IMS Collections
Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková
Vol. 7 (2010) 254–267

DOI: 10.1214/10-IMSCOLL725

Robust error-term-scale estimate

Jan Ámos Víšek[1,*]

Faculty of Social Sciences, Charles University and Institute of Information Theory and Automation

Abstract: A scale-equivariant and regression-invariant estimator of the variance of error terms in the linear regression model is proposed and its consistency proved. The estimator is based on (down)weighting the order statistics of the squared residuals which corresponds to the consistent and scale- and regression-equivariant estimator of the regression coefficients. A small numerical study demonstrating the behaviour of the estimator under the various types of contamination is included.

Let $\mathcal{N}$ denote the set of all positive integers, R the real line and R^p the p-dimensional Euclidean space. For a sequence of $(p+1)$-dimensional random vectors $\{(X_i', e_i)'\}_{i=1}^{\infty}$, for any $n \in \mathcal{N}$ and some fix $\beta^0 \in R^p$ the linear regression model will be considered in the form

$$(1) \qquad Y_i = X_i'\beta^0 + e_i = \sum_{j=1}^{p} X_{ij}\beta_j^0 + e_i, \quad i = 1, 2, \ldots, n \quad \text{or} \quad Y = X\beta^0 + e.$$

To put the introduction which follows in the proper context let us assume:

Conditions $\mathcal{C}1$ The sequence $\left\{(X_i', e_i)'\right\}_{i=1}^{\infty}$ is sequence of independent and identically distributed $(p+1)$-dimensional random variables, distributed according to distribution functions (d.f.) $F_{X,e}(x,r) = F_X(x) \cdot F_e(r)$ where $F_e(r) = F(r\sigma^{-1})$. Moreover, $F(r)$ is absolutely continuous with density $f(r)$ bounded by U and $I\!E_{F_e} e_1 = 0$, $\mathrm{var}_{F_e}(e_1) = \sigma^2$. Finally, $I\!E_{F_X} \|X_1\|^2 < \infty$.

Remark 1 The assumption that the (parent) d.f. $F(r)$ is continuous is not only technical assumption. Possibility that the error terms in regression model are discrete r.v.'s implies problems with treating response variable and it requires special considerations, similar to those which we carry out when studying binary or limited response variable, see e. g. in Judge et al. [16]. Absolute continuity is then a technical assumption. Without the density, even bounded density, we have to assume that $F(r)$ is Lipschitz and it would bring a more complicated form of all what follows.

A general goal of regression analysis is to fit a model (1) to the data. The analysis usually starts with estimating the regression coefficients β_j's, continues by the estimation of the variance σ^2 of the error terms e_i's (sometimes both steps run

*Research was supported by grant of GA ČR number 402/09/0557.
[1]Dept. Macroeconomics and Econometrics, Inst. of Economic Studies, Fac. of Social Sciences, Charles University, Opletalova ulice 26, 110 01 Praha 1 and Dept. Econometrics, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic. e-mail: visek@fsv.cuni.cz
AMS 2000 subject classifications: Primary 62J02; secondary 62F35
Keywords and phrases: robustness, weighting the order statistics of squared residuals, consistency of the scale estimator.

simultaneously, Marazzi [22]), then it includes a validation of the assumptions, etc. The present paper is devoted to the (robust) estimation of σ^2. In the classical LS-analysis we need the estimate of σ (usually assumed as $\sqrt{\hat{\sigma}^2}$) for studentization of the estimates of regression coefficients in order to establish the significance of the explanatory variables. In the robust analysis we employ it at first for studentizing the residuals, in the case when the properties of our estimate depends on the absolute magnitude of residuals, e. g. as in the case of M-estimators. So the estimation of the variance of error terms (in the case of the homoscedasticity of error terms) is one of standard (and important) steps of regression analysis. But it need not be a very simple task.

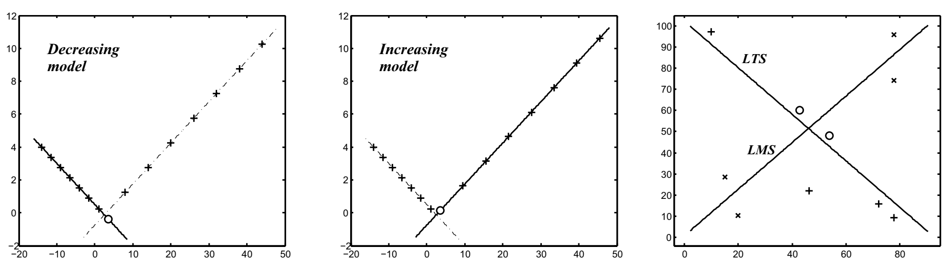
As early as in 1975 Peter Bickel [3] showed that to achieve the *scale-* and *regression-equivariance* of the M-estimates of regression coefficients the studentization of residuals has to be performed by a *scale-equivariant* and *regression-invariant* estimate of the scale of error terms. A proposal of such an estimator by Jana Jurečková and Pranab Kumar Sen [19] is based on regression scores. The idea is derived from the regression quantiles of Roger Koenker and Gilbert Bassett [21] and the evaluation utilizes standard methods of the stochastic linear programming, see Jurečková, Picek [17]. As the regression quantiles are based on L_1 metric (they are in fact M-estimators of the quantiles of d. f. of error terms, provided we know β^0), they can cope with outliers but can be significantly influenced by the presence of leverage points in the data, see Maronna, Yohai [23].

We propose an alternative estimator of σ^2 based on L_2-metric. In fact, our proposal generalizes an LTS-based scale estimator studied by Croux and Rousseeuw [8]. Of course, by a decision how many order statistics of the squared residuals will be taken into account one can adapt the estimator to the contamination level. We shall return to this problem at the end of paper in *Conclusions*. Croux–Rousseeuw estimator was also tested on the economic data by Bramanti and Croux [6]. Later, there appeared the paper by Pison et al. [24] proposing a correction of the estimator for small samples.

Our estimator can be also accommodated to the level and to the character of contamination by selecting an appropriate estimator of regression coefficients (we shall discuss the topic at the end of this section). Similarly as in the classical regression analysis, the evaluation of the estimator proposed here represents the step which follows the estimation of regression coefficients. We assume that the respective estimator of regression coefficients is *scale-* and *regression-equivariant* and *consistent*. Nowadays the robust statistics offer a whole range of such estimators. Let us recall e. g. the *least median of squares (LMS)* (Rousseeuw [26]), the *least trimmed squares (LTS)* (Hampel et al. [11]), the *least weighted squares (LWS)* (Víšek [40]) or the *instrumental weighted variables (IWV)* (Víšek [41]), to give some among many others (*instrumental weighted variables* is the robustified version of classical *instrumental variables* which became in the past (say) three decades the main estimating method in econometrics, being able to cope with the broken orthogonality condition, see Judge et al. [16], Stock, Trebbi [30] or Wooldridge [44]).

