Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Radka Sabolová

Některé testy dobré shody

Katedra pravděpodobnosti a matematické statistiky
Vedoucí diplomové práce: Prof. RNDr. Marie Hušková, DrSc.
Studijní program: Matematika
Studijní obor: Pravděpodobnost, matematická statistika a ekonometrie
Studijní plán: Matematická statistika

2009

Ďakujem vedúcej svojej práce Prof. RNDr. Marii Huškovej, DrSc. za poskytnuté konzultácie, zapožičanie odbornej literatúry a cenné podnety a pripomienky, ktoré mi poskytla pri písaní tejto práce.

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

V Praze dne 7.8.2009 Radka Sabolová

Obsah

Názov práce: Některé testy dobré shody
Autor: Radka Sabolová
Katedra (ústav): Katedra pravděpodobnosti a matematické statistiky
Vedúci diplomovej práce: Prof. RNDr. Marie Hušková, DrSc.
e-mail vedúceho: huskova@karlin.mff.cuni.cz

Abstrakt: Predložená práca sa zaoberá testom dobrej zhody v lineárnom regresnom modeli $\mathbf{Y_n} = \mathbf{X_n}\boldsymbol{\beta} + c\mathbf{e_n}$, kde nás zaujíma distribučná funkcia náhodnej zložky $\mathbf{e_n}$. Využijeme jednoznačnosť vzťahu medzi distribučnou a charakteristickou funkciou a testovú štatistiku založíme na váženej vzdialenosti hypotetickej charakteristickej funkcie a empirickej charakteristickej funkcie reziduí. Časť práce sa venuje aj odhadom parametrov $\boldsymbol{\beta}$ a c a ich asymptotickým vlastnostiam, keďže aj ich voľba ovplyvňuje limitné rozdelenie testovej štatistiky. Ďalej sa zaoberáme asymptotickými vlastnosťami testovej štatistiky za platnosti hypotézy, lokálnej alternatívy a pevnej alternatívy. Práca je doplnená početnými simuláciami pre rôzne voľby váhových funkcií a odhadov parametrov $\boldsymbol{\beta}$ a c, pomocou ktorých overíme silu testu.

Kľúčové slová: empirická charakteristická funkcia, lineárna regresia, test dobrej zhody

Title: Some tests of goodness-of-fit
Author: Radka Sabolová
Department: Department of Probability and Mathematical Statistics
Supervisor: Prof. RNDr. Marie Hušková, DrSc.
Supervisor's e-mail address: huskova@karlin.mff.cuni.cz

Abstract: The presented work deals with the goodness-of-fit test in linear regression model $\mathbf{Y_n} = \mathbf{X_n}\boldsymbol{\beta} + c\mathbf{e_n}$, where we focus upon error distribution. The uniqueness of the relation between the distribution and characteristic function will be used and the test statistics will be based on weighted distance of hypothetical characteristic function and empiric characteristic function of the residuals. A part of this work is concerned with estimating of parametres $\boldsymbol{\beta}$ and c and their asymptotic properties as well, since choosing between them also affects the limit distribution of the test statistics. Further the asymptotic properties of test statistics under null hypothesis, local alternative and fixed alternative are dealt with. The work is supplemented by numerous simulations for various choices of weighing functions and estimation of parametres $\boldsymbol{\beta}$ and c, using which the strenght of test will be verifyied.
Keywords: empirical characteristic function, linear regression, goodness-of-fit test

Keď sa povie test dobrej zhody, mnohým napadne asi ten najznámejší - Pearsonov χ^2 *test dobrej zhody. Niet divu, veď tento test si podľa [5] dokonca zaslúžil miesto medzi 20 najvýznamnejšími objavmi vo vede a technike od roku 1900 . V priebehu nasledujúcich 70 rokov bol tento test jedným z najpoužívanejších nástrojov štatistickej analýzy, predovšetkým vďaka svojmu použitiu v kontingenčných tabuľkách. Bolo to však prvé objektívne kritérium pre zistenie zhody medzi teóriou a realitou a viedol k vzniku mnohých postupov a techník, ktoré sú dnes už bežnou súčasťou štatistiky.*

Jedným z nich je aj napríklad dobre známy Kolmogorov–Smirnovov test či Cramér von Misesov test, ktoré sa zaoberajú vzdialenosťou medzi charakteristickou funkciou a empririckou charakteristickou funkciou. Keďže určitou transformáciou distribučnej funkcie môže vzniknúť charakteristická funkcia, tak sa ponúkajú aj testy dobrej zhody založené na vzdialenosti charakteristickej funkcie a empririckej charakteristickej funkcie. Touto cestou sa uberá aj táto práca. Presnejšie nás bude zaujímať rozdelenie chýb v modeli lineárnej regresie - čiže za prítomnosti rušivých parametrov.

Kapitola 1

Motivácia

V úvode pripomenieme prípad testov dobrej zhody bez rušivých parametrov. Majme náhodný výber $X_1, \dots, X_n$ s distribučnou funkciou F a chceme testovať hypotézu

$$H_0 : F \equiv F_0$$

proti alternatíve

$$H_1 : F \neq F_0.$$

Označíme

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} \mathrm{I}[X_i < x]$$

empirickú distribučnú funkciu výberu a uvedieme vetu pojednávajúcu o jej asymptotickom správaní.

Veta 1 (**Glivenko-Cantelli**). *Nech $F(x)$ označuje skutočnú distribučnú funkciu výberu a $F_n(x)$ empirickú distribučnú funkciu výberu. Označme*

$$D_n = \sup_x |F_n(x) - F(x)|.$$

Potom platí

$$P(\lim_{n\to\infty} D_n = 0) = 1.$$

Uvedieme si niekoľko testov založených na tejto vete - teda testy, v ktorých je testová štatistika založená na vzdialenosti hypotetickej a empirickej distribučnej funkcie. Jedným z najznámejších je Kolmogorov–Smirnovov test, ktorý hypotézu H_0 zamieta, ak $D_n = \sup_x |F_n(x) - F(0)|$ prekročí kritickú hodnotu k_n. Pre veľké hodnoty n môžeme použiť aproximáciu

$$P(D_n < \frac{\lambda}{\sqrt{n}}) \doteq 1 - 2e^{-2\lambda^2},$$

ktorá vyplýva napr. z vety 11.11 uvedenej v [1]. Výhodou tohto testu je, že rozdelenie testovej štatistiky nezáleží na F_0, avšak je vhodný len pre spojité rozdelenia.

Ako ďalší spomenieme Cramer von Misesov test, ktorého testová štatistika má tvar

$$\int_{-\infty}^{\infty} (F_n(x) - F_0(x))^2 \mathrm{d}F_0(x),$$

kde nulovú hypotézu zamietame v prípade, ak testová štatistika prekročí hodnotu c_n. A napokon Anderson–Darlingov test, ktorý je citlivejší na odchýlky na chvostoch než Kolmogorov-Smirnovov test a má kritický obor

$$n \int_{-\infty}^{\infty} \frac{(F_n(x) - F_0(x))^2}{F_n(x)(1 - F_n(x))} \mathrm{d}F_0 \geq a_n,$$

kde a_n závisí na F_0. Hodnoty a_n a c_n sú určené tak, aby testy mali hladinu aspoň α.

Charakteristická funkcia φ príslušná k distribučnej funkcii F je jednoznačne určená vzťahom

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} \mathrm{d}F(x), \qquad t \in \mathbb{R}.$$

Táto funkcia je spojitá, existuje pre všetky reálne t a jej hodnoty v absolútnej hodnote sú menšie alebo rovné 1. Práca [8] sa venovala testom hypotézy H_0 založeným práve na vzdialenosti charakteristickej funkcie $\varphi(t)$ a empirickej charakteristickej funkcie $\varphi_n(t)$. Jednou z výhod použitia charakteristickej funkcie v testoch dobrej zhody je jej spojitosť aj pre diskrétne rozdelenia.

Naša situácia je však trochu iná, avšak testová štatistika bude založená na rovnakej myšlienke. Nech $Y_{1,n}, \ldots, Y_{n,n}$ sú nezávislé náhodné veličiny (označme $\mathbf{Y_n} = (Y_{1,n}, \ldots, Y_{n,n})^T$) spĺňajúce model

$$\mathbf{Y_n} = \mathbf{X_n}\boldsymbol{\beta} + c\mathbf{e_n},$$

kde $\mathbf{X_n}$ je matica regresorov s riadkami $\mathbf{x_{j,n}}^T = (1, x_{j2,n}, \ldots, x_{jp,n}) \in \mathbb{R}^p$, $j = 1, \ldots, n$, $\boldsymbol{\beta} \in \mathbb{R}^p$ a $c > 0$ sú neznáme parametre a $\mathbf{e_n} = (e_{1,n}, \ldots, e_{n,n})^T$ je vektor nezávislých rovnako rozdelených náhodných veličín s distribučnou funkciou F. Chceme testovať nulovú hypotézu

$$H_0 : F \equiv F_0$$

proti alternatíve

$$H_1 : F \neq F_0,$$

kde F_0 je ľubovoľná konkrétna distribučná funkcia symetrická okolo nuly a spĺňajúca ďalšie podmienky.

Pri testovaní nulovej hypotézy použijeme reziduá

$$\hat{e}_{j,n} = \frac{Y_{j,n} - \mathbf{x_{j,n}}^T \hat{\boldsymbol{\beta}}_\mathbf{n}}{\hat{c}_n},$$

kde $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ je odhad parametru $\boldsymbol{\beta}$ a $\hat{c}_n$ je odhad parametru c - parametre $\boldsymbol{\beta}$ a c sú tu rušivé. Spôsobov, akými je možné odhadnúť parametre $\boldsymbol{\beta}$ a c je viacero, tie najpoužívanejšie neskôr pripomenieme.

Asymptotické vlastnosti testovej štatistiky budeme skúmať v Hilbertovom priestore $\mathcal{H} = L^2(\mathbb{R},\mathcal{B},w(t)\mathrm{d}t)$ merateľných funkcií $f : \mathbb{R} \to \mathbb{R}$ takých, že

$$\int_{-\infty}^{\infty} f(t)^2 w(t)\mathrm{d}t < \infty,$$

kde $w(t)$ je nezáporná párna váhová funkcia, pre ktorú platí $\int_{-\infty}^{\infty} w(t)t^2\mathrm{d}t < \infty$. Skalárnym súčinom na priestore $\mathcal{H}$ označujeme

$$\langle f,g\rangle = \int_{-\infty}^{\infty} f(t)g(t)w(t)\mathrm{d}t$$

a normou na priestore $\mathcal{H}$ označujeme

$$||f|| = \sqrt{\int\limits_{-\infty}^{\infty} f(t)^2 w(t)\mathrm{d}t.}$$

V nasledujúcej kapitole uvedieme testovú štatistiku, ale budeme sa zaoberať najmä odhadmi parametrov $\boldsymbol{\beta}$ a c – spomenieme niektoré z nich a ďalej sa zameráme na ich asymptotické správanie. V tretej kapitole sa budeme venovať limitnému rozdeleniu testovej štatistiky – za platnosti hypotézy, pri lokálnej alternatíve a pri pevnej alternatíve. V poslednej – štvrtej kapitole si odvodíme tvar testovej štatistiky pre normálne, Laplaceovo, logistické a Studentovo t rozdelenie pre rôzne váhové funkcie a navrhnutý test aplikujeme na simulované dáta pre rôzne voľby odhadov $\boldsymbol{\beta}$ a c.

Kapitola 2

Základné pojmy a predpoklady

Máme nezávislé pozorovania $Y_{1,n},\ \ldots,\ Y_{n,n}$ spĺňajúce model

$$Y_{j,n} = \mathbf{x}_{j,n}^T \boldsymbol{\beta} + c e_{j,n}, \quad j = 1{,}2,\ \ldots,\ n, \tag{2.1}$$

kde $\mathbf{x}_{j,n} = (1,\ x_{j2,n},\ \ldots,\ x_{jp,n})^T \in \mathbb{R}^p, j = 1,\ \ldots,\ n$ sú známe regresory (riadky matice $\mathbf{X_n}$), $\beta \in \mathbb{R}^p$ a $c > 0$ označujú neznáme parametre a chyby $e_{j,n},\ j = 1,\ \ldots,\ n$ sú nezávislé rovnako rozdelené s distribučnou funkciou $F(\cdot)$. Chceme testovať hypotézu

$$H_0 : F \equiv F_0 \tag{2.2}$$

proti všeobecným alternatívam pre ľubovoľnú F_0, ktorá je symetrická okolo nuly a spĺňa ďalšie podmienky.

Označme $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ odhad parametra $\boldsymbol{\beta}$ a $\hat{c}_n$ odhad parametra c, $\boldsymbol{\beta_0}$ skutočnú hodnotu parametra $\boldsymbol{\beta}$ a c_0 skutočnú hodnotu parametra c. Ďalej označme odhadnuté reziduá

$$\hat{e}_{j,n} = \frac{Y_{j,n} - \mathbf{x_{j,n}^T} \hat{\boldsymbol{\beta}}_{\boldsymbol{n}}}{\hat{c}_n}, \qquad j = 1, \ldots, n.$$

Budeme uvažovať len tie odhady $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$, ktoré spĺňajú

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}(\{\mathbf{x_{j,n}}, Y_{j,n} + \mathbf{x_{j,n}^T v}\}; j = 1,\ \ldots,\ n) = \hat{\boldsymbol{\beta}}_{\boldsymbol{n}}(\{\mathbf{x_{j,n}}, Y_{j,n}\}; j = 1,\ \ldots,\ n) + \mathbf{v} \qquad \forall\, \mathbf{v} \in \mathbb{R}^p \tag{2.3}$$

a odhady $\hat{c}_n$, pre ktoré platí

$$\hat{c}_n(\mathbf{x_{j,n}},\ bY_{j,n}; j = 1,\ \ldots,\ n) = b\hat{c}_n(\mathbf{x_{j,n}},\ Y_{j,n}; j = 1,\ \ldots,\ n) \qquad \forall\, b > 0. \tag{2.4}$$

V takom prípade totiž každá testová štatistika, ktorá závisí na $\{\mathbf{x_{j,n}}, Y_{j,n}\}$ výlučne prostredníctvom $\hat{e}_{j,n},\ j = 1,\ \ldots,\ n$ je regresne a škálovo invariantná, a preto jej rozdelenie nezávisí na neznámych hodnotách parametrov $\boldsymbol{\beta}$ a c. Okrem toho od odhadov $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ a $\hat{c}_n$ požadujeme, aby boli konzistentné – potom totiž pre veľké hodnoty n budú mať $\hat{e}_{1,n},\ \ldots,\ \hat{e}_{n,n}$ približne distribučnú funkciu F_0.

Charakteristickú funkciu príslušnú k F_0 označme $\varphi_0(t)$ a empirickú charakteristickú funkciu odhadnutých reziduí $\varphi_n(t)$, teda

$$\varphi_n(t) = \frac{1}{n}\sum_{i=1}^{n} \exp\left(it\hat{e}_{j,n}\right).$$

Testová štatistika bude založená na vzdialenosti φ_0 a φ_n a má tvar

$$T_{n,w} = n\int_{-\infty}^{\infty} |\varphi_n(t) - \varphi_0(t)|^2 w(t)\mathrm{d}t, \tag{2.5}$$

kde $w(t)$ je nezáporná párna váhová funkcia spĺňajúca $\int_{-\infty}^{\infty} t^2 w(t)\mathrm{d}t < \infty$. Hypotézu H_0 zamietame pre veľké hodnoty testovej štatistiky.

Ďalej sformulujeme predpoklady týkajúce sa nielen $e_{1,n},\ \ldots,\ e_{n,n}$, ale aj odhadov parametrov $\boldsymbol{\beta}$ a c a regresorov. Tie sa využijú v nasledujúcej kapitole vo vetách o limitnom tvare testovej štatistiky.
Nech rozdelenie chýb v regresnom modeli spĺňa:

(A.1) $e_{1,n},\ \ldots,e_{n,n}$ sú nezávislé rovnako rozdelené náhodné veličiny so symetrickou distribučnou funkciou F_0,

(A.2) charakteristická funkcia $\varphi_0(t)$ príslušná k F_0 je absolútne spojitá a funkcia

$$V(a) = \int_{-\infty}^{\infty} t^2(\varphi_0'(at))^2 w(t)\mathrm{d}t$$

je spojitá v $a = 1$ (túto podmienku môžeme vynechať, ak predpokladáme, že rozdelenie chýb má konečný prvý moment).

Odhady $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ a c_n okrem (2.3) a (2.4) spĺňajú aj

(B.1) odhad $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ pre $n \to \infty$ spĺňa

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \mathbf{x}_{\mathbf{j,n}}^{\mathbf{T}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}_{\mathbf{0}}) = \frac{d_\beta}{\sqrt{n}}\sum_{j=1}^{n} \psi_\beta(e_{j,n}) + o_P(1), \tag{2.6}$$

kde ψ_β je merateľná antisymetrická funkcia, pre ktorú platí $\mathsf{E}\psi_\beta(e_{1,n}) = 0$ a $\mathsf{E}\psi_\beta^2(e_{1,n}) < \infty$,

(B.2) odhad $\hat{c}_n$ pre $n \to \infty$ spĺňa

$$\sqrt{n}(\hat{c}_n - c_0) = \frac{d_c}{\sqrt{n}}\sum_{j=1}^{n} \psi_c(e_{j,n}) + o_P(1), \tag{2.7}$$

kde ψ_c je symetrická merateľná funkcia, pre ktorú platí $\mathsf{E}\psi_c(e_{1,n}) = 0$ a $\mathsf{E}\psi_c^2(e_{1,n}) < \infty$.

Požiadavka na regresory

(C.1) $\lim_{n\to\infty} \max_{1\le j\le n, 1\le v\le p} |x_{jv,n}| n^{-\frac{1}{2}} = 0.$

Existuje mnoho spôsobov, ako odhadnúť parametre $\boldsymbol{\beta}$ a c – vo zvyšku tejto kapitoly uvedieme najpoužívanejšie z nich, doplníme predpoklady pre distribučnú funkciu F, odhady stručne popíšeme a nakoniec pre ne overíme podmienky (B.1) a (B.2) – teda nás bude zaujímať asymptotické správanie jednotlivých odhadov. Budeme sa zaoberať nasledujúcimi odhadmi:

- odhad metódou najmenších štvorcov,
- odhad pomocou regresných kvantilov,
- M-odhad.

2.1 Prehľad odhadov

V tejto kapitole si spomenieme niektoré z odhadov parametrov $\boldsymbol{\beta}$ a c, všetky popísané odhady spĺňajú (2.3) a (2.4). Pre tieto odhady neskôr overíme aj podmienky (B.1) a (B.2) a väčšinu z nich použijeme ich v kapitole 4. Pre viac informácií o spomenutých odhadoch viď napríklad [9] alebo [6].

V nasledujúcom predpokladáme, že existuje symetrická pozitívne definitná matica $\mathbf{D}$ taká, že platí

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_{j,n} \mathbf{x}_{j,n}^{\mathrm{T}} = \mathbf{D}.$$

Takisto predpokladáme, že regresory spĺňajú podmienku (C.1) a $e_{1,n}, \ldots, e_{n,n}$ spĺňajú (A.1). Ďalej si označíme

$$\mathbf{H_n} = \mathbf{X_n}(\mathbf{X_n^T X_n})^{-1}\mathbf{X_n^T} \tag{2.8}$$

a $h_{ij,n}$ označíme prvok matice $\mathbf{H_n}$ na pozícii (i,j). Taktiež predpokladáme, že regresory spĺňajú podmienku (C.1) a $e_{1,n}, \ldots, e_{n,n}$ spĺňajú (A.1).

