NBER TECHNICAL WORKING PAPERS SERIES

EFFICIENT TESTS FOR AN AUTOREGRESSIVE UNIT ROOT

Graham Elliott

Thomas J. Rothenberg

James H. Stock

Technical Working Paper No. 130

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 1992

The authors thank Jushan Bai, Maxwell King, Sastry Pantula, Pierre Perron, and Mark Watson for helpful discussions. This research was supported in part by National Science Foundation grant no. SES-91-22463. This paper is part of NBER's research program in Economic Fluctuations. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #130
December 1992

EFFICIENT TESTS FOR AN AUTOREGRESSIVE UNIT ROOT

ABSTRACT

This paper derives the asymptotic power envelope for tests of a unit autoregressive root for various trend specifications and stationary Gaussian autoregressive disturbances. A family of tests is proposed, members of which are asymptotically similar under a general I(1) null (allowing nonnormality and general dependence) and which achieve the Gaussian power envelope. One of these tests, which is asymptotically point optimal at a power of 50%, is found (numerically) to be approximately uniformly most powerful (UMP) in the case of a constant deterministic term, and approximately uniformly most powerful invariant (UMPI) in the case of a linear trend, although strictly no UMP or UMPI test exists. We also examine a modification, suggested by the expression for the power envelope, of the Dickey-Fuller (1979) t-statistic; this test is also found to be approximately UMP (constant deterministic term case) and UMPI (time trend case). The power improvement of both new tests is large: in the demeaned case, the Pitman efficiency of the proposed tests relative to the standard Dickey-Fuller t-test is 1.9 at a power of 50%. A Monte Carlo experiment indicates that both proposed tests, particularly the modified Dickey-Fuller t-test, exhibit good power and small size distortions in finite samples with dependent errors.

Graham Elliott
Department of Economics
Harvard University
Cambridge, MA 02138

James H. Stock
Kennedy School of Government
Harvard University
Cambridge, MA 02138
and NBER

Thomas J. Rothenberg
Department of Economics
University of California
Berkeley, CA 94720

1. Introduction

There is now a large literature on the theory of tests for a unit autoregressive root in univariate time series. In their seminal work, Fuller (1976) and Dickey and Fuller (1979) proposed tests in autoregressive (AR) models of known finite order. These tests were extended to the more general case that the series is integrated of order one (is I(1)) under the null hypothesis and of order zero (is I(0)) under the alternative by Said and Dickey (1984) and, under more general conditions on the disturbances (and using a different approach), by Phillips (1987a) and Phillips and Perron (1988). These papers in turn spurred a profusion of research proposing alternative tests of the general I(1) null against the I(0) alternative.

A natural way to compare these tests is to examine their power in large samples using the local-to-unity asymptotic framework developed by Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), and Phillips (1987b). Nabeya and Tanaka (1990) provide numerical comparisons of the resulting asymptotic power functions for several leading tests (including the Dickey-Fuller (1979) sum-of-coefficients test and the Sargan-Bhargava (1983)/Bhargava (1986) test) and found large power differences among them. However, in the absence of results on asymptotically optimal tests of the unit root hypothesis, power comparisons so far have stressed relative rather than absolute performance. Without an absolute performance standard – an asymptotic power envelope – it is unknown whether there remains room for substantial power improvements over extant asymptotically valid unit root tests. This gap has practical significance: it is widely recognized by practitioners that one of the main limitations of unit root tests is their poor power against alternatives of empirical interest.

This paper undertakes two tasks. The first is to provide the asymptotic power envelope for tests of the hypothesis that the largest autoregressive root (α) equals one, against the alternative $|\alpha| < 1$. These are developed for the general case that the observed series y_t is the sum of a deterministic part, d_t, and a purely stochastic component u_t. The asymptotic power envelope is derived by considering the sequence of Neyman-Pearson tests of the unit root hypothesis against the local alternative $\bar{\alpha} = 1 + \bar{c}/T$ in the finite-sample Gaussian AR(p+1) model, in which all the nuisance parameters except α are assumed known. The envelope of the limiting power functions of this family of tests (indexed by $\bar{c}$) constitutes an asymptotic power bound in the Gaussian AR(p+1) model. It is shown that this bound can be achieved even if the innovation variance and $\{d_t\}$ are unknown, as long as the deterministic component is "slowly varying," a condition which is satisfied for example by a constant deterministic term and a constant with discrete finite breaks or shifts. Thus the Neyman-Pearson power bound is the power envelope in this broader model.

If the deterministic component is a polynomial trend of linear or higher order, this power bound cannot be achieved, so in this case we derive an asymptotic Gaussian power envelope by considering the sequence of most powerful invariant (MPI) tests. The starting point for this analysis is the family of exact MPI tests developed by Dufour and King (1991) in the Gaussian AR(1) model. An explicit asymptotic representation for the Gaussian power envelope is given in the case of a linear trend.

The second task is to develop new test statistics for practical use which achieve the Gaussian power envelope and which are asymptotically similar under a general I(1) null with I(0) disturbances which possess a general unknown dependence structure and which might be nonnormally distributed. These are developed by considering asymptotic versions of the MPI tests in the two leading cases of a constant and a linear trend, modified to be asymptotically similar. If y_t in fact obeys a Gaussian AR(p+1), these tests are asymptotically efficient in the sense that they achieve (at one point) the Gaussian power envelope. One specific test proposed here, based on a statistic P_T defined in section 3, is an asymptotic point optimal invariant (POI) test with power function tangent to the power envelope at a power of 50% and which is asymptotically similar under the general I(1) null.

The theoretical results indicate that a key feature of the POI statistics is their use of data which are detrended by GLS under a local alternative which is chosen by the researcher. This suggests reexamining currently available unit root tests, evaluated instead using the local GLS-demeaned (or local GLS-detrended) data. The specific statistic studied here is the Dickey-Fuller (1979) t-statistic with local GLS detrending, termed the DF-GLS statistic. As it happens, the power functions of both the P_T and the DF-GLS tests nearly fall on the power envelopes in both the demeaned and linearly detrended cases, so that for practical purposes both of the proposed tests can be thought of as approximately asymptotically UMP (demeaned case) and UMPI (linearly detrended case). The power gains from using these tests can be large, particularly in the demeaned case. This parallels the substantial power gains found by Saikkonen and Luukkonen (1992) in their study of asymptotically POI tests for a unit moving average root.

The outline of the paper is as follows. The Gaussian power envelope is derived in section 2. The P_T and DF-GLS tests are developed, and their asymptotic properties are characterized, in section 3. In section 4, the asymptotic Gaussian power envelope is evaluated numerically and is then used as a standard by which to judge the asymptotic power of the proposed P_T and DF-GLS statistics, as well as several leading test statistics previously proposed in the literature (the t-test and sum-of-coefficients test studied by Dickey and Fuller (1979), Said and Dickey (1984), and Phillips and Perron (1988) and the Sargan-Bhargava (1983) test, which is asymptotically equivalent to the symmetric least squares test of Dickey, Hasza and Fuller (1984)). The finite-sample performance of the P_T and DF-GLS tests in the demeaned and detrended cases when the disturbances have nontrivial short-run dependence is studied in Section 5 in a Monte Carlo experiment. Section 6 concludes.

2. The Asymptotic Gaussian Power Envelope

The time series y_t is assumed to have the representation,

(1) $$y_t = d_t + u_t, \quad t = 1, \ldots, T,$$

where $$u_t = \alpha u_{t-1} + v_t, \quad t = 1, \ldots, T$$

where $\{d_t\}$ are deterministic constants (which we refer to as the "trend" component) and v_t is I(0). The object is to test the null hypothesis that $\alpha = 1$ against the alternative $|\alpha|<1$. In this section, v_t is assumed to satisfy a p-th order Gaussian autoregression, so that u_t follows a Gaussian AR(p+1):

(2) Gaussian AR(p+1) model: $v_t = \rho_1 v_{t-1} + \cdots + \rho_p v_{t-p} + \epsilon_t$, ϵ_t i.i.d. $N(0,\sigma^2)$, $t = 0, \pm 1, \pm 2, \ldots$

where the roots of the polynomial $\rho(L) = 1-\sum_{i=1}^{p}\rho_i L^i$ are all outside the unit circle.

This section provides results on the Gaussian power envelope for unit root tests, under the assumptions that $\rho(L)$ is known and $u_0 = 0$. In Section 2A, the asymptotic power function of the Neyman-Pearson test, in which $\{d_t\}$ and σ are assumed to be known, is derived, and these results are extended to the case in which σ and $\{d_t\}$ are unknown but the trend is slowly-varying. Section 2B considers the case where $\{d_t\}$ is a linear time trend.

A. Known trends or unknown but slowly-varying trends

Suppose that $\rho(L)$, $\{d_t\}$, and σ are known. Then the log likelihood for the parameter α is, except for a constant,

$$(3)\qquad \mathcal{L}(\alpha) = A(\alpha) - \tfrac{1}{2}\sigma^{-2}\sum_{t=p+2}^{T}(u_t^{\dagger} - \alpha u_{t-1}^{\dagger})^2$$
$$= A(\alpha) - \tfrac{1}{2}\sigma^{-2}\sum_{t=p+2}^{T}\{\Delta u_t^{\dagger} + (1-\alpha)u_{t-1}^{\dagger}\}^2$$

where $A(\alpha)$ is the log likelihood for the first p+1 observations and, for $t = p+1, p+2, \ldots, T$, $u_t^{\dagger} = y_t^{\dagger} - d_t^{\dagger}$, where $y_t^{\dagger} = \rho(L)y_t$ and $d_t^{\dagger} = \rho(L)d_t$. By the Neyman-Pearson Lemma, the most powerful test of the hypothesis $\alpha = 1$ against the alternative $\alpha = \bar{\alpha}$ rejects for large values of the log likelihood ratio,

$$(4)\qquad \mathcal{L}(\bar{\alpha}) - \mathcal{L}(1) = A(\bar{\alpha}) - A(1) - \tfrac{1}{2}(1-\bar{\alpha})^2\sigma^{-2}\sum_{t=p+2}^{T}u_{t-1}^{\dagger 2} - (1-\bar{\alpha})\sigma^{-2}\sum_{t=p+2}^{T}u_{t-1}^{\dagger}\Delta u_t^{\dagger}.$$

In large samples, only values of $\bar{\alpha}$ close to unity are relevant since distant alternatives will be rejected with probability close to one. Thus we shall consider local alternatives, $\bar{\alpha} = 1+\bar{c}/T$, where $\bar{c}$ is a fixed constant. Furthermore, we shall study the behavior of our tests using the local-to-unity asymptotic nesting, developed and studied by Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), Phillips (1987a, 1987b), and Chan (1988), in which the true value of α is $\alpha = 1+c/T$, where c is a fixed constant. For fixed values of c and $\bar{c}$, $A(1)-A(\bar{\alpha})$ is $O_p(T^{-1})$ because A has a bounded continuous derivative at $\alpha = 1$. In addition, for all c in a compact set, $T^{-1}\sum_{t=p+2}^{T} u_{t-1}^{\dagger}\Delta u_t^{\dagger} = \frac{1}{2}(T^{-1}u_T^{\dagger 2} - T^{-1}u_{p+1}^{\dagger 2} - T^{-1}\sum_{t=p+2}^{T}(\Delta u_t^{\dagger})^2) = \frac{1}{2}T^{-1}u_T^{\dagger 2} - \frac{1}{2}\sigma^2 + o_p(1)$. Moreover, for all c in a compact set, $(T^{-2}\sum_{t=p+2}^{T} u_{t-1}^{\dagger 2}, T^{-1}u_T^{\dagger 2})$ has a nondegenerate limiting distribution. Thus, as $\bar{c}$ ranges over moderate negative values, we obtain a family of asymptotically point-optimal tests with critical regions of the form,

$$\bar{c}^2\sigma^{-2}T^{-2}\sum_{t=p+2}^{T} u_{t-1}^{\dagger 2} - \bar{c}\sigma^{-2}T^{-1}u_T^{\dagger 2} < b(\bar{c}) \tag{5}$$

where the contant $b(\bar{c})$ is determined by the condition that the tests have the desired size.

Because the asymptotic minimal sufficient statistic $(T^{-2}\sum_{t=p+2}^{T} u_{t-1}^{\dagger 2}, T^{-1}u_T^{\dagger 2})$ has dimension greater than the number of unknown parameters, there exists no UMP test even in large samples. This extends the observation by Anderson (1948), that in the Gaussian AR(1) model no exact UMP test of $\alpha=\alpha_0$ against $\alpha<\alpha_0$ exists, to the large-sample case with roots local to one.

The asymptotic size and power of the critical regions defined in (5) can be determined using the asymptotic results in the literature concerning autoregressive roots which are local to unity. Let $[\cdot]$ denote the greatest lesser integer function, so that $T^{-\frac{1}{2}}u_{[T\cdot]}^{\dagger}$ denotes the random function on the unit interval constructed as $T^{-\frac{1}{2}}u_{[Ts]}^{\dagger}$, $0 \le s \le 1$. Also let "=>" denote weak convergence on D[0,1]. Then $\sigma^{-1}T^{-\frac{1}{2}}u_{[T\cdot]}^{\dagger} \Rightarrow W_c(\cdot)$, where $W_c(s)$ is the diffusion process on [0,1] satisfying $dW_c(s) = cW(s)ds + dW(s)$, where W(s) is a standard Brownian motion (see Bobkoski (1983) and Phillips (1987b)). This result and the continuous mapping theorem imply that the asymptotic minimal sufficient statistic has the limiting representation, $(T^{-2}\sum_{t=p+2}^{T} u_{t-1}^{\dagger 2}, T^{-1}u_T^{\dagger 2}) \Rightarrow$

$\sigma^2(\int W_c^2, W_c(1)^2)$. (Throughout we adopt the notational convention that $\int W_c^2$ denotes $\int_{s=0}^1 W_c(s)^2 ds$, etc.) Thus the asymptotic probability of rejecting $\alpha=1$ using the point-optimal rejection region (5) when in fact $\alpha = 1+c/T$ is,

$$(6) \qquad \pi(c,\bar{c}) = \Pr[\bar{c}^2 \int W_c^2 - \bar{c} W_c(1)^2 < b(\bar{c})] .$$

Because the rejection regions defined by (5) are asymptotically equivalent to the rejection regions of the Neyman-Pearson test, the asymptotic Gaussian power envelope $\Pi(c) = \pi(c,c)$ is an upper bound for the local asymptotic power function for any one-sided test for $\alpha = 1$.

In practice it will rarely be the case that the trend $\{d_t\}$ or the scale parameter σ are known, in which case $u_t^\dagger/\sigma = (y_t^\dagger - d_t^\dagger)/\sigma$ is not observable and the critical regions defined in (5) are not feasible. However, if the trend is slowly-varying, then it is possible to construct simple critical regions which are asymptotically equivalent to those in (5) but which do not require knowledge of $\{d_t\}$ or σ. Specifically, suppose that the trend satisfies,

Condition A (slowly-varying trend).
$T^{-1}\sum_{t=1}^T (\Delta d_t)^2 \to 0$ and $T^{-\frac{1}{2}}\max_{t=1,\ldots,T} |d_t| \to 0$.

Define the rejection region,

$$(7) \qquad \{\bar{c}^2 T^{-2} \sum_{t=p+2}^T y_{t-1}^{\dagger 2} - \bar{c} T^{-1} y_T^{\dagger 2}\}/\hat{\sigma}^{\dagger 2} < b(\bar{c})$$

where $\hat{\sigma}^{\dagger 2} = T^{-1}\sum_{t=p+2}^T (\Delta y_t^\dagger)^2$. Note that (7) can be constructed without knowledge of $\{d_t\}$ or σ. Then the critical regions defined in (7) are asymptotically equivalent to those in (5), that is,

$$(8) \qquad \Pr[\{\bar{c}^2 T^{-2} \sum_{t=p+2}^T y_{t-1}^{\dagger 2} - \bar{c} T^{-1} y_T^{\dagger 2}\}/\hat{\sigma}^{\dagger 2} < b(\bar{c})] - \Pr[\{\bar{c}^2 T^{-2} \sum_{t=p+2}^T u_{t-1}^{\dagger 2} - \bar{c} T^{-1} u_T^{\dagger 2}\}/\sigma^2 < b(\bar{c})] \to 0$$

where the convergence is uniform for $(c, \bar{c})$ in a compact set.

