Robust confidence intervals for autoregressive coefficients near one

Samuel Brodsky Thompson*
Harvard University

July 1, 2001

Abstract

We construct outlier robust confidence sets for autoregressive roots near unity. There are a few difficulties in doing this - the asymptotics for robust methods generally involve several poorly estimated nuisance parameters, and robust procedures are more difficult to compute than least squares based methods. We propose a family of "aligned" robust procedures that eliminate the need to estimate some of the nuisance parameters. The procedures are computationally no more burdensome than least squares. In thick-tailed data the robust sets outperform those based on normality.

1 Introduction

A recurring problem in financial econometrics is how to conduct valid inference on a linear mean function estimated from monthly, weekly or daily data. For example, most interest rate models specify the conditional mean to be linear in the previous value of the process. The data typically exhibit outliers and substantial serial dependence, and in most cases standard methods do not reject the presence of a unit root in the autoregressive representation of the series. For empirical problems such as quantifying the effect of parameter uncertainty on short term forecasts and asset pricing formulas, reporting only the unit root test and the parameter estimates are an unsatisfying way to describe the data. For these applications it can be useful to construct confidence sets for the largest autoregressive root of the series.

Both directly and through his students, Thomas Rothenberg has made many contributions to our understanding of inference for integrated and nearly integrated processes. One way to construct a confidence set is to invert a sequence of tests, where each test in the sequence evaluates a particular point null hypothesis. The confidence set contains all the point nulls that are not rejected by the sequence of tests. Elliott and Stock (2000) argued that, since a more powerful test leads to a more accurate

*I thank Thomas Rothenberg.

interval, inverting the asymptotically point optimal tests in Elliott, Rothenberg and Stock (1996) should lead to more accurate confidence sets than had been previously proposed (see Stock (1991), Andrews (1993) and Hansen (1999)). While it is not possible to do better than the tests in Elliott et al. (1996) when the innovations are normal, the outliers present in financial data suggest that other methods may lead to improvements.

This paper proposes "robust" confidence sets which have good accuracy for a variety of error distributions. We create the confidence sets by inverting a sequence of robust t tests. Unlike the tests in Elliott et al. (1996), t tests do not efficiently model the deterministic trend and are not point optimal for any particular distribution. We use t tests because they are robust: Thompson (2001*a*) showed that point optimal tests based on non normal likelihoods can behave very badly when the true error density is unknown and asymmetric. Thus there is a trade-off between Elliott and Stock's (2000) confidence sets, which efficiently handle deterministic trends, and the confidence sets described here, which are robust to outliers.

Since the robust test statistics generally have null distributions that depend on unknown nuisance parameters, finding appropriate critical values is a nontrivial problem. Furthermore, all the robust tests require estimation of more nuisance parameters than do the least squares tests, and the cumulative effect of estimation error can be large. In some cases nonparametric estimators must be used, resulting in slow rates of convergence. An additional problem is that robust tests are somewhat more difficult to compute than least squares-based tests.

We construct a sequence of robust tests which eliminate the need to estimate several of the unknown nuisance parameters. To compute critical values we extend a simple procedure suggested by Rothenberg and Stock (1997). Following Rothenberg and Thompson (2001) we compute "aligned" robust tests, which in some forms are no more difficult to compute than least squares-based tests.

The resulting confidence sets perform well so long as the errors are known to follow an iid process. In this case both asymptotic analysis and a Monte Carlo study show that the robust sets are useful alternatives to the sets proposed by Elliott and Stock (2000). We compare the areas of the various sets and conclude that with normal errors the robust sets do slightly worse than the procedures in Elliott and Stock (2000), and with non normal errors the robust sets do much better. However, when the errors follow an unknown serially correlated process neither the robust sets nor the Elliott and Stock (2000) sets have satisfactory small sample properties.

2 Inverting a sequence of tests

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

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

where $\{\varepsilon_t\}_{t=1}^T$ is an iid sequence with mean 0 and variance σ_ε^2. $\Gamma(L)$ 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 put a confidence region around γ.

Since I am interested in inference when γ is close to zero, 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. With this reparameterization, the augmented Dickey and Fuller (1979) representation of the model is

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

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

A $100(1-a)\,\%$ confidence set $C(y)$ where y is the data has the property that $\Pr_c[c\in C(y)]\geq 1-a$ for all c. Here $\Pr_c$ indicates that the probability is computed assuming c is the true autoregressive parameter. We construct $C(y)$ from a sequence of tests. Suppose, for each point $\overline{c}$ in the parameter space Θ, we construct a test of asymptotic size a for the hypothesis $c=\overline{c}$ versus $c\neq\overline{c}$. We define $C(y)$ as the set of all $\overline{c}$ that we fail to reject. In large samples, $C(y)$ has the desired property that $\Pr_c[c\in C(y)]\geq 1-a$ for all c.

Tests with high power lead to confidence sets with certain optimality properties. $\Pr_{c^*}[\overline{c}\in C(y)]$ is the probability we incorrectly include $\overline{c}$ when c^* is the true value. To make this probability as small as possible, we construct the confidence set using the most powerful test of the null $c=\overline{c}$ against the alternative $c=c^*$.

Stock (1991) constructed confidence sets by inverting a sequence of t-tests based on the ordinary least squares (OLS) estimator for α in equation 1. Each null hypothesis $\alpha=\overline{\alpha}$ versus $\alpha\neq\overline{\alpha}$ is rejected when the statistic $[X'MX]^{1/2}(\widehat{\alpha}_{ls}-\overline{\alpha})$ is too large or too small, where $\widehat{\alpha}_{ls}$ is the OLS estimator and M is the projection matrix $I-Z(Z'Z)^{-1}Z'$. This procedure gives us a confidence region for α which, combined with a consistent estimate of $\Gamma(1)$, leads to a confidence region for c. Andrews (1993) and Hansen (1999) also proposed confidence intervals based on t-tests.

In a stationary autoregressive model with Gaussian errors, the least squares t-test is asymptotically uniformly most powerful against all one-sided alternatives $c<\overline{c}$ or $c>\overline{c}$. Elliott et al. (1996) have shown that when γ is local to zero this result does not hold, and that there does not exist a uniformly most powerful test, even in large samples. Instead, for each null $c=\overline{c}$ we have a family of point optimal tests, each one most powerful only against the point alternative $c=\overline{c}$. While none of the point optimal tests dominate the others, they are generally more powerful than the t-test because the point optimal tests efficiently model the intercept and trend coefficients.

