Ann. Inst. Statist. Math.
Vol. 48, No. 3, 469–495 (1996)

SENSITIVITY ANALYSIS OF *M*-ESTIMATES

JAN ÁMOS VÍŠEK*

Department of Stochastic Informatics, Institute of Information Theory and Automation, Academy of Sciences the Czech Republic,
and
Department of Macroeconomics, Institute of Economic Studies, Charles University, Czech Republic

(Received March 22, 1994; revised June 26, 1995)

Abstract. Bahadur representation of the difference of estimators of regression coefficients for the full data set and for the set from which one observation was deleted is given for the M-estimators which are generated by a continuous ψ-function. The representation is invariant with respect to the scale of residuals and it indicates that the bound of the norm of the difference is proportional to the gross error sensitivity. Then for the ψ-function which corresponds to the median it is shown that the difference of the estimates for the full data and for data without one observation, although being bounded in probability, can be much larger than indicated by the gross error sensitivity.

Key words and phrases: Sensitivity analysis, influential points in M-estimation, scale invariance.

1. Introduction

Diagnostic tools of the regression analysis have attracted, due to the evident reasons, a lot of attention, and presumably in the all recent monographies on the regression analysis they are more or less thoroughly discussed (e.g. Bates and Watts (1988), Rousseeuw and Leroy (1987) or Sen and Srivastava (1990)). Of course, there are also monographies treating only this topic (e.g. Atkinson (1985), Belsley *et al.* (1980), Chatterjee and Hadi (1988) or Cook and Weisberg (1982)).

One of the efficient and simple tools of LS-regression diagnostic has been a formula expressing the difference of estimators of regression coefficients for the full set of data and for the set of data obtained when excluding one observation from the original data. The formula may be written as

$$\hat{\beta}_{\mathrm{LS}}^{(n-1,\ell)} - \hat{\beta}_{\mathrm{LS}}^{(n)} = -\{[X^{(n-1,\ell)}]^{\mathrm{T}} X^{(n-1,\ell)}\}^{-1} X_\ell (Y_\ell - X_\ell^{\mathrm{T}} \hat{\beta}_{\mathrm{LS}}^{(n)}) \tag{1.1}$$

* Mailing address: Pod vodárenskou věží 4, 182 08 Prague, Czech Republic.

(where notation is nearly selfexplaining, nevertheless, $X^{(n-1,\ell)}$ is the design matrix after deletion of the ℓ-th row and X_ℓ is the ℓ-th row (assumed as a column vector) of the design matrix for the full data). For the M-estimators we cannot generally expect that we shall succeed to derive an exact formula for such a difference, not being even able to give an explicit formula for M-estimators themselves. But we may expect to derive an asymptotic representation for it (for the linear models and an absolutely continuous ψ-functions see Víšek (1992*a*, 1992*b*)). This paper gives such representation for the nonlinear model for the M-estimators generated by continuous ψ-functions. It is proved that the upper bound of norm of the difference of estimates is proportional to the gross error sensitivity. The difference of estimators of regression coefficients for the full set of data and for the set of data from which one observation was deleted is also studied for the ψ-function corresponding to median, i.e. for $\psi_m(z) = \text{sign}(z)$. It is shown that in this case, the difference of the estimates, even if bounded in probability, may be much larger than it is indicated by the gross error sensitivity. The approach considers (for continuous ψ-functions) the studentized residuals which reflects the fact that in many statistical packages, the evaluation of M-estimators is based also on studentized residuals. The reason for the studentization of residuals is of course the fact that it allows to use standardized ψ-functions. Moreover, in the special case when the regression model is linear, we obtain as a "premium" the invariance of the estimator; for more complete discussion see Rubio and Víšek (1996). Since we want for the estimator generated by ψ_m to show only its "subsample instability", we shall keep the text as simple as possible and hence we shall not assume studentization of residuals (moreover, since $\psi_m(z) = \text{sign}(z)$, and because the studentization does not change the sign, the value of ψ_m is invariant with respect to scale).

At the end of the paper, we shall offer a numerical study which brings a possibility to create an idea how much may be L_1-estimator, as an example of M-estimators with the discontinuous ψ-function, "subsample instable" in comparison with the other M-estimators.

In the LS-regression analysis the formula (1.1) might have already been used by Sir Francis Galton, when looking for the (most) influential point. In an iterative way, it was utilized even for finding the subsets of the influential observations (although it is not generally the same as searching for the influential subset as a block). The same is possible for our case. We shall return to this problem in the concluding remark when we will be able, with the help of the presented results, to explain the problem better.

Since the estimator $\hat{\beta}_{\text{LS}}^{(n-1,\ell)}$ for $n-1$ observations is included in the formula (1.1), it may seem that the considerations which will follow may be related to those in the jackknifing. But the resemblance is just formal, on the level of the formulas. The diagnostic studies are aimed to construct the tools which will help to reveal the influential point(s) (or influential subsets) among the data, while the goal of the jackknifing is to decrease (the order of) the bias (and possibly to find, as a byproduct, a strongly consistent estimator of the variance of the estimators in question).

2. Notation and setup

Let N denote the set of all positive integers, R^ℓ the ℓ-dimensional Euclidean space (if $\ell = 1$ we shall not write it), R^+ the nonnegative part of the real line and $(\Omega, \mathcal{A}, P)$ a probability space. We shall consider for all $i \in N$ the model

$$Y_i = g(X_i, \beta^0) + e_i \tag{2.1}$$

where for some fix $p, q \in N$, $\{X_i\}_{i=1}^\infty$ is a sequence of vectors from R^q and $\beta^0 = (\beta_1^0, \beta_2^0, \ldots, \beta_p^0)^{\mathrm{T}}$ is a vector of regression parameters ("T" stays for the transposition). Moreover, a function $g : R^{q+p} \to R$ is assumed to be two times differentiable (see Conditions A below) and finally, $\{e_i\}_{i=1}^\infty$, $e_i : \Omega \to R$ is a sequence of independent identically distributed random variables (i.i.d.r.v.). Let us denote $F(z)$ the distribution functions of e_1. Our study will consider the M-estimators of β^0 given as

$$\hat\beta^{(n)} = \operatorname*{argmin}_{\beta \in R^p} \left\{ \sum_{i=1}^n \rho([Y_i - g(X_i, \beta)]\hat\sigma_n^{-1}) \right\} \tag{2.2}$$

and

$$\hat\beta^{(n-1,\ell)} = \operatorname*{argmin}_{\beta \in R^p} \left\{ \sum_{\substack{i=1 \\ i \neq \ell}}^n \rho([Y_i - g(X_i, \beta)]\hat\sigma_n^{-1}) \right\} \tag{2.3}$$

where $\rho : R \to R$ is assumed to be absolutely continuous (denote the derivative—at the points where it exists—by ψ) and $\hat\sigma_n$ is a preliminary estimator of the scale (see Conditions C below).

3. Conditions

We are going to give the conditions we shall need for the preliminary considerations and later in the paper.

CONDITIONS A.

i) There is a positive δ_0 such that for any $\beta \in R^p$, $\|\beta - \beta^0\| < \delta_0$

$$\frac{\partial}{\partial \beta_j} g(x, \beta) \quad (j = 1, 2, \ldots, p) \quad \text{and} \quad \frac{\partial^2}{\partial \beta_j \partial \beta_k} g(x, \beta) \quad (j, k = 1, 2, \ldots, p)$$

exist for any $x \in R^q$. Let us denote the vector of the first partial derivative and the matrix of the second partial derivative simply by $g'(x, \beta)$ and $g''(x, \beta)$, respectively, and their coordinates and elements by $g'_j(x, \beta)$ and $g''_{jk}(x, \beta)$.

ii) The functions $g''_{j_k}(x, \beta)$ $(j, k = 1, 2, \ldots, p)$ are uniformly in $x \in R^q$ Lipschitz of the first order in β in the δ_0-neighborhood of β^0, i.e.

$$\exists (L > 0) \quad \forall (\beta \in R^p, \|\beta - \beta^0\| < \delta_0)$$

$$\max_{1\le j,k\le p} \sup_{x\in R^q} |g''_{jk}(x,\beta) - g''_{jk}(x,\beta^0)| < L \cdot \|\beta - \beta^0\|.$$

Moreover, let

$$\max_{1\le j,k\le p} \sup_{x\in R^q} \max\{|g(x,\beta^0)|, |g'_j(x,\beta^0)|, |g''_{jk}(x,\beta^0)|\} < \infty.$$

iii) There is a regular matrix $Q = \lim \frac{1}{n}\sum_{i=1}^n \{g'(X_i,\beta^0)[g'(X_i,\beta^0)]^{\mathrm{T}}\}$ (denote $(Q)_{ij} = q_{ij}$).

Remark 1. Let us observe that Condition A (ii) implies that there is $J < \infty$ such that

$$\max_{1\le j,k\le p} \sup_{x\in R^q, \beta\in R^p, \|\beta-\beta^0\|<\delta_0} \max\{|g(x,\beta)|, |g'_j(x,\beta)|, |g''_{jk}(x,\beta)|\} < J.$$

Finally, observe that the matrix Q is positive definite.

CONDITIONS B.

i) The function ψ allows decomposition in the form

$$\psi = \psi_a + \psi_c \tag{3.1}$$

where ψ_a has a derivative ψ'_a which is Lipschitz of the first order and ψ_c is a continuous function with derivative ψ'_c being step-function. Let us denote by $D_c = \{r_{c1}, r_{c2}, \ldots, r_{ch_c}\}$, ($h_c$ finite) the points of jumps of ψ'_c.

ii) $\sigma^2 = \mathrm{var}_F\, e_i \in (0,\infty)$ and there is a positive ϑ_0 such that $F(z)$ has a density f which is bounded on $D_c(\vartheta_0) = \bigcup_{i=1}^{h_c}[\sigma\cdot r_{ci} - \vartheta_0, \sigma\cdot r_{ci} + \vartheta_0]$. Let $H < \infty$ be corresponding upper bound of f on $D_c(\vartheta_0)$.

iii) There is a finite K such that $\sup_{z\in D_c(\vartheta_0)} |\psi(z)| < K$ as well as $\sup_{z\in R\setminus D_c} |\psi'(z)| < K$.

iv) $\mathsf{E}_F\psi(\frac{e_1}{\sigma}) = 0$ and $\gamma = \sigma^{-1}\mathsf{E}_F\psi'(\frac{e_1}{\sigma}) > 0$.

Remark 2. Conditions B essentially coincide with those of Hampel *et al.* (1986), Section 2.5a (of course, restricted on the continuous ψ-functions), however, the form of these (especially decomposition (3.1)) follows Jurečková (1988). Some heuristic comments on them may be found also in both references.

