KYBERNETIKA — VOLUME *28* (1992), NUMBER 1, PAGES 26–36

LARGE ADAPTIVE ESTIMATION IN LINEAR REGRESSION MODEL

Part 1. Consistency

Jan Ámos Víšek

Condition of identifiability of linear regression model with symmetric distribution of errors is given. Following Beran's approach for location case consistency and asymptotic normality of this adaptive estimator is proved. The result shows that the estimator is not asymptotically efficient. But it selects model with such distribution function of errors which is (in the sense of Hellinger distance applied on $F(x)$ and $1 - F(-x)$) "as much as possible symmetric" which may be useful when we know that there are no reasons for the asymmetry.

1. INTRODUCTION

An endeavour to robustify the regression analysis yielded in the last twenty years a lot of excellent results. For an insight offering discussion see [5] and for many illustrative examples see [11]. Large attention was devoted to the methods based on L_1-norm or on a combination of L_1 and L_2-norms. For a nice review of results see [3] and references given there. Most of these methods have paid for the robustness by a decrease of efficiency. Moreover some of them were not able to cope with a "heavy" contamination or with leverage points. On the other hand, in some cases highly robust methods may yield an overdetermined model. A hope to solve some of these difficulties seems to be offered by adaptive estimation.

Since the decrease of efficiency is not usually dramatic the main reason to use this adaptive method may be the "symmetry" of residuals of estimated model. Since symmetry of distribution of errors is the (basic) assumption for consistency of many methods of robust regression (e. g. Least Median of Squares) it may be considered also as an attempt to check this assumption. It means that when the model found by the further introduced adaptive procedure is not far from a model obtained by a robust procedure, let us say by the Least Median of Squares, then we may accept the latter model because the assumption under which the model was derived is, at least approximately, fulfilled. In the opposite case we should be more careful and either try to separate the data into (two) groups and build up models for the each group or to accept the "adaptive" model, for numerical example see [18]. (For a detailed discussion of this topic see [16].)

An idea of an adaptive estimation of parameters of unknown type of distribution goes back to [14] and later was discussed in generality by Bickel [2]. For the location model the problem was solved already in seventies by Stone [15] and Beran [1].

We have followed closely the approach of Beran [1] and extended it for the regression model. It revealed (again) the fact that linear regression model is not a mere generalization of location model, compare [4], and hence a difficulty with identifiability of coefficients may occur. It will be shown on an example. Let us mention that alternative approaches to adaptive estimation in linear models were described in [6], [7], [9] and [18].

2. NOTATION

Let us denote by $\mathcal{N}$ the set of all positive integers, by $\mathcal{R}$ the real line and by $\mathcal{R}^n$ the n-dimensional Euclidean space. We shall consider a linear model

$$Y = X \cdot \beta^0 + e, \tag{1}$$

where $Y' = (Y_1, \ldots, Y_n)^T$ is a real vector (response variable), $X = (x_{ij})_{i=1,j=1}^{n\ \ p}$ a known and fixed design matrix, $\beta^{0'} = (\beta_1^0, \ldots, \beta_p^0)^T$ a vector of unknown (but fixed) parameters and $e = (e_1, \ldots, e_n)^T$ a vector of i. i. d. random variables following distribution function (d. f.) G (we implicitly assume that $\{e_i\}_{i=1}^{\infty}$ are defined on a space $(\Omega, \mathcal{A}, P)$). We assume that the intercept, if any, is included in the design matrix, i. e. $x_{i1} = 1$ for $i = 1, \ldots, n$. The d. f. G is assumed to allow a density g with respect to Lebesgue measure which is symmetric around zero, i. e. for any $x \in \mathcal{R}$ $g(x) = g(-x)$. The assumption of symmetry may be omitted, but without it the intercept has to be estimated separately from other coefficients and not adaptively. Naturally, the whole theory have to be modified, too. The data have to be divided into two parts and Hellinger distance of the estimates of density for these halves must be minimized. Although the symmetry is not acceptable in so many situations as it is sometimes believed there are cases which admit symmetry quite well. Let us consider for a while location model and assume that we are in a situation, may be rare, when we do not want to specify type of the parametric model at all. Then only under the assumption of the symmetry the sense of "location" is out of any discussion since modus (if unimodal), median, mean (if exists) and center of symmetry coincide. May be that it is the reason why some practitioner, in the case when data are apparently not symmetric, look for a (one-to-one) transformation which brings (bulk of) data to the symmetry and having estimated location as the center of symmetry they apply inverse transformation. Similar facts are also true for regression analysis, especially in situation when for all data a model cannot be "reasonably" found. When we are able to choose a subsample of data and a regression model (for this subsample) implying approximately symmetric density of residuals we may claim (without any additional assumption on distribution of errors e_i) – in at least intuitively reasonable and clear sense – that the errors e_i have no systematic influence on response variable.

