ASYMPTOTIC EQUIVALENCE OF REGRESSION RANK SCORES ESTIMATORS AND R–ESTIMATORS IN LINEAR MODELS

JANA JUREČKOVÁ AND PRANAB KUMAR SEN

Charles University, Prague, Czechoslovakia,

and University of North Carolina at Chapel Hill, USA

The classical R–estimators in linear models are computationally more cumbersome than the regression rank scores estimators. Under appropriate regularity conditions, both the methods are shown to be asymptotically equivalent. A coordinatewise modification of regression rank scores estimators is also considered in this setup.

1. INTRODUCTION

Consider the usual regression model:

$$\mathbf{Y} = (Y_1,\dots,Y_n)' = \mathbf{X}^o_{(n)}\boldsymbol{\beta}^o + \mathbf{e}; \quad \mathbf{e} = (e_1,\dots,e_n)', \tag{1.1}$$

where

$$\mathbf{X}^o_{(n)} = (\underset{n\times 1}{\mathbf{1}_n}, \underset{n\times p}{\mathbf{X}^{(1)}_{(n)}}, \underset{n\times q}{\mathbf{X}^{(2)}_n}) = (\mathbf{1}_n, \mathbf{X}_{(n)}), \tag{1.2}$$

$\mathbf{1}'_n = (1,\dots,1) \in \mathbb{R}_n$, the $\mathbf{X}^{(j)}_{(n)}$ are matrices of known regression constants, j=1,2,

$$\boldsymbol{\beta}^o = (\underset{1\times 1}{\beta_0}, \underset{1\times p}{\boldsymbol{\beta}^{(1)\prime}}, \underset{1\times q}{\boldsymbol{\beta}^{(2)\prime}})' = (\beta_0, \boldsymbol{\beta})' \tag{1.3}$$

AMS Subject Classification Nos: 62G05, 62J05

Key Words and Phrases: Coordinatewise estimator; linear programming; multi–sample model; regression quantiles; uniform asymptotic linearity.

is a vector of unknown parameters, and the e_i are independent and identically distributed (i.i.d.) random variables (r.v.) having a continuous distribution function (d.f.) F, defined on $\mathbb{R}_1$. The model in (1.1) can be rewritten as

$$\underset{\sim}{Y} = \beta_0 \underset{\sim}{1}_n + \underset{\sim}{X}_{(n)} \underset{\sim}{\beta} + \underset{\sim}{e} = \beta_0 \underset{\sim}{1}_n + \underset{\sim}{X}_{(n)}^{(1)} \underset{\sim}{\beta}^{(1)} + \underset{\sim}{X}_{(n)}^{(2)} \underset{\sim}{\beta}^{(2)} + \underset{\sim}{e}. \tag{1.4}$$

Our goal is to construct a rank–based estimator of $\underset{\sim}{\beta}^{(2)}$, treating β_0, $\underset{\sim}{\beta}^{(1)}$ as nuisance parameters.

Motivated by the work of Adichie (1967) (for p = 0, q = 1), Jurečková (1971) and Koul (1970) constructed different versions of rank (R–) estimators of $\underset{\sim}{\beta}$, while Jaeckel (1972) proposed another version of R–estimators minimizing a specific measure of rank dispersion; all these methods are asymptotically equivalent. Heiler and Willers (1988), using a convexity argument, eliminated one of the regularity conditions on $\underset{\sim}{X}_{(n)}$. Note that the ranks are translation–invariant, and hence, to estimate β_0, one may need to use signed rank statistics. Moreover, in this approach, to estimate $\underset{\sim}{\beta}^{(2)}$, one needs to estimate $\underset{\sim}{\beta}$, so that one can not avoid estimating $\underset{\sim}{\beta}^{(1)}$ simultaneously even when the goal is to estimate $\underset{\sim}{\beta}^{(2)}$ alone. Simultaneous estimation of $\underset{\sim}{\beta}^{(1)}$, $\underset{\sim}{\beta}^{(2)}$ (for $p \geq 1$) generally entails computational complexities of serious concern, and this is one of the reasons why R–estimators, in spite of having global robustness properties, have not met the light of applications (for larger values of p+q).

Based on the concept of regression rank scores developed in Gutenbrunner (1986) and Gutenbrunner and Jurečková (1992), an estimator of $\underset{\sim}{\beta}^{(2)}$ has been proposed by Jurečková (1991). The regression rank scores (being an extension of ordinary rank scores to linear models) are regression–invariant (compared to translation–invariance of the usual ranks). This makes the corresponding estimator computationally simpler than the usual R–estimators.

Our primary task is to establish an asymptotic equivalence of regression rank scores estimators and R–estimators when they involve the common score function. This enables us to examine the relative computational complexities without a compromise of their asymptotic efficiency aspects. This also leads us to an idea of estimating $\underset{\sim}{\beta}^{(2)}$ componentwise (based on regression rank scores) without distorting the asymptotic equivalence results.

Besides the computational aspects, we have a diagnostic point of view too. Avoiding direct esimation of $\beta^{(1)}$ may also avoid eventual leverage on influential points in $X_{(n)}^{(1)}$. This feature is shared by the subhypothesis testing problem (H_0 : $\beta^{(2)} = 0$ vs. $\beta^{(2)} \neq 0$), as has been treated in Gutenbrunner et al. (1992).

Jaeckel's estimator, regression rank scores estimators and Jurečková's estimators are introduced in Section 2. Their asymptotic equivalence, as further extended to a coordinatewise estimator, is established in Section 3. The k–sample model is treated as an illustration in Section 4.

