Libor Mašíček

Optimality of the least weighted squares estimator

Kybernetika, Vol. 40 (2004), No. 6, [715]--734

Persistent URL: http://dml.cz/dmlcz/135629

KYBERNETIKA — VOLUME 40 (2004), NUMBER 6, PAGES 715–734

OPTIMALITY OF THE LEAST WEIGHTED SQUARES ESTIMATOR[1]

LIBOR MAŠÍČEK

The present paper deals with least weighted squares estimator which is a robust estimator and generalizes the classical least trimmed squares. We will prove $\sqrt{n}$-consistency and asymptotic normality for any sequence of roots of normal equation for location model. The influence function for the general case is calculated. Finally optimality of this estimator is discussed and a formula for the most B-robust and most V-robust weights is derived.

Keywords: robust regression, least trimmed squares, least weighted squares, influence function, $\sqrt{n}$-consistency, asymptotic normality, B-robustness, V-robustness
AMS Subject Classification: 62F35, 62J05

1. INTRODUCTION

Let us consider the following regression model

$$Y_i = X_i^T \beta_0 + Z_i \qquad \text{for } i = 1, \dots, n \tag{1.1}$$

where $X_i = (X_{i1}, \dots, X_{ip})^T$ is the $p \times 1$ column vector of explanatory variables, which are random, β_0 is the $p \times 1$ column vector of unknown regression coefficients and Z_i are random fluctuations with continuous distribution and $\mathsf{E} Z_i = 0$. Moreover, the sequence of random vectors $X_1, \dots, X_n$ is independent and identically distributed (i.i.d.), the sequence of random variables $Z_1, \dots, Z_n$ is i.i.d. and the sequences are mutually independent. For the choice $p = 1$ and $X_{i1} \equiv 1$ we obtain

$$Y_i = \beta_0 + Z_i \qquad \text{for } i = 1, \dots, n, \tag{1.2}$$

where $\beta_0 \in \mathbb{R}$ is an unknown parameter. This is known as the *location model.*

In general regression model we denote the ith residuum for any $\beta \in \mathbb{R}^p$ by

$$r_i(\beta) := Y_i - X_i^T \beta = Z_i - X_i^T(\beta - \beta_0) \tag{1.3}$$

and the hth order statistics of squared residuals by $r^2_{(h)}(\beta)$, i.e.

$$0 \le r^2_{(1)}(\beta) \le r^2_{(2)}(\beta) \le \cdots \le r^2_{(n)}(\beta). \tag{1.4}$$

[1]Some results from this paper were presented at 24th European Meeting of Statisticians held on August 19–23, 2002 in Prague.

Similarly for any $\beta \in \mathbb{R}^p$ we denote order statistics of absolute value of residuals by $r_{|h|}(\beta) := \sqrt{r^2_{(h)}(\beta)}$ (i. e. also square root of the hth order statistics of squared residuals).

Now we can define the *least weighted squares estimator* (LWS) as

$$\hat{\beta}_n = \hat{\beta}^{LWS}_{n,w} := \arg\min_{\beta \in \mathbb{R}^p} \sum_{h=1}^{n} w\left(\frac{h-1}{n}\right) r^2_{(h)}(\beta) \tag{1.5}$$

where $w : [0,1] \to [0,\infty)$ is a given *weight function.* Typically we suppose, that w is nonincreasing (i. e. observations with larger residuals have smaller weight). Without loss of generality we suppose $w(1) = 0$.

This estimator was developed by Víšek (see [7] and [8]) and it generalizes classical *least trimmed squares* (LTS) proposed by Rousseeuw (see [6]) which we get for the choice $w(x) = I\{x < \alpha\}$ where $I\{\dots\}$ is an indicator function and $\alpha \in (0,1)$. The main reason for developing this estimator was to improve applicability. In the LTS estimator one can adjust just one constant but in the LWS estimator we can choose the entire weight function. This gives a chance to increase efficiency or decrease gross error sensitivity.

This estimator has some nice properties. First of all the breakdown point comes immediately from the weight function. If $w(\alpha) > 0$ for $\alpha < \overline{\alpha}$ and $w(\alpha) = 0$ for $\alpha > \overline{\alpha}$ then the LWS estimator has breakdown point equal to $\min\{1 - \overline{\alpha}, \overline{\alpha}\}$. This means that we have the breakdown point under control and we can choose it arbitrary up to 0.5.

The most important property of this estimator is the regression and scale equivariance. This is an advantage w.r.t. M-estimators which are regression equivariant but not scale equivariant hence some studentisation of residuals by some robust estimator of scale is needed.

Finally we can multiply the weight function by an arbitrary positive constant and our estimator remains unchanged.

But there are some open questions about the LWS estimator. Under what conditions is this estimator consistent or $\sqrt{n}$-consistent? What is its asymptotic variance? What is its influence function? And presumably the most important question: What is the optimal choice of weights? We will answer some of these questions in this paper.

In the next section we derive the normal equations for the LWS estimator and we rewrite them as a statistical functional (i. e. as a function of empirical distribution function). In Section 3 we restrict ourselves to the case of location model and we provide conditions for $\sqrt{n}$-consistency and asymptotic normality of the LWS estimator for location model. In Section 4 we expresses the influence function of the LWS estimator for general regression. Section 5 combines results of Sections 3 and 4 and the most B-robust and V-robust LWS estimators for location model are expressed. Section 6 provides detailed proofs.

2. NORMAL EQUATIONS

Denote the function that is minimized in (1.5) by

$$\mathrm{MF}_n(\beta) := \sum_{h=1}^{n} w\left(\frac{h-1}{n}\right) r_{(h)}^2(\beta) \tag{2.1}$$

for $\beta \in \mathbb{R}^p$. Now we define random variables $\pi_0(i,\beta)$ for $i = 1,\ldots,n$ and $\beta \in \mathbb{R}^p$ in such a way that $r_i^2(\beta) = r^2_{(\pi_0(i,\beta))}(\beta)$, i.e. for any $\beta \in \mathbb{R}^p$ is $\pi_0(\beta) = \{\pi_0(1,\beta),\ldots,\pi_0(n,\beta)\}$ the random permutation on $\{1,\ldots,n\}$ which converts the ranks of the observations ordered by the squared residuals. Hence we can reorder the summation in (2.1) and rewrite it as

$$\mathrm{MF}_n(\beta) = \sum_{i=1}^{n} w\left(\frac{\pi_0(i,\beta)-1}{n}\right) r_i^2(\beta). \tag{2.2}$$

We see that the MF_n is the same as the minimized function of classical least squares with weights (i.e. weighted least squares, WLS) but in this case weights in (2.2) are not fixed since they depend on $\pi_0(i,\beta)$ and hence on the observations. We can suppose the LWS estimator should satisfy the normal equations of WLS estimator with corresponding weights as follows.

