Asymptotics for extreme R-estimators

Jana Jurečková
Charles University in Prague

1 Introduction. Two-step regression quantiles

Consider the linear regression model

$$Y_i = \beta_0 + \mathbf{x}'_{ni}\boldsymbol{\beta} + e_i, \quad i = 1, \ldots, n \tag{1.1}$$

with observations $Y_1, \ldots, Y_n$, independent errors
$e_1, \ldots, e_n$, identically distributed according to an unknown distribution function F; $\mathbf{x}_{ni} = (x_{i1}, \ldots, x_{in})'$ is the vector of covariates, $i = 1, \ldots, n$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, and $\boldsymbol{\beta}^* = (\beta_0, \boldsymbol{\beta}')'$ are unknown parameters. We shall suppress the subscript n whenever it does not cause a confusion.

The regression α-quantile of Koenker and Bassett
$\widehat{\boldsymbol{\beta}}^*_n(\alpha) = (\hat{\beta}_{n0}(\alpha), \hat{\beta}_{n1}(\alpha), \ldots, \hat{\beta}_{np}(\alpha))'$ is a solution of the minimization

$$\sum_{i=1}^{n} \rho_\alpha(Y_i - b_0 - \mathbf{x}'_i\mathbf{b}) := \min$$

with respect to $(b_0, b_1, \ldots, b_p)' \in I\!R^{p+1}$, where $\mathbf{x}_i'$ is the i-th row of $\mathbf{X}$ and $\rho_\alpha(x) = |x|\{\alpha I[x > 0] + (1-\alpha)I[x < 0]\}, \; x \in I\!R^1$. Denote ψ_α the right-hand derivative of ρ_α, i.e. $\psi_\alpha(x) = \alpha - I[x < 0], \; x \in I\!R$.

Then the R-estimator of $\boldsymbol{\beta}$, generated by the score function

$$\varphi_\alpha(u) = \psi_\alpha(F_\alpha^{-1}(u)) = \alpha - I[u < \alpha], \; 0 < u < 1$$

is asymptotically equivalent to $\widehat{\boldsymbol{\beta}}(\alpha)$ (see Jurečková (1977) result on asymptotic relations of R- and M-estimators).

Generally the relation can be expressed as $\varphi(u) = c \cdot \psi(F^{-1}(u)), \; 0 < u < 1$; however, we profit from the fact that in this special case the dependence on unknown F disappears. The R-estimator can be calculated by means of minimization of Jaeckel's measure of *rank dispersion* (Jaeckel (1972)):

$$\widehat{\boldsymbol{\beta}}_{nR} = \text{argmin}_{\mathbf{b} \in R^p} \mathcal{D}_n(\mathbf{b}), \qquad (1.2)$$

where

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^{n} (Y_i - \mathbf{x}_i'\mathbf{b})\varphi_\alpha\left(\frac{R_{ni}(\mathbf{Y} - \mathbf{X}\mathbf{b})}{n+1}\right) \tag{1.3}$$

Hence, under some conditions on F and $\mathbf{X}$,

$$n^{\frac{1}{2}}\|\widehat{\boldsymbol{\beta}}_{nR} - \widehat{\boldsymbol{\beta}}_n(\alpha)\| = o_p(1) \quad \text{as } n \to \infty. \tag{1.4}$$

[I have just received a message from Keith Knight, that $\widehat{\boldsymbol{\beta}}_{nR}$ and $\widehat{\boldsymbol{\beta}}_n(\alpha)$ are even the second order asymptotically equivalent, and he promised to send a note soon. This can explain the perfect numerical closeness of both estimators.]

Having estimated $\boldsymbol{\beta}$ by R-estimate $\widehat{\boldsymbol{\beta}}_{nR}$, we can further consider the minimization

$$\sum_{i=1}^{n} \rho_\alpha(Y_i - b - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{nR}) := \min \tag{1.5}$$

with respect to $b \in I\!R^1$. Its solution, denoted as $\hat{\beta}_{n0}$, is the $[n\alpha]$-th order statistic of the residuals $Y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{nR}$, $i = 1, \ldots, n$. Jurečková and Picek (2005) showed that $\hat{\beta}_{n0}$ is a consistent estimate of $\beta_0 + F^{-1}(\alpha)$, asymptotically normally distributed, and the variance of its asymptotic distribution concides with that of the sample α-quantile in the location model. Similar results hold also in the linear autoregressive model.

