KYBERNETIKA — VOLUME 28 (1992), NUMBER 5, PAGES 357 - 382

# ADAPTIVE MAXIMUM–LIKELIHOOD–LIKE ESTIMATION IN LINEAR MODELS
## Part 1. Consistency

JAN ÁMOS VÍŠEK

An adaptive estimator of regression model coefficients based on maximization of kernel estimate of likelihood is proposed. Its consistency (in Part 1) and asymptotic normality (in Part 2) is proved. An asymptotic representation of the estimate implies also its asymptotic efficiency.

## 1. INTRODUCTION

This paper is a continuation of [6] and [7] which has shown that the adaptive estimator of regression coefficients based on minimization of Hellinger distance of the density estimate of residuals and the density estimate of "mirror reflection of residuals" is not efficient for dimensions larger than one.

Hence the present paper brings a new proposal of adaptive estimator of linear regression model coefficients based on estimating density of residuals. The estimate of density of residuals uses a preliminary estimate of regression coefficients and than applies maximum likelihood technique. This new estimator is proved to be efficient in the sense given in Corollary 1 at the end of this paper.

One of the main problems lies in proving consistency of proposed estimator. Solution of this problem may be surely given in a similar way as in [4] requiring some rather abstract conditions on probability distribution of (carriers and) errors. This paper preferred to stay conditions in a way which may seem less verifiable but which are more transparent namely that (very) large values of coefficients are not very probable. In applications due to some requirements which are implied by hardware circumstances we usually transform data into some "reasonable" range and hence we have "some feeling" about the physical possibilities how large this or that parameter may be. The paper is rather long since most of steps in proofs were made in details. Only the steps which are standardly made in similar texts were omitted.

## 2. NOTATIONS

Let us denote by $\mathcal{N}$ the set of all positive integers, by $\mathcal{R}$ the real line. We shall consider a linear model

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

where $Y = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ is a real vector (response variable), $X = (x_{ij})_{i=1}^{n}{}_{j=1}^{p}$ a 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. distributed – according to a d. f. $G$ – variables. We suppose that intercept, if any, is included in the model, i. e. when we assume that the model (1) contains an intercept we have $x_{i1} = 1$ for all $i$. D.f. $G$ is assumed to be fixed, unknown, but symmetric (i. e. for any $x \in \mathcal{R}$ $G(x) = 1 - G(x)$) and allowing density with respect to Lebesgue measure.

Throughout the paper, whenever the probabilistic assertions, the mean values etc. are understood with respect to $G$, this will not be emphasized. Only when they will be taken with respect to another distribution it will be marked by a subscript.

**Remark 1.** Symmetry of the d. f. $G$ may be more than technical necessity. Since the adaptive estimator (which will be proposed later on) is based on the estimate of density of residuals we may get into troubles with bias not assuming symmetry. It might be perhaps improved by estimating density of residuals by an estimator having "sufficiently small" bias. It would be however so complicated that it probably hamper any possibility to prove even rather simple property of estimator. Moreover, it seems that without symmetry adaptive estimation is able to estimate consistently only slopes and estimation of intercept (has to) contain(s) some bias.

It implies that another way how to solve the problem of estimating regression model is not to assume symmetry (but some normalization of design matrix, namely $\sum_{i=1}^{n} x_{ij} = 0$ for any $j = 2, \ldots, p$) and estimate only slopes. In a second step we may try to estimate intercept separately (as location parameter). Naturally it may then happen that the first and the second step will have a different efficiency.

Let us denote for any $\beta \in \mathcal{R}^p$ and $i \in \mathcal{N}$ by

$$e_i(\beta) = Y_i - X_i^{\mathrm{T}} \beta \tag{2}$$

$i$th residual where $X_i^{\mathrm{T}} \beta$ stays for $\sum_{j=1}^{p} X_{ij} \beta_j$. For $\beta^0$ we have $e_i(\beta^0) = e_i$ (see (1)). Let $\tilde{\beta}^n$ be a preliminary estimator of $\beta^0$ and let us write simply $\tilde{e}_i$ instead of $e_i(\tilde{\beta}^n)$. Let $\{c_n\}_{n=1}^{\infty} \downarrow 0$ and denote for any $y \in \mathcal{R}$, $Y \in \mathcal{R}^n$ and $\beta \in \mathcal{R}^p$

$$g_n(y, Y, \beta) = \frac{1}{nc_n} \sum_{i=1}^{n} w\left(c_n^{-1}(y - Y_i + X_i^{\mathrm{T}} \beta)\right) = \frac{1}{nc_n} \sum_{i=1}^{n} w\left(c_n^{-1}(y - e_i(\beta))\right).$$

## 3. ASSUMPTIONS

**Condition A.** Let the kernel $w(y)$ be three times differentiable, positive everywhere and symmetric. Suppose that there are constants $K_1$, $K_2$, $K_3$ and $K_4$ such that

$$\sup_{y\in\mathcal{R}} w(y) < K_1, \qquad \sup_{y\in\mathcal{R}} \frac{|w'(y)|}{w(y)} < K_2,$$
$$\sup_{y\in\mathcal{R}} \frac{|w''(y)|}{w(y)} < K_3, \quad \text{and} \quad \sup_{y\in\mathcal{R}} \frac{|w'''(y)|}{w(y)} < K_4.$$

Preliminary estimator $\tilde{\beta}^n$ is assumed to be such that for some $\delta > \frac{1}{4}$ we have

$$n^{\delta}\|\tilde{\beta}^n - \beta^0\| = O_p(1).$$

Moreover let

$$\lim_{n\to\infty} c_n = 0, \qquad \lim_{n\to\infty} nc_n^8 = \infty \tag{4}$$

and

$$\frac{\log w^{-1}(c_n^{-1})}{n^{\frac{\delta}{2}}} = o(1).$$

Further let $g$ be symmetric, having continuous second derivative and for some $M$, $0 < M < \infty$ we have

$$\sup_{y\in\mathcal{R}} |g'(y)| < M. \tag{5}$$

Finally let $g(x)$ be decreasing for $x > 0$.

**Condition B.** Let for any $a \in \mathcal{R}$

$$\lim_{n\to\infty} n^{\frac{1}{2}} c_n^{-2} \int \sup_{|b|<a} w^{-1}\left(c_n^{-1}(z+b-t)\right) g(t)\, g(z)\big)\ \mathrm{d}t\mathrm{d}z = 0.$$

Moreover let us assume that there are positive $\nu$, $D$ such that for any $z_1$, $z_2 \in \mathcal{R}$ such that $|z_1 - z_2| < \nu$ we have $w(z_1)/w(z_2) < D$.

**Remark 2.** Although Condition B may look rather strange it is easy to see that for a kernel with "sufficiently" heavy tails it can be fulfilled even for density $g$ having also rather heavy tails. As an example we may consider $w(z) = \frac{1}{\pi}\frac{1}{1+z^2}$. We obtain for any $a \in \mathcal{R}$

$$\int \sup_{|b|<a} w^{-1}\left(c_n^{-1}(z+b-t)\right)\, g(z)g(t)\,\mathrm{d}z\mathrm{d}t \le \pi\left(1 + 2c_n^{-2}\left[a^2 + \mathsf{E}|e| + \mathrm{var}(e)\right]\right).$$

It is not difficult to verify that the rest of Condition B is fulfilled, too. Although this kernel doesn't fulfill next condition, namely that $\int |x|w(x)\mathrm{d}x < \infty$ it is easy to see that any kernel of type const $\cdot \frac{1}{1+|x|^{2+\gamma}}$ for some $\gamma > 0$ will be acceptable for Condition B as well as for all following.

**Assertion 1.** Let $\lim_{n\to\infty} c_n = 0$ and $\int |z| w(z) \mathrm{d}z < \infty$. Moreover let $f$ be a density such that its derivative exists and $\sup_{y\in\mathcal{R}} |f'(y)| < \infty$. Then there is a sequence $\{d_n\}_{n=1}^{\infty} \downarrow 0$ such that for any $n \in \mathcal{N}$ we have

$$P_f \left\{ \sup_{y\in\mathcal{R}} |g_n(y, Y, \beta^0) - f(y)| > \frac{1}{2} d_n \right\} < d_n.$$

Proof. We may write

$$\begin{aligned} & g_n(y, Y, \beta^0) - \mathsf{E}_f g_n(y, Y, \beta^0) = \\ = \; & \frac{1}{c_n} \left[ \int w\left(c_n^{-1}(y-t)\right) \mathrm{d}F_n(t) - \int w\left(c_n^{-1}(y-t)\right) \mathrm{d}F(t) \right] \end{aligned}$$

where $F_n(t)$ is the empirical distribution function. Hence we have

$$\begin{aligned} & \sup_{y\in\mathcal{R}} \left| g_n(y, Y, \beta^0) - \mathsf{E}_f g_n(y, Y, \beta^0) \right| \le \\ \le \; & n^{-\frac{1}{2}} c_n^{-1} \sup_{y\in\mathcal{R}} \left| \sqrt{n}(F_n(y) - F(y)) \right| \int |w'(t)| \, \mathrm{d}t \end{aligned} \tag{6}$$

(see also [1]). Now let $\{L_m\}_{m=1}^{\infty} \uparrow \infty$. Since $\sup_{y\in\mathcal{R}} \sqrt{n}|F_n(y) - F(y)|$ is bounded in probability we have for $\min\{m, n\} \to \infty$

$$P_f \left\{ \sup_{y\in\mathcal{R}} \sqrt{n}|F_n(y) - F(y)| > L_m \right\} \searrow 0.$$

For every $k \in \mathcal{N}$ find $n_k^* \in \mathcal{N}$ and $m_k \in \mathcal{N}$ such that for all $n \ge n_k^*$

$$P_f \left\{ \sup_{y\in\mathcal{R}} \sqrt{n}|F_n(y) - F(y)| > L_{m_k} \right\} < \frac{1}{k}.$$

Now select $\tilde{n}_k \ge n_k^*$ such that for any $n \ge \tilde{n}_k$

$$n^{-\frac{1}{2}} c_n^{-1} \int |w'(y)| \, \mathrm{d}y < \frac{1}{4kL_{m_k}},$$

i.e.

$$P_f \left\{ n^{-\frac{1}{2}} c_n^{-1} \sup_{y\in\mathcal{R}} \sqrt{n}|F_n(y) - F(y)| \cdot \int |w'(t)| \, \mathrm{d}t > \frac{1}{4k} \right\} < \frac{1}{k}.$$

Further

$$\begin{aligned} & \sup_{y\in\mathcal{R}} \left| \mathsf{E}_f g_n(y, Y, \beta^0) - f(y) \right| = \sup_{y\in\mathcal{R}} \left| c_n^{-1} \int w(c_n^{-1}(y-t)) f(t) \, \mathrm{d}t - f(y) \right| \\ = \; & \sup_{y\in\mathcal{R}} \left| \int w(z) f(y - c_n z) \, \mathrm{d}z - f(y) \right| = c_n \sup_{y\in\mathcal{R}} \left| \int w(z) \left\{ \int_0^z f'(y - c_n t) \, \mathrm{d}t \right\} \mathrm{d}z \right| \\ < \; & c_n L \cdot \int |z| \, w(z) \, \mathrm{d}z = O(c_n), \end{aligned} \tag{7}$$

where $L = \sup_{y\in\mathcal{R}} |f'(y)|$. Hence we may find $n_k > \tilde{n}_k$ such that for any $n \geq n_k$

$$\sup_{y\in\mathcal{R}} \left|\mathsf{E}_f g_n(y, Y, \beta^0) - f(y)\right| < \frac{1}{4k}.$$

