*The Annals of Statistics*
1992, Vol. 20, No. 1, 305–330

# REGRESSION RANK SCORES AND REGRESSION QUANTILES

BY C. GUTENBRUNNER AND J. JUREČKOVÁ

*Universität Marburg and Charles University, Prague*

We show that regression quantiles, which could be computed as solutions of a linear programming problem, and the solutions of the corresponding dual problem, which we call the regression rank-scores, generalize the duality of order statistics and of ranks from the location to the linear model. Noting this fact, we study the regression quantile and regression rank-score processes in the heteroscedastic linear regression model, obtaining some new estimators and interesting comparisons with existing estimators.

**1. Introduction.** Generalizations of $L$-statistics from the one-sample to the linear models, based on regression quantiles (RQ) introduced by Koenker and Bassett (1978), were studied by several authors. With the exception of Koenker and Portnoy (1987), who consider general weight functions, we usually meet the linear combinations of a fixed number of RQ's (systematic statistics) or the generalizations of the trimmed and Winsorized means, respectively.

Koenker and Bassett (1978) characterize the RQ's as solutions of a parametrized family of linear programs. The corresponding family of dual programs was so far considered only as a technical device for the computation of RQ's. We shall show that the dual solutions, which we call the regression rank-scores (RR) have a statistical meaning and applicability, generalizing the duality of order statistics and ranks to the linear model.

We shall first derive an asymptotic representation for the regression rank-score process, generalizing the Hájek–Šidák (1967) result in the one sample model. This representation implies that linear functions of regression rank process are asymptotically equivalent to a class of aligned rank statistics used by Adichie (1978), Sen (1969) and Chiang and Puri (1984) (among others) for nonparametric hypothesis testing in the linear model. However, the finite sample behavior of the two types of statistics can be rather different, as may be illustrated by examples. The result enables to obtain the uniform asymptotic representation of the RQ process [up to the order $o_p(n^{-1/2})$] under weaker and more natural conditions than in Koenker and Portnoy (1987).

We also use the regression rank-score process to generalize the Koenker and Bassett trimmed LSE to general weight functions, obtaining so a new class of linear model $L$-statistics; we then compare this class with that studied by

Received June 1989; revised February 1991.
*AMS* 1980 *subject classifications.* Primary 62G05; secondary 62J05.
*Key words and phrases.* Regression quantile, regression rank-score, trimmed least-squares estimator, $L$-statistic, linear rank statistic.

Koenker and Portnoy (1987). While both types of $L$-statistics are asymptotically equivalent in the homoscedastic linear model, they generally differ asymptotically under nonlocal heteroscedasticity.

**2. Statistical model.** Let us consider the heteroscedastic linear model

$$\begin{aligned} \mathbf{Y}_n &= \mathbf{X}_n \boldsymbol{\beta} + \boldsymbol{\Gamma}_n \mathbf{U}_n, \\ \boldsymbol{\Gamma}_n &= \operatorname{diag}(\mathbf{X}_n \boldsymbol{\gamma}_n), \end{aligned} \tag{2.1}$$

where $\mathbf{Y}_n = (Y_{n1}, \dots, Y_{nn})^T \in \mathbb{R}^n$ is the vector of observations, $\mathbf{X}_n$ is an $(n \times p)$ design matrix, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_n \in \mathbb{R}^p$ are unknown parameters, $\mathbf{U}_n = (U_1, \dots, U_n)^T$ is the vector of errors, independent and identically distributed with an unknown distribution function $F$.

We shall impose the following conditions on $\mathbf{X}_n$, $\boldsymbol{\Gamma}_n$ and $F$:

(A.1) $F$ has a continuous Lebesgue density $f$, which is positive and finite on $\{t\colon 0 < F(t) < 1\}$.

(B.1) The first column of $\mathbf{X}_n$ consists of ones and the other columns are orthogonal to the first one.

For the asymptotic study ($p$ fixed, $n \to \infty$) we moreover impose the following conditions:

$$\|\mathbf{X}_n\|_\infty = o(n^{1/2}), \tag{B.2}$$

(B.3) $\mathbf{Q}_n = n^{-1}\mathbf{X}_n^T\mathbf{X}_n \to \mathbf{Q}$, where $\mathbf{Q}$ is a positive definite $p \times p$ matrix.

For some results in Section 6, this condition will be sometimes replaced by the following stronger condition:

$$\|\mathbf{X}_n\|_4^4 = O(n). \tag{B.4}$$

(C.1) The diagonal elements of $\boldsymbol{\Gamma}_n$ are bounded away from zero and infinity.

(C.2) $\mathbf{D}_n = n^{-1}\mathbf{X}_n^T\boldsymbol{\Gamma}_n^{-1}\mathbf{X}_n \to \mathbf{D}$ as $n \to \infty$, where $\mathbf{D}$ is a $(p \times p)$ matrix.

$\|\mathbf{X}_n\|_\infty$ and $\|\mathbf{X}_n\|_4$ above denote the usual $l_\infty$ and $l_4$ norms in $\mathbb{R}^{np}$.

The most important special case of (2.1) is the homoscedastic linear model which, due to the condition (B.1), corresponds to

$$\boldsymbol{\gamma}_n = \mathbf{e}_1 = (1, 0, \dots, 0)^T \in \mathbb{R}^p. \tag{2.2}$$

Moreover, Koenker and Bassett (1982) considered the local heteroscedasticity with

$$\boldsymbol{\gamma}_n = \mathbf{e}_1 + O(n^{-1/2}) \tag{2.3}$$

which implies $\|\boldsymbol{\Gamma}_n - \mathbf{I}_n\|_\infty \to 0$ as $n \to \infty$.

Two types of $L$-statistics, considered in this paper, are asymptotically equivalent under the homoscedastic as well as under local heteroscedastic models; but this is generally not the case under the general heteroscedastic model (2.1). A class of robust tests of homoscedasticity is studied by Gutenbrunner (1986). For the sake of simplicity, we shall omit the subscript $n$ in $\boldsymbol{\gamma}_n$; the dependence on $n$ will be apparent from the context.

**3. Regression quantiles and regression rank scores.** Nonparametric statistical procedures in the one-sample and several samples location models are usually based either on the order statistics $Y_{n:1}, \dots, Y_{n:n}$ (or sample quantiles) or on the vector $R_{n1}, \dots, R_{nn}$ of ranks of the observation. The complementarity and other interrelations of these two statistics are well-known. While the regression quantiles form the natural counterparts of the sample quantiles in the linear regression model, an analogous counterpart of the ranks has not yet been known. Such statistics, which we shall call the regression rank-scores, will be characterized in the present paper.

Koenker and Bassett (1978) observe that the $\alpha$-sample quantile $(0 < \alpha < 1)$ in the location model may be characterized as a solution of the minimization

$$\sum_{i=1}^{n} \rho_\alpha(Y_i - t) = \min \text{ with respect to } t \in \mathbb{R}^1, \tag{3.1}$$

where

$$\rho_\alpha(x) = |x|\{(1-\alpha)I[x<0] + \alpha I[x>0]\}, \qquad x \in \mathbb{R}^1. \tag{3.2}$$

The same authors carried over this characterization to the linear model and defined the *$\alpha$-regression quantile* $\hat{\boldsymbol{\beta}}_n(\alpha)$ as a solution of

$$\sum_{i=1}^{n} \rho_\alpha\left(Y_i - \mathbf{x}_i^T \mathbf{t}\right) := \min \text{ with respect to } \mathbf{t} \in \mathbb{R}^p, \tag{3.3}$$

where $\mathbf{x}_i^T$ is the $i$th row of $\mathbf{X}_n$. The population counterpart of the $\alpha$-regression quantile is, due to the linear heteroscedasticity in the model (2.1), the *$\alpha$-population regression quantile*

$$\boldsymbol{\beta}(\alpha) = \boldsymbol{\beta} + F^{-1}(\alpha)\boldsymbol{\gamma}, \qquad 0 < \alpha < 1. \tag{3.4}$$

Since the publication of the Koenker and Bassett fundamental paper (1978), more papers have developed, mainly the asymptotic theory of regression quantiles in analogy to that of sample quantiles. Let us mention Ruppert and Carroll (1980), Koenker and Bassett (1982), Bassett and Koenker (1982), Portnoy (1984), Jurečková (1983a, b; 1984), Jurečková and Sen (1989), Antoch and Jurečková (1985), Gutenbrunner (1986), Koenker and Portnoy (1987), Welsh (1987), de Jongh, de Wet and Welsh (1988), among others.

The *regression quantile process*

$$\mathbf{Z}_n = \left\{\mathbf{Z}_n(\alpha) = n^{1/2}\left(\hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}(\alpha)\right), 0 < \alpha < 1\right\} \tag{3.5}$$

was investigated by Portnoy (1984) in the homoscedastic model; they derived a uniform asymptotic representation of $\mathbf{Z}_n$ up to the remainder term of order $o_p(n^{-1/4} \log n)$. The asymptotic results on the rank-score process, derived in the present paper, enable to derive the asymptotic representation of $\mathbf{Z}_n$ up to $o_p(1)$ but under weaker and more natural conditions on $F$ and $\mathbf{X}_n$.

Koenker and Bassett (1978) characterized the regression quantile $\hat{\boldsymbol{\beta}}_n(\alpha)$ as the component $\hat{\boldsymbol{\beta}}$ of the optimal solution $(\hat{\boldsymbol{\beta}}, \mathbf{r}^+, \mathbf{r}^-)$ of the linear program

$$\begin{aligned} &\alpha \mathbf{1}_n^T \mathbf{r}^+ + (1-\alpha)\mathbf{1}_n^T \mathbf{r}^- = \min, \\ (3.6) \qquad &\mathbf{X}_n \hat{\boldsymbol{\beta}} + \mathbf{r}^+ - \mathbf{r}^- = \mathbf{Y}_n, \qquad (P) \\ &(\hat{\boldsymbol{\beta}}, \mathbf{r}^+, \mathbf{r}^-) \in \mathbb{R}^p \times \mathbb{R}_+^{2n}, \end{aligned}$$

where $\mathbf{1}_n = (1, \dots, 1)^T \in \mathbb{R}^n$.

The corresponding dual program,

$$(3.7) \qquad \mathbf{Y}_n^T \hat{\boldsymbol{\Delta}} = \max, \qquad \mathbf{X}_n^T \hat{\boldsymbol{\Delta}} = \mathbf{0}, \qquad \hat{\boldsymbol{\Delta}} \in [\alpha - 1, \alpha]^n, \qquad (D)$$

and its equivalent version

$$(3.8) \qquad \mathbf{Y}_n^T \hat{\mathbf{a}} = \max, \qquad \mathbf{X}_n^T \hat{\mathbf{a}} = (1-\alpha)\mathbf{X}_n^T \mathbf{1}_n, \qquad \hat{\mathbf{a}} \in [0,1]^n, \qquad (\tilde{D})$$

were also mentioned in the Appendix of Koenker and Bassett (1978), but only as a tool for an efficient computation of the whole function $\alpha \to \hat{\boldsymbol{\beta}}_n(\alpha)$.

It is the purpose of the present paper to investigate the statistical properties of the dual optimal solution $\hat{\mathbf{a}}_n(\alpha) = (\hat{a}_{n1}(\alpha), \dots, \hat{a}_{nn}(\alpha))^T \in \mathbb{R}^n$ of $(\tilde{D})$ and to demonstrate that these solutions, which we call the regression rank-scores (RR), may be used to define the rank-scores in the linear model in a very natural manner, allowing for many applications. We shall see that many aspects of the duality of order statistics and ranks carry over from the location to the general linear model. On the other hand, other properties, such as the independence of order statistics and ranks, generalize to the linear model only asymptotically as $n \to \infty$.

