Introduction

Tests of hypotheses in the linear regression model are typically constructed using least-squares statistics. These tests are optimal when the errors are normally distributed, but may have poor power properties when the errors are far from normal. Over the years numerous alternative tests have been developed that possess good power properties for a wide spectrum of error distributions. Koenker and Bassett (1982) describe some robust tests based on least absolute deviation methods; Koenker (1995) surveys some methods based on ranks. Additional examples are given in Jure ková and Sen (1996). Unfortunately, these robust regression tests are somewhat more difficult to compute than least squares tests and seem not to be used much in practice.

Just as there are many asymptotically equivalent versions of the standard likelihood-based test statistics, there are also many asymptotically equivalent versions of the robust tests. Some of these, particularly those analogous to the C(α) versions of the likelihood tests, footnote are very easy to construct. For example, by using least squares to estimate nuisance parameters, one can develop easy-to-compute and easy-to-interpret robust tests that behave quite well in moderate-sized samples. These simple tests may be attractive alternatives to the usual t and F-tests commonly used in econometric research.

A Family of Tests

We consider the linear regression model

$$\mathbf{y} = \mathbf{X}\alpha + \mathbf{Z}\beta + \boldsymbol{\varepsilon} \qquad \#$$

where $\mathbf{y}$ is an n-dimension column vector of observations on some dependent random variable; the $n \times p$ matrix $\mathbf{X}$ and the $n \times q$ matrix $\mathbf{Z}$ contain observations on $p + q$ nonrandom regressors. We assume that the n elements of $\boldsymbol{\varepsilon}$ are i.i.d. random variables with a continuous density function $f(\varepsilon_i)$ having mean zero and variance σ^2. We assume further that the regressor data matrix $\mathbf{W} = [\mathbf{X}\ \ \mathbf{Z}]$ has full column rank and that $\mathbf{Z}$ contains a column of ones with a coefficient representing the intercept. The problem is to test the null hypothesis that $\alpha = 0$ against the alternative hypothesis that $\alpha \neq 0$. footnote

The standard approach to this problem is based on least squares regression methods. Define $\mathbf{b} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$ to be the constrained least squares estimate of β. Under the null hypothesis, $\mathbf{y} - \mathbf{Z}\beta$ is uncorrelated with the columns of $\mathbf{X}$ and hence $\mathbf{X}'(\mathbf{y} - \mathbf{Zb})$ should be near the origin. Indeed, $\mathbf{X}'(\mathbf{y} - \mathbf{Zb})$ has mean $\mathbf{X}'\mathbf{MX}\alpha$ and variance matrix $\sigma^2\mathbf{X}'\mathbf{MX}$, where $\mathbf{M} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. If $\hat{\sigma}^2$ is some consistent estimate of σ^2, one might reject the null hypothesis that $\alpha = 0$ upon observing a large value of the Euclidean distance measure

$$T_{LS} = (\mathbf{y} - \mathbf{Zb})'\mathbf{X}(\mathbf{X}'\mathbf{MX})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{Zb})/\hat{\sigma}^2. \qquad \#$$

Using the facts that $\mathbf{y} - \mathbf{Zb} = \mathbf{My}$ and $\mathbf{MZ} = \mathbf{0}$, we note that T_{LS} can be rewritten in its alternate C(α)-form

$$T_{LS} = (\mathbf{y} - \mathbf{Z}\hat{\beta})'\mathbf{MX}(\mathbf{X}'\mathbf{MX})^{-1}\mathbf{X}'\mathbf{M}(\mathbf{y} - \mathbf{Z}\hat{\beta})/\hat{\sigma}^2 \qquad \#$$

for arbitrary estimate $\hat{\beta}$. If σ^2 is estimated from the unconstrained least-squares residuals so

$$\hat{\sigma}^2 = \frac{y'[\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}']y}{n - p - q},$$

then T_{LS} is p times the usual F-statistic from normal regression theory. If the elements of $\mathbf{W}$ are bounded and $n^{-1}\mathbf{W}'\mathbf{W}$ tends to a positive definite limit as n tends to infinity, the asymptotic null distribution of T_{LS} is chi square with p degrees of freedom even if the errors are not normal.

