Econometric Theory, **20**, 2004, 360–381. Printed in the United States of America.
DOI: 10.1017/S0266466604202055

ROBUST TESTS OF THE UNIT ROOT HYPOTHESIS SHOULD NOT BE "MODIFIED"

SAMUEL B. THOMPSON
Harvard University

The rank-based unit root tests proposed by Hasan and Koenker (1997, *Econometrica* 65, 133–161) have power equal to size for normal innovations. Unit root tests based on *M*-estimators exhibit the same behavior. The problem occurs because the test statistics are transformed to obtain computationally convenient critical values. I describe a convenient way to compute critical values without transforming the test statistics. The resulting tests are almost as powerful as least squares–based tests for normal errors and much more powerful for thicker tailed distributions.

1. INTRODUCTION

Following Dickey and Fuller (1979), econometricians often employ least squares regression to test for a unit root in the autoregressive representation of a time series. Although least squares–based tests should perform well when the innovations are normal, there may be a loss of power relative to other methods when they are not normal. In recent years researchers have proposed a variety of "robust" tests that are designed to have good power for many different error distributions. For example, Cox and Llatas (1991), Lucas (1995a), Lucas (1995b), Hoek, Lucas, and van Dijk (1995), Herce (1996), and Rothenberg and Stock (1997) have proposed robust tests based on *M*-estimators; Hasan and Koenker (1997) have proposed a collection of rank tests.

Because the robust test statistics generally have null distributions that depend on unknown nuisance parameters, finding appropriate critical values is a nontrivial problem. Lucas (1995b) describes a simulation-based method for computing critical values. Herce (1996), Lucas (1995b), and Hasan and Koenker (1997) suggest transforming the test statistics so that the asymptotic null is nuisance-parameter free.

The present paper derives local-to-null asymptotic power functions for the robust tests and uses the asymptotic representations to evaluate the various pro-

I thank Thomas Rothenberg for many useful comments. This paper is based on my dissertation, which he supervised. I also thank Jack Porter, Jim Powell, Richard Stanton, and Jim Stock for good advice. Comments of Bruce Hansen, the editor, and two anonymous referees improved the exposition. Address correspondence to: Samuel B. Thompson, Department of Economics, Harvard University, Cambridge, MA 02138, USA; e-mail: sbthomps@fas.harvard.edu.

 © 2004 Cambridge University Press 0266-4666/04 $12.00

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

cedures for computing critical values. The procedure in Lucas (1995b) leads to tests with incorrect size, even asymptotically. The transformations used in Herce (1996), Lucas (1995b), and Hasan and Koenker (1997) lead to tests with asymptotic power equal to size for normal errors. This result is not entirely new; Lucas (1995b) reports simulation evidence in favor of the result, and Herce (1996) derives a similar result for the test based on the least absolute deviations estimator.

I fit a polynomial approximation to the critical values of the untransformed test statistics. The resulting tests are almost as powerful as least squares for normal errors and much more powerful for thicker tailed distributions. For alternatives far from the null, the tests dominate the transformed statistics used by Herce (1996) and Hasan and Koenker (1997). A Monte Carlo study suggests that the untransformed versions of the Wilcoxon ranks and normal ranks tests of Hasan and Koenker (1997) are superior to the least squares tests for a variety of error distributions.

2. LOCAL-TO-UNITY ASYMPTOTICS FOR ROBUST TESTS

The observations $\{y_t\}_{t=1}^T$ come from the data generating process

$$y_t = \mu_1 + \mu_2 t + u_t,$$
$$\Delta u_t = \gamma u_{t-1} + v_t,$$
$$\Gamma(L) v_t = \varepsilon_t,$$

where $\{\varepsilon_t\}_{t=1}^T$ is an independent and identically distributed (i.i.d.) sequence with mean 0 and variance σ_ε^2. The term $\Gamma(L)$ is the lag polynomial $1 - \Gamma_1 L - \cdots - \Gamma_p L^p$, and we assume that the roots of $\Gamma(z) = 0$ all lie outside the unit circle. For convenience we assume that $u_0 = 0$. We wish to test the unit root hypothesis $\gamma = 0$ versus the alternatives $\gamma < 0$.

The tests are based on the augmented Dickey and Fuller (1979) representation of the model

$$\Delta Y = X\alpha + Z\beta + \epsilon, \tag{1}$$

where $\Delta Y = (y_{p+2} - y_{p+1}, \ldots, y_T - y_{T-1})'$, $X = (y_{p+1}, \ldots, y_{T-1})'$, and Z is the design matrix with row $t - p - 1$ equal to $(1, t, \Delta y_{t-1}, \ldots, \Delta y_{t-p})$. The parameter α is $\gamma\Gamma(1)$, β is a $(p + 2) \times 1$ parameter vector, and $\epsilon = (\varepsilon_{p+2}, \ldots, \varepsilon_T)'$. For the intercept only model (e.g., $\mu_2 = 0$) the design matrix Z has rows $(1, \Delta y_{t-1}, \ldots, \Delta y_{t-p})$.

M-tests are based on *M*-estimators, which may be characterized as solutions to minimization problems. It will be convenient to adopt the notation that, for any function $g(\cdot)$ mapping $\Re$ into $\Re$ and any n-dimensional column vector x with components x_i, $g(x)$ is the n-dimensional column vector with components $g(x_i)$. The *M*-estimators $(\hat{\alpha}, \hat{\beta})$ minimize the objective function

$$1' g(\Delta Y - X\alpha - Z\beta),$$

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

where 1 is a vector of ones and g is a function chosen by the researcher. When $g(x) = x^2$, estimation is by least squares. Table 1 lists some alternative choices for g that "robustify" the estimators against thick-tailed error distributions. The unit root test rejects the null for small values of $\hat{\alpha}$.

Hasan and Koenker (1997) propose an alternative class of unit root tests based on ranks. The null is rejected for small values of $[X'MX]^{-1/2}X'Mb$, where M is the projection matrix $I - Z(Z'Z)^{-1}Z'$ and b is the vector of regression rank scores $b = \int_0^1 \hat{a}(s)\, d\varphi(s)$, where $\hat{a}$ solves the linear programming problem

$$\hat{a}(\tau) = \text{argmax}\{a'\Delta Y | Z'a = (1-\tau)Z'1,\ a \in [0,1]^{T-p}\} \tag{2}$$

and φ is a function chosen by the researcher. This construction for b is a way to handle the nuisance parameters β and allows the researcher to consider general functions of the ranked residuals: when Z is a vector of ones and $\varphi(\tau) = \tau$, each element of b is the rank of the corresponding element of ΔY. Table 1 lists some common choices for φ.

In large samples the critical values and power functions of the various tests depend on several nuisance parameters. For the M-tests the nuisance parameters are determined by a scorelike function $\psi(x)$, equal to $g'(x)$ if g is differentiable at x and zero otherwise (we allow g to have finitely many points of nondifferentiability to handle quantile estimators such as least absolute deviations (LAD)). The nuisance parameters are

$$\rho = \text{Corr}[\varepsilon_1, \psi(\varepsilon_1 - \eta)] \quad \text{and} \quad \lambda = \sigma_\varepsilon \omega / \sigma_\psi,$$

where $\omega = -\int_{\mathbb{R}} \psi(x - \eta)\, df(x)$, $\sigma_\psi^2 = \text{Var}[\psi(\varepsilon_1 - \eta)]$, f is the marginal density of ε_1, and η solves $E\psi(\varepsilon_1 - \eta) = 0$. When ψ is everywhere differentiable ω takes the more familiar form $E\psi'(x - \eta)$. The rank tests depend on the same nuisance parameters, with $\psi(x - \eta)$ replaced by $\varphi(F(x))$, where $F(x) = \Pr[\varepsilon_1 \le x]$. Table 1 lists the g, ψ, and φ functions for some common robust tests, along with the associated formulas for ω.

TABLE 1. Some robust unit root tests

Test	g, ψ, or φ function	ω
Least squares	$g(x) = x^2/2$, $\psi(x) = x$	1
LAD	$g(x) = \lvert x\rvert$, $\psi(x) = \text{sign}(x)$	$2f(\eta)$
Huber's M	$\psi(x) = x1(\lvert x\rvert \le k) + k\,\text{sign}(x)1(\lvert x\rvert > k)$	$\Pr[\lvert\varepsilon_1 - \eta\rvert \le k]$
Student's t	$\psi(x) = (x/\nu)/(1 + x^2/\nu)$	$E\psi'(\varepsilon_1 - \eta)$
Wilcoxon ranks	$\varphi(s) = s - \frac{1}{2}$	$Ef(\varepsilon_1)$
Normal ranks	$\varphi(s) = \Phi^{-1}(s)$	$Ef(\varepsilon_1)/\phi(\Phi^{-1}(F(\varepsilon_1)))$
Sign ranks	$\varphi(s) = \text{sign}(s - \frac{1}{2})$	$2f(\eta)$

Notes: $F(x) = \Pr[\varepsilon_1 \le x]$ and $f(x) = F'(x)$. The parameter k that appears in Huber's M function is chosen by the researcher. The parameter ν that appears in Student's t function is the number of degrees of freedom.