Elliott et al.'s (1996) tests are not point optimal when the error distribution contains outliers. Thompson (2001*a*) investigated whether power improvements could

Table 1: Robust tests

Test	$\psi(x)$ function	ω
Least Squares M-test	x	1
LAD M-test	$\operatorname{sign}(x)$	$2f(\eta)$
Huber's M-test	$x1(\lvert x\rvert \leq k) + k\operatorname{sign}(x)1(\lvert x\rvert > k)$	$\Pr[\lvert\varepsilon_1 - \eta\rvert \leq k]$
Student's t M-test	$(x/n)/(1+x^2/n)$	$\mathrm{E}\,\psi'(\varepsilon_1 - \eta)$
Wilcoxon rank test	$\widehat{F}(x) - 1/2$	$\mathrm{E}\,f(\varepsilon_1)$
Normal rank test	$\Phi^{-1}\left(\widehat{F}(x)\right)$	$\mathrm{E}\,f(\varepsilon_1)/\Phi'\left(\Phi^{-1}(F(\varepsilon_1))\right)$
Sign/Median rank test	$.5\operatorname{sign}\left(\widehat{F}(x) - 1/2\right)$	$2f(m)$

Notes: Φ is the distribution function for a standard normal. m is the median of ε_1. The parameter k which appears in Huber's M function is chosen by the researcher. The parameter n which appears in Student's t function is the number of degrees of freedom. η is defined in section 3.1.

be attained when using robust testing methods that are designed to improve power when the error distribution has thick tails. The improvement occurs when the true error distribution is known or at least is known to be symmetric. If the true error distribution is unknown and asymmetric, the robust point optimal tests can behave very badly. Since in practice we typically cannot exclude asymmetric errors, the viable unit root tests are the traditional robust tests (which inefficiently handle intercepts and trends) and the point-optimal Gaussian tests proposed in Elliott et al. (1996) (which are inefficient in the presence of thick-tailed errors).

Elliott and Stock (2000) constructed confidence sets based on the point optimal Gaussian tests. The present paper proposes a method for creating confidence sets from a sequence of traditional robust t tests which inefficiently handle the deterministic trend. One way to "robustify" the least squares-based t-test is to replace the OLS estimate for α with a robust M-estimate. M-estimators may be characterized as solutions to maximization problems or, equivalently, as solutions to first order conditions. It will be convenient to adopt the notation that, for any function $h(\cdot)$ mapping $\Re$ into $\Re$ and any n-dimensional column vector x with components x_i, $h(x)$ is the n-dimensional column vector with components $h(x_i)$. The M-estimators $(\widehat{\alpha}, \widehat{\beta})$ solve the equations

$$\begin{aligned} Z'\psi\left(\Delta Y - X\widehat{\alpha} - Z\widehat{\beta}\right) &= 0 \\ X'\psi\left(\Delta Y - X\widehat{\alpha} - Z\widehat{\beta}\right) &= 0 \end{aligned} \tag{2}$$

where $\psi(\cdot)$ is a scalar "score"-type function chosen by the researcher. The choice of ψ determines the sensitivity of $\widehat{\alpha}$ to outliers; when $\psi(x) = x$, $\widehat{\alpha}$ is the OLS estimator, and when $\psi(x) = \operatorname{sign}(x)$, $\widehat{\alpha}$ is the least absolute deviations (LAD) estimator. Some common choices for ψ are given in Table 1. Robust t-tests of the unit root hypothesis $\overline{\alpha} = 0$ have been proposed by Lucas (1995) and Herce (1996).

We propose an alternative sequence of tests based on a locally asymptotic representation of the t-statistic. Under both the null $\alpha = \overline{\alpha}$and fixed alternatives $\alpha \neq \overline{\alpha}$,

the t-statistic satisfies the approximation

$$[X'MX]^{1/2}(\widehat{\alpha}-\overline{\alpha}) = \frac{1}{\omega}\frac{X'M\psi\left(\Delta Y - X\overline{\alpha} - Z\widehat{\beta}_R(\overline{\alpha})\right)}{[X'MX]^{1/2}} + o_p(1)$$

where $\widehat{\beta}_R(\overline{\alpha})$ is the restricted estimator that solves the first order conditions with the null hypothesis $\alpha = \overline{\alpha}$ imposed:

$$Z'\psi\left(\Delta Y - X\overline{\alpha} - Z\widehat{\beta}_R(\overline{\alpha})\right) = 0.$$

ω is a nuisance parameter defined in section 3. Some robust tests and the corresponding ω parameters are given in Table 1. When $\psi(x) = x$, $\widehat{\alpha}$ is the OLS estimator and $\omega = 1$, and it is straightforward to show that the approximation holds exactly, with the $o_p(1)$ term equal to zero.

The parameter ω appears in the asymptotic null distribution of the t-statistic, and it must be estimated in order to obtain critical values. In large samples the estimates will converge to their true values and the estimation error will have an asymptotically negligible effect on inference. In small samples estimation error may affect the size and power of the tests. This is especially true if ω is poorly estimated. For example, $\omega = 2f(m)$ for LAD estimation, where $f(\cdot)$ is the density of the errors ϵ and m is their median. The standard kernel estimate of a density at a point does not converge to the true value at root-n speed and is quite variable in small samples.

Since eliminating the $(1/\omega)$ term does not affect asymptotic power, a natural alternative is to reject the null when the ratio

$$Q(\overline{\alpha}) = \frac{X'M\psi\left(\Delta Y - X\overline{\alpha} - Z\widehat{\beta}_R\right)}{[X'MX]^{1/2}}$$

is too large or too small. The asymptotic null distribution of this statistic does not depend on ω.

Computing the sequence of tests requires solving for $\widehat{\beta}_R(\overline{\alpha})$ at each $\overline{\alpha}$. Depending on the choice of ψ, this can be a computationally burdensome procedure. We suggest a slight generalization to the $Q(\overline{\alpha})$ statistic. Instead of computing $\widehat{\beta}_R(\overline{\alpha})$ using ψ, compute it using an alternative set of first order conditions:

$$Z'\phi\left(\Delta Y - X\overline{\alpha} - Z\widehat{\beta}_R(\overline{\alpha})\right) = 0. \tag{3}$$

ϕ is a function chosen by the researcher. We could take $\phi = \psi$, or we could take ϕ to be a function that leads to computationally convenient solutions for $\widehat{\beta}_R(\overline{\alpha})$. For example, $\phi(x) = x$ leads to OLS estimates of $\widehat{\beta}_R(\overline{\alpha})$, and $\phi(x) = \text{sign}(x)$ leads to LAD estimates. If ϕ, ψ and the error density are all symmetric around zero, the asymptotic distribution of $Q(\overline{\alpha})$ does not depend on ϕ. While symmetry of the errors may not hold, in many cases the tests will retain their robustness to outliers. Following Adichie (1986) and Akritas (1991), we call the test based on $Q(\overline{\alpha})$ an "aligned" M-test.

Stock (1991) and Elliott and Stock (2000) proposed confidence sets based on inverting a single test statistic. It is also possible to construct robust confidence sets based on a single statistic, like $Q(0)$. We choose instead to compute a sequence of test statistics, because under each null hypothesis $\alpha = \overline{\alpha}$ the parameter ω does not appear in the null distribution of $Q(\overline{\alpha})$. Inverting a single statistic requires us to calculate its distribution under both the null and under various alternatives. The results in section 3 show that ω appears in the distribution of $Q(\overline{\alpha})$ when $\alpha \neq \overline{\alpha}$. Thus computing a sequence of test statistics allows us to avoid computing the additional nuisance parameter.

We can also construct $Q(\overline{\alpha})$ statistics which are asymptotically equivalent to the rank-based unit root tests described in Hasan and Koenker (1997). Consider a test of the null hypothesis $\alpha = \overline{\alpha}$ using the test statistic $[X'MX]^{-1/2} X'Mb$, where b is the vector of regression rankscores:

$$b = \int_0^1 1\left(\Delta Y - X\overline{\alpha} - Z\widetilde{\beta}(\tau) > 0\right) d\varphi(\tau).$$

$1(\cdot)$ is the indicator function, $\widetilde{\beta}(\tau)$ are the coefficients of the regression quantiles estimated under the null[1], and φ is a function chosen by the researcher. $?\overline{\alpha} = 0$. Under the null, MX is independent of any function of $\Delta Y - X\overline{\alpha} - Z\beta$, including their ranks. When Z is a vector of ones and $\varphi(\tau) = \tau$, each element of b is the rank of the corresponding element of $\Delta Y - X\overline{\alpha}$. The φ function allows the researcher to consider general functions of ranks.