The F-test is known to have poor power properties when the error distribution has thick tails. Since, under the null hypothesis, any function of the elements of $\mathbf{y} - \mathbf{Z}\beta$ should also be uncorrelated with the columns of $\mathbf{X}$, one might wish to base the test on some function of the residuals that downplays outliers, rather than on the residuals themselves. Commonly proposed examples are the sign of the residuals or the rank of the residuals. It will be convenient to follow the notational convention that, for any function $f(\bullet)$ mapping $\mathbf{R}$ into $\mathbf{R}$ and any n-dimensional column vector $\mathbf{a}$ with components a_i, $\mathbf{f}(\mathbf{a})$ is the n-dimensional column vector with components $f(a_i)$. Then, for some nondecreasing scalar function $h_n(\bullet)$ and some preliminary estimate $\widehat{\beta}$, we might by analogy with (3) reject the null hypothesis upon observing large values of

$$T_h(\widehat{\beta}) = \mathbf{h}(\mathbf{y} - \mathbf{Z}\widehat{\beta})'\mathbf{MX}(\mathbf{X}'\mathbf{MX})^{-1}\mathbf{X}'\mathbf{Mh}(\mathbf{y} - \mathbf{Z}\widehat{\beta})/\widehat{\sigma}_{hh} \qquad \#$$

where

$$\widehat{\sigma}_{hh} = \frac{\mathbf{h}(\mathbf{y} - \mathbf{Z}\widehat{\beta})'[\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}']\mathbf{h}(\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})}{n - p - q} . \qquad \#$$

(The function $h_n(\bullet)$ may depend on the data through some estimated parameters; hence the subscript.) Note that $T_h(\widehat{\beta})$ is q times the least squares F-statistic for the hypothesis that, in a regression of $\mathbf{h}(\mathbf{y} - \mathbf{Z}\widehat{\beta})$ on $\mathbf{W} = (\mathbf{X}, \mathbf{Z})$, the coefficients on $\mathbf{X}$ are zero.

A natural class of estimates $\widehat{\beta}_g$ are those that solve moment conditions of the form

$$\mathbf{Z}'\mathbf{g}(\mathbf{y} - \mathbf{Z}\widehat{\beta}_g) = 0 \qquad \#$$

for some nondecreasing function $g_n(\bullet)$. The choice $g_n(\bullet) \equiv h_n(\bullet)$ seems natural, but often leads to difficult computations. An alternative choice for g_n is the identity function so $\widehat{\beta}_g$ is the constrained least-squares estimator $\mathbf{b}$. The latter case leads to very simple computations since the test based on $T_h(\mathbf{b})$ can be performed by the following two-step algorithm:

1. Compute the residuals $\hat{\mathbf{u}}$ from a least-squares regression of $\mathbf{y}$ on $\mathbf{Z}$.
2. Test $\alpha = 0$ using the standard F-statistic from a least-squares regression of $\mathbf{h}(\hat{\mathbf{u}})$ on the full set of regressors $\mathbf{W} = (\mathbf{X}, \mathbf{Z})$.

Although the distribution of $T_h(\widehat{\beta}_g)$ generally will depend on the estimating function $\mathbf{g}$, that dependence will often be small. In the appendix we provide sufficient conditions to guarantee that the null distribution of $T_h(\widehat{\beta}_g)$ is asymptotically chi-square. As long as the regression contains an intercept that is not being tested, these conditions are satisfied for the standard $\mathbf{h}$ and $\mathbf{g}$ functions used in practice.

