Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power

Hansen, Bruce E.

Working Paper No. 378
May 1994

Rethinking the Univariate Approach to Unit Root Testing:
Using Covariates to Increase Power

Bruce E. Hansen

Rochester Center for Economic Research
Working Paper No. 378

Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power

Bruce E. Hansen
Department of Economics
Harkness Hall
University of Rochester
Rochester, NY 14627
(716) 275-7307
hanb@troi.cc.rochester.edu

May 9, 1994

Abstract

In the context of testing for a unit root in a univariate time series, the convention is to ignore information in related time series. This paper shows that this convention is quite costly, as large power gains can be achieved by including correlated stationary covariates in the regression equation.

The paper derives the asymptotic distribution of ordinary least squares (OLS) estimates of the largest autoregressive root and its t statistic. The asymptotic distributions are not the conventional "Dickey-Fuller" distributions, instead they are convex combinations of the Dickey-Fuller distributions and the standard normal, the mixture depending on the correlation between the equation error and the regression covariates. The local asymptotic power functions associated with these test statistics suggest enormous gains over the conventional unit root tests. A simulation study and empirical application illustrate the potential of the new approach.

1 Introduction

A refrain often heard in applied macroeconometric circles is that "Unit root tests have low power." I believe that this view may be partly a result of the convention of testing for unit roots in *univariate* time series. This convention ignores relevant information in multivariate data sets. It turns out to be the case that multivariate data can indeed be very informative in regards to the question of univariate unit roots.

Consider the simple AR(1) model

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

where u_t is iid $(0, \sigma_u^2)$. The hypothesis of a unit root in this model is typically tested by the OLS t-statistic for $\delta = 0$. This test, known as the Dickey-Fuller test, is widely believed to be the best (or approximately so) classical procedure in this context.

It is rare, however, that we observe the time series y_t in isolation. More typically, we observe related (correlated) time series, which we will collect in the vector Δx_t. Our maintained assumption is that x_t is $I(1)$, so that Δx_t is $I(0)$. Let us assume that

$$\begin{pmatrix} u_t \\ \Delta x_t \end{pmatrix} \sim \text{iid} \left(0, \begin{pmatrix} \sigma_u^2 & \sigma_{xu} \\ \sigma_{xu} & \sigma_x^2 \end{pmatrix} \right).$$

Set $b = \sigma_x^{-2}\sigma_{xu}$, and $e_t = u_t - \Delta x_t b$, so that (1), supplemented by the observable Δx_t, becomes

$$\Delta y_t = \delta y_{t-1} + \Delta x_t' b + e_t. \tag{2}$$

Under the assumptions, the parameter δ retains the same meaning as in (1). An important difference in the equations, however, is that the error variance in (2), $\sigma_e^2 = \sigma_u^2 - \sigma_{xu}^2/\sigma_x^2$ will be smaller than σ_u^2 in (1) (unless $\sigma_{xu} = 0$, in which case the variances are equal). This suggests that the regression parameters will be more precisely estimated (at least in large samples) if OLS is applied to (2) rather than (1). This means that confidence intervals will be smaller and test statistics more powerful.

Another interesting question is: What is the distribution theory for the t-statistic for δ in (2)? Sometimes researchers report such statistics, and the presumption has been that the

Dickey-Fuller distribution is appropriate. The standard intuition is that stationary covariates Δx_t do not affect the limiting distribution other than to correct for serial correlation in u_t. We show below that this belief is incorrect.

The results in this paper relate to some previous results in the literature. Kremers, Ericsson and Dolado (1992) discuss the model

$$\Delta z_{1t} = a_0 \Delta z_{2t} + a_1 (z_{1t-1} - z_{2t-1}) + \epsilon_{1t} \tag{3}$$

$$\Delta z_{2t} = \epsilon_{2t}$$

with ϵ_{1t} and ϵ_{2t} iid uncorrelated normal random variables. We can see that (3) falls in the class (2) by setting $y_t = z_{1t} - z_{2t}$, $x_t = z_{2t}$, $\delta = a_1$, and $b = a_0 - 1$. The results which we derive below can therefore be seen as generalizations their results, although it should be emphasized that Kremers, et. al., were discussing tests for cointegration, and not for univariate unit roots.

Horvath and Watson (1993) have recently proposed tests for cointegration when the cointegrating vector is known a priori. While their tests are primarily motivated as tests for cointegration, they could be used to test for stationarity in a particular variable, by setting the "cointegrating vector" equal to the unit vector. The tests and distributional theory they obtain are different from those analyzed in this paper.

A final interesting possibility has been mentioned by Johansen and Juselius (1992). Conditioning on a known cointegrating rank of the data, they propose tests that some of the cointegrating vectors are known. Again setting the known cointegrating vector equal to the unit vector, this allows testing the null hypothesis that a particular series y_t is stationary against the alternative that it is integrated (and cointegrated with some other series x_t). This flips the null and alternative from that considered in our paper, and requires that y_t and x_t are cointegrated when y_t is $I(1)$, which we do not require. Again, it appears that our tests are complementary to those of Johansen-Juselius.

Section 2 introduces a generalized version of (2), allowing for lagged dependent variables and deterministic components. The Gaussian asymptotic power envelope for the test of $\delta = 0$ is derived and compared with the power envelope of model (1). The asymptotic distributions of the OLS estimates of (2) are also found under local alternatives to a unit root. Section 3 discusses test statistics. Both coefficient and t-statistics are introduced, and their asymptotic distributions derived under the null hypothesis and local alternatives. This permits an analysis of asymptotic local power. The sensitivity of the results to misspecification of the order of integration of x_t is also discussed. Section 4 reports a simulation-based study of the finite sample distribution of the test statistics, using data generated from a VAR(1). Section 5 applies the tests to some long time series. We find that real per capita GNP and the unemployment rate are $I(0)$ but highly persistent, that industrial production is $I(1)$. The Appendix contains the mathematical proofs.