2. ALTERNATIVE RANK BASED ESTIMATORS

We write $X'_{(n)} = (x_1,\ldots,x_n)$. For $b \in \mathbb{R}_{p+q}$, let $R_{ni}(b)$ be the rank of $Y_i - x_i'b$ among $Y_1 - x_1'b,\ldots,Y_n - x_n'b$, for $i = 1,\ldots,n$. Let $A_n(i)$, $i = 1,\ldots,n$, be monotone scores derived from a nondecreasing, square–integrable score function $\varphi : [0,1] \to \mathbb{R}_1$, in either of the following manner:

$$A_n(i) = E\, \varphi(U_{n:i}) \text{ or } \varphi\left[\frac{i}{n+1}\right] \text{ or } n\int_{(i-1)/n}^{i/n} \varphi(t)dt, \quad i = 1,\ldots,n, \tag{2.1}$$

where $U_{n:1} \leq \ldots \leq U_{n:n}$ are the order statistics of a sample of size n from the uniform (0,1)d.f. Let $S_n(b) = (S_{n1}(b),\ldots,S_{np+q}(b))'$ be a vector of linear rank statistics of the form

$$S_n(b) = \Sigma_{i=1}^n (x_i - \bar{x}_n) A_n(R_{ni}(b)),\ b \in \mathbb{R}_{p+q}, \tag{2.2}$$

where $\bar{x}_n = n^{-1} \Sigma_{i=1}^n x_i = (\bar{x}_{n1},\ldots,\bar{x}_{np+q})'$. Jurečková (1969, 1971a) established the 'uniform asymptotic linearity of $S_n(b)$ in b' in an $n^{-\frac{1}{2}}$– neighborhood of β, and incorporated the same in the estimation of β based on a L_1–norm on $S_n(b)$. Specifically, the estimator is

$$\hat{\beta}_n = \text{Arg. min}\left\{\Sigma_{j=1}^{p+q} \mid S_{nj}(b) \mid : b \in \mathbb{R}_{p+q}\right\}. \tag{2.3}$$

The related asymptotic distribution theory was developed under appropriate regularity conditions on F, φ and $X_{(n)}$. Heiler and Willers (1988) weakened a concordence–discordance condition on $X_{(n)}$ and derived the same asymptotic linearity results via a convexity argument, so that the asymptotic multinormality result for $\hat{\beta}_n$ holds even under this weakened condition. Jackel (1972) proposed to estimate β by minimizing the dispersion function:

$$D_n(b) = \Sigma_{i=1}^n \left\{A_n(R_{ni}(b)) - \overline{A}_n\right\} (Y_i - x_i' b),\ b \in \mathbb{R}_{p+q}, \tag{2.4}$$

which is continuous, piecewise linear and convex in b; here $\overline{A}_n = n^{-1} \Sigma_{i=1}^n A_n(i)$. The vector of derivatives of $D_n(b)$ with respect to the components of b coincides with $(-1)S_n(b)$ whenever it exists. Hence, under the regularity conditions of Heiler and Willers (1988), the minimization of (2.3) and (2.4) are equivalent. As such, denoting $\hat{\beta}_n = (\hat{\beta}_n^{(1)}, \hat{\beta}_n^{(2)})$, we may use the component $\hat{\beta}_n^{(2)}$ for estimating $\beta^{(2)}$. However, this involves the simultaneous estimation of $\beta^{(1)}$ and $\beta^{(2)}$, and as no algebraic expression is generally available for $\hat{\beta}$, an iterative procedure may have to be employed. The computational scheme becomes quite cumbersome, especially when p+q is not small.

Jurečková (1991) proposed another estimator of $\beta^{(2)}$ based on regression rank scores. To pose this estimator, we consider the full model $Y = X_{(n)}^o \beta^o + e$ where $X_{(n)}^o = (1_n, X_{(n)})$ and $\beta^o = (\beta_0, \beta')'$, so that $\beta^o \in \mathbb{R}_{p+q+1}$. For $\alpha : 0 < \alpha < 1$, let $\hat{\beta}_n^o(\alpha)$ denote the α–regression quantile, proposed by Koenker and Bassett (1978), which can be expressed as a solution of the minimization of

$$\Sigma_{i=1}^n \rho_\alpha(Y_i - x_i^{o\prime} b),\ b \in \mathbb{R}_{p+q+1}, \tag{2.5}$$

where $x_i^{o\prime} = (1, x_i')$, i = 1,...,n, and

$$\rho_\alpha(x) = |x|\ \{\alpha\, I(x > 0) + (1-\alpha)\, I(x < 0)\},\ x \in \mathbb{R}_1. \tag{2.6}$$

Koenker and Bassett also characterized $\hat{\beta}_n^o(\alpha)$ as the component of β^o in the optimal solution (β^o, r^+, r^-) of the linear programme

$$\begin{gathered} \alpha\, 1_n' r^+ + (1-\alpha) 1_n' r^- := \min, \\ X_{(n)}^o\, \beta^o + r^+ - r^- = Y \\ (\beta^o, r^+, r^-) \in \mathbb{R}_{p+q+1} \times \mathbb{R}_n^+ \times \mathbb{R}_n^+,\ 0 < \alpha < 1. \end{gathered} \tag{2.7}$$

