NEJMENŠÍ VÁŽENÉ ČTVERCE V PŘÍKLADECH

PAVEL PLÁT[1]

plat@kmlinux.fjfi.cvut.cz

Katedra matematiky, FJFI, ČVUT Praha

SUMMARY

V této numerické studii je na příkladech analyzováno chování metody nejmenších vážených čtverců (LWS) a výsledky porovnány s odhady regresních koeficientů vypočtených klasickou metodou nejmenších čtverců (LS) a metodou nejmenších usekaných čtverců (LTS_h). Výsledky demonstrují některé žádoucích vlastnosti nejmenších vážené čtverců. LWS je odhad s vysokou odolností vůči kontaminaci dat "hrubými" chybami. S kontaminací se dokáže vypořádat stejně dobře jako nejmenší usekané čtverce. Jeho vlastnosti můžeme navíc ovlivnit volbou váhové funkce. Diky možnosti volby spojité váhové funkce pak mají nejmenší vážené čtverce menší podsouborovou citlivost než nejmenší usekané čtverce. Nabízí se zde srovnání s M-odhady se spojitou ψ-funkcí. Oproti nim mají však nejmenší vážené čtverce tu výhodu, že jsou regresně a škálově ekvivariantní, což M-odhady obecně nejsou.

LINEÁRNÍ REGRESNÍ MODEL

$$Y = X\beta + e$$

NEJMENŠÍ VÁŽENÉ ČTVERCE

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r^2_{(i)}(\beta)$$

ASYMPTOTICKÁ NORMALITA LWS

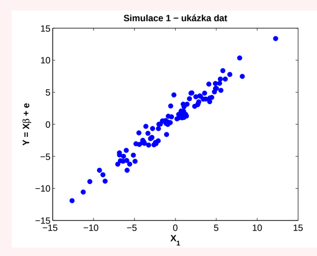
$$\mathcal{L}\left(\sqrt{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta^0\right)\right) \to \mathcal{N}\left(0\,,\,V\left(\hat{\beta}^{(LWS,n,w)}\right)\right)$$

SIMULACE 1 – KONTAMINACE 5%

Počet pozorování $n = 100$, počet opakování v rámci simulace $m = 1000$

$Y = 1 + X_1 + e$, $\mathcal{L}(X_1) = \mathcal{N}(0,5)$, $\mathcal{L}(e_i) = \mathcal{N}(0,1)$, $i = 1, .., 95$, $\mathcal{L}(e_i) = t_1$, $i = 96, .., 100$.

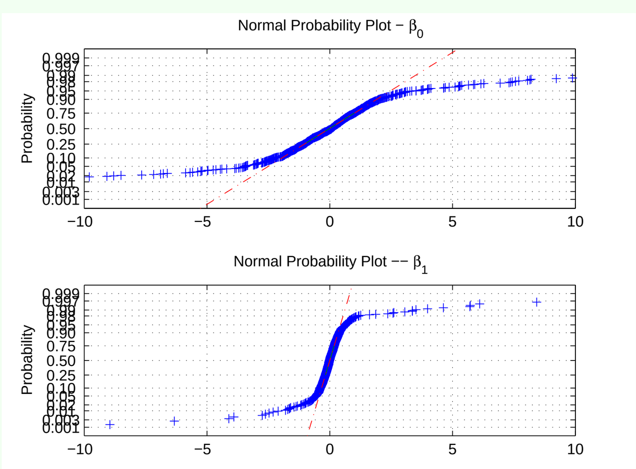
LWS: váhová funkce $w(x) = I\{x \le 0.5\} + (-2x+2) \cdot I\{x > 0.5\}$, LTS: $h = 75$.