2 Regression Framework

2.1 Model and Assumptions

The univariate series y_t consists of a deterministic and stochastic component:

$$y_t = d_t + S_t \tag{4}$$

where the deterministic component is one of the following: $d_t = 0$, $d_t = \mu$, or $d_t = \mu + \theta t$. The stochastic component S_t is modeled by an autoregression with observed covariates x_t :

$$a(L)\Delta S_t = \delta S_{t-1} + b(L)'(\Delta x_t - \mu_x) + e_t \tag{5}$$

where $a(L) = 1 - a_1 L - a_2 L^2 - \cdots - a_p L^p$ is a p-th order polynomial in the lag operator and $\mu_x = E(\Delta x_t)$. Δx_t is an m-vector and $b(L) = b_{q_2} L^{-q_2} + \cdots + b_{q_1} L^{q_1}$ is a lag polynomial allowing for (but not requiring) either or both leads and lags of Δx_t to be included in (5).

Assumption 1 *For some* $p > r > 2$,

1. $\{\Delta x_t, e_t\}$ *is covariance stationary and strong mixing with mixing coefficients* α_m *which satisfy* $\sum_{m=1}^{\infty} \alpha_m^{1/r-1/p}$;
2. $\sup_t E\left[|\Delta x_t|^p + |e_t|^p\right] < \infty$;
3. $E(\Delta x_{t-k} e_t) = 0$ *for* $q_1 \leq k \leq q_2$;
4. $E(e_t e_{t-k}) = 0$ *for all* $k \geq 1$;
5. *The roots of* $a(L)$ *all lie outside the unit circle.*

Assumptions 1.1 and 1.2 are conventional weak dependence and moment restrictions. Assumption 1.3 states that the regressors in (5) are orthogonal to the regression error. This can be achieved simply by appropriate definition for the lag polynomial $b(L)$ (by linear projection). Assumption 1.4 implies that the lag polynomial $a(L)$ is sufficiently large to whiten the errors. It should be possible to extend the analysis to allow for an infinite order polynomial which is approximated in finite samples by a p which grows with sample size, following the technique of Berk (1974) and Said and Dickey (1983).

The specification for $b(L)$ allows for past, current, and even future values of Δx_t to enter the regression equation. In many applications (such as a standard VAR in y_t and Δx_t, in which case (5) is one equation from the VAR) only lagged values of Δx_t will enter the regression, so $b(L)$ will take the form $b_1 L + \cdots + b_{q_1} L^{q_1}$.

2.2 Transformations and Definitions

The behavior of S_t is determined by the random vector

$$u_t = \begin{pmatrix} \Delta x_t - \mu_x \\ e_t \end{pmatrix},$$

which has long-run covariance matrix

$$\Omega_u = \begin{pmatrix} \Omega_{xx} & \Omega'_{xe} \\ \Omega_{xe} & \Omega_{ee} \end{pmatrix} = \sum_{k=-\infty}^{\infty} E\left(u_t u'_{t-k}\right).$$

The vector u_t satisfies the conditions of Herrndorf (1984), so the partial sums of u_t converge weakly to a Brownian motion with covariance matrix Ω_u.

Now define the random variable

$$v_t = b(L)'(\Delta x_t - \mu_x) + e_t$$

and the vector $\eta_t = (v_t \; e_t)'$, which has long-run covariance matrix

$$\Omega_\eta = \begin{pmatrix} b'\Omega_{xx}b + b'\Omega_{xe} + \Omega'_{xe}b + \Omega_{ee} & \Omega'_{xe}b + \Omega_{ee} \\ b'\Omega_{xe} + \Omega_{ee} & \Omega_{ee} \end{pmatrix} \equiv \begin{pmatrix} \sigma_v^2 & \rho\sigma_v\sigma_e \\ \rho\sigma_v\sigma_e & \sigma_e^2 \end{pmatrix}. \tag{6}$$

where $\rho = \sigma_{ve}/\sigma_v\sigma_e$ is the long-run correlation between v_t and e_t, and $b = b(1)$.

Since $b(L)$ is a finite-order polynomial, it follows that the partial sums of η_t converge weakly to a Brownian motion with covariance matrix Ω_η. Furthermore, by the standard decomposition implied by definition (6), we can write this limit as

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \eta_t \Rightarrow \begin{pmatrix} \sigma_v W_1(r) \\ \sigma_e \left(\rho W_1(r) + (1-\rho^2)^{1/2} W_2(r)\right) \end{pmatrix}, \tag{7}$$

where W_1 and W_2 are independent standard Brownian motions.

The zero-frequency squared correlation ρ^2 measures the relative contribution (at the zero frequency) of Δx_t to v_t. One extreme is obtained when $b = 0$, then $v_t = e_t$ and $\rho^2 = 1$. The other extreme is obtained when the regressors x_t explain nearly all the zero-frequency movement in v_t, in which case $\rho^2 \approx 0$. We technically exclude the limiting case of $\rho^2 = 0$ by requiring that $\rho^2 > 0$. We also define the variance ratio

$$R^2 = \frac{\sigma_e^2}{\sigma_v^2}.$$

An important special case arises when e_t is uncorrelated with Δx_{t-k} for all k, for then $\sigma_{ve} = \sigma_e^2$ and $\rho^2 = R^2$. This situation holds when q_1 and q_2 are both large. Since this is a property which we might expect to see in a well-specified regression, we consider $\rho^2 = R^2$ to be a special case of leading interest.

2.3 Preliminary Asymptotic Results

Our asymptotic theory will be based on "local-to-unity asymptotics," following the technique of Phillips (1987b) and Chan and Wei (1987). Model (4)-(5) contains a unit root when $\delta = 0$, which is the null hypothesis of interest:

$$H_0 : \delta = 0.$$

We allow for local departures from the null hypothesis by setting

$$\delta = -ca(1)/T. \tag{8}$$

The null holds when $c = 0$, and holds "locally" as $T \to \infty$ for $c \neq 0$. In a fixed sample, however, (8) is simply a reparameterization.

