ROBUST RANK TESTS OF THE UNIT ROOT HYPOTHESIS

M. N. HASAN AND R. W. KOENKER

ABSTRACT. We consider a family of rank tests based on the regression rank score process introduced by Gutenbrunner and Jurečková (1992) to test the unit root hypothesis in economic time-series. In contrast to tests based on least-squares methods, the rank tests are asymptotically Gaussian under the null hypothesis, and have excellent power – particularly under innovation processes exhibiting heavy tails.

1. INTRODUCTION

Although it is widely recognized that many economic time-series display non-Gaussian features, the econometric literature on inference in time-series models continues to be grounded in optimality under Gaussian conditions. Nowhere is this tendency more pronounced than in the profusion of recent papers concerned with testing the unit-root hypothesis. We seek to redress this imbalance slightly by proposing a new class of rank tests of the unit root hypothesis, a class which offers reliable size and improved power, relative to least squares based tests, when confronted by time-series exhibiting a few large innovations.

Our approach may be viewed as a first step in an attempt to extend rank-based methods of statistical inference to an important class of non-stationary time-series models. The approach exploits a formal duality between rank statistics and quantile regression developed by Gutenbrunner anda Jurečková (1992) and elaborated in Gutenbrunner, Jurečková , Koenker, and Portnoy (1993), hereafter GJKP. This connection enables us to circumvent some of the difficulties associated with preliminary estimation required by aligned rank procedures while at the same time allowing us to accommodate nuisance parameters in a rather general manner.

Despite the fact that rank tests play an important role elsewhere in statistics, they are rarely mentioned in econometrics. We will resist the temptation to speculate on the causes of this neglect, and simply recommend the excellent textbook treatments of Hájek (1969), and Lehmann (1975),and the monographs of Hájek and Šidák (1967),

Revision: January 11, 1996. The authors would like to express their appreciation to their colleagues, Stephen Portnoy, Paul Newbold, and Chung-Ming Kuan for stimulating discussions regarding various aspects of this work. They would also like to thank Peter Robinson and three referees for their constructive comments. This research has been partially supported by NSF Grant SBR 93-20555.

Hettmansperger (1974), and Puri and Sen (1985) as an introduction to these ideas. A recent survey from an econometric standpoint may be found in Koenker (1996).

In an exceptional recent sequence of papers Dufour (1981) and Campbell and Dufour (1993, 1995) have considered several variants of sign and signed rank tests designed to test orthogonality restrictions including the random walk hypothesis in economic time-series models. In their simplest form these tests possess the cardinal virtues of the classical rank tests: a readily computed, exact, finite-sample distribution theory under the null and remarkable robustness under the alternative to departures from the idealized Gaussian conditions which underlie more conventional tests. Extending these results from the context of a simple null hypothesis to composite null hypotheses, however, represents a severe challenge. Campbell and Dufour (1993) suggest an ingenious bounds formulation for the random-walk case in which there is an unknown, scalar drift parameter based on Bonferroni considerations, but as they themselves comment the procedures cannot be readily extended to more complicated testing environments with additional nuisance parameters.

Traditionally, rank tests for simple nulls have been extended to composite situations by preliminary estimation, yielding *aligned* rank tests. Puri and Sen (1985) and Adichie (1984) describe this approach in considerable detail for linear model applications. Recently, Hallin and Puri (1992, 1994) have extended this approach to a broad class of testing problems for ARMA processes. Although the resulting aligned rank analogues of the autocorrelations fail to possess the fundamental invariance to monotone transformations which characterize their unaligned counterparts under white noise, certain linear functions of these autocorrelations are be shown to be asymptotically invariant and consequently asymptotically distribution free, thus providing attractive alternatives to classical, Gaussian-likelihood-based tests for many inference problems in the ARMA context.

While aligned rank tests have proven to be an extremely important device in expanding the applicability of classical rank tests, an alternative approach recently developed by Gutenbrunner and Jurečková (1992) also shows considerable promise in this respect. In this paper we adopt this new approach to construct rank based tests of the unit root hypothesis. These tests have a Gaussian limiting distribution theory in contrast to the least-squares based tests in general use, and they are considerably more powerful than the corresponding least-squares based tests when confronted by time-series with innovations exhibiting heavy tails. We note that Robinson (1993, 1994) has also recently proposed several tests of non-stationarity which possess a limiting Gaussian theory. Koul and Saleh (1995) have studied regression rankscroes for stationary AR-processes.

Section 2 provides a brief introduction to our approach in the familiar context of the linear model. Section 3 contains our primary results on unit root tests. Section 4 briefly treats extensions to models with linear trend. Some Monte Carlo evidence on

the performance of the proposed tests is presented in Section 5. Proofs of results in the body of the paper appear in the appendix.

2. Rank Tests for the Linear Model

The monograph of Hájek and Šidák (1967), which constituted a complete reappraisal of the theory of rank statistics, introduced the rank generating functions

$$\hat{a}_i(t) = \begin{cases} 1 & \text{if } t < (R_i - 1)/n \\ R_i - tn & \text{if } (R_i - 1)/n \leq t < R_i/n \\ 0 & \text{if } R_i/n \leq t. \end{cases} \tag{2.1}$$

corresponding to any sample $\{Y_1, Y_2, ..., Y_n\}$, and associated ranks $\{R_1, R_2, ..., R_n\}$. These functions "generate the ranks" of the Y's in the sense that, integrating with respect to Lesbesgue measure to obtain,

$$\hat{b}_i = \int_0^1 \hat{a}_i(t)\, dt = (R_i - 1/2)/n \tag{2.2}$$

we essentially recover the ranks. 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,

$$\hat{b}_i = \int \hat{a}_i(t)\, d\varphi(t) = \hat{a}_i(1/2) - 1/2 = \begin{cases} +1/2 & \text{if } R_i \geq n/2 + 1 \\ 0 & \text{otherwise} \\ -1/2 & \text{if } R_i \leq n/2. \end{cases} \tag{2.3}$$

The invariance of the ranks to monotone (order-preserving) transformations of the Y_i's means that the R_i's may also be viewed as the ranks of the uniform random variables $\{U_1, ..., U_n\}$ 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 (2.1) Thus the rank generating functions behave like an empirical process as the following result shows.

Theorem 2.1. *(Hájek and Šidák (1967, Thm V.3.5)) Let $c_n = (c_{1i}, ..., c_{nn})$ be a triangular array of real numbers satisfying*

$$\max(c_i - \bar{c})^2 / \sum_{i=1}^{n} (c_i - \bar{c})^2 \to 0 \tag{2.4}$$

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

$$Z_n(t) = [\sum_{i=1}^{n} (c_i - \bar{c})^2]^{-1/2} \sum_{i=1}^{n} (c_i - \bar{c}) \hat{a}_i(t) \tag{2.5}$$

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

A limiting normal theory for a broad class of linear rank statistics of the form

$$S_n = [\sum (c_{ni} - \bar{c}_n)^2]^{-1/2} \sum (c_{ni} - \bar{c}_n)\hat{b}_i \tag{2.6}$$

is immediate. In particular, for square integrable $\varphi : [0,1] \to \mathbf{R}$ we have the linear representation

$$S_n = [\sum (c_{ni} - \bar{c}_n)^2]^{-1/2} \sum (c_{ni} - \bar{c}_n)\varphi(U_i) + o_p(1), \tag{2.7}$$

and consequently, S_n is asymptotically Gaussian under the null with mean 0 and variance $A^2(\varphi) = \int (\varphi(t) - \bar{\varphi})^2 dt$ and $\bar{\varphi} = \int \varphi(t)dt$. An important consequence of the transformation to ranks is the observation that neither S_n, nor its distribution depend upon F.

It may also be noted at this point that the only computation required for the simple linear rank tests described above involves sorting and ranking the sample Y_i's. But to extend these ideas to more general models we require a more sophisticated notion of ranking, a notion provided by the work of Gutenbrunner and Jurečková (1992).

For the classical linear regression model Koenker and Bassett (1978) proposed solving

$$\hat{\beta}_n(\tau) = \mathrm{argmin}_{\mathbf{b} \in \mathbf{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i'\mathbf{b}). \tag{2.8}$$

with $\rho_\tau(u) = u(\tau - I(u < 0))$ to obtain a sequence of p-dimensional "regression quantiles" which may be regarded as estimators of the coefficients of the p-dimensional conditional quantile functions of $y|\mathbf{x}$. Problem (2.8) may be reformulated as a linear program with corresponding dual problem

$$\hat{\mathbf{a}}(\tau) = \mathrm{argmax}\,\{y'\mathbf{a} | \mathbf{X}'\mathbf{a} = (1-\tau)\mathbf{X}'\mathbf{1}_n, \mathbf{a} \in [0,1]^n\} \tag{2.9}$$

In the simple one-sample location problem, where $\mathbf{X} = \mathbf{1}_n$, $\hat{a}_i(\tau)$ specializes to the rank generating functions of Hájek and Šidák . This was first noted by Gutenbrunner and Jurečková (1992), who provided the first 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 (2.1), 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 shall see, they provide a natural means to extend linear rank statistics to more general linear models with nuisance parameters. The price of the expanded applicability of the tests, is the sacrifice of their exact distribution-freeness. Asymptotically distribution free behavior must suffice.

Now consider the partitioned linear model

$$y = \mathbf{X}\beta + \mathbf{Z}\gamma + u$$

where β and γ are p and q dimensional parameters respectively, 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, $H_n : \gamma = \gamma_0/\sqrt{n}$. GJKP propose the test statistic based on the regression rank score process, $\hat{\mathbf{a}}_n(\tau)$, for the *restricted* model,

$$T_n = S_n' \mathbf{Q}_n^{-1} S_n / A^2(\varphi) \tag{2.10}$$

where $S_n = (\mathbf{Z} - \hat{\mathbf{Z}})'\hat{\mathbf{b}}_n$, $\hat{\mathbf{b}}_n = (\int_0^1 \hat{\mathrm{a}}_i(t)\, d\varphi(t))_{i=1}^n$, $\hat{\mathbf{Z}} = \mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}$, $\mathbf{Q}_n = (\mathbf{Z} - \hat{\mathbf{Z}})'(\mathbf{Z} - \hat{\mathbf{Z}})$, with $A^2(\varphi)$ as defined following (2.7). An important feature of the test statistic T_n is that there is no explicit 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. Even tests based on the least-squares estimator require estimation of the nuisance parameter σ^2.

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

$$\eta^2 = [\omega^2(\varphi, F)/A^2(\varphi)]\gamma_0' \mathbf{Q}_0 \gamma_0$$

where $\omega(\varphi, F) = \int_0^1 f(F^{-1}(t))d\varphi(t) = -\int_0^1 \varphi(t)df(F^{-1}(t))$ and $\mathbf{Q}_0 = \lim_{n\to\infty} \mathbf{Q}_n/n$. Thus the asymptotic relative efficiency of these tests with respect to the classical $\mathcal{F}$ test takes the form

$$ARE(T_n^{\varphi}, \mathcal{F}) = \sigma^2(F)A^2(\varphi)/\omega^2(\varphi, f)$$

The following table gives the asymptotic relative efficiency of the Wilcoxon test relative to the $\mathcal{F}$-test for eight well-known F's.

