KYBERNETIKA — VOLUME 42 (2006), NUMBER 1, PAGES 1–36

THE LEAST TRIMMED SQUARES
Part I: Consistency

Jan Ámos Víšek

The consistency of the least trimmed squares estimator (see Rousseeuw [14] or Hampel et al. [8]) is proved under general conditions. The assumptions employed in paper are discussed in details to clarify the consequences for the applications.

Keywords: robust regression, the least trimmed squares, consistency, discussion of assumptions and of algorithm for evaluation of estimator

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

INTRODUCTION AND NOTATIONS

Although *the least trimmed squares* were introduced among the first robust estimators with high breakdown point (Rousseeuw [14] or Hampel et al. [8]), proofs of their consistency and derivation of the asymptotic representation of Bahadur type and asymptotic normality were only sketched or carried out under special circumstances, usually for simple regression. Moreover, the methods used for it did not allow to study sensitivity of the estimator with respect to the deletion of one or several points (for a result describing it see Víšek [29]). The approach employed here enable us to perform such studies (for the analogy for M-estimators see Víšek [21] and [31]) and, due to the possibility of being applied uniformly for all "cutting levels", it will also allow to prove consistency and asymptotic normality for *the least weighted squares*, for the definition of the least weighted squares see again Víšek [29]. To be able to introduce and discuss relevant problems, let us start with notations.

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 *true* regression coefficients and e_i's represent random fluctuations[1] of Y_i's from the mean value $\mathsf{E}Y_i$. (To be complete, let us add that of course $x_i^{\mathrm{T}}\beta = \sum_{j=1}^{p} x_{ij}\beta_j$.) The word *true* means that for any $n \in N$ data are assumed to be generated by the regression model $Y_i = x_i^{\mathrm{T}}\beta^0 + e_i,\ i = 1, 2, \ldots, n$.

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).

Denoting for any $n \in N$ successively $Y^{(n)} = (Y_1, Y_2, \ldots, Y_n)^{\mathrm{T}}$, $X^{(n)} = (x_1, x_2, \ldots \ldots, x_n)^{\mathrm{T}}$ and $e^{(n)} = (e_1, e_2, \ldots, e_n)^{\mathrm{T}}$ the response variable, the design matrix and disturbances (i. e. the vector of random fluctuations), we can rewrite (1) into the form

$$Y^{(n)} = X^{(n)}\beta^0 + e^{(n)} \tag{2}$$

which will be sometimes more convenient. In what follows we shall omit an indication of the dimension of matrix and of vectors which would presumably unnecessarily burden the notation. The indication of the dimension will be used only in the situations when a misinterpretation may appear. Let us notice that in the case when the intercept is included in the model, the first coordinates of all vectors x_i's are assumed to be equal to 1.

For any $\beta \in R^p$ let us put $r_i(\beta) = Y_i - x_i^{\mathrm{T}}\beta$, i. e. $r_i(\beta)$ denotes the ith residual with respect to $\beta \in R^p$. In other words, we shall consider the residuals as the function of $\beta \in R^p$. We emphasize it because sometimes the residuals are assumed in a restricted sense to be $r_i = Y_i - x_i^{\mathrm{T}}\hat{\beta}$ where $\hat{\beta}$ is the estimator in question. Of course, we have $r_i(\beta) = r_i(\beta, \omega)$, i. e. residuals depend also on ω. Finally the order statistics of squared residuals will be denoted by $r_{(i)}^2(\beta),\ i = 1, 2, \ldots, n$. To be more explicit, it means that we have for any $\beta \in R^p$

$$0 \le r_{(1)}^2(\beta) \le r_{(2)}^2(\beta) \le \cdots \le r_{(n)}^2(\beta).$$

Now, let us make 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) $I\{$*the property describing the set* $A\}$ instead of traditional notation $I_{\{the\ property\ describing\ the\ set\ A\}}$. Finally, let us recall that *the least trimmed squares* estimator is usually given as

$$\tilde{\beta}^{(\mathrm{LTS},n,h)} = \mathop{\arg\min}_{\beta \in R^p} \sum_{i=1}^{h} r_{(i)}^2(\beta) \tag{3}$$

where $\frac{n}{2} \le h \le n$. (It is commonly accepted that when we speak about *the least trimmed squares* estimator we use only *the least trimmed squares*. In what follows we shall do the same.)

It is known (and after all it is straightforward) that by selecting the value of h we can control the level of robustness of estimator, namely its breakdown point. Of

[1]Recently, e_i is frequently called "error term" which however may tempt to the interpretation that e_i is the error of measurement of response variable. Such interpretation is however justifiable only sometimes, e. g. in technical applications (see Van Huffel [19]) but not in economics and social sciences (see Víšek [30]). In econometrics e_i's are usually denoted as *disturbances*.

course, for $h = n$ the estimator $\tilde{\beta}^{(\mathrm{LTS},n,h)}$ coincides with *the ordinary least squares* $\hat{\beta}^{(\mathrm{LS},n)}$.

In the applications we have usually an idea about the magnitude of the regression coefficients. The idea may stem from the framework of given problem, from a common idea about magnitudes of coefficients in similar cases or from the data in question.

It may be formally expressed so that we have an idea about a compact set, the *true* value of regression coefficients is to be inside. (Of course, this set may be assumed to be very large.) As we shall see later, without this assumption we have to impose more restrictive assumptions on the distribution of disturbances. So, 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

$$\hat{\beta}^{(\mathrm{LTS},n,h)} = \arg\min_{\beta \in \mathcal{K}} \sum_{i=1}^{h} r_{(i)}^2(\beta) \tag{4}$$

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

Remark 1. Clearly, $\hat{\beta}^{(\mathrm{LTS},n,h)}$ depends on $\mathcal{K}$. Nevertheless, we shall prove that $\hat{\beta}^{(\mathrm{LTS},n,h)}$ is, for any $\mathcal{K}$, consistent and hence asymptotically "nearly" independent of the set $\mathcal{K}$. Moreover, as we have already said, we (at least implicitly) assume that $\mathcal{K}$ is "very large" and hence it does not influence $\hat{\beta}^{(\mathrm{LTS},n,h)}$ too much or not at all. That is why $\mathcal{K}$ did not appear in notation used for *the least trimmed squares.*

It is clear that for given i the squared residual appears in the sum on the right hand side of (4) iff $r_i^2(\beta) \le r_{(h)}^2(\beta)$, so that we can write equivalently

$$\begin{aligned}\hat{\beta}^{(\mathrm{LTS},n,h)} &= \arg\min_{\beta \in \mathcal{K}} \sum_{i=1}^{n} r_i^2(\beta) \cdot I\left\{ r_i^2(\beta) \le r_{(h)}^2(\beta) \right\} \\ &= \arg\min_{\beta \in \mathcal{K}} \sum_{i=1}^{n} (Y_i - x_i^{\mathrm{T}}\beta)^2 \cdot I\left\{ r_i^2(\beta) \le r_{(h)}^2(\beta) \right\}.\end{aligned} \tag{5}$$

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.

$$P(e_1^2 > u_\alpha^2) = 1 - G(u_\alpha^2) = \alpha. \tag{6}$$

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

$$h_n = [(1-\alpha)n]_{\mathrm{int}}. \tag{7}$$

Moreover, for any $a, b \in R$ we shall denote $(a, b)_{\text{ord}} = (\min\{a, b\}, \max\{a, b\})$ 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,

$$\sum_{i=1}^n \|x_i\|^4 = \mathcal{O}(n) \quad \text{and} \quad \mathsf{E}e_1^4 = \kappa_4 \in (0, \infty). \tag{8}$$

Finally,

$$\lim_{n \to \infty} Q_n = Q \tag{9}$$

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

Remark 2. From Assumptions $\mathcal{A}$ it follows that for any fix $\alpha \in (0, \frac{1}{2})$

$$\mathsf{E}\left[e_1 I\left\{e_1^2 \leq u_\alpha^2\right\}\right] = 0. \tag{10}$$

More generally, for any set, say $B(e_1^2)$, which is determined by e_1^2, we have $\mathsf{E}e_1 I\{B(e_1^2)\} = 0$, since $B(e_1^2)$ is symmetric around zero.

Remark 3. It will be clear from what follows that the assumption of positivity of the density $f(z)$ on $(-\infty, \infty)$ is only the technical assumption. In fact, what we need in proofs is that the distribution function is strictly increasing from zero to one on one interval, say $(-a, a)$, so that the density is (strictly) positive on $(-a, a)$ (where a can be infinite). From it then follows that the α-quantile (for any $\alpha \in [0, 1]$) is continuous in $\alpha \in (0, 1)$.

Remark 4. It follows from Assumptions $\mathcal{A}$ that we shall consider the setup with nonrandom explanatory variables (or carriers, if you want). The theory for the setup with random explanatory variables requires some modifications what concerns the assumptions (*orthogonality and sphericality conditions*) as well as what concerns some steps in the proofs, see Víšek [26]. As it can be seen from Víšek [23, 24, 26] some proofs are simpler for random-carriers-framework, some are more complicated. In the case of the least trimmed squares the proof in the framework with random carriers appeared to be simpler than the proof for determinstic explanatory variables.

The absolute continuity of F seems at a first glance rather strong assumption. Nevertheless, let us realize that for *the (ordinary) least squares* we usually assume *(strict)* normality of disturbances since otherwise *the (ordinary) least squares* are optimal[2] only in *the class of linear estimators.* In other words, if we do not ask for normality of disturbances, out of class of linear estimators we can meet estimators which are better than the least squares. Since the restriction on the class of linear estimators may be, at least, impractical, if not drastic (for more details see Hampel et al. [8]), we should assume that the disturbances are normally distributed. But the assumption of normality is (much) stronger than the assumption of absolute continuity.[3]

We shall see later that $\tilde{\beta}^{(\mathrm{LTS},n,h)}$ coincides with *the ordinary least squares* $\hat{\beta}^{(\mathrm{LS},h)}$ applied on appropriate subsample (of size h) of data. Taking into account that $\hat{\beta}^{(\mathrm{LTS},n,h)}$ asymptotically coincides with $\tilde{\beta}^{(\mathrm{LTS},n,h)}$, we conclude that the normality of residuals is plausible for $\hat{\beta}^{(\mathrm{LTS},n,h)}$ due to the same reasons as for *the ordinary least squares.* From this standpoint the absolute continuity of F does not seem to be (very) restrictive.

Further, any study of order statistics assumes the absolute continuity of the underlying distribution, since without this assumption we have got into some technical troubles as the probability that two order statistics attain the same value need not be zero. The same appears in our study.

Let us turn our attention to the assumption that the density is bounded and has bounded in absolute value the first and second derivative everywhere. As we shall see in the next, we need to estimate the probability (for $t \in R^p$, $\|t\| < M$)

$$P\left(e_i \in (u_\alpha, u_\alpha + n^{-\frac{1}{2}} x_i^{\mathrm{T}} t)_{\mathrm{ord}}\right) \tag{11}$$

Then it is clear that some assumptions on $\|x_i\|$ and on $F(z)$ are (probably) necessary (or even inevitable?). If we assume that for some $K < \infty$

$$\sup_{i \in N} \|x_i\| < K, \tag{12}$$

it is evidently sufficient to assume existence of bounded derivatives of density in the neighborhood of $-u_\alpha$ and of u_α (and the same is true about the positivity of the density). However, the assumption (12) is sometimes considered as inadmissibly restrictive while the assumptions of type (8) are accepted. Then, i. e. under (8), the norms $\|x_i\|$, $i = 1, 2, \ldots, n$ are not uniformly bounded and hence to be able to achieve the equality $P(e_i \in (u_\alpha, u_\alpha + n^{-\frac{1}{2}} x_i^{\mathrm{T}} t)_{\mathrm{ord}}) = \|x_i\| \mathcal{O}(n^{-\frac{1}{2}})$, we need some assumption(s) about $F(z)$ to be fulfilled on the whole support of $F(z)$. Of course, under (12) as well as under (8), it is possible to estimate probability (11), nevertheless in the former case it is straightforward while in the latter it requires rather involving considerations.

We have included assertion which hints that from the practical point of view, the difference between (8) and (12) need not be considerable, see Lemma A.2. Moreover,

[2] In the sense that they guarantee minimal variance of estimators of regression coefficients.

[3] Of course, in the applications we try to reach the normality of residuals by adding some explanatory variables, transforming them etc. Such approach has of course also some philosophical consequences which are beyond the scope of this paper, see Prigogine & Stengers [12, 13].

the results in Chatterjee, Hadi [4], Zvára [33] or Víšek [21, 22] and [28] indicate that in the case when the norm of some explanatory vectors is out of control, we cannot guarantee anything about subsample sensitivity, i. e. we cannot say how large is the norm of $\hat{\beta}^{(\mathrm{LTS},n,h)} - \hat{\beta}^{(\mathrm{LTS},n-1,h,\ell)}$ where $\hat{\beta}^{(\mathrm{LTS},n-1,h,\ell)}$ denotes *the least trimmed squares* evaluated for data from which the ℓth observation was deleted, see Víšek [31]. That is why we shall also assume an alternative version of assumptions.

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, (9) holds for some regular matrix Q. Further for any $n \in N$

$$\max_{1 \le i \le n,\ 1 \le j \le p} |x_{ij}| = \mathcal{O}(1). \tag{13}$$

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 $\mathsf{E}e_1^4 = \kappa_4 \in (0,\infty)$.

First of all, we shall look for normal equations for $\hat{\beta}^{(\mathrm{LTS},n,h)}$. So denoting

$$\rho(\beta) = \sum_{i=1}^{n} (Y_i - x_i^{\mathrm{T}}\beta)^2 \cdot I\Big\{ r_i^2(\beta) \le r_{(h)}^2(\beta) \Big\}, \tag{14}$$

(see (5)), we shall study

$$\frac{\partial \rho(\beta)}{\partial \beta} = \sum_{i=1}^{n} \Big[-2(Y_i - x_i^{\mathrm{T}}\beta) x_i\, I\Big\{ r_i^2(\beta) \le r_{(h)}^2(\beta) \Big\} + (Y_i - x_i^{\mathrm{T}}\beta)^2 \frac{\partial}{\partial \beta} I\Big\{ r_i^2(\beta) \le r_{(h)}^2(\beta) \Big\} \Big]. \tag{15}$$

Let us find for $j \in \{1, 2, \ldots, p\}$

$$\frac{\partial}{\partial \beta_j} I\Big\{ r_i^2(\beta) \le r_{(h)}^2(\beta) \Big\} = \lim_{\Delta \to 0} \frac{I\Big\{ r_i^2(\beta^{(\Delta,j)}) \le r_{(h)}^2(\beta^{(\Delta,j)}) \Big\} - I\Big\{ r_i^2(\beta) \le r_{(h)}^2(\beta) \Big\}}{\Delta} \tag{16}$$

where

$$\beta^{(\Delta,j)} = (\beta_1, \beta_2, \ldots, \beta_{j-1}, \beta_j + \Delta, \beta_{j+1}, \ldots, \beta_p)^{\mathrm{T}}. \tag{17}$$

Let us realize that the residuals $r_i(\beta)$'s are continuous and linear (and hence monotone) functions of any given coordinate of β provided the other coordinates and ω are kept fixed (ω is not given in the notation explicitly but, as we have already recalled it, the residuals of course depend on it).

For the order statistics $r_{(h)}^2(\beta)$'s the situation is somewhat more complicated[4]. They are also continuous in any coordinate of β. It follows from the fact that

[4] Please, realize that the vector of order statistics, $r_{(i)}^2(\beta), i = 1, 2, \ldots, n$, represents permutation of squared residuals $r_i^2(\beta), i = 1, 2, \ldots, n$ and the order of indices depends on β as well as on ω.

minimum of a finite set of continuous functions is continuous. Moreover, under Assumptions $\mathcal{A}$ or $\mathcal{B}$, the set $A_n = \{\omega \,:\, e_{(1)} < e_{(2)} < \ldots e_{(n)}\}$ has probability equal to 1. As the probability that there is an index, say j_0, such that $e_{(j_0)} = 0$ is equal to 0, also the set $B_n = \{\omega \,:\, e^2_{(1)} < e^2_{(2)} < \ldots e^2_{(n)}\}$ has probability 1. Let us fix an $\omega \in B_n$ and denote by i_0 such index that $e^2_{i_0} = e^2_{(h)} = r^2_{(h)}(\beta^0)$. Then there is $\delta_\omega > 0$ such that for any $|\Delta| < \delta_\omega$ and any $j \in \{1, 2, \ldots, n\}$

$$r^2_{(h)}(\beta^{(\Delta,j)}) = \left(Y_{i_0} - x^{\mathrm{T}}_{i_0}\beta^{(\Delta,j)}\right)^2.$$

In other words, for given $\omega \in B_n$ there is a neighborhood of $\beta \in R^p$ in which the hth order statistic among the squared residuals is represented by "one fix index". We shall need both these facts, the continuity of order statistics and stability of indices, for the analysis of the limit in (16).