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

The tests come in two versions, an *e*-test and a *t*-test. The *e* and *t* *M*-tests reject the null for small values of the statistics

$$M_e = [\hat{\Gamma}(1)]^{-1}\hat{\lambda}T\hat{\alpha} \quad \text{and} \quad M_t = \hat{\sigma}_\epsilon^{-1}\hat{\lambda}[X'MX]^{1/2}\hat{\alpha},$$

where the variables with the ˆ are consistent estimates of the corresponding nuisance parameters. The *e* and *t* rank tests reject for small

$$R_e = [\hat{\sigma}_\psi\hat{\Gamma}(1)X'MX]^{-1}\hat{\sigma}_\epsilon TX'Mb \quad \text{and} \quad R_t = \hat{\sigma}_\psi^{-1}[X'MX]^{-1/2}X'Mb.$$

Hasan and Koenker (1997) propose only the *t* version of the test, but for completeness we examine the *e* version also. Section 4 provides estimates of the nuisance parameters.

Hasan and Koenker (1997) stress that the nuisance parameter λ does not need to be estimated to normalize a rank test. This may be an advantage of the rank tests over the *M*-tests. In the Monte Carlos in Section 5, the rank tests have more accurate size than the *M*-tests. Thompson (2004b) argues that this occurs because of estimation error associated with $\hat{\omega}$. Although in large samples the estimation error will have an asymptotically negligible effect on inference, in small samples it could affect test size. In some cases ω is especially poorly estimated—for example, for LAD $\omega = 2f(\eta)$, and the estimate of a density at a point converges at less than a root-T rate.

Because I am interested in inference when γ is close to one, I adopt the local-to-zero reparameterization $\gamma = c/T$ so that the parameter space is a shrinking neighborhood of zero as the sample size grows. Following Chan and Wei (1987) and Phillips (1987), I take *c* fixed when making limiting arguments, obtaining asymptotic representations as a function of the local alternative *c*.

Hasan and Koenker (1997) derive local-to-zero asymptotic representations for the rank tests. In Theorem 1 I provide local-to-zero representations for the *M*-tests. The theorem applies to objective functions *g* that are not twice continuously differentiable, such as LAD, Huber's *M* function, and the quantile regression estimator of Koenker and Bassett (1978). We assume that *g* is convex, which rules out Student's *t* objective function. Although the convexity assumption simplifies the proof, alternative methods can be used to show that the theorem holds for Student's *t* function.

The limiting forms of the test statistics are expressed as functionals of Brownian motion. Define $W(\cdot)$ as standard Brownian motion and define $W_c(\cdot)$ as the Ornstein–Uhlenbeck process $W_c(t) = \int_0^t \exp\{c(t-s)\}\, dW(s)$. Define

$$A_c \equiv \int_0^1 D_c(r)\, dW(r),$$

$$B_c \equiv \int_0^1 D_c(r)^2\, dr, \tag{3}$$

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

where $D_c(r) = W_c(r) - \int_0^1 W_c(s)\,ds$ in the intercept only model and $D_c(r) = W_c(r) - 2\int_0^1 (2 - 3s - r(3 - 6s))W_c(s)\,ds$ in the model with a linear time trend. Under the local-to-null asymptotics, the *e*-tests M_e and R_e both converge to the limiting distribution represented by the random variable $\mathcal{E}_c$, and the *t*-tests M_t and R_t converge in distribution to the random variable $\mathcal{T}_c$. The limiting forms are

$$\mathcal{E}_c \equiv \rho \frac{A_c}{B_c} + \sqrt{1-\rho^2}\,\frac{S}{\sqrt{B_c}} + \lambda c,$$

$$\mathcal{T}_c \equiv \rho \frac{A_c}{\sqrt{B_c}} + \sqrt{1-\rho^2}S + \lambda c\sqrt{B_c}, \tag{4}$$

where S is an independent standard normal variable. These results are summarized in the following theorem.

THEOREM 1. *Under Assumptions 1 and 2 in the Appendix, M_e and R_e converge weakly to $\mathcal{E}_c$, and M_t and R_t converge weakly to $\mathcal{T}_c$.*

There are a number of well-known asymptotic results concerning robust inference for a location model (see Rothenberg and Stock, 1997, Appendix). Asymptotic power depends on a single parameter that shifts the limiting distribution of the test statistic. Power is increasing in the parameter, which is maximized under correct specification. The movement of the parameter is asymmetric under incorrect specification, in the sense that with normal data power declines only slightly when using robust methods instead of least squares, but with thick-tailed distributions robust tests lead to significant increases in power.

For the unit root testing problem, large sample power depends on the pair of nuisance parameters ρ and λ. Because we reject for small values of the test statistics, in large samples the power of an *e*-test is given by $\Pr[\mathcal{E}_c \le q_e(\rho, a)]$, where $q_e(\rho, a)$ satisfies $\Pr[\mathcal{E}_0 \le q_e(\rho, a)] = a$ and a is the size of the test. Limiting power for the *t*-test is $\Pr[\mathcal{T}_c \le q_t(\rho, a)]$, where $\Pr[\mathcal{T}_0 \le q_t(\rho, a)] = a$. Power obviously increases with λ, which shifts the distribution to the left when c is negative. The term ρ also affects power by changing the shape of the distribution. The results in Section 3.1 suggest that λ dominates the power function relative to ρ.

In general ρ and λ depend on the test and error density. An exception to this is the least squares–based Dickey–Fuller test, where $\psi(x) = x$ and $\rho = \lambda = 1$ no matter what the error density. For other tests ρ will differ from 1 and the limiting null representations become a linear combination of a "Dickey–Fuller" term ($B_c^{-1}A_c$ for the *e*-test and $B_c^{-1/2}A_c$ for the *t*-test) and another random variable.

In Table 2 we have calculated ρ and λ by numerical integration for various tests and error densities.[1] We include the tests based on LAD, Huber's function, and the Student's *t* density with three degrees of freedom (denoted *t*3). In

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

TABLE 2. Values of ρ and λ

	$A(\tau)\exp(-\lvert x\rvert^{\tau})$				Mixture		
	$\tau = 2.0$ (3.00)	$\tau = 1.8$ (3.23)	$\tau = 1.2$ (4.74)	$\tau = 1.0$ (6.00)	$m = 0.1$ (8.33)	$m = 0.2$ (7.54)	$t3$ (∞)
Values of ρ							
LAD *M*-test	0.80	0.79	0.74	0.71	0.71	0.69	0.64
Huber's *M*-test	0.97	0.97	0.95	0.94	0.92	0.92	0.86
Student's *t*3 *M*-test	0.98	0.97	0.90	0.84	0.82	0.77	0.71
Wilcoxon ranks	0.98	0.97	0.94	0.92	0.91	0.89	0.83
Normal ranks	1.00	1.00	0.99	0.98	0.97	0.97	0.90
Sign ranks	0.80	0.79	0.74	0.71	0.71	0.69	0.64
Values of λ							
LAD *M*-test	0.80	0.84	1.15	1.41	1.00	1.12	1.27
Huber's *M*-test	0.97	0.99	1.08	1.13	1.17	1.22	1.31
Student's *t*3 *M*-test	0.98	1.06	1.12	1.25	1.17	1.28	1.41
Wilcoxon ranks	0.98	1.00	1.12	1.22	1.17	1.26	1.38
Normal ranks	1.00	1.00	1.06	1.13	1.12	1.17	1.28
Sign ranks	0.80	0.84	1.15	1.41	1.00	1.12	1.27

Notes: The numbers in parentheses are the kurtosis of each distribution. All values were calculated either analytically or by numerical integration.

addition we include the three rank tests considered in Hasan and Koenker (1997): the tests based on Wilcoxon, normal, and sign ranks.

To get a sense of how the parameters change as we move from normal errors to thicker tailed densities, we give results for error densities of the form $f(\epsilon) = A(\tau)\exp(-|\epsilon|^{\tau})$ for $\tau = 2.0$, 1.8, 1.2, and 1.0 and for the thick-tailed Student's t distribution with three degrees of freedom (denoted $t3$).[2] When τ is 2 the errors are normal with variance $\frac{1}{2}$, and when τ is 1 the errors are double-exponential with variance 2. We also report values for two mixture distributions (labeled mixture in the table) where a standard normal variable is drawn with probability $1 - m$ and a $N(0,9)$ variable is drawn with probability m. The mixture density has fatter tails than the normal. Results are given for $m = 0.1$ and $m = 0.2$.

The values of ρ fall as the errors become thicker tailed. For normal errors the parameter is close to 1, except for the LAD and sign rank tests. Under normality the null distributions of the other tests are similar to the null for the least squares–based tests. When the errors have thick tails the null distributions are different from the least squares distributions and from each other.

The term λ increases as the errors become thicker tailed, suggesting that thicker tailed errors lead to increased power. Under normal errors the parameters for the LAD and sign-median rank tests are far below 1, so we expect these tests to have poor power relative to the least squares–based tests when

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

errors are normal. For the rest of the tests λ is close to 1 with normal errors, implying that these tests have power almost as good as least squares when errors are normal and better than least squares when errors are thick tailed. Table 4 confirms this analysis; there the robust tests essentially dominate least squares, even for distributions close to Gaussian.