Although the finite-sample distributions of the tests with rejection regions (7) depend on $\{d_t\}$, the asymptotic distributions do not. Thus, under the assumption of Gaussian errors in the AR(p+1) model with $\rho(L)$ known, the family of critical regions (7) constitute an essentially complete class of asymptotically similar and efficient tests of the unit root hypothesis. That is, each member of the family has an asymptotic local power function that is tangent to the (Neyman-Pearson) power envelope $\Pi(c)$ and every point on the power envelope can be attained by a member of the family.

These results are collected in the following theorem.

Theorem 1. Suppose that y_t is generated by the Gaussian AR(p+1) model (1) and (2) with $u_0 = 0$ and with $\rho(L)$ known. Then, under the local-to-unity asymptotics where $c = T(\alpha-1)$ and $\bar{c} = T(\bar{\alpha}-1)$ are fixed as T tends to infinity,

(a) If σ and $\{d_t\}$ are known, then the most powerful test of $\alpha = 1$ vs. $\alpha = \bar{\alpha}$ has local asymptotic power function $\pi(c,\bar{c}) = \Pr[\Psi(c,\bar{c}) < b(\bar{c})]$ where $\Psi(c,\bar{c}) = \bar{c}^2 \int W_c^2 - \bar{c} W_c(1)^2$ and $b(\bar{c})$ is a constant. An upper bound for the local asymptotic power function for any one-sided test of $\alpha = 1$ is given by the power envelope $\Pi(c) = \pi(c,c)$.

(b) If σ and $\{d_t\}$ are unknown but $0<\sigma<\infty$ and $\{d_t\}$ satisfies condition A, then the local asymptotic power envelope for efficient one-sided tests remains $\Pi(c)$. Tests based on the family of feasible regions (7) have local asymptotic power functions $\pi(c,\bar{c})$.

All proofs are given in the appendix.

The results in theorem 1 are valid for c and $\bar{c}$ being finite constants, so they cover tests for locally explosive alternatives as well as locally stationary ones. The focus of the paper, however, is on the locally stationary case $(c, \bar{c} < 0)$.

In addition to the case in which d_t is a constant, condition A is satisfied by a variety of slowly-varying or bounded trends. These include slowly-varying sinusoids (e.g. $d_t = \cos(2\pi kt/T)$ for finite k); slowly increasing time trends (e.g. $d_t = t^\gamma$, $\gamma<\frac{1}{2}$); and a constant with finitely many shifts (e.g. $d_t = \beta_0 + \beta_1 1(t>\tau T)$, where $\tau \in (0,1)$ and $1(\cdot)$ is the indicator function). For each of these types of trends, the power envelope is $\Pi(c) = \pi(c,c)$.

While the assumption $u_0=0$ was made here for simplicity, we note that this condition is stronger than is needed in either the no-deterministic or slowly-varying trend cases. This assumption was used only to argue that $A(\bar{\alpha}) - A(1)$ in (4) is $o_p(1)$ for c in a compact set. However, this holds under weaker conditions on u_0, for example that u_0 is a fixed finite constant. In this sense, the power envelope in theorem 1 applies under more general conditions on u_0.

B. The power envelope with a polynomial time trend

The construction of asymptotically optimal tests of $\alpha = 1$ when d_t is unknown and is not slowly-varying (does not satisfy condition A) is more complicated. Consider, for example, the case of a polynomial time trend,

$$(9) \qquad d_t = \beta' z_t, \quad z_t = (1, t, \ldots, t^k)$$

where β is a k+1 dimensional vector of unknown parameters. Although the critical region (5) still has the local power function $\pi(c,\bar{c})$, the feasible region (7) does not; the statistic $(T^{-2}\sum_{t=p+2}^{T} y_{t-1}^{\dagger 2}, T^{-1}y_T^{\dagger 2})$ does not have the limiting distribution of $(T^{-2}\sum_{t=p+2}^{T} u_{t-1}^{\dagger 2}, T^{-1}u_T^{\dagger 2})$. Indeed, it is not possible to construct feasible tests that, for all β, attain the power bound $\Pi(c)$ in the polynomial trend case for $k\geq 1$.

We therefore restrict attention to a natural class of tests for this problem, those which are invariant to the nuisance parameters β and σ. The construction of finite sample POI tests for general values of an autoregressive root has been studied in detail by Dufour and King (1991), who build on the general results for POI tests in King (1980, 1988). As in the argument leading to (5), ignore the terms involving the asymptotically negligible initial observations. Then the Gaussian AR(p+1) model can be rewritten,

(10) $y_t^\dagger = \beta' z_t^\dagger + u_t^\dagger,\ u_t^\dagger = \alpha u_{t-1}^\dagger + \epsilon_t,\quad t = p+1, p+2, \ldots, T$

where ϵ_t is i.i.d. $N(0,\sigma^2)$ and $z_t^\dagger = \rho(L)z_t$. The assumption $u_0 = 0$ implies that u_{p+1} is distributed $N(0, \kappa\sigma^2)$, where κ is a positive finite constant which depends on $\{\rho_i, \alpha\}$. Dufour and King's (1991) results apply to the model (10) with p=0 and κ a positive constant not depending on α. Because $\alpha = 1+c/T$ and κ has a bounded continuous derivative with respect to α at $\alpha = 1$, the dependence of κ on α in (10) is asymptotically negligible and we can draw on their results to construct asymptotic POI rejection regions for the Gaussian AR(p+1) model with trends of the form (9).

As Dufour and King (1991) discuss, the problem of testing $\alpha = 1$ against the point alternative $\alpha = \bar{\alpha}$ when y_t obeys (10) is invariant to transformations of the form, $y_t^\dagger \to a y_t^\dagger + b' z_t^\dagger$ (t = p+1, . . . , T), where a is a positive scalar and b is a finite (k+1)-vector. Among tests which are invariant to this transformation, the exact MPI test rejects when the sum of squared residuals from the GLS regression under the alternative is small relative to the sum of squared residuals from the GLS regression under the null. That is, after dropping the asymptotically negligible terms involving initial conditions, the MPI test rejects for small values of

(11) $$M_T^\dagger = \textstyle\sum_{t=p+1}^{T}(\tilde{y}_t^\dagger - \tilde{\beta}^{\dagger\prime}\tilde{z}_t^\dagger)^2 / \sum_{t=p+1}^{T}(\hat{y}_t^\dagger - \hat{\beta}^{\dagger\prime}\hat{z}_t^\dagger)^2,$$

where $\tilde{y}_t^\dagger$ and $\tilde{z}_t^\dagger$ are the GLS-transforms of $y_t^\dagger$ and $z_t^\dagger$, respectively, under the local alternative, and $\hat{y}_t^\dagger$ and $\hat{z}_t^\dagger$ are the GLS-transforms of $y_t^\dagger$ and $z_t^\dagger$ under the null (that is, $\tilde{y}_{p+1}^\dagger = y_{p+1}^\dagger$ and, for $t = p+2, \ldots, T$, $\tilde{y}_t^\dagger = y_t^\dagger - \bar{\alpha} y_{t-1}^\dagger$ and similarly for $\tilde{z}_t^\dagger$, and $\hat{y}_{p+1}^\dagger = y_{p+1}^\dagger$ and, for $t = p+2, \ldots, T$, $\hat{y}_t^\dagger = y_t^\dagger - y_{t-1}^\dagger$ and similarly for $\hat{z}_t^\dagger$), and where $\tilde{\beta}^\dagger = (\sum_{t=p+1}^{T}\tilde{z}_t^\dagger\tilde{z}_t^{\dagger\prime})^{-1}\sum_{t=p+1}^{T}\tilde{z}_t^\dagger\tilde{y}_t^\dagger$ and $\hat{\beta}^\dagger = (\sum_{t=p+1}^{T}\hat{z}_t^\dagger\hat{z}_t^{\dagger\prime})^{-1}\sum_{t=p+1}^{T}\hat{z}_t^\dagger\hat{y}_t^\dagger$. Rejection regions based on (11) can be constructed without knowledge of β or σ, but require knowledge of $\rho(L)$.

The two leading cases of the polynomial time trend are a constant-mean (k=0) and a linear time trend (k=1). When k=0 and $\alpha = 1+c/T$, rejection regions based on (11) are asymptotically equivalent to those based on (5), so the POI test in the k=0 case achieves the power bound in theorem 1. (This is a consequence of theorem 3 in the next section.) In the case of a linear trend, the estimation of β_1 is not negligible asymptotically, and a different local power function is obtained. This result is summarized in the next theorem.

Theorem 2. Suppose that y_t is generated by the Gaussian AR(p+1) model (1) and (2) with $u_0 = 0$ and with $\rho(L)$ known. Let $M_T^\dagger$ defined in (11) be the MPI statistic for testing $\alpha = 1$ vs. $\alpha = \bar{\alpha}$ when $d_t = \beta_0 + \beta_1 t$. Then, under local-to-unity asymptotics where $c = T(\alpha-1)$ and $\bar{c} = T(\bar{\alpha}-1)$ are fixed as T tends to infinity, $T(M_T^\dagger - \bar{\alpha})$ converges weakly to the random variable,

$$\Psi^c(c,\bar{c}) = \bar{c}^2 \int V_c^2 + (1-\bar{c})V_c(1)^2$$
$$= \bar{c}^2\{\lambda\int[W_c(s)-sW_c(1)]^2 ds + (1-\lambda)\int[W_c(s)-3s\int rW_c(r)dr]^2 ds\}$$

where $\lambda = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$ and $V_c(s) = W_c(s) - s\{\lambda W_c(1) + (1-\lambda)3\int rW_c(r)dr\}$. The local asymptotic power function is $\pi^\tau(c,\bar{c}) = \Pr[\Psi^\tau(c,\bar{c}) < b^\tau(\bar{c})]$ where $b^\tau(\bar{c})$ is a constant. The envelope $\Pi^\tau(c) = \pi^\tau(c,c)$ is an upper bound for the local asymptotic power function for any one-sided invariant test of $\alpha = 1$ in the presence of a linear trend.

3. Efficient and Nearly-Efficient Tests for a Unit Root

If $\rho(L)$ is unknown, the rejection regions in section 2 are infeasible, since the transformed series $y_t^\dagger = \rho(L)y_t$ and $z_t^\dagger = \rho(L)z_t$ cannot be constructed. More generally, in empirical work there is typically no reason to think that v_t obeys a Gaussian AR(p). Rather, the null and alternative hypotheses of interest are $\alpha=1$ vs. $|\alpha| < 1$, where the maintained hypothesis is that v_t is a general I(0) process.

The operational definition of this maintained hypothesis that v_t is a general I(0) process used henceforth is that the scaled partial-sum process constructed using v_t, say $\nu_T(\cdot) = T^{-\frac{1}{2}}\sum_{s=1}^{[T\cdot]}v_s$, converges to a positive constant times a standard Brownian motion. This will be true under much weaker conditions than those employed in section 2. For the remainder of the paper we therefore assume this convergence, and in addition assume that v_t is covariance stationary with consistent sample autocovariances. (Covariance stationarity is assumed for convenience and could be weakened, cf. Phillips [1987a,b].) Specifically, we henceforth assume that y_t is generated by equation (1) and that v_t satisfies,

Condition B.

(a) $\hat{\gamma}_v(j) \xrightarrow{p} \gamma_v(j)$ for finite fixed j, where $\gamma_v(j) = Ev_tv_{t-j}$ and $\hat{\gamma}_v(j) = T^{-1}\sum_{t=j+1}^{T}v_tv_{t-j}$.

(b) $\nu_T \Rightarrow \omega W$, where W is a standard Brownian motion on [0,1] and ω is a constant with $0 < \omega < \infty$ and $\omega^2 = \sum_{j=-\infty}^{\infty}\gamma_v(j)$.

We also drop the assumption that u_0=0 and instead assume simply that the initial disturbance u_1 has a finite second moment:

Condition C. $Eu_1^2 < \infty$.

This includes includes the special case that u_1 is fixed.[1]

Section 3A considers the construction of tests which have local power functions $\pi(c,\bar{c})$ and, in the detrended case, $\pi^{\tau}(c,\bar{c})$, when v_t obeys condition B. If v_t obeys a Gaussian autoregression with finite but possibly unknown order p, these tests are asymptotically efficient. Section 3B considers the construction of alternative test statistics which seem likely to have asymptotic properties comparable to the efficient tests but which arguably will have better finite-sample performance.

These tests are constructed for the two leading cases of polynomial time trends, $d_t = \beta_0$ and $d_t = \beta_0+\beta_1 t$. A natural requirement for tests of the general I(1) null is that they be invariant to transformations of the form $y_t \rightarrow ay_t + b'z_t$. For example, it is natural to apply unit root tests to the logarithms of the original data; then invariance to a change in the units of measuring the original data is equivalent to invariance to an additive constant in their logarithms. To our knowledge all unit root tests which are used in practice have this invariance property. This section therefore considers only finite-sample invariant tests.

A. Asymptotically POI tests of the general I(1) null

The asymptotically POI tests are constructed in two steps. First, we consider the behavior of Dufour and King's (1991) POI tests of the I(1) null, (incorrectly) constructed under the assumption that v_t is i.i.d. $N(0,\sigma^2)$ and $u_0=0$, when in fact v_t is a general I(0) process and u_1 satisfies condition C. In general these tests will depend on nuisance parameters describing the v_t process, and they will not be asymptotically similar. However, the asymptotic representations of these statistics suggest a natural correction. Second, we provide statistics based on this correction and show that they are asymptotically similar, asymptotically efficient, and consistent against fixed alternatives.

In the special case that $v_t=\epsilon_t$, ϵ_t i.i.d. $N(0,\sigma^2)$, the POI tests (Dufour and King, 1991, theorem 5) are based on the statistics,

(12) $$M_T = \bar{\sigma}^2/\hat{\sigma}^2, \text{ where } \hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\hat{e}_t^2 \text{ and } \bar{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\bar{e}_t^2$$

where $\hat{e}_t$ and $\bar{e}_t$ are the residuals from the GLS regression of y_t onto z_t under $\alpha = 1$ and $\alpha = \bar{\alpha}$, respectively:

(13) $$\hat{e} = \hat{y}_t - \hat{\beta}'\hat{z}_t \quad \text{where } \hat{\beta} = (\sum_{t=1}^{T}\hat{z}_t\hat{z}_t')^{-1}(\sum_{t=1}^{T}\hat{z}_t\hat{y}_t)$$

where $(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_T) = (y_1, \Delta y_2, \ldots, \Delta y_T)$ and $(\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_T) = (z_1, \Delta z_2, \ldots, \Delta z_T)$, and

(14) $\tilde{e} = \tilde{y}_t - \beta'\tilde{z}_t$ where $\beta = (\sum_{t=1}^{T}\tilde{z}_t\tilde{z}_t')^{-1}(\sum_{t=1}^{T}\tilde{z}_t\tilde{y}_t)$

where $(\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_T) = (y_1, (1-\bar{\alpha}L)y_2, \ldots, (1-\bar{\alpha}L)y_T)'$ and $(\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_T) = (z_1, (1-\bar{\alpha}L)z_2, \ldots, (1-\bar{\alpha}L)z_T)'$.[2]

To avoid confusion, henceforth superscripts "μ" and "τ" are used to denote the $z_t = 1$ and $z_t = (1, t)$ cases, respectively. Thus M_T^{μ} and M_T^{τ} respectively refer to (12) evaluated in the demeaned and linearly detrended cases. Similarly, define the GLS-detrended processes $y_t^{\mu} = y_t - \beta_0$ and $y_t^{\tau} = y_t - \beta_0 - \beta_1 t$.

Under the general I(0) null, if $\alpha = 1+c/T$ then $(\omega^{-1}T^{-\frac{1}{2}}\sum_{t=1}^{[T\cdot]}v_t, \omega^{-1}T^{-\frac{1}{2}}u_{[T\cdot]}) \Rightarrow (W(\cdot), W_c(\cdot))$, where W_c is the diffusion defined in section 2, a result proven by Bobkoski (1983) for v_t i.i.d. and by Phillips (1987b) for v_t satisfying mixing conditions. Theorem 3 provides asymptotic representations for the M_T statistics under the general I(1) null and local alternatives.