This procedure can be also modified to obtain an estimate of the extreme error $E_{n:n}$ in the linear regression model (1.1) or in the linear AR model. It can be obtained under the following restriction on the matrix $\mathbf{X}$:

$$\max_{1 \leq i \leq n} \|\mathbf{x}_i\| = \mathcal{O}\left(n^{\frac{1}{2}-\delta}\right) \quad \text{as} \quad n \to \infty, \ 0 < \delta < \tfrac{1}{2}.$$

Actually, let $\widehat{E}_{n:n}$ be the maximum of the residuals,

$$\widehat{E}_{n:n} = \max\{Y_1 - \mathbf{x}_1'\widehat{\boldsymbol{\beta}}_{nR}, \ldots, Y_n - \mathbf{x}_n'\widehat{\boldsymbol{\beta}}_{nR}\}$$

calculated with respect to an suitable R-estimate $\widehat{\boldsymbol{\beta}}_{nR}$ of $\boldsymbol{\beta}$, generated by a score-generating function φ, independent of n. Then $\widehat{E}_{n:n}$ is a consistent estimate of $E_{n:n} + \beta_0$ and has the following properties:

(i) $$|\widehat{E}_{n:n} - E_{n:n} - \beta_0| = \mathcal{O}_p(n^{-\delta}) \quad \text{as} \;\; n \to \infty.$$

(ii) If F belongs to the domain of attraction of the Gumbel distribution, then

$$I\!P\left\{nf(\xi_n)(\widehat{E}_{n:n} - \beta_0 - \xi_n) \leq t\right\} \to \exp\{-\mathrm{e}^{-t}\} \tag{1.6}$$

as $n \to \infty,\; t \in I\!R$, where $\xi_n = F^{-1}\left(1 - \frac{1}{n}\right)$.

(iii) If F belongs to the domain of attraction of the Fréchet distribution and
$1 - F(x) = x^{-m}L(x),\; m > 0$, with a slowly varying function L, then

$$I\!P\left\{\xi_n^{-1}(\widehat{E}_{n:n} - \beta_0) \leq t\right\} \to \exp\left\{-t^{-m}\right\} \tag{1.7}$$

as $n \to \infty, \ t > 0$.

2 Extreme regression quantiles and extreme R-estimators

Smith (1994) was the first who considered an estimator that we today call the *extreme regression quantile* of model (1.1), and found its asymptotic distribution under heavy-tailed distribution F and under some conditions on the $\mathbf{x}_i$. Such estimators were later considered by Portnoy and Jurečková (1999), Knight (2001) and Chernozhukov (2005) (among others), who derived various forms of the asymptotic distributions under other distributions of errors and under different regularity conditions.

Jurečková and Picek (2005) also proposed a *two-step extreme regression quantile* as a special case, starting with a special R-estimator of the slopes. The two-step extreme quantile even coincides with the ordinary extreme regression quantile. Because the ranks are invariant to the shift, the initial R-estimator automatically estimates only the slope parameters in model (1.1). Hence, the two-step version enables to consider the slope and intercept components separately, and it turns out that the "extreme" R-estimator of the slope components even estimates the slope parameters consistently for some distributions of the errors.

The *maximal regression quantile* $\widehat{\boldsymbol{\beta}}_n^*(1)$ is a solution of the following minimization problem:

$$\min_{b_0 \in R, \mathbf{b} \in R^p} \sum_{i=1}^{n} (Y_i - b_0 - \mathbf{x}_i' \mathbf{b})^+ \tag{2.1}$$

where $x^+ = \max(x, 0)$ denotes the positive part of x. The minimization

(2.1) can be alternatively described as any solution to the linear program:

$$\min_{b \in R, \mathbf{b} \in R^p} \sum_{i=1}^{n} \{b_0 + \mathbf{x}_i' \mathbf{b}\}$$
$$\text{s.t. } Y_i \leq b_0 + \mathbf{x}_i' \mathbf{b}, \ i = 1, \dots, n.$$

If $\sum_{i=1}^n x_{ij} = 0, \ j = 1, \dots, p$, then we minimize only b_0 subject to the restrictions.

