Rank Tests for Linear Models *

Roger Koenker

October 20, 1995

Abstract

This paper constitutes a brief, rather idiosyncratic, survey of rank tests stressing their connection in linear model applications to the theory of quantile regression through the formal duality of linear programming.

1 Introduction

Milton Friedman, who was present at the conception of rank tests in the late 1930's, but abandoned his progeny for more lucrative economic pursuits later in his career, left the idea of statistical inference based on ranks in the foster-care of a fortuitous sequence of statisticians who have nourished the infant through a robust adolescence. Friedman's (1937) paper "The use of ranks to avoid the assumption of normality implicit in the analysis of variance" together with the papers of Hotelling and Pabst(1936), and Kendall (1938) are usually credited with initiating the rank based approach to statistical inference. Spearman's(1904) paper is, of course, also fundamental. The appearance of Wilcoxon(1945) and Mann and Whitney(1948) on tests for location shift based on ranks, and the subsequent analysis of these tests by Hodges and Lehmann (1956) and Chernoff and Savage (1958) firmly established the subject as a precocious challenger to classical likelihood based methods of inference. The rigorous reformulation of the asymptotic theory of rank tests introduced by Hájek and developed by Hájek and Šidák (1967) and others,

*This research has been partially supported by NSF Grant SBR-9320555.

in conjunction with the emergence of robustness as a major theme of statistical research in the late 1960's gave a significant impetus to the growth of rank based methods. Work by Jurečková , Jaeckel, Hettmansperger, Adichie, Puri, Sen and others yielded important extensions of rank based inference *and estimation* methods for the linear model.

Nevertheless, like Friedman, econometricians have, for the most part, resisted the allure of rank tests. Of the 355 citations containing the phrase "rank tests" in the *Current Index to Statistics* from 1975 to 1993, none appear in an econometric journal. Nor does the phrase make an appearance in any of the econometrics texts which happen to grace my bookshelves. It may, therefore, appear quixotic to prepare a paper on this subject for an econometric audience. But econometrics has, in recent years, enthusiastically embraced other aspects of nonparametric statistics, and to a lesser degree shown a willingness to consider robustness as an important attribute of statistical procedures, so the time may finally be right for us to recognize an econometric orphan who has wandered so long in the statistical wilderness.

There are many excellent treatments of the vast literature on rank tests ranging from elementary textbooks like Mosteller and Rourke(1973) to more advanced texts like Hájek (1969) and Lehmann(1975) as well as the important monographs of Hájek and Šidák (1967) and Hettmansperger (1984). In addition there are excellent surveys on aligned rank tests for linear model applications by Adichie (1984) and the important monograph by Puri and Sen (1985). My own interest in rank-based inference was stimulated by the thesis work of Cornelius Gutenbrunner (1986) undertaken under the direction of Jana Jurečková . Gutenbrunner's research established an intimate link between the Hájek and Šidák approach to linear rank statistics and formulation of quantile regression appearing in Koenker and Bassett (1978). This link, which is based on a formal linear programming duality between sorting and ranking, will be central focus of this survey. It affords a unified perspective on the construction of rank tests for a broad spectrum of linear model applications.

1.1 An Example

As an introduction to rank tests in an elementary setting, I would like to begin by comparing the performance of the Wilcoxon-Mann-Whitney test for location shift with the classical two sample t-test. Let $X_1, ..., X_n$ be a random sample of "control" observations from the distribution function

F(x) and $Y_1, ..., Y_n$ denote a random sample of "treatment" observations from $F(x - \theta)$. Suppose we wish to test the hypothesis of "no treatment effect" $H_0 : \theta = 0$ against the alternative $H_1 : \theta > 0$.

If F were known to be Gaussian, we would immediately compute the sample means $\bar{X}_n$ and $\bar{Y}_m$ and then compute the test statistic

$$T = (\bar{X}_n - \bar{Y}_m)/s$$

where $s^2 = \sigma^2(n^{-1} + m^{-1})$, replacing σ^2 by

$$\hat{\sigma}^2 = (n + m - 2)^{-1}[\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + \sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2]$$

if necessary. If T exceeded $\Phi^{-1}(1 - \alpha)$ we would reject H_0 at level α, if σ^2 were known, replacing the normal critical value by its corresponding t_{n+m-2} value if σ^2 needed to be estimated.

Mann and Whitney's(1947) proposed alternative to the two sample t-test is based on the statistic,

$$S_1 = \sum_{i=1}^{n}\sum_{j=1}^{m} I(Y_i > X_j).$$

We simple count the number of pairs of observations – one from each sample – for which the treatment observation exceeds the control observation. If S_1 is large it suggests that treatment observations are generally larger than controls, and H_0 should be rejected.

How is this connected to ranks? Wilcoxon(1945) suggests an alternative formulation: pool the two samples and compute the rank of each observation in the pooled sample. Let S_0 denote the sum of the treatment observation ranks, again if S_0 is large it suggests rejection of H_0. It is easy to show that

$$S_1 = S_0 + m(m + 1)/2$$

so the Mann-Whitney and Wilcoxon forms of the test are actually equivalent. One of the most attractive features of this test is that its null distribution is independent of the form of F generating the original observations. To see this, let $Z_1, ..., Z_n$ denote $X_1, ..., X_n$ and $Z_{n+1}, ..., Z_{n+m}$ denote $Y_1, ..., Y_m$. Under H_0, the Z_i are iid. Thus for any permutation $(i_1, ..., i_{n+m})$ of $(1, ..., n + m)$, $(Z_{i_1}, ..., Z_{i_{n+m}})$ has the same distribution as $(Z_1, ..., Z_{n+m})$. So, in the absence of "ties"

$$\{R_1 = r_1, ..., R_{n+m} = r_{n+m}\} = \{Z_{i_1} < ... < Z_{i_{n+m}}\}$$

where R_i is the rank of the ith observation in the pooled sample and $r_{i_j} \equiv j$. It follows that the $(n+m)!$ possible events $\{R_1 = r_1, ..., R_{n+m} = r_{n+m}\}$ are equally likely. Since S_0 is a function solely of these ranks its distribution under H_0 is also independent of F. For modest n, m we can compute exact critical values for the test based on combinatorial considerations, see *e.g.* Mosteller and Rourke(1973) or Lehmann (1975) for details. For large n, m we can rely on the fact that under H_0,

$$\tilde{S}_1 = (S_1 - \mu_1)/\sigma_1 \rightsquigarrow \mathcal{N}(0,1).$$

where $\mu_1 = mn/2$ and $\sigma_1^2 = nm(n+m-2)/12$. Even for modest sample sizes this approximation is quite good.