Lemma 1. Let us denote

$$\mathrm{NR}_n(\beta) := \sum_{i=1}^{n} w\left(\frac{\pi_0(i,\beta)-1}{n}\right) \left(X_i^T\beta - Y_i\right) X_i \tag{2.3}$$

for any $\beta \in \mathbb{R}^p$. Hence the LWS estimator is a solution of equations $\mathrm{NR}_n(\beta) = 0$, i.e.

$$\mathrm{NR}_n(\hat\beta_{n,w}^{LWS}) = 0. \tag{2.4}$$

Proof of Lemma 1. Because the weight function w is nonincreasing we can rewrite MF_n in (2.2) as

$$\mathrm{MF}_n(\beta) = \min_{\pi} \sum_{i=1}^{n} w\left(\frac{\pi(i)-1}{n}\right) r_i^2(\beta) \tag{2.5}$$

where minimization is taken over all permutations π on the set $\{1,\ldots,n\}$ (solving minimization in (2.5) leads (2.2), i.e. to give smaller weights to larger squared residuals). Suppose $\mathrm{NR}_n(\hat\beta_{n,w}^{LWS}) \neq 0$ and denote

$$\hat\beta_{n,w}^{W} := \arg\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} w\left(\frac{\pi_0(i,\hat\beta_{n,w}^{LWS})-1}{n}\right) r_i^2(\beta), \tag{2.6}$$

i.e. $\hat{\beta}^W_{n,w}$ is equal to WLS with weights given by $\hat{\beta}^{LWS}_{n,w}$. Notice $\mathrm{NR}_n(\hat{\beta}^{LWS}_{n,w}) \neq 0$ implies $\hat{\beta}^{LWS}_{n,w}$ does not satisfy normal equations of WLS with weights given by (2.6). Hence $\hat{\beta}^{LWS}_{n,w}$ is not solution of minimization (2.6) and we obtain

$$\begin{aligned}\mathrm{MF}_n(\hat{\beta}^W_{n,w}) &\leq \sum_{i=1}^n w\left(\frac{\pi_0(i,\hat{\beta}^{LWS}_{n,w})-1}{n}\right) r_i^2(\hat{\beta}^W_{n,w}) \\ &< \sum_{i=1}^n w\left(\frac{\pi_0(i,\hat{\beta}^{LWS}_{n,w})-1}{n}\right) r_i^2(\hat{\beta}^{LWS}_{n,w}) = \mathrm{MF}_n(\hat{\beta}^{LWS}_{n,w})\end{aligned} \tag{2.7}$$

where the first inequality comes from (2.5) and the second one from definition of $\hat{\beta}^W_{n,w}$. Hence we have $\mathrm{MF}_n(\hat{\beta}^W_{n,w}) < \mathrm{MF}_n(\hat{\beta}^{LWS}_{n,w})$ which is in contradiction with definition of $\hat{\beta}^{LWS}_{n,w}$ and thus $\mathrm{NR}_n(\hat{\beta}^{LWS}_{n,w}) = 0$. □

Denote by $F(x,z)$ the distribution function (d.f.) of $(p+1)$-dimensional vector $(X_i, Z_i) = (X_{i1}, \dots, X_{ip}, Z_i)$ and $F_n(x,z)$ the corresponding empirical distribution function (e.d.f.) obtained from random vectors $(X_1, Z_1), \dots, (X_n, Z_n)$. In this paper we suppose distribution functions to be left continuous. Now we rewrite the LWS estimator in the form of statistical functional, i.e. as a function of e.d.f.

Denote by $F_t(y)$ the d.f. of random variable $|r_i(\beta_0 + t)| = |X_i^T t - Z_i|$ and corresponding e.d.f. as $F_{n,t}(y)$. We see that $F_{n,t}(|X_i^T t - Z_i|) = (\pi_0(i, \beta_0 + t) - 1)/n$. Hence we can rewrite the function $\mathrm{NR}_n(\beta)$ (see (2.3)) as follows

$$\begin{aligned}\frac{1}{n}\mathrm{NR}_n(\beta) &= \frac{1}{n}\sum_{i=1}^n w\left(\frac{\pi_0(i,\beta)-1}{n}\right)\left(X_i^T\beta - Y_i\right) X_i \\ &= \frac{1}{n}\sum_{i=1}^n w\left(F_{n,t}\left(|X_i^T t - Z_i|\right)\right)(X_i^T t - Z_i) X_i\end{aligned} \tag{2.8}$$

where $t := \beta - \beta_0$. Clearly the last term in (2.8) can be rewritten as an integral with respect to e.d.f. F_n. So we define for an arbitrary $(p+1)$-dimensional d.f. $G(x,z)$ (where $x \in \mathbb{R}^p$ and $z \in \mathbb{R}$) the following statistical functional

$$\mathrm{NR}(\beta_0 + t, G) := \int w\left(G_t\left(|x^T t - z|\right)\right)(x^T t - z) x \,\mathrm{d}G(x,z) \tag{2.9}$$

where G_t is d.f. of random variable $|X_G^T t - Z_G|$ and vector (X_G, Z_G) has d.f. G. Note that $\mathrm{NR}(\beta_0 + t, G)$ is p-dimensional vector.

It is easily seen that for $G := F_n$ the integral in (2.9) is equal to (2.8) and then

$$\frac{1}{n}\mathrm{NR}_n(\beta) = \mathrm{NR}(\beta, F_n). \tag{2.10}$$

Another very useful way how to rewrite $\mathrm{NR}_n(\beta)$ is the following. We can reorder the summation in the second term in (2.8)

$$\frac{1}{n}\mathrm{NR}_n(\beta) = \frac{1}{n}\sum_{h=1}^n w\left(\frac{h-1}{n}\right)\sum_{i=1}^n \left(X_i^T\beta - Y_i\right) X_i \, I\left\{|Y_i - X_i^T\beta| = r_{|h|}(\beta)\right\} \tag{2.11}$$

and hence rewrite it

$$\frac{1}{n}\mathrm{NR}_n(\beta) = \sum_{h=1}^{n} w_{h,n}\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i^T t - Z_i\right) X_i\, I\left\{\left|Z_i - X_i^T t\right| \le r_{|h|}(\beta_0 + t)\right\}\right] \tag{2.12}$$

where $t := \beta - \beta_0$ and

$$w_{h,n} = w\left(\frac{h-1}{n}\right) - w\left(\frac{h}{n}\right) \tag{2.13}$$

(recall that $w(1) = 0$). Since we are working with continuous random errors (see (1.1)) we need not take into account the case $r_{|h|}(\beta) = r_{|l|}(\beta)$ for $h \neq l$.

Now in the same way as in the previous situation we define for any d.f. $G(x, z)$ (where $x \in \mathbb{R}^p$ and $z \in \mathbb{R}$) the following statistical functional

$$\mathrm{NR}_\alpha^*(\beta_0 + t, G) := \int (x^T t - z) x\, I\left\{|z - x^T t| \le G_t^{-1}(\alpha)\right\}\, \mathrm{d}G(x, z) \tag{2.14}$$

where G_t is defined as before. If we choose G equal to e.d.f. F_n we get $G_t^{-1}(\alpha) = F_{n,t}^{-1}(\alpha) = r_{|h|}(\beta_0 + t)$ for $\alpha \in \left(\frac{h-1}{n}, \frac{h}{n}\right]$ and then

$$\mathrm{NR}_\alpha^*(\beta_0 + t, F_n) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^T t - Z_i\right) X_i I\left\{\left|Z_i - X_i^T t\right| \le r_{|h|}(\beta_0 + t)\right\} \tag{2.15}$$

for $\alpha \in \left(\frac{h-1}{n}, \frac{h}{n}\right]$, which is the term in the brackets in (2.12). Define for any d.f. $G(x, z)$ (where $x \in \mathbb{R}^p$ and $z \in \mathbb{R}$)

$$\mathrm{NR}(\beta, G) := \int_0^1 \mathrm{NR}_\alpha^*(\beta, G)\, \mathrm{d}w^*(\alpha) \tag{2.16}$$

where $w^*(\alpha) := w(0) - w(\alpha)$. Hence for $G := F_n$ (2.16) is equal to (2.12). This is because $\mathrm{NR}_\alpha^*(\beta, F_n)$ is piecewise constant with respect to α and therefore

$$\mathrm{NR}(\beta, F_n) = \sum_{h=1}^{n} w_{h,n}\mathrm{NR}_{h/n}^*(\beta, F_n) \tag{2.17}$$

which is equal to (2.12) (see (2.15)).

Finally define the statistical functional $T(G)$ as a solution of normal equations $\mathrm{NR}(\beta, G) = 0$, i. e. it holds $\mathrm{NR}(T(G), G) = 0$ for any d.f. $G(x, z)$ (where $x \in \mathbb{R}^p$ and $z \in \mathbb{R}$). Functional T is not explicitly defined and generally there are more solutions to the normal equations. But we can choose T in such a way, that

$$T(F_n) = \hat{\beta}_n^{LWS}, \tag{2.18}$$

i. e. the statistical functional T represents the LWS estimator.