Such regression quantiles provide a basis for L–estimation in linear models and are themselves extensions of sample quantiles to linear models. We may refer to the work of Ruppert and Carroll (1980), Jurečková (1984), Koenker and Portnoy (1987), Gutenbreunner and Jurečková (1992), among others. The dual programme to (2.7) can be written in the following manner:

$$\underset{\sim}{Y}'\, \hat{\underset{\sim}{a}}(\alpha) := \max, \quad \underset{\sim}{X}^{o\prime}_{(n)}\, \hat{\underset{\sim}{a}}(\alpha) = (1-\alpha)\underset{\sim}{X}^{o\prime}_{(n)}\, \underset{\sim}{1}_n;$$

$$\hat{\underset{\sim}{a}}(\alpha) \in [0,1]^n,\ 0 < \alpha < 1. \tag{2.8}$$

In the location model [i.e., $Y_i = \beta_0 + e_i$, $i \geq 1$], the optimal solution $\hat{\underset{\sim}{a}}_n(\alpha) = (\hat{a}_{n1}(\alpha),\ldots,\hat{a}_{nn}(\alpha))'$ coincides with the vector of rank scores $(a_n^*(R_i,\alpha),\ i=1,\ldots,n)$, where

$$a_n^*(R_i,\alpha) = \begin{cases} 1, & \text{if } 0 \leq \alpha < (R_i-1)/n, \\ a_n^*(R_i,\alpha) = R_i - n\alpha, & \text{if } (R_i-1)/n \leq \alpha \leq R_i/n, \\ 0, & \text{if } R_i/n < \alpha \leq 1, \end{cases} \tag{2.9}$$

and R_i is the rank of Y_i among $Y_1,\ldots,Y_n$, for $i = 1,\ldots,n$ [viz., Hájek and Šidák (1967, Ch. V)]. For this reason, for the linear model too, we find it convenient to call the $\hat{a}_{ni}(\alpha)$, $1 \leq i \leq n$, for $0 < \alpha < 1$, the regression rank scores. We may refer to Gutenbrunner and Jurečková (1992) for a detailed study of various properties of such regression rank scores and for other related references.

An important property of regression rank scores is their invariance with respect to the $\underset{\sim}{X}^o_{(n)}$– regression, i.e.,

$$\hat{\underset{\sim}{a}}_n(\alpha, \underset{\sim}{Y} + \underset{\sim}{X}^o_{(n)}\, \underset{\sim}{b}) = \hat{\underset{\sim}{a}}_n(\alpha,\underset{\sim}{Y}), \quad \forall\, \underset{\sim}{b} \in \mathbb{R}_{p+q+1}, \tag{2.10}$$

and this plays a basic role in the developments to follow. Moreover, $\hat{a}_{ni}(\alpha)$ is a continuous function of bounded variation and is piecewise linear on [0,1], for every i (=1,...,n). Relations of regression quantiles and regression rank scores, following from the duality of linear programmes, extend the duality of order statistics and ranks from the location to the linear model. As such, statistical inference based on regression rank scores is typically based on *linear regression rank scores statistics*, which may be defined as follows:

Take a nondecreasing and square integrable score generating function

$\varphi : (0,1) \to \mathbb{R}_1$ and define the *scores* as

$$\hat{b}_{ni} = -\int_0^1 \varphi(\alpha) d\hat{a}_{ni}(\alpha), \; i = 1,\ldots,n, \tag{2.11}$$

and the linear regression rank scores statistic as

$$Z_n = \Sigma_{i=1}^n c_{ni} \hat{b}_{ni}, \tag{2.12}$$

where $c_{n1},\ldots,c_{nn}$ are given coefficients.

Gutenbrunner et al. (1992) constructed a class of tests of the hypothesis $H_0 : \underset{\sim}{\beta}^{(2)} = \underset{\sim}{0}$ (vs. $\underset{\sim}{\beta}^{(2)} \neq \underset{\sim}{0}$) based on quadratic forms of a vector of statistics of the type (2.12). Such tests do not explicitly involve the estimation of the nuisance parameters β_0, $\underset{\sim}{\beta}^{(1)}$, whereas the usual aligned rank tests, studied in detail by Sen and Puri (1977) and Adichie (1978,1984), among others, rest on the elimination of nuisance parameters through R–estimation.

For $\underset{\sim}{t} \in \mathbb{R}_q$, consider the pseudo–observations $Y_i - \underset{\sim}{x}_i^{(2)'} \underset{\sim}{t}$, $1 \le i \le n$, where $\underset{\sim}{x}_i^{(2)'}$ is the ith row of $\underset{\sim}{X}_{(n)}^{(2)}$, $1 \le i \le n$. Let $\hat{a}_{ni}(\alpha, \underset{\sim}{Y} - \underset{\sim}{X}_{(n)}^{(2)} \underset{\sim}{t})$, $i = 1,\ldots,n$, be the regression rank scores calculated from (2.8) with the Y_i replaced by $Y_i - \underset{\sim}{x}_i^{(2)'} \underset{\sim}{t}$ and $\underset{\sim}{X}_{(n)}^{o}$ replaced by $\underset{\sim}{X}_{(n)}^{o(1)} = (\underset{\sim}{1}_n, \underset{\sim}{X}_n^{(1)})$ (so that $\underset{\sim}{\beta}^{o(1)} = (\beta_0, \underset{\sim}{\beta}^{(1)'})' \in \mathbb{R}_{p+1}$. Jurečková (1991) proposed to estimate $\underset{\sim}{\beta}^{(2)}$ by minimizing $D_n(\underset{\sim}{t})$ with respect to $\underset{\sim}{t} \in \mathbb{R}_q$, where

$$D_n(\underset{\sim}{t}) = \Sigma_{i=1}^n (Y_i - \underset{\sim}{t}' \underset{\sim}{x}_i^{(2)})[b_{ni}(\underset{\sim}{Y} - \underset{\sim}{X}_{(n)}^{(2)} \underset{\sim}{t}) - \overline{\varphi}], \tag{2.13}$$

where

$$\overline{\varphi} = \int_0^1 \varphi(\alpha) d\alpha. \tag{2.14}$$

Thus, we have

$$\tilde{\underset{\sim}{\beta}}_n^{(2)} = \text{Arg. min}\{D_n(\underset{\sim}{t}) : \underset{\sim}{t} \in \mathbb{R}_q\}. \tag{2.15}$$