Although tests based on $T_h(\widehat{\beta}_h)$ have been discussed in the literature, the fact that simpler tests can be formed without having to compute $\widehat{\beta}_h$ has not been emphasized. footnote When the error distribution f is known to be symmetric, it turns out that local asymptotic power is independent of the choice of $\mathbf{g}$ so the proposed two-step algorithm using $T_h(\mathbf{b})$ involves no efficiency loss compared to using alternative estimates $\widehat{\beta}_g$. With asymmetric errors, the best preliminary estimator $\widehat{\beta}_g$ generally depends on the unknown error distribution. Unless something is known about error skewness, the simple test based on $T_h(\mathbf{b})$ appears to be as good as any other.

Rank Tests

An important special case is when h is some ”score” function applied to the ranks of the residuals. Suppose, for example,

$$h_n(u_i) = \psi(R_i/n) \qquad \#$$

where $\mathbf{u} = \mathbf{y} - \mathbf{Z}\widehat{\beta}$, R_i is the rank of u_i among $u_1, \ldots, u_n$, and ψ is some nondecreasing function mapping $\mathbf{R}$ into $\mathbf{R}$. Note that $R_i = nF_n(u_i)$, where F_n is the empirical distribution function for the $u's$. Thus, $h_n(\bullet) = \psi[F_n(\bullet)]$ is random and nondifferentiable. Furthermore, because of the discreteness of the function F_n, there are issues concerning the existence and uniqueness of solutions to equation (ref: est) when $g_n(\bullet)$ is also based on ranks. Nevertheless, the analysis in the previous section still applies since, asymptotically, the empirical distribution function behaves like the population distribution function. Indeed, since ranks are unchanged when a constant is added to each variable, asymmetric errors cause no complication; conditions A1-A3 of the appendix will be satisfied even with skewed error distributions. A detailed argument can be found in Akritas (1991) and in Puri and Sen (1985).

Rank tests were originally developed for simple univariate problems. In that context, they have two very attractive features. First, the rank-based test statistic has a null distribution not depending on the true error density; exact small-sample critical values can be computed. Second, by clever choice of ψ, a test can be found that is about as good as the F-test when the errors are normal, but performs much better when the errors are nonnormal. The first feature holds in our regression model only if α is a scalar and β is known. In our general regression model with covariates, the rank-based version of $T_h(\widehat{\beta}_g)$ is asymptotically, but not exactly, distribution free under the null hypothesis. Although σ_{hh} depends only on the function ψ and hence need not be estimated from the data, exact small-sample critical values are not available. As emphasized by Koenker (1995), the main advantage of rank tests in general regression models is the second feature: robustness can be attained without much loss of efficiency.

When h is of the form (ref: rank), the statistics discussed in the previous section are examples of so-called ”alligned” rank tests analyzed in Chiang and Puri (1984) and elsewhere. The key point is that these tests do not require a complicated preliminary estimate of the nuisance parameter β. At least as far as asymptotic size and power are concerned, any root-n consistent estimate will suffice in constructing $T_h(\widehat{\beta}_g)$. Gutenbrunner, et.al, (1993) propose an alternative construction of regression rank tests, essentially replacing the empirical distribution function F_n by an estimate derived from computed conditional quantiles. Although this procedure may behave better in small samples, it is unnecessary in large samples. Asymptotically, ranking based on the empirical distribution of least squares residuals gives the same answer as ranking based on estimated conditional quantiles. And, for large samples, computing the conditional quantiles is nontrivial.