Odhad metódou najmenších štvorcov

Navyše predpokladáme $\mathsf{E}e_{j,n} = 0$, $\mathsf{E}e_{j,n}^2 = 1$, $\mathsf{E}|e_{j,n}|^4 < \infty$. Potom odhad metódou najmenších štvorcov má tvar

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} = \mathbf{D}_n^{-1} \sum_{i=1}^{n} \mathbf{x}_{j,n} Y_{j,n}$$

a

$$\hat{c}_n^2 = \frac{1}{n-p} \sum_{i=1}^{n} (Y_{j,n} - \mathbf{x_{j,n}^T} \hat{\boldsymbol{\beta}}_{\boldsymbol{n}})^2,$$

kde

$$\mathbf{D}_n = \sum_{j=1}^{n} \mathbf{x_{j,n}} \mathbf{x_{j,n}^T}.$$

Ak je F distribučná funkcia normálneho rozdelenia, potom $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ je maximálne vierohodným odhadom $\boldsymbol{\beta}$. Pre všeobecnú distribučnú funkciu F, ktorá má konečný druhý moment, je $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ podľa Gauss - Markovovej vety najlepším nestranným lineárnym odhadom $\boldsymbol{\beta}$. Avšak je veľmi citlivý na odľahlé pozorovania, na odchýlky od normálneho rozdelenia $e_{1,n}, \ldots, e_{n,n}$ a zlyháva, ak rozdelenie chýb má ťažké chvosty.

Pre odhad parametra $\boldsymbol{\beta}$ platí nasledujúca limitná veta. Postačiteľnosť podmienky

$$\max_{1 \leq i \leq n} h_{ii,n} \xrightarrow[n \to \infty]{} 0$$

je dokázaná napríklad v článku [11].

Veta 2. *Uvažujme model*

$$\mathbf{Y_n} = \mathbf{X_n} \boldsymbol{\beta} + \mathbf{Z_n},$$

kde $\mathbf{X_n}$ *je pevná matica rozmerov* n *x* p*, jej hodnosť je rovná* p*,* $\mathbf{Z_n}$ *je vektor nezávislých rovnako rozdelených náhodných veličín so strednou hodnotou* 0 *a rozptylom* $\sigma^2 < \infty$*. Označme* $\mathbf{H_n}$ *ako v (2.8) a* $h_{ii,n}$ *i-tý diagonálny prvok tejto matice. Nech*

$$\max_{1 \leq i \leq n} h_{ii,n} \xrightarrow[n \to \infty]{} 0.$$

Potom

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}_{\mathbf{0}}) \xrightarrow[n \to \infty]{\mathcal{D}} N_p(0, \sigma^2 \mathbf{D}^{-1}).$$

Regresné kvantily

Regresný q-kvantil je definovaný ako

$$\widehat{\beta_n}(q) = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \sum_{j=1}^{n} \rho_q (Y_{j,n} - \mathbf{x}_{j,n}^T \beta) \right\},$$

kde $\rho_q(z) = |z|(q\mathrm{I}\{z > 0\} + (1-q)\mathrm{I}\{z < 0\})$, $x \in \mathbb{R}$, $0 < q < 1$. Uvažujme odhady

$$\hat{\beta}_n = \widehat{\beta_n} \left(\frac{1}{2} \right), \tag{2.9}$$

$$\hat{c}_n = \widehat{\beta_{n1}} \left(\frac{3}{4} \right) - \widehat{\beta_{n1}} \left(\frac{1}{4} \right), \tag{2.10}$$

kde $\widehat{\beta_{n1}}(q)$ označuje prvú zložku $\widehat{\beta_n}(q)$, $q \in (0,1)$. Odhad (2.9) sa nazýva aj L_1 odhad, či regresný medián.
Je tu nutný predpoklad existencie f_0 a jej spojitosti na okolí $F_0^{-1}(1/4)$, $F_0^{-1}(1/2)$, $F_0^{-1}(3/4)$. Ďalej $f_0(F_0^{-1}(q)) > 0$ pre $q = 1/4,\ 1/2,\ 3/4$.
Pre viac informácií o regresných kvantiloch viď napríklad [9].

M-odhady

Opäť budeme predpokladať $\mathsf{E}e_{1,n}^2 = 1$. M-odhady sú definované ako argument minima

$$\sum_{j=1}^{n} \rho\left(\frac{Y_{j,n} - \mathbf{x}_{j,n}^T\beta}{c}\right),$$

kde $\rho : \mathbb{R} \to \mathbb{R}$ je absolútne spojitá konvexná funkcia, $\rho(0) = 0$. Teda M-odhady sú riešením sústavy rovníc

$$\sum_{j=1}^{n} \mathbf{x}_j \psi\left(\frac{Y_{j,n} - \mathbf{x}_{j,n}^T\beta}{c}\right) = \mathbf{0}, \qquad \sum_{j=1}^{n} \chi\left(\frac{Y_{j,n} - \mathbf{x}_{j,n}^T\beta}{c}\right) = 0, \tag{2.11}$$

kde $\psi(t) = \rho'(t)$ a $\chi(t) = t\psi(t) - \rho(t)$. V tomto prípade je ψ monotónna nepárna funkcia a χ je na $\mathbb{R}$ párna a na intervale $(0,\infty)$ monotónna. Ďalej pre funkcie

$$\lambda_\psi(t) = -\int \psi(x-t)\mathrm{d}F_0(x), \quad t \in \mathbb{R},$$

$$\lambda_\chi(t) = -\int \chi(x/t)\mathrm{d}F_0(x), \quad t \in \mathbb{R},$$

predpokladáme platnosť

$$\lambda_\psi(0) = 0, \quad \lambda'_\psi(0) > 0,$$

derivácia $\lambda'_\beta(t)$ je spojitá na okolí nuly a

$$\int \psi^2(t)\mathrm{d}t < \infty, \qquad \int (\psi(x) - \psi(x+q_n))^2 \mathrm{d}F_0(x) \xrightarrow[n\to\infty]{} 0$$

pre ľubovoľné q_n také, že $q_n - 1 = o_P(1)$. Označíme $d_\psi = \lambda'_\psi(0)$.
Podobne pre funkciu χ predpokladáme

$$\lambda_\chi(1) = 0, \quad \lambda'_\chi(1) > 0,$$

derivácia $\lambda'_\chi(t)$ je spojitá na okolí $t = 1$ a

$$\int \chi^2(t)\mathrm{d}t < \infty, \qquad \int \left(\chi(x) - \chi\left(\frac{x}{q_n}\right)\right)^2 \mathrm{d}F_0(x) \to 0$$

pre ľubovoľné q_n také, že $q_n - 1 = o_P(1)$. Označíme $d_\chi = \lambda'_\chi(1)$.
Ak použijeme

$$\psi = x, \qquad x \in \mathbb{R}$$

dostaneme odhad *metódou najmenších štvorcov* a pre voľbu

$$\psi = \operatorname{sign} x, \qquad x \in \mathbb{R}$$

zase L_1 *odhad.*

Ďalším špeciálnym prípadom je *odhad metódou maximálnej vierohodnosti.* Predpokladajme, že F_0 je absolútne spojitá s hustotou f_0. Ak sú splnené podmienky regularity (viď [1], definícia 7.8, str. 104), potom odhad metódou maximálnej vierohodnosti je ekvivalentný M-odhadu so skórovými funkciami

$$\psi(x) = -\frac{f_0'(x)}{f_0(x)}, \quad \chi(x) = -1 - x\frac{f_0'(x)}{f_0(x)}, \quad x \in \mathbb{R}.$$

Niekedy sa uvažujú M-odhady, ktoré sú riešením sústavy rovníc

$$\sum_{j=1}^{n} \mathbf{x}_j \psi\left(\frac{Y_{j,n} - \mathbf{x}_{j,n}^T \beta}{c}\right) = \mathbf{0}, \qquad \sum_{j=1}^{n} \chi\left(\frac{Y_{j,n} - \mathbf{x}_{j,n}^T \beta}{c}\right) = A_n, \tag{2.12}$$

kde A_n je vhodne zvolená postupnosť konštánt, napr. Huber [6] odporúča

$$A_n = (n-p)\mathsf{E}_{F_0}\chi. \tag{2.13}$$

Ďalším možným prístupom je použitie *studentizovaného M-odhadu.* Uvažujme model

$$\mathbf{Y_n} = \mathbf{X_n}\boldsymbol{\beta} + \mathbf{Z_n}.$$

Studentizovaný M-odhad je potom riešenie sústavy rovníc

$$\sum_{j=1}^{n} \psi\left(\frac{Y_{j,n} - \mathbf{x_{j,n}^T} t}{S_n}\right), \qquad t \in \mathbb{R}^p,$$

kde $S_n > 0$ je vhodná škálová štatistika.

Nakoniec ešte stručne spomenieme jednokrokové M-odhady, pre viac informácií viď [9], str. 318–321. Označme $\tilde{\boldsymbol{\beta}_n}$ odhad parametra $\boldsymbol{\beta}$ spĺňajúci (2.3) a $\tilde{c_n}$ odhad c spĺňajúci (2.4), nech

$$\sqrt{n}(\tilde{\boldsymbol{\beta}_n} - \boldsymbol{\beta}) = O_P(1), \qquad \sqrt{n}(\tilde{c_n} - c) = O_P(1)$$

(môžeme použiť napríklad odhady pomocou regresných kvantilov). Jednokrokové M-odhady $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$, $\hat{c}_n$ sú definované ako

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} = \tilde{\boldsymbol{\beta}_n} + d_\psi \mathbf{D_n}^{-1} \sum_{j=1}^{n} \mathbf{x_{j,n}} \psi(\tilde{e}_{j,n}),$$

$$\hat{c}_n = \tilde{c_n} + d_\chi \sum_{j=1}^{n} \chi(\tilde{e}_{j,n}),$$

kde ψ, χ, d_ψ a d_χ spĺňajú vlastnosti uvažované pre M-odhady a

$$\tilde{e}_{j,n} = \frac{Y_{j,n} - \mathbf{x_{j,n}^T}\tilde{\boldsymbol{\beta}_n}}{\tilde{c_n}}, \qquad j = 1, \ldots, n.$$

2.2 Odhad metódou najmenších štvorcov

V tejto podkapitole budeme okrem už spomenutých predpokladov navyše predpokladať aj $\mathsf{E}e_{1,n} = 0$, $\mathsf{E}e_{1,n}^2 = 1$ a $\mathsf{E}|e_{1,n}|^4 < \infty$.

2.2.1 Asymptotické rozdelenie $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$

Úpravou ľavej strany podmienky (B.1) pre odhad parametra $\boldsymbol{\beta}$ metódou najmenších štvorcov (teda $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} = (\mathbf{X_n^T X_n})^{-1}\mathbf{X_n^T Y_n}$) dostaneme

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}_{\mathbf{0}}) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}((\mathbf{X_n^T X_n})^{-1}\mathbf{X_n^T Y_n} - \boldsymbol{\beta}_{\mathbf{0}}) \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}((\mathbf{X_n^T X_n})^{-1}\mathbf{X_n^T X_n}\boldsymbol{\beta}_{\mathbf{0}} + \\
&\quad + c_0(\mathbf{X_n^T X_n})^{-1}\sum_{k=1}^{n}\mathbf{x_{k,n}}e_{k,n} - \boldsymbol{\beta}_{\mathbf{0}}) \\
&= \frac{c_0}{\sqrt{n}}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbf{x_{j,n}^T}(\mathbf{X_n^T X_n})^{-1}\mathbf{x_{k,n}}e_{k,n} \\
&= \frac{c_0}{\sqrt{n}}\sum_{k=1}^{n}\sum_{j=1}^{n}h_{jk,n}e_{k,n}.
\end{aligned}$$

Pretože v modeli uvažujeme absolútny člen, platí

$$\sum_{j=1}^{n}h_{jk,n} = \sum_{k=1}^{n}h_{jk,n} = 1.$$

Teda podmienka (B.1) je pre odhad metódou najmenších štvorcov splnená a má tvar

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}_{\mathbf{0}}) = \frac{c_0}{\sqrt{n}}\sum_{k=1}^{n}e_{k,n},$$

čiže

$$\psi_\beta(x) = x, \qquad d_\beta = c_0,$$

ψ_β je merateľná nepárna funkcia a z predpokladov spomenutých vyššie máme aj $\mathsf{E}\psi_\beta(e_{1,n}) = 0$ a $\mathsf{E}\psi_\beta^2(e_{1,n}) < \infty$. Navyše podľa Lévy-Lindeberghovej centrálnej limitnej vety pre nezávislé náhodné veličiny $e_{j,n}$, $\mathsf{E}e_{1,n} = 0$, $\mathsf{var}e_{1,n} = 1$ máme

$$\frac{c_0}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}_{\mathbf{0}}) = \frac{c_0}{\sqrt{n}}\sum_{k=1}^{n}e_{k,n} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{Z},$$

kde $\mathcal{Z} \sim$ N(0,1).

2.2.2 Asymptotické rozdelenie $\hat{c}_n$

Najprv si upravíme tvar $\hat{c}_n$ a vyjadríme ho pomocou vektoru chýb $\mathbf{e_n}$

$$
\begin{aligned}
\hat{c}_n^2 &= \frac{1}{n-p}(\mathbf{Y_n} - \mathbf{X_n}\hat{\boldsymbol{\beta}}_{\boldsymbol{n}})^{\mathbf{T}}(\mathbf{Y_n} - \mathbf{X_n}\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}) \\
&= \frac{1}{n-p}(\mathbf{Y_n} - \mathbf{X_n}(\mathbf{X_n^T X_n})^{-1}\mathbf{X_n^T X_n}\boldsymbol{\beta_0} - c_0\mathbf{X_n}(\mathbf{X_n^T X_n})^{-1}\mathbf{X_n^T e_n})^{\mathbf{T}} \\
&\quad (\mathbf{Y_n} - \mathbf{X_n}(\mathbf{X_n^T X_n})^{-1}\mathbf{X_n^T X_n}\boldsymbol{\beta_0} - c_0\mathbf{X_n}(\mathbf{X_n^T X_n})^{-1}\mathbf{X_n^T e_n}) \\
&= \frac{1}{n-p}(c_0\mathbf{e_n} - c_0\mathbf{X_n}(\mathbf{X_n^T X_n})^{-1}\mathbf{X_n^T e_n})^{\mathbf{T}}(c_0\mathbf{e_n} - c_0\mathbf{X_n}(\mathbf{X_n^T X_n})^{-1}\mathbf{X_n^T e_n}) \\
&= \frac{c_0^2}{n-p}\mathbf{e_n^T}(\mathbf{I_n} - \mathbf{H_n})\mathbf{e_n},
\end{aligned}
$$

kde c_0 je skutočná hodnota parametra c. Pretože predpokladáme

$$\mathsf{E}e_{j,n}^2 = 1, \quad j = 1, \dots, n,$$

$$\mathsf{E}|e_{j,n}^4| < \infty, \quad j = 1, \dots, n,$$

a

$$\frac{n}{n-p} \xrightarrow[n\to\infty]{} 1,$$

podľa silného zákona veľkých čísel (viď [4], str. 79) platí

$$\frac{c_0^2}{n-p}\sum_{j=1}^{n} e_{j,n}^2 \xrightarrow[n\to\infty]{} c_0^2 \qquad \text{skoro iste.}$$

Najprv sa pozrieme na výraz

$$\sqrt{n}(\hat{c}_n^2 - c_0^2) = \sqrt{n}\left(\frac{1}{n-p}\sum_{j=1}^{n} e_{j,n}^2 c_0^2 - c_0^2\right) - \frac{\sqrt{n}c_0^2}{n-p}\mathbf{e_n^T H_n e_n}, \qquad (2.14)$$

kde chceme dokázať

$$\frac{\sqrt{n}c_0^2}{n-p}\mathbf{e_n^T H_n e_n} \xrightarrow[n\to\infty]{P} 0.$$

Tento výraz je s pravdepobnosťou 1 nezáporný, pretože

$$\mathbf{e_n^T H_n e_n} = \operatorname{tr}\mathbf{H_n e_n^T e_n} = p\sum_{j=1}^{n} e_{j,n}^2$$

a navyše pre jeho strednú hodnotu platí

$$\mathsf{E}\mathbf{e_n^T H_n e_n} = \operatorname{tr}\mathsf{E}\mathbf{e_n^T H_n e_n} = \operatorname{tr}\mathbf{H_n}\mathsf{E}\mathbf{e_n e_n^T} = \operatorname{tr}\mathbf{H_n} = p.$$

Potom z Čebyševovej nerovnosti ([4], str. 26) pre $\frac{1}{\sqrt{n}}\mathbf{e_n^T H_n e_n}$ získame pre každé $\varepsilon > 0$
$$P\left(\frac{1}{\sqrt{n}}\mathbf{e_n^T H_n e_n} > \varepsilon\right) \le \frac{\mathsf{E}\mathbf{e_n^T H_n e_n}}{\sqrt{n}\varepsilon} \xrightarrow[n\to\infty]{} 0.$$
Teda nakoniec dostávame požadovanú konvergenciu k nule v pravdepodobnosti, čiže
$$\frac{\sqrt{n}c_0^2}{n-p}\mathbf{e_n^T H_n e_n} \xrightarrow[n\to\infty]{P} 0. \tag{2.15}$$

Ďalej platí
$$\sqrt{n}(\hat{c}_n^2 - c_0^2) = \sqrt{n}(\hat{c}_n - c_0)(\hat{c}_n - c_0 + 2c_0)$$
a dokázali sme už, že
$$\hat{c}_n^2 \xrightarrow[n\to\infty]{P} c_0^2.$$
Potom použitím continuous mapping theorem pre funkciu $h(x) = \sqrt{x}$ (spojitá funkcia) dostaneme aj
$$\hat{c}_n \xrightarrow[n\to\infty]{P} c_0. \tag{2.16}$$

Dohromady zo vzťahov (2.14), (2.15) a (2.16) pre $n \to \infty$ získavame
$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(e_{j,n}^2 - 1)c_0^2 + o_P(1) \quad = \quad 2\sqrt{n}c_0(\hat{c}_n - c_0) + o_P(1)$$

Teda pre odhad parametra c
$$\hat{c}_n^2 = \frac{1}{n-p}\sum_{j=1}^{n}(Y_{j,n} - \mathbf{x_{j,n}^T}\hat{\boldsymbol{\beta}}_{\boldsymbol{n}})^2$$
platí podmienka (B.2) v tvare
$$\sqrt{n}(\hat{c}_n - c_0) = \frac{c_0}{2\sqrt{n}}\sum_{j=1}^{n}(e_{j,n}^2 - 1) + o_P(1),$$
takže
$$\psi_c(x) = x^2 - 1, \qquad d_c = \frac{c_0}{2},$$
platí $\mathsf{E}\psi_c(e_{1,n}) = 0$ a $\mathsf{E}\psi_c^2(e_{1,n}) < \infty$, ψ_c je párna merateľná funkcia. Okrem toho, podľa Lévy-Lindeberghovej centrálnej limitnej vety pre $e_{j,n}^2$ s využitím predpokladu $\mathsf{E}e_{j,n}^4 < \infty$ máme
$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}(e_{j,n}^2 - 1)c_0^2\right) \xrightarrow[n\to\infty]{\mathcal{D}} \mathrm{N}(0, c_0^4\mathsf{var}(e_{j,n}^2 - 1)).$$

2.3 Odhad pomocou regresných kvantilov

Predpokladajme, že $F_0^{-1}\left(\frac{1}{2}\right) = 0$ a $F_0^{-1}\left(\frac{3}{4}\right) - F_0^{-1}\left(\frac{1}{4}\right) = 1$. Tieto odhady sú vlastne tiež M–odhadmi s vhodne zvolenými funkciami ψ a χ, presnejšie

$$\psi(x) = \operatorname{sign}(x),$$

$$\chi(x) = \frac{1}{2} - \mathrm{I}\left[F_0^{-1}\left(\frac{1}{4}\right) \leq x \leq F_0^{-1}\left(\frac{3}{4}\right)\right].$$

2.3.1 Asymptotické rozdelenie $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$

Na zistenie asymptotického správania odhadu $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ použijeme nasledujúcu vetu.