The function $D_n(\underset{\sim}{t})$ corresponds to Jaeckel's function in (2.4) (albeit with possibly different score functions), and hence, is continuous, piece–wise linear and convex in $\underset{\sim}{t} \in \mathbb{R}_q$. It is differentiable (with respect to $\underset{\sim}{t}$) a.e., and the vector of its derivatives coincides, whenever it exists, with a special vector of linear regression rank scores statistics. Jurečková (1992) showed that the latter statistics are 'uniformly asymptotically linear in $\underset{\sim}{t}$', and this, in turn, enables us to derive the asymptotic distribution of $n^{\frac{1}{2}}(\tilde{\beta}_n^{(2)} - \beta^{(2)})$.

It may be remarked that the linear regression rank scores statistics in (2.12) involve the scores $\hat{b}_{ni}$, which, in turn, depend on φ and the $\hat{a}_{ni}(\alpha)$. On the other hand, for every $\alpha(0 < \alpha < 1)$, the scores $\hat{a}_{ni}(\alpha)$ as obtained from (2.8) involves (i) $Y - X_n^{(2)} t$ instead of Y and (ii) $X_{(n)}^{(1)o}$ instead of $X_{(n)}^o$. Since (i) involves a linear function of t and (ii) involves a subset of $X_{(n)}^o$, computational complications are less than in (2.8) (when $q > 1$). Note also that $D_n(t)$ and hence, $\tilde{\beta}_n^{(2)}$ are invariant to the $X_{(n)}^{(1)o}$–regression. Further, if in (2.13) we work with $X_{(n)}^{(2)} = X_n$ (i.e., $p = 0$), then we have the Jackel estimation of β, so that it involves comparatively more involved computations. But in either case, we treat β_o as a nuisance parameter, and its estimation can be accomplished by the use of signed–ranks, as has been done in Jurečková (1971b).

Our main contention is to show that under suitable regularity conditions, $\hat{\beta}_n^{(2)}$ and $\tilde{\beta}_n^{(2)}$ are asymptotically (efficiency–) equivalent, and hence, our choice between them can be based mainly on computational and diagnostic aspects. Moreover, one may also estimate $\beta^{(2)}$ coordinatewise, applying Jurečková's (1991) procedure successively for single components of $\beta^{(2)}$ (so that there are p+q nuisance parameter at each stage). Thus, this will involve q linear programmes, each followed by a minimization of a function $D_n(t)$ but with respect to a single argument, so that computationally this will be simpler. On the other hand, the linear programming algorithm will involve p+q instead of p+1 arguments, and may be somewhat more complicated. These opposing views need to be reconciled in choosing such a coordinatewise approach, which will be shown to be asymptotically equivalent to the procedure outlined in (2.13)–(2.15).

3. COMPARISON OF ESTIMATORS

Consider the model in the form (1.4), in which we want to estimate $\beta^{(2)}$.

Because all estimator under consideration are invariant to the translation, they will involve matrices $(I_n - H)X_{(n)}^{(1)}$, $(I_n - H)X_{(n)}^{(2)}$ where $H = n^{-1}(1_n, \ldots, 1_n)$ is the $(n \times n)$ projection matrix corresponding to 1_n. Hence, for simplicity of notation, we shall assume that

$$\mathbf{1}_n' \mathbf{X}_{(n)}^{(1)} = \mathbf{0}', \mathbf{1}_n' \mathbf{X}_{(n)}^{(2)} = \mathbf{0}'. \tag{3.1}$$

If we want to compare two types of estimators, we must assume that our model satisfies a combined set of regularity conditions imposed in Jaeckel (1972), Heiler and Willers (1988) and Jurečková (1991). As a result, we shall impose the following regularity conditions on F, $\mathbf{X}$ and φ:

(F.1) The d.f. F of errors $e_1,\dots,e_n$ in (1.1) has an absolutely continuous density f which is positive and bounded for $A < x < B$ and monotonously decreasing as $x \to A^+$ or $x \to B^-$ where $-\infty \le A = \sup\{x : F(x) = 0\}$ and $+\infty \ge B = \inf\{x : F(x) = 1\}$.
The derivative f′ of f is bounded a.e.

(F.2) $1/f(F^{-1}(\alpha)) \le c(\alpha(1-\alpha))^{-1-a}$ for $0 < \alpha \le \alpha_o$ and $1-\alpha_o \le \alpha < 1$, where $0 < a < \frac{1}{4} - \epsilon$, $\epsilon > 0$, $c > 0$, $0 < \alpha_o < \frac{1}{2}$.

(F.3) $\left|\frac{f'(x)}{f(x)}\right| \le c\,|x|$ for $|x| \ge K(\ge 0)$, $c > 0$.

(X.1) $\lim_{n\to\infty} \mathbf{Q}_n = \mathbf{Q}$ where $\mathbf{Q}_n = n^{-1} \mathbf{X}_{(n)}' \mathbf{X}_{(n)}$ and $\mathbf{Q}$ is a positive definite (p.d.) $(p+q)\times(p+q)$ matrix.