A Monte Carlo Study

We conducted some simulations to check the small-sample properties of alternative robust regression tests based on ranks. We experimented with a number of design matrices and did not find any noticeable differences, so we report here only one case, a design with three explanatory variables (plus intercept) with one coefficient being tested. This design used the 21 observations on variables x_1 (air flow), x_2 (temperature) and x_3 (acid concentration) from the ”stack loss” data set of Daniel and Wood (1980, p 61). Values on the dependent variable were generated as

$$y_i = 2 + 4x_{1i} + x_{2i} + \alpha x_{3i}/10 + \varepsilon_i \qquad \#$$

for various values of α and for i.i.d draws from various error distributions. For each data set, we performed a number of one-tailed tests (at the 5% level) of the hypothesis that $\alpha = 0$ against the alternative that $\alpha > 0$. For larger sample sizes (42 and 84) we just replicated the 21-observation design matrix so the covariance structure remained constant. Rejection probabilities were estimated from 10,000 replications. All computations were done using Ox 2.10, a programming language described in Doornik (1998), and quantile regression programs provided by Roger Koenker.

The test statistics are of the form

$$\frac{\mathbf{x}'\mathbf{M}\mathbf{h}}{s\sqrt{\mathbf{x}'\mathbf{M}\mathbf{x}}}$$

where $\mathbf{x}$ is vector of observations on x_3 and $\mathbf{M}$ is the orthogonal projection matrix for the remaining regressors $\mathbf{Z}$. The vector $\mathbf{h}$ and the scalar s depend on the test. Define the residual vector $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\widehat{\beta}_g$ where $\widehat{\beta}_g$ is obtained by fitting (8) by either least squares or by least absolute deviations, assuming $\alpha = 0$. Then let R_i be the rank of $\hat{u}_i$. We considered seven tests:

- the least-squares t-test with $\mathbf{h} = \mathbf{y}$ and s^2 the usual variance estimate based on the sum of squared unconstrained residuals.
- rank tests r(lad) and r(ols) where $\mathbf{h}$ is the vector whose elements are the normalized ranks R_i/n and $s^2 = 1/12$.
- normal-score tests ns(lad) and ns(ols) where $\mathbf{h}$ is the vector of elements $\Phi^{-1}(R_i/n)$ obtained by applying the inverse of the standard normal distribution function to the normalized ranks; $s = 1$.
- regression quantile versions of the rank and normal score tests using equation (4.7) of Gutenbrunner and Jure ková (1992). footnote These are denoted by r(gjkp) and ns(gjkp) in the tables.

For the t-test, we employed the appropriate Student-t critical value; for the other tests, the asymptotic normal value of 1.645 was used. Rejection probabilities are presented in tables 1-4 for four alternative error distributions:

- standard normal
- Student with three degrees of freedom
- a (0.1, 0.9) mixture of two zero-mean normals with variances 25 and 1, re-standardized to have unit variance. footnote
- lognormal

The t-test is, of course, best when the errors are normal, but the robust tests behave satisfactorily. Size is roughly correct (especially in the larger sample sizes) and the power loss is small. For all three thick-tailed distributions, the robust tests perform much better than the t-test. There is little difference between the alternative robust tests. The r(ols) and ns(ols) tests based on least-squares estimates of nuisance parameters appear to be about as good as those based on more complicated estimates or on the regression quantiles.

Conclusion

Although we have treated here only the case of classical linear regression, our results carry over to other linear models. For example, Hasan and Koenker (1998) apply the regression quantile approach to construct rank tests of the unit-root hypothesis in time series analysis. While conducting the research on robust unit-root tests reported in Thompson (1999), one of us discovered that simple tests based on the ranks of least-squares residuals perform almost identically to the more complicated Hasan-Koenker tests. It was this observation that motivated the current paper.

Estimating nuisance parameters by least squares may seem to be a strange way to produce ”robust” test statistics that are designed to protect against nonnormality. Indeed, in very small samples, fully robust methods may well be superior. But our simulation examples suggest that, in moderate size samples, little is lost by using the simpler procedures. For example, when thick-tailed errors are suspected, the following two-step procedure might be employed to test the hypothesis $\boldsymbol{\alpha} = 0$ in the general regression model (1): after computing the ranks of the residuals from a least squares regression of y on Z, perform the usual F test using a least squares regression of these ranks on the full set of original regressors. This is asymptotically equivalent to the Wilcoxon rank test described by Koenker (1995) and should have satisfactory size and power for a wide spectrum of error distributions.

