$,

¹ V ďalšom texte bude desatinná čiarka označovaná bodkou. V softvéri R-project boli použité tieto package: mvtnorm, VGAM, fGarch, SpatialNP, quantreg, car, ellipse, MASS. V jednotlivých skriptoch sú uvádzané komentáre k jednotlivým vykonaným úkonom.

ďalej v dátach bol k dispozícii 1 bezrizikový cenný papier (T-bill) a trhový index (S&P 500)²:

- z 2363 denných pozorovaní cien³ bolo vytvorených 2362 denných pozorovaní výnosov, preto $T = 2362$ a
- z 113 mesačných pozorovaní cien⁴ bolo vytvorených 112 mesačných pozorovaní výnosov, preto $T = 112$.

K transformácii cien na výnosy bol použitý vzťah:

$$r_{ti} = 100[\log(P_{ti}) - \log(P_{t-1,i})].$$

Článok [11] uvádza, že rozdelenie výnosov jednotlivých cenných papierov sa odlišuje od normálneho rozdelenia. Toto rozdelenie má na rozdiel od normálneho ťažké konce a najčastejšie sa uvažuje Studentovo t -rozdelenie. Vo vyššie zmienenom článku je tiež uvedené, že reziduá majú aproximatívne rovnaké rozdelenie ako cenné papiere vstupujúce do modelu za použitia OLS-odhadu. Preto aj reziduá majú rozdelenie s ťažkými koncami. Ďalej sa v článku uvádza, že stupne voľnosti pre t -rozdelenie⁵ by mali byť od 2 do 10, najčastejšie 3.

Pre vektor r_{tm} bolo (ako jedna varianta) v simuláciách uvažované normované normálne rozdelenie $N(0, 1)$. Ďalej bolo v štúdií predpokladané, že je tento vektor daný $GARCH(1, 1)$ modelom so štandardne nastavenými parametrami v programe R-project⁶. Pre účely simulácií bolo tiež pomocou programu R-project skúmané skutočné rozdelenie r_{tm} . Za prípustné rozdelenia boli považované normálne rozdelenie a t -rozdelenie⁷. Pre maticu chýb ϵ boli uvažované mnohorozmerné normálne rozdelenie (s jednotkovou variančnou maticou a špecifickou variančnou maticou) a mnohorozmerné Studentovo t -rozdelenie s 2, 3, 5 a 10 stupňami voľnosti

² Dáta pochádzajú z knihy „*Statistics and Data Analysis for Financial Engineering*“ profesora Davida Rupperta, dostupné na webovej stránke „legacy.orie.cornell.edu/~davidr/SDAFE/data/capm.csv“.

³ Denné pozorovania boli k dispozícii ku koncu jednotlivých obchodných dní.

⁴ Mesačné pozorovania boli stanovené vždy k poslednému obchodnému dňu mesiaca.

⁵ t_v označuje Studentovo t -rozdelenie s v stupňami voľnosti. V dokumentácii k programu R-project je uvedené, že v nemusí byť celé číslo (v takom prípade sme overili, že generovanie čísel pre takto definované t -rozdelenie nepochádza z rozdelenia so stupňami voľnosti $\lfloor v \rfloor$ ani $\lceil v \rceil$, preto sa v práci môže vyskytnúť t -rozdelenie s necelými stupňami voľnosti). Ďalej sa v dokumentácii uvádza, že hustota t -rozdelenia s v stupňami voľnosti má hustotu v tvare:

$$f(x) = \frac{\Gamma\left(\frac{(n+1)}{2}\right)}{(\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right))} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

pre všetky x reálne.

⁶ U_t sa riadi modelom $GARCH(1, 1)$, ak

$$U_t = h_t \cdot Z_t, \quad t \in \mathbb{Z},$$

kde Z_t sú nezávislé rovnako rozdelené náhodné veličiny, $E[Z_t] = 0$, $var[Z_t] = 1$ a platí:

$$h_t = a_0 + a_1 U_{t-1}^2 + b_1 h_{t-1}^2, \quad a_0, a_1, b_1 > 0, \quad t \in \mathbb{Z}.$$

V programe R-project sú parametre nastavené ako $a_0 = 1e-6$, $a_1 = 0.1$, $b_1 = 0.8$ a $Z_t \sim N(0, 1)$.

⁷ K odhadovaniu parametrov rozdelení $N(\mu, \sigma)$ a t_v bola použitá funkcia „`fitdistr`“. Pre vektor r_{tm} boli ako prípustné rozdelenia testované $N(0, 1)$, $N(\mu, \sigma)$, t_3 a t_v .

(so špecifickou variančnou maticou). V jednotlivých simuláciách bola pre mnohorozmerné rozdelenia špecifikovaná regulárna variančná matica V :

$$V = \begin{pmatrix} 1 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1 \end{pmatrix}.$$

Celkom bolo nasimulovaných 18 kombinácií rozdelení ako pre denné simulované pozorovania, tak aj pre mesačné pozorovania. Použité funkcie v programe R-project a jednotlivé simulované rozdelenia pre r_{tm} a maticu chýb ϵ sú uvedené v nasledujúcej tabuľke (pozri tab. 5.1).

Rozdelenie (model pre) r_{tm} [použitá funkcia R-project]	Mnohorozmerné rozdelenie ϵ [použitá funkcia R-project]	Označenie v ďalšom texte
Normované normálne rozdelenie $N(0, 1)$ [rnorm]	t_2 -rozdelenie [rmvt] t_3 -rozdelenie [rmvt] t_5 -rozdelenie [rmvt] t_{10} -rozdelenie [rmvt] Normálne [rmvnorm] Normované normálne [rmvnorm]	norm_student2 norm_student3 norm_student5 norm_student10 norm_norm norm_nonorm
Model $GARCH(1, 1)$ [garchSim]	t_2 -rozdelenie [rmvt] t_3 -rozdelenie [rmvt] t_5 -rozdelenie [rmvt] t_{10} -rozdelenie [rmvt] Normálne [rmvnorm] Normované normálne [rmvnorm]	norm_student2 norm_student3 norm_student5 norm_student10 norm_norm norm_nonorm
t -rozdelenie odhadnuté z dát pre denné pozorovania: odhadnuté stupne voľnosti boli 3.99 [rt]	t_2 -rozdelenie [rmvt] t_3 -rozdelenie [rmvt] t_5 -rozdelenie [rmvt] t_{10} -rozdelenie [rmvt] Normálne [rmvnorm] Normované normálne [rmvnorm]	datastudent_student2 datastudent_student3 datastudent_student5 datastudent_student10 datastudent_norm datastudent_nonorm
Normálne rozdelenie $N(\mu, \sigma)$ odhadnuté z dát pre mesačné pozorovania: $N(1.49, 7.38^2)$ [rnorm]	t_2 -rozdelenie [rmvt] t_3 -rozdelenie [rmvt] t_5 -rozdelenie [rmvt] t_{10} -rozdelenie [rmvt] Normálne [rmvnorm] Normované normálne [rmvnorm]	datanorm_student2 datanorm_student3 datanorm_student5 datanorm_student10 datanorm_norm datanorm_nonorm

Tabuľka 5.1: Simulované rozdelenia pre r_{tm} a ϵ .

Pre jednoduchosť výpočtu bolo v simulačnej štúdii stanovené, že prvý riadok matice β , t. j. vektor so zložkami α_i , je nulový. Vektor so zložkami β_i bol simulovaný náhodne z intervalu $(0, 2)$. Skonstruovaný počet simulácií sa odvíjal od času výpočtu v programe R-project:

- pre denné pozorovania bolo vytvorených 250 simulácií⁸,
- pre mesačné pozorovania bolo vytvorených 1000 simulácií⁹.

⁸ Doba spracovania bola približne 6 hodín.

⁹ Doba spracovania bola približne 3.5 hodiny.

V každej simulácii bol vytvorený vektor pozorovaní r_{tm} s dĺžkou T , matica reziduí (ϵ_{ti}) typu $T \times n$ a matica parametrov modelu β typu $2 \times n$. Následne bol stanovený vektor¹⁰ r_{ti} a odhadnuté hodnoty α_i a β_i z modelu (5.3).

Funkcie pre odhadovanie parametrov boli zvolené na základe teórie uvedenej v predchádzajúcich kapitolách. Zvolené funkcie aplikované na simulované dáta sú uvedené v tab. 5.2. Pre simultánne odhady bol použitý model prepísaný pomocou *Vec* operátoru a Kroneckerovho súčinu (funkcie „as.vector“ a „kronecker“ v R-project) ako bolo uvedené v kap. 5.2. Odhady po zložkách vychádzali z teórie uvedenej v kap. 5.1. Teória OLS-odhadu bola popísaná v kap. 2.1.1 a 2.2.1, teória M -odhadov bola uvedená v kap. 4.3. Funkcie, ktoré boli pre tieto odhady použité v programe R-project sú taktiež uvedené v nasledujúcej tabuľke (pozri tab. 5.2).

Názov odhadovej funkcie	Popis	Použitá funkcia v R-project
OLSmulti OLS	simultánny OLS-odhad OLS-odhad skonštruovaný po zložkách	lm
L1multi L1	simultánny L_1 -odhad L_1 -odhad skonštruovaný po zložkách	rq
HUBERmulti HUBER	simultánny Huberov M -odhad Huberov M -odhad skonštruovaný po zložkách	rlm
HAMPELmulti HAMPEL	simultánny Hampelov M -odhad Hampelov M -odhad skonštruovaný po zložkách	rlm
TUKEYmulti TUKEY	simultánny Tukeyho M -odhad Tukeyho M -odhad skonštruovaný po zložkách	rlm

Tabuľka 5.2: Funkcie pre odhadovanie parametrov.

Na základe simulácií boli skonštruované výstupné tabuľky (pozri Prílohy pre všetky tabuľky¹¹) v nasledujúcej štruktúre uvedenej v ukázkovej tab. 5.3, kde v ľavom hornom rohu je označenie zvolenej kombinácie rozdelení (pozri tab. 5.1). V softvéri R-project boli porovnané odhady parametrov β_i (odhadnuté hodnoty na základe funkcií z tab. 5.2) s pôvodnými simulovanými hodnotami β_i (skutočné hodnoty), keďže tieto odhady boli primárnym cieľom práce. Výstupom uvedenej tabuľky potom bolo aj porovnanie stredných hodnôt absolútnych odchýliek skutočných a odhadnutých hodnôt (ozn. $AbsSD_{beta_i}$).

Stredné hodnoty absolútnych odchýliek skutočných a odhadnutých hodnôt boli skonštruované podľa vzťahu:

$$AbsSD_{beta_i} = E|(b_i - \beta_i)_k| = \frac{1}{k} \sum_{i=1}^k |(b_i - \beta_i)_k|,$$

¹⁰ Z modelu (5.3), keďže všetky veličiny na pravej strane sú známe.

¹¹ V týchto tabuľkách nie sú uvádzané hodnoty α_i a β_i , ale iba stredné hodnoty absolútnych odchýliek, efíciencia a pridelené skóre. Hodnoty odhadov nie sú pre účely porovnania jednotlivých simulácií vypovedajúce, najmä čo sa týka odhadov β_i , ktorých simulované hodnoty pochádzajú z intervalu $(0, 2)$ a ich hodnota je teda pre každú jednu simuláciu rôzna. Tabuľky v plnej štruktúre možno nájsť na priloženom CD.

kde b_i je odhad simulovaného parametra β_i a k označuje poradie jednotlivých simulácií. Z týchto stredných hodnôt je následne spočítaný pomer udávajúci relatívnu efektivitu pre odhady skonštruované po zložkách:

$$\text{relatívna efektivita } \beta_i = \frac{AbsSD_{beta_i}^{OLS}}{AbsSD_{beta_i}^{MODEL}},$$

kde „MODEL“ predstavuje funkciu pre odhad po zložkách z tab. 5.2, resp. pomer dávajúci relatívnu efektivitu pre simultánne odhady:

$$\text{relatívna efektivita } \beta_i = \frac{AbsSD_{beta_i}^{OLSmulti}}{AbsSD_{beta_i}^{MODELmulti}},$$

kde „MODELmulti“ predstavuje funkciu pre simultánny odhad z tab. 5.2. Pomer „relatívna efektivita β_i “ bol teda vypočítaný zvlášť pre odhady po zložkách a zvlášť pre simultánne odhady¹².

norm_student3	$alfa_1$	$alfa_2$	$alfa_3$	$beta_1$	$beta_2$	$beta_3$	$AbsSD_{beta_1}$
pôvodné (simulované)	0.0000	0.0000	0.0000	0.9946	0.9981	0.9885	0.0000
OLSmulti	-0.0019	0.0017	-0.0009	0.9953	0.9955	0.9869	0.1286
OLS	-0.0019	0.0017	-0.0009	0.9953	0.9955	0.9869	0.1286
L1multi	-0.0013	0.0021	0.0028	0.9881	0.9954	0.9906	0.1069
L1	-0.0013	0.0021	0.0028	0.9881	0.9954	0.9906	0.1069
HUBERmulti	-0.0020	0.0004	0.0007	0.9921	0.9971	0.9902	0.0963
HUBER	-0.0016	0.0006	0.0007	0.9919	0.9971	0.9901	0.0963
HAMPELmulti	-0.0017	0.0012	0.0005	0.9923	0.9982	0.9911	0.0992
HAMPEL	-0.0014	0.0012	0.0007	0.9920	0.9983	0.9912	0.0988
TUKEYmulti	-0.0010	-0.0003	0.0011	0.9906	0.9974	0.9925	0.0945
TUKEY	-0.0006	-0.0001	0.0012	0.9901	0.9972	0.9930	0.0946

	$AbsSD_{beta_2}$	$AbsSD_{beta_3}$	Relatívna efektivita $beta_1$	Relatívna efektivita $beta_2$	Relatívna efektivita $beta_3$	Priemerná relatívna efektivita	Skóre
pôvodné (simulované)	0.0000	0.0000	Inf	Inf	Inf	0.0000	
OLSmulti	0.1314	0.1231	1.0000	1.0000	1.0000	1.0000	0
OLS	0.1314	0.1231	1.0000	1.0000	1.0000	1.0000	0
L1multi	0.1071	0.1041	1.2021	1.2269	1.1824	1.2038	0
L1	0.1071	0.1041	1.2021	1.2269	1.1824	1.2038	0
HUBERmulti	0.0988	0.0943	1.3354	1.3297	1.3056	1.3236	1
HUBER	0.0987	0.0945	1.3350	1.3313	1.3036	1.3233	0
HAMPELmulti	0.1021	0.0976	1.2961	1.2866	1.2612	1.2813	0
HAMPEL	0.1020	0.0976	1.3009	1.2880	1.2619	1.2836	0
TUKEYmulti	0.0982	0.0936	1.3598	1.3377	1.3159	1.3378	3
TUKEY	0.0984	0.0938	1.3594	1.3351	1.3131	1.3359	2

Tabuľka 5.3: Ukážka výstupnej tabuľky pre simulácie modelu.