LS

$$AVG(\hat{\beta}) = \begin{pmatrix} 1.02 \\ 1.01 \end{pmatrix}$$

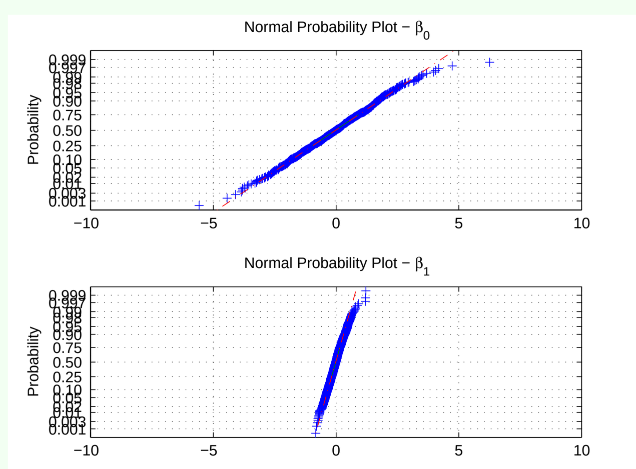
$$\frac{n}{m}\left(\hat{\beta} - \beta^0\right)\left(\hat{\beta} - \beta^0\right)^T = \begin{pmatrix} 220.93 & 11.93 \\ 11.93 & 4.27 \end{pmatrix}$$



LWS

$$AVG(\hat{\beta}) = \begin{pmatrix} 1.00 \\ 1.00 \end{pmatrix}$$

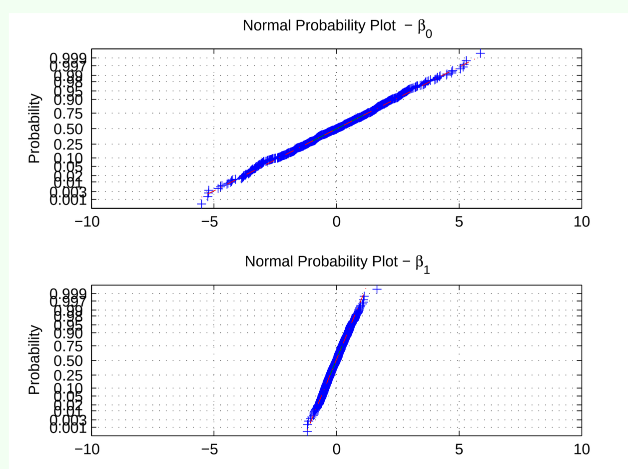
$$\frac{n}{m}\left(\hat{\beta} - \beta^0\right)\left(\hat{\beta} - \beta^0\right)^T = \begin{pmatrix} 2.07 & 0.02 \\ 0.02 & 0.08 \end{pmatrix}$$



LTS

$$AVG(\hat{\beta}) = \begin{pmatrix} 1.00 \\ 1.00 \end{pmatrix}$$

$$\frac{n}{m}\left(\hat{\beta} - \beta^0\right)\left(\hat{\beta} - \beta^0\right)^T = \begin{pmatrix} 3.42 & 0.02 \\ 0.02 & 0.14 \end{pmatrix}$$

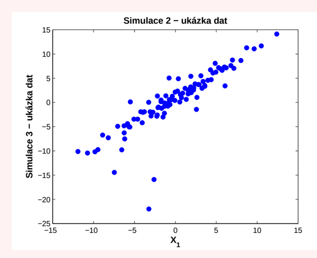


SIMULACE 2 – KONTAMINACE 25%

Počet pozorování $n = 100$, počet opakování v rámci simulace $m = 1000$

$Y = 1 + X_1 + e$, $\mathcal{L}(X_1) = \mathcal{N}(0,5)$, $\mathcal{L}(e_i) = \mathcal{N}(0,1)$, $i = 1, .., 75$, $\mathcal{L}(e_i) = t_1$, $i = 76, .., 100$.