Veta 3. *Uvažujme model*

$$\mathbf{Y_n} = \mathbf{X_n}\boldsymbol{\beta} + \mathbf{Z_n},$$

kde $\mathbf{X_n}$ je matica n x p, $\mathbf{Z_n} = (Z_{1,n}, \ldots, Z_{n,n})$ sú nezávislé náhodné veličiny s distribučnou funkciou F, matica $\mathbf{H_n}$ je definovaná ako v (2.8) a jej prvky označme $h_{ij,n}$. Nech $d_n\sqrt{\log n} \xrightarrow[n\to\infty]{} 0$, kde $d_n = \max\left(h_{ii,n}\right)^{1/2}, 1 \leq i \leq n$, potom za platnosti nulovej hypotézy $\boldsymbol{\beta} = \boldsymbol{\beta_0}$ platí s pravdepodobnosťou 1

$$2f(0)(\mathbf{X_n^T X_n})^{1/2}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta_0}) = \sum_{k=1}^{n}(\mathbf{X_n^T X_n})^{-1/2}\mathbf{x_{k,n}} \operatorname{sign} e_{k,n} + O(d_n^{1/2}(\log n)^{3/4}),$$

kde f je zodpovedajúca hustota príslušná k F.

Dôkaz: uvedený v [2].

Teda pri splnení predpokladu $d_n\sqrt{\log n} \xrightarrow[n\to\infty]{} 0$ za platnosti $\boldsymbol{\beta} = \boldsymbol{\beta_0}$ s pravdepodobnosťou 1 máme

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta_0}) &= \frac{1}{2f_0(0)\sqrt{n}}\sum_{k=1}^{n}\underbrace{\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\mathbf{X_n^T X_n})^{-1}\mathbf{x_{k,n}}}_{=1} \operatorname{sign} e_{k,n} + \\
&\quad + \frac{1}{2f_0(0)\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\mathbf{X_n^T X_n})^{-1/2} O(d_n^{1/2}(\log n)^{3/4}) \\
&= \frac{1}{2f_0(0)\sqrt{n}}\sum_{j=1}^{n}\operatorname{sign} e_{j,n} + \\
&\quad + \frac{1}{2f_0(0)\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\mathbf{X_n^T X_n})^{-1/2} O(d_n^{1/2}(\log n)^{3/4}).
\end{aligned}$$

Potrebujeme zjednodušiť výraz uvedený v poslednom riadku. Pretože

$$\begin{aligned}
\Big|\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\mathbf{X_n^T X_n})^{-1/2}\Big|\Big| &= \sqrt{\frac{1}{n}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\mathbf{X_n^T X_n})^{-1/2}\sum_{k=1}^{n}(\mathbf{X_n^T X_n})^{-1/2}\mathbf{x_{k,n}}} \\
&= \sqrt{\frac{1}{n}\sum_{j=1}^{n}\underbrace{\sum_{k=1}^{n}h_{jk,n}}_{=1}} \\
&= 1,
\end{aligned}$$

tak platí

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\mathbf{X_n^T X_n})^{1/2} = O(1),$$

a teda získame

$$\frac{1}{2f(0)\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\mathbf{X_n^T X_n})^{-1/2}O(d_n^{1/2}(\log n)^{3/4}) = O(d_n^{1/2}(\log n)^{3/4}).$$

Nakoniec ak predpokladáme, že $d_n^{1/2}(\log n)^{3/4} \xrightarrow[n\to\infty]{} 0$, tak pre $n \to \infty$ dostaneme

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}_{\mathbf{0}}) = \frac{1}{2f(0)\sqrt{n}}\sum_{j=1}^{n}\operatorname{sign} e_{j,n} + O(d_n^{1/2}(\log n)^{3/4}) \text{ skoro iste,}$$

čiže pre

$$d_n^{1/2}(\log n)^{3/4} \xrightarrow[n\to\infty]{} 0$$

dostávame vzťah (B.2) pre odhad pomocou regresných kvantilov (dokonca silnejšie tvrdenie, keďže máme konvergenciu skoro iste) s

$$\psi_\beta(x) = \operatorname{sign} x, \qquad d_\beta = \frac{1}{2f_0(0)},$$

kde ψ_β je merateľná nepárna funkcia a pre F_0 symetrickú platí $\mathsf{E}\psi_\beta(e_{1,n}) = 0$ a $\mathsf{E}\psi_\beta^2(e_{1,n}) < \infty$.

Podľa Lévy-Lindeberghovej centrálnej limitnej vety pre $\operatorname{sign} e_{j,n}$, $\mathsf{E}\operatorname{sign} e1,n = 0$, $\mathsf{var}\operatorname{sign} e_{1,n} = 1$ platí

$$\frac{2f_0(0)}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}_{\mathbf{0}}) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{Z},$$

kde $\mathcal{Z} \sim$ N(0,1).

2.3.2 Asymptotické rozdelenie $\hat{c}_n$

Najprv si uvedieme vetu, ktorá pojednáva o asymptotickom správaní sa regresných kvantilov.

Veta 4. *Nech distribučná funkcia F nezávislých rovnako rozdelených náhodných veličín $Z_{1,n}, \ldots, Z_{n,n}$ v modeli*

$$\mathbf{Y_n} = \mathbf{X_n}\boldsymbol{\beta} + \mathbf{Z_n},$$

kde $\mathbf{Y_n} = (Y_1, \ldots, Y_n)^T$, $\mathbf{X_n}$ je matica známych konštánt s riadkami $\mathbf{x_{j,n}^T}$, kde $x_{j1,n} = 1$, $j = 1, \ldots, n$ a $\mathbf{Z} = (Z_1, \ldots, Z_n)^T$, je spojitá a dvakrát diferencovateľná na okolí $F^{-1}(\alpha)$, $F'(F^{-1}(\alpha)) = f(F^{-1}(\alpha)) > 0$, $0 < \alpha < 1$. Nech platia podmienky

- *$lim_{n\to\infty}\frac{1}{n}\mathbf{X_n^T}\mathbf{X_n} = \mathbf{D}$, kde $\mathbf{D}$ je pozitívne definitná matica,*
- *$\frac{1}{n}\sum_{i=1}^{n} x_{ij,n}^4 = O(1)$ pre $n \to \infty$ pre $j = 1, \ldots, p$.*

Potom

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}(\alpha) - \tilde{\boldsymbol{\beta}}(\alpha) = \frac{1}{nf(F^{-1}(\alpha))}\mathbf{D}^{-1}\sum_{j=1}^{n}\mathbf{x_{j,n}}\phi_\alpha(Z_j - F^{-1}(\alpha)) + \mathbf{R}_n(\alpha), \qquad (2.17)$$

kde

$$\tilde{\boldsymbol{\beta}}(\alpha) = (\beta_1 + F^{-1}(\alpha), \beta_2, \ldots, \beta_p)^T,$$
$$||\mathbf{R}_n(\alpha)|| = O_P(n^{-3/4}) \text{ pre } n \to \infty$$

a

$$\phi_\alpha(x) = \alpha - I[x < 0].$$

Dôkaz: uvedený v [9], str 166–169.

Hodnoty α zvolíme $\frac{1}{4}$ a $\frac{3}{4}$ a uvedenú vetu aplikujeme na distribučnú funkciu náhodných veličín $c_0e_{1,n}, \ldots, c_0e_{n,n}$ - ich hypotetickú distribučnú funkciu označme G_0 a hustotu k nej príslušnú g_0. Je zrejmé, že platí

$$F_0(e_{j,n}) = G_0(c_0e_{j,n}).$$

Predpokladajme existenciu g_0 a jej spojitosť na okolí bodov $G_0^{-1}(1/4)$, $G_0^{-1}(1/2)$, $G_0^{-1}(3/4)$. Ďalej $g_0(G_0^{-1}(q)) > 0$ pre $q = 1/4,\ 1/2,\ 3/4$. Ako odhad parametra c vezmeme

$$\hat{c}_n = \widehat{\boldsymbol{\beta}_{n1}}\left(\frac{3}{4}\right) - \widehat{\boldsymbol{\beta}_{n1}}\left(\frac{1}{4}\right),$$

kde $\widehat{\boldsymbol{\beta}_{n1}}(q)$ označuje prvú zložku $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}(q)$.
Potom za predpokladu $\frac{1}{n}\sum_{i=1}^{n} x_{ij,n}^4 = O(1)$ pre $n \to \infty$ pre $j = 1, \ldots, p$, platí veta 4 a z nej dostávame

$$\widehat{\boldsymbol{\beta}_n}\left(\frac{3}{4}\right) - \tilde{\boldsymbol{\beta}}\left(\frac{3}{4}\right) - \left(\widehat{\boldsymbol{\beta}_n}\left(\frac{1}{4}\right) - \tilde{\boldsymbol{\beta}}\left(\frac{1}{4}\right)\right) =$$

$$
\begin{aligned}
&= \frac{1}{n g_0(G_0^{-1}(\frac{3}{4}))}\mathbf{D}^{-1}\sum_{j=1}^{n}\mathbf{x_{j,n}}\left(\frac{3}{4} - I\left[c_0 e_{j,n} - G_0^{-1}\left(\frac{3}{4}\right) < 0\right]\right) - \\
&\quad - \frac{1}{n g_0(G_0^{-1}(\frac{1}{4}))}\mathbf{D}^{-1}\sum_{j=1}^{n}\mathbf{x_{j,n}}\left(\frac{1}{4} - I\left[c_0 e_{j,n} - G_0^{-1}\left(\frac{1}{4}\right) < 0\right]\right) + O_P(n^{-3/4}) = \\
&= \frac{1}{n g_0(G_0^{-1}(\frac{3}{4}))}\mathbf{D}^{-1}\sum_{j=1}^{n}\mathbf{x_{j,n}}\left(\frac{1}{2} - I\left[G_0^{-1}\left(\frac{1}{4}\right) < c_0 e_{j,n} < G_0^{-1}\left(\frac{3}{4}\right)\right]\right) + O_P(n^{-3/4}),
\end{aligned}
$$

kde však

$$
I\left[G_0^{-1}\left(\frac{1}{4}\right) < c_0 e_{j,n} < G_0^{-1}\left(\frac{3}{4}\right)\right] = I\left[F_0^{-1}\left(\frac{1}{4}\right) < e_{j,n} < F_0^{-1}\left(\frac{3}{4}\right)\right]
$$

a

$$
g_0\left(G_0^{-1}\left(\frac{3}{4}\right)\right) = \frac{1}{c_0} f_0\left(F_0^{-1}\left(\frac{3}{4}\right)\right).
$$

Tento výraz si ešte potrebujeme upraviť, aby sme dostali podmienku (B.2). Teda prenásobením riadkovým vektorom $\mathbf{x_{j,n}^T} = (1, x_{j1,n}, \ldots, x_{jp,n})$, sčítaním cez $j = 1, \ldots, n$ a vynásobením $\sqrt{n}$ dostaneme podmienku (B.2) v tvare

$$
\sqrt{n}(\hat{c}_n - c_0) = \frac{1}{\sqrt{n}}\left(\widehat{\boldsymbol{\beta}_{n1}}\left(\frac{3}{4}\right) - \widehat{\boldsymbol{\beta}_{n1}}\left(\frac{1}{4}\right) - \left(G_0^{-1}\left(\frac{3}{4}\right) - G_0^{-1}\left(\frac{1}{4}\right)\right)\right) =
$$

$$
\frac{c_0}{\sqrt{n} f_0(F_0^{-1}(\frac{3}{4}))}\sum_{j=1}^{n}\left(\frac{1}{2} - I\left[F_0^{-1}\left(\frac{1}{4}\right) < e_{j,n} < F_0^{-1}\left(\frac{3}{4}\right)\right]\right) + o_P(1),
$$

keďže platí

$$
\frac{1}{n}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}\left(\widehat{\boldsymbol{\beta}_n}\left(\frac{3}{4}\right) - \widehat{\boldsymbol{\beta}_n}\left(\frac{1}{4}\right)\right) - \widehat{\boldsymbol{\beta}_{n1}}\left(\frac{3}{4}\right) - \widehat{\boldsymbol{\beta}_{n1}}\left(\frac{1}{4}\right) \xrightarrow[n\to\infty]{P} 0.
$$

Čiže podmienka (B.2) je pre tento odhad splnená s

$$
\psi_c(x) = \frac{1}{2} - I\left[F_0^{-1}\left(\frac{1}{4}\right) < x < F_0^{-1}\left(\frac{3}{4}\right)\right], \qquad d_c = \frac{c_0}{f_0(F_0^{-1}(\frac{3}{4}))},
$$

funkcia ψ_c je párna merateľná, $\mathsf{E}\psi_c(e_{1,n}) = 0$, $\mathsf{E}\psi_c^2(e_{1,n}) < \infty$.

2.4 Huberov M-odhad

V tejto podkapitole sa budeme zaoberať dvoma typmi M-odhadov s použitím Huberovej ψ-funkcie. Prvým bude studentizovaný M-odhad, pri ktorom použijeme škálovú štatistiku S_n a v druhom prípade zasa budeme odhadovať sústavou rovníc naraz parameter $\boldsymbol{\beta}$ aj c.

2.4.1 Asymptotické rozdelenie $\hat{\boldsymbol{\beta}}_n$ pri použití studentizovaného M-odhadu

Nech funkciu ψ je možné rozložiť ako

$$\psi = \psi_a + \psi_k + \psi_s,$$

kde ψ_a je absolútne spojitá funkcia s absolútne spojitou deriváciou, ψ_k je spojitá po častiach lineárna funkcia konštantná na okolí $\pm\infty$ a ψ_s je neklesajúca schodovitá funkcia.

Pre získanie asymptotického rozdelenia Huberovho M-odhadu použijeme dôsledok 5.5.3 uvedený v [9], str. 220.

Veta 5. *Uvažujme model*

$$\mathbf{Y_n} = \mathbf{X_n}\boldsymbol{\beta} + \mathbf{Z_n},$$

kde $\mathbf{Y_n}$ je vektor pozorovaní, $\mathbf{X_n}$ je matica konštánt, $\boldsymbol{\beta}$ je neznámy parameter a $\mathbf{Z_n}$ je vektor nezávislých náhodných veličín s distribučnou funkciou F. Nech $F(z)+F(-z) = 1$, $\rho(-z) = \rho(z)$, $z \in \mathbb{R}$. Nech sú splnené nasledujúce predpoklady:

M1 *$S_n(\mathbf{Y_n})$ je regresne a škálovo invariantná štatistika, $S_n > 0$ skoro iste a*

$$\sqrt{n}(S_n - S) = O_P(1)$$

pre nejaký funkcionál $S = S(F) > 0$,

M2 *funkcia $h(t) = \int \rho((z-t)/S)dF(z)$ má jediné minimum v bode $t = 0$*

M3 *pre nejaké $\delta > 0$ a $\eta > 1$ platí*

$$\int_{-\infty}^{\infty} \left\{ |z| \sup_{|u|\leq\delta} \sup_{|v|\leq\delta} |\psi_a''(e^{-v}(z+u)/S)| \right\}^{\eta} dF(z) < \infty$$

a

$$\int_{-\infty}^{\infty} \left\{ |z|^2 \sup_{|u|<\delta} |\psi_a''(z+u)/S| \right\}^{\eta} dF(z) < \infty,$$

kde $\psi_a'(z) = (d/dz)\psi_a(z)$ a $\psi_a''(z) = (d^2/dz^2)\psi_a(z)$,

M4 *ψ_k je spojitá, po častiach lineárna funkcia s uzlami $\mu_1,\ldots,\mu_k$, ktorá je konštantná na okolí $\pm\infty$. Jej derivácia ψ_k' je potom schodovitá funkcia*

$$\psi_k'(z) = \alpha_\nu \text{ pre } \mu_\nu < z < \mu_{\nu+1}, \nu = 0,1,\ldots,k,$$

kde $\alpha_0,\alpha_1,\ldots,\alpha_k \in \mathbb{R}$, $\alpha_0 = \alpha_k = 0$ a $-\infty = \mu_0 < \mu_1 < \ldots < \mu_k < \mu_{k+1} = \infty$. Ďalej predpokladáme, že $f(z) = \frac{dF(z)}{dz}$ je ohraničená na okolí $S\mu_1,\ldots,S\mu_k$,

X1 $x_{j1,n} = 1,\ j = 1,\ldots,n,$

X2 $\frac{1}{n}\sum_{i=1}^{n} x_{ij,n}^4 = O(1)$ *pre* $n \to \infty$ *pre* $j = 1,\ldots,p,$

X3 $lim_{n\to\infty}\frac{1}{n}\mathbf{X_n^T X_n} = \mathbf{D}$, *kde* $\mathbf{D}$ *je pozitívne definitná matica,*

Potom

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta} = \frac{1}{n\gamma_1}\left(\frac{1}{n}\mathbf{X_n^T X_n}\right)^{-1}\sum_{j=1}^{n}\mathbf{x_{j,n}}\psi(Z_{j,n}/S) + \mathbf{R_n},$$

kde

$$\gamma_1 = S^{-1}\int_{-\infty}^{\infty}(\psi_a'(z/S) + \psi_k'(z/S))dF(z) \tag{2.18}$$

a $||\mathbf{R_n}|| = O_p(n^{-1})$, *ak* $\psi_s \equiv 0$ *a* $||\mathbf{R_n}|| = O_p(n^{-3/4})$ *inak. Navyše, ak* $\sigma^2 = \int_{-\infty}^{\infty}\psi^2(x/S)dF(z) < \infty$, *potom* $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta})$ *má asymptoticky normálne rozdelenie*

$$N_p(\mathbf{0},(\sigma^2/\hat{\gamma}_1^2)\mathbf{D}^{-1}).$$

Dôkaz: vyplýva z viet 5.5.1 - 5.5.3 v [9], str 218–219.

Huberova ψ-funkcia má tvar:

$$\psi_H(x) = \begin{cases} x & : |x| \le k \\ k\,\mathrm{sign}(x) & : |x| \ge k \end{cases}$$

$$\rho_H(x) = \begin{cases} \frac{1}{2}x^2 & : |x| \le k \\ k|x| - \frac{1}{2}k^2 & : |x| \ge k \end{cases}$$

V tomto prípade teda $\psi_a \equiv 0$, $\psi_k = \psi_H$ a $\psi_s \equiv 0$. Ďalej

$$\gamma_1 = \frac{1}{S}(2G_0(kS) - 1),$$

kde G_0 opäť označuje distribučnú funkciu $c_0e_{1,n},\ldots,c_0e_{n,n}$.
Overíme predpoklady vety:

X1 splnené - v modeli uvažujeme absolútny člen (viď úvod kapitoly 2),

X3 splnené - predpokladáme v úvode tejto kapitoly,

M1 ako štatistiku S_n použijeme

$$\hat{c}_n = \widehat{\boldsymbol{\beta}_n}\left(\frac{3}{4}\right) - \widehat{\boldsymbol{\beta}_n}\left(\frac{1}{4}\right),$$

ktorá je škálovo a regresne invariantná a platí pre ňu podmienka (B.2), kde

$$S(F) = F^{-1}\left(\frac{3}{4}\right) - F^{-1}\left(\frac{1}{4}\right),$$

M2 integrál nadobúda pre Huberovu ψ-funkciu a F_0 symetrickú iba nezáporné hodnoty, minimum sa nadobúda len v bode $t = 0$,

M3 triviálne splnené, keďže v našom prípade $\psi_a \equiv 0$ a teda oba integrály sú nulové,

M4 funkcia $\psi'_k(x)$ je rovná 0 pre $|x| > k$ a inak je rovná 1.

Ak budeme navyše predpokladať existenciu $g_0(z)$ príslušnej k G_0 a že regresory spĺňajú $\frac{1}{n}\sum_{i=1}^{n} x_{ij,n}^4 = O(1)$ pre $n \to \infty$ pre $j = 1, \ldots, p$, môžeme pre Huberovu ψ-funkciu použiť uvedenú vetu. Podľa nej potom dostávame nasledujúcu asymptotickú reprezentáciu

$$\frac{1}{n}\sum_{j=1}^{n} \mathbf{x_{j,n}^T}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}) = \frac{S}{\sqrt{n}(2G_0(kS) - 1)}\sum_{j=1}^{n} \psi_H(e_{j,n}) + \frac{1}{n}\sum_{j=1}^{n} \mathbf{x_{j,n}^T} O_p(n^{-1}).$$

Výraz

$$\frac{1}{n}\sum_{j=1}^{n} \mathbf{x_{j,n}^T} O_P(n^{-1})$$

vďaka podmienke (C.1) konverguje v pravdepodobnosti k nule pre $n \to \infty$. Nakoniec tak máme splnenú podmienku (B.1) v tvare

$$\frac{1}{n}\sum_{j=1}^{n} \mathbf{x_{j,n}^T}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}) = \frac{S}{\sqrt{n}(2G_0(kS) - 1)}\sum_{j=1}^{n} \psi_H(e_{j,n}) + o_P(1),$$

čiže

$$\psi_\beta(x) = \psi_H(x), \qquad d_\beta = \frac{S}{2G_0(kS) - 1},$$

kde $\mathsf{E}\psi_H(e_{1,n}) = 0$, $\mathsf{E}\psi_H^2(e_{1,n}) < \infty$, funkcia ψ_H je nepárna, merateľná.