Z relatívnej efektivity každého odhadu β_i bola následne vypočítaná priemerná relatívna efektivita, aby bolo možné porovnať výsledky a zvoliť vhodný model

¹² Ak je tento pomer napríklad vo výške 1.2, tak udáva, že odhad metódou najmenších štvorcov bude rovnako dobrý ako robustný odhad práve vtedy, keď bude dátová vzorka pre odhad metódou najmenších štvorcov o 20 % väčšia.

pre reálne dáta. Odhadu, ktorý mal oproti OLS-odhadu najvyššiu priemernú relatívnu efektivitu, t. j. najlepšiemu odhadu bolo priradené skóre „3“, druhému najlepšiemu odhadu bolo priradené skóre „2“ a tretiemu skóre „1“. Ďalším odhadom nebolo priradené žiadne skóre. Na základe priradeného skóre boli porovnané jednotlivé odhady. Toto porovnanie bolo urobené zvlášť pre denné (pozri tab. 5.4) a mesačné simulácie (pozri tab. 5.5). V týchto tabuľkách je uvedený súčet skóre pre:

- všetky simulované kombinácie rozdelení („Súčet skóre“),
- tie kombinácie rozdelení, v ktorých sa neuvažujú chyby simulované z mnoho-rozmerného normálneho rozdelenia („Súčet skóre (bez `_norm/_nonorm`)“),
- tie kombinácie rozdelení, v ktorých majú chyby Studentovo t_3 alebo t_5 -rozdelenie („Súčet skóre (iba `_student3/_student5`)“ a
- tie kombinácie rozdelení, v ktorých majú r_{tm} Studentovo t_v -rozdelenie odhadnuté z reálnych dát („Súčet skóre (iba `datastudent_`)“ pre denné pozorovania, resp. normálne rozdelenie $N(\mu, \sigma)$ odhadnuté z reálnych dát („Súčet skóre (iba `datanorm_`)“ pre mesačné pozorovania.

Porovnanie pre denné pozorovania	Súčet skóre	Súčet skóre (bez <code>_norm/_nonorm</code>)	Súčet skóre (iba <code>_student3/_student5</code>)	Súčet skóre (iba <code>datastudent_</code>)
OLSmulti	16	0	0	6
OLS	10	0	0	4
L1multi	0	0	0	0
L1	0	0	0	0
HUBERmulti	15	15	8	7
HUBER	13	13	6	5
HAMPELmulti	11	5	0	3
HAMPEL	6	2	0	0
TUKEYmulti	20	20	12	7
TUKEY	17	17	10	4

Tabuľka 5.4: Porovnanie odhadov na simulovaných denných pozorovaniach.

Porovnanie pre mesačné pozorovania	Súčet skóre	Súčet skóre (bez <code>_norm/_nonorm</code>)	Súčet skóre (iba <code>_student3/_student5</code>)	Súčet skóre (iba <code>datanorm_</code>)
OLSmulti	18	0	0	6
OLS	12	0	0	4
L1multi	0	0	0	0
L1	0	0	0	0
HUBERmulti	13	13	9	5
HUBER	9	9	5	4
HAMPELmulti	12	9	0	4
HAMPEL	7	4	0	2
TUKEYmulti	21	21	14	6
TUKEY	16	16	8	5

Tabuľka 5.5: Porovnanie odhadov na simulovaných mesačných pozorovaniach.

Vyhodnotenie porovnania jednotlivých odhadov vychádzalo z počtu dosiahnutých skóre pre simulované rozdelenia:

- (1) Po zohľadnení všetkých simulovaných kombinácií rozdelení bol vyhodnotený ako najlepší:
 - pre denné pozorovania odhad „TUKEYmulti“ a následne „TUKEY“,
 - pre mesačné pozorovania odhad „TUKEYmulti“ a následne „OLSmulti“
- (2) Za predpokladu, že chyby nemôžu nadobúdať mnohorozmerné normálne rozdelenie bol vyhodnotený ako najlepší:
 - pre denné pozorovania odhad „TUKEYmulti“ a následne „TUKEY“,
 - pre mesačné pozorovania odhad „TUKEYmulti“ a následne „TUKEY“.
- (3) Za predpokladu, že chyby môžu nadobúdať iba Studentovo rozdelenie s tromi alebo piatimi stupňami voľnosti bol vyhodnotený ako najlepší:
 - pre denné pozorovania odhad „TUKEYmulti“ a následne „TUKEY“,
 - pre mesačné pozorovania odhad „TUKEYmulti“ a následne „HUBERMmulti“.
- (4) Pre skutočné rozdelenie reálnych dát r_{tm} bol po zohľadnení všetkých simulovaných rozdelení matice chýb vyhodnotený ako najlepší:
 - pre denné pozorovania odhad „TUKEYmulti“, „HUBERMmulti“ a následne „OLSmulti“,
 - pre mesačné pozorovania odhad „TUKEYmulti“, „OLSmulti“ a následne „HUBERMmulti“, „TUKEY“.

Keďže bolo v teoretickej časti popísané, že simultánny odhad zohľadňuje všetky korelácie pozorovaní r_{ti} , zameriava sa ďalší text iba na tieto odhady¹³. Z vyhodnotenia je navyše zrejmé, že bez ohľadu na počet pozorovaní v simulačnom vzorku vyšiel jednoznačne ako najlepší Tukeyho simultánny M -odhad. Na reálne dáta bola preto aplikovaná táto odhadová funkcia.

Ďalej boli v rámci simulačnej štúdie pre skutočné rozdelenia vektora r_{tm} skúmané aj stredné kvadratické odchýlky (MSE) podľa vzťahu:

$$MSE_{beta_i} = E[(b_i - \beta_i)_k^2] = \frac{1}{k} \sum_{i=1}^k (b_i - \beta_i)_k^2,$$

kde b_i je odhad simulovaného parametra β_i a k označuje poradie jednotlivých simulácií. Rovnaký vzťah bol aplikovaný aj na odhady α_i .

¹³ To potvrdzuje aj „Súčet skóre“ z tab. a, ktorý pre všetky simultánne odhady dáva pre denné pozorovania súčet 62 a pre mesačné pozorovania súčet 64. Pre odhady po zložkách sú tieto súčty v oboch prípadoch nižšie, pre denné pozorovania to je 46, pre mesačné iba 44 bodov. Podobnú úvahu možno urobiť aj pre posledné stĺpce zohľadňujúce rozdelenie reálnych dát. Navyše pre všetky stĺpce uvedených tabuliek platí, že simultánny odhad („ODHADmulti“) má vždy vyššie skóre ako odhad po zložkách („ODHAD“).

Zo stredných kvadratických odchýliek je následne vypočítaný pomer udávajúci relatívnu efícienciu, a to pre MSE_{α_i} , ale aj pre MSE_{β_i} analogicky podľa vzťahov uvedených vyššie pre $AbsSD_{beta_i}$. Pomer „relatívna efíciencia β_i “ a „relatívna efíciencia α_i “ bol teda opäť vypočítaný zvlášť pre odhady po zložkách a zvlášť pre simultánne odhady. Z relatívnej efíciencie odhadov β_i bola následne spočítaná „priemerná relatívna efíciencia beta“ a z relatívnej efíciencie odhadov matice β bola vypočítaná „priemerná relatívna efíciencia“.

Odhadom bolo potom priradené skóre podľa princípu uvedeného vyššie, a to jednak na základe hodnoty „Skóre beta“, ktoré udáva priemernú efícienciu odhadov parametrov β_i a „Skóre“, ktoré udáva priemernú efícienciu všetkých odhadov matice β . Ukážka výstupnej tabuľky je uvedená v tab. 5.6. V tab. 5.7 sú uvedené súčty dosiahnutých skóre pre všetkých 6 kombinácií simulovaných rozdelení pre jednotlivé odhadové funkcie (pre denné aj mesačné pozorovania).

datanorm- _student3	Relatívna efíciencia α_1	Relatívna efíciencia α_2	Relatívna efíciencia α_3	Relatívna efíciencia β_{a1}	Relatívna efíciencia β_{a2}	Relatívna efíciencia β_{a3}	Priemerná rel. efíciencia β_{beta}	Priemerná relatívna efíciencia	Skóre β_{beta}	Skóre
OLSmulti	1	1	1	1	1	1	1	1	3	0
OLS	1	1	1	1	1	1	1	1	2	0
L1multi	0.0848	49.865	6.6585	0.6257	0.4269	1.5716	0.8747	9.8721	1	0
L1	0.0848	49.865	6.6585	0.6257	0.4269	1.5716	0.8747	9.8721	0	0
HUBERmulti	0.1295	722.7612	2.8084	0.0747	0.6198	1.4274	0.7073	121.3035	0	3
HUBER	0.1412	516.1053	2.9639	0.0729	0.626	1.2824	0.6605	86.8653	0	2
HAMPELmulti	0.8761	152.166	5.6048	0.0565	0.6998	1.3379	0.6981	26.7902	0	1
HAMPEL	1.0547	148.0879	4.6258	0.0569	0.6473	1.3051	0.6697	25.9629	0	0
TUKEYmulti	0.0907	16.2726	4.0903	0.0543	0.5436	1.9237	0.8406	3.8292	0	0
TUKEY	0.1084	17.9487	4.5252	0.0567	0.5317	1.7192	0.7692	4.1483	0	0

Tabuľka 5.6: Ukážka výstupnej tabuľky pre MSE odchýlky.

Porovnanie pre denné pozorovania	Súčet skóre β_i	Súčet skóre	Porovnanie pre me- sačné pozorovania	Súčet skóre β_i	Súčet skóre
OLSmulti	0	0	OLSmulti	3	3
OLS	0	0	OLS	2	2
L1multi	9	9	L1multi	2	6
L1	6	6	L1	0	4
HUBERmulti	4	4	HUBERmulti	7	7
HUBER	2	2	HUBER	7	5
HAMPELmulti	2	2	HAMPELmulti	1	2
HAMPEL	1	1	HAMPEL	3	3
TUKEYmulti	6	6	TUKEYmulti	5	2
TUKEY	6	6	TUKEY	6	2

Tabuľka 5.7: Porovnanie odhadov na základe MSE.

Skóre pridelené jednotlivým odhadom sa veľmi nelíšilo v závislosti na tom, či toto skóre bolo pridelené iba na základe priemernej relatívnej efíciencie parametrov β_i alebo všetkých parametrov (ilustruje to tab. 5.7). Vyhodnotenie porovnania jednotlivých odhadov je preto nasledovné:

- (1) vyhodnotenie iba na základe priemernej efíciencie β_i („Súčet skóre β_i “):

- pre denné pozorovania odhad „L1multi“ a následne „L1“, „TUKEY-multi“, „TUKEY“,
 - pre mesačné pozorovania odhad „HUBERmulti“, „HUBER“ a následne „TUKEY“.
- (2) vyhodnotenie na základe priemernej efektivity všetkých odhadov matice β („Súčet skóre“):
- pre denné pozorovania odhad „L1multi“ a následne „L1“, „TUKEY-multi“, „TUKEY“,
 - pre mesačné pozorovania odhad „HUBERmulti“ a následne „L1multi“.

Na reálne dáta bol sa na základe všetkých uvedených analýz aplikovaný aj Huberov simultánný M -odhad¹⁴. Na reálne dáta sú v ďalších kapitolách aplikované odhady, ktoré uvádza nasledujúca tabuľka:

	Odhad s vyššou prioritou	Odhad s nižšou prioritou
Denné pozorovania, $T = 2362$	TUKEYmulti	HUBERmulti
Mesačné pozorovania, $T = 112$	TUKEYmulti	HUBERmulti

Tabuľka 5.8: Odhady aplikované na reálne dáta.

5.4 Aplikácia odhadov na reálne dáta

V tejto kapitole bolo vybraných niekoľko portfólií, ktoré boli použité v mnohorozmernom CAPM modeli pre odhadovanie miery rizika beta a miery nerovnováhy alfa. K odhadovaniu boli využité funkcie stanovené na základe simulačnej analýzy.

5.4.1 Základné portfólio

Ako bolo uvedené v predchádzajúcej kapitole, v základnom dátovom vzorku (pre ktorý bola vykonaná aj simulačná štúdia) boli k dispozícii pozorovania záverečných cien P_t z obdobia 01.11.1993 - 03.04.2003 pre 3 cenné papiere (Microsoft, GE a Ford), preto $n = 3$, ďalej v dátach bol k dispozícii 1 bezrizikový cenný papier (T-bill) a trhoví index (S&P 500). Z pozorovaní cien boli vytvorené pozorovania výnosov (denné a mesačné). Na tieto výnosy bol následne aplikovaný model (5.2).

Denné pozorovania výnosov

Dáta denných výnosov (jednotlivých CP) boli analyzované pomocou niekoľkých štatistických metód. Prvou použitou metódou bola analýza popisných štatistík. Bola vytvorená tabuľka obsahujúca minimum, maximum, priemer, medián,

¹⁴ Rozhodnutie bolo urobené na základe výsledkov skóre pre tento odhad vo všetkých uvedených tabuľkách - pre denné pozorovania bol tento odhad na základe analýzy „AbsSD“ vyhodnotený ako rovnako dobrý ako Tukeyho simultánný M -odhad, pre mesačné pozorovania voľbu Huberovho simultánneho M -odhadu potvrdila analýza „MSE“. Týmto odhadom ale neprikladáme takú veľkú váhu ako Tukeyho odhadom.

smerodajnú odchýlku, šikmosť a špicatosť jednotlivých analyzovaných CP. Šikmosť udáva, či je rozdelenie symetrické¹⁵, špicatosť udáva, či dáta majú špicatejšie rozdelenie vzhľadom k normálnemu rozdeleniu¹⁶. Táto analýza je uvedená v nasledujúcej tabuľke. Hodnoty výnosov jednotlivých CP viac menej kolíšu okolo nulovej hodnoty (priemer, medián).

Charakteristika	Minimum	Priemer	Medián	Maximum	Smerodajná odchýlka	Šikmosť	Špicatosť
tbill	-19.7531	-0.0437	0.0000	8.7825	1.2592	-2.9938	42.3546
MSFT	-16.9692	0.0983	0.0000	17.8826	2.4546	-0.1031	3.9308
SP500	-7.1127	0.0265	0.0342	5.5744	1.1559	-0.1023	3.2258
GE	-11.2989	0.0597	0.0000	11.7487	1.9232	0.0230	3.0497
Ford	-15.8690	-0.0015	0.0000	14.4914	2.2961	0.1151	3.9339

Tabuľka 5.9: Popisné štatistiky pre denné výnosy jednotlivých CP.

Ďalej bola skonštruovaná matica korelácií. Korelácie medzi jednotlivými výnosmi CP reprezentujúcimi r_{ti} neprekročili hranicu 0.4, teda tu nebola pozorovaná významná závislosť jednotlivých výnosov. Korelácia výnosov a bezrizikového aktíva taktiež nebola významná, bola menšia než 0.11 (pozri tab. 5.10).

	tbill	MSFT	SP500	GE	Ford
tbill	1.0000	0.0465	0.1003	0.1043	0.0889
MSFT	0.0465	1.0000	0.5860	0.3901	0.2657
SP500	0.1003	0.5860	1.0000	0.7478	0.4900
GE	0.1043	0.3901	0.7478	1.0000	0.3972
Ford	0.0889	0.2657	0.4900	0.3972	1.0000

Tabuľka 5.10: Korelácie denných výnosov jednotlivých CP.

Hlavnou časťou analýzy boli hodnoty $r_{ti} - r_{tf}$ a $r_{tm} - r_{tf}$ (výnosy jednotlivých CP očistené o výnos bezrizikového CP), ktoré vstupujú do modelu (5.2):

Charakteristika	Minimum	Priemer	Medián	Maximum	Smerodajná odchýlka	Šikmosť	Špicatosť
MSTF-tbill	-16.6127	0.1421	0.0930	17.8826	2.7062	0.1886	3.4352
GE-tbill	-10.7787	0.1034	0.0713	15.1851	2.1861	0.3920	3.3792
Ford-tbill	-13.0669	0.0422	-0.1047	18.7419	2.5187	0.6497	4.7113
SP500-tbill	-9.2029	0.0702	0.0463	14.7785	1.6217	1.0349	10.4546

Tabuľka 5.11: Popisné štatistiky pre denné pozorovania premenných modelu.

¹⁵ Kladný koeficient šikmosti hovorí, že ide o pozitívne (pravostranne) zošikmené rozdelenie.

¹⁶ Kladný koeficient špicatosti hovorí, že dané rozdelenie pravdepodobnosti je špicatejšie než normálne rozdelenie (vo vzorci pre špicatosť sa za týmto účelom odčíta hodnota „3“).