4. Notation (continued)

In accordance with the given conditions, let us enlarge the notation which was introduced above. Since β^0 will be fix throughout the paper, we shall write instead of $g(X_i, \beta^0 + n^{-1/2}t + n^{-1/2-\tau}u)$ simply $g(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$. When u will be equal to zero we shall write only $g(X_i, n^{-1/2}t)$, however, when also t

will be zero, we shall write $g(X_i, \beta^0)$ instead of $g(X_i, n^{-1/2}0)$. Similarly, we shall abbreviate

$$
\begin{aligned}
&g'_j(X_i, n^{-1/2}t, n^{-1/2-\tau}u) = g'_j(X_i, \beta^0 + n^{-1/2}t + n^{-1/2-\tau}u),\\
&g''_{jk}(X_i, n^{-1/2}t, n^{-1/2-\tau}u) = g''_{jk}(X_i, \beta^0 + n^{-1/2}t + n^{-1/2-\tau}u),\\
&g'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)\\
&\quad = [g'_1(X_i, n^{-1/2}t + n^{-1/2-\tau}u), g'_2(X_i, n^{-1/2}t + n^{-1/2-\tau}u), \ldots,\\
&\qquad\qquad g'_p(X_i, n^{-1/2}t + n^{-1/2-\tau}u)]^{\mathrm{T}},\\
&g''_j(X_i, n^{-1/2}t + n^{-1/2-\tau}u)\\
&\quad = [g''_{j1}(X_i, n^{-1/2}t + n^{-1/2-\tau}u), g''_{j2}(X_i, n^{-1/2}t + n^{-1/2-\tau}u), \ldots,\\
&\qquad\qquad g''_{jn}(X_i, n^{-1/2}t + n^{-1/2-\tau}u)]^{\mathrm{T}},\\
&\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) = g(X_i, \beta^0 + n^{-1/2}t + n^{-1/2-\tau}u) - g(X_i, \beta^0),\\
&s(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})\\
&\quad = \psi([e_i - \delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)]\sigma^{-1}e^{-n^{-1/2}v})\\
&\qquad \times g'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)\\
&\qquad - \psi([e_i - \delta(X_i, n^{-1/2}t)]\sigma^{-1}e^{-n^{-1/2}v})g'(X_i, n^{-1/2}t),
\end{aligned}
$$

and

$$
S(n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v}) = \sum_{i=1}^{n} s(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v}). \tag{4.1}
$$

Similarly, as for the derivatives of the function g, we shall denote by $s_j(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})$ and by $S_j(n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})$ the j-th coordinates of the vectors $s(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})$ and $S(n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})$, respectively. Finally, for any $M > 0$ let us put

$$
\mathcal{T}_M = \{t, u \in R^p, v \in R^+ : \max\{\|t\|, \|u\|, v\} \le M\}. \tag{4.2}
$$

The range of indices or variables (used in just introduced notations) will be clear from the context or it will be indicated at the place where they will be used.

5. Preliminaries

We shall prepare now two lemmas for deriving the Bahadur representation of

$$
n(\hat{\beta}^{(n-1,\ell)} - \hat{\beta}^{(n)}).
$$

LEMMA 5.1. *Let Conditions* A *be fulfilled and let ψ-function have a derivative ψ' which is Lipschitz of the first order, i.e. $\psi = \psi_a$. Moreover, let* $\operatorname{var}_F e_1 = \sigma^2 \in (0, \infty)$, $\mathsf{E}_F\psi(\frac{e_1}{\sigma}) = 0$ *and* $|\mathsf{E}_F\psi'(\frac{e_1}{\sigma})| < \infty$. *Then for any fix $\tau \in [0, \frac{1}{2}]$ there*

are sequences of random matrices $\{\mathcal{U}_n(\tau)\}_{n=1}^{\infty}$ *such that* $\max_{1\le i,j\le p}|(\mathcal{U}_n(\tau))_{ij}| = o(1)$ *a.s. as* $n\to\infty$ *and we have*

$$\sup_{\mathcal{T}_M}\Big\|S(n^{-1/2}t+n^{-1/2-\tau}u,\sigma e^{n^{-1/2}v}) + n^{1/2-\tau}\left[\sigma^{-1}\mathsf{E}_F\psi'\left(\frac{e_1}{\sigma}\right)Q+\mathcal{U}_n\right]u\Big\| = O(n^{-\tau}) \quad a.s. \quad as \quad n\to\infty. \tag{5.1}$$

The proofs of the all lemmas are given in the Appendix because they are mostly of technical character and would only burden the reading of the paper.

Remark 3. Lemma 5.1 was given in a general way although it will be later used only for the specified values of τ, namely 0 and $\frac{1}{2}$. Nevertheless, its present form allows to give the proof in a transparent and simple way simultaneously for both these values (see the Appendix).

LEMMA 5.2. *Let Conditions* A *hold and let the function* ψ *have a derivative* ψ' *such that for* $-\infty = r_0 < r_1 < \cdots < r_h < \infty$ *and real numbers* $\alpha_0,\alpha_1,\ldots,\alpha_{h-1}$, $\psi'(x)=\alpha_k$ *for* $x\in(r_k,r_{k+1}]$ *for* $k=0,1,\ldots,h-1$ *and* $\psi'(x)=\alpha_h$ *for* $x\in(r_h,\infty)$, *i.e.* $\psi=\psi_c$. *Moreover, let* $\operatorname{var}_F e_1=\sigma^2\in(0,\infty)$ *and let in a* ϑ_0*-neighborhood of the points* $\sigma r_1,\sigma r_2,\ldots,\sigma r_h$ *the distribution function* F *have a bounded density* f. *Finally, let* $\mathsf{E}_F\psi(\frac{e_1}{\sigma})=0$. *Then for any fix* $\tau\in[0,\frac{1}{2}]$ *there are sequences of random matrices* $\{\mathcal{U}_n(\tau)\}_{n=1}^{\infty}$ *such that* $\max_{1\le i,j\le p}|(\mathcal{U}_n(\tau))_{ij}| = o(1)$ *a.s. as* $n\to\infty$ *and we have*

$$\sup_{\mathcal{T}_M}\Big\|S(n^{-1/2}t,n^{-1/2-\tau}u,\sigma e^{n^{-1/2}v}) + n^{1/2-\tau}\left[\sigma^{-1}\mathsf{E}_F\psi'\left(\frac{e_1}{\sigma}\right)Q+\mathcal{U}_n(\tau)\right]u\Big\| = O_p(n^{-\tau}) \quad as \quad n\to\infty. \tag{5.2}$$

6. Bahadur's representation for continuous ψ-functions

In this section, we will give Bahadur's representation of $n(\hat\beta^{(n)}-\hat\beta^{(n-1,\ell)})$. The plan on how to do that is simple. At first, using Lemmas 5.1 and 5.2 for $\tau=0$ we shall prove that $n(\hat\beta^{(n)}-\hat\beta^{(n-1,\ell)})=O_p(1)$, and then using the same lemmas for $\tau=\frac{1}{2}$ we derive the representation.

We are now going to specify the conditions on $\hat\beta^{(n)}$, $\hat\beta^{(n-1,\ell)}$ and $\hat\sigma_n$.

CONDITIONS C.

i) The estimators $\hat\beta^{(n)}$ and $\hat\beta^{(n-1,\ell)}$ given by (2.2) and (2.3) are $\sqrt{n}$-consistent in the following sense

$$\forall(\varepsilon>0)\quad \exists(K>0 \text{ and } n_0\in N)\quad \forall(n\in N, n\ge n_0 \text{ and } \ell=1,2,\ldots,n)$$

$$P(\sqrt{n}\|\hat{\beta}^{(n)} - \beta^0\| > K) < \varepsilon \quad \text{and} \quad P(\sqrt{n}\|\hat{\beta}^{(n-1,\ell)} - \beta^0\| > K) < \varepsilon.$$

ii) There is a location invariant and scale equivariant, $\sqrt{n}$-consistent estimator $\hat{\sigma}_n$ of σ, i.e.

$$\sqrt{n}(\hat{\sigma}_n - \sigma) = O_p(1) \quad \text{as} \quad n \to \infty.$$

Remark 4. It is clear that under Conditions A and B, in the case when $\psi_s \equiv 0$, the estimators $\hat{\beta}^{(n)}$ and $\hat{\beta}^{(n-1,\ell)}$ fulfill the following relations (see (2.2) and (2.3)):

$$\sum_{i=1}^{n} \psi([Y_i - g(X_i, \hat{\beta}^{(n)})]\hat{\sigma}_n^{-1}) g'(X_i, \hat{\beta}^{(n)}) = 0 \tag{6.1}$$

and

$$\sum_{\substack{i=1 \\ i \neq \ell}}^{n} \psi([Y_i - g(X_i, \hat{\beta}^{(n-1,\ell)})]\hat{\sigma}_n^{-1}) g'(X_i, \hat{\beta}^{(n-1,\ell)}) = 0, \tag{6.2}$$

respectively (where $\hat{\sigma}_n$ is again a preliminary estimator of scale of residuals). Sometimes the M-estimators (for the linear model) are even defined as solutions of the equations (6.1) and (6.2).

Remark 5. In Rubio and Víšek (1996) it is shown how the result of Liese and Vajda (1995), concerning the consistency of the estimator $\hat{\beta}^{(n)}$ in a nonstudentized framework, can be generalized for the studentized version and then strengthened to $\sqrt{n}$-consistency. Further, in Rubio *et al.* (1994) it is shown that under conditions given here, the $\sqrt{n}$-consistency of $\hat{\beta}^{(n)}$ follows from its consistency. Moreover, also in Rubio and Víšek (1996) it is proved that under conditions given here, there is for the case $\psi = \psi_a + \psi_c$ a $\sqrt{n}$-consistent solution of the equation (6.1) (the result follows from the fix-point theorem and the idea is due to Jana Jurečková). Also the result of Rao and Zhao (1992) seems to be in a straightforward way genaralizable for nonlinear setup (this results applies also for ψ-functions with jumps but on the other hand ψ-function has to be monotone and of course we cannot reach in (6.1) and (6.2) precise equality, see discussion in the Section 8). So, it seems that there may appear very diverse conditions for consistency of the M-estimators for general ψ-functions, and hence we have preferred to give Conditions C in the present form.