Confidence regions for α may be constructed using a sequence of tests based on $[X'MX]^{-1/2} X'Mb$. We instead use the computationally simpler family of aligned rank tests. Given a pair of functions φ and ϕ, for each null hypothesis $\alpha = \overline{\alpha}$ use the ϕ function to compute the residuals $\Delta Y - X\overline{\alpha} - Z\widehat{\beta}_R(\overline{\alpha})$ from equation 3. Letting e_t denote the t^{th} residual, compute the empirical distribution function $\widehat{F}(x) = (\{\#e_s \leq x, s = 1, ..., T\} - .5)/(T - p)$. Form the $Q(\overline{\alpha})$ statistic, taking $\psi(x) = \varphi\left(\widehat{F}(x)\right)$. This test is based on the idea that under the null, functions of the ranked residuals should be approximately independent of MX. To see where the ranked residuals show up in this test, construct the vector $R(\overline{\alpha})$ with each element equal to the rank of the corresponding residual, and notice that

$$\varphi\left(\widehat{F}\left(\Delta Y - X\overline{\alpha} - Z\widehat{\beta}_R(\overline{\alpha})\right)\right) = \varphi\left(\frac{R(\overline{\alpha}) - .5}{T - p}\right).$$

Subject to regularity conditions, every choice for ϕ leads to a $Q(\overline{\alpha})$ test with the same asymptotic null distribution and power function as the test based on $[X'MX]^{-1/2} X'Mb$. Table 1 lists common rank tests and the corresponding choices for ψ.

[1] $\widetilde{B}(\tau)$ solves the minimization problem $\min_\beta \sum_{t=p+1}^{T} \Psi_\tau(\Delta Y - Z\beta)$ with $\Psi_\tau(u) = u(\tau - 1(u < 0))$. See Hasan and Koenker (1997), Koenker and Bassett (1978) and Gutenbrunner and Jureckova (1992) for details.

3 Asymptotic analysis

The limiting distribution of $Q(\overline{\alpha})$ can be expressed as a functional of Brownian motion. Define $W(\cdot)$ to be standard Brownian motion and define $W_c(\cdot)$ to be the Ornstein-Uhlenbeck process $W_c(t) = \int_0^t \exp\{c(t-s)\}\, dW(s)$. Under the null that $\alpha = \overline{\alpha}$, the limiting form of the OLS-based statistic $[X'MX]^{1/2}(\widehat{\alpha}_{ls} - \overline{\alpha})$ is $\sigma_\varepsilon DF(\overline{c})$, where

$$DF(\overline{c}) \equiv \frac{\int_0^1 D_{\overline{c}}(r)\, dW(r)}{\sqrt{\int_0^1 D_{\overline{c}}^2(r)\, dr}}$$

and $\overline{c} = \Gamma^{-1}(1)\,\overline{\alpha}$. The process $D_c(r)$ is defined to be $W_c(r) - \int_0^1 W_c(s)\, ds$ in the intercept only model (e.g., $\mu_2 = 0$ and there is no time trend in the design matrix Z), and $D_c(r)$ equals $W_c(r) - 2\int_0^1 (2 - 3s - r(3-6s)) W_c(s)\, ds$ in the model with a linear time trend. We have the following result.[2]

Proposition 1 *$\widehat{\sigma}_\psi^{-1} Q\left(\widehat{\Gamma}(1)\,\overline{c}\right)$ converges weakly to $\mathcal{Q}(c,\rho) + \lambda(c-\overline{c})\sqrt{\int D_c^2(r)\, dr}$, where*

$$\mathcal{Q}(c,\rho) \equiv \rho DF(c) + \sqrt{1-\rho^2}\, N(0,1).$$

$N(0,1)$ denotes a standard normal variable, independent of $DF(\overline{c})$. ρ and λ are nuisance parameters. $\widehat{\sigma}_\psi$ and $\widehat{\Gamma}(1)$ are consistent estimates of the nuisance parameters σ_ψ and $\Gamma(1)$.

The nuisance parameters ρ, λ and σ_ψ are defined in sections 3.1 and 3.2. In large samples the nuisance parameters ρ and λ determine the null distribution and power function of each test. Under the null $\overline{c} = c$ and λ disappears from the asymptotic representation. Thus ρ controls the null distribution of the test, while ρ and λ together affect power.

3.1 Obtaining critical values

Obtaining a rejection region for $Q(\overline{\alpha})$ requires us to calculate the quantiles of $\mathcal{Q}(\overline{c},\rho)$ and to consistently estimate the nuisance parameters. Let $k_l(\overline{c},\rho)$ and $k_u(\overline{c},\rho)$ denote the quantiles of $\mathcal{Q}(\overline{c},\rho)$ that solve $\Pr[k_l(\overline{c},\rho) \le \mathcal{Q}(\overline{c},\rho) \le k_u(\overline{c},\rho)] = a$. Let $\widehat{\rho}$, $\widehat{\sigma}_\psi$ and $\widehat{\Gamma}(1)$ denote consistent estimates of the nuisance parameters. The null hypothesis $c = \overline{c}$ is rejected in favor of the two sided alternative $c \neq \overline{c}$ when $\widehat{\sigma}_\psi^{-1} Q\left(\widehat{\Gamma}^{-1}(1)\,\overline{c}\right)$ is either below $k_l(\overline{c},\widehat{\rho})$ or above $k_u(\overline{c},\widehat{\rho})$. In large samples this test has size equal to a.

For the aligned M tests ρ and σ_ψ^2 are $\mathrm{Corr}[\varepsilon_1, \psi(\varepsilon_1 - \eta)]$ and $\mathrm{Var}[\psi(\varepsilon_1 - \eta)]$, where η solves $\mathrm{E}\,\phi(\varepsilon_1 - \eta) = 0$. For the aligned rank tests $\rho = \mathrm{Corr}[\varepsilon_1, \varphi(F(\varepsilon_1))]$ and $\sigma_\psi^2 = \mathrm{Var}[\varphi(U)]$, where $F(x)$ is the distribution function $\Pr[\varepsilon_1 \le x]$ and U is a random variable distributed uniformly on the unit interval.

ρ generally depends on ψ, φ and ϕ, and the error density. An exception to this is Stock's (1991) OLS-based test, where $\psi(x) = x$ and $\rho = 1$ no matter what the

[2] All proofs are collected in the Appendix.