Od popisných štatistík výnosov r_{ti} sa významne odlišovali takmer všetky popisné štatistiky (pozri tab. 5.9 a tab. 5.11), ale najmä šikmosť a špicatosť udávajúce tvar rozdelenia.

Ďalej bola pre hodnoty $r_{ti} - r_{tf}$ skonštruovaná matica korelácií. Korelácie medzi jednotlivými výnosmi vykazovali významnejšiu závislosť, než pre jednotlivé CP analyzované vyššie (r_{ti}) (pozri tab. 5.12).

	MSTF-tbill	GE-tbill	Ford-tbill	SP500-tbill
MSTF-tbill	1.0000	0.5123	0.3936	0.6742
GE-tbill	0.5123	1.0000	0.5140	0.8038
Ford-tbill	0.3936	0.5140	1.0000	0.6080
SP500-tbill	0.6742	0.8038	0.6080	1.0000

Tabuľka 5.12: Korelácie pre denné pozorovania premenných modelu.

Rozdelenie, ktoré bolo pre účely ďalšej analýzy smerodajné, bolo rozdelenie vektora $r_{tm} - r_{tf}$, t. j. „SP500-tbill“. Ako bolo uvedené v predchádzajúcej kapitole, prípustné rozdelenia pre tento vektor boli:

- normálne rozdelenie a
- Studentovo t -rozdelenie.

Tieto rozdelenia boli v programe R-project testované pomocou Kolmogorovho-Smirnovho testu (KS-test) pre dva výbery, pričom jedným výberom bol skutočný pozorovaný vektor „SP500-tbill“ (s dĺžkou 2362 pozorovaní) a druhým výberom bol simulovaný vektor z požadovaného rozdelenia (s dĺžkou 10000 pozorovaní). Nasledujúca tabuľka popisuje výsledky týchto testov¹⁷.

Hypotetické rozdelenie $r_{tm} - r_{tf}$	p-hodnota
$N(0.07, 1.62^2)$	2.72E-06
$N(0, 1)$	1.87E-10
t_3	0,0154
$t_{3.99}$	0,0122

Tabuľka 5.13: KS-test skutočného a hypotetického rozdelenia pre denné pozorovania premenných modelu.

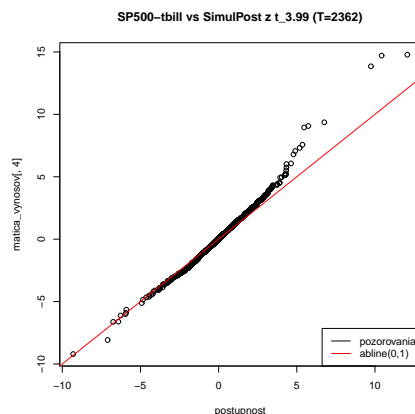
Na základe skonštruovaných testov vyšlo ako jedno z najpriaznivejších rozdelení rozdelenie t_3 a $t_{3.99}$. Keďže p-hodnoty týchto rozdelení sa líšia len nepatrne, bolo za rozdelenie vektora $r_{tm} - r_{tf}$ zvolené práve Studentovo rozdelenie s 3.99 stupňami voľnosti, ktoré bolo odhadnuté programom R-project¹⁸.

Pre $t_{3.99}$ -rozdelenie bol vytvorený tzv. QQ-plot (pozri obr. 5.1), ktorý ukazuje, či hypotetické kvantily zodpovedajú reálnym kvantilom v dátach. Hypotetické

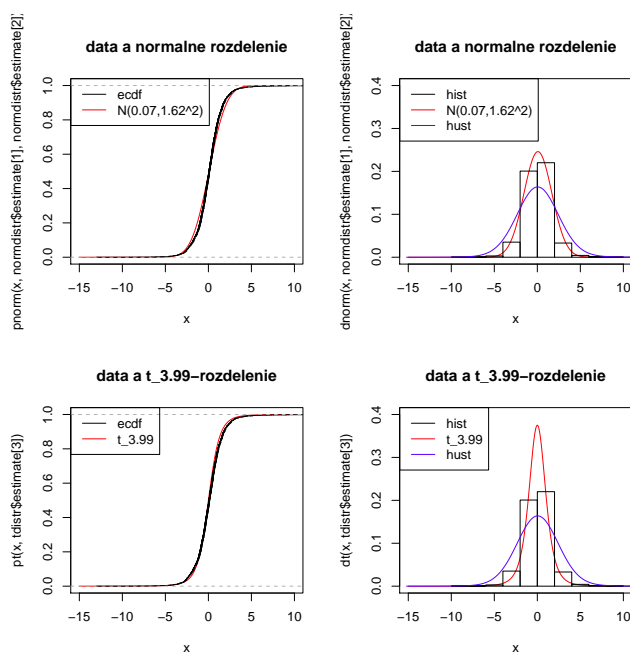
¹⁷ Rozdelenie $N(\mu, \sigma)$ a t_v boli odhadnuté z reálnych dát pomocou funkcie „fitdistr“. Rozdelenia $N(0, 1)$ a t_3 boli zvolené subjektívne.

¹⁸ Ako už bolo uvedené vyššie, stupne voľnosti v programe R-project nemusia predstavovať celé číslo.

rozdelenia vektora $r_{tm} - r_{tf}$, ktoré boli odhadované pomocou funkcie „fitted“ v programe R-project sú zobrazené na obr. 5.2 spolu s rozdeleniami odhadovanými funkciou „ecdf“ (distribučná funkcia) a „density“ (hustota, ozn. „hust“; histogram má v grafe skratku „hist“).



Obr. 5.1: QQ-plot pre denné pozorovania premenných modelu.

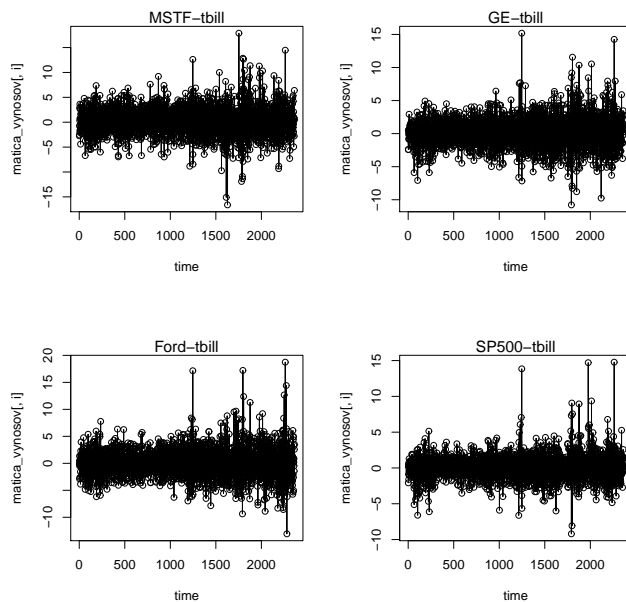


Obr. 5.2: Porovnanie rozdelení pre denné pozorovania premenných modelu.

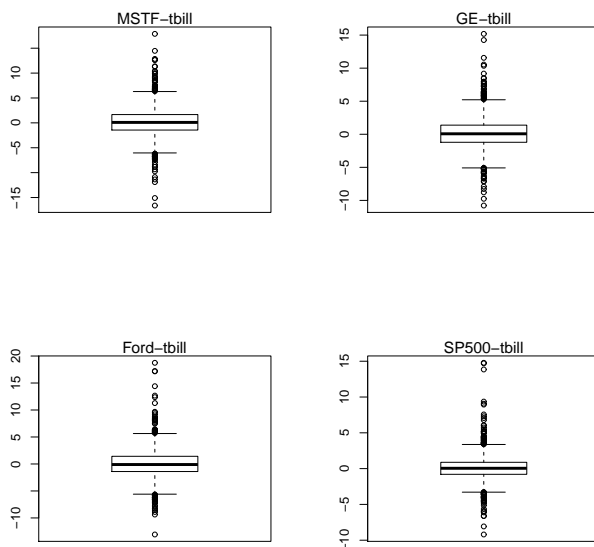
Súčasťou analýzy bolo aj grafické znázornenie jednotlivých pozorovaní:

- Obr. 5.3 ukazuje, že v časových radoch hodnôt $r_{ti} - r_{rf}$ a $r_{tm} - r_{tf}$ sa vyskytujú odľahlé pozorovania. To potvrdzuje aj obr. 5.4.
- Na obr. 5.5 sú zobrazené histogramy premenných vstupujúcich do modelu. Tieto histogramy ukazujú, že rozdelenia jednotlivých výnosov $r_{ti} - r_{tf}$ a $r_{tm} - r_{tf}$ majú ťažké konce.

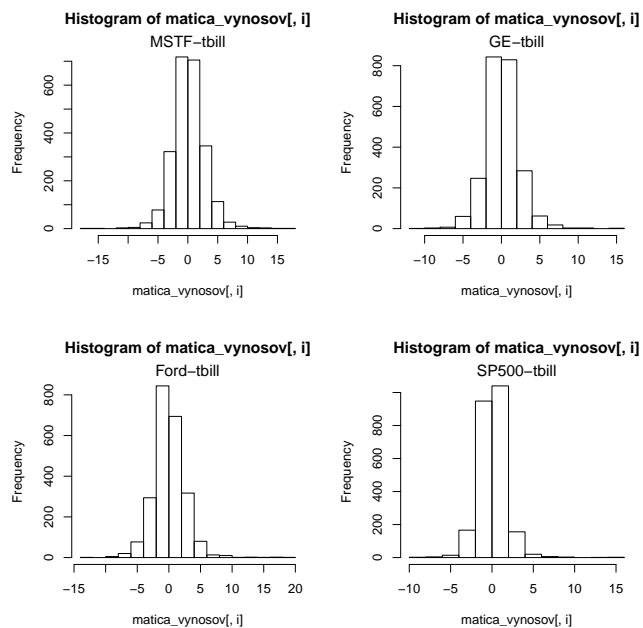
- Závislosti medzi jednotlivými premennými zobrazuje obr. 5.6, pričom na diagonále tohoto obrázku sú znázornené hustoty jednotlivých analyzovaných premenných $r_{ti} - r_{tf}$. Toto grafické zobrazenie potvrdzuje a dopĺňa tabuľku korelácií (pozri tab. 5.12).



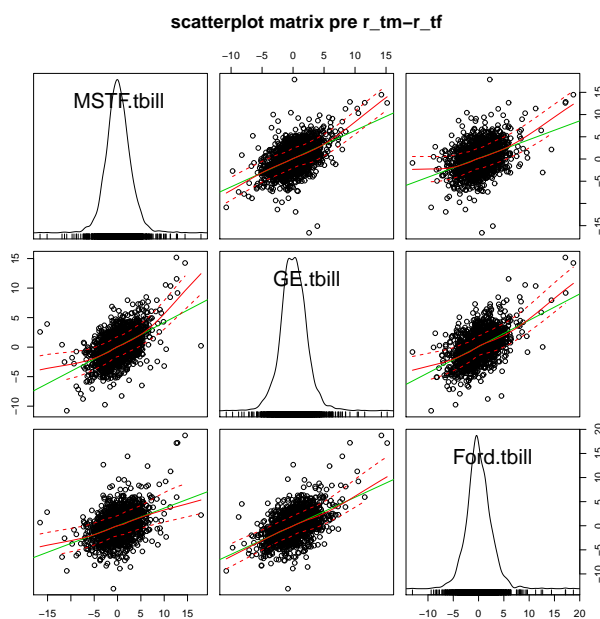
Obr. 5.3: Časové rady pre denné pozorovania premenných modelu.



Obr. 5.4: Odľahlé pozorovania pre denné pozorovania premenných modelu.



Obr. 5.5: Histogramy pre denné pozorovania premenných modelu.



Obr. 5.6: Závislosti pre denné pozorovania premenných modelu.

Odhady parametrov modelu (5.2) sú uvedené v nasledujúcej tabuľke. Tieto odhady sa pre aplikované odhady príliš neodlišujú¹⁹. Pravdepodobne je to spôsobené najmä veľkým počtom pozorovaní v dátovej vzorke a približne rovnako rozloženými odľahlými pozorovaniami v jednotlivých časových radoch, t. j. nízke/vysoké hodnoty $r_{tm} - r_{tf}$ často implikujú nízke/vysoké

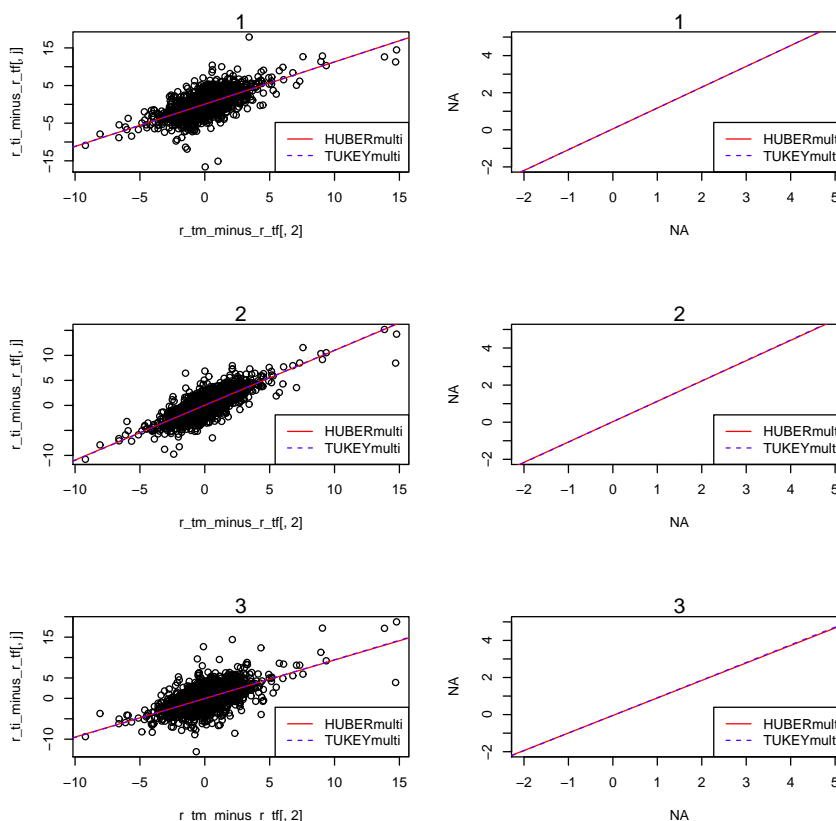
¹⁹ Odhadu „TUKEYmulti“ ale prikladáme väčšiu váhu ako odhadu „HUBERmulti“ (pozri tab. 5.8). V tab. 5.14 uvádzame Huberov simultánny odhad iba pre porovnanie.

hodnoty $r_{ti} - r_{tf}$ (čo potvrdzuje aj obr. 5.3).

	TUKEYmulti miera α	TUKEYmulti miera β		HUBERmulti miera α	HUBERmulti miera β
CP			CP		
MSTF-tbill	0.0515	1.1231	MSTF-tbill	0.0499	1.1227
GE-tbill	0.0255	1.1019	GE-tbill	0.0281	1.0960
Ford-tbill	-0.0438	0.9514	Ford-tbill	-0.0475	0.9438

Tabuľka 5.14: Odhady parametrov modelu pre denné pozorovania.

Tieto odhady zobrazuje nasledujúci graf (na ľavej strane sú uvedené hodnoty časových radov $r_{ti} - r_{tf}$ voči $r_{tm} - r_{tf}$ a lineárna funkcia odhadu podľa parametrov z vyššie uvedenej tabuľky, na pravej strane je uvedené grafické porovnanie odhadov „TUKEYmulti“ a „HUBERmulti“ v grafe s mierkou $[-2, 5] \times [-2, 5]$). V jednotlivých obrázkoch nadpis „1“ označuje „MSTF-tbill“, „2“ značí „GE-tbill“ a „3“ označuje „Ford-tbill“.



