Econometrica, Vol. 64, No. 4 (July, 1996), 813–836

EFFICIENT TESTS FOR AN AUTOREGRESSIVE UNIT ROOT

BY GRAHAM ELLIOTT, THOMAS J. ROTHENBERG, AND JAMES H. STOCK[1]

The asymptotic power envelope is derived for point-optimal tests of a unit root in the autoregressive representation of a Gaussian time series under various trend specifications. We propose a family of tests whose asymptotic power functions are tangent to the power envelope at one point and are never far below the envelope. When the series has no deterministic component, some previously proposed tests are shown to be asymptotically equivalent to members of this family. When the series has an unknown mean or linear trend, commonly used tests are found to be dominated by members of the family of point-optimal invariant tests. We propose a modified version of the Dickey-Fuller t test which has substantially improved power when an unknown mean or trend is present. A Monte Carlo experiment indicates that the modified test works well in small samples.

KEYWORDS: Power envelope, point optimal tests, nonstationarity, Ornstein-Uhlenbeck processes.

1. INTRODUCTION

FOLLOWING THE SEMINAL WORK of Fuller (1976) and Dickey and Fuller (1979), econometricians have developed numerous alternative procedures for testing the hypothesis that a univariate time series is integrated of order one against the hypothesis that it is integrated of order zero. The procedures typically are based on second-order sample moments, but employ various testing principles and a variety of methods to eliminate nuisance parameters. Banerjee et al. (1993) and Stock (1994) survey many of the most popular of these tests. Although numerical calculations (e.g., Nabeya and Tanaka (1990)) suggest that the power functions for the tests can differ substantially, no general optimality theory has been developed. In particular, there are few general results (even asymptotic) concerning the relative merits of the competing testing principles and of the various methods for eliminating trend parameters.

Employing a model common in the previous literature, we assume that the data $y_1, \ldots, y_T$ were generated as

$$(1)\qquad \begin{aligned} y_t &= d_t + u_t \\ u_t &= \alpha u_{t-1} + v_t \end{aligned} \qquad (t = 1, \ldots, T),$$

where $\{d_t\}$ is a deterministic component and $\{v_t\}$ is an unobserved stationary zero-mean error process whose spectral density function is positive at zero frequency. Our interest is in the null hypothesis $\alpha = 1$ (which implies the y_t are integrated of order one) versus $|\alpha| < 1$ (which implies the y_t are integrated of

[1] 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 SES-91-22463.

order zero). Standard asymptotic testing theory, as surveyed for example in Engle (1984), is inapplicable since the data do not give rise to a locally asymptotic normal likelihood.[2] Nevertheless, it is possible to develop an asymptotic framework for comparing alternative tests for a unit root in this model. If the distribution of the data were otherwise known, the Neyman-Pearson Lemma gives us the best test against any given point alternative $\bar{\alpha}$. The power of this optimal test at alternative $\bar{\alpha}$, when plotted against $\bar{\alpha}$, defines the power envelope which is an upper bound for the power function of any test based on the same likelihood. Using large-sample approximations to simplify the analysis, we can then compare the asymptotic power functions of existing tests with this asymptotic bound. In practice, of course, the likelihood function will depend on additional nuisance parameters determining d_t, u_0, and the distribution of $\{v_t\}$. If there exist feasible tests with the same asymptotic power as the Neyman-Pearson point-optimal tests, the comparison will be appropriate in the nuisance parameter case as well.

When the observed time series is Gaussian with constant or slowly evolving deterministic component, we find that, although no test uniformly attains the asymptotic power bound, there exist tests with asymptotic power functions very close to the bound. Furthermore, these tests can be constructed without knowledge of any nuisance parameters. When the deterministic component contains a polynomial trend, no feasible test comes close to attaining the power bound derived under the assumption the trend parameters are known. Nevertheless, the Neyman-Pearson Lemma can still be employed to derive an asymptotic power bound for the natural family of tests that are invariant to the trend parameters. Again, there exist feasible invariant tests with asymptotic power functions very close to this bound, even when there are additional nuisance parameters determining the autocovariances of the v_t.

Our asymptotic power results have implications for tests commonly used in practice. In the case where there is no deterministic component, we find that the asymptotic power curve of the Dickey-Fuller t test virtually equals the bound when power is one-half and is never very far below. In the more relevant case where a deterministic mean or trend is present, power can be improved considerably over the standard Dickey-Fuller test by modifying the method employed to estimate the parameters characterizing the deterministic term.

Our approach is similar to that employed by Dufour and King (1991) in their analysis of exact point-optimal invariant tests in the normal AR(1) model. However, by employing local-to-unity asymptotic approximations, we are able to obtain simpler and more interpretable results that cover a much broader class of models. Saikkonen and Luukkonen (1993) apply a similar analysis in their study of asymptotically point-optimal invariant tests for a unit moving-average root.

[2] Using a different maintained model, Robinson (1994) develops a "standard" asymptotic theory of efficient tests for a unit root. This requires dropping the familiar autoregression framework and assuming, for example, a fractionally differenced process for the data.

2. THE ASYMPTOTIC GAUSSIAN POWER ENVELOPE

In this section we derive an upper bound to the asymptotic power function for tests of the hypothesis $\alpha = 1$ when the data are generated by (1) and the following condition is satisfied.

CONDITION A: *The stationary sequence* $\{v_t\}$ *has a strictly positive spectral density function; it has a moving average representation* $v_t = \sum_{j=0}^{\infty} \delta_j \eta_{t-j}$ *where the* η_t *are independent standard normal random variables and* $\sum_{j=0}^{\infty} j|\delta_j| < \infty$. *The initial* u_0 *is* 0 *and the* δ*'s are known.*

The unrealistic assumption of known u_0, δ's and error distribution is made so we may employ the Neyman-Pearson theory; in Section 3 we show that it may be dropped without any essential change. Our results, however, are quite sensitive to the nature of the deterministic components d_t. Section 2.1 considers the simplest case where the d_t are known. Section 2.2 examines the case where the d_t are "slowly evolving" and Section 2.3 examines the case where d_t is a linear combination of nonrandom trending regressors. Our purpose here is to derive the power bound; tests that might be used in practice are discussed later in the paper. All proofs are given in the Appendix.

2.1. *Known Deterministic Component*

When the d_t are known, u_t is observable and minus two times the log likelihood is (except for an additive constant) given by

$$L(\alpha) = [\Delta u - (\alpha - 1)u_{-1}]' \Sigma^{-1}[\Delta u - (\alpha - 1)u_{-1}] \tag{2}$$

where $\Delta u = (u_1, u_2 - u_1, \ldots, u_T - u_{T-1})'$, $u_{-1} = (0, u_1, \ldots, u_{T-1})'$, and Σ is the non-singular variance-covariance matrix for $v_1, \ldots, v_T$. By the Neyman-Pearson Lemma, the most powerful test of the null hypothesis that $\alpha = 1$ against the alternative that $\alpha = \bar{\alpha}$ rejects for small values of the likelihood ratio statistic $L(\bar{\alpha}) - L(1)$.

When the sample size is large, any reasonable test will have high power unless α is close to one. Thus, in obtaining large-sample approximations, it is natural to employ local-to-unity asymptotics where the parameter space is a shrinking neighborhood of unity as the sample size grows. In our case the appropriate rate to get nondegenerate distributions is T^{-1} so we reparameterize the model writing $c = T(\alpha - 1)$ and take c to be a constant when making limiting arguments. Cf. Chan and Wei (1987), Phillips and Perron (1988). Setting $\bar{c} \equiv T(\bar{\alpha} - 1)$, we can then write the likelihood ratio test statistic as

$$L(\bar{\alpha}) - L(1) = \bar{c}^2 T^{-2} u'_{-1} \Sigma^{-1} u_{-1} - 2\bar{c} T^{-1} u'_{-1} \Sigma^{-1} \Delta u. \tag{3}$$

For any given $\bar{c}$, rejecting when the linear combination (3) is small yields the most powerful test against the alternative that $c = \bar{c}$.

Note that $(T^{-2}u'_{-1}\Sigma^{-1}u_{-1}, T^{-1}u'_{-1}\Sigma^{-1}\Delta u)$ is a pair of minimally sufficient statistics for inference about α when the nuisance parameters are known. Furthermore, since the pair has a nondegenerate joint limiting distribution under local alternatives, the asymptotic minimal sufficient statistic also has dimension two. As a consequence, there exists no uniformly most powerful test of $\alpha = 1$ even in large samples. There is an infinite family of asymptotically admissible tests, indexed by $\bar{c}$, no one dominating the others for all c.

The limiting power functions for the family of Neyman-Pearson tests can be expressed conveniently in terms of stochastic integrals. Let $W_0(\cdot)$ represent standard Brownian motion defined on [0, 1] and let $W_c(\cdot)$ be the related diffusion process $W_c(t) = \int_0^t \exp\{c(t-s)\}\, dW_0(s)$ which satisfies the stochastic differential equation $dW_c(t) = cW_c(t)\,dt + dW_0(t)$ with initial condition $W_c(0) = 0$. In the Appendix we show that the local asymptotic power function for the test indexed by $\bar{c}$ when the significance level is ε is given by

$$\pi(c,\bar{c}) \equiv \Pr\left[\bar{c}^2 \int W_c^2 - \bar{c}W_c^2(1) < b(\bar{c})\right] \tag{4}$$

where $\int W_c^2 \equiv \int_0^1 W_c^2(t)\,dt$ and $b(\bar{c})$ satisfies $\Pr[\bar{c}^2 \int W_0^2 - \bar{c}W_0^2(1) < b(\bar{c})] = \varepsilon$. Because the test indexed by $\bar{c}$ is optimal against the alternative $c = \bar{c}$, the envelope power function for this family of point-optimal tests is $\Pi(c) = \pi(c, c)$.

2.2. *Slowly Evolving Deterministic Component*

Suppose the deterministic components satisfy the following condition.

CONDITION B (Slowly evolving trend): *The Δd_t are bounded with $T^{-1}\Sigma_{t=1}^T(\Delta d_t)^2 \to 0$ and $T^{-1/2}\max|d_t| \to 0$ as $T \to \infty$.*

This will automatically be satisfied if the d_t are constant. It will also be satisfied by a variety of smooth functions of time. These include low frequency sinusoids (e.g., $d_t = \cos(2\pi kt/T)$ for finite k); slowly increasing time trends (e.g., $d_t = \ln(t)$ or $d_t = t^{\delta}$ for $\delta < 1/2$); and step functions with finitely many jumps (e.g., $d_t = \beta_0$ when $t < t_0$ and $d_t = \beta_1$ when $t \geq t_0$). In the slowly evolving trend case, the random component of y_t dominates the deterministic component when t is large. It is tempting therefore to ignore the deterministic term when constructing the test statistic. In the Appendix we show that, if the d_t evolve slowly, replacing u_t by y_t when forming (3) has no effect on the asymptotic size or power of the Neyman-Pearson tests. Under Condition B, there is no efficiency loss from d_t being unknown.