For the test and errors densities considered in Table 2, λ is maximized by choosing $-\psi(x-\eta)$ or $-\varphi(F(x))$ equal to $\partial \log f(x)/dx$. For normal errors, λ is largest with the least squares and normal ranks tests. The same result holds for the LAD and sign ranks tests used with double exponential errors and for the $t3$ test with $t3$ errors. This property holds for any test and error distribution, which is not surprising because it corresponds to testing under correct specifying of the error distribution. To see that it holds for the rank tests, let $f(x)$ equal $A\exp(-\int_{-\infty}^{x}\varphi(F(u))\,du)$, where A makes the density integrate to one. An argument based on integration by parts can be used to show that $\omega=\sigma_{\psi}^{2}$. For any $\tilde{\varphi}$ with associated $\tilde{\sigma}_{\psi}$ and $\tilde{\lambda}$, we have

$$\begin{aligned}\tilde{\lambda} &= -\frac{\sigma_{\varepsilon}}{\tilde{\sigma}_{\psi}}\int \tilde{\varphi}(F(x))\,df(x)\\ &= \frac{\sigma_{\varepsilon}}{\tilde{\sigma}_{\psi}}\int \tilde{\varphi}(F(x))\varphi(F(x))f(x)\,dx\\ &= \lambda\,\mathrm{Corr}[\tilde{\varphi}(F(\varepsilon_1)),\varphi(F(\varepsilon_1))].\end{aligned}$$

Therefore $\tilde{\lambda}\leq\lambda$. The result can be extended to the M-tests by replacing $\varphi(F(x))$ with $\psi(x-\eta)$.

It can be shown, using integration by parts, that $\rho=\lambda^{-1}$ when $-\psi(x-\eta)$ or $-\varphi(F(x))$ equals $\partial \log f(x)/dx$. This may be confirmed in the tables for the least squares and normal ranks tests applied to Gaussian errors. It may also be confirmed for the LAD and sign ranks tests used with double exponential errors and for the $t3$ test with $t3$ errors. Thus, under correct specification of the errors only one nuisance parameter appears in the asymptotic representation of the test statistics.

3. OBTAINING CRITICAL VALUES

It is well known that the asymptotic null distributions of the usual Dickey–Fuller tests do not depend on any nuisance parameters. Critical values can be calculated once and tabulated. For robust, non-least squares–based tests, the critical values cannot be tabulated so simply. The asymptotically correct critical values $q_e(\rho,a)$ and $q_t(\rho,a)$ are the ath quantiles of $\mathcal{E}_0$ and $\mathcal{T}_0$. For a given robust test ρ must be estimated, so the shape of the asymptotic null distribution is not estimable until we see the data.

There are many ways to accurately approximate the quantiles of $\mathcal{E}_0$ and $\mathcal{T}_0$. Shin and So (1999) simulate draws of $\mathcal{E}_0$ and $\mathcal{T}_0$ for each ρ and use the empir-

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

ical quantiles as critical values. Seo (1999) and Lucas (1998) also describe simulation-based procedures. Rothenberg and Stock (1997) and Thompson (2004b) use Cornish–Fisher expansions. Hansen (1997) and Juhl and Xiao (2002) simulate to calculate quantiles for similar test statistics.

A computationally convenient approach is to approximate $q_e(\rho, a)$ and $q_t(\rho, a)$ by a third-order polynomial in $1 - \rho$. For the $\mathcal{E}_0$ statistic the formula is

$$q_e(\rho, a) \approx x_\rho' \gamma_a,$$

where $x_\rho = (1, 1 - \rho, (1 - \rho)^2, (1 - \rho)^3)'$ and γ_a is a four-dimensional coefficient vector. The formula is defined similarly for the $\mathcal{T}_0$-statistic. To compute γ_a I simulated $N = 5$ million draws of $\mathcal{E}_0$ and $\mathcal{T}_0$ for $\rho = \rho_0, \ldots, \rho_{100}$, where $\rho_j = j/100$.[3] I do not consider negative values of ρ because ρ can be shown to be nonnegative. For each ρ_j I sorted the test statistics and computed $\hat{q}_e(\rho_j, a)$ and $\hat{q}_e(\rho_j, a)$ as the (aN)th draws of the simulated statistics. For the $\mathcal{E}_0$ test I compute γ_a by regressing $\hat{q}_e(\rho_j, a)$ on x_{ρ_j}. The minimization problem is

$$\hat{\gamma}_a = \text{argmin}_{\gamma_\alpha} \sum_{j=0}^{100} (\hat{q}_e(\rho_j, a) - x_{\rho_j}' \gamma_a)^2.$$

The same procedure is used for $\mathcal{T}_0$. Table 3 provides the estimated coefficients for a equal to 0.01, 0.025, 0.05, and 0.10.

In practice I calculate critical values using a consistent estimate $\hat{\rho}$ for ρ. So the e-tests reject the null at level a when M_e or R_e are less than $x_{\hat{\rho}}' \hat{\gamma}_a$. Similarly the t-tests reject when M_t or R_t are less than $x_{\hat{\rho}}' \hat{\gamma}_a$. The estimation of $\hat{\rho}$ and the other nuisance parameters is discussed in Section 4.

Figure 1 plots the simulated quantiles along with the polynomial approximations. The quantiles are nearly linear in ρ, and it is difficult to see a difference between the quantiles and the approximating polynomial. I assess the quality of the approximation in terms of the worst large sample size distortion:

$$\max_{0 \le j \le 100} |\widehat{\Pr}[\mathcal{E}_0 \le x_{\rho_j}' \hat{\gamma}_a] - a|.$$

The expression $\widehat{\Pr}[\mathcal{E}_0 \le z]$ is an estimate of $\Pr[\mathcal{E}_0 \le z]$. I calculate $\widehat{\Pr}$ by simulating 5 million draws of $\mathcal{E}_0$, independent of the draws used to estimate γ_α, and calculate $\widehat{\Pr}[\mathcal{E}_0 \le z]$ as the percentage of draws below z. The worst size distortion was calculated for the models with and without a trend and for both the $\mathcal{E}_0$- and $\mathcal{T}_0$-statistics. The maximum over the four cases was 0.0004942. Thus the polynomial approximations are highly accurate both in terms of approximating the quantiles and minimizing asymptotic size distortions.

3.1. Modifying the Tests to Eliminate the Nuisance Parameter

Herce (1996) and Hasan and Koenker (1997) suggest modifying the test statistics to obtain null distributions that are free of ρ. Recall that $\mathcal{E}_0$ and $\mathcal{T}_0$ are

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

TABLE 3. Estimated coefficients $\hat{\gamma}_a$ of polynomial approximation $x'_\rho \hat{\gamma}_a$ to the ath quantiles of the asymptotic null distributions

Test and model	Quantile (%)	Coefficients $\hat{\gamma}_a$ of polynomial			
		Intercept	$1-\rho$	$(1-\rho)^2$	$(1-\rho)^3$
e, intercept only (e.g., $\mu_2 = 0$)	1	−20.392	12.667	−0.823	−0.067
	2.5	−16.745	10.850	−0.673	−0.175
	5	−13.969	9.382	−0.366	−0.368
	10	−11.169	8.050	−0.455	−0.293
t, intercept only (e.g., $\mu_2 = 0$)	1	−3.418	0.556	0.636	−0.098
	2.5	−3.112	0.671	0.598	−0.118
	5	−2.853	0.780	0.548	−0.118
	10	−2.560	0.920	0.494	−0.135
e, time trend (μ_2 unknown)	1	−28.846	18.593	−0.766	−0.318
	2.5	−24.680	16.630	−0.734	−0.355
	5	−21.427	15.106	−0.681	−0.390
	10	−18.037	13.504	−0.672	−0.333
t, time trend (μ_2 unknown)	1	−3.934	0.998	0.656	−0.044
	2.5	−3.637	1.100	0.671	−0.091
	5	−3.389	1.213	0.658	−0.126
	10	−3.109	1.377	0.584	−0.135

Note: The term x_ρ is $(1, 1-\rho, (1-\rho)^2, (1-\rho)^3)'$.

expressed as linear combinations of the "Dickey–Fuller" terms A_0/B_0 and $A_0/\sqrt{B_0}$ and another random variable. Because the least squares test statistics converge to the Dickey–Fuller terms, we can write down a linear combination of the robust test statistics and of the least squares statistics so that the limiting null distribution does not depend on ρ. Let e_{ls} and t_{ls} denote the m_e- and m_t-statistics when estimation is by least squares. The modified tests are based on the relationships

$$\text{Modified } e\text{-test: } (1-\hat{\rho}^2)^{-1/2}[m_e - \hat{\rho} e_{ls}] \Rightarrow S/B_c^{1/2} + (1-\rho^2)^{-1/2}[\lambda - \rho]c,$$

$$\text{Modified } t\text{-test: } (1-\hat{\rho}^2)^{-1/2}[m_t - \hat{\rho} t_{ls}] \Rightarrow S + B_c^{1/2}(1-\rho^2)^{-1/2}[\lambda - \rho]c,$$

where $\hat{\rho}$ is a consistent estimator for ρ. A test that rejects the null when the modified statistics are below the $100a\%$ quantile of $S/B_0^{1/2}$ or S will have asymptotic size equal to a. The rank tests may be similarly modified.