Obr. 5.7: Odhady „TUKEYmulti“ a „HUBERmulti“ pre denné pozorovania.

Ako bolo uvedené vyššie, odhady pre funkcie „TUKEYmulti“ a „HUBERmulti“ sa takmer zhodujú. Tieto rozdiely nemajú vplyv na nasledujúce závery

plynúce z veľkosti jednotlivých odhadov α_i a β_i .

Na základe teórie uvedenej v kap. 1 možno usúdiť, že výnos cenného papiera spoločnosti Microsoft a GE sa pohybuje rovnako, ako výnos trhového indexu, avšak o niečo rýchlejšie. Rizikový parameter beta cenného papiera spoločnosti Ford je niečo málo pod hodnotou 1, preto sa tento výnos pohybuje rovnakým smerom, ako výnos trhového indexu, ale o niečo pomalšie.

Čo sa týka miery nerovnováhy alfa, možno tvrdiť, že cenné papiere spoločnosti Microsoft a GE sú na trhu mierne podhodnotené (čo môže byť najmä u CP Microsoft spôsobené „menom spoločnosti“) a cenné papiere spoločnosti Ford vykazujú mierne podpriemerný výnos v porovnaní s očakávanou hodnotou.

Mesačné pozorovania výnosov

Dáta mesačných výnosov (jednotlivých CP) boli podobne ako pre denné pozorovania analyzované pomocou popisných štatistík (pozri tab. 5.15).

Charakteristika	Minimum	Priemer	Medián	Maximum	Smerodajná odchýlka	Šikmosť	Špicatosť
tbill	-35.4934	-0.9447	-0.3805	12.3379	5.9610	-2.0676	9.3802
MSFT	-42.0881	2.0476	2.0538	34.1810	11.8219	-0.2123	1.1075
SP500	-15.7586	0.5428	1.0782	9.2324	4.6512	-0.6756	0.4038
GE	-19.4471	1.1828	0.8338	17.5835	7.1327	-0.1866	-0.0832
Ford	-24.4639	-0.0463	1.2633	29.6603	9.6716	-0.1867	0.1699

Tabuľka 5.15: Popisné štatistiky pre mesačné výnosy jednotlivých CP.

Hodnoty jednotlivých výnosov už nevykazujú navzájom tak podobné štatistiky, ako tomu bolo u denných pozorovaní. Stredné hodnoty sa od nuly významne odlišujú, pre CP Microsoft je priemer a medián vyšší než 2. Navyiac tento cenný papier vykazuje vysoké absolútne hodnoty minima a maxima, jeho smerodajná odchýlka je taktiež najvyššia zo všetkých analyzovaných CP. Pre CP Ford je priemer výrazne odlišný od mediánu.

V nasledujúcej tabuľke sú uvedené korelácie medzi hodnotami výnosov jednotlivých analyzovaných CP.

	tbill	MSFT	SP500	GE	Ford
tbill	1.0000	-0.0336	0.0402	0.0375	-0.0251
MSFT	-0.0336	1.0000	0.6169	0.4572	0.2525
SP500	0.0402	0.6169	1.0000	0.7196	0.5258
GE	0.0375	0.4572	0.7196	1.0000	0.3484
Ford	-0.0251	0.2525	0.5258	0.3484	1.0000

Tabuľka 5.16: Korelácie mesačných výnosov jednotlivých CP.

Korelácie medzi výnosmi CP reprezentujúcimi r_{ti} neprekročili hranicu 0.5, teda tu podobne ako pre denné pozorovania nebola odhalená významná závislosť jednotlivých výnosov. Korelácia výnosov a bezrizikového aktíva bola zanedbateľná, keďže sa pohybovala pod hranicou ± 0.041 (pozri tab. 5.16).

Dáta boli následne analyzované aj ako jednotlivé vstupné premenné do modelu (5.2), t. j. hodnoty $r_{ti} - r_{tf}$ a $r_{tm} - r_{tf}$:

Charakteristika	Minimum	Priemer	Medián	Maximum	Smerodajná odchýlka	Šikmosť	Špicatosť
MSTF-tbill	-40.8568	2.9923	2.6057	51.0611	13.4176	0.2623	1.1701
GE-tbill	-21.1256	2.1275	1.9198	26.4764	9.1225	0.1332	-0.1747
Ford-tbill	-20.5500	0.8984	1.5615	46.4938	11.4878	0.6835	1.7229
SP500-tbill	-15.3885	1.4875	1.0144	26.9678	7.4121	0.7524	1.2763

Tabuľka 5.17: Popisné štatistiky pre mesačné pozorovania premenných modelu.

Od popisných štatistík výnosov r_{ti} sa významne odlišovali takmer všetky popisné štatistiky, najmä šikmosť a špicatosť udávajúce tvar rozdelenia. Pre hodnoty $r_{ti} - r_{tf}$ sa šikmosť vo väčšine prípadov približuje k nule a špicatosť nenadobúda väčšie hodnoty ako 2. Pre jednotlivé charakteristiky sa ale významne zmenili hodnoty maximálnych a minimálnych hodnôt a smerodajná odchýlka, ktorá sa pre všetky analyzované charakteristiky navýšila.

Korelácie medzi jednotlivými výnosmi $r_{ti} - r_{tf}$ podobne ako pre denné pozorovania vykazovali významnejšiu závislosť, než pre jednotlivé výnosy CP analyzované vyššie (r_{ti}) (pozri tab. 5.18).

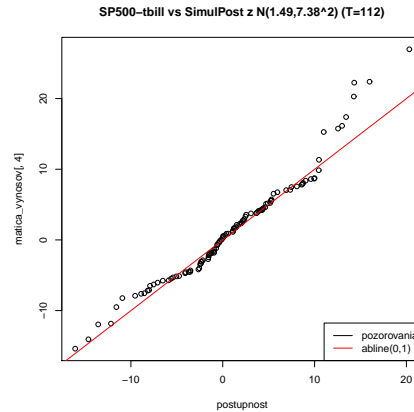
	MSTF-tbill	GE-tbill	Ford-tbill	SP500-tbill
MSTF-tbill	1.0000	0.6116	0.4426	0.7110
GE-tbill	0.6116	1.0000	0.5670	0.8385
Ford-tbill	0.4426	0.5670	1.0000	0.6990
SP500-tbill	0.7110	0.8385	0.6990	1.0000

Tabuľka 5.18: Korelácie pre mesačné pozorovania premenných modelu.

Rozdelenie, ktoré bolo pre účely ďalšej analýzy smerodajné, bolo rozdelenie vektora $r_{tm} - r_{tf}$, t. j. „SP500-tbill“. Prípustné rozdelenia pre tento vektor sa zhodujú s prípustnými rozdeleniami pre denné pozorovania. Kolmogorov-Smirnov test bol vykonaný taktiež rovnako ako pre denné pozorovania. Nasledujúca tabuľka (pozri tab. 5.19) popisuje výsledky týchto testov. Na základe skonštruovaných testov vyšlo ako najvýznamnejšie rozdelenie $N(1.49, 7.38^2)$. Toto rozdelenie bolo odhadnuté programom R-project. Pre rozdelenie $N(1.49, 7.38^2)$ bol teda skonštruovaný tzv. QQ-plot (pozri obr. 5.8).

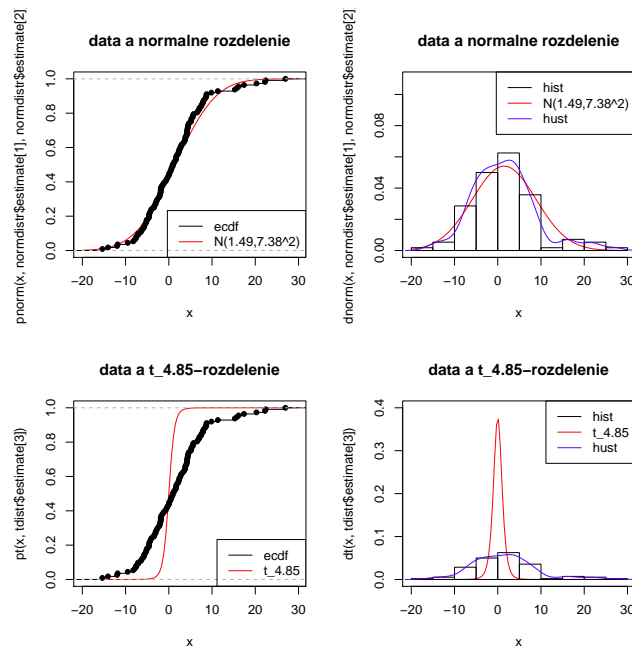
Hypotetické rozdelenie $r_{tm} - r_{tf}$	p-hodnota
$N(1.49, 7.38^2)$	0,5656
$N(0, 1)$	0
t_3	1.68E-14
$t_{4.85}$	3.33E-16

Tabuľka 5.19: KS-test skutočného a hypotetického rozdelenia pre mesačné pozorovania premenných modelu.



Obr. 5.8: QQ-plot pre mesačné pozorovania premenných modelu.

Hypotetické rozdelenia vektora $r_{tm} - r_{tf}$, ktoré boli odhadované pomocou funkcie „fitdistr“ (t. j. okrem rozdelenia $N(1.49, 7.38^2)$ aj rozdelenie $t_{4.85}$) v programe R-project sú zobrazené na obr. 5.9 spolu s rozdeleniami odhadovanými funkciou „ecdf“ a „density“.

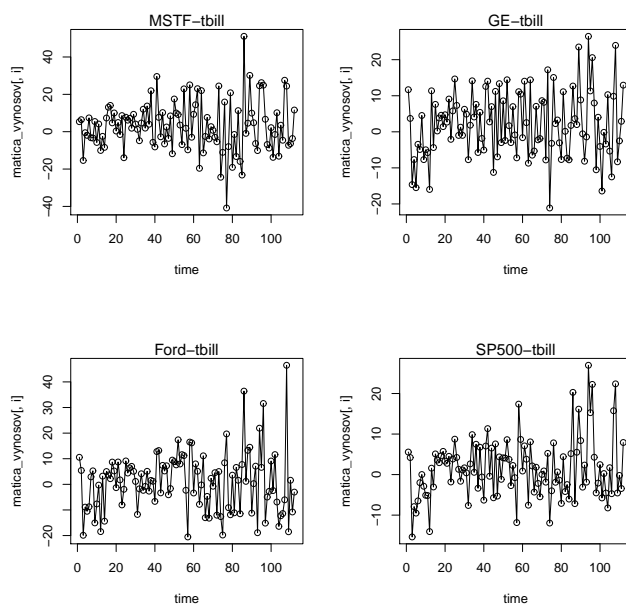


Obr. 5.9: Porovnanie rozdelení pre mesačné pozorovania premenných modelu.

Rozdelenie $N(1.49, 7.38^2)$ je aj podľa grafických znázornení veľmi dobrou aproximáciou skutočného rozdelenia vektora $r_{tm} - r_{tf}$, aj keď pre kladné hodnoty má empirické rozdelenie tohto vektora ťažký koniec²⁰.

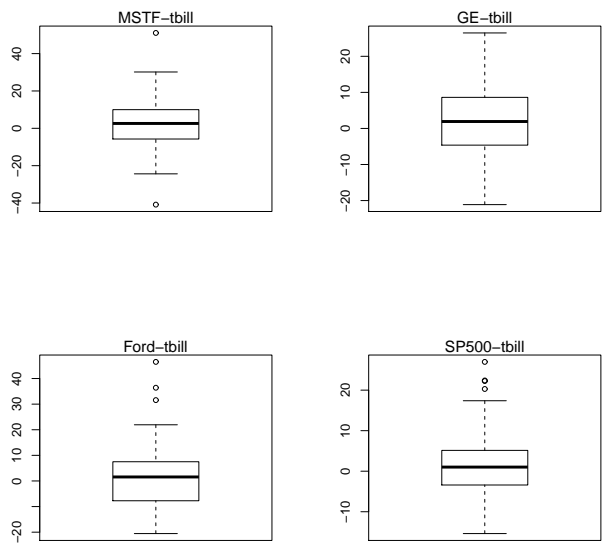
Súčasťou analýzy bolo aj grafické znázornenie jednotlivých pozorovaní:

- Obr. 5.10 ukazuje, že v časových radoch hodnôt $r_{tm} - r_{tf}$ sa vyskytuje niekoľko málo odľahlých pozorovaní. To potvrdzuje aj obr. 5.11 znázorňujúci krabicové (škatuľkové) grafy. Na prvom obrázku je vidieť, že pozorovania časového radu CP GE sa vyvíjajú veľmi podobne, ako pozorovania časového radu trhového indexu S&P 500. Narozdiel od toho, časové rady CP Microsoft a CP Ford nadobúdajú aj hodnoty dvojnásobne väčšie ako hodnoty trhového indexu. Na druhom obrázku možno potom vidieť, že CP GE nevykazuje takmer žiadne výrazné odľahlé pozorovania.
- Na obr. 5.12 sú zobrazené histogramy premenných vstupujúcich do modelu. Tieto histogramy sú na prvý pohľad veľmi nesúmerné a pre niektoré charakteristiky veľmi zošikmené (najmä histogram trhového indexu S&P 500).
- Závislosti medzi premennými zobrazuje obr. 5.13. Toto grafické zobrazenie potvrdzuje a dopĺňa tabuľku korelácií (pozri tab. 5.18). Na diagonále sú zobrazené empirické hustoty jednotlivých analyzovaných premenných vstupujúcich do modelu 5.1.

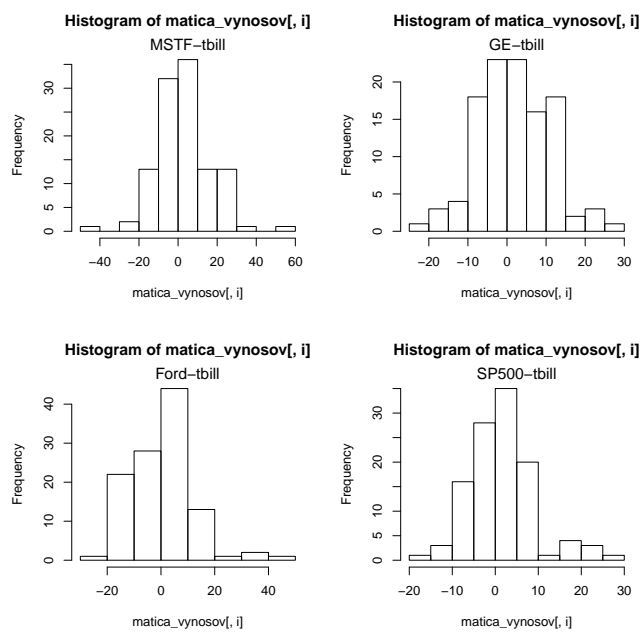


Obr. 5.10: Časové rady pre mesačné pozorovania premenných modelu.

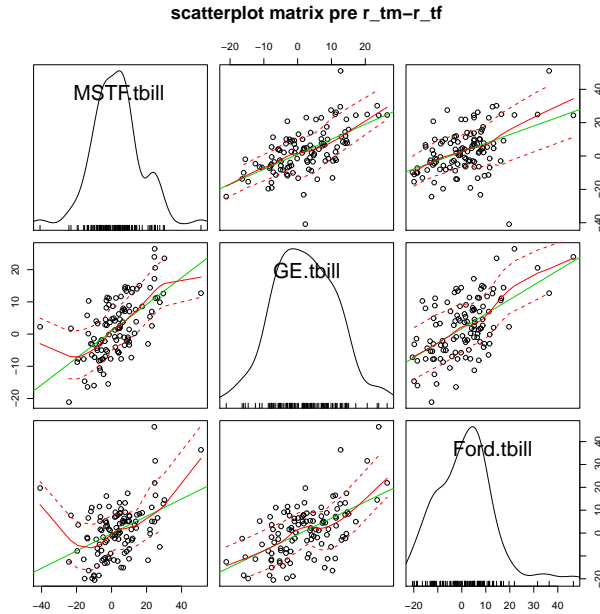
²⁰ Rozdelenie $r_{tm} - r_{tf}$ by teda bolo možné popísať kombináciou rozdelení, napr. normálneho rozdelenia $N(\mu, \sigma)$ a Studentovho t -rozdelenia s v stupňami voľnosti.