It is tempting to think that the main reason for preferring the Wilcoxon test to the t-test is the fact that it has a guaranteed probability of Type I error for quite arbitrary distributions F, while the exact theory of the t-statistic depends on the normality of the observations. However, as long as F has a finite second moment and n and m are reasonably large, the critical values of the t distribution are also reasonably accurate for the t statistic. *It is power considerations that constitute the most compelling case for the Wilcoxon test under non-Gaussian conditions.*

Exact power comparisons are rather impractical for small sample sizes, but asymptotic results are very revealing. In the late 1940's Pitman proposed considering sequences of local alternatives of the form

$$H_n : \theta_n = \theta_0/\sqrt{n}$$

and showed that at the normal model $F = \Phi$, the limiting ratio of sample sizes required to achieve the same size and power with the Wilcoxon and the t-test is $3/\pi \approx .955$. This ratio, which is usually referred to as the asymptotic relative efficiency, or ARE, of the two tests provides a natural measure of their relative performance. If, for example, at the normal model 1000 observations are required to achieve power .95 at level .05 for a given alternative with the Wilcoxon test, this implies that the same power and level would be achievable with the t-test with roughly 955 observations. So transforming to ranks has "wasted" about 5 percent of the observations. This loss of information is hardly surprising since the optimality of the t-test at the normal model is a cornerstone of statistical thinking.

What happens in non-Gaussian situations? The following table presents the asymptotic relative efficiency of the Wilcoxon test relative to the t-test in the location shift problem for eight well-known F's.

TABLE 1.1: ARE of Wilcoxon and t-Tests of Location Shift

F	Normal	Uniform	Logistic	Exp	t_3	Exp	LNormal	t_2
ARE	.955	1.0	1.097	1.5	1.9	3	7.35	∞

The striking feature of this table is not the modest loss of information at the normal model, but the enormous gains achieved by transforming to ranks in certain non-Gaussian situations. We would need 50 percent more observations if we foolishly used the t-test instead of the Wilcoxon test at the double exponential distribution and *seven times* as many observations at the lognormal distribution. In the latter case, note that if we were to do the log transformation to get back to normality the Wilcoxon test is unaffected, since ranks are invariant to monotone transformations. So the transformation to ranks achieves most of what we could have achieved had we known that the log transformation was appropriate. For extremely heavy-tailed distributions the advantage for the rank test is even more pronounced.

The message of Table 1.1 is a familiar one from robust statistics: if we are willing to pay a small insurance premium (5% efficiency loss) at the Gaussian model we can protect ourselves against the extremely poor performance of least-squares based methods in heavy-tailed situations. The 5 percent premium seems very modest, particularly in view of the huge improvement in the lognormal case, but one may wonder whether there are other plausible distributions for which the asymptotic relative efficiency of the two tests is even worse than the .955 in the Gaussian case. This question was answered originally by Hodges and Lehmann(1956) too. To find the f which minimizes ARE, one must minimize the L_2 norm of the density subject to the constraint that the density has variance one. This turns out to be isomorphic to a standard problem in kernel density estimation. The least favorable density turns out to be the one generating the Epanechnikov kernel, and this solution gives a lower bound to the ARE of about .864. Thus, even in non-Gaussian cases, the t-test can never be more than about 15% better than the Wilcoxon test of location shift, but *can be arbitrarily worse* as indicated in the Table.

1.2 Linear Rank Statistics

A more general approach to two sample rank tests is suggested by considering the distribution of the ranks when the X's are generated from $f(x)$ and the Y's from $f(y - \theta)$. Let R denote the vector of ranks from the combined sample $Z = (X_1, ..., X_n, Y_1, ..., Y_m)$, so

$$\begin{aligned} f_R(r;\theta) &= \int_{z_j = z_{(r_j)}} \prod_{k=1}^{n} f(z_k) \prod_{k=n+1}^{n+m} f(z_k - \theta) dz \\ &= \int_{\xi_1 < \ldots < \xi_{n+m}} \prod_{j=1}^{n+m} f(\xi_{r_j}) \prod_{k=1}^{m} \frac{f(\xi_{r_j} - \theta)}{f(\xi_{r_j})} d\xi \end{aligned}$$

where $\xi_{r_j} = z_{(r_j)}, \quad j = 1, ..., n+m$. But the density of the order statistics $Z_{(1)}, ...Z_{(n+m)}$ when $Z_1, ...Z_{n+m}$ are iid with density $f(z)$ is $n\prod f(\xi_{r_j})$ so

$$f_R(r;\theta) = (n)^{-1} E_0 \prod_{k=1}^{m} \frac{f(\xi_{r_j} - \theta)}{f(\xi r_j)}$$

where E_θ designates that the expectation is taken with respect to the density $f(x - \theta)$. For any fixed alternative θ , the optimal rank test could be computed as the likelihood ratio based on this expression for the density. A more practical option involves computing a locally most powerful test based on the score function

$$\frac{\partial \log f_R(R;\theta)}{\partial \theta} = -\sum_{j=1}^{n} E_0 \frac{f'(Z_{(R_j)})}{f(Z_{(R_j)})}.$$

To illustrate, suppose f is logistic, so

$$f(z) = e^z/(1 + e^z)^2 = F(z)(1 - F(z)).$$

Then, $-f'(x)/f(x) = 2F(x) - 1$, and $F(Z(i)) = i/(n + m + 1)$ so the locally most powerful rank test of $H_0 : \theta = 0$ is based on the sum of the ranks and hence *is* the Wilcoxon-Mann-Whitney test. In contrast, if f were Gaussian, so $-f'(x)/f(x) = x$ we may approximate the optimal rank test by using the statistic,

$$\sum_{i=1}^{n} \Phi^{-1}(R_i/(n + m + 1))$$

proposed by van der Waerden. Alternative approximations for $EZ_{(i)}$ with Z_i Gaussian have also been suggested, see e.g. Hettmansperger(1984). In general the approximate test statistic

$$\sum_{i=1}^{m} f'(R_i/(n + m + 1))/f(R_i/(n + m + 1))$$

may be used since $Z_{(i)} = F(U_{(i)})$ where $U_1, ..., U_{n+m}$ are iid uniform on $(0, 1)$, and $EU_{(i)} = i/(n + m + 1)$.

