

UNIVERSITY OF
ILLINOIS LIBRARY
AT URBANA-CHAMPAIGN
BOOKSTACKS

THE HECKMAN BINDERY, INC.
North Manchester, Indiana

KRL

H or V

JUST FONT SLOT TITLE

H CC 1W 22 BEBR
21 FACULTY
20 WORKING
19 PAPER

H CC 1W 8 1990
7 NO.1674-1689

H CC 1W 330
B385<"CV">
no.1674-1689
cop.2

H CC 7W
<IMPRINT>
U. of ILL.
LIBRARY
URBANA

BINDING COPY

PERIODICAL	<input type="checkbox"/> CUSTOM	<input type="checkbox"/> STANDARD	<input type="checkbox"/> ECONOMY	<input type="checkbox"/> THESIS	NO. VOLS.	THIS TITLE	LEAD ATT.
BOOK	<input type="checkbox"/> CUSTOM	<input type="checkbox"/> MUSIC	<input type="checkbox"/> ECONOMY	AUTH 1ST			
ACCOUNT	LIBRARY	NEW	RUBOR	TITLE I.D.	FOIL	COLOR	MATERIAL
	66672 001				6632	WHI	488
	ACCOUNT NAME						

UNIV OF ILLINOIS
ACCOUNT INTERNAL I.D.

ISSN.

P01912400
I.D. #2

NOTES
BINDING
FREQUENCY

WHEEL

SYS I.D.

STX3
COLLATING

1 3

39255

35

ADDITIONAL INSTRUCTIONS

Dept=STX3 Lot=#20 Item=149 HNM=[ZY#
1CR2ST3CR MARK BY # B4 91]

SEP. SHEETS	PTS. BD. PAPER	TAPE STUBS	CLOTH EXT.	GUM	FILLER	STUB
POCKETS PAPER BUCK CLOTH						
SPECIAL PREP.						
LEAF ATTACH.						

INSERT MAT	ACCOUNT LOT NO	JOB NO
PRODUCT TYPE	ACCOUNT PIECE NO.	PIECE NO

HEIGHT	GROUP CARD	VOL THIS
11	1	149
11.2	TITLE	2

COVER SIZE

V X 12

001247520

330
385
D. 1676 COPY 2

STX

BEBR
FACULTY WORKING
PAPER NO. 90-1676

Tests of Linear Hypotheses Based
on Regression Rank Scores

C. Gutenbrunner
J. Jurčková
R. Koenker
S. Portnoy

The Library of JIS

SEP 2 1990

University of Illinois
at Urbana-Champaign



College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 90-1676

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

August 1990

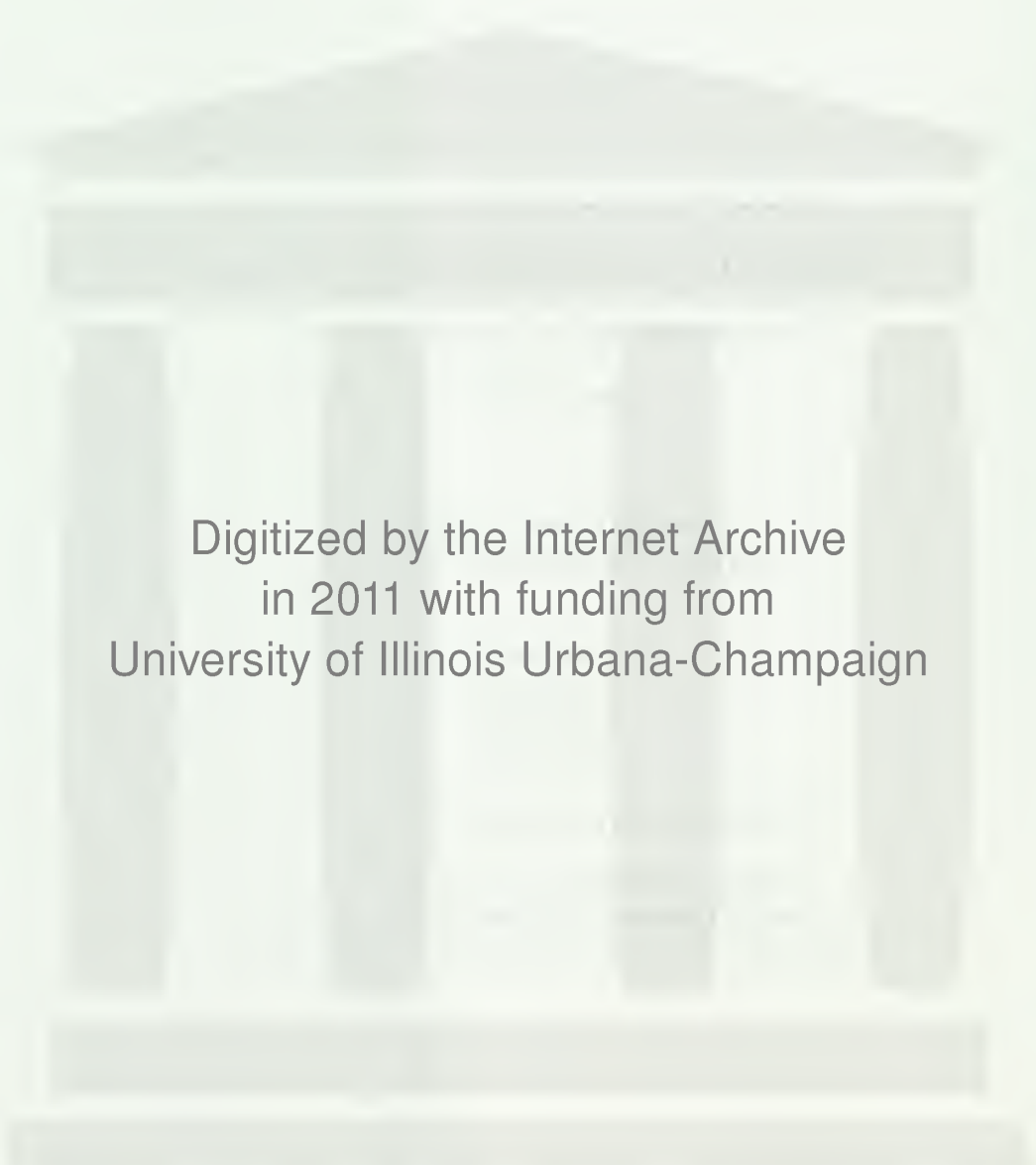
Tests of Linear Hypotheses
Based on Regression Rank Scores

C. Gutenbrunner
J. Jurečková
R. Koenker
S. Portnoy

Dedicated to the memory of Jaroslav Hájek

Philipps Universität, Marburg, Germany
Charles University, Prague, Czechoslovakia
University of Illinois at Urbana-Champaign, USA

The work was partially supported by NSF grants 88-02555 and 89-22472 to S. Portnoy and R. Koenker and by support from the Australian National University to J. Jurečková and R. Koenker.



Digitized by the Internet Archive
in 2011 with funding from
University of Illinois Urbana-Champaign

ABSTRACT

We propose a general class of asymptotically distribution-free tests of a linear hypothesis in the linear regression model. The tests are based on regression rank scores, recently introduced by Gutenbrunner and Jurečková (1990) as dual variables to the regression quantiles of Koenker and Bassett (1978). Their properties are analogous to those of the corresponding rank tests in location model. Unlike the other regression tests based on aligned rank statistics, however, our tests do not require preliminary estimation of nuisance parameters.

AMS 1980 subject classifications: 62G10, 62J05

Key words and phrases: Regression quantile, regression rank scores

TESTS OF LINEAR HYPOTHESES BASED ON REGRESSION RANK SCORES

C. Gutenbrunner, J. Jurečková, R. Koenker, S. Portnoy

Philipps University, Marburg, Germany
Charles University, Prague, Czechoslovakia
University of Illinois at Urbana-Champaign, U.S.A

1. Introduction

Several authors including Koul (1970), Puri and Sen (1985) and Adichie (1978) have proposed and investigated asymptotically distribution-free tests of some types of linear hypothesis for the linear regression model based upon aligned rank statistics. A good review of these results including extensions to multivariate models may be found in Puri and Sen (1985) and Adichie (1984). The hypothesis under consideration typically involves nuisance parameters which should be estimated by a preliminary estimate; the aligned (or signed) rank statistics are then based on residuals from the preliminary estimates. Alternative approaches to inference based on rank *estimation* have been considered by McKean and Hettmansperger (1978), Aubuchon and Hettmansperger (1988) and Draper (1988).

In this paper we explore an alternative approach to the construction of rank tests for the linear regression model based on the regression rank scores introduced in Gutenbrunner and Jurečková (1990). Regression rank scores represent a natural extension of location rank scores introduced in Hájek and Šidák (1967, Section V.3.5) which stand in one-to-one relation to the ranks of sample observations. The tests based on regression rank scores offer a natural extension of rank-based methods of testing to the general linear model and avoid many of the difficulties introduced by preliminary estimation of nuisance parameters in prior proposals.

The next section of the paper surveys our results and establishes notation. Section 3 develops some theory of the regression rank score process. Section 4 treats the theory of

simple linear rank statistics based on this process. Section 5 contains a formal treatment of the proposed tests. And Section 6 describes an example.