Obr. 5.11: Odľahlé pozorovania pre mesačné pozorovania premenných modelu.



Obr. 5.12: Histogramy pre mesačné pozorovania premenných modelu.



Obr. 5.13: Závislosti pre mesačné pozorovania premenných modelu.

Odhady parametrov modelu (5.2) sú uvedené v nasledujúcej tabuľke²¹. Tieto odhady sa pre aplikované odhady pre parametre rizika β_i príliš neodlišujú. Čo sa týka odhadu miery nerovnováhy α_i , jej odhady sú odlišné pre odhad „TUKEYmulti“ a „HUBERmulti“, a to najmä pre CP Microsoft. Tieto rozdiely ale neovplyvňujú nižšie uvedené závery.

CP	TUKEYmulti miera α	TUKEYmulti miera β	CP	HUBERmulti miera α	HUBERmulti miera β
MSTF-tbill	1.2382	1.2033	MSTF-tbill	1.1521	1.2266
GE-tbill	0.4944	1.0502	GE-tbill	0.4921	1.0472
Ford-tbill	-0.4817	1.0836	Ford-tbill	-0.4581	1.0866

Tabuľka 5.20: Odhady parametrov modelu pre mesačné pozorovania.

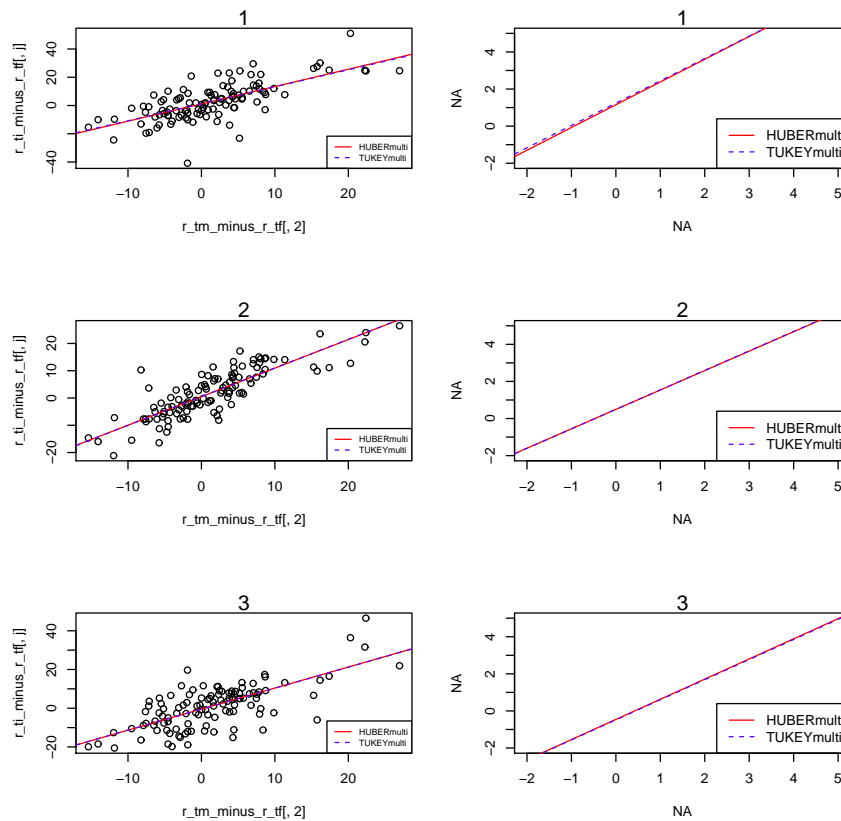
Jednotlivé odhady zobrazuje obr. 5.14²². Na základe teórie uvedenej v kap. 1 sa dá usúdiť, že výnosy všetkých cenných papierov sa pohybujú rovnako, ako výnos trhového indexu, avšak o niečo rýchlejšie. Oproti denným pozorovaniam teda nastala zmena u CP Ford.

Čo sa týka miery nerovnováhy alfa, možno tvrdiť, že cenné papiere spoločnosti Microsoft a GE sú na trhu podhodnotené, pričom CP Microsoft významnejšie

²¹ Na základe tab. 5.8 by odhadom „TUKEYmulti“ mala byť prikladaná väčšia váha.

²² Na ľavej strane sú uvedené hodnoty časových radov $r_{ti} - r_{tf}$ voči $r_{tm} - r_{tf}$ a lineárna funkcia odhadu podľa parametrov z vyššie uvedenej tabuľky, na pravej strane je uvedené grafické porovnanie odhadov „TUKEYmulti“ a „HUBERmulti“ v grafe s mierkou $[-2, 5] \times [-2, 5]$. V obrázkoch nadpis „1“ značí „MSTF-tbill“, „2“ značí „GE-tbill“ a „3“ značí „Ford-tbill“.

než CP GE (pre CP Microsoft sa naviac nedá s určitosťou na základe Tukeyho a Huberovho M -odhadu tvrdiť, ako veľmi je tento cenný papier na trhu podhodnotený), a cenné papiere spoločnosti Ford vykazujú podpriemerný výnos v porovnaní s očakávanou hodnotou.



Obr. 5.14: Odhady „TUKEYmulti“ a „L1multi“ pre mesačné pozorovania.

5.4.2 Upravené portfólio

V tejto kapitole bol v dátach nahradený cenný papier spoločnosti GE cenným papierom spoločnosti Boeing²³. K dispozícii pre CP Boeing bol časový rad pozorovaní od 03.01.2000 do 11.08.2010. Keďže sa dĺžka tohto časového radu nezhdovala s dĺžkou časových radov pre pôvodné dáta, bolo určené obdobie, v ktorom sa tieto časové rady prekrývali. Toto obdobie bolo od 03.01.2000 do 03.04.2003.

Nahradenie CP GE vychádzalo z predchádzajúcej analýzy²⁴, keďže tento CP uvádza len veľmi málo odľahlých pozorovaní, a to jednak pre denné, jednak pre

²³ Zdroj: Bloomberg.

²⁴ Rozhodnutie bolo urobené na základe smerodajných odchýliek GE a GE-tbill, ktoré boli najnižšie spomedzi všetkých analyzovaných CP, a to ako pre denné, tak aj pre mesačné pozorovania. Ďalej bola využitá aj grafická analýza krabicových (škatulkových) grafov jednotlivých pozorovaní.

mesačné pozorovania. K nahradeniu boli použité dáta o CP Boeing, ktoré zachytávajú obdobie teroristického útoku na dvojčiky (World Trade Center) v USA z 11.09.2001. V dátach je zrejmé, že sa po 10.09.2001 až do 17.09.2001 na burze neobchodovalo (pracovné dni 11.09. - 14.09.2001). Po tomto období nastali v dátach CP Boeing určité turbulencie, najmä prepád v cene, ktorý sa ale ďalej ustálil do pôvodného stavu (pozri Prílohy, aj obr. 5.17).

Časové rady pre denné pozorovania boli teda zvolené s dĺžkou $T = 811$ a mesačných pozorovaní bolo k dispozícii iba 39, preto $T = 39$. Pozorovania jednotlivých výnosov r_{ti} a r_{tm} , ako aj výnosov vstupujúcich do regresného modelu $r_{ti} - r_{tf}$ a $r_{tm} - r_{tf}$, boli pre jednotlivé denné/mesačné pozorovania analyzované jednak pomocou popisných štatistík a korelačnej analýzy, jednak pomocou grafického znázornenia. Všetky výstupné tabuľky a obrázky sú uvedené v kap. Prílohy. Finálna tabuľka odhadov a graf porovnávajúci tieto odhady sú uvedené aj priamo v texte a ich obsah je náležite okomentovaný.

Poznámka: Analýza v tejto časti nebola vykonaná s takým detailom ako v predchádzajúcej kapitole. Cieľom bolo poukázať na to, ako sú ovplyvnené odhady odľahlými pozorovaniami.

Denné pozorovania výnosov

Odhady zvolené pre denné pozorovania na základe simulačnej analýzy („TUKEYmulti“, „HUBERmulti“) boli použité aj pre tieto upravené dáta²⁵. Pre porovnanie odchýliek bol skonštruovaný aj simultánny OLS-odhad („OLSmulti“).

Odhady parametrov modelu (5.2) sú uvedené v nasledujúcej tabuľke. Jednotlivé odhady zobrazuje obr. 5.15²⁶.

CP	TUKEYmulti miera <i>alfa</i>	TUKEYmulti miera <i>beta</i>	HUBERmulti miera <i>alfa</i>	HUBERmulti miera <i>beta</i>	OLSmulti miera <i>alfa</i>	OLSmulti miera <i>beta</i>
MSTF-tbill	0.0013	1.1032	-0.0109	1.1125	-0.0547	1.1270
Boeing-tbill	0.0501	0.8341	0.0428	0.8246	0.0399	0.7839
Ford-tbill	-0.0785	0.9131	-0.0731	0.9078	-0.0689	0.9046

Tabuľka 5.21: Odhady parametrov modelu pre denné pozorovania (Boeing).

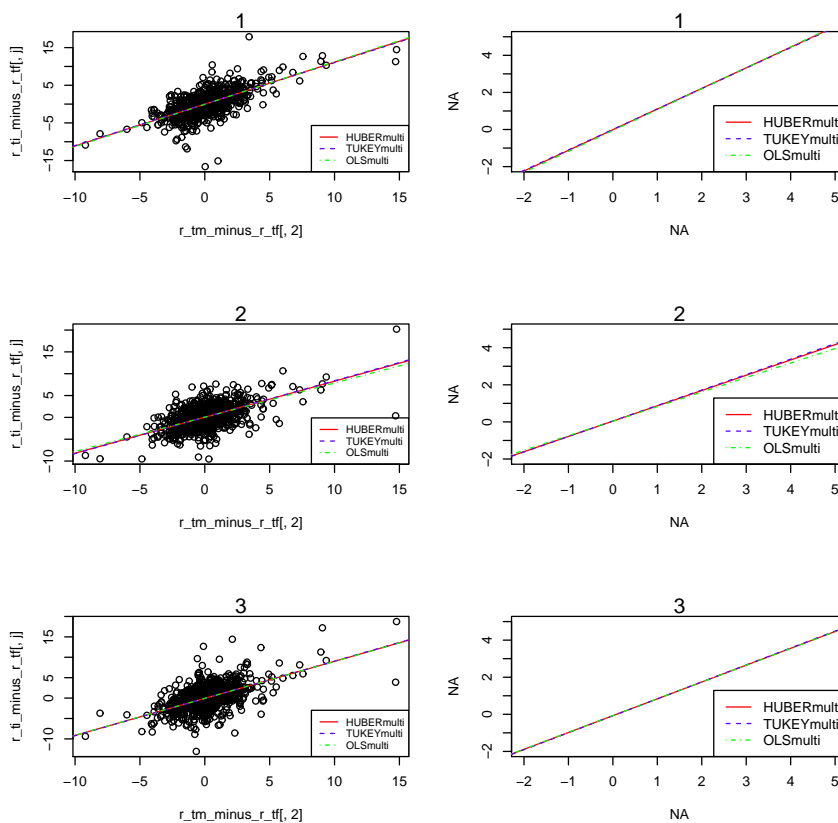
Tieto odhady sa pre aplikované odhadové funkcie iba veľmi málo odlišujú. Pravdepodobne je to spôsobené najmä veľkým počtom pozorovaní v dátovej

²⁵ Zvolené rozdelenia boli aplikované bez ohľadu na to, že tieto dáta nemali rovnakú dĺžku alebo $r_{tm} - r_{tf}$ nemalo rovnaké rozdelenie, ako bolo testované v simuláciách.

²⁶ Na ľavej strane sú uvedené hodnoty časových radov $r_{ti} - r_{tf}$ voči $r_{tm} - r_{tf}$ a lineárna funkcia odhadu podľa parametrov z vyššie uvedenej tabuľky, na pravej strane je uvedené grafické porovnanie odhadov „TUKEYmulti“ a „HUBERmulti“ v grafe s mierkou $[-2, 5] \times [-2, 5]$. V jednotlivých obrázkoch nadpis „1“ označuje „MSFT-tbill“, „2“ značí „Boeing-tbill“ a „3“ označuje „Ford-tbill“.

vzorku (zhhluk pozorovaní udáva parametre lineárnej regresnej priamky). Okrem toho možno pozorovať, že výsledky robustných odhadov sa navzájom viac približujú ako výsledky OLS-odhadu. Odhady sú významnejšie odlišné iba pre odhad miery nerovnováhy alfa cenného papiera spoločnosti Microsoft. Jeho výsledky ovplyvňujú rozhodnutie o tom, či tento CP je alebo nie je na trhu podhodnotený.

Čo sa týka miery rizika beta, jej odhad sa pre CP Boeing príliš nelíši pre robustné simultánne odhady Tukey a Huber, ale OLS-odhad vykazuje nižšie hodnoty. Tento odhad pravdepodobne priložil väčšiu váhu odľahlému pozorovaniu (pozri obr. 5.15, graf „2“ vľavo), pre ktoré neplatí, že s nárastom $r_{tm} - r_{tf}$ rastú aj hodnoty $r_{ti} - r_{tf}$ (predposledné pozorovanie napravo v tomto grafe). Takéto odľahlé pozorovania sú spravidla spôsobené výkyvmi v časovom rade $r_{ti} - r_{tf}$ jednotlivých premenných a opačnými výkyvmi v časovom rade $r_{tm} - r_{tf}$ (pozri Prílohy, obrázok zobrazujúci časové rady).



Obr. 5.15: Odhady „TUKEYmulti“, „HUBERmulti“ a „OLSmulti“ pre denné pozorovania (Boeing).

Na základe teórie uvedenej v kap. 1 sa dá usúdiť, že výnos cenného papiera spoločnosti Microsoft sa pohybuje rovnako, ako výnos trhového indexu, ale o niečo rýchlejšie. Rizikový parameter beta cenného papiera spoločnosti Boeing a Ford je málo pod hodnotou 1, preto sa tento výnos pohybuje rovnakým smerom, ako výnos trhového indexu, no o niečo pomalšie. Túto analýzu potvrdzujú všetky skonštruované odhadové funkcie.

Ďalej je možné uviesť, že CP spoločnosti Boeing sú na trhu mierne podhodnotené a cenné papiere spoločnosti Ford vykazujú mierne podpriemerný výnos v porovnaní s očakávanou hodnotou, t. j. sú nadhodnotené.

Mesačné pozorovania výnosov

Odhady zvolené pre mesačné pozorovania na základe simulačnej analýzy („TUKEYmulti“, „HUBERmulti“) boli použité aj pre tieto upravené dáta²⁷. Pre porovnanie odchýliek bol skonštruovaný aj simultánny OLS-odhad („OLSmulti“).

Odhady parametrov modelu (5.2) sú uvedené v nasledujúcej tabuľke.

CP	TUKEYmulti miera <i>alfa</i>	TUKEYmulti miera <i>beta</i>	HUBERmulti miera <i>alfa</i>	HUBERmulti miera <i>beta</i>	OLSmulti miera <i>alfa</i>	OLSmulti miera <i>beta</i>
MSTF-tbill	-0.9992	1.3855	-1.3771	1.3895	-1.7683	1.4432
Boeing-tbill	1.7860	0.4050	1.7873	0.4159	1.6411	0.4032
Ford-tbill	-2.1392	1.0869	-2.1105	1.0891	-2.0455	1.1165

Tabuľka 5.22: Odhady parametrov modelu pre mesačné pozorovania (Boeing).