What concerns the shape of order statistics of squared residuals (considered as the function of one coordinate of β), it is easy to prove that for any given $\tilde{\beta}$ there is a neighbourhood of $\tilde{\beta}$ so that the given order statistic is either convex or concave or monotone but we shall not need it.

First of all, let us observe that the ratio in the limit (16) will be nonzero only if for all Δ (see (17)) from a neighborhood of zero

$$I\left\{r^2_i(\beta^{(\Delta,j_0)}) \le r^2_{(h)}(\beta^{(\Delta,j_0)})\right\} \ne I\left\{r^2_i(\beta) \le r^2_{(h)}(\beta)\right\}. \tag{18}$$

Now, let us observe that if i is the index for which $r^2_i(\beta) = r^2_{(h)}(\beta)$, (18) does not hold, of course (see the remark a few lines above, saying that hth order statistic among the squared residuals is represented by "one fix index"). In all other cases, i. e. when i is not an index for which $r^2_i(\beta) = r^2_{(h)}(\beta)$, (18) means that (let us repeat, for all Δ from a neighborhood of zero) either $r^2_i(\beta^{(\Delta,j_0)}) \le r^2_{(h)}(\beta^{(\Delta,j_0)})$ and $r^2_i(\beta) > r^2_{(h)}(\beta)$ or $r^2_i(\beta^{(\Delta,j_0)}) > r^2_{(h)}(\beta^{(\Delta,j_0)})$ and $r^2_i(\beta) \le r^2_{(h)}(\beta)$. The first eventuality is impossible at all, because of the continuity of $r^2_i(\beta)$ and of $r^2_{(h)}(\beta)$ in β_{j_0}which implies that there is a $\delta_0 > 0$ so that for all $|\Delta| < \delta_0$ we have $r^2_i(\beta^{(\Delta,j_0)}) > r^2_{(h)}(\beta^{(\Delta,j_0)})$. The second one is possible for some $\Delta > 0$ only if $r^2_i(\beta) = r^2_{(h)}(\beta)$ but this appear only with probability zero (due to the assumption of absolute continuity of underlying distribution function of disturbances). We conclude that $\frac{\partial}{\partial\beta_j} I\{r^2_i(\beta) \le r^2_{(h)}(\beta)\} = 0$ a. e. Since the sum in (15) contains finite number of terms, we have

$$\sum_{i=1}^n \left[(Y_i - x^{\mathrm{T}}_i\beta)^2 \cdot \frac{\partial}{\partial\beta_j} I\left\{r^2_i(\beta) \le r^2_{(h)}(\beta)\right\}\right] = 0 \qquad \text{a. e.}$$

and finally

$$\frac{1}{2}\frac{\partial\rho(\beta)}{\partial\beta} = -\sum_{i=1}^n \left[(Y_i - x^{\mathrm{T}}_i\beta)x_i \cdot I\left\{r^2_i(\beta) \le r^2_{(h)}(\beta)\right\}\right] \qquad \text{a. e.} \tag{19}$$

We shall prove that it means that $\tilde{\beta}^{(\mathrm{LTS},n,h)}$ can be found among solutions of

$$\sum_{i=1}^{n}\left[(Y_i - x_i^{\mathrm{T}}\beta)x_i \cdot I\left\{r_i^2(\beta) \le r_{(h)}^2(\beta)\right\}\right] = 0, \tag{20}$$

i. e. that at the point given as the solution of the extremal problem (3) the relation (20) holds. Notice please that whenever we prove that the estimator given by (4) is consistent (i. e. it exists and converges in probability to β^0), it also solves (20).

It is nearly immediately clear that a solution of the extremal problem (3) exists but we shall clarify it in details since it will hint that $\tilde{\beta}^{(\mathrm{LTS},n,h)}$ can be considered, similarly as $\hat{\beta}^{(\mathrm{LS},n)}$, to be projection of a response variable into the corresponding p-dimensional subset generated by the column of a design matrix.

First of all, let us realize that we have $\binom{n}{h}$ possible subsamples of the size h from a set of size n.

Let us recall (once again) that $r_i(\beta) = r_i(\beta, \omega)$, i. e. that the residuals depend not only on $\beta \in R^p$ but also on $\omega \in \Omega$. Now, let us fix an $\omega_0 \in \Omega$ and an h-tuple of indices $i_1, i_2, \ldots, i_h$. Then, let us apply the ordinary least squares on the data $Y^{(h)} = (Y_{i_1}(\omega_0), Y_{i_2}(\omega_0), \ldots, Y_{i_h}(\omega_0))^{\mathrm{T}}$, $X^{(h)} = (x_{i_1}, x_{i_2}, \ldots, x_{i_h})^{\mathrm{T}}$ and find the estimate of regression coefficients, say $\hat{\beta}^{(\mathrm{LS},h,i_1,i_2,\ldots,i_h)}$. Finally, let us evaluate corresponding sum of squares, say $S(\omega_0, i_1, i_2, \ldots, i_h)$. Keeping ω_0 fixed, we shall repeat this for all h-tuples and we obtain $\binom{n}{h}$ sums of squares of type $S(\omega_0, i_1, i_2, \ldots, i_h)$. Finally we select that h-tuple of indices, for which the corresponding sum of squares is the smallest. It is clear that we have established by this way the value of $\tilde{\beta}^{(\mathrm{LTS},n,h)}$ at ω_0, say $\tilde{\beta}^{(\mathrm{LTS},n,h)}(\omega_0)$. Repeating this for all $\omega \in \Omega$ we obtain the solution of the extremal problem (3). In other words, we have proved assertion:

Assertion 1. There is always a solution of (3).

Now, let us define a mapping $a : \Omega \to \{1, 2, \ldots, n\}^h$ so that $a(\omega) = (i_1, i_2, \ldots, i_h)^{\mathrm{T}}$ is the point of Carthesian product $\{1, 2, \ldots, n\}^h$ for which

$$\min_{\beta \in R^p} \sum_{i \in a(\omega)} (Y_i(\omega) - x_i^{\mathrm{T}}\beta)^2 = \min_{\beta \in R^p} \sum_{i=1}^{h} r_{(i)}^2(\beta, \omega). \tag{21}$$

Further, let us denote by $Y(a) = Y(a(\omega))$ and $X(a) = X(a(\omega))$ corresponding h-dimensional subvector of the vector Y and corresponding submatrix of the matrix X containing all rows of matrix X indices of which fall into the set $a(\omega)$. Finally, denote also by $e(a) = e(a(\omega))$ corresponding subvector of disturbances and by $\hat{\beta}^{(\mathrm{LS},h)}(Y(a), X(a)) = \hat{\beta}^{(\mathrm{LS},h)}(Y(a(\omega)), X(a(\omega)))$ the least squares estimator evaluated just for subpopulation $[Y(a(\omega)), X(a(\omega))]$. Then it is clear that

$$\hat{\beta}^{(\mathrm{LS},h)}(Y(a(\omega)), X(a(\omega))) = \tilde{\beta}^{(\mathrm{LTS},n,h)}(\omega)$$

where we have denoted $\tilde{\beta}^{(\mathrm{LTS},n,h)}(\omega)$ the value of $\tilde{\beta}^{(\mathrm{LTS},n,h)}$ at the point ω. But $\hat{\beta}^{(\mathrm{LS},h)}(Y(a(\omega)), X(a(\omega)))$ is the solution of normal equations

$$X^{\mathrm{T}}(a)X(a)\beta = X^{\mathrm{T}}(a)Y(a)$$

which coincide with the equations given in (20), so that, the equation in (20) has a solution as well. Moreover, we obtain (in the case that $X^{\rm T}(a)X(a)$ is regular, otherwise we have to employ some pseudoinverse)

$$\tilde\beta^{({\rm LTS},n,h)} - \beta^0 = \left(\frac{1}{n}X^{\rm T}(a)X(a)\right)^{-1}\frac{1}{n}X^{\rm T}(a)e(a) \tag{22}$$

which indicates that $\tilde\beta^{({\rm LTS},n,h)}$ can have similar properties as $\hat\beta^{({\rm LS},h)}$.

The snag is that $\binom{n}{h}$ is for large n too large, so that for the practical evaluation this way is not feasible, except of the situations when $n \le 20$. We shall return to the problems with evaluability of $\tilde\beta^{({\rm LTS},n,h)}$ at the Conclusions at the end of Part III of paper.

Prior to considering technicalities, let us stress that $\tilde\beta^{({\rm LTS},n,h)}$ share with other estimators which are based on "geometry" of data, the scale- and regression-equivariance. This represents a substantial distinction from e. g. M-estimators which, in order to achieve the scale- and regression-equivariance, have to utilize the studentized residuals, studentized by an estimator of the scale of disturbances. Due to the fact that this scale estimator, say $\hat\sigma_n$, has to be scale-equivariant and regression-invariant (in order to achieve the scale- and regression-equivariance of the M-estimator $\hat\beta_n^M$, see e. g. Bickel [2] or Jurečková and Sen [7]), it need not be very simple to evaluate $\hat\sigma_n^2$ (although it is not impossible, see again Jurečková and Sen [7] or Víšek [25]) and hence also evaluation of $\hat\beta_n^M$ may be rather complicated.

Another advantage of $\hat\beta^{({\rm LTS},n,h)}$ is that it can serve as preliminary estimator of regression coefficients in the situation when we have no idea about the level of contamination of data and we need to estimate it. Then using this preliminary estimate of regression coefficients, we can estimate density of residuals and employing it we finally estimate contamination level, see Rubio and Víšek [17]. An alternative way to this is to increase successively h by 1 in every step, starting of course with $[\frac{n}{2}]+[\frac{p+1}{2}]$ (see Rousseeuw & Leroy [15]), and to stop this process when the estimate of scale[5] of disturbances begins to increase steeply. For more details about this diagnostic approach see Víšek [27]. More details of the topic we shall offer in the conclusions.

The last but not least, the least trimmed squares have simple realiable algorithm for evaluating the good approximation to the exact value of the estimator. We shall discuss the problem in details at the end of paper, see also Víšek [20] and [28].

Of course, we should admit that the least trimmed squares can be rather sensitive to the deletion of an observation, i. e. the difference between the estimator evaluated for the all n observations and the estimator evaluated after deleting one observation can be rather large (see Víšek [26], consult also Víšek [21] and [31]). We shall also return to the problem at the conclusions.

In the proofs of the next theorems, lemmas and assertions some further notations (for constants, probabilities, sets etc.) will be used. Throughout the paper it will be assumed that they are valid only within given proof.

[5]The scale estimate is of course assumed to be based on residuals with indices in $a(\omega)$.

CONSISTENCY OF LEAST TRIMMED SQUARES

Now we are going to discuss and to prove the consistency of $\hat{\beta}^{(\mathrm{LTS},n,h)}$. The least trimmed squares are based on the same principle as the (ordinary) least squares (OLS) but applied only on a subset of data. That is why the estimator itself is not linear. It implies that we cannot expect that any proof of its consistency will be so simple as for OLS. There are two ways how to study the consistency of the estimator. The first one follows the idea of the uniform law of large numbers (see e. g. Andrews [1]). That is way we shall discuss in details. The second one utilizes the asymptotic linearity of the second derivative of objective function and it proves (only?) an existence of $\sqrt{n}$-consistent solution of equation (20)[6]. For details of this approach see Rubio and Víšek [16].

As we have seen the least trimmed squares select (at every $\omega \in \Omega$) only some subsample of data. Then we can imagine a mixture (or blend or composition or . . . , as you want to call it) of two populations created in such a way that the estimator cannot be consistent. After all, very similar problem was addressed also for M-estimators and it appeared that some assumption on the structure of the points in the factor space seems to be inevitable, see e. g. Liese & Vajda [10]. What we need is a "regular" behavior of the points at which the information about the underlying regression model is available to guarantee the unique solution of the corresponding extremal problem and the consistency of solution, see Lemma 3 bellow. For *the ordinary least squares* the uniqueness is implied by geometry of the problem and the consistency by the assumption (9), see Dhrymes [6]. To see that for $\hat{\beta}^{(\mathrm{LTS},n,h)}$ the assumption (9) need not be sufficient, let as recall that *the ordinary least squares* coincides with $\tilde{\beta}^{(\mathrm{LTS},n,h)}$ for $h = n$. Then (3) can be written as

$$\hat{\beta}^{(\mathrm{LS},n)} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} r_i^2(\beta) \cdot I\left\{ r_i^2(\beta) \le r_{(n)}^2(\beta) \right\} = \underset{\beta \in R^p}{\arg\min} \sum_{i=1}^{n} r_i^2(\beta).$$

So, we observe that for the *the ordinary least squares* $I\{r_i^2(\beta) \le r_{(n)}^2(\beta)\}$ is equal to 1 independently of x_i's and of $\beta \in R^p$ while for *the least trimmed squares* (for $h < n$) $I\{r_i^2(\beta) \le r_{(h)}^2(\beta)\}$ depends on x_i's as well as on $\beta \in R^p$. In other words, it seems that for *the ordinary least squares* it is sufficient to have an assumption on something like the "second moments" of x_i's while for *the least trimmed squares* we need to know something more. A reasonable possibility seems to be to assume that $\{x_i\}_{i=1}^{\infty}$ resembles sampling from a distribution. So, let us consider for any $x \in R^p$

$$\frac{1}{n} \sum_{i=1}^{n} I\left\{ x_i \le x \right\} \tag{23}$$

where of course we assume the inequality to be fulfilled coordinatewise[7]. Interpreting (23) for a while as the empirical distribution function of $x_1, x_2, \ldots, x_n$, i. e.

[6] The question mark behind the word *only* indicates following. Taking into account that we look for $\hat{\beta}^{(\mathrm{LTS},n,h)}$ by an iterative resampling algorithm (see Rousseeuw, Leroy [15] or Víšek [20, 28]), we cannot claim that we have at hand really $\hat{\beta}^{(\mathrm{LTS},n,h)}$ but only better or worse approximation to $\hat{\beta}^{(\mathrm{LTS},n,h)}$ (given as solution of (4)). Then, from the practical point of view, in fact both approaches give the same justification for employing the estimator.

[7] Of course, one can conjecture that (9) already implies that the sums in (23) stabilize for $n \to \infty$.

considering $x_1, x_2, \ldots, x_n$ to be a sampling from a hypothetical distribution, we may assume our deterministic sequence to be selected so that the sum in (23) converges to a distribution function $H(x)$ (say) at the all points of its continuity.

So, let us assume, for a while, that x_i's are i.i.d. random vectors. Then

$$\sup_{x \in R^p} \left| \frac{1}{n} \sum_{i=1}^{n} I\{x_i \le x\} - H(x) \right| = \mathcal{O}_p(n^{-\frac{1}{2}}), \tag{24}$$

see e. g. Csőrgő and Révész [5]. Moreover, if $\mathcal{X} \in R^p$ be a random variable with distribution function $H(x)$, then, denoting for any $\beta \in R^p$ $H^{(\beta)}(t)$, $t \in R$ the distribution function of the random variable $\mathcal{X}^{\mathrm{T}}(\beta - \beta^0)$, we have

$$\sup_{t \in R} \left| \frac{1}{n} \sum_{i=1}^{n} I\left\{x_i^{\mathrm{T}}(\beta - \beta^0) \le t\right\} - H^{(\beta)}(t) \right| = \mathcal{O}_p(n^{-\frac{1}{2}}). \tag{25}$$

It hints that in the case when x_i's are deterministic, it may seem adequate to ask for the same (or an analogous) behaviour as given in (25).

Moreover, we have already recalled (when discussing Assumptions $\mathcal{A}$ and $\mathcal{B}$) that the assumption (8) implies that, except of an arbitrary small fraction of observations, all x_i's fall into a compact set. All these arguments support the idea that the below given Assumptions $\mathcal{C}$ can be reasonable.