2.4.2 Asymptotické rozdelenie $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ a $\hat{c}_n$ pri ich súčasnom odhadovaní

Opäť budeme najprv uvažovať model

$$\mathbf{Y_n} = \mathbf{X_n}\boldsymbol{\beta} + \mathbf{Z_n},$$

pre ktorý uvedieme značenie a vyslovíme vetu, ktoré napokon aplikujeme na model (2.1) a získame asymptotickú reprezentáciu ako v podmienke (B.2).
M-odhad získame ako argument minima funkcie

$$Q(\boldsymbol{\beta},\sigma) = \sum_{j=1}^{n} \rho\left(\frac{Y_{j,n} - \mathbf{X_n}\boldsymbol{\beta}}{\sigma}\right)\sigma + A_n\sigma,$$

teda riešenie sústavy rovníc (2.12).

Nech $\eta < 1 - A\nu^{-1}$, $A > 0$, kde η je najväčší skok v distribučnej funkcii $Z_{1,n}, \ldots, Z_{n,n}$, $\nu = \min\{\chi(-\infty),\chi(\infty)\}$ a $A = \lim_{n\to\infty} A_n/n$. Potom

$$\mathsf{E}[\psi((\mathbf{Z_n} - \mu)/\sigma),\chi((\mathbf{Z_n} - \mu)/\sigma) - A] = 0$$

má riešenie (μ^*,σ^*), kde $\sigma^* > 0$. Označme $\boldsymbol{\theta}^* = (\boldsymbol{\beta},\sigma^*)$,

$$\mathbf{U}_j(\boldsymbol{\theta}) = [\mathbf{x_{j,n}}\psi(Y_{j,n} - \mathbf{x_{j,n}^T}\boldsymbol{\beta})/\sigma,\chi(Y_{j,n} - \mathbf{x_{j,n}^T}\boldsymbol{\beta})/\sigma - n^{-1}A_n],$$

$$\mathbf{B}_n(\boldsymbol{\theta}) = \frac{1}{n}\mathsf{E}\partial^2/\partial\boldsymbol{\theta}^T\partial\boldsymbol{\theta}Q(\boldsymbol{\theta}),$$

$$\mathbf{T}_n(\boldsymbol{\theta}) = -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{U}_j(\boldsymbol{\theta}),$$

$$\mathbf{V}_n(\boldsymbol{\theta}) = \mathsf{var}\mathbf{T}_n(\boldsymbol{\theta}).$$

Ďalej predpokladáme, že platí

$$\mathsf{E}(\mathbf{Z_n}/\sigma^*) = \mathbf{0}. \tag{2.19}$$

Veta 6. *Nech*

- $\lim_{n\to\infty}\frac{1}{n}A_n - A = 0$ *pre* $\frac{1}{n}A_n > 0$, $A < \nu$, *kde* $\nu = \min\{\chi(\infty),\chi(-\infty)\}$ *a* $\sigma^* > 0$,
- $x^2\psi'(x)$ *a* $x^3\psi''(x)$ *sú ohraničené funkcie,*
- *vlastné čísla matice* $\frac{1}{n}\mathbf{X_n^T}\mathbf{X_n}$ *sú kladné a odrazené od nuly,* $\frac{1}{n}\max\{||\mathbf{x_{j,n}}||^2 : 1 \le j \le n\} \to 0$,
- *a,* $ab - c^2$, *u,* $uv - w^2 > 0$, *kde* $a = \mathsf{E}\psi'(\eta) > 0$, $b = \mathsf{E}x^2\psi'(\eta)$, $u = \mathsf{var}\psi(\eta)$, $v = \mathsf{var}\chi(\eta)$ *a* $w = \mathsf{cov}(\psi(\eta),\chi(\eta))$ *pre* $\eta = \mathbf{Z_n}/\sigma^*$.

Potom

$$\mathbf{T}_n(\boldsymbol{\theta}^* + \frac{1}{\sqrt{n}}\gamma) = \mathbf{T}_n(\boldsymbol{\theta}^*) + \gamma\mathbf{B}_n(\boldsymbol{\theta}) + p_n(\gamma), \tag{2.20}$$

kde

$$\sup\{||p_n(\gamma)|| : ||\gamma|| < K\} \to 0 \qquad \textit{pre } n \to \infty \textit{ pre ľub. } K > 0.$$

Dôkaz: uvedený v [10].

Vrátime sa k modelu (2.1). Za predpokladu $\mathsf{E}e_{1,n}^2 = 1$ máme v našom prípade

$$\boldsymbol{\theta} = (\boldsymbol{\beta},c),$$

položme

$$\gamma = \sqrt{n}(\boldsymbol{\theta_n} - \boldsymbol{\theta}^*),$$

potom pri splnení predpokladov podľa tvrdenia vety 6 máme

$$\underbrace{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}[\psi((Y_{j,n}-\hat{\boldsymbol{\beta}}_{\boldsymbol{n}})/\hat{c}_n),\chi((Y_{j,n}-\hat{\boldsymbol{\beta}}_{\boldsymbol{n}})/\hat{c}_n)-A]=}_{=0}$$

$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}[\mathbf{x_{j,n}}\psi((Y_{j,n}-\boldsymbol{\beta_0})/c^*),\chi((Y_{j,n}-\boldsymbol{\beta_0})/c^*)-A]+\sqrt{n}(\boldsymbol{\theta}^*-\boldsymbol{\theta_n})\mathbf{B}_n(\boldsymbol{\theta}^*)+p_n(\boldsymbol{\theta}^*-\boldsymbol{\theta_n}), \tag{2.21}$$

kde

$$\sup\{||p_n(\boldsymbol{\theta}^*-\boldsymbol{\theta_n})||:||(\boldsymbol{\theta}^*-\boldsymbol{\theta_n})||<K\}\to 0 \qquad \text{pre } n\to\infty \text{ pre ľub. } K>0$$

a $(\boldsymbol{\beta}^*,c^*)$ sú riešením

$$\mathsf{E}[\psi((\mathbf{Y_n}-\mathbf{X_n}\boldsymbol{\beta})/c),\chi((\mathbf{Y_n}-\mathbf{X_n}\boldsymbol{\beta})/c)-A]=0$$

a položíme $\boldsymbol{\theta}^*=(\boldsymbol{\beta}_0,c^*)$.
Pre Huberov odhad máme

$$\psi_H(x)=\begin{cases} x & :|x|\le k\\ k\,\mathrm{sign}(x) & :|x|\ge k\end{cases}$$

$$\chi_H(x)=\begin{cases} x^2 & :|x|\le k\\ k|x| & :|x|\ge k\end{cases}$$

a A_n volíme podľa (2.13).
Matica $\mathbf{B}_n(\boldsymbol{\theta}^*)$ má pre $e_{j,n}<k$ tvar

$$\begin{pmatrix} -\frac{1}{nc^*}\mathbf{X_n}\mathbf{X_n^T} & \mathbf{0}\\ \mathbf{0} & -\frac{2c_0^2}{c^{*3}}\mathsf{var}e_{j,n}\end{pmatrix}$$

pre $e_{j,n}>k$

$$\begin{pmatrix} \frac{kf_0(0)}{nc^*}\mathbf{X_n}\mathbf{X_n^T} & \mathbf{0}\\ \mathbf{0} & -\frac{kc_0}{c^{*2}}\mathsf{E}|e_{1,n}|\end{pmatrix}.$$

Teda po dosadení do (2.21) získame

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{x_{j,n}^T}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}-\boldsymbol{\beta_0})=\frac{c^*}{\sqrt{n}}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbf{x_{j,n}^T}(\mathbf{X_n}\mathbf{X_n^T})^{-1}\mathbf{x_{k,n}}e_{k,n}I[e_{k,n}<k]+$$

$$+\frac{c^*}{f_0(0)\sqrt{n}}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbf{x_{j,n}^T}(\mathbf{X_n}\mathbf{X_n^T})^{-1}\mathbf{x_{k,n}}\,\mathrm{sign}(e_{k,n})I[e_{k,n}>k]+o_P(1)=$$

$$=\frac{c^*}{\sqrt{n}}\sum_{j=1}^{n}e_{j,n}I[e_{j,n}<k]+\frac{c^*}{f_0(0)\sqrt{n}}\sum_{j=1}^{n}\mathrm{sign}(e_{j,n})I[e_{j,n}>k]+o_P(1) \tag{2.22}$$

a

$$\sqrt{n}(\hat{c}_n - c^*) = \frac{c^{*3}}{2c_0^2 \mathsf{var} e_{1,n}\sqrt{n}} \sum_{j=1}^{n}((e_{j,n})^2 - A)I[e_{j,n} < k]+$$

$$+\frac{c^{*2}}{kc_0 \mathsf{E}|e_{1,n}|\sqrt{n}} \sum_{j=1}^{n}(k|e_{j,n}| - A)I[e_{j,n} > k] + o_P(1). \tag{2.23}$$

Pretože však z tvaru A vyplýva, že $c^* \xrightarrow[n\to\infty]{P} c_0$, získame splnenie podmienok (B.1) a (B.2) pre M-odhad s použitím Huberovej funkcie v nasledujúcej forme

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \mathbf{x}_{\mathbf{j,n}}^{\mathbf{T}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}_{\mathbf{0}}) = \frac{c_0}{\sqrt{n}} \sum_{j=1}^{n} e_{j,n} I[e_{j,n} < k]+$$

$$+\frac{c_0}{f_0(0)\sqrt{n}} \sum_{j=1}^{n} \operatorname{sign}(e_{j,n}) I[e_{j,n} > k] + o_P(1),$$

$$\sqrt{n}(\hat{c}_n - c_0) = \frac{c_0}{2\mathsf{var} e_{1,n}\sqrt{n}} \sum_{j=1}^{n}((e_{j,n})^2 - A)I[e_{j,n} < k]+$$

$$+\frac{c_0}{\mathsf{E}|e_{1,n}|\sqrt{n}} \sum_{j=1}^{n}(k|e_{j,n}| - A)I[e_{j,n} > k] + o_P(1).$$

Keďže odhad metódou najmenších štvorcov je špeciálnym prípadom Huberovho M-odhadu pre $k = \infty$, môžeme sa pozrieť, či asymptotické výsledky získané pomocou vety 6 sú rovnaké ako výsledky odvodené v kapitole 2.2 priamou úpravou $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ a $\hat{c}_n$. Zrejme $A = 1$ (z predpokladu $\mathsf{E}e_{1,n}^2 = 1$) a teda dosadením do vzťahov (2.22) a (2.23) dostávame

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \mathbf{x}_{\mathbf{j,n}}^{\mathbf{T}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} - \boldsymbol{\beta}_{\mathbf{0}}) = \frac{c_0}{\sqrt{n}} \sum_{j=1}^{n} e_{j,n} + o_P(1),$$

$$\sqrt{n}(\hat{c}_n - c_0) = \frac{c_0}{2\sqrt{n}} \sum_{j=1}^{n}((e_{j,n})^2 - 1) + o_P(1),$$

teda výledky totožné s tými, ktoré sme získali skôr.

Odhady, ktorých asymptotické výsledky (okrem studentizovaného M-odhadu) sme skúmali v tejto kapitole, využijeme aj v simuláciách, kde pomocou nich odhadneme parametre $\boldsymbol{\beta}$ a c a porovnáme silu jednotlivých testov pri ich použití.

Kapitola 3

Limitné rozdelenie testovej štatistiky

V tejto kapitole vyšetríme asymptotické vlastnosti testovej štatistiky $T_{n,w}$. Vety o limitnom rozdelení testovej štatistiky za nulovej hypotézy a pri lokálnej alternatíve sú prevzaté z článku [7].

3.1 Správanie sa za hypotézy

Pripomeňme tvar testovej štatistiky

$$T_{n,w} = n \int_{-\infty}^{\infty} |\varphi_n(t) - \varphi_0(t)|^2 w(t) \mathrm{d}t.$$

Odvoďme ešte iné vyjadrenie testovej štatistiky $T_{n,w}$, ktoré je názornejšie a použije sa napríklad v dôkaze vety o limitnom správaní sa testovej štatistiky za hypotézy, ktorú uvedieme neskôr. Keďže F_0 je symetrická, φ_0 je reálna, párna a váhová funkcia $w(t)$ je párna, platí

$$\int_{-\infty}^{\infty} \cos(t\hat{e}_{j,n}) \sin(t\hat{e}_{k,n}) \varphi_0(t) w(t) \mathrm{d}t = 0,$$

$$\int_{-\infty}^{\infty} \sin(t\hat{e}_{j,n}) \varphi_0(t) w(t) \mathrm{d}t = 0$$

a teda tvar testovej štatistiky môžeme upraviť do nasledujúcej podoby

$$n \int_{-\infty}^{\infty} |\varphi_n(t) - \varphi_0(t)|^2 w(t) \mathrm{d}t =$$

$$
\begin{aligned}
&= n\int_{-\infty}^{\infty}\left[\left(\frac{1}{n}\sum_{j=1}^{n}\cos(t\hat{e}_{j,n})-\varphi_0(t)\right)^2+\left(\frac{1}{n}\sum_{j=1}^{n}\sin(t\hat{e}_{j,n})\right)^2+\right.\\
&\quad\left.+\sum_{j=1}^{n}\sum_{k=1}^{n}\sin(t\hat{e}_{j,n})\cos(t\hat{e}_{k,n})-2\varphi_0(t)\sum_{j=1}^{n}\sin(t\hat{e}_{j,n})\right]w(t)\mathrm{d}t=\\
&= \int_{-\infty}^{\infty}\Bigg(\underbrace{\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\sin(t\hat{e}_{j,n})+\cos(t\hat{e}_{j,n}))-\sqrt{n}\varphi_0(t)}_{Z_n(t)}\Bigg)^2 w(t)\mathrm{d}t.
\end{aligned}
$$

Takže priestore $\mathcal{H}$ môžeme písať

$$T_{n,w}=||Z_n||^2.$$

Veta 7. *Nech $Y_{1,n},\ldots,Y_{n,n}$ spĺňajú model (2.1) pre $e_{j,n}$ spĺňajúce (A.1). Nech sú splnené predpoklady (2.3), (2.4), (A.2), (B.1), (B.2), (C.1). Nech váhová funkcia w je nezáporná symetrická funkcia, pre ktorú platí $\int t^2 w(t)dt<\infty$. Potom*

$$T_{n,\omega}=||W_n||^2+o_P(1), n\to\infty, \tag{3.1}$$

kde

$$W_n(t)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}W_{nj}(t)$$

a

$$W_{nj}(t):=W_n(t,e_{j,n})=(\sin(te_{j,n})+\cos(te_{j,n})-C_0(t))-tC_0(t)d_\beta\psi_\beta(e_{j,n})-tC_0'(t)d_c\psi_c(e_{j,n})$$

pre $j=1,\ldots,n, t\in\mathbb{R}$.
Navyše pre $n\to\infty$ existuje centrovaný gaussovský proces $\mathcal{W}=\{\mathcal{W}(t);t\in\mathbb{R}\}$ taký, že

$$W_n\xrightarrow{\mathcal{D}}\mathcal{W}$$

a

$$T_{n,\omega}\xrightarrow{\mathcal{D}}||\mathcal{W}||^2.$$

Dôkaz: Uvedený v [7].

Vďaka nulovosti strednej hodnoty funkcií ψ_β a ψ_c dostávame tvar autokovariančnej funkcie procesu $\mathcal{W}$, teda

$$
\begin{aligned}
R(s,t) &= \mathsf{E}(\mathcal{W}(s)-\mathsf{E}\mathcal{W}(s))(\mathcal{W}(t)-\mathsf{E}\mathcal{W}(t))\\
&= \varphi_0(s-t)-\varphi_0(s)\varphi_0(t),
\end{aligned}
$$

ktorú použijeme pre výpočet strednej hodnoty a rozptylu $||\mathcal{W}||^2$, kde

$$\mathsf{E}||\mathcal{W}||^2 = \int_{-\infty}^{\infty} R(t,t)w(t)\mathrm{d}t$$

a

$$\mathsf{var}||\mathcal{W}||^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(s,t)w(s)w(t)\mathrm{d}t.$$

Toto vyjadrenie strednej hodnoty a rozptylu $||\mathcal{W}||^2$ je veľmi užitočné, pretože hoci rozdelenie $||\mathcal{W}||^2$ je rovnaké ako rozdelenie

$$\sum_{j=1}^{\infty} \lambda_j N_j^2,$$

kde N_j, $j = 1,1,\ldots$ sú nezávislé rovnako rozdelené náhodné veličiny s normálnym rozdelením, vo väčšine prípadov je nemožné vypočítať hodnoty λ_j, $j = 1,2,\ldots$ (viď [7]).

3.2 Správanie sa za lokálnej alternatívy

Pre úplnosť uvedieme aj limitné správanie sa testovej štatistiky pri lokálnej alternatíve. Zameráme sa na prípad, kedy F_0 je symetrická a absolútne spojitá s hustotou f_0. Uvažujeme túto alternatívu

$$H_n : e_{1,n},\ldots,e_{n,n} \text{ sú i.i.d. s hustotou } g_n,$$

kde

$$g_n = \left(1 + \frac{\kappa}{\sqrt{n}}u(x)\right) f_0(x), \qquad x \in \mathbb{R}, \tag{3.2}$$

$k \neq 0$ a u je merateľná funkcia taká, že

$$\int u(x)f_0(x)\mathrm{d}x = 0, \qquad 0 < \int u^2(x)f_0(x)\mathrm{d}x < \infty. \tag{3.3}$$

Veta 8. *Nech $Y_{1,n},\ldots,Y_{n,n}$ spĺňajú model (2.1) pre $e_{j,n}$ i.i.d. s hustotou g_n (viď (3.2)), kde platí (3.3). Nech sú splnené predpoklady (2.3), (2.4), (A.2), (B.1), (B.2) a (C.1). Potom existuje centrovaný gaussovský proces $\mathcal{W} = \{\mathcal{W}(t); t \in \mathbb{R}\}$ taký, že pre $n \to \infty$*

$$\{W_n(t); t \in \mathbb{R}\} \xrightarrow{\mathcal{D}} \{\mathcal{W}(t) + \kappa\mu(t); t \in \mathbb{R}\}$$

a

$$||W_n||^2 \xrightarrow{\mathcal{D}} \int_{-\infty}^{\infty} (\mathcal{W}(t) + \kappa\mu(t))^2 w(t)\mathrm{d}t.$$

Proces $\{\mathcal{W}(t); t \in \mathbb{R}\}$ je definovaný ako vo vete 7 a

$$\mu(t) = \int (\cos(tx) + \sin(tx) - t\varphi_0(t)d_\beta\psi_\beta(x) - t\varphi_0'(t)d_c\psi_c(x))\, u(x)f_0(x)\mathrm{d}x.$$

Dôkaz: Dôkaz nebudeme uvádzať, keďže ide o náročnú problematiku presahujúcu teoretické zázemie tejto práce. Pre viac informácií viď [7].

Je zrejmé, že pre rastúce κ a $||\mu||^2 \neq 0$ nenáhodná zložka výrazu $T_{n,\omega}$ (teda $\kappa||\mu||^2$) preváži nad náhodnou zložkou a konverguje k nekonečnu. Preto je tento test konzistentný pre $n \to \infty$.