(X.2) $\max_{1\le i\le n} \|\mathbf{x}_i\| = 0(n^{\tau(a,b,\eta)})$ where

$$\tau(a,b,\eta) = (2(b-a)-\eta)/(1+4b),$$

$\eta > 0$ and $b > 0$ is such that $0 < b-a < \epsilon/2$, $\epsilon > 0$ (hence $0 < b < \frac{1}{4} - \frac{\epsilon}{2}$).

Moreover,

$$n^{-1} \sum_{i=1}^{n} \|\mathbf{x}_i\|^3 = O(1).$$

(X.3) $\lim_{n\to\infty} \mathbf{Q}_n^{(2)} = \mathbf{Q}^{(2)}$ where $\mathbf{Q}^{(2)}$ is p.d. and

$$\mathbf{Q}_n^{(2)} = n^{-1}\left[\mathbf{X}_{(n)}^{(2)} - \hat{\mathbf{X}}_{(n)}^{(2)}\right]' \left[\mathbf{X}_{(n)}^{(2)} - \hat{\mathbf{X}}_{(n)}^{(2)}\right]$$

and where $\hat{\mathbf{X}}_{(n)}^{(2)} = \mathbf{X}_n^o(\mathbf{X}_{(n)}^{o\prime} \mathbf{X}_{(n)}^o)^{-1} \mathbf{X}_{(n)}^{o\prime} \mathbf{X}_{(n)}^{(2)}$

is the projection of $\mathbf{X}_{(n)}^{(2)}$ on the space spanned over the columns of $\mathbf{X}_{(n)}^o$.

(φ.1). $\varphi : (0,1) \to \mathbb{R}_1$ is a nondecreasing, square–integrable function which has a derivative $\varphi'(\alpha)$ for $0 < \alpha < \alpha_0$ and $1-\alpha_0 < \alpha < 1$ satisfying

$$|\varphi'(\alpha)| \leq c(\alpha(1-\alpha))^{-1-\eta^*}$$

with $0 < \eta^* < \eta$.

For any $\underset{\sim}{b} \in \mathbb{R}_{p+q}$, let $R_{ni}(\underset{\sim}{b})$ denote the rank of pseudo observation $Y_i - \underset{\sim}{b}'\underset{\sim}{x}_i$, i = 1,...,n. Calculate the scores generated by φ in either of the ways in (2.1) and the Jaeckel estimator $\hat{\underset{\sim}{\beta}}$ of $\underset{\sim}{\beta}$; consider its part $\hat{\underset{\sim}{\beta}}^{(2)}$ as an estimator of $\underset{\sim}{\beta}^{(2)}$.

Moreover, for $\underset{\sim}{t} \in \mathbb{R}_q$, calculate the regression rank scores $\hat{a}_m(\alpha, \underset{\sim}{Y} - \underset{\sim}{X}^{(2)}_{(n)} \underset{\sim}{t})$ from (2.8) with $\underset{\sim}{Y}$ replaced by $\underset{\sim}{Y} - \underset{\sim}{X}^{(2)}_{(n)} \underset{\sim}{t}$ and then calculate the scores

$$\hat{b}_m(\underset{\sim}{Y} - \underset{\sim}{X}^{(2)}_{(n)} \underset{\sim}{t}) = -\int_0^1 \varphi(\alpha)\, d\hat{a}_m(\alpha, \underset{\sim}{Y} - \underset{\sim}{X}^{(2)}_{(n)} \underset{\sim}{t}) \tag{3.2}$$

i=1,...,n. The Jurečková estimator $\tilde{\underset{\sim}{\beta}}^{(2)}$ then minimizes $D_n(\underset{\sim}{t})$ in (2.13) in $\underset{\sim}{t} \in \mathbb{R}_q$.

The following theorem gives the Bahadur–type representations for both $\hat{\underset{\sim}{\beta}}^{(2)}_r$ and $\tilde{\underset{\sim}{\beta}}^{(2)}_n$ which imply their asymptotic equivalence and distributions.

THEOREM 3.1. Assume that F, $\underset{\sim}{X}$ *and* φ *satisfy the above conditions and that*

$$0 < \gamma = -\int_0^1 \varphi(\alpha)\, df(F^{-1}(\alpha)) < \infty. \tag{3.3}$$

Then, as $n \to \infty$,

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\beta}}^{(2)}_n - \underset{\sim}{\beta}^{(2)}) = n^{-\frac{1}{2}}\, \gamma^{-1}(\underset{\sim}{Q}^{(2)})^{-1} \sum_{i=1}^{n} (\underset{\sim}{x}^{(2)}_i - \hat{x}^{(2)}_i)\varphi(F(e_i)) + o_p(1) \tag{3.4}$$

and

$$n^{\frac{1}{2}}(\tilde{\underset{\sim}{\beta}}^{(2)}_n - \underset{\sim}{\beta}^{(2)}) = n^{-\frac{1}{2}}\, \gamma^{-1}(\underset{\sim}{Q}^{(2)})^{-1} \sum_{i=1}^{n} (\underset{\sim}{x}^{(2)}_i - \hat{x}^{(2)}_i)\varphi(F(e_i)) + o_p(1) \tag{3.5}$$

with $\underset{\sim}{x}^{(2)}_i$, *and* $\hat{\underset{\sim}{x}}^{(2)}_i$, *being the i–th row of* $\underset{\sim}{X}^{(2)}_{(n)}$ *and* $\hat{\underset{\sim}{X}}^{(2)}_{(n)}$, *respectively,* i=1,...,n. *Consequently,*

$$n^{\frac{1}{2}}\|\hat{\underset{\sim}{\beta}}^{(2)}_n - \tilde{\underset{\sim}{\beta}}^{(2)}\| \overset{p}{\to} 0, \quad \textit{as } n \to \infty \tag{3.6}$$

and both $n^{\frac{1}{2}}(\hat{\underset{\sim}{\beta}}^{(2)}_n - \underset{\sim}{\beta}^{(2)})$ *and* $n^{\frac{1}{2}}(\tilde{\underset{\sim}{\beta}}^{(2)}_n - \underset{\sim}{\beta}^{(2)})$ *have the same asymptotic q–dimensional normal distribution*

$$p(\underset{\sim}{0}, (\underset{\sim}{Q}^{(2)})^{-1} A^2(\varphi,F)\gamma^{-2}) \tag{3.7}$$

where $A^2(\varphi,F) = \int_0^1 \varphi^2(\alpha)d\alpha - \bar{\varphi}^2$.