Table 2: Values of ρ and λ

	ρ				λ			
	$N(0,1)$	DE	Log Norm	Mix	$N(0,1)$	DE	Log Norm	Mix
Least Squares	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LAD	.80	.71	.63	.47	.80	1.42	.99	1.86
Huber's M	.97	.94	.82	.78	.97	1.13	1.59	1.98
Student's t3	.95	.89	.72	.61	.95	1.21	1.70	2.19
Wilcoxon	.98	.92	.68	.64	.98	1.22	2.72	2.17
Normal ranks	1.00	.98	.76	.77	1.00	1.28	3.55	2.06
Sign ranks	.80	.71	.52	.47	.80	1.42	1.73	1.86

Notes: Expectations were computed as empirical averages of 1,000,000 simulated draws from the error distribution. $N(0,1)$ indicates standard normal draws, DE indicates double exponential, Log Norm is Log Normal, and Mixture is mixture normal. The random draws are normalized to have zero mean and unit variance. For each test $\phi(x) = x$.

error density. For other tests ρ differs from 1 and the limiting null representation becomes a linear combination of the "Dickey-Fuller" term $DF(\overline{c})$ and a standard normal variable. For the M tests σ_ψ^2 depends on both ψ and the error density, and for the rank tests σ_ψ is known and does not need to be estimated.

Table 2 lists values for ρ for various tests and error densities. It includes the aligned M tests based on LAD, Huber's function and the Student's t density with three degrees of freedom (denoted t3).[3] It also includes the three rank tests considered in Hasan and Koenker (1997): the tests based on Wilcoxon, Normal and Sign ranks. The error distributions considered are the standard normal as well as the thicker-tailed double exponential, log normal, and mixture normal distributions. For the mixture distribution a standard normal variable is drawn with probability .95 and a $N(0,36)$ variable is drawn with probability .05. For the M tests $\phi(x)$ equals x. For the rank tests the choice of ϕ does not affect ρ.

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 quantiles of $\mathcal{Q}(\overline{c},\rho)$ are not known in closed form. A number of methods have been proposed for approximating quantiles of statistics similar to $\mathcal{Q}(\overline{c},\rho)$. Stock (1991) calculated quantiles of $\mathcal{Q}(\overline{c},1)$. Lucas (1995), Herce (1996), Hasan and Koenker (1997), Seo (1999) and Thompson (2001*b*) all proposed methods to handle null distributions of the form $\mathcal{Q}(0,\rho)$. All of these methods could be extended to the general statistic $\mathcal{Q}(\overline{c},\rho)$.

We adopt a method proposed by Rothenberg and Stock (1997), who encountered a statistic similar to $\mathcal{Q}(\overline{c},\rho)$. It turns out that, no matter what the trend specification or the value for $\overline{c}$, the "Dickey-Fuller" term $DF(\overline{c})$ is approximately normally

[3]Following Lucas (1995), the scale parameter k in Huber's function is set to $1.345\sigma_\varepsilon$.

distributed. Thus $\mathcal{Q}(\overline{c},\rho)$ is approximately a linear combination of two normal variables, and is approximately normal. Rothenberg and Stock (1997) suggested the Cornish-Fisher expansion as a way to calculate the critical value of a statistic similar to $\mathcal{Q}(\overline{c},\rho)$. The Cornish-Fisher expansion provides good approximations to the quantiles of distributions with approximately Gaussian shapes.

The Cornish-Fisher expansion for the a quantile of a statistic u is

$$\mathrm{E}(u) + [\mathrm{Var}(u)]^{1/2}[q_a + K_3(u)(q_a^2 - 1)/6]$$

where $K_3(u) = \mathrm{E}[u - \mathrm{E}\,u]^3/(\mathrm{Var}(u))^{3/2}$ and q_a is the a-percentile of the standard normal distribution. Correcting the mistake in Rothenberg and Stock (1997), one can show that

$$\begin{aligned}
\mathrm{E}(\mathcal{Q}(\overline{c},\rho)) &= \rho\,\mathrm{E}\,(DF(\overline{c})) \\
\mathrm{Var}(\mathcal{Q}(\overline{c},\rho)) &= \rho^2\,\mathrm{Var}\,(DF(\overline{c})) + \left(1-\rho^2\right) \\
\mathrm{E}\,(\mathcal{Q}(\overline{c},\rho) - \mathrm{E}\,\mathcal{Q}(\overline{c},\rho))^3 &= \rho^3\,\mathrm{E}\,(DF(\overline{c}) - \mathrm{E}\,DF(\overline{c}))^3
\end{aligned}$$

Figure 1 and table 3 contain Monte Carlo estimates of the first three moments of $DF(\overline{c})$ when the model includes an intercept and trend.[4]

3.2 Asymptotic power

Asymptotic power is the probability in large samples that the test rejects the null. Asymptotic power obviously increases with λ, which shifts the distribution of the test statistic to the left when c is negative. ρ also affects power by changing the shape of the distribution. Results in Thompson (2001*b*) suggest that for the unit root hypothesis $c = 0$, the asymptotic power function is much more sensitive to λ than to ρ.

λ is defined equal to $\sigma_\varepsilon\omega/\sigma_\psi$, where ω is $\int_{\mathbb{R}} f(x)\,d\psi(x-\eta)$ for the M tests and $\int_0^1 f(F^{-1}(u))\,d\varphi(u)$ for the rank tests, and f denotes the density function for ε_1. When ψ and φ are everywhere differentiable the formulas have the more familiar form $\mathrm{E}\,\psi'(\varepsilon_1-\eta)$ and $\mathrm{E}\,\varphi'(F(\varepsilon_1))$. Formulas for ω for some robust tests appear in Table 1. It is clear from these formulas that λ generally depends on ψ, ϕ and the error density. For least-squares based tests, $\psi'(x) = 1$ and λ equals 1 no matter what the error density.

Table 2 lists values of λ for various tests and error densities. λ increases as the errors become thicker tailed, suggesting that thicker tailed errors lead to increased power. Under normal errors the parameter 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 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

[4]The simulations which appear in this paper were performed by computing stochastic integrals as normalized sums of 500 successive draws from a discrete time Gaussian AR(1) process with autoregressive parameter $1 + c/T$. There are 25,000 Monte Carlo replications.

