PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 17, Fasc. 2 (1997), pp. 231–257

# DIAGNOSTICS OF REGRESSION SUBSAMPLE STABILITY*

BY

JAN ÁMOS VÍŠEK (PRAGUE)

*Abstract.* An asymptotic representation of Bahadur type of the difference of $M$-estimators $\hat{\beta}_{\psi}^{(n)} - \hat{\beta}_{\psi}^{(n-1,l)}$, i.e. of the difference of estimators of regression coefficients for the full data set and for the set from which the $l$-th observation was deleted, is given under the conditions covering the most of $\psi$-functions which are at the present time used in the robust statistics, including the discontinuous and redescending ones. The representation is invariant with respect to the scale of residuals.

**1. Introduction.** Pre- and post-application diagnostics has become a standard part of any data processing theory, and the statistics is not an exception. In the region of the least squares regression analysis let us mention today nearly classical monographs by Belsley et al. [3], Cook and Weisberg [8], Atkinson [1], Bates and Watts [2], Chatterjee and Hadi [5] or Sen and Srivastava [32] to give at least some among many others. In robust regression the diagnostics are sometimes assumed to be something which is complementary to the robust algorithms (see e.g. Huber [14]). Other authors propose to use the robust algorithms as tools for diagnostics of data (see Hampel et al. [12] or Rousseeuw and Leroy [28]). It is surely possible but some caution is necessary (Hettmansperger and Sheather [13]) because it is even formally easy to show that two consistent estimators may give for an arbitrary large sample size arbitrarily different estimates of the regression model; see Víšek [38] and the references given there. Due to the fact that the robust statistics has offered for the applications large scale of methods we may rather frequently meet a situation when the numerical results of the estimation of regression model by different (highly) robust algorithms gives considerably different models; see Víšek [36]. Then we need to make an idea which of results is acceptable for our data (or adequate for the data, if you prefer this word) and it asks for diagnostics too, see Rubio and Víšek [30] and Nosková [24].

* Research was supported by grant of GA CR number 201/94/0322.

Interesting discussions on diagnostics may be also found in some papers of the transactions by Stahel and Weisberg [34], especially in papers by Ledolter [18], McKean et al. [21], Neykov and Neytchev [23], Portnoy [26] and Simonoff [33].

In Víšek [39] the Bahadur-type representation of the difference of the estimators of regression coefficients for the full data and for the data from which one observation has been deleted is presented in the random-carriers-framework for the continuous $\psi$-functions and for the $\psi$-function which is equal to sign($x$), i.e. which corresponds to the median in the location problem. This paper brings this Bahadur representation for the general type of the discontinuous $\psi$-function, moreover taking into account a practical need of the rescaling of residuals.

The studies of the asymptotic representation of the difference of estimators are of course aimed to find a diagnostic tool for the situation when we look for the most influential point among the data. Let us recall that for the least squares, probably from the very early days of this discipline, the formula

$$\hat{\beta}_{\mathrm{LS}}^{(n-1,l)}-\hat{\beta}_{\mathrm{LS}}^{(n)} = -\{[X^{(n-1,l)}]^{\mathrm{T}} X^{(n-1,l)}\}^{-1} X_l(Y_l - X_l^{\mathrm{T}} \hat{\beta}_{\mathrm{LS}}^{(n)}) \tag{1}$$

was used in the same way. The notation is nearly self-explaining, nevertheless, $X^{(n-1,l)}$ is the design matrix after delection of the $l$-th row from the full design matrix $X$, and $X_l$ is the $l$-th row, assumed as a column vector, of the design matrix for the full data. Sometimes it is referred that the case deletion diagnostics for influential observations were introduced by Cook [6] or [7] (see e.g. Geisser [11]) but the formula (1) has already appeared in Miller [22] and maybe that it has been already used by Sir Francis Galton [10]; see also Chatterjee and Hadi [5].

This paper establishes a formula similar to (1) for the $M$-estimators including those which are generated by discontinuous $\psi$-functions. As already mentioned, it was done in Víšek [39] for the continuous $\psi$-functions. Instead of considering discontinuous functions in full generality Víšek [39] treated only median-type $\psi$-function in the framework of linear model. Generalization on an arbitrary discontinuous $\psi$-function and especially on the nonlinear model is technically somewhat complicated, however principally straightforward. Nevertheless, now we have at hand the Bahadur-type representation of the difference of the $M$-estimators of regression coefficients for the full data and for the data from which one observation has been deleted for all estimators used in robust statistics. This opens the possibility to study a similar type of Bahadur representation for the most influential subset of data. We hope to do this in a forthcoming paper.

**2. Notation and setup.** Let $N$ denote the set of all positive integers, $R^l$ the $l$-dimensional Euclidean space, $R^+$ the nonnegative part of the real line, and

$(\Omega, \mathscr{A}, P)$ a probability space. We shall consider for all $n \in N$ the model

$$Y_n = g(X_n, \beta^0) + e_n, \tag{2}$$

where for some fix $p, q \in N$, $\{X_n\}_{n=1}^{\infty}$ is a fixed sequence of vectors from $R^q$, and $\beta^0 = (\beta_1^0, \beta_2^0, \ldots, \beta_p^0)^T$ is a vector of regression parameters, where $^T$ stands for the transposition. Moreover, a function $g\colon R^{q+p} \to R$ is assumed to be twice differentiable (see Conditions A below) and, finally, $\{e_n\}_{n=1}^{\infty}$, $e_n\colon \Omega \to R$, is a sequence of independent and identically distributed random variables (i.i.d.r.v.), distributed according to the distribution function (d.f.) $F$. We will consider the $M$-estimators of $\beta^0$ given as

$$\hat{\beta}^{(n)} = \arg\min_{\beta \in R^p} \Big\{ \sum_{i=1}^{n} \varrho\big([Y_i - g(X_i, \beta)]\, \hat{\sigma}_n^{-1}\big) \Big\} \tag{3}$$

and

$$\hat{\beta}^{(n-1,l)} = \arg\min_{\beta \in R^p} \Big\{ \sum_{\substack{i=1 \\ i \neq l}}^{n} \varrho\big([Y_i - g(X_i, \beta)]\, \hat{\sigma}_n^{-1}\big) \Big\}, \tag{4}$$

where $\varrho\colon R \to R$ is assumed to be absolutely continuous (denote the derivative — at the points where it exists — by $\psi$) and $\hat{\sigma}_n$ is a preliminary estimator of the scale (see Conditions C below).

Remark 1. As follows from the given setup we shall consider the regression model with the fixed carriers. Although it may seem that it is less general than the setup with random carriers, it is not inevitably so — see e.g. Condition (ii) in Jurečková and Welsh [17] which would need some array of setups with the random carriers to allow an interpretation as a special case of a random setup. On the other hand, the most traditional conditions, in the both types of setups, have some corresponding counterparts, see e.g. Condition B (iii) in Rubio et al. [29] and Conditions A (iii) below. Of course, the setup with random carrier and the noise independent of carriers permits us to treat in a simpler way e.g.

$$n^{-1} \sum_{i=1}^{n} \psi'(e_i/\sigma) \sum_{i=1}^{n} g'(X_i, \beta^0)[g'(X_i, \beta^0)]^T$$

than the setup with nonrandom carriers. In the latter setup, formally a somewhat more complicated version of the law of large numbers for independent but not identically distributed r.v. has to be used although the spirit of the treatment is the same. However, as we shall see in the proof of Lemma 3, sometimes the setup with random carriers may be rather complicated to deal with, at least what concerns the formalism, so that one may prefer to explain the ideas how to reach the asymptotic representation in the framework with the fixed carriers. Of course, a detailed but unfortunately formally really complicated discussion would reveal that the spirit of the treatment is the same for the both setups.

The proofs of Lemmas 1 and 2 below are nearly the same as for the random-carriers-framework which was empoyed in Víšek [39].

**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 $\delta_0$ such that for any $\beta \in R^p$, $\|\beta-\beta^0\| < \delta_0$,

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

exist for any $x \in \{X_n\}_{n=1}^{\infty}$. Let us denote the vector of the first partial derivative and the matrix of the second derivatives 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 $g''_{jk}(x, \beta)(j, k = 1, 2, \ldots, p)$ are uniformly in $x \in \{X_n\}_{n=1}^{\infty}$ Lipschitz (of the first order) in $\beta$ in the $\delta_0$-neighborhood of $\beta^0$, i.e.

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

$$\max_{1 \leq j,k \leq p} \sup_{x \in \{X_n\}_{n=1}^{\infty}} |g''_{jk}(x, \beta) - g''_{jk}(x, \beta^0)| < L \|\beta-\beta^0\|.$$

Moreover, let

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

(iii) There is a regular matrix $Q$ such that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g'(X_i, \beta^0)[g'(X_i, \beta^0)]^{\mathrm{T}} = Q$$

and put $(Q)_{ij} = q_{ij}$.

Remark 2. It is clear that in view of the fact that $\beta_0$ is unknown, to be sure that Condition A (ii) is fulfilled we have to ask if the Lipschitz property of $g''_{jk}(x, \beta)$ holds (at least) in a "reasonable" subset of $R^p$, in dependence on our *a priori* knowledge about possible values of $\beta_0$. A similar statement, *perhaps* somewhat weaker, is true what concerns uniformity in $x \in \{X_n\}_{n=1}^{\infty}$. The word "perhaps" should indicate that our knowledge about possible values of $X$'s is a little bit better due to the fact that at the moment when we apply the results we know the first $n$ elements of the sequence $\{X_n\}_{n=1}^{\infty}$. On the other hand, in the applications we usually use (or if you want, meet with) rather smooth models $g(x, \beta)$. Moreover, for too wild models we would not be able to perform required computations yielding the corresponding estimate.