TABLE 2.1: ARE of Wilcoxon vs. $\mathcal{F}$-Tests

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

The striking feature of this table is how little information is lost by transforming to ranks - at the normal model we sacrifice only 4.5% efficiency, while we would need 50 percent more observations if we foolishly used the $\mathcal{F}$-test instead of the Wilcoxon test at the double exponential distribution and *seven times* as many observations at the lognormal distribution.

For f absolutely continuous with f' absolutely integrable, we can obtain an asymptotically optimal test if we take

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

See Hájek and Šidák (II.4.3, 1967). Thus in the Gaussian case we are led to $\varphi_\Phi(s) = \Phi^{-1}(s)$. These "normal" or van der Waerden scores have the remarkable property that the resulting test has ARE relative to the $\mathcal{F}$-test which exceeds unity for all F, except the normal where ARE = 1. See Chernoff and Savage (1958) and the discussion in Hettemansperger (1984) for details.

Computing the vector $\hat{\mathbf{b}}_n = (\hat{b}_i)_{i=1}^n$ may appear to pose serious problems since it requires the entire sample path of the n-vector $\hat{\mathbf{a}}_n(\tau)$: $[0,1] \to [0,1]^n$. Fortunately, the situation is easily manageable. The function $\hat{\mathbf{a}}_n(\tau)$ is piecewise linear and continuous with breaks at $0 = \tau_0 < \tau_1 < ... < \tau_{J-1} < \tau_J = 1$. 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. Computing the entire sample path of $\hat{\mathbf{a}_n}(\tau)$ for problems with $n = 100$ and $p = 4$ requires about half a second on a Sparc 20, while problems with $n = 1000$ and $p = 7$ require about 70 seconds.

Three score functions play prominent roles in the application of rank tests and we use all three in the Monte Carlo reported below. The piecewise linearity of $\hat{\mathbf{a}}_n(\tau)$ 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,

$$\hat{b}_i = -\int_0^1 (s - 1/2)\, d\,\hat{\mathrm{a}}_i(s) = \sum_{j=1}^{J} 1/2(\hat{\mathrm{a}}_i(\tau_{j+1}) + \hat{\mathrm{a}}_i(\tau_j))\,(\tau_{j+1} - \tau_j) - 1/2.$$

(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 $\hat{\mathrm{a}}_i'(\tau)$ is piecewise constant, we have:

$$\hat{b}_i = -\int_0^1 \Phi^{-1}(s)\, d\,\hat{\mathrm{a}}_i(s) = \sum_{j=1}^{J} \hat{\mathrm{a}}_i'(\tau_j)\big[\phi(\Phi^{-1}(\tau_{j+1})) - \phi(\Phi^{-1}(\tau_j))\big].$$

(iii) Sign-Median scores: $\varphi(s) = \frac{1}{2} sgn(s - \frac{1}{2})$. In this case we have,

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

3. Rank Tests for the Unit Root Hypothesis

Consider discrete time-series $\{y_t\}_{t=1}^{\infty}$ generated by the model

$$y_t = y_{t-1} + u_t \tag{3.1}$$

$$u_t = \phi_1\, u_{t-1} + \cdots + \phi_p\, u_{t-p} + e_t \tag{3.2}$$

where $\{e_t\}_{t=1}^{\infty}$ is an iid sequence with mean 0 and variance σ_e^2. We will assume throughout that the roots of $1 - \phi_1 z - \cdots - \phi_p z^p = 0$ all lie outside the unit circle. For convenience we assume $y_0 = 0$, so we may write $y_t = \sum_{s=1}^{t} u_s$.

Our tests will be based on the augmented Dickey-Fuller(1979) model

$$\Delta y_t = \beta_0 + \gamma y_{t-1} + \sum_{j=1}^{p} \beta_j \Delta y_{t-j} + e_t, \tag{3.3}$$

where we wish to test the hypothesis, $H_0 : \gamma = 0$, versus the local alternatives, $H_T : \gamma_T = \gamma_0/T$. Partitioning the "design" matrix as in the previous section, we will write $\mathbf{x}_t = (1, \Delta y_{t-1}, \cdots, \Delta y_{t-p})', z_t = y_{t-1}$, and in obvious matrix notation $\hat{z} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'z$. Following GJKP, we wish to consider tests based on the regression rank score process $\hat{\mathbf{a}}_T(\tau)$ *for the restricted model of* (3.3) with $\gamma = 0$, and test statistics of the form

$$S_T = \sum (z_t - \hat{z}_t)\hat{b}_t / \sum (z_t - \hat{z}_t)^2)^{1/2} \tag{3.4}$$

where $\hat{b}_t = \int_0^1 \hat{a}_t(\tau) d\varphi(\tau)$ for some appropriately chosen score function φ.

We shall see that under H_0, in contrast to the regression case described above, the limiting behavior of S_T depends upon certain nuisance parameters. Remarkably, a relatively simple modification of S_T can be made which restores its asymptotically Gaussian behavior. The modification required is closely related to the modified l_1-type tests considered by Hercé (1990).

Consider the bivariate process $\{(u_t, v_t)\}$ where $v_t = \varphi(F(e_t))$. Without loss of generality we may choose φ such that $\mathbf{E}v_t = 0$, so $\mathbf{Var}(v_t) = A^2(\varphi)$. Let

$$\begin{aligned} \Sigma &= \lim_{T\to\infty} \mathbf{Var}(T^{-1/2} \sum (u_t, v_t)') \\ &= \lim_{T\to\infty} T^{-1} \mathbf{E}[\sum_{t=1}^{T} (u_t, v_t) \sum_{t=1}^{T} (u_t, v_t)'] \\ &= \Sigma_0 + \Sigma_1 + \Sigma_1' \end{aligned}$$

where

$$\Sigma_0 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \mathbf{E}[(u_t, v_t)'(u_t, v_t)] \tag{3.5}$$

$$\Sigma_1 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{S=1}^{t-1} \mathbf{E}[(u_s, v_s)'(u_t, v_t)]. \tag{3.6}$$

Finally, denote the lower triangular form of the Cholesky decomposition of Σ as

$$\Sigma^{1/2} = \Sigma_{11}^{-1/2} \begin{bmatrix} \Sigma_{11} & 0 \\ \Sigma_{12} & \Delta^{1/2} \end{bmatrix} \tag{3.7}$$

where $\Delta = |\Sigma|$. These quantities will serve to standardize our test statistic.

The familiar augmented Dickey-Fuller(1979) test of H_0 is based on the following OLS estimate of γ in (3.3),

$$\hat{\gamma}_T = [\sum (z_t - \hat{z}_t)^2]^{-1} \sum (z_t - \hat{z}_t)\Delta y_t \tag{3.8}$$

The modification of S_T alluded to above may be viewed as annihilating an unfortunate non-Gaussian term in the linear asymptotic representation of S_T by subtracting a funtion of $\hat{\gamma}_T$. Explicitly, denote as in Hercé (1990) and Phillips (1987),

$$\begin{aligned}
\sigma^2 = \Sigma_{11} &= \lim_{T\to\infty} T^{-1}\mathbf{E}\left(\sum u_t\right)^2 \\
\sigma_u^2 &= \lim_{T\to\infty} T^{-1}\sum \mathbf{E}u_t^2 \\
\sigma_z^2 &= T^{-2}\sum (z_t - \hat{z}_t)^2.
\end{aligned}$$

Then we may write the modified test statistic as

$$\tilde{S}_T = \frac{\sigma}{\Delta^{1/2}} S_T - \frac{\Sigma_{12}\sigma_z}{\Delta^{1/2}\sigma}[T\hat{\gamma}_T - \frac{\sigma^2 - \sigma_u^2}{2\sigma_z^2}] \tag{3.9}$$

The term in square brackets is closely related to the test statistics of Phillips and Perron (1988). The limiting behavior of $\tilde{S}_T$ is given in the following theorem.

Theorem 3.1. *In model (3.3) with $\{y_t\}_{t=1}^T$ generated as in (3.1–2) assume:*

(i.) F has an absolutely continuous density f, and finite Fisher Information;

(ii.) $\varphi : [0,1] \to \mathbf{R}$ is a nondecreasing square integrable function on the interval $[\tau_0, 1-\tau_0]$ for some $\tau_0 \in (0, .5)$ and φ is constant on the set $[0,\tau_0] \cup [1-\tau_0, 1]$.

Then, under H_0, $\tilde{S}_T$ is asymptotically standard normal, and under the sequence of local alternatives H_T, $\tilde{S}_T - \eta_T$ is asymptotically standard normal where,

$$\eta_T = \Delta^{-1/2}\xi_J(\gamma_0)\gamma_0[\omega(\varphi, F)\sigma^2 - \Sigma_{12}]$$

and $\xi_J^2(\delta) = \int_0^1 (J_\delta(r) - \bar{J}_\delta)^2 dr$, $J_\delta(r) = \int_0^1 e^{(r-s)\delta} dW(s)$, $W(s)$ denotes standard Brownian motion, and $\bar{J}_\delta = \int_0^1 J_\delta(r)dr$.

The Ornstein-Uhlenbeck process, $J_\delta(r)$, also plays an important role in the asymptotic theory of least-squares-based tests for unit roots under local alternatives. See e.g., Phillips(1987a, 1988) and Cox and Llatas (1991) for further details.

The limiting Gaussian behavior of $\tilde{S}_T$ may appear surprising in view of the fact that the familiar least squares based test statistics have a notoriously non-Gaussian limiting distribution as tabulated by Dickey and Fuller (1979). Under the local alternatives H_T, the least squares theory is also more complicated, giving rise to limiting Ornstein-Uhlenbeck processes.

The restriction of φ to be constant in the tails avoids certain technical difficulties associated with establishing the uniformity of the empirical process approximation for the tails of the regression rankscore process. In GJKP(1993) this trimming condition is relaxed in the regression setting at the expense of somewhat more stringent conditions on F and considerable added complexity in the proofs.

The restriction on the nature of the dependence in the u_t process to autoregressions of order p could undoubtedly be relaxed at some cost of further complicating the proof of Lemma A.1.