1.3 Asymptotics of Linear Rank Statistics

The monograph of Hájek and Šidák (1967) constituted a complete reappraisal of the theory of rank statistics and provided an elegant general approach to the study of the asymptotic theory of linear rank statistics. For any sample $\{Y_1, Y_2, ..., Y_n\}$, and associated ranks $\{R_1, R_2, ..., R_n\}$, Hájek and Šidák introduced the rank generating functions:

$$\hat{a}_i(t) = \begin{cases} 1 & \text{if } t \leq (R_i - 1)/n \\ R_i - tn & \text{if } (R_i - 1)/n \leq t \leq R_i/n \\ 0 & \text{if } R_i/n \leq t \end{cases} \tag{1}$$

These functions "generate the ranks" of the Y's in the sense that, integrating with respect to Lesbesgue measure,

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

while more general notions of "ranks" may be obtained by replacing Lesbesgue measure by alternative score functions $\varphi(t)$. For example, $\varphi(t) = 1/2 sgn(t - 1/2)$ generates sign scores,

$$\hat{b}_i = \int \hat{a}_i(t)\, d\varphi(t) = \hat{a}_i(1/2) - 1/2 = \begin{cases} +1/2 & \text{if } R_i \geq n/2 + 1 \\ 0 & \text{otherwise} \\ -1/2 & \text{if } R_i \leq n/2 \end{cases}$$

The invariance of the ranks to monotone transformations means that the R_i's may also be viewed as the ranks of the uniform random sample $\{U_1, ..., U_n\}$ with $U_i = F(Y_i)$, and the rank generating functions $\hat{a}_i(t)$ may be seen as replacing the indicator functions $I(Y_i > F^{-1}(t)) = I(U_i > t)$, by the smoother "trapezoidal" form given by ((1)) Thus the rank generating functions behave like an empirical process as the following result shows.

Theorem 1 *(Hájek and Šidák (1967, Thm V.3.5)) Let* $c_n = (c_{1i}, ..., c_{nn})$ *be a triangular array of real numbers satisfying*

$$\max(c_i - \bar{c})^2 / \sum_{i=1}^{n} (c_i - \bar{c})^2 \to 0 \tag{2}$$

and assume that $\{Y_1, ..., Y_n\}$ *constitute a random sample from an absolutely continuous distribution* F*. Then, the process*

$$Z_n(t) = [\sum_{i=1}^{n} (c_i - \bar{c})^2]^{-1/2} \sum_{i=1}^{n} (c_i - \bar{c}) \hat{a}_i(t)$$

converges weakly to a Brownian Bridge process on $C[0,1]$.

In the two sample problem the c_{in}'s may be taken as simply the indicator of which sample the observations belong to, and the "Lindeberg condition" (2) is satisfied as long as n_1/n stays bounded away from 0 and 1. A limiting normal theory for a broad class of linear rank statistics of the form

$$S_n = [\sum(c_{ni} - \bar{c}_n)^2]^{-1/2} \sum(c_{ni} - \bar{c}_n)\hat{b}_i$$

where $\hat{b}_i = \int \hat{a}_i(t) d\varphi(t)$ is immediate. In particular, for square integrable $\varphi : [0,1] \to \mathbf{R}$ we have the linear representation

$$S_n = [\sum(c_{ni} - \bar{c}_n)^2]^{-1/2} \sum(c_{ni} - \bar{c}_n)\varphi(U_i) + o_p(1), \tag{3}$$

and consequently, S_n is asymptotically Gaussian under the null with mean 0 and variance $A^2(\varphi) = \int(\varphi(t) - \bar{\varphi})^2 dt$ where $\bar{\varphi} = \int \varphi(t)dt$. Behavior under sequences of local alternatives can be studied using standard contiguity results.

Thus, for example, in the two-sample location shift model with local alternatives $H_n : \delta_n = \delta_0/\sqrt{n}$ we have S_n asymptotically Gaussian with mean $\omega(\varphi, F)(\sum(c_{in} - \bar{c}_n)^2)^{1/2}\delta_0$ and variance $A^2(\varphi)$ where

$$\omega^2(\varphi, F) = \int f(F^{-1}(t)d\varphi(t).$$

In the Wilcoxon case where φ is Lesbesgue measure $A^2(\varphi) = 1/12$ and $\omega^2 = \int f^2(x)dx$ which yield the expression evaluated to produce Table 1.1. An important virtue of such rank tests is that the test statistic and its limiting behavior under the null hypothesis are independent of the distribution F generating the observations. See Draper(1988) for a detailed discussion of the problems related to the estimation the the nuisance parameter ω^2 in the Wilcoxon case.

1.4 Duality of Ranks and Quantiles

It is worth emphasizing at this point that the only computation required for the simple linear rank tests described above involves sorting and ranking the sample observations. To extend these ideas to more general models it is helpful to embed these operations into an optimization framework as suggested in Koenker and Bassett(1978). Recall that the solutions to the problem

$$\min_{\xi \in \mathbf{R}} \sum_{i=1}^{n} \rho_\tau(y_i - \xi), \tag{4}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ are the τ^{th} quantiles, $\hat{\xi}_n(\tau) = y_{([\tau n]-1)}$, and by varying τ and solving (4), we sort the sample $y's$. This is a parametric linear program of a particularly simple form:

$$\min_{(\xi,u,v)\in\mathbf{R}\times\mathbf{R}_+^{2n}} \{\tau 1_n' u + (1-\tau)1_n' v \mid 1_n\xi + u - v = y\}$$

where 1_n denotes an n-vector of ones. We may call this the primal problem. The corresponding dual problem

$$\max_a \{y'a | 1_n' a = (1-\tau)n, a \in [0,1]^n\} \tag{5}$$

has as solutions the Hájek -Šidák rank generating functions $\hat{a}_i(t)$ defined in (1). So just as the primal problem *sorts* the sample observations, the dual problem *ranks* the observations. Neither problem is terribly interesting in the one-sample model since we already have a painfully obvious way to sort and rank observations. However, for more general parametric models it is far less clear how to sort or rank the observations and the optimization approach suggests a reasonable way to do both operations.