In what follows we shall use kernel estimator of density of residuals. Let us denote by w a kernel which is assumed to be symmetric, twice absolutely continuous, positive everywhere and

$$\sup_{y\in\mathcal{R}} w(y) < K_1,$$

$$\sup_{y\in\mathcal{R}} \frac{|w'(y)|}{w(y)} < K_2$$

and

$$\sup_{y\in\mathcal{R}} \frac{|w''(y)|}{w(y)} < K_3$$

where K_1, K_2 and K_3 are some (positive) real numbers. By $\{c_n\}_{n=1}^{\infty} \searrow 0$ we shall denote the bandwidth of the kernel estimator. Further for $y \in \mathcal{R}$, $Y \in \mathcal{R}^n$ and $\beta \in \mathcal{R}^p$, let us denote by

$$g_n(y, Y, \beta) = \frac{1}{nc_n} \sum_{i=1}^{n} w\left(c_n^{-1}(y - (Y_i - \sum_{j=1}^{p} x_{ij}\beta_j))\right)$$

the above mentioned kernel estimator of density of residuals. In the sequel we shall use $X_i^T\beta$ as an alternative notation for $\sum_{j=1}^{p} x_{ij}\beta_j$. Moreover let $0 \le b(y) \le 1$ be a continuous function with $b(0) = 1$ and $b(y) = 0$ for $|y| > 1$. Then for a sequence of positive constants $\{a_n\}_{n=1}^{\infty} \nearrow \infty$ and for any $y \in \mathcal{R}$ define $\{b_n(y)\}_{n=1}^{\infty}$ as follows:

$$\begin{aligned} b_n(y) &= 1, \qquad |y| \le a_n, \\ &= b\left(\frac{|y| - a_n}{c_n^4}\right) \qquad a_n < |y| \le a_n + c_n^4, \end{aligned}$$

and

$$= 0 \qquad \text{otherwise.}$$

Finally put

$$h_n(y, Y, \beta) = b_n(y) g_n^{1/2}(y, Y, \beta).$$

3. PRELIMINARIES

Let us recall that the "true" value of β was denoted by β^0 (see (1)).

Lemma 1. For any $\beta \in R^p$

$$\int \left[h_n(y, Y, \beta) - \mathsf{E}^{\frac{1}{2}} g_n(y, Y, \beta) \cdot b_n(y)\right]^2 \mathrm{d}y = O_p(n^{-1}c_n^{-1}a_n).$$

P r o o f. Since for any $a \ge 0$ and $b > 0$ we have $(a - b)^2 \le b^{-2}(a^2 - b^2)^2$ we may write

$$\begin{aligned}
& \mathsf{E}\left[h_n(y,Y,\beta) - \mathsf{E}^{\frac{1}{2}} g_n(y,Y,\beta)\cdot b_n(y)\right]^2 \le \\
\le\ & b_n^2(y)\,\mathsf{E}^{-1} g_n(y,Y,\beta)\,\mathsf{E}\left[g_n(y,Y,\beta) - \mathsf{E}g_n(y,Y,\beta)\right]^2 = \\
=\ & b_n^2(y)\mathsf{E}^{-1} g_n(y,Y,\beta)\mathsf{E}\left\{\frac{1}{nc_n}\sum_{i=1}^n \left[w(c_n^{-1}(y-Y_i+X_i^T\beta)) - \mathsf{E}w(c_n^{-1}(y-Y_i+X_i^T\beta))\right]\right\}^2 \\
=\ & b_n^2(y)E^{-1} g_n(y,Y,\beta)\frac{1}{n^2c_n^2}\sum_{i=1}^n \mathsf{E}\left[w(c_n^{-1}(y-Y_i+X_i\beta)) - \mathsf{E}w(c_n^{-1}(y-Y_i+X_i^T\beta))\right]^2 = \\
\le\ & b_n^2(y)\mathsf{E}^{-1} g_n(y,Y,\beta)\,\frac{1}{n^2c_n^2}\sum_{i=1}^n \mathsf{E}w^2(c_n^{-1}(y-Y_i+X_i^T\beta)) \le \\
\le\ & \sup_{z\in\mathcal{R}} w(z)\cdot b_n^2(y)\mathsf{E}^{-1} g_n(y,Y,\beta)\,\frac{1}{n^2c_n^2}\sum_{i=1}^n \mathsf{E}w(c_n^{-1}(y-Y_i+X_i^T\beta)) = \\
=\ & \frac{1}{nc_n}\sup_{z\in\mathcal{R}} w(z)\cdot b_n^2(y).
\end{aligned}$$