Let us observe that A (ii) implies that there is $J < \infty$ such that for

$$\mathscr{L} = \{x \in \{X_n\}_{n=1}^{\infty}, \beta \in R^p, \|\beta-\beta^0\| < \delta_0\}$$

we have

$$\max_{1 \leqslant j,k \leqslant p} \sup_{\mathscr{L}} \max \{|g(x,\beta)|, |g'_j(x,\beta)|, |g''_{jk}(x,\beta)|\} < J.$$

Of course, the boundedness of the model $g(x, \beta)$ for $\beta \in R^p$, $\|\beta - \beta^0\| < \delta$ for some positive $\delta$, is restrictive from the theoretical point of view. From the practical standpoint, the boundedness is nearly irrelevant because any unbounded model would be, in fact, useless. Finally, observe that the matrix $Q$ is positive definite. It is due to the fact that any real symmetric matrix can be written as $A\Lambda A^T$, where $A$ and $\Lambda$ are an orthonormal and a diagonal matrices, respectively. Moreover, for any $z \in R^p$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} z^T g'(X_i, \beta^0)[g'(X_i, \beta^0)]^T z = z^T Q z \geqslant 0,$$

which implies that $\Lambda$ has only nonnegative elements. However, the regularity of $Q$ implies that all diagonal elements of $\Lambda$ are nonzero. Then for any $z \neq 0$

$$z^T Q z = \xi^T \xi > 0, \quad \text{where } \xi = \Lambda^{1/2} A^T z.$$

CONDITIONS B. (i) The function $\psi$ allows decomposition in the form

$$\psi = \psi_a + \psi_c + \psi_s, \tag{5}$$

where $\psi_a$ has a derivative $\psi'_a$ which is Lipschitz of the first order, $\psi_c$ is a continuous function with derivative $\psi'_c$ being a step-function, and $\psi_s$ is a step-function itself. Let us denote by $D_s = \{r_{s1}, r_{s2}, \ldots, r_{sh_s}\}$ ($h_s$ finite) and $D_c = \{r_{c1}, r_{c2}, \ldots, r_{ch_c}\}$ (again $h_c$ finite) the points of jumps of $\psi_s$ and of $\psi'_c$, respectively.

(ii) $\sigma^2 = \operatorname{var} e_1 \in (0, \infty)$ and there is a positive $\vartheta_0$ such that $F(z)$ has a density $f$ which is bounded on

$$D_c(\vartheta_0) = \bigcup_{i=1}^{h_c} [\sigma \cdot r_{ci} - \vartheta_0, \sigma \cdot r_{ci} + \vartheta_0].$$

Moreover, the density $f$ is Lipschitz of the first order on

$$D_s(\vartheta_0) = \bigcup_{i=1}^{h_s} [\sigma \cdot r_{si} - \vartheta_0, \sigma \cdot r_{si} + \vartheta_0].$$

Let us select $H < \infty$ so that it is an upper bound of $f$ on $D_c(\vartheta_0)$ as well as the Lipschitz constant on $D_s(\vartheta_0)$.

(iii) There is a finite $K$ such that

$$\sup_{z \in \{D_s(\vartheta_0) \cup D_c(\vartheta_0)\}} |\psi(z)| < K \quad \text{and} \quad \sup_{z \in R \setminus \{D_s \cup D_c\}} |\psi'(z)| < K.$$

(iv) $\mathsf{E}\,\psi(e_1/\sigma) = 0$ and $\gamma = \sigma^{-1}\,\mathsf{E}\,\psi'(e_1/\sigma)+\theta > 0$ for

$$\theta = \sum_{k=1}^{h_s} f(r_{sk}\,\sigma)\,[\psi(r_{sk}+)-\psi(r_{sk}-)].$$

Remark 3. Conditions B essentially coincide with those of Hampel et al. [12], Section 2.5a, however their form (especially decomposition (5)) follows Jurečková [15]. They were used in [12] to study the change-of-variance function and they cover presumably the most of $\psi$-functions used in the present robust statistics. Some heuristic comments on them may be found also there.

**4. Notation (continued).** In accordance with conditions given above let us enlarge our notation as follows:

$$g'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$$
$$= [g_1'(X_i, \beta^0+n^{-1/2}t+n^{-1/2-\tau}u), g_2'(X_i, \beta^0+n^{-1/2}t+n^{-1/2-\tau}u), \ldots, g_p'(X_i, \beta^0+n^{-1/2}t+n^{-1/2-\tau}u)]^{\mathrm{T}},$$

$$g_j''(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$$
$$= [g_{j1}''(X_i, \beta^0+n^{-1/2}t+n^{-1/2-\tau}u), g_{j2}''(X_i, \beta^0+n^{-1/2}t+n^{-1/2-\tau}u), \ldots, g_{jn}''(X_i, \beta^0+n^{-1/2}t+n^{-1/2-\tau}u)]^{\mathrm{T}},$$

$$\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) = g(X_i, \beta^0+n^{-1/2}t+n^{-1/2-\tau}u)-g(X_i, \beta^0),$$

$$\delta'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) = g'(X_i, \beta^0+n^{-1/2}t+n^{-1/2-\tau}u)-g'(X_i, \beta^0+n^{-1/2}t)$$

(notice that the difference between $\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$ and $\delta'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$ is not only in the sign ′ but also the arguments of the subtracted terms are different). As above we shall write $g'(X_i, n^{-1/2}t)$ instead of $g'(X_i, n^{-1/2}t, 0)$ or $\delta(X_i, n^{-1/2}t)$ instead of $\delta(X_i, n^{-1/2}t, 0)$, etc. Finally, let us put

$$s(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v))$$
$$= \psi([e_i-\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)]\,\sigma^{-1}\exp(-n^{-1/2}v)) \times g'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$$
$$-\psi([e_i-\delta(X_i, n^{-1/2}t)]\,\sigma^{-1}\exp(-n^{-1/2}v))\,g'(X_i, n^{-1/2}t),$$

$$S(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v)) = \sum_{i=1}^{n} s(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v)). \tag{6}$$

Again, as for the derivatives of the function $g$, we denote by

$$s_j(X_i, n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v)) \quad \text{and} \quad S_j(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v))$$

the $j$-th coordinates of the vectors

$$s\big(X_i,\, n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v)\big) \quad \text{and} \quad S\big(n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v)\big),$$

respectively. Finally, for any $M > 0$ let us put

$$\mathcal{T}_M = \{t,\, u\in R^p,\, v\in R^+\colon\ \max\{\|t\|,\, \|u\|,\, v\} \leqslant M\}. \tag{7}$$

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

**5. Preliminaries.** We shall give now three lemmas for deriving the Bahadur representation of

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

LEMMA 1. *Let Conditions* A *be fulfilled and let the $\psi$-function have a derivative $\psi'$ which is Lipschitz of the first order, i.e. $\psi = \psi_a$. Moreover, let* $\operatorname{var} e_1 = \sigma^2 \in (0, \infty)$, $\mathsf{E}\,\psi(e_1/\sigma) = 0$ *and* $|\mathsf{E}\,\psi'(e_1/\sigma)| < \infty$. *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*

$$\max_{1\leqslant i,j\leqslant p} |(\mathcal{U}_n(\tau))_{ij}| = o(1) \text{ a.s.} \quad \text{as } n\to\infty$$

*and we have*

$$\sup_{\mathcal{T}_M} \big\|S\big(n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v)\big) + n^{1/2-\tau}[\sigma^{-1}\,\mathsf{E}\,\psi'(e_1/\sigma)Q + \mathcal{U}_n]u\big\| = O(n^{-\tau}) \text{ a.s.} \quad \text{as } n\to\infty. \tag{8}$$

For the proof of this lemma as well as of the next one see Víšek [39].

LEMMA 2. *Let* $\mathsf{E}\,\psi(e_1/\sigma) = 0$. *Moreover, let Conditions* A *hold and let the function $\psi$ have a derivative $\psi'$ such that for* $-\infty = r_0 < r_1 < \ldots < r_h < \infty$ *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* $x\in(r_h, \infty)$. *Finally, let* $\operatorname{var} e_1 = \sigma^2\in(0, \infty)$ *and let in a $\vartheta_0$-neighborhood of the points* $\sigma r_1, \sigma r_2, \ldots, \sigma r_h$ *the distribution function $F$ have a bounded density $f$ and let us denote its upper bound by $H$. Then for any fixed* $\tau\in[0, \frac{1}{2}]$ *there are sequences of random matrices* $\{\mathcal{U}_n(\tau)\}_{n=1}^{\infty}$ *such that*

$$\max_{1\leqslant i,j\leqslant p} |(\mathcal{U}_n(\tau))_{ij}| = o(1) \text{ a.s.} \quad \text{as } n\to\infty$$

*and we have*

$$\sup_{\mathcal{T}_M} \big\|S\big(n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v)\big) + n^{1/2-\tau}[\sigma^{-1}\,\mathsf{E}\,\psi'(e_1/\sigma)Q + \mathcal{U}_n(\tau)]u\big\| = O_p(n^{-\tau}) \quad \text{as } n\to\infty. \tag{9}$$

We are going to give a similar assertion to those given in Lemmas 1 and 2 for the step $\psi$-function. Unfortunately, the assertion is formally somewhat more complicated than the previous ones because the discontinuity of the function hampers to give a simple approximation of the corresponding sum (compare with the results concerning the asymptotic linearity of the second order $M$-statistics in Rubio and Víšek [31]). As we shall see later, the estimators determined by $\psi$-functions with jumps are really different in character from the estimators determined by the smooth $\psi$-functions.

To prove Lemma 3 we shall need the following assertion:

Assertion 1 (Štepán [35], p. 420, VII.2.8). *Let $a$ and $b$ be positive numbers. Further, let $\xi$ be a random variable such that $P(\xi = -a) = \pi$ and $P(\xi = b) = 1-\pi$ (for a $\pi \in (0, 1)$) and $\mathrm{E}\xi = 0$. Moreover, let $\tau$ be the time for the Wiener process $W(s)$ to exit the interval $(-a, b)$. Then*

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

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

Remark 4. To avoid any misunderstanding let us recall that we assume $W(0) = 0$ (see e.g. Csörgő and Révész [9] or Štepán [35]).

Remark 5. Since the book by Štepán [35] is in Czech, we refer also to Breiman [4] where however 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 [4]. (See also [4], Theorem 13.6, p. 276.)