The efficient computational algorithms for the whole path $\hat{\boldsymbol{\beta}}_n(\cdot)$ are described in Koenker and Portnoy (1987) and Bassett and Koenker (1982). The computation of $\hat{\mathbf{a}}(\cdot)$ is essentially the same activity. If $\{Y_{ni}: i \in M_{n\alpha}\}$ are the observations exactly fitted by $\hat{\boldsymbol{\beta}}_n(\alpha)$, then, by the linear programming theory, the dual variables $\hat{\mathbf{a}}_n(\alpha)$ may be characterized by the inequalities

$$(3.9) \qquad \hat{a}_{ni}(\alpha) = \begin{cases} 1, & \text{if } Y_{ni} > \mathbf{x}_{ni}^T \hat{\boldsymbol{\beta}}_n(\alpha), \\ 0, & \text{if } Y_{ni} < \mathbf{x}_{ni}^T \hat{\boldsymbol{\beta}}_n(\alpha), \end{cases}$$

$i = 1, \dots, n$, and by the linear equations

$$(3.10) \qquad \sum_{i \in M_{n\alpha}} \hat{a}_{ni}(\alpha)\mathbf{x}_{ni} = (1-\alpha)\sum_{i=1}^n \mathbf{x}_{ni} - \sum_{i=1}^n I\left[Y_{ni} > \mathbf{x}_{ni}^T \hat{\boldsymbol{\beta}}_n(\alpha)\right]\mathbf{x}_{ni}.$$

Under continuity of $F$, with probability 1, the $p$ linear equations (3.10) have a unique solution for all $\alpha \in (0,1)$, which is in correspondence with the uniqueness of ranks in the location model.

Computational aspects of regression rank scores are considered in Koenker and d'Orey (1989).

It is easy to see that the rank-scores in the location model reduce to the well-known family of rank-scores introduced in Hájek and Šidák (1967) (Section V.3.5). Namely, let $R_{ni}$ be the rank of $Y_{ni}$ in the sample $\mathbf{Y}_n = (Y_{n1}, \dots, Y_{nn})$ and $\hat{a}_{ni}(\alpha) = a_n(R_{ni}, \alpha)$, $i = 1, \dots, n$, where

$$(3.11) \qquad a_n(R_i, \alpha) = \begin{cases} 0, & \text{if } \dfrac{R_i}{n} < \alpha, \\ R_i - n\alpha, & \text{if } \dfrac{R_i - 1}{n} \le \alpha \le \dfrac{R_i}{n}, \\ 1, & \text{if } \alpha < \dfrac{R_i - 1}{n}. \end{cases}$$

If $\{d_{ni}, i = 1, 2, \dots, n\}$ is a suitably normalized triangular array of constants, then, by Theorem V.3.5 in Hájek and Šidák (1967), the process

$$\left\{ n^{-1/2} \sum_{i=1}^{n} d_{ni}(\hat{a}_{ni}(\alpha) - (1 - \alpha)) : 0 \le \alpha \le 1 \right\}$$

converges weakly to the Brownian bridge in the uniform topology on $C[0, 1]$. We obtain a similar result for the linear model, introducing the *regression rank-score process*

$$(3.12) \quad \hat{\mathbf{W}}_n^d = \left\{ \hat{\mathbf{W}}_n^d(\alpha) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \mathbf{d}_{ni}(\hat{a}_{ni}(\alpha) - (1 - \alpha)) : 0 < \alpha < 1 \right\},$$

where $\hat{a}_{ni}(\alpha)$ are the regression rank-scores of (3.8), $i = 1, \dots, n$ and $\{\mathbf{d}_{ni}, i = 1, \dots, n\}$ is a triangular array of $q$-dimensional vectors, $q$ being a fixed positive integer. Denoting $\mathbf{C}_{nd}$ the $n \times q$ matrix with the rows $\mathbf{d}_{ni}^T$, $i = 1, \dots, n$, we impose the following conditions on $\{\mathbf{d}_{ni}, i = 1, \dots, n\}$:

$$(\text{D.1}) \qquad \|\mathbf{C}_{nd}\|_\infty = o(n^{1/2}),$$

$$(\text{D.2}) \qquad \|\mathbf{C}_{nd}\|_2^2 = O(n) \quad \text{as } n \to \infty.$$

For some results we shall moreover assume the condition

$$(\text{D.3}) \qquad n^{-1}\mathbf{C}_{nd}^T \mathbf{C}_{nd} \to \mathbf{Q}_{[d]},$$

where $\mathbf{Q}_{[d]}$ is a positive definite $q \times q$ matrix.

**4. Statistics derived from regression quantiles and regression rank scores.** Integrating the processes (3.5) and (3.12) with respect to an appropriate signed measure on $(0, 1)$, we obtain three broad classes of statistics which have various applications. We shall impose the following conditions on the corresponding signed measures:

(W.1) $\nu$ is a finite signed measure on the open unit interval $(0,1)$ that has a compact support.

(W.2) $b$ is a function of bounded variation which is constant outside a compact subinterval of $(0,1)$.

(W.3) $J$ is a function of bounded variation, vanishing outside a compact subinterval of $(0,1)$ (a trimming weight function of bounded variation).

We shall distinguish three classes of statistics based on the observations $\mathbf{Y}_n$:

(a) The *linear rank statistics* for the linear model of the form

$$\hat{V}_n^{b,d} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_{ni}\hat{\mathbf{b}}_{ni}, \tag{4.1}$$

where $\hat{b}_{ni}$ are the scores defined in either of the following two ways:

$$\hat{b}_{ni} = -\int_0^1 b(t)\, d\hat{a}_{ni}(t), \qquad i = 1,\dots,n, \tag{4.2}$$

or

$$\hat{b}_{ni} = -\int_0^1 b_n(t)\, d\hat{a}_{ni}(t), \qquad i = 1,\dots,n, \tag{4.3}$$

where

$$b_n(t) = b\left(\frac{k}{n+1}\right), \qquad \frac{k-1}{n} < t \le \frac{k}{n}, \qquad k = 1,\dots,n. \tag{4.4}$$

Let us look in the scores (4.2) and (4.3) in more detail, to illustrate the link to the scores in the location model. The natural conditions on $b$ would be square integrability instead of (W.2), which is a technical restriction. In Gutenbrunner, Jurečková, Koenker and Portnoy (1990), we treat $b$, which satisfies the Chernoff–Savage condition $|b'(\alpha)| \le K(\alpha(1-\alpha))^{-3/2+\delta}$. (4.2) could be also written as

$$\hat{b}_{ni} = -\int b(t)\hat{a}'_{ni}(t)\, dt, \qquad i = 1,\dots,n. \tag{4.5}$$

In the location model, (4.5) reduces to

$$\hat{b}_{ni} = n\int_{(R_i-1)/n}^{R_i/n} b(t)\, dt, \qquad i = 1,\dots,n. \tag{4.6}$$

On the other hand, (4.3) could be written as

$$\hat{b}_{ni} = -\sum_{k=1}^{n} b\left(\frac{k}{n+1}\right)\left[\hat{a}_{ni}\left(\frac{k}{n}\right) - \hat{a}_{ni}\left(\frac{k-1}{n}\right)\right] \tag{4.7}$$

and this in the location model reduces to the simple form

$$\hat{b}_{ni} = b\left(\frac{R_i}{n+1}\right), \qquad i = 1, \dots, n \tag{4.8}$$

[the approximate scores of Hájek and Šidák (1967)]. The *Wilcoxon scores* generated by $b(t) = t - 1/2$, $0 < t < 1$, take the form either

$$\hat{b}_{ni} = -\int_0^1 \left(t - \frac{1}{2}\right) d\hat{a}_{ni}(t) \tag{4.9}$$

or

$$\hat{b}_{ni} = \frac{1}{n+1} \sum_{k=1}^{n} \hat{a}_{ni}\left(\frac{k}{n}\right), \qquad i = 1, \dots, n. \tag{4.10}$$

The *median scores*, generated $b(t) = \operatorname{sign}(t - 1/2)$, $0 < t < 1$, are either

$$\hat{b}_{ni} = 2\hat{a}_{ni}\left(\frac{1}{2}\right) - 1 \tag{4.11}$$

or

$$\hat{b}_{ni} = \begin{cases} 2\hat{a}_{ni}\left(\dfrac{1}{2}\right) - 1, & n \text{ even}, \\ \hat{a}_{ni}\left(\dfrac{n+1}{2n}\right) + \hat{a}_{ni}\left(\dfrac{n-1}{2n}\right) - 1, & n \text{ odd}. \end{cases} \tag{4.12}$$

Similarly, $b(t) = \Phi^{-1}(t)$, $0 < t < 1$ ($\Phi$ being the standard normal distribution function), generates the van der Waerden scores, and so on.

The main application of statistics (4.1) is in testing various hypotheses on the components of $\boldsymbol{\beta}$; the coefficients $\mathbf{d}_{ni}$ are then derived from the matrix $\mathbf{X}_n$. Such tests, simple to perform and asymptotically distribution-free with good asymptotic properties, are a subject of a special study [Gutenbrunner, Jurečková, Koenker and Portnoy (1990)].

For $b$ with (W.2), $\hat{b}_{ni}$ in (4.2) could be rewritten as

$$\hat{b}_{ni} = b(0) + \int_0^1 \hat{a}_{ni}(\alpha)\, db(\alpha), \qquad i = 1, \dots, n \tag{4.13}$$

and $\hat{\mathbf{V}}_n^{b,d}$ could be rewritten as

$$\begin{aligned} \hat{\mathbf{V}}_n^{b,d} &= n^{-1/2} \int_0^1 \hat{\mathbf{W}}_n^d(\alpha)\, db(\alpha) + \bar{b}\bar{\mathbf{d}}_n, \\ \bar{b} &= \int_0^1 b(\alpha)\, d\alpha, \qquad \bar{\mathbf{d}}_n = n^{-1} \sum_{i=1}^{n} \mathbf{d}_{ni}. \end{aligned} \tag{4.14}$$

(b) *L-statistics* for the linear model, *the first type*:

$$\mathbf{T}_n^{\nu} = \int \hat{\boldsymbol{\beta}}_n(\alpha)\, d\nu(\alpha), \tag{4.15}$$

corresponding to the well-known integral representation of $L$-statistic with the weights $\nu_{ni} = \nu[(i-1)/n, i/n)$ in the location case, that is,

$$T_n^{\nu} = \sum_{i=1}^{n} \nu_{ni} Y_{n:i} \qquad (Y_{n:1} \le \cdots \le Y_{n:n}), \tag{4.16}$$

where $\hat{\boldsymbol{\beta}}_n(\alpha)$ reduces to the ordinary quantile process, $\hat{\beta}_n(\alpha) = Y_{n:i}$ if $(i-1)/n \le \alpha < i/n$, $i = 1, \dots, n$.

For finite discrete $\nu$, these statistics were already investigated in the pioneering paper of Koenker and Bassett (1978). Koenker and Portnoy (1987) considered $\nu$ with compact support and a bounded density, using methods quite different from ours. In the present paper, we allow $\nu$ to be any finite signed measure having compact support.

(c) *L-statistics* for the linear model, *the second type*: In the location model, the statistics (4.16) could be rewritten in the dual form of the weighted mean

$$\sum_{i=1}^{n} \nu_{nR_{ni}} Y_{ni}, \tag{4.17}$$

with the random weights $\nu_{nR_{ni}}$ (we may consider the trimmed mean as a prominent example). However, extending the two versions (4.16) and (4.17) to the linear model leads to two generally different classes of statistics.