LEMMA 6.1. *Let Conditions* A, B *and* C *hold. Then*

$$n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,\ell)}) = O_p(1) \quad \textit{as} \quad n \to \infty.$$

THEOREM 6.1. *Let Conditions* A, B *and* C *hold. Then uniformly in* $\ell \in N$ *we have*:

$$\begin{aligned} n(\hat{\beta}^{(n-1,\ell)} - \hat{\beta}^{(n)}) = & -\hat{\sigma}_n \mathbf{E}_F^{-1} \psi'\left(\frac{e_1}{\sigma}\right) Q^{-1} g'(X_\ell, \hat{\beta}^{(n)}) \\ & \cdot \psi([Y_\ell - g(X_\ell, \hat{\beta}^{(n)})]\hat{\sigma}_n^{-1}) + o_p(1) \quad \textit{as} \quad n \to \infty. \end{aligned} \tag{6.3}$$

PROOF. Considering $\tau = \frac{1}{2}$ and taking into account Lemma 6.1 we may substitute $\hat{t}_{n-1} = \sqrt{n-1}(\hat{\beta}^{(n)} - \beta^0)$, $\hat{u}_{n-1} = (n-1) \cdot (\hat{\beta}^{(n-1,\ell)} - \hat{\beta}^{(n)})$ and $\hat{v}_n = \sqrt{n}(\log \hat{\sigma}_n - \log \sigma)$ into (5.1) and (5.2) and we obtain

$$\sum_{i=1, i\neq\ell}^{n} [\psi([Y_i - g(X_i, \hat{\beta}^{(n-1,\ell)})]\sigma_n^{-1}) g'_{(in)}(X_i, \hat{\beta}^{(n-1,\ell)})$$
$$- \psi([Y_i - g(X_i, \hat{\beta}^{(n)})]\sigma_n^{-1}) g'_{(in)}(X_i, \hat{\beta}^{(n)})]$$
$$+ \sigma^{-1} \mathsf{E}_F \psi' \left(\frac{e_1}{\sigma}\right) Q(n-1)(\hat{\beta}^{(n-1,\ell)} - \hat{\beta}^{(n)}) = o_p(1) \quad \text{as} \quad n \to \infty.$$

Then utilizing Lemma 6.1 once again, and employing (6.1) and (6.2) we have

$$\psi([Y_\ell - g(X_\ell, \hat{\beta}^{(n)})]\sigma_n^{-1}) g'(X_\ell, \hat{\beta}^{(n)})$$
$$+ \sigma^{-1} \mathsf{E}_F \psi' \left(\frac{e_1}{\sigma}\right) Q \cdot n(\hat{\beta}^{(n-1,\ell)} - \hat{\beta}^{(n)}) = o_p(1) \quad \text{as} \quad n \to \infty.$$

Finally, taking into account the regularity of the matrix Q, we conclude the proof of the theorem. □

Remark 6. The uniformity in ℓ which has been stated in Theorem 6.1 has to be interpreted (as follows from the proof of the theorem) in the following way:

$$\forall(\varepsilon > 0 \text{ and } \delta > 0) \quad \exists(n_0 \in N) \quad \forall(n \in N, n \geq N_0 \text{ and } \ell = 1, 2, \ldots, n)$$

$$P\Big(\Big\| n(\hat{\beta}^{(n-1,\ell)} - \hat{\beta}^{(n)})$$
$$+ \hat{\sigma}_n \mathsf{E}_F^{-1} \psi' \left(\frac{e_1}{\sigma}\right) Q^{-1} g'(X_\ell, \hat{\beta}^{(n)}) \psi([Y_\ell - g(X_\ell, \hat{\beta}^{(n)})]\hat{\sigma}_n^{-1}) \Big\| > \delta\Big) < \varepsilon,$$

i.e. n_0 is the same for all $\ell = 1, 2, \ldots, n$. It does not mean necessarily that

$$P\Big(\max_{1\leq\ell\leq n} \Big\| n(\hat{\beta}^{(n-1,\ell)} - \hat{\beta}^{(n)})$$
$$+ \hat{\sigma}_n \mathsf{E}_F^{-1} \psi' \left(\frac{e_1}{\sigma}\right) Q^{-1} g'(X_\ell, \hat{\beta}^{(n)}) \psi([(Y_\ell - g(X_\ell, \hat{\beta}^{(n)})]\hat{\sigma}_n^{-1}) \Big\| > \delta\Big) < \varepsilon.$$

Remark 7. From (6.3) it is clear that the upper bound of $n\|\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,\ell)}\|$ is proportional to the gross error sensitivity. In other words, deleting one point from the data may cause only some "reasonable" change of the estimate of regression coefficients. We may interpret this situation by an assertion that the estimator is stable on subsamples.

7. Preliminaries (continued)

Now, we are going to study the behavior of the difference of estimates for ψ_m (let us recall that $\psi_m = \text{sign}(z)$). The goal of this part of paper (as well as of the numerical examples) is to demonstrate that the difference of the estimates for the full data and for data from which we have deleted one observation, can be much higher for the discontinuous ψ-function than for the continuous one. As it was pointed out by the referee, it is sufficient to show it for ψ_m because for any other discontinuous ψ-function, we may expect even a worse behavior of the difference of estimates.

Under L_1-estimator we shall understand

$$\hat{\beta}^{(L_1,n)} = \underset{\beta \in R^p}{\text{argmin}} \left\{ \sum_{i=1}^{n} \rho_m(Y_i - X_i^T \beta) \right\} \tag{7.1}$$

and

$$\hat{\beta}^{(L_1,n-1,\ell)} = \underset{\beta \in R^p}{\text{argmin}} \left\{ \sum_{i=1}^{n} \rho_m(Y_i - X_i^T \beta) \right\} \tag{7.2}$$

where $\rho_m(z) = |z|$. It is clear that the studentization does not play any role here because dividing all residuals by a number, it is the same as to divide the whole sum in (7.1) and (7.2) by that number. That is why (7.1) and (7.2) do not include an estimate of scale.

LEMMA 7.1. *Let* $\sup_{1 \le i \le \infty} \|X_i\| < \infty$ *and let the density* $f(z)$ *exist and is Lipschitz of the first order in a neighborhood of zero. Moreover, let there is a regular matrix* $Q = \lim_{n \leftarrow \infty} \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T$. *Denote again* $(Q)_{ij} = q_{ij}$. *Further denote* $W(s)$, $s \in R^+$ *a Wiener process defined on a space* $(\Omega^*, \mathcal{A}^*)$. *Then for any* $(k = 1, 2, \ldots, p)$ *there are stopping times* $\mu_{ik}(n,t,u)$ *and* $\kappa_{ik}(n,C)$ *such that for any* $t, u \in \mathcal{T}_M$, $\mu_{ik}(n,t,u) < \kappa_{ik}(n,C)$, $\sum_{i=1}^{n} \kappa_{ik}(n,C)$ *is bounded in probability and*

$$\begin{aligned} &\sum_{i=1}^{n} X_{ik} \{ \psi_m(e_i - n^{-1/2} X_i^T t - n^{-1} X_i^T u) - \psi_m(e_i - n^{-1/2} X_i^T t) \} \\ &\qquad + 2f(0) \sum_{j=1}^{p} q_{jk} u_j =_{\mathcal{D}} W\left(\sum_{i=1}^{n} \mu_{ik}(n,t,u) \right). \end{aligned} \tag{7.3}$$

where "$=_{\mathcal{D}}$*" denotes the equality in distributions.*

8. Subsample behavior of L_1-estimator

Remark 8. As we have already mentioned in Remark 6 for the continuous ψ-functions the equations (6.1) and (6.2) hold. Generally, they do not hold for non-smooth ρ-functions, derivative of which is discontinuous. Nevertheless, to be able to apply Lemma 7.1 in a similar way as we have used Lemmas 5.1 and 5.2,

we need to have (6.1) and (6.2) fulfilled at least approximately, say that we would like to have

$$\sum_{i=1}^{n} \psi\left(\frac{Y_i - g(X_i, \hat{\beta}^{(\psi,n)})}{\hat{\sigma}^{(n)}}\right) g'(X_i, \hat{\beta}^{(\psi,n)}) = o_p(1). \tag{8.1}$$

We shall try to make an idea how far we may expect it for M-estimators with the discontinuous ψ-function.

General conditions under which (8.1) is fulfilled for the discontinuous ψ-function are not known, although for some of them, e.g. for ψ_{med}, given by

$$\psi_{\text{med}} = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0 \end{cases}$$

we may reach again even precise equality in (8.1)—under some conditions for symmetry of $g'(X_i, \beta)$ without which it seems questionable to use ψ_{med}.

To create an idea about the problem let us look at first on the much simpler case of estimating location parameter in the case when the central model is assumed to be the standard normal one. After all, in other cases, under assumptions which was used in Huber's paper (Huber (1964)), namely that $-\log \frac{f'(x)}{f(x)}$ is strictly convex, we may, for theoretical considerations, assume that we transform random variables to the normal ones. Let us assume that we shall use skipped Huber's ψ-function $\psi_H(x)$, i.e. $\psi_H(x) = -\psi_H(-x)$ and

$$\psi_H(x) = \begin{cases} x & \text{if } x \in [0, a], \\ a & \text{if } x \in (a, b], \\ 0 & \text{if } x > b \end{cases}$$

for some $0 < a < b < \infty$. Let $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$ be our observation (in fact we may assume $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$ because if any sharp inequality is distorted the (absolute) continuity, is questionable; from the similar reasons we have also $Y_{(i)} - Y_{(j)} \neq 2b$ for $i, j = 1, 2, \ldots, n$ a.e. for any $n \in N$). Now, let us observe that for $t \in (-\infty, Y_{(1)} - b) \cup (Y_{(n)} + b, \infty)$ we have $\sum_{i=1}^{n} \psi_H(Y_{(i)} - t) = 0$.

Since for any $n \in N$ and any $\omega \in \Omega$ we may find $t \in R$ so that $t \in (-\infty, Y_{(1)}(\omega)) \cup (Y_{(n)}(\omega), \infty)$, it is clear that we may obtain inconsistent solution of (8.2) (as well as of (6.1)). In other words, for strongly redescending ψ-function (regardless whether continuous or discontinuous) among the solutions of (8.2) is at least one inconsistent. Nevertheless, for $t = Y_{(1)} - b$ we obtain $\sum_{i=1}^{n} \psi_H(Y_{(i)} - t) = a$ and for $t = Y_{(n)} + b$ we finally get $\sum_{i=1}^{n} \psi_H(Y_{(i)} - t) = -a$. Moreover, $\sum_{i=1}^{n} \psi_H(Y_{(i)} - t)$ is continuous (and nonincreasing) in t except for a finite number of discontinuities, at which it has the positive jumps equal to a. It implies that there is at least one point $\hat{t}^{(n)} \in (Y_{(1)} - b, Y_{(n)} + b)$ such that

$$\sum_{i=1}^{n} \psi_H(Y_{(i)} - \hat{t}^{(n)}) = 0. \tag{8.2}$$

We may observe that the reason why for ψ_H we are able to fulfil (8.2) is a "compensation" of the jump(s) by a decrease of the value of the terms which have argument in the linear part of the ψ-function.