3.1. **Estimating Nuisance Parameters.** Here we describe briefly some suggested procedures for estimating the nuisance parameters appearing in the test statistic $\tilde{S}_T$. We rely on estimators of the form

$$\hat{\Sigma} = \hat{\Sigma}_0 + \sum_{j=1}^{l} \omega(j,l)(\hat{\Sigma}_j + \hat{\Sigma}_j') \tag{3.10}$$

where $\hat{\Sigma}_j = T^{-1}\sum_{t=j+1}^{T} h_t h_{t-j}'$ for $j = 0, 1, \cdots, l$ and $h_t = (\hat{u}_t, \hat{v}_t)$ with $\hat{u}_t = \Delta y_t - \hat{e}_t$ and $v_t = \varphi(\hat{F}(\hat{e}_t))$. In the Monte Carlo reported in Section 4 we have taken the $\hat{e}_t$'s as the residuals from the l_1-fit of the null model, i.e., either (3.3) or (4.1) with $/\gamma = 0$, and $\hat{F}(\hat{e}_t) = (\#\{\hat{e}_s \leq \hat{e}_t, s = 1, ..., T\} - .5)/T$. The integer l, generally referred to as the lag truncation parameter, should grow with sample size, say like $T^{1/3}$. Of many possible weight functions, we chose to use $\omega(j,l) = 1 - j/(l+1)$. We may then write the elements of $\hat{\Sigma}$ explicitly as

$$\begin{aligned}
\hat{\Sigma}_{11} &= \hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T} \hat{u}_t^2 + 2T^{-1}\sum_{j=1}^{l} w(j,l)\sum_{t=1}^{T} \hat{u}_t\hat{u}_{t-j} \\
\hat{\Sigma}_{12} &= T^{-1}\sum_{t=1}^{T} \hat{u}_t\hat{v}_t + T^{-1}\sum_{j=1}^{l} w(j,l)\sum_{t=1}^{T} (\hat{u}_t\hat{v}_{t-j} + \hat{v}_t\hat{u}_{t-j}) \\
\hat{\Sigma}_{22} &= T^{-1}\sum_{t=1}^{T} \hat{v}_t^2 + 2T^{-1}\sum_{j=1}^{l} w(j,l)\sum_{t=1}^{T} \hat{v}_t\hat{v}_{t-j}
\end{aligned}$$

Finally, we estimate σ_u^2 in the obvious way by $\hat{\sigma}_u^2 = T^{-1}\sum \hat{u}_t^2$. These estimators, which may be regarded as kernel estimators of the spectral density matrix of (u_t, v_t) at frequency zero, are consistent as shown, for example, in Newey and West (1987), Phillips (1987a) and Hercé (1990), and may be used in place of the corresponding population parameters in $\tilde{S}_T$ as the following result indicates.

Corollary 3.1. *Under the conditions of Theorem 3.1, the nuisance parameters appearing in the definition of $\tilde{S}_T$ may be replaced by the estimates given above without altering its asymptotic behavior.*

3.2. **Infinite Variance Innovations.** An even more striking contrast between rank tests and their least-squares-based counterparts is seen in the case of infinite variance innovations processes. For least-squares-based tests, Chan and Tran (1989) have studied the model (3.1) with u_t iid with common distribution in the domain of attraction of a stable law with parameter α. For $\alpha \in (0,1)$ they show that the simple OLS estimate $\hat{\gamma}_T = \sum y_{t-1} y_t / \sum y_{t-1}^2$ converges weakly to following limiting distribution,

$$T(\hat{\gamma}_T - 1) \Rightarrow \frac{S^2(1) - v(1)}{2\int_0^1 S^2(t)dt}$$

where $S(t) = \sum_{k=1}^{T} \delta_k Z_k((U_{\mathbf{R}} \leq t)$, $V(t) = \sum Z_k^2 I(U_k < t) \quad Z_k = (\sum \eta_i)^{-1/\alpha}$ and the random variables δ_k, U_k and η_i are iid with Bernoulli (± 1 with probability 1/2), uniform (on $[0, 1]$) and exponential (mean 1), distributions respectively.

In contrast, Knight (1989) establishes that under these conditions the simple l_1 estimator

$$\check{\gamma}_T = \arg\min_\gamma \sum_{t=1}^{T} |y_t - \gamma y_{t-1}|$$

satisfies,

$$\alpha_T \sqrt{T}(\check{\gamma}_T - 1) \Rightarrow \frac{\int_0^1 V(s)dW(s)}{2f(0)\int V^2(s)ds}$$

where $f = F'$ is the density of the innovation sequence and $\alpha_T = T^{1/\alpha} L(T)$ for some slowly varying function L. Thus, for example, in the Cauchy case, where $\alpha = 1$, we have convergence of $\check{\gamma}_T$ to γ at rate $T^{-3/2}$ – considerably faster than T^{-1} rate of the least-squares estimator. The l_1 result can be further simplified, as noted by Knight, if we replace the deterministic norming by $\alpha_T \sqrt{T}$ by stochastic norming using the sequence $\zeta_T = \{(\sum_{t=2}^{T} y_{t-1}^2)^{1/2}\}$. This yields the conclusion that $\zeta_T(\hat{\gamma}_T - 1)/(2f(0))$ is asymptotically standard Gaussian. Qualitatively similar results apply to rank tests in the infinite variance setting with the added advantage that the need to estimate the nuisance parameter $f(0)$ is also avoided. For further details see Hasan (1993).

4. Rank Tests for the Unit Root Hypothesis with Trend

In many econometric applications it is important to permit a deterministic trend as a part of the maintained hypothesis in unit root testing. In the augmented Dickey-Fuller framework this is typically accomplished by estimating the model

$$\Delta y_t = \beta_0 + \gamma y_{t-1} + \delta t + \sum_{s=1}^{p} \beta_j \Delta y_{t-s} + e_t \tag{4.1}$$

in place of the model (3.3) considered in previous section.

The trend effect can be easily accommodated in the rank test approach in the same way simply by incorporating the trend into the x_t vector, $x_t = (1, t, \Delta y_{t-1}, \cdots, \Delta y_{t-p})$

and proceeding as in the previous section. No changes are required in the asymptotic theory described there for the model (3.1-2), with the obvious exception that we must interpret $\hat{z}_t$ as the projection of $z_t = y_{t-1}$ onto the expanded x_t-space which now includes the trend component as is apparent in the Monte Carlo reported in the next section. Of course, finite sample performance of the tests *is* affected by the inclusion of the trend component.

It is also of interest to know how the foregoing theory is affected by the introduction of a drift parameter in (3.1). In particular we may wish to study the behavior of the rank test statistics based on model (4.1) where the process $\{y_t\}_{t=1}^{\infty}$ is generated from the model

$$y_t = \mu + y_{t-1} + u_t \tag{4.2}$$

$$u_t = \phi_1 u_{t-1} + \cdots + \phi_p u_{t-1} + e_t \tag{4.3}$$

and the $\{e_t\}_{t=1}^{\infty}$ sequence is iid as before. Using the invariance result Lemma A.9 it is straightforward to show that S_T is invariant to the drift parameter μ and since the other components of the rank test statistics $\tilde{S}_T$ may also be chosen to be invariant in this way as noted by Phillips(1987), it follows that the asymptotics for $\tilde{S}_T$ also hold for the unit root with drift model (4.2). We summarize the results of this section in the following manner.

Theorem 4.1. *Under conditions (i.-ii.) of Theorem 3.1, with $\{y_t\}_{t=1}^{T}$ generated according to (4.2-3), let $\tilde{S}_T$ denote the rank statistic (3.9) with x_t augmented to include the trend component t and $\hat{z}_t$ modified accordingly. Then $\tilde{S}_T$ is asymptotically standard normal.*

5. Monte Carlo Evidence

In this section we report on a Monte Carlo experiment designed to evaluate the finite sample performance of the rank tests proposed above. Much more detailed Monte Carlo results may be found in Hasan (1993).

We begin by considering the first order autoregressive model,

$$y_t = y_{t-1} + u_t; \qquad t = 1, \cdots, T \tag{5.1}$$

$$u_t = .5u_{t-1} + e_t. \tag{5.2}$$

We set $y_0 = 0$ and generate e_t's as iid from three different distributions: the normal, Student's **t** distribution on three degrees of freedom ($\mathbf{t}_3$) and the Cauchy distribution. These distributions represent a considerable range of tail behavior. Although the Cauchy obviously does not satisfy our regularity condition (i) of Theorem 3.1 it is nevertheless of interest to see how the proposed rank tests perform under such severe conditions.

We report only the results based on the correctly specified ADF model

Innovation Process	Test Statistic	l=4				l=6			
		Power			Size	Power			Size
		.90	.95	.99	1.00	.90	.95	.99	1.00
Normal	ADF	.25	.116	.051	.051	.25	.116	.051	.051
	Wilcoxon	.146	.121	.055	.065	.151	.125	.094	.058
	Normal	.162	.126	.066	.069	.165	.145	.088	.077
	Sign	.106	.116	.076	.076	.110	.101	.089	.066
t_3	ADF	.248	.103	.061	.054	.248	.103	.061	.054
	Wilcoxon	.387	.271	.098	.069	.389	.273	.129	.082
	Normal	.298	.219	.085	.062	.302	.225	.113	.068
	Sign	.386	.267	.112	.097	.388	.266	.139	.065
Cauchy	ADF	.161	.064	.052	.075	.161	.064	.052	.075
	Wilcoxon	.988	.965	.655	.089	.971	.946	.646	.102
	Normal	.977	.946	.595	.088	.952	.925	.613	.112
	Sign	.988	.961	.642	.084	.963	.944	.665	.096

TABLE 5.1. EMPIRICAL POWER AND SIZE OF FOUR UNIT ROOT TESTS: Sample size = 100, number of replications =1000, nominal level of significance of the test is 5%, innovation process : $u_t = .5u_{t-1} + e_t$, null model : $\Delta y_t = \beta_0 + \beta_1 \Delta y_{t-1} + e_t$. Power comparisons are based on stationary alternatives with $1+\gamma \in \{.9, .95, .99\}$. Note that the augmented Dickey-Fuller test statistic is independent of the lag truncation parameter, l.

$$\Delta y_t = \beta_0 + \gamma y_{t-1} + \beta_1 \Delta y_{t-1} + e_t, \tag{5.3}$$

with two choices of the lag truncation parameter $l = 4$ and $l = 6$. Table 5.1 reports the empirical power and size of the three rank tests as well as the standard ADF test based on 1000 replications for $T = 100$. The tables show quite clearly how the tail behavior of the innovation process affects the power of the unit root test. Not surprisingly, at the normal model the ADF test performs quite well, achieving power .25 for $1 + \gamma = .9$. The rank tests are somewhat less powerful at the normal model with the normal score test achieving power .16. The Wilcoxon version of the rank test achieves roughly twice the power of the ADF test at the t_3 distribution, while at the Cauchy the rank tests all have power near one, in sharp contrast to the ADF test which has power roughly equal to size. It may appear surprising that it is actually *easier* to distinguish unit roots from stationary alternatives for longer tailed innovation

Innovation Process	Test Statistics	Lag Truncation Parameter l=4							
		Without Size Corrected				Size Corrected			
		Power			Size	Power			Size
		.90	.95	.99	1.00	.9	.95	.99	1.00
Normal	ADF	.146	.082	.042	.048	.146	.082	.042	.048
	Wilcoxon	.038	.027	.034	.040	.064	.038	.052	.055
	Normal	.054	.030	.041	.043	.062	.039	.049	.053
	Sign	.064	.054	.048	.053	.077	.071	.062	.067
t_3	ADF	.156	.074	.047	.056	.156	.074	.047	.056
	Wilcoxon	.276	.142	.069	.042	.254	.122	.050	.035
	Normal	.202	.103	.054	.046	.189	.092	.053	.036
	Sign	.262	.142	.063	.067	.195	.095	.055	.037
Cauchy	ADF	.107	.076	.064	.053	.107	.076	.064	.053
	Wilcoxon	.954	.833	.491	.057	.950	.873	.473	.047
	Normal	.926	.843	.437	.047	.922	.839	.427	.043
	Sign	.955	.897	.509	.070	.943	.886	.488	.060

TABLE 5.2. EMPIRICAL POWER AND SIZE FOR DETRENDED MODEL: Sample size = 100, number of replications = 1000, nominal significance of the test 5 %, innovation process: $u_t = .5u_{t-1} + e_t$, null model $\Delta y_t = \beta_0 + \delta t + \beta_1 \Delta y_{t-1} + e_t$. Power comparisons are based on stationary alternatives with $1 + \gamma \in \{.9, .95, .99\}$.

processes. This is obviously not the case if we restrict attention to least-squares based tests like the ADF, whose performance deteriorates as the innovations process becomes heavier tailed. Yet further reflection suggests that we may regard heavy-tailed innovations as offering a "more informative natural experiment" concerning γ which the rank tests effectively exploit.