There are nowadays also quick and reliable algorithms for evaluation of the estimates. The research for such algorithms started at very early days of robust statistics (Rousseeuw, Leroy [29]) and it brought a lot of results, see e. g. Marazzi [22]). The research significantly intensified when Thomas Hettmansperger and Simon Sheather [14] discovered a high sensitivity of LMS with respect to a small shift of data (one datum among 80 was changed less than 10% but the estimates changed surprisingly about hundreds – or for some coefficients, even thousands – percents). Fortunately, there appeared a new algorithm by Boček, Lachout [5], based on a modification of the simplex method, which showed that the results by Hettmansperger and Sheather were achieved due to a wrong algorithm they used, see Víšek [35]. The algorithm by Boček and Lachout is (to the knowledge of present author) still superior in the sense of the minimization of corresponding order statistic. Later also an algorithm returning a tight approximation to LTS was proposed (Víšek [34], [35]) and included into XPLORE, see Hărdle et al. [12] or Čížek, Víšek [9]). Several variants of this algorithm was studied for various situations and improved especially for utilization for very large data sets, e. g. Agulló [1], Hawkins [13], Rousseeuw, Driessen [27], [28] and also by Hofmann et al. [16] – for deep theoretical study of the algorithms see Klouda [20]. Recently, the algorithm was generalized for evaluating LWS as well as for IWV, see Víšek [39].

Although Hettmansperger's and Sheather's results appeared misleading, an evaluation of LTS by an exact algorithm (searching through all corresponding subsamples) for their correct and damaged data (the data are nowadays referred to as *Engine Knock Data*, Hettmansperger, Sheather [14]) showed that the two respective estimates of regression coefficients are about hundreds percents different. It "has broken down" a statistical folklore that the robust methods with the high breakdown point – although losing (a lot of) efficiency – can reliably indicate (at least rough) idea about the underlying model. An explanation (by academic data) is given by the next three figures. First two of them indicate that a small change of observation given by the tiny circle (the change may be even arbitrary small – if closer to the intersection of the two lines) can cause a large change of the fitted model, if we use unconsciously an estimator with high breakdown point. The last figure demonstrates that LTS and LMS can give mutually orthogonal models. The observations drawn by circles are taken into account by both estimators while the observations given by '+' and 'x' are considered only by LTS and LMS, respectively. In both cases the curiosities appeared due to the zero-one object function, or in other words, due to the fact that the estimators too much rely on some points and completely reject some others. Hence, some other pairs of estimators with high breakdown point can presumably exhibit a similar behaviour.



A shock caused at the first moment by Hettmansperger's and Sheather's results has also began studies of the sensitivity of robust procedures with respect to (small) changes in the data, which in fact continued the studies by Chatterjee and Hadi [7] or Zvára [45]. It appeared that the estimator with discontinuous object function suffer by large sensitivity with respect of deleting even one point, see Víšek [33], [36], [37]. That is why we offer in the numerical study in the last section as the robust estimator of regression coefficient the *least weighted squares* (LWS) with continuous object function.

Weighting the order statistics of squared residuals

Let us start with recalling definitions of notions we shall need later.

Definition 1 The estimator of regression coefficients, is said to be *scale-equivariant* (*regression-equivariant*) if for any $c \in R^+, b \in R^p, Y \in R^n$ and X – matrix of type $n \times p$ – we have

$$(2) \qquad \hat{\beta}(cY, X) = c\hat{\beta}(Y, X) \qquad \left(\hat{\beta}(Y + Xb, X) = \hat{\beta}(Y, X) + b \right).$$

Definition 2 The estimator $\hat{\sigma}^2$ of the variance σ^2 of error terms is said to be *scale-equivariant* (*regression-invariant*) if for any $c \in R^+, b \in R^p, Y \in R^n$ and X – matrix of type $n \times p$

$$\hat{\sigma}^2(cY, X) = c^2\hat{\sigma}^2(Y, X) \qquad \left(\hat{\sigma}^2(Y + Xb, X) = \hat{\sigma}^2(Y, X) \right).$$

Now we are going to give a proposal of estimator of variance σ^2 of error terms e_i's (see (1)). Let for any $\beta \in R^p \quad r_i(\beta) = Y_i - X_i^{\prime}\beta$ denote the i-th residual and $r^2_{(h)}(\beta)$ the h-th order statistic among the squared residuals, i. e. we have

$$r^2_{(1)}(\beta) \leq r^2_{(2)}(\beta) \leq \quad \cdots \quad \leq r^2_{(n)}(\beta).$$

Finally, let $w(u)$ be a weight function $w : [0, 1] \to [0, 1]$ and put $\gamma = \int w\left(F(|r|)\right) \cdot r^2 f(r) \mathrm{d}r$.

Remark 2 Under Conditions $\mathcal{C}1$ the d. f. $F_e(r)$ has the density $f_e(r) = \sigma^{-1} f(r \cdot \sigma^{-1})$ and hence

$$(3) \qquad \sup_{r \in R} f_e(r) \leq \sigma^{-1} \cdot U.$$

Denote $U_e = \sigma^{-1} \cdot U$. Further, $\int w\left(F_e(|r|)\right) \cdot r^2 \cdot f_e(r) \mathrm{d}r = \sigma^2 \cdot \int w\left(F(|v| \cdot \sigma^{-1})\right) \cdot v^2 f(v) \mathrm{d}v = \gamma \cdot \sigma^2$, i. e.

$$(4) \qquad \gamma^{-1} \cdot \int w\left(F_e(|r|)\right) \cdot r^2 \cdot f_e(r) \mathrm{d}r = \sigma^2.$$

Definition 3 Let $\hat{\beta}^{(n)}$ be an estimator of regression coefficients. Then put

$$(5) \qquad \hat{\sigma}^2_{(n)} = \gamma^{-1} \cdot \frac{1}{n} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r^2_{(i)}(\hat{\beta}^{(n)}).$$

Remark 3 The estimator $\hat{\sigma}^2_{(n)}$ needs to be adjusted to the parent d. f. $F(r)$ by γ. It is similar as e. g. *mean absolute deviation*, see Hampel et al. [11] and Rousseeuw, Leroy [29].