Appendix

A1. Symmetric Errors

The null distribution of $T_h(\widehat{\beta}_g)$ will be approximately chi-square if $(\mathbf{X}'\mathbf{M}\mathbf{X})^{-\mathrm{U}}\,\mathbf{X}'\mathbf{M}\mathbf{h}(\mathbf{y}-\mathbf{Z}\widehat{\beta}_g)/\sqrt{\widehat{\sigma}_{hh}}$ behaves like $(\mathbf{X}'\mathbf{M}\mathbf{X})^{-\mathrm{U}}\,\mathbf{X}'\mathbf{M}\mathbf{h}(\boldsymbol{\varepsilon})/\sqrt{\sigma_{hh}}$ when $\alpha = 0$ and the latter is approximately standard normal. This will hold if $\widehat{\beta}$ is a $\sqrt{n}$-consistent estimate of β, $\mathbf{h}$ can be linearized, and the elements of $\mathbf{h}(\boldsymbol{\varepsilon})$ behave like i.i.d. random variables. More formally, we have

Proposition *Suppose that, under a sequence of local alternatives such that $\alpha_n = n^{-\frac{1}{2}}\boldsymbol{a}$ as n tends to infinity, the following are satisfied*:

A1 *The elements of W are bounded, $n^{-1}W'W$ tends to a positive definite limit as n tends to infinity, and $\sqrt{n}\,(\widehat{\beta}_g-\beta)$ has a limiting distribution.*

A2 $(\mathbf{X}'\mathbf{M}\mathbf{X})^{-\frac{1}{2}}\mathbf{X}'\mathbf{M}\mathbf{h}(\varepsilon)$ *is asymptotically $N(0,\sigma_{hh}\mathbf{I})$ and $\widehat{\sigma}_{hh}$ converges in probability to σ_{hh}.*

A3 *There exists a constant ω_h (not depending on the function g) such that, uniformly in any compact set of a values in $\mathbf{R}^p$,*

$$n^{-\frac{1}{2}}\mathbf{W}'\mathbf{h}(\mathbf{y}-\mathbf{Z}\widehat{\beta}_g) = n^{-\frac{1}{2}}\mathbf{W}'\mathbf{h}(\boldsymbol{\varepsilon}) + \omega_h n^{-1}[\mathbf{W}'\mathbf{X}\mathbf{a} - \mathbf{W}'\mathbf{Z}\sqrt{n}\,(\widehat{\beta}_g-\beta)] + o_p(1).$$

Then, $T_h(\widehat{\beta}_g)$ has a limiting noncentral chi square distribution with p degrees of freedom and noncentrality parameter $\omega_h^2\sigma_{hh}^{-1}\lim n^{-1}\mathbf{a}'\mathbf{X}'\mathbf{M}\mathbf{X}\mathbf{a}$. The limiting distribution does not depend on the particular choice of preliminary estimate $\widehat{\beta}_g$.

Remark A1 and A2 require that $n^{-1/2}\mathbf{Z}'\mathbf{g}(\varepsilon)$ and $n^{-1/2}\mathbf{X}'\mathbf{M}\mathbf{h}(\varepsilon)$ be centered at the origin for all possible error density functions $f(\varepsilon_i)$. This will necessarily be the case if **g** and **h** are odd functions of $\boldsymbol{\varepsilon}$ and the true error density *f* is symmetric about the origin. However, if the unknown error density *f* is allowed to be asymmetric, this cannot be guaranteed for plausible choices for **g** and **h** other than the identify function. Thus, these assumptions essentially rule out the asymmetric case.