2 Regression Quantiles and Rank Scores

For the classical linear regression model

$$y_i = x_i'\beta + u_i, \quad i = 1, 2, ..., n,$$

Koenker and Bassett (1978) proposed solving

$$\hat{\beta}_n(\tau) = \mathrm{argmin}_{\mathrm{b}\in\mathbf{R}^p} \sum_{\mathrm{i}=1}^{\mathrm{n}} \rho_\tau(\mathrm{y_i} - \mathrm{x_i'b}). \tag{6}$$

These p-dimensional "regression quantiles" determine a sequence of hyperplanes which estimate the conditional quantile functions of the response variable y. As in the one sample model, as τ varies over $[0,1]$ we have a parametric linear program which can be solved efficiently. See Koenker and d'Orey (1987, 1993) for details on computation. The dual problem corresponding to (6) is

$$\hat{a}(\tau) = \mathrm{argmax}\,\{\mathrm{y'a} | \mathrm{X'a} = (1-\tau)\mathrm{X}'1_\mathrm{n}, \mathrm{a} \in [0,1]^\mathrm{n}\} \tag{7}$$

In the one sample problem, where $X = 1_n$, $\hat{a}_i(\tau)$ specializes to rank generating functions of Hájek and Šidák . This was first noted by Gutenbrunner and Jurečková (1992), who provided a detailed study for the regression model of $\hat{a}_i(t)$, which they called the regression rankscore process. In regression, these regression rankscore functions are no longer of the simple "trapezoidal" form (1), indeed they can even be non-monotone, but like the primal regression quantile process they are easily computed via parametric linear programming methods, and as we describe in the next section, they provide a natural way to extend linear rank statistics to more general linear models.

2.1 Rank Tests of Linear Hypotheses in Regression

Now consider the partitioned linear model

$$Y = X\beta + Z\gamma + u$$

where β and γ are p and q dimensional parameters, and u is a vector of iid errors with common distribution function F. Suppose we are interested in testing the hypothesis $H_0 : \gamma = 0$, with β unspecified versus the (Pitman) local alternatives, $H_n : \gamma = \gamma_0/\sqrt{n}$. Gutenbrunner, Jurečková , Koenker and Portnoy(1993), hereafter GJKP propose the test statistic,

$$T_n = S_n' Q_n^{-1} S_n / A^2(\varphi) \tag{8}$$

where $S_n = (Z - \hat{Z})'\hat{b}_n$, $\hat{b}_n = (\int_0^1 \hat{a}_i(t)\, d\varphi(t))_{i=1}^n$, $\hat{Z} = X'(X'X)^{-1}X'Z$, $Q_n = (Z - \hat{Z})'(Z - \hat{Z})$, with $A^2(\varphi)$ as defined following (3). An important feature of the test statistic T_n is that it requires no estimation of nuisance parameters, since the functional $A(\varphi)$ 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. Wald and Likelihood ratio tests based on regression quantiles require estimation of the so-called sparsity or quantile density function. Tests based on the least squares estimator of course require estimation of the nuisance parameter σ^2.

The following finite sample invariance result, which follows easily from Theorem 3.2 of Koenker and Bassett (1978), proves to be quite useful in the sequel, extending well-known invariance properties of least squares based tests to the class of rank tests based on S_n.

Lemma 1 *Let $S_n(X, Z, y) = (Z - \hat{Z})'\hat{b}_n$ defined following (2.3), and P be an arbitrary nonsingular $p \times p$ matrix, M an arbitrary $p \times q$ matrix and γ an arbitrary p-vector. Then for any φ such that $\int_0^1 \varphi(t)dt = 0$, $S_n(XP, Z + XM, y + X\gamma) = S_n(X, Z, y)$.*

The asymptotic distribution of T_n under H_n is shown by GJKP(1993) to be non-central χ^2 with q degrees of freedom and noncentrality parameter

$$\eta^2 = [\omega^2(\varphi, F)/A^2(\varphi)]\gamma_0' Q_0 \gamma_0$$

where $Q_0 = \lim_{n\to\infty} Q_n/n$. Thus the asymptotic relative efficiency of these tests with respect to the classical F test is therefore identical to the location shift tests considered earlier. In particular, for the Wilcoxon scores $\varphi(t) = t - 1/2$ we may refer directly to Table 1.1. For f unimodal, we can obtain an asymptotically optimal test if we take

$$\varphi(s) = \varphi_F(s) = -\frac{f'(F^{-1}(s))}{f(F^{-1}(s))}, \quad 0 < s < 1.$$

Thus in the Gaussian case we are led to $\varphi_\Phi(s) = \Phi^{-1}(s)$. Computing the vector $\hat{b}_n = (\hat{b}_i)$ may appear to pose serious problems since it requires the entire sample path of the n-vector $\{\hat{a}_n(\tau) : \tau \in [0, 1]\}$. Fortunately, this problem is easily manageable. The function $\hat{a}_n(\tau)$. is piecewise linear and continuous with breaks at $0 = \tau_0 < \tau_1 < ... < \tau_{J-1} < \tau_J = 1$. Portnoy(1991) has shown that J, is asymptotically $O_p(n \log n)$. The solution, as noted above, at these J breakpoints is easily computed by standard parametric linear programming methods (see Koenker and d'Orey (1987, 1993)) and linear interpolation recovers the rest of the function.

Three score functions play prominent roles in the application of rank tests. The piecewise linearity of $\hat{a}_n(\tau)$ may be exploited to simplify the $\hat{b}_n$ integrals, as the following examples illustrate.