Figure 1: First, second and third moments of $DF(c)$, intercept and trend case

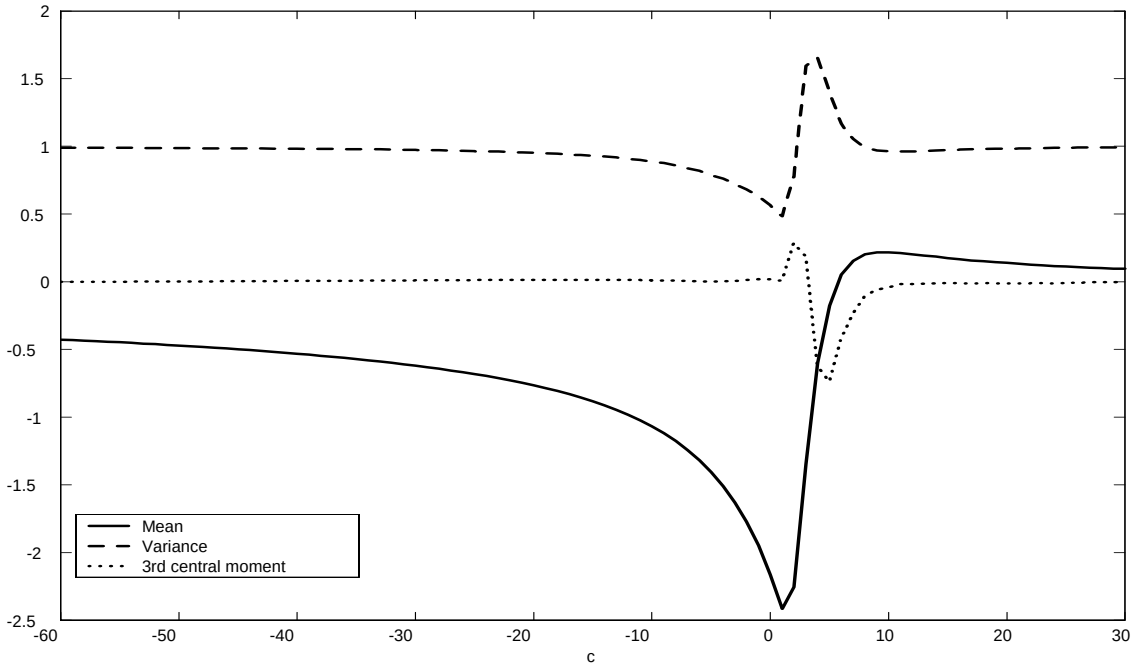


Table 3: Moments of $DF(c)$, model includes a linear time trend

c	$\mathrm{E}\,DF(c)$	$\mathrm{Var}(DF(c))$	$\mathrm{E}(DF(c)-\mathrm{E}\,DF(c))^3$
$c<-60$	$-20(62-c)^{-.8}$	1	0
$-60\leq c\leq 1$	$-2.013-\frac{15.809}{8.1-c}$	$1.112+\frac{.123}{100}c-\frac{3.635}{6.7-c}$	$.0136+\frac{.203}{1000}c$
2	-2.253	.772	.288
3	-1.348	1.594	.188
4	$-.602$	1.658	$-.621$
5	$-.176$	1.411	$-.737$
6	.051	1.169	$-.417$
7	.154	1.056	$-.231$
8	.202	.991	$-.106$
9	.216	.968	$-.056$
10	.215	.961	$-.042$
$11\leq c\leq 30$	$.347-.0147c+\frac{.209}{1000}c^2$	$.913+\frac{.524}{100}c-\frac{.870}{100^2}c^2$	$-.0213+\frac{.591}{1000}c$
$c>30$	$400/(35+c)^2$	1	0

squares when errors are normal and better than least squares when errors are thick tailed. In this sense they are robust to different error densities.

Thompson (2001*b*) showed that for a given error density, λ is maximized by choosing $-\psi$ or $-\varphi(F(x))$ equal to the log density of the errors. For the aligned rank tests this result holds for any choice of ϕ and for the aligned M tests it is true if $\phi = \psi$. The result is not surprising, since it corresponds to testing under correct specification of the error distribution. Thus, to maximize λ with Gaussian errors use the least squares or Normal ranks tests, and with double exponential errors use the LAD or Sign ranks tests.

In figure 2 we compare the large sample properties of the various confidence sets. Six data generating processes are considered: $c = 0$ with iid innovations v_t drawn from standard normal, log normal and mixture normal distributions, and $c = -30$ with iid innovations drawn from the same three distributions. We ignore the possibility of serially correlated innovations because they do not affect the asymptotic representations. The figure includes confidence sets based on seven t tests as well as Elliott and Stock's (2000) P_T test[5]. The P_T test efficiently handles trend coefficients and is point optimal in a Gaussian model. For the robust tests $\phi(x) = x$ so that preliminary estimation is by OLS. The model includes an intercept and time trend.

The figure displays measures of the area covered by the simulated confidence sets. Area is a standard optimality criteria for confidence sets. Sets with smaller areas are generally considered superior. For each test and data generating process, a confidence set was constructed from the sequence of 91 hypothesis tests of the nulls $c = -60$, -59, ..., 30. The area of each confidence set was approximated by the number of null hypotheses which were not rejected. For example, if the confidence set contains only the two values -10 and -11, the area is 2. The figure displays the 25%, 50% and 75% empirical quantiles of the areas of 25000 simulated intervals.

In large samples the robust confidence sets are useful alternatives to the P_T-based sets. Predictably, the P_T sets perform slightly better for Gaussian errors. When $c = 0$ half of the simulated P_T-based areas fall between 11 and 20, while half of the sets based on the Wilcoxon and Normal ranks tests fall between 13 and 23. However, with thick tailed error distributions the robust sets dominate P_T-based sets. When $c = -30$ and the errors are log normal, half of the P_T-based areas range from 32 to 36, compared with areas from 11 to 13 for the Wilcoxon ranks test and 9 to 10 for the Normal ranks test.

The performance of the robust sets is surprising. Thompson (2001*a*) showed that, when testing the unit root hypothesis, the efficiency loss due to using a point optimal test (thus ignoring thick-tailed errors) is often less than that due to using a robust t test (and inefficiently modelling the trend coefficients). For example, the point-optimal Gaussian unit root test is more powerful than many traditional robust unit root tests when the errors are drawn from a Student's t distribution with five or more degrees of freedom. The results in figure 2 suggest that the large power improvements gained from efficiently handling trends translate into small improve-

[5]The asymptotic representation for P_T is given in Elliott and Stock (2000). We use the representation for the the "fixed inital case."

ments in confidence set area. Even the least squares-based t test, which is everywhere dominated by the P_T test, performs only slightly worse in terms of area.

The case for using robust sets is weaker in the intercept only model. In figure 3, which depicts the intercept only case, the P_T-based sets are so short at $c = 0$ that there is little room for improvement. At normal error half of the simulated P_T-based areas fall between 6 and 11, while half of the sets based on the Wilcoxon and Normal ranks tests fall between 8 and 16. At the alternative $c = -30$ the robust sets lead to larger gains with thick tailed distributions. While half of the P_T areas are from 29 to 34 at log normal errors, the numbers for the Wilcoxon and Normal ranks sets are 11 to 13 and 8 to 10.

Figures 2 and 3 provide a number of additional interesting results. All of the sets are larger at $c = -30$ than at $c = 0$. The Wilcoxon ranks and Normal ranks tests have good properties for all the error distributions. In contrast to the other robust sets, the LAD-based sets perform poorly with the thick tailed log normal distribution. The performance of the LAD-based set can be improved by choosing $\phi(x)$ equal to $\text{sign}(x)$ instead of x. The formulas for the nuisance parameters lead immediately to the result that picking $\psi(x) = \phi(x) = \text{sign}(x)$ is asymptotically equivalent to constructing intervals from the Sign ranks test. Thus the much improved large sample results for the LAD-based set appear in the figures.

4 Monte Carlo evidence

We conducted a Monte Carlo study to investigate the small sample properties of the various confidence sets. In all of the simulations that follow we chose $\phi(x)$ equal to x so that estimation is by least squares. With this choice for ϕ the tests can be computed in closed form. Our algorithm for constructing a confidence region for c follows.