2.3. *Trending Regressors*

The construction of a useful asymptotic power bound when the d_t are unknown and not slowly evolving is more complicated. Suppose, for example, the

d_t are modeled as a linear combination of a set of nonrandom regressors so that $d_t = \beta' z_t$ where β is a q-dimensional unknown parameter vector and the z_t are observed q-dimensional data vectors. Unless $\beta' z_t$ happens to satisfy Condition B, the power functions $\pi(c, \bar{c})$ derived in Section 2.1 will not be attainable by any feasible critical region.

In most applications, β is unrelated to α and the testing problem would be unchanged if y_t were replaced by $y_t + \bar{\beta}' z_t$ for arbitrary $\bar{\beta}$. It is therefore natural to restrict attention to the family of tests which are themselves invariant to this group of transformations. This approach is taken by Dufour and King (1991) who build on previous results in King (1980, 1988). (Their additional restriction of scale invariance is ignored here since the tests proposed in Section 3 satisfy this invariance automatically.) Defining the T-dimensional column vector y_a and the $T \times q$ matrix Z_a by

$$(5) \qquad \begin{aligned} y_a &= (y_1, y_2 - a y_1, \ldots, y_T - a y_{T-1})', \\ Z_a &= (z_1, z_2 - a z_1, \ldots, z_T - a z_{T-1})', \end{aligned}$$

we can rewrite (2) as

$$L(\alpha, \beta) = (y_\alpha - Z_\alpha \beta)' \Sigma^{-1} (y_\alpha - Z_\alpha \beta).$$

From the development in Lehmann (1959, p. 249), the most powerful invariant test of $\alpha = 1$ vs. $\alpha = \bar{\alpha}$ rejects for large values of $\int \exp\{-1/2 L(\bar{\alpha}, \beta)\} d\beta / \int \exp\{-1/2 L(1, \beta)\} d\beta$. For our normal likelihood, this is equivalent to rejecting for small values of

$$(6) \qquad L_T^* = \min_\beta L(\bar{\alpha}, \beta) - \min_\beta L(1, \beta).$$

The test statistic is the difference in (weighted) sum of squared residuals from two constrained GLS regressions, one imposing $\alpha = \bar{\alpha}$ and the other imposing $\alpha = 1$.

Asymptotic representations in terms of stochastic integrals can be found for this family of statistics but they depend on the specific z_t. When there is no deterministic term ($z_t = 0$), L_T^* is identical to the statistic defined in (3). Some general results for polynomial trend are given in the Appendix. We present explicit formulas for a constant mean where $z_t = 1$ and a linear trend where $z_t = (1, t)'$. Let $\lambda = (1 - \bar{c})/(1 - \bar{c} + \bar{c}^2/3)$. Defining the process

$$(7) \qquad V_c(t, \bar{c}) = W_c(t) - t\left[\lambda W_c(1) + 3(1-\lambda)\int s W_c(s)\, ds\right],$$

we can summarize our results as follows.

THEOREM 1: *Suppose $\{y_t\}$ is generated by the Gaussian model* (1) *under Condition A. Consider unit-root tests of size ε under local-to-unity asymptotics where both $c = T(\alpha - 1)$ and $\bar{c} = T(\bar{\alpha} - 1)$ are fixed as T tends to infinity.*

a. *When d_t is known or satisfies Condition B, the Neyman-Pearson most powerful test against the alterative $c = \bar{c}$ has asymptotic power function $\pi(c, \bar{c})$ defined in* (4). *An upper bound to the asymptotic power of any unit-root test is given by the power envelope $\Pi(c) \equiv \pi(c, c)$.*

b. *When $d_t = \beta_0$, the most powerful invariant test against the alternative $c = \bar{c}$ has asymptotic power function $\pi(c, \bar{c})$. The asymptotic power envelope for this family of point-optimal invariant tests is $\Pi(c)$.*

c. *When $d_t = \beta_0 + \beta_1 t$, the most powerful invariant test against the alternative $c = \bar{c}$ has asymptotic power function*

$$\pi^{\tau}(c, \bar{c}) = \Pr\left[\bar{c}^2 \int_0^1 V_c^2(t, \bar{c})\, dt + (1 - \bar{c}) V_c^2(1, \bar{c}) < b^{\tau}(\bar{c})\right] \tag{8}$$

where $b^{\tau}(\bar{c})$ satisfies $\Pr[\bar{c}^2 \int V_0^2(t, \bar{c}) + (1 - \bar{c}) V_0^2(1, \bar{c}) < b^{\tau}(\bar{c})] = \varepsilon$. An upper bound to the asymptotic power of any unit-root test invariant to the trend parameters β_0 and β_1 is given by the power envelope $\Pi^{\tau}(c) \equiv \pi^{\tau}(c, c)$.

Our primary interest is in alternatives $\bar{c} < 0$, but the theorem is valid for positive c and $\bar{c}$ as well. There are no simple analytic expressions for the power envelopes $\Pi(c)$ and $\Pi^{\tau}(c)$, but simulations indicate that they are monotonically increasing functions of $|c|$. Some plots and calculations are given in Section 4 where a number of alternative tests are discussed.

3. FEASIBLE POINT-OPTIMAL TESTS

Although the point-optimal test statistics defined in (3) and (6) require Σ and u_0 to be known, it is possible to construct tests having the same large-sample properties even in the absence of this knowledge. Furthermore, the asymptotic theory is valid under less stringent assumptions than those made in Theorem 1. In this section, we continue to assume that equation (1) describes the data generating process but we drop Condition A and consider the properties of some feasible tests under weaker assumptions. For $0 \le s \le 1$, let $[sT]$ be the greatest integer less than or equal to sT and let $\Rightarrow$ denote weak convergence of the underlying probability measures as T tends to infinity.

CONDITION C: *The initial error u_0 has a distribution with bounded second moment for all α in a neighborhood of unity. The zero mean process $\{v_t\}$ is stationary and ergodic with finite autocovariances $\gamma(k) \equiv E v_t v_{t-k}$ such that*

(a) *$\omega^2 = \sum_{k=-\infty}^{\infty} \gamma(k)$ is finite and nonzero;*

(b) *the scaled partial-sum process $T^{-1/2} \sum_{t=1}^{[sT]} v_t \Rightarrow \omega W_0(s)$.*

The assumptions on $\{v_t\}$ are satisfied by stationary and invertible ARMA models under moment conditions and are standard in the literature. The stationary assumption can be relaxed at the cost of a more complex notation. Specific assumptions on v_t which imply (b) are discussed in Phillips (1987) and

Phillips and Solo (1992). When $\alpha = 1 + T^{-1}c$ and Condition C is satisfied, $T^{-1/2}u_{[tT]} \Rightarrow \omega W_c(t)$ and sample moments of the data have limiting representations in terms of stochastic integrals involving W_c.

To generate a convenient family of tests, we note that, when $u_0 = 0$ and the v_t are iid $N(0,1)$, the likelihood ratio statistic L_T^* takes a very simple form. Let $S(a)$ be the sum of squared residuals from a least squares regression of y_a on Z_a, where y_a and Z_a are defined in (5). Then the test statistic is equal to $S(\bar{\alpha}) - S(1)$. If Σ is not in fact the identity matrix, this difference in sum of squared residuals will in general have a limiting distribution depending on the error variances and covariances and thus will not produce a test of the correct size. However, it is easy to construct a modified statistic that does produce a valid large-sample test. For the general problem of testing $\alpha = 1$ vs. $\alpha = \bar{\alpha}$ where $d_t = \beta' z_t$ and the v_t have unknown covariances, consider the feasible statistic

$$(9) \qquad P_T = [S(\bar{\alpha}) - \bar{\alpha}S(1)]/\hat{\omega}^2$$

where $\hat{\omega}^2$ is an estimator for ω^2. If β is known, no regression is needed and $S(a)$ is given by $(y_a - Z_a\beta)'(y_a - Z_a\beta)$.

THEOREM 2: *Suppose* $\{y_t\}$ *is generated by* (1) *where* d_t *is a* (*possibly constant*) *polynomial time trend. Then, if Condition C is satisfied and* $\hat{\omega}^2$ *is a consistent estimator of* ω^2 *when* $c = T(\alpha - 1)$ *is fixed,* P_T *has the same limiting distribution under local-to-unity asymptotics as* $L_T^* - \bar{c}$. *Specifically,* P_T *converges in distribution to* $\bar{c}^2 \int W_c^2 - \bar{c}W_c^2(1)$ *in the zero mean and constant mean cases and to* $\bar{c}^2 \int V_c^2(t,\bar{c}) + (1-\bar{c})V_c^2(1,\bar{c})$ *in the linear trend case.*

Thus the power functions $\pi(c,\bar{c})$ and $\pi^{\tau}(c,\bar{c})$ derived in Theorem 1 for point-optimal tests in the Gaussian model with Σ known can be attained by the simple P_T family of statistics under the much weaker assumptions of this section. This is important, because, in practice, Σ will generally contain unknown parameters and there is often no compelling reason to believe that the data are normally distributed. If the errors are non-normal, tests exploiting the form of the actual likelihood and possessing power higher than $\Pi(c)$ and $\Pi^{\tau}(c)$ could be constructed. In the absence of such information, quasi-likelihood tests based on least-squares regressions are likely to be used in practice. The power bounds derived under normality are still valid when comparing such tests.

Although our analysis is based on relatively weak assumptions, two interesting models considered elsewhere in the literature are ruled out. A problem closely related to ours is to test the null hypothesis that $\{u_t\}$ is an integrated process against the alternative that it is a strictly stationary process. Under that alternative, u_0 will have a variance proportional to $(1-\alpha^2)^{-1}$, a violation of Condition C. The tests studied in Section 2 are not point optimal under this specification and the asymptotic power bounds are no longer valid. Our P_T statistics, however, still have simple local-to-unity limiting representations under the stationary alternative. Suppose, for example, u_0 is normal with mean zero and variance $(1-\alpha^2)^{-1}$ and that the v_t are serially uncorrelated with unit variance. Then $T^{-1/2}u_{[tT]} \Rightarrow W_c^*(t) = W_c(t) + \eta_0 e^{ct}$ where η_0 is a normal variate, independent of $W_c(\cdot)$, with mean zero and variance $(-2c)^{-1}$. The P_T statistics can then be written as functionals of the $W_c^*(t)$ process. Further analysis of the stationary alternative testing problem can be found in Elliott (1993).