2. Notation and preliminary considerations

Consider the classical linear regression model

$$Y = X\beta + E \quad (2.1)$$

which we will partition as

$$Y = X_1\beta_1 + X_2\beta_2 + E \quad (2.2)$$

where β_1 and β_2 are p - and q -dimensional parameters, $X = X_n$ is a known, $n \times (p+q)$ design matrix with rows $x_{ni}' = x_i' = (x_{1i}', x_{2i}') \in R^{p+1}$, $i = 1, \dots, n$. We will assume throughout that $x_{i1} = 1$ for $i = 1, \dots, n$. Y is a vector of observations and E is an $n \times 1$ vector of *i.i.d.* errors with common distribution function F . The precise form of F need not be known but we shall generally assume that F has an absolutely continuous density f on (A, B) where $-\infty \leq A = \sup\{x: F(x) = 0\}$ and $+\infty \geq B = \inf\{x: F(x) = 1\}$. Moreover, we shall impose some conditions on the tails of f assuming, among other conditions, that f monotonically decreases to 0 when $x \rightarrow A+$, or $x \rightarrow B-$. Denoting $D_n = n^{-1}X_1'X_1$ and

$$H_1 = X_1(X_1'X_1)^{-1}X_1' \quad \text{and} \quad Q_n = n^{-1}(X_2 - \hat{X}_2)'(X_2 - \hat{X}_2) \quad (2.3)$$

with $\hat{X}_2 = H_1X_2$ being the projection of X_2 on the space spanned by the columns of X_1 , we shall also assume

$$\lim_{n \rightarrow \infty} D_n = D, \quad \lim_{n \rightarrow \infty} Q_n = Q \quad (2.4)$$

where D and Q are positive definite $(p \times p)$ and $(q \times q)$ matrices, respectively.

We are interested in testing the hypothesis

$$H_0: \beta_2 = 0, \quad \beta_1 \text{ unspecified} \quad (2.5)$$

versus the Pitman (local) alternatives

$$H_n : \beta_{2n} = n^{-1/2} \beta_0 \quad (2.6)$$

with β_0 being a fixed vector from \mathbb{R}^q .

The regression rank scores introduced in Gutenbrunner and Jurečková (1990) arise as a vector of solutions

$$\hat{\mathbf{a}}_n(\alpha) = (\hat{a}_{n1}(\alpha), \dots, \hat{a}_{nn}(\alpha))', \quad 0 < \alpha < 1 \quad (2.7)$$

of the dual form of the linear program required to compute the regression quantile statistics of Koenker and Bassett (1978). More precisely, the vector $\hat{\beta}_1(\alpha) = (\hat{\beta}_1(\alpha), \dots, \hat{\beta}_p(\alpha))' \in \mathbb{R}^p$ of regression quantiles corresponding to the submodel

$$Y = X_1 \beta_1 + E \quad (2.8)$$

is any solution of the minimization

$$\sum_{i=1}^n \rho_\alpha(Y_i - x_{1i}' t), \quad t \in \mathbb{R}^q \quad (2.9)$$

where

$$\rho_\alpha(x) = |x| \{ (1-\alpha)I[x < 0] + \alpha I[x > 0] \}, \quad x \in \mathbb{R}^1. \quad (2.10)$$

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

$$\alpha \mathbf{1}_n' r^+ + (1-\alpha) \mathbf{1}_n' r^- : = \min$$

$$X_1 \hat{\beta} + r^+ - r^- = Y \quad (2.11)$$

$$\beta \in \mathbb{R}^p, \quad r^+, r^- \in \mathbb{R}_+^n$$

and $\mathbf{1}_n = (1, \dots, 1)' \in \mathbb{R}^n$, $0 < \alpha < 1$. Finite-sample as well as asymptotic properties of $\hat{\beta}(\alpha)$ are studied in Koenker and Bassett (1978), Ruppert and Carroll (1980), Jurečková (1984), Gutenbrunner (1986), Koenker and Portnoy (1987), Gutenbrunner and Jurečková (1990) among others.

The dual program to (2.10) can be written in the form

$$Y' \hat{a}(\alpha) := \max$$

$$X_1' \hat{a}(\alpha) = (1-\alpha) X_1' 1_n \quad (2.12)$$

$$\hat{a}(\alpha) \in [0, 1]^n, \quad 0 < \alpha < 1$$

As shown in Gutenbrunner and Jurečková (1990) many aspects of the duality of order statistics and ranks in the location model generalize naturally to the linear model through (2.11) and (2.12).

To motivate the approach, let us illustrate $\{\hat{a}_n(\alpha), 0 < \alpha < 1\}$ in the location model which in the present notation may be viewed as (2.11) with $X_1 = 1_n$. Then $\hat{a}_{ni}(\alpha)$ specializes to

$$a_{ni}^*(\alpha) = a_n^*(R_i, \alpha) = \begin{cases} 1 & \text{if } \alpha \leq (R_i-1)/n \\ R_i - \alpha n & \text{if } (R_i-1)/n < \alpha \leq R_i/n \\ 0 & \text{if } R_i/n < \alpha \end{cases} \quad (2.13)$$

where R_i is the rank of Y_i among Y_1, \dots, Y_n . The function $a_n(j, \alpha)$, $j=1, \dots, n$, $0 < \alpha < 1$, coincides with that introduced in Hájek and Šidák (1967, Section V.3.5). Under the general model (2.1) the finite-sample as well as asymptotic properties of the regression rank scores and of the process $\{\hat{a}_n(\alpha), 0 < \alpha < 1\}$ are studied in the next section.

As in the classical theory of rank tests, we shall consider a score-function $\phi : (0, 1) \rightarrow R^1$ which is nondecreasing and square-integrable on $(0, 1)$. We may then construct scores in the following way:

$$\hat{b}_{ni} = - \int_0^1 \phi(t) d \hat{a}_{ni}(t), \quad i=1, \dots, n. \quad (2.14)$$

Defining

$$S_n = n^{-1/2} (X_{n2} - \hat{X}_{n2})' \hat{b}_n \quad (2.15)$$

where $\hat{b}_n = (\hat{b}_{n1}, \dots, \hat{b}_{nn})'$, we propose the following statistic for testing H_0 against H_n :

$$T_n = S_n' Q_n^{-1} S_n / A^2(\phi) \quad (2.16)$$

with

$$A^2(\phi) = \int_0^1 (\phi(t) - \bar{\phi})^2 dt, \quad \bar{\phi} = \int_0^1 \phi(t) dt \quad (2.17)$$

and with Q_n defined as in (2.3). An important feature of the test statistic T_n is that it requires no estimation of nuisance parameters, since the functional $A(\phi)$ depends only on the score function and not on F . This is familiar from the theory of rank tests, but stands in sharp contrast with other methods of testing in the linear model where typically some estimation of a scale parameter of F is required to compute the test statistic. See for example the discussion in Aubuchon and Hettmansperger (1988) and Draper (1988).

As we shall show in Section 5, the asymptotic distribution of T_n under H_0 is central χ^2 with q degrees of freedom while under H_n it is noncentral χ^2 with q degrees of freedom and noncentrality parameter

$$\eta^2 = [\gamma^2(\phi, F) / A^2(\phi)] \beta_0' Q \beta_0 \quad (2.18)$$

where

$$\gamma(\phi, F) = -\int_0^1 \phi(t) df(F^{-1}(t)) \quad (2.19)$$

The quantities γ and A are familiar from the theory of rank tests. The test based on T_n is asymptotically distribution free in the sense that, under H_0 , neither T_n nor its asymptotic distribution depend on F . Moreover, it follows from (2.18) that the Pitman efficiency of the test based on T_n with respect to the classical F test of H_0 coincides with that of the two-sample rank test of shift in location with respect to the t -test. For f unimodal, we obtain an asymptotically optimal test if we take

$$\phi(t) = \phi_f(t) = -\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}, \quad 0 < t < 1. \quad (2.20)$$

Thus the asymptotic relative efficiency of the test based on T_n , relative to the classical F test is, for Wilcoxon scores ($\phi(u) = u - 1/2$), is $3/\pi = .955$ at the normal distribution and is bounded below by .864 for all F . When F is heavy tailed this asymptotic efficiency is generally greater than one, and can in fact be unbounded. For normal (van der Waerden) scores ($\phi(u) = \Phi^{-1}(u)$) the situation is even more striking. Here the test based on T_n has asymptotic efficiency greater than one, relative to the classical F test, for all symmetric F , attaining one at the normal distribution. See *e.g.* Lehmann (1959, p. 239), and Lehmann(1983, pp 383-87).

Let us now look at the scores (2.14), which can be written as

$$\hat{b}_{ni} = -\int \phi(t) \hat{a}_{ni}'(t) dt \quad i=1, \dots, n \quad (2.21)$$

where the functions $a_{ni}'(t)$ are piecewise constant on $[0,1]$. In the location model, using (2.13) this reduces to

$$\hat{b}_{ni} = n \int_{(R_{i-1})/n}^{R_i/n} \phi(t) dt, \quad i=1, \dots, n$$

There are three typical choices of ϕ :

- (i) *Wilcoxon scores*: $\phi(t) = t - 1/2$, $0 < t < 1$. The scores are

$$\hat{b}_{ni} = -\int_0^1 (t - 1/2) d\hat{a}_i(t) = \int_0^1 \hat{a}_i(t) dt - 1/2$$

while $A^2(\phi) = 1/12$, and $\gamma(\phi, F) = \int f^2(x) dx$. Wilcoxon scores are optimal when f is the logistic distribution.

- (ii) *Normal (van der Waerden) scores*: $\phi(t) = \Phi^{-1}(t)$, $0 < t < 1$, Φ being the *d.f.* of standard normal distribution. Here $A^2(\phi) = 1$ and $\gamma(\phi, F) = \int f(F^{-1}(\Phi(x))) dx$. These scores are asymptotically optimal when f is normal.

- (iii) *Median (sign) scores*: $\phi(t) = \frac{1}{2} \text{sign}(t - \frac{1}{2})$, $0 < t < 1$, then (2.14) leads to the form $\hat{b}_{ni} = \hat{a}_{ni}'(\frac{1}{2})$ which is non-zero if and only if the l_1 estimate passes through the i th observation.

