Jan Ámos Víšek

The least trimmed squares. Part III: Asymptotic normality

Kybernetika, Vol. 42 (2006), No. 2, 203--224

Persistent URL: http://dml.cz/dmlcz/135709

KYBERNETIKA — VOLUME *42* (2006), NUMBER 2, PAGES 203 – 224

THE LEAST TRIMMED SQUARES
Part III: Asymptotic Normality

Jan Ámos Víšek

Asymptotic normality of the least trimmed squares estimator is proved under general conditions. At the end of paper a discussion of applicability of the estimator (including the discussion of algorithm for its evaluation) is offered.

Keywords: robust regression, the least trimmed squares, $\sqrt{n}$-consistency, asymptotic normality

AMS Subject Classification: 62J05, 62F35, 62F12

INTRODUCTION AND NOTATION

The paper is a continuation of the paper of the same name, Part I and II. That is why only brief introduction of notations will be given. For discussion of the definitions and assumptions see Part I. Conclusions of the results of all three parts of paper are given at the end of this part.

Let N denote the set of all positive integers, R the real line and R^p the p-dimensional Euclidean space. Moreover, for any set A let A^o denote the interior of the set (in the topology implied by Euclidean metric). We shall consider for any $n \in N$ the linear regression model

$$Y_i = x_i^{\mathrm{T}} \beta^0 + e_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where Y_i and $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^{\mathrm{T}}$ are values of response and of explanatory variables for the ith case, respectively. β^0 is the vector of regression coefficients and e_i represents random fluctuation (disturbance) of Y_i from the mean value $\mathsf{E}Y_i$. (To be complete, let us add that of course $x_i^{\mathrm{T}} \beta^0 = \sum_{j=1}^{p} x_{ij} \beta_j$.)

Throughout the paper we shall assume that the random variables are defined on a basic probability space $(\Omega, \mathcal{A}, P)$ (other assumptions are given below).

Let us recall that we made (in Part I) one exception from the commonly used notation. Since in what follows we shall use for the description of sets somewhat complicated expressions containing moreover indices, we shall write (in many cases) $I\,\{\mathit{property\ describing\ the\ set\ } A\}$ instead of traditional notation $I_{\{\mathit{property\ describing\ the\ set\ } A\}}$.

In what follows the definition of *the least trimmed squares* will be considered in the form:

Definition 1. For a compact set $\mathcal{K}$ such that the vector of the true regression coefficients $\beta^0 \in \mathcal{K}^o$ the estimator given as

$$\hat{\beta}^{(\mathrm{LTS},n,h)} = \arg\min_{\beta \in \mathcal{K}} \sum_{i=1}^{h} r_{(i)}^2(\beta) \tag{2}$$

will be called *the least trimmed squares* (LTS).

It is clear that for given i the squared residual appears in the sum on the right hand side of (2) iff $r_i^2(\beta) \leq r_{(h)}^2(\beta)$, so that we can write equivalently

$$\begin{aligned}\hat{\beta}^{(\mathrm{LTS},n,h)} &= \arg\min_{\beta \in \mathcal{K}} \sum_{i=1}^{n} r_i^2(\beta) \cdot I\left\{r_i^2(\beta) \leq r_{(h)}^2(\beta)\right\} \\ &= \arg\min_{\beta \in \mathcal{K}} \sum_{i=1}^{n} (Y_i - x_i^{\mathrm{T}}\beta)^2 \cdot I\left\{r_i^2(\beta) \leq r_{(h)}^2(\beta)\right\}. \end{aligned} \tag{3}$$

Now, denote $G(z)$ the distribution function of e_1^2. For any $\alpha \in (0,1)$, u_α^2 will be the upper α-quantile of $G(z)$, i. e.

$$P(e_1^2 > u_\alpha^2) = 1 - G(u_\alpha^2) = \alpha. \tag{4}$$

Further, denote by $[a]_{\mathrm{int}}$ the integer part of a and for any $n \in N$ put

$$h_n = [(1-\alpha)n]_{\mathrm{int}}. \tag{5}$$

Moreover, for any $a, b \in R$ we shall denote $(a,b)_{\mathrm{ord}} = (\min\{a,b\}, \max\{a,b\})$ and the same will be used for the closed intervals. Finally, put $Q_n = \frac{1}{n}\sum_{i=1}^n x_i x_i^{\mathrm{T}}$ and for an arbitrary $\alpha \in (0,1)$ $\;Q_n(\alpha) = \frac{1}{n}\sum_{i=1}^n x_i x_i^{\mathrm{T}} I\{r_i^2(\beta^0) \leq u_\alpha^2\}$.

Prior to continuing the discussion on *the least trimmed squares* it is useful to give the assumptions which will be used in the most assertions.

Assumptions $\mathcal{A}$

The sequences $\{x_i\}_{i=1}^\infty$ $(x_i \in R^p)$ *is a fix sequence of nonrandom vectors from* R^p*. Further, the sequence* $\{e_i\}_{i=1}^\infty$ $(e_i \in R)$ *is a sequence of independent and identically distributed random variables. The distribution function* $F(z)$ *of random fluctuation* e_1 *is symmetric and absolutely continuous with a bounded density* $f(z)$ *which is strictly decreasing on* R^+*. The density is positive on* $(-\infty,\infty)$ *and has bounded in absolute value the first and the second derivative. The second derivative is further Lipschitz of the first order. Moreover,*

$$\sum_{i=1}^{n} \|x_i\|^4 = \mathcal{O}(n) \quad \text{and} \quad \mathsf{E}e_1^4 = \kappa_4 \in (0,\infty). \tag{6}$$

Finally,

$$\lim_{n\to\infty} Q_n = Q \tag{7}$$

where Q is a regular matrix (and convergence is of course assumed coordinatewise).

Alternatively, we shall use the following assumptions (the reasons for it were given in Part I).

Assumptions $\mathcal{B}$

The sequences $\{x_i\}_{i=1}^{\infty}$ ($x_i \in R^p$) is a fix sequence of nonrandom vectors from R^p. Moreover, (7) holds for some regular matrix Q. Further for any $n \in N$

$$\max_{1\le i\le n,\ 1\le j\le p} |x_{ij}| = \mathcal{O}(1). \tag{8}$$

The sequence $\{e_i\}_{i=1}^{\infty}$ ($e_i \in R$) is a sequence of independent and identically distributed random variables with absolutely continuous symmetric distribution function $F(z)$. There is a neighbourhood of u_α in which the distribution $F(z)$ has a bounded density $f(z)$ which is positive and has bounded in absolute value the first and the second derivative. The second derivative is further Lipschitz of the first order. Moreover, the density $f(z)$ is strictly decreasing on R^+ and $\mathsf{E}e_1^4 = \kappa_4 \in (0,\infty)$.

We have proved (in Part I) that

$$\tilde{\beta}^{(\mathrm{LTS},n,h)} = \arg\min_{\beta\in R^p} \sum_{i=1}^{h} r_{(i)}^2(\beta) \tag{9}$$

can be found among solutions of

$$\sum_{i=1}^{n} \left[(Y_i - x_i^{\mathrm{T}}\beta)x_i \cdot I\left\{r_i^2(\beta) \le r_{(h)}^2(\beta)\right\}\right] = 0, \tag{10}$$

i. e. that at the point given as the solution of the extremal problem (9) the relation (10) holds. Notice please that whenever we prove that the estimator given by (2) is consistent (i. e. exists and converges in probability to β^0), it also solves (10).

Assumptions $\mathcal{C}$

There are distribution functions $H^{(\beta)}(t), t \in R, \beta \in R^p$ such that for any compact set $\mathcal{W} \subset R^p$

$$\sup_{\beta\in\mathcal{W}} \sup_{t\in R} \left| \frac{1}{n} \sum_{i=1}^{n} I\left\{x_i^{\mathrm{T}}(\beta - \beta^0) \le t\right\} - H^{(\beta)}(t) \right| = \mathcal{O}(n^{-\frac{1}{2}}). \tag{11}$$

Remark 1. Recently it was found that when X_i's are i.i.d. the first supremum in (11) can be taken over R^p, see Víšek [31].

In what follows let $\#B$ denotes the number of elements of B.

ASYMPTOTIC NORMALITY
OF THE LEAST TRIMMED SQUARES

Lemma 1. Let $\{e_i\}_{i=1}^{\infty}$ ($e_i \in R$) be a sequence of independent and identically distributed random variables with symmetric absolutely continuous distribution function $F(z)$. Moreover, let density $f(z)$ be absolutely continuous and let it have everywhere the derivative bounded in absolute value. Further, let for any $M > 0$

$$\mathcal{T}_M = \left\{ v \in R^+, t \in R^p, \|t\| < M \right\}.$$

Now for any $n \in N$, any $v > 0$, $\tau \in (\frac{1}{2}, \frac{3}{2})$, positive and finite K define $U_n(v, K, \tau)$ so that

$$P(e_1^2 \in (v, U_n(v, K, \tau)) = \min \left\{ n^{-\tau} K, 1 - F(v) \right\}. \tag{12}$$

Now, for any $v, t \in \mathcal{T}_M$, $\tau \in (\frac{1}{2}, \frac{3}{4})$, positive and finite K define

$$m_{n,U}^{(+)}(v,t) = \# \left\{ i \in \{1, 2, \ldots, n\} : r_i(\beta^0) > v \text{ and } r_i(\beta^0 - n^{-\frac{1}{2}} t) \le U_n(v, K, \tau) \right\}, \tag{13}$$

$$m_{n,U}^{(-)}(v,t) = \# \left\{ i \in \{1, 2, \ldots, n\} : r_i^2(\beta^0) \le v^2 \text{ and } r_i(\beta^0 - n^{-\frac{1}{2}} t) \ge U_n(v, K, \tau) \right\}, \tag{14}$$

$$m_{n,L}^{(+)}(v,t) = \# \left\{ i \in \{1, 2, \ldots, n\} : r_i(\beta^0) < -v \text{ and } r_i^2(\beta^0 - n^{-\frac{1}{2}} t) \le U_n^2(v, K, \tau) \right\} \tag{15}$$

and

$$m_{n,L}^{(-)}(v,t) = \# \left\{ i \in \{1, 2, \ldots, n\} : r_i^2(\beta^0) \le v^2 \text{ and } r_i(\beta^0 - n^{-\frac{1}{2}} t) < -U_n(v, K, \tau) \right\}. \tag{16}$$