□

Notice that

$$\mathsf{E}g_n(y,Y,\beta) = \frac{1}{nc_n}\sum_{i=1}^n \int w\left(c_n^{-1}(y-z+X_i^T(\beta^0-\beta))\right) g(z)\mathrm{d}z$$

and

$$\frac{\partial \mathsf{E}g_n(y,Y,\beta)}{\partial\beta_k} = \frac{1}{nc_n^2}\sum_{i=1}^n x_{ik}\int w'\left(c_n^{-1}(y-z+X_i^T(\beta^0-\beta))\right) g(z)\mathrm{d}z.$$

We shall denote $\left[\frac{\partial \mathsf{E}g_n(y,Y,\beta)}{\partial\beta_k}\right]_{\beta=\beta^0}$ by $\frac{\partial \mathsf{E}g_n(y,Y,\beta^0)}{\partial\beta_k}$. Notice also that in fact we have shown that for any $y\in\mathcal{R}$

$$E\left[g_n^{\frac{1}{2}}(y,Y,\beta) - \mathsf{E}^{\frac{1}{2}} g_n(y,Y,\beta)\right]^2 \le (nc_n)^{-1}\sup_{z\in\mathcal{R}} w(z).$$

We shall need it in the proof of

Lemma 2. Let $\lim_{n\to\infty} n^{-1}c_n^{-1}a_n = 0$. Then

$$\int h_n(y,Y,\beta^0)\cdot h_n(-y,Y,\beta^0)\mathrm{d}y \to 1 \ \text{ in probability }.$$

Remark 1. The assumption of Lemma 2 implies a usual requirements that $nc_n\to\infty$ as $n\to\infty$.

P r o o f o f L e m m a 2. We may write

$$\left|\int \left[h_n(y,Y,\beta^0)\cdot h_n(-y,Y,\beta^0) - b_n^2(y)\mathsf{E}^{\frac{1}{2}}g_n(y,Y,\beta^0)\cdot \mathsf{E}^{\frac{1}{2}}g_n(-y,Y,\beta^0)\right]\mathrm{d}y\right|$$

$$\leq \int h_n(y,Y,\beta^0)\cdot \left|h_n(-y,Y,\beta^0) - \mathsf{E}^{\frac{1}{2}}g_n(-y,Y,\beta^0)\cdot b_n(y)\right|\mathrm{d}y +$$

$$+ \int \left|h_n(y,Y,\beta^0) - \mathsf{E}^{\frac{1}{2}}g_n(y,Y,\beta^0)\right|\mathsf{E}^{\frac{1}{2}}g_n(-y,Y,\beta^0)b_n(y)\mathrm{d}y.$$

The first integral is not greater than

$$\left\{\int h_n^2(y,Y,\beta^0)\mathrm{d}y \int \left[h_n(-y,Y,\beta^0) - E^{\frac{1}{2}}g_n(-y,Y,\beta^0)b_n(y)\right]^2\mathrm{d}y\right\}^{\frac{1}{2}}.$$

Similar upper bound is easy to derive for the second integral and hence both are (according to Lemma 1) $O_p(n^{-\frac{1}{2}}c_n^{-\frac{1}{2}}a_n^{\frac{1}{2}})$. A straightforward computation gives (due to the symmetry of w and g)

$$\int \mathsf{E}^{\frac{1}{2}}g_n(y,Y,\beta^0)\cdot \mathsf{E}^{\frac{1}{2}}g_n(-y,Y,\beta^0)\mathrm{d}y = 1$$

which implies

$$\int \mathsf{E}^{\frac{1}{2}}g_n(y,Y,\beta^0)\cdot \mathsf{E}^{\frac{1}{2}}g_n(-y,Y,\beta^0)\ (1-b_n^2(y))\mathrm{d}y \to 0.$$

That concludes the proof. □

Corollary 1. Let $\lim_{n\to\infty} n^{-1}c_n^{-1}a_n = 0$. Then

$$\sup_{\beta\in\mathcal{R}^p}\int h_n(y,Y,\beta)\cdot h_n(-y,Y,\beta)\mathrm{d}y \to 1$$

in probability.

P r o o f. Since from the Cauchy–Schwarz inequality we have

$$\int h_n(y,Y,\beta)\cdot h_n(-y,Y,\beta)\mathrm{d}y \leq 1,$$

the proof follows from Lemma 2. □

Due to Corollary 1 we may give the following definition.