So, we have found a sequence $\{n_k\}_{k=1}^{\infty}$ such that for any $n \geq n_k$ we have

$$P_f\left\{\sup_{y\in\mathcal{R}} \left|g_n(y, Y, \beta^0) - f(y)\right| > \frac{1}{2k}\right\} < \frac{1}{k}$$

and then one may put for any $n \in \mathcal{N}$, $n \in (n_k, n_{k+1}]$ $d_n = \frac{1}{k}$ and the proof follows. □

**Remark 3.** Instead of using (7) one may employ the result of [2] (which is recalled bellow as Lemma 5 in the Part 2 of this paper) that $\sup_{y\in\mathcal{R}} |\mathsf{E}_g g_n(y, Y, \beta^0) - g(y)| = O(c_n)$. From the proof of Assertion 1 it is also possible to see that there is $\{d_n'\}_{n=1}^{\infty} \searrow 0$ so that

$$P_g\left\{\sup_{y\in\mathcal{R}} \left|g_n(y, Y, \beta^0) - \mathsf{E} g_n(y, Y, \beta^0)\right| > \frac{1}{2}\, d_n'\right\} < d_n'.$$

**Definition 1.** For any fixed $\{d_n\}_{n=1}^{\infty} \searrow 0$ let us put

$$G(\{d_n\}_{n=1}^{\infty}) = \Big\{f;\ f \text{ is density such that for any } n \in \mathcal{N}$$
$$P_f\left\{\max\left\{\sup_{y\in\mathcal{R}} \left|g_n(y, Y, \beta^0) - f(y)\right|, \sup_{y\in\mathcal{R}} \left|g_n(y, Y, \beta^0) - \mathsf{E}_f g_n(y, Y, \beta^0)\right|\right\} > \frac{1}{2}\, d_n\right\} < d_n\Big\}.$$

Now for the rest of this paper let us fix some sequence $\{d_n\}_{n=1}^{\infty}$ and we shall assume

**Condition C.** Let

$$\lim_{n\to\infty} \frac{d_n}{c_n} = \infty. \tag{8}$$

Moreover let there be $K_5 < \infty$ such that $\max_{i\in\mathcal{N}, j=1,\ldots,p} |x_{ij}| < K_5$. We shall also assume that the density $g$ is an element of $G(\{d_n\}_{n=1}^{\infty})$. It follows from the assumption that $g(x)$ is decreasing for $x > 0$ that there is a sequence $\{a_n\}_{n=1}^{\infty}$, $a_n > 0$, $a_n \nearrow \infty$ such that

$$(-a_n, a_n) \subset \left\{y \in \mathcal{R}:\ g(y) > d_n^{\frac{1}{2}}\right\}.$$

Then define $b_n(y) = 1$ for $|y| < \frac{1}{2}a_n$ and $b_n(y) = 0$ elsewhere.

In addition to the requirement (4) we will assume that

$$\lim_{n\to\infty} n c_n^6 a_n^{-2} = \infty. \tag{9}$$

**Remark 4.** It is easy to see that to fulfill (9) it requires possibly to make convergence of $c_n$ to zero slower. It may imply that we have to fix a sequence $\{d_n\}_{n=1}^{\infty}$ such that also $d_n$ will have to converge to zero a little slower (see (8)) and it may again cause that $a_n$ will converge to infinity also slower but it improve convergence in (9) and hence (9) is not in a contradiction with any earlier made assumptions.

**Remark 5.** Let us recall that empirical d.f. is given by $F_n(y) = \frac{1}{n}\sum_{i=1}^{n} I_{\{\mathcal{V}_i<y\}}$ where $\mathcal{V}_i$ are i.i.d. distributed random variables. Hence $\mathrm{var}_F F_n(y) = \frac{1}{n} F(y)(1-F(y))$ and therefore the upper bound in (6) may be found uniform for all $f$ (for fixed kernel $w$). The estimate of difference in (7) is uniform for all densities having the same upper bound of its derivative (this is the reason why we have assumed (5)). Hence for a given sequence $\{d_n\}_{n=1}^{\infty}$ the set of all densities belonging to $G(\{d_n\}_{n=1}^{\infty})$ will be rather broad.

**Condition D.** Let us assume that there is $K_6$ such that

$$P\left(\left\| \operatorname*{argmax}_{\beta\in\mathcal{R}^p} \prod_{j=1}^{n} g_n\left(e_j(\beta), Y, \tilde{\beta}^n\right) b_n(\tilde{e}_j)\right\| > K_6\right) \xrightarrow[n\to\infty]{} 0.$$

Let us assume for a while that we know the density of residuals. Then we may estimate regression coefficients by means of maximum likelihood estimator, i.e. as a point (or points) $\hat{\beta}^n$ of $\mathcal{R}^p$ for which

$$\prod_{i=1}^{n} g\left(Y_i - X_i^{\mathrm{T}}\beta\right) = \max!$$

or (due to assumption about existence of $g'$)

$$\sum_{i=1}^{n} x_{ik}\frac{g'(Y_i - X_i^{\mathrm{T}}\beta)}{g(Y_i - X_i^{\mathrm{T}}\beta)} = 0$$

for $k = 1, \ldots, p$. This would lead for normal distribution to the normal equations. Hence using kernel estimate

$$g_n(y, Y, \tilde{\beta}^n) = \frac{1}{nc_n}\sum_{i=1}^{n} w\left(c_n^{-1}(y - \tilde{e}_i)\right)$$

for the estimation of $g(y)$ we may define $\hat{\beta}^n$ as follows.

## 4. DEFINITION OF ESTIMATOR

**Definition 2.** Under $\hat{\beta}^n$ we shall understand a point (or points) of $\mathcal{R}^p$ for which

$$\prod_{j=1}^{n} g_n\left(e_j(\beta), Y, \tilde{\beta}^n\right) b_n(\tilde{e}_j) = \max!$$

or equivalently

$$\hat{\beta}^n = \operatorname*{argmax}_{\beta\in\mathcal{R}^p} \prod_{j=1}^{n} g_n\left(e_j(\beta), Y, \tilde{\beta}^n\right) b_n(\tilde{e}_j).$$

**Remark 6.** Due to assumption about existence of derivative we may look for $\hat{\beta}^n$ by means of equations

$$\sum_{j=1}^{n} x_{jk} \frac{\sum_{i=1}^{n} w'\left(c_n^{-1}(e_j(\beta) - \tilde{e}_i)\right)}{\sum_{i=1}^{n} w\left(c_n^{-1}(e_j(\beta) - \tilde{e}_i)\right)} b_n(\tilde{e}_j) = 0$$

which have to be fulfilled for all $k = 1, \ldots, p$.

Let us assume, for the sake of simplicity that starting from this point all Conditions A, B, C and D hold.

## 5. PRELIMINARIES

**Assertion 2.** Let $\{h_i\}_{i=1}^n$ be positive numbers. Then

$$\left[n^{-1} \sum_{i=1}^{n} h_i\right]^{-1} \leq n^{-1} \sum_{i=1}^{n} h_i^{-1}.$$

A proof follows from the convexity of the function $\frac{1}{x}$.

**Lemma 1.** Let $Q$ be a regular and positive definite symmetric matrix. For any $\delta > 0$ denote $Z_\delta = \{z \in \mathcal{R}^p : \|z\| = \delta\}$. Then

$$\min_{z \in Z_\delta} z^{\mathrm{T}} Q z > 0.$$

Proof. Since $Q$ is regular and symmetric it may be decomposed at $T^{\mathrm{T}}T$ where $T$ is a regular matrix. Moreover $z^{\mathrm{T}}Qz$ is continuous and hence there is a point $z_0 \in Z_\delta$ such that $z^{\mathrm{T}}Qz_0 = \min_{z \in Z_\delta} z^{\mathrm{T}}Qz$. Further for any $z \in Z_\delta$ we have

$$z^{\mathrm{T}} Q z = z^{\mathrm{T}} T^{\mathrm{T}} T z \geq 0.$$

If $z_0 Q z_0 = 0$ then $z_0^{\mathrm{T}} T^{\mathrm{T}} T z_0 = 0$ and therefore also $T z_0 = 0$. But $T$ is regular and it implies that $z_0 = 0$ which contradicts with $z_0 \in Z_\delta$. □

**Lemma 2.** Let $\mathcal{V} = \{v_{kj}\}_{k=1 j=1}^{n p}$ be a matrix such that there is a $H > 0$ such that for any $n \in N$

$$\max_{\substack{k=1,\ldots,n \\ j=1,\ldots,p}} |v_{kj}| < H$$

and $\lim_{n\to\infty} \frac{1}{n} \mathcal{V}^{\mathrm{T}} \mathcal{V} = Q$ where $Q$ is a regular matrix (limit is meant so that for any $k,\ j,\ 1 \leq k,\ j \leq p$ we have $\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} v_{\ell k} v_{\ell j} = q_{kj}$). Then for any $\delta > 0$ there exist $\lambda > 0,\ \tau > 0$ and $n_0 \in \mathcal{N}$ such that for any $z \in \mathcal{R}^p,\ \|z\| \geq \delta$ and $n \geq n_0$ we have

$$\# \left\{ k : \ k \in \{1, \ldots, n\}; \ \left| \sum_{j=1}^{p} v_{kj} z_j \right| > \lambda \right\} \geq \tau \cdot n$$

($\#A$ denotes the number of elements of the set $A$).

Proof. At first we shall prove the assertion of the lemma in a little modified version, namely:
$\forall(\delta > 0)\ \exists(\lambda > 0,\ \tau > 0$ and $n_0 \in \mathcal{N})\ \forall(z \in \mathcal{R}^p,\ \|z\| = \delta$ and $n \geq n_0)$ we have

$$\# \left\{k:\ k \in \{1, \dots, n\};\ \left|\sum_{j=1}^{p} v_{kj} z_j\right| > \lambda\right\} \geq \tau \cdot n.$$

Let us assume that it is not true. Then there is $\delta_0 > 0$ such that for any $\tilde{\lambda} > 0$, $\tilde{\tau} > 0$ and $\tilde{n} \in \mathcal{N}$ there is $z^0 \in \mathcal{R}^p$, $z^0 = z^0(\tilde{\lambda}, \tilde{\tau}, \tilde{n})$, $\|z^0\| = \delta_0$ and $n_0 \geq \tilde{n}$ such that

$$\# \left\{k:\ k \in \{1, \dots, n\};\ \left|\sum_{j=1}^{p} v_{kj} z_\ell\right| > \tilde{\lambda}\right\} < \tilde{\tau} \cdot n_0. \tag{10}$$

Now let $\Delta_0 = \min_{\|z\|=\delta_0} z^{\mathrm{T}} Q z$. Then $\Delta_0 > 0$. Find $n_1 \in \mathcal{N}$ such that for any $n \geq n_1$ and for any $j,\ \ell \in \{1, \dots, p\}$

$$\left|\frac{1}{n} \sum_{k=1}^{n} v_{k\ell} v_{kj} - q_{\ell j}\right| < \frac{\Delta_0}{4 \cdot p^2 \delta_0^2}.$$

Then for any $z$, $\|z\| = \delta_0$ and $n \geq n_1$ we have

$$\left|\sum_{\ell=1}^{p} \sum_{j=1}^{p} z_\ell \left(\frac{1}{n} \sum_{k=1}^{n} v_{k\ell} v_{kj} - q_{\ell j}\right) z_j\right| < p \cdot \delta_0 \cdot \frac{\Delta_0}{4 \cdot p^2 \cdot \delta_0^2} \cdot p \cdot \delta_0 = \frac{\Delta_0}{4}.$$