It is easy to see that the point(s) which solves (8.2) is a (local) minimum of the function $\sum_{i=1}^n \rho(Y_{(i)} - t)$ because $-\psi_H(y - t)$ is increasing in t. Moreover, at any point t^* of jump of $\frac{\partial}{\partial t}\sum_{i=1}^n \rho(Y_{(i)} - t)$ we have

$$\lim_{t \to t^*_-} \frac{\partial}{\partial t} \sum_{i=1}^n \rho(Y_{(i)} - t) > \lim_{t \to t^*_+} \frac{\partial}{\partial t} \sum_{i=1}^n \rho(Y_{(i)} - t)$$

so that the function $\sum_{i=1}^n \rho(Y_{(i)} - t)$ either increases when $t \to t^*_-$, and then for $t > t^*$ again decreases or increases less steeply, or decreases for $t \to t^*_-$, and then for $t > t^*$ it decreases more steeply. Anyway, the function $\sum_{i=1}^n \rho(Y_{(i)} - t)$ cannot have at t^* minimum. So we may conclude that the global minimum is among the points for which (8.2) holds.

More detailed analysis would reveal that a similar situation holds for many ψ-functions, namely that we may hope to fulfill

$$\sum_{i=1}^n \psi(Y_i - t) = o_p(1)$$

for rather large family of ψ-functions.

Some difficulties may appear e.g. for skipped median (or for some other estimators with both types of jumps).

Let us now consider the linear regression. We would want again to show that there is a point $\hat{\beta}$ such that

$$\sum_{i=1}^n \psi_H(Y_{(i)} - X_i^T \hat{\beta}) X_i = 0.$$

Let us consider at first $\sum_{i=1}^n \rho_H(Y_i - L \cdot X_i^T \gamma)$ for $\|\gamma\| = 1$. We easy verify that for any γ

$$-\frac{\partial}{\partial L} \sum_{i=1}^n \rho_H(Y_i - L \cdot X_i^T \gamma) = \sum_{i=1}^n \psi_H(Y_i - L \cdot X_i^T \gamma) X_i^T \gamma$$

is nonincreasing in L (except for finite number ℓ ($\ell \leq 2n$) of positive jumps), and along similar lines as above we again find that there is $L_\gamma^{(1)} < 0$ such that for $L < L_\gamma^{(1)}$ we have $\sum_{i=1}^n \psi_H(Y_i - L \cdot X_i^T \gamma) X_i^T \gamma = 0$ and

$$\sum_{i=1}^n \psi_H(Y_i - L_\gamma^{(1)} \cdot X_i^T \gamma) X_i^T \gamma = \sum_{i \in \mathcal{I}_\gamma^{(1)}} |X_i^T \gamma| \cdot a$$

where $\mathcal{I}_\gamma^{(1)} = \{i \in N : \text{sign}(X_i^T\gamma)\psi_H(Y_i - L_\gamma^{(1)} \cdot X_i^T\gamma) = a\}$. Similarly, we may find an upper "bound" $L_\gamma^{(2)}$. Then there is again at least one $L_\gamma^* \in (L_\gamma^{(1)}, L_\gamma^{(2)})$ such that

$$\sum_{i=1}^{n} \psi_H(Y_i - L_\gamma^* \cdot X_i^T\gamma)X_i^T\gamma = 0. \tag{8.3}$$

Due to similar arguments as above we find that at one of these points (if they are multiple) the function $\sum_{i=1}^n \rho_H(Y_i - L \cdot X_i^T\gamma)$, as the function of L, attains its minimum, and that the points $Y_i - L_\gamma^* \cdot X_i^T\gamma$, $i = 1, 2, \ldots, n$ are not points of discontinuity of the function ψ_H. Let $\rho_0 = \inf_{\|\gamma\|=1} \sum_{i=1}^n \rho_H(Y_i - L_\gamma^* \cdot X_i^T\gamma)$. Taking into account the compactness of the surface of unit ball we find that there is a γ_0, $\|\gamma_0\| = 1$ such that

$$\rho_0 = \sum_{i=1}^{n} \rho_H(Y_i - L_{\gamma_0}^* \cdot X_i^T\gamma_0).$$

Let us recall that the points $Y_i - L_{\gamma_0}^* \cdot X_i^T\gamma_0$, $i = 1, 2, \ldots, n$ are not the points of discontinuity of the function ψ_H, i.e. in the neighborhood of the point $\hat{\beta} = L_{\gamma_0}^*\gamma_0$ the function $\sum_{i=1}^n \rho_H(Y_i - X_i^T\beta)$ has (continuous) partial derivatives, and hence

$$\sum_{i=1}^{n} \psi_H(Y_i - X_i^T\hat{\beta})X_i^T = 0.$$

It is clear that to derive Bahadur representation, we do not need necessarily that (6.1) and (6.2) hold but what we really need is

$$\begin{aligned} &\sum_{i=1}^{n} \psi([Y_i - g(X_i, \hat{\beta}^{(n)})]\hat{\sigma}_n^{-1})g'(X_i, \hat{\beta}^{(n)}) \\ &\quad - \sum_{\substack{i=1 \\ i\neq\ell}}^{n} \psi([Y_i - g(X_i, \hat{\beta}^{(n-1,\ell)})]\hat{\sigma}_n^{-1})g'(X_i, \hat{\beta}^{(n-1,\ell)}) = o_p(1). \end{aligned} \tag{8.4}$$

Numerical experiences say that even (8.4) may be sometimes too optimistic expectation, nevertheless, we have typically (for linear model) the left hand side of (8.4) bounded (coordinatewise) by $\max_{1\leq\ell\leq n} |X_{i\ell}|$. In fact, usually the situation is as follows. Either for "even" number of observations we have difference in (8.4) nearly equal to zero and for "odd" sample size the absolute value of the difference is under the bound $\max_{1\leq\ell\leq n} |X_{i\ell}|$, or vice versa for "odd" sample size it is nearly zero etc. (When we spoke about "even" and "odd" sample size we have meant that we successively delete points from the data.) Nevertheless, for the next theoretical considerations, we shall assume that (8.4) holds.

LEMMA 8.1. *Let* $\sup_{1\leq i\leq\infty} \|X_i\| < \infty$ *and the density* $f(z)$ *exist and is Lipschitz of the first order in a neighborhood of zero. Moreover, let* (8.4) *hold and let*

there is a regular matrix $Q = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i X_i^T$. *Denote again* $(Q)_{ij} = q_{ij}$. *Finally, let for some* $\ell \in \{1, 2, \ldots, n\}$

$$n(\hat{\beta}^{(L_1,n)} - \hat{\beta}^{(L_1,n-1,\ell)}) = O_p(1).$$

Then

$$n(\hat{\beta}^{(L_1,n)} - \hat{\beta}^{(L_1,n-1,\ell)}) = \frac{1}{2} f^{-1}(0) Q^{-1} X_\ell \psi_m(Y_\ell - X_\ell^T, \hat{\beta}^{(L_1,n)}) + \mathcal{R}_n$$

where $\mathcal{R} =_{\mathcal{D}} \frac{1}{2} f^{-1}(0) Q^{-1}[W_n^{(1)} - W_n^{(2)}] + o_p(1)$ *with* $W_n^{(j)} = (W(\sum_{i=1}^{n} \mu_{i1}^{(j)}(n,t,u)), W(\sum_{i=1}^{n} \mu_{i2}^{(j)}(n,t,u)), \ldots, W(\sum_{i=1}^{n} \mu_{ip}^{(j)}(n,t,u)))^T$ *for* $j = 1, 2$ *for some stopping times* $\mu_{ik}^{(j)}(n,t,u)$, $i = 1, 2, \ldots, n$, $k = 1, 2, \ldots, p$, $n \in N$, $t, u \in \mathcal{T}_M$ *and where again* $W(s)$ *is a Wiener process defined on a space* $(\Omega^*, \mathcal{A}^*)$.

PROOF. Taking into account (7.3) and (8.4) one may perform the proof in a nearly the same way as the proof of Theorem 6.1. □

Remark 9. We know already from Lemma 6.1 that for the continuous ψ-function the normed difference of the estimators $n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,\ell)})$ is $O_p(1)$. Theorem 6.1 specifies this information, so that we may give for the function ψ which is bounded, an upper bound for this difference and this upper bound is valid except for a set of small probability. In other words, we may make an idea about stability of the estimation when adding or excluding one observation. (And the numerical experiences say that the approximation works for rather small number of observations, usually about twenty, see Víšek (1992*b*).) On the other hand, for L_1-estimator even if we know that the difference is bounded in probability, the upper bound may be pretty large (and numerical experiences confirm much larger "fluctuation" of L_1-estimator in comparison with the estimators with smooth ψ-functions).

It implies that for the "continuous" case, if the "tuning" constant of the corresponding ψ-function is properly assigned to winsorize really some residuals, $\max_{1\le\ell\le n} \|n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,\ell)})\|$ is nearly deterministically given. For the "discontinuous" case it is not so.

9. Numerical examples

In this section, we shall offer three numerical examples which may illustrate behaviour of M-estimators for continuous and discontinuous ψ-functions. As the discontinuous ψ-function we shall use $\psi_m(z)$ because for it the theoretical result was derived. Earlier than we shall present the promised examples of real data we present one simple example of invented (not simulated) data which enlightens the reason which is behind possible "unstable" behaviour of $\hat{\beta}^{(L_1,n)}$. The data are given in Table 1.

Table 1. L_1-data.

	1	2	3	4	5	6	7	8	9	10	11	12
x	−1.00	−0.75	−0.50	−1.00	−0.65	−0.50	0.50	0.65	1.00	0.25	0.75	1.00
y	1.00	1.25	0.75	−1.00	−0.75	−0.65	0.50	0.75	1.00	−0.40	−1.00	−1.00

Table 2. Results of L_1 analysis of L_1-data.

	Intercept	Slope
Full data	0.0	1.0
Data without point 7	−0.2	−0.8

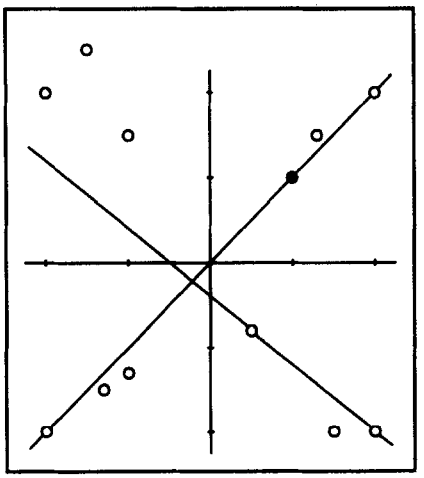

Fig. 1. L_1-data. $\hat{\beta}^{(L_1,n)}$ - the main diagonal, $\hat{\beta}^{(L_1,n-1,\ell)}$ - the other one (point 7 denoted as full circle was deleted in the latter case).