Theorem 3. Suppose $\{v_t\}$ satisfies condition B, u_1 satisfies condition C, $\bar{\alpha} = 1+\bar{c}/T$, and $\alpha = 1+c/T$, where $\bar{c}$ and c are finite constants.

(a) In the no-deterministic case, $\hat{\sigma}^2 \xrightarrow{p} \gamma_v(0)$ and $T(M_T-\bar{\alpha}) \Rightarrow (\omega^2/\gamma_v(0))\Psi(c,\bar{c})$, where $\Psi(c,\bar{c})$ is defined in theorem 1.

(b) (demeaned case) If k = 0, then $\hat{\sigma}^2 \xrightarrow{p} \gamma_v(0)$ and $T(M_T^{\mu}-\bar{\alpha}) \Rightarrow (\omega^2/\gamma_v(0))\Psi^{\mu}(c,\bar{c})$, where $\Psi^{\mu}(c,\bar{c}) = \Psi(c,\bar{c})$.

(c) (linearly detrended case) If k = 1, then $\hat{\sigma}^2 \xrightarrow{p} \gamma_v(0)$ and $T(M_T^{\tau}-\bar{\alpha}) \Rightarrow (\omega^2/\gamma_v(0))\Psi^{\tau}(c,\bar{c})$, where $\Psi^{\tau}(c,\bar{c})$ is defined in theorem 2.

If $\omega^2=\gamma_v(0)$, then the asymptotic representations in theorem 3 are equivalent to those in theorems 1 and 2. However, in general the limiting representations, and thus the critical values for the M_T tests, depend on $\omega^2/\gamma_v(0)$, so that these tests are not similar asymptotically. The representations in theorem 3 nevertheless provide a framework for constructing a test which is asymptotically equivalent to the POI test but is asymptotically similar under the general I(1) null. Let $\hat{\omega}^2$ be a consistent estimator of ω^2, in the sense that $\hat{\omega}^2$ satisfies:

Condition D.

(a) If $\alpha = 1 + c/T$ where c is a finite constant, then $\hat{\omega}^2 \xrightarrow{p} \omega^2$.

(b) If α is fixed and $|\alpha| < 1$, then $\hat{\omega}^2 \xrightarrow{p} d$, where d, $0 < d < \infty$, is some constant.

Two families of estimators satisfying condition D – sum-of-covariance estimators and autoregressive spectral estimators – are discussed in section 5.

The proposed test statistics are,

(15a) $$P_T = (\sum_{t=1}^{T}\tilde{y}_t^2 - \bar{\alpha}\sum_{t=1}^{T}\hat{y}_t^2)/\hat{\omega}^2$$

(15b) $$P_T^{\mu} = (\sum_{t=1}^{T}\tilde{e}_t^2 - \bar{\alpha}\sum_{t=1}^{T}\hat{e}_t^2)/\hat{\omega}^2 \quad \text{where } z_t = 1 \text{ in (13) and (14)}$$

(15c) $$P_T^{\tau} = (\sum_{t=1}^{T}\tilde{e}_t^2 - \bar{\alpha}\sum_{t=1}^{T}\hat{e}_t^2)/\hat{\omega}^2 \quad \text{where } z_t = (1, t) \text{ in (13) and (14)}$$

where $\hat{y}_t$ and $\tilde{y}_t$ are defined following (13) and (14). The statistics P_T, P_T^{μ} and P_T^{τ} differ from $T(M_T - \bar{\alpha})$, $T(M_T^{\mu} - \bar{\alpha})$, and $T(M_T^{\tau} - \bar{\alpha})$ only by an estimated scale factor; that is, $P_T = (\hat{\sigma}^2/\hat{\omega}^2)T(M_T - \bar{\alpha})$, etc. This modification corrects for the factor $\omega^2/\gamma_v(0)$ in theorem 3. It follows from Theorem 3 and condition D that, if $\alpha = 1+c/T$ where c is a fixed finite constant, the test statistics in (15) have the asymptotic representation,

(16a) $$P_T \Rightarrow \Psi(c,\bar{c})$$

(16b) $$P_T^{\mu} \Rightarrow \Psi(c,\bar{c})$$

(16c) $$P_T^{\tau} \Rightarrow \Psi^{\tau}(c,\bar{c})$$

where Ψ and Ψ^{τ} are given in theorems 1 and 2, respectively. These tests are asymptotically similar, and their null representations are obtained by setting c=0 in (16).[3]

Tests based on P_T and P_T^{μ} have power function $\pi(c,\bar{c})$. This is the power function in theorem 1, and it follows that, if v_t in fact obeys a Gaussian AR(p), then the tests are as efficient asymptotically as the Neyman-Pearson test. If $k = 1$, the test has power function $\pi^{\tau}(c,\bar{c})$ and thus achieves the Gaussian power envelope for invariant tests in theorem 2.

An implication of (16) is that the Gaussian power envelope forms a lower bound for the power envelopes of unit root tests when the errors are nonnormally distributed but nonetheless obey condition B. To be concrete, suppose $u_t = \alpha u_{t-1} + \epsilon_t$, $u_0 = 0$, and ϵ_t ~ i.i.d. F_ϵ. If F_ϵ satisfies the Lindeberg condition, then the partial sum process constructed from ϵ_t obeys an invariance principle (e.g. Hall and Heyde (1980), theorem 4.1) and condition B(b) is satisfied with $\omega^2 = \gamma_v(0)$. If condition B(a) is also satisfied, then for this process a test based on M_T will be asymptotically valid and will achieve the asymptotic Gaussian power envelope. It follows that the asymptotic power envelope constructed for the true distribution of ϵ_t cannot lie below the asymptotic Gaussian power envelope.

Although point-optimal invariant tests are not always consistent against fixed alternatives, the proposed tests are. This follows from the next theorem.

Theorem 4. Suppose that conditions B-D hold and that α takes on a fixed value, say α_1, where $|\alpha_1| < 1$. Then $P_T \xrightarrow{p} 0$, $P_T^{\mu} \xrightarrow{p} 0$, and $P_T^{\tau} \xrightarrow{p} 0$.

The consistency of the tests follows from noting that $\Psi(0,\bar{c})$ and $\Psi^{\tau}(0,\bar{c})$ are $O_p(1)$ and are positive with probability one so that the critical values of P_T, P_T^{μ}, and P_T^{τ} are positive.

B. Alternative tests based on GLS detrending

A key feature of the invariant tests is that the trend parameter vector β is estimated by GLS under the local alternative. In the demeaned case, this estimator is $\tilde{\beta}_0 = (y_1 + (1-\bar{\alpha})\sum_{t=2}^{T}(1-\bar{\alpha}L)y_t)/(1 + (T-1)(1-\bar{\alpha})^2)$; because $T(1-\bar{\alpha}) = -\bar{c}$, large weight is placed on the initial observation. A straightforward calculation reveals that, under the null and local alternatives, $T^{-\frac{1}{2}}(\tilde{\beta}_0-\beta_0) \xrightarrow{p} 0$, so the effect of this GLS-detrending is asymptotically negligible on the detrended process (that is, $\max_t T^{-\frac{1}{2}}|y_t^{\mu} - u_t| \xrightarrow{p} 0$, which is an implication of appendix lemma

A2(a)). In contrast, previously proposed unit root tests in the k=0 case, which are necessarily less efficient than P_T^{μ} involve estimators of β_0 such as the sample mean which are not asymptotically negligible. This suggests that the performance of existing unit root tests statistics could be improved by evaluating them using the locally GLS-detrended series y_t^{μ} or y_t^{τ} as appropriate.

Because of its relatively good size properties in Monte Carlo studies (e.g. Schwert (1989)), a natural statistic to modify in this way is the Dickey-Fuller (1979) t-statistic. In the demeaned case the test statistic, which we term the DF-GLS$^{\mu}$ statistic, is the standard t-statistic testing $a_0 = 0$ in the regression,

$$\Delta y_t^{\mu} = a_0 y_{t-1}^{\mu} + \sum_{j=1}^{p} a_j \Delta y_{t-j}^{\mu} + \text{error} \tag{17}$$

where no intercept is included because the data are already demeaned by GLS under the local alternative $\bar{c}$. In the detrended case the statistic, termed the DF-GLS$^{\tau}$ statistic, is the same except that y_t^{μ} is replaced by the locally GLS-detrended series y_t^{τ}.

The local-to-unity asymptotic representations of the DF t-statistic has been derived by Bobkoski (1983), Cavanagh (1985), and Chan and Wei (1987) for p = 0 when the regression is correctly specified (that is, u_t-αu_{t-1} is i.i.d. or a martingale difference sequence); by Phillips (1987b) and Phillips and Perron (1988) for the case that v_t is a general I(0) process but p=0 (so the regression is misspecified); and by Stock (1991) for finite fixed p in the correctly specified case (v_t is an AR(q), $q \leq p$). The results in Stock (1991, appendix A) can be extended to the DF-GLS statistics, with the modification that the GLS-detrended processes have the limits given in our lemma A2. The limiting representations are,

$$\text{DF-GLS}^{\mu} \Rightarrow (\textstyle\int W_c^2)^{½}\{(\int W_c^2)^{-1}\int W_c dW + c\} \tag{18a}$$

$$\text{DF-GLS}^{\tau} \Rightarrow (\textstyle\int V_c^2)^{½}\{(\int V_c^2)^{-1}\int V_c dW + c\} \tag{18b}$$

where V_c is defined in theorem 2.

In the demeaned case (18a), the DF-GLS$^{\mu}$ statistic has the same null and local-to-unity asymptotic representation as the conventional Dickey-Fuller t-statistic in the no-deterministic case or, equivalently, the case that β_0 is known *a-priori.* Thus, as long as $\bar{c}$ is a finite constant, neither the null distribution nor the asymptotic power of the DF-GLS$^{\mu}$ statistic depend on the value of $\bar{c}$ used in the GLS detrending. In the detrended case (18b), the null and local-to-unity asymptotic distributions depend on the detrending value $\bar{c}$, and asymptotic critical values for the DF-GLS$^{\tau}$ statistic must be computed numerically.

4. Critical Values and Asymptotic Power Functions

This section presents numerical results on the asymptotic power envelope, the asymptotic power functions of several leading test statistics in the literature, and the asymptotic power functions of the new P_T and DF-GLS statistics. All tests have asymptotic level 5%. The computations are based on the limiting representations given in theorem 1, theorem 2, (18a), and (18b), in which the limiting functionals of W_c are replaced with discretized realizations generated by Monte Carlo simulation of a Gaussian AR(1) with $\alpha = 1+c/T$ and T=500, using 20,000 Monte Carlo replications.[4]

The resulting power envelope is used to provide an absolute standard by which to assess the performance of three leading tests for a unit root. The first is the Dickey-Fuller sum-of-coefficients test (the $\hat{\rho}$ test). In the AR(1) case, the statistic is $T(\hat{\alpha}-1)$, where $\hat{\alpha}$ is the OLS estimator of α in the regression of y_t onto y_{t-1}, $t = 2, \ldots, T$. This statistic was extended to the general I(1) null by Phillips (1987a) in the no-deterministic case (the Z_α statistic) and by Phillips and Perron (1988) in the constant and linearly detrended cases, using an additive correction in which ω^2 is estimated using a weighted sum-of-covariances (SC) spectral estimator. The results of Said and Dickey (1984) can be used to justify an alternative extension based on a multiplicative correction using a sum of coefficients in an AR(p), the order of which increases with the sample size. The second statistic is the Dickey-Fuller t-statistic ($\hat{\tau}$), which was extended to the general I(1) null by

Said and Dickey (1984) and Phillips (1987a)/Phillips and Perron (1988). The third statistic is the Sargan-Bhargava (1983) statistic in the no-deterministic and demeaned cases and Bhargava's (1986) extension of this statistic to the linearly detrended case (the SB statistic). The SB statistic was extended to the general I(1) null by Stock (1988) using an AR spectral estimator and by Schmidt and Phillips (1992) using a SC spectral estimator. Each of these statistics has an asymptotic representation under the local alternative in terms of functionals of W_c, which are here used to compute the local power functions using the same method as is used to compute the asymptotic power envelope.

Figure 1 presents the Gaussian power envelope in the no-deterministic case. Also presented are the power functions of the $\hat{\rho}$, $\hat{\tau}$, and SB tests and the locally most powerful invariant (LMPI) statistic in the Gaussian AR(1) model, which rejects for small values of $(T^{-\frac{1}{2}}y_T)^2/\hat{\sigma}^2$. With the exception of the LMPI test, which has very low power for all but the the smallest values of c, all the tests have effectively the same asymptotic powers and each lies very close to the power envelope. Consequently we focus on the cases with deterministic terms, which are in any event of greater practical interest.

The Gaussian power envelope and the power functions of the $\hat{\rho}$, $\hat{\tau}$, and SB tests are plotted in figures 2 and 3 for the demeaned and detrended cases, respectively. These figures also respectively plot the asymptotic power functions of the P_T^μ and P_T^τ statistics, where $\bar{c}$ was chosen so that the power functions are tangent to the power envelope at a power of 50%. Numerical investigations indicate that this tangency occurs at approximately $\bar{c} = -7$ in the demeaned case and $\bar{c} = -13.5$ in the detrended case. These statistics will respectively be referred to as the $P_T^\mu(.5)$ and $P_T^\tau(.5)$ statistics.[5] Also plotted are the asymptotic power functions of the DF-GLS$^\mu$ and DF-GLS$^\tau$ statistics.

In contrast to the results for the no-deterministic case, the power functions of the various tests differ widely. A useful way to summarize the power differences is in terms of the Pitman or asymptotic relative efficiencies (ARE's) of the tests. Although the limiting distributions are nonstandard so the ARE's cannot be computed as a ratio of variances as is done when $\sqrt{T}$-Gaussian

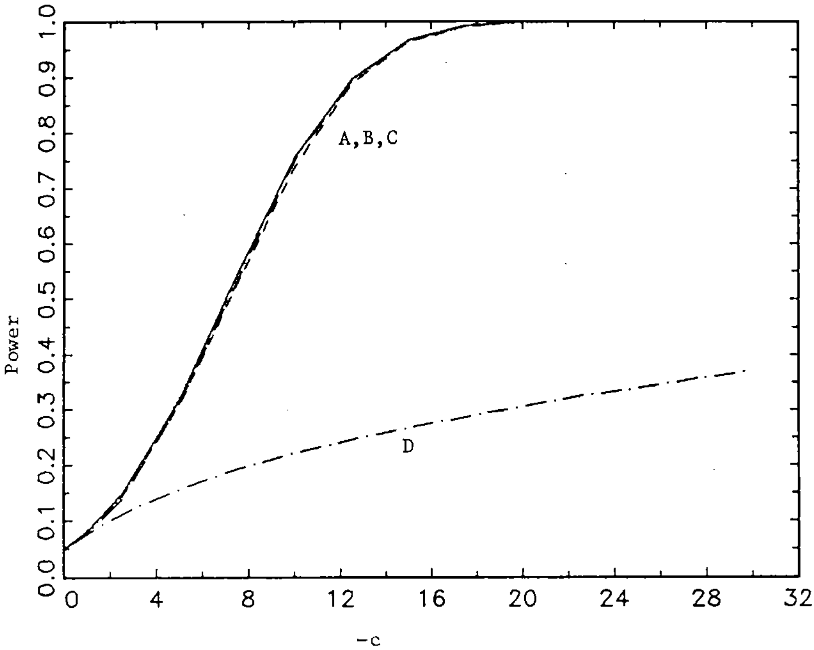

Figure 1

Asymptotic power functions of selected families of unit root tests:
No determistic case