A second, closely related approach to modeling unit roots is also ruled out here. One way to avoid making an assumption about the initial error u_0 is to base the entire statistical analysis on the conditional distribution of the data given the first observation y_1. When d_t is known, there is no difference asymptotically between our analysis based on the full likelihood and analysis based on the conditional likelihood. But when d_t is unknown, the point-optimal invariant test based on the full likelihood is not asymptotically equivalent to the point-optimal test based on the conditional likelihood. Invariance under the transformation $y_t \to y_t + \bar{\beta}'z_t$ is often justified by the argument that adding a constant to the data should not change the analysis. But that argument is not compelling once one has conditioned on the first observation. There appears to be no convincing way to avoid making an assumption about the initial observation when there are unknown nuisance parameters in d_t.

Although our analysis has been based on local-to-unity asymptotics, our tests have good properties when judged by standard fixed parameter asymptotics as well. Specifically, since the power functions for the P_T tests have nondegenerate limits under sequences of local alternatives α approaching unity, one would also expect power to approach one as T tends to infinity for any fixed $\alpha < 1$. This is indeed the case if the estimate $\hat{\omega}^2$ is not only consistent for local alternatives but also well behaved globally.

THEOREM 3: *Suppose the conditions of Theorem* 2 *hold except that* α *is fixed and less than one in absolute value. If* $\Pr[\hat{\omega}^2 > \mu] \to 1$ *for some positive constant* μ, *then the tests which reject for small values of* P_T (*with* $\bar{c} = T(\bar{\alpha} - 1)$ *held fixed*) *have power functions tending to one as* T *tends to infinity.*

Estimators $\hat{\omega}^2$ that are consistent under local alternatives and have nonzero probability limits under fixed alternatives clearly satisfy this condition. Some examples of such estimators are given in Section 5.

4. SOME NEARLY EFFICIENT INVARIANT TESTS

Theorems 1 and 2 imply that, among tests based on second-order sample moments, those that reject for small values of $[S(\bar{\alpha}) - \bar{\alpha}S(1)]/\hat{\omega}^2$ are asymptotically point optimal invariant; each has an asymptotic power curve tangent to the power envelope at one point. It will be convenient to index the test by its power rather than by the value $\bar{\alpha}$. That is, by inverting the envelope power function

$\pi = \Pi[T(\alpha - 1)]$ for $\alpha \leq 1$, we can find that alternative $\bar{\alpha}(\pi, T, \varepsilon)$ which yields (approximate) power π when using the point optimal test of level ε with a sample of size T. Then, for $\varepsilon \leq \pi \leq 1$, the family of test statistics can be written as

$$(10) \qquad P_T(\pi) \equiv \frac{S[\bar{\alpha}(\pi, T, \varepsilon)] - \bar{\alpha}(\pi, T, \varepsilon) S(1)}{\hat{\omega}^2}.$$

(We suppress the dependence of P on ε.) Although every member of this family is admissible, past research suggests that values of π near one-half often yield tests whose power functions lie close to the power envelope over a considerable range. Cf. King (1988).

For the remainder of the paper we restrict attention to the three standard cases discussed in the literature where d_t is either zero, a constant, or a linear trend. To distinguish the cases, we follow Dickey and Fuller (1979) and use a superscript μ when d_t is constant and a superscript τ when it is a linear trend. Since commonly used test statistics have distributions not depending on the parameters determining the d_t, we shall also restrict attention to invariant tests.

When there is no deterministic term, our family of P_T tests includes as special cases many tests previously proposed. Recall that $P_T(\pi)$ has the asymptotic representation $\bar{c}^2(\pi)\int W_c^2 - \bar{c}(\pi) W_c^2(1)$ where $\bar{c}(\pi)$ is a monotonically decreasing function taking the value zero when π is equal to ε (the size of the test) and tending to minus infinity as π approaches one. Sargan and Bhargava (1983) suggest $S(0)/S(1)$ as a test statistic when the v_t are white noise; asymptotically it behaves like $\int W_c^2$ and corresponds to $P_T(1)$. The locally most powerful test described by Dufour and King (1991) behaves asymptotically like $W_c^2(1)$ and corresponds to $P_T(\varepsilon)$. The Dickey-Fuller estimator test (based on their statistic $\hat{\rho}$) is also a member, since its rejection region is determined, asymptotically, by a linear combination of $\int W_c^2$ and $W_c^2(1)$; computations indicate that it has the same limiting distribution as our $P_T(1 - \varepsilon)$. The Dickey-Fuller t statistic (denoted by $\hat{\tau}$) is a nonlinear function of $\int W_c^2$ and $W_c^2(1)$. Nonetheless, computations indicate that the asymptotic power function of their t test is tangent to the power envelope when power is about one-half and behaves like the $P_T(.5)$ test. Likewise, the Z_α and Z_t tests examined in Phillips (1987) and Phillips and Perron (1988) behave like members of the P_T family since they are asymptotically equivalent to the $\hat{\rho}$ and $\hat{\tau}$ tests, respectively.

Figure 1 graphs the asymptotic power functions of these tests along with the power envelope when the tests have size 0.05. These are based on 20,000 Monte Carlo replications where W_c was approximated by its discrete realization from a sample of size 500; simulation standard errors are less than 0.0013. The power envelope is monotonic and equals one-half when $c = -7$. With the exception of the locally most powerful test which puts all the weight on $W_c^2(1)$, all the tests have power functions very close to the power envelope. Indeed, it is hard to distinguish them without vastly changing the scale of the figure. Although none of these tests is uniformly most powerful even asymptotically, our numerical

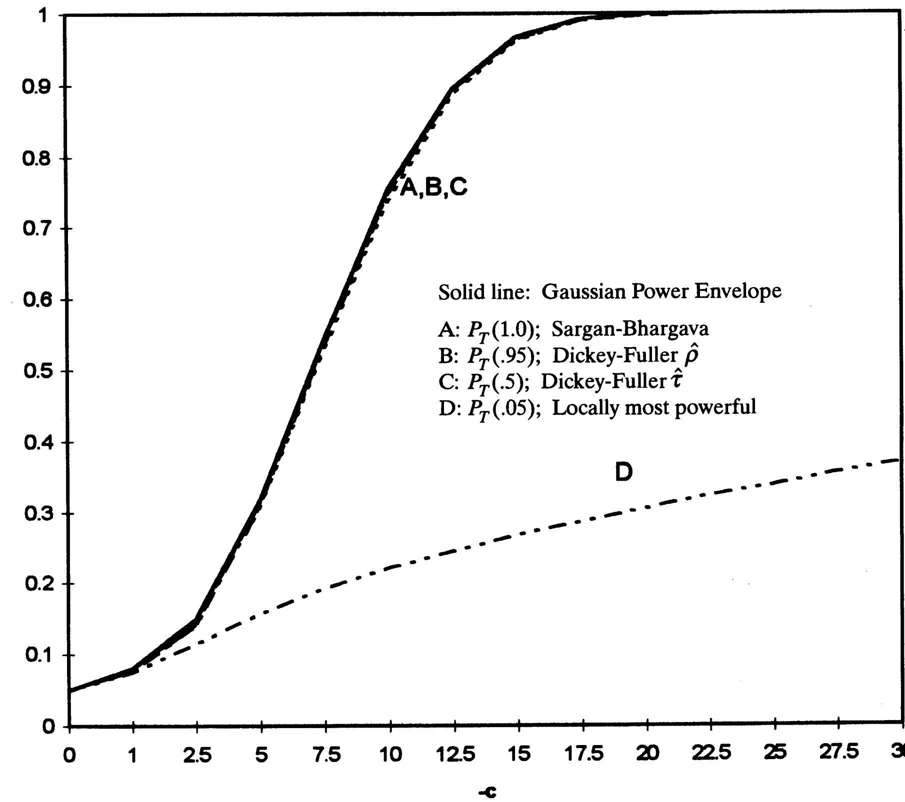

FIGURE 1—Asymptotic power functions of selected unit root tests: no deterministic component.

calculations indicate that, for $.25 \leq \pi \leq .95$, the $P_T(\pi)$ tests are, for all practical purposes, equivalent in large samples and have power functions essentially identical to the power bound. Other calculations not reported here demonstrate that this conclusion carries over to tests at the 1% and 10% significance levels as well.

Things are rather different, however, when d_t contains parameters that have to be estimated. The Sargan-Bhargava (1983) test for the constant mean case, Bhargava's (1986) extension for the linear trend case, the Dickey-Fuller estimator tests (based on their statistics $\hat{\rho}^{\mu}$ and $\hat{\rho}^{\tau}$), the Dickey-Fuller t tests (based on their $\hat{\tau}^{\mu}$ and $\hat{\tau}^{\tau}$), and the Phillips-Perron Z tests are no longer asymptotically equivalent to members of the P_T family since they employ OLS estimates of the β's instead of constrained local-to-unity estimates. The power functions for the $P_T^{\mu}(\pi)$ and $P_T^{\tau}(\pi)$ tests remain very close to the relevant power envelopes $\Pi(c)$ and $\Pi^{\tau}(c)$ for a broad range of π values. The power functions for the tests which use OLS estimates of β are well below the power envelopes. Some results for tests at the 5% level are presented in Figure 2 for the constant