Finally, let us put

$$m_n(v,t) = m_{n,U}^{(+)}(v,t) + m_{n,L}^{(+)}(v,t) - m_{n,U}^{(-)}(v,t) - m_{n,L}^{(-)}(v,t).$$

Then for any $\varepsilon \in (0,1)$ there is $n_\varepsilon \in N$ such that for all $n > n_\varepsilon$

$$P\left(\left\{ \omega \in \Omega : \inf_{v,t \in \mathcal{T}_M} m_n(v,t) \ge 0 \right\} \right) > 1 - \varepsilon. \tag{17}$$

Remark 2. Let us denote

$$\mathcal{B} = \left\{ v \in R, \min \left\{ n^{-\tau} K, 1 - F(v) \right\} = 1 - F(v) \right\}$$

Of course, if $v \in \mathcal{B}$, then $U_n(v, K, \tau) = \infty$. Consequently, any index i for which $r_i^2(\beta^0) > v^2$ appears either in $m_{n,U}^{(+)}(v,t)$ or in $m_{n,L}^{(+)}(v,t)$. Simultaneously, neither $r_i(\beta^0 - n^{-\frac{1}{2}} t) \ge U_n(v, K, \tau)$ nor $r_i(\beta^0 - n^{-\frac{1}{2}} t) \le -U_n(v, K, \tau)$ can be true for any i, so that $m_{n,U}^{(-)}(v,t) = 0$ as well as $m_{n,L}^{(-)}(v,t) = 0$. It finally means that $P(\inf_{v,t \in \mathcal{T}_M \cap \mathcal{B}} m_n(v,t) \ge 0) = 1$ whenever $U_n(v, K, \tau) = \infty$. In other words, in the proof which follows, not restricting generality we may assume that $U_n(v, K, \tau) < \infty$.

P r o o f o f L e m m a 1. Let us recall that $r_i(\beta^0) = e_i$ and $r_i(\beta^0 - n^{-\frac{1}{2}}t) = Y_i - x_i^{\mathrm{T}}\left(\beta^0 - n^{-\frac{1}{2}}t\right) = e_i + n^{-\frac{1}{2}}x_i^{\mathrm{T}}t$ and hence

$$\left\{r_i(\beta^0) > v \text{ and } r_i(\beta^0 - n^{-\frac{1}{2}}t) \le U_n(v,K,\tau)\right\} \Leftrightarrow \left\{v < e_i \le U_n(v,K,\tau) - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t\right\}.$$

So putting

$$b_i^{(+)}(v,t) = I\left\{v < e_i \le U_n(v,K,\tau) - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t\right\}, \tag{18}$$

we have

$$m_{n,U}^{(+)}(v,t) = \sum_{i=1}^{n} b_i^{(+)}(v,t). \tag{19}$$

Similarly

$$\left\{r_i^2(\beta^0) \le v^2 \text{ and } r_i(\beta^0 - n^{-\frac{1}{2}}t) > U_n(v,K,\tau)\right\}$$
$$\Rightarrow \left\{\max\left\{-v, U_n(v,K,\tau) - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t\right\} < e_i \le v\right\}.$$

So putting analogously as in previous

$$b_i^{(-)}(v,t) = I\left\{U_n(v,K,\tau) - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t < e_i \le v\right\}, \tag{20}$$

we have

$$m_{n,U}^{(-)}(v,t) \le \sum_{i=1}^{n} b_i^{(-)}(v,t). \tag{21}$$

Finally, putting

$$c_i^{(+)}(v,t) = I\left\{-U_n(v,K,\tau) - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t \le e_i < -v\right\} \tag{22}$$

and

$$c_i^{(-)}(v,t) = I\left\{-v \le e_i < -U_n(v,K,\tau) - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t\right\}, \tag{23}$$

we have

$$m_n(v,t) \ge \sum_{i=1}^{n}\left[b_i^{(+)}(v,t) - b_i^{(-)}(v,t) + c_i^{(+)}(v,t) - c_i^{(-)}(v,t)\right].$$

Firstly, previous to continuing, let us realize when $b_i^{(+)}(v,t)$, $b_i^{(-)}(v,t)$, $c_i^{(+)}(v,t)$ or $c_i^{(-)}(v,t)$ have chance to be equal to one. Since

$$b_i^{(+)}(v,t) = 1 \quad \Leftrightarrow \quad e_i \in \left(v, U_n(v,K,\tau) - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t\right],$$

$b_i^{(+)}(v,t)$ can be equal to one only if $v < U_n(v,K,\tau) - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t$. Let us denote the case when it holds by $B_{n,i}^{(+)}$. Similarly, let us denote successively by $B_{n,i}^{(-)}$, $C_{n,i}^{(+)}$ and $C_{n,i}^{(-)}$ the situations when $U_n(v,K,\tau) - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t < v$, $-U_n(v,K,\tau) - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t < -v$ and

$-v < -U_n(v,K,\tau) - n^{-\frac{1}{2}} x_i^{\mathrm{T}} t$. In other words, $B_{n,i}^{(+)}$, $B_{n,i}^{(-)}$, $C_{n,i}^{(+)}$ and $C_{n,i}^{(-)}$ denotes the cases when $n^{-\frac{1}{2}} x_i^{\mathrm{T}} t < U_n(v,K,\tau) - v$, $U_n(v,K,\tau) - v < n^{-\frac{1}{2}} x_i^{\mathrm{T}} t$, $v - U_n(v,K,\tau) < n^{-\frac{1}{2}} x_i^{\mathrm{T}} t$ and $n^{-\frac{1}{2}} x_i^{\mathrm{T}} t < v - U_n(v,K,\tau)$. It implies that if

$$v - U_n(v,K,\tau) < n^{-\frac{1}{2}} x_i^{\mathrm{T}} t < U_n(v,K,\tau) - v \;\Rightarrow\; b_i^{(-)}(v,t) = 0 \text{ and } c_i^{(-)}(v,t) = 0,$$

but $b_i^{(+)}(v,t)$ and $c_i^{(+)}(v,t)$ can be equal to one. It implies that only for $B_{n,i}^{(+)} \cap C_{n,i}^{(-)} = C_{n,i}^{(-)}$ both $b_i^{(+)}(v,t)$ and $c_i^{(-)}(v,t)$ can be positive. Similarly only for $B_{n,i}^{(-)} \cap C_{n,i}^{(+)} = B_{n,i}^{(-)}$ both $b_i^{(-)}(v,t)$ and $c_i^{(+)}(v,t)$ can be positive. So, we found that

$$m_n(v,t) \geq \sum_{i=1}^{n} \left(I_{C_{n,i}^{(-)}} + I_{B_{n,i}^{(-)}} \right) \left[b_i^{(+)}(v,t) - b_i^{(-)}(v,t) + c_i^{(+)}(v,t) - c_i^{(-)}(v,t) \right] \quad (24)$$

(notice again that only one of the indicators $I_{C_{n,i}^{(-)}}$ and $I_{B_{n,i}^{(-)}}$ can be nonzero). So, if we prove an analogy of (17) taking into account instead of $m_n(v,t)$ the right hand side of (24), (17) will be verified. Hence let us denote the set of indices for which either $n^{-\frac{1}{2}} x_i^{\mathrm{T}} t < v - U_n(v,K,\tau)$ or $n^{-\frac{1}{2}} x_i^{\mathrm{T}} t > U_n(v,K,\tau) - v$ by $\mathcal{I}$ and restrict ourselves only on them in what follows. As $v - U_n(v,K,\tau) < 0$ and $U_n(v,K,\tau) - v > 0$, we can, e. g. when considering integral in (27), simplify the discussion assuming that

$$n^{-\frac{1}{2}} x_i^{\mathrm{T}} t < 0. \quad (25)$$

Denoting successively

$$\xi_i^{(+)}(v,t) = b_i^{(+)}(v,t) - \mathsf{E} b_i^{(+)}(v,t),$$

$$\xi_i^{(-)}(v,t) = b_i^{(-)}(v,t) - \mathsf{E} b_i^{(-)}(v,t),$$

$$\zeta_i^{(+)}(v,t) = c_i^{(+)}(v,t) - \mathsf{E} c_i^{(+)}(v,t)$$

and

$$\zeta_i^{(-)}(v,t) = c_i^{(-)}(v,t) - \mathsf{E} c_i^{(-)}(v,t),$$

we have

$$\begin{aligned} \inf_{v,t \in \mathcal{T}_M} m_n(v,t) \geq & \inf_{v,t \in \mathcal{T}_M} \sum_{i \in \mathcal{I}} \xi_i^{(+)}(v,t) - \sup_{v,t \in \mathcal{T}_M} \sum_{i \in \mathcal{I}} \xi_i^{(-)}(v,t) \\ & + \inf_{v,t \in \mathcal{T}_M} \sum_{i \in \mathcal{I}} \zeta_i^{(+)}(v,t) - \sup_{v,t \in \mathcal{T}_M} \sum_{i \in \mathcal{I}} \zeta_i^{(-)}(v,t) \\ & + \inf_{v,t \in \mathcal{T}_M} \sum_{i \in \mathcal{I}} \mathsf{E} \left(b_i^{(+)}(v,t) - b_i^{(-)}(v,t) + c_i^{(+)}(v,t) - c_i^{(-)}(v,t) \right). \end{aligned} \quad (26)$$