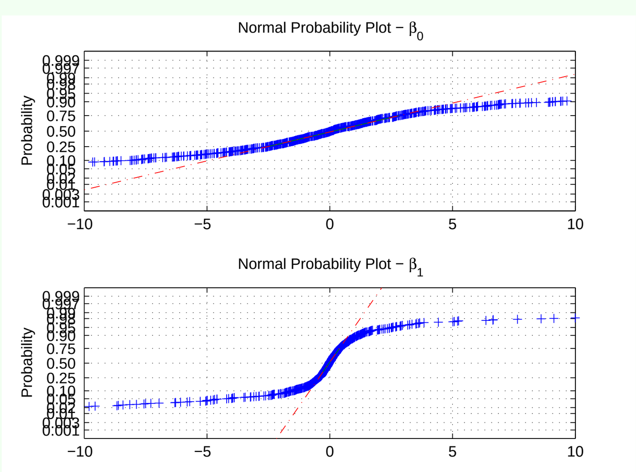
LWS: váhová funkce $w(x) = I\{x \le 0.5\} + (-2x+2) \cdot I\{x > 0.5\}$, LTS: $h = 75$.



LS

$$AVG(\hat{\beta}) = \begin{pmatrix} 0.60 \\ 0.98 \end{pmatrix}$$

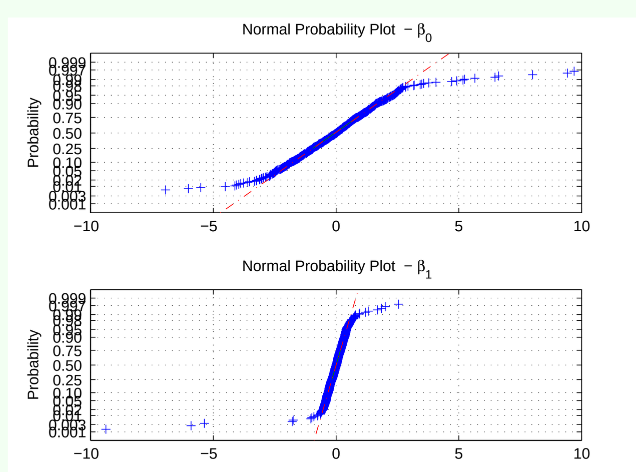
$$\frac{n}{m}\left(\hat{\beta} - \beta^0\right)\left(\hat{\beta} - \beta^0\right)^T = \begin{pmatrix} 27647 & -1464 \\ -1464 & 674 \end{pmatrix}$$



LWS

$$AVG(\hat{\beta}) = \begin{pmatrix} 0.99 \\ 1.00 \end{pmatrix}$$

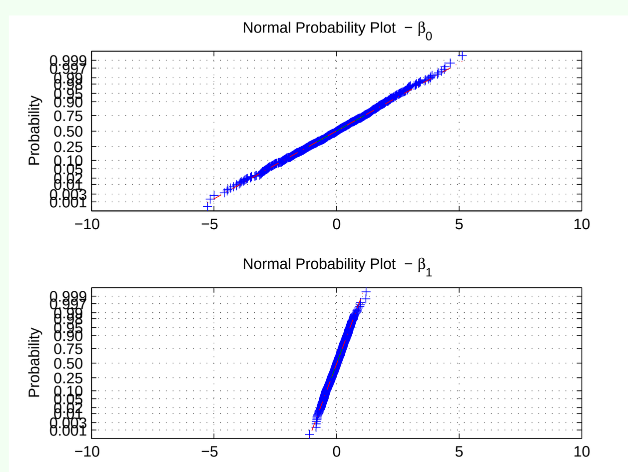
$$\frac{n}{m}\left(\hat{\beta} - \beta^0\right)\left(\hat{\beta} - \beta^0\right)^T = \begin{pmatrix} 13.31 & -1.47 \\ -1.47 & 0.77 \end{pmatrix}$$



LTS

$$AVG(\hat{\beta}) = \begin{pmatrix} 1.00 \\ 1.00 \end{pmatrix}$$

$$\frac{n}{m}\left(\hat{\beta} - \beta^0\right)\left(\hat{\beta} - \beta^0\right)^T = \begin{pmatrix} 2.72 & 0.00 \\ 0.00 & 0.11 \end{pmatrix}$$