In the next examples, the results of L_1 analysis and the regression analysis performed by M-estimators with continuous ψ-function will be presented. It is worthwhile to say before it in a few words.

Firstly, which software was used. For the evaluation of L_1 results we have used software which is due to Jaromír Antoch from Charles University (and we are thankful to have this possibility). The software was checked on many numerical examples (for which correct values of estimator was known from literature) as well as with results obtained by other software products (e.g. Koenker's program for evaluating regression α-quantiles).

For evaluating other M-estimates the software by Alfio Marazzi was used (we

are also very glad for the possibility to use it). This software studentizes the residuals, so that the selection of the tuning constant does not represent a crucial problem.

Secondly, the examples given below should offer a possibility to the reader to create an idea about the changes of the estimates when deleting one (influential) observation from data. Nevertheless, we should remind that the large changes of the intercept can be caused by small changes of the estimates of slopes when data are sufficiently far from the origin. So, although the values of the estimates of intercept were included in the tables (to be completed), please ignore the changes of the intercept.

Example 1. Engine Knock Data (16 cases, firstly used in Mason *et al.* (1989)). The data record dependence of engine knock number on the spark timing, on the air/fuel ratio, on the intake temperature and on the exhaust temperature. The data were used in Hettmansperger and Sheather (1992) to demonstrate the instability of LMS. When the authors of the latter paper included the data in the computer, they wrote wrongly one digit (the value of regressor "Air" for the second observation was written 15.1 rather than 14.1 which is the correct one) and it caused in some sense a surprising behaviour of high breakdown point estimators, see Hettmansperger and Sheather (1992) and Víšek (1992*c*). In Table 3 the results are given for the "contaminated" data. Due to the fact that the Engine Knock Data contains only 16 cases one may object that it is too small number to reveal something about behaviour of an estimator (see e.g. Rousseeuw (1993)). So we have restricted ourselves on two regressors (air/fuel and intake) and intercept to reach the "thumb rule" of 5 cases per dimension, see again Rousseeuw (1993).

Table 3. Results of L_1 analysis of Engine Knock Data (point 3 was deleted).

	Intercept	Air/Fuel	Intake
Full data	31.84	2.471	0.594
Data without point 3	34.10	1.500	0.950

Before we continue further, let us note that in the case of Engine Knock Data deletion of the point 3 causes also the largest possible changes for Huber and Hampel estimator but, as Tables 4 and 5 (see the next page) demonstrate, they are not drastic. For the next two examples, deletion of the point which causes the largest change of L_1 (i.e. points 20 and 3, respectively) does not causes the largest change of Huber's and Hampel's estimator. But the largest change in both cases is not larger than about 5% in comparision with the largest change.

Example 2. Health Club Data (30 cases, Chatterjee and Hadi (1988)). The data describe dependence of time in a one-mile run on the weight, on the resting pulse rate per minute, on the arm and leg strength and on the time in a $\frac{1}{4}$ mile trial run.

Table 4. Results of regression analysis of Engine Knock Data (Huber ψ with tunning constant 1.2 was used, point 3 was deleted).

	Intercept	Air/Fuel	Intake
Full data	31.97	1.785	0.896
Data without point 3	32.71	1.639	0.937

Table 5. Results of regression analysis of Engine Knock Data (Hampel ψ with tunning constant 1.2 was used, point 3 was deleted).

	Intercept	Air/Fuel	Intake
Full data	27.58	2.096	0.885
Data without point 3	28.49	1.934	0.929

Table 6. Results of L_1 analysis of Health Club Data (point 20 was deleted).

	Intercept	Weight	Pulse	Strength	$\frac{1}{4}$ mile
Full data	−57.03	1.090	−0.928	−0.317	4.853
Data without point 20	8.69	0.806	−2.238	−0.365	5.958

Table 7. Results of regression analysis of Health Club Data (Huber ψ with tunning constant 1.2 was used, point 20 was deleted).

	Intercept	Weight	Pulse	Strength	$\frac{1}{4}$ mile
Full data	8.06	1.303	−0.777	−0.538	3.969
Data without point 20	12.18	1.273	−0.868	−0.531	4.048

Table 8. Results of regression analysis of Health Club Data (Hampel ψ with tunning constant 1.2 was used, point 20 was deleted).

	Intercept	Weight	Pulse	Strength	$\frac{1}{4}$ mile
Full data	12.49	1.316	−0.873	−0.553	4.004
Data without point 20	12.82	1.298	−0.849	−0.549	4.019

Table 9. Results of L_1 analysis of U.S. Crime Data (point 3 was deleted).

	Intercept	Age	Education	Police	Income
Full data	450.3	0.426	−0.018	−2.096	−0.795
Data without point 3	389.5	0.507	0.250	−1.818	−0.946

Table 10. Results of regression analysis of U.S. Crime Data (Huber ψ with tunning constant 1.2 was used, point 3 was deleted).

	Intercept	Age	Education	Police	Income
Full data	406.8	0.476	0.241	−2.073	−0.819
Data without point 3	404.6	0.472	0.248	−2.066	−0.811

Table 11. Results of regression analysis of U.S. Crime Data (Hampel ψ with tunning constant 1.2 was used, point 3 was deleted).

	Intercept	Age	Education	Police	Income
Full data	403.1	0.477	0.281	−2.120	−0.781
Data without point 3	399.9	0.471	0.292	−2.107	−0.773

Example 3. U.S. Crime Data (47 cases, Vandaele (1978) or Hand *et al.* (1994)). These data are for the crime in U.S.A., and they concern 47 states. The goal of the investigation was to find how the crime rate (number of offence known to the police per 10^6 population) in 1960 depended on age distribution, on the fact whether the offence was accomplished in southern state, on the educational level, on the police expenditure, on the labour force participation rate, on ratio of males in population, on the total number of population in the state, on the ratio of whites in population, on the unemployment rate, on the median family wealth and on the income inequality. The regressors were selected in the following way: The variables which appeared in the complete LS and in the complete LTS analysis as "highly" insignificant have been deleted (P-value over 0.2). Then LS and LTS analyses were repeated and the variables which were still significant were taken into account (of course, we do not want to claim that it is the only possibility how to choose). So that the variables used in the example are: Age distribution (the number of males aged 14–24 per 10^3 of total state population), Educational level (mean number of years of schooling of the population 25 years old and over), Police expenditure (the per capita expenditure on police protection by state and local government in 1960) and Income inequality (the number of families per 10^3 earnings below one half of the median income). Their P-values for LS model are 0.0^413 (where $0.0^\tau\xi = 10^{-\tau}\cdot 0.\xi$), 0.032714, 0.0^3773, 0.0^6 and 0.0^3137, respectively,

and for the LTS 0.0^6, 0.0^3353, 0.0^6, 0.0^6 and 0.0^6.

10. Concluding remarks

In the LS-regression analysis the formula (1.1) has been frequently used in the studentized form (in the situations when the data are "regressionally equivariant")

$$(10.1) \qquad (\hat\beta_{\mathrm{LS},j}^{(n-1,\ell)} - \hat\beta_{\mathrm{LS},j}^{(n)})[\mathrm{var}(\hat\beta_{\mathrm{LS},j}^{(n-1,\ell)} - \hat\beta_{\mathrm{LS},j}^{(n)})]^{-1/2} = (Y_\ell - X_\ell^{\mathrm{T}}\hat\beta_{\mathrm{LS}}^{(n)})\sigma^{-1}$$

for $j = 1, 2, \ldots, p$, or for the norm of the difference $\|\hat\beta_{\mathrm{LS}}^{(n-1,\ell)} - \hat\beta_{\mathrm{LS}}^{(n)}\|$ in the form

$$\|\hat\beta_{\mathrm{LS},j}^{(n-1,\ell)} - \hat\beta_{\mathrm{LS},j}^{(n)}\|[\mathrm{var}\,\|\hat\beta_{\mathrm{LS},j}^{(n-1,\ell)} - \hat\beta_{\mathrm{LS},j}^{(n)}\|]^{-1/2} = |(Y_\ell - X_\ell^{\mathrm{T}}\hat\beta_{\mathrm{LS}}^{(n)})\sigma^{-1}|.$$

Let us assume the continuous ψ-function. For the M-estimators the presence of $o_p(1)$ in (6.3) generally does not allow to derive directly from (6.3) an approximation to the variance of $\|\hat\beta^{(n-1,\ell)} - \hat\beta^{(n)}\|$. But this fact indicates that the term $o_p(1)$ in the representation (6.3) may cause that the exact variance of $\|\hat\beta^{(n-1,\ell)} - \hat\beta^{(n)}\|$ (or of $\hat\beta_j^{(n-1,\ell)} - \hat\beta_j^{(n)}$) is much larger than

$$(10.2) \qquad \sigma^2_{\|\cdot\|} = \mathsf{E}_F^{-2}\psi'\left(\frac{e_1}{\sigma}\right)\mathrm{trace}\{Q^{-1}\}\,\mathrm{var}\{\psi([Y_\ell - g(X_\ell, \hat\beta^{(n)})]\hat\sigma_n^{-1})\}\sigma^2$$

(or than

$$(10.3) \qquad \sigma^2_{(j)} = \mathsf{E}_F^{-2}\psi'\left(\frac{e_1}{\sigma}\right)\{Q^{-1}\}_{jj}\,\mathrm{var}\{\psi([Y_\ell - g(X_\ell, \hat\beta^{(n)})]\hat\sigma_n^{-1})\}\sigma^2).$$

It means that the large values of $\mathrm{var}\,\|\hat\beta^{(n-1,\ell)} - \hat\beta^{(n)}\|$ (and of $\mathrm{var}(\hat\beta_j^{(n-1,\ell)} - \hat\beta_j^{(n)})$) might be caused by the fluctuation of $\|\hat\beta^{(n-1,\ell)} - \hat\beta^{(n)}\|$ (and of $\hat\beta_j^{(n-1,\ell)} - \hat\beta_j^{(n)}$) on a set of (very) small probability. But then we may prefer to "studentize" $\|\hat\beta^{(n)} - \hat\beta^{(n-1,\ell)}\|$ by (10.2) (or $\hat\beta_j^{(n-1,\ell)} - \hat\beta_j^{(n)}$ by (10.3)), i.e. by the asymptotic variance of $\|\hat\beta^{(n)} - \hat\beta^{(n-1,\ell)}\|$ rather than by an approximation to the exact variance. We obtain

$$\|\hat\beta^{(n-1,\ell)} - \hat\beta^{(n)}\| \cdot \sigma^{-1}_{\|\cdot\|} = \frac{|Y_\ell - g(X_\ell, \hat\beta^{(n)})|}{\mathrm{var}^{1/2}\{\psi([Y_\ell - g(X_\ell, \beta)]\hat\sigma_n^{-1})\}}$$

or (independently for any $j = 1, 2, \ldots, p$)

$$(\hat\beta_j^{(n-1,\ell)} - \hat\beta_j^{(n)})\sigma^{-1}_{(j)} = \frac{\psi(Y_\ell - g(X_\ell, \hat\beta^{(n)}))}{\mathrm{var}^{1/2}\{\psi([Y_\ell - g(X_\ell, \beta)]\hat\sigma_n^{-1})\}}.$$