3. ASYMPTOTIC PROPERTIES OF LWS FOR LOCATION MODEL

Let us restrict ourselves to the location model (1.2) in this section. The following assumptions will be needed throughout the paper.

A1: The weight function w is nonincreasing and bounded with derivative existing almost everywhere. Moreover, it is positive on some neighbourhood of zero, $w(\alpha) = 0$ for $\alpha \in (\overline{\alpha}, 1)$ where $0 < \overline{\alpha} < 1$ and $\int_0^1 w(\alpha)\,\mathrm{d}\alpha > 0$.

A2: Random errors $Z_1, \dots, Z_n$ are i.i.d. and have continuous distribution with distribution function F_Z and density f_Z. This density is bounded, symmetric, strictly decreasing on $(0, \infty)$ and $f_Z(x) > 0$ for $x \in \mathbb{R}$. Random errors have finite second moments and f'_Z exists everywhere.

The substantial condition is **A2**. Symmetric and unimodal density is a very important condition for consistency. The counter example is obvious. Suppose just one dimensional data with symmetric density of observations which have two sharp peaks – one around -1 and one around 1. I. e. approximately one half of data is around -1 and one half around 1. Hence the LTS with $\alpha = 0.5$ estimates value close to -1 or 1 because the LTS tries to fit 50% of data. But we expect the value around zero, which is the expectation value of observations. Classical least squares will be consistent – it estimates value close to zero.

Under the proposed conditions we will prove not only $\sqrt{n}$-consistency of the LWS estimator for the location model but also $\sqrt{n}$-consistency of any sequence of solutions of normal equation. We will prove asymptotic normality under stronger conditions.

Theorem 1. Let $\hat{\beta}^*_n$ be an arbitrary sequence of solutions of normal equations for the location model (i. e. $\mathrm{NR}_n(\hat{\beta}^*_n) = 0$). Then under **A1** and **A2** this sequence is a $\sqrt{n}$-consistent estimator of β_0.

Moreover, if the weight function is piecewise constant, i. e.

$$w(\alpha) = \sum_{j=1}^{J} \lambda_j I\{\alpha \le \alpha_j\} \tag{3.1}$$

for some $J \in \{1, 2, \dots\}$, $\lambda_j > 0$ and $\alpha_j \in (0, 1)$ then

$$\sqrt{n}\left(\hat{\beta}^*_n - \beta_0\right) \to_D N(0, V^2_\infty) \tag{3.2}$$

where the asymptotic variance is

$$V^2_\infty = \frac{\displaystyle\int x^2 w^2\left(F_{|Z|}(|x|)\right) f_Z(x)\,\mathrm{d}x}{\left(\displaystyle\int x\, w\left(F_{|Z|}(|x|)\right) f'_Z(x)\,\mathrm{d}x\right)^2} \tag{3.3}$$

and $F_{|Z|}$ is the distribution function of $|Z_i|$.

The proof of Theorem 1 is provided in Section 6. Theorem 1 obviously implies $\sqrt{n}$-consistency and asymptotic normality of LWS for the location model.

4. INFLUENCE FUNCTION FOR GENERAL REGRESSION

In this section we derive *the influence function* of the LWS estimator for a general regression model (1.1). Recall that the influence function is defined as the directional derivative of statistical functional $T(F)$ at F in the direction of one-point distribution function Δ_{x_0,z_0} (i. e. the Dirac measure at point (x_0, z_0))

$$\mathrm{IF}(x_0, z_0; T, F) = \lim_{\varepsilon \to 0^+} \frac{T\left((1-\varepsilon)F + \varepsilon\Delta_{x_0,z_0}\right) - T(F)}{\varepsilon}. \tag{4.1}$$

Influence function describes the effect of a contamination at the point (x_0, z_0) on the estimate, standardized by the mass of the contamination.

Theorem 2. Let conditions **A1** and **A2** be satisfied. Moreover, suppose that the $p \times p$ matrix $\mathsf{E}X_1X_1^T$ is positive definite. Then the influence function of LWS estimator defined in (4.1) is

$$\mathrm{IF}(x_0, z_0; T, F) = \left[\mathsf{E}X_1X_1^T\right]^{-1} \cdot x_0 \cdot \frac{z_0 \cdot w\left(F_{|Z|}(|z_0|)\right)}{-\int \left[z \cdot w\left(F_{|Z|}(|z|)\right) \cdot f_Z'(z)\right] \mathrm{d}z} \tag{4.2}$$

where $x_0 \in \mathbb{R}^p$ and $z_0 \in \mathbb{R}$.

The proof of Theorem 2 is provided in Section 6. It immediately follows that under conditions of Theorem 2 the right hand side of (4.2) is well defined and that similarly as for M-estimators the influence function of the LWS estimator can be bounded with respect to z_0, but it cannot be bounded with respect to x_0.

5. OPTIMALITY FOR THE LOCATION MODEL

In this section we want to answer the question: How shall we choose the weights? Theorems 1 and 2 indicate a close relation between LWS estimator and M-estimators for location model (1.2). This is because LWS for the location model has the same asymptotic variance V_∞^2 and similarly the influence function as M-estimator for the location model with score function (i. e. the function which generates the normal equation of M-estimator)

$$\psi(x) = x \cdot w\left(F_{|Z|}(|x|)\right). \tag{5.1}$$

Both estimators are, of course, different, they have, however, just the same asymptotic variance and influence function. In some cases we can also find an inverse formula to (5.1), i.e. we can calculate weight function w from a given ψ. If ψ is an antisymmetric function then the weight function w which satisfies equation (5.1) is

$$w(\alpha) = \frac{\psi\left(F_{|Z|}^{-1}(\alpha)\right)}{F_{|Z|}^{-1}(\alpha)}. \tag{5.2}$$

This relation together with results for optimality of redescending M-estimators helps us in expressing the optimal weight functions. Recall that the score function of redescending M-estimators satisfies $\psi(x) = 0$ for $|x| > r$ where r is a given constant. Noticing (5.2) we obtain $w(\alpha) = 0$ for $\alpha > \overline{\alpha}$ where $\overline{\alpha} = F_{|Z|}(r)$, i.e. we have relation between LWS estimators with a given breakdown point (i.e. given $\overline{\alpha}$) and redescending M-estimators with a given r.

Now under conditions **A1** and **A2** we can easily express the most B-robust estimator. Recall that the most B-robust estimator minimizes gross error sensitivity which is supremum of absolute value of the influence function. The most B-robust redescending M-estimator with given $r > 0$ is the ***skipped median*** (see Lemma A2)

$$\psi_{\mathrm{med}(r)}(x) := \mathrm{sign}(x) \cdot I\{|x| < r\}. \tag{5.3}$$

Noticing (5.2) we realize that the most B-robust LWS estimator in the set of all LWS estimators with given $\overline{\alpha}$ (recall $w(\alpha) = 0$ for $\alpha \geq \overline{\alpha}$) has weight function

$$w_1(\alpha) := \frac{1}{F_{|Z|}^{-1}(\alpha)} \cdot I\{\alpha < \overline{\alpha}\}, \tag{5.4}$$

where $\min\{\overline{\alpha}, 1-\overline{\alpha}\}$ represents the breakdown point. This is because LWS estimator with weight function w_1 and M-estimator with score function $\psi_{\mathrm{med}(r)}$ have the same influence function and hence the same gross error sensitivity. For any other weight function w with given $\overline{\alpha}$ the LWS estimator has the gross error sensitivity the same as redescending M-estimator with ψ given by (5.1) which is larger than the gross error sensitivity of M-estimator with $\psi_{\mathrm{med}(r)}$. Hence for the weight function w_1 the LWS estimator has minimal gross error sensitivity.

Unfortunately, the function w_1 is unbounded and hence for this type of weight function consistency can not be proven by our method (**A1** is not satisfied). But we can take $\min\{w_1(\alpha), K\}$ where K is any given positive constant. For large K we get an estimator which has the gross error sensitivity very close to the minimal value.