But it implies that for any $z$, $\|z\| = \delta_0$, $n \geq n_1$, any $\lambda > 0$ and $m = \#\ \{k:\ k \in \{1, \dots, n\}$ ; $|\sum_{\ell=1}^{p} x_{k\ell} z_\ell| > \lambda\}$ we have

$$-\frac{\Delta_0}{4} + \sum_{\ell=1}^{p} \sum_{j=1}^{p} z_\ell q_{\ell j} z_j < \frac{1}{n} \sum_{\ell=1}^{p} \sum_{j=1}^{p} z_\ell \left\{\sum_{k=1}^{n} v_{k\ell} v_{kj}\right\} z_j =$$

$$= \frac{1}{n} \sum_{k=1}^{n} \left\{\sum_{\ell=1}^{p} z_\ell v_{k\ell}\right\} \left\{\sum_{j=1}^{p} z_j v_{kj}\right\} \leq \frac{n-m}{n} \lambda^2 + \frac{m}{n} [H \cdot \delta_0 \cdot p]^2.$$

Put now $\tilde{\lambda} = \sqrt{\frac{\Delta_0}{4}}$, $\tilde{\tau} = [H \cdot \delta_0 \cdot p]^{-2} \cdot \frac{\Delta_0}{4}$ and $\tilde{n} = n_1$. According to (10) then there exists $n_0 \geq \tilde{n} = n_1$ and $z^0 \in \mathcal{R}^p$ $\left(z^0 = z^0(\tilde{\lambda}, \tilde{\tau}, \tilde{n})\right)$, $\|z^0\| = \delta_0$ such that $\# \left\{k:\ k \in \{1, \dots, n\};\ |\sum_{\ell=1}^{p} v_{k\ell} z_\ell^0| > \sqrt{\frac{\Delta_0}{4}}\right\} < \frac{\Delta_0\, n_0}{4 \cdot [H \cdot p \cdot \delta_0]^2}$. But then we have for this $z^0$ and $n_0$ (remember that $\Delta_0 = \min_{\|z\|=\delta_0} z^{\mathrm{T}} Q z$)

$$-\frac{\Delta_0}{4} + \Delta_0 \leq -\frac{\Delta_0}{4} + z_0^{\mathrm{T}} Q\, z_0 < \frac{n_0 - m}{n_0} \cdot \frac{\Delta_0}{4} + \frac{\Delta_0\, n_0}{4 n_0 \cdot [H \cdot p \cdot \delta_0]^2} \cdot [H \cdot p \cdot \delta_0]^2 \leq \frac{\Delta_0}{2}$$

which is a contradiction.

Now let $z \in \mathcal{R}^p$ be arbitrary point with $\|z\| \geq \delta$. Put $y = z \cdot \frac{\delta}{\|z\|} \in \mathcal{R}^p$. Then $\|y\| = \delta$ and we may apply previous part of proof. □

**Lemma 3.** Let $f(u)$ be a convex function on $(0,\infty)$. Then there is a nondecreasing function $\varphi(x)$ on $(0,\infty)$ such that for any pair $g_1$ and $g_2$ of densities on $(-\infty,\infty)$ we have

$$\mathsf{E}_{g_1}\left\{f\left(\frac{g_2}{g_1}\right)\right\} = f(1) + \int_0^1 (t-1)\,\mathrm{d}\varphi(t) + \frac{1}{2}\int_0^\infty \left\{1 - t + \mathsf{E}_{g_1}\left|\frac{g_2(x)}{g_1(x)} - t\right|\right\}\,\mathrm{d}\varphi(t).$$

Proof. Since $f(v)$ is convex we may write

$$f(v) = \left\langle \begin{array}{ll} f(1) + \int_1^v \varphi(t)\,\mathrm{d}t & v \geq 1 \\ f(1) - \int_v^1 \varphi(t)\,\mathrm{d}t & 0 < v < 1 \end{array}\right.$$

where $\varphi(t)$ is a nondecreasing function (see [3], 18.43). Denote $g_2(x)/g_1(x)$ by $D(x)$ and by $P_i$ probability measure generated by $g_i$. Then

$$\mathsf{E}_{g_1} f(D(x)) = \int_{-\infty}^\infty f(D(x))g_1(x)\,\mathrm{d}x = \int_{-\infty}^{T_0} f(D(x))g_1(x)\,\mathrm{d}x + \int_{T_0}^\infty f(D(x))g_1(x)\,\mathrm{d}x$$

where we have defined $T_0$ so that $D(T_0) = 1$. Then we have

$$\mathsf{E}_{g_1} f(D(x)) = \int_{-\infty}^{T_0}\left\{f(1) - \int_{D(x)}^1 \varphi(t)\,dt\right\} g_1(x)\,\mathrm{d}x + \int_{T_0}^\infty \left\{f(1) + \int_1^{D(x)} \varphi(t)\,\mathrm{d}t\right\} g_1(x)\,\mathrm{d}x.$$

Let us study at first the second term of the right hand side. Now $\int_{T_0}^\infty f(1)g_1(x)\,\mathrm{d}x = P_1(D(x) > 1)\cdot f(1)$. Further

$$\begin{aligned} &\int_{T_0}^\infty \left\{\int_1^{D(x)} \varphi(t)\,\mathrm{d}t\, g_1(x)\right\}\mathrm{d}x \\ = &\int_{T_0}^\infty \left\{[(t - D(x))\varphi(t)]_1^{D(x)} g_1(x)\right\}\mathrm{d}x - \int_{T_0}^\infty \left\{\int_1^{D(x)} (t - D(x))\,\mathrm{d}\varphi(t)\right\} g_1(x)\mathrm{d}x = \\ = &\ \varphi(1)\int_{T_0}^\infty (D(x) - 1)\,g_1(x)\,\mathrm{d}x + \int_1^\infty \left\{\int_{\{D(x)>t\}} (D(x) - t)g_1(x)\,\mathrm{d}x\right\}\mathrm{d}\varphi(t). \end{aligned}$$

Moreover

$$1 - t = \int_{-\infty}^\infty (D(x) - t)g_1(x)\,\mathrm{d}x = -\int_{-\infty}^\infty |D(x) - t|\,g_1(x)\,\mathrm{d}x + 2\int_{\{D(x)>t\}} (D(x) - t)g_1(x)\,\mathrm{d}x.$$

Together it gives

$$\begin{aligned} &\int_{T_0}^\infty f(D(x))g_1(x)\,\mathrm{d}x \\ = &\ f(1)P_1(D(x) \geq 1) + \varphi(1)\,[P_2(D(x) \geq 1) - P_1(D(x) \geq 1)] \\ &+ \frac{1}{2}\int_1^\infty \{1 - t + \mathsf{E}_{g_1}|D(x) - t|\}\,\mathrm{d}\varphi(t). \end{aligned}$$

Similarly for

$$
\begin{aligned}
&\int_{-\infty}^{T_0}\left\{f(1)-\int_{D(x)}^{1}\varphi(t)\,\mathrm{d}t\right\}g_1(x)\,\mathrm{d}x\\
=\;& f(1)\,P_1(D(x)<1)-\int_{-\infty}^{T_0}\int_{D(x)}^{1}\varphi(t)\,\mathrm{d}t g_1(x)\,\mathrm{d}x\\
=\;& f(1)\,P_1(D(x)<1)-\int_{-\infty}^{T_0}\left\{[(D(x)-t)\,\varphi(t)]_{D(x)}^{1}\right\}g_1(x)\,\mathrm{d}x\\
&+\int_{-\infty}^{T_0}\int_{D(x)}^{1}(t-D(x))\,\mathrm{d}\varphi(t)\,g_1(x)\,\mathrm{d}x=\\
=\;& f(1)\,P_1(D(x)<1)+\varphi(1)\,(P_2(D(x)<1)-P_1(D(x)<1))\\
&+\int_0^1\int_{\{D(x)<t\}}(t-D(x))\,g_1(x)\,\mathrm{d}x\,\mathrm{d}\varphi(t).
\end{aligned}
$$

But the last integral may be written as

$$
\begin{aligned}
&\int_0^1\int_{-\infty}^{\infty}(t-D(x))\,g_1(x)\,\mathrm{d}x\,\mathrm{d}\varphi(t)+\int_0^1\int_{\{D(x)>t\}}(D(x)-t)\,g_1(x)\,\mathrm{d}x\,\mathrm{d}\varphi(t)\\
=\;&\int_0^1(t-1)\,\mathrm{d}\varphi(t)+\frac{1}{2}\int_0^1\{1-t+\mathsf{E}_{g_1}|D(x)-t|\}\,\mathrm{d}\varphi(t)
\end{aligned}
$$

and the proof follows. □

**Lemma 4.** For any $\theta_1$, $\theta_2$, $0\le\theta_1<\theta_2$ and $c>0$ we have

$$
\mathsf{E}_g\log\frac{c^{-1}\int w(c^{-1}(t-\theta_1-z))g(z)\mathrm{d}z}{g(t)}\ge\mathsf{E}_g\log\frac{c^{-1}\int w(c^{-1}(t-\theta_2-z))g(z)\mathrm{d}z}{g(t)}.
$$

P r o o f. We need to show that

$$
\mathsf{E}_g\left\{-\log\frac{c^{-1}\int w(c^{-1}(t-\theta_1-z))g(z)\mathrm{d}z}{g(t)}\right\}\le\mathsf{E}_g\left\{-\log\frac{c^{-1}\int w(c^{-1}(t-\theta_2-z))g(z)\,dz}{g(t)}\right\}.
$$

Let us denote $g_c(x)=c^{-1}\int w(c^{-1}(x-y))g(y)\mathrm{d}y$ and $g_{ci}(x)=g_c(x-\theta_i)$ for $i=1,2$. Due to the fact that the function $\varphi(t)$ from Lemma 3 is nondecreasing it is sufficient to verify that for any $t\in\mathcal{R}$

$$
\mathsf{E}_g\left|\frac{g_{c1}(x)}{g(x)}-t\right|\le\mathsf{E}_g\left|\frac{g_{c2}(x)}{g(x)}-t\right|.
$$

For an arbitrary density $\tilde g$ we have

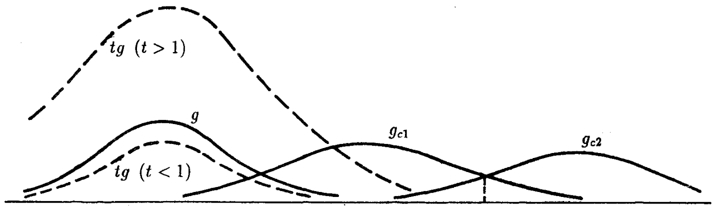
$$
\mathsf{E}_g\left|\frac{\tilde g}{g}-t\right|=\int_{-\infty}^{\infty}\left|\frac{\tilde g(x)}{g(x)}-t\right|g(x)\mathrm{d}x=1-t+2t\cdot\int_{\tilde g<t\cdot g}g\,\mathrm{d}x-2\int_{\tilde g<t\cdot g}\tilde g\mathrm{d}x.
$$

Hence

$$\frac{1}{2}\left\{\mathsf{E}_g\left|\frac{g_{c2}}{g}-t\right|-\mathsf{E}_g\left|\frac{g_{c1}}{g}-t\right|\right\}=$$

$$= t\cdot\left\{\int_{\{g_{c2}<t\cdot g\}} g\,\mathrm{d}x-\int_{\{g_{c1}<t\cdot g\}} g\,\mathrm{d}x\right\}+\int_{\{g_{c1}<t\cdot g\}} g_{c1}\,\mathrm{d}x-\int_{\{g_{c2}<t\cdot g\}} g_{c1}\,\mathrm{d}x. \qquad (11)$$