REMARK. Using the standard reduction to canonical form *e.g.* Scheffé (1959, Section 2.6) or Amemiya (1985, Section 1.4.2), we may consider a more general form of the linear hypothesis

$$\mathbf{R}'\boldsymbol{\beta} = \mathbf{r} \in \mathbf{R}^q \quad (2.22)$$

where \mathbf{R} is a $(p + q) \times q$ matrix of rank $q < p$. Let \mathbf{V} be a $(p + q) \times p$ matrix such that $\mathbf{A} = [\mathbf{V} : \mathbf{R}]'$ is nonsingular and $\mathbf{R}'\mathbf{V} = 0$. Set $\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\beta}$ and $\mathbf{Z} = \mathbf{X}\mathbf{A}^{-1}$. Partitioning $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1' : \boldsymbol{\gamma}_2']'$ where $\boldsymbol{\gamma}_1 = \mathbf{V}'\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_2 = \mathbf{R}'\boldsymbol{\beta}$, under the hypothesis (2.22) we have

$$\mathbf{Y} - \mathbf{X}\mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{r} = \mathbf{X}\mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\boldsymbol{\gamma}_1 + \mathbf{E}.$$

Thus, in view of the equivariance of regression quantiles we may define $\tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{r}$, $\tilde{\mathbf{X}}_1 = \mathbf{X}\mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}$, $\tilde{\mathbf{X}}_2 = \mathbf{X}\mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}$, and proceed as previously discussed with $(\tilde{\mathbf{Y}}, \tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2)$ playing the roles of $(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)$. By this device the tests described above and detailed in Section 5 below may be extended to a wide range of applications including, for example, the hypotheses of parallelism and coincidence of regression lines discussed by Adichie (1984) and others.

3. Properties of regression rank scores

Consider the linear regression model (2.1) with design \mathbf{X}_n of dimension $n \times p$. Let $\hat{\boldsymbol{\beta}}(\alpha) \in \mathbf{R}^p$ be the α -regression quantile in (2.11) and $\hat{\mathbf{a}}(\alpha) \in \mathbf{R}^n$ be the vector of α th regression rank scores defined in (2.15). We see from (2.12) that the regression rank scores are *regression invariant*, i.e.,

$$\hat{\mathbf{a}}_n(\alpha, \mathbf{Y} + \mathbf{X}\mathbf{b}) = \hat{\mathbf{a}}_n(\alpha, \mathbf{Y}), \quad \mathbf{b} \in \mathbf{R}^p. \quad (3.1)$$

Moreover, in view of (2.12), we may assume

$$\sum_{i=1}^n x_{ij} = 0, \quad j=2, \dots, p \quad (3.2)$$

without loss of generality. The formal duality between $\hat{\boldsymbol{\beta}}(\alpha)$ and $\hat{\mathbf{a}}(\alpha)$ implies that for $i=1, \dots, n$

$$\hat{\mathbf{a}}_{rs}(\alpha) = \begin{cases} 1 & \text{if } Y_i > \sum_{j=1}^p x_{ij} \hat{\beta}_j(\alpha) \\ 0 & \text{if } Y_i < \sum_{j=1}^p x_{ij} \hat{\beta}_j(\alpha) \end{cases} \quad (3.3)$$

while the components of $\hat{\mathbf{a}}_n(\alpha)$ corresponding to $\{i \mid Y_i = \mathbf{x}_i' \hat{\boldsymbol{\beta}}(\alpha)\}$ are determined by the equality constraints of (2.12).

Our primary interest in this section will be the properties of the regression rank scores process

$$\{\hat{\mathbf{a}}_n(t) : 0 \leq t \leq 1\}. \quad (3.4)$$

Gutenbrunner and Jurečková(1990) studied the process

$$\mathbf{W}_n^d = \{\mathbf{W}_n^d(t) = \sqrt{n} \sum_{i=1}^n \mathbf{d}_{ri} \hat{\mathbf{a}}_{ri}(t) : 0 \leq t \leq 1\} \quad (3.5)$$

and showed that $\mathbf{W}_n^d(t) = \mathbf{U}_n^d(t) + o_p(1)$ where

$$\mathbf{U}_n^d(t) = n^{-1/2} \sum_{i=1}^n \mathbf{d}_{ri} I[E_i > F^{-1}(t)] \quad (3.6)$$

as $n \rightarrow \infty$ uniformly on any fixed interval $[\epsilon, 1-\epsilon]$, where $0 < \epsilon < 1/2$ for any properly standardized triangular array $\{\mathbf{d}_{ri} : i=1, \dots, n\}$ of vectors from \mathbf{R}^q . They also showed that the process (3.4) (and hence (3.5)) has continuous trajectories and, under the standardization

$\sum_{i=1}^n \mathbf{d}_{ri} = 0$, (3.5) is tied-down to 0 at $t = 0$, and $t = 1$. The same authors also established the

weak convergence of (3.5) to the Brownian bridge over $[\epsilon, 1-\epsilon]$. Note however that Theorem V.3.5 in Hájek and Šidák (1967) establishes the weak convergence of (3.5) to the Brownian bridge over the entire interval $[0, 1]$ in the special case of location submodel. Here we extend the results of Gutenbrunner and Jurečková (1990) into the tails of $[0, 1]$, in order to find the asymptotic behavior of the rank scores and the test statistics (2.14) and (2.15).

To this end, we will assume that the errors E_1, \dots, E_n in (2.1) are independent and identically distributed according to the distribution function $F(x)$ which has an absolutely

continuous density f . We will assume that f is positive for $A < x < B$ and decreases monotonically as $x \rightarrow A+$, and $x \rightarrow B-$ where $-\infty \leq A = \sup \{x: F(x) = 0\}$ and $+\infty \geq B = \inf \{x: F(x) = 1\}$. For $0 < \alpha < 1$ denote ψ_α the score function corresponding to (2.10)

$$\psi_\alpha(x) = \alpha - I[x < 0], \quad x \in \mathbb{R}^1. \quad (3.7)$$

We shall impose the following conditions on F :

$$(F.1) \quad |F^{-1}(\alpha)| \leq c(\alpha(1-\alpha))^{-a} \text{ for } 0 < \alpha \leq \alpha_0, \quad 1-\alpha_0 \leq \alpha < 1, \text{ where } 0 < a \leq 1/4 - \epsilon, \quad \epsilon > 0$$

and $c > 0$.

$$(F.2) \quad 1/f(F^{-1}(\alpha)) \leq c(\alpha(1-\alpha))^{-1-a} \text{ for } 0 < \alpha \leq \alpha_0 \text{ and } 1-\alpha_0 \leq \alpha < 1, \quad c > 0.$$

$$(F.3) \quad f(x) > 0 \text{ is absolutely continuous, bounded and monotonically decreasing as } x \rightarrow A+ \text{ and } x \rightarrow B-. \text{ The derivative } f' \text{ is bounded a.e.}$$

$$(F.4) \quad \left| \frac{f'(x)}{f(x)} \right| \leq c|x| \text{ for } |x| \geq K \geq 0, \quad c > 0.$$

REMARK. These conditions are satisfied, for example, by the normal, logistic, double exponential and t distributions with 5, or more, degrees of freedom.

The following design assumptions will also be employed.

$$(X.1) \quad x_{i1} = 1, \quad i=1, \dots, n$$

$$(X.2) \quad \lim_{n \rightarrow \infty} D_n = D \text{ where } D_n = n^{-1} X_n' X_n \text{ and } D \text{ is a positive definite } p \times p \text{ matrix.}$$

$$(X.3) \quad n^{-1} \sum_{i=1}^n \|x_i\|^3 = O(1) \text{ as } n \rightarrow \infty.$$

$$(X.4) \quad \max_{1 \leq i \leq n} \|x_i\| = O(n^{(2(b-a)-\delta)/(1+4b)}) \text{ for some } b > 0 \text{ and } \delta > 0 \text{ such that } 0 < b-a < \epsilon/2 \text{ (hence } 0 < b < 1/4 - \epsilon/2).$$

We may now define

$$\alpha_n^* = n^{-1/(1+4b)} \quad (3.8)$$

and

$$\sigma_\alpha = \frac{(\alpha(1-\alpha))^{1/2}}{f(F^{-1}(\alpha))}, \quad 0 < \alpha < 1. \quad (3.9)$$

and prove the following crucial lemma.

LEMMA 3.1 Assume that F satisfies (F.1) - (F.4) and that X_n satisfies (X.1) - (X.3). Then, as $n \rightarrow \infty$,

$$\sup\{|r_n(t, \alpha)| : \|t\| \leq C, \alpha_n^* \leq \alpha \leq 1 - \alpha_n^*\} \xrightarrow{P} 0 \quad (3.10)$$

for any fixed $C > 0$, where

$$\begin{aligned} r_n(t, \alpha) &= (\alpha(1-\alpha))^{-1/2} \sigma_\alpha^{-1} \sum_{i=1}^n [\rho_\alpha(E_{i\alpha} - n^{1/2} \sigma_\alpha x_i' t) - \rho_\alpha(E_{i\alpha})] \\ &\quad + n^{-1/2} (\alpha(1-\alpha))^{-1/2} \sum_{i=1}^n x_i' t \psi_\alpha(E_{i\alpha}) - \frac{1}{2} t' D_n t \end{aligned} \quad (3.11)$$

and

$$E_{i\alpha} = E_i - F^{-1}(\alpha), \quad i = 1, \dots, n. \quad (3.12)$$

PROOF.

(i) First fix $\alpha \in [\alpha_n^*, 1 - \alpha_n^*]$ and t such that $\|t\| \leq C$.

Define

$$B_n = \min(n^{-(b-a)/(2(1+4b))}, n^{-2a/(1+4b)}) \quad (3.13)$$

We wish to show that for any $\lambda > 0$

$$P(|r_n(t, \alpha)| \geq (\lambda+1)B_n) \leq K n^{-\lambda} \quad (3.14)$$

with a fixed $K > 0$. To do this, we will use the Markov inequality

$$P(|r_n(t, \alpha)| \geq s_n) \leq \exp(-u s_n)(M(u) + M(-u)), \quad u > 0 \quad (3.15)$$

where $M(u) = E \exp(ur_n(t, \alpha))$.