(i) **Wilcoxon scores:** $\varphi(s) = s - 1/2$. Integration by parts yields,

$$\begin{aligned}
\hat{b}_i &= -\int_0^1 (s - 1/2)\, d\,\hat{a}_i(s) \\
&= \int_0^1 \hat{a}_i(s)\, ds - 1/2 \\
&= \sum_{j=1}^{J} 1/2(\hat{a}_i(\tau_{j+1}) + \hat{a}_i(\tau_j))\,(\tau_{j+1} - \tau_j) - 1/2
\end{aligned}$$

(ii) **Normal (van der Waerden) scores:** $\varphi(s) = \Phi^{-1}(s); 0 < s < 1$. Denoting the standard normal df and density by Φ and ϕ respectively, and using the fact that the function $\hat{a}_i'(\tau)$ is piecewise constant, we have:

$$\begin{aligned}
\hat{b}_i &= -\int_0^1 \Phi^{-1}(s)\, d\,\hat{a}_i(s) \\
&= -\int_0^1 \Phi^{-1}(s) \sum_{j=1}^{J} \hat{a}_i'(\tau_j)\, ds \\
&= \sum_{j=1}^{J} \hat{a}_i'(\tau_j)\Big[\phi(\Phi^{-1}(\tau_{j+1})) - \phi(\Phi^{-1}(\tau_j))\Big]
\end{aligned}$$

(iii) **Sign-Median scores:** $\varphi(s) = \frac{1}{2} sgn(s - \frac{1}{2})$.

$$\hat{b}_i = -\int_0^1 \varphi(s)\, d\,\hat{a}_i(s) = \hat{a}_i(1/2) - 1/2$$

The use of the sign scores of T_n was already considered in Koenker and Bassett (1982), however there was some ambiguity there as to how the signs corresponding to the p zero residuals of the restricted model should be evaluated. The regression rank score approach resolves this point by evaluating $\hat{a}_i(1/2) - 1/2$ which lies in the open interval $(-1/2, 1/2)$ for these observations.

2.2 Confidence Intervals via Inversion of Rank Tests

An interesting application of this rank test approach involves the construction of confidence intervals for the parameters of linear quantile regression estimators. This problem has received considerable recent attention with several authors focusing on bootstrap methods. See for example Buchinsky (1994), Hahn (1995) and Parzen, Wei, Ying (1994). The emphasis on resampling methods reflects a dissatisfaction with earlier methods, e.g., Koenker and Bassett (1982), Welsh (1988), based on estimation of the sparsity function.

Hušková (1994) has proposed sequential procedures based on the regression rankscore process and suggested inverting rank score tests to obtain sequential, fixed length confidence intervals. For general score functions this inversion is computationally rather difficult; however, in one important special case it turns out to be very tractable. This is the case of constructing a confidence interval for a single parameter in a simple linear quantile regression model. Consider the special case of the previous testing problem in which X_2 consists of a single covariate, say x_2, and following Koenker (1994) write

$$y = X_1\beta_1 + x_2\beta_2 + u.$$

Suppose we would like to test the hypothesis

$$H_0 : \beta_2 = \xi$$

at the τ-th quantile regression. The score function appropriate at the τ-th quantile is $\varphi_\tau(t) = \tau - I(t < \tau)$ which is the obvious generalization of sign scores for the case of median (l_1) regression. Our test would be based on the regression rank score process

$$\max_{a\in[0,1]^n}\{(y - x_2\xi)'a|X_1'a = (1-\tau)X_1'1\} \tag{9}$$

which corresponds to fitting the restricted model under the null hypothesis. By the foregoing theory, under H_0,

$$S_n(\xi) = n^{-1/2}x_2'\hat{b}_n(\xi) \rightsquigarrow \mathcal{N}(0, A^2(\varphi)q_n^2)$$

where $\hat{b}_n = (\hat{b}_{ni})$,

$$\hat{b}_{ni}(\xi) = \int_0^1 \hat{a}_{ni}(t;\xi)d\varphi_\tau(t) = \hat{a}_{ni}(\tau;\xi) - \tau$$

$\hat{a}_{ni}(\tau;\xi)$ denotes elements of the solution to ((9)) and

$$q_n^2 = n^{-1}x_2'(I - X_1(X_1'X_1)^{-1}X_1')x_2.$$

Once again the computational burden of the test may appear daunting, but like the problem of computing the sample path of $\hat{a}_n(\tau)$, parametric linear programming may be employed. The function $\hat{a}_{ni}(\tau;\xi)$ is piecewise constant in ξ for fixed τ; ξ may be gradually altered without compromising the optimality of an initial $\xi = \xi_0$ as long as the signs of the residuals in the primal quantile regression problem don't change, that is as long as the new $y - x_2\xi$ doesn't cross the τ-th quantile regression plane determined by the response vector $y - x_2\xi_0$. When ξ eventually hits this boundary the solution is updated by making a single simplex pivot. The process continues in this way until $S_n(\xi)$ exceeds a chosen critical value. Since $S_n(\xi)$ is piecewise constant it seems useful to adopt a linear interpolation of the function to construct the confidence interval. See Beran and Hall (1993) for a detailed analysis of such interpolation schemes for the case of ordinary sample quantiles.

The resulting intervals unlike the more conventional Wald-type intervals based on estimation of the sparsity function are not symmetric. They are however centered on the point estimate $\hat{\beta}_2(\tau)$ in the sense that $S_n(\hat{\beta}_2(\tau)) = 0$, a fact that follows immediately from the constraint $X'\hat{a}(\tau) = (1-\tau)X'1$ in the unconstrained problem. Since the steps taken in the parametric programming implementation of the computations are extremely simple, the confidence intervals obtained by this rank test inversion are comparable to those based on sparsity estimation in terms of computational effort. Resampling methods are obviously considerably more demanding in this respect.

3 Rank Tests for Heteroscedasticity

As the confidence interval example illustrates, it is fruitful to consider alternative forms of the rank test score function specificly tailored to interesting alternative hypotheses. Another example of this sort involves tests for heteroscedasticity. There is a well established rank test literature on two-sample tests for equality of scale. These tests can be easily adapted to the linear model to provide an interesting class of new tests for heteroscedasticity.