<u>Key</u>: Solid line: Power Envelope

A: Sargan-Bhargava/modified Sargan-Bhargava tests

B: Dickey-Fuller/Phillips-Perron $\hat{\rho}$ tests

C: Dickey-Fuller/Phillips-Perron $\hat{\tau}$ tests

D: LMPI test

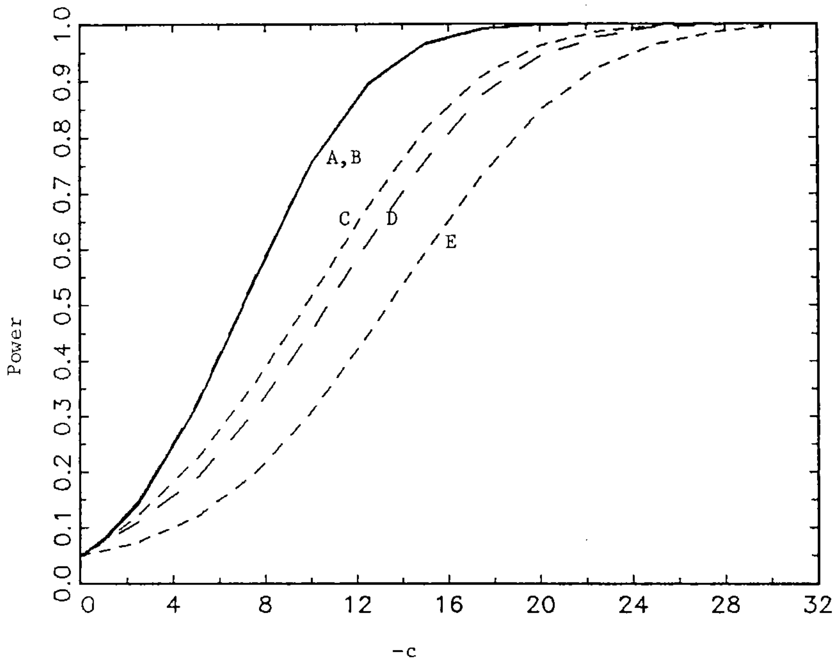

Figure 2

Asymptotic power functions of selected families of unit root tests: Demeaned case ($z_t = 1$)

<u>Key</u>: Solid line: Power Envelope

A: P_T^{μ} test

B: DF-GLS$^{\mu}$ test

C: Bhargava/modified Bhargava tests

D: Dickey-Fuller/Phillips-Perron $\hat{\rho}^{\mu}$ tests

E: Dickey-Fuller/Phillips-Perron $\hat{\tau}^{\mu}$ tests

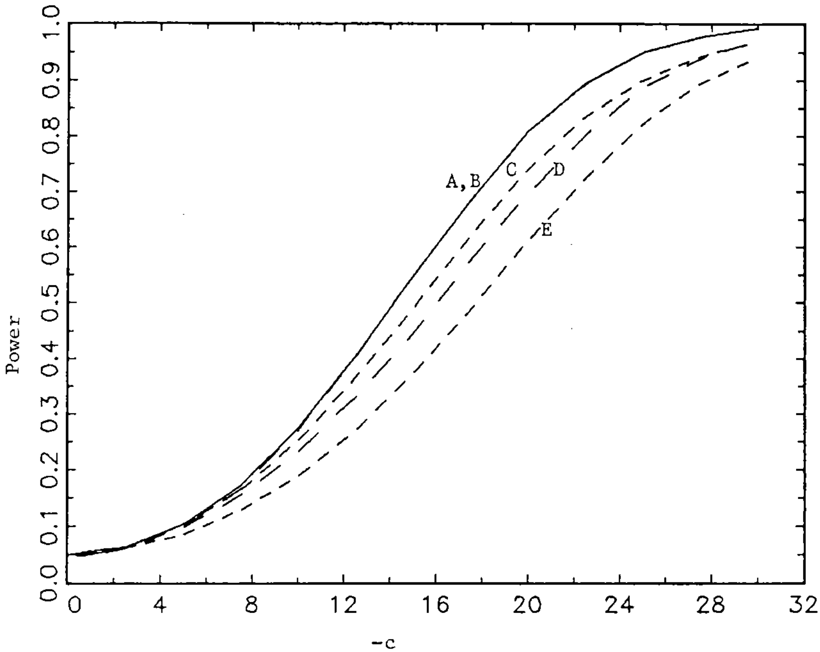

Figure 3

Asymptotic power functions of selected families of unit root tests: Linearly detrended case ($z_t = (1, t)$)

<u>Key</u>: Solid line: Power Envelope

A: P_T^{τ} test

B: DF-GLS$^{\tau}$ test

C: Bhargava/modified Bhargava tests

D: Dickey-Fuller/Phillips-Perron $\hat{\rho}^{\tau}$ tests

E: Dickey-Fuller/Phillips-Perron $\hat{\tau}^{\tau}$ tests asymptotic theory prevails, the ARE's at a given power can nonetheless be computed as the ratio of the values of c at which the various tests achieve the specified power (see Nyblom and Makelainen (1983), appendix A.1). In the demeaned case at a power of 50%, the ARE's of the SB, $\hat{\rho}$, and $\hat{\tau}$ tests are, respectively, 1.40, 1.53, and 1.91, relative to the power envelope. That is, in the demeaned case the asymptotic power loss from using the $\hat{\tau}$ family of tests (the Dickey-Fuller $\hat{\tau}$ test or the Phillips-Perron Z_τ test), relative to the POI test, is equivalent to discarding almost half the sample; the asymptotic power loss from using the $\hat{\rho}$ family of tests is equivalent to discarding one-third of the sample.

The $P_T^\mu(.5)$, DF-GLS$^\mu$, $P_T^\tau(.5)$, and DF-GLS$^\tau$ statistics have power functions that are very close to the power envelope, so close that they are difficult to distinguish visually in the plots. By construction, the ARE of the $P_T^\mu(.5)$ and $P_T^\tau(.5)$ statistics is 1 at a power of 50%. This ARE is very close to one at other powers as well.

Table 1 presents asymptotic critical values for the $P_T^\mu(.5)$, $P_T^\tau(.5)$, and DF-GLS$^\tau$ statistics, respectively computed by Monte Carlo integration of $\Psi(0,\bar{c})$ and $\Psi^\tau(0,\bar{c})$ and of (18b) with c=0. For the $P_T^\mu(.5)$ statistic, $\bar{c}$ = -7.0 was used; for the $P_T^\tau(.5)$ and DF-GLS$^\tau$ statistics, $\bar{c}$=-13.5 was used. Also included in Table 1 are finite-sample critical values of this statistic under the Gaussian AR(1) null with $\alpha = 1$, when the $P_T(.5)$ statistics are computed using $\hat{\omega}^2 = \hat{\sigma}^2$ and the DF-GLS statistic is computed with p=0 in (17). Although the T = 50 and T = 100 quantiles differ somewhat, there are only small differences between the T = 200 and T = ∞ cases. The limiting local-to-unity distribution of the DF-GLS$^\mu$ statistic does not depend on $\bar{c}$ and its asymptotic critical values can be found in the first panel of Fuller's (1976) table 8.5.2.

5. Finite Sample Performance

This section studies the finite-sample size and power of the P_T and DF-GLS statistics and, to provide a basis for comparison, the Dickey-Fuller (1979) t- and sum-of-coefficients tests, in a

Table 1. Critical Values for the $P_T^{\mu}(.5)$, $P_T^{\tau}(.5)$, and DF-GLS$^{\tau}$ Tests

T	Level 1%	2.5%	5%	10%
	A. Demeaned: $P_T^{\mu}(.5)$			
50	1.87	2.39	2.97	3.91
100	1.95	2.47	3.11	4.17
200	1.91	2.47	3.17	4.33
500	1.95	2.53	3.22	4.38
∞	1.99	2.55	3.26	4.48
	B. Detrended: $P_T^{\tau}(.5)$			
50	4.22	4.94	5.72	6.77
100	4.26	4.90	5.64	6.79
200	4.05	4.83	5.66	6.86
500	4.05	4.80	5.62	6.93
∞	3.96	4.78	5.62	6.89
	C. Detrended: DF-GLS$^{\tau}$			
50	-3.77	-3.46	-3.19	-2.89
100	-3.58	-3.29	-3.03	-2.74
200	-3.46	-3.18	-2.93	-2.64
500	-3.47	-3.15	-2.89	-2.59
∞	-3.48	-3.15	-2.89	-2.57

Notes: The $P_T^{\mu}(.5)$ statistic is given by (15b) with $\bar{c}=-7.0$; the $P_T^{\tau}(.5)$ statistic is given by (15c) with $\bar{c}=-13.5$. The critical values for T = 50, 100, 200, and 500 were computed by Monte Carlo simulation of these statistics with $\hat{\omega}^2 = \hat{\sigma}^2$, in which case P_T^{μ} simplifies to $P_T^{\mu}=T(M_T^{\mu}-\bar{\alpha})$ and P_T^{τ} simplifies to $P_T^{\tau}=T(M_T^{\tau}-\bar{\alpha})$. The DF-GLS$^{\tau}$ statistic is computed as described in the text surrounding (17); its finite-sample critical values were computed with p=0 in (17). The pseudo-data were generated according to a Gaussian random walk. The "T = ∞" line provides asymptotic critical values, computed using the asymptotic representations (16) (P_T statistics) and (18) (DF-GLS$^{\tau}$ statistic), where the functionals of Brownian motion were computed using their discretized counterparts based on a Gaussian random walk with T=500. Entries are based on 20,000 Monte Carlo replications.

Monte Carlo experiment consisting of 13 designs (data generating processes or DGP's). In each, $y_t = u_t$, where $u_t = \alpha u_{t-1} + v_t$, $t = 1, 2, \ldots, T$. Five values of α were considered, corresponding to $c = 0, -5, -10, -20$, and -30 under the nesting $\alpha = 1+c/T$. The DGP's are largely ones studied elsewhere in the literature, and the specific design is taken from Stock (1992), to which the reader is referred for details and for results for test statistics not considered here.

Four classes of DGP's are examined:

(19a) Gaussian MA(1): $v_t = \epsilon_t - \theta\epsilon_{t-1}$, $u_0=0$, $\theta = .8, .5, 0, -.5, -.8$

(19b) Gaussian AR(1): $v_t = \phi v_{t-1} + \epsilon_t$, $u_0=0$, $\phi = .5, -.5$

(19c) GARCH MA(1): $v_t = \varsigma_t - \theta\varsigma_{t-1}$, $\varsigma_t = h_t^{\frac{1}{2}}\epsilon_t$, $u_0=0$, $h_t = \omega_0 + .65h_{t-1} + .25\varsigma_{t-1}^2$

(19d) Gaussian MA(1), u_0 unconditional: $v_t = \epsilon_t - \theta\epsilon_{t-1}$, $u_1 \sim N(0, \gamma_u(0))$, $\theta = .5, 0, -.5$

where in each case $\epsilon_t \sim$ i.i.d. N(0, 1). The Gaussian MA(1) DGP (19a) has received the most attention in the literature and was the focus of Schwert's (1989) and Phillips and Perron's (1988) studies. DeJong, Nankervis, Savin and Whiteman (1992) suggested studying the Gaussian AR(1) model, in part because it can capture more realistically the nonzero sample autocovariances found in macroeconomic time series data. The GARCH DGP introduces conditional heteroskedasticity in the innovations, a phenomenon often found in financial data. The GARCH parameters sum to 0.9, so this DGP introduces considerable persistence in the heteroskedasticity although fourth moments of ς_t still exist. This model is taken from Lumsdaine (1991), who chose it as representative of the empirical models in Baillie and Bollerslev (1989). For the simulations, $\omega_0 = 1$ and $h_0 = 1$ (the unit root test statistics considered here are invariant to ω_0, which is in effect a scale parameter).

Because a constant (or constant and trend) are included, the null distributions of the test statistics in this Monte Carlo study are invariant to the initial condition. However, as pointed out by Evans and Savin (1981, 1984) and studied by Schmidt and Phillips (1992) and DeJong, Nankervis, Savin and Whiteman (1992), the power of unit root tests typically depends on u_1. This dependence is investigated in the final design by drawing u_1 from its unconditional distribution, $N(0, \gamma_u(0))$, where $\gamma_u(0)=(1+\theta^2-2\theta\alpha)/(1-\alpha^2)$ and only values of $|\alpha|<1$ are examined.

Two spectral estimators for the P_T^{μ} and P_T^{τ} tests are considered: a SC estimator and an AR spectral estimator. The AR estimator is given by,

(20) $$\hat{\omega}_{AR}^2 = \hat{\sigma}_{\epsilon}^2/(1-\sum_{j=1}^{p}\hat{a}_j)^2,$$

where $\{\hat{a}_j\}$ and $\hat{\sigma}_{\epsilon}^2$ are obtained from the OLS regression,

(21) $$\Delta y_t = \delta_0 + a_0 y_{t-1} + \sum_{j=1}^{p} a_j \Delta y_{t-j} + \epsilon_t, \quad t = 2, 3, \ldots, T.$$

Three choices of lag length p are studied: p = 4, p = 8, and p chosen using the BIC, where the maximum number of lags is 8 and the minimum is 3. The resulting estimators are respectively denoted AR(4), AR(12), and AR(BIC). (The Dickey-Fuller tests were also evaluated for these three choices of p.) Stock (1988, lemma 1) shows that, when p is chosen to satisfy Berk's (1974) rate conditions, the AR estimator satisfies condition D(b).

The SC estimator is given by,

(22) $$\hat{\omega}_{SC}^2 = \sum_{m=-\ell_T}^{\ell_T} k(m/\ell_T)\hat{\gamma}_{\Delta\hat{u}}(m)$$

where $\hat{\gamma}_x(m) = (T-m)^{-1}\sum_{t=|m|+1}^{T}(x_t-\bar{x})(x_{t-m}-\bar{x})$, $k(\cdot)$ is a kernel weighting function, and $\Delta\hat{u}$ is the residual from the regression of y_t onto $(1, y_{t-1})$ (demeaned case) or y_t onto $(1, t, y_{t-1})$ (linearly detrended case), and where the Parzen kernel is used. Three methods for the selection of the truncation parameter ℓ_T are used: $\ell_T = 4$, $\ell_T = 12$, and ℓ_T estimated according to Andrews' (1991) optimal automatic procedure (his equations (6.2) and (6.4)). The SC estimators are respectively denoted SC(4), SC(12), and SC(auto). It follows from Phillips and Ouliaris's (1990, theorem 5.1) more general result that under suitable rate conditions on ℓ_T, the SC estimator satisfies condition D(b).

The results are summarized in Table 2 for the demeaned statistics and in Table 3 for the detrended statistics. In both cases, tests with asymptotic level 5% are studied and T = 100. For each statistic, the first column provides the asymptotic size (always 5%) and power. The remaining entries for α=1 are the empirical size, that is, the Monte Carlo rejection rate based on the asymptotic critical value. The entries for $|\alpha|<1$ are the size-adjusted power.

These results suggest six conclusions. First, the main prediction of the asymptotic results in the previous section – that the P_T and DF-GLS statistics have significantly better (size-adjusted) power – is borne out by the Monte Carlo results. As suggested by the asymptotic results, this power improvement is greatest for the demeaned statistics. For example, in the $\theta = 0$ case (model (19a); table 2, third column of results), the $\hat{\tau}^\mu$, $\hat{\rho}^\mu$, P_T^μ and DF-GLS$^\mu$ tests (AR(BIC) choice of p) respectively have power of .22, .37, .59, and .60 against $\alpha = .9$; against $\alpha = .8$, these powers are .59, .76, .91, and .93. The values of the size-adjusted powers depend somewhat on the value of the nuisance parameter θ in model (19a); for example, for $\theta = .8$, the power of the DF-GLS$^\mu$ test rises to .68, while for $\theta = -.8$ it drops to .56. However, the relative improvements in size-adjusted powers for the P_T and DF-GLS statistics over the $\hat{\rho}$ and $\hat{\tau}$ statistics are present typically for all values of the nuisance parameters in each of the conditional models (models (19a-c)). In the detrended case, the improvements in size-adjusted power are smaller but remain substantial. For example, in the $\theta = 0$, $\alpha = .9$ case, the $\hat{\rho}^\tau$ and $\hat{\tau}^\tau$ statistics (AR(BIC)) have powers of .20 and .15, respectively, while the DF-GLS$^\tau$ and P_T^τ(AR(BIC)) tests have powers of .24 and .36, respectively. This improvement in size-adjusted power is present for the other values of the nuisance parameters as well (models (19a-c)).

Second, the choice of spectral estimator has a large effect on the size of the P_T^μ and P_T^τ tests, with the AR estimator exhibiting much smaller size distortions than the SC estimator. For example, in model (19a) in the linearly detrended case, the P_T^τ test with the SC(auto) estimator has sizes of 49% and 1% with $\theta = .5$ and $\theta = -.5$, respectively, while the size for the AR(BIC)-based test is 5% and 10% in these two cases. These findings mirror similar results found for other unit root test statistics; see DeJong, Nankervis, Savin and Whiteman (1992) and Perron (1991) for discussions.