Denote

$$\epsilon_{ni} = \epsilon_{ni}(t, \alpha) = n^{-1/2} \sigma_\alpha x_i' t \quad (3.16)$$

and

$$\begin{aligned} R_i(t, \alpha) &= (\alpha(1-\alpha))^{-1/2} \sigma_\alpha^{-1} [\rho_\alpha(E_{i\alpha} - n^{-1/2} \sigma_\alpha x_i' t) - \rho_\alpha(E_{i\alpha})] \\ &\quad + n^{-1/2} (\alpha(1-\alpha))^{-1/2} x_i' t \psi_\alpha(E_{i\alpha}) - \frac{1}{2} n^{-1} (x_i' t)^2 \quad i=1, \dots, n. \end{aligned} \quad (3.17)$$

By definition of $E_{i\alpha}$, σ_α , ρ_α and ψ_α ,

$$\begin{aligned} R_i(t, \alpha) + \frac{1}{2} n^{-1} (x_i' t)^2 &= (\alpha(1-\alpha))^{-1/2} \sigma_\alpha^{-1} \{ (E_{i\alpha} - \epsilon_{ni}) I[\epsilon_{ni} < E_{i\alpha} < 0] \\ &\quad + (\epsilon_{ni} - E_{i\alpha}) I[0 < E_{i\alpha} < \epsilon_{ni}] \} \end{aligned} \quad (3.18)$$

and hence, uniformly for $\alpha_n^* \leq \alpha \leq 1 - \alpha_n^*$, $\|t\| \leq C$ and $i=1, \dots, n$,

$$|R_i(t, \alpha) + \frac{1}{2} n^{-1} (x_i' t)^2| \leq 2n^{-1/2} (\alpha(1-\alpha))^{-1/2} |x_i' t| = O(n^{-(2a+\delta)/(1+4b)}). \quad (3.19)$$

If uR_i is bdd, that is $0 < u < n^{(2a+\delta)/(1+4b)}$ may be expanded to obtain.

$$\log M_{R_i}(u) \leq uER_i(t, \alpha) + cu^2 \text{Var}(R_i(t, \alpha)) \quad (3.20)$$

for some constant $c > 0$. By (3.18) and conditions F.1-F.4, for $\epsilon_{ni} > 0$ and for $\alpha_n^* \leq \alpha \leq \alpha_0$,

$1 - \alpha_n^* \geq \alpha \geq 1 - \alpha_0$,

$$\begin{aligned} ER_i(t, \alpha) + \frac{1}{2} n^{-1} (x_i' t)^2 &\leq (\alpha(1-\alpha))^{-1/2} \sigma_\alpha^{-1} \int_0^{\epsilon_{ni}} (\epsilon_{ni} - z) \int_0^z (|F^{-1}(\alpha)| + u) f(F^{-1}(\alpha)) du dz \\ &\leq c(\alpha(1-\alpha))^{-1/2-2a} n^{-3/2} |x_i' t|^3 + c(\alpha(1-\alpha))^{-1-2a} n^{-2} |x_i' t|^4 \end{aligned} \quad (3.21)$$

and we get the same inequality for $\epsilon_{ni} < 0$. The same expressions are $O(n^{-3/2} |x_i' t|^3) + O(n^{-2} |x_i' t|^4)$ if $\alpha_0 \leq \alpha \leq 1 - \alpha_0$. Hence,

$$\sum_{i=1}^n ER_i(t, \alpha) + \frac{1}{2} t' D_n t = O\left(n^{\frac{-2(b-a)}{1+4b}}\right). \quad (3.22)$$

Similarly,

$$\begin{aligned}
\sum_{i=1}^n \text{Var} R_i(t, \alpha) &\leq \sum_{i=1}^n E R_i^2(t, \alpha) \\
&\leq \left[\frac{f(F^{-1}(\alpha))}{\alpha(1-\alpha)} \right]^2 \sum_{i=1}^n \int_0^{|\epsilon_{ni}|} (|\epsilon_{ni}| - y)^2 f(F^{-1}(\alpha) + y) dy \\
&= O(n^{-1/2} (\alpha(1-\alpha))^{1/2} / f(F^{-1}(\alpha))) = O(n^{\frac{-(2b-a)}{1+4b}}).
\end{aligned} \tag{3.23}$$

These results hold uniformly in α, t .

Hence, using (3.15) and (3.13) with $u = \log n / B_n = O(n^{2a/(1+4b)})$, so that 3.20 holds,

$$\begin{aligned}
P(|r_n(t, \alpha)| \geq (\lambda+1)B_n) &\leq \exp \{-(\lambda+1)\log n + (K \log n / B_n) \cdot n^{\frac{-2(b-a)}{1+4b}} + (K \log^2 n / B_n^2) \cdot n^{\frac{-(2b-a)}{1+4b}}\} \\
&\leq n^{-\lambda}
\end{aligned} \tag{3.24}$$

for $n \geq n_0$ where $K > 0$ and n_0 do not depend on α and t .

(ii) Following the proof of Lemma A.2 in Koenker and Portnoy (1987), choose intervals $[\alpha_l, \alpha_{l+1}]$ of length $1/n^5$ covering $[\alpha_n^*, 1-\alpha_n^*]$ and balls of radius $1/n^5$ covering $\{t: \|t\| \leq C\}$.

Let $(\alpha_1, \alpha_2) \subset [\alpha_l, \alpha_{l+1}]$ and t_1, t_2 lie in one of the balls covering $\{t: \|t\| \leq C\}$. For $i \notin \{l_1, l_2\}$ we use (3.17) (and the boundedness of f and f') to obtain

$$\begin{aligned}
|R_i(t_1, \alpha_1) - R_i(t_2, \alpha_2)| &\leq \frac{|f(F^{-1}(\alpha_1)) - f(F^{-1}(\alpha_2))|}{\alpha_1(1-\alpha_2)} |(F^{-1}(\alpha^*) + \epsilon_{ni})| \\
&\quad + |f(F^{-1}(\alpha^*) + \epsilon_{ni})F^{-1}(\alpha^*)| \left| \frac{1}{\alpha_1(1-\alpha_1)} - \frac{1}{\alpha_2(1-\alpha_2)} \right| \\
&\quad + \frac{f(F^{-1}(\alpha^*))}{\alpha_1(1-\alpha_2)} |F^{-1}(\alpha_1) - F^{-1}(\alpha_2)| \\
&\quad + n^{-1/2} (\alpha_1(1-\alpha_1))^{-1/2} \|x_i\| \|t_1 - t_2\|
\end{aligned}$$

$$+ Cn^{-1/2}\|x_i\| |(\alpha_1(1-\alpha_1))^{-1/2} - (\alpha_2(1-\alpha_2))^{-1/2}| = O(n^{-2});$$

here $\alpha^* \in (\alpha_l, \alpha_{l+1})$ and we have used the inequalities $|F^{-1}(\alpha_1) - F^{-1}(\alpha_2)| \leq \frac{|\alpha_1 - \alpha_2|}{f(F^{-1}(\alpha^*))}$

$$= O(n^{-5 + \frac{1+\alpha}{1+4b}}) = O(n^{-3.75}). \text{ Hence, on any ball in the covering, } |r_n(t_1, \alpha_1) - r_n(t_2, \alpha_2)| \leq K/n.$$

Since the number of sets needed to cover the set $S \equiv [\alpha_n^*, 1 - \alpha_n^*] \times \{t : \|t\| \leq C\}$ is bounded by $n^{5(p+1)}$ we obtain from (3.14) for $\lambda > 5(p+1)$

$$P\left(\sup_{(\alpha, t) \in S} |r_n(t, \alpha)| \geq (\lambda+1)B_n + K/n\right) \leq n^{5(p+1)}n^{-\lambda} \rightarrow 0 \quad \blacksquare$$

LEMMA 3.2. Assume the conditions of Lemma 3.1 and let $d_n = (d_{n1}, \dots, d_{nn})'$ be a sequence of vectors satisfying

$$X_n' d_n = 0, \quad \frac{1}{n} \sum_{i=1}^n d_{ni}^2 \rightarrow \Delta^2, \quad 0 < \Delta^2 < \infty \quad (\text{D.1})$$

$$n^{-1} \sum_{i=1}^n |d_{ni}|^3 = O(1) \quad \text{as } n \rightarrow \infty \quad (\text{D.2})$$

$$\max_{1 \leq i \leq n} |d_{ni}| = O\left(n^{(2(b-a)-\delta)/(1+4b)}\right) \quad (\text{D.3})$$

then

$$\sup_{(\alpha, t) \in S} \{(\alpha(1-\alpha))^{-1/2} n^{-1/2} \sum_{i=1}^n |d_{ni} [\psi_\alpha(E_{i\alpha} - n^{-1/2} \sigma_\alpha x_i' t) - \psi_\alpha(E_{i\alpha}) + n^{-1/2} \sigma_\alpha x_i' t]| \} \xrightarrow{P} 0 \quad (3.25)$$

as $n \rightarrow \infty$ for any fixed $C > 0$ and for α_n^* given in (3.8).

PROOF. Consider the model

$$Y = X^* \beta^* + E \quad (3.26)$$

where $X^* = (X_n : d_n)$, $\beta^* = (\beta_1, \dots, \beta_p, \beta_{p+1})$. Then

$$\mathbf{X}^*{}'\mathbf{X}^* = \begin{pmatrix} \mathbf{X}_n'\mathbf{X}_n & 0 \\ 0 & \mathbf{d}_n'\mathbf{d}_n \end{pmatrix}$$

and the conditions of Lemma 3.1 are satisfied even when replacing \mathbf{X} by \mathbf{X}^* . Computing the right derivative of (3.11) with respect to $\mathbf{t} \in \mathbf{R}_{p+1}$, we arrive at (3.25). ■

Let $\hat{\beta}_n(\alpha)$ be the α -regression quantile corresponding to the model (2.1) with the design matrix of order $(n \times p)$; i.e., $\hat{\beta}_n(\alpha)$ is a solution of the minimization

$$\sum_{i=1}^n \rho_\alpha(Y_i - \mathbf{x}_i'\mathbf{t}) := \min, \quad \mathbf{t} \in \mathbf{R}^p. \quad (3.27)$$

The following theorem establishes the rate of consistency of regression quantiles, and is needed for the representation of the dual process.