4. DEFINITION OF ESTIMATOR

Definition 1. For any $Y \in \mathcal{R}^\infty$ let us denote by $\hat\beta_{(n)}(Y)$ points $\beta\in\mathcal{R}^p$ for which $\int h_n(y,Y,\beta)h_n(-y,Y,\beta)\mathrm{d}y$ reaches its maximum. If there is no such point, let us understand under $\hat\beta_{(n)}(Y)$ a point(s) $\beta^*\in\mathcal{R}^p$ for which

$$\int h_n(y,Y,\beta^*)\cdot h_n(-y,Y,\beta^*)\mathrm{d}y > \sup_{\beta\in\mathcal{R}^p}\int h_n(y,Y,\beta)\cdot h_n(-y,Y,\beta)\mathrm{d}y - \frac{1}{n}.$$

Remark 2. For evaluation of $\hat{\beta}_{(n)}(Y)$ only first n coordinates of Y are used, so that the above definition may start with $Y \in \mathcal{R}^n$. On the other hand in the following it will be more convenient to assume in every assertion the infinitely dimensional space $\mathcal{R}^\infty$.

It will be shown in the next section that the design matrix X has to fulfill some conditions to allow us to prove consistency of $\beta_n(Y)$.

5. IDENTIFIABILITY CONDITION

Let us consider a very simple example with $p = 1$ and $X_{i1} = (-1)^{i+1}$ for every $i = 1, \ldots, n$ and $n \in \mathcal{N}$. Further let the sequence of r. v.'s $\{e_i\}_{i=1}^\infty$ be i. i. distributed according to standard normal law. Now let us fix $\beta^1 \in \mathcal{R}$, $\beta^1 \neq 0$ and assume $Y_i = X_{i1} \cdot \beta^1 + e_i$ for $i = 1, \ldots, n$. Finally consider our estimator which is based on residuals $Y_i - X_{i1} \cdot \beta^1$ $(= e_i)$ and which utilizes the assumption of their symmetric distribution. Let us put ourselves a question: Is (or are) there any other $\tilde{\beta}$ (or $\tilde{\beta}$'s) $\in \mathcal{R}$ such that the residuals $Y_i - X_{i1} \cdot \tilde{\beta}$ may have a symmetric distribution and hence our estimator cannot distinguish between β^1 and $\tilde{\beta}$? The answer is, unfortunately, positive. We see that even for any $\beta \in \mathcal{R}$ we obtain for the odd indexes i

$$e_i' = Y_i - X_{i1} \cdot \beta = Y_i - X_{i1} \cdot \beta^1 + X_{i1}(\beta^1 - \beta) = e_i + \beta^1 - \beta$$

and for the even ones $e_i' = e_i - \beta^1 + \beta$ and therefore any reasonable density estimator applied on the sequence $\{e\}_{i=1}^\infty$ will yield an estimate converging to the density corresponding to the mixture

$$\frac{1}{2}\left[N(\beta^1 - \beta, 1) + N(\beta - \beta^1, 1)\right]$$

where $N(\mu, \sigma^2)$ denotes normal distribution with mean μ and variance σ^2.

This simple example shows that under the mere assumption of symmetry of d. f. of e_i we cannot prove such property as consistency of $\hat{\beta}_n(Y)$. There exist a few different remedies. We may for instance require not only the minimization of Hellinger distance but also minimization of variance of r. v. corresponding to estimated density of errors. It is clear that it may be misleading since it may change the "true" variance of residuals (at least). On the other hand, we may arrive at a pragmatical model with better predictive, and maybe even explaining properties than the "true" model.

Let us return to our example. We see that the source of the described difficulties lies in symmetry of the design matrix which may be interpreted as realization of a sequence of i. i. d. r. v.'s $\{Z_i\}_{i=1}^\infty$ such that

$$P(Z_i = -1) = P(Z_i = 1) = \frac{1}{2}.$$

We may then consider Y_i being sum of two random variables, both symmetrically distributed and hence unseparable by our estimator.

Moreover, the design matrix is – in some sense – a tool, say a microscope, through which we observe response variable and in many – not at all – cases we may prescribe its properties (and check them). Hence it seems (quite) natural to assume something about it. Even in the case when X represents a (realization of) sequence of random vectors we may sometimes prefer to restrict character of this sequence than to restrict character of errors.

So our condition has to remove the "symmetry" of the design matrix. Another thing which we need when proving the consistency of estimator is some compactness restriction which holds without any assumption for the case of location parameter (see [1], Lemma 2) but which, as it is easy to show, may generally fail for regression model.