For the model with trend, we adopt the ADF model,

$$\Delta y_t = \beta_0 + \delta t + \beta_1 \Delta y_{t-1} + e_t, \tag{5.4}$$

and generate data under the null from (5.1-2) with $p = 1$ and the innovation sequence again generated from the normal, t_3, and Cauchy family. Recall that the asymptotic behavior of the rank tests of the unit root hypothesis are invariant to the drift parameter in (4.2) so there is no loss of generality in setting it equal to 0 here. However, it is important to note that we must ensure that the nuisance parameters are also estimated in a fashion that is invariant to the drift and to this end we use the l_1-

Innovation Process	Test Statistic	Lag Truncation Parameter l=6							
		Without Size Correction				With Size Correction			
		Power			Size	Power			Size
		.90	.95	.99	1.00	.9	.95	.99	1.00
Normal	ADF	.146	.082	.042	.048	.146	.082	.042	.048
	Wilcoxon	.135	.096	.070	.070	.111	.077	.057	.053
	Normal	.162	.102	.088	.088	.120	.077	.057	.060
	Sign	.114	.079	.066	.067	.108	.072	.060	.065
t_3	ADF	.156	.074	.047	.056	.156	.074	.047	.056
	Wilcoxon	.344	.192	.081	.095	.256	.116	.045	.061
	Normal	.291	.156	.080	.097	.191	.087	.038	.056
	Sign	.310	.171	.080	.072	.209	.096	.040	.041
Cauchy	ADF	.107	.076	.064	.053	.107	.076	.064	.053
	Wilcoxon	.962	.881	.500	.067	.948	.864	.457	.056
	Normal	.925	.840	.439	.066	.912	.829	.418	.055
	Sign	.938	.894	.521	.067	.901	.865	.479	.052

TABLE 5.3. EMPIRICAL POWER AND SIZE FOR DETRENDED MODEL: Configuration as in previous table.

residuals $\hat{e}_t$ from (5.4) and set

$$\begin{aligned} \hat{u}_1 &= \hat{\delta} + \hat{e}_1 \\ \hat{u}_t &= \hat{\delta} t + \hat{\beta}_1 \hat{u}_{t-1} + \hat{e}_t \end{aligned} \tag{5.5}$$

for $t = 2, \ldots, T$. The consistency and invariance properties of $\hat{u}_t$ and hence the estimated nuisance parameters based on them follow from the properties of $\hat{e}_t$. In Table 5.2 and 5.3 we report the size and power performance of the rank tests *vis à vis* the corresponding ADF test for the lag truncation 4 and 6 cases respectively.

These tables exhibit similar features to the previous simple case without trend with one important exception: for $l = 4$ the power of the rank tests for normal innovations is very poor and estimated size is also low. This loss of power due to detrending is familiar in the unit root literature and can also be seen for ADF test. To adjust for the size distortion we also report size corrected results for the rank tests. The size corrected power of the rank tests is quite close to ADF for normal innovations, at least for $l = 6$, and substantially outperforms the ADF test for t_3 and Cauchy errors.

DEPARTMENT OF ECONOMICS, ILLINOIS STATE UNIVERSITY, NORMAL, IL, 61790
DEPARTMENT OF ECONOMICS, UNIVERSITY OF ILLINOIS, CHAMPAIGN, IL, 61820

Appendix A.

We begin with a lemma which adapts Lemma 3.1 of GJKP to the present setting and provides a quadratic approximation of the quantile regression objective function uniformly in τ and the unknown parameters of the model. It is essential to the development of the asymptotic linear representation of our test statistic S_T. Consider the augmented Dickey-Fuller model (3.3) with the sequence $\{y_t\}_{t=1}^{\infty}$ generated according to (3.1). Let

$$\tilde{\mathbf{x}}_t = (\mathbf{x}_t, z_t) = (1, \Delta y_{t-1}, \ldots, \Delta y_{t-p}, y_{t-1}/\sqrt{T}) \tag{A.1}$$

and note that

$$\max_{1\le t\le T} \|\tilde{\mathbf{x}}_t\| = o_p(T^{1/2}) \tag{A.2}$$

and

$$\mathbf{d}_T = T^{-1}\sum \tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t' = O_p(1). \tag{A.3}$$

Lemma A.1. *Under Condition (i.) of Theorem 3.1, for any fixed $\tau_0 \in (0,.5)$, and $C \ge 0$, as $T \to \infty$,*

$$\sup\{|R_T(\mathbf{b},\tau)| : \|\mathbf{b}\| < C; \tau_0 \le \tau \le 1-\tau_0\} \to^p 0 \tag{A.4}$$

where

$$R_T(\mathbf{b},\tau) = \kappa(\tau)\sigma_\tau^{-1}\sum_{t=1}^{T}\left[\rho_\tau(e_{t\tau} - T^{-1/2}\sigma_\tau\tilde{\mathbf{x}}_t'\mathbf{b}) - \rho_\tau(e_{t\tau})\right]$$
$$+ T^{-1/2}\kappa(\tau)\sum_{t=1}^{T}\tilde{\mathbf{x}}_t'\mathbf{b}\,\psi_\tau(e_{t\tau}) - \frac{1}{2}\mathbf{b}'\mathbf{d}_T\mathbf{b},$$

$e_{t\tau} = e_t - F^{-1}(\tau)$, $\kappa(\tau) = (\tau(1-\tau))^{-1/2}$, $\psi_\tau(x) = \tau - I[x<0]$, *and* $\sigma_\tau = \kappa(\tau)/f(F^{-1}(\tau))$.

Proof: The proof employs a truncation method. First we show

$$\sup_S |R_T(\mathbf{b},\tau)| I[\|\tilde{\mathbf{x}}_t\| \le B] \to^p 0 \tag{A.5}$$

for $S \equiv \{(\mathbf{b},\tau) : \tau_0 \le \tau \le 1-\tau_0; \|\mathbf{b}\| \le C\}$. Then we evaluate the approximation error due to the truncation.

We begin by proving the asymptotic negligibility of truncated R_T for a fixed τ and $\mathbf{b}$ such that $\|\mathbf{b}\| \le C$. We then extend the result showing uniformity in τ and $\mathbf{b}$ by applying a chaining argument. Define $\delta_{Tt} = T^{-1/2}\sigma_\tau\tilde{\mathbf{x}}_t'\mathbf{b}$ and

$$r_t(\mathbf{b},\tau) = \kappa(\tau)\sigma_\tau^{-1}\left[\rho_\tau(e_{t\tau} - \delta_{Tt}) - \rho_\tau(e_{t\tau})\right] + T^{-1/2}\kappa(\tau)\tilde{\mathbf{x}}_t'\mathbf{b}\psi_\tau(e_{t\tau}) - \frac{1}{2T}(\tilde{\mathbf{x}}_t'\mathbf{b})^2 \tag{A.6}$$

so that $R_T(\mathbf{b},\tau) = \sum_{t=1}^{T} r_t(\mathbf{b},\tau)$. Thus,

$$r_t(\mathbf{b},\tau) = \kappa(\tau)\sigma_\tau^{-1}\{(e_{t\tau} - \delta_{Tt})I[\delta_{Tt} < e_{t\tau} < 0] + (\delta_{Tt} - e_{t\tau})I[0 < e_{t\tau} < \delta_{Tt}]\} - \frac{1}{2T}(\tilde{\mathbf{x}}_t'\mathbf{b})^2 \tag{A.7}$$

Now define the following truncated variables:

$$A_{Tt} = -\frac{1}{2T}(\tilde{\mathbf{x}}_t'\mathbf{b})^2 I(\|\tilde{\mathbf{x}}_t\| \leq B) \tag{A.8}$$

$$B_{Tt} = \kappa(\tau)\sigma_\tau^{-1}(e_{t\tau} - \delta_{Tt})I[\delta_{Tt} < e_{t\tau} < 0]I(\|\tilde{\mathbf{x}}_t\| \leq B) \tag{A.9}$$

$$C_{Tt} = \kappa(\tau)\sigma_\tau^{-1}(\delta_{Tt} - e_{t\tau})I[0 < e_{t\tau} < \delta_{Tt}]\}I(\|\tilde{\mathbf{x}}_t\| \leq B) \tag{A.10}$$

so that we can write

$$\tilde{r}_t(\mathbf{b},\tau) = -A_{Tt} + B_{Tt} + C_{Tt} \tag{A.11}$$

$$\tilde{R}_T(\mathbf{b},\tau) = \sum_{t=1}^{T} \tilde{r}_t(\mathbf{b},\tau) \tag{A.12}$$

We wish to show that for any $\lambda > 0$, there exists a constant $c > 0$, and $0 \leq \delta \leq 1/4$, such that

$$P(|\tilde{R}_T(\mathbf{b},\tau)| \geq (\lambda+1)T^{-\delta}) \leq cT^{-\lambda} \tag{A.13}$$

To do this, we use the Markov inequality

$$P(|\tilde{R}_T(\mathbf{b},\tau)| \geq s_T) \leq exp(-vs_T)(M(v) + M(-v)), \quad v > 0 \tag{A.14}$$

where $M(v) = \mathbf{E}\{\exp(v\tilde{R}_T(\mathbf{b},\tau))\}$ is the moment generating function of $\tilde{R}_T$. Note that from (A.8)

$$A_{Tt} \leq \frac{1}{2T}\|\tilde{\mathbf{x}}_t\|^2\|\mathbf{b}\|^2 I(\|\tilde{\mathbf{x}}_t\| \leq B) \leq \frac{1}{2T}B^2C^2 \leq O(T^{-1}) \tag{A.15}$$

Using the definitions in (A.8) – (A.11) we can see that

$$|\tilde{r}_t(\mathbf{b},\tau)| \leq A_{Tt} + 2\kappa(\tau)T^{-1/2}|\tilde{\mathbf{x}}_t\mathbf{b}|I(\|\tilde{\mathbf{x}}_t\| \leq B) \leq O(T^{-1}) + O(T^{-1/2}) = O(T^{-1/2}) \tag{A.16}$$

Therefore, if $v\tilde{r}_t$ is bounded in probability, i.e. $0 < v < T^{1/2}$, Taylor series expansion yields

$$\log M_{\tilde{r}_t}(v) \leq v\mathbf{E}\left[\tilde{r}_t(\mathbf{b},\tau)\right] + cv^2\,\mathbf{Var}\left[\tilde{r}_t(\mathbf{b},\tau)\right] \tag{A.17}$$

for some constant c. Denote by $\mathcal{F}_{t-1} = \sigma(y_0, y_1, \cdots, y_{t-1})$, the smallest σ-algebra generated by $(y_0, y_1, \cdots, y_{t-1})$ so that from (A.17)

$$\log\left(M_{\tilde{r}_t}(v)|\mathcal{F}_{t-1}\right) \leq v\mathbf{E}\left[\tilde{r}_t|\mathcal{F}_{t-1}\right] + cv^2\,\mathbf{Var}\left[\tilde{r}_t|\mathcal{F}_{t-1}\right]. \tag{A.18}$$

Therefore,

$$\left(M_{\tilde{r}_t}(v)|\mathcal{F}_{t-1}\right) \leq \exp\left\{v\mathbf{E}\left[\tilde{r}_t|\mathcal{F}_{t-1}\right] + cv^2\,\mathbf{Var}\left[\tilde{r}_t|\mathcal{F}_{t-1}\right]\right\}. \tag{A.19}$$

Now to compute conditional expectations and conditional variances of $\tilde{r}_t$, we observe the following:

(i) $\mathbf{E}\left[A_{Tt}|\mathcal{F}_{t-1}\right] = A_{Tt}$, because A_{Tt} is $\mathcal{F}_{t-1}$-measurable.