Actually, let $J$ be a function satisfying (W.3) and

$$J \ge 0, \qquad \bar{J} = \int_0^1 J(\alpha)\, d\alpha > 0$$

and define the random weights as either

$$\hat{J}_{ni} = -\int_0^1 J(t)\, d\hat{a}_{ni}(t) \tag{4.18}$$

or

$$\hat{J}_{ni} = -\int_0^1 J_n(t)\, d\hat{a}_{ni}(t), \qquad i = 1, \dots, n, \tag{4.19}$$

where

$$J_n(t) = J\left(\frac{k}{n+1}\right), \qquad \frac{k-1}{n} < t < \frac{k}{n}, \qquad k = 1, \dots, n \tag{4.20}$$

[compare with (4.2) and (4.3)]. The extension of (4.17) to the linear model with the above weights is the weighted least-squared estimator (LSE)

$$\mathbf{L}_n^{\nu} = \bar{J}\left(\mathbf{X}_n^T \hat{\mathbf{J}}_n \mathbf{X}_n\right)^{-1} \mathbf{X}_n^T \hat{\mathbf{J}}_n \mathbf{Y}_n, \tag{4.21}$$

with

$$\hat{\mathbf{J}}_n = \begin{cases} \mathbf{I}_n, & \text{if } \mathbf{X}_n^T \operatorname{diag}(\hat{J}_{ni}) \mathbf{X}_n \text{ is singular,} \\ \operatorname{diag}(\hat{J}_{ni}), & \text{otherwise,} \end{cases} \tag{4.22}$$

and

$$\nu(A) = \int_A J(\alpha)\, d\alpha, \qquad A \in \mathscr{B}_{(0,1)}.$$

Note that the probability of $\mathbf{X}_n^T \operatorname{diag}(\hat{J}_{ni})\mathbf{X}_n$ in (4.22) being singular tends to zero as $n \to \infty$. We set $\hat{\mathbf{J}}_n = \mathbf{I}_n$ in this exceptional case in order to have a proper unique inverse $(\mathbf{X}_n^T \hat{\mathbf{J}}_n \mathbf{X}_n)^{-1}$ which guarantees the desirable finite sample equivariance and invariance properties.

In the location case, $\mathbf{L}_n^{\nu}$ coincides with (4.17) and hence also with $\mathbf{T}_n^{\nu}$ in (4.16). If $J(t) = I[\alpha_1 \le t \le \alpha_2]$, $t \in (0,1)$, then $\mathbf{L}_n^{\nu}$ is a version of the trimmed LSE proposed in Koenker and Bassett (1978) and studied in Ruppert and Carroll (1980), Jurečková (1983b, 1984) and Koenker and Portnoy (1987).

For a general $J$ taking on positive and negative values, write $J = J^+ - J^-$ and set

$$\mathbf{L}_n^{\nu} = \mathbf{L}_n^{\nu+} - \mathbf{L}_n^{\nu-}. \tag{4.23}$$

Besides the statistics mentioned above other $L$-statistics in the linear model were also studied by more authors. Bickel (1973) defined a class of iteratively computed statistics; however, they are not invariant to a reparametrization. Ruppert and Carroll (1980) compared the Koenker and Bassett trimmed LSE with another estimator which was, like Bickel's one, based on a preliminary estimator of $\boldsymbol{\beta}$. The influence function of two latter statistics turned out to depend heavily on the preliminary estimator. Recently Welsh (1987) showed that a special type of Winsorizing rather than the trimming leads to the desired influence function. Bounded influence versions of the trimmed LSE were proposed in Antoch and Jurečková (1985) and in de Jongh, de Wet and Welsh (1988).

The definitions in (a)–(c) above still need some justification concerning the existence, computational aspects and applicability. Let us first mention that the regression quantile $\hat{\boldsymbol{\beta}}_n(\alpha)$, being implicitly defined, is generally not uniquely determined. Nevertheless, it is possible to select solutions such that $\hat{\boldsymbol{\beta}}_n(\cdot)$ is a stochastic process with the paths in the Skorokhod $(D(0,1))^q$ space (notice the open unit interval). It could be shown [e.g., Gutenbrunner (1986)], that the unique lexicographic maximal solution has the desirable properties. On the other hand, under continuity of $F$, with probability 1, the regression rank-scores $\hat{\mathbf{a}}_n(\alpha)$ are unique for all $\alpha$. $\hat{\mathbf{a}}_n(\cdot)$ is a process with paths in $(C(0,1))^n$, the paths not only being continuous but also piecewise linear with finitely many vertices.

The right-hand derivative $\hat{\mathbf{a}}'_n(\cdot)$ of $\hat{\mathbf{a}}_n(\cdot)$, which is a step-function, may be characterized in the form [compare with (3.10)]

$$\sum_{i \in M_{n\alpha}} \hat{a}'_{ni}(\alpha)\mathbf{x}_{ni} = -\sum_{i=1}^{n} \mathbf{x}_{ni} \qquad (= -n\mathbf{e}_1) \tag{4.24}$$

and

$$\hat{a}'_{ni}(\alpha) = 0 \quad \text{if } i \notin M_{n\alpha}. \tag{4.25}$$

Also, integrating by parts, we may write

$$\hat{J}_{ni} = \int \hat{a}_{ni}(\alpha)\, dJ(\alpha) = -\int \hat{a}'_{ni}(\alpha)\, d\nu(\alpha), \qquad i = 1, \dots, n \tag{4.26}$$

with $J$ being the density of $\nu$.

Both types of $L$-statistics are regression equivariant, that is,

$$\mathbf{S}_n^{\nu}(\sigma \mathbf{Y}_n + \mathbf{X}_n \boldsymbol{\beta}) = \sigma \mathbf{S}_n^{\nu}(\mathbf{Y}_n) + \nu(0,1)\boldsymbol{\beta} \tag{4.27}$$

for $\sigma > 0$ and $\boldsymbol{\beta} \in \mathbb{R}^p$ in the sense that (4.27) holds for $\mathbf{S}_n^{\nu} = \mathbf{L}_n^{\nu}$ and there exists a version of $\mathbf{T}_n^{\nu}$ such that (4.27) holds for $\mathbf{S}_n^{\nu} = \mathbf{T}_n^{\nu}$. Moreover, if $\nu$ is symmetric around $1/2$, then $\mathbf{L}_n^{\nu}$ is symmetric in the sense

$$\mathbf{L}_n^{\nu}(-\mathbf{Y}_n) = -\mathbf{L}_n^{\nu}(\mathbf{Y}_n), \tag{4.28}$$

and this implies that $\mathbf{L}_n^{\nu}$ is a median unbiased estimator of $(\nu(0,1) \cdot \boldsymbol{\beta})$ provided $F$ is symmetric around zero. To obtain the same property for $\mathbf{T}_n^{\nu}$, one must use a very special version of regression quantiles, following the lines of Farebrother (1985).

**5. Asymptotic representations.** When approximating the processes of Section 3, we shall use the topology of uniform convergence on compact subsets of (0, 1). For two processes $\mathbf{A}_n, \mathbf{B}_n$ with realizations in $(D(0,1))^q$, we shall write

$$\mathbf{A}_n = \mathbf{B}_n + o_p^*(1) \quad \left[\text{or } O_p^*(1), o^*(1), O^*(1), \text{respectively}\right],$$

if

$$\begin{aligned} \|\mathbf{A}_n - \mathbf{B}_n\|_{(\varepsilon)} &= \sup_{\varepsilon \le \alpha \le 1-\varepsilon} \|\mathbf{A}_n(\alpha) - \mathbf{B}_n(\alpha)\| \\ &= o_p(1) \quad \left[\text{or } O_p(1), o(1), O(1), \text{respectively}\right] \end{aligned} \tag{5.1}$$

for all $\varepsilon \in (0, 1/2)$. If $\mathbf{B}_n \to_{\mathscr{D}} \mathbf{B}$ and $\mathbf{A}_n = \mathbf{B}_n + o_p^*(1)$, then $\mathbf{A}_n \to_{\mathscr{D}} \mathbf{B}$.

To obtain the weak convergence results on $D[0,1]$ from those on $D(0,1)$, one needs to study the tail behavior of the quantile process; however, no such results are at disposal for the regression quantile process.

Let $\mathbf{W}_n^d = \{\mathbf{W}_n^d(\alpha)\colon 0 < \alpha < 1\}$ denote the process

$$\mathbf{W}_n^d(\alpha) = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \mathbf{d}_{ni}(a_i^*(\alpha) - 1 + \alpha), \tag{5.2}$$

where

$$a_i^*(\alpha) = I\left[U_i > F^{-1}(\alpha)\right], \qquad i = 1, \dots, n. \tag{5.3}$$

The following theorem provides the asymptotic representations for regression quantile and regression rank-score processes.

THEOREM 1. (i) *Under the conditions* (A.1), (B.1–3), (C.1–2), *the regression quantile process* $\mathbf{Z}_n$ *of* (3.5) *admits the representation*

$$\mathbf{Z}_n(\alpha) = \left(f(F^{-1}(\alpha))\right)^{-1} \mathbf{D}_n^{-1} \mathbf{W}_n^x(\alpha) + o_p^*(1). \tag{5.4}$$

*Moreover*,

$$\mathbf{Z}_n \to_{\mathscr{D}} \left(f \circ F^{-1}\right)^{-1} \mathbf{D}^{-1} \mathbf{W}^*_{(p)} \quad \textit{as } n \to \infty, \tag{5.5}$$

*where* $\mathbf{W}^*_{(p)}$ *is a vector of* $p$ *independent Brownian bridges on* $(0,1)$, *and* $f \circ F^{-1}(\alpha) = f(F^{-1}(\alpha))$.

(ii) *Under the additional conditions* (D.1–2), *the regression rank-score process* $\hat{\mathbf{W}}_n^d$ *of* (3.12) *admits the representation*

$$\hat{\mathbf{W}}_n^d = \mathbf{W}_n^{d-d^*} + o_p^*(1), \tag{5.6}$$

*with*

$$\begin{aligned} \mathbf{d}^*_{ni} &= \mathbf{D}_{nd} \mathbf{D}_n^{-1} \mathbf{x}_{ni}, \qquad i = 1, \dots, n, \\ \mathbf{D}_{nd} &= n^{-1} \mathbf{C}_{nd}^T \mathbf{\Gamma}_n^{-1} \mathbf{X}_n, \end{aligned} \tag{5.7}$$

*and with* $\mathbf{W}_n^d$ *of* (5.2).