Consider the R-estimate $\widehat{\boldsymbol{\beta}}_{nR}^{+}$ of $\boldsymbol{\beta}$ generated by the score function

$$\varphi_n(u) = 1 - \tfrac{1}{n} - I[u \leq 1 - \tfrac{1}{n}], \quad 0 \leq u \leq 1 \tag{2.2}$$

with the approximate rank scores

$$a_n(i) = \varphi_n\left(\frac{i}{n+1}\right) = I[i = n] - \frac{1}{n}, \quad i = 1, \dots, n. \tag{2.3}$$

Let $R_{ni}(\mathbf{b})$ denote the rank of the residual $Y_i - \mathbf{x}_i'\mathbf{b}$ among $Y_1 - \mathbf{x}_1'\mathbf{b}, \dots, Y_n - \mathbf{x}_n'\mathbf{b}, \ \mathbf{b} \in \mathbb{R}, \quad i = 1, \dots, n$. Then $\widehat{\boldsymbol{\beta}}_{nR}^{+}$ can be defined as the minimizer of the Jaeckel (1972) measure of the rank dispersion; in this special case the Jaeckel measure takes on the form

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^{n} (Y_i - \mathbf{x}_i'\mathbf{b})\varphi_n\left(\frac{R_{ni}(\mathbf{b})}{n+1}\right) \tag{2.4}$$

$$= \sum_{i=1}^{n} (Y_i - \mathbf{x}_i'\mathbf{b}) I[R_{ni}(\mathbf{b}) = n] - (\bar{Y}_n - \bar{\mathbf{x}}_n'\mathbf{b}),$$

$$= \max_{1 \le i \le n} \{Y_i - (\mathbf{x}_i - \bar{\mathbf{x}}_n)'\mathbf{b}\} - \bar{Y}_n = (Y_i - (\mathbf{x}_i - \bar{\mathbf{x}}_n)'\mathbf{b})_{n:n} - \bar{Y}_n,$$

where $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and $\bar{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$. Hence, $\widehat{\boldsymbol{\beta}}_{nR}^{+}$ minimizes the extreme residual $(Y_i - (\mathbf{x}_i - \bar{\mathbf{x}}_n)'\mathbf{b})_{n:n}$ with respect to $\mathbf{b} \in I\!R^p$. Moreover, if we define $\widehat{\beta}_0^{+}$ as

$$\widehat{\beta}_0^{+} = \max\{Y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{nR}^{+},\ 1 \le i \le n\} \tag{2.5}$$

then $\widehat{\beta}_0^+ + \bar{\mathbf{x}}'\widehat{\boldsymbol{\beta}}_{nR}^+$ is the maximum of $\{Y_i - (\mathbf{x}_i - \bar{\mathbf{x}})'\widehat{\boldsymbol{\beta}}_{nR}^+,\ 1 \le i \le n\}$. It follows from (2.5) that

$$\widehat{\beta}_0^+ + \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{nR}^+ \ge Y_i, \quad i = 1, \ldots, n \tag{2.6}$$

while for some i_0 the inequality converts in an equality, hence $\widehat{\beta}_0^+ + \bar{\mathbf{x}}'\widehat{\boldsymbol{\beta}}_{nR}^+$ is minimized subject to the restriction (2.6). Hence, the two-step maximal regression quantile $\left(\widehat{\beta}_0^+, \widehat{\boldsymbol{\beta}}_{nR}^{+\prime}\right)'$ coincides with the extreme regression quantile defined in (2.2).