Denote by

$$\mathcal{C}_p(a, \beta^0) = \{\beta \in \mathcal{R}^p : \ \|\beta - \beta^0\| > a\}$$

and by

$$\mathcal{C}_p(a, b, \beta^0) = \mathcal{C}_p(a, \beta^0) \cap \mathcal{C}_p^c(b, \beta^0).$$

Condition A. For any $\delta > 0$ there exist a $\Delta \in (0, 1)$ and $K_\Delta \in \mathcal{R}$ such that

i)

$$\limsup_{n\to\infty} \sup_{\beta \in \mathcal{C}_p(\delta, K_\Delta, \beta^0)} \int \mathsf{E}^{\frac{1}{2}} g_n(y, Y, \beta) \cdot \mathsf{E}^{\frac{1}{2}} g_n(-y, Y, \beta) \mathrm{d}y < \Delta$$

and

ii)

$$\limsup_{n\to\infty} \sup_{\beta \in \mathcal{C}_p(K_\Delta, \beta^0)} \int h_n(y, Y, \beta) h_n(-y, Y, \beta) \mathrm{d}y < \Delta \quad \text{in probability.}$$

Moreover let $K_4 \in \mathcal{R}$ be such that

$$\sup_{i \in \mathcal{N}} \sup_{j=1,\dots,p} |x_{ij}| < K_4.$$

Remark 3. The problem of identifiability may be probably solved also under another conditions similar to those of [10]. We have preferred more "direct" ones. It is easy to see that the first condition guarantees that the large values of estimator are senseless. The second assures that the kernel estimate behaves similarly as the "true" density.

6. CONSISTENCY OF ESTIMATOR

Now we are going to give the main result of the paper.

Theorem 1. Let the Condition A be fulfilled and

$$\lim_{n\to\infty} nc_n^{4p} a_n^{-2p} = \infty.$$

Then $\hat{\beta}_{(n)}(Y)$ is a consistent estimator of β^0.

Proof. Let us fix an $\varepsilon \in (0,1)$ and $\delta > 0$ and find $\Delta \in (0,1)$, K_4 and K_Δ from Condition A.

Since β^0 is the fixed ("true") value we may find $\tilde{K}$ such that for any $\beta \in \mathcal{R}^p$ such that $\sup_{j=1,2,\dots,p} |\beta_j| > \tilde{K}$ we have $\|\beta - \beta^0\| > K_\Delta$. From ii) of Condition A it follows that there exists $n_1 \in \mathcal{N}$, so that for $n > n_1$ we have

$$P\left\{ \sup_{\beta \in \mathcal{C}(K_\Delta, \beta^0)} \int h_n(y,Y,\beta)\, h_n(-y,Y,\beta)\, \mathrm{d}y > \Delta \right\} \le \frac{\varepsilon}{4}. \tag{2}$$

Similarly from Corollary 1 follow that there exists $n_2 > n_1$ such that for any $n \in \mathcal{N},\ n > n_2$ we have

$$P\left\{ \sup_{\beta \in \mathcal{R}^p} \int h_n(y,Y,\beta)\, h_n(-y,Y,\beta)\, \mathrm{d}y < \Delta + \left(\tfrac{1-\Delta}{2}\right) \right\} < \frac{\varepsilon}{4}. \tag{3}$$

(In fact (2) and (3) implies that

$$P\left\{ \sup_{j=1,\dots,p} |\hat{\beta}_{(n)j}| > \tilde{K} \right\} < \frac{\varepsilon}{2}.)$$

Denote by $\mathcal{K} = \{\beta \in \mathcal{R}^p : \sup_{j=1,\dots,p} |\beta_j| < \tilde{K}\}$. Now for every $n \in \mathcal{N}$ find a set of points from $\mathcal{R}^p$ say $\{\beta^1, \beta^2, \dots, \beta^r\}$, such that for any $\beta \in \mathcal{K}$ there is an $\ell_0 \in \{1,2,\dots,r\}$ such that $\|\beta - \beta^{\ell_0}\| < n^{-\frac{1}{2p}}$ and r is the smallest possible integer. Then we have

$$\begin{aligned} &\left| \int h_n(y,Y,\beta)\, h_n(-y,Y,\beta) \mathrm{d}y - \int h_n(y,Y,\beta^{\ell_0})\, h_n(-y,Y,\beta^{\ell_0})\, \mathrm{d}y \right| \quad (4) \\ \le\ & \left\{ \int h_n^2(y,Y,\beta) dy \int \left[h_n(-y,Y,\beta) - h_n(y,Y,\beta^{\ell_0})\right]^2 \mathrm{d}y \right\}^{\frac{1}{2}} + \\ +\ & \left\{ \int \left[h_n(y,Y,\beta) - h_n(y,Y,\beta^{\ell_0})\right]^2 \mathrm{d}y \int h_n^2(-y,Y,\beta^{\ell_0}) \mathrm{d}y \right\}^{\frac{1}{2}}. \end{aligned}$$