(iii) *If*, *moreover*,

$$\begin{aligned} \mathbf{Q}_{n[d-d^*]} &= n^{-1}(\mathbf{C}_{nd} - \mathbf{C}_{nd^*})^T(\mathbf{C}_{nd} - \mathbf{C}^*_{nd}) \\ &\to Q_{[d-d^*]} \quad \textit{positive definite of order } q \times q \textit{ as } n \to \infty, \end{aligned} \tag{5.8}$$

*then*

$$\hat{\mathbf{W}}_n^d \to_{\mathscr{D}} \mathbf{Q}^{1/2}_{[d-d^*]} \mathbf{W}^*_{(q)} \quad \textit{as } n \to \infty. \tag{5.9}$$

Corollary 1. *Under the local heteroscedasticity* (1.3) *and under* (A.1), (B.1–3) *and* (D.1–2), *the regression quantile and regression rank-score processes are asymptotically independent.*

Proof of Theorem 1. Put

$$\hat{\boldsymbol{\delta}}_n(\alpha) = \hat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}, \qquad \boldsymbol{\delta}(\alpha) = \boldsymbol{\beta}(\alpha) - \boldsymbol{\beta} = F^{-1}(\alpha)\boldsymbol{\gamma}. \tag{5.10}$$

Then $\mathbf{Z}_n = n^{1/2}(\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta})$. We shall further use the empirical processes

$$\hat{\mathbf{G}}_n^d(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n \mathbf{d}_{ni} I\left[U_i \le \mathbf{x}_{ni}^T \mathbf{t}/\sigma_{ni}\right], \tag{5.11}$$

$$\mathbf{G}_n^d(\mathbf{t}) = E\hat{\mathbf{G}}_n^d(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n \mathbf{d}_{ni} F\left(\mathbf{x}_{ni}^T \mathbf{t}/\sigma_{ni}\right), \qquad \mathbf{t} \in \mathbb{R}^p, \tag{5.12}$$

and

$$\mathbf{Y}_n^d = n^{1/2}\left(\hat{\mathbf{G}}_n^d - \mathbf{G}_n^d\right), \tag{5.13}$$

where we denoted

$$\sigma_{ni} = \mathbf{x}_{ni}^T \boldsymbol{\gamma}, \qquad i = 1, \dots, n. \tag{5.14}$$

Regarding (3.9) and (3.10), we see that the difference $\hat{\mathbf{W}}_n^d - \mathbf{W}_n^d$ is essentially $-n^{-1/2}(\hat{\mathbf{G}}_n^d \circ \hat{\boldsymbol{\delta}}_n - \hat{\mathbf{G}}_n^d \circ \boldsymbol{\delta}_n)$. Linearizing the latter expression, we may replace it by

$$(f \circ F^{-1}) \mathbf{D}_{nd} \mathbf{Z}_n,$$

arriving at the important basic relation

$$\hat{\mathbf{W}}_n^d = \mathbf{W}_n^d - (f \circ F^{-1})\mathbf{D}_{nd}\mathbf{Z}_n - \mathbf{R}_{1n}^d + \mathbf{R}_{2n}^d, \tag{5.15}$$

where

$$\mathbf{R}_{1n}^d = n^{1/2}\Big(\hat{\mathbf{G}}_n^d \circ \hat{\boldsymbol{\delta}}_n - \hat{\mathbf{G}}_n^d \circ \boldsymbol{\delta}_n\Big) - (f \circ F^{-1})\mathbf{D}_{nd}\mathbf{Z}_n \tag{5.16}$$

and

$$\mathbf{R}_{2n}^d = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \mathbf{d}_{ni} I\Big[Y_{ni} = \mathbf{x}_{ni}^T \hat{\boldsymbol{\beta}}_n(\cdot)\Big]\hat{a}_{ni}(\cdot). \tag{5.17}$$

By (D.1) and regarding that $|M_{n\alpha}| \le p$ and $|\hat{a}_{ni}(\alpha)| \le 1$, we obtain that

$$\big\|\mathbf{R}_{2n}^d(\alpha)\big\| \le pn^{-1/2} \max_{1 \le i \le n} \|\mathbf{d}_{ni}\| \to 0 \quad \text{as } n \to \infty \tag{5.18}$$

uniformly in $\alpha$. Concerning $\mathbf{R}_{1n}^d$, we shall split it in $\mathbf{R}_{1n}^d = \mathbf{R}_{3n}^d + \mathbf{R}_{4n}^d$, where

$$\mathbf{R}_{3n}^d = \mathbf{Y}_n^d \circ \hat{\boldsymbol{\delta}}_n - \mathbf{Y}_n^d \circ \boldsymbol{\delta} \tag{5.19}$$

and

$$\mathbf{R}_{4n}^d = n^{1/2}\Big(\mathbf{G}_n^d \circ \hat{\boldsymbol{\delta}}_n - \mathbf{G}_n^d \circ \boldsymbol{\delta}\Big) - (f \circ F^{-1})\mathbf{D}_{nd}\mathbf{Z}_n. \tag{5.20}$$

We shall treat $\mathbf{R}_{3n}^d$ using a uniform continuity argument [see, e.g., Koul (1969)] while $\mathbf{R}_{4n}^d$ will be treated using a uniform differentiability property of the mapping $\mathbf{G}_n^d$. This will be the contents of the following lemmas which will be proved in the Appendix.

LEMMA 1. (i) *Under the conditions* (A.1), (B.1–3), (C.1–2) *and* (D.2),

$$\begin{aligned} &\sup\Big\| n^{1/2}\big[\mathbf{G}_n^d(\boldsymbol{\delta}(\alpha) + n^{-1/2}\mathbf{t}) - \mathbf{G}_n^d(\boldsymbol{\delta}(\alpha))\big] - f(F^{-1}(\alpha))\mathbf{D}_{nd}\mathbf{t}\Big\| \\ &\qquad = o(1) \quad \textit{for any } K > 0 \textit{ and } \varepsilon > 0, \end{aligned} \tag{5.21}$$

*where the supremum is taken over all* $(\mathbf{t}, \alpha)$ *with* $\|\mathbf{t}\| \le K$, $\varepsilon \le \alpha \le 1 - \varepsilon$.

(ii) *If, moreover,* (D.1) *is satisfied, then also*

$$\sup\Big\|\mathbf{Y}_n^d(\boldsymbol{\delta}(\alpha) + n^{-1/2}\mathbf{t}) - \mathbf{Y}_n^d(\boldsymbol{\delta}(\alpha))\Big\| = o_p(1), \tag{5.22}$$

*with the supremum taken over the same set* $(\mathbf{t}, \alpha)$ *as in* (i).

The next lemma is a corollary of Theorem 2.2 of Shorack (1979).

LEMMA 2 (Shorack). (i) *Under* (D.1–2), *the sequence* $\{\mathbf{W}_n^d\}$ *is uniformly tight on* $(D[0,1])^q$.

(ii) *Under* (D.1–3),

$$\mathbf{W}_n^d \to \mathbf{Q}_{[d]}^{1/2}\mathbf{W}_{(q)}^* \quad \textit{on } (D[0,1])^q \tag{5.23}$$

*with* $\mathbf{W}_{(q)}^*$ *being the vector of q independent Brownian bridges.*

LEMMA 3. *Under* (A.1), (B.1–3) *and* (C.1–3),

$$\mathbf{Z}_n = O_p^*(1). \tag{5.24}$$

COMPLETION OF THE PROOF OF THEOREM 1. Notice that $\hat{\mathbf{W}}_n^x \equiv \mathbf{0}$ by (3.8). Then, under $\mathbf{d}_{ni} = \mathbf{x}_{ni}$, $i = 1, \dots, n$, (5.15) and Lemmas 1 and 2 imply

$$(f \circ F^{-1})\mathbf{D}_n \mathbf{Z}_n = \mathbf{W}_n^x + o_p^*(1) \to_{\mathscr{D}} \mathbf{Q}^{1/2}\mathbf{W}_{(p)}^* \tag{5.25}$$

and this further implies the proposition (i). Inserting (5.4) in (5.15) yields

$$\hat{\mathbf{W}}_n^d = \mathbf{W}_n^d - \mathbf{D}_{nd}\mathbf{D}_n^{-1}\mathbf{W}_n^x + o_p^*(1) = \mathbf{W}_n^{d-d^*} + o_p^*(1) \tag{5.26}$$

and hence (ii).

The sequence $\{\mathbf{W}_n^{d-d^*}\}$ is uniformly tight by Lemma 2; actually, $\{\mathbf{d}_{ni} - \mathbf{d}_{ni}^*\}$ satisfies (D.1) and (D.2) due to the inequalities

$$\|\mathbf{d}_{ni} - \mathbf{d}_{ni}^*\| \le \|\mathbf{d}_{ni}\| + \|\mathbf{D}_{nd}\|\,\|\mathbf{D}_n^{-1}\|\,\|\mathbf{x}_{ni}\|, \qquad i = 1, \dots, n,$$

and

$$\|\mathbf{D}_{nd}\| \le \frac{c}{n}\sum_{i=1}^{n} \|\mathbf{d}_{ni}\|\,\|\mathbf{x}_{ni}\| = O(1).$$

This together with (i), (ii) and Lemma 2(ii) implies (iii). □

Note that Theorem 1 was proved under mild conditions, which are necessary even for the consistency and asymptotic normality of the ordinary LSE [only (A.1) should be then replaced by $\sigma^2(F) < \infty$]. Moreover, $f$ does not need to be differentiable and symmetric. The proofs use the elementary techniques (this applied also to the proofs of Lemmas 1–3).

PROOF OF COROLLARY 1. Notice that, under the local heteroscedasticity (1.3), $\mathbf{X}_n$ and $\mathbf{C}_{nd} - \mathbf{C}_{nd^*} = \mathbf{C}_{n,d-d^*}$ are asymptotically orthogonal in the sense that

$$n^{-1}\mathbf{X}_n^T \mathbf{C}_{n,d-d^*} \to \mathbf{0} \quad \text{as } n \to \infty. \tag{5.27}$$

It follows from the proof of Theorem 1(ii) that the sequence $\{\mathbf{Q}_{n[d-d^*]}\}$ is relatively compact, hence (5.8) holds for some subsequence $\{n_m\}$ of positive integers. Denoting

$$\tilde{\mathbf{d}}_{ni} = \left(\mathbf{x}_{ni}^T, (\mathbf{d}_{ni} - \mathbf{d}_{ni}^*)^T\right)^T \in \mathbb{R}^{p+q},$$

then, by (5.27), $n_m^{-1}\mathbf{C}_{n_m\tilde{d}}^T \cdot \mathbf{C}_{n_m\tilde{d}}$ converges to the matrix

$$\mathbf{Q}_{[d]} = \begin{pmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & Q_{[d-d^*]} \end{pmatrix}$$

as $m \to \infty$. Then, by Theorem 1,

$$\begin{pmatrix} (f \circ F^{-1}) \mathbf{D}_{n_m} \mathbf{Z}_{n_m} \\ \hat{\mathbf{W}}^d_{n_m} \end{pmatrix} = \mathbf{W}^{\bar{d}}_{n_m} + o_p^*(1) \to_{\mathscr{D}} \mathbf{Q}^{1/2}_{[\bar{d}]} \mathbf{W}^*_{(p+q)} = \begin{pmatrix} \mathbf{Q}^{1/2}, \mathbf{W}^*_{(p)} \\ \mathbf{Q}^{1/2}_{d-d^*}, \mathbf{W}^*_{(q)} \end{pmatrix}.$$

Hence, the Prokhorov distance between the simultaneous distribution of $([(f \circ F^{-1})\mathbf{D}_{n_m}\mathbf{Z}_{n_m}]^T, [\hat{\mathbf{W}}^d_{n_m}]^T)^T$ and the product of its marginal distribution tends to zero; the same holds also for $(\mathbf{Z}^T_{n_m}, (\hat{\mathbf{W}}^d_{n_m})^T)^T$. Since every sequence of positive integers contains a subsequence with this property, the assertion follows. □

The next theorem provides the asymptotic representations of $L$-statistics of the first type and of the linear rank statistics.