We will need some conditions on the weight function.

Conditions $\mathcal{C}2$ The weight function $w(u)$ is continuous nonincreasing, $w : [0, 1] \to [0, 1]$ with $w(0) = 1$. Moreover, $w(u)$ is Lipschitz in absolute value, i. e. there is L such that for any pair $u_1, u_2 \in [0, 1]$ we have $|w(u_1) - w(u_2)| \leq L \cdot |u_1 - u_2|$.

Following Hájek and Šidák [10] for any $i \in \{1, 2, \ldots, n\}$ and any $\beta \in R^p$ let us define *regression ranks* as

$$(6) \qquad \pi(\beta, i) = j \in \{1, 2, \ldots, n\} \qquad \Leftrightarrow \qquad r^2_i(\beta) = r^2_{(j)}(\beta).$$

Let us denote the empirical distribution function (e.d.f.) of the absolute value of residual as

$$F_\beta^{(n)}(r) = \frac{1}{n}\sum_{j=1}^{n} I\left\{|r_j(\beta)| < r\right\} = \frac{1}{n}\sum_{j=1}^{n} I\left\{|Y_j - X_j^{'}\beta| < r\right\}. \tag{7}$$

Due to (6), $r_i^2(\beta)$ is the $\pi(\beta,i)$-th smallest value among the squared residuals, i. e. $|r_i(\beta)|$ is the $\pi(\beta,i)$-th smallest value among the absolute values of the residuals. Hence e. d. f. has at $|r_i(\beta)|$ its $\pi(\beta,i)$-th jump (of magnitude $\frac{1}{n}$), nevertheless due to the sharp inequality in the definition of e. d. f. (see (7)) we have

$$F_\beta^{(n)}(|r_i(\beta)|) = \frac{\pi(\beta,i)-1}{n}. \tag{8}$$

Then we have from (5)

$$\begin{aligned} \hat{\sigma}^2_{(n)} &= \gamma^{-1}\cdot\frac{1}{n}\sum_{i=1}^{n} w\left(\frac{\pi(\beta,i)-1}{n}\right) r_i^2(\beta) \\ &= \gamma^{-1}\cdot\frac{1}{n}\sum_{i=1}^{n} w\left(F_{\hat\beta}^{(n)}(|r_i(\hat\beta)|)\right) r_i^2(\hat\beta). \end{aligned} \tag{9}$$

Putting moreover

$$F_\beta(r) = P\left(|Y_1 - X_1^{'}\beta| < r\right) = P\left(|e_1 - X_1^{'}\left(\beta-\beta^0\right)| < r\right), \tag{10}$$

we can give key lemmas for reaching the consistency of $\hat{\sigma}^2_{(n)}$.

Lemma 1 Let Conditions $\mathcal{C}1$ hold. Then for any $\varepsilon > 0$ there is K_ε and $n_\varepsilon \in \mathcal{N}$ so that for all $n > n_\varepsilon$

$$P\left(\left\{\omega\in\Omega : \sup_{r\in R^+,\ \beta\in R^p} \sqrt{n}\left|F_\beta^{(n)}(r) - F_\beta(r)\right| < K_\varepsilon\right\}\right) > 1-\varepsilon. \tag{11}$$

For the proof see Víšek [38] (the proof is based on generalization of result by Kolmogorov and Smirnov). An alternative way how to prove (11) is to employ Skorohod embedding (see Breiman [4] or Štěpán [31] for the method and e. g. Portnoy [25], Jurečková, Sen [19] or Víšek [42] for examples of employing this technique).

Lemma 2 Under Conditions $\mathcal{C}1$ there is $K < \infty$ so that for any pair $\beta^{(1)}, \beta^{(2)} \in R^p$ we have

$$\sup_{r\in R}\left|F_{\beta^{(1)}}(r) - F_{\beta^{(2)}}(r)\right| \le K\cdot\left\|\beta^{(1)}-\beta^{(2)}\right\|.$$

Proof: We have

$$F_\beta(r) = P\left(\left|e_1 - X_1^{'}\left(\beta-\beta^0\right)\right| < r\right) = \int I\{\left|s - x^{'}\left(\beta-\beta^0\right)\right| < r\} \mathrm{d}F_{X,e}(x,s)$$

(see (10)). Then

$$\sup_{r\in R}\left|F_{\beta^{(1)}}(r) - F_{(\beta^{(2)})}(r)\right|$$

$$\le \sup_{r\in R}\int\left|I\{|s-x^{'}\left(\beta^{(1)}-\beta^0\right)| < r\} - I\{|s-x^{'}\left(\beta^{(2)}-\beta^0\right)| < r\}\right| f_e(s)\mathrm{d}s\,\mathrm{d}F_X(x).$$

Further, recalling that $\sup_{r\in R} f_e(r) \le U_e$ (see Remark 2), we have

$$\int\left|I\left\{|s-x^{'}\left(\beta^{(1)}-\beta^0\right)| < r\right\} - I\left\{|s-x^{'}\left(\beta^{(2)}-\beta^0\right)| < r\right\}\right| f_e(s)\mathrm{d}s$$

$$\leq \int_{\min\{-r+x'\left(\beta^{(1)}-\beta^{0}\right),-r+x'\left(\beta^{(2)}-\beta^{0}\right)\}}^{\max\{-r+x'\left(\beta^{(1)}-\beta^{0}\right),-r+x'\left(\beta^{(2)}-\beta^{0}\right)\}} f_e(s)\mathrm{d}s$$

$$+\int_{\min\{r+x'\left(\beta^{(1)}-\beta^{0}\right),r+x'\left(\beta^{(2)}-\beta^{0}\right)\}}^{\max\{r+x'\left(\beta^{(1)}-\beta^{0}\right),r+x'\left(\beta^{(2)}-\beta^{0}\right)\}} f_e(s)\mathrm{d}s$$

$$\leq 2\cdot U_e\cdot\left|x'\left(\beta^{(1)}-\beta^{(2)}\right)\right|.$$

Hence putting $K = 2\cdot U_e\cdot I\!E\,\|X_1\|$, for any $\beta^{(1)},\beta^{(2)}\in R^p$ we have

$$\sup_{r\in R}\left|F_{\beta^{(1)}}(r)-F_{\beta^{(2)}}(r)\right|\leq 2\cdot U_e\int\left|x'\left(\beta^{(1)}-\beta^{(2)}\right)\right|f_X(x)\mathrm{d}x$$

$$\leq 2\cdot U_e\cdot I\!E\,\|X_1\|\cdot\left\|\beta^{(1)}-\beta^{(2)}\right\|\leq K\cdot\left\|\beta^{(1)}-\beta^{(2)}\right\|. \qquad \square$$