Now, for any $\beta \in R^p$ and any $\delta \in (0,1)$ put $B(\beta, \delta) = \left\{\tilde{\beta} \in R^p \ : \ \|\tilde{\beta} - \beta\| < \delta\right\}$, i. e. $B(\beta, \delta)$ is a ball with the center β and radius δ.

Assumptions $\mathcal{C}$

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

$$\sup_{\beta \in \mathcal{W}} \sup_{t \in R} \left| \frac{1}{n} \sum_{i=1}^{n} I\left\{x_i^{\mathrm{T}}(\beta - \beta^0) \le t\right\} - H^{(\beta)}(t) \right| = \mathcal{O}(n^{-\frac{1}{2}}). \tag{26}$$

Remark 5. Recently it was found that when X_i's are i.i.d. the first supremum in (26) can be taken over R^p, see Víšek (2005).

Now we are going to give an assertion. Although it is (very) simple, since the proof of it is just a straightforward computation, we shall need it later.

However, that is not true. It is sufficient to consider two sequences of i.i.d. random variables, say $\{X_i^{(1)}\}_{i=1}^{\infty}$ and $\{X_i^{(2)}\}_{i=1}^{\infty}$ with $\mathsf{E}X_1^{(1)} = \mathsf{E}X_1^{(2)}$ and $\mathrm{var}X_1^{(1)} = \mathrm{var}X_1^{(2)}$ but with $F_{X_1^{(1)}} \neq F_{X_1^{(2)}}$. Then there is a set of probability 1 on which both empirical distribution functions converge to the theoretical ones. Selecting a point ω_0 of this set and putting $x_i = X_1^{(1)}(\omega_0)$ for some indices and $x_i = X_1^{(2)}(\omega_0)$ for others we conclude the counterexample.

Assertion 2. Let Assumptions $\mathcal{A}$ or $\mathcal{B}$ and $\mathcal{C}$ be fulfilled and let $\mathcal{X} \in R^p$ be a random variable with distribution function $H(x)$. Moreover, let $\mathcal{X}$ be independent from the sequence $\{e_i\}_{i=1}^{\infty}$ and $H_e^{(\beta)}(z)$ denote the distribution function of $e_1 - \mathcal{X}^{\mathrm{T}}(\beta - \beta^0)$. Then for any compact set $\mathcal{W}$

$$\sup_{\beta \in \mathcal{W}} \sup_{z \in R} \left| \frac{1}{n} \sum_{i=1}^{n} I\left\{r_i(\beta) \leq z\right\} - H_e^{(\beta)}(z) \right| = \mathcal{O}_p(n^{-\frac{1}{2}}) \tag{27}$$

where, of course, $r_i(\beta) = e_i - x_i^{\mathrm{T}}(\beta - \beta^0)$. Moreover, the distribution $H_e^{(\beta)}(z)$ is absolutely continuous and strictly increasing.

For any compact set $\mathcal{W}$ and any $\delta > 0$ such that $\mathcal{W} \setminus B(\beta^0, \delta) \neq \emptyset$ there is $\gamma_\delta > 0$ and $n_0 \in N$ such that for all $n > n_0$

$$\inf_{\beta \in \mathcal{W} \setminus B(\beta^0, \delta)} \frac{1}{n} \sum_{i=1}^{n} \left[x_i^{\mathrm{T}}(\beta - \beta^0)\right]^2 > \gamma_\delta. \tag{28}$$

Proof. The assertion (27) follows from the fact that $r_i(\beta) = e_i - x_i^{\mathrm{T}}(\beta - \beta^0)$ and from that we have assumed that $\mathcal{X}$ is independent from the sequence $\{e_i\}_{i=1}^{\infty}$. Then also $\mathcal{X}^{\mathrm{T}}(\beta - \beta^0)$ for $\beta \neq \beta^0$ is independent from the sequence $\{e_i\}_{i=1}^{\infty}$.

Finally, the absolute continuity and strict monotonicity of $H_e^{(\beta)}(z)$ is due to the same properties of $F(z)$ (the distribution of disturbances).

The second part of the assertion follows from the fact that Q is regular and hence positive definite. It implies that

$$\tau_\delta = \min_{\beta \in \mathcal{W} \setminus B(\beta^0, \delta)} (\beta - \beta^0)^{\mathrm{T}} Q (\beta - \beta^0) > 0.$$

Notice that due to the fact that the set $\mathcal{W} \setminus B(\beta^0, \delta)$ is compact we can write *min* instead of *inf* which in turn is the reason why the *min* is strictly positive. Then putting

$$d = \max_{\beta, \tilde{\beta} \in \mathcal{W}} \left\| \beta - \tilde{\beta} \right\|, \tag{29}$$

let us find $n_0 \in N$ so that

$$\|Q - \frac{1}{n} X^{\mathrm{T}} X\| = \max_{\ell, j \in \{1, 2, \ldots, p\}} |q_{\ell j} - \frac{1}{n} \sum_{i=1}^{n} x_{i\ell} x_{ij}| < \tfrac{1}{2} \tau_\delta \cdot d^{-2}.$$

Finally, taking into account that

$$X^{\mathrm{T}} X = \sum_{i=1}^{n} x_i x_i^{\mathrm{T}},$$

the proof follows for $\gamma_\delta = \frac{1}{2} \tau_\delta$. □

Remark 6. Let us notice that in contrast to the assumptions on the disturbances, e. g. that they are independent and identically distributed, which can be checked at least a posteriori by the well-known diagnostic tools, the assumptions on the explanatory variables, as (8) or (26) cannot be (usually[8]) verified at all. Then (26) gives only a hint for which character of data we may expect reasonable results.

Now, let us put

$$G^{(\beta)}(z) = H_e^{(\beta)}(\sqrt{z}) - H_e^{(\beta)}(-\sqrt{z}), \tag{30}$$

i. e. $G^{(\beta)}(z)$ is the distribution function of $\left[e_i - \mathcal{X}^{\mathrm{T}}(\beta - \beta^0)\right]^2$. Moreover, let for any $\beta \in R^p$ and any $\alpha \in [0,1]$ $u_\alpha^2(\beta)$ be the upper α-quantile of the d.f.

$$G^{(\beta)}(z), \qquad \text{i. e.} \qquad G^{(\beta)}(u_\alpha^2(\beta)) = 1 - \alpha. \tag{31}$$

Since the crucial role in the definition of $\hat{\beta}^{(\mathrm{LTS},n,h)}$ plays $r_{(h_n)}^2(\beta)$, the next lemma gives the idea about it (for h_n see (7)).

Lemma 1. Let $\alpha \in (0, \frac{1}{2})$ and let Assumptions $\mathcal{A}$ or $\mathcal{B}$ and $\mathcal{C}$ be fulfilled. Then for any $\varepsilon > 0$ there is $K^{(\varepsilon)} < \infty$ and $n_\varepsilon \in N$ such that for all $n > n_\varepsilon$

$$P\left(\sup_{\beta \in \mathcal{K}} \left| r_{(h_n)}^2(\beta) - u_\alpha^2(\beta)\} \right| < n^{-\frac{1}{2}} K^{(\varepsilon)} \right) > 1 - \varepsilon.$$

P r o o f. At first, we shall show that for any $\varepsilon > 0$ there is $K_\varepsilon < \infty$ so that

$$P\left(\sup_{\beta \in \mathcal{K}} \left[r_{(h_n)}^2(\beta) - u_\alpha^2(\beta) \right] < n^{-\frac{1}{2}} \cdot K_\varepsilon \right) > 1 - \tfrac{1}{2}\varepsilon. \tag{32}$$

Since $H_e^{(\beta)}(z)$ is absolutely continuous and strictly monotone, there is a positive density $h_e^{(\beta)}(z)$ almost everywhere. Now, due to the fact that $\alpha \in (0, \frac{1}{2})$ there is $\delta > 0$ and $L_{h_e^{(\beta)}} > 0$ so that

$$\inf_{z \in (u_\alpha^2(\beta) - \delta, u_\alpha^2(\beta) + \delta)} \text{ess} \; h_e^{(\beta)}(z) > L_{h_e^{(\beta)}}. \tag{33}$$

According to (27), for already fixed ε there is $K^{(1)} < \infty$ and $n^{(1)} \in N$ so that for all $n > n^{(1)}$ there is a set B_n such that $P(B_n) > 1 - \frac{1}{4}\varepsilon$ and for all $\omega \in B_n$

$$\sup_{\beta \in \mathcal{K}} \sup_{z \in R} \left| \frac{1}{n} \sum_{i=1}^n I\left\{ r_i(\beta) \le z \right\} - H_e^{(\beta)}(z) \right| < n^{-\frac{1}{2}} K^{(1)}. \tag{34}$$

Let us put

$$K_\varepsilon = 2\, \frac{K^{(1)}}{L_{h_e^{(\beta)}}} + 1. \tag{35}$$

[8] Of course, there are situations when the design of experiment allows to check such assumptions. E. g. in some situations we can decide at which point we shall carry out the experiment. Then we may give a rule by which we select systematically points from the factor space in such a way that (26) holds.

According to (34) (and according to the previous lines), for any $\omega \in B_n$ and any $z \in R$ (and hence also for $z_n = -\sqrt{u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon}$)

$$\sup_{\beta \in \mathcal{K}} \left| \frac{1}{n} \sum_{i=1}^{n} I\left\{ r_i(\beta) \leq -\sqrt{u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon} \right\} - H_e^{(\beta)}\left(-\sqrt{u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon} \right) \right| < n^{-\frac{1}{2}} K^{(1)}$$

as well as for $z_n = \sqrt{u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon}$

$$\sup_{\beta \in \mathcal{K}} \left| \frac{1}{n} \sum_{i=1}^{n} I\left\{ r_i(\beta) \leq \sqrt{u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon} \right\} - H_e^{(\beta)}\left(\sqrt{u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon} \right) \right| < n^{-\frac{1}{2}} K^{(1)},$$

i. e.

$$\sup_{\beta \in \mathcal{K}} \left| \frac{1}{n} \sum_{i=1}^{n} I\left\{ r_i^2(\beta) \leq u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon \right\} \right.$$
$$\left. - \left[H_e^{(\beta)}\left(\sqrt{u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon} \right) - H_e^{(\beta)}\left(-\sqrt{u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon} \right) \right] \right| < 2n^{-\frac{1}{2}} \cdot K^{(1)}. \tag{36}$$

It means that due to (33) for any $n > n^{(1)}$, any $\beta \in \mathcal{K}$ and any $\omega \in B_n$

$$\frac{1}{n} \sum_{i=1}^{n} I\left\{ r_i^2(\beta) \leq u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon \right\} \geq 1 - \alpha + n^{-\frac{1}{2}} L_{h_e^{(\beta)}} K_\varepsilon - 2n^{-\frac{1}{2}} \cdot K^{(1)}.$$

Taking into account (35), we conclude that there is $K^{(2)} > 0$ such that

$$n(1-\alpha) + n^{\frac{1}{2}} K^{(2)} \leq \sum_{i=1}^{n} I\left\{ r_i^2(\beta) \leq u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon \right\}.$$

In other words, it means that for all $n > n^{(1)}$ on the set B_n, we have for all $\beta \in \mathcal{K}$ more than $h_n = [(1-\alpha)n]$ of the squared residuals $r_i^2(\beta)$'s smaller than $u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon$. But it implies that for $\beta \in \mathcal{K}$, the h_nth order statistics among $r_i^2(\beta)$'s are smaller than $u_\alpha^2(\beta) + n^{-\frac{1}{2}} K_\varepsilon$. Since it holds for all $n > n^{(1)}$ on the set of probability at least $1 - \varepsilon$, simultaneously for all $\beta \in \mathcal{K}$, we have proved (32). The proof of

$$P\left(\sup_{\beta \in \mathcal{K}} \left[u_\alpha^2(\beta) - r_{(h_n)}^2(\beta) \right] < n^{-\frac{1}{2}} \cdot K_\varepsilon \right) > 1 - \tfrac{1}{2}\varepsilon.$$

can be carried out along similar lines. □

Remark 7. It may be of interest that Lemma 1 can be proved without Assumption $\mathcal{C}$ but then the proof is considerably more complicated, including Skorohod embeding into Wiener process in a similar way as it will be employed in the proof of the next lemma. As however we shall need Assumption $\mathcal{C}$ in what follows, we gave the proof in this, simpler form.

Lemma 2. Let $\alpha \in (0, \frac{1}{2})$ and let Assumptions $\mathcal{A}$ or $\mathcal{B}$ be fulfilled and $\mathcal{K}$ be a compact subset of R^p, $\beta^0 \in \mathcal{K}^o$. Then

$$\sup_{\beta\in\mathcal{K}} \left| \frac{1}{n}\sum_{i=1}^{n} \left(r_i^2(\beta) I\left\{ r_i^2(\beta) \le u_\alpha^2(\beta)\right\} - \mathsf{E}\left[r_i^2(\beta) I\left\{ r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right] \right) \right| \to 0$$

$$\text{a. s. as } n \to \infty. \tag{37}$$

P r o o f. First of all, similarly as in (29), let us put (throughout this proof)

$$d = \max_{\beta,\tilde{\beta}\in\mathcal{W}} \left\| \beta - \tilde{\beta} \right\|. \tag{38}$$

For any $\beta \in R^p$ and any $\delta \in (0,1)$ put

$$r_i^{2*}(\beta,\delta) = \sup_{\tilde{\beta}\in B(\beta,\delta)\cap\mathcal{K}} r_i^2(\tilde{\beta}) I\left\{ r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\}$$

and

$$r_{i*}^{2}(\beta,\delta) = \inf_{\tilde{\beta}\in B(\beta,\delta)\cap\mathcal{K}} r_i^2(\tilde{\beta}) I\left\{ r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\}.$$

Let us recall once again that

$$r_i(\tilde{\beta}) = Y_i - x_i^{\mathrm{T}}\tilde{\beta} = r_i(\beta^0) - x_i^{\mathrm{T}}(\tilde{\beta} - \beta^0) = e_i - x_i^{\mathrm{T}}(\tilde{\beta} - \beta^0). \tag{39}$$

Since $x_i^{\mathrm{T}}(\beta - \beta^0)$ does not depend on $\omega \in \Omega$, (39) hints that $r_i^{2*}(\beta,\delta)$ as well as $r_{i*}^2(\beta,\delta)$ are measurable and hence random variables. Then the inequalities

$$0 \le r_{i*}^2(\beta,\delta) \le r_i^{2*}(\beta,\delta)$$

and (recalling (38))

$$r_i^{2*}(\beta,\delta) = \sup_{\tilde{\beta}\in B(\beta,\delta)\cap\mathcal{K}} r_i^2(\tilde{\beta}) I\left\{ r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\} \le 2\left(e_i^2 + d^2\|x_i\|^2\right)$$

imply that $\mathsf{E}r_i^{2*}(\beta,\delta)$ and $\mathsf{E}r_{i*}^2(\beta,\delta)$ exist and that

$$\max\left\{ \mathrm{var}\left(r_i^{2*}(\beta,\delta)\right), \mathrm{var}\left(r_{i*}^2(\beta,\delta)\right)\right\} \le 4\left(\mathsf{E}e_i^4 + d^4\|x_i\|^4\right).$$

Moreover, since $(a+b)^4 \le 8(a^4+b^4)$

$$\mathsf{E}r_i^4(\tilde{\beta}) \le 8\left[\mathsf{E}e_i^4 + \|x_i\|^4 \left\|\tilde{\beta} - \beta^0\right\|^4\right] \le 8\left[\kappa_4 + d^4\|x_i\|^4\right]. \tag{40}$$

Now taking into account (8) and employing the strong law of large numbers (an appropriate version can be found e. g. in Breiman [3]), we find that the law holds pointwise for the sequence $\left\{r_i^{2*}(\beta,\delta)\right\}_{i=1}^{\infty}$, i. e. for any fix $\beta \in \mathcal{K}$ and any fix $\delta \in (0,1)$

$$\frac{1}{n}\sum_{i=1}^{n}\left(r_i^{2*}(\beta,\delta) - \mathsf{E}r_i^{2*}(\beta,\delta)\right) \to 0 \quad \text{a. s. as } n \to \infty \tag{41}$$

and the same is true for $r_{i*}^2(\beta,\delta)$. Further, from $r_i(\tilde{\beta}) = Y_i - x_i^{\mathrm{T}}\tilde{\beta} = r_i(\beta) - x_i^{\mathrm{T}}(\tilde{\beta} - \beta)$ it follows that

$$\left| r_i(\tilde{\beta})\right| \le |r_i(\beta)| + \|x_i\| \left\|\tilde{\beta} - \beta\right\| \tag{42}$$

and hence

$$r_i^2(\tilde\beta) \le r_i^2(\beta) + 2\,|r_i(\beta)|\,\|x_i\|\,\left\|\tilde\beta - \beta\right\| + \|x_i\|^2\left\|\tilde\beta - \beta\right\|^2. \tag{43}$$

Applying (42) on β and β^0 and plugging such upper bound of $|r_i(\beta)|$ into (43), we obtain

$$\mathsf{E}\left|r_i^2(\tilde\beta) - r_i^2(\beta)\right| \le 2\left\{\mathsf{E}|e_i| + \|x_i\|\cdot\left\|\beta - \beta^0\right\|\right\}\|x_i\|\cdot\left\|\tilde\beta - \beta\right\| + \|x_i\|^2\left\|\tilde\beta - \beta\right\|^2. \tag{44}$$

To be able to find an upper bound of $|\mathsf{E}r_i^{2*}(\beta,\delta) - \mathsf{E}r_i^2(\beta)I\{r_i^2(\beta) \le u_\alpha^2(\beta)\}|$, we need to consider in more details the mutual relation of $u_\alpha^2(\tilde\beta)$ and $u_\alpha^2(\beta)$.