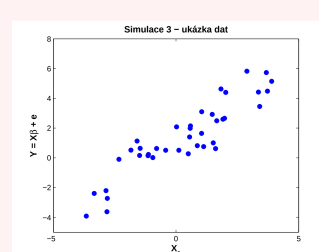


SIMULACE 3 – PODSOUBOROVÁ STABILITA

Počet pozorování $n = 20$, počet opakování v rámci simulace $m = 1000$

$Y = 1 + X_1 + e$, $\mathcal{L}(X_1) = \mathcal{N}(0, 2.5)$, $\mathcal{L}(e) = \mathcal{N}(0,1)$

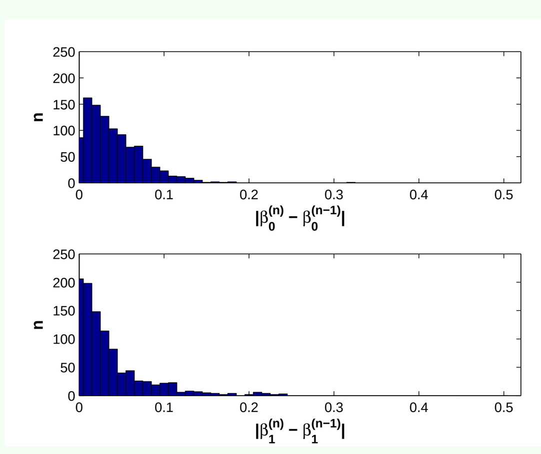
LWS: váhová funkce $w(x) = I\{x \le 0.5\} + (-2x+2) \cdot I\{x > 0.5\}$, LTS: $h = 15$.



LS

$$AVG(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1)}) = \begin{pmatrix} 0.041 \\ 0.036 \end{pmatrix}$$

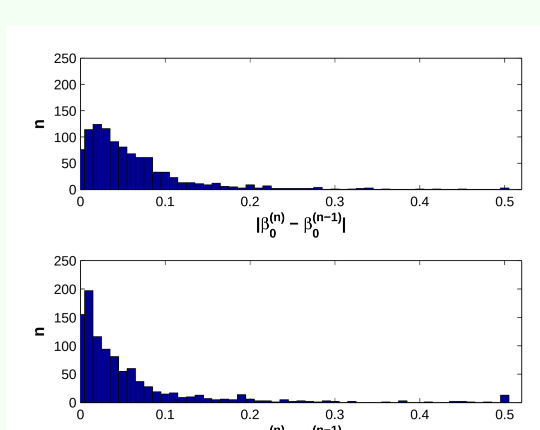
$$STD(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1)}) = \begin{pmatrix} 0.033 \\ 0.043 \end{pmatrix}$$



LWS

$$AVG(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1)}) = \begin{pmatrix} 0.061 \\ 0.059 \end{pmatrix}$$

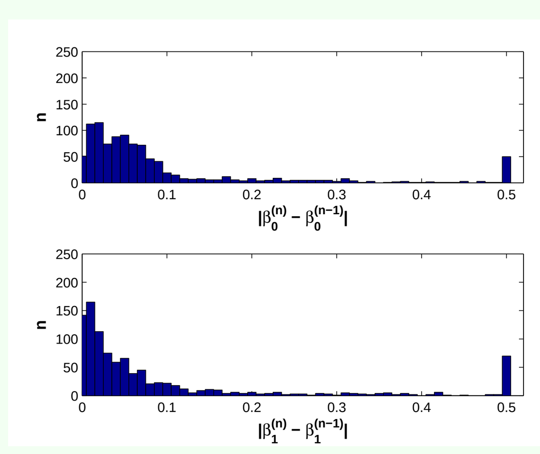
$$STD(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1)}) = \begin{pmatrix} 0.069 \\ 0.097 \end{pmatrix}$$



LTS

$$AVG(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1)}) = \begin{pmatrix} 0.109 \\ 0.122 \end{pmatrix}$$

$$STD(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1)}) = \begin{pmatrix} 0.183 \\ 0.246 \end{pmatrix}$$



[1]Výzkum byl prováděn v rámci grantu GA ČR number 402/06/0408.