The asymptotic theory for near-integrated processes utilizes diffusion representations. We will use the following notation. For any continuous stochastic process $Z(r)$ and any constant c, we define the stochastic process $Z^c(r)$ as the solution to the stochastic differential equation

$$dZ^c(r) = -cZ^c(r) + dZ(r).$$

We can now derive asymptotic representations for the stochastic component S_t and its sample moments.

Lemma 1

1. $\frac{1}{\sqrt{T}} S_{[Tr]} \Rightarrow a(1)^{-1}\sigma_v W_1^c(r)$.

2. $\frac{1}{T^2}\sum_{t=2}^{T} S_{t-1}^2 \Rightarrow a(1)^{-2}\sigma_v^2 \int_0^1 (W_1^c)^2$.

3. $\frac{1}{T}\sum_{t=2}^{T} S_{t-1}e_t \Rightarrow a(1)^{-1}\sigma_v\sigma_e \left(\rho\int_0^1 W_1^c dW_1 + (1-\rho^2)^{1/2}\int_0^1 W_1^c dW_2\right)$.

2.4 Power Envelope

The Gaussian power envelope for the unit root testing problem in model (5) can be easily derived. Start by assuming that the nuisance parameters a, $b(L)$, μ, μ_x, and θ are known, and fix a point alternative $\overline{\delta}$. Then assume that the error e_t is iid $N(0,\sigma_e^2)$ and is independent of Δx_t at all leads and lags. Finally, assume that the initial condition S_0 is fixed. This allows the construction of a Gaussian likelihood. The likelihood ratio test for a unit root ($\delta = 0$ versus $\delta < 0$) rejects for small values of

$$LR = \frac{1}{\sigma_v^2}\sum_{t=2}^{T}\left[\left(a(L)\Delta S_t - \overline{\delta}S_{t-1} - b(L)'\Delta x_t\right)^2 - \left(a(L)\Delta S_t - b(L)'\Delta x_t\right)^2\right], \tag{9}$$

and the Neyman-Pearson Lemma shows that this is the most powerful test of the simple hypothesis. Setting $\overline{\delta} = -\overline{c}a(1)/T$, the large sample distribution of $LR(\overline{\delta})$ can be found fairly directly from Lemma 3.

Theorem 1

$$LR \Rightarrow \left(\overline{c}^2 - 2\overline{c}c\right)\int_0^1 (W_1^c)^2 + 2R\overline{c}\left(\rho\int_0^1 W_1^c dW_1 + \left(1-\rho^2\right)^{1/2}\int_0^1 W_1^c dW_2\right).$$

Note that the limiting distribution of the likelihood ratio statistic depends on the parameters $(c, \overline{c}, R^2, \rho^2)$. Under the null ($c = 0$) the distribution depends on $(\overline{c}, R^2, \rho^2)$. The point optimal likelihood ratio statistic sets $\overline{c} = c$, so the asymptotic distribution under the alternative depends on (c, R^2, ρ^2). This means that the Gaussian power envelope (maximal rejection frequency for a test of fixed size, traced out as a function of the alternative c) for this testing problem depends on two nuisance parameters, R^2 and ρ^2.

Note that when $\rho^2 = R^2 = 1$ the limiting distribution of Theorem 1 simplifies to

$$LR \Rightarrow \bar{c}^2 \int_0^1 (W_1^c)^2 + \bar{c} W_1^c(1)^2 - \bar{c}$$

which is the distribution found for the point optimal Gaussian likelihood ratio in a autoregressive model without covariates (see Elliott, Rothenberg and Stock (1992)). In this case, the power envelope for model (5) equals that of the Dickey-Fuller model

Figure 1 plots the power envelope[1] for the leading case $R^2 = \rho^2$ and for a range of values of ρ^2. The lowest curve is for $\rho^2 = 1$. The curves are strictly increasing as ρ^2 falls. In fact, the increase in the power envelope due to a decrease in ρ^2 is quite dramatic. Take the alternative $c = 5$, which corresponds to an autoregressive root of .95 when $T = 100$. The power envelope for the standard autoregressive model (when $\rho^2 = 1$) is 33%, increasing to 51% when $\rho^2 = .7$, and to 90% when $\rho^2 = .3$. By itself, Figure 1 does not demonstrate an increase in power of feasible tests (we leave this to Section 3), but it does show the enormous potential of allowing for covariates in unit root tests.

2.5 Least Squares Estimation

When $d_t = 0$, we have

$$a(L)\Delta y_t = \delta y_{t-1} + b(L)'\Delta x_t + e_t. \tag{10}$$

When $d_t = \mu$, the model is

$$a(L)\Delta y_t = \mu^* + \delta y_{t-1} + b(L)'\Delta x_t + e_t, \tag{11}$$

where $\mu^* = -\delta\mu - b'\mu_x$, and when $d_t = \mu + \theta t$, the model is

$$a(L)\Delta y_t = \mu^* + \theta^* t + \delta y_{t-1} + b(L)'\Delta x_t + e_t, \tag{12}$$

where $\mu^* = a(1)\theta - \delta\mu - b'\mu_x$ and $\theta^* = -\delta\theta$. (10), (11) or (12) can be estimated by ordinary least squares (OLS). Let $\hat{\delta}$, $\hat{\delta}^\mu$, and $\hat{\delta}^\tau$ denote the OLS estimates of δ from these three regressions.

[1]The power envelope was calculated for each value of ρ^2 shown and at $c = 1, 2, ..., 16$. The distributions were approximated by calculations from samples of size 1000 with iid Gaussian innovations. To calculate the envelope at each ρ^2 and c, 40,000 draws were made under the null to compute the 5% critical value, and 20,000 draws were made under the alternative to compute the power.