Let us fix an $\varepsilon > 0$ and denote $W(s)$ the Wiener process. Further, let us find $K^{(1)} < \infty$ such that

$$P\left(\sup_{0\le s\le 2}|W(s)| > K^{(1)}\right) < \tfrac{1}{2}\varepsilon, \tag{45}$$

see e. g. Breiman (1968) or Csöőrgő, Révész (1981). Finally, let us denote for $\alpha \in (0,1)$ and $\beta \in R^p$

$$v_i^{(\alpha)}(\beta) = I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\} - \mathsf{E}I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}.$$

Recalling that $r_i(\beta) = e_i - x_i^{\mathrm T}(\beta - \beta^0)$, we conclude that $v_i^{(\alpha)}(\beta)$ are independent r.v.'s. Due to the fact that we have assumed that the support of $F(x)$ is the whole R, we have for any $\alpha \in (0,1)$ and for all $i \in N$, $a_i^{(\alpha)}(\beta) = \mathsf{E}I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\} \in (0,1)$. Then in the case when $r_i^2(\beta) \le u_\alpha^2(\beta)$ we have $v_i^{(\alpha)}(\beta) = 1 - a_i^{(\alpha)}(\beta) > 0$ otherwise $v_i^{(\alpha)}(\beta) = -a_i^{(\alpha)}(\beta) < 0$. Now, let $\{W_i(s)\}_{i=1}^\infty$ be a sequence of independent Wiener processes and let

$$\tau_i^{(\alpha)}(\beta) = \text{time for } W_i(s) \text{ to exit the interval } (-a_i^{(\alpha)}(\beta), 1 - a_i^{(\alpha)}(\beta)).$$

Then applying Lemma A.1 we obtain

$$v_i^{(\alpha)}(\beta) =_{\mathcal D} W_i(\tau_i^{(\alpha)}(\beta))$$

and

$$n^{-\frac12}\sum_{i=1}^n v_i^{(\alpha)}(\beta) =_{\mathcal D} n^{-\frac12}\sum_{i=1}^n W_i(\tau_i^{(\alpha)}(\beta)) =_{\mathcal D} W_n\left(n^{-1}\sum_{i=1}^n \tau_i^{(\alpha)}(\beta)\right). \tag{46}$$

Further, let us define a sequence of i.i.d. r.v.'s $\{V_i\}_{i=1}^\infty$

$$V_i = \text{time for } W_i(s) \text{ to exit the interval } (-1,1)$$

and applying Lemma A.1 once again we have $\mathsf{E}V_i = 1$ for all i. Moreover, since for any $\beta \in \mathcal K$ $(-a_i^{(\alpha)}(\beta), 1 - a_i^{(\alpha)}(\beta)) \subset (-1,1)$, we have

$$n^{-1}\sum_{i=1}^n \tau_i^{(\alpha)}(\beta) \le n^{-1}\sum_{i=1}^n V_i \tag{47}$$

for all $\beta \in \mathcal{K}$ and any $\omega \in \Omega$. Now, for $\varepsilon > 0$ (which was fixed at the beginning of the proof), employing the law of large numbers on the sequence $\{V_i\}_{i=1}^{\infty}$ of i.i.d. r.v.'s, let us find $n^{(1)} \in N$ such that for any $n > n^{(1)}$, putting

$$C_n = \left\{ \omega \in \Omega \ : \ n^{-1} \sum_{i=1}^{n} V_i > 2 \right\}, \tag{48}$$

we have

$$P(C_n) < \tfrac{1}{2}\varepsilon. \tag{49}$$

Notice, please that C_n does not depend on $\beta \in \mathcal{K}$. It implies that due to (47) and (48) for any $\omega \in C_n^c$ and any $n > n^{(1)}$

$$n^{-1} \sum_{i=1}^{n} \tau_i^{(\alpha)}(\beta) \leq 2. \tag{50}$$

Taking into account (46), (49) and Portnoy [11], we have for any $n > n^{(1)}$

$$P\left(n^{-\frac{1}{2}} \sup_{\beta \in \mathcal{K}} \left| \sum_{i=1}^{n} v_i^{(\alpha)}(\beta) \right| > K^{(1)} \right)$$

$$= P\left(\left\{ n^{-\frac{1}{2}} \sup_{\beta \in \mathcal{K}} \left| \sum_{i=1}^{n} v_i^{(\alpha)}(\beta) \right| > K^{(1)} \right\} \cap C_n^c \right) + P\left(C_n\right)$$

$$\leq P\left(\left\{ n^{-\frac{1}{2}} \sup_{\beta \in \mathcal{K}} \left| \sum_{i=1}^{n} W_i(\tau_i^{(\alpha)}(\beta)) \right| > K^{(1)} \right\} \cap C_n^c \right) + \tfrac{1}{2}\varepsilon$$

$$= P\left(\left\{ \sup_{\beta \in \mathcal{K}} \left| W_n(n^{-1} \sum_{i=1}^{n} \tau_i^{(\alpha)}(\beta)) \right| > K^{(1)} \right\} \cap C_n^c \right) + \tfrac{1}{2}\varepsilon.$$

Employing now (45) and (50), we obtain

$$P\left(\left\{ \sup_{\beta \in \mathcal{K}} \left| W_n(n^{-1} \sum_{i=1}^{n} \tau_i^{(\alpha)}(\beta)) \right| > K^{(1)} \right\} \cap C_n^c \right) + \tfrac{1}{2}\varepsilon$$

$$\leq P\left(\left\{ \sup_{0 \leq s \leq 2} \ |W(s)| > K^{(1)} \right\} \cap C_n^c \right) + \tfrac{1}{2}\varepsilon < \varepsilon$$

so that

$$\frac{1}{n} \sup_{\beta \in \mathcal{K}} \left| \sum_{i=1}^{n} \left[I\left\{ r_i(\beta) \leq u_\alpha^2(\beta) \right\} - \mathsf{E} I\left\{ r_i^2(\beta) \leq u_\alpha^2(\beta) \right] \right\} \right| \leq n^{-\frac{1}{2}} K^{(1)} \tag{51}$$

with probability at least $1 - \varepsilon$. Now, similarly as we have derived (36), we can find $K^{(2)} < \infty$ so that

$$\left| \frac{1}{n} \sum_{i=1}^{n} I\left\{ r_i^2(\beta) \leq u_\alpha^2(\beta) \right\} - \left[H_e^{(\beta)}(u_\alpha(\beta)) - H_e^{(\beta)}(-u_\alpha(\beta)) \right] \right| < n^{-\frac{1}{2}} K^{(2)}, \tag{52}$$

again with probability at least $1-\varepsilon$. Since

$$G^{(\beta)}(u_\alpha^2(\beta)) = H_e^{(\beta)}(u_\alpha(\beta)) - H_e^{(\beta)}(-u_\alpha(\beta)),$$

taking into account (51) and (52) we may conclude that there is a constant $K^{(3)} < \infty$ so that

$$\frac{1}{n}\left|\sum_{i=1}^n \mathsf{E} I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\} - G^{(\beta)}(u_\alpha^2(\beta))\right| \le n^{-\frac{1}{2}}\, K^{(3)}. \tag{53}$$

(Notice please that the last inequality does not contain any random term.) As $G^{(\beta)}(u_\alpha^2(\beta)) = 1-\alpha = G^{(\tilde\beta)}(u_\alpha^2(\tilde\beta))$, (53) implies that for $K^{(4)} = 2\cdot K^{(3)}$

$$\frac{1}{n}\left|\sum_{i=1}^n \left[\mathsf{E} I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\} - \mathsf{E} I\left\{r_i^2(\tilde\beta) \le u_\alpha^2(\tilde\beta)\right\}\right]\right| < n^{-\frac{1}{2}}\, K^{(4)}$$

which may be written as

$$\frac{1}{n}\left|\sum_{i=1}^n \left[P\left(r_i^2(\beta) \le u_\alpha^2(\beta)\right) - P\left(r_i^2(\tilde\beta) \le u_\alpha^2(\tilde\beta)\right)\right]\right| < n^{-\frac{1}{2}}\, K^{(4)}. \tag{54}$$

As $r_i(\beta) = e_i - x_i^{\mathrm{T}}(\beta-\beta^0)$, $\quad r_i^2(\beta) \le u_\alpha^2(\beta)$ is equivalent to $\left[e_i - x_i^{\mathrm{T}}(\beta-\beta^0)\right]^2 \le u_\alpha^2(\beta)$ which gives

$$-u_\alpha(\beta) + x_i^{\mathrm{T}}(\beta-\beta^0) \le e_i \le u_\alpha(\beta) + x_i^{\mathrm{T}}(\beta-\beta^0).$$

Then of course,

$$P\left(r_i^2(\beta) \le u_\alpha^2(\beta)\right) = F(u_\alpha(\beta) + x_i^{\mathrm{T}}\left(\beta-\beta^0\right)) - F(-u_\alpha(\beta) + x_i^{\mathrm{T}}\left(\beta-\beta^0\right)) \tag{55}$$

and similarly for $P\left(r_i^2(\tilde\beta) \le u_\alpha^2(\tilde\beta)\right)$. Due to the assumption that the density exist everywhere, we can find point u_i so that $u_i \in (a_i, \tilde a_i)_{\mathrm{ord}}$ where $a_i = u_\alpha(\beta) + x_i^{\mathrm{T}}\left(\beta-\beta^0\right)$ and $\tilde a_i = u_\alpha(\tilde\beta) + x_i^{\mathrm{T}}\left(\tilde\beta-\beta^0\right)$ and

$$F\left(u_\alpha(\beta) + x_i^{\mathrm{T}}\left(\beta-\beta^0\right)\right) - F\left(u_\alpha(\tilde\beta) + x_i^{\mathrm{T}}(\tilde\beta-\beta^0)\right)$$

$$= f(u_i)\left[u_\alpha(\beta) - u_\alpha(\tilde\beta) + x_i^{\mathrm{T}}\left(\beta-\tilde\beta\right)\right] \tag{56}$$

and $\tilde u_i$ so that $\tilde u_i \in (b_i, \tilde b_i)_{\mathrm{ord}}$ where $b_i = -u_\alpha(\beta) + x_i^{\mathrm{T}}\left(\beta-\beta^0\right)$ and $\tilde b_i = -u_\alpha(\tilde\beta) + x_i^{\mathrm{T}}\left(\tilde\beta-\beta^0\right)$ and

$$F(-u_\alpha(\beta) + x_i^{\mathrm{T}}\left(\beta-\beta^0\right)) - F(-u_\alpha(\tilde\beta) + x_i^{\mathrm{T}}\left(\tilde\beta-\beta^0\right))$$

$$= f(\tilde u_i)\left[u_\alpha(\tilde\beta) - u_\alpha(\beta) + x_i^{\mathrm{T}}\left(\beta-\tilde\beta\right)\right]. \tag{57}$$

Now, taking into account (54), (55), (56) and (57), we arrive at

$$\frac{1}{n}\left|\sum_{i=1}^{n}\left\{[f(u_i)+f(\tilde{u}_i)]\cdot\left[u_\alpha(\beta)-u_\alpha(\tilde{\beta})\right]+[f(u_i)-f(\tilde{u}_i)]\,x_i^{\mathrm{T}}\left(\beta-\tilde{\beta}\right)\right\}\right|<2n^{-\frac{1}{2}}\,K^{(4)}$$

which yields

$$-2n^{-\frac{1}{2}}\,K^{(4)}-\frac{1}{n}\sum_{i=1}^{n}[f(u_i)-f(\tilde{u}_i)]\,x_i^{\mathrm{T}}\left(\beta-\tilde{\beta}\right)$$
$$<\frac{1}{n}\sum_{i=1}^{n}[f(u_i)+f(\tilde{u}_i)]\cdot\left[u_\alpha(\beta)-u_\alpha(\tilde{\beta})\right]$$
$$<-\frac{1}{n}\sum_{i=1}^{n}[f(u_i)-f(\tilde{u}_i)]\,x_i^{\mathrm{T}}\left(\beta-\tilde{\beta}\right)+2n^{-\frac{1}{2}}\,K^{(4)}.$$

Since $-|a|\leq a\leq|a|$ and

$$\left|\frac{1}{n}\sum_{i=1}^{n}[f(u_i)-f(\tilde{u}_i)]\,x_i^{\mathrm{T}}\left(\beta-\tilde{\beta}\right)\right|\leq\frac{1}{n}\sum_{i=1}^{n}\left|f(u_i)-f(\tilde{u}_i)\right|\cdot\left|x_i^{\mathrm{T}}\left(\beta-\tilde{\beta}\right)\right|,$$

we have

$$-2\cdot n^{-\frac{1}{2}}\,K^{(4)}-\frac{1}{n}\sum_{i=1}^{n}|f(u_i)-f(\tilde{u}_i)|\left|x_i^{\mathrm{T}}\left(\beta-\tilde{\beta}\right)\right|$$
$$<\frac{1}{n}\left[u_\alpha(\beta)-u_\alpha(\tilde{\beta})\right]\cdot\sum_{i=1}^{n}[f(u_i)+f(\tilde{u}_i)]$$
$$<\frac{1}{n}\sum_{i=1}^{n}|f(u_i)-f(\tilde{u}_i)|\left|x_i^{\mathrm{T}}\left(\beta-\tilde{\beta}\right)\right|+2\cdot n^{-\frac{1}{2}}\,K^{(4)}.$$

It means that

$$\frac{1}{n}\left|u_\alpha(\beta)-u_\alpha(\tilde{\beta})\right|\sum_{i=1}^{n}[f(u_i)+f(\tilde{u}_i)]\leq\frac{1}{n}\sum_{i=1}^{n}\left|f(u_i)-f(\tilde{u}_i)\right|\cdot\left|x_i^{\mathrm{T}}\left(\beta-\tilde{\beta}\right)\right|+2\,n^{-\frac{1}{2}}K^{(4)}.\tag{58}$$