Let us denote $A_t = \{g_{c2} < tg\}$ and $B_t = \{g_{c1} < tg\}$ and draw an exhibit




Let $a_0$ be a point of intersection of $g_{c1}$ and $g_{c2}$, i. e. $g_{c1}(a_0) = g_{c2}(a_0)$. Let us recall that $g_{c1} = g_c(x-\theta_1), g_{c2} = g_c(x-\theta_2)$ and hence

$$\begin{aligned} g_c(a_0-\theta_1) &= g_c(a_0-\theta_2), \\ a_0 &= \frac{1}{2}(\theta_1+\theta_2). \end{aligned}$$

Moreover we have $g_{c1}(x) > g_{c2}(x)$ for any $x < a_0$. The expression in (11) may be written as

$$\int_{A_t}(tg-g_{c2})\,\mathrm{d}x-\int_{B_t}(tg-g_{c1})\,\mathrm{d}x.$$

Let us consider the situation when $A_t \subset (-\infty, a_0)$. Then $B_t \subset (-\infty, a_0)$ (even $B_t \subset A_t$) and

$$\int_{A_t}(tg-g_{c2})\,\mathrm{d}x-\int_{B_t}(tg-g_{c1})\,\mathrm{d}x \geq \int_{B_t}(g_{c1}-g_{c2})\,\mathrm{d}x \geq 0.$$

For the case when $A_t$ is not subset of $(-\infty, a_0)$ let us realize that

$$\int_{-\infty}^{a_0}(g_{c1}-g_{c2})\,\mathrm{d}x=\int_{a_0}^{\infty}(g_{c2}-g_{c1})\,\mathrm{d}x$$

and even $g_{c1}(a_0-r)-g_{c2}(a_0-r) = g_{c2}(a_0+r)-g_{c1}(a_0+r)$ for any $r \in \mathcal{R}$. Moreover

$$\begin{aligned} &\int_{A_t}(tg-g_{c2})\,\mathrm{d}x-\int_{B_t}(tg-g_{c1})\,\mathrm{d}x \\ =\ &\int_{A_t\cap(-\infty,a_0)}(tg-g_{c2})\,\mathrm{d}x-\int_{B_t\cap(-\infty,a_0)}(tg-g_{c1})\,\mathrm{d}x \\ +\ &\int_{A_t\cap(a_0,+\infty)}(tg-g_{c2})\,\mathrm{d}x-\int_{B_t\cap(a_0,+\infty)}(tg-g_{c1})\,\mathrm{d}x. \end{aligned}$$

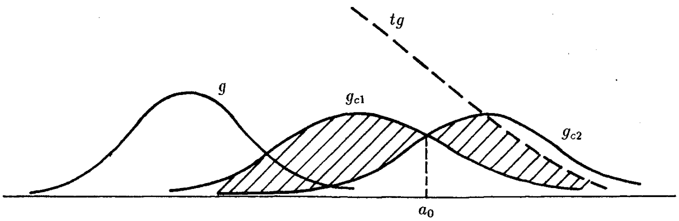
Now (keep in mind that $B_t \cap (-\infty, a_n) \subset A_t \cap (-\infty, a_0)$)

$$\begin{aligned}
&\int_{A_t \cap (-\infty, a_0)} (tg - g_{c2})\ \mathrm{d}x - \int_{B_t \cap (-\infty, a_0)} (tg - g_{c1})\ \mathrm{d}x = \\
= &\int_{A_t \cap B_t \cap (-\infty, a_0)} (g_{c1} - g_{c2})\ \mathrm{d}x + \int_{(A_t \setminus B_t) \cap (-\infty, a_0)} (tg - g_{c2})\ \mathrm{d}x = \\
= &\int_{A_t \cap (-\infty, a_0)} (\min\{tg, g_{c1}\} - g_{c2})\ \mathrm{d}x = \\
= &\int_{-\infty}^{a_0} (g_{c1} - g_{c2})\ \mathrm{d}x - \int_{B_t^0 \cap (-\infty, a_0} (g_{c1} - \max\{g_{c2}, tg\})\ \mathrm{d}x.
\end{aligned} \tag{12}$$

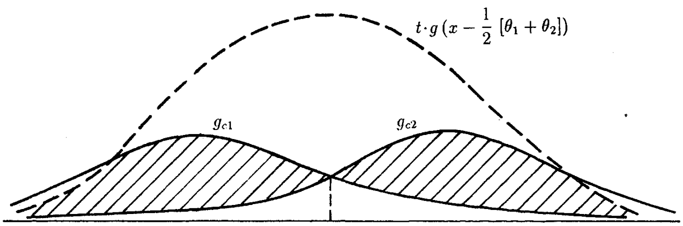
Similarly

$$\begin{aligned}
&\int_{A_t \cap (a_0, +\infty)} (tg - g_{c2})\ \mathrm{d}x - \int_{B_t \cap (a_0, +\infty)} (tg - g_{c1})\ \mathrm{d}x = \\
= &\int_{a_0}^{\infty} (g_{c1} - g_{c2}) + \int_{A_t^c \cap (a_0, +\infty)} (g_{c2} - \max\{tg, g_{c1}\})\ \mathrm{d}x.
\end{aligned} \tag{13}$$

Let us take into account that (12) represents the shadow square given left from $a_0$ in the next exhibit while (13) is equal to the shadow square right from $a_0$ but with minus sign.




Let us assume for a while the above exhibit modified as follows.




Then the shadow square left from $a_0$ is equal to the shadow square right from $a_0$. The last but one picture differ from the last one only in position of dashed curve which is (in last but one) shifted to the left. But it means that the square left from $a_0$ increases and vice versa the square right from $a_0$ decreases (it follows directly from the assumption that $g$ is decreasing for $x > 0$). On the other hand (as already mentioned above) "left" square represents a positive part of (11) namely (12) and "right" square contributes negatively to (11). It is equal to (13) with minus sign. It concludes the proof. □

## 6. CONSISTENCY

We are going to give now the main result of this paper. Alghough the proof is rather long we have prefered to present it in full details for convenience of reader.

**Theorem 1.** Under Conditions A, B, C and D the estimator $\hat{\beta}_n$ is (weakly) consistent.

Proof. The proof will be based on a finite sequence of comparatively simple approximations. The first step will be to show that

$$\frac{1}{n} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \log \frac{\sum_{i=1}^{n} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))}{\sum_{i=1}^{n} w(c_n^{-1}(e_j - \tilde{e}_i))} \{b_n(\tilde{e}_j) - b_n(e_j)\} = o_p(1).$$

It is clear that the above expression is nonzero (and hence it may be larger than some $\varepsilon$) only for those $\omega$'s $\in \Omega$ for which $b_n(\tilde{e}_j) \neq b_n(e_j)$, i.e. for the case when $|\tilde{e}_j| \leq \frac{1}{2}a_n$ and $|e_j| > \frac{1}{2}a_n$ or $|\tilde{e}_j| \geq \frac{1}{2}a_n$ and $|e_j| < \frac{1}{2}a_n$. Let us realize that

$$e_j - \tilde{e}_j = Y_j - X_j^{\mathrm{T}}\beta^0 - Y_j + X_j^{\mathrm{T}}\tilde{\beta}^n = X_j^{\mathrm{T}}(\tilde{\beta}^n - \beta^0).$$

And hence

$$c_n^{-1}(e_j - \tilde{e}_j) = \frac{n^{\delta} X_j^{\mathrm{T}}(\tilde{\beta} - \beta^0)}{n^{\delta} c_n}.$$

Due to this we have, uniformly in $j = 1, \ldots, n$, $c_n^{-1}|e_j - \tilde{e}_j| < \tau$ (for some $\tau > 0$ starting with some $n_0 \in N$) with probability $1 - \varepsilon$ (for apriori given $\varepsilon > 0$). Let us restrict ourselves on the set on which $c_n^{-1}|e_j - \tilde{e}_j| < \tau$. To have a possibility to obtain then $b_n(e_j) \neq b_n(\tilde{e}_j)$ we must have $e_j \in (\frac{1}{2}a_n - \delta, \frac{1}{2}a_n + \delta)$. Let us bound at first just studied expression from above. Let us keep in mind that for $c_n^{-1}|e_j - \tilde{e}_j| < \tau$ we have for $j = 1, \ldots, n$ $w(c_n^{-1}(e_j - \tilde{e}_j)) > \kappa$ where $\kappa$ is a positive number. It implies that

$$\log \frac{\sum_{i=1}^{n} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))}{\sum_{i=1}^{n} w(c_n^{-1}(e_j - \tilde{e}_i))} < \log \frac{n \cdot K_1}{\kappa}$$

(remember that $w$ is bounded) with probability at least $1 - \varepsilon$. Let us use Chebyshev's inequality saying that (for $\varepsilon > 0$)

$$P(X > \varepsilon) \leq \frac{1}{\varepsilon} \mathsf{E} \max\{X, 0\}.$$

The probability that there are $k$ indexes such that $b_n(\tilde{e}_j) = 1$ and $b_n(e_j) = 0$ is not larger than

$$\binom{n}{k} \nu_n^k (1 - \nu_n)^{n-k}$$

where $\nu_n = P(|e_j| \in (a_n - c_n^{-1}(e_j - \tilde{e}_j),\, a_n + c_n^{-1}(e_j - \tilde{e}_j))) = O\left(n^{-\frac{\delta}{2}}\right)$. Hence

$$\mathsf{E}\max\left\{\frac{1}{n}\sup_{\beta\in\mathcal{R}^p}\sum_{j=1}^{n}\log\frac{\sum_{i=1}^{n} w(c_n^{-1}(e_j(\beta) - \tilde{e}_j))}{\sum_{i=1}^{n} w(c_n^{-1}(e_j - \tilde{e}_j))}\{b_n(\tilde{e}_j) - b_n(e_j),\, 0\}\right\}$$

$$\leq \frac{1}{n}\log\frac{n\cdot K_1}{\kappa}\sum_{k=1}^{n} k\cdot\binom{n}{k}\nu_n^k(1-\nu_n)^{n-k} = \nu_n\cdot\log\frac{n\cdot K_1}{\kappa}$$

which tends to zero as $n \to \infty$.

Similarly for a lower bound. We should consider case that $\{b_n(e_j) = 1$ and $b_n(\tilde{e}_j) = 0$ for some $j$'s$\}$ and evaluate conditional mean value of

$$\frac{1}{n}\sup_{\|\beta-\beta^0\|<K_6}\sum_{j=1}^{n}\log\frac{\sum_{i=1}^{n} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))}{\sum_{i=1}^{n} w(c_n^{-1}(e_j - \tilde{e}_i))}.$$

Due to the fact that $\beta$'s over which we take infimum are such that $\|\beta - \beta^0\| < K_6$ the mean value over these cases will be of order

$$\sum_{k=1}^{n} k\cdot\log\frac{w(c_n^{-1}\cdot 2p\cdot K_5\cdot K_6)}{K_1}\cdot\binom{n}{k}\nu_n^k(1-\nu_n)^{n-k}.$$

Under a straightforward computation we find that it is of order $\nu_n\log w(c_n^{-1})$ and taking into account that $\nu_n = O(n^{-\frac{\delta}{2}})$ we obtain that the conditional mean value converges to zero.