3.3 Správanie sa za pevnej alternatívy

V tejto podkapitole odvodíme správanie testovej štatistiky pri pevnej alternatíve

$$H_1 : F = F_1$$

so zodpovedajúcou charakteristickou funkciou φ_1 (môže byť aj komplexná) pre odhad metódou najmenších štvorcov a pre odhad pomocou regresných kvantilov v prípade $c = 1$.

Veta 9. *Nech pozorovania $Y_{1,n}, \ldots, Y_{n,n}$ spĺňajú model $Y_{j,n} = \mathbf{x_{j,n}}\boldsymbol{\beta} + e_{j,n}$, kde $e_{1,n}, \ldots, e_{n,n}$ sú nezávislé náhodné veličiny s distribučnou funkciou F_1 a charakteristickou funkciou k nej príslušnej φ_1. Nech platia podmienky (C.1), $\mathsf{E}e_{j,n} = 0$, $\mathsf{E}|e_{j,n}|^4 < \infty$ a navyše ešte*

$$\max\nolimits_{1\leq j\leq n} \mathbf{x_{j,n}^T}(\mathbf{X_n^T X_n})^{-1}\mathbf{x_{j,n}} = \max\nolimits_{1\leq j\leq n} h_{jj,n} \xrightarrow[n\to\infty]{} 0,$$

$$\frac{1}{n}(\mathbf{X_n^T X_n}) = \mathbf{D_n} \xrightarrow[n\to\infty]{} \mathbf{D},$$

kde $\mathbf{D}$ je pozitívne definitná matica. Nech φ_0 je charakteristická funkcia príslušná distribučnej funkcii F_0 z hypotézy (2.2). Nech $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ je odhad parametra $\boldsymbol{\beta}$ metódou najmenších štvorcov. Potom

$$n^{-1}T_{n,w} \xrightarrow[n\to\infty]{P} \int_{-\infty}^{\infty} |\varphi_1(t) - \varphi_0(t)|^2 w(t)\,dt.$$

Dôkaz: Najprv si upravíme tvar testovej štatistiky

$$\begin{aligned}
n^{-1}T_{n,w} &= \int_{-\infty}^{\infty} |\varphi_n(t) - \varphi_0(t)|^2 w(t)\mathrm{d}t \qquad (3.4)\\
&= \int_{-\infty}^{\infty} \left|\frac{1}{n}\sum_{j=1}^{n}\exp(it\hat{e}_{j,n}) - \varphi_0(t)\right|^2 w(t)\mathrm{d}t\\
&= \int_{-\infty}^{\infty} \left|\frac{1}{n}\sum_{j=1}^{n}\exp(it\hat{e}_{j,n}) - \frac{1}{n}\sum_{j=1}^{n}\exp(ite_{j,n}) + \frac{1}{n}\sum_{j=1}^{n}\exp(ite_{j,n}) - \varphi_1(t) + \right.\\
&\quad \left. + \varphi_1(t) - \varphi_0(t)\right|^2 w(t)\mathrm{d}t = \\
&= \int_{-\infty}^{\infty} |A_n(t) + B_n(t) + C(t)|^2 w(t)\mathrm{d}t, \qquad (3.5)
\end{aligned}$$

kde sme si označili

$$\begin{aligned}
A_n(t) &= \frac{1}{n}\sum_{j=1}^{n}\exp(it\hat{e}_{j,n}) - \frac{1}{n}\sum_{j=1}^{n}\exp(ite_{j,n}) \\
B_n(t) &= \frac{1}{n}\sum_{j=1}^{n}\exp(ite_{j,n}) - \varphi_1(t) \\
C(t) &= \varphi_1(t) - \varphi_0(t).
\end{aligned}$$

Integrand upravíme

$$\begin{aligned}
|A_n(t) + B_n(t) + C(t)|^2 &= (A_n(t) + B_n(t) + C(t))(\overline{A_n(t) + B_n(t) + C(t)}) \\
&= |A_n(t)|^2 + |B_n(t)|^2 + |C(t)|^2 + \\
&\quad +2\mathrm{Re}[A_n(t)]\mathrm{Re}[C(t)] + 2\mathrm{Re}[B_n(t)]\mathrm{Re}[C(t)] + \\
&\quad +2\mathrm{Re}[A_n(t)]\mathrm{Re}[B_n(t)] + 2\mathrm{Im}[A_n(t)]\mathrm{Im}[B_n(t)] + \\
&\quad +2\mathrm{Im}[B_n(t)]\mathrm{Im}[C(t)] + 2\mathrm{Im}[A_n(t)]\mathrm{Im}[C(t)]. \qquad (3.6)
\end{aligned}$$

Keďže platí

$$\mathrm{Re}[X]\mathrm{Re}[Y] \leq \sqrt{|\mathrm{Re}[X]|^2|\mathrm{Re}[Y]|^2},$$

potom z Hölderovej nerovnosti dostávame

$$\int \sqrt{|\mathrm{Re}[X]|^2|\mathrm{Re}[Y]|^2}w(t)\mathrm{d}t \leq \sqrt{\int |\mathrm{Re}[X]|^2 w(t)\mathrm{d}t \int |\mathrm{Re}[X]|^2 w(t)\mathrm{d}t},$$

analogicky aj pre imaginárnu časť.
Potom dosadením do (3.4) dostávame nasledujúcu nerovnosť

$$\int_{-\infty}^{\infty} |A_n(t) + B_n(t) + C(t)|^2 \leq$$

$$
\begin{aligned}
\leq\ & \int_{-\infty}^{\infty}|A_n(t)|^2w(t)\mathrm{d}t + \int_{-\infty}^{\infty}|B_n(t)|^2w(t)\mathrm{d}t + \int_{-\infty}^{\infty}|C(t)|^2w(t)\mathrm{d}t + \\
+\ & \sqrt{\int_{-\infty}^{\infty}|\mathrm{Re}[A_n(t)]|^2w(t)\mathrm{d}t}\sqrt{\int_{-\infty}^{\infty}|\mathrm{Re}[B_n(t)]|^2w(t)\mathrm{d}t} + \\
+\ & \sqrt{\int_{-\infty}^{\infty}|\mathrm{Im}[A_n(t)]|^2w(t)\mathrm{d}t}\sqrt{\int_{-\infty}^{\infty}|\mathrm{Im}[B_n(t)]|^2w(t)\mathrm{d}t} + \\
+\ & \sqrt{\int_{-\infty}^{\infty}|\mathrm{Re}[B_n(t)]|^2w(t)\mathrm{d}t}\sqrt{\int_{-\infty}^{\infty}|\mathrm{Re}[C(t)]|^2w(t)\mathrm{d}t} + \\
+\ & \sqrt{\int_{-\infty}^{\infty}|\mathrm{Im}[B_n(t)]|^2w(t)\mathrm{d}t}\sqrt{\int_{-\infty}^{\infty}|\mathrm{Im}[C(t)]|^2w(t)\mathrm{d}t} + \\
+\ & \sqrt{\int_{-\infty}^{\infty}|\mathrm{Re}[A_n(t)]|^2w(t)\mathrm{d}t}\sqrt{\int_{-\infty}^{\infty}|\mathrm{Re}[C(t)]|^2w(t)\mathrm{d}t} + \\
+\ & \sqrt{\int_{-\infty}^{\infty}|\mathrm{Im}[A_n(t)]|^2w(t)\mathrm{d}t}\sqrt{\int_{-\infty}^{\infty}|\mathrm{Im}[C(t)]|^2w(t)\mathrm{d}t}. \qquad (3.7)
\end{aligned}
$$

Naším cieľom je postupne upraviť jednotlivé sčítance v uvedenej nerovnosti. Začneme úpravou výrazu $|A_n(t)|^2$

$$
\begin{aligned}
|A_n(t)|^2 &= \left|\frac{1}{n}\sum_{j=1}^{n}\exp(it\hat{e}_{j,n}) - \frac{1}{n}\sum_{j=1}^{n}\exp(ite_{j,n})\right|^2 = \\
&= \frac{1}{n^2}\left|\sum_{j=1}^{n}(\cos(t\hat{e}_{j,n}) + i\sin(t\hat{e}_{j,n})) - \sum_{j=1}^{n}(\cos(te_{j,n}) + i\sin(te_{j,n}))\right|^2 = \\
&= \frac{1}{n^2}\left|\sum_{j=1}^{n}(\cos(t\hat{e}_{j,n}) - \cos(te_{j,n})) + i\sum_{j=1}^{n}(\sin(t\hat{e}_{j,n}) - \sin(te_{j,n}))\right|^2 = \\
&= \frac{4}{n^2}\left(\sum_{j=1}^{n}\sin\left(\frac{t}{2}(\hat{e}_{j,n} - e_{j,n})\right)\sin\left(\frac{t}{2}(\hat{e}_{j,n} + e_{j,n})\right)\right)^2 + \\
&\quad + \frac{4}{n^2}\left(\sum_{j=1}^{n}\sin\left(\frac{t}{2}(\hat{e}_{j,n} - e_{j,n})\right)\cos\left(\frac{t}{2}(\hat{e}_{j,n} + e_{j,n})\right)\right)^2 \leq
\end{aligned}
$$

$$\begin{aligned}
&\leq \frac{4}{n^2}\sum_{j=1}^{n}\sin^2\left(\frac{t}{2}(\hat{e}_{j,n}-e_{j,n})\right)\sum_{j=1}^{n}\sin^2\left(\frac{t}{2}(\hat{e}_{j,n}+e_{j,n})\right)+ && (3.8)\\
&\quad +\frac{4}{n^2}\sum_{j=1}^{n}\sin^2\left(\frac{t}{2}(\hat{e}_{j,n}-e_{j,n})\right)\sum_{j=1}^{n}\cos^2\left(\frac{t}{2}(\hat{e}_{j,n}+e_{j,n})\right)= && (3.9)\\
&= \frac{4}{n}\sum_{j=1}^{n}\sin^2\left(\frac{t}{2}(\hat{e}_{j,n}-e_{j,n})\right)\\
&\leq \frac{t^2}{n}\sum_{j=1}^{n}(\hat{e}_{j,n}-e_{j,n})^2 && (3.10)
\end{aligned}$$

Platí
$$\hat{e}_{j,n}-e_{j,n}=\mathbf{x}_{\mathbf{j,n}}^{\mathbf{T}}(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}).$$

Keďže $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ je odhad metódou najmenších štvorcov, teda $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}=(\mathbf{X_n^TX_n})^{-\mathbf{1}}\mathbf{X_n^TY_n}$, $\mathsf{var}\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}=\sigma^2(\mathbf{X_n^TX_n})^{-1}$, ďalej máme

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}-\boldsymbol{\beta_0} &= (\mathbf{X_n^TX_n})^{-1}(\mathbf{X_n^TY_n})-\boldsymbol{\beta_0}\\
&= (\mathbf{X_n^TX_n})^{-1}(\mathbf{X_n^TX_n}\boldsymbol{\beta_0}+\mathbf{X_n^Te_n})-\boldsymbol{\beta_0}\\
&= (\mathbf{X_n^TX_n})^{-1}\mathbf{X_n^Te_n}.
\end{aligned}$$

Platí, že
$$P\left(\frac{1}{n}\sum_{j=1}^{n}(\hat{e}_{j,n}-e_{j,n})^2\geq 0\right)=1.$$

Pre výraz
$$\frac{1}{n}\sum_{j=1}^{n}(\hat{e}_{j,n}-e_{j,n})^2$$

použijeme Čebyševovu nerovnosť (viď [4], str. 26). Vypočítame jeho strednú hodnotu

$$\begin{aligned}
\mathsf{E}\frac{1}{n}\sum_{j=1}^{n}(\hat{e}_{j,n}-e_{j,n})^2 &= \mathsf{E}\frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_{\mathbf{j,n}}^{\mathbf{T}}(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}))^2=\\
&= \frac{1}{n}\sum_{j=1}^{n}\sigma^2\mathbf{x}_{\mathbf{j,n}}^{\mathbf{T}}\mathsf{var}\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}\mathbf{x}_{\mathbf{j,n}}=\\
&= \frac{\sigma^2}{n}\sum_{j=1}^{n}\mathrm{tr}(\mathbf{x}_{\mathbf{j,n}}^{\mathbf{T}}(\mathbf{X_n^TX_n})^{-1}\mathbf{x}_{\mathbf{j,n}})\\
&= \sigma^2\frac{p}{n}\xrightarrow[n\to\infty]{}0. && (3.11)
\end{aligned}$$

Potom s použitím tvrdenia (ii) vety 2.10 v [4] pre $r = 1$ dostávame $\forall\varepsilon > 0$

$$P\left(\frac{1}{n}\sum_{j=1}^{n}(\hat{e}_{j,n} - e_{j,n})^2 \geq \varepsilon\right) \leq \frac{\mathsf{E}\frac{1}{n}\sum_{j=1}^{n}(\hat{e}_{j,n} - e_{j,n})^2}{\varepsilon},$$

a pretože platí (3.11), dostávame

$$\frac{1}{n}\sum_{j=1}^{n}(\hat{e}_{j,n} - e_{j,n})^2 \xrightarrow[n\to\infty]{P} 0.$$

Dosadením získame konvergenciu výrazu $\int_{-\infty}^{\infty}|A_n(t)|^2 w(t)\mathrm{d}t$, pretože platí

$$\begin{aligned} \int_{-\infty}^{\infty}|A_n(t)|^2 w(t)\mathrm{d}t &\leq \frac{1}{n}\int_{-\infty}^{\infty}\sum_{j=1}^{n}(\hat{e}_{j,n} - e_{j,n})^2 t^2 w(t)\mathrm{d}t = \\ &= \underbrace{\frac{1}{n}\sum_{j=1}^{n}(\hat{e}_{j,n} - e_{j,n})^2}_{\xrightarrow[n\to\infty]{P} 0}\underbrace{\int_{-\infty}^{\infty}t^2 w(t)\mathrm{d}t}_{<\infty} \xrightarrow[n\to\infty]{P} 0. \end{aligned} \tag{3.12}$$

Keďže môžeme napísať

$$\int_{-\infty}^{\infty}|A_n(t)|^2 w(t)\mathrm{d}t = \int_{-\infty}^{\infty}(\underbrace{(\mathrm{Re}[A_n(t)])^2}_{\geq 0} + \underbrace{(\mathrm{Im}[A_n(t)])^2}_{\geq 0})w(t)\mathrm{d}t \xrightarrow[n\to\infty]{P} 0,$$

zo vzťahu (3.12) potom vyplýva aj

$$\int_{-\infty}^{\infty}(\mathrm{Re}[A_n(t)])^2 w(t)dt \xrightarrow[n\to\infty]{P} 0, \tag{3.13}$$

$$\int_{-\infty}^{\infty}(\mathrm{Im}[A_n(t)])^2 w(t)dt \xrightarrow[n\to\infty]{P} 0. \tag{3.14}$$

Pozrieme na sa ďalšie výrazy z (3.7). Podľa silného zákona veľkých čísel pre pevné $t \in \mathbb{R}$

$$\frac{1}{n}\sum_{j=1}^{n}\exp(ite_{j,n}) \xrightarrow[n\to\infty]{} \varphi_1(t) \qquad \forall t \in \mathbb{R} \qquad \text{skoro iste.}$$

Je zrejmé, že existuje $K > 0$, že

$$|B_n(t)| \leq K \qquad \forall t \in \mathbb{R} \qquad \forall n$$

a funkcia $B_n(t)$ je spojitá na $\mathbb{R}$ $\forall n$. Pretože predpokladáme $\int_{-\infty}^{\infty} t^2 w(t)\mathrm{d}t$, tak

$$\forall \varepsilon > 0 \qquad \exists A_\varepsilon \qquad \int_{|t|>A_\varepsilon} t^2 w(t)\mathrm{d}t < \varepsilon,$$

čiže aj

$$\forall \varepsilon > 0 \qquad \exists B_\varepsilon \qquad \int_{|t|>B_\varepsilon} |B_n(t)|^2 w(t)\mathrm{d}t < \varepsilon.$$

Teda pre nejaké pevné $\varepsilon > 0$ nám stačí skúmať

$$\int_{-B_\varepsilon}^{B_\varepsilon} |B_n(t)|^2 w(t)\mathrm{d}t.$$

Pretože však vieme, že funkcia $B_n(t)$ je spojitá, ohraničená a zaujíma nás na uzavretom konečnom intervale, potom podľa [3] (veta 7.3) dostávame aj

$$\int_{-\infty}^{\infty} |B_n(t)|^2 w(t)\mathrm{d}t \xrightarrow[n\to\infty]{P} 0. \tag{3.15}$$

Ďalej sa pozrieme na strednú hodnotu výrazu $\int_{-\infty}^{\infty} (\mathrm{Re}[B_n(t)])^2 w(t)\mathrm{d}t$

$$\begin{aligned}
&\mathsf{E}\int_{-\infty}^{\infty} (\mathrm{Re}[B_n(t)])^2 w(t)\mathrm{d}t = \mathsf{E}\int_{-\infty}^{\infty} \left(\frac{1}{n}\sum_{j=1}^{n}(\cos(te_{j,n}) - \mathsf{E}\cos(te_{j,n}))\right)^2 w(t)\mathrm{d}t = \\
&= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty} \underbrace{\mathsf{E}((\cos(te_{j,n}) - \mathsf{E}\cos(te_{j,n}))(\cos(te_{k,n}) - \mathsf{E}\cos(te_{k,n})))}_{=0} w(t)\mathrm{d}t + \\
&+ \frac{1}{n^2}\sum_{j=1}^{n}\int_{-\infty}^{\infty} \underbrace{\mathsf{E}(\cos(te_{j,n}) - \mathsf{E}\cos(te_{j,n}))^2}_{<\infty} w(t)\mathrm{d}t \xrightarrow[n\to\infty]{} 0.
\end{aligned} \tag{3.16}$$

Z toho potom vyplýva

$$\int_{-\infty}^{\infty} (\mathrm{Re}[B_n(t)])^2 w(t)\mathrm{d}t \xrightarrow[n\to\infty]{P} 0.$$

Analogicky aj

$$\int_{-\infty}^{\infty} (\mathrm{Im}[B_n(t)])^2 w(t)\mathrm{d}t \xrightarrow[n\to\infty]{P} 0.$$

A nakoniec výraz

$$\int_{-\infty}^{\infty} |C(t)|^2 w(t)\mathrm{d}t = \int_{-\infty}^{\infty} |\varphi_1(t) - \varphi_0(t)|^2 w(t)\mathrm{d}t$$

je rovný konštante. Teda máme

$$\underbrace{\sqrt{\int_{-\infty}^{\infty}|\mathrm{Re}[B_n(t)]|^2 w(t)\mathrm{d}t}}_{\xrightarrow[n\to\infty]{P}0}\underbrace{\sqrt{\int_{-\infty}^{\infty}|\mathrm{Re}[C(t)]|^2 w(t)\mathrm{d}t}}_{=\text{ konštanta}}\xrightarrow[n\to\infty]{P}0,$$

$$\underbrace{\sqrt{\int_{-\infty}^{\infty}|\mathrm{Im}[B_n(t)]|^2 w(t)\mathrm{d}t}}_{\xrightarrow[n\to\infty]{P}0}\underbrace{\sqrt{\int_{-\infty}^{\infty}|\mathrm{Im}[C(t)]|^2 w(t)\mathrm{d}t}}_{=\text{ konštanta}}\xrightarrow[n\to\infty]{P}0.$$

Nakoniec dosadením (3.15), (3.16), (3.12), (3.13), (3.14) do (3.7) tak dostávame limitný tvar testovej štatistiky pri pevnej alternatíve pri použití odhadu metódou najmenších štvorcov

$$n^{-1}T_{n,w}\xrightarrow[n\to\infty]{P}\int_{-\infty}^{\infty}|\varphi_1(t)-\varphi_0(t)|^2 w(t)\mathrm{d}t. \qquad \square$$