Now

$$f(u_i)=f(u_\alpha(\beta))+f'(v_i)x_i^{\mathrm{T}}\left(\beta-\tilde{\beta}\right)\tag{59}$$

and

$$f(\tilde{u}_i)=f(u_\alpha(\beta))+f'(\tilde{v}_i)x_i^{\mathrm{T}}\left(\beta-\tilde{\beta}\right)\tag{60}$$

where v_i and $\tilde{v}_i$ were appropriately selected (similarly as u_i and $\tilde{u}_i$ a few lines above). Denoting the upper bound of the absolute value of the derivative f' by $U_{f'}$, we get from (59) and (60)

$$\frac{1}{n}\sum_{i=1}^{n}\left|f(u_i)-f(\tilde{u}_i)\right|\cdot\left|x_i^{\mathrm{T}}\cdot\left(\beta-\tilde{\beta}\right)\right|\leq\frac{1}{n}\sum_{i=1}^{n}\left|f'(v_i)-f'(\tilde{v}_i)\right|\cdot\left[x_i^{\mathrm{T}}\cdot\left(\beta-\tilde{\beta}\right)\right]^2$$

$$\leq \frac{2}{n} U_{f'} \sum_{i=1}^{n} \left[x_i^{\mathrm{T}} \cdot \left(\beta - \tilde{\beta} \right) \right]^2 \leq U_{f'} \left\| \beta - \tilde{\beta} \right\|^2 \cdot \frac{2}{n} \sum_{i=1}^{n} \|x_i\|^2 \tag{61}$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \left[f(u_i) + f(\tilde{u}_i) \right] \geq 2 \cdot f(u_\alpha(\beta)) - U_{f'} \left\| \beta - \tilde{\beta} \right\|^2 \frac{2}{n} \sum_{i=1}^{n} \|x_i\|^2. \tag{62}$$

Since we have assumed that $f(z)$ is positive everywhere, we have $f(u_\alpha(\beta)) > 0$. Moreover, due to (8) there is a constant $K^{(5)}$ and $n^{(2)} \in N, n^{(2)} > n^{(1)}$ such that for all $n > n^{(2)}$ we have

$$\frac{1}{n} \sum_{i=1}^{n} \|x_i\|^2 < K^{(5)}.$$

Let us put $\delta_0 = \min \left\{ \frac{1}{2} \frac{f(u_\alpha(\beta))}{U_f \cdot K^{(5)}}, 1 \right\}$ and consider any positive $\delta < \delta_0$. Then for any $\beta, \tilde{\beta} \in R^p$ such that $\left\| \beta - \tilde{\beta} \right\| < \delta$, we have

$$U_{f'} \left\| \beta - \tilde{\beta} \right\|^2 \frac{2}{n} \sum_{i=1}^{n} \|x_i\|^2 \leq f(u_\alpha(\beta))$$

and hence (62) gives

$$\frac{1}{n} \sum_{i=1}^{n} \left[f(u_i) + f(\tilde{u}_i) \right] \geq f(u_\alpha(\beta)). \tag{63}$$

Now taking into account (58), (61) and (63) we conclude that

$$\left| u_\alpha(\beta) - u_\alpha(\tilde{\beta}) \right| \leq f^{-1}(u_\alpha(\beta)) \left\{ 2K^{(5)} U_{f'} \left\| \beta - \tilde{\beta} \right\|^2 + 2n^{-\frac{1}{2}} K^{(4)} \right\}.$$

Since the left hand side of the previous inequality does not depend on n, we conclude that there is a constant $K^{(6)} < \infty$ such that for all $\left\| \beta - \tilde{\beta} \right\| < \delta$ (remember, we have considered $\delta \in (0, 1)$)

$$\left| u_\alpha(\beta) - u_\alpha(\tilde{\beta}) \right| \leq K^{(6)} \left\| \beta - \tilde{\beta} \right\|^2. \tag{64}$$

Now,

$$r_i^2(\beta) \leq u_\alpha^2(\beta) \quad \Leftrightarrow \quad -u_\alpha(\beta) \leq r_i(\beta) \leq u_\alpha(\beta) \quad \Leftrightarrow$$
$$\Leftrightarrow \quad -u_\alpha(\beta) + x_i^{\mathrm{T}} \cdot \left(\beta - \beta^0 \right) \leq e_i \leq u_\alpha(\beta) + x_i^{\mathrm{T}} \cdot \left(\beta - \beta^0 \right).$$

Hence $I \left\{ r_i^2(\tilde{\beta}) \leq u_\alpha^2(\tilde{\beta}) \right\} \neq I \left\{ r_i^2(\beta) \leq u_\alpha^2(\beta) \right\}$ appears when either

$$e_i \in \left(-u_\alpha(\beta) + x_i^{\mathrm{T}} \cdot \left(\beta - \beta^0 \right), -u_\alpha(\tilde{\beta}) + x_i^{\mathrm{T}} \cdot \left(\tilde{\beta} - \beta^0 \right) \right)_{\mathrm{ord}} \tag{65}$$

or

$$e_i \in \left(u_\alpha(\beta) + x_i^{\mathrm{T}} \cdot \left(\beta - \beta^0 \right), u_\alpha(\tilde{\beta}) + x_i^{\mathrm{T}} \cdot \left(\tilde{\beta} - \beta^0 \right) \right)_{\mathrm{ord}}. \tag{66}$$

As we have assumed the density $f(z)$ to be bounded (say by B_f), we conclude from (64), (65) and (66) that there is a constant $K^{(7)} < \infty$ such that for any pair $\beta, \tilde{\beta}$ such that $\|\beta - \tilde{\beta}\| < \delta$

$$P\left(I\left\{r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\} \neq I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right) \le B_f \left\{\|x_i\| \left\|\beta - \tilde{\beta}\right\| + K^{(7)} \left\|\beta - \tilde{\beta}\right\|^2\right\}. \tag{67}$$

Further, we have

$$\left[I\left\{r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\} - I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right]^2 = \left|I\left\{r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\} - I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right|$$

and hence

$$\left|\mathsf{E}\left\{r_i^2(\tilde{\beta}) I\left\{r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\}\right\} - \mathsf{E}\left\{r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right\}\right|$$
$$\le \mathsf{E}\left\{\left|r_i^2(\tilde{\beta}) - r_i^2(\beta)\right| \cdot I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right\}$$
$$+ \mathsf{E}\left\{r_i^2(\tilde{\beta}) \cdot \left|I\left\{r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\} - I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right|\right\}$$
$$\le \mathsf{E}\left\{\left|r_i^2(\tilde{\beta}) - r_i^2(\beta)\right| \cdot I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right\}$$
$$+ \left\{\mathsf{E} r_i^4(\tilde{\beta}) \cdot \mathsf{E}\left|I\left\{r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\} - I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right|\right\}^{\frac{1}{2}}.$$

Now recalling that

$$\mathsf{E}\left|I\left\{r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\} - I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right| = P\left(I\left\{r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\} \neq I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right),$$

taking into account that and (67) together with (40) and (44), we conclude that there is a constant $K^{(9)}$ such that for any $\beta \in \mathcal{K}$ and $\tilde{\beta} \in \mathcal{K}$, $\|\beta - \tilde{\beta}\| < \delta$

$$\left|\mathsf{E}\left\{r_i^2(\tilde{\beta}) I\left\{r_i^2(\tilde{\beta}) \le u_\alpha^2(\tilde{\beta})\right\}\right\} - \mathsf{E}\left\{r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right\}\right|$$
$$\le K^{(9)} \left[\|x_i\| + \|x_i\|^2 + \|x_i\|^3 + \|x_i\|^4\right] \cdot \left\|\tilde{\beta} - \beta\right\|. \tag{68}$$

But (68) implies that there is a constant $K^{(10)}$ such that for any $\beta \in \mathcal{K}$ and any $\delta < \delta_0$ we have

$$\left|\mathsf{E} r_i^{2*}(\beta, \delta) - \mathsf{E}\left\{r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right\}\right| \le K^{(10)} \left[\|x_i\| + \|x_i\|^2 + \|x_i\|^3 + \|x_i\|^4\right] \delta.$$

Now taking into account assumption (8), we conclude that for any $\varepsilon > 0$ there is a $\delta'_\varepsilon \in (0,1)$ so that

$$\sup_{n \in N} \left|\frac{1}{n} \sum_{i=1}^{n} \left(\mathsf{E} r_i^{2*}(\beta, \delta'_\varepsilon) - \mathsf{E}\left[r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right]\right)\right| < \varepsilon. \tag{69}$$

Along similar lines we can obtain for the same ε and some $\delta''_\varepsilon \in (0,1)$ so that

$$\sup_{n\in N} \left| \frac{1}{n}\sum_{i=1}^n \left(\mathsf{E} r_{i*}^2(\beta,\delta''_\varepsilon) - \mathsf{E}\left[r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right]\right)\right| < \varepsilon. \tag{70}$$

Let us put $\delta_\varepsilon = \min\{\delta'_\varepsilon, \delta''_\varepsilon\}$. Employing now the inequalities (69) and (70), we obtain for given ε and δ_ε

$$\frac{1}{n}\sum_{i=1}^n \mathsf{E}\left[r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right] - \varepsilon \le \frac{1}{n}\sum_{i=1}^n \mathsf{E} r_{i*}^2(\beta,\delta_\varepsilon)$$

$$\le \frac{1}{n}\sum_{i=1}^n \mathsf{E} r_i^{2*}(\beta,\delta_\varepsilon) \le \frac{1}{n}\sum_{i=1}^n \mathsf{E}\left[r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right] + \varepsilon. \tag{71}$$

(Notice please that δ_ε does not depend on β but we shall not need it.) Now, for given ε the system of the balls $\{B(\beta,\delta_\varepsilon),\ \beta \in \mathcal{K}\}$ represents an open cover of the compact set $\mathcal{K}$. Hence there is a finite subsystem, say $\{B(\beta_\ell,\delta_\varepsilon),\ \ell = 1,2,\ldots,L\}$ which also covers $\mathcal{K}$. For any $\beta \in B(\beta_\ell,\delta_\varepsilon)$ we have

$$\frac{1}{n}\sum_{i=1}^n \left(r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\} - \mathsf{E}\left[r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right]\right)$$

$$\le \frac{1}{n}\sum_{i=1}^n \left(r_i^{2*}(\beta_\ell,\delta_\varepsilon) - \mathsf{E} r_{i*}^2(\beta_\ell,\delta_\varepsilon)\right)$$

$$\le \frac{1}{n}\sum_{i=1}^n \left(r_i^{2*}(\beta_\ell,\delta_\varepsilon) - \mathsf{E} r_i^{2*}(\beta_\ell,\delta_\varepsilon)\right) + 2\varepsilon \tag{72}$$

and

$$\frac{1}{n}\sum_{i=1}^n \left(r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\} - \mathsf{E}\left[r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right]\right)$$

$$\ge \frac{1}{n}\sum_{i=1}^n \left(r_{i*}^2(\beta_\ell,\delta_\varepsilon) - \mathsf{E} r_i^{2*}(\beta_\ell,\delta_\varepsilon)\right)$$

$$\ge \frac{1}{n}\sum_{i=1}^n \left(r_{i*}^2(\beta_\ell,\delta_\varepsilon) - \mathsf{E} r_{i*}^2(\beta_\ell,\delta_\varepsilon)\right) - 2\varepsilon. \tag{73}$$

Now for any $\beta \in \mathcal{K}$ (72) and (73) give

$$\min_{1\le\ell\le L} \left\{\frac{1}{n}\sum_{i=1}^n \left(r_{i*}^2(\beta_\ell,\delta_\varepsilon) - \mathsf{E} r_{i*}^2(\beta_\ell,\delta_\varepsilon)\right)\right\} - 2\varepsilon$$

$$\le \frac{1}{n}\sum_{i=1}^n \left(r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\} - \mathsf{E}\left[r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\}\right]\right)$$

$$\le \max_{1\le\ell\le L} \left\{\frac{1}{n}\sum_{i=1}^n \left(r_i^{2*}(\beta_\ell,\delta_\varepsilon) - \mathsf{E} r_i^{2*}(\beta_\ell,\delta_\varepsilon)\right)\right\} + 2\varepsilon.$$

Taking into account (41) (and the same fact as (41) for $r_{i*}^2(\beta,\delta)$), we conclude the proof. □

Corollary 1. Let assumptions of Lemma 1 and 2 are fulfilled. Then

$$\sup_{\beta\in\mathcal{K}}\left|\frac{1}{n}\sum_{i=1}^{n}\left(r_i^2(\beta)I\left\{r_i^2(\beta)\le r_{(h_n)}^2(\beta)\right\}-\mathsf{E}\left[r_i^2(\beta)I\left\{r_i^2(\beta)\le u_\alpha^2(\beta)\right\}\right]\right)\right|\to 0$$

in probability as $n\to\infty$.

The proof follows immediately from Lemma 1 and 2. □

Lemma 3. Let $\alpha\in(0,\frac{1}{2})$ and let Assumptions $\mathcal{A}$ or $\mathcal{B}$ be fulfilled and $\mathcal{K}$ be a compact subset of R^p, $\beta^0\in\mathcal{K}^o$. Then for any $\delta>0$ there is $n_\delta\in N$ and $\gamma_\delta>0$ such that for any $n>n_\delta$

$$\min_{\beta\in\mathcal{K}-B(\beta^0,\delta)}\frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\left[r_i^2(\beta)I\left\{r_i^2(\beta)\le u_\alpha^2(\beta)\right\}\right]-\mathsf{E}\left[r_1^2(\beta^0)I\left\{r_1^2(\beta^0)\le u_\alpha^2\right\}\right]>\gamma_\delta. \tag{74}$$

P r o o f. Let us denote the boundary of $B(\beta^0,\delta)$ by $B^b(\beta^0,\delta)$, i.e. $B^b(\beta^0,\delta)=\bar{B}(\beta^0,\delta)\setminus B(\beta^0,\delta)$ where $\bar{B}(\beta^0,\delta)$ is the closure of $B(\beta^0,\delta)$.

The idea of proof is as follows. Firstly to show that there is a $\delta^*\in(0,1)$ such that for any $0<\delta<\delta^*$ there is a $\gamma_\delta>0$ such that

$$\min_{\beta\in B^b(\beta^0,\delta)}\frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\left[r_i^2(\beta)I\left\{r_i^2(\beta)\le u_\alpha^2(\beta)\right\}\right]-\mathsf{E}\left[r_1^2(\beta^0)I\left\{r_1^2(\beta^0)\le u_\alpha^2\right\}\right]>\gamma_\delta>0.$$

Secondly, we show that the expression $\frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\left[r_i^2(\beta)I\left\{r_i^2(\beta)\le u_\alpha^2(\beta)\right\}\right]$ is not decreasing when we move along any direction from β^0. Let us make the first step. Please, keep in mind that we consider $\delta^*\in(0,1)$ and hence for any $\delta\in(0,\delta^*)$ and any $\beta\in B^b(\beta^0,\delta)$ we have $\left\|\beta-\beta^0\right\|^k<\left\|\beta-\beta^0\right\|$ for any $k>1$.

Throughout the first part of the proof let us write Δ_i instead of $x_i^{\mathrm{T}}\left(\beta-\beta^0\right)$. We have

$$\frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\left[r_i^2(\beta)I\left\{r_i^2(\beta)\le u_\alpha^2(\beta)\right\}\right]-\mathsf{E}\left[r_1^2(\beta^0)I\left\{r_1^2(\beta^0)\le u_\alpha^2\right\}\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\left\{\left[r_i^2(\beta)-r_i^2(\beta^0)\right]I\left\{r_i^2(\beta^0)\le u_\alpha^2\right\}\right\} \tag{75}$$

$$+\frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\left\{r_i^2(\beta)\left[I\left\{r_i^2(\beta)\le u_\alpha^2(\beta)\right\}-I\left\{r_i^2(\beta^0)\le u_\alpha^2\right\}\right]\right\}. \tag{76}$$

Let us try to evaluate the expression in (75) and let us recall that $r_i(\beta^0)=e_i$ and $r_i(\beta)=e_i-\Delta_i$. Then due to (10)

$$\mathsf{E}\left\{\left[r_i^2(\beta)-r_i^2(\beta^0)\right]I\left\{r_i^2(\beta^0)\le u_\alpha^2\right\}\right\}=\mathsf{E}\left\{\left[\Delta_i^2-2\Delta_ie_i\right]I\left\{r_i^2(\beta^0)\le u_\alpha^2\right\}\right\}$$

$$=(1-\alpha)\Delta_i^2=(1-\alpha)\left(x_i^{\mathrm{T}}\left(\beta-\beta^0\right)\right)^2. \tag{77}$$

Now, let us turn to (76). We shall find a lower bound of

$$\mathsf{E}\left\{r_i^2(\beta)\left[I\left\{r_i^2(\beta)\leq u_\alpha^2(\beta)\right\}-I\left\{r_i^2(\beta^0)\leq u_\alpha^2\right\}\right]\right\}=\int\limits_{B_i^{(1)}} r_i^2(\beta)\mathrm{d}P(\omega)-\int\limits_{B_i^{(2)}} r_i^2(\beta)\mathrm{d}P(\omega) \tag{78}$$

where

$$B_i^{(1)}=\left\{\omega\in\Omega\ :\ I\left\{r_i^2(\beta)\leq u_\alpha^2(\beta)\right\}-I\left\{r_i^2(\beta^0)\leq u_\alpha\right\}=1\right\}$$

and

$$B_i^{(2)}=\left\{\omega\in\Omega\ :\ I\left\{r_i^2(\beta)\leq u_\alpha^2(\beta)\right\}-I\left\{r_i^2(\beta^0)\leq u_\alpha\right\}=-1\right\}.$$

To be able to evaluate integrals in (78) we need to make an idea about $B_i^{(1)}$, $B_i^{(2)}$ and $u_\alpha(\beta)$. Let us start with $u_\alpha(\beta)$.