The next problem is that w_1 depends on the distribution of random errors, which is of course typically unknown. But on the other hand it does not change if we multiply random errors by some positive constant. Hence for normally distributed errors it does not depend on variance (if we change variance we only multiply w_1 by some positive constant). Figure 1 (left figure) shows w_1 for normally distributed errors and $\overline{\alpha} = 1$. For $\overline{\alpha} < 1$ we just cut the weight function at an appropriate point, i.e. we multiply it by $I\{\alpha < \overline{\alpha}\}$ (see (5.4)).

Next we will minimize asymptotic variance under conditions **A1** and **A2**. The redescending M-estimator (with given $r \in \mathbb{R}$) which minimizes asymptotic variance

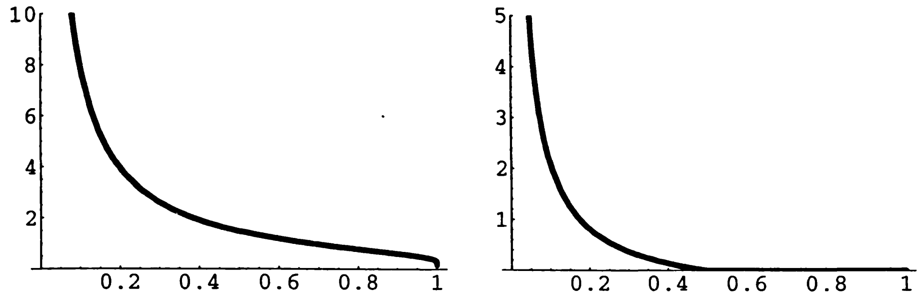

Fig. 1. Weight function w_1 for $\overline{\alpha} = 1$ (left) and w_3 for $\overline{\alpha} = 0.5$ (right).

has ψ-function (see Lemma A2)

$$\psi_r(x) := -\frac{f_Z'(x)}{f_Z(x)} \cdot I\{|x| < r\}. \tag{5.5}$$

For $r = \infty$ we get the maximum likelihood estimator. Using (5.2) we obtain the weight function of the corresponding LWS estimator (with given $\overline{\alpha}$) which minimizes the asymptotic variance

$$w_2\left(F_{|Z|}(x)\right) := -\frac{f_Z'(x)}{x \cdot f_Z(x)} \cdot I\{x < r\} \tag{5.6}$$

where $r > 0$. This follows by the same method as for the most B-robust LWS estimator.

For normally distributed errors is $f_Z'(x)/f_Z(x) = -x$ and we obtain $w_2(\alpha) = I\{\alpha < \overline{\alpha}\}$, i. e. the weight function of the LTS estimator. Hence the LTS estimator minimizes asymptotic variance in the group of LWS estimators with given breakdown point for normally distributed errors.

The last case is the most V-robust LWS estimator, i. e. we minimize the maximum of change of variance function divided by asymptotic variance. Recall that the change of variance function is the directional derivative of asymptotic variance V_∞^2 at F in the direction of one-point distribution function (i. e. it is an analogue of influence function in case of the asymptotic variance). For detailed definition of the change of variance function see [1], Section 2.5. The most V-robust redescending M-estimator (with given $r \in \mathbb{R}$) has the score function (see Lemma A2)

$$\psi_{\tanh(r)}(x) := (\kappa_r - 1)^{\frac{1}{2}} \tanh\left[\frac{1}{2}(\kappa_r - 1)^{\frac{1}{2}} B_r(r - |x|)\right] \operatorname{sign}(x) \cdot I\{|x| < r\} \tag{5.7}$$

where κ_r and B_r are appropriate constants (see Lemma A2). Using (5.2) we get the weight function of the most V-robust LWS estimator

$$w_3(\alpha) := \frac{\tanh\left[C_{\overline{\alpha}}\left(F_{|Z|}^{-1}(\overline{\alpha}) - F_{|Z|}^{-1}(\alpha)\right)\right]}{F_{|Z|}^{-1}(\alpha)} \cdot I\{\alpha < \overline{\alpha}\} \tag{5.8}$$

where $C_{\overline{\alpha}}$ is an appropriate constant. Some values of $C_{\overline{\alpha}}$ for normally distributed errors are in Table 1. Estimator generated by w_3 has continuous weight function. For this type of weight function we have not proved asymptotic normality, but we can take piecewise constant weight function which is close to w_3. For normally distributed errors and $\overline{\alpha} = 0.5$ is w_3 shown in Figure 1 (right figure).

Table 1. Values of $C_{\overline{\alpha}}$ for normally distributed errors.

$\overline{\alpha}$	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
$C_{\overline{\alpha}}$	0.507	0.500	0.496	0.494	0.493	0.491	0.489	0.486	0.482	0.475

Let us now compare three types of optimal weight functions. We will examine them from three points of view: breakdown point, asymptotic efficiency and gross error sensitivity. We denote the limit of variance of least squares estimator divided by variance of corresponding estimator as asymptotic efficiency.

In Tables 2 and 3 asymptotic efficiency and gross error sensitivity depending on $\overline{\alpha}$ for the LWS estimator with weight functions w_1, w_2 and w_3 and normally distributed errors are given. In the last column there are limit values for $\overline{\alpha} \to 1$.

Table 2. Asymptotic efficiency [%].

$\overline{\alpha}$	0.50	0.60	0.70	0.80	0.90	1
w_1	5.3	9.4	15.7	25.0	38.9	63.7
w_2	7.1	12.9	21.7	35.0	56.1	100.0
w_3	1.8	3.4	6.0	10.2	17.9	63.7

Table 3. Gross error sensitivity.

$\overline{\alpha}$	0.50	0.60	0.70	0.80	0.90	1
w_1	6.2	4.2	3.0	2.2	1.7	1.3
w_2	9.5	6.5	4.8	3.7	2.9	∞
w_3	19.0	11.4	7.7	5.3	3.6	1.3

We see that for weight functions w_1 and w_2 asymptotic efficiency and gross error sensitivity are similar. Estimator with w_1 is of course better with respect to the gross error sensitivity and with w_2 in the asymptotic efficiency.

We can also use a combination of these weight functions. For example for

$$w(x) = \min\{w_1(x), K\} \tag{5.9}$$

we get an estimator which has the gross error sensitivity and asymptotic efficiency between values for w_1 and w_2. For large K the weight function is close to w_1, for small K it is constant on $(0, \overline{\alpha})$ and hence equal to w_2.

Finally, recall that these three types of weight functions give possibilities which are optimal from different points of view. In real life situation we should use one of them or a combination of them (for example (5.9)) depending on the data set.

6. PROOFS

This section contains proofs of Theorems 1 and 2. First we prove Theorem 2. Particularly we derive the influence function for general regression. In the next part we restrict ourselves to the location model and we prove Theorem 1 (i.e. $\sqrt{n}$-consistency and asymptotic normality).

Proof of Theorem 2. The proof follows the lines of an analogous proof for M-estimator (see [1]). Namely, we plug into the normal equations (2.9) instead of G the contaminated distribution $F^\varepsilon := (1-\varepsilon)F + \varepsilon\Delta_{x_0,z_0}$, where $\varepsilon > 0$, $x_0 \in \mathbb{R}^p$, $z_0 \in \mathbb{R}$ and denote by Δ_{x_0,z_0} the d.f. of Dirac measure at point (x_0, z_0). Now we differentiate the equation with respect to ε in $\varepsilon = 0^+$ and calculate the influence function. Let us do this process step by step.