Lemma 3 Let Conditions $\mathcal{C}1$ and $\mathcal{C}2$ hold. Then there is $K<\infty$ so that for any pair $\beta^{(1)},\beta^{(2)}\in R^p$ and any $i=1,2,\ldots,n$ we have

$$\left|w\left(F_{\beta^0}\left(\left|r_i(\beta^{(1)})\right|\right)\right)-w\left(F_{\beta^0}\left(\left|r_i(\beta^{(2)})\right|\right)\right)\right|\leq K\cdot\left\|\beta^{(1)}-\beta^{(2)}\right\|\cdot\|X_i\|.$$

Proof: Let us recall once again that

$$F_\beta(r)=P\left(\left|e_1-X_1'\beta\right|<r\right)=\int I\{\left|s-x'\beta\right|<r\}f_e(s)\mathrm{d}s\,\mathrm{d}F_X(x)$$

and that $\sup_{r\in R} f_e(r)\leq U_e$ (see Remark 2). Then

$$\left|F_{\beta^0}\left(\left|r_i(\beta^{(1)})\right|\right)-F_{\beta^0}\left(\left|r_i(\beta^{(2)})\right|\right)\right|$$

$$\leq\int\left|I\left\{|s-x'\beta^0|<\left|r_i(\beta^{(1)})\right|\right\}-I\left\{|s-x'\beta^0|<\left|r_i(\beta^{(2)})\right|\right\}\right|f_e(s)\mathrm{d}s\,\mathrm{d}F_X(x).$$

Further

$$\int\left|I\left\{|s-x'\beta^0|<\left|r_i(\beta^{(1)})\right|\right\}-I\left\{|s-x'\beta^0|<\left|r_i(\beta^{(2)})\right|\right\}\right|f_e(s)\mathrm{d}s$$

$$\leq\int_{\min\{-\left|r_i(\beta^{(1)})\right|+x'\beta^0,-\left|r_i(\beta^{(2)})\right|+x'\beta^0\}}^{\max\{-\left|r_i(\beta^{(1)})\right|+x'\beta^0,-\left|r_i(\beta^{(2)})\right|+x'\beta^0\}} f_e(s)\mathrm{d}s$$

$$+\int_{\min\{\left|r_i(\beta^{(1)})\right|+x'\beta^0,\left|r_i(\beta^{(2)})\right|+x'\beta^0\}}^{\max\{\left|r_i(\beta^{(1)})\right|+x'\beta^0,\left|r_i(\beta^{(2)})\right|+x'\beta^0\}} f_e(s)\mathrm{d}s$$

$$\leq 2\cdot U_e\cdot\left|r_i(\beta^{(1)})-r_i(\beta^{(2)})\right|\leq 2\cdot U_e\cdot\|X_i\|\cdot\left\|\beta^{(1)}-\beta^{(2)}\right\|$$

where we have used $\left||a|-|b|\right|\leq|a-b|$. Hence putting $K=2\cdot L\cdot U_e$, we have

$$\left|w\left(F_{\beta^0}\left(\left|r_i(\beta^{(1)})\right|\right)\right)-w\left(F_{\beta^0}\left(\left|r_i(\beta^{(2)})\right|\right)\right)\right|\leq K\cdot\left\|\beta^{(1)}-\beta^{(2)}\right\|\|X_i\|. \qquad \square$$

Assertion 1 We have

$$\sum_{i=1}^n\left|r_i^2(\hat\beta)-e_i^2\right|\leq 2\cdot\left\|\beta^0-\hat\beta\right\|\cdot\sum_{i=1}^n|e_i|\cdot\|X_i\|+\left\|\beta^0-\hat\beta\right\|^2\cdot\sum_{i=1}^n\|X_i\|^2. \tag{12}$$

Proof: Straightforward steps gives

$$\left|r_i^2(\hat\beta)-e_i^2\right|=\left|\left[e_i-X_i'\left(\hat\beta-\beta^0\right)\right]^2-e_i^2\right|\le 2\cdot|e_i|\cdot\|X_i\|\cdot\left\|\hat\beta-\beta^0\right\|+\|X_i\|^2\cdot\left\|\hat\beta-\beta^0\right\|^2. \square$$

Conditions $\mathcal{C}3$ The estimator of regression coefficients $\hat\beta^{(n)}$ is *scale-* and *regression-equivariant* and *consistent*.

Corollary 1 Under Conditions $\mathcal{C}1$ and $\mathcal{C}3$ we have

$$(13)\qquad \frac{1}{n}\sum_{i=1}^n\left|r_i^2(\hat\beta)-e_i^2\right|=o_p(1) \quad\text{and hence also}\quad \frac{1}{n}\sum_{i=1}^n r_i^2(\hat\beta)=\mathcal{O}_p(1).$$

Proof: Under Conditions $\mathcal{C}1$ we have $I\!\!E\left\{|e_1|\cdot\|X_1\|\right\}<\infty$ as well as $I\!\!E\left\{\|X_1\|^2\right\}<\infty$. Hence $\frac{1}{n}\sum_{i=1}^n|e_i|\cdot\|X_i\|=\mathcal{O}_p(1)$ and also $\frac{1}{n}\sum_{i=1}^n\|X_i\|^2=\mathcal{O}_p(1)$. As $\left\|\hat\beta-\beta^0\right\|=o_p(1)$, applying Assertion 1, we prove the left hand side of (13). Then

$$\frac{1}{n}\sum_{i=1}^n r_i^2(\hat\beta)\le\frac{1}{n}\sum_{i=1}^n\left|r_i^2(\hat\beta)-e_i^2\right|+\frac{1}{n}\sum_{i=1}^n e_i^2=\mathcal{O}_p(1). \qquad\square$$

Theorem 1 Let Conditions $\mathcal{C}1$, $\mathcal{C}2$ and $\mathcal{C}3$ hold. Then the estimator $\hat\sigma^2_{(n)}$ is weakly consistent, scale-equivariant and regression-invariant.