1. Choose a finite list of points $\{\overline{c}_i\}_{i=1}^r$. We pick $\overline{c}_i = -60 + i$ for $i = 0, ..., 90$.

2. For each $\overline{c}_i$, compute the hypothesis test of the null $c = \overline{c}_i$ against the two-sided alternative $c \neq \overline{c}_i$:

   (a) Compute the estimates $\widehat{\rho}$, $\widehat{\sigma}_\psi$ and $\widehat{\Gamma}(1)$. Following Elliott and Stock (2000), we use autoregressive estimates from least squares estimation of the equation

   $$\Delta y_t^d = a_0 y_{t-1}^d + \sum_{i=1}^{p} a_i \Delta y_{t-i}^d + e_t$$

   where y_t^d is obtained by GLS detrending.[6] Let $\widehat{e}_t$ denote the residuals from this regression. For the M tests, $\widehat{\sigma}_\psi^2$ is $T^{-1}\sum\left(\psi(\widehat{e}_t) - \overline{\psi}\right)^2$, $\widehat{\sigma}_\varepsilon^2$ is

[6] $y_t^d = y_t - (1,t)\beta^{dt}$, where β^{dt} is obtained by regressing $\left(y_1, y_2 - T^{-1}\overline{c}_i y_1, ..., y_T - T^{-1}\overline{c}_i y_{T-1}\right)'$ on the $T \times 2$ matrix with first row equal to $(1,1)$ and t^{th} row equal to $\left(-cT^{-1}, t - (t-1)cT^{-1}\right)$ for $t > 1$ The number of lags p is chosen by Ng and Perron's (2000) MAIC procedure with a maximum of four lags.

Figure 2: Asymptotic 25^{th}, 50^{th}, and 75^{th} quantiles of the area covered by each confidence set. The model includes an intercept and time trend, with $(\mu_1, \mu_2) = (1, 1)$. (Key to tests: 1 = P_T, 2 =least squares, 3 =LAD, 4 =Huber, 5 =t3, 6 =Wilcoxon, 7 =Normal ranks, 8 =Sign ranks.)

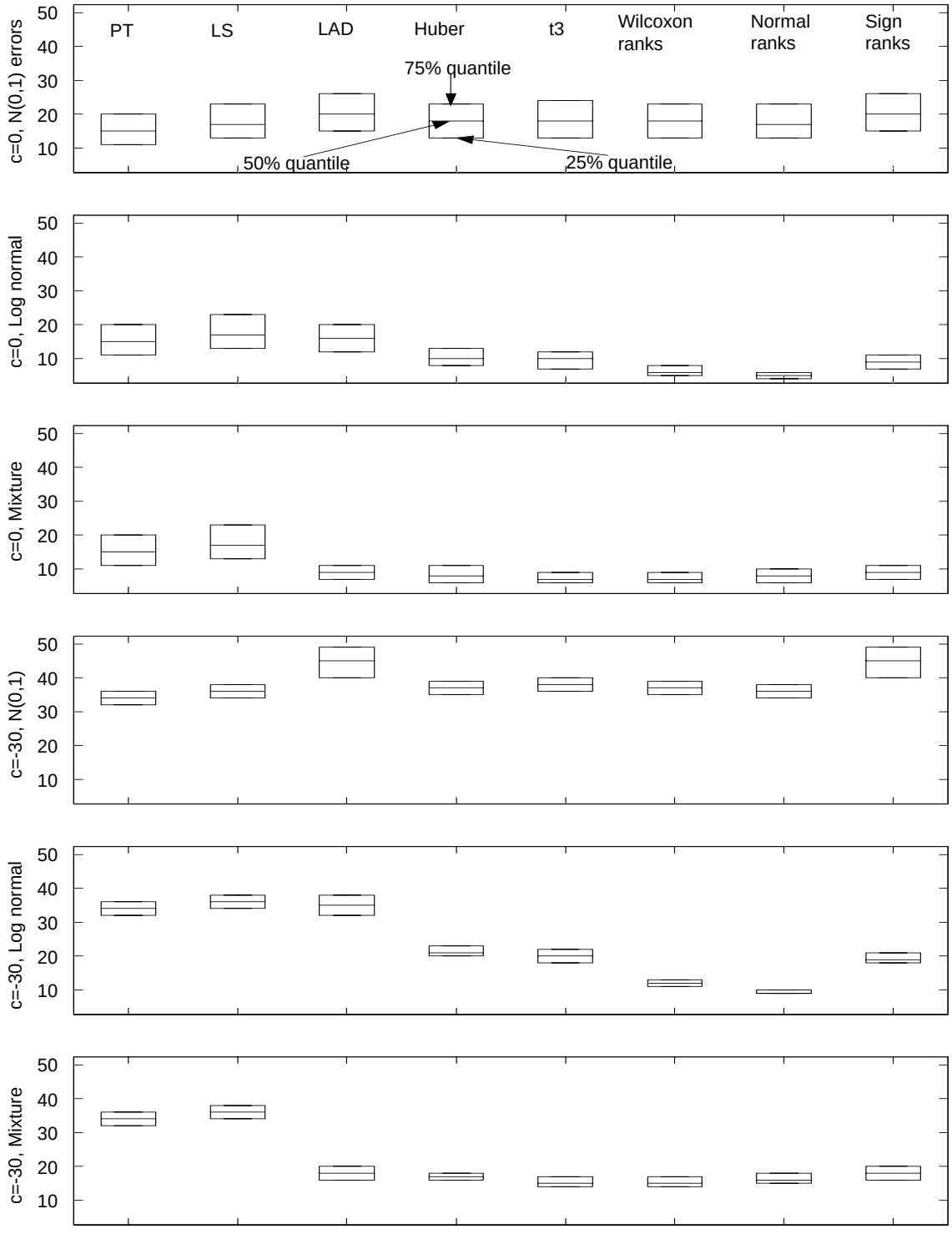
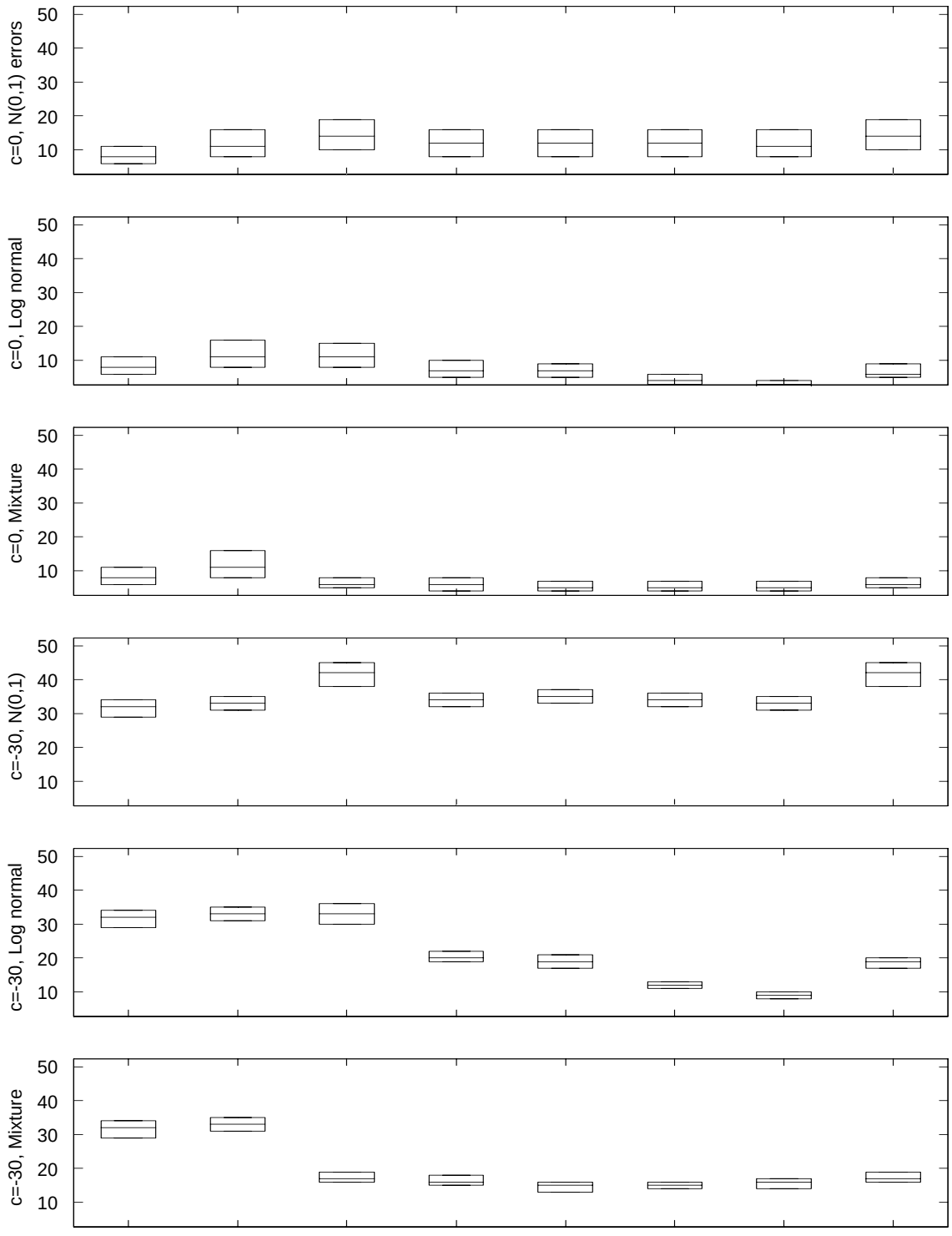