A2 will be satisfied if the $h_n(\varepsilon_i)$ are i.i.d. with mean zero and variance σ_{hh}. If the function is also differentiable, the linearization of A3 can be established by Taylor's theorem. If, for example, h_n' is the derivative of h_n, then ω_h will equal $E[h_n'(\varepsilon_i)]$. But the linearization may also work when *h* is nondifferentiable as in the case where $h_n(\varepsilon_i) = sgn(\varepsilon_i)$. If $h_n(\bullet)$ depends on the data through some parameter estimates, then σ_{hh} and the ω_h are typically obtained by taking expectations after the parameter estimates are replaced by the true values. For the special case where the true error density *f* is known to be symmetric, A1-A3 have been verified for the common robust functions **g** and **h** used in practice.

A2. Asymmetric Errors

As already noted, Assumptions A1-A3 will not typically be satisfied if the unknown error density *f* is asymmetric. However, it is easy to extend the theory to the case where asymmetric error distributions are permitted. Let **1** denote a column vector of *n* ones and let **e** denote a q-dimensional column vector with one as the first element and the rest of the elements being zeros. Although it may not be the case that *plim* $n^{-1}\mathbf{Z}'\mathbf{g}(\boldsymbol{\varepsilon}) = \mathbf{0}$, there generally will exist a scalar θ_g (depending on both the error density *f* and the function **g**) such that *plim* $n^{-1}\mathbf{Z}'\mathbf{g}(\boldsymbol{\varepsilon}-\theta_g\mathbf{1}) = \mathbf{0}$. Furthermore, since we are testing slope coefficients in a regression with an intercept, we may assume that the first column of **Z** is **1**. Then, $\mathbf{M1} = \mathbf{0}$ and $n^{-\frac{1}{2}}\mathbf{X}'\mathbf{M}\mathbf{h}(\boldsymbol{\varepsilon}-\theta_g\mathbf{1})$ will be centered at zero as long as $Eh_n(\varepsilon_i-\theta_g)$ does not depend on *i*. The following weaker conditions are typically satisfied even when asymmetric errors are contemplated:

A1' *The elements of W are bounded and $n^{-1}W'W$ tends to a positive definite limit as n tends to infinity. There exists a scalar θ_g* such that $\sqrt{n}\,(\widehat{\beta}_g-\beta-\theta_g\mathbf{e})$ *has a limiting distribution.*

A2' $(\mathbf{X}'\mathbf{M}\mathbf{X})^{-\frac{1}{2}}\mathbf{X}'\mathbf{M}\mathbf{h}(\boldsymbol{\varepsilon}-\theta_g\mathbf{1})$ *is asymptotically $N(0,\sigma_{hh}\mathbf{I})$ and $\widehat{\sigma}_{hh}$ converges in probability to*

σ_{hh}.

A3' *There exists a constant* ω_h *such that, uniformly in any compact set of* **a** *values in* $\mathbf{R}^p$,

$$n^{-\frac{1}{2}}\mathbf{W}'\mathbf{h}(\mathbf{y}-\mathbf{Z}\hat{\beta}_g) = n^{-\frac{1}{2}}\mathbf{W}'\mathbf{h}(\boldsymbol{\varepsilon}-\theta_g\mathbf{1})$$
$$+\,\omega_h n^{-1}[\mathbf{W}'\mathbf{Xa}-\mathbf{W}'\mathbf{Z}\sqrt{n}\,(\hat{\beta}_g-\beta-\theta_g\mathbf{e})] + o_p(1).$$

In the case where $h_n(\bullet)$ is nonrandom and differentiable, $\omega_h = E_f[h_n'(\varepsilon_i-\theta_g)]$ and $\sigma_{hh} = var_f[h_n(\varepsilon_i-\theta_g)]$. We find:

Proposition *When testing slope coefficients in a regression with intercept, suppose conditions* $A1'$*-*$A3'$ *are satisfied. Then, under a sequence of local alternatives,* $T_h(\hat{\beta}_g)$ *has a limiting noncentral chi square distribution with p degrees of freedom and noncentrality parameter* $\omega_h^2\sigma_{hh}^{-1}\, lim\; n^{-1}\boldsymbol{a}'\boldsymbol{X}'\boldsymbol{MXa}$.