Proof. By Heiler and Willers (1988) and Jurečková (1971, 1977),

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}) = \gamma^{-1} \underset{\sim}{Q}_n^{-1} \underset{\sim}{S}_n(\underset{\sim}{\beta}) + o_p(1) \tag{3.8}$$

where

$$\underset{\sim}{S}_n(\underset{\sim}{\beta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \underset{\sim}{x}_i A_n(R_{ni}(\underset{\sim}{\beta})). \tag{3.9}$$

Writing

$$\underset{\sim}{X}'_{(n)} \underset{\sim}{X}_{(n)} = \begin{bmatrix} \underset{\sim}{V}_{11} & \underset{\sim}{V}_{12} \\ \underset{\sim}{V}_{21} & \underset{\sim}{V}_{22} \end{bmatrix} \tag{3.10}$$

where $\underset{\sim}{V}_{ij} = \underset{\sim}{X}^{(i)'}_{(n)} \underset{\sim}{X}^{(j)}_{(n)}$, i,j = 1,2, we obtain (for $n \geq n_0$)

$$(\underset{\sim}{X}'_{(n)} \underset{\sim}{X}_{(n)})^{-1} = \begin{bmatrix} \underset{\sim}{V}^{11} & \underset{\sim}{V}^{12} \\ \underset{\sim}{V}^{21} & \underset{\sim}{V}^{22} \end{bmatrix} \tag{3.11}$$

where

$$\underset{\sim}{V}^{21} = -\underset{\sim}{V}_0^{-1} \underset{\sim}{V}_{21} \underset{\sim}{V}_{11}^{-1}, \quad \underset{\sim}{V}^{22} = \underset{\sim}{V}_0^{-1} \tag{3.12}$$

and

$$\underset{\sim}{V}_0 = \underset{\sim}{V}_{22} - \underset{\sim}{V}_{21} \underset{\sim}{V}_{11}^{-1} \underset{\sim}{V}_{12} = \left[\underset{\sim}{X}^{(2)}_{(n)} - \hat{\underset{\sim}{X}}^{(2)}_{(n)}\right]' \left[X^{(2)}_{(n)} - \hat{X}^{(2)}_{(n)}\right] \tag{3.13}$$

Writing (3.9) as

$$\underset{\sim}{S}_n(\underset{\sim}{\beta}) = \begin{bmatrix} S_n^{(1)}(\underset{\sim}{\beta}) \\ S_n^{(2)}(\underset{\sim}{\beta}) \end{bmatrix} = \begin{bmatrix} n^{-\frac{1}{2}} \sum_{i=1}^{n} \underset{\sim}{x}_i^{(1)} A_n(R_{ni}(\underset{\sim}{\beta})) \\ n^{-\frac{1}{2}} \sum_{i=1}^{n} \underset{\sim}{x}_i^{(2)} A_n(R_{ni}(\underset{\sim}{\beta})) \end{bmatrix} \tag{3.14}$$

we obtain, in view of (3.8), (3.12) and (3.13),

$$n^{\frac{1}{2}} \gamma(\hat{\underset{\sim}{\beta}}_n^{(2)} - \underset{\sim}{\beta}^{(2)}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} (\underset{\sim}{V}^{21} \underset{\sim}{x}_i^{(1)} + \underset{\sim}{V}^{22} \underset{\sim}{x}_i^{(2)}) A_n(R_{ni}(\underset{\sim}{\beta})) + o_p(1)$$

$$= n^{-\frac{1}{2}}(Q_n^{(2)})^{-1} \sum_{i=1}^{n} (x_i^{(2)} - \hat{x}_i^{(2)}) A_n(R_{ni}(\underset{\sim}{\beta})) + o_p(1)$$

and this leads to (3.4) with the aid of results of Hájek and Sidák (1967), Chapter V.

For the estimator $\tilde{\beta}_n^{(2)}$, Jurečková (1991) proved

$$n^{\frac{1}{2}} \gamma(\tilde{\beta}^{(2)}-\beta^{(2)}) = n^{-\frac{1}{2}}(Q_n^{(2)})^{-1} \sum_{i=1}^{n} (x_i^{(2)}-\hat{x}_i^{(2)})\, \hat{b}_{ni}(Y-X\beta) + o_p(1)$$

and this, by Theorem 4.1 in Gutenbrunner et al. (1992) implies

$$n^{\frac{1}{2}} \gamma(\tilde{\beta}_n^{(2)}-\beta^{(2)}) = n^{-\frac{1}{2}}(Q_n^{(2)})^{-1} \sum_{i=1}^{n} (x_i^{(2)}-\hat{x}_i^{(2)})\, \varphi(F(e_i)) + o_p(1)$$

and hence (3.5) follows. The remaining propositions easily follow from (3.4) and (3.5).