Third, among the tests which use the autoregressive spectral estimator, using the AR(8) rather than AR(4) reduces the size distortion for the $\hat{\tau}$ and DF-GLS tests but increases it for the $\hat{\rho}$ and

Table 2. Size and Size-Adjusted Power of Selected Tests of the I(1) null: Monte Carlo Results

5% level tests, demeaned case (z_t = 1), T = 100

Test Statistic	α	Asy. Power	MA(1), θ= -0.8	-0.5	0.0	0.5	0.8	AR(1), ϕ= 0.5	-0.5	GARCH MA(1), θ= -0.5	0.0	0.5	Uncond'l MA(1), θ= -0.5	0.0	0.5
$P_T^\mu(.5)$ AR(4)	1.00	.05	0.08	0.11	0.11	0.10	0.31	0.13	0.11	0.11	0.12	0.11	0.11	0.11	0.10
	.95	.32	0.25	0.24	0.24	0.26	0.19	0.22	0.25	0.25	0.25	0.25	0.15	0.16	0.17
	.90	.76	0.50	0.49	0.51	0.54	0.41	0.45	0.51	0.50	0.50	0.51	0.32	0.33	0.35
	.80	1.00	0.81	0.81	0.84	0.86	0.78	0.74	0.80	0.80	0.82	0.82	0.59	0.62	0.65
	.70	1.00	0.92	0.92	0.93	0.94	0.92	0.87	0.90	0.91	0.91	0.92	0.73	0.76	0.80
$P_T^\mu(.5)$ AR(8)	1.00	.05	0.18	0.20	0.20	0.18	0.20	0.22	0.20	0.21	0.20	0.18	0.20	0.20	0.18
	.95	.32	0.18	0.18	0.19	0.18	0.15	0.18	0.17	0.18	0.18	0.17	0.13	0.13	0.13
	.90	.76	0.31	0.31	0.32	0.32	0.30	0.31	0.29	0.30	0.30	0.30	0.24	0.25	0.25
	.80	1.00	0.47	0.48	0.50	0.51	0.51	0.46	0.46	0.48	0.48	0.49	0.40	0.42	0.43
	.70	1.00	0.56	0.57	0.59	0.60	0.47	0.55	0.53	0.56	0.57	0.58	0.49	0.51	0.52
$P_T^\mu(.5)$ AR(BIC)	1.00	.05	0.14	0.11	0.10	0.11	0.42	0.11	0.10	0.13	0.11	0.12	0.11	0.10	0.11
	.95	.32	0.24	0.27	0.28	0.28	0.19	0.26	0.27	0.26	0.26	0.27	0.17	0.17	0.17
	.90	.76	0.50	0.57	0.59	0.59	0.41	0.52	0.56	0.54	0.56	0.57	0.37	0.37	0.36
	.80	1.00	0.82	0.89	0.91	0.92	0.79	0.83	0.88	0.86	0.88	0.89	0.67	0.69	0.69
	.70	1.00	0.92	0.97	0.98	0.98	0.94	0.93	0.96	0.95	0.96	0.96	0.81	0.83	0.84
$P_T^\mu(.5)$ SC(4)	1.00	.05	0.04	0.04	0.06	0.37	0.93	0.02	0.23	0.04	0.07	0.38	0.04	0.06	0.37
	.95	.32	0.30	0.31	0.32	0.29	0.23	0.29	0.31	0.29	0.30	0.29	0.18	0.18	0.17
	.90	.76	0.68	0.69	0.72	0.70	0.28	0.66	0.73	0.67	0.70	0.68	0.40	0.44	0.38
	.80	1.00	0.98	0.98	0.99	0.98	0.12	0.97	0.99	0.97	0.98	0.97	0.74	0.79	0.67
	.70	1.00	1.00	1.00	1.00	0.99	0.02	1.00	1.00	1.00	1.00	0.99	0.86	0.90	0.73
$P_T^\mu(.5)$ SC(12)	1.00	.05	0.02	0.02	0.07	0.50	0.98	0.01	0.33	0.03	0.08	0.51	0.02	0.07	0.50
	.95	.32	0.29	0.29	0.29	0.27	0.15	0.29	0.29	0.29	0.30	0.28	0.17	0.17	0.15
	.90	.76	0.64	0.65	0.70	0.62	0.09	0.59	0.68	0.64	0.68	0.59	0.36	0.40	0.33
	.80	1.00	0.96	0.97	0.99	0.84	0.01	0.92	0.97	0.95	0.98	0.80	0.63	0.73	0.49
	.70	1.00	1.00	1.00	1.00	0.78	0.00	0.98	0.99	0.99	1.00	0.74	0.76	0.85	0.46
$P_T^\mu(.5)$ SC(auto)	1.00	.05	0.04	0.04	0.06	0.31	0.88	0.03	0.18	0.04	0.07	0.34	0.04	0.06	0.31
	.95	.32	0.30	0.30	0.32	0.31	0.30	0.29	0.31	0.30	0.31	0.31	0.17	0.19	0.18
	.90	.76	0.67	0.68	0.74	0.73	0.59	0.64	0.73	0.68	0.69	0.70	0.39	0.44	0.41
	.80	1.00	0.97	0.98	0.99	0.99	0.72	0.96	0.99	0.97	0.98	0.98	0.72	0.80	0.70
	.70	1.00	1.00	1.00	1.00	1.00	0.71	0.99	1.00	1.00	1.00	1.00	0.85	0.91	0.79
DF-GLS$^\mu$ AR(4)	1.00	.05	0.05	0.07	0.07	0.08	0.33	0.07	0.07	0.06	0.07	0.09	0.07	0.07	0.08
	.95	.32	0.25	0.25	0.27	0.30	0.31	0.25	0.27	0.27	0.26	0.26	0.15	0.16	0.17
	.90	.75	0.53	0.54	0.57	0.65	0.69	0.51	0.59	0.55	0.56	0.60	0.32	0.35	0.37
	.80	1.00	0.87	0.88	0.91	0.96	0.97	0.83	0.92	0.88	0.89	0.93	0.60	0.62	0.63
	.70	1.00	0.96	0.97	0.98	1.00	1.00	0.94	0.99	0.96	0.97	0.99	0.71	0.74	0.71
DF-GLS$^\mu$ AR(8)	1.00	.05	0.05	0.06	0.06	0.06	0.12	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.06
	.95	.32	0.21	0.23	0.23	0.23	0.30	0.23	0.24	0.22	0.22	0.23	0.14	0.14	0.13
	.90	.75	0.42	0.43	0.45	0.47	0.62	0.42	0.46	0.43	0.44	0.47	0.25	0.25	0.24
	.80	1.00	0.68	0.70	0.72	0.79	0.92	0.66	0.74	0.69	0.71	0.78	0.40	0.40	0.37
	.70	1.00	0.80	0.82	0.84	0.91	0.98	0.77	0.87	0.82	0.83	0.90	0.46	0.47	0.42

Table 2, continued

Test		Asy.	- - - - - MA(1), θ= - - - - -					AR(1), φ=		GARCH MA(1), θ=			Uncond'l MA(1), θ=		
Statistic	α	Power	-0.8	-0.5	0.0	0.5	0.8	-0.5	0.5	-0.5	0.0	0.5	-0.5	0.0	0.5
DF-GLS$^{\mu}$	1.00	.05	0.10	0.08	0.07	0.11	0.45	0.07	0.08	0.09	0.08	0.11	0.08	0.07	0.11
AR(BIC)	.95	.32	0.26	0.27	0.28	0.30	0.30	0.26	0.28	0.26	0.27	0.26	0.17	0.16	0.17
	.90	.75	0.56	0.59	0.60	0.67	0.68	0.54	0.62	0.57	0.59	0.61	0.37	0.37	0.37
	.80	1.00	0.87	0.92	0.93	0.97	0.98	0.86	0.95	0.90	0.92	0.95	0.66	0.68	0.68
	.70	1.00	0.96	0.98	0.99	1.00	1.00	0.95	1.00	0.98	0.98	1.00	0.79	0.80	0.76
DF - ρ^{μ}	1.00	.05	0.07	0.09	0.10	0.12	0.50	0.10	0.10	0.10	0.11	0.14	0.09	0.10	0.12
AR(4)	.95	.19	0.16	0.16	0.16	0.16	0.14	0.16	0.17	0.16	0.16	0.18	0.14	0.14	0.15
	.90	.45	0.32	0.32	0.33	0.37	0.33	0.29	0.35	0.33	0.34	0.38	0.32	0.33	0.35
	.80	.94	0.63	0.65	0.69	0.79	0.73	0.56	0.72	0.65	0.67	0.78	0.65	0.69	0.79
	.70	1.00	0.83	0.84	0.88	0.95	0.74	0.74	0.91	0.83	0.86	0.94	0.84	0.88	0.94
DF - ρ^{μ}	1.00	.05	0.15	0.17	0.17	0.17	0.30	0.17	0.17	0.17	0.16	0.17	0.17	0.17	0.17
AR(8)	.95	.19	0.14	0.14	0.14	0.14	0.12	0.15	0.15	0.15	0.14	0.14	0.12	0.13	0.13
	.90	.45	0.23	0.22	0.23	0.26	0.25	0.23	0.25	0.24	0.24	0.25	0.22	0.24	0.25
	.80	.94	0.38	0.37	0.39	0.45	0.38	0.37	0.42	0.39	0.40	0.44	0.38	0.41	0.46
	.70	1.00	0.47	0.45	0.48	0.54	0.30	0.44	0.50	0.47	0.49	0.52	0.45	0.49	0.54
DF - ρ^{μ}	1.00	.05	0.13	0.10	0.08	0.13	0.62	0.09	0.09	0.11	0.10	0.14	0.10	0.08	0.13
AR(BIC)	.95	.19	0.17	0.18	0.17	0.17	0.13	0.17	0.17	0.17	0.17	0.17	0.15	0.16	0.15
	.90	.45	0.33	0.35	0.37	0.40	0.31	0.32	0.36	0.34	0.36	0.39	0.35	0.36	0.38
	.80	.94	0.67	0.73	0.76	0.85	0.77	0.64	0.78	0.69	0.74	0.82	0.73	0.76	0.85
	.70	1.00	0.85	0.91	0.93	0.98	0.91	0.81	0.95	0.88	0.91	0.97	0.90	0.93	0.98
DF-τ^{μ}	1.00	.05	0.04	0.05	0.05	0.06	0.30	0.05	0.05	0.05	0.05	0.06	0.05	0.05	0.06
AR(4)	.95	.12	0.07	0.09	0.09	0.11	0.13	0.09	0.11	0.10	0.11	0.12	0.09	0.10	0.12
	.90	.31	0.14	0.18	0.20	0.25	0.32	0.18	0.23	0.21	0.22	0.27	0.19	0.21	0.26
	.80	.85	0.37	0.44	0.49	0.66	0.77	0.39	0.57	0.49	0.53	0.67	0.46	0.51	0.67
	.70	1.00	0.58	0.66	0.73	0.89	0.95	0.57	0.81	0.70	0.76	0.89	0.67	0.74	0.90
DF-τ^{μ}	1.00	.05	0.05	0.05	0.05	0.05	0.09	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
AR(8)	.95	.12	0.07	0.08	0.08	0.08	0.12	0.08	0.09	0.08	0.08	0.09	0.09	0.09	0.09
	.90	.31	0.12	0.14	0.14	0.16	0.26	0.13	0.16	0.14	0.14	0.17	0.14	0.15	0.18
	.80	.85	0.24	0.26	0.28	0.36	0.59	0.24	0.32	0.28	0.29	0.37	0.28	0.29	0.38
	.70	1.00	0.33	0.37	0.39	0.53	0.81	0.33	0.47	0.39	0.42	0.55	0.38	0.41	0.55
DF-τ^{μ}	1.00	.05	0.08	0.06	0.06	0.08	0.46	0.06	0.05	0.07	0.06	0.08	0.06	0.06	0.08
AR(BIC)	.95	.12	0.11	0.10	0.10	0.13	0.13	0.10	0.11	0.10	0.10	0.13	0.11	0.11	0.13
	.90	.31	0.23	0.22	0.22	0.31	0.31	0.20	0.25	0.23	0.23	0.29	0.24	0.24	0.32
	.80	.85	0.55	0.56	0.59	0.77	0.78	0.46	0.65	0.54	0.58	0.73	0.57	0.60	0.77
	.70	1.00	0.76	0.79	0.83	0.96	0.96	0.67	0.89	0.77	0.82	0.93	0.80	0.84	0.96

Notes to table 2: For each statistic, the first row of entries are the empirical rejection rates under the null, that is, the empirical size of the test. The remaining entries are the size-adjusted power under the model described in the column heading. The column, "Asy. Power," gives the local-to-unity approximation to the asymptotic power function for each statistic. The entry below the name of each statistic indicates the spectral density estimator used. SC(auto) is the Andrews (1991) automatic bandwidth estimator with the Parzen kernel. AR(4), AR(8), and AR(BIC) are the AR estimator with p=4, p=8, and p selected by the BIC automatic selector (subject to $3 \le p \le 8$), respectively. The size in the final three columns was computed with $u_0 = 0$; for the $|\alpha|<1$ cases in the final three columns, u_0 was drawn from its unconditional distribution. Based on 5000 Monte Carlo replications.

Table 3. Size and Size-Adjusted Power of Selected Tests of the I(1) null: Monte Carlo Results