Veta 10. *Nech pozorovania $Y_{1,n},\ldots,Y_{n,n}$ spĺňajú model $Y_{j,n}=\mathbf{x_{j,n}}\boldsymbol{\beta}+e_{j,n}$, kde $e_{1,n},\ldots,e_{n,n}$ sú nezávislé náhodné veličiny s distribučnou funkciou F_1, kde F_1 je symetrická okolo 0 a charakteristickou funkciou k nej príslušnej φ_1. Nech*

$$\max_{1\le i\le n}\sqrt{h_{ii,n}}(\log n)^{1/2}\to 0 \textit{ pre } n\to\infty,$$

$$\frac{1}{n}(\mathbf{X_n^T X_n})=\mathbf{D_n}\xrightarrow[n\to\infty]{}\mathbf{D},$$

kde $\mathbf{D}$ je pozitívne definitná matica a nech platí (C.1). Nech φ_0 je charakteristická funkcia príslušná distribučnej funkcii F_0 z hypotézy (2.2). Nech $\hat{\boldsymbol{\beta}}_{\mathbf{n}}$ je odhad parametra $\boldsymbol{\beta}$ pomocou regresných kvantilov. Potom

$$n^{-1}T_{n,w}\xrightarrow[n\to\infty]{P}\int_{-\infty}^{\infty}|\varphi_1(t)-\varphi_0(t)|^2 w(t)\,dt.$$

Dôkaz: Dôkaz tejto vety sa od dôkazu predchádzajúcej vety bude líšiť len úpravou výrazu $|A_n(t)|^2$. Použijeme vetu 3, ktorej predpoklady sú splnené, a dostávame

$$\begin{aligned}
\sum_{j=1}^{n}(\hat{e}_{j,n}-e_{j,n})^2 &= \sum_{j=1}^{n}(\mathbf{x_{j,n}^T}(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}_{\mathbf{n}}))^2\\
&= \sum_{j=1}^{n}(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}_{\mathbf{n}})^T\mathbf{x_{j,n}}\mathbf{x_{j,n}^T}(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}_{\mathbf{n}})\\
&= \frac{1}{4f(0)^2}((\mathbf{X_n^T X_n})^{-1/2}\sum_{k=1}^{n}\mathbf{x_{k,n}}\,\mathrm{sign}\,e_{k,n}+o_P(1))^T\\
&\quad ((\mathbf{X_n^T X_n})^{-1/2}\sum_{l=1}^{n}\mathbf{x_{l,n}}\,\mathrm{sign}\,e_{l,n}+o_P(1))\\
&= O_P(1),
\end{aligned}$$

pretože

$$\mathsf{E}(\mathbf{X_n^T X_n})^{-1/2}\sum_{k=1}^{n}\mathbf{x_{k,n}}\,\mathrm{sign}\,e_{k,n}=\mathbf{0}$$

a

$$\mathsf{var}(\mathbf{X_n^T X_n})^{-1/2}\sum_{k=1}^{n}\mathbf{x_{k,n}}\,\mathrm{sign}\,e_{k,n}=(\mathbf{X_n^T X_n})^{-1/2}\sum_{k=1}^{n}\mathbf{x_{k,n}}\mathsf{var}\,\mathrm{sign}\,e_{k,n}\mathbf{x_{k,n}^T}(\mathbf{X_n^T X_n})^{-1/2}=$$

$$=\mathsf{var}\,\mathrm{sign}\,e_{k,n}\mathbf{I}_p,$$

kde $\mathbf{I}_p$ je jednotková matica rádu p. Teda po vydelení n získame

$$\frac{1}{n}\sum_{j=1}^{n}(\hat{e}_{j,n}-e_{j,n})^2 \xrightarrow[n\to\infty]{P} 0. \tag{3.17}$$

Čiže rovnako ako pre odhad metódou najmenších štvorcov máme

$$\int_{-\infty}^{\infty}|A_n(t)|^2 w(t)\mathrm{d}t \xrightarrow[n\to\infty]{P} 0,$$

z čoho nám opäť vyplýva

$$\int_{-\infty}^{\infty}(\mathrm{Re}[A_n(t)])^2 w(t)dt \quad \xrightarrow[n\to\infty]{P} \quad 0, \tag{3.18}$$

$$\int_{-\infty}^{\infty}(\mathrm{Im}[A_n(t)])^2 w(t)dt \quad \xrightarrow[n\to\infty]{P} \quad 0. \tag{3.19}$$

Dosadením (3.15), (3.16), (3.17), (3.18), (3.19) do (3.7) pre L_1 odhad tiež dostávame

$$n^{-1}T_{n,w} \xrightarrow[n\to\infty]{P} \int_{-\infty}^{\infty}|\varphi_1(t)-\varphi_0(t)|^2 w(t)\mathrm{d}t. \qquad \square$$

Kapitola 4

Simulácie

V tejto kapitole aplikujeme navrhnutý test na simulované dáta, porovnáme niekoľko volieb odhadov parametrov $\boldsymbol{\beta}$ a c a testové štatistiky pre rôzne váhové funkcie. V testoch budeme porovnávať tieto rozdelenia: normálne rozdelenie N(0, 1), Laplaceovo rozdelenie Lap(0, 1), logistické rozdelenie Log(0, 1) a Studentovo t-rozdelenie so 4 stupňami voľnosti. Všetky numerické výpočty a simulácie boli vykonané v štatistickom programe `R`.

Na určenie kritických hodnôt sme použili parametrický bootstrap. Popíšeme si ho v niekoľkých krokoch:

1. Zvolíme parametre $\boldsymbol{\beta}$ a c. Vygenerujeme jednu sadu dát $Y_{1,n}, \dots, Y_{n,n}$, kde

$$Y_{j,n} = \mathbf{X_n}\boldsymbol{\beta} + ce_{j,n}, \quad j = 1, \dots, n$$

   kde $e_{j,n}$, $j = 1, \dots, n$ je výber z nejakého rozdelenia (použili sme rozdelenia N(0, 1), Lap(0, 1), Log(0, 1), t_4). Z tohto modelu odhadneme parametre $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ a $\hat{c}_n$ a vypočítame hodnotu testovej štatistiky.

2. Vygenerujeme náhodný výber $e^*_{1,n}, \dots, e^*_{n,n}$ z hypotetického rozdelenia, teda s distribučnou funkciou F_0. Vytvoríme pozorovania $Y^*_{1,n}, \dots, Y^*_{n,n}$, kde

$$Y^*_{j,n} = \mathbf{X_n}\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} + \hat{c}_n e^*_{j,n}, \quad j = 1, \dots, n.$$

   Odhadneme parametre $\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}$ a $\hat{c}_n$ a vypočítame hodnotu testovej štatistiky T^*. Tento krok opakujeme N-krát.

3. Z hodnôt testovej štatistiky $T^*_1, \dots, T^*_N$ vyberieme $(1-\alpha)$ kvantil, ktorý bude aproximovať kritickú hodnotu pre testovú štatistiku.

Zvolili sme $N = 2000$ a všetky testy sme vykonali na hladine $\alpha = 0.05$. Hodnoty parametra a sme zvolili nasledujúce: $\frac{1}{4}$, $\frac{1}{2}$, 1, 2, 3 a π. Jedine pre Studentovo t_4 rozdelenie sme nepoužili hodnotu $a = 2$, keďže v tomto prípade boli hodnoty testovej štatistiky niekedy nedefinované. Pred generovaním náhodného výberu z hypoetického rozdelenia $e^*_{1,n}, \ldots, e^*_{n,n}$ sme si vždy nastavili `set.seed(9)`, teda z daného rozdelenia bol vždy generovaný rovnaký náhodný výber. Pretože naše odhady spĺňajú (2.3) a (2.4), získané kritické hodnoty sú rovnaké bez ohľadu na rozdelenie $e_{j,n}$, $j = 1, \ldots, n$. Pri testovaní sme vždy použili rovnaké náhodné výbery pomocou nastavenia `set.seed(9)`.

Parametre $\boldsymbol{\beta}$ a c sme odhadovali pomocou metódy najmenších štvorcov (v tabuľkách označenie L_2), pomocou M-odhadu s použitím Huberovej funkcie (v tabuľkách M) a pomocou regresných kvantilov. V prípade odhadu pomocou regresných kvantilov sme použili dva rôzne odhady parametra c: pomocou regresných kvantilov – viď (2.10) (v tabuľkách L_1) a pomocou reziduálneho súčtu štvorcov (v tabuľkách L_1^*), teda

$$\frac{1}{n-p}\sum_{j=1}^{n}(Y_{j,n} - \mathbf{x}_{\mathbf{j},\mathbf{n}}^{\mathbf{T}}\hat{\boldsymbol{\beta}}_{\boldsymbol{n}})^2. \tag{4.1}$$

Pretože nie všetky uvažované rozdelenia majú rozptyl jednotkový a žiadne nemá jednotkové medzikvartilové rozpätie, odhady parametra c sme pred výpočtom odhadu reziduí ešte vydelili smerodajnou odchýlkou (prípadne medzikvartilovým rozpätím) hypotetického rozdelenia.

Použili sme nasledujúce váhové funkcie

$$\begin{aligned} w_1(t) &= e^{-at^2}, \\ w_2(t) &= e^{-a|t|}, \\ w_3(t) &= (1+t^2)e^{-at^2}, \\ w_4(t) &= (1+t^2)e^{-a|t|}. \end{aligned}$$

Pre každé rozdelenie uvedieme aspoň jeden konkrétny tvar testovej štatistiky a potom výsledky testov. Najprv však ešte popíšeme použitú maticu regresorov $\mathbf{X_n}$ a voľby parametrov použité v simuláciách.

V simuláciách použijeme maticu $\mathbf{X_n}$ s riadkami

$$\mathbf{x}_{\mathbf{j},\mathbf{n}}^{\mathbf{T}} = \left(1, \frac{j}{n}, \frac{j^2}{n^2}\right),$$

ktorá spĺňa podmienku (C.1), teda

$$\lim_{n\to\infty} \max_{1\le j\le n, 1\le v\le p} |x_{jv,n}| n^{-\frac{1}{2}} = 0$$

a platí, že

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_{j,n}\mathbf{x}_{j,n}^{\mathrm{T}} = \mathbf{D},$$

kde

$$\mathbf{D} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}$$

je pozitívne definitná matica.

Pri aproximácii kritických hodnôt sme volili $\boldsymbol{\beta} = (2{,}3{,}1)^T$ a $c = 1$, avšak vďaka vlastnostiam použitých odhadov, vypočítané kritické hodnoty nezávisia na týchto hodnotách parametrov. Pri testovaní sme volili $\boldsymbol{\beta} = (1{,}2{,}3)^T$ a $c = 1$.

V nasledujúcich kapitolách si vždy uvedieme najprv tvar testovej štatistiky pre danú voľbu váhovej funkcie – tá vždy závisí aj na voľbe parametra a, ktorý ovplyvňuje konvergenciu váhovej funkcie k nule, preto testovú štatistiku budeme označovať $T_{n,a}$. Ďalej uvedieme tabuľku kritických hodnôt a nakoniec aj tabuľky, v ktorých sú uvedené pravdepodobnosti zamietnutia (v percentách) nulovej hypotézy – v prvom stĺpci tabuľky je uvedené skutočné rozdelenie $e_{1,n}, \dots, e_{n,n}$ a v poslednom stĺpci tabuliek (označenom KS) sú výsledky Kolmogorov–Smirnovovho testu aplikovaného na odhadnuté reziduá.

4.1 Normálne rozdelenie

Testujeme hypotézu

$$H_0 : F \equiv F_0,$$

kde F_0 je distribučná funkcia normálneho rozdelenia N(0, 1) proti alternatíve

$$H_1 : F \neq F_0.$$

Najprv si upravíme tvar testovej štatistiky $T_{n,w}$ dosadením charakteristickej funkcie normálneho rozdelenia $\varphi_0(t) = e^{-\frac{t^2}{2}}$, teda

$$\begin{aligned}
T_{n,w} &= n\int_{-\infty}^{\infty} |\varphi_n(t) - \varphi_0(t)|^2 w(t)dt \\
&= n\int_{-\infty}^{\infty} \left| \frac{1}{n}\sum_{j=1}^{n} e^{it\hat{e}_{j,n}} - e^{-\frac{t^2}{2}} \right|^2 w(t)dt \\
&= n\int_{-\infty}^{\infty} \left| \frac{1}{n}\sum_{j=1}^{n} \cos(t\hat{e}_{j,n}) + \frac{i}{n}\sum_{j=1}^{n} \sin(t\hat{e}_{j,n}) - e^{-\frac{t^2}{2}} \right|^2 w(t)dt \\
&= n\int_{-\infty}^{\infty} \left(\frac{1}{n^2}\left(\sum_{j=1}^{n} \cos(t\hat{e}_{j,n})\right)^2 + \frac{1}{n^2}\left(\sum_{j=1}^{n} \sin(t\hat{e}_{j,n})\right)^2 - \right. \\
&\quad \left. -\frac{2}{n}\sum_{j=1}^{n} \cos(t\hat{e}_{j,n})e^{-\frac{t^2}{2}} + e^{-t^2} \right) w(t)dt \\
&= \int_{-\infty}^{\infty} \left(\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \cos(t(\hat{e}_{j,n} - \hat{e}_{k,n})) - 2\sum_{j=1}^{n} cos(t\hat{e}_{j,n})e^{-\frac{t^2}{2}} + ne^{-t^2} \right) w(t)dt
\end{aligned}$$

$$= \int_{-\infty}^{\infty} \left(\frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \mathrm{Re}[\ e^{it\hat{e}_{j,n}-\hat{e}_{k,n}}] - 2 \sum_{j=1}^{n} \mathrm{Re}[\ e^{it\hat{e}_{j,n}-\frac{t^2}{2}}] + ne^{-\frac{t^2}{2}} \right) w(t)dt.$$

Váhová funkcia $w_1(t) = e^{-at^2}$

Pre váhovú funkciu $w_1(t) = e^{-at^2}$ dostaneme tvar testovej štatistiky

$$T_{n,a} = \frac{1}{n}\sqrt{\frac{\pi}{a}} \left(\sum_{j,k=1}^{n} e^{-(\hat{e}_{j,n}-\hat{e}_{k,n})^2/4a} \right) + n\sqrt{\frac{\pi}{1+a}} - 2\sqrt{\frac{2\pi}{1+2a}} \left(\sum_{j=1}^{n} e^{-\hat{e}_{j,n}^2/(2+4a)} \right).$$

Pomocou parametrického bootstrapu sme získali túto aproximáciu kritických hodnôt:

	$a = \frac{1}{4}$	$a = \frac{1}{2}$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$
L_2	2,333	0,906	0,259	0,053	0,019	0,017
M	1,809	0,612	0,180	0,063	0,033	0,030
L_1	41.668	35.321	27.653	19.161	14.155	13.610
L_1^*	2.623	1.120	0.518	0.275	0.185	0.177

Tabuľka 4.1: Aproximácia kritických hodnôt, váhová funkcia $w_1(t) = e^{-at^2}$

Test pri použití odhadov parametra $\boldsymbol{\beta}$ a c založený na regresných kvantiloch (tabuľka 4.4) najmenej reagoval na zmenu rozdelenia (zároveň aj kritické hodnoty sa výrazne líšili od kritických hodnôt pre iné odhady). Zvyšné použité odhady nám dali podobné výsledky – pri normálnej hypotéze bolo najťažšie odlíšiť logistické rozdelenie. Avšak zatiaľ čo v prípade metódy najmenších štvorcov a Huberovho odhadu s rastúcim parametrom a takmer vždy rástla aj pravdepodobnosť zamietnutia, ak sme použili L_1 odhad parametra $\boldsymbol{\beta}$ a c sme odhadli pomocou reziduálneho súčtu štvorcov, to bolo presne naopak. Najvyššie pravdepodobnosti zamietnutia boli v prípade použitia odhadu metódou najmenších štvorcov.

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0, 1)	5,15	4,85	4,70	5,50	5,15	5,15	0,05
Lap(0, 1)	52,10	53,00	50,60	48,90	46,95	46,75	2,35
Log(0, 1)	17,60	18,45	20,60	21,80	21,00	21,10	0,25
t_4	40,30	44,00	46,00	46,75	46,15	46,20	2,95

Tabuľka 4.2: Pravdepodobnosti zamietnutia (v %) - metóda najmenších štvorcov, váhová funkcia $w_1(t) = e^{-at^2}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0, 1)	4,40	5,40	5,75	4,80	4,50	4,50	0,00
Lap(0, 1)	28,50	35,55	39,25	38,50	41,55	42,25	0,00
Log(0, 1)	8,25	10,90	11,60	10,05	11,10	11,30	0,00
t_4	13,55	20,75	26,60	26,55	30,50	31,10	0,00

Tabuľka 4.3: Pravdepodobnosti zamietnutia (v %) - M-odhad, Huberova ψ funkcia, váhová funkcia $w_1(t) = e^{-at^2}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	3.95	3.95	4.05	3.90	4.00	4.00	29.40
Lap(0,1)	5.55	5.65	5.70	5.90	6.35	6.35	31.85
Log(0,1)	4.95	5.00	5.00	5.25	5.45	5.45	28.80
t_4	4.65	4.65	4.85	5.00	5.10	5,10	29,25

Tabuľka 4.4: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou regresných kvantilov, váhová funkcia $w_1(t) = e^{-at^2}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	4,60	4,70	5,40	6,05	6,45	6,40	0,30
Lap(0,1)	56,45	52,75	29,25	5,95	2,85	2,65	5,05
Log(0,1)	18,30	17,50	10,45	4,60	3,95	3,90	1,20
t_4	43,45	43,80	31,50	11,65	5,85	5,35	5,25

Tabuľka 4.5: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou reziduálneho súčtu štvorcov, váhová funkcia $w_1(t) = e^{-at^2}$

Váhová funkcia $w_3(t) = (1+t^2)e^{-at^2}$

Pre váhovú funkciu $w_3(t) = (1+t^2)e^{-at^2}$ dostaneme tvar testovej štatistiky

$$\begin{aligned} T_{n,a} &= \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} -\frac{e^{-\frac{(\hat{e}_{j,n}-\hat{e}_{k,n})^2}{4a}}\sqrt{\pi}\left((\hat{e}_{j,n}-\hat{e}_{k,n})^2 - 2a(1+2a)\right)}{4a^{5/2}} \\ &\quad -2\sum_{j=1}^{n} \frac{e^{-\frac{\hat{e}_{j,n}^2}{2+4a}}\sqrt{2\pi}\left(2-\hat{e}_{j,n}^2+6a+4a^2\right)}{(1+2a)^{5/2}} \\ &\quad +n\frac{\sqrt{\pi}(3+2a)}{2(1+a)^{3/2}}. \end{aligned}$$

Získané kritické hodnoty (pre odhad $\boldsymbol{\beta}$ i c pomocou regresných kvantilov sú kritické hodnoty opäť výrazne iné):

	$a=\frac{1}{4}$	$a=\frac{1}{2}$	$a=1$	$a=2$	$a=3$	$a=\pi$
L_2	14,463	3,799	0,799	0,119	0,035	0,030
M	11,970	2,762	0,511	0,113	0,051	0,047
L_1	83,851	56,081	39,733	25,240	18,081	17,342
L_1^*	15,365	4,072	1,064	0,404	0,248	0,234

Tabuľka 4.6: Aproximácia kritických hodnôt, váhová funkcia $w_3(t) = (1+t^2)e^{-at^2}$

Výsledky získané na základe testov pri použití váhovej funkcie $w_3(t) = (1+t^2)e^{-at^2}$ sú podobné ako výsledky získané pri použití $w_1(t) = e^{-at^2}$. Použitie metódy najmenších štvorcov nám opäť dávalo najvyššie pravdepodobnosti zamietnutia a použitie odhadu pomocou regresných kvantilov sa zasa ukázalo ako najmenej vhodné zo skúmaných odhadov. Takisto ak $e_{1,n}, \ldots, e_{n,n}$ pochádzali z logistického rozdelenia, pravdepodobnosť zamietnutia nulovej hypotézy bola z uvažovaných rozdelení najmenšia – teda rovnako ako pri použití váhovej funkcie $w_1(t)$.