This approach has been developed by Gutenbrunner (1994) who considered a very general class of heteroscedasticity tests, for the linear model

$$y = X\beta + u.$$

Following Gutenbrunner, we may consider a family of error densities of the form

$$f_\lambda(u) = e^{-\lambda} f_0(ue^{-\lambda})$$

where $\sigma_i = e^{\lambda_i}$ may be regarded as scale parameter. We will partition X as $X = [1 \vdots X_1 \vdots X_2]$ with a third component of X, X_3, reserved as a possible determinant of λ, but not appearing in X. Let p_i denote the column rank of X_i. The model takes the form

$$y = X\beta + \Sigma v$$

where $\Sigma = \text{diag}\ (\sigma_i)$ and the $\{v_i\}$'s are iid. We may formulate the hypothesis of homoscedasticity by expressing λ as the linear function of the

$$\lambda = X_2\delta_2 + X_3\delta_3$$

and we will test

$$H_0 : \delta_2 = 0, \quad \delta_3 = 0$$

versus

$$H_n : \delta_2 = \eta_2/\sqrt{n}, \quad \delta_3 = \eta_3/\sqrt{n}.$$

Note that the decomposition of the covariates permits X_1 to influence only location, X_3 only scale, and X_2 both location and scale. For the δ_3 component of the test we require a score function appropriate to testing equality of scale in the two sample problem. For given f, the optimal rank score function for this purpose is given by, see *e.g.* Hájek and Šidák (1967),

$$\varphi_3(u) = -1 - F^{-1}(u)\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}$$

In the normal case the optimal score function is thus $\varphi(u) = (\Phi^{-1}(u))^2$, a U-shaped function characteristic of the score functions for rank tests of scale. Ignoring for the moment the δ_2 component of the alternative hypothesis we could base a test of H_0 versus H_n on the statistic

$$T_{3n} = S_n' Q_n^{-1} S_n / A^2(\varphi)$$

where as in ((8)), for an appropriate choice of φ,

$$\begin{aligned} S_n &= (X_3 - \hat{X}_3)'\hat{b}_n \\ \hat{b}_n &= (\int_0^1 \hat{a}_i(t)d\varphi(t))_{i=1}^n \\ \hat{X}_3 &= X(X'X)^{-1}X'X_3 \\ Q_n &= (X_3 - \hat{X}_3)'(X_3 - \hat{X}_3). \end{aligned}$$

To deal with the δ_2 component of the alternative hypothesis we have two options. One is to make some preliminary estimate the regression effect of X_2 on location and construct a rank test based on ranks of the recentered observations, i.e., residuals. Such aligned rank tests have an extensive literature, and are particularly attractive in problems of testing for regression shift. See Adichie (1984) and Puri and Sen (1985) for detailed discussions. However, as noted in Gutenbrunner, Jurečková and Koenker (1995) and elsewhere, the aligned rank approach to testing for regression-in-scale is well-suited to models in which the underlying error density is symmetric, but in the event of asymmetric densities alignment tends to confound location and scale effects.

Following Gutenbrunner (1994) we may construct tests sensitive to the δ_2 component based on the L-statistic

$$\tilde{\beta}_n^{\nu} = \int \hat{\beta}(t) d\nu(t)$$

where ν is a finite signed measure designed to provide an L-estimator of scale. The choice of a good ν is obviously also dependent on f. It is well known, see, e.g., Serfling (1980), that the optimal choice of ν is that generated by the "signed density"

$$J(u) = \frac{d}{dx}\left(x \ \frac{f'(x)}{f(x)}\right)_{x=F^{-1}(u)}$$

so for example at the normal model we would choose $J(u) = \Phi^{-1}(u)$. As long as we choose J to satisfy $J = \int J(u)du = 0$ so $\nu(0,1) = 0$, the estimator is invariant to regression shift, since the transformation $y \to y + X\gamma$ implies $\hat{\beta}(t) \to \hat{\beta}(t) + \gamma$.

Partitioning $(\tilde{\beta}_n^{\nu})' = (\tilde{\beta}_{0n}^{\nu}, (\tilde{\beta}_{1n}^{\nu})', (\tilde{\beta}_{2n}^{\nu})')$ to conform with our original partitioning of β we may take $\tilde{\beta}_{2n}^{\nu}$ as an estimator of δ_2 for suitable choice of the measure ν. Gutenbrunner and Jurečková (1992) can then be used to show that

$$\sqrt{n}(\tilde{\beta}_{2n}^{\nu} - \beta_2(\nu)) \rightsquigarrow \mathcal{N}(0, \sigma^2(\nu, F)Q_2^{-1})$$

where

$$\beta_2(\nu) = \nu(0,1)\beta_2 + \delta_2 \int_0^1 F^{-1}(u)d\nu(u)$$

$$\sigma^2(\nu, F) \int\int \frac{u \wedge v - uv}{f(F^{-1}(u))f(F^{-1}(v))} d\nu(u)d\nu(v)$$

$$Q_2^{-1} = \lim_{n\to\infty} n^{-1}(X_2 - \hat{X}_2)'(X_2 - \hat{X}_2)$$

and $\hat{X}_2$ is the projection of X_2 on $[1 \vdots X_1]$. Therefore at the null, having chosen ν such that $\nu(0,1) = 0$, $\beta_2(\nu) = 0$ and we have

$$T_{2n} \equiv n\sigma^{-2}(\nu, F)(\tilde{\beta}_{2n}^{\nu})'Q_2\tilde{\beta}_{2n}^{\nu} \sim \chi^2_{p_2}.$$

While under H_n, T_{2n} is noncentral $\chi^2_{p_2}$ with non centrality parameter, $\zeta = \sigma^{-2}(\nu, F)\eta_2' Q_2 \eta_2$.