Denoting

$$\pi_{i,n}(v,t) = \mathsf{E} b_i^{(+)}(v,t),$$

we have

$$P\left\{ \xi_i^{(+)}(v,t) = 1 - \pi_{i,n}(v,t) \right\} = \pi_{i,n}(v,t)$$

and

$$P\left\{\xi_i^{(+)}(v,t) = -\pi_{i,n}(v,t)\right\} = 1 - \pi_{i,n}(v,t).$$

Now, we are going to employ Lemma A.1 and principle of invariance (see [3] Theorem 13.12). Let us recall that due to the definition of $\xi_i^{(+)}(v,t)$'s, they are independent. Similarly as in Part I of this paper let us denote by $W(s)$ the Wiener process and let us define $\tau_i^{(+)}(v,t)$ to be the time for the Wiener process to exit interval $(-\pi_{i,n}(v,t), 1-\pi_{i,n}(v,t))$. Then $\xi_i^{(+)}(v,t) =_{\mathcal{D}} W(\tau_i^{(+)}(v,t))$ and hence

$$n^{-\frac{1}{4}} \sum_{i\in\mathcal{I}} \xi_i^{(+)}(v,t) =_{\mathcal{D}} n^{-\frac{1}{4}} \sum_{i\in\mathcal{I}} W(\tau_i^{(+)}(v,t)) =_{\mathcal{D}} W\left(n^{-\frac{1}{2}} \sum_{i\in\mathcal{I}} \tau_i^{(+)}(v,t)\right).$$

Now we need to estimate $\mathsf{E}b_i^{(+)}(v,t)$. Having denoted the upper and the lower bounds of the derivative $f'(z)$ by $U_{f'}$ and $L_{f'}$, respectively, taking into account the remark above (25) and expanding the density as

$$f(v) = f(y) + f'(\eta_i)(v-y)$$

for an appropriate $\eta_i \in (y,v)_{\text{ord}}$, we have (for $i \in \mathcal{I}$), we have

$$\begin{aligned}
&\mathsf{E}b_i^{(+)}(v,t) = \pi_{i,n}(v,t) &(27)\\
&= P\left(v < e_i \le U_n(v,K,\tau) - n^{-\frac{1}{2}} x_i^{\mathrm{T}} t\right) \le I_{\left\{n^{-\frac{1}{2}} x_i^{\mathrm{T}} t<0\right\}} \int_v^{U_n(v,K,\tau)-n^{-\frac{1}{2}} x_i^{\mathrm{T}} t} f(z)\mathrm{d}z\\
&= I_{\left\{n^{-\frac{1}{2}} x_i^{\mathrm{T}} t<0\right\}} \int_v^{U_n(v,K,\tau)} f(z)\mathrm{d}z + I_{\left\{n^{-\frac{1}{2}} x_i^{\mathrm{T}} t<0\right\}} \int_{U_n(v,K,\tau)}^{U_n(v,K,\tau)-n^{-\frac{1}{2}} x_i^{\mathrm{T}} t} f(z)\mathrm{d}z\\
&\le I_{\left\{n^{-\frac{1}{2}} x_i^{\mathrm{T}} t<0\right\}} \left[n^{-\tau}\cdot K - n^{-\frac{1}{2}} x_i^{\mathrm{T}} t f(v)\right] + U_{f'}\left\{\frac{1}{2} n^{-1}\left[x_i^{\mathrm{T}} t\right]^2 + n^{-2\tau} K\right\}
\end{aligned}$$

and also

$$\begin{aligned}
\pi_{i,n}(v,t) \quad \ge \quad & I_{\left\{n^{-\frac{1}{2}} x_i^{\mathrm{T}} t< v-U_n(v,K,\tau)\right\}} \left[n^{-\tau}\cdot K - n^{-\frac{1}{2}} x_i^{\mathrm{T}} t f(v)\right]\\
& + L_{f'}\left\{\frac{1}{2} n^{-1}\left[x_i^{\mathrm{T}} t\right]^2 + n^{-2\tau} K\right\}.
\end{aligned}$$

It gives

$$\pi_{i,n}(v,t) = I_{\left\{n^{-\frac{1}{2}} x_i^{\mathrm{T}} t< v-U_n(v,K,\tau)\right\}} \left[n^{-\tau}\cdot K - n^{-\frac{1}{2}} x_i^{\mathrm{T}} t f(v)\right] + R_i(v,t) \tag{28}$$

where, due to the fact that $v,t \in \mathcal{T}_M$, there is $K^{(1)} < \infty$ so that

$$\sup_{u,t\in\mathcal{T}_M} |R_i(v,t)| < K^{(1)}\left[n^{-1} + 2n^{-2\tau} K\right] \cdot \|x_i\|^2. \tag{29}$$

So, denoting the upper bound of the density $f(z)$ by U_f, we have

$$\sup_{u,t\in\mathcal{T}_M} \pi_{i,n}(v,t) \le n^{-\frac{1}{2}} p^{\frac{1}{2}} U_f \|x_i\| M + n^{-\tau} K^{(2)} + n^{-1} \|x_i\|^2 K^{(3)}$$

where $K^{(2)}$ and $K^{(3)}$ are positive finite constants. Now, let V_i be the time for the Wiener process to exit the interval $(-n^{-\frac{1}{2}} p^{\frac{1}{2}} U_f \|x_i\| M + n^{-\tau} K^{(2)} + n^{-1} \|x_i\|^2 K^{(3)}, 1)$. Then

$$\frac{1}{\sqrt{n}} \sum_{i\in\mathcal{I}} \tau_i^{(+)}(v,t) \le \frac{1}{\sqrt{n}} \sum_{i\in\mathcal{I}} V_i$$

and hence – by invariance principle (see [3], the proof of Proposition 13.15)

$$n^{-\frac{1}{4}} \sup_{u,t\in\mathcal{T}_M} \left| \sum_{i\in\mathcal{I}} \xi_i^{(+)}(v,t) \right| =_{\mathcal{D}} \sup_{u,t\in\mathcal{T}_M} \left| W\left(\frac{1}{\sqrt{n}} \sum_{i\in\mathcal{I}} \tau_i^{(+)}(v,t) \right) \right|$$

$$\le \sup \left\{ |W(s)| \ : \ 0 \le s \le \frac{1}{\sqrt{n}} \sum_{i\in\mathcal{I}} V_i \right\}. \tag{30}$$

Moreover, employing Lemma A.1 once again we arrive at

$$\mathsf{E}\left\{ \frac{1}{\sqrt{n}} \sum_{i\in\mathcal{I}} V_i \right\} \le \left(n^{-1} p^{\frac{1}{2}} U_f M + n^{-\frac{3}{2}} K^{(3)} \right) \sum_{i\in\mathcal{I}} \left[\|x_i\| + \|x_i\|^2 \right] + K^{(4)} = O(1)$$

(where $K^{(4)}$ was selected so that $n^{1-\frac{1}{2}-\tau} K^{(2)} < K^{(4)}$; remember that $\tau \in \left(\frac{1}{2}, \frac{3}{4}\right)$). Let us fix an $\varepsilon > 0$. As all V_i's are nonnegative

$$P\left(\frac{1}{\sqrt{n}} \sum_{i\in\mathcal{I}} V_i > L \right) < \frac{1}{L} \mathsf{E}\left\{ \frac{1}{\sqrt{n}} \sum_{i\in\mathcal{I}} V_i \right\}$$

and hence there is a constant $K^{(5)}$ such that for all $n \in N$ we have

$$P\left(\frac{1}{\sqrt{n}} \sum_{i\in\mathcal{I}} V_i > K^{(5)} \right) < \varepsilon. \tag{31}$$

Now, from the character of the Wiener process it follows that there is a constant $K^{(6)}$ such that

$$P\left(\sup\left\{ |W(s)| : 0 \le s \le K^{(5)} \right\} > K^{(6)} \right) \le \varepsilon. \tag{32}$$

Finally, utilizing (30), (31) and (32), we obtain

$$P\left(n^{-\frac{1}{4}} \sup_{u,t\in\mathcal{T}_M} \left| \sum_{i\in\mathcal{I}} \xi_i^{(+)}(v,t) \right| > K^{(6)} \right) < 2\varepsilon. \tag{33}$$

Along similar lines we can show that, for an appropriate $K^{(7)}$

$$P\left(n^{-\frac{1}{4}}\sup_{u,t\in\mathcal{T}_M}\left|\sum_{i\in\mathcal{I}}\xi_i^{(-)}(v,t)\right|>K^{(7)}\right)<2\varepsilon \tag{34}$$

etc. for $\sum_{i\in\mathcal{I}}\zeta_i^{(+)}(v,t)$ and $\sum_{i\in\mathcal{I}}\zeta_i^{(-)}(v,t)$. In other words, the first four terms on the right hand side of (26) are $O_p(n^{\frac{1}{4}})$. On the other hand, we can derive similarly as in (28) and (29) that there is a positive and finite constant $K^{(8)}$ such that

$$\mathsf{E}b_i^{(+)}(v,t)\cdot I_{B_{n,i}^{(+)}\cap C_{n,i}^{(-)}}=\left[n^{-\tau}\cdot K-n^{-\frac{1}{2}}x_i^{\mathrm{T}}tf(v)\right]\cdot I_{B_{n,i}^{(+)}\cap C_{n,i}^{(-)}}+R_{i,b^{(+)}}(v,t), \tag{35}$$

$$\mathsf{E}b_i^{(-)}(v,t)\cdot I_{B_{n,i}^{(-)}\cap C_{n,i}^{(+)}}=\left[-n^{-\tau}\cdot K-n^{-\frac{1}{2}}x_i^{\mathrm{T}}tf(v)\right]\cdot I_{B_{n,i}^{(-)}\cap C_{n,i}^{(+)}}+R_{i,b^{(-)}}(v,t), \tag{36}$$

$$\mathsf{E}c_i^{(+)}(v,t)\cdot I_{B_{n,i}^{(-)}\cap C_{n,i}^{(+)}}=\left[n^{-\tau}\cdot K-n^{-\frac{1}{2}}x_i^{\mathrm{T}}tf(v)\right]\cdot I_{B_{n,i}^{(-)}\cap C_{n,i}^{(+)}}+R_{i,c^{(+)}}(v,t) \tag{37}$$

and

$$\mathsf{E}c_i^{(-)}(v,t)\cdot I_{B_{n,i}^{(+)}\cap C_{n,i}^{(-)}}=\left[-n^{-\tau}\cdot K-n^{-\frac{1}{2}}x_i^{\mathrm{T}}tf(v)\right]\cdot I_{B_{n,i}^{(+)}\cap C_{n,i}^{(-)}}+R_{i,c^{(-)}}(v,t) \tag{38}$$

with

$$\sup_{u,t\in\mathcal{T}_M}\left[|R_{i,b^{(+)}}(v,t)|+|R_{i,b^{(-)}}(v,t)|+|R_{i,c^{(+)}}(v,t)|+|R_{i,c^{(-)}}(v,t)|\right]$$
$$<n^{-1}K^{(8)}\left[\|x_i\|+\|x_i\|^2\right]. \tag{39}$$