Let us show now that

$$\frac{1}{n}\sup_{\|\beta-\beta^0\|<K_6}\left|\sum_{j=1}^{n}\left\{\log\frac{\sum_{i=1}^{n} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))}{\sum_{i=1}^{n} w(c_n^{-1}(e_j - \tilde{e}_i))}\right.\right.$$

$$\left.\left. - \log\frac{\sum_{i\neq j}^{n} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))}{\sum_{r\neq j}^{n} w(c_n^{-1}(e_j - \tilde{e}_r))}\right\} b_n(e_j)\right| = o_p(1). \qquad (14)$$

Having rewritten this expression into the form

$$\frac{1}{n}\sup_{\|\beta-\beta^0\|<K_6}\left|\sum_{j=1}^{n}\left\{\left[\log\sum_{i=1}^{n} w(c^{-1}(e_j(\beta) - \tilde{e}_i)) - \log\sum_{i\neq j}^{n} w(c^{-1}(e_j(\beta) - \tilde{e}_i))\right]\right.\right.$$

$$\left.\left. - \left[\log\sum_{i=1}^{n} w(c^{-1}(e_j(\beta) - \tilde{e}_i)) - \log\sum_{i\neq j}^{n} w(c^{-1}(e_j(\beta) - \tilde{e}_i))\right]\right\} b_n(e_j)\right|$$

and using Taylor's expansion we obtain

$$\frac{1}{n} \sup_{\|\beta-\beta^0\|<K_6} \left| \sum_{j=1}^{n} \left\{ \frac{w(c_n^{-1}(e_j(\beta) - \tilde{e}_j))}{\xi_j} - \frac{w(c_n^{-1}(e_j - \tilde{e}_j))}{\eta_j} \right\} b_n(e_j) \right|$$

where

$$\xi_j \in \left( \sum_{i \neq j} w(c^{-1}(e_j(\beta) - \tilde{e}_i)), \sum_{i=1} w(c^{-1}(e_j(\beta) - \tilde{e}_i)) \right)$$

and

$$\eta_j \in \left( \sum_{i \neq j} w(c^{-1}(e_j - \tilde{e}_i)), \sum_{i=1} w(c^{-1}(e_j - \tilde{e}_i)) \right).$$

Hence (14) can be bounded by

$$\begin{aligned}
&\frac{1}{n} \sup_{\|\beta-\beta^0\|<K_6} \left| \sum_{j=1}^{n} \left\{ \frac{w(c_n^{-1}(e_j(\beta) - \tilde{e}_j))}{\sum_{i\neq j} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))} + \frac{w(c_n^{-1}(e_j - \tilde{e}_j))}{\sum_{i\neq j} w(c_n^{-1}(e_j - \tilde{e}_i))} \right\} b_n(e_j) \right| \\
\leq\ &\frac{1}{n} \sup_{\|\beta-\beta^0\|<K_6} \left| \sum_{j=1}^{n} \frac{\frac{1}{n} w(c_n^{-1}(e_j(\beta) - \tilde{e}_j))}{\frac{1}{n}\sum_{i\neq j} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))} b_n(e_j) \right| \\
+\ &\frac{1}{n} \sup_{\|\beta-\beta^0\|<K_6} \left| \sum_{j=1}^{n} \frac{\frac{1}{n} w(c_n^{-1}(e_j - \tilde{e}_j))}{\frac{1}{n}\sum_{i\neq j} w(c_n^{-1}(e_j - \tilde{e}_i))} b_n(e_j) \right| \qquad (15)
\end{aligned}$$

Let us consider the first member of (15). It is not greater than

$$n^{-2} K_1 \sup_{\|\beta-\beta^0\|<K_6} \cdot \sum_{j=1}^{n} \left[ \frac{1}{n} \sum_{i\neq j} w\left(c_n^{-1}(e_j(\beta) - \tilde{e}_i)\right) \right]^{-1} \qquad (16)$$

Using Assertion 2 we obtain as an upper bound of (16) the expression

$$n^{-3} K_1 \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \sum_{i\neq j} w^{-1}\left(c_n^{-1}(e_j(\beta) - \tilde{e}_i)\right).$$

Now for any $\varepsilon > 0$ (notice that in that follows residua are without "$\tilde{}$")

$$\begin{aligned}
&P\left\{ n^{-3} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \sum_{i\neq j} w^{-1}\left(c_n^{-1}(e_j(\beta) - e_i)\right) > \frac{\varepsilon}{2} \right\} \\
=\ &P\left\{ n^{-3} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \sum_{i\neq j} w^{-1}\left(c_n^{-1}(e_j - X_j^{\mathrm{T}}(\beta - \beta^0) - e_i)\right) > \frac{\varepsilon}{2} \right\} \\
\leq\ &\frac{2}{\varepsilon} n^{-2} \sum_{j=1}^{n} \int \sup_{\|\beta-\beta^0\|<K_6} w^{-1}\left(c_n^{-1}(z - X_j^{\mathrm{T}}(\beta - \beta^0) - t)\right) g(z) g(t)\, \mathrm{d}z\, \mathrm{d}t
\end{aligned}$$

which converges to zero as $n \to \infty$ (see Condition B). Let us fix a $\Delta > 0$. Moreover denote by $B_n$ the set

$$\left\{\omega \in \Omega : \; n^{-3} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \sum_{i \neq j} \frac{1}{w(c_n^{-1}(e_j - e_i - X_j^T(\beta - \beta^0))} > \frac{\varepsilon}{2}\right\}.$$

Then find $n_0$ so that for any $n \geq n_0$ $P(B_n) < \Delta$. For $w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))$ write

$$w\left(c_n^{-1}(e_j(\beta) - \tilde{e}_i)\right) = w\left(c_n^{-1}(e_j - X_i^T(\beta^0 - \tilde{\beta}^n) - e_i)\right) + w'(\xi_{jin}) \cdot X_i^T(\beta^0 - \tilde{\beta}^n)c_n^{-1} \quad (17)$$

where $\xi_{jin}$ is appropriately selected point which will be specified later. Since we have assumed that $n^\delta\|\tilde{\beta}^n - \beta^0\| = O_p(1)$ we may find $n_1 \in \mathcal{N}$, $n_1 \geq n_0$ and $L > 0$ such that for any $n \geq n_1$

$$P\left\{n^\delta\|\beta^0 - \tilde{\beta}^n\| > L\right\} < \Delta.$$

Finally let us denote by $C_n = \left\{\omega \in \Omega : \; n^\delta\|\beta^0 - \tilde{\beta}^n\| > L\right\}$ and

$$E_n = \left\{\omega \in \Omega : \; n^{-3} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \sum_{i \neq j} w^{-1}(c_n^{-1}(e_j - X_j(\beta - \beta^0) - \tilde{e}_i)) > \varepsilon\right\}$$

(notice that $e$'s in definition of $E_n$ are with "$\tilde{\ }$" in difference with $B_n$). Now, find $n_2 \in \mathcal{N}$, $n_2 \geq n_1$ such that for any $n \geq n_2$ we have $c_n^{-1} \cdot n^{-\delta} \cdot K_2 \cdot K_4 \cdot D \cdot L \cdot p < \frac{1}{2}$ and $c_n^{-1} n^{-\delta} L \cdot K_4 \cdot p < \nu$ (see Condition A and B). Since we have for any $j \in \mathcal{N}$

$$\left|e_j - X_j^T(\beta - \beta^0) - \tilde{e}_i - (e_j - X_j^T(\beta - \beta^0) - e_i)\right| = |e_i - \tilde{e}_i| = |X_i^T(\tilde{\beta}^n - \beta^0)|$$

and for $\xi_{jin}$ from (17) we have

$$\begin{aligned}\xi_{jin} \in \big[&c_n^{-1}\min\{e_j - X_j^T(\beta - \beta^0) - e_i,\, e_j - X_j^T(\beta - \beta^0) - \tilde{e}_i\},\\ &c_n^{-1}\max\{e_j - X_j^T(\beta - \beta^0) - e_i,\, e_j - X_j^T(\beta - \beta^0) - \tilde{e}_i\}\big]\end{aligned}$$

it holds for any $n \geq n_2$ and $\omega \in C_n^c$

$$\begin{aligned}|w'(\xi_{jin})| &= \left|\frac{w'(\xi_{jin})}{w(\xi_{jin})} \cdot \frac{w(\xi_{jin})}{w(c_n^{-1}(e_j - X_j^T(\beta - \beta^0) - e_i))} \cdot w\left(c_n^{-1}(e_j - X_j^T(\beta - \beta^0) - e_i)\right)\right|\\ &\leq K_2 \cdot D \cdot w\left(c_n^{-1}(e_j - X_j^T(\beta - \beta^0) - e_i)\right).\end{aligned}$$

Taking into account (17) it implies that for $\omega \in C_n^c$ and $n \geq n_2$ we have

$$\begin{aligned}w\left(c_n^{-1}(e_j(\beta) - \tilde{e}_i)\right) &\geq w\left(c_n^{-1}(e_j(\beta) - e_i)\right)\left[1 - n^{-\delta}c_n^{-1} \cdot K_2 \cdot K_5 \cdot D \cdot L \cdot p\right]\\ &\geq \frac{1}{2} w\left(c_n^{-1}(e_j(\beta) - e_i)\right).\end{aligned}$$

Now for any $n \leq n_2$ and $\omega \in E_n \cap C_n^c$ we have

$$\varepsilon < n^{-3} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \sum_{i \neq j} w^{-1} \left(c_n^{-1}(e_j - (\beta) - \tilde{e}_i)\right)$$

$$< \; 2n^{-3} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \sum_{i \neq j} w^{-1} \left(c_n^{-1}(e_j(\beta) - e_i)\right),$$

i.e. $\omega \in B_n$. So, since we may write

$$P(E_n) = P(E_n \cap C_n) + P(E_n \cap C_n^c) \leq P(C_n) + P(B_n) \leq 2\Delta,$$

we have proved that the first supremum in (15) is small in probability. The second supremum may be treated in a similar way. Now we would like to prove that also supremum of the difference

$$\frac{1}{n} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \left\{ \log \frac{\sum_{i\neq j} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))}{\sum_{i\neq j} w(c_n^{-1}(e_j - \tilde{e}_i))} \right.$$
$$\left. - \log \frac{\sum_{i\neq j} w(c_n^{-1}(e_j(\beta) - e_i))}{\sum_{i\neq j} w(c_n^{-1}(e_j - e_i))} \right\} b_n(e_j) = o_p(1).$$

Analogously as above we may write this difference in a form

$$\frac{1}{n} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \left\{ \left[ \log \sum_{i\neq j} w(c^{-1}(e_j(\beta) - \tilde{e}_i)) - \log \sum_{i\neq j} w(c^{-1}(e_j(\beta) - e_i)) \right] \right.$$
$$\left. - \left[ \log \sum_{i\neq j} w(c^{-1}(e_j - \tilde{e}_i)) - \log \sum_{i\neq j} w(c^{-1}(e_j - e_i)) \right] \right\} b_n(e_j).$$