Lemma 3. *Let Conditions* A *hold and let $\psi$ be a step-function with the steps at the points $r_1, r_2, \ldots, r_h$. Moreover, let* $\operatorname{var} e_1 = \sigma^2 \in (0, \infty)$ *and assume that in a $\vartheta_0$-neighborhood of the points $\sigma r_1, \sigma r_2, \ldots, \sigma r_h$ the distribution function $F$ has a density which is Lipschitz (of the first order). Then for any fix $\tau \in [0, \frac{1}{2}]$ there are the Wiener processes $\mathscr{W}_{j\tau} = W_{j\tau}(s)$, a system of stopping times $\mu_{ijn}(n^{-1/2} t, n^{-1/2-\tau} u, \sigma \exp(n^{-1/2} v))$, a sequence of random variables $\kappa_{jkn}(\tau)$ and a sequence of processes $\mathscr{K}_{jn}(n^{-1/2} t, n^{-1/2-\tau} u, \sigma \exp(n^{-1/2} v))$ with $j, k = 1, 2, \ldots, p$, $s \in R^+$, $n \in N$, $t, u, v \in \mathscr{T}_M$, so that*

$$\max_{1 \le j,k \le p} |\kappa_{jkn}(\tau)| = o_p(1), \tag{10}$$

$$\max_{1 \le j \le p} \sup_{\mathscr{T}_M} \left|\mathscr{K}_{jn}(n^{-1/2} t, n^{-1/2-\tau} u, \sigma \exp(n^{-1/2} v))\right| = O_p(n^{-\tau}), \tag{11}$$

$$\max_{1 \le j \le p} \sup_{\mathscr{T}_M} \|u\|^{-1} \sum_{i=1}^{n} \mu_{ijn}(n^{-1/2} t, n^{-1/2-\tau} u, \sigma \exp(n^{-1/2} v)) = O_p(n^{1/2-\tau}) \quad \text{as } n \to \infty, \tag{12}$$

*and for* $t, u, v \in \mathscr{T}_M$

$$(13)\quad S_j(n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v)) + n^{1/2-\tau}\theta\sum_{k=1}^{p}[q_{jk}+\kappa_{jkn}(\tau)]\,u_k + \mathscr{K}_{jn}(n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v)) =_{\mathscr{D}} W_{j\tau}\Big(\sum_{i=1}^{n}\mu_{ijn}(n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v))\Big),$$

*where* $=_{\mathscr{D}}$ *denotes the equality in distribution.*

Remark 6. The Wiener processes $\mathscr{W}_{j\tau}$ and the stopping times $\mu_{ijn}(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v))$ with $i = 1, 2, \ldots, n$, $n \in N$, $j = 1, 2, \ldots, p$, $\tau \in [0, \frac{1}{2}]$, $t, u, v \in \mathscr{T}_M$, are defined on a probability space $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$ which is generally different from $(\Omega, \mathscr{A}, P)$. So it is clear that $O_p(n^{1/2-\tau})$ in (12) refers to $\tilde{P}$.

The proof of Lemma 3 below seems to be technically a little bit more complicated than the proof of the previous assertions, however its idea is as simple as in the previous cases. Let us also note that in Víšek [39] the proof of a similar assertion to Lemma 3 is given, and it is simpler due to the fact that $g(X_i, \beta) = X_i^T\beta$ and the rescaling of residuals is not considered. So to read that proof may be helpful in understanding the following one. Finally, let us mention that the technique which will be used in the proof is due to Portnoy [25] (see also Jurečková and Sen [16]).

Remark 7. In the proofs of next lemmas we shall need some constants $C_m$, $m = 1, 2, \ldots, \Delta$, $\varepsilon$, etc., definitions of which will be straightforward. The definitions of the constants will hold only within the given proof.

Proof of Lemma 3. Notice that due to the fact that $\psi$ is bounded we may assume that $\mathsf{E}\psi(e_1/\sigma) = 0$. Moreover, without any loss of generality we may assume that $\sigma^2 = 1$ and $h = 1$ (we shall write $r$ instead of $r_1$). Finally, let $\alpha_0 < \alpha_1$, and $\Delta = \max\{|\alpha_0|, |\alpha_1|\}$ and let $n_0$ be the smallest integer such that $M < \delta_0 n_0^{1/2}$ (see A (i) and (7)), and let us assume only that $n \in N$, $n > n_0$. Let us write for $\tau \in [0, \frac{1}{2}]$

$$A_{in}(\tau) = \{\tilde{t}, \tilde{u} \in R^p\colon \delta(X_i, n^{-1/2}\tilde{t}, n^{-1/2-\tau}\tilde{u}) \geqslant \delta(X_i, n^{-1/2}\tilde{t})\}.$$

Then denoting successively by $S^{(\zeta)}(n^{-1/2}t, n^{-1/2-\tau}u, r\exp(n^{-1/2}v))$, $\zeta = 1, 2, 3$, the expressions

$$(14)\quad \sum_{i=1}^{n}\{[\alpha_0 g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - \alpha_1 g_1'(X_i, n^{-1/2}t)] \times \big[I_{\{\delta(X_i,n^{-1/2}t)+r\exp(n^{-1/2}v)<e_i<\delta(X_i,n^{-1/2}t,n^{-1/2-\tau}u)+r\exp(n^{-1/2}v)\}} - F(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v)) + F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v))\big]\, I_{A_{in}(\tau)}(t, u)$$

$$+[\alpha_1 g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - \alpha_0 g_1'(X_i, n^{-1/2}t)]$$
$$\times\big[I_{\{\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)+r\exp(n^{-1/2}v) < e_i < \delta(X_i, n^{-1/2}t)+r\exp(n^{-1/2}v)\}}$$
$$-F\big(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v)\big)$$
$$+F\big(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v)\big)\big]\, I_{A^c_{in}(\tau)}(t, u)\big\},$$

$$(15)\quad \sum_{i=1}^{n} \alpha_1 \delta_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)\big\{\big[I_{\{\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)+r\exp(n^{-1/2}v)\leqslant e_i\}} - 1$$
$$+F\big(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v)\big)\big]\, I_{A_{in}(\tau)}(t, u) + \big[I_{\{\delta(X_i, n^{-1/2}t)+r\exp(n^{-1/2}v)\leqslant e_i\}} - 1$$
$$+F\big(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v)\big)\big]\, I_{A^c_{in}(\tau)}(t, u)\big\}$$

and

$$(16)\quad \sum_{i=1}^{n} \alpha_0 \delta_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)\big\{\big[I_{\{e_i \leqslant \delta(X_i, n^{-1/2}t)+r\exp(n^{-1/2}v)\}}$$
$$-F\big(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v)\big)\big]\, I_{A_{in}(\tau)}(t, u) + \big[I_{\{e_i \leqslant \delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)+r\exp(n^{-1/2}v)\}}$$
$$-F\big(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v)\big)\big]\, I_{A^c_{in}(\tau)}(t, u)\big\},$$

we have (keep in mind please that we have assumed $\sigma = 1$)

$$S_1\big(n^{-1/2}t, n^{-1/2-\tau}u, \exp(n^{-1/2}v)\big) - \mathsf{E}\, S_1\big(n^{-1/2}t, n^{-1/2-\tau}u, \exp(n^{-1/2}v)\big)$$
$$= \sum_{\zeta=1}^{3} S^{(\zeta)}\big(n^{-1/2}t, n^{-1/2-\tau}u, r\exp(n^{-1/2}v)\big).$$

Let us consider at first (again for $\tau \in [0, \frac{1}{2}]$) $S^{(1)}\big(n^{-1/2}t, n^{-1/2-\tau}u, r\exp(n^{-1/2}v)\big)$ and let us put for $i = 1, 2, \ldots, n$

$$c_{in}^{(1)}(\tau) = \min\big\{[\alpha_0 g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - \alpha_1 g_1'(X_i, n^{-1/2}t)]$$
$$\times\big[1 - F\big(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v)\big) + F\big(\delta(X_i, n^{-1/2}t)$$
$$+ r\exp(n^{-1/2}v)\big)\big],$$
$$[\alpha_0 g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - \alpha_1 g_1'(X_i, n^{-1/2}t)]$$
$$\times\big[F\big(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v)\big) - F\big(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$$
$$+ r\exp(n^{-1/2}v)\big)\big]\big\}\, I_{A_{in}(\tau)}(t, u)$$
$$+\min\big\{[\alpha_1 g_1'(X_i, \beta^0 + n^{-1/2}t + n^{-1/2-\tau}u) - \alpha_0 g_1'(X_i, n^{-1/2}t)]$$
$$\times\big[1 - F\big(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v)\big) + F\big(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$$
$$+ r\exp(n^{-1/2}v)\big)\big],$$

$$[\alpha_1 g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - \alpha_0 g_1'(X_i, n^{-1/2}t)]$$
$$\times \left[F\left(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v)\right) - F\left(\delta(X_i, n^{-1/2}t)\right.\right.$$
$$\left.\left. + r\exp(n^{-1/2}v)\right)\right]\} \, I_{A_{in}^c(\tau)}(t, u)$$

and

$$d_{in}^{(1)}(\tau) = \max\{\left[\alpha_0 g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - \alpha_1 g_1'(X_i, n^{-1/2}t)\right]$$
$$\times \left[1 - F\left(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v)\right) + F\left(\delta(X_i, n^{-1/2}t)\right.\right.$$
$$\left.\left. + r\exp(n^{-1/2}v)\right)\right],$$
$$[\alpha_0 g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - \alpha_1 g_1'(X_i, n^{-1/2}t)]$$
$$\times \left[F\left(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v)\right) - F\left(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)\right.\right.$$
$$\left.\left. + r\exp(n^{-1/2}v)\right)\right]\} \, I_{A_{in}(\tau)}(t, u)$$
$$+\max\{[\alpha_1 g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - \alpha_0 g_1'(X_i, n^{-1/2}t)]$$
$$\times \left[1 - F\left(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v)\right) + F\left(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)\right.\right.$$
$$\left.\left. + r\exp(n^{-1/2}v)\right)\right],$$
$$[\alpha_1 g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - \alpha_0 g_1'(X_i, n^{-1/2}t)]$$
$$\times \left[F\left(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v)\right)\right.$$
$$\left. - F\left(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v)\right)\right]\} \, I_{A_{in}^c(\tau)}(t, u)$$

and denote by $\mathscr{W} = (W(s), s \in R)$ a Wiener process. Finally, for $i = 1, 2, \ldots, n$ let us define $\mu_{in}^{(1)}(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v))$ to be the time for $W(s)$ to exit the interval $(c_{in}^{(1)}(\tau), d_{in}^{(1)}(\tau))$.