	$a=0{,}25$	$a=0{,}5$	$a=1$	$a=2$	$a=3$	$a=\pi$	KS
N(0,1)	4,30	5,25	4,60	5,05	5,30	5,35	0,05
Lap(0,1)	44,25	52,05	51,40	49,75	48,10	48,05	2,35
Log(0,1)	12,35	17,85	19,50	21,85	21,30	21,20	0,25
t_4	32,50	40,90	45,15	46,75	46,70	46,70	2,95

Tabuľka 4.7: Pravdepodobnosti zamietnutia (v %) - metóda najmenších štvorcov, váhová funkcia $w_3(t) = (1+t^2)e^{-at^2}$

	$a=0{,}25$	$a=0{,}5$	$a=1$	$a=2$	$a=3$	$a=\pi$	KS
N(0,1)	4,10	5,10	5,05	4,75	4,60	4,60	0,00
Lap(0,1)	19,85	29,50	36,20	36,15	39,25	39,75	0,00
Log(0,1)	6,65	9,10	11,45	9,35	10,30	10,35	0,00
t_4	9,35	15,05	22,85	24,15	27,40	28,15	0,00

Tabuľka 4.8: Pravdepodobnosti zamietnutia (v %) - M-odhad, Huberova ψ funkcia, váhová funkcia $w_3(t) = (1+t^2)e^{-at^2}$

	$a=0{,}25$	$a=0{,}5$	$a=1$	$a=2$	$a=3$	$a=\pi$	KS
N(0,1)	3,25	4,00	4,05	3,95	3,90	3,90	29,40
Lap(0,1)	5,40	5,40	5,65	5,75	6,05	6,10	31,85
Log(0,1)	5,20	5,00	5,00	5,05	5,25	5,25	28,80
t_4	5,00	4,80	4,65	4,90	5,05	5,05	29,25

Tabuľka 4.9: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou regresných kvantilov, váhová funkcia $w_3(t) = (1+t^2)e^{-at^2}$

	$a=0{,}25$	$a=0{,}5$	$a=1$	$a=2$	$a=3$	$a=\pi$	KS
N(0,1)	4,90	4,85	4,75	5,95	6,20	6,25	0,30
Lap(0,1)	52,85	57,85	47,75	12,60	3,90	3,60	5,05
Log(0,1)	16,20	19,40	16,30	5,70	4,20	4,05	1,20
t_4	37,90	45,15	41,75	18,80	8,40	7,60	5,25

Tabuľka 4.10: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou reziduálneho súčtu štvorcov, váhová funkcia $w_3(t) = (1+t^2)e^{-at^2}$

Váhová funkcia $w_2(t) = e^{-a|t|}$

Dosadením váhovej funkcie $w_2(t) = e^{-a|t|}$ dostaneme tvar testovej štatistiky

$$
\begin{aligned}
T_{n,a} \;=\; & \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{2a}{(\hat{e}_{j,n}-\hat{e}_{k,n})^2+a^2} \\
& -2\sum_{j=1}^{n}\mathrm{Re}\left(e^{-\frac{1}{2}(j+i\alpha)^2}\sqrt{\frac{\pi}{2}}\left(1+e^{2ij\alpha}-ie^{2ij\alpha}\mathrm{Erfi}\left[\frac{j-i\alpha}{\sqrt{2}}\right]+i\mathrm{Erfi}\left[\frac{j+i\alpha}{\sqrt{2}}\right]\right)\right) \\
& +ne^{\frac{\alpha^2}{4}}\sqrt{\pi}\mathrm{Erfc}\left(\frac{a}{2}\right),
\end{aligned}
$$

kde $\mathrm{Erfi}(x) = \frac{Erf(ix)}{i}$ a $\mathrm{Erfc}(x) = 1 - \mathrm{Erf}(x)$.
Pomocou parametrického bootstrapu sme získali tieto kritické hodnoty:

	$a=\frac{1}{4}$	$a=\frac{1}{2}$	$a=1$	$a=2$	$a=3$	$a=\pi$
L_2	9,776	3,777	1,016	0,149	0,033	0,027
M	9,115	3,306	0,812	0,108	0,029	0,025
L_1	47,670	35,243	24,098	13,292	7,611	7,129
L_1^*	11,165	4,316	1,200	0,259	0,113	0,103

Tabuľka 4.11: Aproximácia kritických hodnôt, váhová funkcia $w_2(t) = e^{-a|t|}$

Aj v tomto prípade sme získali obdobné výsledky ako pri použití predchádzajúcich dvoch váhových funkcií.

Keď zhrnieme test „normálnej hypotézy“, tak v našom prípade sa ako nevhodný ukázal odhad parametrov $\boldsymbol{\beta}$ a c pomocou regresných kvantilov. Zároveň išlo o jediný odhad, pri ktorého použití boli pravdepodobnosti zamietnutia nižšie než pravdepodobnosti zamietnutia pri použití Kolmogorov–Smirnovovho testu. Zvyšné tri odhady, ktoré sme použili, dávali podobné výsledky: najmenej často zamietali nulovú hypotézu, ak skutočné rozdelenie chýb bolo logistické Log(0, 1). Ostatne, podobný výsledok dávali aj simulácie v [7]. Na základe týchto simulácií sa zdá, že na použití váhovej funkcii nezáleží, ale pri použití odhadu metódou najmenších štvorcov je test citlivejší.

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	4,45	4,05	4,50	4,55	4,75	4,85	0,05
Lap(0,1)	37,00	44,55	50,65	52,85	50,50	50,35	2,35
Log(0,1)	9,80	12,55	16,25	18,80	20,50	21,00	0,25
t_4	25,65	33,30	39,10	44,65	46,00	46,25	2,95

Tabuľka 4.12: Pravdepodobnosti zamietnutia (v %) - metóda najmenších štvorcov, váhová funkcia $w_2(t) = e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	4,20	4,35	4,10	5,10	5,45	5,15	0,00
Lap(0,1)	15,05	20,05	27,70	39,20	42,65	42,60	0,00
Log(0,1)	5,50	6,55	7,65	11,10	12,40	12,40	0,00
t_4	6,70	9,70	13,70	24,65	30,40	30,95	0,00

Tabuľka 4.13: Pravdepodobnosti zamietnutia (v %) - M-odhad, Huberova ψ funkcia, váhová funkcia $w_2(t) = e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	2,20	2,40	2,50	2,35	2,50	2,50	29,40
Lap(0,1)	2,70	2,25	2,30	2,35	2,70	2,70	31,85
Log(0,1)	2,55	2,45	2,40	2,55	2,90	2,90	28,80
t_4	2,75	2,25	2,35	2,45	2,55	2,55	29,25

Tabuľka 4.14: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou regresných kvantilov, váhová funkcia $w_2(t) = e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	4,20	4,20	4,55	5,25	5,95	5,95	0,30
Lap(0,1)	44,00	49,75	53,30	37,20	12,75	10,80	5,05
Log(0,1)	11,05	14,10	17,25	12,75	5,65	5,45	1,20
t_4	28,55	35,75	40,95	33,95	18,20	15,75	5,25

Tabuľka 4.15: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou reziduálneho súčtu štvorcov, váhová funkcia $w_2(t) = e^{-a|t|}$

4.2 Laplaceovo rozdelenie

Testujeme hypotézu

$$H_0 : F \equiv F_0,$$

kde F_0 je distribučná funkcia Laplaceovho rozdelenia Lap(0, 1) proti alternatíve

$$H_1 : F \neq F_0.$$

Podobne ako v prípade normálneho rozdelenia, aj tu si upravíme tvar testovej štatistiky $T_{n,w}$ dosadením charakteristickej funkcie Laplaceovho rozdelenia $\varphi_0(t) = \frac{1}{1+t^2}$

$$\begin{aligned} T_{n,w} &= n \int_{-\infty}^{\infty} |\varphi_n(t) - \varphi_0(t)|^2 w(t) dt = \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \mathrm{Re}[\ e^{it\hat{e}_{j,n} - \hat{e}_{k,n}}] - \frac{2}{1+t^2} \sum_{j=1}^{n} \mathrm{Re}[\ e^{it\hat{e}_{j,n}}] + \frac{n}{(1+t^2)^2} \right) w(t) dt. \end{aligned}$$

Váhová funkcia $w_3(t) = (1+t^2)e^{-at^2}$

Pri použití váhovej funkcie $w_3(t) = (1+t^2)e^{-at^2}$ má testová štatistika tvar

$$\begin{aligned} T_{n,a} &= \frac{1}{n}\sqrt{\frac{\pi}{16a^5}} \left(\sum_{j,k=1}^{n} [2a(1+2a) - (\hat{e}_{j,n} - \hat{e}_{k,n})^2 e^{-(\hat{e}_{j,n} - \hat{e}_{k,n})^2/4a}] \right) \\ &\quad + n\pi[1 - \mathrm{Erf}(\sqrt{a})]e^a - 2\sqrt{\frac{\pi}{a}} \left(\sum_{j=1}^{n} e^{-\hat{e}_{j,n}^2/4a} \right), \end{aligned}$$

kde $\mathrm{Erf}(x) = (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt$.

	$a=\frac{1}{4}$	$a=\frac{1}{2}$	$a=1$	$a=2$	$a=3$	$a=\pi$
L_2	21,702	8,146	2,531	0,644	0,266	0,239
M	20,456	6,979	2,029	0,731	0,450	0,424
L_1	71,371	46,006	32,285	22,294	16,939	16,465
L_1^*	15,888	5,751	1,859	0,561	0,299	0,278

Tabuľka 4.16: Aproximácia kritických hodnôt, váhová funkcia $w_3(t) = (1+t^2)e^{-at^2}$

Použitie odhadu pomocou regresných kvantilov sa opäť ukázalo ako najmenej vhodné. V ostatných prípadoch bolo najčastejšie „odhalené“ normálne rozdelenie, avšak pravdepodobnosť zamietnutia nulovej hypotézy klesala s rastúcim parametrom a.

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	27,45	29,95	25,75	9,65	0,95	0,70	1,30
Lap(0,1)	5,70	5,10	4,90	5,10	4,80	4,75	0,85
Log(0,1)	15,40	14,80	11,45	4,05	1,40	1,30	0,60
t_4	9,70	9,70	10,25	9,55	8,85	8,75	0,95

Tabuľka 4.17: Pravdepodobnosti zamietnutia (v %) - metóda najmenších štvorcov, váhová funkcia $w_3(t) = (1 + t^2)e^{-at^2}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	23,65	23,20	9,15	0,30	0,00	0,00	0,35
Lap(0,1)	6,00	5,50	4,90	4,85	5,05	5,15	0,20
Log(0,1)	15,05	13,55	5,55	0,80	0,50	0,50	0,20
t_4	12,35	9,35	5,30	1,75	1,55	1,45	0,05

Tabuľka 4.18: Pravdepodobnosti zamietnutia (v %) - M-odhad, Huberova ψ funkcia, váhová funkcia $w_3(t) = (1 + t^2)e^{-at^2}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	2,95	3,65	3,85	3,65	3,55	3,55	22,35
Lap(0,1)	5,00	5,25	5,45	5,05	5,20	5,20	23,85
Log(0,1)	4,65	4,90	4,85	4,85	4,80	4,80	22,70
t_4	4,65	4,55	4,45	4,30	4,20	4,15	23,20

Tabuľka 4.19: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou regresných kvantilov, váhová funkcia $w_3(t) = (1 + t^2)e^{-at^2}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	21,65	26,95	25,45	14,35	10,15	10,20	0,60
Lap(0,1)	5,85	5,30	4,85	4,80	4,50	4,55	0,40
Log(0,1)	10,05	12,45	10,80	6,45	5,95	6,05	0,40
t_4	7,90	9,35	10,40	10,10	10,35	10,40	0,70

Tabuľka 4.20: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou reziduálneho súčtu štvorcov, váhová funkcia $w_3(t) = (1 + t^2)e^{-at^2}$

Váhová funkcia $w_4(t) = (1+t^2)e^{-a|t|}$

Pre váhovú funkciu $w_4(t) = (1+t^2)e^{-a|t|}$ dostaneme tvar testovej štatistiky

$$\begin{aligned} T_{n,a} &= \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} 2a\left(\frac{\hat{e}_{j,n}^4 - 4j^3\hat{e}_{k,n} + \hat{e}_{k,n}^4 + 2\hat{e}_{k,n}^2\left(-3+a^2\right) + a^2\left(2+a^2\right)}{((\hat{e}_{j,n}-\hat{e}_{k,n})^2+a^2)^3} - \right. \\ &\quad \left. - \frac{4\hat{e}_{j,n}\hat{e}_{k,n}\left(-3+\hat{e}_{k,n}^2+a^2\right) + 2\hat{e}_{j,n}^2\left(-3+3\hat{e}_{k,n}^2+a^2\right)}{((\hat{e}_{j,n}-\hat{e}_{k,n})^2+a^2)^3}\right) \\ &\quad -2\sum_{j=1}^{n}\frac{2a}{\hat{e}_{j,n}^2+a^2} \\ &\quad +n(2\int_0^a \frac{\cos(t)}{t}\mathrm{d}t\sin(a) + \cos(a)(\pi - 2\int_0^a \frac{\sin(t)}{t}\mathrm{d}t)). \end{aligned}$$

V testoch použijeme tieto kritické hodnoty získané pomocou parametrického bootstrapu:

	$a=\frac{1}{4}$	$a=\frac{1}{2}$	$a=1$	$a=2$	$a=3$	$a=\pi$
L_2	399,411	61,158	10,411	1,395	0,339	0,287
M	384,806	60,429	9,697	1,186	0,3301	0,292
L_1	490,811	102,677	37,873	15,801	9,483	8,939
L_1^*	409,491	56,145	8,057	1,012	0,275	0,237

Tabuľka 4.21: Aproximácia kritických hodnôt, váhová funkcia $w_4(t) = (1+t^2)e^{-a|t|}$

Aj pri použití váhovej funkcie $w_4(t) = (1+t^2)e^{-a|t|}$ bolo v najväčšom počte prípadov „odhalené“ normálne rozdelenie. Avšak na rozdiel od predchádzajúceho prípadu, pravdepodobnosť zamietnutia stúpala s rastúcim a. Odhad pomocou regresných kvantilov sa opäť neosvedčil.

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	6,30	16,60	25,05	27,90	20,90	19,10	1,30
Lap(0,1)	4,25	5,75	5,30	4,95	4,40	4,35	0,85
Log(0,1)	5,00	9,65	13,25	13,20	9,90	8,95	0,60
t_4	4,75	9,30	9,75	10,75	10,30	9,95	0,95

Tabuľka 4.22: Pravdepodobnosti zamietnutia (v %) - metóda najmenších štvorcov, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	8,40	15,35	23,05	19,20	1,90	1,05	0,35
Lap(0,1)	6,45	5,85	6,45	5,70	5,45	5,25	0,20
Log(0,1)	7,60	10,75	14,75	11,25	1,80	1,40	0,20
t_4	6,00	9,90	12,40	8,65	2,40	2,00	0,05

Tabuľka 4.23: Pravdepodobnosti zamietnutia (v %) - M-odhad, Huberova ψ funkcia, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	2,20	1,80	3,10	3,65	3,55	3,55	22,35
Lap(0,1)	5,20	5,00	5,20	5,35	5,15	5,15	23,85
Log(0,1)	2,90	3,10	4,55	4,85	4,85	4,80	22,70
t_4	3,35	2,95	4,75	4,40	4,20	4,05	23,20

Tabuľka 4.24: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou regresných kvantilov, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	5,10	7,70	17,85	26,65	22,90	21,85	0,60
Lap(0,1)	6,20	6,20	5,20	5,20	4,80	4,75	0,40
Log(0,1)	3,85	5,95	7,90	12,25	10,25	9,75	0,40
t_4	5,30	6,45	7,70	10,10	10,50	10,45	0,70

Tabuľka 4.25: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou reziduálneho súčtu štvorcov, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

Pri testovaní nulovej hypotézy, že rozdelenie chýb má Laplaceovo rozdelenie Lap(0, 1) sme túto hypotézu najčastejšie zamietali, ak skutočné rozdelenie chýb bolo normálne N(0, 1). Test bol najmä pre vyššie hodnoty parametra a citlivejší na zmenu rozdelenia pri použití metódy najmenších štvorcov a kombinácie

L_1 odhadu a odhadu parametra c pomocou reziduálneho súčtu štvorcov. Hoci pravdepodobnosti zamietnutia boli nižšie než v prípade testovania normálneho rozdelenia, boli výsledky stále lepšie než pri použití Kolmogorov–Smirnovovho testu (ak neberieme do úvahy test pri použití odhadu $\boldsymbol{\beta}$ aj c pomocou regresných kvantilov).

4.3 Logistické rozdelenie

Testujeme hypotézu

$$H_0 : F \equiv F_0,$$

kde F_0 je distribučná funkcia logistického rozdelenia log(0, 1) proti alternatíve

$$H_1 : F \neq F_0.$$

Pre logistické rozdelenie s charakteristickou funkciou $\varphi_0(t) = \frac{\pi t}{\sinh(\pi t)}$ upravíme tvar testovej štatistiky $T_{n,w}$

$$\begin{aligned} T_{n,w} &= n \int_{-\infty}^{\infty} |\varphi_n(t) - \varphi_0(t)|^2 w(t) dt = \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \mathrm{Re}[\ e^{it\hat{e}_{j,n} - \hat{e}_{k,n}}] - \frac{2\pi t}{\sinh(\pi t)} \sum_{j=1}^{n} \mathrm{Re}[\ e^{it\hat{e}_{j,n}}] + \frac{n\pi^2 t^2}{(sinh(\pi t))^2} \right) w(t) dt. \end{aligned}$$

Váhová funkcia $w_2(t) = e^{-a|t|}$

Pre váhovú funkciu $w_2(t) = e^{-a|t|}$ má testová štatistika tvar

$$\begin{aligned} T_{n,a} &= \frac{2a}{n} \left(\sum_{j,k=1}^{n} \frac{1}{a^2 + (\hat{e}_{j,n} - \hat{e}_{k,n})^2} \right) + \frac{n}{\pi} \left[2\xi(2{,}1 + \frac{a}{2\pi}) - \frac{a}{\pi} \xi(3{,}1 + \frac{a}{2\pi}) \right] \\ & -\frac{2}{\pi} \sum_{j=1}^{n} \left[S^{(1)}(a{,}\hat{e}_{j,n}) - \frac{\hat{e}_{j,n}^2 2\pi^2}{S}^{(2)}(a{,}\hat{e}_{j,n}) \right], \end{aligned}$$

kde

$$\xi(a{,}x) = \sum_{k=0}^{\infty} \frac{1}{(k+x)^a}$$

a

$$S^{(m)}(a{,}x) = \sum_{k=0}^{\infty} \left[\left(\frac{x}{2\pi} \right)^2 + \left(\frac{a+\pi}{2\pi} + k \right)^2 \right]^{-m}, \quad m = 1,\ 2.$$

Pri testovaní nulovej hypotézy, že chyby majú logistické rozdelenie, sa najmenej často zamietala nulová hypotéza, ak chyby mali v skutočnosti normálne rozdelenie. Okrem odhadu pomocou regresných kvantilov, ktorý opäť zamietal nulovú hypotézu iba v mále prípadov, pre malé hodnoty parametra a sa veľmi neosvedčil ani M-odhad s použitím Huberovej ψ funkcie.