(ii) It is shown in Appendix A.5 of Hasan (1993), that for $\delta_{Tt} < 0$,

$$\mathbf{E}\left[B_{Tt}|\mathcal{F}_{t-1}\right] \leq \frac{1}{2T}(\tilde{\mathbf{x}}_t'\mathbf{b})^2 I[\|\tilde{\mathbf{x}}_t\| \leq B]c[T^{-3/2}\|\tilde{\mathbf{x}}_t\|^3 + T^{-2}\|\tilde{\mathbf{x}}_t\|^4]I[\|\tilde{\mathbf{x}}_t\| \leq B] \tag{A.20}$$

(iii) Similarly for $\delta_{Tt} > 0$, the same inequality holds for $\mathbf{E}\left[C_{Tt}|\mathcal{F}_{t-1}\right]$.

Thus using (A.8), (A.20) we can conclude that

(A.21)
$$\mathbf{E}\left[\tilde{r}_t|\mathcal{F}_{t-1}\right] \leq c\left[T^{-3/2}\|\tilde{\mathbf{x}}_t\|^3 + T^{-2}\|\tilde{\mathbf{x}}_t\|^4\right] I(\|\tilde{\mathbf{x}}_t\| \leq B) \leq c\left[T^{-3/2} + T^{-2}\right] = O(T^{-3/2})$$

since $\|\tilde{\mathbf{x}}_t\| I(\|\tilde{\mathbf{x}}_t\| \leq B) < \infty$. We may write

$$\begin{aligned} . \mathbf{Var}\left(\tilde{r}_t|\mathcal{F}_{t-1}\right) &= \mathbf{Var}(B_{Tt}|\mathcal{F}_{t-1}) + \mathbf{Var}(C_{Tt}|\mathcal{F}_{t-1}) \\ &\leq \mathbf{E}\left[(B_{Tt}^2 + C_{Tt}^2)|\mathcal{F}_{t-1}\right] \\ &\leq T^{-3/2} c\|\tilde{\mathbf{x}}_t\|\|\mathbf{b}\| I[\|\tilde{\mathbf{x}}_t\| \leq B] \\ &\leq cT^{-3/2} \end{aligned} \tag{A.22}$$

Substituting (A.22) and (A.21) in (A.19), we have

$$M_{\tilde{r}_t}(v) \leq \exp\left\{vcT^{-3/2} + cv^2T^{-3/2}\right\}. \tag{A.23}$$

for $t = 1, 2, \ldots, T$. Finally, we compute the moment generating function of $\tilde{R}_T$

$$\begin{aligned} \mathbf{E}\left[\exp\left\{v\tilde{R}_T\right\}\right] &= \mathbf{E}\left[\exp\left\{v\sum_{t=1}^{T}\tilde{r}_t\right\}\right] \\ &= \mathbf{E}\left[\mathbf{E}\left[\exp\left\{v\tilde{R}_T\right\}\exp\left\{v\sum_{t=1}^{T-1}\tilde{r}_t\right\}|\mathcal{F}_{T-1}\right]\right] \\ &= \mathbf{E}\left[\exp\left\{v\sum_{t=1}^{T-1}\tilde{r}_t\right\}\mathbf{E}\left[\exp\left\{v\tilde{r}_t\right\}|\mathcal{F}_{T-1}\right]\right] \\ &= \mathbf{E}\left[\exp\left\{v\sum_{t=1}^{T-1}\tilde{r}_t\right\}\left(M_{\tilde{r}_t}(v)|\mathcal{F}_{T-1}\right)\right] \end{aligned} \tag{A.24}$$

since $\sum_{t=1}^{T-1}\tilde{r}_t$ is $\mathcal{F}_{T-1}$-measurable. Thus,

$$\begin{aligned} \mathbf{E}\left[\exp\left\{v\tilde{R}_T\right\}\right] &\leq \mathbf{E}\left[\exp\left\{v\sum_{t=1}^{T-1}\tilde{r}_t\right\}\exp\{vcT^{-3/2} + cv^2cT^{-3/2}\}\right] \\ &= \exp\{vcT^{-3/2} + cv^2cT^{-3/2}\}\mathbf{E}\left[\exp\left\{v\sum_{t=1}^{T-1}\tilde{r}_t\right\}\right] \end{aligned} \tag{A.25}$$

And, recursively substituting (A.23) for each t conditional on its respective $\mathcal{F}_{t-1}$, i.e. following the same kind of argument as (A.24) for $(M_{\tilde{R}_{T-1}}|\mathcal{F}_{T-2})$, $(M_{\tilde{R}_{T-2}}|\mathcal{F}_{T-3}), \ldots$ we can write

(A.26)
$$\mathbf{E}\left[\exp\left\{v\tilde{R}_T\right\}\right] = \exp\left\{\sum_{t=1}^{T}vcT^{-3/2} + \sum_{t=1}^{T}v^2cT^{-3/2}\right\} \leq \exp\left\{vcT^{-1/2} + v^2cT^{-1/2}\right\}$$

Now choosing $v = T^{\delta}\log T < O(T^{1/2})$, where $0 < \delta < 1/4$, we have established (A.17). Hence using (A.14) and setting $s_T = (\lambda+1)T^{-\delta}$

(A.27)

$$\begin{aligned} P(|\tilde{R}_T(\mathbf{b},\tau)| \geq (\lambda+1)T^{-\delta}) &\leq \exp\left\{-(\lambda+1)\log T \quad + (c\frac{\log T}{T^{-\delta}})\cdot T^{-\frac{1}{2}} + (c\frac{\log^2 T}{T^{-2\delta}})T^{-\frac{1}{2}}\right\} \\ &= T^{-\lambda}\cdot\frac{T^{\lambda-1}}{\exp\{\lambda\log T\}}\exp\left\{(c\frac{\log T}{T^{-\delta}})\cdot T^{-\frac{1}{2}} + (c\frac{\log^2 T}{T^{-2\delta}})T^{-\frac{1}{2}}\right\} \\ &\leq cT^{-\lambda} \end{aligned}$$

for $T \geq T_0$, where $c > 0$ and T_0 do not depend on τ and $\mathbf{b}$.

To complete the proof of Lemma A.1 we need to extend the result in (A.27) to hold uniformly for $(\mathbf{b},\tau) \in S$. Here we apply a chaining argument: following the proof of Lemma A.2 in Koenker and Portnoy (1987), we choose intervals of length $1/T^5$ covering $[\tau_0, 1-\tau_0]$ and balls of radius $1/T^5$ covering $\{\mathbf{b} : \|\mathbf{b}\| \leq C\}$. Let $\{\tau_1,\tau_2\}$ lie in one of the intervals and $\{\mathbf{b}_1,\mathbf{b}_2\}$ lie in one of the balls covering $\{\mathbf{b} : \|\mathbf{b}\| \leq C\}$. We would like to bound $\|r_t(\mathbf{b}_1,\tau_1) - r_t(\mathbf{b}_2,\tau_2)\|$. For notational convenience we define ϵ_l, K_l, and δ_l for $l = 1,2$ as follows:

$$K_l = \kappa(\tau_l)/\sigma_{\tau_l} = \kappa^2(\tau_l)f(F^{-1}(\tau_l)) \tag{A.28}$$

$$T^{-1/2}\sigma_{\tau_l}\tilde{\mathbf{x}}_t'\mathbf{b}_l = \delta_l, \quad e_{t\tau_l} = e_t - F^{-1}(\tau(l)) = \epsilon(l) \tag{A.29}$$

$$I_l = I(\delta_l < \epsilon(l) < 0) + I(0 < \epsilon(l) < \delta_l) \tag{A.30}$$

We observe the following relationship between $\rho_\tau(\cdot)$ and $\psi_\tau(\cdot)$

$$\rho_\tau(\epsilon-\delta) - \rho_\tau(\epsilon) = -\psi_\tau(\epsilon)\delta + |\epsilon-\delta|\left[(I(\delta<\epsilon<0) + I(0<\epsilon<\delta)\right] \tag{A.31}$$

Using the definition of $r_t(\mathbf{b},\tau)$ in (A.6), and (A.28) – (A.30) after some algebra we can write

$$\begin{aligned} |\tilde{r}_t(\mathbf{b}_1,\tau_1) - \tilde{r}_t(\mathbf{b}_2,\tau_2)| \leq \big\{&|K_1\epsilon(1) - K_2\epsilon(2)|I_1 + |K_2\delta_2 - K_1\delta_1|I_1 \\ &+ K_2|\epsilon(2)-\delta_2||I_1 - I_2| + \frac{1}{2T}|[(\tilde{\mathbf{x}}_t'\mathbf{b}_2)^2 - (\tilde{\mathbf{x}}_t'\mathbf{b}_1)^2]|\big\}I(\|\tilde{\mathbf{x}}_t \leq B) \end{aligned} \tag{A.32}$$