Since $\int h_n^2(y,Y,\beta)\mathrm{d}y$ is not greater than one it suffices to find an upper bound for $\int \left[h_n(y,Y,\beta) - h_n(y,Y,\beta^{\ell_0})\right]^2 \mathrm{d}y$. Making use of the inequality $[a-b]^2 \le 2|a^2 - b^2|$ valid for nonnegative a and b we obtain

$$\begin{aligned} &\int \left[h_n(y,Y,\beta) - h_n(y,Y,\beta^{\ell_0})\right]^2 \mathrm{d}y \le \\ \le\ & \frac{2}{nc_n} \sum_{i=1}^{n} \int \left| w\left(c_n^{-1}(y - Y + X_i\beta)\right) - w\left(c_n^{-1}(y - Y + X_i\beta^{\ell_0})\right) \right| \mathrm{d}y. \end{aligned}$$

But

$$\begin{aligned} & \left| w\left(c_n^{-1}(y - Y + X_i\beta)\right) - w\left(c_n^{-1}(y - Y + X_i\beta^{\ell_0})\right)\right| = \\ = \; & \left| c_n^{-1} w'(c_n^{-1}\xi_i) \cdot X_i[\beta - \beta^{\ell_0}]\right| = \\ = \; & \left| c_n^{-1} \frac{w'(c_n^{-1}\xi_i)}{w(c_n^{-1}\xi_i)} \cdot w(c_n^{-1}\xi_i) \cdot X_i[\beta - \beta^{\ell_0}]\right| \le \\ \le \; & c_n^{-1} \cdot K_1 \cdot K_2 \cdot K_4 \cdot p \cdot n^{-\frac{1}{2p}}. \end{aligned}$$

Hence

$$\begin{aligned} & \sup_{Y \in \mathcal{R}^\infty} \sup_{\beta \in \mathcal{K}} \min_{\ell = 1, \dots r} \left| \int h_n(y, Y, \beta) h_n(-y, Y, \beta) \mathrm{d}y - \right. \\ & \left. - \int h_n(y, Y, \beta^\ell) h_n(-y, Y, \beta^\ell) \mathrm{d}y \right| = O\left(n^{-\frac{1}{2p}} c_n^{-2} a_n\right). \end{aligned}$$

i.e. it converges to zero. Let us choose $n_3 > n_2$ so that for any $n \in \mathcal{N}, n > n_3$ the just studied difference (4) is less than $(1 - \Delta)/8$. Further we have for any $\ell \in \{1, \dots, r\}$

$$\begin{aligned} & \left| \int h_n(y, Y, \beta^\ell) h_n(-y, Y, \beta^\ell) \mathrm{d}y - \int \mathsf{E}^{\frac{1}{2}} h_n^2(y, Y, \beta^\ell) \mathrm{d}y \cdot \mathsf{E}^{\frac{1}{2}} h_n^2(-y, Y, \beta^\ell) \mathrm{d}y \right| \\ \le \; & \left| \int h_n(y, Y, \beta^\ell) \left[h_n(-y, Y, \beta^\ell) - \mathsf{E}^{\frac{1}{2}} h_n^2(-y, Y, \beta^\ell)\right] \mathrm{d}y \right| + \\ + \; & \left| \int \mathsf{E}^{\frac{1}{2}} h_n^2(-y, Y, \beta^\ell) \left[h_n(y, Y, \beta^\ell) - \mathsf{E}^{\frac{1}{2}} h_n^2(y, Y, \beta^\ell)\right] \mathrm{d}y \right|. \end{aligned} \tag{5}$$

Now

$$\begin{aligned} & \left[\int h_n(y, Y, \beta^\ell) \left[h_n(-y, Y, \beta^\ell) - \mathsf{E}^{\frac{1}{2}} h_n^2(-y, Y, \beta^\ell)\right] \mathrm{d}y \right]^2 \le \\ \le \; & \int h_n^2(y, Y, \beta^\ell) \mathrm{d}y \int \left[h_n(-y, Y, \beta^\ell) - \mathsf{E}^{\frac{1}{2}} h_n^2(-y, Y, \beta^\ell)\right]^2 \mathrm{d}y \le \\ \le \; & \int \left[h_n(-y, Y, \beta^\ell) - \mathsf{E}^{\frac{1}{2}} h_n^2(-y, Y, \beta^\ell)\right]^2 \mathrm{d}y \end{aligned}$$

and similarly for the second term of the right-hand side of (5). Using once again inequality from the proof of Lemma 1 we obtain