The R-estimator $\widehat{\boldsymbol{\beta}}_{nR}^+$ can be even a consistent estimate of $\boldsymbol{\beta}$ under some distributions of $e_1, \ldots, e_n$. This will be demonstrated and the asymptotic distribution of the consistent $\widehat{\boldsymbol{\beta}}_{nR}^+$ will be derived under the following conditions:

(A1) The errors $e_1, \ldots, e_n$ in model (1.1) are independent and identically distributed with density f that is absolutely continuous and has finite Fisher's information, i.e.

$$0 < \mathcal{I}(f) = \int_R \left(\frac{f'(x)}{f(x)} \right)^2 f(x)dx < \infty.$$

We assume that

$$\lim_{n\to\infty} nf(\xi_n) = \infty \tag{2.7}$$

where $\xi_n = F^{-1}\left(1 - \frac{1}{n}\right)$ and F is the distribution function corresponding to f.

(A2) Density f belongs to the domain of attraction of the Gumbel distribution with the distribution function $G(t) = \exp\left\{-\mathrm{e}^{-t}\right\}$.

(A3) The triangular array of p-dimensional vectors $\{\mathbf{x}_{n1}, \ldots, \mathbf{x}_{nn}\}_{n=1}^{\infty}$ satisfies either of the following two conditions:

(A3.a) $\{\mathbf{x}_{n1},\ldots,\mathbf{x}_{nn}\}$ are known p-dimensional vectors satisfying

$$\lim_{n\to\infty}\mathbf{D}_n=\mathbf{D}\quad\text{where } \mathbf{D}_n=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_{ni}-\bar{\mathbf{x}}_n)(\mathbf{x}_{ni}-\bar{\mathbf{x}})' \tag{2.8}$$

and $\mathbf{D}$ is a positively definite $p\times p$ matrix. Moreover,

$$\lim_{n\to\infty}\max_{1\le i\le n} n^{-1}(\mathbf{x}_{ni}-\bar{\mathbf{x}}_n)'\mathbf{D}_n^{-1}(\mathbf{x}_{ni}-\bar{\mathbf{x}}_n)\to 0\quad\text{and}$$

$$n^{-1}\sum_{i=1}^{n}\|\mathbf{x}_i-\bar{\mathbf{x}}\|^3=O_p(1)\quad\text{as}\quad n\to\infty. \tag{2.9}$$

(A3.b) $\{\mathbf{x}_{n1},\ldots,\mathbf{x}_{nn}\}$ are independent p-dimensional random vectors, independent of errors $e_1,\ldots,e_n$, possessing finite moments up to order 3, identically distributed with distribution function H_n. We assume that

$$I\!E(\mathbf{x}_{ni})=\mathbf{0},\ i=1,\ldots,n;\ n=1,2,\ldots,$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' = \mathbf{Q}_n \xrightarrow{p} \mathbf{Q} \quad \text{as} \quad n \to \infty,$$

where $\mathbf{Q}$ is a positively definite $p \times p$ matrix, and that

$$\lim_{n\to\infty} H_n(\mathbf{z}) = H(\mathbf{z}), \quad \mathbf{z} \in I\!R^p, \tag{2.10}$$

where H is a continuous p-dimensional distribution function.

Denote $\kappa_n = (nf(\xi_n))^{-1}$. Then $\kappa_n \downarrow 0$ as $n \to \infty$. Let $R_i(\kappa_n\mathbf{b})$ be the rank of $e_i - \kappa_n\mathbf{x}_i'\mathbf{b}$ among $e_1 - \kappa_n\mathbf{x}_1'\mathbf{b}, \ldots, e_n - \kappa_n\mathbf{x}_n'\mathbf{b}, \quad i = 1, \ldots, n$. Consider the vector of linear rank statistics

$$\mathbf{S}_n(\mathbf{b}) = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}}_n)a_n(R_i(\kappa_n\mathbf{b})). \tag{2.11}$$

The following theorem gives the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{nR}^{+}$:

Theorem 1 *Let* $\widehat{\boldsymbol{\beta}}^{+}_{nR}$ *be the R-estimator of* $\boldsymbol{\beta}$ *defined by the minimization of* $\mathcal{D}_n(\mathbf{b})$ *in (2.4). Then*
(i) Under the conditions **(A1)**, **(A2)** *and* **(A3.a)**, *the sequence of random vectors* $\left\{nf(\xi_n)(\widehat{\boldsymbol{\beta}}^{+}_{nR} - \boldsymbol{\beta})\right\}$ *admits the asymptotic representation, as* $n \to \infty$,

$$nf(\xi_n)\left(\widehat{\boldsymbol{\beta}}^{+}_{nR} - \boldsymbol{\beta}\right) = \mathbf{D}_n^{-1}\mathbf{S}_n(\mathbf{0}) + o_p(1) \tag{2.12}$$

$$= \mathbf{D}_n^{-1}\sum_{i=1}^{n}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\left[a_n^*(R_i(\mathbf{0}), 1 - \tfrac{1}{n})\right] + o_p(1)\,[= O_p(1)]$$

where for $j = 1, \ldots, n$ *and* $0 \le \alpha \le 1$

$$a_n^*(j,\alpha) = \begin{cases} 0, & j \le n\alpha, \\ j - n\alpha, & n\alpha \le j \le n\alpha + 1, \\ 1, & n\alpha + 1 \le j, \end{cases} \tag{2.13}$$

are Hájek's rank scores (Hájek (1965)). Hence, $nf(\xi_n)(\widehat{\boldsymbol{\beta}}^+_{nR} - \boldsymbol{\beta})$ *has asymptotic p-dimensional normal distribution* $\mathcal{N}_p(\mathbf{0}, \mathbf{D}^{-1})$.
(ii) Under the conditions **(A1)**, **(A2)**, **(A3.b)**, *the asymptotic distribution function of the sequence of random vectors*

$$nf(\xi_n)\mathbf{Q}_n(\widehat{\boldsymbol{\beta}}^+_{nR} - \boldsymbol{\beta}) \tag{2.14}$$

coincides with the limiting distribution function H *of* $\{(x_{n1}, \ldots, x_{np})'\}$, *defined in (2.10).*

Remark 1 *Notice that (2.12) implies that* $nf(\xi_n)\left(\widehat{\boldsymbol{\beta}}^+_{nR} - \boldsymbol{\beta}\right)$ *asymptotically coincides with the least squares estimator of the vector* $(nI[R_{n1}(\mathbf{0}) = n], \ldots, nI[R_{nn}(\mathbf{0}) = n])'$ *in the linear model with the design matrix*

$$\left[(x_{ij} - \bar{x}_j)\right]_{i=1,\ldots,n}^{j=1,\ldots,p}.$$

Some ideas of the proof of Theorem 1.1. Let $a_n(i,f),\ 1\leq i\leq n$ be the locally and asymptotically optimal rank scores for distribution f, i.e.

$$a_n(i,f)=I\!E\left\{-\frac{f'(U_{n:i})}{f(U_{n:i})}\right\},\qquad i=1\ldots,n \tag{2.15}$$

where $U_{n:1}\leq\ldots\leq U_{n:n}$ are the order statistics corresponding to the sample from the uniform $R(0,1)$ distribution. Notice that, under conditions **(A1)**, **(A2)**,

$$-\frac{f(x)}{1-F(x)}\approx\frac{f'(x)}{f(x)}\approx\frac{f''(x)}{f'(x)}\approx\ldots \tag{2.16}$$

where $a(x)\approx b(x)$ means that $\lim_{x\to\infty}\frac{a(x)}{b(x)}=1$. This further implies

$$\lim_{n\to\infty}\kappa_n a_n(n,f)=\lim_{n\to\infty}\frac{a_n(n,f)}{nf(F^{-1}(1-\frac{1}{n}))}=1. \tag{2.17}$$

Actually, because $(n-1)u^{n-2}$ is a density on $(0,1)$ with the expectation $1-\frac{1}{n}$, we get by Jenssen inequality

$$a(n,f) = n\int_0^1 -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}u^{n-1}du = n(n-1)\int_0^1 f(F^{-1}(u))u^{n-2}du$$
$$\leq nf(F^{-1}(1-\frac{1}{n})) \tag{2.18}$$

and the unimodality of f implies that $f(F^{-1}(u))$ is concave on $(0,1)$. On the other hand, (2.16) implies that

$$\frac{f(F^{-1}(u))}{1-u} \approx -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}$$

as $u \uparrow 1$, hence the unimodality again implies that $\frac{f(F^{-1}(u))}{1-u}$ is ultimately increasing as $u \uparrow 1$. Thus, to any $\varepsilon > 0$ there exists an n_0 such that, for