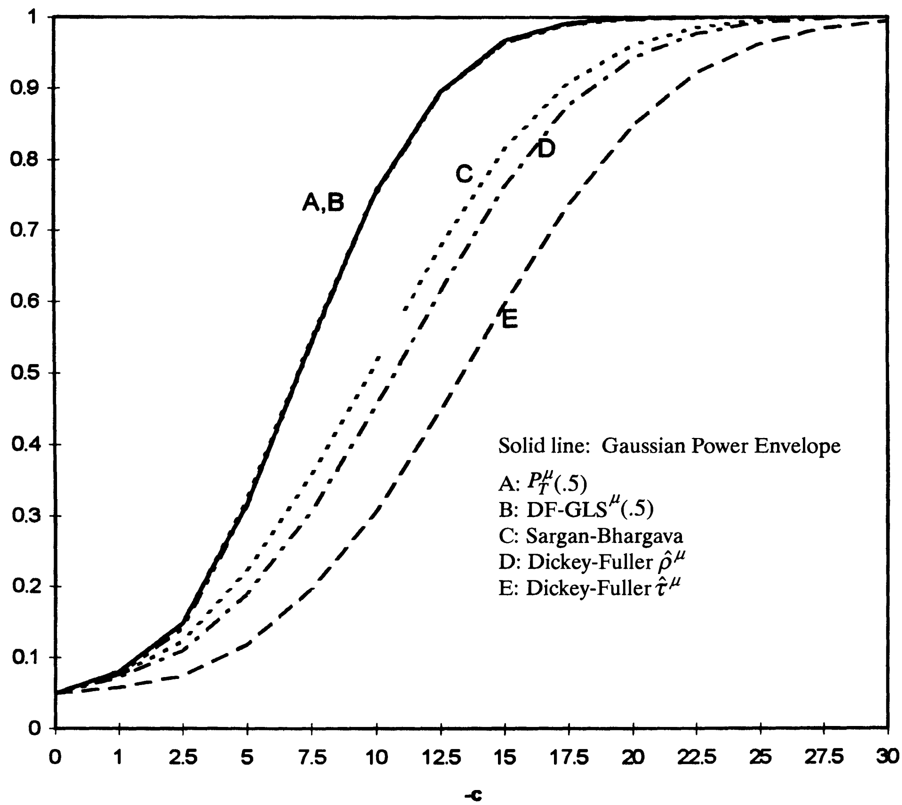

FIGURE 2—Asymptotic power functions of selected unit root tests: constant mean ($z_t = 1$).

mean case and in Figure 3 for the linear trend case. The envelope power curve $\Pi^{\tau}(c)$ has the same shape as $\Pi(c)$, but now takes the value one-half when $c = -13.5$. The power loss of the commonly used tests is particularly dramatic in the constant mean case. The same pattern is found for tests at the 1% and 10% significance levels.

A measure of the difference between two tests is Pitman asymptotic relative efficiency (ARE), defined as the ratio of the values of c at which the tests achieve a specified power. Evaluating efficiency at power one-half and using 5% level tests, we find in the constant mean case the ARE's of the Sargan-Bhargava, $\hat{\rho}^{\mu}$ and $\hat{\tau}^{\mu}$ tests relative to the power envelope are, respectively, 1.40, 1.53, and 1.91. Since c is proportional to T, this implies that using the Dickey-Fuller t test instead of the $P_t(.5)$ test is equivalent in large samples to discarding almost half of the observations. The corresponding ARE's for the linear trend case are 1.07, 1.13, and 1.25.

Since the difficulties with the standard tests are associated with inefficient estimates of the trend parameters, it is reasonable to expect that modified

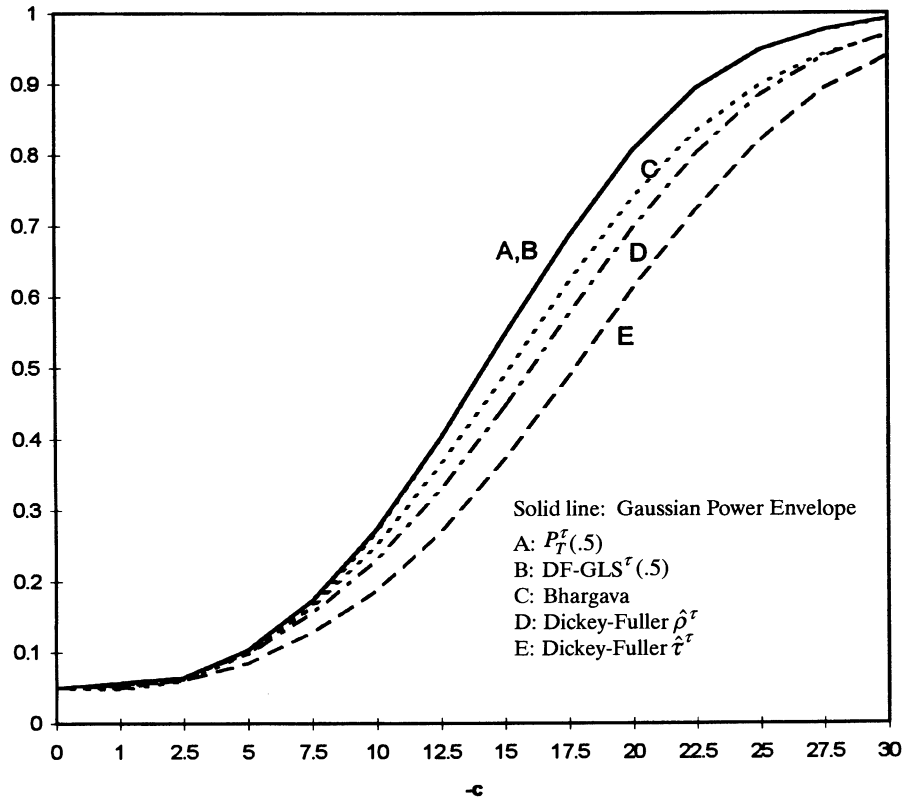

FIGURE 3—Asymptotic power functions of selected unit root tests: linear trend ($z_t = (1, t)'$).

estimates could improve their performance. Because of their relatively good size properties found in small-sample Monte Carlo studies (e.g., Schwert (1989)), natural tests to modify are those based on the Dickey-Fuller t statistics $\hat{\tau}^{\mu}$ and $\hat{\tau}^{\tau}$. Choosing $\bar{\alpha}$ to be that alternative where maximal power is approximately one-half, we propose regressing $y_{\bar{\alpha}}$ on $Z_{\bar{\alpha}}$ to obtain the estimate $\hat{\beta}$. Then one can perform the usual augmented Dickey-Fuller t test (without deterministic regressors) using the residual series $y_t^d \equiv y_t - \hat{\beta}' z_t$ in place of y_t. Thus the modified test statistic (denoted by DF-GLS(π) in the tables and figures) is the t statistic for testing $a_0 = 0$ in the regression

$$(11) \qquad \Delta y_t^d = a_0 y_{t-1}^d + a_1 \Delta y_{t-1}^d + \cdots + a_p \Delta y_{t-p}^d + \text{error}.$$

When $d_t = \beta_0$, the estimate $\hat{\beta}_0$ is stochastically bounded and $T^{-1/2}(y_{[sT]} - \hat{\beta}_0) \Rightarrow \omega W_c(s)$; the t statistic calculated from the demeaned data has the limiting representation $0.5(\int W_c^2)^{-1/2}[W_c^2(1) - 1]$, which is identical to that of $\hat{\tau}$. Critical values and asymptotic power are those of the conventional Dickey-Fuller t statistic when there is no intercept. In the linear trend case, the detrended series

TABLE I

CRITICAL VALUES[a]

T	Level 1%	2.5%	5%	10%
	A. Constant Mean: P_T^{μ} with $\bar{c} = -7$			
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
∞	1.99	2.55	3.26	4.48
	B. Linear Trend: P_T^{τ} with $\bar{c} = -13.5$			
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
∞	3.96	4.78	5.62	6.89
	C. Linear Trend: DF-GLS$^{\tau}$ with $\bar{c} = -13.5$			
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
∞	−3.48	−3.15	−2.89	−2.57

[a] Entries are based on 20,000 Monte Carlo replications. Data were generated by (1) with $\alpha = 1$ and Gaussian white noise $\{v_t\}$. The P_T statistic is given by (9) with $\hat{\omega}^2 = S(1)/T$; DF-GLS$^{\tau}$ is the t statistic calculated from (11) with $p = 0$ and $y^d = y - Z\hat{\beta}$ where $\hat{\beta}$ is obtained by regressing $y_{\bar{\alpha}}$ on $Z_{\bar{\alpha}}$. In both cases, $\bar{\alpha} = 1 + \bar{c}/T$. The line $T \to \infty$ was calculated using a discrete approximation to the relevant stochastic integrals.

$y_t^{\tau} = y_t - \hat{\beta}_0 - \hat{\beta}_1 t$ plays the role of y_t^d. It is shown in the Appendix that $T^{-1/2} y_{[sT]}^{\tau} \Rightarrow \omega V_c(s, \bar{c})$ when $\bar{\alpha} = 1 + \bar{c}/T$ is used for the estimation of β; the t statistic then has the limiting representation $0.5[\int V_c^2(s,\bar{c})]^{-1/2}[V_c^2(1,\bar{c}) - 1]$. Figure 3 graphs the asymptotic power function of the locally detrended t test when $\bar{c} = -13.5$. It is indistinguishable from the power envelope.

The $\bar{\alpha}$ that produces a given asymptotic power π depends on the size of the test, so critical values for the $P_T(\pi)$ tests and for the Dickey-Fuller t test applied to locally detrended data depend on ε. This is inconvenient as it requires an extensive set of tables and, if marginal significance levels are calculated, recomputing the test statistic for different ε. Since the power curves of the tests are not sensitive to $\bar{\alpha}(\pi)$ in the range $.25 \leq \pi \leq .95$, a simpler approach is to fix $\bar{\alpha} = 1 + \bar{c}/T$ independently of ε. Our calculations indicate that, if $\bar{c} = -7$ is chosen for the constant mean case and $\bar{c} = -13.5$ for the linear trend case, the limiting power functions of the resulting P_T tests and for the Dickey-Fuller t test applied to locally detrended data are within 0.01 of the power envelope for $.01 \leq \varepsilon \leq .10$. Some critical values for this choice of $\bar{c}$ are given in Table I. Note that, although the small-sample values are valid only for

Gaussian white noise $\{v_t\}$, the large-sample critical values do not depend on Σ or normality.

5. FINITE SAMPLE PERFORMANCE

A Monte Carlo experiment was conducted to see how well the asymptotic theory describes the small-sample properties of our tests. We investigated tests based on $P_T^{\mu}(.5)$, the standard Dickey-Fuller t statistic (denoted DF-$\hat{\tau}^{\mu}$), and the modified Dickey-Fuller t statistic (denoted DF-GLS$^{\mu}$) for the constant mean case and the corresponding three tests (based on $P_T^{\tau}(.5)$, DF-$\hat{\tau}^{\tau}$, and DF-GLS$^{\tau}$) in the linear trend case. Data generating processes considered elsewhere in the literature (e.g., Phillips and Perron (1988), Schwert (1989), DeJong et al. (1992), and Lumsdaine (1994)) were employed. Specifically, letting $\{\eta_t\}$ be a set of independent standard normal variables, we used the following three models for the $\{v_t\}$ process:

I. MA(1): $v_t = \eta_t - \theta\eta_{t-1}$ $\quad (\theta = .8, .5, 0, -.5, -.8)$,

II. AR(1): $v_t = \phi v_{t-1} + \eta_t$ $\quad (\phi = .5, -.5)$,

III. GARCH MA(1): $v_t = \zeta_t - \theta\zeta_{t-1}, \quad \zeta_t = h_t^{1/2}\eta_t,$

$$h_t = 1 + .65h_{t-1} + .25\zeta_{t-1}^2, \quad h_0 = 0 \quad (\theta = .5, 0, -.5).$$

In each of these models the initial condition was $u_0 = 0$. Although the null distribution of the test statistics considered here are invariant to the initial condition, small-sample power typically depends on u_0. This dependence is investigated by considering a variant of the first model where the $\{u_t\}$ are strictly stationary under the alternative hypothesis. That is, u_0 is normal with mean zero and variance equal to $(1 + \theta^2 - 2\theta\alpha)/(1 - \alpha^2)$, $\alpha \neq 1$. This design violates our Condition C and is intended to shed light on the importance of that assumption.