Proof: Fix $\varepsilon>0$ and according to Lemma 1 find $K_\varepsilon>0$ and $n_\varepsilon\in\mathcal{N}$ so that for any $n>n_\varepsilon$ we have

$$(14)\qquad P\left(\left\{\omega\in\Omega:\sup_{r\in R^+,\ \beta\in R^p}\sqrt{n}\left|F_\beta^{(n)}(r)-F_\beta(r)\right|<K_\varepsilon\right\}\right)>1-\varepsilon.$$

Denote the set

$$(15)\qquad B_n=\left\{\omega\in\Omega:\sup_{r\in R^+,\ \beta\in R^p}\sqrt{n}\left|F_\beta^{(n)}(r)-F_\beta(r)\right|<K_\varepsilon\right\}.$$

Then for any $\omega\in B_n$ we have

$$\left|\gamma\cdot\hat\sigma^2_{(n)}-\frac{1}{n}\sum_{i=1}^n w\left(F_{\hat\beta}(|r_i(\hat\beta)|)\right)r_i^2(\hat\beta)\right|$$

$$=\left|\frac{1}{n}\sum_{i=1}^n\left[w\left(F_{\hat\beta}^{(n)}(|r_i(\hat\beta)|)\right)-w\left(F_{\hat\beta}(|r_i(\hat\beta)|)\right)\right]r_i^2(\hat\beta)\right|$$

$$\le\frac{1}{n}\sum_{i=1}^n\left|w\left(F_{\hat\beta}^{(n)}(|r_i(\hat\beta)|)\right)-w\left(F_{\hat\beta}(|r_i(\hat\beta)|)\right)\right|r_i^2(\hat\beta)$$

$$\le L\cdot\sup_{r\in R^+,\ \beta\in R^p}\sqrt{n}\left|F_\beta^{(n)}(r)-F_\beta(r)\right|n^{-\frac{3}{2}}\sum_{i=1}^n\left|r_i^2(\hat\beta)\right|.$$

Due to (13) we have $n^{-\frac{3}{2}}\sum_{i=1}^n\left|r_i^2(\hat\beta)\right|=o_p(1)$ and hence, due to (14),

$$\gamma \cdot \hat{\sigma}^2_{(n)} - \frac{1}{n}\sum_{i=1}^{n} w\left(F_{\hat{\beta}}(|r_i(\hat{\beta})|)\right) \cdot r_i^2(\hat{\beta}) = o_p(1). \tag{16}$$

Now, taking into account Condition $\mathcal{C}2$, we have

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left[w\left(F_{\hat{\beta}}(|r_i(\hat{\beta})|)\right) - w\left(F_{\beta^0}(|r_i(\hat{\beta})|)\right)\right] \cdot r_i^2(\hat{\beta})\right| \tag{17}$$
$$\leq \quad L \cdot \frac{1}{n}\sum_{i=1}^{n}\left|F_{\hat{\beta}}(|r_i(\hat{\beta})|) - F_{\beta^0}(|r_i(\hat{\beta})|)\right| \cdot r_i^2(\hat{\beta}).$$

Now, employing Lemma 2, we have (write for a while r_i instead of $r_i(\hat{\beta})$)

$$\frac{1}{n}\sum_{i=1}^{n}\left|F_{\hat{\beta}}(|r_i|) - F_{\beta^0}(|r_i|)\right| \cdot r_i^2 \leq \sup_{r\in R}\left|F_{\hat{\beta}}(r) - F_{\beta^0}(r)\right| \frac{1}{n}\sum_{i=1}^{n}\left|r_i^2\right| \tag{18}$$
$$\leq \quad K \cdot \left\|\hat{\beta} - \beta^0\right\| \cdot \frac{1}{n}\sum_{i=1}^{n}\left|r_i^2\right|.$$

Under Condition $\mathcal{C}1$, due to the consistency of $\hat{\beta}$, (18) is $o_p(1)$. Similarly, employing Lemma 3 and once again Condition $\mathcal{C}1$ and $\mathcal{C}2$, we have (remember that $r_i(\beta^0) = e_i$)

$$\frac{1}{n}\sum_{i=1}^{n} w\left(F_{\beta^0}(|r_i(\hat{\beta})|)\right) \cdot r_i^2(\hat{\beta}) - \frac{1}{n}\sum_{i=1}^{n} w\left(F_{\beta^0}(|e_i|)\right) \cdot r_i^2(\hat{\beta}) = o_p(1). \tag{19}$$

Employing Corollary 1, due to Conditions $\mathcal{C}1$, $\mathcal{C}2$ and $\mathcal{C}3$ we have (for $\|\hat{\beta} - \beta^0\| \leq 1$)

$$\left|\frac{1}{n}\sum_{i=1}^{n} w\left(F_{\beta^0}(|e_i|)\right) \cdot \left(r_i^2(\hat{\beta}) - e_i^2\right)\right| \tag{20}$$
$$\leq \quad 2 \cdot \left\|\hat{\beta} - \beta^0\right\| \frac{1}{n}\sum_{i=1}^{n}\left[|e_i| \cdot \|X_i\| + \|X_i\|^2\right] = o_p(1).$$

Finally, (16), (17), (19) and (20) implies that

$$\gamma \cdot \hat{\sigma}^2_{(n)} = \frac{1}{n}\sum_{i=1}^{n} w\left(F_{\beta^0}(|e_i|)\right) \cdot e_i^2 + o_p(1). \tag{21}$$

Taking into account (4), the weak consistency of $\hat{\sigma}^2_{(n)}$ follows from (21).

The scale-equivariance and the regression-invariance of $\hat{\sigma}^2_{(n)}$ follows directly from two facts. Firstly, estimator $\hat{\sigma}^2_{(n)}$ is based on the squared residuals of the estimator $\hat{\beta}$ of regression coefficients. As the estimator $\hat{\beta}$ is scale- and regression-equivariant, the residuals are scale-equivariant and regression-invariant, see (2). Since the weights depend on the empirical d. f., they are scale- and regression-invariant. □

Conditions $\mathcal{C}4$ The estimator of regression coefficients $\hat{\beta}^{(n)}$ is *scale-* and *regression-equivariant* and $\sqrt{n}$*-consistent.*

Corollary 2 Under Conditions $\mathcal{C}1$ and $\mathcal{C}4$

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\left|r_i^2(\hat{\beta}) - e_i^2\right| = \mathcal{O}_p(1). \tag{22}$$

Proof: Similarly as in (12) we have

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\left|r_i^2(\hat{\beta})-e_i^2\right| \le 2\cdot\sqrt{n}\left\|\beta^0-\hat{\beta}\right\|\cdot\frac{1}{n}\sum_{i=1}^{n}|e_i|\cdot\|X_i\|+\sqrt{n}\left\|\beta^0-\hat{\beta}\right\|^2\cdot\frac{1}{n}\sum_{i=1}^{n}\|X_i\|^2. \tag{23}$$

Using similar arguments as in the proof of Corollary 1, we conclude the proof. □

Theorem 2 Let the Conditions $\mathcal{C}1$, $\mathcal{C}2$ and $\mathcal{C}4$ hold. Then the estimator $\hat{\sigma}^2_{(n)}$ is $\sqrt{n}$-consistent.