5% level tests, detrended case ($z_t = (1, t)$), T = 100

Test Statistic	α	Asy. Power	MA(1), θ= -0.8	-0.5	0.0	0.5	0.8	AR(1), ϕ= 0.5	-0.5	GARCH MA(1), θ= -0.5	0.0	0.5	Uncond'l MA(1), θ= -0.5	0.0	0.5
$P_T^\tau(.5)$ AR(4)	1.00	.05	0.05	0.08	0.09	0.05	0.19	0.11	0.07	0.09	0.09	0.06	0.08	0.09	0.05
	.95	.10	0.17	0.18	0.18	0.18	0.14	0.17	0.17	0.17	0.17	0.18	0.13	0.14	0.14
	.90	.27	0.32	0.33	0.34	0.37	0.32	0.30	0.34	0.33	0.34	0.36	0.28	0.29	0.31
	.80	.81	0.59	0.62	0.65	0.70	0.69	0.55	0.61	0.61	0.63	0.66	0.55	0.57	0.62
	.70	.99	0.75	0.78	0.81	0.85	0.89	0.70	0.75	0.77	0.78	0.82	0.71	0.73	0.78
$P_T^\tau(.5)$ AR(8)	1.00	.05	0.16	0.18	0.18	0.14	0.13	0.21	0.17	0.19	0.19	0.13	0.18	0.18	0.14
	.95	.10	0.15	0.16	0.16	0.16	0.13	0.15	0.15	0.14	0.15	0.15	0.13	0.13	0.13
	.90	.27	0.26	0.26	0.26	0.27	0.25	0.25	0.24	0.23	0.25	0.25	0.23	0.24	0.25
	.80	.81	0.40	0.41	0.41	0.44	0.45	0.38	0.38	0.37	0.40	0.42	0.38	0.39	0.42
	.70	.99	0.48	0.49	0.50	0.53	0.42	0.45	0.46	0.46	0.48	0.51	0.46	0.48	0.50
$P_T^\tau(.5)$ AR(BIC)	1.00	.05	0.13	0.10	0.07	0.05	0.29	0.10	0.05	0.11	0.08	0.06	0.10	0.07	0.05
	.95	.10	0.18	0.17	0.17	0.18	0.15	0.16	0.16	0.17	0.16	0.17	0.14	0.14	0.14
	.90	.27	0.36	0.36	0.36	0.39	0.32	0.31	0.35	0.34	0.34	0.37	0.29	0.30	0.32
	.80	.81	0.65	0.69	0.72	0.77	0.70	0.60	0.68	0.66	0.68	0.73	0.61	0.63	0.68
	.70	.99	0.82	0.86	0.88	0.92	0.90	0.76	0.84	0.83	0.84	0.89	0.78	0.81	0.86
$P_T^\tau(.5)$ SC(4)	1.00	.05	0.01	0.01	0.04	0.59	1.00	0.01	0.39	0.02	0.05	0.62	0.01	0.04	0.59
	.95	.10	0.11	0.11	0.12	0.11	0.08	0.12	0.11	0.11	0.11	0.11	0.10	0.10	0.11
	.90	.27	0.30	0.30	0.32	0.30	0.08	0.28	0.30	0.28	0.30	0.27	0.23	0.25	0.22
	.80	.81	0.76	0.78	0.85	0.74	0.02	0.72	0.83	0.74	0.80	0.66	0.62	0.70	0.53
	.70	.99	0.96	0.97	0.99	0.86	0.00	0.94	0.98	0.96	0.98	0.80	0.85	0.92	0.64
$P_T^\tau(.5)$ SC(12)	1.00	.05	0.00	0.00	0.03	0.77	1.00	0.00	0.57	0.00	0.03	0.77	0.00	0.03	0.77
	.95	.10	0.10	0.11	0.12	0.11	0.05	0.11	0.11	0.11	0.12	0.11	0.09	0.10	0.09
	.90	.27	0.25	0.26	0.32	0.22	0.02	0.24	0.26	0.27	0.29	0.20	0.21	0.25	0.16
	.80	.81	0.65	0.69	0.81	0.35	0.00	0.57	0.61	0.68	0.74	0.31	0.51	0.64	0.25
	.70	.99	0.87	0.91	0.96	0.29	0.00	0.83	0.71	0.89	0.93	0.25	0.73	0.85	0.20
$P_T^\tau(.5)$ SC(auto)	1.00	.05	0.01	0.01	0.04	0.49	0.99	0.00	0.26	0.01	0.05	0.51	0.01	0.04	0.49
	.95	.10	0.12	0.11	0.12	0.12	0.10	0.11	0.12	0.11	0.11	0.11	0.10	0.10	0.10
	.90	.27	0.30	0.30	0.32	0.33	0.22	0.27	0.32	0.30	0.30	0.30	0.23	0.25	0.24
	.80	.81	0.77	0.79	0.85	0.85	0.41	0.69	0.86	0.76	0.81	0.80	0.62	0.70	0.63
	.70	.99	0.97	0.97	0.99	0.98	0.39	0.93	0.99	0.97	0.98	0.97	0.87	0.93	0.82
DF-GLS$^\tau$ AR(4)	1.00	.05	0.03	0.05	0.06	0.08	0.37	0.06	0.06	0.05	0.06	0.08	0.05	0.06	0.08
	.95	.10	0.10	0.11	0.11	0.10	0.11	0.10	0.11	0.10	0.10	0.10	0.09	0.09	0.09
	.90	.27	0.22	0.22	0.23	0.25	0.27	0.19	0.25	0.22	0.23	0.24	0.17	0.18	0.19
	.80	.81	0.50	0.51	0.55	0.64	0.67	0.42	0.59	0.51	0.53	0.61	0.40	0.42	0.46
	.70	.99	0.71	0.73	0.77	0.88	0.88	0.60	0.82	0.72	0.76	0.86	0.58	0.61	0.65
DF-GLS$^\tau$ AR(8)	1.00	.05	0.04	0.05	0.05	0.04	0.09	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04
	.95	.10	0.08	0.09	0.09	0.10	0.11	0.09	0.09	0.08	0.09	0.09	0.08	0.08	0.09
	.90	.27	0.16	0.17	0.17	0.20	0.25	0.15	0.18	0.15	0.16	0.18	0.13	0.13	0.15
	.80	.81	0.30	0.31	0.33	0.40	0.53	0.28	0.35	0.31	0.32	0.38	0.22	0.23	0.26
	.70	.99	0.41	0.42	0.45	0.56	0.68	0.37	0.48	0.43	0.45	0.53	0.30	0.31	0.33

Table 3, continued

Test Statistic	α	Asy. Power	MA(1), θ= -0.8	-0.5	0.0	0.5	0.8	AR(1), φ= -0.5	0.5	GARCH MA(1), θ= -0.5	0.0	0.5	Uncond'l MA(1), θ= -0.5	0.0	0.5
DF-GLS$^{\tau}$	1.00	.05	0.11	0.08	0.07	0.11	0.58	0.06	0.07	0.08	0.06	0.11	0.08	0.07	0.11
AR(BIC)	.95	.10	0.11	0.10	0.10	0.11	0.12	0.10	0.10	0.10	0.10	0.11	0.09	0.09	0.09
	.90	.27	0.23	0.23	0.24	0.28	0.27	0.22	0.25	0.23	0.24	0.26	0.19	0.19	0.21
	.80	.81	0.53	0.57	0.61	0.72	0.70	0.48	0.63	0.56	0.59	0.69	0.46	0.49	0.54
	.70	.99	0.75	0.80	0.84	0.94	0.91	0.69	0.88	0.78	0.82	0.91	0.67	0.71	0.76
DF - ρ^{τ}	1.00	.05	0.11	0.15	0.16	0.20	0.70	0.16	0.17	0.15	0.17	0.21	0.15	0.16	0.20
AR(4)	.95	.10	0.09	0.09	0.09	0.09	0.09	0.09	0.10	0.09	0.09	0.10	0.08	0.09	0.09
	.90	.23	0.17	0.17	0.18	0.19	0.17	0.16	0.19	0.17	0.17	0.20	0.16	0.17	0.18
	.80	.70	0.37	0.38	0.43	0.51	0.42	0.33	0.45	0.39	0.39	0.49	0.37	0.41	0.48
	.70	.97	0.55	0.57	0.64	0.76	0.53	0.48	0.68	0.58	0.60	0.75	0.57	0.63	0.75
DF - ρ^{τ}	1.00	.05	0.28	0.31	0.31	0.31	0.46	0.30	0.32	0.30	0.30	0.31	0.31	0.31	0.31
AR(8)	.95	.10	0.08	0.08	0.08	0.08	0.08	0.07	0.08	0.08	0.08	0.07	0.07	0.08	0.08
	.90	.23	0.12	0.11	0.12	0.12	0.12	0.11	0.12	0.12	0.12	0.12	0.11	0.12	0.13
	.80	.70	0.18	0.16	0.18	0.20	0.15	0.15	0.18	0.19	0.20	0.20	0.16	0.18	0.22
	.70	.97	0.22	0.20	0.22	0.25	0.14	0.18	0.22	0.22	0.23	0.23	0.20	0.22	0.23
DF - ρ^{τ}	1.00	.05	0.21	0.16	0.13	0.21	0.81	0.13	0.13	0.16	0.13	0.21	0.16	0.13	0.21
AR(BIC)	.95	.10	0.09	0.09	0.10	0.10	0.08	0.10	0.09	0.09	0.10	0.10	0.09	0.09	0.10
	.90	.23	0.18	0.18	0.20	0.22	0.17	0.18	0.20	0.18	0.19	0.21	0.18	0.19	0.20
	.80	.70	0.42	0.45	0.49	0.58	0.47	0.39	0.50	0.43	0.47	0.56	0.43	0.48	0.57
	.70	.97	0.62	0.67	0.74	0.87	0.73	0.57	0.78	0.65	0.71	0.84	0.66	0.73	0.86
DF-τ^{τ}	1.00	.05	0.03	0.05	0.05	0.06	0.37	0.05	0.05	0.05	0.05	0.07	0.05	0.05	0.06
AR(4)	.95	.09	0.07	0.07	0.07	0.08	0.09	0.08	0.09	0.07	0.07	0.08	0.07	0.08	0.08
	.90	.19	0.10	0.12	0.13	0.16	0.18	0.13	0.15	0.11	0.13	0.15	0.12	0.14	0.17
	.80	.61	0.24	0.28	0.32	0.43	0.49	0.25	0.36	0.27	0.30	0.41	0.28	0.32	0.44
	.70	.94	0.40	0.45	0.53	0.71	0.78	0.39	0.59	0.44	0.51	0.68	0.45	0.52	0.72
DF-τ^{τ}	1.00	.05	0.04	0.05	0.05	0.04	0.10	0.04	0.04	0.04	0.05	0.04	0.05	0.05	0.04
AR(8)	.95	.09	0.06	0.07	0.07	0.07	0.08	0.06	0.07	0.06	0.06	0.07	0.07	0.07	0.07
	.90	.19	0.09	0.09	0.10	0.11	0.14	0.09	0.10	0.09	0.08	0.11	0.10	0.11	0.11
	.80	.61	0.15	0.16	0.17	0.22	0.33	0.15	0.20	0.16	0.16	0.23	0.17	0.19	0.24
	.70	.94	0.21	0.22	0.25	0.34	0.53	0.20	0.28	0.23	0.23	0.36	0.23	0.26	0.36
DF-τ^{τ}	1.00	.05	0.10	0.07	0.05	0.09	0.58	0.05	0.06	0.07	0.06	0.09	0.07	0.05	0.09
AR(BIC)	.95	.09	0.09	0.08	0.08	0.09	0.08	0.08	0.08	0.08	0.08	0.09	0.08	0.08	0.09
	.90	.19	0.16	0.14	0.15	0.18	0.17	0.14	0.15	0.14	0.14	0.18	0.15	0.15	0.18
	.80	.61	0.36	0.36	0.39	0.51	0.50	0.30	0.42	0.34	0.37	0.48	0.36	0.39	0.52
	.70	.94	0.57	0.58	0.64	0.81	0.80	0.48	0.69	0.55	0.60	0.78	0.58	0.64	0.81

Notes: see the notes to Table 2.

P_T tests. Using more lags reduces size-adjusted power in each case. The AR(BIC) estimator seems to provide a good compromise, with sizes which usually fall between the AR(4)- and AR(8)-based tests and with substantially better size-adjusted power than the AR(8)-based tests.[6]

Fourth, although there is no uniform ranking of the tests in terms of their size distortions, it is possible to make some general observations about relative size performance. We restrict the comparisons across tests to the tests using the AR(BIC) lag choice. Among these tests, in both the demeaned and detrended cases the $\hat{\tau}$ and DF-GLS tests have the lowest size distortions, with the $\hat{\tau}$ test typically best. In the demeaned case, the P_T^μ and $\hat{\rho}^\mu$ tests have sizes between 8% and 14% for all cases in models (19a-c), except for the extreme case of $\theta = -.8$ for which all tests have size exceeding 40%. In the detrended case, the $\hat{\rho}^\tau$ test has large size distortions in all cases, with size of at least 13%, but the P_T^τ test has smaller distortions, with size almost as good as DF-GLS$^\tau$.

Fifth, in the demeaned case, the size-adjusted power is typically lower for each statistic than implied by the asymptotic power. This gap is greatest for alternatives farthest from the null hypothesis. Similar results are found in the detrended case, with the exception of the P_T^τ statistic, which has size-adjusted power above the asymptotic power envelope for alternatives $\alpha \geq .9$. For all tests in all cases, the introduction of GARCH errors has little effect on the size or size-adjusted power.

Sixth, the powers of the $\hat{\rho}$, P_T and DF-GLS tests deteriorate substantially when u_0 is drawn from its unconditional distribution (model (19d)). Under the local nesting for T large, $T^{-\frac{1}{2}}u_0 = O_p(1)$, so condition C is violated, the results in theorem 3, (16), and (18) do not obtain, and in particular the power functions of the $P_T(.5)$ statistic need not achieve the power envelope.[7] Even so, in the detrended case (model (19d), $\theta = 0$), the size-adjusted power of P_T^τ and DF-GLS$^\tau$ exceeds that of $\hat{\rho}^\tau$ and $\hat{\tau}^\tau$ for all values of α considered. In the demeaned case, the size-adjusted power of P_T^μ and DF-GLS$^\mu$ exceeds that of $\hat{\rho}^\mu$ and $\hat{\tau}^\mu$ for close but not distant alternatives.

Because (21) does not contain t as a regressor, the AR spectral estimator (20) is invariant in the demeaned case, but it is not invariant to changes in β_1 in the detrended case under the alternative; the P_T^τ/AR statistic in the detrend case as calculated for table 3 is not strictly invariant (although

P_T^μ in table 2 is). A strictly invariant version of P_T^τ can be constructed by including t as a regressor in (21). For c near zero, the size-adjusted Monte Carlo power of this alternative P_T^τ statistic is near the envelope, falling off for more distant alternatives (for model (19a), $\theta = 0$, the size-adjusted powers are .10, .20, .43, and .61 for $\alpha = .95, .90, .80, .70$, respectively, for the same design parameters as in tables 3, AR(4) spectral estimator). The size of this alternative P_T^τ statistic, however, is unsatisfactory, 20% for the $\theta = 0$, model (19a), AR(4) estimator case. Although the P_T^τ statistic reported in table 3 is not invariant, these results (and additional simulations) suggest it might be preferable to the strictly invariant version if β_1 is thought to be small.

These Monte Carlo results suggests that the DF-GLS statistic with BIC lag choice represents a reasonable tradeoff between size and power. Except in the extreme $\theta = -.8$ case, its size is between 6% and 11% in the both the demeaned and detrended cases. The size-adjusted power of the DF-GLS statistic is in general comparable to that of the P_T statistic, except for alternatives close to zero in the detrended case when the P_T^τ test is more powerful. The size-adjusted power of the DF-GLS statistics always exceeds that of the $\hat{\rho}$ and $\hat{\tau}$ statistics (except for distant alternatives in model (19d)), in some cases by a large margin, particularly in the demeaned case. The Monte Carlo results indicate that the size and power of both the DF-GLS and P_T statistics are sensitive to the choice of AR parameter p. This suggests that investigations of alternative methods for selecting p might result in improved performance of these two tests.

6. Conclusions and Extensions

Although their asymptotic representations might seem complicated, it should be stressed that the P_T, P_T^μ and P_T^τ statistics can be computed in three simple steps. First, the sum of squared residuals is computed from the GLS regression of y_t onto z_t under the null α=1. Second, y_t is regressed onto z_t using GLS under the alternative $\bar{\alpha} = 1+\bar{c}/T$, where $\bar{c} = -7$ for the demeaned case and $\bar{c} = -13.5$ for the detrended case. Third, the AR spectral estimator is computed using (20) after running the "Dickey-Fuller" regression (21). The test statistic is then constructed according to (15). The DF-GLS$^{\mu}$ and DF-GLS$^{\tau}$ statistics are even simpler to compute: after GLS-demeaning or GLS-detrending with $\bar{\alpha} = 1+\bar{c}/T$, the modified Dickey-Fuller t-statistic is computed by running the regression (17).

The numerical finding that, as a practical matter, the power functions of the $P_T(.5)$ and DF-GLS tests effectively lie on the asymptotic Gaussian power envelopes in both the demeaned and linearly detrended cases indicates that there is little room for improvement upon these tests, as measured in terms of asymptotic power, at least if v_t obeys a Gaussian AR(p). In this sense, these tests can be thought of as approximately asymptotically UMP in the demeaned case and UMPI in the detrended case.

The Monte Carlo results suggest that, on balance, the DF-GLS$^{\mu}$ and DF-GLS$^{\tau}$ statistics have the best overall performance in terms of size and size-adjusted power. The results also confirm previous findings that test performance in finite samples is sensitive to the choice of spectral estimator. This suggests that more work on the small-sample properties of these statistics, such as the recent investigations by Perron (1991) and Nabeya and Perron (1991) of estimators of the AR root, and in particular on the choice of spectral estimator, could result in improvements in both size and size-adjusted power.

Appendix

Proofs of Theorems

Proof of Theorem 1.

(a) To show the $o_p(1)$ result in the text preceding (5), use $T^{-1}\sum_{t=p+2}^{T}(\Delta u_t^{\dagger})^2 =$ $T^{-1}\sum_{t=p+2}^{T}(\epsilon_t + cT^{-1}u_{t-1}^{\dagger})^2 = T^{-1}\sum_{t=p+2}^{T}\epsilon_t^2 + 2cT^{-2}\sum_{t=p+2}^{T}u_{t-1}^{\dagger}\epsilon_t + c^2T^{-3}\sum_{t=p+2}^{T}(u_{t-1}^{\dagger})^2$. Because $(T^{-1}\sum_{t=p+2}^{T}u_{t-1}^{\dagger}\epsilon_t, T^{-2}\sum_{t=p+2}^{T}(u_{t-1}^{\dagger})^2) \Rightarrow \sigma^2(\int W_c dW, \int W_c^2)$, the second two terms in the expression converge to zero in probability so $T^{-1}\sum_{t=p+2}^{T}(\Delta u_t^{\dagger})^2 = T^{-1}\sum_{t=p+2}^{T}\epsilon_t^2 + o_p(1) = \sigma^2 + o_p(1)$. As discussed in the text, the result then follows from the Neyman-Pearson Lemma and the continuous mapping theorem.