The autocorrelation structure of $\{v_t\}$ was assumed to be unknown to the investigator, so two types of loosely parameterized estimators, autoregressive (AR) and sum-of-covariances (SC), were used for $\hat{\omega}^2$. The AR estimators are given by

$$\hat{\omega}_{AR}^2 = \hat{\sigma}_\eta^2 \Big/ \left(1 - \sum_{i=1}^{p} \hat{a}_i\right)^2 \tag{13}$$

where $\hat{\sigma}_\eta^2$ and the $\hat{a}_i$ are OLS estimates from the regression

$$\Delta y_t = a_0 y_{t-1} + a_1 \Delta y_{t-1} + \cdots + a_p \Delta y_{t-p} + a_{p+1} + \eta_t. \tag{14}$$

Two choices of lag length were employed: the AR(8) estimator used $p = 8$ and the AR(BIC) estimator used p chosen by the Schwarz (1978) Bayesian information criterion constrained so $3 \leq p \leq 8$. The SC estimators are given by

$$\hat{\omega}_{SC}^2 = \sum_{m=-l_T}^{l_T} K(m/l_T)\hat{\gamma}(m)$$

where $K(\cdot)$ is the Parzen kernel, $\hat{\gamma}(m) = T^{-1}\sum_{t=1}^{T-m} e_t e_{t+m}$, and e_t is the residual from an OLS regression of y_t on (y_{t-1}, z_t). Two variants were employed: SC(12) using $l_T = 12$ and SC(auto) using Andrews' (1991) optimal automatic procedure (his equations (6.2) and (6.4)).

The results are summarized in Table II for a constant mean and in Table III for a linear trend. Tests were at the 5% asymptotic significance level and the sample size T was 100. For $\alpha = 1$, the tables report the observed rejection rates from 5000 Monte Carlo replications when critical values were based on the limiting distributions. For $\alpha < 1$, the tables report size-adjusted power; this is the rejection rate when critical values are estimated from the $\alpha = 1$ Monte Carlo trials.

The results suggest three conclusions. First, the predicted superiority of the tests using local-to-unity estimates of the mean and trend parameters is borne out by the Monte Carlo study. The P_T and modified Dickey-Fuller tests have higher size-adjusted power than the standard Dickey-Fuller t test for almost all of the data generating processes and all choices of $\hat{\omega}^2$. The improvement is largest in the constant mean case. Although the observed power curves tend to be somewhat below the asymptotic power curves, the results are generally consistent with the predictions of the asymptotic theory. The main exception is the poor performance of the point-optimal tests using SC estimates of ω^2 when the MA parameter θ is large.

Second, the choice of estimator for ω^2 has a large effect on the size of the P_T tests, with the AR estimator exhibiting much smaller distortions than the SC estimator. This mirrors similar results found for other unit-root statistics; see, for example, DeJong et al. (1992) and Perron (1996). The AR(8) and AR(BIC) tests have moderate size distortion except in the MA model with large θ. The modified Dickey-Fuller tests have notably smaller size distortions than those based on P_T. In addition, the tests based on the AR(BIC) estimator have better size-adjusted power than those based on the AR(8) estimator, which typically estimates more nuisance parameters. Other experiments not reported in Tables II or III indicate that the AR(BIC) tests also dominate the ones based on the AR(4) estimator for ω^2. Lag length selection based on sequential likelihood ratio statistics was also tried; no general improvement over AR(BIC) was found, although the LR selector appears to improve the size-adjusted power of the modified Dickey-Fuller test relative to BIC in the linear trend case, at least for small values of θ.

Third, the powers of the P_T and modified Dickey-Fuller tests deteriorate substantially when the u_t are stationary. Even so, in the linear trend case with

TABLE II

SIZE AND SIZE-ADJUSTED POWER OF SELECTED TESTS OF THE I(1) NULL: MONTE CARLO RESULTS
5% LEVEL TESTS, CONSTANT MEAN ($z_t = 1$), $T = 100$

Test Statistic	α	Asymptotic Power	MA(1), $\theta =$					AR(1), $\phi =$		GARCH MA(1), $\theta =$			Stationary MA(1), $\theta =$		
			−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
$P_T^\mu(.5)$	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
AR(8)	.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
$P_T^\mu(.5)$	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
AR(BIC)	.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
$P_T^\mu(.5)$	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
SC(12)	.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
$P_T^\mu(.5)$	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
SC(auto)	.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$^\mu$(.5)	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
AR(8)	.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
DF-GLS$^\mu$(.5)	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-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: For each statistic, entries in the first row are the empirical rejection rate under the null (the size). The remaining entries are the size-adjusted power under the model described in the column heading. The column, "Asymptotic Power," is the local-to-unity asymptotic power for each statistic. The entry below the name of each statistic indicates the estimator of ω^2 used (see Section 5). For the $|\alpha| < 1$ cases in the final three columns, u_0 was drawn from its stationary distribution. Based on 5000 Monte Carlo replications.

TABLE III

SIZE AND SIZE-ADJUSTED POWER OF SELECTED TESTS OF THE I(1) NULL: MONTE CARLO RESULTS
5% LEVEL TESTS, LINEAR TREND ($z_t = (1,t)'$), $T = 100$

Test Statistic	α	Asymptotic Power	MA(1), $\theta =$ −0.8	MA(1), $\theta =$ −0.5	MA(1), $\theta =$ 0.0	MA(1), $\theta =$ 0.5	MA(1), $\theta =$ 0.8	AR(1), $\phi =$ 0.5	AR(1), $\phi =$ −0.5	GARCH MA(1), $\theta =$ −0.5	GARCH MA(1), $\theta =$ 0.0	GARCH MA(1), $\theta =$ 0.5	Stationary MA(1), $\theta =$ −0.5	Stationary MA(1), $\theta =$ 0.0	Stationary MA(1), $\theta =$ 0.5
$P_T^\tau(.5)$	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
AR(8)	.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
$P_T^\tau(.5)$	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
AR(BIC)	.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
$P_T^\tau(.5)$	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
SC(12)	.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
$P_T^\tau(.5)$	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
SC(auto)	.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$^\tau$(.5)	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
AR(8)	.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
DF-GLS$^\tau$(.5)	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-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 II.

$\theta = 0$, the size-adjusted powers of the tests using local detrending exceed that of DF-$\hat{\tau}^{\tau}$. In the constant mean case, the size-adjusted powers of tests using local demeaning exceed that of DF-$\hat{\tau}^{\mu}$ for close but not distant alternatives. The gains from employing local-to-unit estimates of the intercept appear to depend crucially on the assumption that, under both the null and the alternative hypotheses, only the early observations are informative about that parameter.

6. CONCLUSIONS

The P_T and modified Dickey-Fuller t statistics are easily computed from least squares regressions. If the sample size is large enough so the effects of residual autocorrelation are captured by $\hat{\omega}^2$ and the asymptotic approximations are accurate, these tests are essentially optimal among tests based on second-order sample moments and should perform considerably better than tests which employ OLS estimates of the parameters determining d_t. Our Monte Carlo results suggest that the Dickey-Fuller t test applied to a locally demeaned or detrended time series, using a data-dependent lag length selection procedure, has the best overall performance in terms of small-sample size and power.

The numerical finding that, as a practical matter, the asymptotic power functions of the $P_T(.5)$ and the modified Dickey-Fuller t tests effectively lie on the Gaussian power envelope indicates that, in large samples, there is little room for improvement under the stochastic specification made here. Of course, if the errors have a known non-normal distribution or if the initial error u_0 is large compared to ω, better tests could be constructed. Furthermore, the Monte Carlo evidence suggests that autocorrelation in the v_t can have very substantial effects in small samples. Nevertheless, it appears that, when parameters in the deterministic component of a series have to be estimated, the proposed tests for a unit root dominate those currently in common use.

Dept. of Economics, University of California—San Diego, 9500 Gilman Dr., La Jolla, CA 92093, U.S.A.,

Dept. of Economics, University of California—Berkeley, Evans Hall, Berkeley, CA 94720, U.S.A.,

and

Kennedy School of Government, Harvard University, 79 John F. Kennedy St., Cambridge, MA 02138, U.S.A.

Manuscript received November, 1992; final revision received June, 1995.

APPENDIX A: PRELIMINARY LEMMAS

Let $\gamma(k)$ be the autocovariance function and $f(\lambda)$ the spectral density function for the stationary process $\{v_t\}$ satisfying Condition A. The rs element of the $T \times T$ covariance matrix Σ is $\gamma(r-s) = \int_{-\pi}^{\pi} e^{i(r-s)\lambda} f(\lambda)\, d\lambda$. We shall approximate Σ^{-1} by the $T \times T$ matrix Ψ with rs element $\rho(r-s) \equiv \int_{-\pi}^{\pi} e^{i(r-s)\lambda}[4\pi^2 f(\lambda)]^{-1}\, d\lambda$. The rs element of $D \equiv I_T - \Psi\Sigma$ is given by Davies (1973) and Dzhaparidze (1985) as

$$d_{rs} = \sum_{k=-\infty}^{s-1-T} \gamma(k)\rho(s-r-k) + \sum_{k=s}^{\infty} \gamma(k)\rho(s-r-k).$$

For real $p \times q$ matrix $B = [b_{ij}]$, let $r(B)$ be the square root of the largest characteristic root of $B'B$, let $\|B\| \equiv \Sigma_{i=1}^{p} \Sigma_{j=1}^{q} |b_{ij}|$, and let $|B| \equiv \mathrm{tr}^{1/2}(B'B)$. Then, $r(B) \le |B| \le \|B\|$ and, if B and C are conformable, $|\mathrm{tr}(BC)| \le |B||C|$ and $|BC| \le |B| r(C)$. Cf. Davies (1973). Since $\Sigma_{j=-\infty}^{\infty} |\delta_j j| < \infty$ implies $\Sigma_{k=-\infty}^{\infty} |\gamma(k)k| < \infty$ and (by Theorem 5.2 in Zygmund (1968, p. 245)) $\Sigma_{j=-\infty}^{\infty} |\rho(j)| < \infty$, we find

$$\text{(A1)} \qquad \|D\| = \sum_{r=1}^{T} \sum_{s=1}^{T} |d_{rs}| \le 2 \sum_{k=1}^{\infty} |\gamma(k)| k \sum_{j=-\infty}^{\infty} |\rho(j)| < \infty.$$

As a consequence, we have the following four lemmas.