One may also observe that the difference in the prediction of the response variable based on the estimate $\hat\beta^{(n)}$ or on $\hat\beta^{(n-1,\ell)}$ is proportional to the same quantity. In fact, for any $\ell = 1, 2, \ldots, n$ and some $X \in R^p$ we obtain

$$\hat Y^{(n)} - \hat Y^{(n-1,\ell)} = X^{\mathrm{T}}\{\hat\beta^{(n)} - \hat\beta^{(n-1,\ell)}\}$$

and hence

$$\sup_{\|X\|=1} \{|\hat{Y}^{(n)} - \hat{Y}^{(n-1,\ell)}| \|X\|^{-1}\} = \|\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,\ell)}\|.$$

It also follows from (6.3) *that for* $\psi_s \equiv 0$ *the largest change of the estimates of regression coefficients cannot overcome some bound which is proportional to* $\sup_{z\in R} |\psi(z)|$. *Whenever* $\psi_s \not\equiv 0$ *the change may be much larger* (*see* (7.3)). *It hints that it is presumably better to avoid discontinuous* ψ*-functions.*

Further, as we have already observed in Remark 9 the only random factor in (6.3) which depends on the d.f. F is $\psi(Y_\ell - g(X_\ell, \hat{\beta}^{(n)}))$, range of which is bounded by $\inf_{z\in R} \psi(z)$ and $\sup_{z\in R} \psi(z)$. It means that in the case when the gross error sensitivity of the estimator is properly assigned, i.e. when some outliers are actually winsorized, $\max_{1\le\ell\le n} \|\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,\ell)}\|$ is nearly always equal to $\sup_{z\in R} |\psi(z)|$ multiplied by some constant. So, it seems somewhat strange to try to test significance of the largest change of the estimates. It implies that to create a possibility to test significance of the change, we need to exclude some, sufficiently large subsample of data. The percentage of the excluded observations has to be larger than the contamination level. We hope the problem will be treated in the forthcoming paper.

Acknowledgements

We would like to express our gratitude to the anonymous referee for carefully reading the manuscript. In fact, the present form of the proof of Lemma A.1 is due to him/her. We also thank the valuable suggestions which led to the general improvement of the paper, in order that this, as we hope, is easier to read.

Appendix

Remark 10. In the proofs we shall need some constants C_m, $m = 1, 2, \ldots$, definitions of which will be straightforward. The definitions of the constants will hold only within the given proof.

PROOF OF LEMMA 5.1. First of all, let us put

$$\begin{aligned}
&\xi(X_i, n^{-1/2}t, n^{-1/2-\tau}u, e^{-n^{-1/2}v}) \\
&\quad = \min\{e_i - \delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u), e_i - \delta(X_i, n^{-1/2}t)\}\sigma^{-1}e^{-n^{-1/2}v}, \\
&\zeta(X_i, n^{-1/2}t, n^{-1/2-\tau}u, e^{-n^{-1/2}v}) \\
&\quad = \max\{e_i - \delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u), e_i - \delta(X_i, n^{-1/2}t)\}\sigma^{-1}e^{-n^{-1/2}v}.
\end{aligned}$$

Under the assumptions of the lemma we have for $j = 1, 2, \ldots, p$

$$\begin{aligned}
&S_j(n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v}) \\
&\quad = -n^{-1/2-\tau} \sum_{i=1}^{n} \{\psi'(\tilde{\eta}_i) g_j'(X_i, n^{-1/2}t, n^{-1/2-\tau}\tilde{u}) \\
&\qquad \cdot [g(X_i, n^{-1/2}t, n^{-1/2-\tau}\tilde{u})]^{\mathrm{T}} \sigma^{-1} e^{-n^{-1/2}v} \\
&\qquad - \psi(\tilde{\eta}_i)[g_{j\cdot}''(X_i, n^{-1/2}t, n^{-1/2-\tau}\tilde{u})]^{\mathrm{T}}\} u
\end{aligned}$$

where $\tilde{u} = \tilde{u}(\tau)$ and $\tilde{\eta}_i = \tilde{\eta}_i(\tau, t, \tilde{u}, v)$ are appropriately selected and we have $\|\tilde{u}\| < \|u\|$ and

$$\tilde{\eta}_i \in (\xi(X_i, n^{-1/2}t, n^{-1/2-\tau}u, e^{-n^{-1/2}v}), \zeta(X_i, n^{-1/2}t, n^{-1/2-\tau}u, e^{-n^{-1/2}v})).$$

Using the fact that the derivatives ψ', g' and g'' are Lipschitz we obtain (uniformly for $j,k = 1,2,\ldots,p$, $n > 4\delta_0^{-2}M^2$, uniformly in $X_i \in R^q$, $i = 1,2,\ldots,n$ and $t,u \in \mathcal{T}_M$)

$$\left|\psi'(\tilde{\eta}_i) - \psi'\left(\frac{e_i}{\sigma}\right)\right| \le n^{-1/2} \cdot C_1,$$
$$|g_j'(X_i, n^{-1/2}t, n^{-1/2-\tau}\tilde{u}) - g_j'(X_i, \beta^0)| \le n^{-1/2} \cdot C_2$$

and

$$|g_{jk}''(X_i, n^{-1/2}t, n^{-1/2-\tau}\tilde{u}) - g_{jk}''(X_i, \beta^0)| \le n^{-1/2} \cdot C_3$$

where C_1, C_2 and C_3 are finite constants. It implies that

$$\sup_{\mathcal{T}_M} \Bigg| S_j(n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})$$
$$+ n^{-1/2-\tau} \sum_{i=1}^{n} \Bigg\{ \psi'\left(\frac{e_i}{\sigma}\right) g_j'(X_i, \beta^0)[g'(X_i, \beta^0)]^{\mathrm{T}} \sigma^{-1}$$
$$- \psi\left(\frac{e_i}{\sigma}\right)[g_j''(X_i, \beta^0)]^{\mathrm{T}} \Bigg\} u \Bigg| = O(n^{-\tau}) \quad \text{as} \quad n \to \infty.$$

The application of the central limit theorem and of the law of large numbers concludes the proof of lemma. □

PROOF OF LEMMA 5.2. From the character of the ψ-function it follows that $|\mathsf{E}_F \psi'(\frac{e_1}{\sigma})| < \infty$. Let us denote for any $u \in R^p$ and $k \in \{1,2,\ldots,p\}$

$$z_n^{(k)} = (u_1, u_2, \ldots, u_{k-1}, z, 0, \ldots, 0)^{\mathrm{T}}$$

and

$$\mathcal{H}_n(n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})$$
$$= \{i \in N : (\xi(X_i, n^{-1/2}t, n^{-1/2-\tau}u, e^{-n^{-1/2}v}),$$
$$\zeta(X_i, n^{-1/2}t, n^{-1/2-\tau}u, e^{-n^{-1/2}v}))$$
$$\cap \{r_1, r_2, \ldots, r_h\} \neq \emptyset\}.$$

Now, we may write for any $i = 1,2,\ldots,n$, $t,u \in R^p$ and $v \in R^+$

$$s_j(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})$$
$$= n^{-1/2-\tau} \sum_{k=1}^{p} \int_0^{u_k} \{-\psi'([e_i - \delta(X_i, n^{-1/2}t, n^{-1/2-\tau}z_n^{(k)})]\sigma^{-1}e^{-n^{-1/2}v})$$
$$\times g_j'(X_i, n^{-1/2}t, n^{-1/2-\tau}z_n^{(k)})$$
$$\times g_k'(X_i, n^{-1/2}t, n^{-1/2-\tau}z_n^{(k)}) \cdot \sigma^{-1}e^{-n^{-1/2}v}$$
$$+ \psi([e_i - \delta(X_i, n^{-1/2}t, n^{-1/2-\tau}z_n^{(k)})]\sigma^{-1}e^{-n^{-1/2}v})$$
$$\times g_{jk}''(X_i, n^{-1/2}t, n^{-1/2-\tau}z_n^{(k)})\}\mathrm{d}z.$$

So, for $i \in \mathcal{H}_n(n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})$ and for $t, u, v \in \mathcal{T}_M$

$$|s_j(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})| \leq n^{-1/2-\tau} \cdot K \cdot J \cdot (J+1)p^{1/2}\|u\|.$$

Moreover, $I_{\{i\in\mathcal{H}_n\}} = 1$ implies that there is $\ell \in \{1, 2, \ldots, h\}$ such that $r_\ell \in (\xi(X_i, n^{-1/2}t, n^{-1/2-\tau}u, e^{-n^{-1/2}v}), \zeta(X_i, n^{-1/2}t, n^{-1/2-\tau}u, e^{-n^{-1/2}v}))$, i.e. either

$$r_\ell \sigma e^{n^{-1/2}v} + \delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) \leq e_i \leq r_\ell \sigma e^{n^{-1/2}v} + \delta(X_i, n^{-1/2}t)$$

or

$$r_\ell \sigma e^{n^{-1/2}v} + \delta(X_i, n^{-1/2}t) \leq e_i \leq r_\ell \sigma e^{n^{-1/2}v} + \delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u).$$

Assuming $\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) \geq \delta(X_i, n^{-1/2}t)$ we have

$$\begin{aligned}(\text{A.1}) \quad & F(r_\ell \sigma e^{n^{-1/2}v} + \delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)) \\ & \quad - F(r_\ell \sigma e^{n^{-1(1/2)}v} + \delta(X_i, n^{-1/2}t)) \\ & = n^{-1/2-\tau} \sum_{k=1}^{p} \int_0^{u_k} f(r_\ell \sigma e^{n^{-1(1/2)}v} + \delta(X_i, n^{-1/2}t, n^{-1/2-\tau}uz_n^{(k)})) \\ & \qquad \times g_k'(X_i, n^{-1/2}t, n^{-1/2-\tau}z_n^{(k)})\mathrm{d}z\end{aligned}$$

and so

$$P(I_{\{i\in\mathcal{H}_n\}} = 1) \leq 2n^{-1/2-\tau} \cdot H \cdot J \cdot h \cdot p^{1/2}\|u\|$$