Theorem 2

$$T\left(\hat{\delta}-\delta\right) \Rightarrow a(1)R\left(\rho\frac{\int_0^1 W_1^c dW_1}{\int_0^1 (W_1^c)^2} + \left(1-\rho^2\right)^{1/2}\frac{\int_0^1 W_1^c dW_2}{\int_0^1 (W_1^c)^2}\right). \tag{13}$$

The asymptotic distributions for $T\left(\hat{\delta}^{\mu}-\delta\right)$ *and* $T\left(\hat{\delta}^{\tau}-\delta\right)$ *are the same as (13), except that* $W_1^c(r)$ *is replaced by*

$$W_1^{c\mu}(r) = W_1^c(r) - \int_0^1 W_1^c,$$

and

$$W_1^{c\tau}(r) = W_1^c(r) - \left(4\int_0^1 W_1^c - 6\int_0^1 W_1^c s\right) + \left(12\int_0^1 W_1^c s - 6\int_0^1 W_1^c\right) r,$$

respectively.

Given that the asymptotic distribution given by (13) depends on the nuisance parameters R^2 and ρ^2, it is important to be able to estimate them consistently. Natural estimators are given by

$$\hat{\rho}^2 = \frac{\hat{\sigma}_{ve}^2}{\hat{\sigma}_v^2\hat{\sigma}_e^2} \tag{14}$$

and

$$\hat{R}^2 = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_v^2}$$

where

$$\hat{\Omega} = \begin{pmatrix} \hat{\sigma}_v^2 & \hat{\sigma}_{ve} \\ \hat{\sigma}_v^2 & \hat{\sigma}_e^2 \end{pmatrix} = \sum_{k=-M}^{M} w(k/M)\frac{1}{T}\sum_t \hat{\eta}_{t-k}\hat{\eta}_t',$$

$\hat{\eta}_t = (\hat{v}_t\ \hat{e}_t)'$, $\hat{e}_t = \hat{a}(L)\Delta y_t - \hat{\delta}y_{t-1} - \Delta x_t'\hat{b}$, and $\hat{v}_t = \Delta x_t'\hat{b} + \hat{e}_t$. The function $w(\cdot)$ may be any kernel weight function which produces positive semi-definite covariance matrices, such as the Bartlett or Parzen kernels, and M is a bandwidth selected to diverge to infinity slowly with sample size. Conditions under which these estimates are consistent are given in Hansen (1992b), and selection rules for M which minimize asymptotic mean squared error are given in Andrews (1991).

3 Testing for a Unit Root

3.1 Test Statistics

There are two natural test statistics for the hypothesis of a unit root

$$H_0 : \delta = 0$$

in models (10)-(12). One is the t-statistic

$$t(\hat{\delta}) = \frac{\hat{\delta}}{s(\hat{\delta})}, \tag{15}$$

where $s(\hat{\delta})$ is the OLS standard error for $\hat{\delta}$. Alternatively, for models (11) or (12), we denote the t-statistic for δ by $t(\hat{\delta}^{\mu})$ or $t(\hat{\delta}^{\tau})$, respectively. The other statistic is the normalized coefficient

$$z(\hat{\delta}) = \frac{T\hat{\delta}}{\hat{a}(1)\hat{R}}. \tag{16}$$

For models (11) and (12), we denote the coefficient statistic by $z(\hat{\delta}^{\mu})$ and $z(\hat{\delta}^{\tau})$, respectively. We now give the asymptotic distributions of these test statistics under the local-to-unity structure (8).

Theorem 3

1.

$$z(\hat{\delta}) \Rightarrow -\frac{c}{R} + \rho \frac{\int_0^1 W_1^c dW_1}{\int_0^1 (W_1^c)^2} + \left(1-\rho^2\right)^{1/2} \frac{\int_0^1 W_1^c dW_2}{\int_0^1 W_c^2}, \tag{17}$$

2.

$$t(\hat{\delta}) \Rightarrow -\frac{c}{R}\left(\int_0^1 (W_1^c)^2\right)^{1/2} + \rho \frac{\int_0^1 W_1^c dW_1}{\left(\int_0^1 (W_1^c)^2\right)^{1/2}} + \left(1-\rho^2\right)^{1/2} N(0,1), \tag{18}$$

where in (18) the $N(0,1)$ variable is independent of W_1. The asymptotic distributions for $z(\hat{\delta}^{\mu})$, $z(\hat{\delta}^{\tau})$, $t(\hat{\delta}^{\mu})$ and $t(\hat{\delta}^{\tau})$ are the same as (17) and (18), except that W_1^c is replaced by $W_1^{c\mu}$ and $W_1^{c\tau}$, respectively.

3.2 Asymptotic Null Distributions

Corollary 1 *Under the null hypothesis $\delta = 0$,*

1.

$$z(\hat{\delta}) \Rightarrow \rho \frac{\int_0^1 W_1 dW_1}{\int_0^1 W_1^2} + \left(1-\rho^2\right)^{1/2} \frac{\int_0^1 W_1 dW_2}{\int_0^1 W_1^2}, \tag{19}$$

2.

$$t(\hat{\delta}) \Rightarrow \rho \frac{\int_0^1 W_1 dW_1}{\left(\int_0^1 W_1^2\right)^{1/2}} + \left(1-\rho^2\right)^{1/2} N(0,1). \tag{20}$$

The asymptotic distributions for $z(\hat{\delta}^{\mu})$, $z(\hat{\delta}^{\tau})$, $t(\hat{\delta}^{\mu})$, and $t(\hat{\delta}^{\tau})$ are the same as (19) and (20), except that W_1 is replaced by

3.

$$W_1^{\mu}(r) = W_1(r) - \int_0^1 W_1,$$

and

$$W_1^{\tau}(r) = W_1(r) - \left(4\int_0^1 W_1 - 6\int_0^1 W_1 s\right) + \left(12\int_0^1 W_1 s - 6\int_0^1 W_1\right) r,$$

respectively.