We are going to show that due to the fact that $H_e^{(\beta)}(z)$ is the convolution of $F(z)$ and $H^{(\beta)}(t)$ (see Assumptions $\mathcal{A}$ or $\mathcal{B}$ and $\mathcal{C}$, and also Assertion 2), the interval $(-u_\alpha,u_\alpha)$ is shorter for any $\alpha\in(0,1)$ and any $\beta\in R^p$ than $(-u_\alpha(\beta),u_\alpha(\beta))$, i.e. $u_\alpha^2\leq u_\alpha^2(\beta)$. First of all, let us recall that u_α^2 was defined as the upper α-quantile of the d.f. $G(z)$, i.e. $\alpha=1-G(u_\alpha^2)=P\left(e_1^2>u_\alpha^2\right)=P\left(e_1<-u_\alpha\right)+P\left(e_1>u_\alpha\right)=\alpha$, see (6). Similarly, $u_\alpha^2(\beta)$ was given as the upper α-quantile of $G^{(\beta)}(z)$, i.e. $G^{(\beta)}(u_\alpha^2(\beta))=1-\alpha$. In other words, see (30) and (31) $H_e^{(\beta)}(u_\alpha(\beta))-H_e^{(\beta)}(-u_\alpha(\beta))=1-\alpha$. As we have said

$$H_e^{(\beta)}(z)=\int_{-\infty}^{-\infty}F(z-t)\mathrm{d}H^{(\beta)}(t)$$

and hence we look for such a value $u_\alpha(\beta)$ that

$$\int_{-\infty}^{-\infty}\left[F(u_\alpha(\beta)-t)-F(-u_\alpha(\beta)-t)\right]\mathrm{d}H^{(\beta)}(t)=1-\alpha.$$

Let us recall that $f(z)$ is symmetric around 0, strictly decreasing on R^+ and $F(u_\alpha)-F(-u_\alpha)=1-\alpha$. Then for any $a\in(0,u_\alpha)$ we have

$$F(a)-F(-a)<1-\alpha$$

and hence for any $t\in R$

$$F(a-t)-F(-a-t)<1-\alpha$$

(after all, the last inequality is immediately visible from a picture of density f). But it means that for any $a\in(0,u_\alpha)$

$$\int_{-\infty}^{-\infty}\left[F(a-t)-F(-a-t)\right]\mathrm{d}H^{(\beta)}(t)<\int_{-\infty}^{-\infty}(1-\alpha)\mathrm{d}H^{(\beta)}(t)=1-\alpha$$

and hence $u_\alpha(\beta)\geq u_\alpha$.

It immediately implies that $|u_\alpha(\beta) - u_\alpha| = u_\alpha(\beta) - u_\alpha$ and then according to (64) there is a constant $K^{(1)} \in (1, \infty)$ and $\delta > 0$ such that for any $\|\beta - \beta^0\| < \delta$

$$u_\alpha(\beta) - u_\alpha < K^{(1)} \left\|\beta - \beta^0\right\|^2 . \tag{79}$$

It allows to find a direct approximation to $u_\alpha(\beta) - u_\alpha$. Put $\tilde{\delta} = \min\{[K^{(1)}]^{-1}, \delta\}$ and in what follows assume only $\beta \in R^p$ such that $\|\beta - \beta^0\| < \tilde{\delta}$. Then of course, due to (79), $|u_\alpha(\beta) - u_\alpha|^k < |u_\alpha(\beta) - u_\alpha| < 1$ for any $k > 1$. Employing (54) for $\tilde{\beta} = \beta^0$ we obtain (for some $K^{(2)} < \infty$)

$$\frac{1}{n}\left|\sum_{i=1}^{n}\left[P\left(r_i^2(\beta) \le u_\alpha^2(\beta)\right) - P\left(r_i^2(\beta^0) \le u_\alpha^2(\beta^0)\right)\right]\right| < n^{-\frac{1}{2}}\,K^{(2)}$$

i. e.

$$\frac{1}{n}\left|\sum_{i=1}^{n}\{F(u_\alpha(\beta) + \Delta_i) - F(-u_\alpha(\beta) + \Delta_i) - [F(u_\alpha) - F(-u_\alpha)]\}\right| < n^{-\frac{1}{2}}\,K^{(2)}. \tag{80}$$

Let us consider at first

$$F(u_\alpha(\beta) + \Delta_i) - F(-u_\alpha(\beta) + \Delta_i) - [F(u_\alpha) - F(-u_\alpha)]$$

$$= F(u_\alpha(\beta) + \Delta_i) - F(u_\alpha) - [F(-u_\alpha(\beta) + \Delta_i) - F(-u_\alpha)]\,. \tag{81}$$

We are going to utilize for $z \in (u_\alpha, u_\alpha(\beta) + \Delta_i)_{\text{ord}}$ the expansion

$$f(z) = f(u_\alpha) + f'(u_\alpha)(z - u_\alpha) + \tfrac{1}{2}f''(a_1)(z - u_\alpha)^2$$

where $a_1 = a_1(z) \in (u_\alpha, z)_{\text{ord}}$ and for $z \in (-u_\alpha, -u_\alpha(\beta) + \Delta_i)_{\text{ord}}$

$$f(z) = f(u_\alpha) - f'(u_\alpha)(z + u_\alpha) + \tfrac{1}{2}f''(a_2)(z + u_\alpha)^2$$

where $a_2 = a_2(z) \in (-u_\alpha, z)_{\text{ord}}$. Let us denote J_1 the (upper) bound of the absolute value of the second derivative of density, i. e.

$$J_1 = \sup_{z \in R} |f''(z)|\,.$$

Then

$$F(u_\alpha(\beta)+\Delta_i) - F(u_\alpha) = f(u_\alpha)\left[u_\alpha(\beta) + \Delta_i - u_\alpha\right] + \tfrac{1}{2}f'(u_\alpha)\left[u_\alpha(\beta) + \Delta_i - u_\alpha\right]^2 + R_i$$

where

$$|R_i| = \tfrac{1}{2}\left|\int_{u_\alpha}^{u_\alpha(\beta)+\Delta_i} f''(a_1(z))\,(z - u_\alpha)^2\,\mathrm{d}z\right|$$

$$\le \frac{1}{6}J_1\left[u_\alpha(\beta) + \Delta_i - u_\alpha\right]^3 \le K^{(3)}\left\{|u_\alpha(\beta) - u_\alpha|^3 + |\Delta_i|^3\right\}$$

for some $K^{(3)} < \infty$. Taking into account (8), (79) and (80), we can find $K^{(4)} < \infty$ so that

$$\frac{1}{n}\left|\sum_{i=1}^{n}\left\{f(u_\alpha)\left[u_\alpha(\beta)+\Delta_i-u_\alpha\right]+\tfrac{1}{2}f'(u_\alpha)\left[u_\alpha(\beta)+\Delta_i-u_\alpha\right]^2\right.\right.$$
$$\left.\left.-f(u_\alpha)\left[-u_\alpha(\beta)+\Delta_i+u_\alpha\right]+\tfrac{1}{2}f'(u_\alpha)\left[-u_\alpha(\beta)+\Delta_i+u_\alpha\right]^2\right\}\right|$$
$$< \quad K^{(4)}\cdot\left\|\beta-\beta^0\right\|^3+n^{-\frac{1}{2}}\,K^{(2)}.$$

The last inequality may be written as

$$\frac{1}{n}\left|\sum_{i=1}^{n}\left\{2f(u_\alpha)\left[u_\alpha(\beta)-u_\alpha\right]+f'(u_\alpha)\left[\left(u_\alpha(\beta)-u_\alpha\right)^2+\Delta_i^2\right]\right\}\right|$$
$$< \quad K^{(4)}\cdot\left\|\beta-\beta^0\right\|^3+n^{-\frac{1}{2}}\,K^{(2)}.$$

Taking into account that the derivative $f'(z)$ is bounded in absolute value and employing (79) once again, we obtain from the previous line (for some $K^{(5)} < \infty$; remember that we consider only such $\beta \in R^p$ that $\|\beta-\beta^0\| < \tilde{\delta}$)

$$\frac{1}{n}\left|\sum_{i=1}^{n}\left\{2\cdot f(u_\alpha)\left[u_\alpha(\beta)-u_\alpha\right]+f'(u_\alpha)\Delta_i^2\right\}\right| < K^{(5)}\left\|\beta-\beta^0\right\|^3+n^{-\frac{1}{2}}K^{(2)}$$

which, due to the fact that $f(u_\alpha) > 0$, finally gives for some $K^{(6)} < \infty$

$$\left|u_\alpha(\beta)-u_\alpha+\frac{f'(u_\alpha)}{2f(u_\alpha)}\cdot\frac{1}{n}\sum_{i=1}^{n}\Delta_i^2\right| \le K^{(6)}\left\{\left\|\beta-\beta^0\right\|^3+n^{-\frac{1}{2}}\right\}. \tag{82}$$

Now, let us return to $B_i^{(1)}$ and $B_i^{(2)}$. The event $B_i^{(1)}$ appears whenever

$$\{-u_\alpha(\beta)+\Delta_i \le e_i \le u_\alpha(\beta)+\Delta_i\} \cap \{\{e_i < -u_\alpha\} \cup \{u_\alpha < e_i\}\}$$

which is equivalent to

$$\{-u_\alpha(\beta)+\Delta_i \le e_i \le -u_\alpha\} \cup \{u_\alpha \le e_i \le u_\alpha(\beta)+\Delta_i\}\,. \tag{83}$$

Similarly for $B_i^{(2)}$ we have

$$\Big\{\{e_i < -u_\alpha(\beta)+\Delta_i\} \cup \{u_\alpha(\beta)+\Delta_i < e_i\}\Big\} \cap \Big\{-u_\alpha \le e_i \le u_\alpha\Big\}$$

which is again equivalent to

$$\{-u_\alpha \le e_i < -u_\alpha(\beta)+\Delta_i\} \cup \{u_\alpha(\beta)+\Delta_i \le e_i < u_\alpha\}\,. \tag{84}$$

First of all, let us assume that $\Delta_i \ge 0$. Then, if $u_\alpha(\beta)-u_\alpha < \Delta_i$, the first interval in (83) as well as the second interval in (84) are empty. For $0 < \Delta_i < u_\alpha(\beta)-u_\alpha$ both intervals in (83) are nonempty while both intervals in (84) are empty. For $\Delta_i < 0$ the situation is symmetric. Please, keep it in mind.

So, let us assume that $\Delta_i \geq 0$. Then similarly as in previous for $u_\alpha \leq z \leq u_\alpha + \Delta_i$ let us write the density $f(z)$ as

$$f(z) = f(u_\alpha) + f'(u_\alpha) \cdot (z - u_\alpha) + \frac{1}{2} f''(a_3) \cdot (z - u_\alpha)^2$$

where $a_3 = a_3(z),\ \ u_\alpha \leq a_3 \leq z$ and define Δ_i^* by $u_\alpha(\beta) + \Delta_i = u_\alpha + \Delta_i^*$. Then (79) implies that for $\|\beta - \beta^0\| < \tilde{\delta}$

$$\Delta_i - K^{(1)} \cdot \left\|\beta - \beta^0\right\|^2 \leq \Delta_i^* = u_\alpha(\beta) - u_\alpha + \Delta_i \leq \Delta_i + K^{(1)} \cdot \left\|\beta - \beta^0\right\|^2 . \quad (85)$$

A chain of routine steps in evaluation of the integral $\int_{B_i^{(1)}} r_i^2(\beta) \mathrm{d}P(\omega)$ leads to the conclusion that for some $K^{(7)} < \infty$ (remember that $r_i(\beta) = e_i - \Delta_i$)

$$\int_{B_i^{(1)}} r_i^2(\beta) \mathrm{d}P(\omega) \geq \int_{u_\alpha}^{u_\alpha + \Delta_i^*} (z - \Delta_i)^2 \left[f(u_\alpha) + f'(u_\alpha) \cdot (z - u_\alpha) - J_1 \cdot (z - u_\alpha)^2 \right] \mathrm{d}z$$

$$\geq f(u_\alpha) \left[u_\alpha^2 \Delta_i^* - u_\alpha \Delta_i^2 \right] + \tfrac{1}{2} f'(u_\alpha) u_\alpha^2 \Delta_i^2 - K^{(7)} \cdot \left\|\beta - \beta^0\right\|^3 . \quad (86)$$

(By the way, along similar lines, for the case when $\Delta_i > u_\alpha(\beta) - u_\alpha$, we can arrive to the opposite inequality

$$\int_{B_i^{(1)}} r_i^2(\beta) \mathrm{d}P(\omega) \leq f(u_\alpha) \left[u_\alpha^2 \Delta_i^* - u_\alpha \Delta_i^2 \right] + \tfrac{1}{2} f'(u_\alpha) u_\alpha^2 \Delta_i^2 + K^{(7)} \cdot \left\|\beta - \beta^0\right\|^3 .$$

But we shall not need it.)

Let us turn our attention to the integral over $B_i^{(2)}$. In this case we need to treat at first the case of $\Delta_i > u_\alpha(\beta) - u_\alpha$, since then (as we have already mention it) the first interval in (84) is nonempty while the second one is empty. In other words, it means that $B_i^{(2)} = \{-u_\alpha \leq e_i < -u_\alpha(\beta) + \Delta_i\} \neq \emptyset$. Similarly as in the previous steps, when we treated $u_\alpha \leq e_i < u_\alpha(\beta) + \Delta_i$, we have now for $-u_\alpha \leq e_i < -u_\alpha(\beta) + \Delta_i$

$$f(z) = f(-u_\alpha) + f'(-u_\alpha) \cdot (z + u_\alpha) + \frac{1}{2} f''(a_4) \cdot (z + u_\alpha)^2$$
$$= f(u_\alpha) - f'(u_\alpha) \cdot (z + u_\alpha) + \frac{1}{2} f''(a_4) \cdot (z + u_\alpha)^2$$

where $-u_\alpha \leq a_4 \leq z$. Putting in this case $-u_\alpha(\beta) + \Delta_i = -u_\alpha + \Delta_{*i}$, we have again due to (79) for some $K^{(8)} < \infty$

$$\Delta_i - K^{(8)} \cdot \left\|\beta - \beta^0\right\|^2 \leq \Delta_{*i} = u_\alpha - u_\alpha(\beta) + \Delta_i \leq \Delta_i + K^{(8)} \cdot \left\|\beta - \beta^0\right\|^2 \quad (87)$$

and, carrying out again a chain of calculations, we arrive at

$$\int_{B_i^{(2)}} r_i^2(\beta) \mathrm{d}P(\omega) \leq \int_{-u_\alpha}^{-u_\alpha + \Delta_{*i}} (z - \Delta_i)^2 \left[f(u_\alpha) - f'(u_\alpha) \cdot (z + u_\alpha) + J_1 \cdot (z + u_\alpha)^2 \right] \mathrm{d}z$$

$$\leq f(u_\alpha)\left[u_\alpha^2\Delta_{*i} + u_\alpha\Delta_i^2\right] - \tfrac{1}{2}f'(u_\alpha)u_\alpha^2\Delta_i^2 + K^{(9)}\cdot\left\{\left\|\beta-\beta^0\right\|^3 + n^{-\frac{1}{2}}\right\}$$

for some $K^{(9)} < \infty$. Taking into account (86), we conclude that for some $K^{(10)} < \infty$

$$-K^{(10)}\cdot\left\|\beta-\beta^0\right\|^3$$

$$< \int_{B_i^{(1)}} r_i^2(\beta)\mathrm{d}P(\omega) - \int_{B_i^{(2)}} r_i^2(\beta)\mathrm{d}P(\omega) - f(u_\alpha)u_\alpha^2\left[\Delta_i^* - \Delta_{*i}\right] - f'(u_\alpha)u_\alpha^2\Delta_i^2 + 2\cdot f(u_\alpha)u_\alpha\Delta_i^2. \tag{88}$$

For the case when $0 < \Delta_i < u_\alpha(\beta) - u_\alpha$ both intervals in (84) are empty. Recalling that for $0 < \Delta_i < u_\alpha(\beta) - u_\alpha$ even the first interval in (83) is nonempty and hence repeating the previous lines, without estimating the integral over $B_i^{(2)}$, we conclude that (88) holds also for $0 < \Delta_i < u_\alpha(\beta) - u_\alpha$. Since the situation is symmetric in Δ_i, we obtain the same result for the case when $\Delta_i < 0$.