(b) To show that (5) and (7) define asymptotically equivalent regions, it suffices to show that: (i) $T^{-2}\sum_{t=p+2}^{T}(y_{t-1}^{\dagger 2} - u_{t-1}^{\dagger 2}) \xrightarrow{p} 0$; (ii) $T^{-1}(y_T^{\dagger 2} - u_T^{\dagger 2}) \xrightarrow{p} 0$; and (iii) $\hat{\sigma}^{\dagger 2} - \sigma^2 \xrightarrow{p} 0$. But $|y_t^{\dagger 2} - u_t^{\dagger 2}| = |d_t^{\dagger 2} + 2u_t^{\dagger}d_t^{\dagger}| \le (\max_{t=p+1,\ldots,T}|d_t^{\dagger}|)^2 + 2|u_t^{\dagger}|\max_{t=p+1,\ldots,T}|d_t^{\dagger}|$, $t = p+1, \ldots, T$. Moreover, since $|d_t^{\dagger}| < (1 + |\rho_1| + \cdots + |\rho_p|)\max_{s=t-p,\ldots,t}|d_s|$, condition A implies $T^{-\frac{1}{2}}\max_{t=p+1,\ldots,T}|d_t^{\dagger}| \to 0$. Since $T^{-\frac{1}{2}}u_T^{\dagger} \Rightarrow W_c(1)$ and $T^{-3/2}\sum_{t=p+2}^{T}|u_{t-1}^{\dagger}| \Rightarrow$ $\int|W_c|$, (i) and (ii) follow. As for (iii), note that, for fixed c, $T^{-1}\sum_{t=p+2}^{T}(\Delta u_t^{\dagger})^2 \xrightarrow{p} \sigma^2$ from the proof of part (a). But

$$T^{-1}|\sum_{t=p+2}^{T}(\Delta y_t^{\dagger})^2 - \sum_{t=p+2}^{T}(\Delta u_t^{\dagger})^2| = |2T^{-1}\sum_{t=p+2}^{T}\Delta u_t^{\dagger}\Delta d_t^{\dagger} + T^{-1}\sum_{t=p+2}^{T}(\Delta d_t^{\dagger})^2|$$
$$\le 2[T^{-1}\sum_{t=p+2}^{T}(\Delta u_t^{\dagger})^2]^{\frac{1}{2}}[T^{-1}\sum_{t=p+2}^{T}(\Delta d_t^{\dagger})^2]^{\frac{1}{2}} + T^{-1}\sum_{t=p+2}^{T}(\Delta d_t^{\dagger})^2.$$

Thus, because $T^{-1}\sum_{t=p+2}^{T}(\Delta d_t^{\dagger})^2 \to 0$ by condition A, (iii) follows. □

Our proofs of theorems 2-4 rely on two preliminary lemmas. The first provides an alternative expression for the statistic M_T defined in (12) in terms of $\hat{\beta}$-β and the GLS-detrended series

$y_t^d = y_t - \hat{\beta}'z_t$. The second lemma provides asymptotic representations for $\hat{\beta}$ and the corresponding detrended series y_t^d for general z_t, as long as z_t includes a constant. This includes the two leading cases $\beta'z_t = \beta_0$ and $\beta'z_t = \beta_0 + \beta_1 t$.

Lemma A1.

(a) If $d_t = 0$, then $T(M_T - \bar{\alpha}) = [\bar{c}^2 T^{-2} \sum_{t=2}^T y_{t-1}^2 - \bar{c}T^{-1} y_T^2]/\hat{\sigma}^2$.

(b) if $d_t = \beta'z_t$ where the first component of z_t is equal to unity, then

$T(M_T - \bar{\alpha}) = (B_{1T} + \bar{\alpha}B_{2T} + B_{3T})/\hat{\sigma}^2$, where

$B_{1T} = (y_1^d)^2 = (\hat{\beta}-\beta)'z_1 z_1'(\hat{\beta}-\beta)$,

$B_{2T} = (\hat{\beta}-\beta)'(\sum_{t=2}^T \Delta z_t \Delta z_t')(\hat{\beta}-\beta)$, and

$B_{3T} = \bar{c}^2 T^{-2} \sum_{t=2}^T (y_{t-1}^d)^2 - \bar{c}T^{-1}[(y_T^d)^2 - (y_1^d)^2]$.

Proof of lemma A1.

(a) Write $T(M_T - \bar{\alpha}) = T(\tilde{\sigma}^2 - \hat{\sigma}^2)/\hat{\sigma}^2 - \bar{c}$. When $d_t = 0$, $T(\tilde{\sigma}^2 - \hat{\sigma}^2) = \bar{c}^2 T^{-2} \sum_{t=2}^T y_{t-1}^2 - 2\bar{c}T^{-1} \sum_{t=2}^T y_{t-1}\Delta y_t$ and $T\hat{\sigma}^2 = y_1^2 + \sum_{t=2}^T (\Delta y_t)^2$. The result then follows from the identity, $y_T^2 - y_1^2 = 2\sum_{t=2}^T y_{t-1}\Delta y_t + \sum_{t=2}^T (\Delta y_t)^2$.

(b) If the first component of z_t is unity, then $\hat{z}_1$ has first component equal to one and $\hat{z}_t$ (t>1) has first component equal to zero. Thus, by the orthogonality conditions of least squares, $\hat{e}_1 = y_1 - \hat{\beta}'z_1 = 0$ and $\sum_{t=2}^T \Delta z_t \hat{e}_t = 0$. Hence $T\hat{\sigma}^2$ can be written as,

$$T\hat{\sigma}^2 = \sum_{t=2}^T (\Delta y_t - \hat{\beta}'\Delta z_t)^2 = \sum_{t=2}^T (\Delta y_t - \beta'\Delta z_t)^2 - (\hat{\beta}-\beta)'(\sum_{t=2}^T \Delta z_t \Delta z_t')(\hat{\beta}-\beta).$$

The result follows from straightforward algebra. □

Lemma A2.

Suppose y_t is generated according to (1) where d_t is given by (9), v_t satisfies condition B, u_1 satisfies condition C, and $\alpha = 1+c/T$, where c is a fixed constant.

(a) If k = 0, then $T^{\frac{1}{2}}(\hat{\beta}_0-\beta_0-u_1) => \omega[\bar{c}^2\int W_c - \bar{c}W_c(1)]$ and $T^{-\frac{1}{2}}y^{\mu}_{[T\cdot]} => \omega W_c(\cdot)$.

(b) If k = 1, then $T^{\frac{1}{2}}(\hat{\beta}_0-\beta_0-u_1, \hat{\beta}_1-\beta_1) => \omega(\beta_0^*, \beta_1^*)$ and $T^{-\frac{1}{2}}y^{\tau}_{[T\cdot]} => \omega V_c(\cdot)$, where $V_c(s) = W_c(s)-\beta_1^* s$ and

$$\beta_0^* = \bar{c}^2\int W_c - \bar{c}W_c(1) - (1-\bar{c}+\tfrac{1}{2}\bar{c}^2)[\lambda W_c(1) + (1-\lambda)3\int rW_c(r)dr],$$

$$\beta_1^* = \lambda W_c(1) + (1-\lambda)3\int rW_c(r)dr,$$

and $\lambda = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$. Furthermore, $T^{\frac{1}{2}}(\hat{\beta}_1-\beta_1) => \omega(1-\lambda)[3\int rW_c(r)dr - W_c(1)]$.

Proof of lemma A2.

(a) Under condition C with $\alpha = 1+c/T$, $T^{-\frac{1}{2}}u_{[T\cdot]} => \omega W_c(\cdot)$. By direct calculation,

$$\hat{\beta}_0 = [y_1 - \bar{c}T^{-1}\textstyle\sum_{t=2}^{T}(\Delta y_t-\bar{c}T^{-1}y_{t-1})]/[1 + \bar{c}^2T^{-2}(T-1)]$$

$$= \beta_0 + [u_1 - \bar{c}T^{-1}\textstyle\sum_{t=2}^{T}(\Delta u_t-\bar{c}T^{-1}u_{t-1})]/[1 + \bar{c}^2T^{-2}(T-1)]$$

$$= \beta_0 + u_1 + \{\bar{c}^2T^{-2}\textstyle\sum_{t=2}^{T}u_{t-1} - \bar{c}T^{-1}(u_T-u_1) - \bar{c}^2T^{-2}(T-1)u_1\}/[1 + \bar{c}^2T^{-2}(T-1)]\,. \tag{A.1}$$

Thus, $T^{\frac{1}{2}}(\hat{\beta}_0-\beta_0-u_1) => \omega[\bar{c}^2\int W_c - \bar{c}W_c(1)]$. It follows that $T^{-\frac{1}{2}}(\hat{\beta}_0-\beta_0) \xrightarrow{p} 0$ so $T^{-\frac{1}{2}}y^{\mu}_{[T\cdot]} = T^{-\frac{1}{2}}u_{[T\cdot]} - T^{-\frac{1}{2}}(\hat{\beta}_0-\beta_0) => \omega W_c(\cdot)$.

(b) Defining $(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_T) = (u_1, (1-\bar{\alpha}L)u_2, \ldots, (1-\bar{\alpha}L)u_T)$, we can write the least squares normal equations as $\sum_{t=1}^{T}\tilde{z}_t\tilde{z}_t'(\hat{\beta}-\beta) = \sum_{t=1}^{T}\tilde{z}_t\tilde{u}_t$. After multiplying the first equation by $T^{\frac{1}{2}}$ and the second equation by $T^{-\frac{1}{2}}$, performing the summations, and grouping terms according to magnitude, we can rewrite this system as the pair of equations,

$$T^{\frac{1}{2}}(\hat{\beta}_0-\beta_0) + (1-\bar{c}+\tfrac{1}{2}\bar{c}^2)T^{\frac{1}{2}}(\hat{\beta}_1-\beta_1) = T^{\frac{1}{2}}u_1 + \bar{c}^2T^{-3/2}\textstyle\sum_{t=2}^{T}u_{t-1} - \bar{c}T^{-\frac{1}{2}}u_T + o_p(1) \tag{A.2}$$

$$(1-\bar{c}+\bar{c}^2/3)T^{\frac{1}{2}}(\hat{\beta}_1-\beta_1) = T^{-\frac{1}{2}}u_T - \bar{c}T^{-3/2}\textstyle\sum_{t=2}^{T}(u_{t-1}+t\Delta u_t) + \bar{c}^2T^{-5/2}\textstyle\sum_{t=2}^{T}tu_{t-1} + o_p(1). \tag{A.3}$$

But using the identity $\sum_{t=2}^{T} t\Delta u_t = \sum_{t=2}^{T} tu_t - \sum_{t=2}^{T}(t-1+1)u_{t-1} = Tu_T - u_1 - \sum_{t=2}^{T} u_{t-1}$, equation (A.3) can be rewritten as,

$$(A.4) \qquad (1-\bar{c}+\bar{c}^2/3)T^{½}(\hat{\beta}_1-\beta_1) = (1-\bar{c})T^{-½}u_T + \bar{c}^2 T^{-5/2}\sum_{t=2}^{T} tu_{t-1} + o_p(1).$$

Employing the continuous mapping theorem, we obtain the limiting representations β_0^* and β_1^*. Because $T^{½}(\hat{\beta}_1-\beta_1) \Rightarrow \omega\beta_1^*$ with $\bar{c}=0$, the representation for $T^{½}(\hat{\beta}_1-\hat{\beta}_1)$ follows immediately. □

Except for asymptotically negligible initial terms, the likelihood function for $y_t^\dagger$, $t = p+1, \ldots, T$, is the same as the likelihood function for y_t, $t = 1, \ldots, T$ when $p = 0$ and $\omega = \sigma$. Hence, theorem 2 is essentially a special case of theorem 3.

Proof of Theorem 3.

(a) When there is no deterministic component, $y_t = u_t$. But, under local-to-unity asymptotics where $c = T(\alpha-1)$ and $\bar{c} = T(\bar{\alpha}-1)$ are fixed as T tends to infinity, $T^{-½}u_{[\cdot]} \Rightarrow \omega W_c(\cdot)$. Thus, by the representation given in lemma A1(a) and the continuous mapping theorem,

$$\hat{\sigma}^2 T(M_T-\bar{\alpha}) \Rightarrow \omega^2[\bar{c}^2\int W_c^2 - \bar{c}W_c(1)^2] = \omega^2\Psi(c,\bar{c}).$$

Also,

$$\hat{\sigma}^2 = T^{-1}[u_1^2 + \sum_{t=2}^{T}(\Delta u_t)^2] = T^{-1}[u_1^2 + \sum_{t=2}^{T}(v_t + cT^{-1}u_{t-1})^2]$$
$$= T^{-1}u_1^2 + T^{-1}\sum_{t=2}^{T} v_t^2 + 2cT^{-2}\sum_{t=2}^{T} u_{t-1}v_t + c^2T^{-3}\sum_{t=2}^{T} u_{t-1}^2 .$$

The final two terms in this expression converge in probability to zero. Under condition C, $T^{-1}u_1^2 \xrightarrow{p} 0$ and, under condition B, $T^{-1}\sum_{t=2}^{T} v_t^2 \xrightarrow{p} \gamma_v(0)$. Thus $T(M_T-\bar{\alpha}) \Rightarrow (\omega^2/\gamma_v(0))\Psi(c,\bar{c})$.

(b) In the representation $T(M_T-\bar{\alpha}) = (B_{1T} + \bar{\alpha}B_{2T} + B_{3T})/\hat{\sigma}^2$ given in lemma A1(b), note that when k=0, $B_{1T} = (\hat{\beta}_0-\beta_0-u_1)^2$ and $B_{2T} = 0$. The result in lemma A2(a) implies that $B_{1T} \xrightarrow{p} 0$ and that $T^{-\frac{1}{2}}y^{\mu}_{[T\cdot]} \Rightarrow \omega W_c(\cdot)$. Because $T^{-\frac{1}{2}}y_1^d \xrightarrow{p} 0$, it follows from the continuous mapping theorem that $\hat{\sigma}^2T(M_T-\bar{\alpha}) \Rightarrow \omega^2\Psi(c,\bar{c})$. From the proof of lemma A1(b), we have $\hat{\sigma}^2 = T^{-1}\sum_{t=2}^{T}(\Delta y_t-\hat{\beta}'\Delta z_t)^2 = T^{-1}\sum_{t=2}^{T}(\Delta u_t)^2$ (where the second equality follows from $\Delta z_t = 0$, $t = 2, \ldots, T$), which by the argument in part (a) converges in probability to $\gamma_v(0)$.