The null distributions are not the conventional "Dickey Fuller" distributions for $\rho \neq 1$. For the t-statistic, the null distribution is convex mixture of the standard normal and the standard "Dickey-Fuller distribution," with the weights determined by ρ^2. As $\rho^2 \to 1$, we find the standard Dickey-Fuller, and as $\rho^2 \to 0$ we obtain the standard normal distribution. Estimated[2] 1%, 5%, and 10% critical values for $t(\hat{\delta})$ and $z(\hat{\delta})$ are given in Tables 1 and 2. To use the statistics (15) and (16), the estimator $\hat{\rho}^2$ from (14) must be used to select the appropriate row from the Tables. Since $\hat{\rho}^2$ is consistent for ρ^2, it can be used to select the row.

The observation that the conventional Dickey-Fuller critical values are inappropriate when a regression has stochastic covariates has not been made before. This alone is a useful implication of our analysis. We can see from the form of the asymptotic distributions that

[2]The critical values were calculated from 60,000 draws generated from samples of size 1000 with iid Gaussian innovations.

Table 1: Asymptotic Critical Values of Covariate t-Tests

	Standard			Demeaned			Detrended		
ρ^2	1%	5%	10%	1%	5%	10%	1%	5%	10%
1.0	-2.57	-1.94	-1.62	-3.43	-2.86	-2.57	-3.96	-3.41	-3.13
0.9	-2.57	-1.94	-1.61	-3.39	-2.81	-2.50	-3.88	-3.33	-3.04
0.8	-2.57	-1.94	-1.60	-3.36	-2.75	-2.46	-3.83	-3.27	-2.97
0.7	-2.55	-1.93	-1.59	-3.30	-2.72	-2.41	-3.76	-3.18	-2.87
0.6	-2.55	-1.90	-1.56	-3.24	-2.64	-2.32	-3.68	-3.10	-2.78
0.5	-2.55	-1.89	-1.54	-3.19	-2.58	-2.25	-3.60	-2.99	-2.67
0.4	-2.55	-1.89	-1.53	-3.14	-2.51	-2.17	-3.49	-2.87	-2.53
0.3	-2.52	-1.85	-1.51	-3.06	-2.40	-2.06	-3.37	-2.73	-2.38
0.2	-2.49	-1.82	-1.46	-2.91	-2.28	-1.92	-3.19	-2.55	-2.20
0.1	-2.46	-1.78	-1.42	-2.78	-2.12	-1.75	-2.97	-2.31	-1.95

Table 2: Asymptotic Critical Values of Covariate Coefficient Tests

	Standard			Demeaned			Detrended		
ρ^2	1%	5%	10%	1%	5%	10%	1%	5%	10%
1.0	-13.8	-8.1	-5.7	-20.7	-14.1	-11.3	-29.5	-21.8	-18.3
0.9	-13.6	-7.8	-5.5	-20.0	-13.7	-10.9	-28.3	-21.0	-17.6
0.8	-13.0	-7.5	-5.3	-19.0	-13.1	-10.3	-27.2	-20.0	-16.7
0.7	-12.4	-7.3	-5.1	-18.8	-12.4	-9.9	-26.1	-19.2	-15.9
0.6	-11.9	-7.0	-5.0	-18.0	-12.0	-9.4	-25.2	-18.4	-15.2
0.5	-11.7	-6.8	-4.8	-17.0	-11.4	-9.0	-24.0	-17.2	-14.2
0.4	-11.0	-6.4	-4.5	-16.2	-10.7	-8.4	-22.5	-16.2	-13.3
0.3	-10.3	-6.0	-4.2	-15.0	-9.9	-7.7	-20.8	-14.9	-12.2
0.2	-9.6	-5.5	-3.8	-13.7	-9.0	-7.0	-18.8	-13.4	-10.9
0.1	-8.7	-5.0	-3.5	-12.2	-7.9	-6.0	-16.7	-11.7	-9.3

the conventional asymptotic critical values are conservative, implying that tests mistakenly based on the Dickey-Fuller critical values will have less power than they could.

3.3 Asymptotic Local Power Functions

Theorem 3 gives the asymptotic distribution of the t-statistic and coefficient test statistic under the local-to-unity alternative (8). The expressions show that the asymptotic distributions, and hence the asymptotic local power, depends on c, ρ^2, and R^2. Figure 2 displays the power functions[3] and power envelope for these two tests when $d_t = 0$, for $\rho^2 = 1.0$, 0.7,

[3]The power functions were calculated for each ρ^2 and $c = 1, 2, ..., 16$ from 20,000 samples of size 1000 with iid Gaussian innovations, using the asymptotic critical values from Tables 1 and 2.

0.4, and 0.1, setting $R^2 = \rho^2$. We can see that the power of the two tests are quite close to one another and the power envelope. Interestingly, for $\rho^2 < 1$ and small c, the power of $t(\hat{\delta})$ exceeds that of $z(\hat{\delta})$, and lies very close to the power envelope, but the reverse is true for large c. (For $\rho^2 = 1$, all three curves are virtually identical.)

Figures 3 and 4 display the same set of power functions, but for the cases $d_t = \mu$ and $d_t = \mu + \theta t$, respectively. In the case of a mean correction ($d_t = \mu$), the coefficient test $z(\hat{\delta}^\mu)$ does much better than the t-test $t(\hat{\delta}^\mu)$ for ρ^2 close to 1, but the t-test has higher power for small ρ^2 and small c. The same pattern is seen for the case of a trend correction ($d_t = \mu + \theta t$) in Figure 4, but the t-test has higher power for a wider range of values of c and ρ^2. For both test statistics, the power curves for small values of ρ^2 are far above those of the conventional Dickey-Fuller tests, which are given by the curves for $\rho^2 = 1$.