Proof: Similarly as above, (13) and (14) yields

$$\sqrt{n}\cdot\gamma\cdot\hat{\sigma}^2_{(n)}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w\left(F_{\hat{\beta}}(|r_i(\hat{\beta})|)\right)\cdot r_i^2(\hat{\beta})=\mathcal{O}_p(1). \tag{24}$$

Employing again Lemma 1 and Condition $\mathcal{C}1$ and $\mathcal{C}2$, we have

$$\begin{aligned}&\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[w\left(F_{\hat{\beta}}(|r_i(\hat{\beta})|)\right)-w\left(F_{\beta^0}(|r_i(\hat{\beta})|)\right)\right]r_i^2(\hat{\beta})\right|\\ &\qquad\le L\cdot\sup_{r\in R^+,\ \beta\in R^p}\sqrt{n}\left|F_\beta^{(n)}(r)-F_\beta(r)\right|\frac{1}{n}\sum_{i=1}^{n}r_i^2(\hat{\beta})=\mathcal{O}_p(1).\end{aligned} \tag{25}$$

Similarly, utilizing Lemma 3 and once again Condition $\mathcal{C}1$ and $\mathcal{C}2$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w\left(F_{\beta^0}(|r_i(\hat{\beta})|)\right)\cdot r_i^2(\hat{\beta})-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w\left(F_{\beta^0}(|r_i(\beta^0)|)\right)\cdot r_i^2(\hat{\beta})=\mathcal{O}_p(1). \tag{26}$$

Using Corollary 2, due to Conditions $\mathcal{C}1$, $\mathcal{C}2$ and $\mathcal{C}4$ we have (for $\left\|\beta^0-\hat{\beta}\right\|\le 1$)

$$\begin{aligned}&\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w\left(F_{\beta^0}(|e_i|)\right)\cdot\left(r_i^2(\hat{\beta})-e_i^2\right)\right|\\ \le\ &2\sqrt{n}\left\|\beta^0-\hat{\beta}\right\|\frac{1}{n}\sum_{i=1}^{n}\left[|e_i|\cdot\|X_i\|+\|X_i\|^2\right]=\mathcal{O}_p(1).\end{aligned} \tag{27}$$

Finally, (25), (26) and (27) implies that

$$\sqrt{n}\cdot\gamma\cdot\left(\hat{\sigma}^2_{(n)}-\sigma^2\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(w\left(F_{\beta^0}(|e_i|)\right)\cdot e_i^2-\gamma\cdot\sigma^2\right)+\mathcal{O}_p(1)$$

and the $\sqrt{n}$-consistency of $\hat{\sigma}^2_{(n)}$ follows from the Central Limit Theorem and Remark 2. □

In the next chapter we offer a numerical study of the proposed scale estimator $\hat{\sigma}^2$. We shall use $\hat{\beta}^{(LWS,n,w)}$, given as solution of extremal problem

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

see Víšek [35], in the role of the robust, scale- and regression-equivariant estimator of regression coefficient. We shall need following conditions:

Conditions $\mathcal{C}5$ There is the only solution of

$$\beta' I\!\!E \left[w\left(F_\beta(|r(\beta)|)\right) X_1 \left(e - X_1' \left(\beta - \beta^0\right)\right)\right] = 0 \tag{28}$$

namely β^0 (the equation (28) is assumed as a vector equation in $\beta \in R^p$).

Conditions $\mathcal{NC}\,1$ The derivative $f'(r)$ exists and is bounded in absolute value by $B_e < \infty$. The derivative $w'(\alpha)$ exists and is Lipschitz of the first order (with the corresponding constant $J_w < \infty$).

Theorem 3 Under Conditions $\mathcal{C}1$, $\mathcal{C}2$ and $\mathcal{C}5$ $\hat{\beta}^{(LWS,n,w)}$ is consistent, scale- and regression-equivariant. Similarly, under Conditions $\mathcal{C}1, \mathcal{C}2, \mathcal{C}5$ and $\mathcal{NC}\,1$ $\hat{\beta}^{(LWS,n,w)}$ is $\sqrt{n}$-consistent.

Proof can be found in Víšek [40], [43].

Hence $\hat{\beta}^{(LWS,n,w)}$ can be used as the estimator we have considered in the construction of $\hat{\sigma}^2$.

Numerical study

The model (1) was employed with coefficients given in the first row of tables presented below. The *explanatory variables* were generated as sample from *3-dimensional normal population with zero means and diagonal covariance matrix (diagonal elements equal to 9).*

The *error terms* were generated as *normal with zero mean and variance equal to 2.*

We have generated 100 datasets, each of them containing 100 observations. As the robust, scale- and regression-equivariant estimator we have used $\hat{\beta}^{(LWS,n,w)}$, see the end of the previous chapter. The weight function was given for processing a *mild contamination* (see below) as

$$w(u) = 1 \quad \text{for } u \in [0, 0.8], \qquad w(u) = 20 \cdot (0.8 - u) + 1 \; u \in [0.8, 0.85], \tag{29}$$
$$w(u) = 0 \text{ otherwise.}$$

For processing a *heavy contamination* (see again below) we have began with a weight function of type (29) but with the upper bound of the first interval equal to 0.4 (instead of 0.8) and with much slower slope. Then we increased (step by step equal to 0.01) the upper bound of interval $[0, 0.4]$. The estimate of the scale of error term $\hat{\beta}^{(LWS,n,w)}$ as well as of regression coefficients were stable and they lost a stability when we overcame the value of the upper bound) 0.45. Hence we used

$$w(u) = 1 \quad \text{for } u \in [0, 0.45], \quad w(u) = 2.5 \cdot (0.45 - u) + 1 \; u \in [0.45, 0.85],$$
$$w(u) = 0 \text{ otherwise.}$$

As a benchmark we offer results of the *ordinary least squares* $\hat{\beta}^{(OLS,n)}$ and of the *least weighted squares* $\hat{\beta}^{(LWS,n,w)}$ for data without any contamination (the first table). The following tables collect results of the estimation of model by $\hat{\beta}^{(OLS,n)}$ and $\hat{\beta}^{(LWS,n,w)}$ under various types of contamination specified in the captions of tables (inside the frames).