Now using the fact that $|\tau_1 - \tau_2| \leq T^{-5}$ and $\|\mathbf{b}_1 - \mathbf{b}_2\| \leq T^{-5}$, it is straightforward to show that all the four terms in the above inequality are $O(T^{-2})$. (For details see page 33-34 of Hasan (1993)). Hence if S_ν denotes the covering set containing $(\tau_1,\mathbf{b}_1)$ and $(\tau_2,\mathbf{b}_2)$, then

$$\begin{aligned} \sup_{S_\nu}\left|\sum_{t=1}^{T}\tilde{r}_t(\tau_1,\mathbf{b}_1) - \sum_{t=1}^{T}\tilde{r}_t(\tau_2,\mathbf{b}_2)\right| &\leq T^{-1} + Lc_5T^{-1/2}\max_{1\leq t\leq T}\|\tilde{\mathbf{x}}_t\|I(\|\tilde{\mathbf{x}}_t\| \leq B) \\ &\leq o(1) + Lo(1) \end{aligned} \tag{A.33}$$

where L is the number of times one of the e_t's lies in the set $\{(A_1 \cup A_2)\backslash(A_1 \cap A_2)\}$, where $A_l = [F^{-1}(\alpha_l) - \delta_l, F^{-1}(\alpha_l) + \delta_l]$, for $l = 1,2$. Now, $L \sim b(T,p^*)$ where p^* is the probability that e_t lies in the above set. Thus, since f is bounded by Condition (i), and making use of

the fact that $|\tau_1 - \tau_2| \leq T^{-5}$ and $\|\mathbf{b}_1 - \mathbf{b}_2\| \leq T^{-5}$, it can be shown (see Appendix A.7 of Hasan (1993)) that $p^* \leq cT^{-2}$ Therefore,

(A.34)

$$P\left\{\sup_{S_\nu} |R_t(\mathbf{b}_1, \tau_1) - R_t(\mathbf{b}_2, \tau_2)| \geq T^{-1} + \lambda o(1)\right\} \leq \sum_{k=\lambda}^{T} \left(\begin{array}{c} T \\ k \end{array} \right) \left[\frac{c}{T^2}\right]^k \left[1 - \frac{c^*}{T^2}\right]^{T-k}$$
$$\leq c' T^{-\lambda}$$

Since the number of sets needed to cover the set $S \equiv [\tau_0, 1-\tau_0] \times \{\mathbf{b} : \|\mathbf{b}\| \leq C\}$ is bounded by $T^{5(p+1)}$ (where p is the number of parameters to be estimated under the null model) we obtain for $\lambda > 5(p+1)$

$$P\left\{\sup_{(\tau,\mathbf{b})} |R_t(\mathbf{b}, \tau)| \geq (\lambda+1)B_T + o(1) + \lambda o(1)\right\} \leq T^{5(p+1)} T^{-\lambda} \to 0. \tag{A.35}$$

Finally, we need to deal with the approximation error due to the truncation at B. It suffices to show that

$$\sup_{(\mathbf{b},\tau)\in S} |R_T(\mathbf{b}, \tau)| I[\|\tilde{\mathbf{x}}_t\| > B] \to^p 0 \tag{A.36}$$

To accomplish this we observe that for any $\delta > 0$,

$$P\left[\{|R_T(\mathbf{b}, \tau)| I[\|\tilde{\mathbf{x}}_t\| > B]\} > \delta\right] \leq P\left[\max_{1\leq t\leq T} \|\tilde{\mathbf{x}}_t\| > B\right] \to 0 \tag{A.37}$$

using (A.2), by first letting $T \to \infty$ and then $B \to \infty$. Since the right hand side of the above inequality does not depend on $(\tau, \mathbf{b}) \in S$,

$$\sup_{(\mathbf{b},\tau)\in S)} P\left[\{|R_T(\mathbf{b}, \tau)| I[\|\tilde{\mathbf{x}}_t\| > B]\} > \delta\right] \to 0 \tag{A.38}$$

which completes the proof of Lemma A.1.

Now recalling the notation in Section 3, let $\tilde{z}_t = (z_t - \hat{z}_t)$. The following lemma follows using the argument for Lemma 3.2 in GJKP.

Lemma A.2. *Under Condition (i) of Theorem 3.1, as $T \to \infty$, for $S = \{\|\mathbf{b}\| \leq C, \tau_0 \leq \tau \leq 1 - \tau_0\}$,*

$$\sup_{(\tau,\mathbf{b})\in S} \{|\kappa(\tau) T^{-1} \sum_{t=1}^{T} \tilde{z}_t [\psi_\tau(e_{t\tau} - T^{-1/2}\sigma_\tau \mathbf{x}_t'\mathbf{b}) - \psi_\tau(e_{t\tau})]|\} \to^p 0. \tag{A.39}$$

Remark. Note that the conditions of Lemma 3.2 in GJKP follow immediately in the present circumstances since $\sum \mathbf{x}_t' \tilde{z}_t = 0$ and $T^{-2} \sum \tilde{z}_t^2 = \sigma_z^2 = O_p(1)$.

We may now approximate the regression rank score process by an empirical process.

Lemma A.3. *Assume Condition (i) of Theorem 3.1, and let $\tilde{\mathrm{a}}_t(\tau) = I(e_t \geq F^{-1}(t))$, for $t = 1, \ldots, T$, then as $T \to \infty$,*

$$\sup_{\tau_0\leq\tau\leq 1-\tau_0} \left\{ |T^{-1}\kappa(\tau) \sum_{t=1}^{T} \tilde{z}_t (\hat{\mathrm{a}}_t(\tau) - \tilde{\mathrm{a}}_t(\tau))| \right\} \quad \to^p 0. \tag{A.40}$$

Proof: Let $\hat{\beta}_T(\tau)$ be the τ^{th}– regression quantile corresponding to the null model; i.e. $\hat{\beta}_T(\tau)$ is a solution of the minimization

$$\sum_{t=1}^{T} \rho_\tau(\Delta y_t - \mathbf{x}_t'\mathbf{b}) := \min, \qquad \mathbf{b} \in R^{p+1}. \tag{A.41}$$

Analogously, define $\beta(\tau) = \beta + (F^{-1}(\tau), 0, 0, \ldots, 0)'$; that is, the solution to (A.41) when the summation is replaced by expectation. Also, let $\hat{a}_T(\tau)$ be the vector of τth regression rank scores defined in (2.9). The formal duality between $\hat{\beta}(\tau)$ and $\hat{\mathrm{a}}_t(\tau)$ implies that for $t = 1, \ldots, T$

$$\hat{\mathrm{a}}_t(\tau) = \begin{cases} 1 & \text{if } \Delta y_t > x_t'\hat{\beta}(\tau) \\ 0 & \text{if } \Delta y_t < x_t'\hat{\beta}(\tau) \end{cases} \tag{A.42}$$

while the components of $\hat{a}_T(\tau)$ corresponding to $\{t|\Delta y_t = x_t'\hat{\beta}(\tau)\}$ are determined by the equality constraints of (2.9).

Now using the result on the rate of consistency of regression quantiles from Theorem 3.1 of GJKP, we substitute

$$\mathbf{b}_T = \left[T^{1/2}\sigma_\tau^{-1}\left(\hat{\beta}_T(\tau) - \beta(\tau)\right)\right] = O_p(1) \tag{A.43}$$

and the definition of ψ_τ in Lemma A.1 to get

$$\begin{aligned} &\sup_{\tau_0 \le \tau \le 1-\tau_0} \kappa(\tau) T^{-1} \sum_{t=1}^{T} |\tilde{z}_t[\psi_\tau(e_{t\tau} - T^{-1/2}\sigma_\tau \mathbf{x}_t'\mathbf{b}_T) - \psi_\tau(e_{t\tau})] \\ &\quad = \sup_{\tau_0 \le \tau \le 1-\tau_0} \kappa(\tau) T^{-1} \sum_{t=1}^{T} |\tilde{z}_t[I(e_{t\tau} \ge \mathbf{x}_t'(\hat{\beta}(\tau) - \beta(\tau))) - \tilde{\mathrm{a}}_t(\tau)]| \\ &\quad = \sup_{\tau_0 \le \tau \le 1-\tau_0} \kappa(\tau) T^{-1} \sum_{t=1}^{T} |\tilde{z}_t[I(\Delta y_t \ge \mathbf{x}_t'\hat{\beta}(\tau)) - \tilde{\mathrm{a}}_t(\tau)]| \to^p 0. \end{aligned} \tag{A.44}$$

Finally, the lemma follows from (A.44) using (A.42) and the fact that

$$|T^{-1}\kappa(\tau)\sum_{t=1}^{T} \tilde{z}_t I(\Delta y_t = \mathbf{x}_t'\hat{\beta}(\tau))\,\hat{\mathrm{a}}_t| \le (p+1)\kappa(\tau)T^{-1/2} \max_{1\le t\le T} |T^{-1/2}\tilde{z}_t| \quad \to^p 0 \tag{A.45}$$

since $|\hat{\mathrm{a}}_t| \le 1$, for all $t = 1, \ldots T$, $T^{-1/2}\max_{1\le t\le T}|T^{-1/2}\tilde{z}_t| = o_p(1)$, and at most (p+1) of the total observations can satisfy the equation $\Delta y_t = \mathbf{x}_t'\hat{\beta}(\tau)$.

Our next objective is to obtain an asymptotic representation for linear rank statistics of the form

$$L_T = T^{-1}\sum_{t=1}^{T} \tilde{z}_t \varphi(F(e_t)) + o_p(1) \tag{A.46}$$

The following lemma is essential.

Lemma A.4. *Under the Conditions of Theorem 3.1,*

$$T^{-1}\sum_{t=1}^{T}\tilde{z}_t\int_0^1[\hat{a}_t(s)-\tilde{a}_t(s)]\,d\varphi(s)\quad\to^p 0. \tag{A.47}$$

Proof: First observe that since φ is constant in the tails

$$\sum_{t=1}^{T}\tilde{z}_t\int_0^1[\hat{a}_t(s)-\tilde{a}_t(s)]\,d\varphi(s)=T^{-1}\sum_{t=1}^{T}\tilde{z}_t\int_{\tau_0}^{1-\tau_0}(\hat{a}_t(s)-\tilde{a}_t(s))\,d\varphi(s) \tag{A.48}$$

Applying the Lebesgue dominated convergence theorem and Lemma A.3 we can see that

(A.49)

$$T^{-1}\sum_{t=1}^{T}\tilde{z}_t\int_{\tau_0}^{1-\tau_0}(\hat{a}_t(s)-\tilde{a}_t(s))\,d\varphi(s)\le\int_{\tau_0}^{1-\tau_0}|\kappa(\tau)T^{-1}\sum_{t=1}^{T}\tilde{z}_t(\hat{a}_t(s)-\tilde{a}_t(s))|\,d\varphi(s)\to^p 0$$

since $\varphi(s)$ is a bounded square integrable function in the interval $(\tau_0,1-\tau_0)$. Hence the lemma follows.