Making use of Assertion 1, i.e. employing Skorohod's embedding of the Wiener process, we obtain

$$S^{(1)}(n^{-1/2}t, n^{-1/2-\tau}u, r\exp(n^{-1/2}v))$$
$$=_{\mathscr{D}} W\Big(\sum_{i=1}^{n} \mu_{in}^{(1)}(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v))\Big).$$

Starting with some $n_1 \geqslant n_0$ we have for $t, u, v \in \mathscr{T}_M$ and $i = 1, 2, \ldots, n$

$$\max\{|g_1'(X_i, \beta^0 + n^{-1/2}t + n^{-1/2-\tau}u)|, |g_1'(X_i, n^{-1/2}t)|\}$$
$$\times \left| F\left(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v)\right)\right.$$
$$\left. - F\left(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v)\right)\right| \leqslant n^{-1/2-\tau} \cdot J \cdot C_1 \cdot \|u\|,$$

where $C_1$ is a positive constant. Thus for $k = 0, 1$ we denote by $V_k(\tau, u)$ the time for $W(s)$ to exit the interval $(a_k(\tau, u), b_k(\tau, u))$ with

(17) $a_k(\tau, u) = \min\{(-1)^{k+1} n^{-1/2-\tau} \cdot 2\Delta \cdot J \cdot C_1 \cdot \|u\|, (-1)^k \cdot 2\Delta \cdot J \cdot C_1 \cdot \|u\|\}$

and

(18) $b_k(\tau, u) = \max\{(-1)^{k+1} n^{-1/2-\tau} \cdot 2\Delta \cdot J \cdot C_1 \cdot \|u\|, (-1)^k \cdot 2\Delta \cdot J \cdot C_1 \cdot \|u\|\}$.

Consequently, we obtain

$$\mathsf{E}\left[\sup_{\mathcal{T}_M} \sum_{i=1}^{n} \frac{V_0(\tau, u) + V_1(\tau, u)}{\|u\|}\right] \leqslant n^{1/2-\tau} \cdot C_2 \quad \text{for } C_2 = 2\Delta \cdot J \cdot C_1 > 0.$$

So for any fix $\varepsilon > 0$ the Chebyshev inequality for nonnegative random variables gives

$$\tilde{P}\Big(\sup_{\mathcal{T}_M} \|u\|^{-1} \sum_{i=1}^{n} [V_0(\tau, u) + V_1(\tau, u)] > \varepsilon^{-1} n^{1/2-\tau} \cdot C_2\Big) < \varepsilon.$$

Since

$$\mu_{in}^{(1)}(n^{-1/2} t, n^{-1/2-\tau} u, \sigma \exp(n^{-1/2} v)) \leqslant V_0(\tau, u) + V_1(\tau, u) \quad \text{for } i = 1, 2, \ldots, n,$$

we obtain also

$$(19) \quad \tilde{P}\Big(\sup_{\mathcal{T}_M} \|u\|^{-1} \sum_{i=1}^{n} \mu_{in}^{(1)}(n^{-1/2} t, n^{-1/2-\tau} u, \sigma \exp(n^{-1/2} v)) > \varepsilon^{-1} n^{1/2-\tau} \cdot C_2\Big) < \varepsilon.$$

We shall now consider, again for $\tau \in [0, \frac{1}{2}]$, $S^{(2)}(n^{-1/2} t, n^{-1/2-\tau} u, r \exp(n^{-1/2} v))$. Recalling that

$$\delta'(X_i, n^{-1/2} t, n^{-1/2-\tau} u) = g'(X_i, n^{-1/2} t, n^{-1/2-\tau} u) - g'(X_i, n^{-1/2} t)$$

and keeping in mind A (ii), let us write

$$(20) \quad \|\delta'(X_i, n^{-1/2} t, n^{-1/2-\tau} u)\|^2 = n^{-1/2-\tau} \sum_{k=1}^{p} \Big[\sum_{j=1}^{p} g''_{jk}(X_i, \tilde{\beta}^{(j)}) u_j\Big]^2$$

$$\leqslant n^{-1/2-\tau} \sum_{k=1}^{p} \Big\{\sum_{j=1}^{p} [g''_{jk}(X_i, \tilde{\beta}^{(j)})]^2\Big\}^{1/2} \|u\| \leqslant n^{-1/2-\tau} \cdot p^{3/2} \cdot J \cdot M,$$

where $\tilde{\beta}^{(j)}$ are appropriate points from the neighborhood of $\beta^0$ and the following inequality holds:

$$\max_{1 \leqslant j \leqslant p} \|\tilde{\beta}^{(j)} - \beta^0\| \leqslant n^{-1/2} M.$$

Let us put similarly as above

$$c_{in}^{(2)}(\tau) = \min\{\alpha_1 \delta'(X_i, n^{-1/2} t, n^{-1/2-\tau} u) \times [F(\delta(X_i, n^{-1/2} t, n^{-1/2-\tau} u) + r \exp(n^{-1/2} v)) - 1],$$
$$\alpha_1 \delta'(X_i, n^{-1/2} t, n^{-1/2-\tau} u) F(\delta(X_i, n^{-1/2} t, n^{-1/2-\tau} u) + r \exp(n^{-1/2} v))\} I_{A_{in}(\tau)}(t, u),$$

$$+\min\{\alpha_1 \delta'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)[F(\delta(X_i, n^{-1/2}t)+r\exp(n^{-1/2}v))-1],$$
$$\alpha_1 \delta'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)F(\delta(X_i, n^{-1/2}t)+r\exp(n^{-1/2}v))\}I_{A^c_{in}(\tau)}(t, u)$$

and

$$d^{(2)}_{in}(\tau) = \max\{\alpha_1 \delta'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$$
$$\times[F(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)+ r\exp(n^{-1/2}v))-1],$$
$$\alpha_1 \delta'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)F(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$$
$$+r\exp(n^{-1/2}v))\}I_{A_{in}(\tau)}(t, u)$$
$$+\max\{\alpha_1 \delta'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)[F(\delta(X_i, n^{-1/2}t)+r\exp(n^{-1/2}v))-1],$$
$$\alpha_1 \delta'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)F(\delta(X_i, n^{-1/2}t)+r\exp(n^{-1/2}v))\}I_{A^c_{in}(\tau)}(t, u).$$

Repeating the steps from the previous part of the proof and making use of (20) we obtain

(21) $$S^{(2)}(n^{-1/2}t, n^{-1/2-\tau}u, r\exp(n^{-1/2}v)) =_{\mathscr{D}} W\Big(\sum_{i=1}^{n} \mu^{(2)}_{in}(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v))\Big)$$

with

(22) $$\tilde{P}\Big(\sup_{\mathscr{T}_M}\|u\|^{-1}\sum_{i=1}^{n} \mu^{(2)}_{in}(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v)) > \varepsilon^{-1}n^{1/2-\tau}\cdot C_3\Big) < \varepsilon,$$

where $C_3$ is an appropriate constant, $\mu^{(2)}_{in}(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v))$, $i = 1, 2, \ldots, n$, are corresponding stopping times, and $\varepsilon$ any positive number. Further, modifying slightly a few previous lines we obtain also

(23) $$S^{(3)}(n^{-1/2}t, n^{-1/2-\tau}u, r\exp(n^{-1/2}v)) =_{\mathscr{D}} W\Big(\sum_{i=1}^{n} \mu^{(3)}_{in}(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v))\Big)$$

with

(24) $$\tilde{P}\Big(\sup_{\mathscr{T}_M}\|u\|^{-1}\sum_{i=1}^{n} \mu^{(3)}_{in}(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v)) > \varepsilon^{-1}n^{1/2-\tau}\cdot C_4\Big) < \varepsilon$$

and appropriately defined $c^{(3)}_{in}(\tau)$, $d^{(3)}_{in}(\tau)$, $\mu^{(3)}_{in}(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v))$ and $C_4$. Putting now

$$\mu_{i1n}(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v)) = \sum_{\zeta=1}^{3} \mu^{(\zeta)}_{in}(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v)),$$

from (19), (22) and (24) we get

$$\sup_{\mathcal{T}_M} \|u\|^{-1} \sum_{i=1}^{n} \mu_{i1n}(n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v)) = O_p(n^{1/2-\tau}) \quad \text{as } n\to\infty$$

and

$$S_1(n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v)) - \mathsf{E}\,S_1(n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v))$$
$$=_{\mathscr{D}} W\Big(\sum_{i=1}^{n} \mu_{i1n}(n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v))\Big).$$

For $j=2, 3, \ldots, p$ the proof can be carried out along the same lines. To conclude the assertion of the lemma we need to carry out an approximation to the mean values

$$\mathsf{E}\,S_j(n^{-1/2}t,\, n^{-1/2-\tau}u,\, \sigma\exp(n^{-1/2}v)) = \sum_{\zeta=1}^{3} \pi^{(\zeta)}_{(n,\tau)j}(t, u, v), \quad j=1, 2, \ldots, p,$$

with $\pi^{(\zeta)}_{(n,\tau)j}(t, u, v)$, $\zeta = 1, 2, 3$, given successively by the following expressions:

$$(25)\quad \sum_{i=1}^{n} \{[\alpha_0 g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - \alpha_1 g_1'(X_i, n^{-1/2}t)]$$
$$\times [F(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v))$$
$$- F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v))]\, I_{A_{in}(\tau)}(t, u)$$
$$+ [\alpha_1 g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - \alpha_0 g_1'(X_i, n^{-1/2}t)]$$
$$\times [F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v))$$
$$- F(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v))]\, I_{A^c_{in}(\tau)}(t, u)\},$$

$$(26)\quad \sum_{i=1}^{n} \{\alpha_1 \delta_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$$
$$\times [1 - F(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v))\, I_{A_{in}(\tau)}(t, u)$$
$$- F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v))\, I_{A^c_{in}(\tau)}(t, u)]\}$$

and

$$(27)\quad \sum_{i=1}^{n} \{\alpha_0 \delta_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)$$
$$\times [F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v))\, I_{A_{in}(\tau)}(t, u)$$
$$+ F(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v))\, I_{A^c_{in}(\tau)}(t, u)]\}$$

(see (14)–(16)). One can verify that $\pi^{(2)}_{(n,\tau)1}(t, u, v) + \pi^{(3)}_{(n,\tau)1}(t, u, v)$ may be written as

$$(28)\quad \sum_{i=1}^{n} \delta'(X_i, n^{-1/2}t, n^{-1/2-\tau}u)\{\alpha_1[F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v))$$
$$-F(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v))]\, I_{A_{in}(\tau)}(t, u)$$
$$+\alpha_1[1 - F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v))]$$
$$+\alpha_0 F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v))$$
$$+\alpha_0 F[(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v))$$
$$-F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v))]\, I_{A^c_{in}(\tau)}(t, u)\}.$$