Figure 5 explores the impact of letting ρ^2 differ from R^2, displaying six power functions for $t(\hat{\delta}^\mu)$, setting $\rho^2 = .1$ and $\rho^2 = .7$, and letting R^2 take three values above, equal, and below ρ^2. First, examine the curves for $\rho^2 = .7$, which corresponds to mildly successful regressors. Allowing $R^2 = .35$ achieves a major improvement in power relative to $R^2 = .7$, while setting $R^2 = 1$ shows a substantial decline in power, nearing the power function for the Dickey-Fuller t-test. Next, examine the curves for $\rho^2 = .1$. Here the impact of $R^2 \neq \rho^2$ appears less dramatic, although the impact is still substantial. As the asymptotic theory predicts, a lower R^2 implies a higher local power function. The coefficient tests had similar behavior to the t-tests and are not displayed.

Figures 2 through 5 show that the regression tests come close to the power envelope for the case without a deterministic component, and that neither the t-test nor coefficient test strictly dominate the other. The most important message is that enormous power gains can be achieved by finding appropriate covariates x_t which produce a low ρ^2 and/or a low R^2.

3.4 Over-Differenced Regressors

The theory of the previous sections is based upon the strong assumption that the series Δx_t is stationary. Due to the classic spurious regression problem, it is fairly obvious that it is important to not include in the regression equation an integrated process, hence most included variables will be first-differenced to induce stationarity (and hence the notation

Δx_t). This of course raises the possibility that Δx_t could be "over-differenced." Suppose, for example, that x_t is $I(0)$, and the regression included Δx_t, which is thus $I(-1)$. It is easy to show that $\rho^2 = 1$, so the asymptotic power of our tests is equivalent to that of the Dickey-Fuller tests.

This conclusion appears more pessimistic than warranted. The reason why x_t is differenced is because it is highly serially correlated. To develop a better finite sample approximation to the power function, let us assume that x_t is near-integrated. Specifically, assume that x_t satisfies

$$\Delta x_t = -\frac{g}{T} x_{t-1} + x_t^* \tag{21}$$

with x_t^* satisfying the assumptions we previously made about Δx_t. When $g = 0$, x_t is $I(1)$ and our model is not misspecified. As we allow g to depart from zero, we induce a continuous distortion away from the model's assumptions, and can examine the impact of misspecification. When $g = \infty$, x_t is $I(0)$ and the power function for the regression tests should equal that of the Dickey-Fuller tests.

The model is essentially the same as before, except that there is one more parameter, g. The asymptotic local power function for the regression test will therefore depend on ρ^2, R^2 and g. Figure 6 plots the power functions for $\rho^2 = .1$ and g equal to 0, 2, 4, 8 and 16, and the power function for $\rho^2 = 1$ and $g = 0$, which is should be the same as setting $\rho^2 = .1$ and $g = \infty$. The qualitative impact of letting $g > 0$ is as anticipated: increasing g moves the power function down and away from that when $g = 0$, eventually reaching the power function of the Dickey-Fuller test. What is somewhat surprising (at least to the author) is the quantitative finding that the magnitude of the power loss is fairly mild. Even setting $g = 16$ does not lead to a major loss of power. The asymptotic analysis suggests that over-differencing is not likely to be a major specification error.

4 Small Sample Distributions

To demonstrate the performance of the new test statistics in a small sample, I performed a simulation experiment. Data were generated from the model

$$\Delta y_t = -(c/T)y_{t-1} + u_t,$$

$$\begin{pmatrix} u_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} u_{t-1} \\ \Delta x_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix},$$

and

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{21} \\ \sigma_{21} & 1 \end{pmatrix} \right).$$

Each replication discarded the first 100 observations to eliminate start-up effects.

Throughout the experiment, I set $a_{22} = 0$ and $a_{11} = 0$, since theses parameters do not affect the nuisance parameters ρ^2 and R^2 (so long as the VAR is stationary). This leaves three free parameters (σ_{21}, a_{21}, a_{12}) which control the degree of correlation between Δx_t and u_t. The first experiment set all three to zero, the remaining 16 set $\sigma_{21} = 0.4$, and varied a_{21} and a_{12} among $\{-.3,\ 0,\ .3,\ .6\}$.

One Augmented Dickey-Fuller regression and three covariate regressions were considered. All regressions included a constant and two lags of the dependent variable. The three covariate regressions differed in their choice of lags of x_t. The selections were:

Case A: $\{\Delta x_t\}$.

Case B: $\{\Delta x_t,\ \Delta x_{t-1}\}$.

Case C: $\{\Delta x_t,\ \Delta x_{t+1}\}$.

Case D: $\{\Delta x_t,\ \Delta x_{t-1},\ \Delta x_{t+1}\}$.

The asymptotic theory suggests that (to a first approximation) the power of the tests will depend on ρ^2 and R^2, which are complicated functions of the model parameters and the choice of regressors. To calculate these parameters, I used a simulation technique. 10 samples of length 10,000 were generated from each parameterization, and the estimates $\hat{\rho}^2$ and $\hat{R}^2$ calculated using a Parzen kernel and Andrews' (1991) automatic bandwidth estimator, and are reported[4] in Table 3. The results are quite interesting. We can see that it is possible for R^2 to exceed one, and for the addition of extra covariates to increase ρ^2, which may run counter to intuition. For the parameterizations where $a_{12} < 0$ or $a_{21} < 0$, there is no major decrease in ρ^2 or R^2 by inclusion of $\Delta x_t' s$, and there may even be an increase. This points out that simply the presence of correlation between two variables does not mean that the ρ^2 and R^2 measures will be low. It will depend on the nature of the correlation.