(for H see Condition B (ii)). Moreover, $\mathsf{E}\{n^{-1/2-\tau}\sum_{i=1}^n |I_{\{i\in\mathcal{H}_n\}}|\} \leq 2n^{-2\tau}H \cdot J \cdot h \cdot p^{1/2}\|u\|$ and the Chebyshev inequality for nonnegative random variable gives for any $C_1 > 0$

$$\begin{aligned} & P\left(\sup_{\mathcal{T}_M} \left|\sum_{i=1}^n s_j(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})I_{\{i\in\mathcal{H}_n\}}\right| > C_1\right) \\ & \quad \leq P_G\left(\sup_{\mathcal{T}_M} \sum_{i=1}^n \left|s_j(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})I_{\{i\in\mathcal{H}_n\}}\right| > C_1\right) \\ & \quad \leq C_1^{-1}\mathsf{E}\left\{n^{-1/2-\tau}\sum_{i=1}^n I_{\{i\in\mathcal{H}_n\}}\right\},\end{aligned}$$

i.e.

$$\sum_{i=1}^n s_j(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})I_{\{i\in\mathcal{H}_n\}} = O_p(n^{-2\tau}).$$

Along the same lines we may prove that

$$n^{-1/2-\tau}\sup_{\mathcal{T}_M}\left|\sum_{i=1}^n \left\{\psi'\left(\frac{e_i}{\sigma}\right) g_j'(X_i, \beta^0)[g'(X_i, \beta^0)]^{\mathrm{T}} I_{\{i\in\mathcal{H}_n\}}\right\} u\right| = O_p(n^{-2\tau}),$$

and

$$n^{-1/2-\tau}\sup_{\mathcal{T}_M}\left|\sum_{i=1}^{n}\left\{\psi\left(\frac{e_i}{\sigma}\right)g_{j\cdot}''(X_i,\beta^0)I_{\{i\in\mathcal{H}_n\}}\right\}u\right| = O_p(n^{-2\tau}),$$

and since

$$\begin{aligned}S_j(n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v}) &= \sum_{i=1}^{n} s_j(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})I_{\{i\notin\mathcal{H}_n\}} \\ &\quad + \sum_{i=1}^{n} s_j(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma e^{n^{-1/2}v})I_{\{i\in\mathcal{H}_n\}},\end{aligned}$$

the same steps which were performed in the proof of Lemma 5.1 conclude the proof of Lemma 5.2. □

LEMMA A.1. *Let for some* $p \in N$, $\{\mathcal{V}^{(n)}\}_{n=1}^{\infty}$, $\mathcal{V}^{(n)} = \{v_{ij}^{(n)}\}_{i=1,2,\dots,p}^{j=1,2,\dots,p}$ *be a sequence of* $(p\times p)$ *matrixes such that for* $i = 1,2,\dots,p$ *and* $j = 1,2,\dots,p$

$$\lim_{n\to\infty} v_{ij}^{(n)} = q_{ij} \quad \textit{in probability} \tag{A.2}$$

where $Q = \{q_{ij}\}_{i=1,2,\dots,p}^{j=1,2,\dots,p}$ *is a fixed nonrandom regular matrix. Moreover, let* $\{\theta^{(n)}\}_{n=1}^{\infty}$ *be a sequence of p-dimensional random vectors such that*

$$\exists(\varepsilon>0)\quad \forall(K>0)\quad \limsup_{n\to\infty} P(\|\theta^{(n)}\|>K)>\varepsilon. \tag{A.3}$$

Then

$$\exists(\delta>0)\quad \forall(L>0)$$

so that

$$\limsup_{n\to\infty} P(\|\mathcal{V}^{(n)}\theta^{(n)}\|>L)>\delta.$$

PROOF. Due to (A.2) the matrix $\mathcal{V}^{(n)}$ is regular in probability. Let then $0<\lambda_{1n}<\lambda_{2n}<\cdots<\lambda_{pn}$ and $z_{1n}, z_{2n},\dots,z_{pn}$ be eigenvalues and corresponding eigenvectors (selected to be mutually orthogonal) of the matrix $[\mathcal{V}^{(n)}]^{\mathrm{T}}\mathcal{V}^{(n)}$. Let us write $\theta^{(n)} = \sum_{j=1}^{p} a_{jn}z_{jn}$ (for an appropriate vector $a_n = (a_{1n}, a_{2n},\dots,a_{pn})^{\mathrm{T}}$). Then we have

$$\|\mathcal{V}^{(n)}\theta^{(n)}\|^2 = \sum_{j=1}^{p}[a_{jn}]^2\lambda_{jn}\|z_{jn}\|^2 \geq \lambda_{1n}\|\theta^{(n)}\|^2. \tag{A.4}$$

Moreover, denoting λ_1 the smallest eigenvalue of the matrix $Q^{\mathrm{T}}Q$, we have $\lambda_{1n}\to\lambda_1$ as $n\to\infty$. The assertion of the lemma then follows from (A.4). □

PROOF OF LEMMA 6.1. Let us recall that according to Conditions C we have

$$\sqrt{n}(\hat\beta^{(n)} - \hat\beta^{(n-1,\ell)}) = O_p(1) \quad \text{as} \quad n\to\infty$$

and let us put $\tilde{t}_n = \sqrt{n}(\hat{\beta}^{(n)} - \beta^0)$, $\tilde{u}_n = \sqrt{n}(\hat{\beta}^{(n-1,\ell)} - \hat{\beta}^{(n)})$ and $\tilde{v}_n = \sqrt{n}(\log \hat{\sigma}_n - \log \sigma)$. Then there is a constant $C_1 > 0$ so that starting with some n_ε we have

$$P(\max\{\|\tilde{t}\|, \|\tilde{u}\|\|\tilde{v}|\} < C_1) > 1 - \varepsilon.$$

Considering $\tau = 0$ and substituting $\tilde{t}_n$, $\tilde{u}_n$ and $\tilde{v}_n$ into Lemmas 5.1 and 5.2 we obtain

$$\begin{aligned}\sum_{i=1}^{n} &[\psi([Y_i - g(X_i, \hat{\beta}^{(n-1,\ell)})]\sigma_n^{-1})g'(X_i, \hat{\beta}^{(n-1,\ell)}) \\ &\quad - \psi([Y_i - g(X_i, \hat{\beta}^{(n)})]\sigma_n^{-1})g'(X_i, \hat{\beta}^{(n)})] \\ &\quad - [\sigma_n^{-1}\gamma Q + o_p(1)]n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,\ell)}) = O_p(1) \quad \text{as} \quad n \to \infty.\end{aligned}$$

Finally, taking into account (6.1) and (6.2)

$$\begin{aligned}&[\sigma_n^{-1}\gamma Q + o_p(1)]n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,\ell)}) \\ &\quad = \psi([Y_\ell - g(X_\ell, \hat{\beta}^{(n-1,\ell)})]\sigma_n^{-1})g'(X_\ell, \hat{\beta}^{(n-1,\ell)}) + O_p(1) \quad \text{as} \quad n \to \infty.\end{aligned}$$

The application of Lemma A.1 concludes the proof. □

LEMMA A.2. (Štěpán (1987), p. 420, VII.2.8) *Let a and b be positive numbers. Further let ξ be a random variable such that $P(\xi = -a) = \pi$ and $P(\xi = b) = 1 - \pi$ (for a $\pi \in (0,1)$) and $E\xi = 0$. Moreover let τ be the time for the Wiener process $W(s)$ to exit the interval $(-a, b)$. Then*

$$\xi =_{\mathcal{D}} W(\tau)$$

where "$=_{\mathcal{D}}$" denotes the equality of distributions of the corresponding random variables. Moreover, $E\tau = a \cdot b = \operatorname{var}\xi$.

Remark 11. Since the book of Štěpán (1987) is in Czech language we refer also to Breiman (1968) where this simple assertion is not isolated. Nevertheless, the assertion can be found directly in the first lines of the proof of Proposition 13.7 (p. 277) of Breiman's book. (See also Theorem 13.6 on the p. 276.)

PROOF OF LEMMA 7.1. First of all, we shall consider

$$S_n(t,u) = \sum_{i=1}^{n}[\psi_m(e_i - n^{-1/2}X_i^T t - n^{-1}X_i^T u) - \psi_m(e_i - n^{-1/2}X_i^T t)]X_i.$$

According to the assumptions, there is a $\Delta > 0$ and $C_1 < \infty$ such that for any $|z| < \Delta$ we have $|f(z)| < C_1$. Let us denote $\xi_i(n,t,u) = \psi_m(e_i - n^{-1/2}X_i^T t - n^{-1}X_i^T u) - \psi_m(e_i - n^{-1/2}X_i^T t)$. It is clear that $\xi_i(n,t,u) \neq 0$ with positive probability only if either

$$\begin{aligned}(A.5) \qquad &e_i - n^{-1/2}X_i^T t < 0 < e_i - n^{-1/2}X_i^T t - n^{-1}X_i^T u \\ &\quad \leftrightarrow n^{-1/2}X_i^T t + n^{-1}X_i^T u < e_i < n^{-1/2}X_i^T t\end{aligned}$$

or

$$\text{(A.6)} \qquad \begin{aligned} & e_i - n^{-1/2} X_i^T t - n^{-1} X_i^T u < 0 < e_i - n^{-1/2} X_i^T t \\ & \quad \leftrightarrow n^{-1/2} X_i^T t < e_i < n^{-1/2} X_i^T t + n^{-1} X_i^T u. \end{aligned}$$

Let us denote the probability of this event by $\pi_i(n,t,u)$ (observe that of course, for given i only one of the events (A.5) or (A.6) can appear). We have, starting with $n > 4pK^2C^2\Delta^{-1}$, in the case (A.5)

$$\pi_i(n,t,u) = \int_{n^{-1/2}X_i^T t + n^{-1}X_i^T u}^{n^{-1/2}X_i^T t} f(z)\mathrm{d}z \leq n^{-1}C_2$$

for some finite positive C_2. The same is true for (A.6). Further, we shall assume $\sum_{i=1}^n X_{i\ell}[\xi_i - E\xi_i]$ and $t, u \in \mathcal{T}_M$. As a first possibility let us consider that $X_i^T u < 0$ and $X_{i\ell} > 0$. We easy find that

$$\begin{aligned} X_{i\ell}[\xi_i(n,t,u) - E_F\xi_i(n,t,u)] &= 2X_{i\ell}(1 - \pi_i(n,t,u)) \\ &= 2|X_{i\ell}|(1 - \pi_i(n,t,u)) < 2|X_{i\ell}| \\ &\qquad \text{with probability } \pi_i(n,t,u), \\ &= -2X_{i\ell}\pi_i(n,t,u) \\ &= -2|X_{i\ell}|\pi_i(n,t,u) > -2n^{-1}|X_{i\ell}|C_2 \\ &\qquad \text{with probability } 1 - \pi_i(n,t,u). \end{aligned}$$