The normal equations for contaminated distribution F^ε are (see (2.9))

$$\begin{aligned} 0 = (1-\varepsilon) \int w\left(F_T^\varepsilon\left(|z - x^T T|\right)\right)(x^T T - z)x \,\mathrm{d}F(x,z) \\ + \varepsilon w\left(F_T^\varepsilon\left(|z_0 - x_0^T T|\right)\right)(x_0^T T - z_0)x_0 \end{aligned} \tag{6.1}$$

where $0 \in \mathbb{R}^p$ and $T = T(F^\varepsilon)$ is the solution of normal equations. Notice that symmetry of random errors implies $T(F^\varepsilon) = T(F) = 0$ for $\varepsilon = 0$. The F_T^ε can be rewritten as follows (see (2.9), definition of G_t)

$$\begin{aligned} F_T^\varepsilon(u) &= \int I\left\{|x_1^T T - z| < u\right\} \mathrm{d}F^\varepsilon(x_1, z) \\ &= (1-\varepsilon)\int \left[F_Z(x_1^T T + u) - F_Z(x_1^T T - u)\right] \mathrm{d}F_X(x_1) \\ &\quad + \varepsilon \cdot I\left\{|x_0^T T - z_0| < u\right\} \end{aligned} \tag{6.2}$$

for $u > 0$ where F_Z and F_X are marginal d.f. from F. For $\varepsilon = 0$ is $F_T^\varepsilon = F_{|Z|}$.

Let us now differentiate (6.1) with respect to ε and then put $\varepsilon = 0^+$ and recall that the definition of influence function implies

$$\mathrm{IF}(x_0, z_0; T, F) := \left[\frac{\partial}{\partial\varepsilon} T(F^\varepsilon)\right]_{\varepsilon=0^+}. \tag{6.3}$$

Finally we obtain

$$\begin{aligned} 0 = &\int w\left(F_{|Z|}(|z|)\right) \cdot x \cdot x^T \cdot \mathrm{IF}(x_0, z_0; T, F)\,\mathrm{d}F(x,z) \\ &- \int w'\left(F_{|Z|}(|z|)\right) \cdot z \cdot x \cdot \left[\frac{\partial}{\partial\varepsilon} F_T^\varepsilon(|x^T T - z|)\right]_{\varepsilon=0^+} \mathrm{d}F(x,z) \\ &- w\left(F_{|Z|}(|z_0|)\right) z_0 x_0. \end{aligned} \tag{6.4}$$

To finish our proof we should calculate corresponding derivative in the second term on the right hand side of (6.4). This we get by differentiating (6.2) for $u := |x^T T - z|$

and putting $\varepsilon = 0$. Finally we obtain

$$\left[\frac{\partial}{\partial\varepsilon} F_T^{\varepsilon}(|x^T T - z|)\right]_{\varepsilon=0+} = -2\,\mathrm{sign}(z) f_Z(|z|) x^T \mathrm{IF}(x_0, z_0; T, F) - F_{|Z|}(|z|) + I\left\{|z_0| < |z|\right\}. \tag{6.5}$$

Substituting (6.5) into (6.4) the second and third term in (6.5) vanish because they are symmetric functions with respect to z. Hence we obtain

$$w\left(F_{|Z|}(|z_0|)\right) z_0 x_0 = \int \left[w\left(F_{|Z|}(|z|)\right) \cdot z\right]' \cdot x x^T \cdot \mathrm{IF}(x_0, z_0; T, F)\, \mathrm{d}F(x, z) \tag{6.6}$$

because

$$\left[w\left(F_{|Z|}(|z|)\right) \cdot z\right]' = w\left(F_{|Z|}(|z|)\right) + w'\left(F_{|Z|}(|z|)\right) 2\,\mathrm{sign}(z) f_Z(|z|) z. \tag{6.7}$$

Using independency of random errors and regressors together with integration by parts implies

$$w\left(F_{|Z|}(|z_0|)\right) z_0 x_0 = \left[\mathsf{E} X_1 X_1^T\right] \mathrm{IF}(x_0, z_0; T, F) \int w\left(F_{|Z|}(|z|)\right) z f_Z'(x)\, \mathrm{d}z. \tag{6.8}$$

Now we can easily express $\mathrm{IF}(x_0, z_0; T, F)$ and finish the proof. □

Next we prove Theorem 1, first $\sqrt{n}$-consistency, next asymptotic normality. To prove $\sqrt{n}$-consistency we approximate the function $\mathrm{NR}(\beta, F_n)$ by $\mathrm{NR}(\beta, F)$. If we knew that the only solution of $\mathrm{NR}(\beta, F) = 0$ is $\beta = \beta_0$ and $\mathrm{NR}(\beta, F_n)$ is close to $\mathrm{NR}(\beta, F)$ for large n, then we would get that solution of $\mathrm{NR}(\beta, F_n) = 0$ is close to β_0. The following lemma shows that $\mathrm{NR}(\beta, F) = 0$ has the only solution for $\beta = \beta_0$ and that $\mathrm{NR}(\beta, F)$ is increasing at least linearly in some neighbourhood of β_0.

Lemma 2. Under conditions **A1** and **A2** for any $K > 0$ there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$|\mathrm{NR}(\beta_0 + t, F)| \geq \min\{\delta_1 |t|, \delta_2\} \qquad \text{for } t \in [-K, K]. \tag{6.9}$$

Proof of Lemma 2. For simplicity denote the distribution function and density of random variables Z_i by F and f respectively. By symmetry of density of random errors Z_i we get $\mathrm{NR}(\beta_0, F) = 0$ immediately. To finish the proof it suffices to show

$$\left[\frac{\partial}{\partial t} \mathrm{NR}(\beta_0 + t, F)\right]_{t=0} > 0 \tag{6.10}$$

and

$$\mathrm{NR}(\beta_0 + t, F) \neq 0 \qquad \text{for } t \neq 0. \tag{6.11}$$

Since $\mathrm{NR}(\beta_0, F) = 0$ (6.10) implies existence of constants $\varepsilon > 0$ and $\delta_1 > 0$ such that

$$|\mathrm{NR}(\beta_0 + t, F)| \geq \delta_1 |t| \qquad \text{for } |t| \leq \varepsilon. \tag{6.12}$$

Because the set $M := [-K, -\varepsilon] \cup [\varepsilon, K]$ is a compact subset of $\mathbb{R}$ and $\mathrm{NR}(\beta_0 + t, F)$ is continuous with respect to t (moreover, it is differentiable) then $|\mathrm{NR}(\beta_0 + t, F)|$ (restricted to M) attains minimum at some point $t_0 \in M$.

Define $\delta_2 := |\mathrm{NR}(\beta_0 + t_0, F)|$. Because $t_0 \neq 0$ we get from (6.11) that $\delta_2 > 0$. Hence

$$|\mathrm{NR}(\beta_0 + t, F)| \geq |\mathrm{NR}(\beta_0 + t_0, F)| = \delta_2 > 0 \qquad \text{for } \varepsilon \leq |t| \leq K. \tag{6.13}$$

Inequality (6.12) together with (6.13) imply (6.9).

Let us now prove (6.10) and (6.11). Noticing (2.9) and using substitution $z := t - y$ we realize that (6.10) follows from

$$\begin{aligned} \left[\frac{\partial}{\partial t}\mathrm{NR}(\beta_0 + t, F)\right]_{t=0} &= \left[\int_{-\infty}^{\infty} w(F_t(|t - z|))(t - z)\,\mathrm{d}F(z)\right]_{t=0} \\ = \left[\int_{-\infty}^{\infty} w(F_t(|y|))\,y f(t - y)\,\mathrm{d}y\right]_{t=0} &= \int_{-\infty}^{\infty} w(F_Z(|y|))(-y f'(y))\,\mathrm{d}y > 0. \end{aligned} \tag{6.14}$$

Finally, we prove (6.11). Suppose $t > 0$. Equation (2.9) and the same substitution yields

$$\begin{aligned} \mathrm{NR}(\beta_0 + t, F) &= \int_{-\infty}^{\infty} w(F_t(|t - z|))(t - z)\,\mathrm{d}F(z) \\ &= \int_{-\infty}^{\infty} w(F_t(|y|))\,y f(t - y)\,\mathrm{d}y = \int_{0}^{\infty} w(F_t(y))\,y\,[f(t - y) - f(t + y)]\,\mathrm{d}y. \end{aligned} \tag{6.15}$$