Recalling that we have assumed that $\mathsf{E}\,\psi(e_i) = 0$, we obtain

$$(29)\quad \mathsf{E}\,\psi(e_i) = \alpha_0[1 - F(r)] + \alpha_1 F(r) = 0.$$

Taking into account (20) and the fact that the density $f$ is Lipschitz of the first order in $\vartheta_0$-neighborhood of the point $r$, we may find constants $C_5$ and $C_6$ such that

$$(30)\quad |F(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v))$$
$$-F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v))| \leqslant n^{-1/2-\tau}\cdot C_5\cdot M,$$

and

$$|F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v)) - F(r)| \leqslant n^{-1/2}\cdot C_6\cdot M,$$

which together with (29) implies that the absolute value of the expression in (28) (and hence also $|\pi^{(2)}_{(n,\tau)1}(t, u, v) + \pi^{(3)}_{(n,\tau)1}(t, u, v)|$) is for any $t, u, v \in \mathcal{T}_M$ bounded by $n^{-\tau}\cdot C_7 < \infty$. Secondly, we have

$$|g_1'(X_i, n^{-1/2}t, n^{-1/2-\tau}u) - g_1'(X_i, n^{-1/2}t)| \leqslant n^{-1/2-\tau}\cdot J\cdot p^{1/2}\cdot \|u\|$$

together with

$$|g_1'(X_i, n^{-1/2}t) - g_1'(X_i, \beta^0)| \leqslant n^{-1/2}\cdot J\cdot p^{1/2}\cdot \|t\|,$$

and taking into account (30) once again we obtain

$$\sup_{\mathcal{T}_M}\Big|\mathsf{E}\,S_1(n^{-1/2}t, n^{-1/2-\tau}u, \sigma\exp(n^{-1/2}v))$$
$$-(\alpha_1 - \alpha_0)\sum_{i=1}^{n} g_1'(X_i, \beta^0)\,|F(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2}v))$$
$$-F(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v))|\Big| = O_p(n^{-\tau}).$$

Putting $z_n^{(k)} = (0, \ldots, 0, z, 0, \ldots, 0)$ (where $z$ is the $k$-th coordinate of $z_n^{(k)}$), we may write

$$(31) \quad F\big(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}u) + r\exp(n^{-1/2}v)\big) - F\big(\delta(X_i, n^{-1/2}t) + r\exp(n^{-1/2-\tau}v)\big) - n^{-1/2-\tau}f(r)[g'(X_i, \beta^0)]^T u$$

$$= n^{-1/2-\tau}\sum_{k=1}^{p}\int_0^{u_j}\big[f\big(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}z_n^{(k)}) + r\exp(n^{-1/2}v)\big)g_j'(X_i, \beta^0 + n^{-1/2}t + n^{-1/2-\tau}z_n^{(k)}) - f(r)g_j'(X_i, \beta^0)\big]\,dz.$$

Making use of the assumption that $f(z)$ in the $\vartheta_0$-neighborhood of $r$ is Lipschitz of the first order as well as of the fact that $g''(x, \beta)$ in the neighborhood of $\beta^0$ exists and is bounded (uniformly in $x \in X_1, X_2, \ldots, X_{(n)}$, we conclude that

$$\big|f\big(\delta(X_i, n^{-1/2}t, n^{-1/2-\tau}z_n^{(k)}) + r\exp(n^{-1/2}v)\big) - f(r)\big| \leq n^{-1/2}\cdot C_8$$

as well as

$$\|g'(X_i, n^{-1/2}t, n^{-1/2-\tau}z_n^{(k)}) - g'(X_i, \beta^0)\| \leq n^{-1/2}\cdot C_9$$

for some constants $C_8$ and $C_9$. Hence (31) is (starting with some $n_3$) bounded in an absolute value by $n^{-1-\tau}\cdot C_{10}$, and hence

$$\sup_{\mathcal{T}_M}\Big\|\mathbf{E}\,S_1\big(n^{-1/2}t, n^{-1/2-\tau}u, \sigma(\exp(n^{-1/2}v))\big) + n^{-1/2-\tau}(\alpha_1 - \alpha_0)f(r)\sum_{i=1}^{n} g_1'(X_i, \beta^0)[g'(X_i, \beta^0)]^T u\Big\| < n^{-\tau}\cdot C_{10}$$

for $n = n_3$

uniformly in $X_i \in S$, $i = 1, 2, \ldots, n$. This completes the proof of the lemma. ∎

**6. Bahadur's representation.** In this section we will give Bahadur's representation of $n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,l)})$. The plan how to do this is simple. At first, using Lemmas 1, 2 and 3 for $\tau = 0$ we shall prove that $n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,l)}) = O_p(1)$, and then using the same lemmas for $\tau = \frac{1}{2}$ we derive the representation.

Remark 8. It is clear that under Conditions A and B, in the case when $\psi_s \equiv 0$, the estimators $\hat{\beta}^{(n)}$ and $\hat{\beta}^{(n-1,l)}$ fulfil the following relations:

$$(32) \quad \sum_{i=1}^{n} \psi\big([Y_i - g(X_i, \hat{\beta}^{(n)})]\,\hat{\sigma}_n^{-1}\big)g'(X_i, \hat{\beta}^{(n)}) = 0$$

and

$$(33) \quad \sum_{\substack{i=1\\ i\neq l}}^{n} \psi\big([Y_i - g(X_i, \hat{\beta}^{(n-1,l)})]\,\hat{\sigma}_n^{-1}\big)g'(X_i, \hat{\beta}^{(n-1,l)}) = 0,$$

respectively (where $\hat{\sigma}_n$ is a preliminary estimator of the scale of residuals; see (3) and (4), and also Conditions C below). Sometimes the $M$-estimators (for the linear model) are even defined as solutions of the equations (32) and (33); see e.g. Jurečková and Welsh [17] or Rao and Zhao [27].

Generally this is not possible for nonsmooth $\varrho$-functions, the derivative of which is discontinuous. Then (32) and (33) need not necessarily hold for $\hat{\beta}^{(n)}$ and $\hat{\beta}^{(n-1,l)}$ given by (3) and (4). As we shall see later to be able to apply Lemmas 1, 2 and 3 (in order to derive Bahadur's representation) the relations (32) and (33) are to be fulfilled at least approximately. In Víšek [39] one may find a rather large discussion of such a requirement. It is shown that sometimes even for a discontinuous function $\psi$ (when it is simple or even symmetric) we may reach equality in (32) and (33). The discussion also hints that we may hope that for frequently used $\psi$-functions and for the case when $g(x, \beta)$ is not too "wild," we may recognize wheather (32) and (33) are approximately (in the sense of (34) and (35)) fulfilled.

That is why we shall assume (and specify this in the following Conditions C) that the left-hand sides of the equations (32) and (33) are small in probability even for the discontinuous $\psi$-functions.

CONDITIONS C. (i) The estimators $\hat{\beta}^{(n)}$ and $\hat{\beta}^{(n-1,l)}$ given by (3) and (4) are $\sqrt{n}$-consistent in the following sense:

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

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

and fulfil the relations

$$\sum_{i=1}^{n} \psi([Y_i - g(X_i, \hat{\beta}^{(n)})]\,\hat{\sigma}_n^{-1})\,g'(X_i, \hat{\beta}^{(n)}) = o_p(1) \tag{34}$$

and

$$\sum_{i=1}^{n} \psi([Y_i - g(X_i, \hat{\beta}^{(n-1,l)})]\,\hat{\sigma}_n^{-1})\,g'(X_i, \hat{\beta}^{(n-1,l)}) = o_p(1). \tag{35}$$

(ii) There is an $\sqrt{n}$-consistent estimator $\hat{\sigma}_n$ of $\sigma$, i.e.

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

which is affine invariant, i.e. for any $b \in R^p$

$$\hat{\sigma}_n(Y + X \cdot b) = \hat{\sigma}_n(Y)$$

and scale equivariant, i.e. for any $c > 0$

$$\hat{\sigma}_n(cY) = c \cdot \hat{\sigma}_n.$$

Remark 9. It is obviously only a technical matter to show how the result of Liese and Vajda [20], concerning the consistency of the estimator $\hat{\beta}^{(n)}$ in a nonstudentized framework, can be generalized for the studentized version. Further, in Rubio et al. [29] it is shown that under conditions given here the $\sqrt{n}$-consistency of $\hat{\beta}^{(n)}$ follows from their consistency. Moreover, also in Rubio and Víšek [31] it is proved that under conditions given here there is for the case $\psi_s \equiv 0$ an $\sqrt{n}$-consistent solution of the equation (32). Also the result of Rao and Zhao [27] seems to be in a straightforward way generalizable for a nonlinear setup (their result applies also for $\psi$-functions with jumps but on the other hand the $\psi$-function has to be monotone). So, it seems that there may appear very diverse conditions for the consistency of the $M$-estimators for general $\psi$-functions, and hence we have preferred to give Conditions C in the present form.

Before proving the $n$-consistency of $(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,l)})$, let us recall one result from Csörgő and Révész [9].

Assertion 2. *Lévy's law of iterated logarithm:*

$$\limsup_{s\to\infty} \sup_{0\le t\le s} \frac{|W(t)|}{\sqrt{2s\log\log s}} = 1 \ \text{a.s.} \quad \text{as } n \to \infty,$$

*where $W(t)$ is the Wiener process* (see Assertion 1).