Unfortunately, this modification seriously affects the power of the tests. Let Φ and ϕ denote the distribution function and density function of a standard normal variable. When the errors are i.i.d. normal, the rank test formula for λ is

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

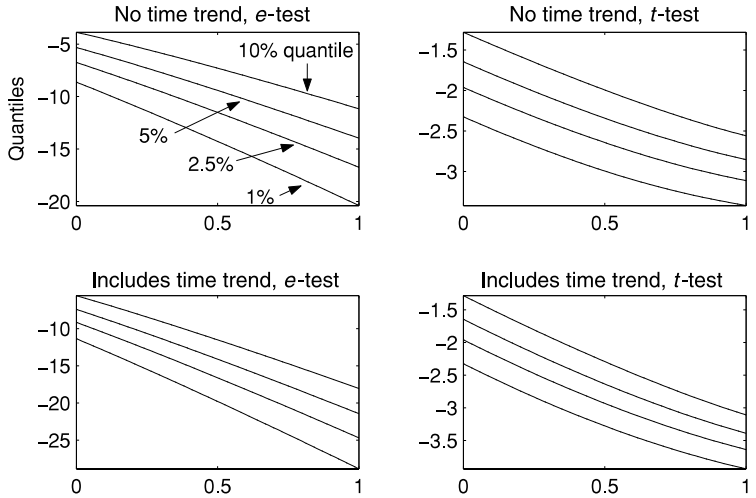


FIGURE 1. Simulated 1%, 2.5%, 5%, and 10% quantiles (solid lines) plotted along with the polynomial approximations from Table 3 (dotted lines). The approximations are so close that it is difficult to distinguish the solid from the dotted lines. The horizontal axes give values of ρ.

$$\lambda = -\frac{\sigma_\varepsilon}{\sigma_\psi}\int \varphi\left(\Phi\left(\frac{x}{\sigma_\varepsilon}\right)\right) d\left[\frac{1}{\sigma_\varepsilon}\phi\left(\frac{x}{\sigma_\varepsilon}\right)\right]$$

$$= \frac{1}{\sigma_\psi \sigma_\varepsilon}\int x\varphi\left(\Phi\left(\frac{x}{\sigma_\varepsilon}\right)\right)\frac{1}{\sigma_\varepsilon}\phi\left(\frac{x}{\sigma_\varepsilon}\right) dx$$

$$= \mathrm{Corr}[\varepsilon_1, \varphi(F(\varepsilon_1))].$$

The same result holds for the *M*-tests. Thus, under normal errors $\lambda = \rho$ no matter what the test, so the modified *t*-test has local power everywhere equal to its size. This may be confirmed in Table 2, where under normality λ equals ρ for every test considered. In Table 4 we also show that the modified *e*-test also has very little power with normal errors. Lucas (1995b) reports simulation results for *M*-tests that are consistent with these results. Herce (1996) obtains a stronger result for the LAD estimator—under normal errors the modified LAD *t*-test has power equal to its size. Although our results apply only to the local power function, in Section 5 we provide Monte Carlo evidence that the actual power function exhibits similar behavior for empirically relevant sample sizes. Robust methods are meant to perform well for plausible misspecifications of the model. By this measure the traditional Dickey–Fuller methods are robust relative to the modified statistics.

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

TABLE 4. Local alternative where asymptotic power is approximately 50%

		Innovation Process						
		$A(\tau)\exp(-\lvert x\rvert^{\tau})$				Mixture		
Test	Test type	$\tau = 2.0$	$\tau = 1.8$	$\tau = 1.2$	$\tau = 1.0$	$m = 0.1$	$m = 0.2$	$t3$
LAD or sign-median ranks	e	−17.50	−16.50	−11.50	−9.25	−13.00	−11.50	−9.75
	Modified e	−100	−95	−12.25	−7.50	−18.00	−12.50	−9.00
	t	−30	−30	−11.75	−8.25	−15.25	−12.25	−9.50
	Modified t	−100	−100	−18.50	−9.00	−40	−19.50	−12.00
Huber's M	e	−16.00	−15.75	−14.00	−13.50	−12.75	−12.25	−11.00
	Modified e	−100	−90	−17.75	−13.25	−11.75	−9.75	−8.50
	t	−18.25	−17.50	−14.75	−13.25	−12.25	−11.50	−9.75
	Modified t	−100	−100	−35	−30	−17.25	−13.25	−10.75
Student's $t3$	e	−16.00	−14.50	−13.25	−11.25	−12.00	−10.50	−9.25
	Modified e	−100	−20	−14.75	−10.00	−12.25	−9.25	−7.50
	t	−18.00	−15.25	−13.50	−10.50	−12.00	−9.75	−8.25
	Modified t	−100	−45	−30	−13.50	−18.50	−12.25	−9.00
Wilcoxon ranks	e	−16.00	−15.50	−13.50	−12.25	−12.75	−11.75	−10.25
	Modified e	−100	−60	−14.00	−9.75	−11.75	−9.50	−7.50
	t	−18.00	−17.25	−13.50	−11.50	−12.25	−10.75	−8.75
	Modified t	−100	−100	−30	−13.25	−17.75	−12.50	−9.25
Normal ranks	e	−15.75	−15.75	−14.75	−13.75	−13.75	−13.25	−11.50
	Modified e	−100	−100	−15.00	−10.00	−12.00	−9.00	−8.50
	t	−17.50	−17.50	−15.50	−13.50	−13.75	−12.50	−10.25
	Modified t	−100	−100	−30	−13.50	−18.25	−11.75	−11.00

Notes: The ratio of local alternatives for two different tests is their asymptotic relative efficiency at 50% power. In nearly all cases the unmodified tests are more efficient. Entries are largest local alternatives where asymptotic power is at least 50%. Local power was calculated for alternatives from $c = 0$ to -100, at increments of -0.25 for alternatives 0 through -20 and increments of -5 for alternatives -25 through -100 (so $c = 0, -0.25, -0.5, \ldots, -20, -25, -30, \ldots, -100$). Size is 5%. Results for the LAD and sign ranks tests are given on the same line because the two tests are asymptotically equivalent. To see this, notice from Table 1 that for these tests η is the median of ε_1 and $\psi(x - \eta) = \varphi(F(x))$.

Table 4 provides some evidence to compare the modified tests with the unmodified tests that use the polynomial approximation to the critical value. The table lists the local alternative where asymptotic power is approximately $\frac{1}{2}$ in the model that includes a linear time trend (e.g., no restrictions on μ_2). King (1988) and Elliot, Rothenberg, and Stock (1996) make the case for preferring tests that achieve 50% power at alternatives closer to zero. By definition, tests that achieve 50% power at lower values of c have better asymptotic relative efficiency at 50% power (see Stock, 1994, p. 2775).

The modified tests have virtually no power for distributions close to the Gaussian, but sometimes they beat the unmodified tests for the thick-tailed double exponential and $t3$ errors. Asymptotic power was calculated for local alterna-

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

tives as small as -100, so a value of -100 in the table means that the test did not achieve 50% power for any alternative greater than -100. Although the modified *e*-test has power greater than size for Gaussian errors, the modified *e*-tests considered here do not achieve 50% power for any alternative greater than -100. Both the modified and unmodified *e*-tests beat the corresponding *t*-tests for error distributions close to normality, and the *t*-tests are more powerful for thicker tailed error distributions.

The unmodified robust tests essentially dominate least squares. The least squares–based *e*- and *t*-tests of Dickey and Fuller (1979) achieve 50% power at $c = -15.75$ and -17.5; the unmodified robust tests do slightly worse for normal errors and much better for nonnormal errors. For each robust test the presence of thick-tailed errors leads to greater power relative to normal innovations. There is a larger decline in power under normality for the LAD and sign ranks tests, but even in these two cases the decrease is small relative to the gain in power for thicker tailed distributions.

The normal ranks test is at least as efficient as the least squares test for every error distribution in the table. In the classical location model we can make the stronger statement that the normal ranks test is asymptotically at least as powerful as the least squares *t*-test no matter what the error distribution. This result is shown by Chernoff and Savage (1958) (also see the discussion in Hettmansperger, 1984, Sec. 2.9). The result may hold for the unit root problem, but I was not able to prove it. At normal errors both the Dickey–Fuller and the normal ranks tests have $\lambda = \rho = 1$, so their power functions are identical. I am able to show that for the normal ranks test, $\lambda \geq 1$ and $\rho \geq 1/\lambda$ for any error distribution. Power obviously increases in λ, but simulations show that power is not monotonic in ρ. However, I conducted simulations[4] suggesting that in the model with a time trend, large sample power of the *e*- and *t*-tests for $\lambda \geq 1$ and $\rho \geq 1/\lambda$ is always at least as high as when $\lambda = \rho = 1$. In the model with no time trend the result holds only for the *e*-test. In that model the *t*-test has about 1% less power than the Dickey–Fuller test in the range $0.90 \leq \rho \leq 0.92$ and $1.06 \leq \lambda \leq 1.11$. Although I am unable to find an error distribution with these nuisance parameters, one may exist.