Notice $[f(t - y) - f(t + y)] > 0$ for $t > 0$ and $y > 0$. This is because $|t-y| < |t+y|$ for $t > 0$ and $y > 0$ and because density f is symmetric and decreasing on $(0, \infty)$. The weight function is nonnegative and on some neighbourhood of zero positive, hence the integrand in (6.15) is nonnegative and on some interval positive which implies $\mathrm{NR}(\beta_0 + t, F) > 0$ for $t > 0$. In the same way we can prove $\mathrm{NR}(\beta_0 + t, F) < 0$ for $t < 0$ and so we omit it here. □

Now we can make the first step in proving Theorem 1, i. e. $\sqrt{n}$-consistency. The proof of $n^{\frac{1}{4}}$-consistency of LWS for location model was shown in [5]. We will use the same method but instead of working with function MF (which is statistical functional based on MF_n defined in an analogical way like NR) we will use function NR. The function MF is quadratic in the neighbourhood of β_0 for the theoretical distribution F . Hence proving

$$\sup\{|\mathrm{MF}(\beta, F_n) - \mathrm{MF}(\beta, F)|, |\beta| \leq K\} = \mathcal{O}_p(n^{-\frac{1}{2}}) \tag{6.16}$$

gives just $n^{\frac{1}{4}}$-consistency. Because the function NR is linear in a neighbourhood of β_0 for the theoretical distribution, analogous approximation for normal equation gives the $\sqrt{n}$-consistency.

Proof of Theorem 1 – $\sqrt{n}$-consistency. For simplicity let us denote the distribution function and density of random variables Z_i by F and f respectively. We denote the empirical distribution function based on $Z_1, \dots, Z_n$ by F_n.

To prove consistency we will use the following invariance principle result

$$\sqrt{n}\|F_n - F\|_\infty = \mathcal{O}_p(1) \tag{6.17}$$

where $\|.\|_\infty$ is the supremum norm (for details see [3], Section 2.5.13).

Fix $\varepsilon > 0$ and a probability space $(\Omega, \mathcal{A}, \mathcal{P})$. Equality (6.17) together with the weak law of large numbers gives the following. There exist positive constants K_1, K_2 and n_1 (depending on ε) such that for any $n > n_1$ is

$$\mathbb{P}(B_n) > 1 - \varepsilon \tag{6.18}$$

where

$$B_n = \left[\frac{1}{n}\sum_{i=1}^{n}|Z_i| \le K_1 \text{ and } |F_n(z) - F(z)| \le n^{-\frac{1}{2}}K_2 \text{ for } z \in \mathbb{R}\right] \subset \Omega. \tag{6.19}$$

Now we find constants K_3 and n_0 such that for any sequence $\hat{\beta}_n^*$ which satisfies $\mathrm{NR}_n(\hat{\beta}_n^*) = 0$ the inequality $|\hat{\beta}_n^* - \beta_0| \le n^{-\frac{1}{2}}K_3$ holds on B_n for $n \ge n_0$. This implies the $\sqrt{n}$-consistency of sequence $\hat{\beta}_n^*$. Let us do so and find K_3.

First we find constants K_4 and n_2 such that for any $t \in \mathbb{R}$, $|t| > K_4$ and $n > n_2$ it holds $\mathrm{NR}(\beta_0 + t, F_n) \ne 0$ on B_n. I. e. for any $\omega \in B_n$ and large n all solutions t of the normal equation $\mathrm{NR}(\beta_0 + t, F_n) = 0$ are in the interval $[-K_4, K_4]$.

Combining definition (2.3) and equality (2.10) gives

$$\mathrm{NR}(\beta_0 + t, F_n) = \frac{1}{n}\mathrm{NR}_n(\beta_0 + t) = \frac{1}{n}\sum_{i=1}^{n} w\left(\frac{\pi_0(i,\beta) - 1}{n}\right)(t - Z_i) \tag{6.20}$$

and hence

$$\mathrm{NR}(\beta_0 + t, F_n) = t\frac{1}{n}\sum_{h=1}^{n} w\left(\frac{h-1}{n}\right) - \frac{1}{n}\sum_{i=1}^{n} w\left(\frac{\pi_0(i,\beta) - 1}{n}\right) Z_i. \tag{6.21}$$

We find upper bound for the absolute value of the second summand of (6.21) and lower bound for the absolute value of the first one. Because the weight function is nonincreasing we can write

$$\frac{1}{n}\left|\sum_{i=1}^{n} w\left(\frac{\pi_0(i,\beta) - 1}{n}\right) Z_i\right| \le w(0)\frac{1}{n}\sum_{i=1}^{n}|Z_i| \le w(0)K_1 \tag{6.22}$$

where the last inequality holds on B_n. By condition **A1**

$$\frac{1}{n}\sum_{h=1}^{n} w\left(\frac{h-1}{n}\right) \to \int_0^1 w(\alpha)\,\mathrm{d}\alpha = K_5 > 0 \tag{6.23}$$

therefore there exists n_2 such that

$$\frac{1}{n}\sum_{h=1}^{n} w\left(\frac{h-1}{n}\right) \geq \frac{1}{2}K_5 \tag{6.24}$$

for any $n > n_2$. Combining (6.21), (6.22) and (6.24) we have that for $n > n_2$ and $|t| > 2w(0)K_1/K_5$

$$|\mathrm{NR}(\beta_0 + t, F_n)| \geq \frac{1}{2}|t|K_5 - w(0)K_1 > 0 \tag{6.25}$$

on B_n. So the desired constant is $K_4 := 2w(0)K_1/K_5$.

Now we will look at the behaviour of the function $\mathrm{NR}(\beta_0+t, F_n)$ for $t \in [-K_4, K_4]$. Suppose that there exists a constant K_6 such that on B_n for NR^*_α (see (2.14)) the inequality

$$|\mathrm{NR}^*_\alpha(\beta_0 + t, F_n) - \mathrm{NR}^*_\alpha(\beta_0 + t, F)| \leq n^{-\frac{1}{2}}K_6 \tag{6.26}$$

holds for $|t| \leq K_4$ and $\alpha \in (0, \overline{\alpha})$. Formula (2.16) implies

$$|\mathrm{NR}(\beta_0 + t, F_n) - \mathrm{NR}(\beta_0 + t, F)| \leq n^{-\frac{1}{2}}K_6 w(0) \tag{6.27}$$

(notice NR does not depend on NR^*_α for $\alpha \geq \overline{\alpha}$ because $w(\alpha) = 0$ for $\alpha \geq \overline{\alpha}$). By Lemma 2 (for $K := K_4$) there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$|\mathrm{NR}(\beta_0 + t, F)| \geq \min\{\delta_1|t|, \delta_2\} \tag{6.28}$$

for $|t| \leq K_4$. Combining (6.27) with (6.28) yields

$$\begin{aligned} |\mathrm{NR}(\beta_0 + t, F_n)| &\geq |\mathrm{NR}(\beta_0 + t, F)| - |\mathrm{NR}(\beta_0 + t, F_n) - \mathrm{NR}(\beta_0 + t, F)| \\ &\geq \min\{\delta_1|t|, \delta_2\} - n^{-\frac{1}{2}}K_6 w(0) \end{aligned} \tag{6.29}$$

where the last inequality holds on B_n for $|t| \leq K_4$. Let us define $K_7 := w(0)K_6/\delta_1$ and $n_3 := (w(0)K_6/\delta_2)^2$. We see that

$$\min\{\delta_1|t|, \delta_2\} - n^{-\frac{1}{2}}K_6 w(0) > 0 \tag{6.30}$$

for $|t| > n^{-\frac{1}{2}}K_7$ and $n > n_3$. Now (6.29) together with (6.30) imply that on B_n there is no solution of normal equation $\mathrm{NR}(\beta_0 + t, F_n) = 0$ for which it would hold $n^{-\frac{1}{2}}K_7 < |t| \leq K_4$.