THEOREM 2. (i) *Under the conditions* (A.1), (B.1–3), (C.1–3) *and* (W.1) *as* $n \to \infty$,

$$n^{1/2}(\mathbf{T}^\nu_n - \boldsymbol{\beta}(\nu, F)) = n^{-1/2} \sum_{i=1}^n \psi_{\nu, F}(U_i) \mathbf{D}_n^{-1} \mathbf{x}_{ni} + o_p(1). \tag{5.28}$$

*Hence*, $n^{1/2}(\mathbf{T}^\nu_n - \boldsymbol{\beta}(\nu, F))$ *is asymptotically normally distributed*

$$N_p(\mathbf{0}, \mathbf{D}^{-1}\mathbf{Q}\mathbf{D}^{-1}\sigma^2(\nu, F)), \tag{5.29}$$

*where*

$$\boldsymbol{\beta}(\nu, F) = \nu(0,1)\boldsymbol{\beta} + \mu(\nu, F)\boldsymbol{\gamma} \quad \textit{with } \mu(\nu, F) = \int_0^1 F^{-1}(\alpha)\, d\nu(\alpha), \tag{5.30}$$

$$\psi_{\nu, F}(t) = \int_0^1 (\alpha - I[F(t) \le \alpha]) [f(F^{-1}(\alpha))]^{-1} \, d\nu(\alpha), \qquad t \in \mathbb{R}, \tag{5.31}$$

*and*

$$\begin{aligned} \sigma^2(\nu, F) &= \int \psi^2_{\nu, F}(t)\, dF(t) \\ &= \int_0^1 \int_0^1 (u \wedge v - uv) \big( f(F^{-1}(u)) f(F^{-1}(v)) \big)^{-1} \, d\nu(u)\, d\nu(v). \end{aligned}$$

*If* $\boldsymbol{\gamma} = \boldsymbol{\gamma}_n = \mathbf{e}_1 + O(n^{-1/2})$, *then* $\mathbf{D}_n$ *and* $\mathbf{D}$ *in* (5.28) *and* (5.29) *could be replaced by* $\mathbf{Q}_n$ *and* $\mathbf{Q}$, *respectively.*

(ii) *Under* (A.1), (B.1–2), (C.1–3), (W.2), (D.1–2),

$$n^{1/2}\left(\hat{\mathbf{V}}^{b,d}_n - \bar{b}\bar{\mathbf{d}}_n\right) = \frac{1}{n^{1/2}} \sum_{i=1}^n (\mathbf{d}_{ni} - \mathbf{d}^*_{ni})(b \circ F(U_i)) + o_p(1). \tag{5.32}$$

*Furthermore, if* (5.8) *holds, then* $n^{1/2}(\hat{\mathbf{V}}^{b,d}_n - \bar{b}\bar{\mathbf{d}}_n)$ *is asymptotically normally distributed*

$$N_q\left(\mathbf{0}, \mathbf{Q}_{[d-d^*]} \int_0^1 \left(b(u) - \bar{b}\right)^2 du\right). \tag{5.33}$$

PROOF. Theorem 2 follows immediately from Theorem 1 through integrating with respect to $\nu$ and $b$, respectively, and using relations of the type $\int(A_n + o_p^*(1))\,d\nu = \int A_n\,d\nu + o_p(1)$ together with the central limit theorem and the Cramér–Wold theorem. The possible replacement of $\mathbf{D}_n^{-1}$ by $\mathbf{Q}_n^{-1}$ under local heteroscedasticity follows from the fact that then $\mathbf{D}_n^{-1} - \mathbf{Q}_n^{-1} \to \mathbf{0}$ and that the right-hand side of (5.28) is bounded in probability. □

REMARKS. (i) In the special case with $\boldsymbol{\gamma} = \mathbf{e}_1$, $\nu$ possessing a bounded Lebesgue density and with $\mathbf{X}_n$ and $F$ satisfying some additional restrictions [namely, (X.1–4) and F of Koenker and Portnoy (1987)], the proposition (i) is covered by Theorem 3.1 of Koenker and Portnoy (1987).

(ii) Using the known representation of $M$- and $R$-estimators (in the homoscedastic model with $\mathbf{D}_n = \mathbf{Q}_n$), we obtain asymptotic relations of $L$-, $M$- and $R$-statistics which are analogous to those in the location model.

(iii) The asymptotic representation (5.32) of $\hat{\mathbf{V}}_n^{b,d}$ in the homoscedastic case coincides with that of the aligned rank statistics $\hat{S}_n^{b,d}$ used for testing by Adichie (1978), Sen (1969), Chiang and Puri (1984), among others. However, the finite-sample behavior of the two types of statistics may be rather different, as illustrated by the following example.

EXAMPLE. Assuming a two-factor model with interaction and a balanced design, $n = 4m$, we may have the following design matrix:

| | $x_{ni1}$ | $x_{ni2}$ | $d_{ni1}$ | $d_{ni2}$ |
|---|---|---|---|---|
| $1 \le i \le m$ | 1 | $-1$ | $-1$ | 1 |
| $m < i \le 2m$ | 1 | $-1$ | 1 | $-1$ |
| $2m < i \le 3m$ | 1 | 1 | $-1$ | $-1$ |
| $3m < i \le 4m$ | 1 | 1 | 1 | 1 |

Under the homoscedastic model,

$$Y_{ni} = x_{xi1}\beta_{n1} + x_{ni2}\beta_{n2} + d_{ni1}\beta_{n1}^d + d_{n2}\beta_{n2}^d + U_i,$$

thus $\beta_{n2}$ corresponds to the main effect of the first factor, $\beta_{n1}^d$ to the main effect of the second factor and $\beta_{n2}^d$ to the interaction, respectively. We may test the hypothesis

$$H_0\colon \beta_{n1}^d = \beta_{n2}^d = 0 \text{ (no effect of the second factor)}$$

by means of $\hat{V}_n^{b,d}$ as well of $\hat{S}_n^{b,d}$, respectively. Since, for this special design, the regression quantiles reduce to simple functions of the order statistics in the separate subsamples

$$\mathbf{Y}_1 = \{Y_{ni}, l \le i \le 2m\} \quad \text{and} \quad \mathbf{Y}_2 = \{Y_{ni}, 2m < i \le 4m\},$$

the statistic $\hat{V}_n^{b,d}$ turns out to be a simple linear function of a pair of two ordinary two-sample (location) rank statistics based on the ranks of $Y_{ni}$ in $\mathbf{Y}_1$ or $\mathbf{Y}_2$, $1 \le i \le 2m$ or $2m < i \le 4m$, respectively. On the other hand, the aligned rank statistic $\hat{S}_n^{b,d}$ is based on the ranks of aligned observations with respect to the whole sample of size $4m$. Hence, in this example, $\hat{V}_n^{b,d}$ corresponds to blockwise ranking rather than to ranking after an alignment.

The asymptotic behavior of the aligned rank statistics under nonlocal heteroscedasticity is not yet known; their representations may differ from (5.32). □

The asymptotic representation (5.28) of $\mathbf{T}_n^\nu$ could be further rewritten in the following form

$$\mathbf{T}_n^\nu = \mathbf{K}_{n,\gamma}\mathbf{Y}_n^*(\nu, F) + o_p(n^{-1/2}), \tag{5.34}$$

and under the local heteroscedasticity (1.3) in the form

$$\mathbf{T}_n^\nu = \mathbf{K}_n\mathbf{Y}_n^*(\nu, F) + o_p(n^{-1/2}), \tag{5.35}$$

where

$$\mathbf{K}_{n,\gamma} = \left(\mathbf{X}_n^T\mathbf{\Gamma}_n^{-1}\mathbf{X}_n\right)^{-1}\mathbf{X}_n^T\mathbf{\Gamma}_n^{-1}, \tag{5.36}$$

$$\mathbf{K}_n = \left(\mathbf{X}_n^T\mathbf{X}_n\right)^{-1}\mathbf{X}_n^T \tag{5.37}$$

and

$$Y_{ni}^*(\nu, F) = \mathbf{x}_{ni}^T\boldsymbol{\beta} + \mathbf{x}_{ni}^T\boldsymbol{\gamma}\big(\psi_{\nu,F}(U_i) + \mu(\nu, F)\big), \qquad i = 1,\dots,n. \tag{5.38}$$

Hence, (5.34) and (5.35) represent $\mathbf{T}_n^\nu$ as the weighted or ordinary LSE, respectively, corresponding to the pseudo-observations (5.38), up to $o_p(n^{-1/2})$. This, being combined with the next Theorem 3, shows that while both types of $L$-statistics ($\mathbf{T}_n^\nu$ and $\mathbf{L}_n^\nu$) are asymptotically equivalent under the local heteroscedasticity (1.3), their difference is nondegenerate under the general heteroscedasticity.

Concerning the representation (5.32) of the linear rank statistic, notice that it reduces to the well-known representation in the location model [see (15) in Section V.1.5 of Hájek and Šidák (1967)]. Actually,

$$\mathbf{C}_{nd^*} = \begin{pmatrix} \mathbf{d}_{n1}^{*T} \\ \vdots \\ \mathbf{d}_{nn}^{*T} \end{pmatrix} = \mathbf{X}_n\mathbf{K}_{n,\gamma}\mathbf{C}_{nd}$$

and hence $\mathbf{C}_{nd^*}$ is a projection of $\mathbf{C}_{nd}$ onto the column space of $\mathbf{X}_n$ (in the heteroscedastic case nonorthogonal).

**6. Asymptotic relations of two kinds of $L$-statistics.** We shall now consider the asymptotic relations of $\mathbf{T}_n^\nu$ and $\mathbf{L}_n^\nu$, two kinds of $L$-statistics introduced in Section 4. While they are asymptotically equivalent under the homoscedastic or under the local heteroscedastic models, generally this is the case only for the intercept components; the difference of the slope component vectors has a nondegenerate asymptotic $(p-1)$-dimensional normal distribution.

THEOREM 3. (i) *Under* (A.1), (B.1–4), (C.1–2) *and* (W.3),

$$\mathbf{L}_n^{\nu} = \mathbf{K}_n \mathbf{Y}_n^{**}(\nu, F) + \mathbf{K}_{n,\gamma} \tilde{\mathbf{Y}}_n(\nu, F) + o_p(n^{-1/2}), \tag{6.1}$$

*where*

$$Y_{ni}^{**}(\nu, F) = \mathbf{x}_{ni}^T \boldsymbol{\beta} + \mathbf{x}_{ni}^T \boldsymbol{\gamma} \psi_{\nu,F}^{**}(U_i), \tag{6.2}$$

$$\tilde{\mathbf{Y}}_n(\nu, F) = \mathbf{Y}_n^*(\nu, F) - \mathbf{Y}_n^{**}(\nu, F) \qquad [\mathbf{Y}_n^* \text{ of } (5.38)] \tag{6.3}$$

$$\psi_{\nu,F}^{**}(t) = \begin{cases} J(F(t)) \cdot \left[t - \mu(\nu, F)(\nu(0,1))^{-1}\right], & \text{if } \nu \text{ is positive}, \\ \psi_{\nu^+,F}^{**}(t) - \psi_{\nu^-,F}^{**}(t), & \text{otherwise}. \end{cases} \tag{6.4}$$

(ii) *Under the additional assumption* (1.3) *of local heteroscedasticity*,

$$\mathbf{L}_n^{\nu} = \mathbf{K}_n \mathbf{Y}_n^*(\nu, F) + O_p(n^{-1}) = \mathbf{T}_n^{\nu} + o_p(n^{-1/2}). \tag{6.5}$$

PROOF. We may assume that $\nu$ is a probability measure, $\nu(0,1) = \int J(u)\,du = 1$, without loss of generality. We shall first show that

$$\mathbf{L}_n^{\nu} = \mathbf{K}_n \mathbf{Y}_n^*(\nu, F) + \mathbf{Q}_n^{-1} \hat{\mathbf{V}}_n^{b,g} + O_p(n^{-1}), \tag{6.6}$$

where $\hat{\mathbf{V}}_n^{b,g}$ has score function $b = b_{J,F} - \bar{b}_{J,F}$ with

$$b_{J,F}(\alpha) = \int_{(0,\alpha]} \left[F^{-1}(u) - \mu(\nu, F)\right] dJ(u) \tag{6.7}$$

and the regression constants

$$\mathbf{g}_{ni} = \mathbf{x}_{ni}^T \boldsymbol{\gamma} \mathbf{x}_{ni} = \sigma_{ni} \mathbf{x}_{ni}, \qquad i = 1, \dots, n. \tag{6.8}$$

(6.6) could be rewritten as

$$\begin{aligned} n^{1/2}(\mathbf{L}_n^{\nu} - \boldsymbol{\beta}(\nu, F)) = \mathbf{Q}_n^{-1} \left\{ n^{-1/2} \sum_{i=1}^{n} \psi_{\nu,F}(U_i) \mathbf{g}_{ni} + \int \hat{\mathbf{W}}_n^g \, db_{J,F} \right\} \\ + O_p(n^{-1/2}). \end{aligned} \tag{6.9}$$