Chernoff and Savage (1958) establish that $\lambda \geq 1$ for the normal ranks test; see for example their Theorem 3 or Theorem 2.9.2 of Hettmansperger (1984). To see that $\rho \geq 1/\lambda$, note that $\sigma_\psi^2 = 1$ and assume without loss of generality that $\sigma_\epsilon^2 = 1$. Thus

$$\rho = \int y\Phi^{-1}(F(y))f(y)\,dy = \int F^{-1}(\Phi(x))x\phi(x)\,dx = -\int F^{-1}(\Phi(x))\,d\phi(x),$$

where f is the density of the errors and ϕ is a standard normal density. The second equality follows from the change of variables $x = \Phi^{-1}(F(y))$, and the third equality follows from the fact that $\partial\phi(x)/\partial x = -x\phi(x)$. Integration by parts and a change of variables leads to the expressions

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

$$\rho = \int \phi(\Phi^{-1}(t))\, dF^{-1}(t) = \int \frac{\phi(\Phi^{-1}(F(x)))}{f(x)}\, f(x)\, dx = E\left[\frac{\phi(\Phi^{-1}(F(\varepsilon_1)))}{f(\varepsilon_1)}\right],$$

$$\lambda = \int f(\Phi(t))\, d\Phi^{-1}(t) = \int \frac{f(x)}{\phi(\Phi^{-1}(F(x)))}\, f(x)\, dx = E\left[\frac{f(\varepsilon_1)}{\phi(\Phi^{-1}(F(\varepsilon_1)))}\right].$$

Let $W = f(\varepsilon_1)/\phi(\Phi^{-1}(F(\varepsilon_1)))$. By Jensen's inequality,

$\rho = E[1/W] \geq 1/E[W] = 1/\lambda$.

3.2. Lucas's (1995b) Method for Obtaining Critical Values

Lucas (1995b) obtains critical values by computing the distribution of the test statistics from realizations of simulated data. Lucas simulates the data by generating a time series from the data generating process with i.i.d. standard normal innovations v_t and with d_t and γ set to zero. Because the parameter ρ depends on both the choice of estimation method and the error density, this approach will give asymptotically incorrect critical values for the robust tests unless the errors are in fact normal. For the tests based on the $t3$ estimator, $\rho =$ 0.98 for normal errors and the 5% quantiles of $\mathcal{E}_0$ and $\mathcal{T}_0$ are -21.8 and -3.42 when the model has a time trend. For $t3$ errors $\rho = 0.71$ and the quantiles are -17.5 and -3.01. A simulation shows that when the errors are distributed $t3$, using the quantiles based on normal errors instead of the correct values results in e- and t-tests with asymptotic sizes of 1.59% and 1.68%, respectively.[5]

4. CALCULATING THE TESTS

To calculate the test statistics we need consistent estimates for the various nuisance parameters. Likewise, the polynomial approximation to the critical values, $x'_{\hat{\rho}}\hat{\gamma}_\alpha$, is evaluated at a consistent estimate $\hat{\rho}$ for ρ.

Following Hasan and Koenker (1997), the nuisance parameters are estimated under the null. For a given M-test with associated functions g and ψ, let $\tilde{\epsilon}$ denote the vector of residuals $\Delta Y - Z\tilde{\beta}$, where $\tilde{\beta}$ minimizes $1'g(\Delta Y - Z\beta)$. Theorem 1 implies that under the null, $\tilde{\epsilon}$ will converge to the recentered errors $\epsilon - \eta$. So we estimate the two variances $\sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_1)$ and $\sigma_\psi^2 = \mathrm{Var}(\psi(\varepsilon_1 - \eta))$ with

$$\hat{\sigma}_\varepsilon^2 = (\tilde{\epsilon} - \bar{\tilde{\epsilon}})'(\tilde{\epsilon} - \bar{\tilde{\epsilon}})/T \quad \text{and} \quad \hat{\sigma}_\psi^2 = (\psi(\tilde{\epsilon}) - \bar{\psi})'(\psi(\tilde{\epsilon}) - \bar{\psi})/T,$$

where $\bar{\tilde{\epsilon}}$ and $\bar{\psi}$ are the sample averages of $\tilde{\epsilon}$ and $\psi(\tilde{\epsilon})$. Similarly, the correlation $\rho = \mathrm{Corr}[\varepsilon_1, \psi(\varepsilon_1 - \eta)]$ is estimated by

$$\hat{\rho} = (\tilde{\epsilon} - \bar{\tilde{\epsilon}})'(\psi(\tilde{\epsilon}) - \bar{\psi})/(T\hat{\sigma}_\varepsilon\hat{\sigma}_\psi).$$

For Huber's M-test and the $t3$-test ω is estimated by $1'\nabla\psi(\tilde{\epsilon})/T$, where $\nabla\psi(x)$ equals $\psi'(x)$ if ψ is differentiable at x and zero otherwise. Following Herce

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

(1996), for the LAD test I estimate $\omega = 2f(\eta)$ using the kernel estimator $\hat{f}(\eta) = 1'\phi(\tilde{\epsilon}/h)/(Th)$, where $h = 1.06\hat{\sigma}_\varepsilon T^{-1/5}$. The bandwidth h was chosen based on a recommendation in Silverman (1986, Ch. 2). The term $\hat{f}(\eta)$ will converge in probability to $f(\eta)$ because $\tilde{\epsilon}$ is converging to $\epsilon - \eta$. Following Lucas (1995b), the scale parameter k appearing in Huber's M function is $1.345(\hat{\epsilon}'(\hat{\epsilon} - 1'\hat{\epsilon}/T)/T)^{1/2}$, where $\hat{\epsilon} = \Delta Y - X\hat{\alpha} - Z\hat{\beta}$ are the residuals from a preliminary LAD regression.

For the rank tests the nuisance parameters are calculated using the same formulas. I follow Hasan and Koenker (1997) and set the residuals $\tilde{\epsilon}$ equal to $\Delta Y - Z\tilde{\beta}$, where $\tilde{\beta}$ minimizes the LAD objective function $1'|\Delta Y - Z\beta|$. For the rank tests $\psi(x)$ is replaced by $\varphi(\tilde{F}(x))$, where $\tilde{F}$ is an estimate of F. Following Hasan and Koenker (1997), I take $\tilde{F}(x)$ to be the empirical distribution function of $\tilde{\epsilon}$, equal to $(\{\#\text{residuals} \leq x\} - 0.5)/(T - p)$. The term σ_ψ^2 is known in closed form for the rank tests and does not need to be estimated; σ_ψ^2 is $\frac{1}{12}$ for the Wilcoxon ranks test and 1 for the normal and sign ranks tests.

Computing the rank tests requires calculating the vector $b = \int_0^1 \hat{a}(s)\, d\varphi(s)$, where $\hat{a}$ solves the linear programming problem in (2). Hasan and Koenker (1997) use an algorithm that calculates $\hat{a}(s)$ for all s. I instead solve the linear programming problem in expression (2) at the points $\{\tau_i\}_{i=1}^{50}$ (with $\tau_i = (i-1)/49$) and use numerical integration. For the Wilcoxon ranks test, $b = \sum_{i=1}^{50} \hat{a}(\tau_i)/50 - 0.5$. For the normal ranks test, b is $\sum_{i=1}^{50} \hat{a}(\tau_i)/(50\phi(\Phi^{-1}(\tau_i)))$ 1and for the sign ranks tests it is $0.5\,\text{sign}(\hat{a}(\tau_{26}) - 0.5)$. I use the discretized version of the $\hat{a}$ process because canned algorithms for solving linear programming problems are widely available—for example, I used the Matlab function linprog.

5. MONTE CARLO EVIDENCE

We carried out a Monte Carlo study to evaluate the small sample performance of the various tests. The calculation of the various tests was described in Section 4.

Table 5 lists empirical rejection rates under the null ($\gamma = 0$) and for a distant alternative ($\gamma = -0.5$). The errors ε_t are i.i.d. draws from either the $N(0,1)$ or Student's $t3$ distribution. We make the unrealistic assumption that the researcher knows that the errors are i.i.d. and sets the lag length p equal to zero. Results are given for both the intercept only model and for the model with a time trend. Small sample size looks good for all the tests except the modified e. The LAD test exhibits some size distortions. Thompson (2004b) argues that this results from a poor estimate of ω, because $\hat{f}(\eta)$ does not converge at a root-T rate.

Previously we have shown that the modified tests have poor power against the local alternative c. The Monte Carlo results suggest that power is also poor against the distant alternative $\gamma = -0.5$. Test power is very close to one for most of the unmodified tests, whereas the modified tests suffer from large power