Now utilizing (26), the fact that the first four terms on the right hand side of (26) are $O_p(n^{\frac{1}{4}})$, (35)–(39) and (6), we can find $K^{(9)}$, $K^{(10)}\in(0,\infty)$ and $n_\varepsilon\in N$, so that for all $n>n_\varepsilon$

$$P\left(\inf_{u,t\in\mathcal{T}_M}m_n(v,t)\geq -n^{\frac{1}{4}}\cdot K^{(9)}+n^{1-\tau}\cdot K^{(10)}+\mathcal{O}(1)\right)>1-8\varepsilon. \tag{40}$$

Since $1-\tau>\frac{1}{4}$, the expression $-n^{\frac{1}{4}}\cdot K^{(9)}+n^{1-\tau}\cdot K^{(10)}+\mathcal{O}(1)$ converges to ∞. That concludes the proof. □

Corollary 1. Let the assumptions of Lemma 1 hold and $1\leq\ell\leq n$. Then for any $\varepsilon>0$, any $\tau\in(\frac{1}{2},\frac{3}{4})$ and $K\in(0,\infty)$ there is an $n_{\varepsilon,K,\tau}\in N$ so that for any $n>n_{\varepsilon,K,\tau}$ there a set is $B_n\subset\Omega$ such that $P\left(B_n\right)>1-\varepsilon$ and

$$P\left(e_1^2\in\left(r_{(\ell)}^2(\beta^0),r_{(\ell)}^2(\beta^0-n^{-\frac{1}{2}}t)\right)_{\mathrm{ord}}\cap B_n\right)\leq n^{-\tau}K. \tag{41}$$

Proof. According to Lemma 1 we can for any positive ε, any $\tau\in(\frac{1}{2},\frac{3}{4})$ and any positive and finite K find n_ε so that for any $n>n_\varepsilon$

$$B_n=\left\{\omega\in\Omega:\inf_{v,t\in\mathcal{T}_M}m_n(v,t)\geq 0\right\}$$

is such that $P(B_n) > 1 - \varepsilon$ and

$$\sup_{v \in R} P(e_1^2 \in (v, U_n(v,K,\tau)) \le n^{-\tau} K.$$

Now we may write

$$\begin{aligned} & P\left(e_1^2 \in \left(r^2_{(\ell)}(\beta^0), r^2_{(\ell)}(\beta^0 - n^{-\frac{1}{2}}t)\right)_{\text{ord}} \cap B_n\right) \\ = \; & P\left(e_1^2 \in \left(r^2_{(\ell)}(\beta^0), r^2_{(\ell)}(\beta^0 - n^{-\frac{1}{2}}t)\right)_{\text{ord}} \cap B_n \cap \{\left\{e_1^2 = r^2_{(\ell)}(\beta^0)\right\}\right) \\ & + P\left(e_1^2 \in \left(r^2_{(\ell)}(\beta^0), r^2_{(\ell)}(\beta^0 - n^{-\frac{1}{2}}t)\right)_{\text{ord}} \cap B_n \cap \left\{e_1^2 \neq r^2_{(\ell)}(\beta^0)\right\}\right). \end{aligned}$$

Assertion 1 of Part II then implies that

$$P\left(e_1^2 \in \left(r^2_{(\ell)}(\beta^0), r^2_{(\ell)}(\beta^0 - n^{-\frac{1}{2}}t)\right)_{\text{ord}} \cap B_n \cap \left\{e_1^2 = r^2_{(\ell)}(\beta^0)\right\}\right) \le \frac{1}{n}.$$

For any $v \in R^+$ let us denote by $\mathcal{C}(v) = \left\{r^2_{(\ell)}(\beta^0) = v\right\}$. Then

$$\begin{aligned} & P\left(e_1^2 \in \left(r^2_{(\ell)}(\beta^0), r^2_{(\ell)}(\beta^0 - n^{-\frac{1}{2}}t)\right)_{\text{ord}} \right. \\ & \qquad \left. \cap B_n \cap \left\{e_1^2 \neq r^2_{(\ell)}(\beta^0)\right\}\right) \\ = \; & \mathsf{E}_{\mathcal{C}(v)} \left\{P\left(e_1^2 \in \left(r^2_{(\ell)}(\beta^0), r^2_{(\ell)}(\beta^0 - n^{-\frac{1}{2}}t)\right)_{\text{ord}} \right.\right. \\ & \qquad \left.\left. \cap B_n \cap \left\{e_1^2 \neq r^2_{(\ell)}(\beta^0)\right\} \mid \mathcal{C}(v)\right)\right\} \\ = \; & \mathsf{E}_{\mathcal{C}(v)} \left\{P\left(e_1^2 \in \left(r^2_{(\ell)}(\beta^0), r^2_{(\ell)}(\beta^0 - n^{-\frac{1}{2}}t)\right)_{\text{ord}} \right.\right. \\ & \qquad \left.\left. \cap B_n \cap \left\{e_1^2 \neq r^2_{(\ell)}(\beta^0)\right\} \mid r^2_{(\ell)}(\beta^0) = v\right)\right\}. \end{aligned}$$

Considerations similar to those we made in the proof of Lemma 2 of Part II hint that e_1 is on the set $\{e_1^2 \neq r^2_{(\ell)}(\beta^0)\}$ independent from $r^2_{(\ell)}(\beta^0)$ (an alternative way how to see it is to consider the $\ell - 1$th order statistic among $e_2^2, e_3^2, \ldots, e_n^2$ for the case when $e_1^2 < r^2_{(\ell)}(\beta^0)$ and the ℓth order statistic among $e_2^2, e_3^2, \ldots, e_n^2$ for the case when $e_1^2 > r^2_{(\ell)}(\beta^0)$). Now, let us realize once again that $m_n(v,t)$ represents the lower bound for the difference between the number of squared residuals $r_i^2(\beta^0 - n{-}\frac{1}{2}t)$'s which are smaller than $U_n(v,K,\tau)$ and number of squared residuals $r_i^2(\beta^0)$'s which are smaller than v. Since we assume that $r^2_{(\ell)}(\beta^0) = v$ (i. e. number of squared residuals $r_i^2(\beta^0)$'s which are smaller than v is equal to ℓ), (17) implies that

$$\begin{aligned} & P\left(e_1^2 \in \left(r^2_{(\ell)}(\beta^0), r^2_{(\ell)}(\beta^0 - n^{-\frac{1}{2}}t)\right)_{\text{ord}} \cap B_n \cap \left\{e_1^2 \neq r^2_{(\ell)}(\beta^0)\right\} \mid r^2_{(\ell)}(\beta^0) = v\right) \\ & \qquad \le P(e_1^2 \in (v, U_n(v,K,\tau)) \le n^{-\tau} K. \end{aligned}$$

Taking the mean value over $\mathcal{C}(v)$ we arrive at

$$P\left(e_1^2 \in \left(r^2_{(\ell)}(\beta^0), r^2_{(\ell)}(\beta^0 - n^{-\frac{1}{2}}t)\right)_{\text{ord}} \cap B_n \cap \left\{e_1 \neq r^2_{(\ell)}(\beta^0)\right\}\right) \le n^{-\tau} K$$

and the proof follows. $\square$

Remark 3. Notice that the upper bound given in the previous corollary does not depend on ℓ. So we can claim that even

$$\sup_{\ell=1,2,\ldots,n} P\left(e_1^2 \in \left(r^2_{(\ell)}(\beta^0), r^2_{(\ell)}(\beta^0 - n^{-\frac{1}{2}}t)\right)_{\text{ord}} \cap B_n\right) \leq n^{-\tau}K$$

but we shall not need it.

Lemma 2. Let $\{e_i\}_{i=1}^{\infty}$ $(e_i \in R)$ be a sequence of independent and identically distributed random variables with symmetric absolutely continuous distribution function $F(z)$. Moreover, let density $f(z)$ be absolutely continuous and bounded. Let $\alpha \in \left[0, \frac{1}{2}\right]$ and u_α^2 be the corresponding upper quantile of the distribution $G(z)$. Further, let for any $M > 0$

$$\mathcal{T}_M = \left\{u, v, \tau \in R, t \in R^p, |u| < M, |v| < M, \tau \in \left(\frac{1}{2}, \frac{3}{4}\right), \|t\| < M\right\}.$$

Now, for any $u, v, \tau, t \in \mathcal{T}_M$, define

$$\begin{aligned} S(u,v,\tau,t) \quad = \quad & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ x_i \cdot \left[\left| I\left\{ r_i^2(\beta^0 - n^{-\frac{1}{2}}t) \leq \left(u_\alpha + n^{-\frac{1}{2}}u + n^{-\tau}v\right)^2 \right\} \right.\right.\right. \\ & \left. - I\left\{ r_i^2(\beta^0) \leq \left(u_\alpha + n^{-\frac{1}{2}}u\right)^2 \right\} \right| \\ & - \mathsf{E} \left| I\left\{ r_i^2(\beta^0 - n^{-\frac{1}{2}}t) \leq \left(u_\alpha + n^{-\frac{1}{2}}u + n^{-\tau}v\right)^2 \right\} \right. \\ & \left.\left.\left. - I\left\{ r_i^2(\beta^0) \leq \left(u_\alpha + n^{-\frac{1}{2}}u\right)^2 \right\} \right| \right] \right\}. \end{aligned} \tag{42}$$