Putting together the results for $|t| > K_4$ and for $t \in [-K_4, K_4]$ we obtain

$$\mathrm{NR}(\beta_0 + t, F_n) \neq 0 \tag{6.31}$$

on B_n for $|t| > n^{-\frac{1}{2}} K_7$ and $n > n_0 := \max\{n_1, n_2, n_3\}$. Hence we see that any solution of normal equation is close to β_0 on B_n and the desired constant is $K_3 := K_7$.

To finish the proof we have to find a constant K_6 such that (6.26) holds. The rest of the proof follows the lines of a similar proof in [5]. Since $|t| \le K_4$ and on B_n the empirical distribution is approximated by theoretical one, we have existence of constant K_8 such that

$$|F_{n,t}^{-1}(\alpha)| \le K_8, \qquad |F_t^{-1}(\alpha)| \le K_8 \tag{6.32}$$

for $\alpha \le \overline{\alpha}$ and $n > n_4$ and existence of K_9 such that

$$|F_{n,t}^{-1}(\alpha) - F_t^{-1}(\alpha)| \le n^{-\frac{1}{2}} K_9 \tag{6.33}$$

for $\alpha \le \overline{\alpha}$ and $n > n_5$. Since the theoretical d.f. F and the e.d.f. F_n are close, the theoretical d.f. F is strictly increasing (density f is positive) and t is bounded ($|t| \le K_4$). Hence the quantiles in (6.33) are close for $\alpha \in (0, \overline{\alpha})$. That is since α is not close to 1 ($\alpha \le \overline{\alpha} < 1$).

Denote for simplicity $u_n := F_{n,t}^{-1}(\alpha)$ and $u := F_t^{-1}(\alpha)$ and rewrite the definition (2.14) of NR_α^* for $\alpha \le \overline{\alpha}$

$$\begin{aligned}\mathrm{NR}_\alpha^*(t, F_n) - \mathrm{NR}_\alpha^*(t, F) &= \int_{t-u_n}^{t+u_n} (t-z)\,\mathrm{d}F_n(z) - \int_{t-u}^{t+u} (t-z)\,\mathrm{d}F(z) \\ = \int_{t-u_n}^{t-u} (t-z)\,\mathrm{d}F(z) &+ \int_{t+u}^{t+u_n} (t-z)\,\mathrm{d}F(z) + \int_{t-u_n}^{t+u_n} (t-z)\,\mathrm{d}(F_n - F)(z).\end{aligned} \tag{6.34}$$

Now we find an upper bound for each of the three terms in (6.34). Using (6.32) and (6.33) the first term can be bounded as

$$\begin{aligned}\left|\int_{t-u_n}^{t-u} (t-z)\,\mathrm{d}F(z)\right| &\le \int_{t-u_n}^{t-u} |t-z| f(z)\,\mathrm{d}z \\ &\le |u_n - u| \max\{u_n, u\} M_f \le n^{-\frac{1}{2}} K_9 K_8 M_f\end{aligned} \tag{6.35}$$

where $M_f := \sup\{f(x), x \in \mathbb{R}\} < \infty$. The same upper bound can be used for the second term.

Finally, we find an upper bound for the third term in (6.34). We use the following general formula which holds for any d.f. G and $a, b \in \mathbb{R}$

$$\begin{aligned}\int_a^b z\,\mathrm{d}G(z) &= \int_a^b \int_a^z 1\,\mathrm{d}y\,\mathrm{d}G(z) + a(G(b) - G(a)) \\ = \int_a^b (G(b) - G(y))\,\mathrm{d}y &+ a(G(b) - G(a)) = bG(b) - aG(a) - \int_a^b G(y)\,\mathrm{d}y.\end{aligned} \tag{6.36}$$

Taking difference of equation (6.36) for $G := G_1$ and $G := G_2$ implies

$$\left|\int_a^b z\,\mathrm{d}(G_1 - G_2)(z)\right| \le 2(|b| + |a|) \sup\{|G_1(x) - G_2(x)| : x \in \mathbb{R}\}. \tag{6.37}$$

Using (6.37) for $G_1 := F_n$ and $G_2 := F$ in the third term in (6.34) implies (remind $|t| \leq K_4$, $u \leq K_8$, $u_n \leq K_8$ and $\|F - F_n\|_\infty \leq n^{-\frac{1}{2}} K_2$)

$$\begin{aligned}
&\left| \int_{t-u_n}^{t+u_n} (t-z)\, \mathrm{d}(F_n - F)(z) \right| \\
&\quad \leq \left| t \int_{t-u_n}^{t+u_n} 1\, \mathrm{d}(F_n - F)(z) \right| + \left| \int_{t-u_n}^{t+u_n} z\, \mathrm{d}(F_n - F)(z) \right| \\
&\quad \leq 2|t|\, \|F - F_n\|_\infty + 2(|t + u_n| + |t - u_n|) \|F - F_n\|_\infty \\
&\quad \leq (6K_4 + 4K_8) n^{-\frac{1}{2}} K_2 \qquad (6.38)
\end{aligned}$$

on B_n. Now we put together the upper bounds of all three terms in (6.34) (see (6.35) and (6.38)) and obtain

$$|\mathrm{NR}_\alpha^*(t, F_n) - \mathrm{NR}_\alpha^*(t, F)| \leq n^{-\frac{1}{2}} (2K_9 K_8 M_f + K_2(6K_4 + 4K_8)). \qquad (6.39)$$

Hence we can take $K_6 := 2K_9K_8M_f + K_2(6K_4 + 4K_8)$ in (6.26) which finishes the proof. □

To prove asymptotic normality we will use the same method as was used for M-estimators in [2]. We will use asymptotic linearity which was provided in [5] (see Lemma A1). Asymptotic linearity together with $\sqrt{n}$-consistency gives us asymptotic normality.

Proof of Theorem 1 – asymptotic normality. Fix $\varepsilon > 0$. Because the sequence $\hat{\beta}_n^*$ is $\sqrt{n}$-consistent there exist constants K_1 and n_1 such that

$$\mathbb{P}\left(\sqrt{n} \left(\hat{\beta}_n^* - \beta_0 \right) \geq K_1 \right) \leq \varepsilon \qquad (6.40)$$

for $n > n_1$. By Lemma A1 for $M := K_1$ there exist constants K_2 and n_2 such that

$$\mathbb{P}\left(n^{-\frac{1}{4}} \sup_{|t| < K_1} \left| n\, \mathrm{NR}\left(\beta_0 + n^{-\frac{1}{2}} t, F_n \right) - n\, \mathrm{NR}(\beta_0, F_n) - n^{\frac{1}{2}} t R_w \right| > K_2 \right) \leq \varepsilon \qquad (6.41)$$

for $n > n_2$. Now we choose

$$t := t_n = \sqrt{n} \left(\hat{\beta}_n^* - \beta_0 \right). \qquad (6.42)$$

Combining (6.40), (6.41), (6.42) and $\mathrm{NR}(\hat{\beta}_n^*, F_n) = 0$ gives

$$\mathbb{P}\left(n^{-\frac{1}{4}} \left| n\, \mathrm{NR}(\beta_0, F_n) + n \left(\hat{\beta}_n^* - \beta_0 \right) R_w \right| > K_2 \right) \leq 2\varepsilon \qquad (6.43)$$

for $n > n_0 := \max\{n_1, n_2\}$.

Because $\varepsilon > 0$ was arbitrary we obtain

$$n\, \mathrm{NR}(\beta_0, F_n) + n \left(\hat{\beta}_n^* - \beta_0 \right) R_w = \mathcal{O}_p\left(n^{\frac{1}{4}} \right). \qquad (6.44)$$

The last equation can be rewritten as

$$\sqrt{n}\left(\hat{\beta}_n^* - \beta_0\right) = -R_w^{-1}\sqrt{n}\,\mathrm{NR}(\beta_0, F_n) + \mathcal{O}_p\left(n^{-\frac{1}{4}}\right). \tag{6.45}$$

Hence asymptotic behaviour of the left hand side of (6.45) depends only on behaviour of random variable $\sqrt{n}\,\mathrm{NR}(\beta_0, F_n)$.