$n \geq n_0$,

$$a(n,f) = n(n-1)\int_0^1 f(F^{-1}(u))u^{n-2}du \tag{2.19}$$
$$\geq n(n-1)\int_{1-\frac{1}{n}}^1 f(F^{-1}(u))u^{n-2}du$$
$$\geq n(n-1)nf(F^{-1}(1-\tfrac{1}{n}))\int_{1-\frac{1}{n}}^1 (1-u)u^{n-2}du$$
$$\geq nf(F^{-1}(1-\tfrac{1}{n}))(1-\varepsilon).$$

(2.18) and (2.19) imply (2.17).

It follows from Jurečková and Milhaud (2003) that

$$\begin{aligned} & I\!\!P_n\left\{(R_1(\kappa_n \mathbf{b}), \ldots, R_n(\kappa_n \mathbf{b})) = (r_1, \ldots, r_n)\right\} \\ & = \frac{1}{n!} + \frac{1}{n!}\kappa_n \mathbf{b}' \sum_{i=1}^{n} \mathbf{x}_i a_n(r_i, f) + o(\kappa_n) \\ & = \frac{1}{n!} + \frac{1}{n!}\kappa_n \mathbf{b}' \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}}) a_n(r_i, f) + o(\kappa_n) \end{aligned} \tag{2.20}$$

for any permutation $(r_1, \ldots, r_n)$ of $(1, \ldots, n)$ as $n \to \infty$. The last equality in (2.20) is true because $\sum_{k=1}^{n} a_n(k, f) = 0$.

This further implies for $k = 1, \ldots, n$

$$\mathbb{P}_n\{R_{ni}(\kappa_n \mathbf{b}) = k\} = \frac{1}{n} + \frac{1}{n}\kappa_n(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b}a_n(k, f) + o(\kappa_n), \tag{2.21}$$

$$\mathbb{P}_n\{R_{n1}(\kappa_n \mathbf{b}) = k, R_{n2}(\kappa_n \mathbf{b}) = \ell\} = \frac{1}{n(n-1)}[1 + \kappa_n((\mathbf{x}_1 - \bar{\mathbf{x}})'\mathbf{b}a_n(k, f)$$
$$+(\mathbf{x}_2 - \bar{\mathbf{x}})'\mathbf{b}a_n(\ell, f))] + o(\kappa_n), \ k \neq \ell,$$

$$\mathbb{P}_n\{R_{n1}(\kappa_n \mathbf{b}) = k, R_{n2}(\mathbf{0}) = \ell\}$$
$$= \frac{1}{n(n-1)}[1 + \kappa_n(\mathbf{x}_1 - \bar{\mathbf{x}})'\mathbf{b}a_n(k, f)] + o(\kappa_n).$$

Hence, we obtain for $k \neq n$ from (2.21), as $n \to \infty$,

$$\mathbb{P}_n\{R_{ni}(\kappa_n \mathbf{b}) = n\} = \frac{1}{n}[1 + (\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b}] + o(1) \tag{2.22}$$

$$\mathbb{P}_n\{R_{n1}(\kappa_n\mathbf{b}) = n, R_{n2}(\kappa_n\mathbf{b}) = k\} = \frac{1}{n(n-1)}[1 + (\mathbf{x}_1 - \bar{\mathbf{x}})'\mathbf{b}$$
$$+\kappa_n(\mathbf{x}_2 - \bar{\mathbf{x}})'\mathbf{b}a_n(k, f)] + o(1),$$

$$\mathbb{P}_n\{R_{n1}(\kappa_n\mathbf{b}) = n, R_{n2}(\mathbf{0}) = k\} = \frac{1}{n(n-1)}[1 + (\mathbf{x}_1 - \bar{\mathbf{x}})'\mathbf{b}] + o(1),$$

$$\mathbb{P}_n\{R_{n1}(\kappa_n\mathbf{b}) = k, R_{n2}(\mathbf{0}) = n\}$$
$$= \frac{1}{n(n-1)}[1 + \kappa_n(\mathbf{x}_2 - \bar{\mathbf{x}})'\mathbf{b}a_n(k, f)] + o(1).$$