For the proof see Lévy [19] or Csörgő and Révész [9], Theorem 1.3.1. We will need also the following

Lemma 4. *Let for some $p \in N$, $\{\mathscr{V}^{(n)}\}_{n=1}^{\infty}$, $\mathscr{V}^{(n)} = \{v_{ij}^{(n)}\}_{i=1,2,\ldots,p}^{j=1,2,\ldots,p}$ be a sequence of $(p\times p)$ random matrices such that for $i, j = 1, 2, \ldots, p$*

$$\lim_{n\to\infty} v_{ij}^{(n)} = q_{ij} \ \textit{in probability}, \tag{36}$$

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

$$\exists(\varepsilon > 0)\ \forall(K > 0)\ \limsup_{n\to\infty} P(\|\theta^{(n)}\| > K) > \varepsilon. \tag{37}$$

*Then*

$$\exists(\delta > 0)\ \forall(L > 0)\ \limsup_{n\to\infty} P(\|\mathscr{V}^{(n)}\theta^{(n)}\| > L) > \delta.$$

Proof. Due to (36) the matrix $\mathscr{V}^{(n)}$ is regular in probability. Let then $0 < \lambda_{1n} < \lambda_{2n} < \ldots < \lambda_{pn}$ and $z_{1n}, z_{2n}, \ldots, z_{pn}$ be eigenvalues and corresponding eigenvectors (selected to be mutually orthogonal) of the matrix $[\mathscr{V}^{(n)}]^T \mathscr{V}^{(n)}$. Let us write

$$\theta^{(n)} = \sum_{j=1}^{p} a_{jn} z_{jn}$$

(for an appropriate vector $a_n = (a_{1n}, a_{2n}, \ldots, a_{pn})^T$). Then we have

(38) $$\|\mathscr{V}^{(n)} \theta^{(n)}\|^2 = \sum_{j=1}^{p} [a_{jn}]^2 \lambda_{jn}^2 \|z_{jn}\|^2 \geqslant \lambda_{1n}^2 \|\theta^{(n)}\|^2.$$

Moreover, denoting by $\lambda_1$ the smallest eigenvalue of the matrix $Q^T Q$, we have $\lambda_{1n} \to \lambda_1$ in probability as $n \to \infty$. The assertion of the lemma then follows from (38). ∎

LEMMA 5. *Let Conditions* A, B *and* C *be satisfied. Then*

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

Proof. First of all, let us recall that the Wiener processes as well as the corresponding stopping times are defined on a probability space $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{P})$. Let us also recall that for any $z \in R$ we may write

$$\psi(z) = \psi_a(z) + \psi_c(z) + \psi_s(z).$$

Now, let us fix $\Delta > 0$ and $\varepsilon > 0$, and making use of (12) for $\tau = 0$ let us find $C_1 < \infty$ and $n_1 \in N$ so that for the stopping times from Lemma 3 we have for any $n > n_1$

(39) $$\tilde{P}(\{\max_{1 \leqslant j \leqslant p} \sup_{\mathscr{T}_M} \|u\|^{-1} \sum_{i=1}^{n} \mu_{ijn}(n^{-1/2} t, n^{-1/2-\tau} u, \sigma \exp(n^{-1/2} v)) > n^{1/2} C_1\}) \leqslant \varepsilon.$$

Further, by the strong law of large numbers we have

(40) $$\frac{|W_j(s)|}{s} \to 0 \text{ a.s.} \quad \text{for } s \to \infty$$

(see also Assertion 2), where $W_j(s)$ are the Wiener processes introduced in Lemma 3 (for $\tau = 0$), and hence also

$$\sup_{s>t} \frac{|W_j(s)|}{s} \to 0 \text{ a.s.} \quad \text{for } t \to \infty.$$

Since the stochastic convergence follows from the a.s. one, let us use (40) and establish $C_2 < \infty$ so that we have

(41) $$\tilde{P}\left(\max_{1 \leqslant j \leqslant p} \sup_{s > C_2} \frac{|W_j(s)|}{s} > \Delta \cdot C_1^{-1}\right) < \varepsilon.$$

Now applying Assertion 2 let us select $C_3 < \infty$ so that

(42) $$\tilde{P}(\max_{1 \leqslant j \leqslant p} \sup_{0 \leqslant s \leqslant C_2} |W_j(s)| > C_3) < \varepsilon.$$

Now, using (8)–(13) again for $\tau = 0$ let us find sequences of random variables $\{\kappa_{jkn}^{(r)}(0)\}_{n=1}^{\infty}$ and of random processes $\{\mathscr{K}_{jn}(n^{-1/2} t, n^{-1/2} u, \sigma \exp(n^{-1/2} v))\}_{n=1}^{\infty}$

with $r = 1, 2, 3, j, k = 1, 2, \ldots, p$, a constant $C_4$, and $n_2 > n_1$ so that for any $n > n_2$ we have

$$P\big(\max_{1\leqslant j,k\leqslant p} |\kappa^{(r)}_{jkn}(0)| > \Delta\big) < \varepsilon, \tag{43}$$

$$P\big(\max_{1\leqslant j\leqslant p} \sup_{\mathscr{T}_M} |\mathscr{K}_{jkn}(n^{-1/2} t, n^{-1/2} u, \sigma \exp(n^{-1/2} v))| > C_4\big) < \varepsilon, \tag{44}$$

$$P\big(\max_{1\leqslant j\leqslant p} \sup_{\mathscr{T}_M} |S^{\psi_a}_{(n,0)j}(t, u, v) + n^{1/2} [\sum_{k=1}^{p} (\sigma^{-1} \mathsf{E} \psi'_a(e_1/\sigma) q_{jk} + \kappa^{(1)}_{jkn}(0)) u_k]| > C_4\big) < \varepsilon, \tag{45}$$

and

$$P\big(\max_{1\leqslant j\leqslant p} \sup_{\mathscr{T}_M} |S^{\psi_c}_{(n,0)j}(t, u, v) + n^{1/2} [\sum_{k=1}^{p} (\sigma^{-1} \mathsf{E} \psi'_c(e_1/\sigma) q_{jk} + \kappa^{(2)}_{jkn}(0)) u_k]| > C_4\big) < \varepsilon, \tag{46}$$

where we have denoted the corresponding sums for $\psi_a$ and $\psi_c$ (see (6)) by $S^{\psi_a}_{(n,0)}(t, u, v)$ and $S^{\psi_c}_{(n,0)}(t, u, v)$, respectively. Let us denote similarly the corresponding sum for $\psi_s$ by $S^{\psi_s}_{(n,0)}(t, u, v)$ and let us put

$$B_{1n} = \{\tilde{\omega}: \max_{1\leqslant j\leqslant p} \sup_{\mathscr{T}_M} \|u\|^{-1} \sum_{i=1}^{n} \mu_{ijn}(n^{-1/2} t, n^{-1/2} v, \sigma \exp(n^{-1/2} v)) > n^{1/2} \cdot C_1\}, \tag{47}$$

$$B_{2n} = \left\{\tilde{\omega}: \max_{1\leqslant j\leqslant p} \frac{|W_j(s)|}{s} > \Delta \cdot C_1^{-1}\right\} \tag{48}$$

and

$$B_{3n} = \left\{\tilde{\omega}: \max_{1\leqslant j\leqslant p} \sup_{0\leqslant s\leqslant C_2} |W_j(s)| > C_3\right\}. \tag{49}$$

Putting $D_n = B^c_{1n} \cap B^c_{2n} \cap B^c_{3n}$ we have

$$\tilde{P}(D_n) \geqslant 1 - 3\varepsilon.$$

For any $\tilde{\omega} \in D_n$, any $j = 1, 2, \ldots, p$ and for any $t, u, v \in \mathscr{T}_M$ we have either

$$\sum_{i=1}^{n} \mu_{ijn}(n^{-1/2} t, n^{-1/2} u, \sigma \exp(n^{-1/2} v)) > C_2 \tag{50}$$

or

$$\sum_{i=1}^{n} \mu_{ijn}(n^{-1/2} t, n^{-1/2} u, \sigma \exp(n^{-1/2} v)) \leqslant C_2. \tag{51}$$

Let us assume at first that (50) holds. Then, due to (47) and (48) we obtain

$$|W_j(\sum_{i=1}^{n} \mu_{ijn}(n^{-1/2} t, n^{-1/2} u, \sigma \exp(n^{-1/2} v)))| < \Delta \cdot C_1^{-1} \cdot \sum_{i=1}^{n} \mu_{ijn}(n^{-1/2} t, n^{-1/2} u, \sigma \exp(n^{-1/2} v)) < n^{1/2} \cdot \Delta \|u\|.$$

For the case when $\sum_{i=1}^{n} \mu_{ijn}(n^{-1/2}t, n^{-1/2}u, \sigma\exp(n^{-1/2}v)) \leqslant C_2$ we have (see (49))

$$\Big|W_j\Big(\sum_{i=1}^{n} \mu_{ijn}(n^{-1/2}t, n^{-1/2}u, \sigma\exp(n^{-1/2}v))\Big)\Big| < C_3.$$

This means that for any $\tilde{\omega} \in D_n$, any $j = 1, 2, \ldots, p$ and for any $t, u, v \in \mathscr{T}_M$ we have

$$\Big|W_j\Big(\sum_{i=1}^{n} \mu_{ijn}(n^{-1/2}t, n^{-1/2}u, \sigma\exp(n^{-1/2}v))\Big)\Big| < n^{1/2}\cdot\Delta\,\|u\| + C_3,$$

i.e. for any $\tilde{\omega} \in D_n$, any $j = 1, 2, \ldots, p$ and any $t, u, v \in \mathscr{T}_M$ we have

$$W_j\Big(\sum_{i=1}^{n} \mu_{ijn}(n^{-1/2}t, n^{-1/2}u, \sigma\exp(n^{-1/2}v))\Big) - n^{1/2}\cdot\Delta\,\|u\| < C_3$$

and

$$W_j\Big(\sum_{i=1}^{n} \mu_{ijn}(n^{-1/2}t, n^{-1/2}u, \sigma\exp(n^{-1/2}v))\Big) + n^{1/2}\cdot\Delta\,\|u\| > -C_3.$$

Thus defining for $j = 1, 2, \ldots, p$

$$\Delta_{jkn}^{*}(t, u, v) = \Delta\cdot u_k\cdot\operatorname{sign}\big(S_{(n,0)j}^{\psi_s}(t, u, v) + n^{1/2}\sum_{l=1}^{p}[\theta q_{jl} + \kappa_{jln}^{(3)}]\,u_l + \mathscr{K}_{jn}(n^{-1/2}t, n^{-1/2}u, \sigma\exp(n^{-1/2}v)\big)$$

and taking into account that $\Delta$ was an arbitrary positive number we have

$$(52)\quad P\Big(\max_{1\leqslant j\leqslant p}\sup_{\mathscr{T}_M}\Big|S_{(n,0)j}^{\psi_s}(n^{-1/2}t, n^{-1/2}u, \sigma\exp(n^{-1/2}v)) + n^{1/2}\sum_{k=1}^{p}[\theta q_{jk} + \kappa_{njk}^{(3)} - \Delta_{jkn}^{*}(t, u, v)]\,u_k + \mathscr{K}_{jn}(n^{-1/2}t, n^{-1/2}u, \sigma\exp(n^{-1/2}v))\Big| > C_3\Big) < 3\varepsilon.$$