Then

$$\sup_{u,v,\tau,t \in \mathcal{T}_M} \|S(u,v,t,\tau)\| = o_p(1). \tag{43}$$

P r o o f. Recalling that $r_i(\beta^0 - n^{-\frac{1}{2}}t) = e_i + n^{-\frac{1}{2}}x_i^{\mathrm{T}}t$, we can verify that

$$\begin{aligned} \zeta_{in}(u,v,\tau,t) \quad = \quad & \left| I\left\{ r_i^2(\beta^0 - n^{-\frac{1}{2}}t) \leq (u_\alpha + n^{-\frac{1}{2}}u + n^{-\tau}v)^2 \right\} \right. \\ & \left. - I\left\{ r_i^2(\beta^0) \leq (u_\alpha + n^{-\frac{1}{2}}u)^2 \right\} \right| = 1 \end{aligned}$$

if either

$$-u_\alpha - n^{-\frac{1}{2}}u - n^{-\tau}v - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t \leq e_i < -u_\alpha - n^{-\frac{1}{2}}u$$

or

$$u_\alpha + n^{-\frac{1}{2}}u < e_i \leq u_\alpha + n^{-\frac{1}{2}}u + n^{-\tau}v + n^{-\frac{1}{2}}x_i^{\mathrm{T}}t$$

or

$$-u_\alpha - n^{-\frac{1}{2}}u \leq e_i < -u_\alpha - n^{-\frac{1}{2}}u - n^{-\tau}v - n^{-\frac{1}{2}}x_i^{\mathrm{T}}t$$

or

$$u_\alpha + n^{-\frac{1}{2}}u + n^{-\tau}v + n^{-\frac{1}{2}}x_i^{\mathrm{T}}t < e_i \leq u_\alpha + n^{-\frac{1}{2}}u.$$

So, denoting

$$P\left(\left|I\left\{r_i^2(\beta^0-n^{-\frac{1}{2}}t)\le u_\alpha+n^{-\frac{1}{2}}u+n^{-\tau}v\right\}\right.\right.$$
$$\left.\left.-I\left\{r_i^2(\beta^0)\le u_\alpha+n^{-\frac{1}{2}}u\right\}\right|=1\right)=\pi_{in}(u,v,\tau,t)$$

and by $U_f=\sup_{z\in R} f(z)$, we find that

$$\pi_{in}(u,v,\tau,t)\le\left(n^{-\tau}+n^{-\frac{1}{2}}\|x_i\|\right)U_f M$$

and hence

$$\sup_{u,v,\tau,t\in\mathcal{T}_M}\pi_{in}(u,v,\tau,t)\le\left(n^{-\tau}+n^{-\frac{1}{2}}\|x_i\|\right)U_f M<n^{-\frac{1}{2}}\left(1+\|x_i\|\right)U_f M. \tag{44}$$

We have of course

$$\mathsf{E}\zeta_{in}(u,v,\tau,t)=\pi_{in}(u,v,\tau,t).$$

Let us fix $j_0\in\{1,2,\ldots,p\}$ and put

$$\xi_{in}(u,v,\tau,t)=|x_{ij_0}|\left(\zeta_{in}(u,v,\tau,t)-\mathsf{E}\zeta_{in}(u,v,\tau,t)\right)$$

and consider

$$n^{-\frac{1}{4}}\sum_{i=1}^{n}\xi_{in}(u,v,\tau,t).$$

We observe that $\xi_{in}(u,v,\tau,t)$ attains value $|x_{ij_0}|\,(1-\pi_{in}(u,v,\tau,t))$ with probability $\pi_{in}(u,v,\tau,t)$ and value $-|x_{ij_0}|\,\pi_{in}(u,v,\tau,t)$ with probability $1-\pi_{in}(u,v,\tau,t)$. Similarly as in previous let us denote by $\{W_i(s)\}_{i=1}^{\infty}$ the sequence of Wiener process and let us define $\tau_{in}(u,v,\tau,t)$ to be the time for the Wiener process to exit interval $(-|x_{ij_0}|\,\pi_{i,n}(u,v,\tau,t),|x_{ij_0}|\,(1-\pi_{in}(u,v,\tau,t)))$. Then
$\xi_{in}(u,v,\tau,t)=_{\mathcal{D}}W_i(\tau_{in}(u,v,\tau,t)(u,t))$. Moreover,

$$\begin{aligned}n^{-\frac{1}{4}}\sum_{i=1}^{n}\xi_{in}(u,v,\tau,t)\quad&=_{\mathcal{D}}\quad n^{-\frac{1}{4}}\sum_{i=1}^{n}W_i(\tau_{in}(u,v,\tau,t)(u,t))\\&=_{\mathcal{D}}\quad W_n\left(n^{-\frac{1}{2}}\sum_{i=1}^{n}\tau_{in}(u,v,\tau,t)(u,t)\right).\end{aligned}$$

Employing (44), let us define V_i be the time for the Wiener process $W_i$to exit the interval

$$(-n^{-\frac{1}{2}}\,|x_{ij_0}|\,(1+\|x_i\|)\,U_f M,|x_{ij_0}|).$$

Then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tau_{in}(u,v,\tau,t)(u,t)\le\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i$$

and hence (using invariance principle – see also [3] the proof of Proposition 13.15)

$$n^{-\frac{1}{4}}\sup_{u,v,\tau,t\in\mathcal{T}_M}\left|\sum_{i=1}^{n}\xi_{in}(u,v,\tau,t)\right|=_{\mathcal{D}}\sup_{u,v,\tau,t\in\mathcal{T}_M}\left|W_n\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tau_{in}(u,v,\tau,t)(u,t)\right)\right|$$

$$\leq \sup\left\{|W(s)| \;:\; 0 \leq s \leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\right\}. \tag{45}$$

Moreover, employing Lemma A.1 once again we arrive at

$$\mathsf{E}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\right\} \leq n^{-1}\sum_{i=1}^{n} x_{ij_0}^2\left(1 + \|x_i\|\right) U_f M = O(1).$$

Let us fix an $\varepsilon > 0$. As all V_i's are nonnegative

$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i > L\right) < \frac{1}{L}\mathsf{E}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\right\}$$

and hence there is a constant $K^{(5)}$ such that for all $n \in N$ we have

$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i > K^{(1)}\right) < \varepsilon. \tag{46}$$

Now, from the character of the Wiener process it follows that there is a constant $K^{(2)} < \infty$ such that

$$P\left(\sup\left\{|W(s)| : 0 \leq s \leq K^{(1)}\right\} > K^{(2)}\right) \leq \varepsilon. \tag{47}$$

Finally, utilizing (45), (46) and (47), we obtain

$$P\left(n^{-\frac{1}{4}}\sup_{u,v,\tau,t\in\mathcal{T}_M}\left|\sum_{i=1}^{n}\xi_{in}(u,v,\tau,t)\right| > K^{(2)}\right) < 2\varepsilon.$$

and the proof follows. □

Theorem 1. Let Assumptions $\mathcal{A}$ or $\mathcal{B}$ and $\mathcal{C}$ be fulfilled. Moreover, let $\mathcal{K} \subset R^p$ be compact and $\beta^0 \in \mathcal{K}$. Then

$$\sqrt{n}\left(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0\right) = n^{-\frac{1}{2}} Q_n^{-1}\left[(1-\alpha) - 2\cdot u_\alpha f(u_\alpha)\right]^{-1} \times$$
$$\times \sum_{i=1}^{n}\left(Y_i - x_i^{\mathrm{T}}\beta^0\right) x_i \cdot I\left\{e_i^2 \leq u_\alpha^2\right\} + o_p(1)$$

and $\hat{\beta}^{(\mathrm{LTS},n,h)}$ is asymptotically normal with mean value equal to β^0 and covariance matrix

$$V(F,\alpha) = Q^{-1}\left[(1-\alpha) - 2\cdot u_\alpha f(u_\alpha)\right]^{-2}\int_{-u_\alpha}^{u_\alpha} z^2 \mathrm{d}F(z),$$

i. e.

$$\mathcal{L}\left(\sqrt{n}\left(\hat{\beta}^{(\mathrm{LTS},n,h_n)} - \beta^0\right)\right) \to \mathcal{N}(0, V(F,\alpha)) \qquad \text{as } n \to \infty$$

where $h_n = [(1-\alpha)n]_{\text{int}}$ (see (7) of Part I).