Let us use again Taylor's expansion. We obtain that this difference is bounded by

$$\frac{1}{n} \left\{ \sup_{\beta\in\mathcal{R}^p} \sum_{j=1}^{n} \frac{\sum_{i\neq j}\{w(c_n^{-1}(e_j(\beta) - \tilde{e}_i)) - w(c_n^{-1}(e_j(\beta) - e_i))\}}{\tau_{jn}} \right.$$
$$\left. + \sup_{\beta\in\mathcal{R}^p} \sum_{j=1}^{n} \frac{\sum_{i\neq j}\{w(c_n^{-1}(e_j - \tilde{e}_i)) - w(c_n^{-1}(e_j - e_i))\}}{\lambda_{jn}} \right\} b_n(\tilde{e}_j) \tag{18}$$

where

$$\tau_{jn} \in \left[ \min\left\{ \sum_{i\neq j} w(c^{-1}(e_j(\beta) - \tilde{e}_i)), \sum_{i\neq j} w(c^{-1}(e_j(\beta) - e_i)) \right\}, \right.$$
$$\left. \max\left\{ \sum_{i\neq j} w(c^{-1}(e_j(\beta) - \tilde{e}_i)), \sum_{i\neq j} w(c^{-1}(e_j(\beta) - e_i)) \right\} \right]$$

and

$$\lambda_{jn} \in \left[\min\left\{\sum_{i\neq j} w(c^{-1}(e_j - \tilde{e}_i)), \sum_{i\neq j} w(c^{-1}(e_j - e_i))\right\},\right.$$
$$\left.\max\left\{\sum_{i\neq j} w(c^{-1}(e_j - \tilde{e}_i)), \sum_{i\neq j} w(c^{-1}(e_j - e_i))\right\}\right].$$

Now let us fix again some $\Delta > 0$ and $L \in \mathcal{R}$ and find an $n_0 \in \mathcal{N}$ such that for any $n \geq n_0$

$$P\left\{n^{\delta}\|\beta^0 - \tilde{\beta}^n\| > L\right\} < \Delta$$

and again denote

$$C_n = \left\{\omega \in \Omega :\ n^{\delta}\|\beta^0 - \tilde{\beta}^n\| > L\right\}.$$

Similarly as above find $n_1 \geq n_0$ such that for any $n \geq n_1$ we have $c_n^{-1} n^{-\delta} K_2 \cdot K_4 \cdot D \cdot L \cdot p < \frac{1}{2}$ and $c_n^{-1} n^{-\delta} L \cdot K_4 \cdot p < \nu$ (see again Condition A). Then we have for $n \geq n_1$ and $\omega \in C_n^c$

$$|e_j(\beta) - \tilde{e}_i - e_j(\beta) - e_i| = \left|X_i^{\mathrm{T}}(\beta^0 - \tilde{\beta}^n)\right| \leq n^{-\delta} \cdot L \cdot K_4 \cdot p$$

and hence again for any $\omega \in C_n^c$ and for any

$$y \in \left[c_n^{-1} \min\{e_j(\beta) - e_i,\, e_j(\beta) - \tilde{e}_i\},\quad c_n^{-1} \max\{e_j(\beta) - e_i,\, e_j(\beta) - \tilde{e}_i\}\right]$$

$$|w'(y)| \leq K_2 \cdot D \cdot w\left(c_n^{-1}(e_j(\beta) - e_i)\right)$$

which implies that $\sum_{i\neq j} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i)) \geq \frac{1}{2}\sum_{i\neq j} w(c_n^{-1}(e_j(\beta) - e_i))$. Therefore we have for $n \geq n_1$ and $\omega \in C_n^c$

$$\tau_{jn} \geq \frac{1}{2}\sum_{i\neq j} w\left(c_n^{-1}(e_j(\beta) - e_i)\right).$$

It allows to bound from above the first member of (18) by

$$\frac{2}{n}\sup_{\beta\in\mathcal{R}^p}\sum_{j=1}^{n}\frac{\sum_{i\neq j}|w(c_n^{-1}(e_j(\beta) - \tilde{e}_i)) - w(c_n^{-1}(e_j(\beta) - e_i))|}{\sum_{i\neq j} w(c_n^{-1}(e_j(\beta) - e_i))}. \tag{19}$$

But we have also

$$w\left(c_n^{-1}(e_j(\beta) - \tilde{e}_i)\right) - w\left(c_n^{-1}(e_j(\beta) - e_i)\right) = w'(\xi_{jin}).X_i(\beta^0 - \tilde{\beta}_n)$$

(see (17)) and hence for $n \geq n_1$ and $\omega \in C_n^c$

$$\left|w\left(c_n^{-1}(e_j(\beta) - \tilde{e}_i)\right) - w\left(c_n^{-1}(e_j(\beta) - e_i)\right)\right| \leq$$
$$\leq\ n^{-\delta} \cdot K_2 \cdot K_5 \cdot L \cdot p \cdot D \cdot w\left(c_n^{-1}(e_j(\beta) - e_i)\right).$$

Now fix any $\varepsilon > 0$ and denote by

$$F_n \left\{ \omega \in \Omega : \left| \frac{1}{n} \sum_{j=1}^{n} \log \frac{\sum_{i \neq j} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))}{\sum_{i \neq j} w(c_n^{-1}(e_j(\beta) - e_i))} \right| > \varepsilon \right\}.$$

Finally find $n_2 \geq n_1$ such that

$$2 \cdot n_2^{-\delta} \cdot K_2 \cdot K_5 \cdot L \cdot p \cdot D < \varepsilon. \tag{20}$$

Since for any $n \geq n_2$ and $\omega \in C_n^c$ we have from (19) and (20)

$$\frac{1}{n} \sup_{\beta \in R^p} \left| \sum_{j=1}^{n} \log \frac{\sum_{i \neq j} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))}{\sum_{i \neq j} w(c_n^{-1}(e_j(\beta) - e_i))} \right| < \varepsilon,$$

$\omega \in C_n^c$ implies $\omega \in F_n^c$, i.e. $C_n^c \subset F_n^c$ and therefore $F_n \subset C_n$ and this is the same as

$$P(F_n) < \Delta$$

for any $n \geq n_2$. The second member of (18) may be proved to be small in probability along similar lines.

Now we shall show that

$$\begin{aligned} S_n &= n^{-1} \sup_{\|\beta - \beta^0\| < K_6} \left| \sum_{j=1}^{n} \left[ \log \frac{\sum_{i \neq j} w(c_n^{-1}(e_j(\beta) - e_i))}{\sum_{i \neq j} w(c_n^{-1}(e_j - e_i))} - \right. \right. \\ &\quad - \mathsf{E} \left\{ \log \frac{\sum_{i \neq j} w(c_n^{-1}(e_j(\beta) - e_i))}{\sum_{i \neq j} w(c_n^{-1}(e_j - e_i))} \right| e_1 = z_1, \ldots, e_{j-1} = z_{j-1}, e_{j+1} = z_{j+1}, \ldots \\ &\quad \left. \left. \ldots, e_n = z_n \right\} \right] \Bigg| = o_p(1). \end{aligned} \tag{21}$$

Notice that

$$\begin{aligned} &\mathsf{E} \left\{ \log \frac{\sum_{i \neq j} w(c_n^{-1}(e_j(\beta) - e_i))}{\sum_{i \neq j} w(c_n^{-1}(e_j - e_i))} \right| e_1 = z_1, \ldots, e_{j-1} = z_{j-1}, e_{j+1} = z_{j+1}, \ldots e_n = z_n \right\} = \\ &= \int \log \frac{\sum_{i \neq j} w(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta - \beta^0) - z_i))}{\sum_{i \neq j} w(c_n^{-1}(y - z_i))} g(y)\, dy. \end{aligned}$$

To prove (21) we shall start with proving

$$\begin{aligned} V_n(\beta) &= n^{-1} \sum_{j=1}^{n} \left[ \log \sum_{i \neq j} w\left(c_n^{-1}(e_j(\beta) - e_i)\right) \right. \\ &\quad - \mathsf{E} \left\{ \log \sum_{i \neq j} w\left(c_n^{-1}(e_j(\beta) - e_i)\right) \right| e_1 = z_1, \ldots, e_{j-1} = z_{j-1}, e_{j+1} = z_{j+1}, \ldots \\ &\quad \left. \ldots e_n = z_n \right\} \Bigg] = o_p(1). \end{aligned}$$

Using Chebyshev's inequality for some fix (positive) $\varepsilon$ we arrive

$$P\left(|V_n(\beta)| > \varepsilon\right) \le \frac{1}{\varepsilon^2} \mathsf{E}\, V_n^2(\beta) = \frac{1}{\varepsilon^2} \sum_{h=1}^{4} \mathcal{E}_h$$

where

$$\mathcal{E}_1 = \mathsf{E}\left\{ n^{-2} \sum_{j=1}^{n} \left[ \log \sum_{i \ne j} w\left(c_n^{-1}(e_j(\beta) - e_i)\right) - \mathsf{E}\left\{ \log \sum_{i \ne j} w\left(c_n^{-1}(e_j(\beta) - e_i)\right) \right| \right. \right.$$

$$\left.\left.\left. \left| e_1 = z_1, \ldots, e_{j-1} = z_{j-1}, e_{j+1} = z_{j+1}, \ldots e_n = z_n \right\} \right]^2 \right\},$$

$$\mathcal{E}_2 = 2\mathsf{E}\left\{ n^{-2} \sum_{j=1}^{n} \sum_{s>j} \left[ \log \sum_{i \ne j} w\left(c_n^{-1}(e_j(\beta) - e_i)\right) - \mathsf{E}\left\{ \log \sum_{i \ne j} w\left(c_n^{-1}(e_j(\beta) - e_i)\right) \right| \right. \right.$$

$$\left. \left| e_1 = z_1, \ldots, e_{j-1} = z_{j-1}, e_{j+1} = z_{j+1}, \ldots e_n = z_n \right\} \right]$$

$$\left[ \log \sum_{\substack{v \ne s \\ v \ne j}} w\left(c_n^{-1}(e_s(\beta) - e_v)\right) - \mathsf{E}\left\{ \log \sum_{\substack{v \ne s \\ v \ne j}} w\left(c_n^{-1}(e_s(\beta) - e_v)\right) \right| \right.$$

$$\left.\left.\left. \left| e_1 = z_1, \ldots, e_{s-1} = z_{s-1}, e_{s+1} = z_{s+1}, \ldots e_n = z_n \right\} \right] \right\},$$

$$\mathcal{E}_3 = 2\mathsf{E}\left\{ n^{-2} \sum_{j=1}^{n} \sum_{s>j} \left[ \log \sum_{i \ne j} w\left(c_n^{-1}(e_j(\beta) - e_i)\right) - \mathsf{E}\left\{ \log \sum_{i \ne j} w\left(c_n^{-1}(e_j(\beta) - e_i)\right) \right| \right. \right.$$

$$\left. \left| e_1 = z_1, \ldots, e_{j-1} = z_{j-1}, e_{j+1} = z_{j+1}, \ldots e_n = z_n \right\} \right]$$

$$\left.\left[ \log \sum_{\substack{v \ne s \\ v \ne j}} w\left(c_n^{-1}(e_s(\beta) - e_v)\right) - \log \sum_{v \ne s} w\left(c_n^{-1}(e_s(\beta) - e_v)\right) \right] \right\}$$

and finally

$$\mathcal{E}_4 = 2\mathsf{E}\left\{ n^{-2} \sum_{j=1}^{n} \sum_{s>j} \left\{ \log \sum_{i \ne j} w\left(c_n^{-1}(e_j(\beta) - e_i)\right) - \right. \right.$$

$$\left. - \mathsf{E}\left\{ \log \sum_{i \ne j} w\left(c_n^{-1}(e_j(\beta) - e_i)\right) \right| e_1 = z_1, \ldots, e_{j-1} = z_{j-1}, e_{j+1} = z_{j+1}, \ldots e_n = z_n \right\} \Bigg]$$

$$\left[ \mathsf{E}\left\{ \log \sum_{v \ne s} w\left(c_n^{-1}(e_s(\beta) - e_v)\right) \right| e_1 = z_1, \ldots, e_{s-1} = z_{s-1}, e_{s+1} = z_{s+1}, \ldots e_n = z_n \right\} \Bigg]$$

$$\left.\left.\left. - \mathsf{E}\left\{ \log \sum_{\substack{v \ne s \\ v \ne j}} w\left(c_n^{-1}(e_s(\beta e_v)\right) \right| e_1 = z_1, \ldots, e_{s-1} = z_{s-1}, e_{s+1} = z_{s+1}, \ldots e_n = z_n \right\} \right] \right\} \right\}.$$