Gutenbrunner (1994) suggests two alternative estimators of the nuisance parameter $\sigma^2(\nu, F)$, one for the case in which ν has a signed density J, the other intended for the case in which ν is discrete. In the former case we have, integrating by parts,

$$\sigma^2(\nu, F) = -\int_0^1 \int_0^\nu (F^{-1}(u) - F^{-1}(v))^2 d\tilde{J}(u) dJ(v) - (\int F^{-1}(u) d\tilde{J}(u))^2$$

where $\tilde{J}(u) = uJ(u)$. Under the null this can be estimated by replacing $F^{-1}(u)$ by $\hat{\beta}_0(u)$, i.e. by

$$\hat{\sigma}_n^2(\nu) = -\int_0^1 \int_0^\nu (\hat{\beta}_0(u) - \hat{\beta}_0(\nu)) d\tilde{J}(u) dJ(u) - (\int \hat{\beta}_0(u) d\tilde{J}(u))^2.$$

In the latter case we require direct estimation of the sparsity function

$$s(u) = \frac{d}{du} F^{-1}(u) = 1/f(F^{-1}(u))$$

by, for example, the difference quotient,

$$s_n(u) = (2h_n)^{-1}(\hat{\beta}_0(u + h_n) - \hat{\beta}_0(u - h_n))$$

which can be substituted into our initial expression to obtain

$$\hat{\sigma}_n(\nu) = \int \int (u \wedge v - uv) s_n(u) s_n(v) d\nu(u) d\nu(v).$$

As noted by Gutenbrunner, the rate of convergence of the former approach is $O(n^{-1/2})$, while the latter is slower due to the estimation of the sparsity function. Tests related to the discrete approach were proposed by Koenker and Bassett (1982), but it now seems preferable to avoid the sparsity estimation problem by adopting the smoother form of the measure ν.

The classical theory of L-statistics admits two equivalent formulations. One may either consider weighted averages of the order statistics with weights generated by a fixed function, say J, or equivalently, we may consider randomly weighted averages of the original observations with the weights determined by the ranks of the observations. Thus, we have

$$\sum_{i=1}^n J(i/(n+1)) X_{(i)} = \sum_{i=1}^n J(R_i/(n+1)) X_i.$$

These two approaches diverge in the linear model yielding two distinct approaches to L-statistics, as noted by Gutenbrunner and Jurečková (1992). To see this consider weights defined by

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

for J of bounded variation with $J \geq 0, \int J(t)dt = 1$. Gutenbrunner and Jurečková (1992) consider the weighted least squares estimator

$$\hat{\beta}_n^J = (X'WX)^{-1}X'Wy$$

with $W = \text{diag } (w_i)$. In the simple case of the J function corresponding to the trimmed mean, $J_\alpha(u) = (1-2\alpha)^{-1}I(\alpha \leq u \leq (1-\alpha))$ were already proposed in Koenker and Bassett (1978) and studied by Ruppert and Carroll (1980), and correspond to least squares estimation based on the observations lying between the α^{th} and $(1-\alpha)^{\text{th}}$ regression quantile planes. By relating this approach to rank statistics via the regression rank score process the formulation of Gutenbrunner and Jurečková has unified the theory. And from a practical standpoint it has clarified the role of the zero-residual observations for this version of trimmed least squares. For general J taking both positive and negative values we may write $J = J^+ - J^-$, and $w = w^+ - w^-$. This is essential for scale estimators. Thus we may replace the L-estimators *of type 1* discussed above with these L-estimators *of type 2* and proceed as before. This approach is elaborated in Gutenbrunner, Jurečková and Koenker (1995).

Combining the R and L components of the heteroscedasticity tests described above is straightforward particularly if we have orthogonality among the components of X which implies independent χ^2 components. The orthogonal case actually involves no sacrifice in generality, since the invariance lemma of Section 2.1 may be used to transform to this case.

4 Rank Tests for Time Series Models

Much of the theory surveyed in the previous sections may be applied directly to linear time-series models. Koul and Saleh (1995) have studied applications to stationary $AR(p)$ models. Koul and Mukerjee (1994) have studied applications to long-memory processes. There has been considerable recent research on aligned rank tests in time-series applications, see e.g., Hallin and Puri (1992,1994), and Campbell and Dufour (1995). It would be very interesting to compare the performance of various rank based tests in this context carefully evaluating the influence of preliminary estimators for the aligned tests, as well as the effect of the choice of score functions. There is a strong sense, reading the empirical literature in time series econometrics, that robust methods of inference are needed, particularly because innovation distributions appear to be long-tailed. Rank-based methods which traditionally have offered reliable size and greatly improved power in such circumstances would seem to be very promising.

Two particularly exciting new arenas of application for rank-based inference in econometrics are tests of nonstationary hypotheses and inference for ARCH-type models. In Hasan and Koenker (1994) tests of the unit root hypothesis are proposed using the regression rank score approach of Gutenbrunner, Jurečková , Koenker and Portnoy (1993). Related tests have also been suggested by Hercé (1995). Monte Carlo simulations corroborate the attractive robustness of these methods relative to the well-established least-squares based tests proposed by Dickey and Fuller (1979) and others.

ARCH and related stochastic volatility models also provide a fertile field for rank based inference since they extend in an important way the more traditional heteroscedastic models of economics discussed in Section 3. Some initial steps in this direction have been taken in Koenker and Zhao (1995).

5 Conclusion

Statistical inference based on ranks continues to provide an extremely attractive alternative to classical, likelihood-based, inference methods. The approach to rank-based inference introduced by Gutenbrunner and Jurečková (1992) has significantly expanded the scope of these methods by providing an elegant generalization of the Hájek and Šidák rankscore functions to linear models with nuisance parameters, thus circumventing alignment and preliminary estimation problems. Many interesting opportunities remain. We have focused exclusively on hypotheses related to location and scale shift, but broader classes of hypotheses including those of Kolmogorov-Smirnov form could be considered. See, e.g., Jurečková (1992). And there many potentially important extensions of these methods to the realm of nonlinear models.

References

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.

Bassett, G.W. and Koenker, R.W. (1982). An empirical quantile function for linear models with iid errors. *Journal of the American Statistical Association.* **77**, 407-415.

Bassett, G.W. and Koenker, R.W. (1986). Strong consistency of regression quantiles and related empirical processes *Econometric Theory*, **2**, 191-201.