$$\begin{aligned} & P\left\{ \left| \int \left[h_n(y, Y, \beta^\ell) h_n(-y, Y, \beta^\ell) - \mathsf{E}^{\frac{1}{2}} h_n^2(y, Y, \beta^\ell) \cdot \mathsf{E}^{\frac{1}{2}} h_n^2(-y, Y, \beta^\ell)\right] \mathrm{d}y \right| > \left(\tfrac{1-\Delta}{8}\right) \right\} \le \\ \le \; & P\left\{ \int \left[\mathsf{E}^{-1} h_n^2(-y, Y, \beta^\ell) \cdot \left[h_n^2(-y, Y, \beta^\ell) - \mathsf{E} h_n^2(-y, Y, \beta^\ell)\right]\right]^2 \mathrm{d}y > \frac{1}{4} \left(\left(\tfrac{1-\Delta}{8}\right)\right)^2 \right\} + \\ + \; & P\left\{ \int \left[\mathsf{E}^{-1} h_n^2(y, Y, \beta^\ell) \cdot \left[h_n^2(y, Y, \beta^\ell) - \mathsf{E} h_n^2(y, Y, \beta^\ell)\right]\right]^2 \mathrm{d}y > \frac{1}{4} \left(\left(\tfrac{1-\Delta}{8}\right)\right)^2 \right\}. \end{aligned}$$

But the probabilities may be bounded by

$$4 \cdot (8/(1-\Delta))^2 \, \mathsf{E} \left\{ \int \mathsf{E}^{-1} h_n^2(y,Y,\beta^\ell). \left[h_n^2(y,Y,\beta^\ell) - \mathsf{E} h_n^2(y,Y,\beta^\ell) \right]^2 \mathrm{d}y \right\} =$$

$$= \quad 4 \cdot (8/(1-\Delta))^2 \int \left\{ \mathsf{E}^{-1} h_n^2(y,Y,\beta^\ell) \cdot \mathsf{E} \left[h_n^2(y,Y,\beta^\ell) - \mathsf{E} h_n^2(y,Y,\beta^\ell) \right]^2 \right\} \mathrm{d}y$$

and proceeding further as in the proof of Lemma 1 one obtains

$$P \left\{ \left| \int \left[h_n(y,Y,\beta^\ell) \cdot h_n(-y,Y,\beta^\ell) - \mathsf{E}^{\frac{1}{2}} h_n^2(y,Y,\beta^\ell) \cdot \mathsf{E}^{\frac{1}{2}} h_n^2(-y,Y,\beta^\ell)^2 \right] \mathrm{d}y \right| > \left(\left(\tfrac{1-\Delta}{8} \right) \right) \right\} <$$

$$< \quad 8 \cdot (8/(1-\Delta))^2 \cdot \frac{1}{nc_n} \sup_{z \in \mathcal{R}} w(z) \cdot 2(a_n+1).$$

Notice that the upper bound does not depend on $\ell \in \{1, \ldots, r\}$. Hence

$$\begin{aligned} & P \left\{ \sup_{\ell=1,\ldots r} \left| \int \left[h_n(y,Y,\beta^\ell) . h_n(-y,Y,\beta^\ell) - \right. \right. \right. \\ & \left. \left. \left. - \mathsf{E}^{\frac{1}{2}} h_n^2(y,Y,\beta^\ell) \cdot \mathsf{E}^{\frac{1}{2}} h_n^2(-y,Y,\beta^\ell) \right] \mathrm{d}y \right| < \left(\left(\tfrac{1-\Delta}{8} \right) \right) \right\} > \\ < \quad & 1 - r \cdot 8. \left(\tfrac{8}{1-\Delta} \right)^2 \cdot \frac{1}{nc_n} \cdot \sup_{z \in \mathcal{R}} w(z) \cdot 2(a_n+1) > \\ > \quad & 1 - \left[2\tilde{K} \cdot n^{\frac{1}{2p}} \right]^p \cdot 8 \cdot \tfrac{8}{1-\Delta}^2 \cdot \frac{1}{nc_n} \cdot \sup_{z \in \mathcal{R}} w(z) \cdot 2(a_n+1) = \\ = \quad & 1 - O\left(n^{\frac{1}{2}} c_n^{-1} \cdot a_n \right). \end{aligned}$$

Find an $n_4 \in \mathcal{N}, \ n_4 > n_3$ such that this probability is less than $\frac{\varepsilon}{2}$. Finally using i) of Condition A select $n_5 > n_4$ so that for any $n \in \mathcal{N}, \ n > n_5$

$$\sup_{\beta \in \mathcal{C}_p(\delta, K_\Delta, \beta^0)} \int \mathsf{E}^{\frac{1}{2}} g_n(y,Y,\beta) \cdot \mathsf{E}^{\frac{1}{2}} g_n(-y,Y,\beta) dy < \Delta + (1-\Delta)/8.$$