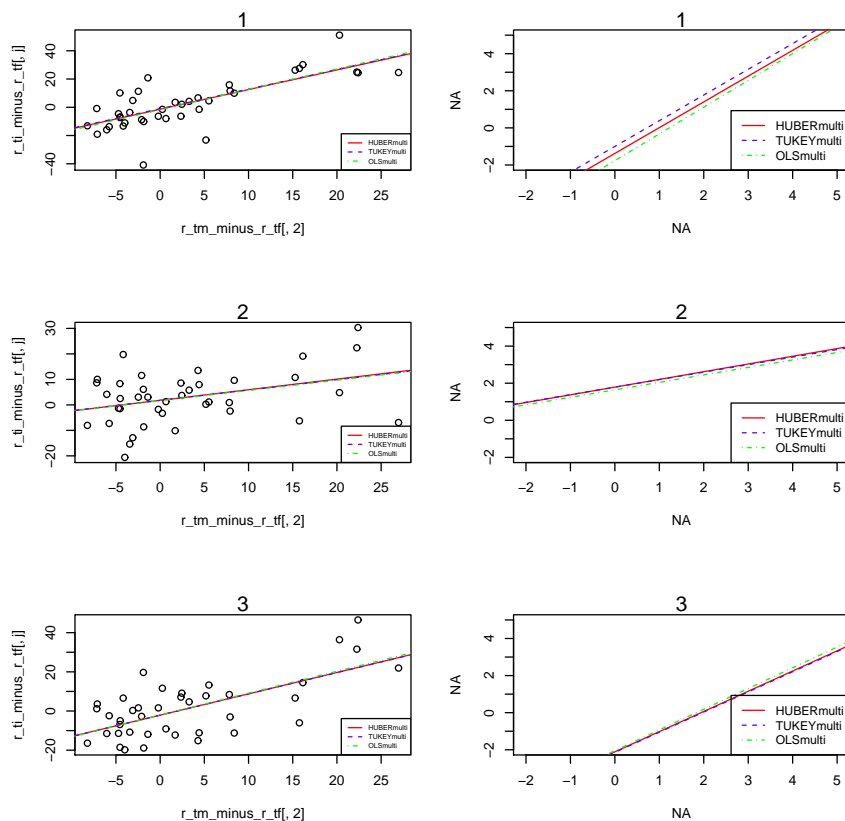
Odhady uvedené v tab. 5.22 treba brať s určitým nadhľadom, keďže v dátovej vzorke bolo iba 39 pozorovaní. Výsledky odhadov sa síce nezhodujú, ale závery z nich plynúce na základe pravidiel pre mieru rizika beta a mieru nerovnováhy alfa uvedených v kap. 1 sú pre všetky odhady zhodné²⁸. Z výsledkov vyššie uvedenej tabuľky možno usudzovať, že odhad miery rizika beta pre CP Ford je iba mierne nad hladinou 1, preto by výnos tohto CP mohol približne zodpovedať výnosu trhového indexu. CP Ford má teda priemerné systematické riziko.

Jednotlivé odhady zobrazuje obr. 5.16²⁹. Z grafickej analýzy sa dá usúdiť, že odhady parametrov sú významnejšie ovplyvnené odľahlými pozorovaniami (najmä pre CP Microsoft). Navyiac u CP Microsoft je možné pozorovať, že robustné odhady miery beta sa takmer vôbec nelíšia, avšak odlišnosti miery alfa sú značné. Grafické znázornenie tejto situácie je vidieť aj na obr. 5.16.

²⁷ Zvolené rozdelenia boli aplikované bez ohľadu na to, že tieto dáta nemali rovnakú dĺžku alebo $r_{tm} - r_{tf}$ nemalo rovnaké rozdelenie, ako bolo testované v simuláciách.

²⁸ Toto vyhodnotenie už bolo uvedené v práci niekoľkokrát, preto v tejto kapitole už nebude uvedené.

²⁹ Na ľavej strane sú uvedené hodnoty časových radov $r_{ti} - r_{tf}$ voči $r_{tm} - r_{tf}$ a lineárna funkcia odhadu podľa parametrov z vyššie uvedenej tabuľky, na pravej strane je uvedené grafické porovnanie odhadov „TUKEYmulti“ a „HUBERmulti“ v grafe s mierkou $[-2, 5] \times [-2, 5]$. V jednotlivých obrázkoch nadpis „1“ označuje „MSFT-tbill“, „2“ značí „Boeing-tbill“ a „3“ označuje „Ford-tbill“.



Obr. 5.16: Odhady „TUKEYmulti“, „L1multi“ a „OLSmulti“ pre mesačné pozorovania (Boeing).

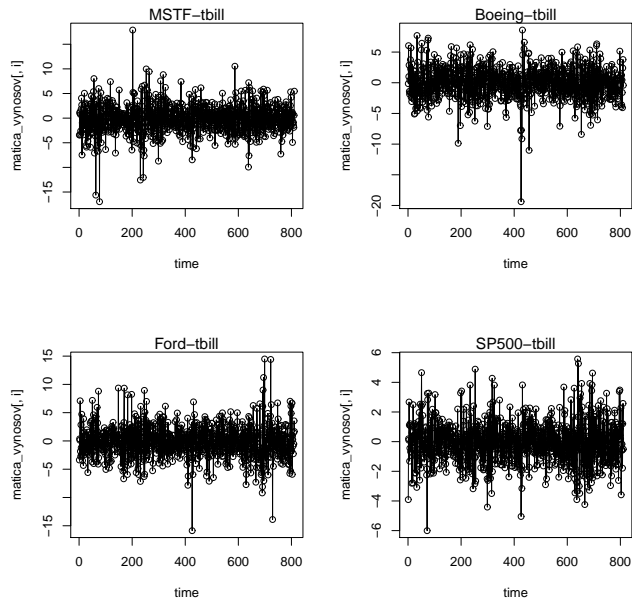
5.4.3 Zjednodušené portfólio

V tejto kapitole boli stanovené rovnaké odhady ako v predchádzajúcej kapitole (pomocou rovnakých odhadových funkcií pre denné/mesačné pozorovania), avšak s tým rozdielom, že tieto odhady boli aplikované priamo na výnosy jednotlivých CP, t. j. na hodnoty r_{tm} a r_{ti} , a nie na výnosy upravené o výnos bezrizikového CP T-bill. Predpokladá sa teda, že vektor r_{tf} je nulový.

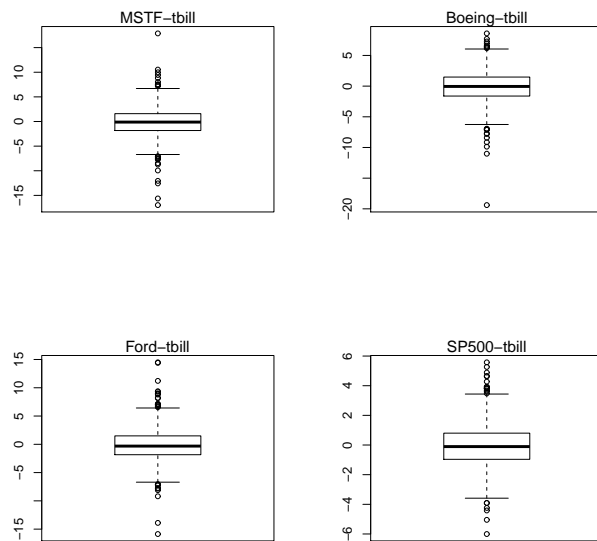
Poznámka: Analýza v tejto časti bola vykonaná ešte s menšou mierou detailu ako analýzy uvedené v predchádzajúcich kapitolách. Graficky boli analyzované jednotlivé časové rady, odľahlé pozorovania a výsledky jednotlivých odhadov. Cieľom bolo znovu len poukázať na to, ako veľmi sú ovplyvnené odhady odľahlými pozorovaniami.

Denné pozorovania výnosov

Grafická analýza časových radov uvedená na nasledujúcich obrázkoch ukazuje, že jednotlivé výnosy analyzovaných CP (v tomto konkrétnom príklade ide iba o hodnoty r_{ti} , keďže sa predpokladá, že r_{tf} je rovné nule) vykazujú niekoľko málo odľahlých pozorovaní, ktoré sa ale javia ako významné, napr. prepád výnosu CP Boeing po teroristickom útoku 11.09.2001. Dá sa predpokladať, že odhad metódou najmenších štvorcov bude skreslený práve týmito hodnotami.



Obr. 5.17: Časové rady pre denné pozorovania (Boeing, $r_{tf} = 0$).



Obr. 5.18: Odľahlé pozorovania pre denné pozorovania (Boeing, $r_{tf} = 0$).

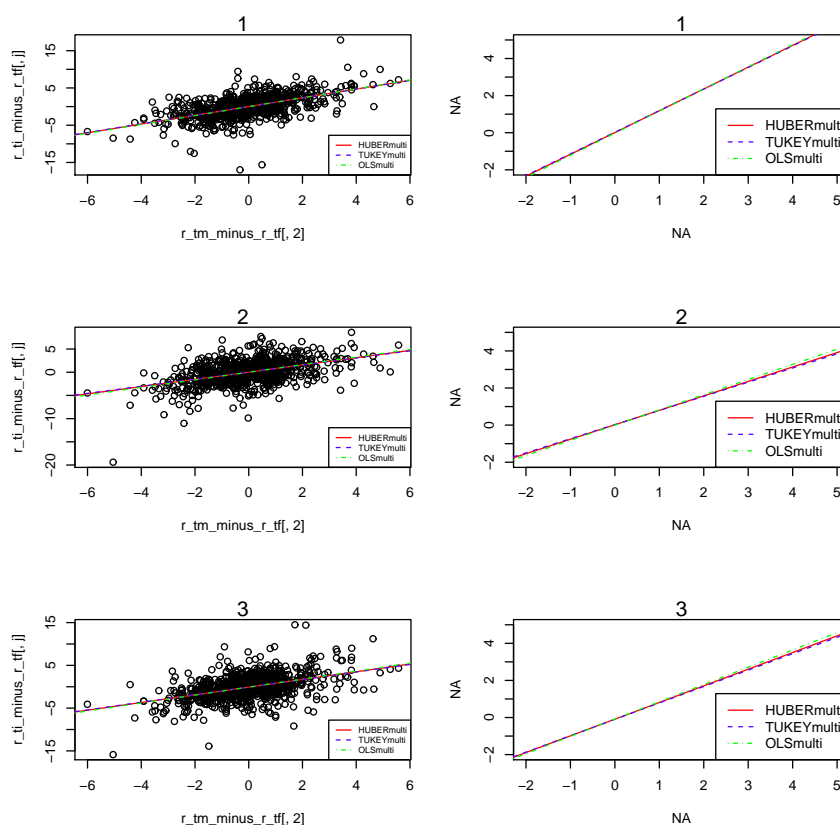
Ako ukazuje tab. 5.23, odhady parametrov modelu (5.2) sa pre aplikované odhadové funkcie navzájom takmer nelíšia. Domnievame sa, že je to spôsobené väčším počtom pozorovaní v hlavnom dátovom zhluku, ktorý udáva trend lineárnej regresnej priamky. Odhady sú významnejšie odlišné iba pre odhad miery nerovnováhy alfa CP spoločnosti Microsoft. Jeho výsledky ovplyvňujú rozhodnutie o tom, či tento CP je alebo nie je na trhu podhodnotený. Závbery plynúce

z ďalších odhadov sú pre všetky odhadové funkcie zhodné³⁰.

CP	TUKEYmulti miera <i>alfa</i>	TUKEYmulti miera <i>beta</i>	HUBERmulti miera <i>alfa</i>	HUBERmulti miera <i>beta</i>	OLSmulti miera <i>alfa</i>	OLSmulti miera <i>beta</i>
MSTF-tbill	0.0230	1.1647	0.0084	1.1748	-0.0259	1.1928
Boeing-tbill	0.0284	0.7627	0.0171	0.7791	0.0001	0.8188
Ford-tbill	-0.0986	0.8846	-0.0907	0.8987	-0.0857	0.9325

Tabuľka 5.23: Odhady parametrov modelu pre denné pozorovania (Boeing, $r_{tf} = 0$).

Odhady sú ilustrované na obr. 5.19. Je zaujímavé, že tieto odhady sa príliš neodlišujú od odhadov uvedených v predchádzajúcej kapitole pre denné pozorovania, t. j. bezrizikový CP nemá výrazný vplyv na odhady miery alfa a beta jednotlivých analyzovaných CP.

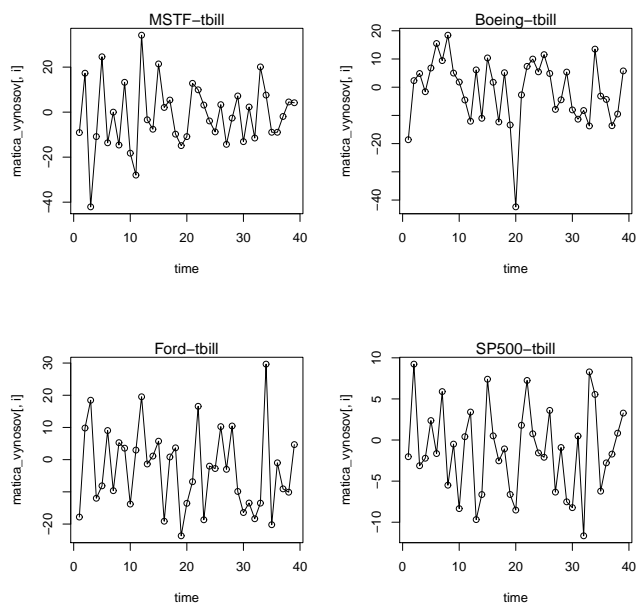


Obr. 5.19: Odhady „TUKEYmulti“, „HUBERmulti“ a „OLSmulti“ pre denné pozorovania (Boeing, $r_{tf} = 0$).

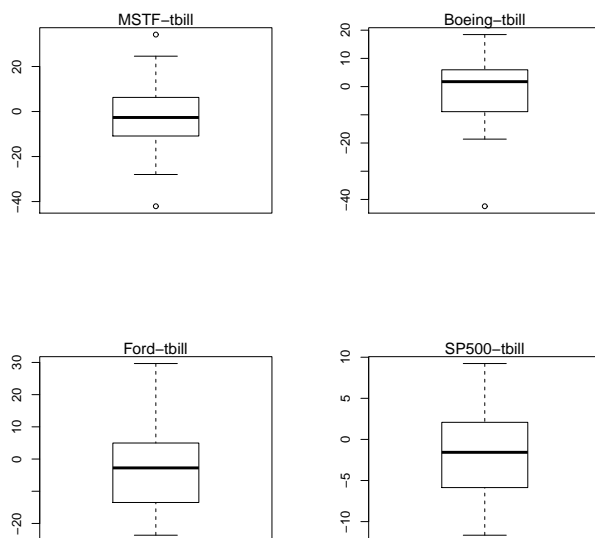
³⁰ Komentár výsledkov je analogický ako v predchádzajúcich kapitolách, preto ho na tomto mieste neuvádzam.

Mesačné pozorovania výnosov

Grafická analýza na nasledujúcich obrázkoch ukazuje, že jednotlivé časové rady výnosov vykazujú niekoľko odľahlých pozorovaní, ktoré sa javia ako významné.



Obr. 5.20: Časové rady pre mesačné pozorovania (Boeing, $r_{tf} = 0$).



Obr. 5.21: Odľahlé pozorovania pre mesačné pozorovania (Boeing, $r_{tf} = 0$).