Hence, for $a_n(n) = 1 - \frac{1}{n}, a_n(k) = -\frac{1}{n}, k \neq n,$

$$\mathbb{E}[a_n(R_i(\kappa_n\mathbf{b}))] \approx (1 - \tfrac{1}{n})\tfrac{1}{n}[1 + \tfrac{1}{n}(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b}]$$
$$-\frac{1}{n}\sum_{k\neq n}\frac{1}{n}[1 - \kappa_n(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b}a_n(k, f)] = \frac{1}{n}(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b}(1 + o(1)). \quad (2.23)$$

Analogously,

$$I\!\!E\left[a_n(R_i(\kappa_n \mathbf{b}))\right]^2 = \tfrac{1}{n} + \tfrac{1}{n}(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b}\left(1 - \tfrac{1}{n}(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b}\right) + o(1),$$

$$I\!\!E\left[a_n(R_i(\kappa_n \mathbf{b})a_n(R_j(\kappa_n \mathbf{b}))\right]$$
$$= -\left(\tfrac{1}{n}(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b}\right)\cdot\left(\tfrac{1}{n}(\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{b}\right) + o(1),\ i \neq j,$$

$$I\!\!E\left[a_n(R_i(\kappa_n \mathbf{b}))a_n(R_j(0))\right] = \tfrac{1}{n(n-1)}\left(1 - \tfrac{1}{n}(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b}\right) + o(1),$$

$$I\!\!E\left[a_n(R_i(\kappa_n \mathbf{b}))a_n(R_i(0))\right] = \tfrac{1}{n}\left[1 + \left(1 - \tfrac{1}{n}\right)(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{b}\right].$$

Thus, (2.22) – (2.24) imply that, as $n \to \infty$,

$$I\!\!E\left[(\mathbf{S}_n(\mathbf{b}) - \mathbf{S}_n(\mathbf{0}))(\mathbf{S}_n(\mathbf{b}) - \mathbf{S}_n(\mathbf{0}))'\right] \tag{2.24}$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' \cdot I\!\!E\{[a_n(R_i(\kappa_n \mathbf{b})) - a_n(R_i(0))] \to 0$$

and because $\mathbf{S}_n(\mathbf{b})$ is a gradient of a convex function $\sum_{i=1}^n (Y_i - \mathbf{x}_i'\mathbf{b}) a_n(R_i(\kappa\mathbf{b}))$, the convergence (2.24) is uniform in $\mathbf{b}$ (see Heiler and Willers (1988)). Finally, (2.23) and (2.24) imply that

$$\sup_{\|\mathbf{b}\| \leq C} \|\mathbf{S}_n(\mathbf{b}) - \mathbf{S}_n(\mathbf{0}) + \mathbf{D}_n \mathbf{b}\| = o_p(1) \tag{2.25}$$

as $n \to \infty$. Inserting $\mathbf{b} \mapsto nf(\xi_n)(\widehat{\boldsymbol{\beta}}_{nR}^{+} - \boldsymbol{\beta})$ into (2.25), we obtain

$$nf(\xi_n)(\widehat{\boldsymbol{\beta}}_{nR}^{+} - \boldsymbol{\beta}) = \mathbf{D}_n{}^{-1}\mathbf{S}_n(\mathbf{0}) + o_p(1) \tag{2.26}$$

as $n \to \infty$, and this in turn implies (2.12). Hájek (1965) proved the convergence of the standardized process $\{\sum_{i=1}^n (x_{ij} - \bar{x}_j) a_n^*(i,t) : \; 0 \leq t \leq 1\}$ of the scores (2.13) to the Brownian bridge, $j = 1, \ldots, p$; this in turn implies the asymptotic normality of $nf(\xi_n)\left(\widehat{\boldsymbol{\beta}}_{nR}^{+} - \boldsymbol{\beta}\right)$. This completes the proof of proposition (i).