Finally, using (39), (41)–(43), (45), (46) and (52), we may find for any $\Delta > 0$ such an $n_\Delta \in N$ that for any $n \in N$, $n > n_\Delta$,

$$(53)\quad P\Big(\sup_{\mathscr{T}_M}\Big\|S(n^{-1/2}t, n^{-1/2}u, \sigma\exp(n^{-1/2}v)) + n^{1/2}\gamma\Big[Q - \Delta_n^{*}(t, u, v) + \sum_{r=1}^{3}\tilde{\kappa}_{jn}^{(r)}(0)\Big]u\Big\| > 3C_4 + C_3\Big) < 6\varepsilon,$$

where $\tilde{\kappa}_{jn}^{(r)}(0) = (\kappa_{j1n}^{(r)}(0), \kappa_{j2n}^{(r)}(0), \ldots, \kappa_{jpn}^{(r)}(0))^{\mathrm{T}}$. Let us recall that according to Conditions C we have

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

and let us put

$$\tilde{t}_n = \sqrt{n}\,(\hat{\beta}^{(n)} - \beta^{0}),\qquad \tilde{u}_n = \sqrt{n}\,(\hat{\beta}^{(n-1,l)} - \hat{\beta}^{(n)}),\qquad \tilde{v}_n = \sqrt{n}\,(\log\hat{\sigma}_n - \log\sigma).$$

Then there is a constant $C_5 > 0$ so that starting with some $n_\varepsilon$ we have

$$P(\max\{\|\tilde{t}\|, \|\tilde{u}\|, |\tilde{v}|\} < C_5) > 1-\varepsilon.$$

Putting $\tilde{t}_n$, $\tilde{u}_n$ and $\tilde{v}_n$ into (53) we obtain

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

Finally, taking into account (34) and (35) we get

$$[\sigma_n^{-1}\gamma Q + o_p(1)]\,n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,l)})$$
$$= \psi([Y_l - g(X_l, \hat{\beta}^{(n-1,l)})]\,\sigma_n^{-1})\,g'(X_l, \hat{\beta}^{(n-1,l)}) + O_p(1) \quad \text{as } n \to \infty.$$

The application of Lemma 4 completes the proof. ∎

THEOREM 1. *Assume that Conditions* A, B *and* C *hold and* $\psi_s \equiv 0$. *Then uniformly in* $l \in N$ *we have*

$$(54) \quad n(\hat{\beta}^{(n-1,l)} - \hat{\beta}^{(n)}) = -\hat{\sigma}_n \mathsf{E}^{-1}\psi'(e_1/\sigma)\,Q^{-1} g'(X_l, \hat{\beta}^{(n)})$$
$$\times \psi([Y_l - g(X_l, \hat{\beta}^{(n)})]\,\hat{\sigma}_n^{-1}) + o_p(1) \quad \textit{as } n \to \infty.$$

Proof. Let at first $\psi_s \equiv 0$. Considering $\tau = \frac{1}{2}$ and taking into account Lemma 5 we may put

$$(55) \quad \hat{t}_{n-1} = \sqrt{n-1}\,(\hat{\beta}^{(n)} - \beta^0), \qquad \hat{u}_{n-1} = (n-1)(\hat{\beta}^{(n-1,l)} - \hat{\beta}^{(n)}),$$
$$\hat{v}_n = \sqrt{n}\,(\log\hat{\sigma}_n - \log\sigma)$$

into (8) and (9) and we obtain

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

Then using Lemma 5 once again, and employing (34) and (35) we have

$$\psi([Y_l - g(X_l, \hat{\beta}^{(n)})]\,\sigma_n^{-1})\,g'(X_l, \hat{\beta}^{(n)})$$
$$+\sigma^{-1}\mathsf{E}\psi'(e_1/\sigma)\,Q \cdot n(\hat{\beta}^{(n-1,l)} - \hat{\beta}^{(n)}) = o_p(1) \quad \text{as } n \to \infty.$$

Finally, taking into account the regularity of the matrix $Q$, we conclude the assertion of the theorem.

LEMMA 6. *Let $\hat{t}_{n-1}$, $\hat{u}_{n-1}$, and $\hat{v}_n$ be defined as in* (55) *and let us put*

$$(U_n)_j = S_j(\tfrac{1}{2}, \hat{t}_{n-1}, \hat{u}_{n-1}, \hat{v}_n)+\theta \sum_{k=1}^{p} q_{jk}(\hat{u}_{n-1})_k.$$

*Then*

$$(U_n)_j =_{\mathscr{D}} W_j\Big(\sum_{i=1}^{n} \mu_{in}(\hat{t}_{n-1}, \hat{u}_{n-1}, \hat{v}_n)\Big) \tag{56}$$

*and*

$$U_n = O_p(1). \tag{57}$$

Proof. First of all, let us observe that for $\tau = \frac{1}{2}$ we have

$$\max_{1\leqslant j\leqslant p} \Big|\sum_{k=1}^{p} \kappa_{jkn}(\tfrac{1}{2}\hat{u}_{n-1,k})\Big| = o_p(1)$$

(where $\hat{u}_{n-1,k}$ is the $k$-th coordinate of $\hat{u}_{n-1}$) and

$$\max_{1\leqslant j,k\leqslant p} \sup_{\mathscr{T}_M} \big|\mathscr{K}_{jn}(n^{-1/2}t, n^{-1}u, \sigma\exp(n^{-1/2}v))\big| = o_p(1).$$

Now, using the main assertion of Lemma 3 for $\tau = \frac{1}{2}$, we obtain (56). Applying (12) for $\tau = \frac{1}{2}$, we obtain

$$\max_{1\leqslant j\leqslant p} \sup_{\mathscr{T}_M} \|u\|^{-1} \sum_{i=1}^{n} \mu_{ijn}(n^{-1/2}t, n^{-1}u, \sigma\exp(n^{-1/2}v)) = O_p(n^{-1}) \quad \text{as } n\to\infty,$$

and hence also

$$\max_{1\leqslant j\leqslant p} \sup_{\mathscr{T}_M} \sum_{i=1}^{n} \mu_{ijn}(n^{-1/2}t, n^{-1}u, \sigma\exp(n^{-1/2}v)) = O_p(n^{-1}) \quad \text{as } n\to\infty.$$

Finally,

$$\max_{1\leqslant j\leqslant p} \sum_{i=1}^{n} \mu_{ijn}(n^{-1/2}\hat{t}_{n-1}, n^{-1}\hat{u}_{n-1}, \sigma\exp(n^{-1/2}\hat{v}_n)) = O_p(n^{-1}) \quad \text{as } n\to\infty.$$

Since the supremum of the Wiener process over a bounded interval is bounded in probability, we get (57). ∎

THEOREM 2. *Let Conditions* A, B *and* C *be satisfied and* $\psi_s \not\equiv 0$. *Then uniformly in* $l\in N$ *we have*

$$n(\beta^{(n-1,l)}-\beta^{(n)}) = \gamma^{-1}Q^{-1}\{g'(X_l, \beta^{(n)})\psi([Y_l-g(X_l, \beta^{(n)})]\hat{\sigma}_n^{-1})+U_n\} + o_p(1) \quad \text{as } n\to\infty, \tag{58}$$

*where $U_n$ was defined in Lemma* 6.

Proof. Using (34), (35), and Lemmas 3 and 5 we obtain

$$\psi([Y_l - g(X_l, \hat{\beta}^{(n)})]\,\sigma_n^{-1})\,g'(X_l, \hat{\beta}^{(n)}) + \gamma Q \cdot n(\hat{\beta}^{(n-1,l)} - \hat{\beta}^{(n)}) = U_n + o_p(1) \quad \text{as } n \to \infty,$$

and the proof of the theorem is complete. ∎

Remark 10. We have already known from Lemma 5 that the normed difference of the estimators $n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,l)})$ is $O_p(1)$. Theorems 1 and 2 specify this information for the cases when $\psi_s \equiv 0$ and $\psi_s \not\equiv 0$. For the former case we may give, if the function $\psi$ is bounded, with a large probability a nonrandom upper bound for this difference, so that we have an idea about stability of the estimation when adding or excluding one observation. Since the most of the functions which are used in the robust considerations are bounded, this information is useful from the computational point of view. (And the numerical experience says that the approximation works for a rather small number of observations, usually about twenty; see Víšek [36].) On the other hand, for the latter case we of course also know that the difference is bounded in probability but the upper bound may be pretty large and the numerical example given in Víšek [39] confirms much larger "fluctuation" of the $L_1$-estimator, i.e. of the estimator generated by a $\psi$-function which contains jump, in comparison with the estimators with smooth $\psi$-functions.

In other words, for the former case if the "tuning" constant of the corresponding $\psi$-function is properly assigned to "winsorize" really some residuals, the $\max_{1 \leqslant l \leqslant n} \|n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,l)})\|$ is nearly deterministically given. For the latter case it is not so, and hence it may be preferable to avoid discontinuous $\psi$-functions.

Remark 11. The uniformity in $l$ which has been stated in Theorem 1 has to be interpreted (as follows from the proof of the theorem) in the following way (let us consider for simplicity the case $\psi_s \equiv 0$):

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

$$P\Big(\big\|n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,l)}) - \hat{\sigma}_n \mathsf{E}^{-1}\psi'(e_1/\sigma)\,Q^{-1} g'(X_l, \hat{\beta}^{(n)})\,\psi([Y_l - g(X_l, \hat{\beta}^{(n)})]\,\hat{\sigma}_n^{-1})\big\| > \delta\Big) < \varepsilon,$$

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

$$P\Big(\max_{1 \leqslant l \leqslant n} \big\|n(\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,l)}) - \hat{\sigma}_n \mathsf{E}^{-1}\psi'(e_1/\sigma)\,Q^{-1} g'(X_l, \hat{\beta}^{(n)})\,\psi([Y_l - g(X_l, \hat{\beta}^{(n)})]\,\hat{\sigma}_n^{-1})\big\| > \delta\Big) < \varepsilon.$$

Similarly for the case $\psi_s \not\equiv 0$.