P r o o f . We are going to employ (see the proof of Theorem 1 of Part II)

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[(Y_i - x_i^{\mathrm{T}}\beta^0)x_i \cdot I\Big\{r_i^2(\beta^0) \le r_{(h)}^2(\beta^0)\Big\}\right] \\
= \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}\left(\hat\beta^{(\mathrm{LTS},n,h)} - \beta^0\right) I\Big\{r_i^2(\beta^0) \le r_{(h)}^2(\beta^0)\Big\} \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Y_i - x_i^{\mathrm{T}}\hat\beta^{(\mathrm{LTS},n,h)}\right)x_i\left[I\Big\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)}) \le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\Big\}\right. \\
&\left.-I\Big\{r_i^2(\beta^0) \le r_{(h)}^2(\beta^0)\Big\}\right] \\
= \;& \frac{1}{n}\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}\left[I\{e_i^2 \le u_\alpha^2\}\right]\cdot\sqrt{n}\left(\hat\beta^{(\mathrm{LTS},n,h)} - \beta^0\right) \qquad (48)\\
&+\frac{1}{n}\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}\left[I\Big\{e_i^2 \le e_{(h)}^2\Big\} - I\{e_i^2 \le u_\alpha^2\}\right]\sqrt{n}\left(\hat\beta^{(\mathrm{LTS},n,h)} - \beta^0\right) \\
&+\frac{1}{n}\sum_{i=1}^{n} x_i x_i^{\mathrm{T}}\left[I\Big\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)}) \le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\Big\}\right. \\
&\left.-I\Big\{r_i^2(\beta^0) \le r_{(h)}^2(\beta^0)\Big\}\right]\cdot\sqrt{n}\left(\hat\beta^{(\mathrm{LTS},n,h)} - \beta^0\right) \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i x_i\left[I\Big\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)}) \le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\Big\} - I\Big\{r_i^2(\beta^0) \le r_{(h)}^2(\beta^0)\Big\}\right].
\end{aligned}
$$

The analysis of the left hand side as well as of the first three terms of right hand side can be the same as in the proof of Theorem 1 of Part II. The only difference will be in the analysis of the last term of the right hand side, namely of

$$
I\Big\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)}) \le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\Big\} - I\Big\{r_i^2(\beta^0) \le r_{(h)}^2(\beta^0)\Big\}. \qquad (49)
$$

First of all, let us fix an $\varepsilon \in (0,1)$ and employ Lemma 1 of Part I, Theorem 1 of Part II and Corollary 1 in order to find $K^{(1)} < \infty$ and $n_1 \in N$ so that for all $n > n_1$ there is a set B_n such that $P(B_n) > 1-\varepsilon$ and for any $\omega \in B_n$

$$
\left|\sqrt{e_{(h_n)}^2} - u_\alpha\right| < n^{-\frac{1}{2}}K^{(1)}, \qquad \sqrt{n}\left\|\hat\beta^{(\mathrm{LTS},n,h)} - \beta^0\right\| < K^{(1)} \qquad (50)
$$

and moreover for any $i = 1, 2, \ldots, n$

$$
P\left(e_i^2 \in \left(r_{(h_n)}^2(\beta^0), r_{(h_n)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\right)_{\mathrm{ord}} \cap B_n\right) \le n^{-\tau}\cdot K^{(1)} \qquad (51)
$$

with $\tau \in \left(\frac{1}{2}, \frac{3}{4}\right)$. In what follows (up to the end of the proof) let us restrict ourselves, without repeating it, in all considerations on the set B_n. It means that we shall analyze e. g. instead of (49)

$$
\left[I\Big\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)}) \le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\Big\} - I\Big\{r_i^2(\beta^0) \le r_{(h)}^2(\beta^0)\Big\}\right] I_{B_n},
$$

etc. Let us recall that the difference in (49) is equal to one if either

$$-\sqrt{r^2_{(h_n)}(\hat{\beta}^{(\mathrm{LTS},n,h)})} + x_i^{\mathrm{T}}\left(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0\right) \le e_i < -\sqrt{r^2_{(h_n)}(\beta^0)} \tag{52}$$

or

$$\sqrt{r^2_{(h_n)}(\beta^0)} < e_i \le \sqrt{r^2_{(h_n)}(\hat{\beta}^{(\mathrm{LTS},n,h)})} + x_i^{\mathrm{T}}\left(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0\right) \tag{53}$$

and is equal to minus one if

$$-\sqrt{r^2_{(h_n)}(\beta^0)} \le e_i < -\sqrt{r^2_{(h_n)}(\hat{\beta}^{(\mathrm{LTS},n,h)})} + x_i^{\mathrm{T}}\left(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0\right) \tag{54}$$

or

$$\sqrt{r^2_{(h_n)}(\hat{\beta}^{(\mathrm{LTS},n,h)})} + x_i^{\mathrm{T}}\left(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0\right) < e_i \le \sqrt{r^2_{(h_n)}(\beta^0)}. \tag{55}$$

Now, let us observe that from (51) it follows that whenever $x_i^{\mathrm{T}}(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0) \ge 0$ the probability of the event in (52) as well as of event in (55) is not larger than $n^{-\tau} \cdot K^{(1)}$ (please, keep in mind that in fact we speak about the event given by (52) intersected with B_n as well as about the event given by (55) intersected with B_n). Similarly, when $x_i^{\mathrm{T}}(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0) < 0$ the same is true about the events in (53) and in (54). Let us denote the event when $\omega \in B_n$, $x_i^{\mathrm{T}}(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0) \ge 0$, e_i falls into one of the intervals (53) or (54) by $C_{in}^{(+)}$ and similarly when $\omega \in B_n$, $x_i^{\mathrm{T}}(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0) < 0$, e_i falls in one of the intervals (52) or (55) by $C_{in}^{(-)}$ and finally when $\omega \in B_n$ and either $x_i^{\mathrm{T}}(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0) \ge 0$ and e_i falls in one of the intervals (52) or (55), or $x_i^{\mathrm{T}}(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0) < 0$ and e_i falls in one of (53) or (54) by D_{in} (i. e. $D_{in} = B_n \setminus (C_{in}^{(+)} \cup C_{in}^{(-)})$). Let us recall that we have assumed that there is a finite upper bound of density and that we have shown in the proof of Theorem 1 of Part II that $\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0$ is symmetrically distributed, i.e. $P(x_i^{\mathrm{T}}(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0) \ge 0) = P(x_i^{\mathrm{T}}(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0) < 0) = \frac{1}{2}$. Then we conclude that there is a constant $K^{(2)} \in (K^{(1)}, \infty)$ so that for all $n > n_1$ (for n_1 see the beginning of the proof)

$$P\left(D_{in}\right) < n^{-\tau} \cdot K^{(2)}, \tag{56}$$

$$P\left(C_{in}^{(+)}\right) = f(u_\alpha) \cdot x_i^{\mathrm{T}}\left(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0\right) + \nu_i^{(1)} + \eta_i^{(1)} \tag{57}$$

and

$$P\left(C_{in}^{(-)}\right) = -f(u_\alpha) \cdot x_i^{\mathrm{T}}\left(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0\right) + \nu_i^{(2)} + \eta_i^{(2)} \tag{58}$$

(remember that $x_i^{\mathrm{T}}\left(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0\right) < 0$ on $C_{in}^{(-)}$) where

$$|\nu_i^{(j)}| < n^{-\tau} \cdot K^{(2)} \qquad \text{and} \qquad |\eta_i^{(j)}| < n^{-1} \cdot K^{(2)}\left(1 + \|x_i\|^2\right) \quad \text{for } j = 1, 2. \tag{59}$$

The consequence of (56) is that

$$\mathsf{E}\left| I\left\{r_i^2(\hat{\beta}^{(\mathrm{LTS},n,h)}) \le r^2_{(h)}(\hat{\beta}^{(\mathrm{LTS},n,h)})\right\} - I\left\{r_i^2(\beta^0) \le r^2_{(h)}(\beta^0)\right\}\right| I_{D_{in}} < n^{-\tau} \cdot K^{(2)}. \tag{60}$$

Now, let us write the last term of (**??**) as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{e_i x_i\left[I\left\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)})\le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\right\}-I\left\{r_i^2(\beta^0)\le r_{(h)}^2(\beta^0)\right\}\right]I_{D_{in}}\right\} \tag{61}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{e_i x_i\left[I\left\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)})\le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\right\}-I\left\{r_i^2(\beta^0)\le r_{(h)}^2(\beta^0)\right\}\right]I_{C_{in}^{(+)}}\right\} \tag{62}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{e_i x_i\left[I\left\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)})\le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\right\}-I\left\{r_i^2(\beta^0)\le r_{(h)}^2(\beta^0)\right\}\right]I_{C_{in}^{(-)}}\right\}. \tag{63}$$

Prior to continuing, let us realize that whenever $\omega \in B_n \cap D_{in}$, we have $e_i^2 \in (r_{(h)}^2(\beta^0), r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)}))_{\mathrm{ord}}$. Then we can show – utilizing both inequalities in (50) – that there is a constant $K^{(3)} \in (1,\infty)$ so that for $n > n_1$

$$|e_i| \le K^{(3)}.$$

Let us find $n_2 \in N$ so that for all $n > n_2$ we have $n^{-(\frac{1}{2}+\tau)}\cdot K^{(2)}\cdot K^{(3)}\cdot\sum_{i=1}^{n}\|x_i\| < \varepsilon^2$ (please keep in mind that $\tau \in \left(\frac{1}{2},\frac{3}{4}\right)$). Then (60) immediately implies that

$$\begin{aligned}
&P\left(\frac{1}{\sqrt{n}}\left\|\sum_{i=1}^{n}\left\{e_i x_i\left[I\left\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)})\le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\right\}\right.\right.\right.\right.\\
&\qquad\left.\left.\left.\left.-I\left\{r_i^2(\beta^0)\le r_{(h)}^2(\beta^0)\right\}\right]I_{D_{in}}\right\}\right\|>\varepsilon\right)\\
\le\ &P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\|x_i\|\,|e_i|\cdot\left|I\left\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)})\le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\right\}\right.\right.\\
&\qquad\left.\left.-I\left\{r_i^2(\beta^0)\le r_{(h)}^2(\beta^0)\right\}\right|I_{D_{in}}>\varepsilon\right)\\
\le\ &\frac{1}{\sqrt{n}}\varepsilon^{-1}\sum_{i=1}^{n}\|x_i\|\,K^{(3)}\cdot\mathsf{E}\left\{\left|I\left\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)})\le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\right\}\right.\right.\\
&\qquad\left.\left.-I\left\{r_i^2(\beta^0)\le r_{(h)}^2(\beta^0)\right\}\right|I_{D_{in}}\right\}\\
\le\ &2n^{-(\frac{1}{2}+\tau)}\varepsilon^{-1}\cdot K^{(2)}\cdot K^{(3)}\cdot\sum_{i=1}^{n}\|x_i\|<\varepsilon.
\end{aligned} \tag{64}$$

So it suffices to study (62) and (63). Let us start with (62). Notice that if e_i falls into the interval in (53), the difference of the indicators is equal to one and $e_i \ge 0$ and vice versa, if e_i falls into the interval in (54), the difference of the indicators is equal to minus one and $e_i < 0$. It means that (62) can be written as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{|e_i|\cdot x_i\cdot\left|I\left\{r_i^2(\hat\beta^{(\mathrm{LTS},n,h)})\le r_{(h)}^2(\hat\beta^{(\mathrm{LTS},n,h)})\right\}-I\left\{r_i^2(\beta^0)\le r_{(h)}^2(\beta^0)\right\}\right|I_{C_n^{(+)}}\right\}.$$