THEOREM 3.1. Under the conditions (F.1) - (F.4) and (X.1) - (X.4),

$$n^{1/2}\sigma_\alpha^{-1}(\hat{\beta}_n(\alpha) - \beta(\alpha)) = n^{1/2}(\alpha(1-\alpha))^{-1/2}\mathbf{D}_n \sum_{i=1}^n \psi_\alpha(E_{i\alpha}) + o_p(1) \quad (3.28)$$

uniformly in $\alpha_n^* \leq \alpha \leq 1-\alpha_n^*$. Consequently,

$$\sup_{\alpha_n^* \leq \alpha \leq 1-\alpha_n^*} \|n^{1/2}\sigma_\alpha^{-1}(\hat{\beta}_n(\alpha) - \beta(\alpha))\| = O_p(1). \quad (3.29)$$

PROOF. If $\hat{\beta}_n(\alpha)$ minimizes (3.27), then

$$\mathbf{T}_{n\alpha} = n^{1/2}\sigma_\alpha^{-1}(\hat{\beta}_n(\alpha) - \beta(\alpha)) \quad (3.30)$$

minimizes the convex function

$$G_{n\alpha}(\mathbf{t}) = (\alpha(1-\alpha))^{-1/2}\sigma_\alpha^{-1} \sum_{i=1}^n [\rho_\alpha(E_{i\alpha} - n^{-1/2}\sigma_\alpha \mathbf{x}_i'\mathbf{t}) - \rho_\alpha(E_{i\alpha})] \quad (3.31)$$

with respect to $\mathbf{t} \in \mathbf{R}^p$. By Lemma 3.1, for any fixed $C > 0$

$$\min_{\|\mathbf{t}\| < C} G_{n\alpha}(\mathbf{t}) = \min_{\|\mathbf{t}\| < C} \{-\mathbf{t}'\mathbf{Z}_{n\alpha} + \frac{1}{2}\mathbf{t}'\mathbf{D}_n\mathbf{t}\} + o_p(1) \quad (3.32)$$

uniformly in $\alpha_n^* \leq \alpha \leq 1 - \alpha_n^*$, where

$$Z_{n\alpha} = n^{-1/2}(\alpha(1-\alpha))^{-1/2} \sum_{i=1}^n x_i \psi_\alpha(E_{i\alpha}). \quad (3.33)$$

Denoting

$$U_{n\alpha} = \arg \min_{t \in R^p} \{-t' Z_{n\alpha} + \frac{1}{2} t' D_n t\}, \quad (3.34)$$

we immediately get

$$U_{n\alpha} = D_n^{-1} Z_{n\alpha} = O_p(1) \quad (3.35)$$

uniformly in $\alpha_n^* \leq \alpha \leq 1 - \alpha_n^*$ and

$$\min_{t \in R^p} \{-t' Z_{n\alpha} + \frac{1}{2} t' D_n t\} = -\frac{1}{2} Z_{n\alpha}' D_n^{-1} Z_{n\alpha}. \quad (3.36)$$

From (3.35) and (3.36), we can write

$$t' Z_{n\alpha} + \frac{1}{2} t' D_n t = \frac{1}{2} \{(t - U_{n\alpha})' D_n (t - U_{n\alpha}) - U_{n\alpha}' D_n U_{n\alpha}\} \quad (3.37)$$

and hence we could rewrite (3.10) in the form

$$\sup_{(\alpha, t) \in S} |r_n(t, \alpha)| = \sup_{(\alpha, t) \in S} \{G_{n\alpha}(t) - \frac{1}{2} [(t - U_{n\alpha})' D_n (t - U_{n\alpha}) - U_{n\alpha}' D_n U_{n\alpha}]\} \xrightarrow{p} 0. \quad (3.38)$$

Inserting $U_{n\alpha} = O_p(1)$, for t , we further obtain

$$\sup_{\alpha_n^* \leq \alpha \leq 1 - \alpha_n^*} \{ |G_{n\alpha}(U_{n\alpha}) + \frac{1}{2} U_{n\alpha}' D_n U_{n\alpha} | \} = o_p(1). \quad (3.39)$$

We would like to show that

$$\sup_{\alpha_n^* \leq \alpha \leq 1 - \alpha_n^*} \{ \|T_{n\alpha} - U_{n\alpha}\| \} = o_p(1). \quad (3.40)$$

Consider the ball $B_{n\alpha}$ with center $U_{n\alpha}$ and radius $\delta > 0$. This ball lies in a compact set with probability exceeding $(1 - \epsilon)$ for $n \geq n_0$; actually, for $t \in B_{n\alpha}$,

$$\|t\| \leq \|t - U_{n\alpha}\| + \|U_{n\alpha}\| \leq \delta + K_1$$

for some K_1 with probability exceeding $1 - \epsilon$ for $n \geq n_0$. Hence, by (3.10),

$$\Delta_{n\alpha} = \sup_{\alpha_n^* \leq \alpha \leq 1 - \alpha_n^*} \sup_{t \in B_{n\alpha}} |r_n(t, \alpha)| \xrightarrow{P} 0. \quad (3.41)$$

Following Pollard (1988), consider the behavior of $G_{n\alpha}(t)$ outside $B_{n\alpha}$. Suppose $t_\alpha = U_{n\alpha} + k\xi$, $k > \delta$ and $\|\xi\| = 1$. Let t_α^* be the boundary point of $B_{n\alpha}$ that lies on the line from $U_{n\alpha}$ to t_α , i.e., $t_\alpha^* = U_{n\alpha} + \delta\xi$. Then $t_\alpha^* = (1 - (\delta/k))U_{n\alpha} + (\delta/k)t_\alpha$ and hence, by (3.38) and (3.39),

$$\delta/k G_{n\alpha}(t) + (1 - \delta/k) G_{n\alpha}(U_{n\alpha}) \geq G_{n\alpha}(t_\alpha^*) \geq \frac{1}{2} \delta^2 \lambda_0 + G_{n\alpha}(U_{n\alpha}) - 2\Delta_{n\alpha}$$

where λ_0 is the minimal eigenvalue of \mathbf{D} . Hence,

$$\inf_{\|t - U_{n\alpha}\| \geq \delta} G_{n\alpha}(t) \geq G_{n\alpha}(U_{n\alpha}) + (k/\delta)(\frac{1}{2} \delta^2 \lambda_0 - 2\Delta_{n\alpha}). \quad (3.42)$$

Using (3.39) the last term is positive with probability tending to one uniformly in α for any fixed $\delta > 0$. Hence, given $\delta > 0$ and $\epsilon > 0$, there exist n_0 and $\eta > 0$ such that for $n \geq n_0$,

$$P(\inf_{\alpha_n^* \leq \alpha \leq 1 - \alpha_n^*} \int_{\|t - U_{n\alpha}\| \geq \delta} G_{n\alpha}(t) - G_{n\alpha}(U_{n\alpha}) > \eta) > 1 - \epsilon \quad (3.43)$$

and hence (since the event in (3.43) implies that $G_{n\alpha}$ must be minimized inside the ball of radius δ) $P(\sup_{\alpha_n^* \leq \alpha \leq 1 - \alpha_n^*} \|T_{n\alpha} - U_{n\alpha}\| \leq \delta) \rightarrow 1$ for any fixed $\delta > 0$, as $n \rightarrow \infty$. ■

The following theorem approximates the regression rank score process by an empirical process.

THEOREM 3.2. Let d_n satisfy (D.1) - (D.3), X_n satisfy (X.1) - (X.4) and F satisfy (F.1) - (F.4). Then

$$\sup_{\alpha_n^* \leq \alpha \leq 1 - \alpha_n^*} \{ |n^{-1/2}(\alpha(1-\alpha))^{-1/2} \sum_{i=1}^n d_{ni}(\hat{a}_{ni}(\alpha) - \tilde{a}_i(\alpha))| \} \xrightarrow{P} 0 \quad (3.44)$$

as $n \rightarrow \infty$, where

$$\tilde{a}_i(\alpha) = I[E_i \geq F^{-1}(\alpha)], \quad i = 1, \dots, n. \quad (3.45)$$

PROOF. Insert $n^{1/2} \sigma_\alpha^{-1}(\hat{\beta}_n(\alpha) - \beta(\alpha))$ for t in (3.25) and notice (3.29) and the fact that

$$\sup_{\alpha_n \leq \alpha \leq 1-\alpha_n} \{n^{-1/2}(\alpha(1-\alpha))^{-1/2} \sum_{i=1}^n d_{ni} I[Y_i = x_i, \hat{\beta}(\alpha)]\} \xrightarrow{P} 0, \quad (3.46)$$

from which (3.44) follows. ■

The following theorem which follows from Theorem 3.2 is an extension of Theorem V.3.5 in Hájek and Šidák (1967) to the regression rank scores. Some applications of this result to Kolmogorow-Smirnov type tests will appear in Jurečková(1990).

THEOREM 3.3. Under the conditions of Theorem 3.2, as $n \rightarrow \infty$,

$$\sup_{0 \leq \alpha \leq 1} \{ |n^{-1/2} \sum_{i=1}^n d_{ni} (\hat{a}_{ni}(\alpha) - \tilde{a}_{ni}(\alpha))| \} \xrightarrow{P} 0 \quad (3.47)$$

Moreover, the process

$$\{ \Delta^{-1} n^{-1/2} \sum_{i=1}^n d_{ni} \hat{a}_{ni}(\alpha) : 0 \leq \alpha \leq 1 \} \quad (3.48)$$

converges to the Brownian bridge in the Prokhorov topology on $C[0, 1]$.