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

TABLE 5. Empirical rejection rates, i.i.d. errors

	e-tests		*t*-tests		Modified *e*-tests		Modified *t*-tests	
Test	$N(0,1)$	$t3$	$N(0,1)$	$t3$	$N(0,1)$	$t3$	$N(0,1)$	$t3$
True $\gamma = 0$, intercept only model, $(\mu_1, \mu_2) = (1,0)$								
Least squares	0.06	0.06	0.05	0.05	—	—	—	—
LAD	0.05	0.04	0.03	0.02	0.04	0.03	0.03	0.01
Huber's *M*	0.06	0.07	0.05	0.06	0.04	0.08	0.03	0.05
Student's *t*3	0.06	0.07	0.05	0.06	0.04	0.08	0.04	0.06
Wilcoxon ranks	0.06	0.06	0.05	0.04	0.08	0.07	0.03	0.04
Normal ranks	0.06	0.06	0.06	0.05	0.15	0.10	0.06	0.06
Sign ranks	0.06	0.07	0.05	0.06	0.07	0.08	0.05	0.06
True $\gamma = 0$, model includes intercept and time trend, $(\mu_1, \mu_2) = (1,1)$								
Least squares	0.06	0.06	0.05	0.05	—	—	—	—
LAD	0.06	0.04	0.03	0.02	0.05	0.02	0.02	0.01
Huber's *M*	0.07	0.08	0.06	0.07	0.04	0.08	0.03	0.06
Student's *t*3	0.06	0.08	0.06	0.08	0.06	0.09	0.04	0.06
Wilcoxon ranks	0.06	0.06	0.05	0.04	0.11	0.08	0.03	0.04
Normal ranks	0.07	0.06	0.06	0.05	0.17	0.12	0.04	0.06
Sign ranks	0.08	0.07	0.06	0.05	0.08	0.08	0.06	0.05
True $\gamma = -0.5$, intercept only model, $(\mu_1, \mu_2) = (1,0)$								
Least squares	0.99	1.00	0.98	0.99	—	—	—	—
LAD	1.00	1.00	0.78	0.95	0.22	0.61	0.02	0.22
Huber's *M*	0.99	1.00	0.98	1.00	0.28	0.92	0.03	0.69
Student's *t*3	0.99	1.00	0.97	1.00	0.31	0.94	0.04	0.73
Wilcoxon ranks	0.99	1.00	0.96	1.00	0.32	0.78	0.01	0.50
Normal ranks	0.99	1.00	0.98	1.00	0.47	0.76	0.06	0.55
Sign ranks	0.98	1.00	0.82	0.98	0.34	0.79	0.04	0.37
True $\gamma = -0.5$, model includes intercept and time trend, $(\mu_1, \mu_2) = (1,1)$								
Least squares	0.97	0.97	0.95	0.94	—	—	—	—
LAD	0.92	0.98	0.59	0.86	0.17	0.52	0.02	0.22
Huber's *M*	0.97	0.99	0.93	0.99	0.23	0.87	0.03	0.68
Student's *t*3	0.97	0.99	0.91	0.99	0.26	0.90	0.05	0.75
Wilcoxon ranks	0.96	0.99	0.88	0.98	0.28	0.68	0.01	0.44
Normal ranks	0.97	0.99	0.93	0.97	0.40	0.65	0.04	0.47
Sign ranks	0.92	0.98	0.68	0.93	0.26	0.69	0.05	0.36

Notes: There are 1,000 Monte Carlo replications. Sample size is $n = 100$. Under normality the modified *t*-tests have virtually no power against the distant alternative $\gamma = -0.5$.

losses. As the asymptotics predict, the modified *t*-tests have power equal to size for normal errors. The asymptotics also predict that the modified tests will beat the unmodified ones for the thick-tailed Student's *t*3 distribution, and this

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

does not seem to be the case. One explanation for this is that the local-to-null asymptotic approximations become less accurate for distant alternatives such as $\gamma = -0.5$. Thus the Monte Carlo results make a stronger case against the modified tests than the asymptotics.

The calculations in Table 5 assume that the errors are independent. To investigate the impact of serially correlated errors, Table 6 provides evidence about the small sample behavior of the unmodified tests when the errors exhibit serial dependence of unknown form. The data generating processes are

$$\text{AR}(1): v_t = 0.5v_{t-1} + \zeta_t,$$

$$\text{MA}(1): v_t = .\zeta_t - 0.5\zeta_{t-1},$$

$$\text{GARCH MA}: v_t = \vartheta_t - 0.5\vartheta_{t-1}, \qquad \vartheta_t = h_t^{1/2}\zeta_t,$$

$$h_t = 1 + 0.65h_{t-1} + 0.25\vartheta_{t-1}^2, \qquad h_0 = 0.$$

The fundamental innovations ζ_t have initial condition $\zeta_0 = 0$ and were generated from either the $N(0,1)$ or the $t3$ distribution. The table gives results for the intercept only model and also for the model that includes a time trend.

Bringing the tests to serially correlated data requires an algorithm for selecting the lag length p. The term p is chosen by the Bayesian information criterion of Schwarz (1978) constrained so $2 \leq p \leq 6$. Because the tests are designed for thick-tailed data, we calculate Schwarz's criterion from a double-exponential likelihood with constant variance.

Size is generally accurate for AR errors, less accurate for MA errors, and worse for GARCH errors. The *t*-tests exhibit more accurate size than the *e*-tests. For the rank *t*-tests, rejection rates under the null range from 0.04 to 0.11. The rank tests in general exhibit better size properties than the *M*-tests. Thompson (2004b) argues that this occurs because of estimation error associated with $\hat{\omega}$. Computing the rank tests does not require an estimate of $\hat{\omega}$. Although in large samples the estimation error will have an asymptotically negligible effect on inference, in small samples it could affect test size.

As the asymptotic analysis predicted, the robust tests essentially dominate least squares at the alternative $\gamma = -0.1$. Except for the LAD and sign ranks tests, the robust tests are about as good as least squares at normal errors and are much more powerful at $t3$ errors. The asymptotics also suggested that the *e*-tests would be more powerful than the *t*-tests for normal errors but less powerful for $t3$ errors, but in the Monte Carlo the *e*-tests are at least as powerful as the *t*-tests for both distributions. Taking the size and power results together, the unmodified Wilcoxon ranks and normal ranks *t*-tests are nice alternatives to least squares: both rank tests have accurate size, and both do at least as well as least squares over a range of error distributions.

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

TABLE 6. Empirical rejection rates with serially correlated errors

	e-tests						*t*-tests					
	$N(0,1)$ errors			$t3$ errors			$N(0,1)$ errors			$t3$ errors		
Test	AR	MA	GH	AR	MA	GH	AR	MA	GH	AR	MA	GH
True $\gamma = 0$, intercept only model, $(\mu_1, \mu_2) = (1,0)$												
Least squares	0.06	0.11	0.13	0.06	0.11	0.16	0.05	0.09	0.09	0.05	0.09	0.08
LAD	0.05	0.09	0.12	0.03	0.08	0.17	0.03	0.04	0.06	0.02	0.04	0.13
Huber's *M*	0.06	0.12	0.14	0.07	0.13	0.22	0.05	0.09	0.11	0.06	0.11	0.18
Student's *t*3	0.06	0.12	0.18	0.07	0.14	0.48	0.06	0.09	0.15	0.07	0.11	0.48
Wilcoxon ranks	0.07	0.11	0.12	0.07	0.12	0.16	0.05	0.09	0.07	0.05	0.09	0.08
Normal ranks	0.08	0.12	0.13	0.07	0.12	0.16	0.06	0.10	0.08	0.05	0.10	0.10
Sign ranks	0.06	0.12	0.12	0.07	0.13	0.15	0.05	0.08	0.07	0.06	0.09	0.07
True $\gamma = 0$, model includes intercept and time trend, $(\mu_1, \mu_2) = (1,1)$												
Least squares	0.05	0.11	0.13	0.05	0.12	0.14	0.05	0.10	0.10	0.05	0.11	0.11
LAD	0.06	0.11	0.13	0.05	0.08	0.21	0.03	0.06	0.09	0.01	0.04	0.16
Huber's *M*	0.05	0.11	0.16	0.07	0.15	0.24	0.05	0.09	0.14	0.07	0.16	0.24
Student's *t*3	0.05	0.11	0.24	0.09	0.17	0.62	0.05	0.11	0.25	0.07	0.16	0.64
Wilcoxon ranks	0.07	0.12	0.13	0.06	0.12	0.14	0.05	0.09	0.06	0.04	0.10	0.08
Normal ranks	0.07	0.12	0.13	0.07	0.12	0.15	0.06	0.11	0.08	0.05	0.11	0.09
Sign ranks	0.08	0.13	0.13	0.08	0.14	0.16	0.06	0.11	0.07	0.05	0.11	0.06
True $\gamma = -0.1$, intercept only model, $(\mu_1, \mu_2) = (1,0)$												
Least squares	0.35	0.64	0.64	0.35	0.63	0.58	0.23	0.49	0.48	0.22	0.47	0.44
LAD	0.30	0.54	0.59	0.39	0.70	0.72	0.11	0.23	0.29	0.18	0.43	0.54
Huber's *M*	0.37	0.65	0.68	0.51	0.79	0.72	0.24	0.48	0.53	0.44	0.74	0.68
Student's *t*3	0.36	0.64	0.70	0.61	0.88	0.92	0.23	0.44	0.50	0.54	0.80	0.89
Wilcoxon ranks	0.38	0.62	0.60	0.53	0.81	0.63	0.20	0.42	0.37	0.46	0.73	0.39
Normal ranks	0.39	0.65	0.64	0.49	0.77	0.63	0.24	0.47	0.44	0.42	0.71	0.45
Sign ranks	0.37	0.58	0.55	0.55	0.82	0.58	0.20	0.31	0.25	0.42	0.67	0.24
True $\gamma = -0.1$, model includes intercept and time trend, $(\mu_1, \mu_2) = (1,1)$												
Least squares	0.18	0.40	0.42	0.17	0.38	0.42	0.13	0.32	0.33	0.12	0.31	0.32
LAD	0.17	0.33	0.40	0.20	0.44	0.58	0.08	0.13	0.23	0.10	0.26	0.47
Huber's *M*	0.19	0.41	0.46	0.28	0.56	0.59	0.14	0.32	0.40	0.30	0.59	0.61
Student's *t*3	0.20	0.41	0.59	0.39	0.68	0.91	0.13	0.32	0.48	0.42	0.70	0.91
Wilcoxon ranks	0.21	0.39	0.38	0.28	0.56	0.40	0.10	0.26	0.21	0.28	0.54	0.22
Normal ranks	0.21	0.42	0.41	0.26	0.52	0.39	0.13	0.30	0.28	0.25	0.51	0.24
Sign ranks	0.23	0.37	0.35	0.33	0.61	0.39	0.12	0.24	0.15	0.31	0.52	0.16

Notes: There are 1,000 Monte Carlo replications. Sample size is $n = 100$. GH denotes the GARCH MA data generating process. The robust tests are slightly less powerful than least squares for normal errors and much more powerful for $t3$ errors.