LEMMA A1: *Let Σ and Ψ be $T \times T$ Toeplitz matrices formed from $\gamma(k)$ and $\rho(k)$, the Fourier coefficients of $2\pi f(\lambda)$ and $[2\pi f(\lambda)]^{-1}$, respectively. Let x be a $T \times 1$ vector such that $\lim_{T\to\infty} T^{-1}|x| = 0$. If the elements of the $T \times 1$ vector z and of the $T \times T$ matrix A are bounded in absolute value, then, under Condition A,*

$$\text{(A2)} \qquad \lim_{T\to\infty} T^{-1} x'(\Sigma^{-1} - \Psi) z = \lim_{T\to\infty} T^{-2} \mathrm{tr}[A'(\Sigma^{-1} - \Psi)A] = 0.$$

PROOF: Since $f(\lambda)$ is continuous and positive on $[-\pi, \pi]$, there exist positive m and M such that $m \le f(\lambda) \le M$; hence, $r(\Sigma) \le 2\pi M$ and $r(\Sigma^{-1}) \le (2\pi m)^{-1}$. For some constant K, $|D'z| \le K\|D\|$, $|D'A| \le KT^{1/2}\|D\|$, and $|A| \le KT$. Using (A1), we find

$$T^{-1}|x'(\Sigma^{-1} - \Psi)z| = T^{-1}|x'\Sigma^{-1}D'z| \le T^{-1}|x'\Sigma^{-1}||D'z| \le T^{-1}|x| r(\Sigma^{-1}) K\|D\| \to 0,$$

$$T^{-2}|\mathrm{tr}[A'(\Sigma^{-1} - \Psi)A]| = T^{-2}|\mathrm{tr}[A'\Sigma^{-1}D'A]| \le T^{-2} r(\Sigma^{-1})|A||D'A|$$
$$\le T^{-1/2}K^2 r(\Sigma^{-1})\|D\| \to 0.$$

LEMMA A2: *Suppose the data are generated by* (1) *under Condition A. Then $\omega^2 \equiv \Sigma_{k=-\infty}^{\infty}\gamma(k)$ is positive and, if $c = T(\alpha - 1)$ is fixed as $T \to \infty$,*

$$T^{-2}u'_{-1}\Sigma^{-1}u_{-1} - \omega^{-2}T^{-2}u'_{-1}u_{-1} \xrightarrow{p} 0,$$

$$2T^{-1}u'_{-1}\Sigma^{-1}\Delta u + 1 - \omega^{-2}[2T^{-1}u'_{-1}\Delta u + \gamma(0)] \xrightarrow{p} 0.$$

PROOF: Since $2\pi f(\lambda) = \Sigma_{-\infty}^{\infty}\gamma(k)e^{ik\lambda}$ and $[2\pi f(\lambda)]^{-1} = \Sigma_{-\infty}^{\infty}\rho(k)e^{ik\lambda}$, we find that $\omega^2 = 2\pi f(0) > 0$ and $\Sigma_{-\infty}^{\infty}\rho(k) = \omega^{-2}$. As $\Delta u = v + cT^{-1}u_{-1}$, it will suffice to show that $S_1 \equiv T^{-2}u'_{-1}(\Sigma^{-1} - \omega^{-2}I)u_{-1} \xrightarrow{p} 0$ and $S_2 \equiv 2T^{-1}u'_{-1}(\Sigma^{-1} - \omega^{-2}I)v \xrightarrow{p} \omega^{-2}\gamma(0) - 1$. When $u_0 = 0$, $u_{-1} = Av$, where $A = [a_{rs}]$ is a $T \times T$ matrix with a_{rs} equal to α^{r-s-1} when $r > s$ and zero otherwise. Note that $|a_{rs}| \le e^{|c|}$ and, for nonrandom square matrix B, $E(v'Bv) = \mathrm{tr}(B\Sigma)$ and $\mathrm{var}(v'Bv) = \mathrm{tr}(B\Sigma B\Sigma + B\Sigma B'\Sigma)$. Defining $R = (\omega\Sigma^{-1/2} - \omega^{-1}\Sigma^{1/2})$, we find

$$|E(S_1)| = T^{-2}\omega^{-1}|\mathrm{tr}[A'R\Sigma^{-1/2}A\Sigma]| \le T^{-1}\omega^{-1}r(\Sigma)r^{1/2}(\Sigma^{-1})e^{|c|}|RA|,$$

$$\mathrm{Var}(S_1) = T^{-4}\omega^{-2}2\,\mathrm{tr}[A'R\Sigma^{-1/2}A\Sigma A'\Sigma^{-1/2}RA\Sigma] \le 2T^{-2}\omega^{-2}r^2(\Sigma)r(\Sigma^{-1})e^{2|c|}|RA|^2,$$

$$E(S_2) = 2T^{-1}[\mathrm{tr}(A) - \omega^{-2}\mathrm{tr}(\Sigma A)]$$
$$= -2\omega^{-2}T^{-1}\Sigma_{k=1}^{T}\gamma(k)(T-k)\alpha^{k-1} \to \omega^{-2}\gamma(0) - 1,$$

$$\mathrm{Var}(S_2) = T^{-2}\omega^{-2}4\,\mathrm{tr}[A'R\Sigma^{1/2}A'R\Sigma^{1/2} + RA\Sigma A'R] \le 8T^{-2}\omega^{-2}r(\Sigma)|RA|^2.$$

To complete the proof, we need to show that $T^{-1}|RA| \to 0$ as $T \to \infty$. Define

$$a_T(k) \equiv T^{-2}\sum_{t=1}^{T-k}\sum_{s=1}^{T} a_{t,s}a_{t+k,s} = T^{-2}(1 + cT^{-1})^k \sum_{r=1}^{T-k}(1 + cT^{-1})^{2(r-1)}(T - k - r).$$

Note that, for fixed k, $a_T(k) \to (e^{2c} - 1 - 2c)/(2c)^2$ when $c \ne 0$ and $a_T(k) \to 1/2$ when $c = 0$, where the limits are independent of k. Moreover, the $a_T(k)$ are bounded by $e^{|2c|}$ and the sequences

$\gamma(k)$ and $\rho(k)$ are absolutely summable. Since $s_T \equiv \sum_{k=-T+1}^{T-1}\gamma(k) \to \omega^2$ and $r_T \equiv \sum_{k=-T+1}^{T-1}\rho(k) \to \omega^{-2}$, it follows that

$$T^{-2}\operatorname{tr}[A'(\Sigma - s_T I)A] = 2\sum_{k=1}^{T-1}\gamma(k)[a_T(k) - a_T(0)] \to 0,$$

$$T^{-2}\operatorname{tr}[A'(\Psi - r_T I)A] = 2\sum_{k=1}^{T-1}\rho(k)[a_T(k) - a_T(0)] \to 0.$$

Cf. Anderson (1971, 10.2.3). Using (A2), we have

(A3) $$T^{-2}|RA|^2 = T^{-2}\operatorname{tr} A'R^2A = T^{-2}\operatorname{tr}[\omega^2 A'\Sigma^{-1}A + \omega^{-2}A'\Sigma A - 2A'A] \to 0.$$

LEMMA A3: *If the data are generated by* (1) *under Conditions A and B, then*

(A4) $$\lim T^{-2}d_{-1}'\Sigma^{-1}d_{-1} = \lim T^{-1}\Delta d'\,\Sigma^{-1}d_{-1} = 0,$$

(A5) $$\operatorname{plim} T^{-2}d_{-1}'\Sigma^{-1}u_{-1} = \operatorname{plim} T^{-1}d_{-1}'\Sigma^{-1}\Delta u = \operatorname{plim} T^{-1}\Delta d'\,\Sigma^{-1}u_{-1} = 0,$$

where $d_{-1} = (0, d_1, \ldots, d_{T-1})'$ *and* $\Delta d = (d_1, d_2 - d_1, \ldots, d_T - d_{T-1})'$.

PROOF: Under Condition B, $T^{-2}d_{-1}'\Sigma^{-1}d_{-1} \le r(\Sigma^{-1})T^{-2}d_{-1}'d_{-1} \le r(\Sigma^{-1})\max_{t<T} d_t^2/T \to 0$. By Lemma A1, $\lim T^{-1}\Delta d'(\Sigma^{-1} - \Psi)d_{-1} = 0$. But, defining $d_0 \equiv 0$, we have

(A6) $$\begin{aligned}|2T^{-1}\Delta d'\Psi d_{-1}| &= T^{-1}|d'\Psi d - d_{-1}'\Psi d_{-1} - \Delta d'\Psi\Delta d| \\ &= \Big|\rho(0)T^{-1}\sum_{t=1}^{T}[d_t^2 - d_{t-1}^2] \\ &\quad + 2\sum_{k=1}^{T-1}\rho(k)T^{-1}\sum_{t=1}^{T-k}[d_t d_{t+k} - d_{t-1}d_{t-1+k}] - T^{-1}\Delta d'\Psi\Delta d\Big| \\ &\le \max_{t\le T} T^{-1}d_t^2 \sum_{k=-\infty}^{\infty}|\rho(k)| + r(\Psi)T^{-1}\Delta d'\,\Delta d \to 0.\end{aligned}$$

Using the notation developed in the proof of Lemma A2, we have

$$E[T^{-2}d_{-1}'\Sigma^{-1}u_{-1}]^2 = T^{-4}d_{-1}'\Sigma^{-1}A\Sigma A'\Sigma^{-1}d_{-1} \le T^{-2}e^{2|c|}r(\Sigma)r^2(\Sigma^{-1})|d_{-1}|^2 \to 0$$

which implies that $\operatorname{plim} T^{-2}d_{-1}'\Sigma^{-1}u_{-1} = 0$. For the second part of (A5), note that $\Delta u = v + cT^{-1}u_{-1}$ so it suffices to verify that $\operatorname{var}[T^{-1}d_{-1}'\Sigma^{-1}v] = T^{-2}d_{-1}'\Sigma^{-1}d_{-1} \to 0$. Finally, $T^{-1}\Delta d'\,\Delta d \to 0$ and $|D'A| \le T^{1/2}e^{|c|}\|D\|$ imply $T^{-1}\Delta d'(\Sigma^{-1} - \Psi)u_{-1} \xrightarrow{p} 0$ since

$$\begin{aligned}\operatorname{Var}[T^{-1}\Delta d'(\Sigma^{-1} - \Psi)u_{-1}] &= T^{-2}\Delta d'\,\Sigma^{-1}D'A\Sigma A'D\Sigma^{-1}\Delta d \\ &\le T^{-1}r(\Sigma)r^2(\Sigma^{-1})|\Delta d|^2\|D\|^2e^{2|c|} \to 0.\end{aligned}$$

But $T^{-1}\Delta d'\Psi u_{-1} \equiv T^{-1}[d'\Psi u - d_{-1}'\Psi u_{-1}] - T^{-1}d'\Psi\Delta u$. The last term tends to zero by the same argument used for $T^{-1}d_{-1}'\Sigma^{-1}\Delta u$. Since $\max T^{-1/2}u_t$ is stochastically bounded, the algebra of (A6) shows that

$$T^{-1}|d'\Psi u - d_{-1}'\Psi u_{-1}| \le T^{-1}\max_{t\le T}|d_t|\max_{t\le T}|u_t|\sum_{k=-\infty}^{\infty}|\rho(k)| \xrightarrow{p} 0,$$

thus completing the proof.