PROOF. By Theorem 3.2,

$$\sup_{\alpha_n \leq \alpha \leq 1-\alpha_n} |n^{-1/2} \sum_{i=1}^n d_{ni} (\hat{a}_{ni}(\alpha) - \tilde{a}_{ni}(\alpha))| \xrightarrow{P} 0. \quad (3.49)$$

Further,

$$\begin{aligned} \sup_{0 \leq \alpha \leq \alpha_n} |n^{-1/2} \sum_{i=1}^n d_{ni} \hat{a}_{ni}(\alpha)| &= \sup_{0 \leq \alpha \leq \alpha_n} |n^{-1/2} \sum_{i=1}^n d_{ni} (1 - \hat{a}_{ni}(\alpha))| \leq n^{1/2} \max_{1 \leq i \leq n} |d_{ni}| \alpha_n^\delta \\ &= O \left(n^{1/2 + \frac{2(b-\alpha)-\delta}{1+4b} - \frac{1}{1+4b}} \right) = O(n^{-2\delta}) \end{aligned} \quad (3.50)$$

and we obtain an analogous conclusion for $\sup_{1-\alpha_n \leq \alpha \leq 1} |n^{-1/2} \sum_{i=1}^n d_{ni} \hat{a}_{ni}(\alpha)|$. On the other hand,

$$\sup_{0 \leq \alpha \leq \alpha_n} |n^{-1/2} \sum_{i=1}^n d_{ni} \tilde{a}_{ni}(\alpha)| = \sup_{0 \leq \alpha \leq \alpha_n} |n^{-1/2} \sum_{i=1}^n d_{ni} (I[E_i < F^{-1}(\alpha)] - \alpha)|$$

$$\leq \max_{1 \leq i \leq n} |d_{ni}| \cdot O_p(\alpha_n^*(1-\alpha_n^*))^{1/2} = o_p(1) \quad (3.51)$$

and analogously

$$\sup_{1-\alpha_n^* \leq \alpha \leq 1} |n^{-1/2} \sum_{i=1}^n d_{ni} \tilde{a}_i(\alpha)| = o_p(1).$$

Thus (3.47) follows, and consequently (3.48). ■

4. Asymptotic properties of simple linear regression rank scores statistics

Maintaining the notation of Section 3, let $\phi(t): 0 < t < 1$ be a nondecreasing square-integrable score-generating function and let $\hat{b}_{ni}, i=1, \dots, n$ be the scores defined by (2.14). Let $\{d_n\}$ be a sequence of vectors satisfying (D.1) - (D.3).

Following Hájek and Šidák (1967), we shall call the statistics

$$S_n = n^{-1/2} \sum_{i=1}^n d_{ni} \hat{b}_{ni} \quad (4.1)$$

the *simple linear regression rank-score statistics*, or just *simple linear rank statistics*. Our primary objective in this section is to investigate the conditions on ϕ under which we may integrate (3.47) and obtain an asymptotic representation for S_n of the form

$$S_n = n^{-1/2} \sum_{i=1}^n d_{ni} \phi(F(E_i)) + o_p(1). \quad (4.2)$$

We shall prove (4.2) for ϕ satisfying a condition of the Chernoff-Savage(1958) type; thus our results will cover Wilcoxon, van der Waerden, and median scores, among others.

THEOREM 4.1. Let $\phi(t): 0 < t < 1$, be a nondecreasing square integrable function such that $\phi'(t)$ exists for $0 < t < \alpha_0$, $1-\alpha_0 < t < 1$ and satisfies

$$|\phi'(t)| \leq c(t(1-t))^{-1-\delta^*} \quad (4.3)$$

for some $\delta^* < \delta$ where δ is given in condition (X.4), and for $t \in (0, \alpha_0) \cup (1 - \alpha_0, 1)$. Then, under (F.1) - (F.4), (X.1) - (X.4) and (D.1) - (D.3), the statistic S_n admits the representation (4.2) and hence is asymptotically normally distributed with zero expectation and with variance

$$\Delta^2(\int_0^1 \phi^2(t)dt - \bar{\phi}^2), \quad \bar{\phi} = \int_0^1 \phi(t)dt. \quad (4.4)$$

PROOF. Let us consider S_n defined in (4.1) with the scores (2.14). Integrating by parts (notice that $\hat{a}_{ni}(t) - \tilde{a}_i(t) = 0$ for $t = 0, 1$), we obtain

$$-n^{-1/2} \sum_{i=1}^n d_{ni} \int_0^1 \phi(t) d(\hat{a}_{ni}(t) - \tilde{a}_i(t)) = n^{-1/2} \sum_{i=1}^n d_{ni} \int_0^1 (\hat{a}_{ni}(t) - \tilde{a}_i(t)) d\phi(t). \quad (4.5)$$

which we must show is $o_p(1)$. We shall split the domain of integration into the intervals $(0, \alpha_n^*)$, (α_n^*, α_0) , $[\alpha_0, 1 - \alpha_0]$, $(1 - \alpha_0, 1 - \alpha_n^*)$, $[1 - \alpha_n^*, 1)$ and denote the respective integrals by I_1, \dots, I_5 . Regarding Theorem 3.2, we immediately get that $I_3 \xrightarrow{P} 0$ by the dominated convergence theorem. Similarly,

$$\begin{aligned} |I_2| &\leq \int_{\alpha_n^*}^{\alpha_0} |\phi'(t)| |n^{-1/2} \sum_{i=1}^n d_{ni} (\hat{a}_{ni}(t) - \tilde{a}_i(t))| dt \\ &\leq c \int_{\alpha_n^*}^{\alpha_0} (t(1-t))^{-1+\delta^*} (t(1-t))^{1/2} \cdot |n^{-1/2} (t(1-t))^{-1/2} \sum_{i=1}^n d_{ni} (\hat{a}_{ni}(t) - \tilde{a}_i(t))| dt \\ &= c \int_{\alpha_n^*}^{\alpha_0} (t(1-t))^{\delta^*-1/2} dt \cdot o_p(1) = o_p(1). \end{aligned}$$

Finally,

$$|I_1| \leq n^{-1/2} \max_{1 \leq i \leq n} |d_{ni}| \int_0^{\alpha_n^*} |\phi'(t)| \sum_{i=1}^n |\hat{a}_{ni}(t) - \tilde{a}_i(t)| dt \leq I_{11} + I_{12}$$

where

$$I_{11} = n^{-1/2} \max_{1 \leq i \leq n} |d_{ni}| \int_0^{\alpha_n^*} |\phi'(t)| \sum_{i=1}^n (1 - \hat{a}_{ni}(t)) dt \quad (4.6)$$

and

$$I_{12} = n^{-1/2} \max_{1 \leq i \leq n} |d_{ni}| \int_0^{\alpha_n^*} |\phi'(t)| \sum_{i=1}^n (1 - \tilde{\alpha}_i(t)) dt. \quad (4.7)$$

Then

$$I_{11} \leq n^{1/2} \max_{1 \leq i \leq n} |d_{ni}| \int_0^{\alpha_n^*} t^{-\delta^*} dt = O(n^{1/2 + \frac{2(b-a)-\delta}{1+4b} - \frac{(1-\delta^*)}{1+4b}}) = O(n^{-2(\delta-\delta^*)}).$$

Finally,

$$I_{12} = n^{-1/2} \sum_{i=1}^n d_{ni} \int_0^{\alpha_n^*} \phi'(t) I[t > F(E_i)] dt = n^{-1/2} \sum_{i=1}^n d_{ni} [\phi(\alpha_n^*) - \phi(F(E_i))] I(F(E_i) < \alpha_n^*)$$

Now we may assume that $\phi(\alpha_n^*) < 0$ for $n \geq n_0$, since otherwise if ϕ were bounded from below

then $I_{12} \xrightarrow{P} 0$. Hence

$$\text{Var}(I_{12}) \leq n^{-1} \sum_{i=1}^n d_{ni}^2 E([2\phi(F(E_i))]^2 I[F(E_i) < \alpha_n^*]) \leq \int_0^{\alpha_n^*} \phi^2(u) du \cdot O(1) \rightarrow 0$$

due to the square-integrability of ϕ . Treating the integrals I_4, I_5 analogously, we arrive at (4.5) and this proves the representation (4.2). ■

5. Tests of linear subhypotheses based on regression rank scores

Returning to the model (2.2), assume that the design matrix $X = (X_1 : X_2)$ satisfies the conditions (X.1) - (X.4), (2.3) and (2.4). We want to test the hypothesis $H_0 : \beta_2 = 0$ (β_1 unspecified) against the alternative $H_n : \beta_{2n} = n^{-1/2} \beta_0$ ($\beta_0 \in R^q$ fixed).

Let $\hat{\alpha}_n(\alpha) = (\hat{\alpha}_{n1}(\alpha), \dots, \hat{\alpha}_{nn}(\alpha))$ denote the regression rank scores corresponding to the submodel

$$Y = X_1 \beta_1 + E \quad \text{under } H_0. \quad (5.1)$$

Let $\phi(t) : (0, 1) \rightarrow R^1$ be a nondecreasing and square integrable score-generating function.

Define the scores $\hat{b}_{ni}, i=1, \dots, n$ by the relation (2.14), and consider the test statistic

$$T_n = S_n' Q_n^{-1} S_n / A^2(\phi) \quad (5.2)$$

where

$$S_n = n^{-1/2}(X_{n2} - \hat{X}_{n2})' \hat{b}_n \quad (5.3)$$

and where Q_n and $A^2(\phi)$ are defined in (2.4) and (2.17), respectively. The test is based on the asymptotic distribution of T_n under H_0 , given in the following theorem. Thus, we shall reject H_0 provided $T_n \geq \chi_q^2(\omega)$, i.e. provided T_n exceeds the ω critical value of the χ^2 distribution with q d.f. The same theorem gives the asymptotic distribution of T_n under H_n and thus shows that the Pitman efficiency of the test coincides with that of the classical rank test.

THEOREM 5.1. Assume that X_1 satisfies (X.1) - (X.4) and $(X_1 : X_2)$ satisfies (2.3) and (2.4). Further assume that F satisfies (F.1) - (F.4). Let T_n defined in (5.3) and (5.4) be generated by the score function ϕ satisfying (4.3), and nondecreasing and square-integrable on $(0, 1)$.