So, we have

$$-K^{(10)}\cdot\left\|\beta-\beta^0\right\|^3 < \frac{1}{n}\sum_{i=1}^n\left\{\int_{B_i^{(1)}} r_i^2(\beta)\mathrm{d}P(\omega) - \int_{B_i^{(2)}} r_i^2(\beta)\mathrm{d}P(\omega)\right\}$$
$$-f(u_\alpha)u_\alpha^2\frac{1}{n}\sum_{i=1}^n\left[\Delta_i^* - \Delta_{*i}\right] - f'(u_\alpha)u_\alpha^2\frac{1}{n}\sum_{i=1}^n\Delta_i^2 + 2\cdot f(u_\alpha)u_\alpha\frac{1}{n}\sum_{i=1}^n\Delta_i^2.$$

Now (82) and the equalities in (85) and (87) imply that there is a constant $K^{(11)} < \infty$

$$-\frac{f'(u_\alpha)}{f(u_\alpha)}\frac{1}{n}\sum_{i=1}^n\Delta_i^2 - K^{(11)}\cdot\left\{\left\|\beta-\beta^0\right\|^3 + n^{-\frac{1}{2}}\right\}$$
$$\leq \frac{1}{n}\sum_{i=1}^n\left(\Delta_i^* - \Delta_{*i}\right) \leq -\frac{f'(u_\alpha)}{f(u_\alpha)}\frac{1}{n}\sum_{i=1}^n\Delta_i^2 + K^{(11)}\cdot\left\{\left\|\beta-\beta^0\right\|^3 + n^{-\frac{1}{2}}\right\}.$$

Finally, we arrive at

$$\frac{1}{n}\sum_{i=1}^n\left\{\int_{B_i^{(1)}} r_i^2(\beta)\mathrm{d}P(\omega) - \int_{B_i^{(2)}} r_i^2(\beta)\mathrm{d}P(\omega)\right\}$$
$$\geq -2\cdot f(u_\alpha)u_\alpha\frac{1}{n}\sum_{i=1}^n\Delta_i^2 - K^{(12)}\cdot\left\{\left\|\beta-\beta^0\right\|^3 + n^{-\frac{1}{2}}\right\} \tag{89}$$

for some $K^{(12)} < \infty$.

So, taking into account (77) and (89), we conclude that

$$\frac{1}{n}\sum_{i=1}^n \mathsf{E}r_i^2(\beta)I\left\{r_i^2(\beta)\leq u_\alpha^2(\beta)\right\} - \mathsf{E}r_1^2(\beta^0)I\left\{r_1^2(\beta^0)\leq u_\alpha^2\right\}$$
$$\geq \frac{1}{n}\sum_{i=1}^n\left[x_i^{\mathrm{T}}\left(\beta-\beta^0\right)\right]^2\cdot\left[(1-\alpha) - 2\cdot f(u_\alpha)u_\alpha\right] + \mathcal{O}(\left\|\beta-\beta^0\right\|^3 + n^{-\frac{1}{2}}).$$

Since (due to the symmetry of $F(z)$ and strict monotonicity of $f(z)$ on R^+) for any $\alpha \in (0,1)$ we have $1-\alpha = \int_{-u_\alpha}^{u_\alpha} f(z)\mathrm{d}z > 2\cdot u_\alpha f(u_\alpha)$, we have $(1-\alpha)-2u_\alpha f(u_\alpha) > 0$. In other words, we have shown that there is a $\delta^* > 0$ such that for all $0 < \delta < \delta^*$ there is $n_\delta \in N$ and $\gamma_\delta > 0$ such that for any $n > n_\delta$

$$\min_{\beta \in B^b(\beta^0,\delta)} \frac{1}{n}\sum_{i=1}^n \mathsf{E} r_i^2(\beta) I\left\{r_i^2(\beta) \le u_\alpha^2(\beta)\right\} - \mathsf{E} r_1^2(\beta^0) I\left\{r_1^2(\beta^0) \le u_\alpha^2\right\} > \gamma_\delta.$$

So, the first part of the proof is done.

Now let us consider $\beta^{(2)} = \beta^0 + \lambda\left(\beta^{(1)} - \beta^0\right)$ for $\lambda > 1$.

First of all we show that $u_\alpha(\beta^{(1)}) \le u_\alpha(\beta^{(2)})$. Let us recall that $u_\alpha(\beta)$ is such a point that $H_e^{(\beta)}(u_\alpha(\beta)) - H_e^{(\beta)}(-u_\alpha(\beta)) = 1-\alpha$. Moreover, let us recall that

$$H_e^{(\beta)}(z) = \int_{-\infty}^{\infty} H^{(\beta)}(z-t)\mathrm{d}F(t) = \int_{-\infty}^{\infty} F(z-t)\mathrm{d}H^{(\beta)}(t).$$

Now employing Assumption $\mathcal{C}$, we can find $K^{(13)}$ and n_1 so that for all $n > n_1$

$$\sup_{\beta\in\mathcal{K}} \sup_{t\in R} \left| \frac{1}{n}\sum_{i=1}^n I\left\{x_i^{\mathrm{T}}(\beta-\beta^0) \le t\right\} - H^{(\beta)}(t)\right| < n^{-\frac{1}{2}} K^{(13)}.$$

Since $\beta^{(2)} - \beta^0 = \lambda\left(\beta^{(1)} - \beta^0\right)$, we have for all $n > n_1$

$$\sup_{t\in R} \left| H^{(\beta^{(1)})}(\frac{t}{\lambda}) - H^{(\beta^{(2)})}(t)\right| \tag{90}$$

$$\le \sup_{t\in R} \left| \frac{1}{n}\sum_{i=1}^n I\left\{x_i^{\mathrm{T}}(\beta^{(1)}-\beta^0) \le \frac{t}{\lambda}\right\} - H^{(\beta^{(1)})}(\frac{t}{\lambda})\right|$$

$$+ \sup_{t\in R} \left| \frac{1}{n}\sum_{i=1}^n I\left\{x_i^{\mathrm{T}}(\beta^{(2)}-\beta^0) \le t\right\} - H^{(\beta^{(2)})}(t)\right| < 2n^{-\frac{1}{2}} K^{(13)}.$$

Since (90) does not depend on n, we conclude $H^{(\beta^{(1)})}(\frac{t}{\lambda}) = H^{(\beta^{(2)})}(t)$ for $t \in R$. Further, due to the fact that the density of disturbances $f(z)$ is symmetric around zero and strictly decreasing on R^+, we are going to show that for any $\lambda > 1$ and $y \in R$

$$F\left(u_\alpha(\beta^{(1)}) - \lambda\cdot y\right) - F\left(-u_\alpha(\beta^{(1)}) - \lambda\cdot y\right) < F\left(u_\alpha(\beta^{(1)}) - y\right) - F\left(-u_\alpha(\beta^{(1)}) - y\right). \tag{91}$$

Consider at first that $y > 0$ and remember that $u_\alpha(\beta) > 0$. Then $\lambda\cdot y > y$,

$$u_\alpha(\beta^{(1)}) - y > u_\alpha(\beta^{(1)}) - \lambda\cdot y > -u_\alpha(\beta^{(1)}) - \lambda\cdot y$$

and $-u_\alpha(\beta^{(1)}) - y > -u_\alpha(\beta^{(1)}) - \lambda\cdot y$. Due to the fact that $f(u_\alpha(\beta^{(1)})) = f(-u_\alpha(\beta^{(1)}))$ and $f(z)$ is strictly decreasing on R^+, we conclude that

$$F(u_\alpha(\beta^{(1)}) - y) - F(u_\alpha(\beta^{(1)}) - \lambda y) > F(-u_\alpha(\beta^{(1)}) - y) - F(-u_\alpha(\beta^{(1)}) - \lambda y)$$

(it is again immediately clear from the picture of density f), i. e. (91) holds. For $y < 0$ the situation is symmetric. Then

$$\int_{-\infty}^{-\infty} \left[F(u_\alpha(\beta^{(1)}) - t) - F(-u_\alpha(\beta^{(1)}) - t)\right] \mathrm{d}H^{(\beta^{(2)})}(t)$$

$$= \int_{-\infty}^{-\infty} \left[F(u_\alpha(\beta^{(1)}) - t) - F(-u_\alpha(\beta^{(1)}) - t)\right] \mathrm{d}H^{(\beta^{(1)})}(\frac{t}{\lambda})$$

$$= \int_{-\infty}^{-\infty} \left[F(u_\alpha(\beta^{(1)}) - \lambda \cdot y) - F(-u_\alpha(\beta^{(1)}) - \lambda \cdot y)\right] \mathrm{d}H^{(\beta^{(1)})}(y)$$

$$\leq \int_{-\infty}^{-\infty} \left[F(u_\alpha(\beta^{(1)}) - y) - F(-u_\alpha(\beta^{(1)}) - y)\right] \mathrm{d}H^{(\beta^{(1)})}(y) = 1 - \alpha.$$

(Notice that in the previous chain of modifications we used correctly the substitution since (symbolically) $\mathrm{d}H^{(\beta^{(1)})}(\frac{t}{\lambda}) = h^{(\beta^{(1)})}(\frac{t}{\lambda}) \cdot \frac{1}{\lambda}\mathrm{d}t = h^{(\beta^{(1)})}(y)\mathrm{d}y = \mathrm{d}H^{(\beta^{(1)})}(y)$ and that even the strict inequality is preserved – although we shall not need it.) But it means that $u_\alpha(\beta^{(1)}) < u_\alpha(\beta^{(2)})$.

Finally let us recall that for any $\beta \in R^p$ $\quad r_i(\beta) = e_i - x_i^{\mathrm{T}}\left(\beta - \beta^0\right)$ and hence $r_i^2(\beta) \leq u_\alpha^2(\beta)$ is equivalent to

$$-u_\alpha(\beta) + x_i^{\mathrm{T}}\left(\beta - \beta^0\right) \leq e_i \leq u_\alpha(\beta) + x_i^{\mathrm{T}}\left(\beta - \beta^0\right).$$

Then

$$\mathsf{E} r_i^2(\beta) I\left\{r_i^2(\beta) \leq u_\alpha^2(\beta)\right\} = \int_{-u_\alpha(\beta)+x_i^{\mathrm{T}}(\beta-\beta^0)}^{u_\alpha(\beta)+x_i^{\mathrm{T}}\left(\beta-\beta^0\right)} \left[z - x_i^{\mathrm{T}}\left(\beta - \beta^0\right)\right]^2 \mathrm{d}F(z).$$

It means that

$$\frac{1}{n}\sum_{i=1}^{n} \mathsf{E}\left\{r_i^2(\beta^{(2)}) I\left\{r_i^2(\beta^{(2)}) \leq u_\alpha^2(\beta^{(2)})\right\} - r_i^2(\beta^{(1)}) I\left\{r_i^2(\beta^{(1)}) \leq u_\alpha^2(\beta^{(1)})\right\}\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\int_{-u_\alpha(\beta^{(2)})}^{u_\alpha(\beta^{(2)})} y^2 f(y + x_i^{\mathrm{T}}(\beta^{(2)} - \beta^0))\mathrm{d}y - \int_{-u_\alpha(\beta^{(1)})}^{u_\alpha(\beta^{(1)})} y^2 f(y + x_i^{\mathrm{T}}(\beta^{(1)} - \beta^0))\mathrm{d}y\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\int_{-u_\alpha(\beta^{(2)})}^{-u_\alpha(\beta^{(1)})} y^2 f(y + x_i^{\mathrm{T}}(\beta^{(2)} - \beta^0)\mathrm{d}y + \int_{u_\alpha(\beta^{(1)})}^{u_\alpha(\beta^{(2)})} y^2 f(y + x_i^{\mathrm{T}}(\beta^{(2)} - \beta^0)\mathrm{d}y \right. \tag{92}$$

$$\left. + \int_{-u_\alpha(\beta^{(1)})}^{u_\alpha(\beta^{(1)})} y^2 \left[f(y + x_i^{\mathrm{T}}\left(\beta^{(2)} - \beta^0\right)) - f(y + x_i^{\mathrm{T}}\left(\beta^{(1)} - \beta^0\right))\right] \mathrm{d}y\right\}. \tag{93}$$

Both integrals in (92) are nonnegative. Now let us write

$$f(y + x_i^{\mathrm{T}}\left(\beta^{(2)} - \beta^0\right)) - f(y + x_i^{\mathrm{T}}\left(\beta^{(1)} - \beta^0\right))$$

$$= f'(y + x_i^{\mathrm{T}} \left(\beta^{(1)} - \beta^{0}\right)) \cdot x_i^{\mathrm{T}} \left(\beta^{(2)} - \beta^{(1)}\right) + \frac{1}{2} f''(\xi_i) \left[x_i^{\mathrm{T}} \left(\beta^{(2)} - \beta^{(1)}\right)\right]^2$$

where $\xi_i \in \left(y + x_i^{\mathrm{T}} \left(\beta^{(2)} - \beta^{0}\right), y + x_i^{\mathrm{T}} \left(\beta^{(1)} - \beta^{0}\right)\right)_{\mathrm{ord}}$ was appropriately selected. Further

$$f'(y + x_i^{\mathrm{T}}(\beta^{(1)} - \beta^{0})) \cdot x_i^{\mathrm{T}}(\beta^{(2)} - \beta^{(1)}) = f'(y) x_i^{\mathrm{T}}(\beta^{(2)} - \beta^{(1)}) + \frac{1}{2} f''(\zeta_i) \left[x_i^{\mathrm{T}}(\beta^{(2)} - \beta^{(1)})\right]^2$$

where $\zeta_i \in \left(y, y + x_i^{\mathrm{T}} \left(\beta^{(2)} - \beta^{(1)}\right)\right)_{\mathrm{ord}}$ was again appropriately selected. Now

$$\int_{-u_\alpha(\beta^{(1)})}^{u_\alpha(\beta^{(1)})} y^2 f'(y) \mathrm{d}y = 0 \tag{94}$$

so that to finish the proof we have to cope with

$$[f''(\zeta_i) + f''(\xi_i)] \left[x_i^{\mathrm{T}} \left(\beta^{(2)} - \beta^{(1)}\right)\right]^2.$$

It can be written as

$$2 f''(y) \left[x_i^{\mathrm{T}} \left(\beta^{(2)} - \beta^{(1)}\right)\right]^2$$
$$+ \left[(f''(\zeta_i) - f''(y)) + (f''(\xi_i) - f''(y))\right] \left[x_i^{\mathrm{T}} \left(\beta^{(2)} - \beta^{(1)}\right)\right]^2.$$

The last two terms, due to Assumptions $\mathcal{A}$ or $\mathcal{B}$ and due to (64), can be bounded in absolute value by $K^{(14)} \cdot \left[x_i^{\mathrm{T}} \left(\beta^{(2)} - \beta^{(1)}\right)\right]^3$ with $K^{(14)} < \infty$. But it means that there is a constant $K^{(15)} < \infty$ so that

$$\left| \frac{1}{n} \sum_{i=1}^{n} \int_{-u_\alpha(\beta^{(1)})}^{u_\alpha(\beta^{(1)})} y^2 \left[f''(\zeta_i) - f''(y) + f''(\xi_i) - f''(y)\right] \left[x_i^{\mathrm{T}} \left(\beta^{(2)} - \beta^{(1)}\right)\right]^2 \mathrm{d}y \right|$$
$$\leq 2 u_\alpha^3(\beta^{(1)}) K^{(15)} (\lambda - 1)^3 \frac{1}{n} \sum_{i=1}^{n} \left[x_i^{\mathrm{T}} \left(\beta^{(1)} - \beta^{0}\right)\right]^3. \tag{95}$$