∎

If we apply the Jurečková procedure separately to every component of $\beta^{(2)}$, we obtain the coordinate estimator of $\beta^{(2)}$ which we shall denote $\breve{\beta}_n^{(2)}$. More precisely, put $\beta^{(2)} = \beta_{p+j}$ (the $(p+j)^{th}$ component of β) and $X_{(n)}^{(2)} = x^{(p+j)}$ (the $(p+j)^{th}$ column of $X_{(n)}$), $j = 1,\dots,q$. Calculate the regression rank scores cooresponding to the matrix $X_{(n)}^{*} \backslash x_{(n)}^{(p+j)}$ and to the pseudo observations $Y_i - t\, x_{i,p+q}$ and the scores generated by $\varphi_{(j)}$ by (3.2), which we shall denote by $\hat{b}_{ni}^{(j)}(Y - t\, x_{(n)}^{(p+j)})$, $i = 1,\dots,n$; $j=1,\dots,q$; $t \in \mathbb{R}_1$. The estimator β_{p+j}^{0} of β_{p+j} is then a solution of the minimization

$$D_n^{(j)}(t) = \sum_{i=1}^{n} (Y_i - t\, x_{i,p+j})(\hat{b}_{ni}^{(j)}(Y - t\, x_{(n)}^{(p+j)}) - \bar{\varphi}) := \min, \; t \in \mathbb{R}_1. \tag{3.15}$$

The function $D_n^{(j)}(t)$ is convex and piecewise linear; its right–hand derivative

$$Z_n^{(j)}(t) = -\sum_{i=1}^{n} x_{i,j+p}[\hat{b}_{ni}^{(j)}(Y - t\, x_{(n)}^{(p+j)}) - \bar{\varphi}] \tag{3.16}$$

is nondecreasing in t. Hence, $\breve{\beta}_{p+j}$ could be characterized in the following way

$$\breve{\beta}_{p+j} = \tfrac{1}{2}(t_j^- + t_j^+)$$

$$t_j^- = \sup\{t : Z_n^{(j)}(t) > 0\}$$

$$t_j^+ = \inf\{t : Z_n^{(j)}(t) < 0\}. \tag{3.17}$$

We propose $\tilde{\beta}^{(2)} = (\tilde{\beta}_{p+1},\ldots,\tilde{\beta}_{p+q})$, as an estimator of $\beta^{(2)}$.

Surprisingly, even $\tilde{\beta}^{(2)}$ is asymptotically equivalent to $\hat{\beta}^{(2)}$ and to $\tilde{\beta}^{(2)}$, as it is formulated in the following theorem. Analogously, we could estimate $\beta^{(2)}$ not only componentwise but also in arbitrary groups of components. This provides us with a variety of alternative possibilities.

THEOREM 3.2. Consider the model (1.4) with the design matrix satisfying conditions (X.1) and (X.2) and with every column of $X_{(n)}^{(2)}$ *satisfying (X.3). Then, under (F.1)–(F.3), (φ.1) and (3.3),*

$$n^{\frac{1}{2}}\|\tilde{\beta}_n^{(2)} - \hat{\beta}_n^{(2)}\| = o_p(1),\ n^{\frac{1}{2}}\|\tilde{\beta}_n^{(2)} - \tilde{\beta}_n^{(2)}\| = o_p(1) \tag{3.18}$$

as $n\to\infty$, *and* $\tilde{\beta}_n^{(2)}$ *admits the asymptotic representation (3.4) and is asymptotically normally distributed according to (3.7).*

<u>Proof</u>. It follows from Theorem 3.1 that

$$n^{\frac{1}{2}}(\tilde{\beta}_{p+j,n} - \hat{\beta}_{p+j,n}) = o_p(1),\ j=1,\ldots,q,$$

hence

$$n^{\frac{1}{2}}\|\tilde{\beta}_n^{(2)} - \hat{\beta}_n^{(2)}\| = o_p(1)$$

and this further implies the remaining propositions. ■

4. ESTIMATION IN k–SAMPLE MODEL

As an illustration, apply the above results to the special model

$$Y_{ij} = \beta_o + \beta_i + e_{ij},\quad j = 1,\ldots,n_i;\ i = 1,\ldots,k \tag{4.1}$$

with $\beta_1 = 0$. Consider the problem of estimating β_k.

The regression rank scores corresponding to the submodel (4.1) under $\beta_1 = \beta_k = 0$ satisfy

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}\,\hat{a}_{ij}(\alpha) := \max$$

$$\sum_{j=1}^{n_i} \hat{a}_{ij}(\alpha) = n_i(1-\alpha),\ i = 2,3,\ldots,k-1 \tag{4.2}$$

$$\sum_{j=1}^{n_1} \hat{a}_{1j}(\alpha) + \sum_{j=1}^{n_k} \hat{a}_{kj}(\alpha) = (n_1+n_k)(1-\alpha),\ 0 < \alpha < 1.$$

This implies that $(\hat{a}_{i1}(\alpha),\ldots,\hat{a}_{in_i}(\alpha))$ generate the ranks of the i^{th} sample $Y_{i1},\ldots,Y_{in_i}$, $i = 2,\ldots,k-1$, in the same manner as in (2.9). On the other hand, $(\hat{a}_{11}(\alpha),\ldots,\hat{a}_{1n_1}(\alpha), \hat{a}_{k1}(\alpha),\ldots,\hat{a}_{kn_k}(\alpha))$ analogously generate the ranks in the combined 1^{st} and k^{th} samples.