- (i) Then, under H_0 , the statistic T_n is asymptotically central χ^2 with q degrees of freedom.
- (ii) Under H_n , T_n is asymptotically noncentral χ^2 with q degrees of freedom and with non-centrality parameter,

$$\eta^2 = \beta_0' Q \beta_0 \cdot \gamma^2(\phi, F) / A^2(\phi) \quad (5.4)$$

with

$$\gamma(\phi, F) = -\int_0^1 \phi(t) df(F^{-1}(t)). \quad (5.5)$$

PROOF.

- (i) It follows from Theorem 4.1 that, under H_0 , S_n has the same asymptotic distribution as

$$\tilde{S}_n = n^{-1/2}(X_{n2} - \hat{X}_{n2})' \tilde{b}_n$$

where $\tilde{b}_n = (\tilde{b}_{n1}, \dots, \tilde{b}_{nm})'$ and $\tilde{b}_{ni} = \phi(F(E_i))$, $i=1, \dots, n$. The asymptotic distribution of \tilde{S}_n follows from the central limit theorem and coincides with q -dimensional normal distribution with expectation 0 and the covariance matrix $Q \cdot A^2(\phi)$.

- (ii) The sequence of local alternatives H_n is contiguous with respect to the sequence of null distributions with the densities $\{\prod_{i=1}^n f(e_i)\}$. Hence, (4.1) holds also under H_n and the asymptotic distributions of \tilde{S}_n under H_n coincide. The proposition then follows from the fact that the asymptotic distribution of \tilde{S}_n under H_n is normal $N_q(\gamma(\phi, F)Q\beta_0, QA^2(\phi))$.

■

6. An Example

To illustrate the tests proposed above we consider briefly an example taken from Adichie (1984, Example 3). The log of the leaf burn (in seconds) of 30 batches of tobacco is thought to depend upon the percent composition of nitrogen, chlorine, and potassium. Adichie suggests testing the potassium effect and describes an aligned rank version of the test. We are unable to reproduce some details of his calculations, however, using his approach we get least squares estimates of the nitrogen and chlorine effects of -.529 and -.290 with an intercept of 2.653. With these preliminary estimates we obtain aligned ranks

7	17	2	18	6	1	11	3	30	13
25	16	4	29	26	27	21	23	19	12
28	10	8	15	24	20	22	5	14	9

which yield a test statistic of 13.59 highly significant relative to the 1% χ_1^2 critical value of 6.63.

In contrast the Wilcoxon regression rank scores computed as

$$\hat{b}_i = -\int_0^1 (t - 1/2) d\hat{a}_i(t) = \int_0^1 \hat{a}_i(t) dt - 1/2$$

and based on the restricted model excluding potassium, are

-0.27	0.06	-0.41	0.09	-0.32	-0.48	-0.17	-0.38	0.48	-0.06
0.23	0.04	-0.37	0.42	0.28	0.37	0.19	0.41	0.15	-0.26
0.38	-0.16	-0.23	-0.01	0.33	0.12	0.15	-0.42	-0.10	-0.06

and yield a test statistic of 13.17. The full set of regression rank scores $\hat{a}_i(t)$ for this data are illustrated in Figure 6.1 with the plots ordered according to their Wilcoxon rank score. Note

that as a practical matter when $\bar{\phi} = \int_0^1 \phi(t)dt = 0$, we may omit the \hat{X}_2 term in the computation of S_n in (5.3) since \hat{b}_n is orthogonal to X_1 . This is in contrast with the aligned rank situation where the use of $X_2 - \hat{X}_2$ is essential.

Corresponding calculations for the normal scores using

$$\hat{b}_i = -\int_0^1 \Phi^{-1}(t) d\hat{a}_i(t) = \sum_{j=1}^{J_n} \hat{a}_i'(t_j) [\phi(\Phi^{-1}(t_j)) - \phi(\Phi^{-1}(t_{j-1}))]$$

yields

-0.74	0.15	-1.41	0.23	-0.91	-2.13	-0.45	-1.17	2.08	-0.15
0.63	0.10	-1.25	1.44	0.78	1.15	0.50	1.35	0.40	-0.72
1.41	-0.40	-0.61	-0.03	0.94	0.30	0.39	-1.45	-0.26	-0.18

and a test statistic of 12.87. The corresponding normal score aligned rank statistic is 11.72.

Finally, regression rank score version of the sign test yields the scores

-1.00	1.00	-1.00	1.00	-1.00	-1.00	-1.00	-1.00	1.00	-1.00
1.00	1.00	-1.00	1.00	1.00	1.00	1.00	1.00	0.16	-1.00
1.00	-1.00	-1.00	-0.37	1.00	1.00	1.00	-1.00	-1.00	-0.79

and a test statistic of 8.42 while the aligned rank sign scores yield 10.20. Obviously, all versions of the tests lead to a decisive rejection of the null. Note that for the sign scores the test coincides with the l_1 Lagrange multiplier test discussed in Koenker and Bassett(1982).

Since an important objective of the proposed rank tests is robustness to outlying observations, it is interesting to observe the effect of perturbing the first y observation of the Adichie data set on the aligned and rank scores versions of the test statistic. This sensitivity analysis is illustrated in Figure 6.2. Even a modest perturbation in y_1 is enough to confound the initial least squares estimate and reverse the conclusion of the aligned rank test. However the regression rank score version of the test is seen to be relatively insensitive to such perturbations. One should be aware that comparable perturbations in the design observations may wreck havoc even with the rank score form of the test. Recent work of Antoch and Jurečková (1985) and deJongh, deWet, and Welsh (1988) contain suggestions on robustifying

regression quantiles to the effect of influential design points.

Computation of the tests was carried out in *S+* using the algorithm described in Koenker and d'Orey (1988, 1990) to compute regression quantiles.

REFERENCES

- Adichie, J.N. (1978). "Rank tests for sub-hypotheses in the general linear regression," *Am. Statist.*, 6, 1012-1026.
- Adichie, J.N. (1984). "Rank tests in linear models," in P.R. Krishnaiah and P.K. Sen (eds.), *Handbook of Statistics*, vol. 4, Elsevier, New York.
- Amemiya, T. (1985). *Advanced econometrics*, Harvard University Press, Cambridge.
- Antoch, J. and Jurečková, J. (1985). "Trimmed LSE resistant to leverage points," *Comp. Statist. Quarterly*, 4, 329-339.
- Aubuchon, J.C. and Hettmansperger, T.P. (1989). "Rank based inference for linear models: Asymmetric errors," *Statistics and Probability Letters*, 8, 97-107.
- Chernoff, H. and Savage, I.R. (1958). "Asymptotic normality and efficiency of certain non-parametric test statistics," *Ann. Math. Statist.*, 29, 972-994.
- Draper, D. (1988). "Rank-based robust analysis of linear models. I. Exposition and Review," *Statistical Science*, 3, 239-271.
- Gutenbrunner, C. (1986). Zur Asymptotik von Regression quantileprozessen und daraus abgeleiten Statistiken, Ph.D. Dissertation, Universität Freiburg.
- Gutenbrunner, C. and Jurečková, J. (1990). "Regression rank-scores and regression quantiles," *Ann. Statist* (to appear).
- Hájek, J. and Šidák, Z. (1967). Sidak, Z. (1967). "Theory of rank tests," *Academia*, Prague.
- deJongh, P.J., deWet, T. and Welsh, A.H. (1988). "Mallows type bounded-influence-regression trimmed means," *J. Amer. Statist. Assoc.*, 83, 805-810.
- Jurečková, J. (1971). "Nonparametric estimate of regression coefficients," *Ann. Math. Statist.*, 42, 1328-1338.
- Jurečková, J. and Sen, P.K. (1989). "Uniform second order asymptotic linearity of M -statistics in linear models," *Statistics & Decisions*, 7, 263-276.
- Koenker, R. and Bassett, G. (1978). "Regression quantiles," *Econometrica*, 46, 33-50.
- Koenker, R. and Bassett, G. (1982). "Tests of linear hypotheses and l_1 -estimation," *Econometrica*, 50, 1577-83.

- Koenker, R. and d'Orey, V. (1987). "Computing regression quantiles," *Applied Statistics*, 36, 383-393.
- Koenker, R. and d'Orey V. (1990). "Remark on algorithm 229," preprint.
- Koul, H.L. (1970). "A class of ADF tests for subhypotheses in the multiple linear regression," *Ann. Math. Statist.*, 41, 1273-1281.
- Lehmann, E.L. (1959). *Testing Statistical Hypotheses*, Wiley, New York.
- Lehmann, E.L. (1983). *Theory of Point Estimation*, Wiley, New York.
- McKean, J.W. and Hettmansperger, T.P. (1978). "A robust analysis of the general linear model based on one-step R -estimates," *Biometrika*, 65, 571-79.
- Pollard, D. (1988). "Asymptotics for least absolute deviation regression estimators," *Econometric Theory*, to appear.
- Puri, M.L. and Sen, P.K. (1985). *Nonparametric Methods in General Linear Models*. J. Wiley, New York.
- Ruppert, D. and Carroll, R.J. (1980). "Trimmed least squares estimation in the linear model," *J. Amer. Statist. Assoc.*, 75, 828-838.
- Scheffé, H. (1959). *The Analysis of Variance*, Wiley, New York.