Denote $Z_{|h|} := r_{|h|}(\beta_0)$ for $h = 1, \dots, n$ (i. e. the order statistics of $|Z_1|, \dots, |Z_n|$) and $S_i := \mathrm{sign}(Z_i)$ for $i = 1, \dots, n$. Notice that random variables S_i and $|Z_i|$ are independent because distribution of Z_i is symmetric. Hence by (2.10) and (2.12) we have

$$\sqrt{n}\,\mathrm{NR}(\beta_0, F_n) = -\frac{1}{\sqrt{n}}\sum_{h=1}^{n}\sum_{i=1}^{n} w_{h,n} S_i |Z_i| I\left\{|Z_i| \le Z_{|h|}\right\}. \tag{6.46}$$

Because sequences $S_1, \dots, S_n$ and $|Z_1|, \dots, |Z_n|$ are independent we can order random variables $|Z_i|$ in the inner summation (i. e. replace $|Z_i|$ by $Z_{|i|}$) and the distribution of the right hand side of (6.46) remains unchanged. Finally we obtain (see also (2.13), denote $=_D$ equation of distributions)

$$\begin{aligned}\sqrt{n}\,\mathrm{NR}(\beta_0, F_n) &=_D -\frac{1}{\sqrt{n}}\sum_{h=1}^{n}\sum_{i=1}^{n} w_{h,n} S_i\, Z_{|i|}\, I\left\{Z_{|i|} \le Z_{|h|}\right\} \\ &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{h=i}^{n} w_{h,n} S_i\, Z_{|i|} = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) S_i\, Z_{|i|}.\end{aligned} \tag{6.47}$$

Define a σ-algebra $\mathcal{A}_Z := \sigma\{|Z_i|, i = 1, 2, \dots\}$. Notice that the summands in (6.47) are conditionally independent therefore we can use central limit theorem

$$\left[\frac{\sqrt{n}}{V_{n,w}}\mathrm{NR}(\beta_0, F_n)\middle|\mathcal{A}_Z\right] \to \mathcal{N}(0,1) \tag{6.48}$$

where

$$V_{n,w}^2 := \mathrm{var}\left(\sqrt{n}\,\mathrm{NR}(\beta_0, F_n)\middle|\mathcal{A}_Z\right) = \frac{1}{n}\sum_{i=1}^{n} w^2\left(\frac{i-1}{n}\right) Z_{|i|}^2. \tag{6.49}$$

Using the same method as in the proof of $\sqrt{n}$-consistency (i. e. (6.17) and consecutive steps) we obtain

$$V_{n,w}^2 = \int x^2 w^2\left(F_n^{|Z|}(x)\right)\,\mathrm{d}F_n(x) \to_P V_w^2 := \int x^2 w^2\left(F_{|Z|}(x)\right)\,\mathrm{d}F(x) \tag{6.50}$$

where $F_n^{|Z|}$ is e.d.f. based on random variables $|Z_1|, \dots, |Z_n|$ and $\to_P$ denotes convergence in probability.

Because the limit distribution in (6.48) does not depend on $\mathcal{A}_Z$ the convergence (6.48) together with (6.50) imply

$$\sqrt{n}\,\mathrm{NR}(\beta_0, F_n) \to \mathcal{N}(0, V_w^2). \tag{6.51}$$

The last convergence together with (6.45) finish the proof and give us the formula for the asymptotic variance $V_\infty^2 = R_w^{-2} \cdot V_w^2$ (see (3.3)). □

APPENDIX

Lemma A1. Under conditions **A1** and **A2** for piecewise constant weight function the normal equation of LWS estimator for the location model is asymptotically linear in the following sense: For any $M \in (0, \infty)$ it holds

$$n^{-\frac{1}{4}} \sup_{|t|<M} \left| n\mathrm{NR}\left(\beta_0 + n^{-\frac{1}{2}}t, F_n\right) - n\mathrm{NR}\left(\beta_0, F_n\right) - n^{\frac{1}{2}} t R_w \right| = \mathcal{O}_p(1) \tag{A.1}$$

where

$$\begin{aligned} R_w &= \int_0^1 \left[\alpha - 2F_{|Z|}^{-1}(\alpha) \cdot f_Z\left(F_{|Z|}^{-1}(\alpha)\right)\right] \mathrm{d}w^*(\alpha) \\ &= -\int_{-\infty}^{\infty} x\, w\left(F_{|Z|}(|x|)\right) f_Z'(x)\, \mathrm{d}x \end{aligned} \tag{A.2}$$

and $w^*(\alpha) = w(0) - w(\alpha)$.

Proof of Lemma A1 is based on the following principle. We split the main term in (A.1) into several parts that can be written as stochastic processes in t. For each part the convergence in distribution is proved. The result was presented in [5] and a detailed proof can be found in [4].

Lemma A2. a) The most B-robust (i. e. minimizing gross error sensitivity) redescending M-estimator is

$$\psi_{\mathrm{med}(r)}(x) := \mathrm{sign}(x) \cdot I\left\{|x| < r\right\}. \tag{A.3}$$

b) The redescending M-estimator which minimizes asymptotic variance is

$$\psi_r(x) := -\frac{f_Z'(x)}{f_Z(x)} \cdot I\left\{|x| < r\right\}. \tag{A.4}$$

c) The most V-robust (i. e. minimizes maximum of change of variance function divided by asymptotic variance) redescending M-estimator is

$$\psi_{\tanh(r)}(x) := (\kappa_r - 1)^{\frac{1}{2}} \tanh\left[\frac{1}{2}(\kappa_r - 1)^{\frac{1}{2}} B_r(r - |x|)\right] \mathrm{sign}(x) \cdot I\left\{|x| < r\right\} \tag{A.5}$$

where κ_r and B_r are given constants such that

$$\int \psi_{\tanh(r)}^2(x)\, \mathrm{d}F(x) = 1, \qquad \int \psi_{\tanh(r)}'(x)\, \mathrm{d}F(x) = B_r. \tag{A.6}$$

For proof of Lemma A2 see [1], Section 2.6.

ACKNOWLEDGEMENT

The research was supported by the Grant Agency of the Czech Republic under Grant 402/03/0084 and by the Ministry of Education, Youth and Sports of the Czech Republic under Project MS 113200008.

(Received December 16, 2002.)

REFERENCES

[1] F. R. Hampel, E. M. Ronchetti, R. J. Rousseeuw, and W. A. Stahel: Robust Statistics – The Approach Based on Influence Function. Wiley, New York 1986.

[2] J. Jurečková: Asymptotic representation of M-estimators of location. Math. Operationsforsch. Statist., Ser. Statistics *11* (1980), 1, 61–73.

[3] J. Jurečková and P. K. Sen: Robust Statistical Procedures. Wiley, New York 1996.

[4] L. Mašíček: Konzistence odhadu LWS pro parametr polohy (Consistency of LWS estimator for location model). KPMS Preprint *25*, Department of Probability and Mathematical Statistics, Faculty of Mathemetics and Physics, Charles University, Prague 2002.

[5] L. Mašíček: Konzistence odhadu LWS pro parametr polohy (Consistency of LWS estimator for location model). In: ROBUST'2002 (J. Antoch, G. Dohnal, and J. Klaschka, eds.), Union of the Czech Mathematicians and Physicists 2002, pp. 240–246.

[6] P. J. Rousseeuw and A. M. Leroy: Robust Regression and Outlier Detection. Wiley, New York 1987.

[7] J. Á. Víšek: Regression with high breakdown point. In: ROBUST'2000 (J. Antoch and G. Dohnal, eds.), Union of the Czech Mathematicians and Physicists 2001, pp. 324–356.

[8] J. Á. Víšek: A new paradigm of point estimation. In: Data Analysis 2000 – Modern Statistical Methods – Modelling, Regression, Classification and Data Mining (K. Kupka, ed.), TRILOBYTE Software 2001, pp. 195–230.

Libor Mašíček, Department of Probability and Statistics, Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mails: masicek@karlin.mff.cuni.cz, libor@matfyz.cz