Define

$$\text{(A7)} \qquad Q_T(a) = u_a' Z_a [Z_a' Z_a]^{-1} Z_a' u_a, \qquad Q_T^M(a) = u_a' \Sigma^{-1} Z_a [Z_a' \Sigma^{-1} Z_a]^{-1} Z_a' \Sigma^{-1} u_a$$

where $u_a = (u_1, u_2 - au_1, \ldots, u_T - au_{T-1})'$ and Z_a is given in (5).

LEMMA A4: *Suppose the data are generated by* (1) *under Condition A and* $d_t = \beta' z_t$, *where* $z_t' = (1, t, t^2, \ldots, t^{q-1})$. *Then* $Q_T(\bar{\alpha}) - Q_T(1)$ *has a limiting distribution when* $c = T(\alpha - 1)$ *and* $\bar{c} = T(\bar{\alpha} - 1)$ *are fixed as T tends to infinity. Furthermore,*

$$\text{(A8)} \qquad Q_T^M(\bar{\alpha}) - Q_T^M(1) - \omega^{-2}[Q_T(\bar{\alpha}) - Q_T(1)] \xrightarrow{p} 0.$$

PROOF: Define the $q \times q$ diagonal matrix $N_T = \text{diag}(T^{1/2}, 1, T^{-1}, \ldots, T^{-q-2})$ and the $T \times q$ matrix $\bar{Z} = Z_{\bar{\alpha}} N_T$. Then, setting $\xi \equiv \Delta u - \bar{c}T^{-1}u_{-1} = v + (c - \bar{c})T^{-1}u_{-1}$ we have $Q_T(\bar{\alpha}) = \xi'\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'\xi$ and $Q_T^M(\bar{\alpha}) = \xi'\Sigma^{-1}\bar{Z}(\bar{Z}'\Sigma^{-1}\bar{Z})^{-1}\bar{Z}'\Sigma^{-1}\xi$. Write $\bar{Z} = [x, X]$, where x is the first column of $\bar{Z}$ and $X = [x_{tj}]$ consists of the other $q - 1$ columns. Defining e_1 to be the first column of I_T and ι to be a vector of T ones, we can write $x = (T^{1/2} + \bar{c}T^{-1/2})e_1 - \bar{c}T^{-1/2}\iota$; hence, $T^{-1}x'x \to 1$ and $T^{-1/2}x'\xi = u_1 + o_p(1)$. For $j = 1, \ldots, q - 1$, $x_{1j} = 1$ and $x_{tj} = T^{-j}(T\Delta t^j - \bar{c}t^j)$ when $t > 1$. For all t and j, $|x_{tj}| \le q + \bar{c}$ and hence $T^{-1}X'x \to 0$. Define the continuous function $h_j(s, \bar{c}) = (j - \bar{c}s)s^{j-1}$ for $0 \le s \le 1$. Then $T^{-1}X'X$ converges to the matrix $G(\bar{c}) = [g_{ij}] = [\int_0^1 h_i(s, \bar{c})h_j(s, \bar{c})\,ds]$.

The identity $X'u - X_{-1}'u_{-1} = X'\Delta u + \Delta X'u_{-1}$ implies

$$\text{(A9)} \qquad T^{-1/2}X'\xi \equiv T^{-1/2}X'(\Delta u - \bar{c}T^{-1}u_{-1}) = T^{-1/2}x_T u_T - T^{-3/2}(T\Delta X + \bar{c}X)'u_{-1}$$

where x_T is the last column of X'. Under Condition A, $T^{-1/2}u_{[sT]} \Rightarrow \omega W_c(s)$. By the continuous mapping theorem and the Ito calculus,

$$\omega^{-1}T^{-1/2}X'\xi \Rightarrow h(1, \bar{c})W_c(1) - \int_0^1 [H(s, \bar{c}) + \bar{c}h(s, \bar{c})]W_c(s)\,ds = \int h(\bar{c})[dW_c - \bar{c}W_c],$$

where $h(s, \bar{c})$ is the vector consisting of the $q - 1$ functions $h_j(s, \bar{c})$, $H(s, \bar{c})$ is the vector of first derivatives $dh_j(s, \bar{c})/ds$, and the time index is dropped in the final term for notational convenience. Since $\lim T^{-1}\bar{Z}'\bar{Z}$ is block diagonal, we find that $Q_T(\bar{\alpha}) \Rightarrow u_1^2 + \omega^2 Q(\bar{c})$ where

$$\text{(A10)} \qquad Q(\bar{c}) \equiv \left[\int h(\bar{c})(dW_c - \bar{c}W_c)\right]' \left[\int h(\bar{c})h'(\bar{c})\right]^{-1} \left[\int h(\bar{c})(dW_c - \bar{c}W_c)\right].$$

Setting $\bar{\alpha}$ to one, the same argument shows that $Q_T(\bar{\alpha}) - Q_T(1) \Rightarrow \omega^2[Q(\bar{c}) - Q(0)]$.

The argument now follows the proof of Lemma A2. Because the elements $\bar{z}_{tj}$ of $\bar{Z}$ are polynomials in t, for all k and for all (i, j) pairs except $i = j = 1$, the terms

$$b_T^{ij}(k) \equiv T^{-1} \sum_{t=1}^{T-k} [\bar{z}_{t,i}\bar{z}_{t+k,j} + \bar{z}_{t,j}\bar{z}_{t+k,i}]$$

are uniformly bounded and tend to constants independent of k. The ij element of $T^{-1}(\bar{Z}'\Psi\bar{Z} - r_T\bar{Z}'\bar{Z})$ is given by $\Sigma_{k=1}^{T-1}\rho(k)[b_T^{ij}(k) - b_T^{ij}(0)]$ and hence tends to zero except when $i = j = 1$. Since $T^{-1}X'(\Sigma^{-1} - \Psi)X \to 0$ and $T^{-1}X'(\Sigma^{-1} - \Psi)x \to 0$ by Lemma A1, we conclude that $T^{-1}X'\Sigma^{-1}x \to 0$ and $\omega^2 T^{-1}X'\Sigma^{-1}X \to G(\bar{c})$. Similarly, $T^{-1}(X'\Sigma X - \omega^2 X'X) \to 0$, so $T^{-1}|RX|^2 \to 0$, where R is defined in the proof of Lemma A2. It follows that

$$S_3 = T^{-3/2}X'(\Sigma^{-1} - \omega^{-2}I)u_{-1} \xrightarrow{p} 0, \qquad S_4 = T^{-1/2}X'(\Sigma^{-1} - \omega^{-2}I)v \xrightarrow{p} 0,$$

since $\text{tr}[E(S_3S_3')] = \omega^{-2}T^{-3}\text{tr}[X'R\Sigma^{-1/2}A\Sigma A'\Sigma^{-1/2}RX] \le \omega^{-2}T^{-1}r(\Sigma)r(\Sigma^{-1})e^{2|c|}|RX|^2 \to 0$ and $\text{tr}[E(S_4S_4')] = \omega^{-2}T^{-1}\text{tr}[X'R^2X] = \omega^{-2}T^{-1}|RX|^2 \to 0$. This implies $T^{-1/2}X'(\Sigma^{-1} - \omega^{-2}I)\xi \xrightarrow{p} 0$. By the same argument, $T^{-2}\iota'\Sigma^{-1}\iota \to 0$, $T^{-1}\iota'\Sigma^{-1}e_1 \to 0$, $T^{-1}\iota'\Sigma^{-1}\xi \xrightarrow{p} 0$, and $T^{-1}e_1'\Sigma^{-1}u_{-1} \xrightarrow{p} 0$.

Defining the chi-square variate $\chi^2(v) \equiv (e_1'\Sigma^{-1}v)^2/e_1'\Sigma^{-1}e_1$, we find that $(x'\Sigma^{-1}\xi)^2/x'\Sigma^{-1}x - \chi^2(v) \xrightarrow{p} 0$. Since $\lim T^{-1}\bar{Z}'\Sigma^{-1}\bar{Z}$ is block diagonal, we have

(A11) $$Q_T^M(\bar{\alpha}) = \omega^{-2}Q_T(\bar{\alpha}) + \chi^2(v) - \omega^{-2}u_1^2 + o_p(1).$$

Since (A11) holds for all $\bar{\alpha}$ in a neighborhood of unity and $\chi^2(v) - \omega^{-2}u_1^2$ does not depend on $\bar{\alpha}$, (A8) follows.