Moreover, recalling once again that $e^2_{(h_n)} = r^2_{(h_n)}(\beta^0)$, we observe that whenever $I_{C_n^{(+)}} = 1$ and $n > n_1$, (50) and (51) imply that

$$||e_i| - u_\alpha| < \left(n^{-\frac{1}{2}} + n^{-\tau}\right) \cdot K^{(1)} \le 2 \cdot n^{-\frac{1}{2}} \cdot K^{(1)}$$

and hence, due to (57) and (58),

$$\begin{aligned}
&P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ ||e_i| - u_\alpha| \cdot \|x_i\| \cdot \left| I\left\{ r_i^2(\hat\beta^{(\mathrm{LTS},n,h)}) \le r^2_{(h)}(\hat\beta^{(\mathrm{LTS},n,h)}) \right\} \right.\right.\right. \\
&\qquad \left.\left.\left. - I\left\{ r_i^2(\beta^0) \le r^2_{(h)}(\beta^0) \right\} \right| I\left\{ C_n^{(+)} \right\} \right\} > \varepsilon \right) \\
&\le \varepsilon^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathsf{E} \left\{ ||e_i| - u_\alpha| \cdot \|x_i\| \cdot \left| I\left\{ r_i^2(\hat\beta^{(\mathrm{LTS},n,h)}) \le r^2_{(h)}(\hat\beta^{(\mathrm{LTS},n,h)}) \right\} \right.\right. \\
&\qquad \left.\left. - I\left\{ r_i^2(\beta^0) \le r^2_{(h)}(\beta^0) \right\} \right| I\left\{ C_n^{(+)} \right\} \right\} \\
&\le 2\varepsilon^{-1} \cdot n^{-1} \cdot K^{(1)} \sum_{i=1}^{n} \|x_i\| \cdot \left\{ f(u_\alpha) \cdot \left[x_i^{\mathrm{T}} \left(\hat\beta^{(\mathrm{LTS},n,h)} - \beta^0 \right) + \nu_i^{(1)} \right] + \eta_i^{(1)} \right\}. \qquad (65)
\end{aligned}$$

Employing now (50) and (59), we find the upper bound of (65) as

$$8\varepsilon^{-1} \cdot n^{-\frac{3}{2}} \cdot f(u_\alpha) \cdot \left[K^{(1)}\right]^2 \sum_{i=1}^{n} \|x_i\|^2 \left(1 + \|x_i\|^2\right).$$

So we can consider instead of (62) the sum

$$\begin{aligned}
&\frac{u_\alpha}{\sqrt{n}} \sum_{i=1}^{n} \left\{ x_i \cdot \left[\left| I\left\{ r_i^2(\hat\beta^{(\mathrm{LTS},n,h)}) \le r^2_{(h)}(\hat\beta^{(\mathrm{LTS},n,h)}) \right\} \right.\right.\right. \\
&\qquad \left. - I\left\{ r_i^2(\beta^0) \le r^2_{(h)}(\beta^0) \right\} \right| I\left\{ C_n^{(+)} \right\} \\
&- \mathsf{E} \left| I\left\{ r_i^2(\hat\beta^{(\mathrm{LTS},n,h)}) \le r^2_{(h)}(\hat\beta^{(\mathrm{LTS},n,h)}) \right\} \right. \\
&\qquad \left.\left.\left. - I\left\{ r_i^2(\beta^0) \le r^2_{(h)}(\beta^0) \right\} \right| I\left\{ C_n^{(+)} \right\} \right] \right\} \qquad (66) \\
&+ \frac{u_\alpha}{\sqrt{n}} \sum_{i=1}^{n} \left\{ x_i \cdot \mathsf{E} \left| I\left\{ r_i^2(\hat\beta^{(\mathrm{LTS},n,h)}) \le r^2_{(h)}(\hat\beta^{(\mathrm{LTS},n,h)}) \right\} \right.\right. \\
&\qquad \left.\left. - I\left\{ r_i^2(\beta^0) \le r^2_{(h)}(\beta^0) \right\} \right| I\left\{ C_n^{(+)} \right\} \right\}. \qquad (67)
\end{aligned}$$

We are going to utilize Lemma 2. Due to (50) and (51) we can put $u = r^2_{(h_n)}(\beta^0) - u_\alpha$, for a $\tau \in (\frac{1}{2}, \frac{3}{4})$, $v = n^\tau (r^2_{(h_n)}(\hat\beta^{(\mathrm{LTS},n,h)}) - r^2_{(h_n)}(\beta^0))$ and $t = n^{\frac{1}{2}}(\hat\beta^{(\mathrm{LTS},n,h)} - \beta^0)$. Plugging these values into (42), we obtain from (43) that (66) is $o_p(1)$. Let us turn to (67). In a similar way as we derived (57) we obtain

$$x_i \cdot \mathsf{E} \left| I\left\{ r_i^2(\hat\beta^{(\mathrm{LTS},n,h)}) \le r^2_{(h)}(\hat\beta^{(\mathrm{LTS},n,h)}) \right\} - I\left\{ r_i^2(\beta^0) \le r^2_{(h)}(\beta^0) \right\} \right| I\left\{ C_n^{(+)} \right\}$$

$$= f(u_\alpha) \cdot x_i \cdot x_i^{\mathrm{T}} \left(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0 \right) + o(n^{-\tau}).$$

Since (63) may be treated along similar lines we finally arrive at

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[e_i x_i \cdot I\{e_i^2 \le u_\alpha^2\} \right]$$
$$= \quad \{Q_n \cdot [(1-\alpha) - 2 \cdot u_\alpha f(u_\alpha)]\} \cdot \sqrt{n} \left(\hat{\beta}^{(\mathrm{LTS},n,h)} - \beta^0 \right) + o_p(1).$$

It concludes the proof. □

CONCLUSIONS

The paper, in Parts I, II and III, offers proofs of consistency and asymptotic normality of $\hat{\beta}^{(\mathrm{LTS},n,h)}$. It also gives the asymptotic representation of the estimator[1]. Due to the already enormous length of paper it was not possible to include the result on the sensitivity of $\hat{\beta}^{(\mathrm{LTS},n,h)}$ (in the sense of finding asymptotic representation for the difference $\hat{\beta}^{(\mathrm{LTS},n,h)} - \hat{\beta}^{(\mathrm{LTS},n-1,h,\ell)}$ or $\hat{\beta}^{(\mathrm{LTS},n,h)} - \hat{\beta}^{(\mathrm{LTS},n-1,I_{k_n})}$ where $\hat{\beta}^{(\mathrm{LTS},n-1,h,\ell)}$ denote the estimate for the data from which the ℓth observation was deleted and similarly $\hat{\beta}^{(\mathrm{LTS},n-1,I_{k_n})}$ denote the estimate for data from which the group of observations – given by a set of indices I_k – was again deleted, see Víšek [21] or [28]; the representation of $\hat{\beta}^{(\mathrm{LTS},n,h)} - \hat{\beta}^{(\mathrm{LTS},n-1,h,\ell)}$ can be found in [32]; for an extensive discussion see also Chatterjee and Hadi [4] or Zvára [33]). The result describing sensitivity of $\hat{\beta}^{(\mathrm{LTS},n,h)}$ with respect to deletion of an observation, in the framework with random carriers, can be found in Víšek [23]. It indicates that the difference $\hat{\beta}^{(\mathrm{LTS},n,h)} - \hat{\beta}^{(\mathrm{LTS},n-1,h,\ell)}$ can be large. A proposal removing this disadvantage was given in Víšek [27]. It was called there *the least weighted squares* and defined as

$$\hat{\beta}^{(LWS,n,w)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} w_i r_{(i)}^2(\beta) \tag{68}$$

(later it appeared under different name e. g. in Čížek [5]). The consistency, asymptotic normality and Bahadur representation for the framework with random carriers (which appeared to be easier to deal with than the framework with deterministic carriers) can be found in Víšek [29, 30].

As we have seen the proofs of the $\sqrt{n}$-consistency and asymptotic normality of $\hat{\beta}^{(\mathrm{LTS},n,h)}$ are not short and easy. On the other hand, the estimator is (highly) nonlinear and hence we cannot expect the proofs to be very short. Nevertheless, the proofs consist of (sometimes rather long) chains of simple steps. May be that the Skorohod embedding into Wiener process and the application of invariance principle can be assumed as an exception from it. However, the embedding of binary random variables into Wiener process is already becoming to belong among basic

[1] There was of course a proof of consistency and asymptotic normality of $\hat{\beta}^{(\mathrm{LTS},n,h)}$ by Rousseeuw and Leroy [16] for one dimensional case, i. e. for estimating location parameter. They gave also a hint of proof for general regression model, referring to paper by Maronna and Yohai [13]. This paper however seems to be applicable for estimators of the type as M-estimators.

proving technique due to its transparency, constructive character and ability to cope with situations when we need to reach uniformity over some set of parameters, (see Portnoy [15], Jurečková and Sen [7] or Víšek [21, 22], see also [3], Proof of Proposition 13.15). Although some alternative techniques can sometimes serve for the (same) purposes as well, see e. g. Hampel et al. [8], Huber [11], Liese and Vajda [12] or Pollard [14], the papers by Víšek [21] and [28] demonstrated that this technique can offer something more, namely it is able to carry out the sensitivity studies of the estimator in question.