Denoting

$$\hat{\mathbf{Q}}_n = n^{-1} \mathbf{X}_n^T \hat{\mathbf{J}}_n \mathbf{X}_n, \tag{6.10}$$

then, regarding (4.9), $\mathbf{L}_n^{\nu} = n^{-1} \hat{\mathbf{Q}}_n^{-1} \mathbf{X}_n^T \hat{\mathbf{J}}_n \mathbf{Y}_n$ and this, regarding the form (5.30) of $\boldsymbol{\beta}(\nu, F)$ and the identity $n^{-1} \hat{\mathbf{Q}}_n^{-1} \mathbf{X}_n^T \hat{\mathbf{J}}_n \mathbf{X}_n = \mathbf{I}_p$, could be further rewritten as

$$n^{1/2}(\mathbf{L}_n^{\nu} - \boldsymbol{\beta}(\nu, F)) = n^{-1} \hat{\mathbf{Q}}_n^{-1} \mathbf{X}_n^T \hat{J}_n(\mathbf{Y}_n - \mathbf{X}_n \boldsymbol{\beta} - \mu(\nu, F) \mathbf{X}_n \boldsymbol{\gamma}). \tag{6.11}$$

By Lemma 6 [see (L6.2)] of the Appendix, $\hat{\mathbf{Q}}_n^{-1} = \mathbf{Q}_n^{-1} + O_p(n^{-1/2})$ and, by (L6.1),

$$\hat{\mathbf{J}}_n = \operatorname{diag} \hat{\mathbf{J}}_{ni} + o_p(0),$$

where we say that $A_n = B_n + o_p(0)$ if $P(A_n \neq B_n) \to 0$ as $n \to \infty$. Hence,

$$
\begin{aligned}
&n^{1/2}\hat{\mathbf{Q}}_n(\mathbf{L}_n^{\nu} - \boldsymbol{\beta}(\nu, F)) \\
&\quad = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \mathbf{x}_{ni}\hat{J}_{ni}\big(Y_{ni} - \mathbf{x}_{ni}^T\boldsymbol{\beta} - \mu(\nu, F)\mathbf{x}_{ni}^T\boldsymbol{\gamma}\big) + o_p(0) \\
&\quad = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \mathbf{x}_{ni}^T\boldsymbol{\gamma}\mathbf{x}_{ni}\hat{J}_{ni}(U_i - \mu(\nu, F)) + o_p(0) \\
&\quad = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \mathbf{g}_{ni}\hat{J}_{ni}(U_i - \mu(\nu, F)) + o_p(0).
\end{aligned} \tag{6.12}
$$

Introduce the interpolating quantities

$$
\hat{U}_{ni}(\alpha) = \hat{a}_{ni}(\alpha)U_i + \big(1 - \hat{a}_{ni}(\alpha)\big)F^{-1}(\alpha), \qquad i = 1, \dots, n; \qquad 0 < \alpha < 1. \tag{6.13}
$$

Regarding the identity

$$
U_i \vee F^{-1}(\alpha) = a_i^*(\alpha)U_i + \big(1 - a_i^*(\alpha)\big)F^{-1}(\alpha) \tag{6.14}
$$

with $a_i^*(\alpha)$ as in (5.3), $i = 1, \dots, n$; $0 < \alpha < 1$, we get

$$
\hat{U}_{ni}(\alpha) - \big(U_i \vee F^{-1}(\alpha)\big) = \big(\hat{a}_{ni}(\alpha) - a_i^*(\alpha)\big)\big(U_i - F^{-1}(\alpha)\big), \tag{6.15}
$$

$i = 1, \dots, n$; $0 < \alpha < 1$, and Lemma 5 of the Appendix yields

$$
\sum_{i=1}^{n} \mathbf{g}_{ni} \int \big[\hat{U}_{ni}(\alpha) - \big(U_i \vee F^{-1}(\alpha)\big)\big]\, dJ(\alpha) = O_p(1). \tag{6.16}
$$

Moreover, we have the identity

$$
\begin{aligned}
&\big[\hat{a}_{ni}(\alpha)(U_i - \mu(\nu, F))\big] - \big[U_i \vee F^{-1}(\alpha) - \alpha\big(F^{-1}(\alpha) - \mu(\nu, F)\big)\big] \\
&\quad = \big[\hat{U}_{ni}(\alpha) - U_i \vee F^{-1}(\alpha)\big] \\
&\qquad + \big[\big(\hat{a}_{ni}(\alpha) - 1 + \alpha\big)\big(F^{-1}(\alpha) - \mu(\nu, F)\big)\big] - \mu(\nu, F).
\end{aligned} \tag{6.17}
$$

By (W.3), it holds that

$$
\int dJ = J(1-) - J(0+) = 0 \tag{6.18}
$$

and integrating by parts in (5.31) we obtain

$$
\psi_{\nu, F}(t) = \int \big[\big(t \vee F^{-1}(\alpha)\big) - \alpha\big(F^{-1}(\alpha) - \mu(\nu, F)\big)\big]\, dJ(\alpha). \tag{6.19}
$$

Combining (6.16)–(6.19) we obtain

$$
\begin{aligned}
&\frac{1}{n^{1/2}} \sum_{i=1}^{n} \mathbf{g}_{ni}\big[\hat{J}_{ni}(U_i - \mu(\nu, F)) - \psi_{\nu, F}(U_i)\big] \\
&\quad = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \mathbf{g}_{ni} \int \big(\hat{U}_{ni}(\alpha) - \big(U_i \vee F^{-1}(\alpha)\big)\big)\, dJ(\alpha) + \int \hat{\mathbf{W}}_n^g \, db_{J, F} \\
&\quad = \int \hat{\mathbf{W}}_n^g \, db_{J, F} + O_p(n^{-1/2})
\end{aligned} \tag{6.20}
$$

and this, being combined with (6.12), yields

$$
\begin{aligned}
&n^{-1/2}\mathbf{X}_n^T\hat{\mathbf{J}}_n(\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta} - \mathbf{X}_n\boldsymbol{\gamma}\mu(\nu, F)) \\
&\qquad = n^{-1/2}\sum_{i=1}^{n} \mathbf{g}_{ni}\psi_{\nu,F}(U_i) + \int \hat{\mathbf{W}}_n^g \, db_{J,F} + O_p(n^{-1/2}).
\end{aligned}
\tag{6.21}
$$

Due to Lemma 6 and due to the fact that the right-hand side (r.h.s.) of (6.21) is stochastically bounded, we may multiply the l.h.s. and r.h.s. of (6.21) by $\hat{\mathbf{Q}}_n^{-1}$ and $\mathbf{Q}_n^{-1}$, respectively, without changing the order $O_p(n^{-1/2})$. This proves (6.9) and hence (6.6).

The behavior of the term $\int \hat{\mathbf{W}}_n^g \, db = n^{1/2}\hat{\mathbf{V}}_n^{b,g}$ is different in homo- and heteroscedastic cases: In the former case it follows from (6.8) that $\mathbf{g}_{ni} = \mathbf{x}_{ni}$, $i = 1, \dots, n$ and hence $\hat{\mathbf{W}}_n^g = \hat{\mathbf{W}}_n^x = \mathbf{0}$. Under the local heteroscedasticity, $\boldsymbol{\gamma}_n = \mathbf{e} + n^{1/2}\boldsymbol{\gamma}_{on}$, $\|\boldsymbol{\gamma}_{on}\| = O(1)$ setting

$$
\mathbf{d}_{ni} := \mathbf{x}_{ni}^T\boldsymbol{\gamma}_{on}\mathbf{x}_{ni} = n^{-1/2}(\mathbf{g}_{ni} - \mathbf{x}_{ni}), \qquad i = 1, \dots, n,
$$

we obtain from Lemma 5 in the Appendix,

$$
\begin{aligned}
\int \hat{\mathbf{W}}_n^g \, db &= \int \hat{\mathbf{W}}_n^{g-x} \, db = n^{-1/2}\int \hat{\mathbf{W}}_n^d \, db \\
&= \frac{1}{n^{1/2}} O_p\left(\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{d}_{ni}\|^2\right)^{1/2} \\
&= \frac{1}{n^{1/2}} O_p\left(\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{x}_{ni}\|^4\right)^{1/2} = O_p(n^{-1/2})
\end{aligned}
$$

and, on the other hand,

$$
\frac{1}{n^{1/2}}\sum_{i=1}^{n} (\mathbf{g}_{ni} - \mathbf{x}_{ni})\psi_{\nu,F}(U_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_{ni}\psi_{\nu,F}(U_i) = O_p(n^{-1/2}),
$$

and this completes the proof of part (ii).

In the general case, applying Theorem 2 to $\hat{\mathbf{V}}_n^{b,g}$, regarding

$$
b_{J,F}(F(t)) = \psi_{\nu,F}^{**}(t) - \psi_{\nu,F}(t) + \text{const.} = -\tilde{\psi}_{\nu,F}(t) + \text{const.},
$$

and $\sum_{i=1}^{n}(\mathbf{g}_{ni} - \mathbf{g}_{ni}^*) = 0$ for $g_{ni}^* = \mathbf{Q}_n\mathbf{D}_n^{-1}\mathbf{x}_{ni}$, $i = 1, \dots, n$, we get

$$
\mathbf{Q}_n^{-1}\int \hat{\mathbf{W}}_n^g \, db_{J,F} = n^{-1/2}\sum_{i=1}^{n} \tilde{\psi}_{\nu,F}(U_i)\big(\mathbf{D}_n^{-1}\mathbf{x}_{ni} - \mathbf{Q}_n^{-1}\mathbf{g}_{ni}\big) + o_p(1)
$$

This, together with (6.9), yields

$$
\begin{aligned}
&n^{-1/2}(\mathbf{L}_n^\nu - \boldsymbol{\beta}_n(\nu, F)) \\
&\qquad = n^{-1/2}\sum_{i=1}^{n} \Big[\mathbf{Q}_n^{-1}\mathbf{g}_{ni}\big(\psi_{\nu,F}(U_i) - \tilde{\psi}_{\nu,F}(U_i)\big) \\
&\qquad\qquad + \mathbf{D}_n^{-1}\mathbf{x}_{ni}\tilde{\psi}_{\nu,F}(U_i)\Big] + o_p(1)
\end{aligned}
\tag{6.22}
$$

and this is equivalent to (6.1). □

Theorem 3 shows that the first components of $\mathbf{T}_n^\nu$ and $\mathbf{L}_n^\nu$ are asymptotically equivalent even under the general heteroscedastic model. This is in correspondence with fact that the first component estimates the location. However, the slope components of both estimators generally have different asymptotic distributions, as we shall illustrate in the following example.

EXAMPLE. Fix $\alpha \in (0, 1/2)$ and consider $J(u) = (1 - 2\alpha)^{-1} I[\alpha \le u \le 1 - \alpha]$. Then $\mathbf{L}_n^\nu$ is a version of the trimmed LSE introduced by Koenker and Bassett (1978). It was shown in Gutenbrunner (1986) that, under the local heteroscedasticity model, all versions of the trimmed LSE are equivalent up to the order $O_p(n^{-1}\|\mathbf{X}_n\|_\infty)$.

If we could assume

$$\mathbf{S}_n = n^{-1}\mathbf{X}_n^T \boldsymbol{\Gamma}_n^2 \mathbf{X}_n \to \mathbf{S}, \qquad (p \times p),$$

then the asymptotic covariance matrix of $\mathbf{L}_n^\nu$ under general model and for symmetric $f$ takes on the form

$$\boldsymbol{\Sigma}_L = h_1 \mathbf{D}^{-1}\mathbf{Q}\mathbf{D}^{-1} + h_2 \mathbf{Q}^{-1}\mathbf{S}\mathbf{Q}^{-1},$$

where $h_1 = 2\alpha c_\alpha^2$, $h_2 = \int_{|t| \le c_\alpha} t^2\, dF(t)$, $c_\alpha = F^{-1}(1 - \alpha)$. Indeed, in this case, $\psi_{\nu, F}$ coincides with Huber function $\psi_{\nu, t}(t) = -c_\alpha \vee t \wedge c_\alpha$, $\psi^{**}(t) = tI[|t| \le c_\alpha]$ and $\tilde{\psi}_{\nu, F} = c_\alpha(I[t > c_\alpha] - I[t < -c_\alpha])$, $t \in \mathbb{R}^1$; in this case $\tilde{\psi} \cdot \psi^{**} \equiv 0$ and $E\psi^{**}(U_1) = 0$.