6. CONCLUSION

A goal of the robustness literature is to produce a test that is almost as powerful as least squares–based tests for normal errors and much more powerful for

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

thicker tailed distributions. Although the modified tests have poor power, unmodified tests can achieve this goal. Shin and So (1999) document large power gains from using tests that adapt to the shape of the error distribution. Thompson (2004a) provides examples where the robust tests are more powerful than the point-optimal DF-GLS test of Elliot et al. (1996). Thompson (2004b) constructs confidence intervals for the parameter c by inverting the rank t-test R_t. Under normality the robust intervals are almost as short as the least squares–based intervals of Stock (1991) and Elliot and Stock (2001), and under nonnormal distributions the robust intervals tend to be much shorter.

NOTES

1. Numerical integration was carried out using the Matlab function quad, which implements a quadrature algorithm. All integrals were solved between -40 and 40, so, for example, ω was calculated as $\int_{-40}^{40}[f(x)]^2\,dx$ for the Wilcoxon scores tests. Kurtosis is defined to be $E[(\epsilon - E\epsilon)^4]/\mathrm{Var}(\epsilon)^2$. We calculate the quantity by numerical integration.

2. The normalizing constants $A(\tau)$ are chosen to make the densities integrate to one.

3. Throughout the paper simulations from the distributions of A_c and B_c were computed as the realizations of normalized sums of 500 successive draws from a discrete time Gaussian AR(1) process with autoregressive parameter $1 - c/T$.

4. I carried out the following simulation. Set $\rho_i = i/50$, $\lambda_j = 50/j$, and $c_k = -k$. For each $(\rho, \lambda, c) = (\rho_i, \lambda_j, c_k)$ triple over $i = 1,\ldots,50$, $j = 1,\ldots,50$, $k = 1,\ldots,50$, I simulated 100,000 Monte Carlo draws of $\mathcal{E}_c$ and $\mathcal{T}_c$. For each c I computed power for a 5% level test using the polynomial approximations to the critical values given in Table 3.

5. There are 50,000 Monte Carlo replications.

REFERENCES

Chan, N.H. & C.Z. Wei (1987) Asymptotic inference for nearly nonstationary Ar(1) processes. *Annals of Statistics* 15, 1050–1063.

Chernoff, H. & R. Savage (1958) Asymptotic normality and efficiency of certain nonparametric test statistics. *Annals of Mathematical Statistics* 29, 972–994.

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

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

Elliott, G., T. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

Elliott, G. & J.H. Stock (2001) Confidence intervals for autoregressive coefficients near one. *Journal of Econometrics* 103, 155–181.

Hansen, B.E. (1997) Approximate asymptotic p values for structural-change tests. *Journal of Business and Economic Statistics* 15, 60–67.

Hasan, M.N. & R.W. Koenker (1997) Robust rank tests of the unit root hypothesis. *Econometrica* 65, 133–161.

Herce, M.A. (1996) Asymptotic theory of LAD estimation in a unit root process with finite variance errors. *Econometric Theory* 12, 129–153.

Hettmansperger, T.P. (1984) *Statistical Inference Based on Ranks*. Wiley.

Hjort, H.L. & D. Pollard (1993) Asymptotics for Minimizers of Convex Processes. Preprint, Department of Statistics, Yale University.

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

Hoek, H., A. Lucas, & H.K. van Dijk (1995) Classical and Bayesian aspects of robust unit root inference. *Journal of Econometrics* 69, 27–59.
Juhl, T. & Z. Xiao (2002) Partially Linear Models with Unit Roots. Discussion paper 1359, Cowles Foundation.
King, M.L. (1988) Towards a theory of point optimal testing. *Econometric Reviews* 6, 169–218.
Koenker, R. & G. Bassett (1978) Regression quantiles. *Econometrica* 46, 33–50.
Lucas, A. (1995a) An outlier robust unit root test with an application to the extended Nelson-Plosser data. *Journal of Econometrics* 66, 153–173.
Lucas, A. (1995b) Unit root tests based on *m* estimators. *Econometric Theory* 11, 331–346.
Lucas, A. (1998) Inference on cointegrating ranks with LR and LM tests based on pseudo-likelihoods. *Econometric Reviews* 17, 185–214.
Phillips, P.C.B. (1987) Toward a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.
Phillips, P.C.B. (1988) Regression theory for near-integrated time series. *Econometrica* 56, 1021–1043.
Pollard, D. (1985) New ways to prove central limit theorems. *Econometric Theory* 1, 295–314.
Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.
Rothenberg, T.J. & J.H. Stock (1997) Inference in a nearly integrated autoregressive model with nonnormal innovations. *Journal of Econometrics* 80, 269–286.
Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics* 6, 461–464.
Seo, B. (1999) Distribution theory for unit root tests with conditional heteroskedasticity. *Journal of Econometrics* 91, 113–144.
Shin, D.W. & B.S. So (1999) Unit root tests based on adaptive maximum likelihood estimation. *Econometric Theory* 15, 1–23.
Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
Stock, J.H. (1991) Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics* 28, 435–459.
Stock, J.H. (1994) Unit roots, structural breaks, and trends. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2739–2841. Elsevier.
Thompson, S.B. (2004a) Optimal versus robust inference in nearly integrated non-Gaussian models. *Econometric Theory* 20, 23–55.
Thompson, S.B. (2004b) Robust confidence intervals for autoregressive coefficients near one. In *Fetschrift in Honor of Thomas Rothenberg.* Cambridge University Press, forthcoming.

APPENDIX

Here we prove Theorem 1. We will derive the results for *M*-estimators for the model with a linear time trend. All of the results may be extended to the model with an intercept only (e.g., $\mu_1 = 0$). The required results for the rank tests are contained in the Appendix to Hasan and Koenker (1997). Throughout the Appendix $\|\cdot\|$ denotes the norm: $\|b_{ij}\| = (\sum_{i,j} b_{ij}^2)^{1/2}$.

To incorporate estimators based on the LAD and Huber's *M* function, we allow g to have finitely many points of nondifferentiability if continuity and several growth conditions hold. We assume that g is convex because it simplifies the proofs. The proofs rely on an asymptotic approximation to the objective function, and assuming convexity allows

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

us to conclude that our approximation, which is proved keeping the parameters fixed, holds uniformly over the parameters in a neighborhood of their true values. Convexity also greatly simplifies the demonstration of the rate of convergence of the estimators. Our extensive use of convexity is due to results in Pollard (1991) and Hjort and Pollard (1993). Some objective functions, such as the Student's t, are not convex. Because this objective function is three times differentiable, the usual Taylor series–based arguments can be used to show that Theorem 1 holds for this objective function.

Assumption 1 (Objective Function). $g(x)$ is convex and strictly increasing in $|x|$; $g(x)$ is everywhere twice differentiable except for x in P, where P contains the D points $p_1,\ldots,p_D$. g satisfies the Lipshitz condition $|g(x) - g(y)| \leq H|x - y|$. $|\psi(x)| \leq H$, and there exist two finite points x_L and x_H so that $\psi(x)$ is constant for $x > x_H$ and for $x < x_L$. The term $|g''(x)| < H$ for x not in P.

This assumption rules out the least squares estimator, because the squared error function $g(x) = x^2$ does not satisfy the Lipschitz condition.

Because g may not be everywhere differentiable, we cannot use the usual Taylor series expansions to approximate the objective function. Instead of pointwise differentiability, the proofs make use of stochastic differentiability as in Pollard (1985). This requires that I impose smoothness conditions on the error density to make up for the lack of smoothness in the objective function.

Assumption 2 (Smoothness of the Error Density). The errors $\{\varepsilon_t\}_{t=1}^{T}$ are i.i.d. mean zero with $E|\varepsilon_1|^{2+\delta} < H$ for some $\delta > 0$. The term ε_1 has a density function $f(z)$ that is bounded and uniformly continuous.

It will prove convenient to reparameterize the objective function. Consider the estimator $\hat{\theta}$ that minimizes $\sum_{t=p+2} q_t(\theta)$, where

$$q_t(\theta) = g(\varepsilon_t - \eta - w_t'\theta) - g(\varepsilon_t - \eta).$$

Here $w_t' = T^{-1/2}(u_{t-1}/T^{1/2}, 1, t/T, \Delta u_{t-1},\ldots,\Delta u_{t-p})'$ and $\hat{\theta}$ is a $3 + p$-dimensional parameter vector given by

$$\hat{\theta}_1 = T\hat{\alpha} - c\Gamma(1),$$

$$\hat{\theta}_2 = \sqrt{T}\left[(\hat{\alpha} - c\Gamma(1)/T)(\mu_1 - \mu_2) + \mu_2 \sum_{i=3}^{p}(\hat{\beta}_i - \beta_i) + (\hat{\beta}_1 - \beta_1) - \eta\right],$$

$$\hat{\theta}_3 = T^{3/2}[(\hat{\alpha} - c\Gamma(1)/T)\mu_2 + (\hat{\beta}_2 - \beta_2)],$$

$$\hat{\theta}_{3+i} = \sqrt{T}(\hat{\beta}_i - \beta_i)$$

for $i = 1,\ldots,p$. Because there is a one-to-one mapping between $\hat{\theta}$ and $(\hat{\alpha}, \hat{\beta})$, we will describe the asymptotic properties of $\hat{\theta}$.