The least trimmed squares are utilized not only as the estimator of regression coefficients (or "preliminary" estimator) but also as an efficient diagnostic tool, especially with h near to the assumed level of contamination (see Rubio and Víšek [17]). In fact, if we star with $h = \left[\frac{n}{2}\right] + \left[\frac{p+1}{2}\right]$ – see Rousseeuw and Leroy [16] – and increase in every step h by one, it may happen (and it frequently does happen) that we arrive into an interval of h's for which corresponding subpopulations of "original" data are nested, the estimates of regression coefficients are stable and the estimate of variance of disturbances increases only slowly. When we however "overcome" some size of subpopulation, we notice that the estimate of regression model changes a lot and the estimate of variance of disturbances jumpes up. It is easy to imagine in such a case that the data probably consist of two (or more), hopefully homogeneous, subpolulations. For an example of economic application leading to a decomposition of data on two subpopulations, each of them allowing a reasonable modeling see Benáček et al. [1] or Víšek [24]. In these papers the determinants of the export of goods produced by the Czech economy and of the foreigner direct investments into the Czech republic were studied. The results discovered that there are already some industries of economy which behave as industries in market economies, namely that they follow Cobb–Douglas production function. Unfortunately (in 1994), a large part of the Czech economy still behaved like centrally planned economy. From the economic point of view, it may seem completely strange but the (normed) labor and the (normed) capital are in direct linear dependence in this part of the Czech economy (although we are used that in market oriented economy they are, at least partially, substitutable).

In the past there were objections that the estimators of this type, as the least trimmed squares, require extensive computation. Moreover, we typically obtain "only" an approximation to the value of estimate and what is even worse, we are not able directly to verify, how good approximation we obtained. So it is claimed that these estimators are automatically disqualified for any practical goals. It is not already true, at least not for the least trimmed squares. There is a simple algorithm which can be characterized as an improved algorithm of *repeated selection of subsamples.* Although a simple version of the *repeated selection of subsamples* appeared to be unreliable (see Hettmansperger and Sheather [10] together with [19]), the improved form appeared to be reliable and gives very good approximation to the estimate (see Víšek [20] and [26]). The reliability of the algorithm was confirmed by mutual comparison with the results of an algorithm for the least median of squares $\hat{\beta}^{(\mathrm{LMS},n,h)}$ which is due to Boček and Lachout [2]. In Víšek [20] and [26] we used data of such size that inspection of all subsets of required size was possible and hence $\hat{\beta}^{(\mathrm{LTS},n,h)}$ was evaluated precisely. Nevertheless, even then the algorithm by Boček and Lachout gave smaller hth order statistics among the squared residuals than $\hat{\beta}^{(\mathrm{LTS},n,h)}$ (while the latter gave of course smaller sum of h smallest squared residuals than $\hat{\beta}^{(\mathrm{LMS},n,h)}$). Then we used both algorithms on several (nowadays already) "classical" data sets. For all of them $\hat{\beta}^{(\mathrm{LTS},n,h)}$ gave smaller sum of h smallest squared residuals than the sum of the h smallest squared residuals in model estimated by $\hat{\beta}^{(\mathrm{LMS},n,h)}$ appeared to be. On the other hand, in all cases, i. e. for all models estimated by $\hat{\beta}^{(\mathrm{LMS},n,h)}$, the hth order statistic among squared residuals was smaller than hth order statistic in the models estimated by $\hat{\beta}^{(\mathrm{LTS},n,h)}$. We may say that the estimates were "mutually consistent". The implementation of just described algorithm is available from author on request. Another implementation, in MATLAB is also available from Libor Mašíček – on the same address as present author, and finally there is commertially available implementation in XPLORE, see Čížek and Víšek [5]. A similar algorithm appeared recently in Hawkins and Olive [9].

Moreover, this algorithm allows to make an idea about the complexity of the structure of data. If the approximation is evaluated either in a reasonable time or the most of the evaluated models with small values of sum of squared residuals are similar, we may suppose that the structure of data is (relatively) simple. In an opposite case the data may have rather complicated structure or even they do not allow, for any reasonably large subpopulation, a linear regression model at all. In other words, in the former case there is e. g. a "main cloud" of data (which we may consider as proper data) and a rest which is clearly separated from that "main cloud". The rest may be then assumed to be a contamination.

APPENDIX

([18], page 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 $\mathsf{E}\xi = 0$. 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, $\mathsf{E}\tau = a \cdot b = var\,\xi$.

Remark A.1. Since the book of Štěpán [18] is in Czech language we refer also to Breiman [3] 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 (page 277) of Breiman's book. (See also Theorem 13.6, page 276.)

Assertion A1. Let ζ_1 and ζ_2 be (mutually) independent random variables and $u > 0$. Then $\zeta_1 \cdot I\{|\zeta_1| < u\}$ and $\zeta_2 I\{|\zeta_2| < u\}$ are again independent random variables.

Proof is a straightforward computation. Let a_1 and a_2 be real numbers. Then

$$P(\zeta_1 \cdot I\{|\zeta_1| < u\} \leq a_1, \zeta_2 \cdot I\{|\zeta_2| < u\} \leq a_2)$$

$$= P(-u \le \zeta_1 \le \min\{a_1, u\}, -u \le \zeta_2 \le \min\{a_2, u\})$$
$$= P(-u \le \zeta_1 \cdot \le \min\{a_1, u\}) \cdot P(-u \le \zeta_2 \cdot \le \min\{a_2, u\})$$
$$= P(\zeta_1 \cdot I\{|\zeta_1| < u\} \le a_1) \cdot P(\zeta_2 \cdot I\{|\zeta_2| < u\} \le a_2). \qquad \square$$

ACKNOWLEDGEMENT

Research was supported by by the Grant Agency of the Czech Republic under Grant 402/06/0408.

(Received May 7, 2004.)

REFERENCES

[1] V. J. Benáček, M. Jarolím, and J. Á. Víšek: Supply-side characteristics and the industrial structure of Czech foreign trade. In: Proc. Business and Economic Development in Central and Eastern Erupe: Implications for Economic Integration into wider Europe. Technical University in Brno and University of Wisconsin, Whitewater, and the Nottingham Trent University 1998, pp. 51–68.

[2] P. Boček and P. Lachout: Linear programming approach to LMS-estimation. Comput. Statist. Data Anal. (Memorial volume) *19*(1995), 129–134.

[3] L. Breiman: Probability. Addison–Wesley Publishing Company, London 1968.

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

[5] P. Čížek: Robust estimation with discrete explanatory variables. In: COMPSTAT 2003, pp. 509–514.

[6] P. Čížek and J. Á. Víšek: Least trimmed squares. In: EXPLORE, Application Guide (W. Härdle, Z. Hlavka, and S. Klinke, eds.), Springer–Verlag, Berlin 2000, pp. 49–64.

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

[8] 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.

[9] D. M. Hawkins and D. J. Olive: Improved feasible solution algorithms for breakdown estimation. Comput. Statist. Data Anal. *30* (1999), 1, 1–12.

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

[11] P. J. Huber: Robust Statistics. Wiley, New York 1981.

[12] F. Liese and I. Vajda: Consistency of M-estimators in general models. J. Multivar. Anal. *50* (1994), 93–114.

[13] R. A. Maronna and V. J. Yohai: Asymptotic behaviour of general M-estimates for regression and scale with random carriers. Z. Wahrscheinlichkeitstheorie verw. Gebiete *58* (1981), 7–20.

[14] D. Pollard: Asymptotics for least absolute deviation regression estimator. Econometric Theory *7* (1991), 186–199.

[15] 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–Verlag, New York 1983, pp. 231–246.

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

[17] A. M. Rubio and J. Á. Víšek: Estimating the contamination level of data in the framework of linear regression analysis. Qűestiió *21* (1997), 9–36.
[18] J. Štěpán: Teorie pravděpodobnosti (Probability Theory). Academia, Prague 1987.
[19] J. Á. Víšek: A cautionary note on the method of Least Median of Squares reconsidered. In: Trans. Twelfth Prague Conference on Inform. Theory, Statist. Dec. Functions and Random Processes, Prague 1994, pp. 254–259.
[20] J. Á. Víšek: On high breakdown point estimation. Comput. Statistics *11* (1996), 137–146.
[21] J. Á. Víšek: Sensitivity analysis M-estimates. Ann. Inst. Statist. Math. *48* (1996), 469–495.
[22] J. Á. Víšek: Diagnostics of regression subsample stability. Probab. Math. Statist. *17* (1997), 2, 231–257.
[23] J. Á. Víšek: Robust estimation of regression model. Bull. Czech Econometric Society *9* (1999), 57–79.
[24] J. Á. Víšek: The least trimmed squares – random carriers. Bull. Czech Econometric Society *10* (1999), 1–30.
[25] J. Á. Víšek: The robust regression and the experiences from its application on estimation of parameters in a dual economy. In: Proc. Macromodels'99, Rydzyna 1999,pp. 424–445.
[26] J. Á. Víšek: On the diversity of estimates. Comput. Statist. Data Anal. *34* (2000) 67–89.
[27] J. Á. Víšek: Regression with high breakdown point. In: Robust 2000 (J. Antoch and G. Dohnal, eds.), Union of the Czechoslovak Mathematicians and Physicists, Prague 2001, 324–356.
[28] J. Á. Víšek: Sensitivity analysis of M-estimates of nonlinear regression model: Influence of data subsets. Ann. Inst. Statist. Math. *54* (2002), 2, 261–290.
[29] J. Á. Víšek: The least weighted squares I. The asymptotic linearity of normal equation. Bull. Czech Econometric Society *9* (2002), 15, 31–58.
[30] J. Á. Víšek: The least weighted squares II. Consistency and asymptotic normality. Bull. Czech Econometric Society *9* (2002), 16, 1–28.
[31] J. Á. Víšek: Kolmogorov–Smirnov statistics in linear regression. In: Proc. ROBUST 2006, submitted.
[32] J. Á. Víšek: Least trimmed squares – sensitivity study. In: Proc. Prague Stochastics 2006, submitted.
[33] K. Zvára: Regresní analýza (Regression Analysis – in Czech). Academia, Prague 1989.

Jan Ámos Víšek, Department of Macroeconomics and Econometrics, Institute of Economic Studies, Faculty of Social Sciences, Charles University, Smetanovo nábřeží 6, 110 01 Praha 1 and Institute of Information Theory and Automation — Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 09 Praha 8. Czech Republic.
e-mail: visek@mbox.fsv.cuni.cz