The estimates were evaluated by algorithm discussed in Víšek [39] and implemented in MATLAB (the implementation is available on request). Every table contains in its first row the true values of regression model. The second and the third row of tables contain the empirical means from hundred $\hat{\beta}^{(OLS,n)}$'s and $\hat{\beta}^{(LWS,n,w)}$'s, respectively, evaluated for the (above mentioned) 100 datasets. The type and level of contamination is given in the first line of respective frame.

The adjusting constant γ was evaluated by numerical integration. Finally,

$$\hat{\sigma}^2_{OLS} = \frac{1}{n-p} \sum_{i=1}^{n} r_i^2(\hat{\beta}^{(OLS,n)})$$

and

$$\hat{\sigma}^2_{LWS} = \gamma^{-1} \cdot \frac{1}{n} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r^2_{(i)}(\hat{\beta}^{(LWS,n,w)}).$$

The results of estimating the variance of the error terms by these estimators are given on the second and on the third line of the frames, respectively.

Regression without contamination

For this case we have started with the weight function given in (29) and we have shifted the interval [0.8,0.85] to the right – step by step (equal 0.01) – so long while the results were stable, so that we have used finally

$$w(u) = 1 \text{ for } u \in [0, 0.95] \quad \textit{and} \quad w(u) = 20 \cdot (0.95 - u) + 1 \text{ for } u \in [0.95, 1].$$

$\hat{\sigma}^2_{OLS} = 1.99_{(.0641)}$ $\hat{\sigma}^2_{LWS} = 1.99_{(.0647)}$

β^0	1.5	4.3	−3.2
$\hat{\beta}^{(OLS,n)}$	$1.49_{(.0040)}$	$4.28_{.0039)}$	$-3.20_{(.0060)}$
$\hat{\beta}^{(LWS,n,w)}$	$1.49_{(.0042)}$	$4.28_{.0044)}$	$-3.20_{(.0063)}$

Regression with mild contamination

Contamination: For the first 5 observations we changed:
(let us recall that the true values of coefficients are in the first row of tables, while the second and the third ones contain $\hat{\beta}^{(OLS,n)}$ and $\hat{\beta}^{(LWS,n,w)}$, respectively; variances of estimates are in parenthesis)

Y_i to $2 * Y_i$

$\hat{\sigma}^2_{OLS} = 7.17_{(9.91)}$
$\hat{\sigma}^2_{LWS} = 2.30_{(.059)}$

1.5	4.3	−3.2
$1.55_{(.016)}$	$4.43_{(.017)}$	$-3.33_{(.022)}$
$1.49_{(.007)}$	$4.30_{(.006)}$	$-3.20_{(.007)}$

Y_i to $2 * Y_i$
and X_i to $2 * X_i$

$\hat{\sigma}^2_{OLS} = 72.52_{(1775.0)}$
$\hat{\sigma}^2_{LWS} = 2.29_{(0.082)}$

1.5	4.3	−3.2
$1.04_{(.490)}$	$3.17_{(.646)}$	$-2.29_{(.644)}$
$1.49_{(.007)}$	$4.30_{(.006)}$	$-3.21_{(.007)}$

Regression with heavy contamination but with inappropriate weight function

Contamination: For the first 45 observations we changed:

Y_i to $2 * Y_i$

$\hat{\sigma}^2_{OLS} = 34.90_{(26.26)}$
$\hat{\sigma}^2_{LWS} = 31.23_{(20.76)}$

1.5	4.3	-3.2
$2.16_{(.072)}$	$6.18_{(.091)}$	$-4.61_{(.010)}$
$1.89_{(.125)}$	$5.41_{(.283)}$	$-4.06_{(.181)}$

Y_i to $2 * Y_i$
and X_i to $2 * X_i$

$\hat{\sigma}^2_{OLS} = 237.52_{(1144.6)}$
$\hat{\sigma}^2_{LWS} = 214.1_{(1097.1)}$

1.5	4.3	-3.2
$-.77_{(.206)}$	$-2.14_{(.248)}$	$1.67_{(.211)}$
$-1.1_{(.176)}$	$-3.03_{(.579)}$	$2.26_{(.432)}$

Regression with heavy contamination and accommodated weight function

Contamination: For the first 45 observations we changed:

Y_i to $2 * Y_i$

$\hat{\sigma}^2_{OLS} = 333.99_{(24.66)}$
$\hat{\sigma}^2_{LWS} = 1.89_{(0.057)}$

1.5	4.3	-3.2
$2.16_{(.087)}$	$6.18_{(.101)}$	$-4.60_{(.104)}$
$1.54_{(.112)}$	$4.52_{(.682)}$	$-3.34_{(.394)}$

Y_i to $2 * Y_i$
and X_i to $2 * X_i$

$\hat{\sigma}^2_{OLS} = 232.4_{(899.9)}$
$\hat{\sigma}^2_{LWS} = 2.62_{(0.104)}$

1.5	4.3	-3.2
$-.71_{(.188)}$	$-2.1_{(.244)}$	$1.54_{(.242)}$
$1.5_{(.109)}$	$4.20_{(.736)}$	$-3.14_{(.41)}$

Conclusions of numerical study. It is clear that the outliers have a small influence on the estimates while the “combined” contamination (simultaneously by outliers and leverage points) much larger. Nevertheless, both $\hat{\beta}^{(LWS,n,w)}$ as well as $\hat{\sigma}^2_{LWS}$ have copped with contamination quite well - if the weight function was properly accommodated to the level of contamination. In practice we do not know the level of contamination. Then we may keep a (rather general) rule saying that starting with the “highest possible” robustness of $\hat{\sigma}^2_{(n)}$ and of $\hat{\beta}^{(LWS,n,w)}$, we can decrease their robustness so long when the estimates lose their stability, see e. g. Benáček, Víšek [2].

Acknowledgement. We would like to thank to two anonymous referees. Their comments indicated very precisely what was to be modified to make the text clear and easier to understand.

References

[1] Agulló, J. (2001). New algorithms for computing the least trimmed squares regression estimators. *Computational Statistics and Data Analysis* **36** 425–439.

[2] Benáček, V. and Víšek, J. Á. (2002). Determining factors of trade specialization and growth of a small economy in transition. Impact of the EU

opening-up on Czech exports and imports. *IIASA, Austria, IR series* no. IR-03-001 1–41.