APPENDIX B: PROOFS OF THEOREMS 1–3

The statistical theory underlying Theorem 1 can be found in Lehmann (1959, Chapter 6). The limiting representations for the test statistics are derived as follows:

(a) Under Condition A, $T^{-1/2}u_{[sT]} \Rightarrow \omega W_c(s)$ and hence $\omega^{-2}(T^{-2}u_{-1}'u_{-1}, T^{-1}u_T^2) \Rightarrow (\int W_c^2, W_c^2(1))$. Lemma A2 implies that $(T^{-2}u_{-1}'\Sigma^{-1}u_{-1}, 2T^{-1}u_{-1}'\Sigma^{-1}\Delta u)$ converges in probability to $\omega^{-2}(T^{-2}u_{-1}'u_{-1}, 2T^{-1}u_{-1}'\Delta u + \gamma(0) - \omega^2)$. But

(B1) $$2T^{-1}u_{-1}'\Delta u \equiv T^{-1}[u_T^2 - \Delta u'\Delta u] = T^{-1}u_T^2 - \gamma(0) + o_p(1)$$

since $T^{-1}\Delta u'\Delta u = T^{-1}v'v + c^2T^{-3}u_{-1}'u_{-1} + 2cT^{-2}u_{-1}'v \xrightarrow{p} \gamma(0)$. Thus,

(B2) $$\bar{c}^2T^{-2}u_{-1}'\Sigma^{-1}u_{-1} - 2\bar{c}T^{-1}u_{-1}'\Sigma^{-1}\Delta u \Rightarrow \bar{c}^2\int W_c^2 - \bar{c}[W_c^2(1) - 1].$$

If y is used in place of u in the slowly evolving trend case, we must add

$$\bar{c}^2T^{-2}(d_{-1}'\Sigma^{-1}d_{-1} + 2d_{-1}'\Sigma^{-1}u_{-1}) - 2\bar{c}T^{-1}(\Delta d'\Sigma^{-1}d_{-1} + d_{-1}'\Sigma^{-1}\Delta u + \Delta d'\Sigma^{-1}u_{-1})$$

to the statistic on the left of (B2). But, from Lemma A3, these terms converge in probability to zero under Condition B, so the limiting distribution is unchanged.

(b) From standard GLS projection theory and (A7)

$$\min_\beta L(a, \beta) = y_a'\left[\Sigma^{-1} - \Sigma^{-1}Z_a(Z_a'\Sigma^{-1}Z_a)^{-1}Z_a'\Sigma^{-1}\right]y_a = u_a'\Sigma^{-1}u_a - Q_T^M(a).$$

Recalling that $L_T^* = \min_\beta L(\bar{\alpha}, \beta) - \min_\beta L(1, \beta)$ and $\bar{\alpha} = 1 + T^{-1}\bar{c}$, we find

$$L_T^* = \bar{c}^2T^{-2}u_{-1}'\Sigma^{-1}u_{-1} - 2\bar{c}T^{-1}u_{-1}'\Sigma^{-1}\Delta u + Q_T^M(1) - Q_T^M(\bar{\alpha}).$$

For the polynomial trend of Lemma A4, we have from the continuous mapping theorem,

(B3) $$(T^{-2}u_{-1}'u_{-1}, T^{-1}u_T^2, Q_T(\bar{\alpha}) - Q_T(1)) \Rightarrow \omega^2\left(\int W_c^2, W_c^2(1), Q(\bar{c}) - Q(0)\right),$$

where Q is defined in (A10). Thus, from (A8) and the argument leading to (B2),

(B4) $$L_T^* \Rightarrow L^* \equiv \bar{c}^2\int W_c^2 - \bar{c}W_c^2(1) + Q(0) - Q(\bar{c}) + \bar{c}.$$

For polynomial trend, L^* is a function of c, $\bar{c}$ and q; it does not depend on Σ at all. When $q = 1$ so $z_t = 1$, $Q(\bar{c}) = 0$ and the result is the same as in part (a).

(c) When $y_t = \beta_0 + \beta_1 t + u_t$, the function $h(s, \bar{c})$ in (A10) is given by $(1 - \bar{c}s)$ so $\int h(\bar{c})^2 = 1 - \bar{c} + \bar{c}^2/3$ and $\int h(\bar{c})(dW_c - \bar{c}W_c) = (1 - \bar{c})W_c(1) + \bar{c}^2\int sW_c(s)$. After considerable algebraic manipulation, we find the following alternative expressions for (B4):

(B5) $$\begin{aligned} L^* - \bar{c} &= \bar{c}^2\int W_c^2 + (1 - \bar{c})W_c^2(1) - (1 - \bar{c} + \bar{c}^2/3)^{-1}\left[(1 - \bar{c})W_c(1) + \bar{c}^2\int sW_c(s)\right]^2 \\ &= \bar{c}^2\left[\lambda\int[W_c(s) - sW_c(1)]^2\,ds + (1 - \lambda)\int\left[W_c(s) - s3\int sW_c(s)\right]^2 ds\right] \\ &= \bar{c}^2\int[W_c(s) - sb_1]^2\,ds + (1 - \bar{c})[W_c(1) - b_1]^2 = \bar{c}^2\int V_c^2(s, \bar{c}) + (1 - \bar{c})V_c^2(1, \bar{c}) \end{aligned}$$

where $\lambda = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$, $b_1 = \lambda W_c(1) + (1-\lambda)3\int sW_c(s)$, and $V_c(s,\bar{c}) = W_c(s) - sb_1$. The process $V_c(s,\bar{c})$ has a simple interpretation. Let y_t^μ be the detrended series

$$y_t^\mu = y_t - \hat{\beta}_0^M - \hat{\beta}_1^M t = u_t - (\hat{\beta}_0^M - \beta_0) - (\hat{\beta}_1^M - \beta_1)t$$

where $\hat{\beta}_0^M$ and $\hat{\beta}_1^M$ are the estimates that minimize $L(\bar{\alpha}, \beta)$ when $\bar{\alpha} = 1 + T^{-1}\bar{c}$. From the algebra leading to (A11), we find that $\hat{\beta}_0^M$ is stochastically bounded and $T^{1/2}(\hat{\beta}_1^M - \beta_1) \Rightarrow \omega b_1$. Hence, $V_c(s,\bar{c})$ is the limiting representation of the standardized detrended series $T^{-1/2}\omega^{-1}y_{[sT]}^\mu$. By Lemma A4, OLS estimates of β in the polynomial trend case are asymptotically equivalent to GLS estimates. Thus the same interpretation holds when detrending is done with OLS.

PROOF OF THEOREM 2: From (B1) and the fact that $S(a) = u_a'u_a - Q_T(a)$,

$$\text{(B6)} \qquad \hat{\omega}^2 P_T = S(\bar{\alpha}) - S(1)(1+\bar{c}T^{-1}) = \bar{c}^2 T^{-2} u_{-1}'u_{-1} - \bar{c}T^{-1}u_T^2 + Q_T(1)(1+\bar{c}T^{-1}) - Q_T(\bar{\alpha}).$$

The limiting results (A10) and (B3) follow from the fact that $T^{-1/2}u_{[sT]} \Rightarrow \omega W_c(s)$ and that $T^{-1}v'v \xrightarrow{p} \gamma(0)$. Since these limits also follow from Condition C, we have

$$\text{(B7)} \qquad P_T \Rightarrow \hat{\omega}^{-2}\omega^2\left[\bar{c}^2\int W_c^2 - \bar{c}W_c^2(1) + Q(0) - Q(\bar{c})\right].$$

Comparing (B4) and (B7), we see that $P_T + \bar{c} \Rightarrow L^*$ as long as plim $\hat{\omega}^2 = \omega^2$.

PROOF OF THEOREM 3: It suffices to show that $\hat{\omega}^2 P_T \xrightarrow{p} 0$ when $T \to \infty$ with $|\alpha| < 1$ and fixed. Since the initial condition is asymptotically negligible, $\{u_t\}$ behaves like a stationary process. Cf. Anderson (1971, Section 5.5.2). Thus $T^{-2}u_{-1}'u_{-1} \xrightarrow{p} 0$ and $T^{-1}u_T^2 \xrightarrow{p} 0$. From (A9), $T^{-1/2}X'\xi = T^{-1/2}x_T u_T - T^{-3/2}(T\Delta X + \bar{c}X)'u_{-1} \xrightarrow{p} 0$, so both $Q_T(1)$ and $Q_T(\bar{\alpha})$ converge to u_1^2. It follows from (B6) that $\hat{\omega}^2 P_T \xrightarrow{p} 0$.

REFERENCES

ANDERSON, T. W. (1971): *The Statistical Analysis of Time Series*. New York: Wiley.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

BANERJEE, A., J. J. DOLADO, J. W. GALBRAITH, AND D. F. HENDRY (1993): *Co-integration, Error Correction, and the Econometric Analysis of Non-stationary Data*. Oxford: Oxford University Press.

BHARGAVA, A. (1986): "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies*, 53, 369–384.

CHAN, N. H., AND C. Z. WEI (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics*, 15, 1050–1063.

DAVIES, R. B. (1973): "Asymptotic Inference in Stationary Gaussian Time-Series," *Advances in Applied Probability*, 5, 469–497.

DEJONG, D. N., J. C. NANKERVIS, N. E. SAVIN, AND C. H. WHITEMAN (1992): "The Power Problems of Unit Root Tests in Time Series with Autoregressive Errors," *Journal of Econometrics*, 53, 323–343.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

DUFOUR, J.-M., 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.

DZHAPARIDZE, K. (1985): *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. New York: Springer-Verlag.

ELLIOTT, G. (1993): "Efficient Tests for a Unit Root when the Initial Observation is Drawn from its Unconditional Distribution," unpublished manuscript.

ENGLE, R. F. (1984): "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in *Handbook of Econometrics, Vol. II*, ed. by Z. Griliches and M. Intriligator. New York: North Holland.

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

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

——— (1988): "Towards a Theory of Point Optimal Testing," *Econometric Reviews*, 6, 169–218.

LEHMANN, E. (1959): *Testing Statistical Hypotheses*. New York: Wiley.

LUMSDAINE, R. L. (1995): "Finite Sample Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 13, 1–10.

NABEYA, S., AND K. TANAKA (1990): "Limiting Power of Unit-Root Tests in Time-Series Regression," *Journal of Econometrics*, 46, 247–271.

PERRON, P. (1996): "The Adequacy of Asymptotic Approximations in the Near-Integrated Autoregressive Model with Dependent Errors," *Journal of Econometrics*, 70, 317–350.

PHILLIPS, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for a Unit Root 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, 971–1001.

ROBINSON, P. M. (1994): "Efficient Tests of Nonstationary Hypotheses," *Journal of the American Statistical Association*, 89, 1420–1437.

SAIKKONEN, P., AND R. LUUKKONEN (1993): "Point Optimal Tests for Testing the Order to Differencing in ARIMA Models," *Econometric Theory*, 9, 343–362.

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.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

SCHWERT, G. W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147–159.

STOCK, J. H. (1994): "Unit Roots and Trend Breaks in Econometrics," in *Handbook of Econometrics*, Vol. 4, ed. by R. F. Engle and D. McFadden. New York: North Holland, pp. 2740–2841.

ZYGMUND, A. (1968): *Trigonometric Series, Vol. 1*. Cambridge: Cambridge University Press.