(c) Again we use the representation $T(M_T-\bar{\alpha}) = (B_{1T} + \bar{\alpha}B_{2T} + B_{3T})/\hat{\sigma}^2$ given in lemma A1(b) and the limiting distributions given in lemma A2. For the linear time trend case, we have $B_{1T} = (\hat{\beta}_0-\beta_0+\hat{\beta}_1-\beta_1-u_1)^2 \xrightarrow{p} 0$ from lemma A2(b). Moreover the 2x2 matrix $\sum_{t=2}^{T}\Delta z_t\Delta z_t'$ is zero except for the (2,2) element, which is T-1. Thus, from lemma A2(b), $B_{2T} = (T-1)(\hat{\beta}_1-\beta_1)^2 \Rightarrow \omega^2(1-\lambda)^2[3\int rW_c(r)dr - W_c(1)]^2$. Finally,

$$B_{3T} \Rightarrow \omega^2\bar{c}^2\int\{W_c(s) - s[\lambda W_c(1) + (1-\lambda)3\int rW_c(r)dr]\}^2ds - \omega^2\bar{c}(1-\lambda)^2[3\int rW_c(r)dr - W_c(1)]^2.$$

Collecting terms, we find $\hat{\sigma}^2T(M_T-\bar{\alpha}) \Rightarrow \omega^2\Psi^{\tau}(c,\bar{c})$, where

$$\begin{aligned}\Psi^{\tau}(c,\bar{c}) &= \bar{c}^2\int V_c^2 + (1-\bar{c})V_c(1)^2 \\ &= \bar{c}^2\{\lambda\int[W_c(s)-sW_c(1)]^2ds + (1-\lambda)\int[W_c(s) - 3s\int rW_c(r)dr]^2ds\}.\end{aligned}$$

where $V_c(s) = W_c(s)-s\beta_1^*$. Finally,

$$\begin{aligned}\hat{\sigma}^2 &= T^{-1}\sum_{t=2}^{T}(\Delta u_t)^2 - 2T^{-1}\sum_{t=2}^{T}\Delta u_t\Delta z_t'(\hat{\beta}-\beta) + (\hat{\beta}-\beta)'(T^{-1}\sum_{t=2}^{T}\Delta z_t\Delta z_t')(\hat{\beta}-\beta) \\ &= T^{-1}\sum_{t=2}^{T}(\Delta u_t)^2 - 2T^{-1}(u_T-u_1)(\hat{\beta}_1-\beta_1) + (\hat{\beta}_1-\beta_1)^2(T-1)/T.\end{aligned}$$

The last two terms converge in probability to zero so again $\hat{\sigma}^2 \xrightarrow{p} \gamma_v(0)$. □

Proof of Theorem 4.

(i) $d_t = 0$ case. From lemma A1(a), we have $P_T = [\bar{c}^2 T^{-2}\sum_{t=2}^{T} u_{t-1}^2 - \bar{c}T^{-1}u_T^2]/\hat{\omega}^2$ since, with no deterministic terms, $y_t = u_t$. Under the fixed alternative, $u_T = O_p(1)$ and $T^{-1}\sum_{t=2}^{T} u_{t-1}^2$ converges in probability to a constant. Under condition D, $\hat{\omega}^2$ converges to a nonzero constant. Hence, $P_T \xrightarrow{p} 0$.

(ii) $d_t = \beta_0$ case. From lemma A1(b) and because $B_{2T} = 0$ when k=0, $P_T^{\mu} = (B_{1T}+B_{3T})/\hat{\omega}^2$. From (A.1), $B_{1T} = (\hat{\beta}_0-\beta_0-u_1)^2 = [\bar{c}^2 T^{-2}\sum_{t=2}^{T} u_{t-1} - \bar{c}T^{-1}(u_T-u_1)]^2 + O_p(T^{-1})$. Since $T^{-1}\sum_{t=2}^{T} u_{t-1}$, u_T, and u_1 are stochastically bounded, $B_{1T} \xrightarrow{p} 0$.

To show that $B_{3T} \xrightarrow{p} 0$, it is sufficient to show that $(T^{-1/2}y_1^{\mu}, T^{-1/2}y_T^{\mu}, T^{-2}\sum_{t=2}^{T}(y_{t-1}^{\mu})^2) \xrightarrow{p} 0$. Now $T^{-1/2}y_t^{\mu} = T^{-1/2}u_t - T^{-1/2}(\hat{\beta}_0-\beta_0)$, $t = 1, \ldots, T$. In addition $T^{-1/2}(\hat{\beta}_0-\beta_0) = T^{-1/2}(\beta_0-\hat{\beta}_0) + T^{-1/2}(\hat{\beta}_0-\beta_0) = (T^{-1}B_{1T})^{1/2} + T^{-1/2}u_1 \xrightarrow{p} 0$ from the previous argument concerning B_{1T}. Because u_t is stochastically bounded, we thus have $(T^{-1/2}y_1^{\mu}, T^{-1/2}y_T^{\mu}) \xrightarrow{p} 0$. Turning to the final term, $T^{-2}\sum_{t=2}^{T}(y_{t-1}^{\mu})^2 = T^{-2}\sum_{t=2}^{T}u_t^2 + 2(\hat{\beta}_0-\beta_0)T^{-2}\sum_{t=2}^{T}u_t + (T-1)T^{-2}(\hat{\beta}_0-\beta_0)^2 \xrightarrow{p} 0$ from the previous results; thus $B_{3T} \xrightarrow{p} 0$. Because $\hat{\omega}^2 \xrightarrow{p} d > 0$ by condition D, it follows that $P_T^{\mu} \xrightarrow{p} 0$.

(iii) $d_t = \beta_0 + \beta_1 t$ case. The proof for the linear trend case follows that of the k=0 case. Again we write $P_T^{\tau} = (B_{1T} + \bar{\alpha}B_{2T} + B_{3T})/\hat{\omega}^2$. It is readily verified that the "$o_p(1)$" terms in (A.2) and (A.4) in the proof of lemma A2 are $o_p(1)$ under a fixed alternative as well as under the local-to-unity assumption of lemma A2. From (A.2) and (A.4) we see that $T^{1/2}(\hat{\beta}_0-\beta_0-u_1)$, $T^{1/2}(\hat{\beta}_1-\beta_1)$, and $T^{1/2}(\hat{\beta}_1-\hat{\beta}_1)$ all converge in probability to zero for fixed $|\alpha|<1$ since u_t is stochastically bounded. Thus $B_{1T} = (u_1+\beta_0-\hat{\beta}_0+\beta_1-\hat{\beta}_1)^2 \xrightarrow{p} 0$ and $B_{2T} = (T-1)(\hat{\beta}_1-\beta_1)^2 \xrightarrow{p} 0$. Finally, $T^{-1/2}y_t^{\tau} = T^{-1/2}u_t - T^{-1/2}(\hat{\beta}_0-\beta_0) - T^{1/2}(\hat{\beta}_1-\beta_1)t/T$, so $(T^{-1/2}y_1^{\tau}, T^{-1/2}y_T^{\tau}, T^{-2}\sum_{t=2}^{T}(y_{t-1}^{\tau})^2) \xrightarrow{p} 0$. Thus $B_{3T} \xrightarrow{p} 0$ and, with condition D, $P_T^{\tau} \xrightarrow{p} 0$. □

Footnotes

1. More specific assumptions on v_t which imply condition B are provided by Herndorff (1984) for the case that v_t is a mixingale (see Phillips (1987a) for a discussion) and by Hall and Heyde (1980) and Phillips and Solo (1992) for the case that v_t is a linear process. A condition like assumption C, which ensures that the initial condition is $O_p(1)$, is standard in this literature. Assumptions with the same effect are made by, for example, Chan and Wei (1987), Phillips (1987b), Perron (1989), and Nabeya and Tanaka (1990).

2. Dufour and King's (1991) theorem 5 applies to their invariance group G2. For the null $\alpha = 1$ their G2 is equivalent to the invariance group for the transformation $y_t \rightarrow ay_t + b'z_t$ in the case that $z_t = (1, t)$. If $z_t = 1$ then the two invariance groups are equal only if Dufour and King's (1991) γ_{k+1} is restricted to be zero, although the MPI test (12) nonetheless obtains from their theorem 5 and in any event is readily derived from first principles in the $z_t = 1$ case. Also note that Dufour and King's (1991) augmented regression under the alternative (with projection matrix $\bar{M}$ in their notation) reduces to our GLS equation (14) when a constant is included in the regression and in their notation $d_1 = 1$ (which is equivalent to $u_0 = 0$ here); see the discussion preceding their equation (16).

3. Inder and Grose (1990) proposed an alternative generalization of the Dufour-King statistics to the AR(p+1) case. Their procedure involves estimating $\rho(L)$ under the null hypothesis and using the $\hat{\rho}(L)$-filtered series (an estimate of $u_t^{\dagger}$) to construct the M_T statistic. They do not, however, provide any distribution theory, either finite-sample or asymptotic, they provide no results concerning efficiency or consistency, and they only consider the no-deterministic case d_t=0. Their procedure represents an alternative which merits further investigation.

4. Monte Carlo standard errors are less than .0013. In his comparison of three methods for evaluating large-sample local-to-unity distributions, Chan (1988) concluded that the Monte Carlo method was preferred. Presumably greater accuracy could be obtained by extending the techniques in Perron (1989) or Nabeya and Tanaka (1990) to the statistics here, but such an extension is beyond

the scope of this paper.

5. One motivation for the choice of 50% as the tangency point is that it is about such alternatives that previous evidence is likely to be the least conclusive, so it is here that the researcher would like to maximize power. In the unit roots problem, DeJong, Nankervis, Savin and Whiteman (1992) examined the alternative $\alpha = .85$, which they considered representative of the stationary alternatives of practical interest. With T =100, this corresponds to $\bar{c} = -15$, which is close to the value of $\bar{c}$ for the POI(.5) test in the detrended case. There is precedent in the POI testing literature for choosing 50% tangency points (e.g. Shively (1988)). An alternative choice, used by Davies (1969) and Shively (1988), is tangency at power of 80%. King (1988, section 3.3) lists several alternative approaches for selecting among POI tests (also see the comment on King (1988) by Potscher (1988)). Although this ambiguity about the choice of tangency point is present in theory, as discussed in this section it turns out to have little practical impact in our problem, at least asymptotically.

6. In personal communication, Pierre Perron suggested we examine lag length selection using sequential likelihood ratio statistics. The results in tables 2 and 3 were also produced for this selector, with maximal lag length 8, minimal lag length 1, and 10% rejection level. In most cases, the sizes of these tests were slightly larger, and the power slightly less, than with the AR(BIC) selector, for both the demeaned and the detrended statistics. These tendencies were more pronounced for P_T and $\hat{\rho}$ than $\hat{\tau}$ and DF-GLS. Notably, however, the power of the LR selector appears to improve the size-adjusted power of the DF-GLS$^{\tau}$ statistic relative to BIC, at least for small values of θ.

7. Replace condition C by the assumption that $T^{-1/2}u_0 \xrightarrow{p} \eta$, where η can be random or fixed, and consider the demeaned case. Then instead of the limit given in lemma A2(a), $T^{-1/2}y^{\mu}_{[T\cdot]} \Rightarrow M_c(\cdot)$, where $M_c(s) = \omega W_c(s)+(\exp(cs)-1)\eta$. If u_0 is drawn from its unconditional distribution and $v_t = \epsilon_t$, ϵ_t i.i.d. $N(0,\sigma^2)$, then $\eta \sim N(0, -\sigma^2/2c)$ and η and W_c are independent. Then $\omega = \sigma$ and the local-to-unity representations of the P^{μ}_T and DF-GLS$^{\mu}$ statistics are obtained by replacing W_c by M_c in the expression for $\Psi(c,\bar{c})$ and in (18a), respectively.

References

Anderson, T.W. (1948), "On the Theory of Testing Serial Correlation," *Skandinavisk Aktuariedskrift*, 31, 88-116.

Andrews, D.W.K. (1991), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.

Baillie, R.T. and T. Bollerslev (1989), "The Message in Daily Exchange Rates: A Conditional Variance Tale", *Journal of Business and Economic Statistics*, 7, 297-305.

Berk, K.N. (1974), "Consistent Autoregressive Spectral Estimates," *Annals of Statistics*, 2, 489-502.

Bhargava, A. (1986), "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies*, 53, no. 174, 369-384.

Bobkoski, M.J. (1983), *Hypothesis Testing in Nonstationary Time Series*, unpublished Ph.D. thesis, Department of Statistics, University of Wisconsin.

Cavanagh, C. (1985), "Roots local to unity," manuscript, Department of Economics, Harvard University.

Chan, N.H., (1988), "On The Parameter Inference for Nearly Nonstationary Time Series," *Journal of the American Statistical Association* 83, no. 403, 857-862.

Chan, N.H. and C.Z. Wei, (1987), "Asymptotic Inference For Nearly Nonstationary AR(1) Processes," *Annals of Statistics* 15, 1050-1063.

DeJong, D.N. J.C. Nankervis, N.E. Savin, and C.H. Whiteman (1992), "The Power Problems of Unit Root Tests for Time Series with Autoregressive Errors," forthcoming, *Journal of Econometrics.*

Davies, R.B. (1969), "Beta-optimal Tests and an Application to the Summary Evaluation of Experiments," *Journal of the Royal Statistical Society B*, 31, 524-538.

Dickey, D.A. and Fuller, W.A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, no. 366, 427-431.

Dickey, D.A., D.P. Hasza, and W.A. Fuller (1984), "Testing for Unit Roots in Seasonal Time Series," *Journal of the American Statistical Association*, 79, no. 386, 355-367.

Dufour, J.-M., and M.L. King (1991), "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors," *Journal of Econometrics*, 47, 115-143.

Evans, G.B.A. and N.E. Savin (1981), "Testing for Unit Roots: I," *Econometrica*, 49, 753-779.

Evans, G.B.A. and N.E. Savin (1984), "Testing for Unit Roots: II," *Econometrica*, 52, 1241-1260.

Fuller, W.A. (1976), *Introduction to Statistical Time Series.* New York: Wiley.

Hall, P. and C.C. Heyde (1980), *Martingale Limit Theory and its Applications.* New York: Academic Press.

Herrndorf, N. A. (1984), "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *Annals of Probability*, 12, 141-153.

Inder, B. and S. Grose (1990), "The Search for a Robust and Powerful Test for Unit Roots," notes, Department of Econometrics, Monash University, Australia.

King, M.L. (1980), "Robust Tests for Spherical Symmetry and their Application to Least Squares Regression," *Annals of Statistics*, 8, 1265-1271.

King, M.L. (1988), "Towards a Theory of Point Optimal Testing", *Econometric Reviews*, 6, 169-218.

Lumsdaine, R.L. (1991), *Essays on Time Series Econometrics*, Ph.D. Dissertation, Department of Economics, Harvard University.

Nabeya, S. and P. Perron (1991), "Local Asymptotic Distributions Related to the AR(1) Model with Dependent Errors," Econometric Research Program Memorandum No. 362, Princeton University.

Nabeya, S. and K. Tanaka (1990), "Limiting Powers of Unit Root Tests in Time Series Regression," *Journal of Econometrics*, 46, 247- 271.

Nyblom, J. and T. Makelainen (1983), "Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model," *Journal of the American Statistical Association*, 78, 856-864.

Perron, P. (1989), "The Calculation of the Limiting Distribution of the Least-Squares Estimator in a Near-Integrated Model," *Econometric Theory*, 5, 241-255.

Perron, P. (1991), "The Adequacy of Asymptotic Approximations in the Near-Integrated Model with Dependent Errors," manuscript, Princeton University.

Phillips, P.C.B. (1987a), "Time Series Regression with a Unit Root," *Econometrica*, 55, 277-302.

Phillips, P.C.B. (1987b), "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535-547.

Phillips, P.C.B. and S. Ouliaris (1990), "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165-94.

Phillips, P.C.B and P. Perron (1988) "Testing for a Unit Roots in a Time Series Regression," *Biometrika*, 75, 335-346.

Phillips, P.C.B. and V. Solo (1992), "Asymptotics for Linear Processes," *Annals of Statistics*, 20, no. 2, 971-1001.

Potscher, B.M. (1988), "Towards a Theory of Point Optimal Testing - Comment," *Econometric Reviews*, 6, 235-239.

Said, S.E. and D.A. Dickey (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599-608.

Saikkonen, P. and R. Luukkonen (1992), "Point Optimal Tests for the Moving Average Unit Root Hypothesis", manuscript, University of Helsinki.

Sargan, J.D. and A. Bhargava (1983), "Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," *Econometrica*, 51, 153-174.

Schmidt, P. and P.C.B Phillips (1992), "Testing for Unit Root in the Presence of Deterministic Trends," *Oxford Bulletin of Economics and Statistics*, forthcoming.

Schwert, G.W. (1989), "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147-160.

Shively, T.S. (1988), "An Exact Text for A Stochastic Coefficient in A Time Series Regression Model", *Journal of Time Series Analysis*, 9, 81-88.

Stock, J.H. (1988), "A Class of Tests for Integration and Cointegration," manuscript, Kennedy School of Government, Harvard University.

Stock, J.H. (1991), "Confidence Intervals for the Largest Autoregressive Root in U.S. Economic Time Series," *Journal of Monetary Economics*, 28, no. 3, 435-460.

Stock, J.H. (1992), "Unit Roots and Trend Breaks in Econometrics", prepared for *The Handbook of Econometrics, vol. IV*.