Now, let us calculate regression rank scores analogously, with the only difference that the k^{th} sample in (4.2) is replaced by pseudo observations $Y_{k1}-t_k,\ldots,Y_{kn_k}-t_k$. Calculate the Wilcoxon scores corresponding to $\varphi(\alpha) = \alpha - \frac{1}{2}$, $0 \le \alpha \le 1$, i.e.

$$\hat{b}_{ij}(t_k) = -\int_0^1 (\alpha - \tfrac{1}{2})\, d\, \hat{a}_{ij}(\alpha, t_k) \tag{4.3}$$

$j = 1,\ldots,n_i$; $i=1,\ldots,k$. The estimator of β_k then minimizes the function

$$D_n(t_k) = \sum_{i=1}^{k-1} \sum_{j=1}^{n_i} \hat{b}_{ij}(t_k) Y_{ij} + \sum_{j=1}^{n_k} \hat{b}_{kj}(t_k)(Y_{kj}-t_k) \tag{4.4}$$

with respect to $t_k \in \mathbb{R}_1$. The right–hand derivative of (4.4) in t_k is equal to $-\sum_{j=1}^{n_k} \hat{b}_{kj}(t_k)$, while the above remark shows that

$$(n\hat{b}_{11}(t_k),\ldots,n\hat{b}_{1n_1}(t_k),\, n\hat{b}_{k1}(t_k),\ldots,n\hat{b}_{kn_k}(t_k))$$

are the ranks of $(Y_{11},\ldots,Y_{1n_1}, Y_{k1}-t_k,\ldots,Y_{kn_k}-t_k)$ centered by $\frac{1}{2}(n_1 + n_k + 1)$. Hence, we immediately obtain that the estimator $\hat{\beta}_k$ of β_k based on regression rank scores coincide with the Wilcox scores estimator of the shift between the k^{th} and the 1^{st} samples, i.e.

$$\hat{\beta}_k = \text{med}\{Y_{kr}-Y_{1s} : 1 \le r \le u_k,\ 1 \le s \le n_1\}, \tag{4.5}$$

considered by Hodges and Lehmann (1963) and Sen (1963). When we estimate $\beta_2,\ldots,\beta_k$ analogously we obtain the coordinatewise estimator $(\hat{\beta}_2,\ldots,\hat{\beta}_k)$ of $(\beta_2,\ldots\beta_k)$. The j^{th} component $\hat{\beta}_j$ is the Wilcoxon scores estimator between the j^{th} and the 1^{st} samples. It follows from Theorem 3.2 that $(\hat{\beta}_2,\ldots,\hat{\beta}_k)$ is asymptotically equivalent to Jaeckel's estimator of $(\beta_2,\ldots,\beta_k)$, corresponding to the Wilcoxon scores.

REFERENCES

[1] ADICHIE, J.N. (1967). Estimating of regression coefficients based on rank tests. *Ann. Math. Statist.* 38, 894–904.

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

[3] ADICHIE, J.N. (1984). Rank tests in linear models. Handbook of Statistics, Vol. 4, Nonparametric Methods. (P.R. Krishnaiah and P.K. Sen, eds.), pp. 229–257. Elsevier, New York.

[4] GUTENBRUNNER, C. (1986). Zur Asymptotik von Regressionquantileprozessen und daraus abgeleiten Statistiken. Ph.D. Dissertation, Universitat Freiburg.

[5] GUTENBRUNNER, C. and JUREČKOVÁ, J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* 20, No. 1.

[6] GUTENBRUNNER, C., JUREČKOVÁ, J., KOENKER, R. and PORTNOY, S. (1992). Tests of linear hypotheses based on regression rank scores. Submitted.

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

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

[9] HODGES, J.L. and LEHMANN, E.L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* 34, 598–611.

[10] JAECKEL, L.A. (1972). Estimating regression coefficients by minimizing the dispersion of residuals. *Ann. Math. Statist.* 43, 1449–1458.

[11] JUREČKOVÁ, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. *Ann. Math. Statist.* 40, 1889–1900.

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

[13] JUREČKOVÁ, J. (1971b). Asymptotic independence of rank test statistic for testing symmetry on regression. *Sankhyā A*, 33, 1–18.

[14] JUREČKOVÁ, J. (1977). Asymptotic relations of M–estimates and R–estimates in linear regression model. *Ann. Statist.* 5, 664–672.

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

[16] JUREČKOVÁ, J. (1991). Estimation in a linear model based on regression rank scores. *Nonparametric Statist.* 1, 197–203.

[17] JUREČKOVÁ, J. (1992). Uniform asymptotic linearity of regression rank scores process. Proc. Intern. Symp. on Nonparametric Statistics and Related Topics (A.K. Md. E. Saleh, ed.), pp.   , Elsevier, New York.

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

[19] KOENKER, R. and d'OREY, V. (1987). Computing regression quantiles. *Applied Statist.* 36, 383–393.

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

[21] KOUL, H.L. (1970). A class of ADF tests for subhypothesis in the multiple linear regression. *Ann. Math. Statist.* 41, 1273–1281.

[22] PURI, M.L. and SEN, P.K. (1985). Nonparametric Methods in General Linear Models. J. Wiley, New York.

[23] RUPPERT, D. and CARROLL, R.J. (1980). Trimmed least squares estiamtion in the linear model. *J. Amer. Statist. Assoc.* 75, 256–272.

[24] SEN, P.K. (1963). On the estimation of relative potency in dilution (direct) assays by distribution–free methods. *Biometrics* 19, 532–552.

[25] SEN, P.K. (1968). Estimates of regression based on Kendall's tau. *Jour. Amer. Statist. Assoc.* 63, 1379–1389.

[26] SEN, P.K. and PURI, M.L. (1977). Asymptotically distribution–free aligned rank order tests for composite hypotheses for general linear models. *Zeit. Warsch. Verw. Geb.* 39, 175–186.