On the other hand, the asymptotic covariance matrix $\boldsymbol{\Sigma}_T$ of $\mathbf{T}_n^\nu$ has the form $\boldsymbol{\Sigma}_T = (h_1 + h_2)\mathbf{D}^{-1}\mathbf{Q}\mathbf{D}^{-1}$ and hence

$$\boldsymbol{\Sigma}_L - \boldsymbol{\Sigma}_T = h_2(\mathbf{Q}^{-1}\mathbf{S}\mathbf{Q}^{-1} - \mathbf{D}^{-1}\mathbf{Q}\mathbf{D}^{-1}). \tag{6.23}$$

More specifically, let $p = 2$, $n = 3m$, $m \in N$, hence $\mathbf{X}_n$ is of order $n \times 2$ with the first column of units. Let the second column of $\mathbf{X}_n$ be as follows:

$$x_{in2} = \begin{cases} \dfrac{-d}{3}, & \text{if } 1 \le i \le m, \\ 0, & \text{if } m < i < 2m, \\ \dfrac{d}{3}, & \text{if } 2m < i \le 3m, \end{cases}$$

with $d = 3\sqrt{3/2}$; the heteroscedasticity is fixed to $\gamma = (1, 1/d)^T$. Then $\mathbf{Q} = \mathbf{I}_2$, thus $\boldsymbol{\Sigma}_L - \boldsymbol{\Sigma}_T = h_2(\mathbf{S} - \mathbf{D}^{-2})$, where

$$\mathbf{S} = \frac{1}{27}\begin{pmatrix} 29 & 4d \\ 4d & \dfrac{20d^2}{9} \end{pmatrix} = \frac{1}{27}\begin{pmatrix} 29 & 12\sqrt{\dfrac{3}{2}} \\ 12\sqrt{\dfrac{3}{2}} & 30 \end{pmatrix}$$

and

$$\mathbf{D}^{-2} = \frac{1}{27}\begin{pmatrix} 29 & \dfrac{53}{d} \\ \dfrac{53}{d} & 2\left(1 + \left(\dfrac{13}{d}\right)^2\right) \end{pmatrix} = \frac{1}{27}\begin{pmatrix} 29 & \dfrac{53}{3}\sqrt{\dfrac{2}{3}} \\ \dfrac{53}{3}\sqrt{\dfrac{2}{3}} & \dfrac{730}{27} \end{pmatrix}.$$

If we consider the second components of the trimmed LSE $\mathbf{L}_n^\nu$ and of the trimmed average $\mathbf{T}_n^\nu$ of regression quantiles [proposed by Koenker and Portnoy (1987)] as the respective estimators of the slope parameter $\beta_2$, then their relative asymptotic efficiency is

$$\begin{aligned}\frac{\sigma_{L22}}{\sigma_{T22}} &= \frac{h_1}{h_1 + h_2} + \frac{h_2}{h_1 + h_2}\frac{10}{9}\frac{729}{730} \\ &= 1 + \frac{h_2}{h_1 + h_2}\frac{72}{657} \in (1, 1 + 0.11(1 - 2\alpha)],\end{aligned}$$

since $0 < h_1 < \infty$, $0 < h_2 < ((1 - 2\alpha)/2\alpha)h_1$. Thus, $T_{n2}^\nu$ is more efficient estimator of $\beta_2$ than $L_{n2}^\nu$ but, on the other hand, $\det(\mathbf{\Sigma}_L - \mathbf{\Sigma}_T) < 0$ and hence $\mathbf{\Sigma}_L - \mathbf{\Sigma}_T$ is not positive semidefinite. □

## APPENDIX

Let us first note that the proofs of Theorem 1 and 2 use only Lemma 1–3; the proof of Theorem 3 uses Lemma 4–6 and Theorem 1. More detailed proofs of Lemmas 1 and 3 may be found in Gutenbrunner (1986).

Proof of Lemma 1. We shall first prove (5.21). The derivative of $\mathbf{G}_n^d$ at $\mathbf{t} \in \mathbb{R}^p$ equals

$$\nabla \mathbf{G}_n^d(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^n \sigma_{ni}^{-1}\mathbf{d}_{ni}\mathbf{x}_{ni}^T f(\mathbf{x}_{ni}^T\mathbf{t}/\sigma_{ni}) \tag{L1.1}$$

and specifically for $\mathbf{t} = \boldsymbol{\delta}(\alpha) = F^{-1}(\alpha)\boldsymbol{\gamma}$,

$$\nabla \mathbf{G}_n^d(\boldsymbol{\delta}(\alpha)) = f(F^{-1}(\alpha))\mathbf{D}_{nd}, \tag{L1.2}$$

hence

$$\begin{aligned}&\left\|\mathbf{G}_n^d(\boldsymbol{\delta}(\alpha) + n^{-1/2}\mathbf{t}) - \mathbf{G}_n^d(\boldsymbol{\delta}(\alpha)) - n^{-1/2}f(F^{-1}(\alpha))\mathbf{D}_{nd}\mathbf{t}\right\| \\ &\quad = \left\| \frac{1}{n}\sum_{i=1}^n \mathbf{d}_{ni}\frac{1}{n^{1/2}}\left(\frac{\mathbf{x}_{ni}^T\mathbf{t}}{\sigma_{ni}}\right) \right. \\ &\qquad \left. \times \int_0^1 \left[ f\left(F^{-1}(\alpha) + \frac{1}{n^{1/2}}\left(\frac{\mathbf{x}_{ni}^T\mathbf{t}}{\sigma_{ni}}\right)s\right) - f(F^{-1}(\alpha))\right] ds \right\| \\ &\quad = o(n^{-1/2})\end{aligned} \tag{L1.3}$$

uniformly in $\mathbf{t}$ and $\alpha$, $\|\mathbf{t}\| \le K$ and $\varepsilon \le \alpha \le 1 - \varepsilon$, as follows from (B.2), (C.1) and the uniform continuity of $f$ on compact sets. This proves (5.21).

(5.22) is proved using a slight modification of Lemmas A.5 and A.6 of Koul (1969) [see Gutenbrunner (1986) for details]. □

PROOF OF LEMMA 2. The lemma is a consequence of Theorem 2.1 of Shorack (1979). □

PROOF OF LEMMA 3. The proof follows the lines of the proof of Lemma 5.2 in Jurečková (1977); we use Lemma 1.2 and the monotonicity of the function $\mathbf{e}^T\hat{\mathbf{G}}_n^x(\mathbf{z} + t\mathbf{e})$ in $t \in \mathbb{R}^1$ with fixed $\mathbf{e}, \mathbf{z} \in \mathbb{R}^p$. Moreover, we utilize the inequality

$$\begin{aligned}\left\|\hat{\mathbf{G}}_n^x\big(\hat{\boldsymbol{\delta}}_n(\alpha)\big) - \mathbf{G}_n^x(\boldsymbol{\delta}(\alpha))\right\| &= \left\|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_{ni} I\left[Y_{ni} = \mathbf{x}_{ni}^T\hat{\boldsymbol{\beta}}_n(\alpha)\right]\right\| \\ &\le pn^{-1}\|\mathbf{X}_n\|_\infty = o(n^{-1/2})\end{aligned}$$

with $\hat{\boldsymbol{\delta}}_n(\alpha) = \boldsymbol{\delta}_n(\alpha) + n^{-1/2}\|\mathbf{Z}_n(\alpha)\|(\mathbf{Z}_n(\alpha)/\|\mathbf{Z}_n(\alpha)\|)$ and the fact that $f(F^{-1}(\alpha))\mathbf{e}^T\mathbf{D}_n\mathbf{e}$ is bounded away from 0 for $\|\mathbf{e}\| = 1$, $\alpha \in [\varepsilon, 1 - \varepsilon]$ and $n \ge n_0$. □

LEMMA 4. *Denote*

$$I_n(K, \varepsilon) = I\left[\|\mathbf{Z}_n\|_{(\varepsilon)} \le K\right], \qquad K > 0, \qquad 0 < \varepsilon < 1. \tag{L4.1}$$

*Then, under the conditions* (A.1), (B.1) *and* (C.1),

$$\sup_{\varepsilon \le \alpha \le 1-\varepsilon} E\left\{I_n(K, \varepsilon)\left\|\hat{\mathbf{W}}_n^d(\alpha) - \mathbf{W}_n^d(\alpha)\right\|\right\} = O\left(\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_{ni}\|\,\|\mathbf{d}_{ni}\|\right) \tag{L4.2}$$

and

$$\begin{aligned}&\sup_{\varepsilon \le \alpha \le 1-\varepsilon} E\left\{I_n(K, \varepsilon)\left\|\sum_{i=1}^n \mathbf{d}_{ni}\left[\hat{U}_{ni}(\alpha) - \left(U_i \vee F^{-1}(\alpha)\right)\right]\right\|\right\} \\ &\qquad = O\left(\frac{1}{n}\sum_{i=1}^n \left(\|\mathbf{x}_{ni}\|^2\|\mathbf{d}_{ni}\|\right)\right)\end{aligned} \tag{L4.3}$$

*for any fixed* $K > 0$ *and* $\varepsilon > 0$.

PROOF. $\hat{a}_{ni}(\alpha)$ of (3.9) could be rewritten for $\varepsilon \le \alpha \le 1 - \varepsilon$ as

$$\hat{a}_{ni}(\alpha) = \begin{cases} 1, & \text{if } U_i > F^{-1}(\alpha) + n^{-1/2}\mathbf{x}_{ni}^T\mathbf{Z}_n(\alpha)/\sigma_{ni}, \\ 0, & \text{if } U_i < F^{-1}(\alpha) + n^{-1/2}\mathbf{x}_{ni}^T\mathbf{Z}_n(\alpha)/\sigma_{ni}. \end{cases}$$

Then, using (6.15), we get

$$\begin{aligned}
&E\Big(I_n(K,\varepsilon)\big|\hat{U}_{ni}(\alpha) - \big(U_i \vee F^{-1}(\alpha)\big)\big|\Big)\\
&\quad\le E\big(I_n(K,\varepsilon)|\hat{a}_{ni}(\alpha) - a_i^*(\alpha)||U_i - F^{-1}(\alpha)|\big)\\
&\quad\le E\Big(I_n(K,\varepsilon)I\big[|U_i - F^{-1}(\alpha)| \le n^{-1/2}\|\mathbf{x}_{ni}\|(K/c)\big]|U_i - F^{-1}(\alpha)|\Big)\\
&\quad\le K^* n^{-1}\|\mathbf{x}_{ni}\|^2,
\end{aligned}\tag{L4.4}$$

hence, for $\alpha \in [\varepsilon, 1-\varepsilon]$,

$$E\left(I_n(K,\varepsilon)\left\|\sum_{i=1}^n \mathbf{d}_{ni}\big(\hat{U}_{ni}(\alpha) - \big(U_i \vee F^{-1}(\alpha)\big)\big)\right\|\right) \le K^* n^{-1}\sum_{i=1}^n \|\mathbf{x}_{ni}\|^2\|\mathbf{d}_{ni}\|.$$

The proof of (L4.2) is analogous. □

LEMMA 5. *Under* (A.1), (B.1–3), (C.1–2) *and* (W.2–3),

$$\int \hat{\mathbf{W}}_n^d\,db = O_p\left(\left(\frac{1}{n}\sum_{i=1}^n \|\mathbf{d}_{ni}\|^2\right)^{1/2}\right)\tag{L5.1}$$

*and*

$$\begin{aligned}
&\sum_{i=1}^n \mathbf{d}_{ni}\int\big[\hat{U}_{ni}(\alpha) - \big(U_{ni} \vee F^{-1}(\alpha)\big)\big]\,dJ(\alpha)\\
&\qquad = O_p\left(\left(\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_{ni}\|^2\|\mathbf{d}_{ni}\|\right)^{1/2}\right).
\end{aligned}\tag{L5.2}$$

REMARK. Notice that (L5.1) is not a consequence of Theorem 1 because it does not impose condition (D.1) on $\{\mathbf{d}_{ni}, i = 1,\dots,n\}$.