Lemma A.5. *Under the conditions of Lemma A.4, L_T admits the linear representation*

$$L_T=T^{-1}\sum_{t=1}^{T}\tilde{z}_t\hat{b}_{Tt}=T^{-1}\sum_{t=1}^{T}\tilde{z}_t\varphi(F(e_t))+o_p(1). \tag{A.50}$$

Proof: Integrating by parts, and using Lemma A.4

$$T^{-1}\sum_{t=1}^{T}\tilde{z}_t\int_0^1\varphi(s)d\,(\hat{a}_t(s)-\tilde{a}_t(s))=-T^{-1}\sum_{t=1}^{T}\tilde{z}_t\int_0^1[\hat{a}_t-\tilde{a}_t(\tau)]\,d\varphi(s)\quad\to^p 0 \tag{A.51}$$

Thus using the definition of $\hat{b}_{Tt}$

$$\begin{aligned}L_T=T^{-1}\sum_{t=1}^{T}\tilde{z}_t\hat{b}_{Tt}&=-T^{-1}\sum_{t=1}^{T}\tilde{z}_t\int_0^1\varphi(s)\,d(\,\hat{a}_t(s))\\&=-T^{-1}\sum_{t=1}^{T}\tilde{z}_t\int_0^1\varphi(s)\,d(\,\tilde{a}_t(s))+o_p(1)\\&=T^{-1}\sum_{t=1}^{T}\tilde{z}_t\,\varphi(F(e_t))+o_p(1).\end{aligned} \tag{A.52}$$

We are now interested in the limiting distribution of $T^{-1}\sum_{t=1}^{T}z_t\varphi(F(e_t))$, which is the analogue of $T^{-1}\sum_{t=1}^{T}z_tu_t$ in the least-squares setup and $T^{-1}\sum_{t=1}^{T}z_tsign(u_t)$ in the l_1 test statistics. To do so we make use of a *bivariate invariance principle* as in Hercé (1990).

Lemma A.6. *Under the conditions of Theorem 3.1*

$$T^{-1}\sum_{t=1}^{T}z_tv_t\Rightarrow\Sigma_{12}\int_0^1W_1(s)\,dW_1(s)+\triangle^{1/2}\int_0^1W_1(s)\,dW_2(s). \tag{A.53}$$

Proof: We write the partial sums of v_t with $V_0 = 0$ as

$$V_t = \sum_{i=1}^{t} v_i \tag{A.54}$$

Recalling the definition of $\Sigma^{-1/2}$, we consider the standardized bivariate process

$$W_T(s) = \Sigma^{-1/2} T^{-1/2} \sum_{t=1}^{[Ts]} (u_t, v_t)' \tag{A.55}$$

which for $(j-1)/T \leq s < j/T$ reduces to $\Sigma^{-1/2} T^{-1/2} (z_j, V_{j-1})'$. From the multivariate invariance principle of Wooldridge and White (1988) or Phillips and Durlauf (1986), we obtain $W_T(s) \Rightarrow W(s)$, a standard Brownian motion on $C(0,1) \times C(0,1)$.

Furthermore, using the Lemma 3.1(e) of Phillips and Durlauf (1986), we have

$$T^{-1} \sum_{t=1}^{T} (z_t, V_{t-1})' (u_t, v_t) \Rightarrow \Sigma^{1/2} \int_0^1 W(s)\, dW(s)' (\Sigma^{1/2})' + \Sigma_1 \tag{A.56}$$

where $\Sigma^{1/2}$ and Σ_1 are as defined in (3.7) and (3.6) respectively. Clearly, the off diagonal element of the above matrix in the right hand side of (A.56) gives us the following weak convergence result

$$T^{-1} \sum_{t=1}^{T} z_t v_t \Rightarrow \Sigma_{12} \int_0^1 W_1(s)\, dW_1(s) + \triangle^{1/2} \int_0^1 W_1(s)\, dW_2(s).$$

where W_1 and W_2 are the two elements of the bivariate Brownian motion. Note that the off diagonal elements of the matrix Σ_1 are zero in our setting. Thus the proof of the above lemma follows immediately.

Proof of Theorem 3.1: Recall the test statistic

$$\tilde{S}_T = \frac{\sigma}{\hat{\triangle}^{1/2}} S_T - \frac{\Sigma_{12} \sigma_z}{\triangle^{1/2} \sigma} \left[T \hat{\gamma}_T - \frac{\sigma^2 - \sigma_u^2}{2\sigma_z^2} \right]. \tag{A.57}$$

We will show that

$$\tilde{S}_T \Rightarrow \frac{\int_0^1 (W_1 - \bar{W}_1)\, dW_2}{(\int_0^1 (W_1 - \bar{W}_1)^2)\, ds)^{1/2}} \tag{A.58}$$

where $\bar{W}_1 = \int W_1\, ds$. Conditional on $\sum y_{t-1}$, it is clear that $\int_0^1 (W_1 - \bar{W}_1)\, dW_2$ is normally distributed with mean zero and variance equal to $\sigma^2 \int_0^1 (W_1(s) - \bar{W})^2\, ds$. Furthermore, since the conditional distribution is independent of $\sum y_{t-1}$, we also get the unconditional normal distribution of the test statistic with mean zero and variance 1. It remains to show (A.58).

Consider the first term $S_T = L_T / \sigma_z$. By Lemma A.5

$$\frac{\sigma}{\triangle^{1/2}} S_T = \frac{\sigma}{\triangle^{1/2}} \frac{\sum \tilde{z}_t v_t}{\sigma_z} + o_p(1). \tag{A.59}$$

Note that

$$\sigma_z^2 = T^{-2}\sum (z_t - \hat{z}_t)^2 \Rightarrow \sigma^2 \int_0^1 (W_1(s) - \bar{W}_1)^2\, ds, \tag{A.60}$$

where $\bar{W}_1 = \int_0^1 W_1\, ds$. Therefore, by Lemma A.6, collecting terms,

$$\frac{\sigma}{\triangle^{1/2}} S_T \Rightarrow \frac{\Sigma_{12}}{\triangle^{1/2}} \frac{\int_0^1 (W_1 - \bar{W}_1) dW_1}{(\int_0^1 (W_1 - \bar{W}_1)^2 ds)^{1/2}} + \frac{\int_0^1 (W_1 - \bar{W}_1) dW_2}{(\int_0^1 (W_1 - \bar{W}_1)^2 ds)^{1/2}} \tag{A.61}$$

Observing that all the first difference terms, $\Delta y_{t-j} = u_{t-j}$ of the design matrix X of the null model are asymptotically orthogonal to z, the second component of $\tilde{S}_T$ is asymptotically equivalent to the $Z(\hat{\alpha})$ test statistic of Phillips and Perron (1988). Thus, following them it is easy to see that

$$\left[T\hat{\gamma}_T - \frac{\sigma^2 - \sigma_u^2}{2\sigma_z^2} \right] \Rightarrow \frac{\int_0^1 (W_1 - \bar{W}_1) dW_1}{(\int_0^1 (W_1 - \bar{W}_1)^2 ds)} \tag{A.62}$$

Therefore combining (A.61) and (A.62) we can conclude that (A.58) holds.

Test Statistics under Local Alternative: Here we wish to derive the distribution of the test statistic under the sequence of local alternative hypotheses.

Lemma A.7. *Under condition (i) of Theorem 3.1, and local alternatives* H_T

$$T^{-1}\sum_{t=1}^{T} (z_t, V_{t-1})'(u_t, v_t) \Rightarrow \Sigma^{1/2} \int_0^1 \mathbf{K}_\delta(r)\, dW(r)' (\Sigma^{1/2})' + \Sigma_1 \tag{A.63}$$

where $\Sigma^{1/2}$ *and* Σ_1 *are as defined in* (3.7) *and* (3.6) *respectively,*

$$\mathbf{K}_\delta(r) = \begin{bmatrix} J_\delta(r) \\ \int_0^1 dW_2(s) \end{bmatrix} \tag{A.64}$$

where $J_\delta(r) = \int_0^r e^{(r-s)\delta}\, dW_1(s), \delta = \gamma_0$, $W_1(r)$ *and* $W_2(r)$ *are two independent Brownian motions.*

Proof: See Lemma 3.1(e) of Phillips (1988).

Lemma A.8. *Under the same conditions as in Lemma A.7,*

$$T^{-1}\sum \tilde{z}_t v_t \Rightarrow \Sigma_{12}\left[\int_0^1 (J_\delta(s) - \bar{J}_\delta)\, dW_1(s) \right] + \triangle^{1/2} \int_0^1 (J_\delta(s) - \bar{J}_\delta)\, dW_2(s) \tag{A.65}$$

where $\bar{J}_\delta = \int_0^1 J_\delta(r)\, dr$.

Proof: The off diagonal element of (A.63) gives

$$T^{-1}\sum_{t=1}^{T} z_t v_t \Rightarrow \Sigma_{12} \int_0^1 J_\delta(r)\, dW_1(r) + \triangle^{1/2} \int_0^1 J_\delta(r)\, dW_2(r) \tag{A.66}$$

since the off diagonal element of Σ_1 is zero in our setting. And the result follows by the remark following (A.61).

Following Shorack and Wellner (1986), it can be shown that our local alternative model is contiguous to the null model. For details see Theorem 2.7.1 of Hasan (1993). Thus, by Lemma A.5, and contiguity of the local alternative model to the null model, we can write

$$L_T(\delta) = T^{-1}\sum_{t=1}^{T} \tilde{z}_t \varphi(F(e_t + \delta T^{-1} z_t)) + o_p(1) \equiv \tilde{L}_T(\delta) + o_p(1) \tag{A.67}$$

We will show that

$$M_T(\delta) = |\tilde{L}_T(\delta) - \tilde{L}_T(0) - E(\tilde{L}_T(\delta) - \tilde{L}_T(0))| = o_p(1). \tag{A.68}$$

Since the e_t's are iid,

$$\begin{aligned} \mathbf{Var}(M_T(\delta) &\le E(\tilde{L}_T(\delta) - \tilde{L}_T(0))^2 \\ &= \mathbf{E}[T^{-1}\sum \tilde{z}_t(\varphi(F(e_t + \delta T^{-1} z_t)) - \varphi(F(e_t)))]^2 \\ &= \mathbf{E}T^{-2}\sum \tilde{z}_t^2(\varphi(F(e_t + \delta T^{-1} z_t)) - \varphi(F(e_t)))^2 \end{aligned} \tag{A.69}$$

so conditioning on $\mathcal{F}_{t-1}$, we have,

$$\mathbf{Var}(M_T(\delta) \le \mathbf{E}[T^{-2}\sum \tilde{z}_t^2] \max_t \{\int_{-\infty}^{\infty} [\varphi(F(e_t + \delta T^{-1} z_t)) - \varphi(F(e_t))]^2 f(s) ds\} \tag{A.70}$$

But $T^{-2}\sum \tilde{z}_t^2 = O_p(1)$ and the second factor is $o_p(1)$ by the argument for A.2 in Bickel(1975). Thus, unconditioning on $\mathcal{F}_{t-1}$, (A.68) follows. Now, noting that $\mathbf{E}\tilde{L}_T(0) = 0$, and again conditioning on $\mathcal{F}_{t-1}$, we have,

$$\mathbf{E}\tilde{L}_T(\delta) = \mathbf{E}_{\mathcal{F}_{t-1}} T^{-1} \sum \tilde{z}_t \mathbf{E}\varphi(F(e_t + \delta T^{-1} z_t))), \tag{A.71}$$

so, expanding f,

$$\begin{aligned} \mathbf{E}\tilde{L}_T(\delta) &= \mathbf{E}_{\mathcal{F}_{t-1}} T^{-1} \sum \tilde{z}_t \int_{-\infty}^{\infty} \varphi(F(u)) f(u - \delta T^{-1} z_t) du \\ &= -\mathbf{E}_{\mathcal{F}_{t-1}} T^{-2} \sum \tilde{z}_t^2 \delta \int_{-\infty}^{\infty} \varphi(F(u)) f'(u) du + o_p(1) \\ &\Rightarrow \delta \omega \sigma^2 \xi_J^2(\delta) \end{aligned} \tag{A.72}$$

where $\xi_J^2(\delta) = \int_0^1 (J_\delta(r) - \bar{J}_\delta)^2 \, dr$, and $\omega = -\int_0^1 \varphi(u) \, df(F^{-1}(u))$. By Lemma A.8,

$$L_T \Rightarrow \Sigma_{12} \int_0^1 (J_\delta(r) - \bar{J}_\delta) \, dW_1(r) + \triangle^{1/2} \int_0^1 (J_\delta(r) - \bar{J}_\delta) \, dW_2(r) + \delta \omega \sigma^2 \xi_J^2(\delta). \tag{A.73}$$