Finally, let us consider

$$\int_{-u_\alpha(\beta^{(1)})}^{u_\alpha(\beta^{(1)})} y^2 f''(y) \mathrm{d}y.$$

It is equal to

$$\left[y^2 f'(y)\right]_{-u_\alpha(\beta^{(1)})}^{u_\alpha(\beta^{(1)})} - 2 \int_{-u_\alpha(\beta^{(1)})}^{u_\alpha(\beta^{(1)})} y f'(y) \mathrm{d}y.$$

First of all, let us observe that the first term of the previous expression is equal to zero. Due to the assumption that the distribution F is symmetric and the density $f(y)$ is decreasing on the positive part of the real line, the expression $y f'(y)$ is negative (except of the value at point $y = 0$). It implies that

$$-\frac{2}{n} \sum_{i=1}^{n} \left[x_i^{\mathrm{T}} \left(\beta^{(1)} - \beta^{0}\right)\right]^2 \int_{-u_\alpha(\beta^{(1)})}^{u_\alpha(\beta^{(1)})} y f'(y) \mathrm{d}y > 0 \tag{96}$$

and is of order $(\lambda-1)^2$. So, writing κ instead of $\lambda-1$, we may conclude that (92), (94), (95) and (96) imply that

$$\lim_{\kappa\to 0_+}\frac{1}{\kappa}\mathsf{E}\left[r_i^2(\beta^{(2)})I\left\{r_i^2(\beta^{(2)})\le u_\alpha^2(\beta^{(2)})\right\}-r_i^2(\beta^{(1)})I\left\{r_i^2(\beta^{(1)})\le u_\alpha^2(\beta^{(1)})\right\}\right]$$

$$=\lim_{\kappa\to 0_+}\frac{1}{\kappa}\mathsf{E}\left[r_i^2(\kappa(\beta^{(1)}-\beta^0)+\beta^{(1)})\times\right.$$

$$\times\, I\left\{r_i^2(\kappa(\beta^{(1)}-\beta^0)+\beta^{(1)})\le u_\alpha^2(\kappa(\beta^{(1)}-\beta^0)+\beta^{(1)})\right\}$$

$$\left.-r_i^2(\beta^{(1)})I\left\{r_i^2(\beta^{(1)})\le u_\alpha^2(\beta^{(1)})\right\}\right]\ge 0 \qquad (97)$$

for any $\beta^{(1)}\in\mathcal{K}$ (notice that we have used in proof the relation (64) which holds, as we can trace out from the proof of Lemma 2, for $\beta\in\mathcal{K}$). In other words, derivative of

$$\mathsf{E}\left[r_i^2(\kappa(\beta^{(1)}-\beta^0)+\beta^{(1)})I\left\{r_i^2(\kappa(\beta^{(1)}-\beta^0)+\beta^{(1)})\le u_\alpha^2(\kappa(\beta^{(1)}-\beta^0)+\beta^{(1)})\right\}\right]$$

as the function of κ is nonnegative at any $\beta^{(1)}\in\mathcal{K}$ and $\kappa=0$.

That concludes the proof. □

Remark 8. Notice please, that it follows from the proof of Lemma 3 that γ_δ in (74) does not depend on $\mathcal{K}$.

Theorem 1. Let $\alpha\in(0,\frac{1}{2})$ and let Assumptions $\mathcal{A}$ or $\mathcal{B}$ and $\mathcal{C}$ hold and $\mathcal{K}$ be a compact subset of R^p, $\beta^0\in\mathcal{K}^o$. Then $\hat\beta^{(\mathrm{LTS},n,h)}$ is consistent, i. e.

$$\hat\beta^{(\mathrm{LTS},n,h)}\to\beta^0 \quad \text{in probability} \quad \text{as } n\to\infty.$$

P r o o f. Let us put

$$I_n(\beta)=\frac{1}{n}\sum_{i=1}^n r_i^2(\beta)I\left\{r_i^2(\beta)\le r_{(h_n)}^2(\beta)\right\}$$

and

$$J_n(\beta)=\frac{1}{n}\sum_{i=1}^n \mathsf{E}\left[r_i^2(\beta)I\left\{r_i^2(\beta)\le u_\alpha^2(\beta)\right\}\right].$$

Then we have for all $\omega\in\Omega$

$$I_n(\hat\beta^{(\mathrm{LTS},n,h)})\le I_n(\beta^0)$$

and hence for any $\delta>0$ we may write

$$1=P\left(\left\{\omega\in\Omega\ :\ I_n(\hat\beta^{(\mathrm{LTS},n,h)})\le I_n(\beta^0)\right\}\right)$$

$$= P\left(\left\{\omega \in \Omega \;:\; I_n(\hat{\beta}^{(\mathrm{LTS},n,h)}) \le I_n(\beta^0)\right\} \cap \left\{\omega \in \Omega \;:\; \hat{\beta}^{(\mathrm{LTS},n,h)}) \in B(\beta^0,\delta)\right\}\right)$$
$$+P\left(\left\{\omega \in \Omega \;:\; I_n(\hat{\beta}^{(\mathrm{LTS},n,h)}) \le I_n(\beta^0)\right\} \cap \left\{\omega \in \Omega \;:\; \hat{\beta}^{(\mathrm{LTS},n,h)}) \in \mathcal{K} - B(\beta^0,\delta)\right\}\right)$$
$$\le P\left(\left\{\omega \in \Omega \;:\; \hat{\beta}^{(\mathrm{LTS},n,h)}) \in B(\beta^0,\delta)\right\}\right)$$
$$+ P\left(\left\{\omega \in \Omega \;:\; \inf_{\beta \in \mathcal{K} - B(\beta^0,\delta)} I_n(\beta) < I_n(\beta^0)\right\}\right).$$

It means that if for any $\varepsilon > 0$ there is $n^{(1)}$ such that for any $n > n^{(1)}$

$$P\left(\left\{\omega \in \Omega \;:\; \inf_{\beta \in \mathcal{K} - B(\beta^0,\delta)} I_n(\beta) < I_n(\beta^0)\right\}\right) < \varepsilon,$$

also for all $n > n^{(1)}$

$$P\left(\left\{\omega \in \Omega \;:\; \hat{\beta}^{(\mathrm{LTS},n,h)}) \in B(\beta^0,\delta)\right\}\right) > 1 - \varepsilon.$$

On the other hand, we have

$$P\left(\left\{\omega \in \Omega \;:\; \inf_{\beta \in \mathcal{K} - B(\beta^0,\delta)} I_n(\beta) < I_n(\beta^0)\right\}\right)$$
$$= P\left(\left\{\omega \in \Omega \;:\; \inf_{\beta \in \mathcal{K} - B(\beta^0,\delta)} [I_n(\beta) - J_n(\beta) + J_n(\beta)] < I_n(\beta^0)\right\}\right)$$
$$\le P\left(\left\{\omega \in \Omega \;:\; \inf_{\beta \in \mathcal{K} - B(\beta^0,\delta)} [I_n(\beta) - J_n(\beta)] + \inf_{\beta \in \mathcal{K} - B(\beta^0,\delta)} J_n(\beta) < I_n(\beta^0)\right\}\right)$$
$$\le P\left(\left\{\omega \in \Omega \;:\; \inf_{\beta \in \mathcal{K}} [I_n(\beta) - J_n(\beta)] + \inf_{\beta \in \mathcal{K} - B(\beta^0,\delta)} J_n(\beta) < I_n(\beta^0)\right\}\right)$$
$$\le P\left(\left\{\omega \in \Omega \;:\; -\sup_{\beta \in \mathcal{K}} |I_n(\beta) - J_n(\beta)|\right.\right.$$
$$\left.\left. < I_n(\beta^0) - J_n(\beta^0) + J_n(\beta^0) - \inf_{\beta \in \mathcal{K} - B(\beta^0,\delta)} J_n(\beta)\right\}\right)$$
$$\le P\left(\left\{\omega \in \Omega \;:\; \sup_{\beta \in \mathcal{K}} |I_n(\beta) - J_n(\beta)|\right.\right.$$
$$\left.\left. > \inf_{\beta \in \mathcal{K} - B(\beta^0,\delta)} J_n(\beta) - J_n(\beta^0) - \left|I_n(\beta^0) - J_n(\beta^0)\right|\right\}\right). \tag{98}$$

Due to Lemma 3 for our δ there are $n_\delta \in N$ and $\gamma_\delta > 0$ such that for $n > n_\delta$

$$\inf_{\beta \in \mathcal{K} - B(\beta^0,\delta)} J_n(\beta) - J_n(\beta^0) > \gamma_\delta.$$

Now employing Corollary 1 for this γ_δ and for any $\varepsilon > 0$, we may find $n_{\varepsilon,\gamma_\delta} > n_\delta$ such that for all $n > n_{\varepsilon,\gamma_\delta}$

$$P\left(\left\{\omega \in \Omega \;:\; \sup_{\beta \in \mathcal{K}} |I_n(\beta) - J_n(\beta)| < \tfrac{1}{2}\gamma_\delta\right\}\right) > 1 - \varepsilon. \tag{99}$$

Notice, please, that (99) includes also $I_n(\beta^0) - J_n(\beta^0)$. But it means that the probability in (98) is at most equal to ε and the proof follows. □

Remark 9. Notice again that $n_{\varepsilon,\gamma_\delta}$ may depend on the compact $\mathcal{K}$ but $\mathcal{B}(\beta^0,\delta)$ inside of which $\hat{\beta}^{(\mathrm{LTS},n,h)}$ appears for $n > n_{\varepsilon,\gamma_\delta}$ with probability $1-\varepsilon$ is not influenced by $\mathcal{K}$.

The first part of paper offers the proof of consistency of $\hat{\beta}^{(\mathrm{LTS},n,h)}$. The two others (which will follow) bring $\sqrt{n}$-consistency and asymptotic representation of Bahadur type which immediately implies asymptotic normality. At the end of them we will give a discussion of results together with some considerations on the algorithms of evaluating the estimator. A brief comment on the *least weighted squares* will be given, too.

APPENDIX

Lemma A.1. (Štěpán [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 the 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 on page 276.)

Assertion A.1. 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.

P r o o f . is a straightforward computation. Let a_1 and a_2 be real numbers. Then

$$\begin{aligned} & P(\zeta_1 \cdot I\{|\zeta_1| < u\} \le a_1, \quad \zeta_2 \cdot I\{|\zeta_2| < u\} \le a_2) \\ = \; & P(-u \le \zeta_1 \le \min\{a_1, u\}, -u \le \zeta_2 \le \min\{a_2, u\}) \\ = \; & P(-u \le \zeta_1 \cdot \le \min\{a_1, u\}) \cdot P(-u \le \zeta_2 \cdot \le \min\{a_2, u\}) \\ = \; & P(\zeta_1 \cdot I\{|\zeta_1| < u\} \le a_1) \cdot P(\zeta_2 \cdot I\{|\zeta_2| < u\} \le a_2). \end{aligned}$$ □

Lemma A.2. Let us have $\sum_{i=1}^n \|x_i\| = \mathcal{O}(n)$. Then for any $\Delta \in (0,1]$ there is a $K_\Delta < \infty$ such that denoting for any $n \in N$

$$m_n = \#\{i : 1 \le i \le n, \|x_i\| > K_\Delta\}$$

we have $m_n < \Delta \cdot n$ (where "$\#A$" denotes the number of elements of the set A).

P r o o f. Due to the assumptions of lemma there is C such that for all $n \in N$ we have $\frac{1}{n}\sum_{i=1}^{n} \|x_i\| < C$. Fix $\Delta \in (0,1]$ and put $K_\Delta = \frac{C}{\Delta} + 1$. Then

$$C > \frac{1}{n}\sum_{i=1}^{n} \|x_i\| = \frac{1}{n}\left\{ \sum_{\{i:\|x_i\| \le K_\Delta\}} \|x_i\| + \sum_{\{i:\|x_i\| > K_\Delta\}} \|x_i\| \right\} > \frac{1}{n} m_n K_\Delta$$

and hence $m_n < n \cdot \frac{C}{K_\Delta} < n \cdot \Delta$. □

ACKNOWLEDGEMENT

Research was supported by the Grant Agency of the Czech Republic under Grant 02/03/0084.

(Received May 7, 2004.)

REFERENCES

[1] D. W. K. Andrews: Consistency in nonlinear econometric models: A generic uniform law of large numbers. Econometrica *55* (1987), 1465–1471.
[2] P. J. Bickel: One-step Huber estimates in the linear model. J. Amer. Statist. Assoc. *70* (1975), 428–433.
[3] L. Breiman: Probability. Addison–Wesley, London 1968.
[4] S. Chatterjee and A. S. Hadi: Sensitivity Analysis in Linear Regression. Wiley, New York 1988.
[5] M. Csőrgő and P. Révész: Strong Approximation in Probability and Statistics. Akademia Kiadó, Budapest 1981.
[6] P. J. Dhrymes: Introductory Econometrics. Springer–Verlag, New York 1978.
[7] J. Jurečková and P. K. Sen: 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] T. P. Hettmansperger and S. J. Sheather: A cautionary note on the method of least median squares. Amer. Statist. *46* (1992), 79–83.
[10] F. Liese and I. Vajda: Consistency of M-estimators in general models. J. Multivar. Anal. *50* (1994), 93–114.
[11] 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.
[12] I. Prigogine and I. Stengers: La Nouvelle Alliance. SCIENTIA 1977, Issues 5–12.
[13] I. Prigogine and I. Stengers: Out of Chaos. William Heinemann Ltd, London 1984.
[14] P. J. Rousseeuw: Least median of square regression. J. Amer. Statist. Assoc. *79* (1984), 871–880.
[15] P. J. Rousseeuw and A. M. Leroy: Robust Regression and Outlier Detection. Wiley, New York 1987.
[16] A. M. Rubio and J. Á. Víšek: A note on asymptotic linearity of M-statistics in nonlinear models. Kybernetika *32* (1996), 353–374.
[17] A. M. Rubio and J. Á. Víšek: Estimating the contamination level of data in the framework of linear regression analysis. Qüestiió *21* (1997), 9–36.
[18] J. Štěpán: Teorie pravděpodobnosti (Probability Theory). Academia, Prague 1987.

[19] S. Van Huffel: Total least squares and error-in-variables modelling: Bridging the gap between statistics, computational mathematics and enginnering. In: Proc. Computational Statistics, COMPSTAT 2004 (J. Antoch, ed.), Physica–Verlag, Springer 2004, pp. 539–555.
[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: Ekonometrie I (Econometrics I). Carolinum, Publishing House of Charles University, Prague 1997.
[23] J. Á. Víšek: Robust specification test. In: Proc. Prague Stochastics'98 (M. Hušková, P. Lachout, and J. Á. Víšek, eds.), Union of Czechoslovak Mathematicians and Physicists, Prague 1998, pp. 581–586.
[24] J. Á. Víšek: Robust instruments. In: Robust'98 (J. Antoch and G. Dohnal, eds.), Union of Czechoslovak Mathematicians and Physicists, Prague 1998, pp. 195–224.
[25] J. Á. Víšek: Robust estimation of regression model. Bull. Czech Econometric Society *9* (1999), 57–79.
[26] J. Á. Víšek: The least trimmed squares – random carriers. Bull. Czech Econometric Society *10* (1999), 1–30.
[27] 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.
[28] J. Á. Víšek: On the diversity of estimates. Comput. Statist. Data Anal. *34* (2000) 67–89.
[29] J. Á. Víšek: Regression with high breakdown point. In: Robust 2000 (J. Antoch and G. Dohnal, eds.), Union of Czechoslovak Mathematicians and Physicists, Prague 2001, pp. 324–356.
[30] J. Á. Víšek: A new paradigm of point estimation. In: Proc. Data Analysis 2000/II, Modern Statistical Methods – Modelling, Regression, Classification and Data Mining (K. Kupka, ed.), TRYLOBITE, Pardubice 2000, 195–230.
[31] 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.
[32] J. Á. Víšek: $\sqrt{n}$-consistency of empirical distribution function of residuals in linear regression model. Probab. Lett., submitted.
[33] K. Zvára: Regresní analýza (Regression Analysis). 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 08 Praha 8. Czech Republic.
e-mail: visek@mbox.fsv.cuni.cz