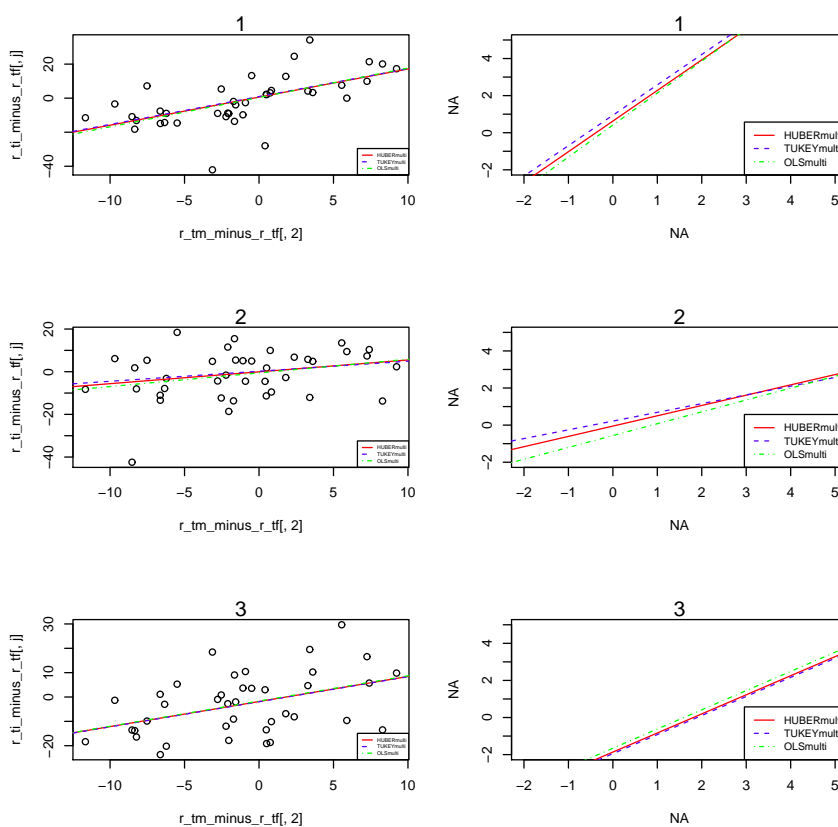
Časové rady obsahujú iba 39 pozorovaní, preto sa dá predpokladať, že odhad metódou najmenších štvorcov bude významne skreslený. Z obr. 5.20 je zřejmé,

že odľahlé pozorovania sa vyskytujú práve tam, kde sa časové rady pre CP Microsoft, Boeing či Ford pohybujú výrazne opačným smerom ako CP S&P 500. Pozorovanie v čase $t = 20$ pre CP Boeing je podľa grafickej analýzy vyhodnotené ako odľahlé, hodnota výnosu CP Boeing je menej ako -40 , pokiaľ čo hodnota vysvetľujúcej premennej S&P 500 je približne -9 .

Odhady parametrov modelu (5.2) sú uvedené v nasledujúcej tabuľke. Jednotlivé odhady ilustruje obr. 5.22.

CP	TUKEYmulti miera α	TUKEYmulti miera β	HUBERmulti miera α	HUBERmulti miera β	OLSmulti miera α	OLSmulti miera β
MSTF-tbill	0.9418	1.6382	0.6275	1.6509	0.4131	1.7252
Boeing-tbill	0.2167	0.4704	-0.0520	0.5556	-0.5633	0.6391
Ford-tbill	-1.9497	1.0287	-1.8629	1.0292	-1.6562	1.0359

Tabuľka 5.24: Odhady parametrov modelu pre mesačné pozorovania (Boeing, $r_{tf} = 0$).



Obr. 5.22: Odhady „TUKEYmulti“, „HUBERmulti“ a „OLSmulti“ pre mesačné pozorovania (Boeing, $r_{tf} = 0$).

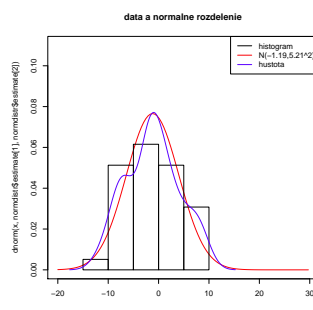
Týmto posledným príkladom bol dosiahnutý cieľ práce, a to, že bolo ukázané, ako veľmi dokážu odľahlé pozorovania a odchýlky od predpokladov OLS-odhadu ovplyvniť odhad parametrov lineárneho regresného modelu.

Odhady miery rizika beta pre CP Boeing sa veľmi významne odlišujú naprieč konštruovanými odhadovými funkciami. Pravdepodobne sa odhady pre CP Microsoft neodlišujú až tak významne (až na odhady miery nerovnováhy alfa) práve preto, že odľahlé pozorovania sú rovnomerne roztrúsené ako u menších, tak aj u väčších hodnôt grafu r_{tm} voči r_{ti} , a teda sa ich celkový vplyv na odhady modelu vzájomne ruší. Odhady miery rizika beta pre CP Ford sú relatívne stabilné. Navyiac v týchto dátach je možné pre všetky odhady pozorovať nekonzistenciu odhadu miery nerovnováhy alfa. Pre CP Boeing sa navyiac odhady miery nerovnováhy alfa naprieč jednotlivými odhadovými funkciami nezhodujú v znamienkach, čo mení rozhodnutie o podhodnotení/nadhodnotení CP. Záver teda nie je jednoznačný.

Z podrobnejšej analýzy jednotlivých odhadov uvedených v tejto a v predchádzajúcej kapitole sa dá navyše usudzovať, že odhady parametrov pre mesačné pozorovania sa až tak veľmi nelíšia pre mieru rizika beta, ale vôbec sa nezhodujú pre mieru nerovnováhy alfa. Z toho je možné usudzovať, že berzizikový CP veľmi neovplyvňuje odhad parametrov beta, ale výrazne vplýva na odhad miery nerovnováhy alfa. Preto nemožno spoľahlivo stanoviť, či sú cenné papiere na trhu nadhodnotené (vykazujú podpriemerný výnos) alebo podhodnotené (vykazujú nadpriemerný výnos).

Poznámka: Na záver bola ešte vytvorená jednoduchá simulačná analýza pre jednorozmerný lineárny model, aby bolo potvrdené rozhodnutie o výbere Tuke-yho M -odhadu. Model CAPM bol pre tieto účely uvažovaný v zjednodušenom tvare (5.3), kde sa predpokladá $n = 1$ (jednorozmerný model) a $T = 39$ (ako v poslednom príklade). Pre tento jednoduchý model bola urobená simulačná analýza, pričom v každej simulácii bolo zvolené:

- $\alpha = 0$, $\beta = 1$,
- vektor pozorovaní r_{tm} z normálneho rozdelenia $N(-1.19, 5.21^2)$, ktoré bolo pozorované na reálnych dátach výnosov CP S&P 500 pre $T = 39$,



Obr. 5.23: Rozdelenie vektora r_{tm} , $T = 39$.

- vektor ϵ zo Studentovho t -rozdelenia s tromi stupňami voľnosti.

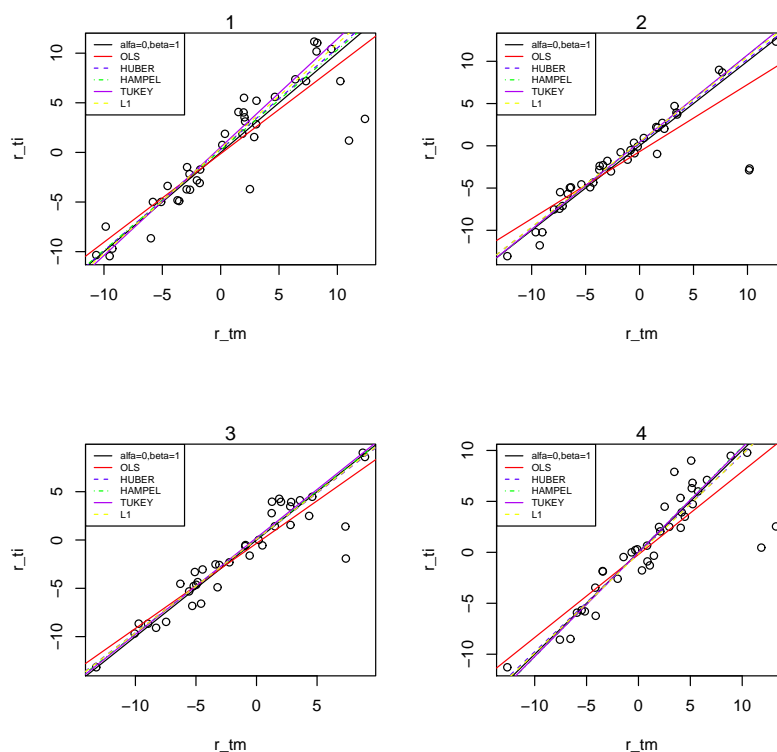
Následne boli dopočítané hodnoty vektora r_{ti} .

Za účelom dosiahnuť situáciu, ktorá často nastáva v reálnych dátach, bola nahradená hodnota mediánu r_{tm} v simulovanom vektore hodnotou mediánu navýšenou o hodnotu maxima tohto vektora (hodnoty r_{ti} už samozrejme neboli prepočítavane). Tak bolo skonštruované odľahlé pozorovanie, ktoré by mohlo skresliť odhad miery alfa a beta. Na takto skonštruované hodnoty r_{tm} a r_{ti} boli aplikované tieto odhadové funkcie (pozri tab. 5.2): OLS, HUBER, HAMPEL, TUKEY a L1.

Celý vyššie uvedený proces bol opakovaný 100-krát a následne boli spočítané priemerné absolútne odchýlky odhadov od nastavených hodnôt $\alpha = 0$ a $\beta = 1$ ($AbsSD_{alfa}$ a $AbsSD_{beta}$, pozri tab. 5.25). Pre ilustráciu bol skonštruovaný aj obrázok zobrazujúci prvé štyri simulácie.

	OLS	HUBER	HAMPEL	TUKEY	L1
$AbsSD_{alfa}$	0.5235	0.2249	0.2406	0.1885	0.2075
$AbsSD_{beta}$	0.1375	0.0524	0.0549	0.0388	0.0475
Rozhodnutie podľa alfa	1	3	2	5	4
Rozhodnutie podľa beta	1	3	2	5	4
Priemerná $AbsSD$	0.3305	0.1386	0.1478	0.1137	0.1275
Rozhodnutie	1	3	2	5	4

Tabuľka 5.25: Porovnanie odhadov v zjednodušenom simulačnom modeli.



Obr. 5.24: Porovnanie simulácií v zjednodušenom modeli.

Z vyššie uvedenej tab. 5.25 a obr. 5.24 je zrejmé, že robustné odhady umelému odľahlému pozorovaniu nepriložili veľkú váhu (odhady sa takmer trafili na hodnoty $\alpha = 0$ a $\beta = 1$), avšak OLS-odhad toto pozorovanie zohľadnil (odhady sú viac ako dvojnásobne skreslené oproti robustným odhadom). Je dôležité podotknúť, že z robustných odhadov v teste najlepšie obstál Tukeyho M -odhad s najvyšším skóre.

Záver

Účelom práce bolo aplikovať teóriu robustných odhadov na mnohorozmerný CAPM model. Za týmto účelom boli použité M -odhady. Potrebná teória je uvedená v prvej polovici práce. V druhej polovici práce je uvedený simulačný experiment a aplikácia jeho výsledkov na reálne portfóliá dát.

V simulačnej časti práce bolo urobené rozhodnutie o výbere vhodnej odhadovej funkcie na reálne pozorovania z dátového vzorku. Zo simulácií sa ako pre denné, tak aj pre mesačné pozorovania javil ako najlepší simultánny Tukeyho M -odhad. Tento odhad bol teda aplikovaný na reálne dáta. Ako ďalší odhad bol pre porovnanie skonštruovaný aj Huberov simultánny M -odhad.

Na základe odhadových funkcií stanovených v simulačnej štúdií pre pôvodnú dátovú vzorku (pozorovania záverečných cien CP Microsoft, GE, Ford, S&P 500 a T-bill, počet pozorovaní cien bol 2363) boli urobené približne rovnaké závery pre dátové vzorky denných a mesačných pozorovaní výnosov (výnosy boli skonštruované z cien pomocou logaritmickéj transformácie, bolo vytvorených 2362 pozorovaní denných výnosov a 112 pozorovaní mesačných výnosov jednotlivých CP):

- Pri denných pozorovaniach bolo ukázané, že výnosy CP Microsoft a CP GE sa pohybujú rovnakým smerom ako trhoví index, ale o niečo rýchlejšie. Výnosy CP Ford podľa analýzy vykazujú pomalší rast, avšak stále v smere trhového výnosu. Dokonca by bolo možné podľa odhadov miery rizika CP GE a CP Ford tvrdiť, že ich systematické riziko je priemerné, lebo ich hodnoty iba mierne kolíšu okolo 1.

Z odhadov bolo ďalej odvodené, že CP Microsoft a CP GE sú na trhu mierne podhodnotené (miera alfa je väčšia než nula) a CP Ford zas mierne nadhodnotený (miera alfa je menšia než nula), takže naopak vykazuje podpriemerný výnos v porovnaní s očakávanou hodnotou. Všetky odhady miery alfa ale významne kolíšu okolo nulovej hodnoty, čo by indikovalo, že tieto CP sú na trhu ocenené správne a nie nadhodnotené/podhodnotené.

- Pri mesačných pozorovaniach bolo ukázané, že čo sa týka záverov pre denné pozorovania, sú rovnaké aj pre mesačné pozorovania, akurát s tým rozdielom, že odhady miery rizika beta sa pre CP GE a CP Ford viac priblížili k 1, teda bolo potvrdené priemerné systematické riziko.

Miera nerovnováhy sa pre jednotlivé CP značne diverzifikovala a potvrdila „mierne“ výsledky analýzy denných pozorovaní.

Ďalej bola v práci dátová vzorka upravovaná tak, aby bola dosiahnutá čo najväčšia miera odľahlých pozorovaní. Bola uvažovaná nová dátová vzorka (pozorovania cien cenných papierov Microsoft, Boeing, Ford, S&P 500 a T-bill, počet pozorovaní cien bol 812), v ktorej boli nahradené pozorovania CP GE pozorovaniami CP Boeing z obdobia teroristického útoku na World Trade Center v USA dňa 11.09.2001. Tieto pozorovania vykazovali dostatok odľahlých pozorovaní a pre ďalšiu analýzu boli vyhodnotené ako vhodné. Z dátovej vzorky boli opäť skonštruované medzidenné (počet pozorovaní 811) a medzimesačné výnosy (počet pozorovaní 39).

- Pre denné pozorovania boli potvrdené predchádzajúce výsledky pre CP Microsoft a CP Ford. CP Boeing bol vyhodnotený ako CP s menším výnosom ako vykazuje trhovú index a podľa odhadov miery nerovnováhy alfa sa zdá, že je na trhu ocenený správne. Navyše bolo v tejto analýze ukázané, že robustné odhady poskytujú navzájom približne rovnaké výsledky a OLS-odhad sa od nich líši (odlišuje sa iba mierne, najmä z hľadiska veľkého počtu pozorovaní v hlavnej dátovej vzorke, ktorá určuje parametre lineárnej regresnej priamky).
- Pre mesačné pozorovania nemožno tvrdiť, že došlo k zhode výsledkov s predchádzajúcou analýzou. Miera nerovnováhy sa pre CP Microsoft a CP Ford značne zmenila (u CP Microsoft sa jej výška v okolí hodnoty 1 síce nezmenila, avšak jej znamienko, udávajúce podhodnotenie/nadhodnotenie výnosov, je odlišné, čo teda výrazne ovplyvňuje závery analýzy), ale miera rizika beta sa hlavne pre CP Ford priblížila predchádzajúcim hodnotám. Pre CP Boeing bolo pozorované, že jeho výnos sa pohybuje o niečo pomalšie ako výnos trhového indexu (rovnakým smerom) a tento CP je na trhu výrazne podhodnotený, takže možno usudzovať, že vykazuje nadpriemerný výnos v porovnaní s rovnovážnou hodnotou danou SML priamkou.