Figure 3: Asymptotic 25^{th}, 50^{th}, and 75^{th} quantiles of the area covered by each confidence set. The model includes an intercept only, with $\mu_1 = 1$. (Key to tests: $1 = P_T$, 2 =least squares, 3 =LAD, 4 =Huber, 5 =t3, 6 =Wilcoxon, 7 =Normal ranks, 8 =Sign ranks.)



$T^{-1}\sum(\widehat{e}_t-\overline{e})^2$ and $\widehat{\rho}$ equals $T^{-1}\sum\psi(\widehat{e}_t)(\widehat{e}_t-\overline{e})/\widehat{\sigma}_\varepsilon\widehat{\sigma}_\psi$. $\overline{\psi}$ and $\overline{e}$ denote the usual sample averages. For the rank tests the same formulas hold with $\psi(x)$ replaced by $\varphi(\widetilde{F}(x))$, where $\widetilde{F}(x)$ is the empirical distribution function for $\widehat{e}_t$. $\widehat{\Gamma}(1)$ is estimated by $1-\sum_{i=1}^{p}\widehat{a}_i$.

(b) Calculate $\widehat{\beta}_R\left(\overline{c}_i\widehat{\Gamma}(1)\right)$ by regressing $Y-\overline{c}_i\widehat{\Gamma}(1)X$ on Z. Calculate the residuals MX from a regression of X on Z.

(c) Choose a ψ or φ function and compute the test statistic $Q\left(\overline{c}_i\widehat{\Gamma}(1)\right)$.

(d) Use the method in section 3.1 to calculate the a_l^{th} and a_h^{th} quantiles of $\mathcal{Q}(\overline{c},\rho)$. In the Monte Carlo study we use the 2.5% and 97.5% quantiles. The null is rejected when $\widehat{\sigma}_\psi^{-1}Q\left(\overline{c}_i\widehat{\Gamma}(1)\right)$ is greater than the a_h^{th} quantile or less than the a_l^{th} quantile.

3. The $100(a_h-a_l)$-percent confidence region consists of all $\overline{c}_i$ which the hypothesis tests fail to reject.

Figure 4 provides encouraging results about the small sample coverage probabilities of the robust intervals. The four graphs depict empirical coverage probabilities from 2000 Monte carlo replications of sample size 100 from the model with iid errors v_t drawn from the standard normal, double exponential, log normal and mixture normal distributions. The true values for $1-\gamma$ range from 0 to 1.2. We make the unrealistic assumption that the reseacher knows that the errors are iid and includes no lagged Δy_t terms in the design matrix. The model includes an intercept and time trend.

In this simple setup the small sample coverage probabilities are quite close to the nominal probabilities of .95. This holds for each test and error distribution, even for values of γ far from the integrated case $\gamma=0$. Thus the local to zero asymptotics provide good small sample approximations for both integrated and stationary data. This occurs because under the null the test statistic $Q(\overline{\alpha})$ is stochastically bounded for both the integrated and stationary cases. In the stationary case $Q(\overline{\alpha})$ has a limiting standard normal null distribution, and the local to zero asymptotics lead to complicated representations of standard normal variables.

These methods do not work nearly as well when the dependence structure of the error is unknown. Table 4 gives empirical coverage probabilities of the various confidence sets for several types of serial correlation:

$$\begin{array}{ll}\text{IID:} & v_t=\varepsilon_t\\ \text{AR:} & v_t=.3v_{t-1}+\varepsilon_t\\ \text{MA:} & v_t=\varepsilon_t-.3\varepsilon_{t-1}.\end{array}$$

The numbers are reasonably close to the nominal coverage rate of .95 so long as the true value of c is 0. For $c=-30$, the coverage rates are strongly biased downward. This is true of all the tests, and the effect is strongest for the P_T test. IID errors lead to the best coverage rates, with probabilities close to .9 in many cases. A possible

Table 4: Empirical coverage probablities

	N(0,1) errors			DE errors			Log Norm errors			Mixture errors		
Test	IID	AR	MA	IID	AR	MA	IID	AR	MA	IID	AR	MA
					True $c = 0$							
P_T	.94	.91	.91	.94	.92	.91	.95	.93	.92	.95	.93	.91
Least squares	.95	.94	.94	.95	.94	.93	.96	.95	.95	.94	.93	.93
LAD	.96	.95	.96	.96	.95	.95	.95	.95	.95	.97	.96	.95
Huber's M	.95	.94	.94	.96	.94	.94	.96	.95	.94	.96	.95	.93
Student's t3	.96	.94	.95	.95	.94	.94	.96	.95	.95	.96	.95	.94
Wilcoxon	.96	.95	.95	.96	.95	.94	.96	.95	.95	.96	.95	.95
Normal ranks	.96	.95	.94	.96	.95	.94	.96	.95	.94	.96	.95	.95
Sign ranks	.96	.96	.94	.95	.95	.94	.95	.95	.95	.96	.96	.96
					True $c = -30$							
P_T	.61	.31	.36	.61	.30	.34	.66	.26	.34	.71	.28	.32
Least squares	.88	.63	.74	.88	.63	.75	.87	.61	.74	.92	.58	.73
LAD	.90	.75	.82	.88	.57	.71	.86	.46	.62	.90	.48	.71
Huber's M	.88	.65	.76	.87	.54	.69	.84	.34	.51	.88	.33	.45
Student's t3	.88	.67	.78	.87	.52	.66	.83	.33	.51	.87	.32	.47
Wilcoxon	.88	.66	.77	.87	.53	.68	.81	.28	.50	.87	.36	.55
Normal ranks	.89	.64	.75	.88	.58	.71	.81	.29	.52	.88	.35	.52
Sign ranks	.91	.75	.82	.88	.57	.71	.84	.38	.59	.90	.49	.73

Notes: The model includes an intercept and time trend, with $(\mu_1, \mu_2) = (1, 1)$. There are 2000 Monte Carlo replications. Sample size is $n = 100$.

explanation for these results is that the estimator for $\Gamma(1)$ performs poorly at alternatives far from $c = 0$. Another explanation is that the asymptotic approximations work only when the largest root of the autoregressive polynomial for u_t dominates the other roots. As the alternative range from $c = 0$ to $c = -30$, the largest root shrinks relative to the other roots. While these results are discouraging, there may be ways to improve the nuisance parameter estimates or the asymptotic approximations.