[3] Bickel, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–433.

[4] Breiman, L. (1968). *Probability.* Addison-Wesley Publishing Company, London.

[5] Boček, P. and Lachout, P. (1993). Linear programming approach to *LMS*-estimation. *Memorial volume of Comput. Statist. & Data Analysis* **19** 129–134.

[6] Bramanti, M. C. and Croux, C. (2007). Robust estimators for the fixed effects panel data model. *The Econometrics Journal* **10** 321–540.

[7] Chatterjee, S. and Hadi A. S. (1988). *Sensitivity Analysis in Linear Regression.* J. Wiley & Sons, New York.

[8] Croux, C. and Rousseeuw, P. J. (1992). A class of high-breakdown scale estimators based on subranges. *Communications in Statistics – Theory and Methods* **21** 1935 –1951.

[9] Čížek, P. and Víšek, J. Á. (2000). The least trimmed squares. *User Guide of Explore.*

[10] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Test.* Academic Press, New York.

[11] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel W. A. (1986). *Robust Statistics – The Approach Based on Influence Functions.* J. Wiley & Sons, New York.

[12] Hărdle, W., Hlávka, Z, and Klinke, S. (2000). *XploRe Application Guide.* Springer Verlag, Heilderberg.

[13] Hawkins, D. M. (1994). The feasible solution algorithm for least trimmed squares regression. *Computational Statistics and Data Analysis* **17**, 185–196.

[14] Hettmansperger, T. P. and Sheather, S. J. (1992). A cautionary note on the method of Least Median Squares. *The American Statistician* **46** 79–83.

[15] Hofmann, M., Gatu, C., and Kontoghiorghes E. J. (2010). An exact least trimmed squares algorithm for a range of coverage values. *J. of Computational and Graphical Statistics* **19** 191–204.

[16] Judge, G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., and Lee, T. C. (1982). *Introduction to the Theory and Practice of Econometrics.* J. Wiley & Sons, New York.

[17] Jurečková, J. and Picek J. (2006). *Robust Statistical Methods with R.* Chapman & Hall, New York.

[18] Jurečková, J. and Sen, P. K. (1984). On adaptive scale-equivariant *M*-estimators in linear models. *Statistics and Decisions* **2**, Suppl. Issue No. 1.

[19] Jurečková, J. and Sen, P. K. (1993). Regression rank scores scale statistics and studentization in linear models. *Proc. Fifth Prague Symposium on Asymptotic Statistics*, Physica Verlag 111–121.

[20] Klouda, K. (2007). *Algorithms for computing robust regression estimates.* Diploma Thesis, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague.

[21] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

[22] Marazzi, A. (1992). *Algorithms, Routines and S Functions for Robust Statistics.* Wadsworth & Brooks/Cole Publishing Company, Belmont.

[23] Maronna, R. A. and Yohai, V. J. (1981). The breakdown point of simultaneous general *M*-estimates of regression and scale. *J. of Amer. Statist. Association* **86** (415), 699–704.

[24] Pison, G., Van Aelst, S., and Willems, G. (2002). Small sample correc-

tions for LTS and MCD. *Metrika* **55** 111–123.
[25] Portnoy, S. (1983). Tightness of the sequence of empiric c. d. f. processes defined from regression fractiles. In *Robust and Nonlinear Time-Series Analysis* (J. Franke, W. Hărdle, D. Martin, eds.), 231–246. Springer-Verlag, New York.
[26] Rousseeuw, P. J. (1984). Least median of square regression. *J. Amer. Statist. Association* **79** 871–880.
[27] Rousseeuw, P. J. and Driessen, K. (2000). An algorithm for positive-breakdown regression based on concentration steps. In *Data Analysis: Scientific Modeling and Practical Application* (W. Gaul, O. Opitz, M. Schader, eds.), 335 - 346. Springer-Verlag, Berlin.
[28] Rousseeuw, P. J. and Driessen, K. (2002). *Fast-LTS in Matlab*, code revision 20/04/2006.
[29] Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* J. Wiley, New York.
[30] Stock, J. H. and Trebbi, F. (2003). Who invented instrumental variable regression? *Journal of Economic Perspectives* **17** 177–194.
[31] Štěpán, J. (1987). *Teorie pravděpodobnosti.* Academia, Praha.
[32] Víšek, J. Á. (1994). A cautionary note on the method of the Least Median of Squares reconsidered. *Trans. Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Academy of Sciences of the Czech Republic, 254–259.
[33] Víšek, J. Á. (1996). Sensitivity analysis of M-estimates. *Annals of the Institute of Statistical Mathematics* **48** 469–495.
[34] Víšek, J. Á. (1996). On high breakdown point estimation. *Computational Statistics* 137–146.
[35] Víšek, J. Á. (2000). Regression with high breakdown point. *Robust 2000* (J. Antoch & G. Dohnal, eds.) Union of the Czech Mathematicians and Physicists, 324–356.
[36] Víšek, J. Á. (2002). Sensitivity analysis of M-estimates of nonlinear regression model: Influence of data subsets. *Annals of the Institute of Statistical Mathematics* **54** 261–290.
[37] Víšek, J. Á. (2006). The least trimmed squares. Sensitivity study. *Proc. Prague Stochastics 2006* (M. Hušková & M. Janžura, eds.), matfyzpress, 728–738.
[38] Víšek, J. Á. (2006). Kolmogorov–Smirnov statistics in multiple regression. *Proc. ROBUST 2006* (J. Antoch & G. Dohnal, eds.), Union of the Czech Mathematicians and Physicists, 367–374.
[39] Víšek, J. Á. (2006). Instrumental weighted variables – algorithm. *Proc. COMPSTAT 2006* 777–786.
[40] Víšek, J. Á. (2009). Consistency of the least weighted squares under heteroscedasticity. Submitted to the *Kybernetika.*
[41] Víšek, J. Á. (2009). Consistency of the instrumental weighted variables. *Annals of the Institute of Statistical Mathematics* **61** 543–578.
[42] Víšek, J. Á. (2010). Empirical distribution function under heteroscedasticity. To appear in *Statistics.*
[43] Víšek, J. Á. (2010). Weak $\sqrt{n}$-consistency of the least weighted squares under heteroscedasticity. Submitted to *Acta Universitatis Carolinae – Mathematica et Physica.*
[44] Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data.* MIT Press, Cambridge, Massachusetts.
[45] Zvára, K. (1989). *Regresní analýza* (Regression Analysis – in Czech). Academia, Praha.