Posledná dátová vzorka bola skonštruovaná tak, že bol uvažovaný nulový vektor pozorovaní T-bill, t. j. v modeli nebol uvažovaný bezrizikový cenný papier. Vzorky denných a mesačných pozorovaní ďalších cenných papierov zostali nezmenené oproti druhej dátovej vzorke.

- Pri denných aj mesačných pozorovaní sa viac-menej potvrdili výsledky analýzy na upravenej dátovej vzorke (s CP Boeing miesto CP GE), najmä čo sa týka odhadov miery rizika beta jednotlivých analyzovaných CP. Vplyv úpravy premenných CAPM modelu o bezrizikový výnos teda nemal veľký vplyv na odhady miery alfa a beta.

Zoznam použitej literatúry

- [1] Anděl, J.: *Základy matematické statistiky*. 1. vydanie. Praha: Matfyzpress, 2005. ISBN 80-86732-40-1.
- [2] Beaulieu, M.-C.; Dufour, J.-M.; Khalaf, L.: *Identification-robust estimation and testing of the zero-beta CAPM*. Série Scientifique. Montréal: Février, 2011. ISSN 1198-8177.
- [3] Birkes, D.; Dodge, Y.: *Alternative methods of regression*. 1. vydanie. New York: John Wiley & Sons, 1993. ISBN 0-471-56881-3.
- [4] Brewer, J. W.: *Kronecker products and matrix calculus in system theory*. IEEE Transactions on circuits and systems, 1978. 772 - 781. ISSN 0098-4094.
- [5] Cipra, T.: *Finanční ekonometrie*. 1. vydanie. Praha: Ekopress, 2008. ISBN 978-80-86929-43-9.
- [6] Cipra, T.: *Finanční matematika v praxi*. 1. vydanie. Praha: HZ, 1993. ISBN 80-901495-1-0.
- [7] Cipra, T.: *Matematika cenných papírů*. 1. vydanie. Praha: HZ, 2000. ISBN 80-86009-35-1.
- [8] Cipra, T.: *Praktický průvodce finanční a pojistnou matematikou*. 2. vydanie. Praha. Ekopress, 2005. ISBN 80-86119-91-2.
- [9] Hoaglin, D. C.; Mosteller, F.; Tukey, J. W.: *Understanding robust and exploratory data analysis*. New York: John Wiley & Sons, 2000. ISBN 978-0-471-38491-5.
- [10] Huber, P. J.: *Robust statistics*. New York: John Wiley & Sons, 1981. ISBN 0-471-41805-6.
- [11] Chan, L. K. C.; Lakonishok, J.: *Robust measurement of beta risk*. Journal of financial and quantitative analysis, vol. 27, no 2, 1992. 265 - 282.
- [12] Christensen, R.: *Advanced linear modeling*. 2. vydanie. New York: Springer - Verlag, 2001. ISBN 0-387-95296-9.
- [13] Jurečková, J.: *Robustní statistické metody*. 1. vydanie. Praha: Karolinum, 2001. ISBN 80-246-0259-8.
- [14] Jurečková, J.; Picek, J.: *Robust statistical methods with R*. 1. vydanie. London: Taylor & Francis Group, 2006. ISBN 978-1-58488-454-5.

- [15] Jurečková, J.; Sen, P. K.: *Robust statistical procedures: Asymptotics and interrelations*. 1. vydanie. New York: John Wiley & Sons, 1996. Str. 215–224. ISBN 0-471-82221-3.
- [16] Koenker, R.; Portnoy, S.: *M estimation of multivariate regressions*. Journal of the American statistical association, vol. 85, no. 412, 1990. 1060 - 1068.
- [17] Maimon, O.; Rokach, L.: *Data mining and knowledge discovery handbook*. 2. vydanie. New York: Springer, 2010. ISBN 978-0-387-09822-7. 117-130.
- [18] Maronna, R. A.; Martin, R. D.; Yohai, V. J.: *Robust statistics: Theory and methods*. 1. vydanie. New York: John Wiley & Sons, 2006. ISBN 0-470-01092-4.
- [19] R Development Core Team: *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2012. ISBN 3-900051-07-0. Dostupné z www: www.R-project.org
- [20] Rousseeuw, P. J.; Leroy, A. M.: *Robust regression and outlier detection*. 1. vydanie. New York: John Wiley & Sons, 1987. ISBN 0-471-85233-3.
- [21] Wong, W.; Bian, G.: *Robust estimation in capital asset pricing model*. Journal of applied mathematics & Decision sciences, 4, 2000. 65 - 82.
- [22] www.cran.r-project.org [online]. 21.2.2005 [cit. 24.5.2012]. Ricci, V.: *Fitting Distributions with R*. Dostupné z www: www.cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf
- [23] www.dst.unive.it [online]. 28.10.2005 [cit. 23.4.2012]. Bellio, R.; Ventura, L.: *An Introduction to Robust Estimation with R functions*, University of Udine and University of Padova. Dostupné z www: www.dst.unive.it/rsr/BenVelTutorial.pdf
- [24] www.snee.org [online]. 17.10.2008 [cit. 19.8.2011]. Zhang, J.; Wihlborg, C.: *Unconditional and Conditional CAPM: Evidence from European Emerging Markets*. Dostupné z www: www.snee.org/filer/papers/266.pdf
- [25] www.stats.ox.ac.uk [online]. 16.11.2004 [cit. 18.8.2011]. Ripley, B. D.: *Robust statistics*. Dostupné z www: <http://www.stats.ox.ac.uk/pub/StatMeth/Robust.pdf>
- [26] Zichová, J.: *Plánování experimentů a predikční vícerozměrová analýza*. 1. vydanie. Praha: Karolinum, 2007. ISBN 978-80-246-1407-6.

Zoznam obrázkov

1.1	CML priamka.	7
1.2	Odvodenie SML vzťahu.	8
1.3	SML priamka - kovariancia σ_{iM}	9
1.4	SML priamka - miera beta.	10
1.5	Miera beta.	11
1.6	Miera alfa.	12
5.1	QQ-plot pre denné pozorovania premenných modelu.	58
5.2	Porovnanie rozdelení pre denné pozorovania premenných modelu.	58
5.3	Časové rady pre denné pozorovania premenných modelu.	59
5.4	Odláhlé pozorovania pre denné pozorovania premenných modelu.	59
5.5	Histogramy pre denné pozorovania premenných modelu.	60
5.6	Závislosti pre denné pozorovania premenných modelu.	60
5.7	Odhady pre denné pozorovania.	61
5.8	QQ-plot pre mesačné pozorovania premenných modelu.	64
5.9	Porovnanie rozdelení pre mesačné pozorovania premenných modelu.	64
5.10	Časové rady pre mesačné pozorovania premenných modelu.	65
5.11	Odláhlé pozorovania pre mesačné pozorovania premenných modelu.	66
5.12	Histogramy pre mesačné pozorovania premenných modelu.	66
5.13	Závislosti pre mesačné pozorovania premenných modelu.	67
5.14	Odhady pre mesačné pozorovania.	68
5.15	Odhady pre denné pozorovania (Boeing).	70
5.16	Odhady pre mesačné pozorovania (Boeing).	72
5.17	Časové rady pre denné pozorovania (Boeing, $r_{tf} = 0$).	73
5.18	Odláhlé pozorovania pre denné pozorovania (Boeing, $r_{tf} = 0$).	73
5.19	Odhady pre denné pozorovania (Boeing, $r_{tf} = 0$).	74
5.20	Časové rady pre mesačné pozorovania (Boeing, $r_{tf} = 0$).	75
5.21	Odláhlé pozorovania pre mesačné pozorovania (Boeing, $r_{tf} = 0$).	75
5.22	Odhady pre mesačné pozorovania (Boeing, $r_{tf} = 0$).	76
5.23	Rozdelenie vektora r_{tm} , $T = 39$	77
5.24	Porovnanie simulácií v zjednodušenom modeli.	78

Zoznam tabuliek

1.1	Vývoj výnosu analyzovaného aktíva na základe koeficientu beta.	11
1.2	Vývoj výnosu analyzovaného aktíva na základe koeficientu alfa.	12
4.1	Najpoužívanejšie M -odhady.	37
4.2	Najpoužívanejšie L -odhady.	41
5.1	Simulované rozdelenia pre r_{tm} a ϵ	49
5.2	Funkcie pre odhadovanie parametrov.	50
5.3	Ukážka výstupnej tabuľky pre simulácie modelu.	51
5.4	Porovnanie odhadov na simulovaných denných pozorovaniach.	52
5.5	Porovnanie odhadov na simulovaných mesačných pozorovaniach.	52
5.6	Ukážka výstupnej tabuľky pre MSE odchýlky.	54
5.7	Porovnanie odhadov na základe MSE.	54
5.8	Odhady aplikované na reálne dáta.	55
5.9	Popisné štatistiky pre denné výnosy jednotlivých CP.	56
5.10	Korelácie denných výnosov jednotlivých CP.	56
5.11	Popisné štatistiky pre denné pozorovania premenných modelu.	56
5.12	Korelácie pre denné pozorovania premenných modelu.	57
5.13	KS-test pre denné pozorovania premenných modelu.	57
5.14	Odhady parametrov modelu pre denné pozorovania.	61
5.15	Popisné štatistiky pre mesačné výnosy jednotlivých CP.	62
5.16	Korelácie mesačných výnosov jednotlivých CP.	62
5.17	Popisné štatistiky pre mesačné pozorovania premenných modelu.	63
5.18	Korelácie pre mesačné pozorovania premenných modelu.	63
5.19	KS-test pre mesačné pozorovania premenných modelu.	64
5.20	Odhady parametrov modelu pre mesačné pozorovania.	67
5.21	Odhady parametrov modelu pre denné pozorovania (Boeing).	69
5.22	Odhady parametrov modelu pre mesačné pozorovania (Boeing).	71
5.23	Odhady parametrov pre denné pozorovania (Boeing, $r_{tf} = 0$).	74
5.24	Odhady parametrov pre mesačné pozorovania (Boeing, $r_{tf} = 0$).	76
5.25	Porovnanie odhadov v zjednodušenom simulačnom modeli.	78

Zoznam použitých skratiek

Skratka	Popis
α	miera nerovnováhy alfa
β	miera rizika beta
CAPM model	Capital Asset Pricing Model
CLV	Centrálna limitná veta
CML	Capital Market Line (priamka kapitálového trhu)
CP	Cenný papier
<i>GARCH</i>	Generalized Autoregressive Conditionally Heteroskedastic proces
GLS-odhad	Generalized Least Squares
L -odhad	Linear Combinations of Order Statistics
LAD-odhad	Least Absolute Deviation (L_1 -odhad)
KS-test	Kolmogorov-Smirnov dvojvýberový test
M -odhad	Maximum Likelihood Type Estimate
OLS-odhad	Ordinary Least Squares
$O_p(1)$	ohraničenosť v pravdepodobnosti (pozri kap. 4.2.3)
R -odhad	Estimate Derived from Rank Test
S_e	reziduálny súčet štvorcov
SML	Security Market Line (priamka trhu cenných papierov)

Prílohy

V nasledujúcich kapitolách su uvedené prílohy k diplomovej práci. Všetky skripty vytvorené v softvéri R-project, použité zdrojové údaje, výstupné tabuľky a obrázky sú uložené na priloženom CD.

Príloha č. 1

Táto príloha diplomovej práce sa týka výstupných tabuliek simulácií k denným, resp. mesačným pozorovaniám, kde sa pre model (5.3) predpokladá $T = 2362$, resp. $T = 112$. V ľavom hornom rohu každej tabuľky je vždy uvedená kombinácia rozdelenia r_{tm} a ϵ (pre podrobnosti pozri tab. 5.1). V jednotlivých tabuľkách sú uvedené nasledujúce hodnoty:

- $alfa_i$, resp. $beta_i$... priemerné hodnoty odhadov parametrov za jednotlivé simulácie v danom modeli (5.3) (pre podrobnosti k použitým odhadovým funkciám pozri tab. 5.2,
- $AbsSD_{beta_i}$... stredná hodnota absolútnej odchýlky skutočnej simulovanej hodnoty β_i a jej odhadu,
- Relatívna efciencia $alfa_i$, resp. $beta_i$... priemer relatívnych efciencií odhadov $alfa_i$ resp. $beta_i$,
- Priemerná relatívna efciencia ... v prípade skúmania $AbsSD_{beta_i}$ ide o priemer relatívnych efciencií $beta_i$ (prvých 18 tabuliek), v prípade skúmania MSE ide o priemer relatívnej efciencie všetkých odhadov (posledných 6 tabuliek),
- Priemerná relatívna efciencia $alfa$... priemer relatívnej efciencie odhadov $alfa_i$ v prípade skúmania MSE (posledných 6 tabuliek),
- Priemerná relatívna efciencia $beta$... priemer relatívnej efciencie odhadov $beta_i$ v prípade skúmania MSE (posledných 6 tabuliek),
- Skóre ... skóre pridelené odhadu na základe hodnoty priemernej relatívnej efciencie,
- Skóre $beta$... skóre pridelené odhadu na základe hodnoty priemernej relatívnej efciencie $beta$.

Výstupné tabuľky sú uvedené v zložke s názvom „Prílohy“ na priloženom CD. V priečinku „Príloha 1“ sa nachádzajú dokumenty „Príloha_c_1.xlsx“ a „Príloha_c_1.xls“. Na jednotlivých listoch dokumentov sú uvedené:

- list č. 1 s názvom „dni AbsSD“ obsahuje výstupné tabuľky pre denné pozorovania a porovnanie priemernej relatívnej efektivity odhadov na základe priemerných absolútnych odchýliek odhadov,
- list č. 2 s názvom „dni MSE“ obsahuje výstupné tabuľky pre denné pozorovania a porovnanie priemernej relatívnej efektivity odhadov na základe priemerných kvadratických odchýliek odhadov,
- list č. 3 s názvom „mesiace AbsSD“ obsahuje výstupné tabuľky pre mesačné pozorovania a porovnanie priemernej relatívnej efektivity odhadov na základe priemerných absolútnych odchýliek odhadov,
- list č. 4 s názvom „mesiace MSE“ obsahuje výstupné tabuľky pre mesačné pozorovania a porovnanie priemernej relatívnej efektivity odhadov na základe priemerných kvadratických odchýliek odhadov.

Príloha č. 2

Táto príloha obsahuje odkazy na výstupné tabuľky a obrázky pre dátovú vzorku denných/mesačných pozorovaní, v ktorej bol nahradený cenný papier spoločnosti GE cenným papierom spoločnosti Boeing (príloha ku kap. 5.4.2).

Výstupné tabuľky a obrázky sú uvedené v zložke s názvom „Prilohy“ na príložnom CD, v priečinku „Priloha 2“. Koncovky názvov jednotlivých súborov udávajú, či sú tieto obrázky konštruované pre denné alebo pre mesačné pozorovania (pre denné pozorovania je to koncovka „-dni-boeing“, pre mesačné pozorovania „-mes-boeing“).

Obrázky a tabuľky boli konštruované rovnako ako pre účely kap. 5.4.1, preto je z ich názvov možné jednoducho odvodiť ich obsah. Príloha obsahuje pre denné a mesačné pozorovania premenných vstupujúcich do modelu CAPM tieto výstupy:

- tabuľku popisných štatistík,
- korelačnú maticu,
- tabuľku KS-testov,
- graf časových radov,
- krabicové (škatuľkové) grafy,
- histogramy,
- porovnanie odhadnutých rozdelení pre $r_{tm} - r_{tf}$ a
- QQ-plot pre zvolené odhadnuté rozdelenie.