5 Appendix

A Sketches of Proofs

To be completed....

References

Adichie, J. N. (1986), Rank tests in linear models, *in* P. Krishnaiah and P. Sen, eds, 'Handbook of Statistics', Vol. 4, Elsevier, New York.

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

Andrews, D. W. K. (1993), 'Exactly median-unbiased estimation of first order autoregressive/unit root models', *Econometrica* **61**(1), 139–165.

Chan, N. H. and Wei, C. Z. (1987), 'Asymptotic inference for nearly nonstationary ar(1) processes', *The Annals of Statistics* **15**(3), 1050–1063.

Dickey, D. and Fuller, W. (1979), 'Distribution of estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association* **74**(366), 427–431.

Elliott, G., Rothenberg, T. and Stock, J. H. (1996), 'Efficient tests for an autoregressive unit root', *Econometrica* **64**(4), 813–836.

Elliott, G. and Stock, J. H. (2000), 'Confidence intervals for autoregressive coefficients near one', *Journal of Econometrics* . forthcoming.

Gutenbrunner, C. and Jureckova, J. (1992), 'Regression rank scores and regression quantiles', *Annals of Statistics* **20**(1), 305–330.

Hansen, B. E. (1999), 'The grid bootstrap and the autoregressive model', *Review of Economics and Statistics* **81**, 594–607.

Hasan, M. N. and Koenker, R. W. (1997), 'Robust rank tests of the unit root hypothesis', *Econometrica* **65**(1), 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.

Koenker, R. and Bassett, G. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

Lucas, A. (1995), 'Unit root tests based on m estimators', *Econometric Theory* **11**, 331–346.

Ng, S. and Perron, P. (2000), 'Lag length selection and the construction of unit root tests with good size and power', *Working Paper* .

Phillips, P. C. B. (1987), 'Toward a unified asymptotic theory for autoregression', *Biometrika* **74**, 535–547.

Rothenberg, T. J. and Stock, J. H. (1997), 'Inference in a nearly integrated autoregressive model with nonnormal innovations', *Journal of Econometrics* **80**, 269–286.

Rothenberg, T. J. and Thompson, S. B. (2001), 'Robust tets in the linear model', *Working Paper* .

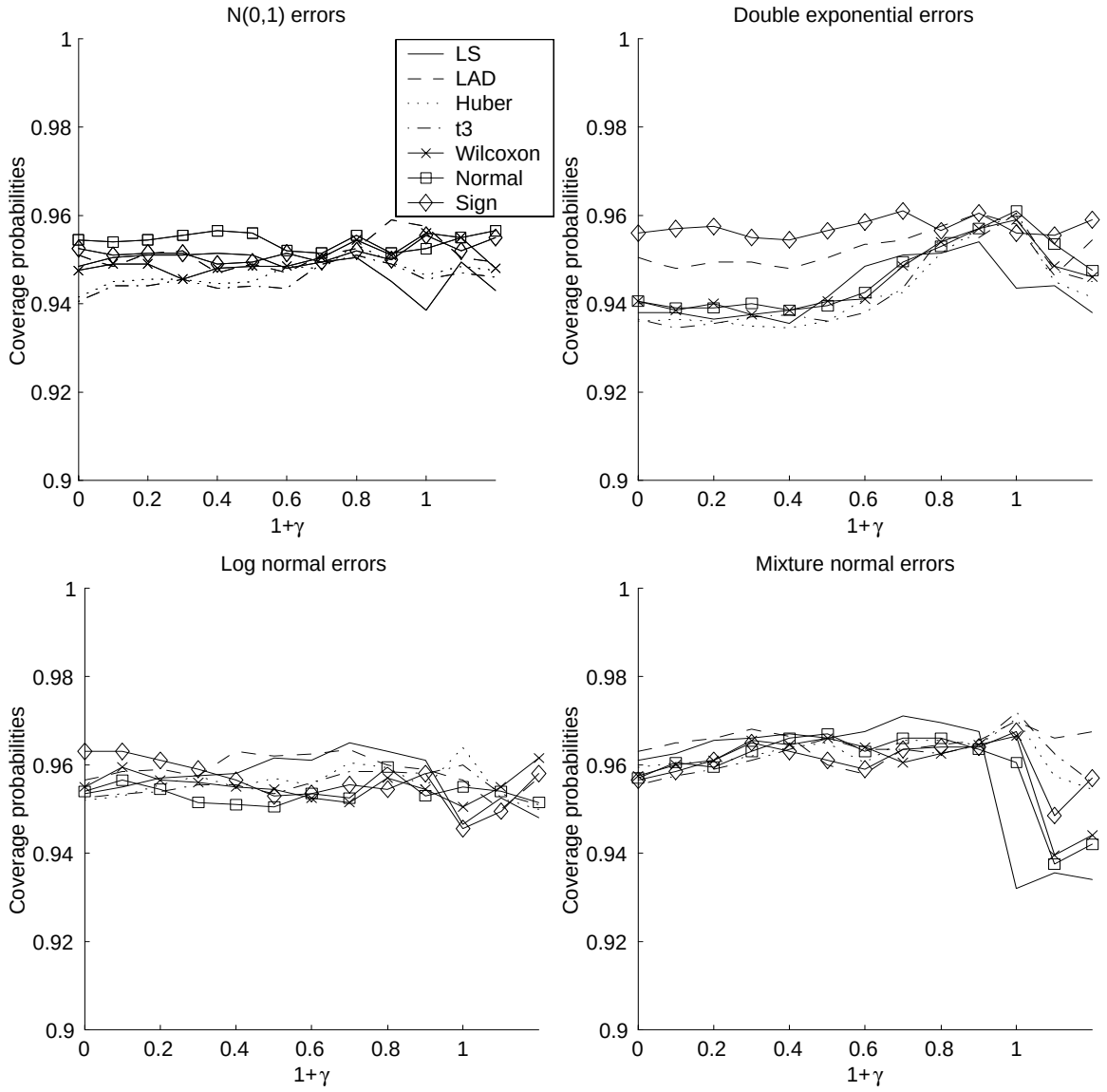
Seo, B. (1999), 'Distribution theory for unit root tests with conditional heteroskedasticity', *Journal of Econometrics* **91**, 113–144.

Stock, J. H. (1991), 'Confidence intervals for the largest autoregressive root in u.s. macroeconomic time series', *Journal of Monetary Economics* **28**, 435–459.

Thompson, S. B. (2001*a*), 'Optimal versus robust inference in nearly integrated non gaussian models', *Working Paper* .

Thompson, S. B. (2001*b*), 'Robust unit root testing with correct size and good power', *Working Paper* .

Figure 4: Empirical coverage probabilities when the researcher knows that v_t are iid.



Notes: The model includes an intercept and time trend, with $(\mu_1, \mu_2) = (1, 1)$. There are 2000 Monte Carlo replications. Sample size is $n = 100$.