**7. Concluding remarks.** In the LS-regression analysis the formula

$$\hat{\beta}_{LS}^{(n-1,l)}-\hat{\beta}_{LS}^{(n)} = -\{[X^{(n-1,l)}]^T X^{(n-1,l)}\}^{-1} X_l(Y_l-X_l^T\hat{\beta}_{LS}^{(n)})$$

has been frequently used in the studentized form (in the situations when the data are "regressionally equivariant"), i.e. it was applied in the form

$$(59)\qquad (\hat{\beta}_{LS,j}^{(n-1,l)}-\hat{\beta}_{LS,j}^{(n)})[\operatorname{var}(\hat{\beta}_{LS,j}^{(n-1,l)}-\hat{\beta}_{LS,j}^{(n)})]^{-1/2} = (Y_l-X_l^T\hat{\beta}_{LS}^{(n)})\sigma^{-1}$$

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

$$\|\hat{\beta}_{LS}^{(n-1,l)}-\hat{\beta}_{LS}^{(n)}\|\,[\operatorname{var}\|\hat{\beta}_{LS}^{(n-1,l)}-\hat{\beta}_{LS}^{(n)}\|]^{-1/2} = |(Y_l-X_l^T\hat{\beta}_{LS}^{(n)})\sigma^{-1}|.$$

(In what follows for the sake of simplicity of explanation we shall assume for a while that $\psi_s \equiv 0$.)

For the $M$-estimators the presence of $o_p(1)$ in (54) generally does not allow to derive directly from (54) an approximation to the variance of $\|\hat{\beta}^{(n-1,l)}-\hat{\beta}^{(n)}\|$. But it indicates that the term $o_p(1)$ in the representation (54) may cause that the exact variance of $\|\hat{\beta}^{(n-1,l)}-\hat{\beta}^{(n)}\|$ (or of $\hat{\beta}_j^{(n-1,l)}-\hat{\beta}_j^{(n)}$) is much greater than

$$(60)\qquad \Sigma_{\|\cdot\|}^2 = \mathsf{E}^{-2}\psi'(e_1/\sigma)\operatorname{tr}\{Q^{-1}\}\operatorname{var}\{\psi([Y_l-g(X_l,\hat{\beta}^{(n)})]\sigma_n^{-1})\}\sigma^2$$

or than

$$(61)\qquad \Sigma_j^2 = \mathsf{E}^{-2}\psi'(e_1/\sigma)\{Q^{-1}\}_{jj}\operatorname{var}\{\psi([Y_l-g(X_l,\hat{\beta}^{(n)})]\sigma_n^{-1})\}\sigma^2,$$

respectively. This means that the large values of $\operatorname{var}\|\hat{\beta}^{(n-1,l)}-\hat{\beta}^{(n)}\|$ (and of $\operatorname{var}(\hat{\beta}_j^{(n-1,l)}-\hat{\beta}_j^{(n)})$) might be caused by the fluctuation of $\|\hat{\beta}^{(n-1,l)}-\hat{\beta}^{(n)}\|$ (and of $\hat{\beta}_j^{(n-1,l)}-\hat{\beta}_j^{(n)}$) on a set of (very) small probability. But then we may prefer to "studentize" $\|\hat{\beta}^{(n)}-\hat{\beta}^{(n-1,l)}\|$ by (60) (or $\hat{\beta}_j^{(n-1,l)}-\hat{\beta}_j^{(n)}$ by (61)), i.e. by the asymptotic variance of $\|\hat{\beta}^{(n)}-\hat{\beta}^{(n-1,l)}\|$ rather than by an approximation to the exact variance. We obtain

$$\|\hat{\beta}^{(n-1,l)}-\hat{\beta}^{(n)}\|\,\Sigma_{\|\cdot\|}^{-1} = \frac{|Y_l-g(X_l,\hat{\beta}^{(n)})|}{\operatorname{var}^{1/2}\{\psi([Y_l-g(X_l,\beta)]\sigma_n^{-1})\}}+o_p(1)\qquad \text{as } n\to\infty$$

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

$$(\hat{\beta}_j^{(n-1,l)}-\hat{\beta}_j^{(n)})\Sigma_j^{-1} = \frac{\psi(Y_l-g(X_l,\hat{\beta}^{(n)}))}{\operatorname{var}^{1/2}\{\psi([Y_l-g(X_l,\beta)]\sigma_n^{-1})\}}+o_p(1)\qquad \text{as } n\to\infty.$$

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

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

and hence

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

Further, as we have already observed in Remark 10, the only random factor in (54) which depends on the d.f. $F$ is $\psi(Y_i - g(X_i, \hat{\beta}^{(n)}))$, the range of which is bounded by $\inf_{z\in R} \psi(z)$ and $\sup_{z\in R} \psi(z)$. This means that in the case when the great error sensitivity of the estimator is properly assigned, i.e. when some outliers are actually "winsorized," $\max_{1\leqslant l\leqslant n} \|\hat{\beta}^{(n)} - \hat{\beta}^{(n-1,l)}\|$ 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. This 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 is the contamination level. We hope the problem will be treated in the forthcoming paper.

## REFERENCES

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

[2] D. M. Bates and D. G. Watts, *Nonlinear Regression Analysis and Its Applications*, Wiley, New York 1988.

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

[4] L. Breiman, *Probability*, Addison–Wesley, Reading, Massachusetts, 1968.

[5] S. Chatterjee and A. S. Hadi, *Sensitivity Analysis in Linear Regression*, Wiley, New York 1988.

[6] R. D. Cook, *Detection of influential observations in linear regression*, Technometrics 19 (1977), pp. 15–18.

[7] – *Influential observations in linear regression*, J. Amer. Statist. Assoc. 74 (1979), pp. 169–174.

[8] – and S. Weisberg, *Residuals and Influence in Regression*, Chapman and Hall, London 1982.

[9] M. Csörgő and P. Révész, *Strong Approximations in Probability and Statistics*, Akadémiai Kiadí, Budapest 1981.

[10] F. Galton, *Regression towards mediocrity in hereditary stature*, J. Antropological Institute 15 (1886), pp. 246–263.

[11] S. Geisser, *Diagnostics, divergence and perturbation analysis*, in: *Directions in Robust Statistics and Diagnostics*, Part I, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 89–100.

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

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

[14] P. J. Huber, *Between robustness and diagnostics*, in: *Directions in Robust Statistics and Diagnostics*, Part I, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 121–131.

[15] J. Jurečková, *Consistency of M-estimators of vector parameters*, in: *Proceedings of the Fourth Prague Symposium on Asymptotic Statistics*, Charles University, Prague 1988, pp. 305–312.

[16] – and P. K. Sen, *Uniform second order asymptotic linearity of M-statistics in linear models*, Statist. Decisions 7 (1989), pp. 263–276.

[17] J. Jurečková and A. H. Welsh, *Asymptotic relations between L- and M-estimators in the linear model*, Ann. Inst. Statist. Math. 42 (1990), pp. 671–698.

[18] J. Ledolter, *Outliers in time series analysis: Some comments on their impact and their detection*, in: *Directions in Robust Statistics and Diagnostics*, Part I, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 159–166.

[19] P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris 1937.
[20] F. Liese and I. Vajda, *M-estimators in regression with i.i.d. errors*, J. Multivariate Anal. 50 (1994), pp. 93–114.
[21] J. W. McKean, S. J. Sheather and T. P. Hettmansperger, *Regression diagnostics for rank-based method II*, in: *Directions in Robust Statistics and Diagnostics*, Part II, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 21–32.
[22] R. G. Miller, *An unbalanced jackknife*, Ann. Statist. 2 (1974), pp. 880–891.
[23] N. M. Neykov and P. N. Neytchev, *Unmasking multivariate outliers and leverage points by means of BMDP3R*, in: *Directions in Robust Statistics and Diagnostics*, Part II, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 115–128.
[24] J. Nosková, *M-odhady v lineární regresi* (*M-estimators in linear regression*, in Czech), 1991, unpublished doctor thesis.
[25] S. Portnoy, *Tightness of the sequence of empiric c.d.f. processes defined from regression fractiles*, in: *Robust and Nonlinear Time-Series Analysis*, J. Franke, W. Härdle and D. Martin (Eds.), Springer, New York 1983, pp. 231–246.
[26] — *Regression quantile diagnostics for multiple outliers*, in: *Directions in Robust Statistics and Diagnostics*, Part II, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 145–158.
[27] C. R. Rao and L. C. Zhao, *On the consistency of M-estimate in linear model obtained through an estimating equation*, Statist. Probab. Lett. 14 (1992), pp. 79–84.
[28] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley, New York 1987.
[29] A. M. Rubio, F. Quintana and J. Á. Víšek, *Test for differences between M-estimates of non-linear regression model*, Probab. Math Statist. 14 (1993), pp. 189–206.
[30] A. M. Rubio and J. Á. Víšek, *Diagnostics of regression model: Test of goodness of fit*, in: *Proceedings of the Fifth Prague Symposium on Asymptotic Statistics*, M. Hušková and P. Mandl (Eds.), Physica Verlag, 1994, pp. 423–432.
[31] — *A note on asymptotic linearity of M-statistics in nonlinear models*, Kybernetika 32 (1996), pp. 353–374.
[32] A. Sen and M. Srivastava, *Regression Analysis. Theory, Methods and Applications*, Springer, Berlin 1990.
[33] J. S. Simonoff, *General approaches to stepwise identification of unusual values in data analysis*, in: *Directions in Robust Statistics and Diagnostics*, Part II, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991, pp. 223–242.
[34] W. Stahel and S. Weisberg, *Directions in Robust Statistics and Diagnostics*, Parts I and II, W. Stahel and S. Weisberg (Eds.), Springer, New York 1991.
[35] J. Štěpán, *Teorie pravděpodobnosti* (*Theory of Probability*, in Czech) Academia, Praha 1987.
[36] J. Á. Víšek, *A proposal of model selection: Stability of linear regression model*, in: *Transactions of the Eleventh Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Volume B, ACADEMIA, Prague 1992, pp. 451–461.
[37] — *Stability of regression estimates with respect to subsamples*, Comput. Statist. 7 (1992), pp. 183–203.
[38] — *On high breakdown point estimation*, ibidem 11 (1996), pp. 137–146.
[39] — *Sensitivity analysis of M-estimates*, Ann. Inst. Statist. Math. 48 (1996), pp. 469–495.

Department of Stochastic Informatics
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
and
Department of Macroeconomics and Econometrics
Institute of Economic Studies, Faculty of Social Sciences, Charles University
Mailing address: Pod vodárenskou věží 4, 182 08 Prague, Czech Republic

*Received on 29.1.1996;*
*revised version on 10.12.1996*