Thus, the statistic $T_h(\hat{\beta}_g)$ defined in (ref: tee) still yields a valid chi-square test even in the presence of asymmetric errors. However, since ω_h and σ_{hh} will typically depend on the choice of preliminary estimator, it is no longer the case that local asymptotic power is the same for all $\hat{\beta}_g$. In the absence of information about the unknown error distribution, it is natural to use the simplest test statistic $T_h(\mathbf{b})$ where $\theta_g = 0$.

A3. **Relationship with Wald Tests**

The test statistics discussed in Section 3 are based on the correlation between **X** and $\mathbf{h}(\mathbf{y}-\mathbf{Z}\hat{\beta}_g)$. One could instead construct asymptotically equivalent statistics based on the estimate $\hat{\alpha}_h$ satisfying $\mathbf{W}'\mathbf{h}(\mathbf{y}-\mathbf{X}\hat{\alpha}_h-\mathbf{Z}\hat{\beta}_h) = \mathbf{0}$. Under the regularity conditions A1'-A3', $(\mathbf{X}'\mathbf{MX})^{-\mathrm{U}}\,(\hat{\alpha}_h-\alpha)$ is asymptotically normal with zero mean and variance matrix $(\sigma_{hh}/\omega_h^2)\mathbf{I}$ when the null hypothesis is true. The alternative Wald-type test statistic

$$(\hat{\alpha}_h-\alpha)'(\mathbf{X}'\mathbf{MX})^{-1}(\hat{\alpha}_h-\alpha)\hat{\omega}_h^2/\hat{\sigma}_{hh}$$

has the same limiting distribution under local alternatives as $T_h(\hat{\beta}_h)$. Besides being computationally more difficult, this Wald-type statistic requires the estimation of the additional parameter ω_h and appears to converge more slowly to the asymptotic distribution.

bibitem Akritas, M. (1991). An alternative derivation of aligned rank tests for regression, *Journal of Statistical Planning and Inference*, **27**, 171-186.

bibitem Chiang, C. and Puri, M. (1984). Rank Procedures for Testing Subhypotheses in Linear Regression, *Annals of the Institute of Statistical Mathematics*, **36**, Part A, 35-50.

bibitem Daniel, C. and Wood, F. (1980). *Fitting Equations to Data,* J. Wiley, New York.

bibitem Doornik, J.A. (1998), Object-Oriented Matrix Programming using Ox 2.0, London: Timberlake Consultants Ltd and Oxford: www.nuff.ox.ac. uk/Users/Doornik.

bibitem Gourieroux, C. and Monfort, A. (1995). *Statistics and Econometric Models, vol*. 2, Cambridge University Press, Cambridge.

bibitem Gutenbrunner, C. and Jure ková , J., (1992). Regression quantile and regression rank score proces in the linear model and derived statistics, *Annals of Statistics*, **20**, 305-330.

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

bibitem Hasan, M. and Koenker, R. (1997). Robust rank tests of the unit root hypothesis, *Econometrica*, **65**, 133-161.

bibitem Jure kova , J. and Sen, P. K. (1996). *Robust Statistical Procedures*: *Asymptotics and Interrelations*, J. Wiley, New York.

bibitem Koenker, R. and Bassett, G. (1982). Test of linear hypotheses and l_1 estimation, *Econometrica*, **50**, 1577-1584.

bibitem Koenker, R. (1995). Rank tests for linear models, in C.R. Rao and G.S.Madalla, eds. *The Handbook of Statistics*, *vol* 15, Elsevier, New York.

bibitem Thompson, S. (1999). Robust unit root testing with correct size and good power, unpublished working paper.