Beran, R. and Hall, P. (1993). Interpolated nonparametric prediction intervals and confidence intervals, J. Royal Stat. Soc. (B), **55**, 643-652.

Buchinsky, M. (1994) Changes in the U.S. wage structure 1963-1987: Application of quantile regression, *Econometrica*, **62**, 405-458.

Campbell, B. and Dufour, J.-M. (1995) Exact nonparametric orthogonality and random walk tests, *Review of Econ and Stat.*, **77**, 1-16.

Chernoff, H. and Savage, I.R. (1958). Asymptotic normality and efficiency of certain non-parametric test statistics, *Ann. Math. Statist.*, **29**, 972-994.

Dickey, D.A. and Fuller, W.A. (1979) Distribution of the estimators for autoregressive time series with a unit root, *J. of Am. Stat. Assoc.* **74**, 427-431.

Draper, D. (1988). "Rank-based robust analysis of linear models. I. Exposition and Review," *Statistical Science*, **3**, 239-271.

Friedman, M. (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance, *J. of Am. Stat. Assoc*, **32**, 675-701.

Gutenbrunner, C. (1986). *Zur Asymptotik von Regression quantileprozessen und da raus abgeleiten Statistiken*, Ph.D. Dissertation, Univ. Freiburg.

Gutenbrunner, C. and Jurečková , J. (1992) Regression quantile and regression rank score process in the linear model and derived statistics, *Ann. Statist.* **20**, 305-330.

Gutenbrunner, C., Jurečková , J., Koenker, R. and Portnoy, S. (1993). Tests of linear hypotheses based on regression rank scores, *J. of Nonparametric Statistics*, **2**, 307-331.

Gutenbrunner, C. (1994). Tests for heteroscedasticity based on regression quantiles and regression rank scores, in Mandl, P. and Hušková , M. (eds.) *Asymptotic Statistics: Proceedings of the 5th Prague Symposium*

Physica-Verlag, Heidleberg.

Gutenbrunner, C., Jurečková , J., and Koenker, R. (1995). Regression rank tests for heteroscedasticity, preprint.

Hahn, J. (1995) Bootstrapping quantile regression estimators, *Econometric Theory*, **11**, 105-121.

Hájek , J. (1969) *A Course in Nonparametric Statistics* Holden-Day, San Francisco.

Hájek , J. and Šidák , Z. (1967). *Theory of Rank Tests*, Academia, Prague.

Hallin, M. and Puri, M.L. (1992) Rank Tests for time-series analysis: a survey, in Brillinger, *et al* (eds.) *New Directions in Time Series Analysis*, **45**, 111-145.

Hallin, M. and Puri, M.L. (1994) Aligned Rank Tests for linear models with autocorrelated error terms, *J. Mult. Anal.*, **51**, 318-337.

Hasan, M.N. and Koenker, R. (1994) Robust rank tests of the unit root hypothesis, preprint.

Hercé, M. (1995) Asymptotic theory of LAD estimation in a unit root process with finite variance errors, *Econometric Theory*, forthcoming.

Hettmansperger T.P. (1984) *Statistical inference based on ranks.* Wiley, New York.

Hodges, J.L. and Lehmann, E.L. (1956) The efficiency of some nonparametric competitors of the t-test, *Ann. Math. Stat.*, **27**, 324-335.

Hotelling, H. and Pabst, M.R. (1936), Rank correlation and tests of significance involving no assumption of normality, *Annals of Math. Stat*, **7**, 29-43.

Hušková , M. (1994) Some sequential procedures based on regression rank scores, *J. of Nonparametric Statistics*, **3**, 285-298.

Jureckova, J. (1992) Tests of Kolmogorov-Smirnov type based on regression rank scores, preprint.

Kendall, M.G. (1938) A new measure of rank correlation, *Biometrika*, **30**, 81-93.

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 Bassett, G (1982). Tests of linear hypotheses and l_1 estimation, *Econometrica*, **50**, 1577-1584.

Koenker, R. and d'Orey, V. (1987, 1993) Computing regression quantiles, *Applied Statistics*, **36**, 383-393 and **43**, 410-414.

Koenker, R.W. and Portnoy, S. (1987). L-Estimation for the Linear Model., *Journal of the American Statistical Association*, **82**, 851-857.

Koenker, R. (1994) Confidence intervals for regression quantiles, in Mandl, P. and Hušková , M. (eds.) *Asymptotic Statistics: Proceedings of the 5th Prague Symposium* Physica-Verlag, Heidleberg.

Koenker, R. and Zhao, Q. (1995) Conditional quantile estimation and inference for ARCH models, *Econometric Theory*, forthcoming.

Koul, H.L. and Saleh, (1995) Autoregression quantiles and related rank score processes, *Annals of Statistics*, **23**, 670-689.

Koul, H.L. and Mukherjee (1994) Regression quantiles and related processes under long range dependent errors, *J. Mult. Anal.*, **51**, 318-337.

Lehmann, E.L. (1975) *Nonparametrics*, Holden-Day, San Francisco.

Mann, H.B. and Whitney, D.R. (1947) On a test of whether one of two random variables is stochastically larger than the other, *Ann. Math. Stat.* **18**, 50-60.

Mosteller, F. and Rourke, R.E.K. (1973) *Sturdy Statistics*, Addison-Wesley, Reading, Mass.

Parzen, M. I. Wei, L.J., and Ying, Z. (1994) A resampling method based on pivotal estimating functions, *Biometrika*, **81**, 341-350.

Puri, M.L. and Sen, P.K. (1985). *Nonparametric Methods in General Linear Models.* J. Wiley, New York.

Serfling, R. J. (1980) *Approximation Theorems in Mathematical Statistics*, Wiley, New York.

Spearman, C. (1904) The proof and measurement of association between two things, *Am. J. Psychol.* **15**, 72-101.

Welsh, A.H. (1988). Asymptotically efficient estimation of the sparsity function at a point. *Stat. and Prob. Letters*, **6**, 427-432.

Wilcoxon, F. (1945) Individual comparisons by ranking methods, *Biometrics*, **1**, 80-83.