Since $\sum_{i\neq j} w\left(c_n^{-1}(e_j(\beta)-e_i)\right) \leq n \cdot K_1$ and $n^{-1} \cdot \log^2 n \to 0$ for $n \to \infty$, $\mathcal{E}_1 \to 0$ for $n \to \infty$. (Notice that convergence to zero is a consequence of boundedness of the kernel $w$ and doesn't depend on $\beta$.) $\mathcal{E}_2$ may be rewritten into the form

$$2\mathsf{E}\left\{ n^{-2}\sum_{j=1}^{n}\sum_{s>j}\left|\log\sum_{\substack{v\neq s\\ v\neq j}} w\left(c_n^{-1}(e_s(\beta)-e_v)\right)\right.\right.$$
$$- \left.\mathsf{E}\left\{\log\sum_{v\neq s} w\left(c_n^{-1}(e_s(\beta)-e_v)\right)\,\middle|\, e_1=z_1,\ldots,e_{s-1}=z_{s-1},e_{s+1}=z_{s+1},\ldots e_n=z_n\right\}\right|$$
$$\times\mathsf{E}\left\{\left[\log\sum_{i\neq j} w\left(c_n^{-1}(e_j(\beta)-e_i)\right)-\right.\right.$$
$$- \left.\left.\mathsf{E}\left\{\log\sum_{i\neq j} w\left(c_n^{-1}(e_j(\beta)-e_i)\right)\,\middle|\, e_1=z_1,\ldots,e_{j-1}=z_{j-1},e_{j+1}=z_{j+1},\ldots e_n=z_n\right\}\right]\right|$$
$$\left.\left.\middle|\, e_1=z_1,\ldots,e_{j-1}=z_{j-1},e_{j+1}=z_{j+1},\ldots e_n=z_n\right\}\right\}.$$

(Remember that $e_j(\beta) = Y_j - X_j^{\mathrm{T}}\beta = Y_j - X_j^{\mathrm{T}}\beta^0 - X_j^{\mathrm{T}}(\beta-\beta^0) = e_j - X_j^{\mathrm{T}}(\beta-\beta^0)$, and hence $e_j(\beta)$ doesn't depend on $e_1, e_2, \ldots, e_{j-1}, e_{j+1}, \ldots, e_n$). The last modification is possible due to fact that the expression

$$\log\sum_{\substack{v\neq s\\ v\neq j}} w\left(c_n^{-1}(e_s(\beta)-e_v)\right)$$

as well as its conditional mean value depends only on random variables which are "fixed" by the set in condition, namely $\{e_1 = z_1, \ldots, e_{j-1} = z_{j-1}, e_{j+1} = z_{j+1}, \ldots e_n = z_n\}$. But

$$\mathsf{E}\left\{\left[\log\sum_{i\neq j} w\left(c_n^{-1}(e_j(\beta)-e_i)\right)\right.\right.$$
$$- \left.\left.\mathsf{E}\left\{\log\sum_{i\neq j} w\left(c_n^{-1}(e_j(\beta)-e_i)\right)\,\middle|\, e_1=z_1,\ldots,e_{j-1}=z_{j-1},e_{j+1}=z_{j+1},\ldots e_n=z_n\right\}\right]\right|$$
$$\left.\middle|\, e_1=z_1,\ldots,e_{j-1}=z_{j-1},e_{j+1}=z_{j+1},\ldots e_n=z_n\right\} = 0.$$

(Notice again that the last mean value is equal to zero without any dependence on $\beta$). Hence $\mathcal{E}_2 = 0$. The expression $\mathcal{E}_3$ may be bounded by

$$4n^{-2}\cdot\log(K_1.n)\cdot\sum_{j=1}^{n}\sum_{s>j}\mathsf{E}\left|\log\sum_{v\neq s} w\left(c_n^{-1}(e_s(\beta)-e_v)\right)-\log\sum_{\substack{v\neq s\\ v\neq j}} w\left(c_n^{-1}(e_s(\beta)-e_v)\right)\right|$$

$$\leq 4n^{-2}\cdot\log(K_1\cdot n)\cdot\sum_{j=1}^{n}\sum_{s>j}\mathsf{E}\frac{w(c_n^{-1}(e_s(\beta)-e_j))}{\sum_{v\neq s, v\neq j} w(c_n^{-1}(e_s(\beta)-e_v))}$$

$$\leq 4n^{-4}\cdot K_1\cdot\log(K_1\cdot n)\cdot\sum_{j=1}^{n}\sum_{s>j}\sum_{v\neq s}\mathsf{E}w^{-1}\left(c_n^{-1}(e_s(\beta)-e_v)\right)$$

$$= 4n^{-2}\cdot\log(K_1\cdot n)\cdot\sum_{s=1}^{n}\mathsf{E}w^{-1}\left(c_n^{-1}(e_s-X_s^{\mathrm{T}}(\beta-\beta^0)-e_1)\right)$$

$$= 4n^{-2}\cdot\log(K_1\cdot n)\cdot\sum_{s=1}^{n}\int w^{-1}\left(c_n^{-1}(z-X_s^{\mathrm{T}}(\beta-\beta^0)-t)\right)g(z)\,g(t)\,\mathrm{d}z\,\mathrm{d}t$$

which converges to zero for $n\to\infty$ (even uniformly for $\|\beta-\beta^0\|<K_6$). The expression $\mathcal{E}_4$ may be treated similarly.

Now we have to make use of the fact that $V_n(\beta)$ is continuous in $\beta$, uniformly continuous for $\beta\in\{\beta\in\mathcal{R}^p:\ \|\beta-\beta^0\|<K_6\}$ and by means of standard technique of covering the ball $\{\beta\in\mathcal{R}^p:\ \|\beta-\beta^0\|<K_6\}$ by a finite set of balls $\{\beta\in\mathcal{R}^p:\ \|\beta-\beta^i\|<\gamma_i\}_{i=1}^{S}$ we may find, using the law of large numbers, (for any apriori fixed $\varepsilon$ and $\tau$, $\varepsilon>0,\tau>0$) a set $A_n$ and $n_0\in\mathcal{N}$ such that for any $n\geq n_0$ and $\omega\in A_n$

$$\left|\frac{1}{n}\sup_{\|\beta-\beta^0\|<K_6}\sum_{j=1}^{n}\left[\log\sum_{i\neq j}w(c_n^{-1}(e_j(\beta)-e_i))\right.\right.$$
$$\left.\left.-\ \mathsf{E}\log\sum_{i\neq j}w(c_n^{-1}(e_j(\beta)-e_i))\,\middle|\, e_1=z_1,\ldots,e_{j-1}=z_{j-1},e_{j+1}=z_{j+1},\ldots,e_n=z_n\right\}\right]\right|<\varepsilon$$

and $P(A_n)>1-\tau$. Similarly we may show that also

$$W_n = n^{-1}\sup_{\|\beta-\beta^0\|<K_6}\left[\log\sum_{i\neq j}w\left(c_n^{-1}(e_j-e_i)\right)\right.$$
$$\left.-\ \mathsf{E}\left\{\log\sum_{i\neq j}w\left(c_n^{-1}(e_j-e_i)\right|e_1=z_1,\ldots,e_{j-1}=z_{j-1},e_{j+1}=z_{j+1},\ldots,e_n=z_n\right\}\right]=o_p(1).$$

So we have proved that

$$\frac{1}{n}\sup_{\|\beta-\beta^0\|<K_6}\sum_{j=1}^{n}\left\{\log\frac{\sum_{i=1}^{n}w(c_n^{-1}(e_j(\beta)-\tilde{e}_i))}{\sum_{i=1}^{n}w(c_n^{-1}(e_j-\tilde{e}_i))}\,b_n(\tilde{e}_j)\right.$$
$$\left.-\int\log\frac{\sum_{i\neq j}w(c_n^{-1}(t-X_j^{\mathrm{T}}(\beta-\beta^0)-e_i))}{\sum_{i\neq j}w(c_n^{-1}(t-e_i))}\,g(t)\,b_n(t)\,\mathrm{d}t\right\}=o_p(1).$$

Now we will prove that

$$\int\log\frac{\sum_{i\neq j}w(c_n^{-1}(t-X_j^{\mathrm{T}}(\beta-\beta^0)-e_i))}{\sum_{i\neq j}w(c_n^{-1}(t-e_i))}\,g(t)\,b_n(t)\,\mathrm{d}t$$

may be substituted by

$$\int \log \frac{\sum_{i=1}^{n} w(c_n^{-1}(t - X_j^{\mathrm{T}}(\beta - \beta^0) - e_i))}{\sum_{i=1}^{n} w(c_n^{-1}(t - e_i))} g(t)\, b_n(t)\, \mathrm{d}t.$$

To do it, let us consider at first the difference

$$\frac{1}{n} \sup_{\|\beta-\beta^0\|<K_6} \left| \sum_{j=1}^{n} \int \left[ \log \sum_{i=1}^{n} w\left(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta - \beta^0) - e_i)\right) \right.\right.$$
$$\left.\left. - \quad \log \sum_{i \neq j} w\left(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta - \beta^0) - e_i)\right) \right] b_n(y)\, g(y)\, \mathrm{d}y \right|.$$

The absolute value of this expression may be bounded by

$$\frac{1}{n} \sup_{\|\beta-\beta^0\|<K_6} \left| \sum_{j=1}^{n} \int \frac{\frac{1}{n} w(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta - \beta_0) - e_j))}{\frac{1}{n} \sum_{i \neq j} w(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta - \beta^0) - e_i))} g(y)\, b_n(y)\, \mathrm{d}y \right|$$

(compare (14) and (15)). This is not larger than (see Assertion 2)

$$\frac{1}{n^3} K_1 \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \sum_{i \neq j} \int w^{-1}\left(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta - \beta^0) - e_i)\right) g(y)\, b_n(y)\, \mathrm{d}y$$

and for some fixed $\varepsilon > 0$

$$P\left( n^{-3} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \sum_{i \neq j} \int w^{-1}\left(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta - \beta^0) - e_i)\right) g(y)\, b_n(y)\, \mathrm{d}y > \varepsilon \right)$$
$$\leq \frac{1}{\varepsilon} n^{-2} \int \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \int w^{-1}\left(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta - \beta^0) - t)\right) g(y)\, b_n(y)\, g(t)\, \mathrm{d}y\, \mathrm{d}t$$

which converges to zero according to Condition B.