Figure 6.1

Regression Rank Scores for Tobacco Data

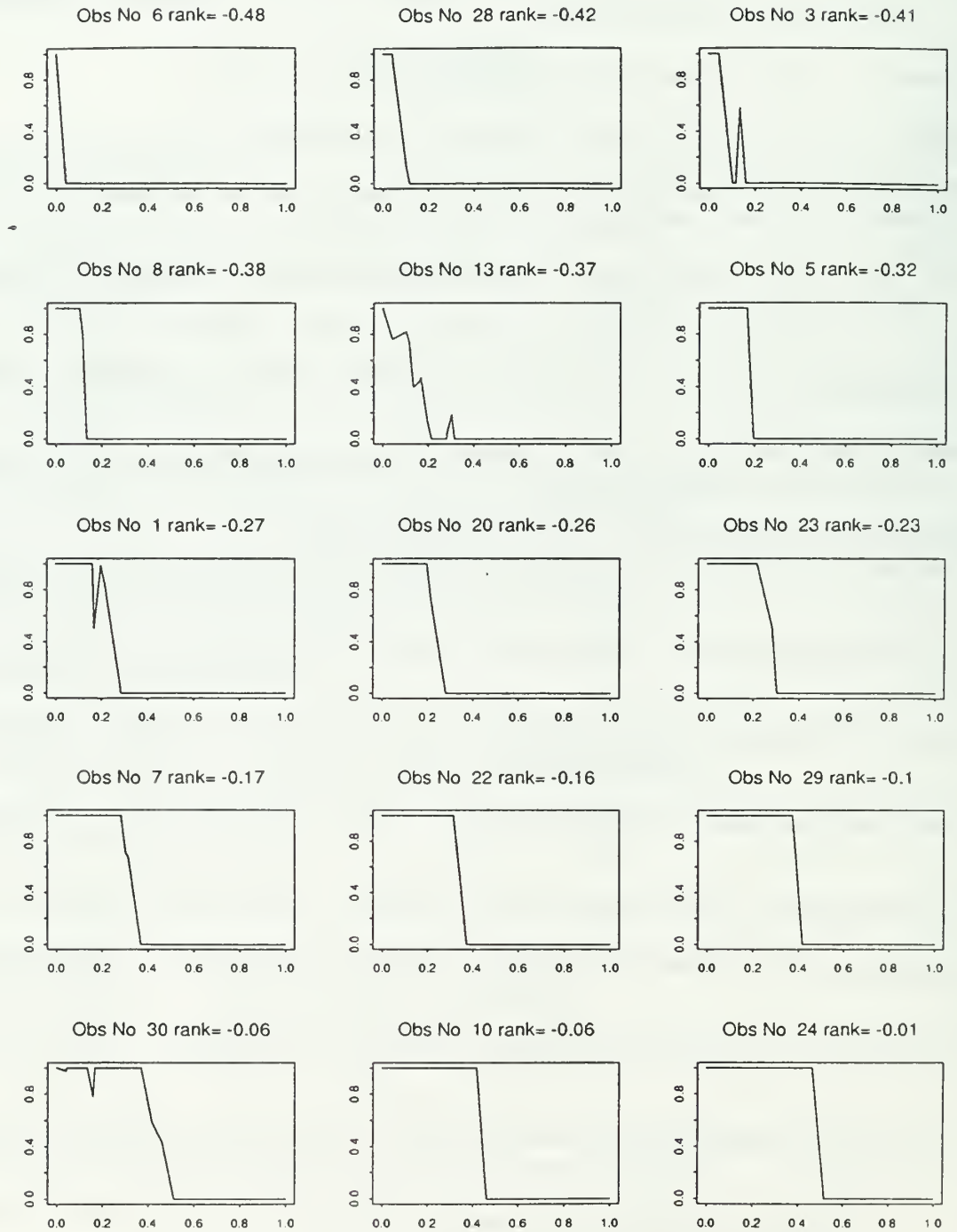


Figure 6.1
(continued)

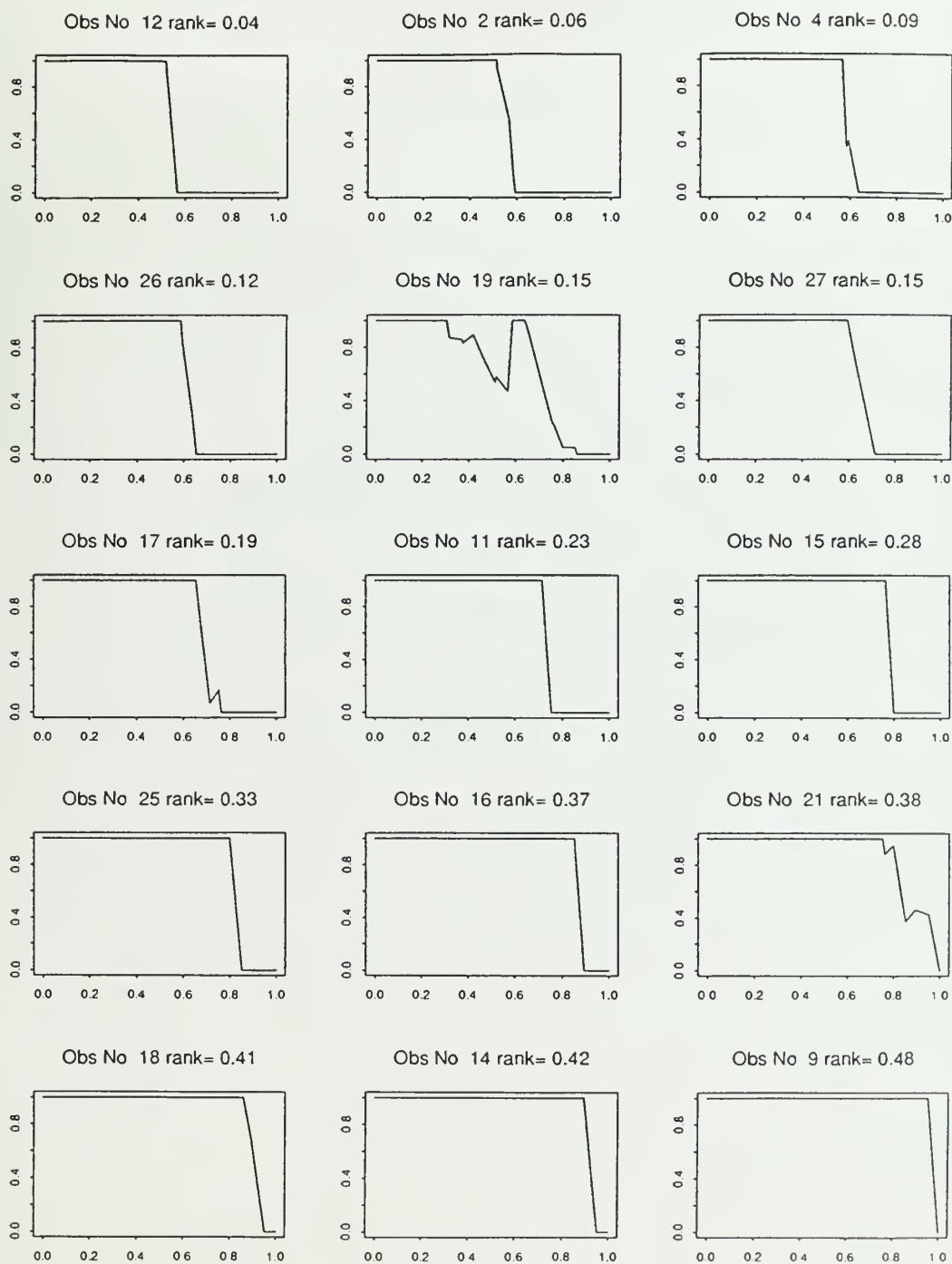
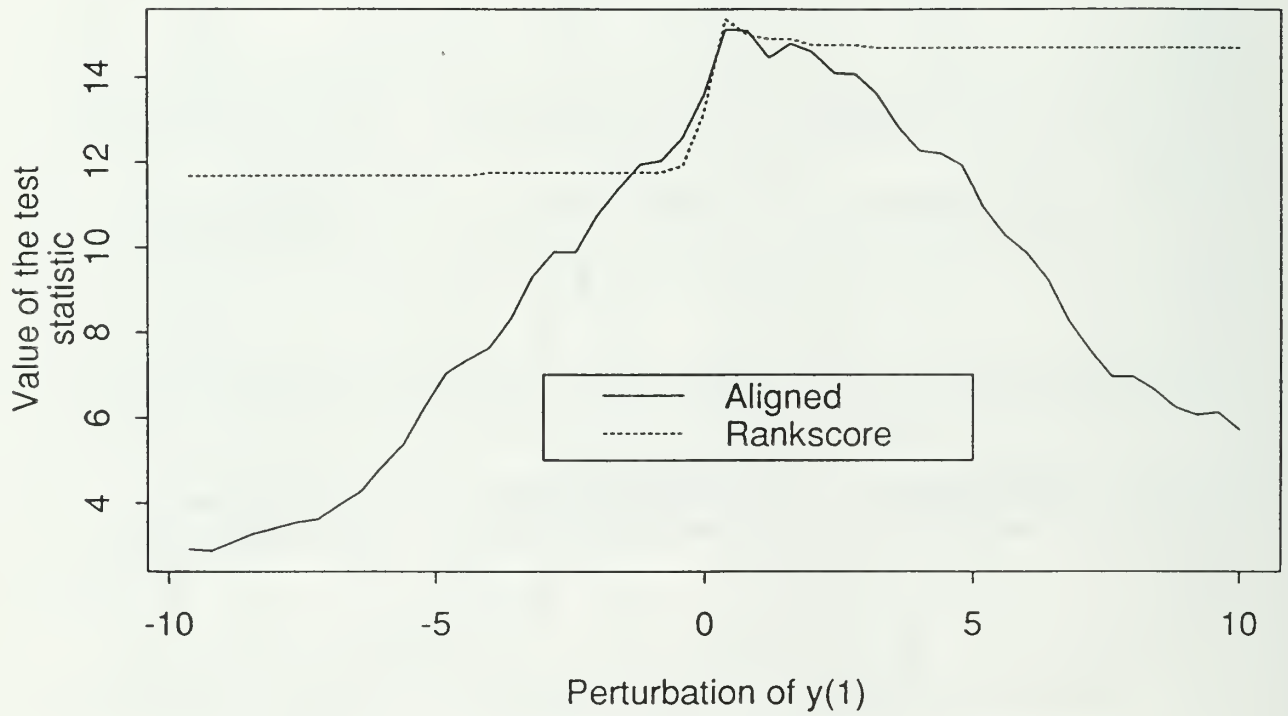


Figure 6.2

Sensitivity Curves for Rank Tests



HECKMAN
BINDERY INC.



JUN 95

Bound To Please® N. MANCHESTER,
INDIANA 46962

UNIVERSITY OF ILLINOIS-URBANA



3 0112 060295935