Therefore,

$$\frac{\sigma}{\triangle^{1/2}} S_T \Rightarrow \frac{\Sigma_{12}}{\triangle^{1/2}} \frac{\int_0^1 (J_\delta(r) - \bar{J}_\delta) \, dW_1(r)}{\xi_J(\delta)} + \frac{\int_0^1 (J_\delta(r) - \bar{J}_\delta) \, dW_2(r)}{\xi_J(\delta)} + \triangle^{-1/2} \delta \omega \sigma^2 \xi_J(\delta). \tag{A.74}$$

From Theorem 3 of Phillips and Perron (1988), we can obtain the limiting behavior of the second component of $\tilde{S}_T$ which is asymptotically equivalent to their $Z(\hat{\alpha})$, i.e.

$$\text{(A.75)} \qquad \frac{\Sigma_{12}\sigma_z}{\triangle^{1/2}\sigma}\left[T\hat{\gamma}_T - \frac{\sigma^2-\sigma_u^2}{2\sigma_z^2}\right] \Rightarrow \triangle^{-1/2}\delta\Sigma_{12}\xi_J(\delta) + \frac{\Sigma_{12}}{\triangle^{1/2}}\frac{\int_0^1 (J_\delta(r)-\bar{J}_\delta)\,dW_1(r)}{\xi_J(\delta)}$$

Therefore, combining (A.74) and (A.75) we can conclude that

$$\text{(A.76)} \qquad \tilde{S}_T - \eta_T \Rightarrow \frac{\int_0^1 (J_\delta(r)-\bar{J}_\delta)\,dW_2(r)}{\xi_J(\delta)}$$

where

$$\text{(A.77)} \qquad \eta_T = \sigma^{-1}\triangle^{-1/2}\delta\sigma_z[\sigma^2\omega - \Sigma_{12}] \Rightarrow \triangle^{-1/2}\delta\xi_J(\delta)[\sigma^2\omega-\Sigma_{12}].$$

The same conditioning argument as for (A.58) yields the limiting normality of $\tilde{S}_T - \eta_T$.

Proof of Corollary 3.1: Replacing the nuisance parameters appearing in $\tilde{S}_T$ by the consistent estimates as discussed in Section 3.1, and applying continuous mapping theorem, we get the same limiting distribution of $\tilde{S}_T$.

The following invariance lemma is needed in the proof of Theorem 4.1 but may be of independent interest.

Lemma A.9. *Let* $S_n(\mathbf{X},\mathbf{Z},y) = (\mathbf{Z}-\hat{\mathbf{Z}})'\hat{\mathbf{b}}_n$ *as in (2.10), and* $\mathbf{P}$ *be an arbitrary nonsingular* $p\times p$ *matrix,* $\mathbf{Q}$ *an arbitrary* $p\times q$ *matrix and* γ *an arbitrary* p*-vector. Then for any* φ *such that* $\int_0^1 \varphi(t)dt = 0$, $S_n(\mathbf{XP},\mathbf{Z}+\mathbf{XQ},y+\mathbf{X}\gamma) = S_n(\mathbf{X},\mathbf{Z},y)$.

Proof of Lemma A.9: Let $\hat{\beta}(\tau,\mathbf{X},y)$ denote the τ^{th} regression quantile, then Koenker and Bassett (1978, Theorem 3.2) implies $\hat{\beta}(\tau,\mathbf{XP},y+\mathbf{X}\gamma) = \mathbf{P}^{-1}(\hat{\beta}(\tau,\mathbf{X},y)+\gamma)$, and consequently $\hat{u}(\tau,\mathbf{X},y) \equiv y - \mathbf{X}\hat{\beta}(\tau,\mathbf{X},y)$ satisfies $\hat{\mathbf{u}}(\tau,\mathbf{XP},y+\mathbf{X}\gamma) = \hat{\mathbf{u}}(\tau,\mathbf{X},y)$. Thus $\hat{\mathbf{a}}(\tau,\mathbf{XP},y+\mathbf{X}\gamma) = \hat{\mathbf{a}}(\tau,\mathbf{X},y)$ which in turn implies the invariance of $\hat{\mathbf{b}}$. It remains to show that $\mathbf{X}'\hat{\mathbf{b}} = 0$ for any φ satisfying $\int_0^1 \varphi(t)\,dt = 0$. Integrating by parts we have

$$\text{(A.78)} \qquad \hat{b}_i = -\int_0^1 \varphi(s)\,d\hat{\mathrm{a}}_i(s)$$

and since $\hat{\mathrm{a}}_i(s)$ is continuous and piecewise linear we have using the constraint in (1.15)

$$\text{(A.79)} \qquad \frac{d}{ds}\mathbf{X}'\hat{\mathbf{a}}_n(s) = -\mathbf{X}_n'\mathbf{1}_n$$

and therefore,

$$\text{(A.80)} \qquad \mathbf{X}'\hat{\mathbf{b}} = \mathbf{X}'\mathbf{1}_n\int_0^1 \varphi(s)\,ds = 0.$$

Proof of Theorem 4.1: The proof for the trend model follows exactly the same line of argument as for Theorem 2.1 with the following definition of augmented $\tilde{x}_t$,

$$\text{(A.81)} \qquad \tilde{\mathbf{x}}_t = (\mathbf{x}_t, z_t) = (1, t/T, \Delta y_{t-1},\ldots,\Delta y_{t-p}, y_{t-1}/\sqrt{T})$$

satisfying (A.2) and (A.3). Since all the components except the intercept and the trend component of the null design matrix X, are asymptotically orthogonal to $z_t = y_{t-1}$, the test statistic has the following limiting behavior,

$$\tilde{S}_T \Rightarrow \frac{\int_0^1 (W_1(s) - \hat{W})\, dW_2}{[\int_0^1 (W_1(s) - \hat{W})^2\, ds]^{1/2}} \tag{A.82}$$

where

$$\hat{W} = \int_0^1 (4 - 6r) W_1(r) dr - r \int_0^1 (6 - 12r) W_1(r) dr \tag{A.83}$$

Again, conditioning yields the asymptotic normality for $\tilde{S}_T$ under H_0.

REFERENCES

Adichie, J.N. (1991) Rank tests in linear models, in P.R. Krishnaiah and P.K. Sen (eds.), *Handbook of Statistics*, **4**, New York: Elsevier.

Bickel, P.J. (1975) One-Step Huber estimators in the Linear Model, *Journal of the American Statistical Association*, **70**, 428-434.

Campbell, B. and Dufour, J.-M. (1993) Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter, Concordia University DP-9313.

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

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

Cox D.D., and Llatas, I. (1991) Maximum likelihood type estimation for nearly nonstationary autoregressive time series *Annals of Statistics*, **19** , 1109-1128.

Chan, N.H. and Tran T. (1989) On the first order autoregressive process with infinite variance, *Econometric Theory* **5**, 354-362.

Dickey, D.A. and Fuller, W.A. (1979) Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, 427-431.

Dufour, J.-M. (1981) Rank tests for serial dependence, *Journal of Time Series Analysis*, **2**, 117-128.

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

Gutenbrunner, C. and Jurečková , J. (1992) Regression rank-scores and regression quantiles, *Annals of Statistics*, **20**, 305-330.

Hájek , J. (1969) *Nonparametric Statistics*. San Francisco: Holden-Day.

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

Hallin, M. and Puri, M.L. (1992) Rank Tests for time-series analysis: a survey, in D. 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. *Journal of Multivariate Analysis*, **51**, 318-337.

HASAN, M. N. (1993) Robust testing for unit roots based on regression rank scores, Ph.D. Dissertation. University of Illinois at Urbana-Champaign.

HERCÉ , M.A. (1990) The asymptotic theory of L_1-norm estimation of a unit root, Ph.D Dissertation. University of California, San Diego.

HETTMANSPERGER, T.P. (1984) *Statistical Inference Based on Ranks*. New York: Wiley.

KNIGHT, K. (1989) Limit theory for autoregressive parameters in an infinite variance random walk. *Canadian Journal of Statistics*, **17**, 261-278.

KOENKER, R. (1996) Rank tests for Linear Models, in Rao, C.R. and Maddala, G.S. (eds.), *Handbook of Statistics*, **15**, New York: Elsevier.

KOENKER, R.W. AND BASSETT, G. (1978) Regression quantiles. *Econometrica*, **46** 33-50.

KOENKER, R.W. AND D'OREY, V. (1987) Computing regression quantiles, *Applied Statistics*, **36**, 383-393.

KOENKER, R.W. AND D'OREY, V. (1993) Remark on algorithm 229, Computing dual regression quantiles and regression rankscores, *Applied Statistics*, **42**, 432-428.

KOENKER, R.W. AND PORTNOY, S. (1987) L-estimation for the linear model. *Journal of the American Statistical Association*, **82**, 851-857.

KOUL, H. AND SALEH, A.K.M.E. (1995) Autoregression quantiles and related rank scores processes, *Annals of Statistics*, **23**, 670-689.

LEHMANN, E.L. (1975) *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.

NEWEY, K. AND K.D. WEST, (1987) A simple positive definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703-708.

PHILLIPS, P.C.B. (1987) Time series regression with a unit root. *Econometrica*, **55**, 277-301.

PHILLIPS, P.C.B. (1988) Regression theory for near-integrated time series. *Econometrica*, **56**, 1021-1043.

PHILLIPS, P.C.B. AND DURLAUF, S.N. (1986) Multiple time series regression with integrated processes. *Review of Economic Studies*, **25**, 303-325.

PHILLIPS, P.C.B. AND PERRON, P. (1988) Testing for a unit root in time series regression. *Biometrika*, **75**, 335-346.

PURI, M.L. AND SEN, P.K. (1985) *Nonparametric Methods in General Linear Models*. New York: Wiley.

ROBINSON, P.M. (1993) Highly insignificant F-ratios, *Econometrica*, **61**, 687-696.

ROBINSON, P.M. (1994) Efficient Tests of nonstationary hypotheses, *Journal of the American Statistical Association,* **89**, 1420-37.

SHORACK, G. AND WELLNER, J. (1986) *Empirical Processes with Applications to Statistics*. New York: Wiley.

WOOLDRIDGE, J. AND WHITE, H. (1988) Some invariance principles and central limit theorems for dependent heterogenous processes. *Econometric Theory*, **4**, 210-230.