Now we shall use Condition C. Let us fix some positive $\varepsilon$ and positive $\Delta$ and find $n_0 \in \mathcal{N}$ so that $d_{n_0} < \min\left\{\varepsilon^2, \frac{\Delta}{2}\right\}$ and $d_{n_0} < \frac{1}{2} d_{n_0}^{\frac{1}{2}}$. Denote

$$S_{\varepsilon,\Delta,n} = \left\{ \omega \in \Omega : \max\left\{ \sup_{y \in \mathcal{R}} \left| g_n(y, Y, \beta^0) - g(y) \right|, \sup_{y \in \mathcal{R}} \left| g_n(y, Y, \beta^0) - \mathsf{E} g_n(y, Y, \beta^0) \right| \right\} < \frac{1}{2} d_n \right\}.$$

Then for any $n \geq n_0$ and $\omega \in S_{\varepsilon,\Delta,n}$ we have (notice that supremum in the next expression is in fact taken over $\left[\frac{1}{2}a_n, \frac{1}{2}a_n\right]$ and hence $g(y) > d_n^{\frac{1}{2}}$)

$$\sup_{y \in \mathcal{R}} \frac{\left| g_n(y, Y, \beta^0) - \mathsf{E}\, g_n(y, Y, \beta^0) \right|}{\mathsf{E}\, g_n(y, Y, \beta^0)} b_n(y) \leq \frac{d_n/2}{-d_n/2 - d_n/2 + d_n^{\frac{1}{2}}} < d_n^{\frac{1}{2}} < \varepsilon$$

and also

$$\sup_{y\in\mathcal{R}} \frac{|g_n(y,Y,\beta^0) - \mathsf{E}\, g_n(y,Y,\beta^0)|}{g_n(y,Y,\beta^0)}\, b_n(y)$$
$$= \sup_{y\in\mathcal{R}} \frac{|g_n(y,Y,\beta^0) - \mathsf{E}\, g_n(y,Y,\beta^0)|b_n(y)}{g_n(y,Y,\beta^0) - g(y) + g(y)}\, b_n(y)$$
$$\le \frac{d_n/2}{-d_n/2 - d_n/2 + d_n^{\frac{1}{2}}} < d_n^{\frac{1}{2}} < \varepsilon.$$

So we have (remember that $g \in \mathcal{G}(\{d_n\}_{n=1}^{\infty})$)

$$P\left(\max\left(\sup_{y\in\mathcal{R}} \frac{|g_n(y,Y,\beta^0) - \mathsf{E}\, g_n(y,Y,\beta^0)|}{\mathsf{E}\, g_n(y,Y,\beta^0)}\, b_n(y),\right.\right.$$
$$\left.\left.\sup_{y\in\mathcal{R}} \frac{|g_n(y,Y,\beta^0) - \mathsf{E}\, g_n(y,Y,\beta^0)|}{\mathsf{E}\, g_n(y,Y,\beta^0)}\, b_n(y)\right\} > \varepsilon\right) < \Delta.$$

Starting with some $n_0$ we have $K_5 \cdot K_6 \cdot p < \frac{1}{2}a_n$ which implies that we have also $\sup_{j=1,\dots,n} \sup_{\|\beta-\beta^0\|<K_6} \|X_j^{\mathrm{T}}(\beta - \beta^0)\| < \frac{1}{2}a_n$ and hence for any $y$ for which $b_n(y) = 1$, i.e. $|y| < \frac{1}{2}a_n$, we have $\sup_{j=1,\dots,n} \sup_{\|\beta-\beta^0\|<K_6} |y - X_j^{\mathrm{T}}(\beta - \beta^0)| < a_n$. Therefore also $\inf_{j=1,\dots,n} \inf_{\|\beta-\beta^0\|<K_6} g(y - X_j^{\mathrm{T}}(\beta-\beta^0)) > d_n^{\frac{1}{2}}$ (see Condition C) and carrying out similar step as a few lines above we obtain for any $n > n_0$

$$P\left(\max\left\{\sup_{j=1,\dots,n}\sup_{y\in\mathcal{R}} \frac{|g_n(y - X_j^{\mathrm{T}}(\beta-\beta^0),Y,\beta^0) - \mathsf{E}g_n(y - X_j^{\mathrm{T}}(\beta-\beta^0),Y,\beta^0)|}{g_n(y - X_j^{\mathrm{T}}(\beta-\beta^0),Y,\beta^0)}\, b_n(y),\right.\right.$$
$$\left.\left.\sup_{j=1,\dots,n}\sup_{y\in\mathcal{R}} \frac{|g_n(y - X_j^{\mathrm{T}}(\beta-\beta^0) - \mathsf{E}g(y - X_j^{\mathrm{T}}(\beta-\beta^0),Y,\beta^0)|}{\mathsf{E}g(y - X_j^{\mathrm{T}}(\beta-\beta^0),Y,\beta^0)}\, b_n(Y)\right\} > \varepsilon\right) < \Delta.$$

Now using inequality $(a, b > 0)$

$$\log a - \log b < \frac{a-b}{\min\{a,b\}}$$

we may show that

$$\frac{1}{n}\sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n}\left\{\int \log \frac{\sum_{i=1}^n w(c_n^{-1}(t - X_j^{\mathrm{T}}(\tilde\beta - \beta^0) - e_i)}{\sum_{i=1}^n w(c_n^{-1}(t - e_i))}\, g(t)\, b_n(t)\, \mathrm{d}t\right.$$
$$\left. - \int \log \frac{\sum_{i\ne j} w(c_n^{-1}(t - X_j^{\mathrm{T}}(\beta-\beta^0) - z))g(z)\mathrm{d}z}{\sum_{i\ne j} w(c_n^{-1}(t-z))g(z)\mathrm{d}z}\, g(t)\, b_n(t)\, \mathrm{d}t\right\} = o_p(1).$$

So we have proved that

$$\frac{1}{n}\sup_{\|\beta-\beta^0\|<K_6} \left|\sum_{j=1}^{n}\left\{\log \frac{\sum_{i=1}^n w(c_n^{-1}(e_j(\beta) - \tilde e_i))}{\sum_{i=1}^n w(c_n^{-1}(e_j - \tilde e_i))}\, b_n(\tilde e_j) - \right.\right.$$
$$\left.\left. - \int \log \frac{\int w(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta-\beta^0) - z))g(z)\mathrm{d}z}{\int w(c_n^{-1}(y-z))g(z)\mathrm{d}z} g(y) b_n(y) \mathrm{d}y\right\}\right| = o_p(1).$$

Now using Lemma 2 and 4 we may find for any $\gamma > 0$ some $\tau > 0$ and $n_0 \in \mathcal{N}$ such that for any $n \geq n_0$ we have

$$\frac{1}{n} \sup_{\gamma<\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \int \log \frac{\int w(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta - \beta^0) - z))g(z)\mathrm{d}z}{\int w(c_n^{-1}(y - z))g(z)\mathrm{d}z} \, g(y)\, b_n(y)\,\mathrm{d}y < -\tau.$$

It may be shown as follows. One may order the absolute values $\left\{|X_j^{\mathrm{T}}(\beta - \beta^0)|\right\}_{j=1}^{n}$. Lemma 2 then says that for the above given $\gamma$ there are $\lambda > 0$ and $\xi > 0$ such that starting with some $n_1 \in \mathcal{N}$ for all $n > n_1$ and any $\beta$, $\|\beta - \beta^0\| > \gamma$, the number of indices for which $|X_j^{\mathrm{T}}(\beta - \beta^0)| > \lambda$ is larger than $n \cdot \xi$. Let us denote by $I$ the set of corresponding indices. From Condition C it follows that there is an $n_0 \in \mathcal{N}$, $n_0 > n_1$ and $\Delta < 0$ such that for any $n \in \mathcal{N}$, $n \geq n_0$ we have

$$\mathsf{E}_g \left\{ \log \frac{c_n^{-1} \int w(c_n^{-1}(t - \lambda - z))g(z)\mathrm{d}z}{g(t)} - \log \frac{c_n^{-1} \int w(c_n^{-1}(t - z))g(z)\mathrm{d}z}{g(t)} \right\} < \Delta.$$

Using Lemma 4 we have for the above given integral

$$\sum_{j=1}^{n} \int \log \frac{\int w(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta - \beta^0) - z))g(z)\mathrm{d}(z)}{\int w(c_n^{-1}(y - z))g(z)\mathrm{d}z} \, g(y)\, b_n(y)\,\mathrm{d}y$$

$$= \sum_{j=1}^{n} \int \left\{ \log \frac{\int w(c_n^{-1}(y - X_j^{\mathrm{T}}(\beta - \beta^0) - z))g(z)\mathrm{d}(z)}{g(y)} \right.$$

$$\left. - \log \frac{\int w(c_n^{-1}(y - z))g(z)\mathrm{d}(z)}{g(y)} \right\} g(y)\, b_n(y)\,\mathrm{d}y$$

$$< \sum_{j\in I} \int \left\{ \log \frac{\int w(c_n^{-1}(y - \lambda - z))g(z)\mathrm{d}(z)}{g(y)} - \log \frac{\int w(c_n^{-1}(y - z))g(z)\mathrm{d}(z)}{g(y)} \right\} g(y)\, b_n(y)\,\mathrm{d}y.$$

Since $b_n(y) \to 1$ for $n \to \infty$ the last integral is – starting with some $n_0 (> n_1)$ – bounded by $\frac{1}{2} n \cdot \xi \cdot \Delta$. So it is sufficient to put $\tau = -\frac{1}{2} \cdot \xi \cdot \Delta$.

On the other hand for $\beta = \beta^0$ we have $e_j(\beta^0) = e_j$ and hence

$$\frac{1}{n} \sum_{j=1}^{n} \log \frac{\sum_{i=1}^{n} w(c_n^{-1}(e_j(\beta^0) - \tilde{e}_i))}{\sum_{i=1}^{n} w(c_n^{-1}(e_j - \tilde{e}_i))} \, b_n(\tilde{e}_j) = 0$$

which implies that for any $n \in \mathcal{N}$ we have

$$\frac{1}{n} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \log \frac{\sum_{i=1}^{n} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))}{\sum_{i=1}^{n} w(c_n^{-1}(e_j - \tilde{e}_i))} \, b_n(\tilde{e}_j) \geq 0.$$

Due to continuity of all functions and compactness of the ball $\{\beta \in \mathcal{R}^p : \ \|\beta - \beta^0\| < K_6\}$ we have

$$\frac{1}{n} \sup_{\|\beta-\beta^0\|<K_6} \sum_{j=1}^{n} \log \frac{\sum_{i=1}^{n} w(c_n^{-1}(e_j(\beta) - \tilde{e}_i))}{\sum_{i=1}^{n} w(c_n^{-1}(e_j - \tilde{e}_i))} \, b_n(\tilde{e}_j)$$

$$= \frac{1}{n} \sum_{j=1}^{n} \log \frac{\sum_{i=1}^{n} w(c_n^{-1}(e(\hat{\beta}^n) - \tilde{e}_i))}{\sum_{i=1}^{n} w(c_n^{-1}(e_j - \tilde{e}_i))} \, b_n(\tilde{e}_j).$$

Assuming that $\hat{\beta}^n$ is not consistent (together with Condition C) one finds a convergent subsequence which is – starting from some $n_1$ – out of the ball $\{\beta \in \mathcal{R}^p : \ 2\|\beta - \beta^0\| \leq \gamma\}$. This lead to contradiction. □

The asymptotic normality of $\hat{\beta}_n(Y)$ will be proved in the second part of this paper.

(Received March 7, 1991.)

## REFERENCES

[1] R. Beran: An efficient and robust adaptive estimator of location. Ann. Statist. *6* (1978), 292–313.

[2] M. Csörgö and P. Révész: Strong Approximations in Probability and Statistics. Akadémiai Kiadó, Budapest 1981.

[3] E. Hewitt and K. Stromberg: Real and Abstract Analysis. Springer-Verlag, Berlin – Heidelberg – New York 1965.

[4] 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.

[5] R. C. Rao: Linear Statistical Inference and Its Applications. J. Wiley, New York 1973.

[6] J. Á. Víšek: Adaptive estimation in linear regression model. Part 1. Consistency. Kybernetika *28* (1991), 1, 26–36.

[7] J. Á. Víšek: Adaptive estimation in linear regression model. Part 2. Asymptotic normality. Kybernetika *28* (1991), 2, 100–119.

*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.*