For $X_{i\ell} < 0$ we have

$$\begin{aligned} X_{i\ell}[\xi_i(n,t,u) - E_F\xi_i(n,t,u)] &= 2X_{i\ell}(1 - \pi_i(n,t,u)) \\ &= -2|X_{i\ell}|(1 - \pi_i(n,t,u)) > -2|X_{i\ell}| \\ &\qquad \text{with probability } \pi_i(n,t,u), \\ &= -2X_{i\ell}\pi_i(n,t,u) \\ &= 2|X_{i\ell}|\pi_i(n,t,u) < 2n^{-1}|X_{i\ell}|C_2, \\ &\qquad \text{with probability } 1 - \pi_i(n,t,u). \end{aligned}$$

Having analyzed in the same way the case $X_i^T u > 0$, we find that for $X_i^T u \cdot X_{i\ell} > 0$

$$\text{(A.7)} \quad X_{i\ell}[\xi_i(n,t,u) - E_F\xi_i(n,t,u)] = -2|X_{i\ell}|(1 - \pi_i(n,t,u)) > -2|X_{i\ell}|,$$
$$\text{with probability } \pi_i(n,t,u),$$
$$\text{(A.8)} \qquad = 2|X_{i\ell}|\pi_i(n,t,u) < 2n^{-1}|X_{i\ell}|C_2,$$
$$\text{with probability } 1 - \pi_i(n,t,u),$$

and for $X_i^T u \cdot X_{i\ell} < 0$

$$\text{(A.9)} \quad X_{i\ell}[\xi_i(n,t,u) - E_F\xi_i(n,t,u)] = 2|X_{i\ell}|(1 - \pi_i(n,t,u)) < 2|X_{i\ell}|$$
$$\text{with probability } \pi_i(n,t,u),$$

(A.10) $$= -2|X_{i\ell}|\pi_i(n,t,u) > -2n^{-1}|X_{i\ell}|C_2$$
with probability $1 - \pi_i(n,t,u)$.

So, putting for $X_i^T u \cdot X_{i\ell} > 0$ $a_{i\ell}(t,u) = 2|X_{i\ell}|(1-\pi_i(n,t,u))$ and $b_{i\ell}(t,u) = 2|X_{i\ell}|\pi_i(n,t,u)$, and for $X_i^T t \cdot X_{i\ell} < 0$ $a_{i\ell}(t,u) = 2n^{-1}|X_{i\ell}|\pi_i(n,t,u)$ and $b_{i\ell}(t,u) = 2|X_{i\ell}|(1-\pi_i(n,t,u))$, we may utilize Lemma A.2 and define

$\mu_{i\ell}(n,t,u)$ the time for Wiener process to exit the interval $(-a_{i\ell}(t,u), b_{i\ell}(t,u))$

and we obtain (since $W(\mu_{n\ell}(n,t,u)) =_{\mathcal{D}} W(\sum_{i=1}^n \mu_{i\ell}(n,t,u)) - W(\sum_{i=1}^{n-1} \cdot \mu_{i\ell}(n,t,u))$

$$[S_{n\ell}(t,u) - \mathsf{E}_F S_{n\ell}(t,u)] =_{\mathcal{D}} \sum_{i=1}^n W(\mu_{i\ell}(n,t,u)) =_{\mathcal{D}} W\left(\sum_{i=1}^n \mu_{i\ell}(n,t,u)\right)$$

(see again Theorem 13.6 of Breiman (1968)). Now, due to inequalities which are given in (A.7), (A.8), (A.9) and (A.10), putting $c_{i\ell} = 2n^{-1}|X_{i\ell}|C_2$ and $d_{i\ell} = 2|X_{i\ell}|$ and defining

(A.11) $\kappa_i^+(n,M)$ the time for Wiener process to exit the interval $(-c_{i\ell}, d_{i\ell})$

and

(A.12) $\kappa_i^-(n,M)$ the time for Wiener process to exit the interval $(-d_{i\ell}, c_{i\ell})$

we obtain

$$\mu_i(n,t) \le \kappa_i^+(n,M) + \kappa_i^-(n,M) = \kappa_i(n,M),$$

so that

(A.13) $$\sup_{\mathcal{T}_M} |S_{n\ell}(t) - E_F S_{n\ell}(t)| =_{\mathcal{D}} \sup_{\mathcal{T}_M} \left| W\left(\sum_{i=1}^n \mu_i(n,t,u)\right)\right|$$
$$\le \sup\left\{|W(s)| : 0 \le s \le \sum_{i=1}^n \kappa_i(n,M)\right\}.$$

Moreover, see again Lemma A.2, we have from (A.11) and (A.12) for any $t, u \in \mathcal{T}_M$

$$E_F \kappa_i(n,C) \le 4n^{-1} X_{i\ell}^2 C_2 \le n^{-1} C_3$$

for some $C_3 \in (0,\infty)$ for all $n \in N$, i.e.

$$E_F\left\{\sum_{i=1}^n \kappa_i(n,C)\right\} \le C_3.$$

It means that for any positive ε there is a constant C_4 and $n_\varepsilon \in N$ so that for any $n > n_\varepsilon$

(A.14) $$P\left\{\sum_{i=1}^n \kappa_i(n,C) > C_4\right\} < \frac{\varepsilon}{2}$$

and then there is also $C_5 \in (0, \infty)$ such that

$$P\{\sup\{|W(s)| : 0 \leq s \leq C_4\} > C_5\} < \frac{\varepsilon}{2}, \tag{A.15}$$

see e.g. Csörgö and Révész (1981). Taking into account (A.13), (A.14) and (A.15), we arrive at

$$P\left\{\sup_{\mathcal{T}_M} |S_{n\ell}(t) - E_F S_{n\ell}(t)| > C_5\right\} < \varepsilon$$

and it means that also

$$\sup_{\mathcal{T}_M} \|S_n(t) - E_F S_n(t)\|$$

is bounded in probability. We shall finish the proof if we show that

$$\sup_{\mathcal{T}_M} \|E_F S_n(t) - 2f(0)Qu\| = \mathcal{O}(1).$$

Let $X_i^T u > 0$. Then

$$\begin{aligned} E_F X_i \xi_i(n,t,u) &= 2X_i \int_0^{n^{-1}X_i^T u} f(z + n^{-1/2} X_i^T t)\mathrm{d}z \\ &= 2X_i \int_0^{n^{-1}X_i^T u} [f(0) + f(z + n^{-1/2}X_i^T t) - f(0)]\mathrm{d}z \\ &= n^{-1} X_i X_i^T u f(0) + R_{int}^*. \end{aligned}$$

Since $f(z)$ is Lipschitz, we have $|f(z + n^{-1/2}X_i^T t) - f(0)| < n^{-1/2}C_6$ and hence

$$\|R_{int}^*\| \leq n^{-1/2} \left\| C_6 \cdot X_i \cdot \int_0^{n^{-1}X_i^T u} \mathrm{d}z \right\| = n^{-3/2} C_6 \|X_i X_i^T u\|$$

and it implies that (due to (2.3))

$$\left\| \sum_{i=1}^n E_F X_i \xi_i(n,t,u) - Q_n \cdot u \cdot f(0) \right\| \leq n^{-1/2} \|C_6 \cdot Q_n \cdot u\| = o_p(1)$$

which concludes the proof. □

References

Atkinson, A. C. (1985). *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Claredon Press, Oxford.

Bates, D. M. and Watts D. G. (1988). *Nonlinear Regression Analysis and Its Applications*, Wiley, New York.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley, New York.

Breiman, L. (1968). *Probability*, Addison-Wesley, Reading, Massachusetts.

Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*, Wiley, New York.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman and Hall, London.

Csörgö, M. and Révész, P. (1981). *Strong Approximation in Probability and Statistics*, Akademia Kiadó, Budapest.

Hampel, F. R., Rousseeuw, P. Y., Ronchetti, E. M. and Stahel, W. A. (1986). *Robust Statistics—The Approach Based on Influence Functions*, Wiley, New York.

Hand, D. J., Daly, F., Lunn, A. D., McConway, K. J. and Ostrowski, E. (1994). *Handbook of Small Data Sets*, Chapman & Hall, London.

Hettmansperger, T. P. and Sheather, S. J. (1992). A cautionary note on the method of least median squares, *Amer. Statist.*, **46**, 79–83.

Huber, P. J. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, **35**, 73–101.

Jurečková, J. (1988). Constistency of M-estimators of vector parameters, *Proceedings of the Fourth Prague Symposium on Asymptotic Statistics*, 305–312, Charles University, Prague.

Liese, F. and Vajda, I. (1995). Consistency of M-estimators in general models, *J. Multivariat. Anal.*, **50**, 93–114.

Mason, R. L., Gunst, R. F. and Hess, J. L. (1989). *Statistical Design and Analysis of Experiments*, Wiley, New York.

Rao, C. R. and Zhao, L. C. (1992). On the consistency of M-estimate in linear model obtained through an estimating equation, *Statist. Probab. Lett.*, **14**, 79–84.

Rousseeuw, P. J. (1993). Unconventional features of positive-breakdown estimators, Report no. 93-26 of University of Antwerp, Department of Mathematics & Computer Science, UIA, Belgium.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*, Wiley, New York.

Rubio, A. M. and Víšek, J. Á. (1996). A note on asymptotic linearity of M-statistics in nonlinear models, *Kybernetika* (to appear).

Rubio, A. M., Quintana, F. and Víšek, J. Á. (1994). Test for differences between M-estimates of non-linear regression model, *Probab. Math. Statist.*, **14**, Fasc. 2, 189–206.

Sen, A. and Srivastava, M. (1990). *Regression Analysis: Theory, Methods and Applications*, Springer, Berlin.

Štěpán, J. (1987). *Teorie pravděpodobnosti* (*Probability Theory*), Academia, Prague.

Vandaele, W. (1978). Participation in illegitimate activities: Erlich revisted, *Deterrence and Incapacitation* (eds. A. Blumstein, J. Cohen and D. Nagin), 270–335, National Academy of Sciences, Washington, D.C.

Víšek, J. Á. (1992*a*). A proposal of model selection: Stability of linear regression model, *Transactions of the Eleventh Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Volume B, 451–461, Academia, 1990, Prague.

Víšek, J. Á. (1992*b*). Stability of regression estimates with respect to subsamples, *Computational Statistics*, **7**, 183–203.

Víšek, J. Á. (1994). A cautionary note on the method of least median of squares reconsidered, *Transactions of the Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, 254–259, Prague.