	$a=\frac{1}{4}$	$a=\frac{1}{2}$	$a=1$	$a=2$	$a=3$	$a=\pi$
L_2	10,542	4,938	1,975	0,525	0,187	0,162
M	9,625	4,263	1,521	0,371	0,130	0,116
L_1	33,692	25,278	18,657	12,430	8,567	8,155
L_1^*	11,419	5,141	1,908	0,514	0,213	0,194

Tabuľka 4.26: Aproximácia kritických hodnôt, váhová funkcia $w_2(t) = e^{-a|t|}$

	$a=0{,}25$	$a=0{,}5$	$a=1$	$a=2$	$a=3$	$a=\pi$	KS
N(0,1)	6,05	6,35	5,65	5,30	3,75	3,60	0,10
Lap(0,1)	17,35	19,55	19,65	21,30	21,30	21,90	1,10
Log(0,1)	5,95	5,90	5,00	4,90	5,20	5,55	0,10
t_4	11,45	13,85	15,20	18,50	21,00	21,80	1,70

Tabuľka 4.27: Pravdepodobnosti zamietnutia (v %) - metóda najmenších štvorcov, váhová funkcia $w_2(t) = e^{-a|t|}$

	$a=0{,}25$	$a=0{,}5$	$a=1$	$a=2$	$a=3$	$a=\pi$	KS
N(0,1)	6,75	6,00	5,50	3,75	2,65	2,20	0,00
Lap(0,1)	9,00	9,55	10,55	14,40	21,35	21,65	0,00
Log(0,1)	6,55	5,90	5,30	4,40	5,60	5,40	0,00
t_4	5,50	6,05	6,30	6,75	10,15	10,20	0,00

Tabuľka 4.28: Pravdepodobnosti zamietnutia (v %) - M-odhad, Huberova ψ funkcia, váhová funkcia $w_2(t) = e^{-a|t|}$

	$a=0{,}25$	$a=0{,}5$	$a=1$	$a=2$	$a=3$	$a=\pi$	KS
N(0,1)	3,95	3,80	3,90	3,75	3,90	3,90	27,20
Lap(0,1)	7,15	6,05	5,50	5,50	5,65	5,70	29,65
Log(0,1)	5,90	5,00	4,90	4,95	5,00	5,00	26,65
t_4	5,50	5,05	4,65	4,40	4,80	4,90	27,15

Tabuľka 4.29: Výsledky testov - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou regresných kvantilov, váhová funkcia $w_2(t) = e^{-a|t|}$

	$a=0{,}25$	$a=0{,}5$	$a=1$	$a=2$	$a=3$	$a=\pi$	KS
N(0,1)	3,90	3,30	2,70	3,05	3,90	3,90	0,30
Lap(0,1)	24,00	26,55	28,20	26,00	19,95	18,60	2,25
Log(0,1)	5,50	5,40	5,50	5,00	5,00	4,95	0,50
t_4	14,05	17,05	19,65	21,50	20,55	19,85	2,20

Tabuľka 4.30: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou reziduálneho súčtu štvorcov, váhová funkcia $w_2(t) = e^{-a|t|}$

Váhová funkcia $w_4(t) = (1+t^2)e^{-a|t|}$

Pre váhovú funkciu $w_4(t) = (1+t^2)e^{-a|t|}$ dostaneme tvar testovej štatistiky

$$\begin{aligned} T_{n,a} &= \frac{2a}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\frac{\hat{e}_{j,n}^4 - 4\hat{e}_{j,n}^3\hat{e}_{k,n} + \hat{e}_{k,n}^4 + 2\hat{e}_{k,n}^2\left(-3+a^2\right) + a^2\left(2+a^2\right)}{\left(\hat{e}_{j,n}^2 - 2\hat{e}_{j,n}\hat{e}_{k,n} + \hat{e}_{k,n}^2 + a^2\right)^3}\right. \\ &\left. + \frac{-4\hat{e}_{j,n}\hat{e}_{k,n}\left(-3+\hat{e}_{k,n}^2+a^2\right) + 2\hat{e}_{j,n}^2\left(-3+3\hat{e}_{k,n}^2+a^2\right)}{\left(\hat{e}_{j,n}^2 - 2\hat{e}_{j,n}\hat{e}_{k,n} + \hat{e}_{k,n}^2 + a^2\right)^3}\right) \\ &- \sum_{j=1}^{n}\sum_{m=0}^{\infty}\left(\frac{8\pi\left(-\hat{e}_{j,n}^2 + (\pi+2m\pi+a)^2\right)}{\left(\hat{e}_{j,n}^2 + (\pi+2m\pi+a)^2\right)^2} + \right. \\ &\left. + \frac{48\pi\left(\hat{e}_{j,n}^4 - 6\hat{e}_{j,n}^2(\pi+2m\pi+a)^2 + (\pi+2m\pi+a)^4\right)}{\left(\hat{e}_{j,n}^2 + (\pi+2m\pi+a)^2\right)^4}\right) \\ &+ n\frac{16\pi^3\zeta\left(2,\frac{a}{2\pi}\right) - 8\pi^2 a\zeta\left(3,\frac{a}{2\pi}\right) + 48\pi\zeta\left(4,\frac{a}{2\pi}\right) - 24a\zeta\left(5,\frac{a}{2\pi}\right)}{8\pi^4} \end{aligned}$$

Získali sme tieto kritické hodnoty:

	$a=\frac{1}{4}$	$a=\frac{1}{2}$	$a=1$	$a=2$	$a=3$	$a=\pi$
L_2	361,504	53,605	8,471	1,363	0,388	0,326
M	355,393	52,449	7,640	1,032	0,272	0,233
L_1	427,454	79,602	29,084	14,844	9,978	9,462
L_1^*	410,042	59,316	9,251	1,288	0,380	0,331

Tabuľka 4.31: Aproximácia kritických hodnôt, váhová funkcia $w_4(t) = (1+t^2)e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	5,95	5,45	6,45	5,70	5,25	5,10	0,10
Lap(0,1)	7,70	9,95	18,30	19,55	21,20	21,25	1,10
Log(0,1)	6,35	5,70	5,85	4,80	4,95	5,00	0,10
t_4	6,00	8,20	12,15	14,95	18,65	19,35	1,70

Tabuľka 4.32: Pravdepodobnosti zamietnutia (v %) - metóda najmenších štvorcov, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	5,90	6,30	6,75	5,35	3,65	3,25	0,00
Lap(0,1)	7,60	6,60	8,75	10,50	15,05	15,65	0,00
Log(0,1)	6,80	5,80	6,45	5,15	4,40	4,55	0,00
t_4	5,60	4,65	6,40	5,95	6,85	6,95	0,00

Tabuľka 4.33: Pravdepodobnosti zamietnutia (v %) - M-odhad, Huberova ψ funkcia, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	4,80	3,60	4,00	3,85	3,80	3,85	27,20
Lap(0,1)	7,55	9,70	7,50	5,45	5,55	5,60	29,65
Log(0,1)	5,35	5,45	6,10	4,90	4,95	4,95	26,65
t_4	5,90	5,90	5,80	4,65	4,50	4,60	27,15

Tabuľka 4.34: Výsledky testov - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou regresných kvantilov, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 2$	$a = 3$	$a = \pi$	KS
N(0,1)	5,80	5,00	3,65	2,95	3,15	3,05	0,30
Lap(0,1)	8,70	15,90	24,00	28,65	25,40	24,95	2,25
Log(0,1)	5,10	6,00	5,20	5,45	5,10	5,15	0,50
t_4	7,35	10,30	14,15	19,80	21,55	21,50	2,20

Tabuľka 4.35: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou reziduálneho súčtu štvorcov, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

Výsledky získané pri voľbe váhovej funkcie $w(t) = (1 + t^2)e^{-a|t|}$ sú podobné ako tie, ktoré sme získali pri váhovej funkcii $w(t) = e^{-a|t|}$.

V prípade nulovej hypotézy, že chyby pochádzajú z logistického rozdelenia, bolo (z uvažovaných rozdelení) najťažšie odlíšiť prípad, keď chyby v skutočnosti mali normálne rozdelenie. K rovnakému výsledku na základe simulácii dospeli aj v článku [7]. Odhad pomocou regresných kvantilov sa neosvedčil ani v tomto prípade, test využívajúci M-odhad s Huberovou ψ funkciou zamietal nulovú hypotézu menej často než zvyšné dva odhady.

4.4 Studentovo t-rozdelenie

Testujeme hypotézu
$$H_0 : F \equiv F_0,$$
kde F_0 je distribučná funkcia Studentovho t rozdelenia so 4 stupňami voľnosti proti alternatíve
$$H_1 : F \neq F_0.$$

Charakteristická funkcia Studentovho t-rozdelenia so 4 stupňami voľnosti má tvar
$$\varphi_0(t) = 2|x|^2\, K_2(2|x|),$$
kde K_m označuje modifikovanú Besselovu funkciu druhého rádu. Po dosadení váhovej funkcie $w_4(t) = (1+t^2)e^{-a|t|}$ dostaneme nasledujúci tvar testovej štatistiky

$$\begin{aligned}
T_{n,a} &= \frac{2a}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(\frac{\hat{e}_{j,n}^4 - 4\hat{e}_{j,n}^3\hat{e}_{k,n} + \hat{e}_{k,n}^4 + 2\hat{e}_{k,n}^2\left(-3+a^2\right) + a^2\left(2+a^2\right)}{\left(\hat{e}_{j,n}^2 - 2\hat{e}_{j,n}\hat{e}_{k,n} + \hat{e}_{k,n}^2 + a^2\right)^3}\right.\\
&- \left.\frac{4\hat{e}_{j,n}\hat{e}_{k,n}\left(-3+\hat{e}_{k,n}^2+a^2\right) + 2j^2\left(-3+3\hat{e}_{k,n}^2+a^2\right)}{\left(\hat{e}_{j,n}^2 - 2\hat{e}_{j,n}\hat{e}_{k,n} + \hat{e}_{k,n}^2 + a^2\right)^3}\right)\\
&\quad -2\sum_{j=1}^{n}\left[\frac{2a}{(a^4 + a^2(\hat{e}_{j,n}^2 - 4) + (\hat{e}_{j,n}^2+4)^2)^4}\right.\\
&\quad +(a^{14} + a^{12}(7\hat{e}_{j,n}^2 - 32) + a^{10}(21\hat{e}_{j,n}^4 - 128\hat{e}_{j,n}^2 + 392) +\\
&\quad +a^8(35\hat{e}_{j,n}^6 - 160\hat{e}_{j,n}^4 + 1096\hat{e}_{j,n}^2 - 3240) +\\
&\quad +(4+\hat{e}_{j,n}^2)^3(\hat{e}_{j,n}^8 + 20\hat{e}_{j,n}^6 - 120\hat{e}_{j,n}^4 + 5176\hat{e}_{j,n}^2 - 3232) +\\
&\quad +a^6(35\hat{e}_{j,n}^8 + 1104\hat{e}_{j,n}^4 + 12064\hat{e}_{j,n}^2 + 21120) +\\
&\quad +a^4(21\hat{e}_{j,n}^{10} + 160\hat{e}_{j,n}^8 + 656\hat{e}_{j,n}^6 - 13104\hat{e}_{j,n}^4 - 114304\hat{e}_{j,n}^2 - 91904) +\\
&\quad +a^2(7\hat{e}_{j,n}^{12} + 128\hat{e}_{j,n}^{10} + 424\hat{e}_{j,n}^8 - 23648\hat{e}_{j,n}^6 - 64640\hat{e}_{j,n}^4 + 175616\hat{e}_{j,n}^2 + 217088))+
\end{aligned}$$

$$
\begin{aligned}
&+\frac{12(36+a^4+4ia^3\hat{e}_{j,n}-22\hat{e}_{j,n}^2+\hat{e}_{j,n}^4+a^2(22-6\hat{e}_{j,n}^2)-4ia\hat{e}_{j,n}(\hat{e}_{j,n}^2-11))\pi}{(4-a^2-2ia\hat{e}_{j,n}+\hat{e}_{j,n}^2)^{9/2}} \\
&+\frac{12(36+a^4-4ia^3\hat{e}_{j,n}-22\hat{e}_{j,n}^2+\hat{e}_{j,n}^4+a^2(22-6\hat{e}_{j,n}^2)+4ia\hat{e}_{j,n}(\hat{e}_{j,n}^2-11))\pi}{(4-a^2+2ia\hat{e}_{j,n}+\hat{e}_{j,n}^2)^{9/2}} \\
&-\frac{\arcsin(\frac{1}{2}(a-i\hat{e}_{j,n}))}{(4-a^2+2ia\hat{e}_{j,n}+\hat{e}_{j,n}^2)^{9/2}} \\
&\quad 24(36+a^4-4ia^3\hat{e}_{j,n}-22\hat{e}_{j,n}^2+\hat{e}_{j,n}^4+a^2(22-6\hat{e}_{j,n}^2)+4ia\hat{e}_{j,n}(\hat{e}_{j,n}^2-11)) \\
&-\frac{\arcsin(\frac{1}{2}(a-i\hat{e}_{j,n}))}{(4-a^2-2ia\hat{e}_{j,n}+\hat{e}_{j,n}^2)^{9/2}} \\
&\quad 24(36+a^4+4ia^3\hat{e}_{j,n}-22\hat{e}_{j,n}^2+\hat{e}_{j,n}^4+a^2(22-6\hat{e}_{j,n}^2)-4ia\hat{e}_{j,n}(\hat{e}_{j,n}^2-11))\Big] \\
+\quad & nc_a,
\end{aligned}
$$

kde

$$c_a = \int_{-\infty}^{\infty} \varphi_0^2(t) w(t) \mathrm{d}t.$$

Najprv uvedieme získané odhadnuté kritické hodnoty, ktoré potom použijeme v testoch:

	$a=\frac{1}{4}$	$a=\frac{1}{2}$	$a=1$	$a=3$	$a=\pi$
L_2	387,417	60,184	10,010	0,393	0,333
M	379,342	57,635	8,912	0,292	0,255
L_1	473,754	99,799	36,690	8,680	8,084
L_1^*	430,740	64,230	9,611	0,347	0,300

Tabuľka 4.36: Aproximácia kritických hodnôt, váhová funkcia $w_4(t) = (1+t^2)e^{-a|t|}$

Pri testovaní hypotézy, že chyby pochádzajú zo Studentovho t rozdelenia so 4 stupňami voľnosti boli pravdepodobnosti zamietnutia nulovej hypotézy veľmi nízke pre všetky voľby odhadov a všetky uvažované rozdelenia.

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 3$	$a = \pi$	KS
N(0,1)	4,90	6,65	12,30	10,05	9,25	0,00
Lap(0,1)	5,45	5,80	5,80	3,00	2,90	0,00
Log(0,1)	4,30	4,30	5,85	3,80	3,50	0,00
t_4	4,60	5,60	5,80	5,55	5,40	0,00

Tabuľka 4.37: Pravdepodobnosti zamietnutia (v %) - metóda najmenších štvorcov, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 3$	$a = \pi$	KS
N(0,1)	6,05	7,05	9,95	2,10	1,50	0,00
Lap(0,1)	6,10	4,70	4,05	11,35	12,30	0,00
Log(0,1)	5,75	5,45	6,50	3,25	2,85	0,00
t_4	4,55	4,70	6,45	5,15	5,20	0,00

Tabuľka 4.38: Pravdepodobnosti zamietnutia (v %) - M-odhad, Huberova ψ funkcia, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 3$	$a = \pi$	KS
N(0,1)	4,20	3,25	3,85	4,25	4,15	0,00
Lap(0,1)	4,20	3,25	3,85	4,25	4,15	0,00
Log(0,1)	4,80	5,20	5,65	5,50	5,50	0,00
t_4	5,90	5,25	5,65	5,20	5,20	0,00

Tabuľka 4.39: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou regresných kvantilov, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

	$a = 0{,}25$	$a = 0{,}5$	$a = 1$	$a = 3$	$a = \pi$	KS
N(0,1)	5,05	3,85	4,50	8,60	8,40	0,00
Lap(0,1)	5,05	3,85	4,50	8,60	8,40	0,00
Log(0,1)	4,60	3,70	3,10	3,15	3,25	0,00
t_4	6,15	6,40	5,65	5,55	5,50	0,00

Tabuľka 4.40: Pravdepodobnosti zamietnutia (v %) - odhad $\boldsymbol{\beta}$ pomocou regresných kvantilov, parameter c odhadnutý pomocou reziduálneho súčtu štvorcov, váhová funkcia $w_4(t) = (1 + t^2)e^{-a|t|}$

Všetky testy pomerne verne kopírovali zvolenú hladinu $\alpha = 0{,}05$. V našom prípade sa neosvedčil odhad parametrov $\boldsymbol{\beta}$ a c pomocou regresných kvantilov, avšak ak sme c odhadli pomocou reziduálneho súčtu štvorcov, výsledky boli omnoho lepšie a nulovú hypotézu zamietali približne v rovnakom počte prípadov ako zvyšné odhady. Takisto sme nezistili veľké rozdiely medzi výsledkami získanými pre rôzne voľby váhovej funkcie.

Ešte zhrnieme prípady, ktoré bolo pri testovaní nulovej hypotézy najťažšie odlíšiť: pri testovaní nulovej hypotézy, že chyby majú normálne rozdelenie, sa hypotéza najmenej často zamietala pre logistické rozdelenie, pre Laplaceovu nulovú hypotézu to bolo v prípade, ak chyby mali v skutočnosti t_4 alebo logistické rozdelenie. Pri logistickej nulovej hypotéze mali najnižšiu pravdepodobnosť zamietnutia výbery z normálneho rozdelenia. A napokon pri testovaní Studentovho t rozdelenia so 4 stupňami voľnosti boli pravdepodobnosti zamietnutia nízke pre všetky uvažované rozdelenia. Kolmogorov–Smirnovov test zamietal nulovú hypotézu oveľa menej často ako nami navrhovaný test (okrem prípadov, keď sme oba parametre $\boldsymbol{\beta}$ a c odhadovali pomocou regresných kvantilov).

Bolo by možno zaujímavé zopakovať testy pre inú voľbu matice $\mathbf{X_n}$, žiaľ, z časových dôvodov to nebolo možné. Z rovnakých dôvodov bol obmedzený aj počet opakovaní na 2000.

Záver

V práci sme sa venovali testu dobrej zhody za prítomnosti rušivých parametrov – v lineárnom modeli $\mathbf{Y_n} = \mathbf{X_n}\boldsymbol{\beta} + c\mathbf{e_n}$, ktorý je založený na vzdialenosti hypotetickej charakteristickej funkcie a empirickej charakteristickej funkcie odhadnutých reziduí. Pretože rozdelenie testovej štatistiky závisí na voľbe odhadov parametrov $\boldsymbol{\beta}$ a c, značná časť práce bola venovaná týmto odhadom a ich asymptotickému správaniu – zaoberali sme sa odhadom metódou najmenších štvorcov, odhadom získaným pomocou regresných kvantilov a M-odhadom s použitím Huberovej ψ funkcie. Ďalej sme sa venovali limitnému rozdeleniu testovej štatistiky, a to nielen za hypotézy či lokálnej alternatívy, ale odvodili sme jej rozdelenie aj za pevnej alternatívy pri použití odhadu metódou najmenších štvorcov a odhadu pomocou regresných kvantilov. Nakoniec sme uvažovaný test použili pre nasimulované dáta. V našej situácii sa neosvedčil odhad parametrov $\boldsymbol{\beta}$ a c pomocou regresných kvantilov, zvyšné testy však zamietali nulovú hypotézu omnoho častejšie než Kolmogorov–Smirnovov test aplikovaný na odhadnuté reziduá.

Literatúra

[1] Anděl J. (2005): Základy matematické statistiky. MFF UK, Praha.

[2] Babu G. J. (1989): Strong representations for LAD estimators in linear models,*Probability Theory and Related Fields* **83**, 547–558.

[3] Billingsley P. (1999): Convergence of probability measures. John Wiley & Sons, New York.

[4] Dupač V., Hušková M. (2003): Pravděpodobnost a matematická statistika. MFF UK, Praha.

[5] Hacking I. (1984): Trial by numbers, *Science* **5**, 69–73.

[6] Huber P. J. (1981): Robust statistics. John Wiley & Sons, New York.

[7] Hušková M., Meintanis S. G. (2007): Omnibus tests for the error distribution in the linear regression model, *Statistics* **41:5**, 363–376.

[8] Jirmanová Z. (2007): Testy založené na empirických charakteristických funkcích, Diplomová práce, Univerzita Karlova v Praze.

[9] Jurečková J., Sen P. K. (1996): Robust statistical procedures: Asymptotics and interrelations, John Wiley & Sons, New York.

[10] Silvapulle M. J. (1985): Asymptotic behaviour of robust estimators of regression and scale parameters with fixed carriers,*Tha Annals of Statistics*, **13:4**, 1490–1497.

[11] Srivastava M. S., Srivastava V. K. (1986): Asymptotic distribution of least squares estimator and a test statistic in linear regression models, *Economic Letters* **21**, 173–176.