(ii) The proof of the second part is analogous to that of (i); the approximations are conditional under given $\mathbf{x}_{n1}, \ldots, \mathbf{x}_{np}$. In such a way we obtain

$$n f(\xi_n) \mathbf{Q}_n (\widehat{\boldsymbol{\beta}}^{+}_{nR} - \boldsymbol{\beta}) = \mathbf{S}_{n0} + o_p(1) \quad \text{as} \quad n \to \infty$$

and $\mathbf{S}_{n0} = \mathbf{x}_{ni}$ with probability $\frac{1}{n}$, $i = 1, \ldots, n$; hence, for $B \in \mathcal{B}^p$,

$$I\!P\left(\mathbf{S}_{n0} \in B\right\} = \frac{1}{n} \sum_{i=1}^{n} I\!P(\mathbf{x}_{ni} \in B) = H_n(B) \to H(B)$$

as $n \to \infty$. □

Let $\widehat{\beta}_0^+ = \max\{Y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_{nR}^+,\ 1 \le i \le n\}$ be the intercept component of the extreme regression quantile (cf. (2.5). Then

$$E_{n:n} - \mathbf{x}_{d_n}'\left(\widehat{\boldsymbol{\beta}}_{nR}^+ - \boldsymbol{\beta}\right) \le \widehat{\beta}_0^+ - \beta_0 \le E_{n:n} - \min_{1\le i\le n}\left[\mathbf{x}_i'\left(\widehat{\boldsymbol{\beta}}_{nR}^+ - \boldsymbol{\beta}\right)\right],$$

where d_n is the antirank of $E_{n:n}$, i.e.

$$d_r = i \Longleftrightarrow E_{n:r} = E_i,\ \ i, r = 1, \ldots, n.$$

On the other hand, for some $r = 1, \ldots, n$,

$$\widehat{\beta}_0^+ - \beta_0 = E_{n:r} - \mathbf{x}_{d_r}'\left(\widehat{\boldsymbol{\beta}}_{nR}^+ - \boldsymbol{\beta}\right).$$

To determine the asymptotic distribution of $nf(\xi_n)(\widehat{\beta}_0^+ - \beta_0 - \xi)$ needs to impose some more conditions on the $\mathbf{x}_i$.

References

[1] Chernozhukov, V. (2005). Extremal quantile regression. *Ann. Statist.* **30**/2.

[2] Hájek, J. (1965). Extension of the Kolmogorov-Smirnov test to regression alternatives. *Proc. of Bernoulli-Bayes-Laplace Seminar* (L. LeCam, ed.), pp. 45–60. Univ. of California Press.

[3] Heiler, S. and R. Willers (1988). Asymptotic normality of R-estimates in the linear model. *Statistics* **19**, 173–184.

[4] Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43**, 1449–1459.

[5] Jurečková, J. (1971). Nonparametric estimate of regression coefficients. *Ann. Math. Statist.* **42**, 1328–1338.

[6] Jurečková, J. (1977). Asymptotic relation of M-estimates and R-estimates

in the linear regression model. *Ann. Statist.* **5**, 464–472.

[7] Jurečková, J. and X. Milhaud (2003). Derivative in the mean of a density and statistical applications. *Mathematical Statistics and Applications. Festschrift for Constance van Eeden* (M. Moore, C. Leger, S. Froda, eds.) *IMS Lecture Notes* **42**, 217–232.

[8] Jurečková, J. and J. Picek (2005). Two-step regression quantiles. *Sankhya* (to appear).

[9] Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica* **46**, 33–50.

[10] Knight, K. (2001). Limiting distributions for linear programming estimators. *Extremes* **4:2**, 87–103.

[11] Portnoy, S. and J. Jurečková (1999). On extreme regression quantiles. *Extremes* **2:3**, 227–243.

[12] Ruppert, D. and R. J. Carroll (1980). Trimmed least square estimation in the linear model. *J. Amer. Statist. Assoc.* **75**, 828–838.

[13] Smith, R. (1994). Nonregular regression. *Biometrika* **81**, 173-183.