To examine the size and power of the test, I ran experiments with samples of size 50, 100, and 250. The null hypothesis obtains by setting $c = 0$, and power was examined by setting $c = 4$, 8, and 15. To conserve space, only the results for $T = 100$ and $c = 0$ and $c = 8$ are reported, as the results for the other sample sizes and alternatives were predictably similar. All experiments used 5000 replication.

To examine size, the 10 tests (t-test and coefficient test for the Dickey-Fuller and four covariate regressions) were compared against asymptotic critical values. For the covariate tests, $\hat{\rho}^2$ was calculated using a Parzen kernel and Andrews' (1991) automatic bandwidth estimator. We find a substantial range of size behavior, with some parameter designs producing over-rejection, and others producing under-rejection. In general, the new covariate tests have more size distortion than the Dickey-Fuller test, and the coefficient tests reject more frequently than the t-tests.

Finally, the power of the tests were examined. Because some of the tests displayed size distortion, the power calculations were done with finite sample critical values, obtained from the simulated data generated under the null hypothesis. The Dickey-Fuller tests have power which is roughly independent of the design, ranging from 17-22% for the t-test and 26-33% for the coefficient test. The power of the covariate tests is much higher than the ADF tests, and is well predicted by ρ^2 and R^2. Indeed, the power gains from inclusion of covariates is quite substantial, reaching to 99% power in one case. Note that it is important to get the "correct" covariates, for major losses in power can be obtained by inclusion or exclusion of

[4]Standard errors (not shown) indicate that the estimates are quite precise.

Table 3: Simulation Design

Design	Parameters			Case A		Case B		Case C		Case D	
	σ_{21}	a_{12}	a_{21}	ρ^2	R^2	ρ^2	R^2	ρ^2	R^2	ρ^2	R^2
1	0	0	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	.4	-.3	-.3	.87	1.31	.97	1.06	.98	1.06	.98	.97
3	.4	-.3	0	.86	1.04	.96	.96	.86	1.04	.96	.96
4	.4	-.3	.3	.85	.86	.86	.78	.66	.90	.71	.72
5	.4	-.3	.6	.85	.76	.97	.84	.48	.77	.59	.68
6	.4	0	-.3	.84	1.05	.84	1.05	.96	.97	.96	.97
7	.4	0	0	.84	.84	.84	.84	.84	.84	.84	.84
8	.4	0	.3	.83	.67	.83	.67	.61	.63	.61	.63
9	.4	0	.6	.84	.55	.84	.55	.38	.46	.38	.46
10	.4	.3	-.3	.86	.89	.68	.80	.98	.90	.73	.72
11	.4	.3	0	.87	.71	.65	.66	.87	.71	.65	.66
12	.4	.3	.3	.86	.53	.63	.42	.59	.43	.38	.40
13	.4	.3	.6	.85	.40	.62	.29	.27	.26	.17	.25
14	.4	.6	-.3	.90	.79	.45	.72	.99	.90	.66	.67
15	.4	.6	0	.91	.62	.47	.46	.90	.62	.46	.46
16	.4	.6	.3	.90	.46	.46	.26	.59	.30	.22	.24
17	.4	.6	.6	.89	.31	.44	.13	.18	.12	.07	.11

covariates.

This experiment suggests that the asymptotic theory provides useful predictions in moderate samples for the performance of the tests. Since the asymptotic theory predicts major power gains from the inclusion of covariates which decrease ρ^2 and R^2, this can be used as a guide in applications to increase the power of unit root tests.

Table 4: Finite Sample Size, $T = 100$

	t Tests					Coefficient Tests				
	Dickey	Covariate Tests				Dickey	Covariate Tests			
Design	Fuller	A	B	C	D	Fuller	A	B	C	D
1	5	5	5	5	5	6	7	7	7	7
2	5	10	6	7	5	6	8	8	7	7
3	5	10	6	10	6	7	7	9	8	9
4	6	10	6	12	6	8	7	10	9	11
5	6	11	6	12	6	7	7	11	9	11
6	5	5	5	5	5	6	7	7	8	8
7	5	5	5	5	5	7	6	7	6	7
8	5	5	5	5	5	7	6	6	6	6
9	5	5	5	4	4	7	6	6	5	5
10	5	2	5	5	6	7	7	9	9	12
11	5	2	5	2	5	7	5	5	6	6
12	6	2	5	1	4	7	5	5	5	5
13	6	3	5	1	4	8	5	5	3	4
14	4	1	4	5	6	6	5	9	9	12
15	5	1	5	1	5	6	4	5	5	5
16	6	2	5	0	5	7	5	5	3	5
17	7	2	6	0	3	8	5	5	3	4

Table 5: Finite Sample Power, $T = 100$, $c = 8$

	t Tests					Coefficient Tests				
	Dickey	Covariate Tests				Dickey	Covariate Tests			
Design	Fuller	A	B	C	D	Fuller	A	B	C	D
1	19	19	19	18	19	29	28	29	28	28
2	19	14	18	17	22	30	26	29	28	30
3	20	20	20	20	20	30	30	29	30	28
4	20	28	23	28	25	29	36	30	36	31
5	21	35	22	40	34	33	42	34	48	44
6	20	20	20	20	19	30	29	29	28	28
7	19	28	27	28	27	29	36	36	36	36
8	20	38	37	45	43	28	40	41	46	46
9	20	52	48	65	63	29	51	51	69	69
10	19	23	26	20	26	27	32	33	28	31
11	20	37	44	34	42	29	39	48	38	47
12	17	51	68	67	72	27	48	65	60	72
13	19	69	86	88	88	26	59	83	90	93
14	21	26	38	18	35	29	34	41	27	36
15	20	39	64	35	63	29	43	66	40	65
16	19	59	88	82	88	27	51	88	73	91
17	18	81	99	99	99	27	70	98	99	99

5 Application to U.S. Time Series

Nelson and Plosser (1982) challenged the prevailing wisdom of the 1970s by showing that the Augmented Dickey-Fuller (1979) test did not reject the hypothesis of a unit root for many U.S. annual macroeconomic time series. The original Nelson-Plosser data ran up to 1970, but was recently extended to 1988 by Schotman and van Dijk (1991). We apply our covariate tests to three series: Real GNP per Capita (1913-1988), Industrial Production (1891-1988), and the Unemployment Rate (1894-1988). All are measured in logs. Since the null hypothesis is that each series has one unit root, each variable was first-differenced before used as a covariate. All regressions included a constant and linear time trend, and three lags of the dependent variable ($p = 3$). In addition to the ADF tests, we report three specifications with covariates, setting k_1 and k_2 each equal to 0 and 2. Note that in each case the contemporaneous value of the covariate is included.