Define the remainder term

$$r_t(\theta) = q_t(\theta) + w_t'\theta\psi(\varepsilon_t - \eta) - (\omega/2)\theta' w_t w_t'\theta.$$

The following two lemmas establish that $r_t(\theta)$ is asymptotically negligible, implying that the objective function is approximately quadratic.

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

LEMMA 1. *Define* $b_t = 1(T^{1/2}\|w_t\| > M)$. *For any* $\varepsilon > 0$, *we can take* M *large enough so that* $\lim_{T\to\infty} \Pr[|\sum r_t(\theta)b_t| > \varepsilon] < \varepsilon$.

Proof. Write the remainder as $\sum r_t(\theta)b_t = \sum E_{t-1}r_t(\theta)b_t + R_{0,T}$, where $R_{0,T} = \sum(r_t(\theta) - E_{t-1}r_t(\theta))b_t$. To handle the $E_{t-1}r_t(\theta)b_t$ term, define the function $h(u) = E_{t-1}[g(\varepsilon_t - \eta + u) - g(\varepsilon_t - \eta)]$, or equivalently

$$h(u) = \int_{\mathbb{R}} [g(s - \eta + u) - g(s - \eta)] f(s)\, ds. \tag{A.1}$$

We have imposed enough smoothness on g, ψ, and f so that $h(u)$ is twice continuously differentiable. By a Taylor series expansion,

$$h(-w_t'\theta) = h(0) - h'(0)w_t'\theta + (h''(u_t^*)/2)\theta' w_t w_t'\theta,$$

where $|u_t^*| \le |w_t'\theta|$. Notice that $h(0) = 0$, $h'(u) = E_{t-1}[\psi(\varepsilon_t - \eta + u)]$, and $h'(0) = 0$. The term $h''(u)$ can be shown to be $-\int \psi(x - \eta + u)\, df(x)$ by an argument based on integration by parts. The boundedness of f along with the conditions on ψ in Assumption 1 imply that there exists a bound B so that $\sup_u |h''(u)| < B$. Therefore

$$E\sum |E_{t-1}r_t(\theta)b_t| = E\sum \frac{|h''(u_t^*)| + \omega}{2}(\theta' w_t)^2 b_t \le \frac{B + \omega}{2} E\sum (\theta' w_t)^2 b_t.$$

By Hölder's inequality,

$$\begin{aligned} E\sum (\theta' w_t)^2 b_t &\le \sum [E(\theta' w_t)^{2+\delta}]^{1/(1+\delta/2)} (\Pr[T^{1/2}\|w_t\| > M])^{(2+\delta)/\delta} \\ &\le \epsilon(M)^{(2+\delta)/\delta} \left(T^{-1}\sum [E(T^{1/2}\theta' w_t)^{2+\delta}]^{1/(1+\delta/2)}\right), \end{aligned} \tag{A.2}$$

where $\epsilon(M) = \sup_t \Pr[T^{1/2}\|w_t\| > M]$. For all t, $T^{1/2}w_t$ converges to an $O_p(1)$ random variable, so $E(T^{1/2}\theta' w_t)^{2+\delta}$ is bounded and $\epsilon(M) \to 0$ as M increases. Thus we can pick M large enough to make the term small with high probability.

Next we handle the $R_{0,T}$ term:

$$ER_{0,T}^2 \le \sum Er_t^2(\theta)b_t.$$

The Lipshitz condition on g implies that $|r_t(\theta)| \le (2H + |\psi(\varepsilon_t - \eta)|)|w_t'\theta|$. Therefore

$$ER_{0,T}^2 \le (2H + \sigma_\psi)^2 E\sum (w_t'\theta)^2 b_t.$$

We showed in equation (A.2) that we can pick M large enough to make $E\sum(w_t'\theta)^2 b_t$ asymptotically negligible. ■

LEMMA 2. *For any fixed* M, $\sum r_t 1(T^{1/2}\|w_t\| < M) \xrightarrow{p} 0$.

Proof. It will prove convenient to examine the remainder over regions where it is differentiable and regions where it is not differentiable. Let $\widetilde{w}_t$ denote $w_t 1(T^{1/2}\|w_t\| < M)$ and let A_t denote the event that $|\varepsilon_t - \eta - p_i| > |\widetilde{w}_t'\theta|$ for all i. So $1(A_t)g(\varepsilon_t - \eta - \widetilde{w}_t'\theta)$ is differentiable in θ. Write the remainder as $\sum r_t 1(T^{1/2}\|w_t\| < M) = R_{2,T} + R_{3,T} + R_{4,T}$, where $R_{2,T} = \sum[1(A_t)z_t(\theta) - E_{t-1}1(A_t)z_t(\theta)]$, $R_{3,T} = \sum[1(A_t^c)z_t(\theta) - E_{t-1}1(A_t^c)z_t(\theta)]$,

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press

$$R_{4,T} = \sum E_{t-1}[g(\varepsilon_t - \eta - \widetilde{w}_t'\theta) - g(\varepsilon_t - \eta)] - (\omega/2)\theta'\widetilde{w}_t\widetilde{w}_t'\theta,$$

and $z_t(\theta) = g(\varepsilon_t - \eta - \widetilde{w}_t'\theta) - g(\varepsilon_t - \eta) + \widetilde{w}_t'\theta\psi(\varepsilon_t - \eta)$.

A Taylor series expansion yields

$$R_{2,T} = \sum \theta\widetilde{w}_t\widetilde{w}_t'\theta[1(A_t)\psi'(\varepsilon_t - \eta - u_t^*) - E_{t-1}1(A_t)\psi'(\varepsilon_t - \eta - u_t^*)],$$

where $|u_t^*| \le |\widetilde{w}_t'\theta|$. Because $R_{2,T}$ is a sum of martingale difference sequences,

$$ER_{2,T}^2 \le E\sum(\widetilde{w}_t'\theta)^4[\psi'(\varepsilon_t - \eta - u_t^*)]^2 \le T^{-1}H^2\|\theta\|^4M^4.$$

Therefore $\lim_{T\to\infty} ER_{2,T}^2 = 0$, so $R_{2,T}$ converges in probability to zero.

The condition $1(A_t^c) = 1$ puts a bound on the values ε_t can take. Therefore the moment $E1(A_t^c)z_t^2$ exists. For large enough T,

$$ER_{3,T}^2 \le E\sum E_{t-1}1(A_t^c)z_t^2$$

$$\le \|\theta\|^2H^2M^2\sum_{i=1}^{D}\Pr[|\varepsilon_1 - \eta - p_i| \le T^{-1/2}M\|\theta\|].$$

Because the density of ε_1 is bounded, $\lim_{T\to 0} ER_{3,T}^2 = 0$ and $|R_{3,T}|$ converges to zero in probability.

To handle $R_{4,T}$, define $h(u)$ as in (A.1). By a Taylor series expansion of h around 0,

$$|R_{4,T}| \le \frac{1}{2}\left\{|\omega - h''(0)| + \sup_t|h''(0) - h''(u_t^*)|\right\}M^2\|\theta\|^2.$$

The continuity of $h''(u)$ leads to the result $\lim_{T\to\infty}\{\sup_t|h''(0) - h''(u_t^*)|\} = 0$, with $\sup_t|u_t^*| < T^{-1/2}M\|\theta\|$. The term ω can be shown to be equal to $h''(0)$ using an argument based on integration by parts. Thus $R_{4,T}$ converges to zero in probability. ■

Proof of Theorem 1. Hasan and Koenker (1997) prove the result for the rank test statistics R_e and R_t. To deal with the M-statistics M_e and M_t, combine Lemmas 1 and 2 to obtain the pointwise convergence result $\sum r_t(\theta) \xrightarrow{p} 0$. By the convexity of g and Lemma 1 of Hjort and Pollard (1993), if $\sum r_t(\theta) \xrightarrow{p} 0$ for fixed θ then the convergence is uniform over θ in compact sets. Uniform convergence of the remainder, combined with the argument in Section 3 of Pollard (1991), leads to the asymptotic approximation

$$\hat{\theta} = \left[\omega\sum w_t w_t'\right]^{-1}\sum w_t\psi(\varepsilon_t - \eta) + o_p(1).$$

Lemma 3.1 of Phillips (1988) implies the weak convergence result

$$\left(\frac{u_{[rT]}}{\sqrt{T}}, \sum_{t\le[rT]}\frac{\psi(\varepsilon_t - \eta)}{\sqrt{T}}\right)' \Rightarrow \left(\frac{\sigma_\varepsilon}{\Gamma(1)}W_c(r), \sigma_\psi[\rho W_0(r) + \sqrt{1-\rho^2}\widetilde{W}(r)]\right)',$$

where $\widetilde{W}$ is standard Brownian motion, independent of W. The desired results follow from the usual applications of the continuous mapping theorem. ■

https://doi.org/10.1017/S0266466604202055 Published online by Cambridge University Press