So we have derived that for any $n \in \mathcal{N}, n > n_5$ there is a set, say C, such that $P(C) > 1 - \frac{\varepsilon}{2}$ and for any $\omega \in C$ we have

$$\sup_{\beta \in \mathcal{C}_p(\delta, K_\Delta, \beta^0)} \int h_n(y,Y,\beta) \cdot h_n(-y,Y,\beta) \mathrm{d}y < \Delta + 3(1-\Delta)/8. \tag{6}$$

Finally find $n_6 \in \mathcal{N}, \ n_6 > n_5$ such that $n_6^{-1} < (1-\Delta)/8$. Then taking into account (2), (3), (6) and the way how n_3 and n_6 were selected, we obtain for any $n \in \mathcal{N}, \ n > n_6$

$$P \left\{ \sup_{\beta \in \mathcal{C}_p(\delta, \beta^0)} \int h_n(y,Y,\beta) \cdot h_n(-y,Y,\beta) \mathrm{d}y < \Delta + 3(1-\Delta)/8 \text{ and} \right.$$

$$\left. \sup_{\beta \in \mathcal{R}^p} \int h_n(y,Y,\beta) \cdot h_n(-y,Y,\beta) \mathrm{d}y > \Delta + (1-\Delta)/2 \right\} > 1 - \varepsilon,$$

which concludes the proof. □

The asymptotic normality of $\hat{\beta}_n(Y)$ together with numerical examples will be presented in the second part of this paper.

(Received June 19, 1990.)

REFERENCES

[1] R. Beran: An efficient and robust adaptive estimator of location. Ann. Statist. *6* (1978), 292 – 313.
[2] P. J. Bickel: The 1980 Wald Memorial Lectures – On adaptive estimation. Ann. Statist. *10* (1982), 647 – 671.
[3] Y. Dodge: An introduction to statistical data analysis L_1-norm based. In: Statistical Data Analysis Based on the L_1-norm and Related Methods (Y. Dodge, ed.), North-Holland, Amsterdam 1987.
[4] J. Jurečková: Regression quantiles and trimmed least square estimator under a general design. Kybernetika *20* (1984), 345 – 357.
[5] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel: Robust Statistics: The Approach Based on Influence Functions. J. Wiley, New York 1986.
[6] G. Heimann: Adaptive und robuste Schätzer in Regression Modellen. Ph. D. Diseration, Institut für Math. Stochastik, Universität Hamburg 1988.
[7] R. Koenker: A lecture read on Charles University during his visit in Prague 1989.
[8] R. Koenker and G. Basset: Regression quantiles. Econometrica *46* (1978), 33 – 50.
[9] H. Koul and F. DeWet: Minimum distance estimation in a linear regression model. Ann. Statist. *11* (1983), 921 – 932.
[10] R. A. Maronna and V. J. Yohai: Asymptotic behaviour of general M-estimates for regression and scale with random carriers. Z. Wahrsch. verw. Gebiete *58* (1981), 7 – 20.
[11] P. J. Rousseeuw and A. M. Leroy: Robust Regression and Outlier Detection. J. Wiley, New York 1987.
[12] D. Ruppert and R. J. Carroll: Trimmed least squares estimation in linear model. J. Amer. Statist. Assoc. *75* (1980), 828 – 838.
[13] A. V. Skorokhod: Limit theorems for stochastic processes. Teor. Veroyatnost. i Primenen. *1* (1956), 261 – 290.
[14] C. Stein: Efficient nonparametric testing and estimation. In: Proc. Third Berkeley Symp. Math. Statist. Prob. *1* (1956), 187 – 196. Univ. of California Press, Berkeley, Calif. 1956.
[15] C. Stone: Adaptive maximum likelihood estimators of a location parameter. Ann. Statist. *3* (1975), 267 – 284.
[16] J. Á. Víšek: What is adaptivity of regression analysis intended for? Transactions of ROBUST'90, JČMF, Prague 1990, pp. 160 – 181.
[17] J. Á. Víšek: Adaptive Estimation in Linear Regression Model. Research Report No. 1642, Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, Prague 1990.
[18] J. Á. Víšek: Adaptive Maximum-likelihood-like Estimation in Linear Model. Research Report No. 1654, Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, Prague 1990.
[19] J. Á. Víšek: Adaptive estimation in linear regression model and test of symmetry of residuals. To appear in the Proceedings of the Second International Workshop on Model-Oriented Data Analysis, Springer-Verlag, Wien.

RNDr. Jan Ámos Víšek, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation – Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.