PROOF. First, notice that

$$\int \mathbf{W}_n^b\,db = \frac{1}{n^{1/2}}\sum_{i=1}^n \mathbf{d}_{ni}\big[b(F(U_i)) - \bar{b}\big] = O_p\left[\left(\frac{1}{n}\sum_{i=1}^n \|\mathbf{d}_{ni}\|^2\right)^{1/2}\right].\tag{L5.3}$$

Moreover, (L4.1) together with condition (B.1) implies

$$\begin{aligned}
&E\left\{I_n(K,\varepsilon)\left\|\int\big(\hat{\mathbf{W}}_n^d - \mathbf{W}_n^d\big)\,db\right\|\right\}\\
&\qquad\le E\left\{I_n(K,\varepsilon)\int\big\|\hat{\mathbf{W}}_n^d - \mathbf{W}_n^d\big\|\,|db|\right\}\\
&\qquad\le \|b\|_{TV}\sup_{\varepsilon\le\alpha\le 1-\varepsilon} E\big(I_n(K,\varepsilon)\big\|\hat{\mathbf{W}}_n^d - \mathbf{W}_n^d\big\|\big)\\
&\qquad = O\left(\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_{ni}\|\,\|\mathbf{d}_{ni}\|\right) = O\left(\left(\frac{1}{n}\sum_{i=1}^n \|\mathbf{d}_{ni}\|^2\right)^{1/2}\right),
\end{aligned}\tag{L5.4}$$

where $\int \cdot |db|$ denotes the integration with respect to the total variation of $b$ and $\|b\|_{TV}$ denotes the total variation norm of $b$.

By Lemma 3, given $\varepsilon, \eta > 0$, there exists $K > 0$ so that $P(I_n(K, \varepsilon) = 1) > 1 - \eta$ for all $n$; hence if for all $K$, $E(I_n(K, \varepsilon)A_n) = O(1)$, then $A_n = O_p(1)$. Regarding that, (L5.3) and (L5.4) lead to (L5.1). Analogously, (L4.3) leads to (L5.2). □

LEMMA 6. *Let $\nu(A) = \int_A J(u)\,du$ be a positive measure. Then, under the conditions of Theorem* 3,

$$P\left\{\mathbf{X}_n^T \operatorname{diag}\left(\hat{J}_{ni}\right)\mathbf{X}_n \ singular\right\} \to 0 \quad as\ n \to \infty \tag{L6.1}$$

*and, for* $\hat{\mathbf{Q}}_n = (n\bar{J})^{-1}\mathbf{X}_n^T\hat{\mathbf{J}}_n\mathbf{X}_n$ *of* (6.10),

$$\hat{\mathbf{Q}}_n^{-1} = \mathbf{Q}_n^{-1} + O_p(n^{-1/2}). \tag{L6.2}$$

PROOF. We may take $\nu$ as a probability measure, $\nu(0,1) = \bar{J} = 1$. Put

$$\hat{\mathbf{Q}}_n = \frac{1}{n}\mathbf{X}_n^T \operatorname{diag}\left(\hat{J}_{ni}\right)\mathbf{X}_n = \frac{1}{n}\sum_{i=1}^n \hat{J}_{ni}\mathbf{x}_{ni}\mathbf{x}_{ni}^T. \tag{L6.3}$$

Then $\hat{\mathbf{Q}}_n = \tilde{\mathbf{Q}}_n$ if and only if $\tilde{\mathbf{Q}}_n$ is nonsingular, $\hat{\mathbf{Q}}_n = \mathbf{Q}_n$ otherwise. Considering $\mathbf{x}_{ni}\mathbf{x}_{ni}^T = \mathbf{d}_{ni}$ as a vector in $\mathbb{R}^{p^2}$, we see that $\tilde{\mathbf{Q}}_n$ is of the type $\hat{\mathbf{V}}_n^{J,d}$. Then, using (L5.1) with

$$\bar{J} = \nu(0,1) = 1, \bar{\mathbf{d}}_n = \mathbf{Q}_n, \frac{1}{n}\sum_{i=1}^n \|\mathbf{d}_{ni}\|^2 = \frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_{ni}\|^4,$$

and regarding (B.4), we get

$$\tilde{\mathbf{Q}}_n - \mathbf{Q}_n = n^{-1/2}\left(O_p\left(\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_{ni}\|^4\right)^{1/2}\right) = O_p(n^{-1/2}).$$

Since $\mathbf{Q} = \lim \mathbf{Q}_n$ is nonsingular, we get (L6.1) and, moreover,

$$\hat{\mathbf{Q}}_n = \tilde{\mathbf{Q}}_n + o_p(0) = \mathbf{Q}_n + O_p(n^{-1/2}).$$

Denote

$$\mathbf{B}_n = \left(\hat{\mathbf{Q}}_n - \mathbf{Q}_n\right)I\left[\left\|\hat{\mathbf{Q}}_n - \mathbf{Q}_n\right\| \le \tfrac{1}{2}\left\|\mathbf{Q}_n^{-1}\right\|^{-1}\right];$$

then $\|\mathbf{B}_n\| = O_p(n^{-1/2})$ and $\hat{\mathbf{Q}}_n = \mathbf{Q}_n + \mathbf{B}_n + o_p(0)$. Regarding (L6.4), (L6.2) follows from the Taylor expansion

$$\left\|(\mathbf{Q}_n + \mathbf{B}_n)^{-1} - \mathbf{Q}_n^{-1} + \mathbf{Q}_n^{-1}\mathbf{B}_n\mathbf{Q}_n^{-1}\right\| \le 2\|\mathbf{B}_n\|^2\left\|\mathbf{Q}_n^{-1}\right\|^3$$

which implies $\hat{\mathbf{Q}}_n^{-1} - \mathbf{Q}_n^{-1} = \mathbf{Q}_n^{-1}(\mathbf{Q}_n - \hat{\mathbf{Q}}_n)\mathbf{Q}_n^{-1} + O_p(n^{-1})$. □

**Acknowledgments.** The authors wish to thank Professor H. Witting for initiating their cooperation and helpful discussions. They also highly appreciate the discussions with Professor R. Koenker and thank him for his interest in the subject. The authors gratefully acknowledge the constructive comments of the referees and of an Associate Editor.

## REFERENCES

ADICHIE, J. N. (1978). Rank tests for sub-hypotheses in the general linear regression. *Ann. Statist.* **6** 1012–1026.

ANTOCH, J. and JUREČKOVÁ, J. (1985). Trimmed LSE resistant to leverage points. *Computational Statistics Quarterly* **4** 329–339.

BASSETT, G. and KOENKER, R. (1982). An empirical quantile function for linear models with iid errors. *J. Amer. Statist. Assoc.* **77** 407–415.

BICKEL, P. J. (1967). Some contributions to the theory of order statistics. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 575–592. Univ. California Press, Berkeley.

BICKEL, P. J. (1973). On some analogues to linear combinations of order statistics in the linear model. *Ann. Statist.* **1** 597–616.

CHIANG, C. Y. and PURI, M. L. (1984). Rank-order tests for the parallelism of several regression surfaces. *J. Statist. Plann. Inference* **10** 43–57.

DE JONGH, P. J., DE WET, T. and WELSH, A. H. (1988). Mallows type bounded-influence-regression trimmed means. *J. Amer. Statist. Assoc.* **83** 805–810.

FAREBROTHER, R. W. (1985). Unbiased $L_1$ and $L$ estimation. *Comm. Statist. Theory Methods* **14** 1941–1962.

GUTENBRUNNER, C. (1986). Zur Asymptotik von Regressions quantil prozessen und da raus abgeleiten Statistiken. Ph.D. dissertation, Univ. Freiburg.

GUTENBRUNNER, C., JUREČKOVÁ, J., KOENKER, R. and PORTNOY, S. (1990). Tests of linear hypotheses based on regression rank scores. Unpublished manuscript.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academia, Prague.

HEILER, S. and WILLERS, R. (1988). Asymptotic normality of $R$-estimates in the linear model. *Statistics* **19** 173–184.

JUREČKOVÁ, J. (1977). Asymptotic relations of $M$-estimates and $R$-estimates in linear regression model. *Ann. Statist.* **5** 464–472.

JUREČKOVÁ, J. (1980). Asymptotic representation of $M$-estimates of location. *Mathematische Operationsforschung und Statistik* **11** 61–73.

JUREČKOVÁ, J. (1983a). Robust estimators of location and regression parameter and their second order asymptotic relations. In *Transactions of the Ninth Prague Conference on Information Theory, Random Processes, and Statistical Decision Functions* (J. Á. Víšek, ed.) 19–21. Reidel, Dordrecht.

JUREČKOVÁ, J. (1983b). Winsorized least-squares estimator and its $M$-estimator counterpart. In *Contributions to Statistics: Essays in Honour of Norman Lloyd Johnson* (P. K. Sen, ed.) 237–245. North-Holland, Amsterdam.

JUREČKOVÁ, J. (1984). Regression quantiles and trimmed least squares estimator under a general design. *Kybernetika* **20** 345–357.

JUREČKOVÁ, J. (1985). Robust estimators of location and their second order asymptotic relations. In *A Celebration of Statistics: The ISI Centenary Volume* (A. C. Atkinson and S. E. Fienberg, eds.) 377–392. Springer, New York.

JUREČKOVÁ, J. and SEN, P. K. (1989). Uniform second order asymptotic linearity of $M$-statistics in linear model. *Statist. Decisions* **7** 263–276.

KOENKER, R. and BASSETT, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

KOENKER, R. and BASSETT, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50** 43–61.

KOENKER, R. and D'OREY, V. (1989). A note on computing dual regression quantiles and regression rank scores. Remark on algorithm 229. Preprint.

KOENKER, R. and PORTNOY, S. (1987). $L$-estimation for linear models. *J. Amer. Statist. Assoc.* **82** 851–857.

KOUL, H. L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Ann. Math. Statist.* **40** 1950–1979.

PORTNOY, S. (1984). Tightness of the sequence of empiric c.d.f. processes defined from regression fractiles. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 231–246. Springer, New York.

RUPPERT, D. and CARROLL, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838.

SEN, P. K. (1969). On a class of rank order tests for the parallelism of several regression lines. *Ann. Math. Statist.* **40** 1668–1683.

SHORACK, G. R. (1979). The weighted empirical process of row independent random variables with arbitrary distribution functions. *Statist. Neerlandica* **33** 169–189.

WELSH, A. H. (1987). Trimmed mean in the linear model. *Ann. Statist.* **15** 20–45.

WITTING, H. (1974). *Mathematische Statistik*. Teubner, Stuttgart.

UNIVERSITÄTKLINIK FÜR KINDER
UND JUGENDPSYCHIATRIE
DER PHILIPPS UNIVERSITÄT
HANS-SACHS-STR. 6
D-3550 MARBURG
GERMANY

DEPARTMENT OF PROBABILITY AND STATISTICS
CHARLES UNIVERSITY
SOKOLOVSKÁ 83
18600 PRAGUE 8
CZECHOSLOVAKIA