Table 6 presents the results. For GNP and Industrial Production, the first difference of the unlogged Unemployment Rate was used as Δx_t, and for the Unemployment Rate, the growth rate of Industrial Production was used as Δx_t. The OLS estimate of $\hat{\delta}$, its standard error, $t(\hat{\delta})$, and $z(\hat{\delta})$ are presented for all cases, and the estimated $\hat{\rho}^2$ and $\hat{R}^2$ for the covariate regressions.

First examine GNP. The ADF t-test is not significant, while the z-test is. The point estimate for the coefficient on lagged GNP is about -.20, with large standard errors. The message from the ADF regressions is that the data are uninformative regarding the order of integration of GNP.

The covariate tests are much more revealing. For every lag specification, both tests are significant at the asymptotic 5% level, with all but one test rejected at the asymptotic 1% level. The estimated $\hat{\rho}^2$ are extremely low, ranging from .06 to .08, and the estimated $\hat{R}^2$ are similar. This indicates that the estimates should be quite precisely estimated and the power of the unit root tests considerably higher than for the ADF tests. Interestingly, while the t-statistics show that δ is significantly different from 0, the point estimates (about -.08)

Table 6: Unit Root Tests for Extended Nelson-Plosser Data

GNP, Using UR as Covariate

	ADF	Covariate Tests			
		$k_2 = 0$		$k_2 = 2$	
		$k_1 = 0$	$k_1 = 2$	$k_1 = 0$	$k_1 = 2$
$\hat{\delta}$	-.20	-.09	-.08	-.09	-.08
$s(\hat{\delta})$	.06	.03	.03	.03	.03
$t(\hat{\delta})$	-3.3	-3.4**	-3.2**	-3.1**	-2.9*
$z(\hat{\delta})$	-25.5*	-24.3**	-21.0**	-24.6**	-21.1**
$\hat{\rho}^2$		.06	.08	.07	.08
$\hat{R}^2$		.09	.09	.08	.08

IP, Using UR as Covariate

	ADF	Covariate Tests			
		$k_2 = 0$		$k_2 = 2$	
		$k_1 = 0$	$k_1 = 2$	$k_1 = 0$	$k_1 = 2$
$\hat{\delta}$	-.24	-.06	-.06	-.05	-.06
$s(\hat{\delta})$	.07	.04	.04	.05	.04
$t(\hat{\delta})$	-3.3	-1.4	-1.5	-1.1	-1.3
$z(\hat{\delta})$	-29.6**	-9.9	-8.9	-8.4	-8.4
$\hat{\rho}^2$		.21	.16	.17	.15
$\hat{R}^2$		.20	.19	.20	.19

UR, Using IP as Covariate

	ADF	Covariate Tests			
		$k_2 = 0$		$k_2 = 2$	
		$k_1 = 0$	$k_1 = 2$	$k_1 = 0$	$k_1 = 2$
$\hat{\delta}$	-.28	-.14	-.11	-.19	-.15
$s(\hat{\delta})$	.07	.06	.06	.07	.07
$t(\hat{\delta})$	-3.9*	-2.2	-1.7	-2.8	-2.3
$z(\hat{\delta})$	-45.6**	-26.6**	-19.7**	-35.9**	-26.9**
$\hat{\rho}^2$		.53	.20	.58	.39
$\hat{R}^2$		.50	.27	.47	.30

GNP: Real per Capita Gross National Product
IP: Industrial Production Index
UR: Unemployment Rate

* Significant at the asymptotic 5% level
** Significant at the asymptotic 1% level

are much closer to 0 than the ADF estimates. Hence, while real per capita GNP appears to be I(0), it is highly persistent.

Second, examine the Industrial Production series. The ADF t-test statistic lies quite close to the asymptotic 5% critical value, the ADF z-test is significant, and the point estimate and standard error of δ indicates considerable uncertainty. The covariate tests, however, strongly support the unit root hypothesis. The point estimates of δ are about -.06, with insignificant t-statistics and z-statistics.

Third, turn to the unemployment rate. Both ADF tests strongly suggest that the series is $I(0)$. Examining the estimated ρ^2 for the covariate regressions, we see that there is considerable diversity between the specifications, as setting $k_1 = 0$ yields $\hat{\rho}^2$ around .55, while $k_1 = 2$ yields $\hat{\rho}^2 \in [.20, .39]$. For the specification with the lowest $\hat{\rho}^2$, the z-stat is significant at the asymptotic 1% level, while the t-stat is not statistically significant. For this specification, the estimated δ is quite small, at -.11, suggesting considerable persistence, even if the data is $I(0)$. It is hard to know how to decisively interpret these conflicting results. The ADF and covariate z-statistics suggest stationarity, while the covariate t-tests suggest caution. On balance, it seems prudent to infer that the unemployment rate is stationary, but highly persistent.

In all three cases, the inclusion of covariates in the regression equation moved the point estimate of δ much closer to 0. This is a likely consequence of the more precise and less biased sampling distributions implied by Theorem 2. The ADF estimator is highly biased and considerably imprecise, while the covariate estimator is less biased and more precise. Thus if a series is highly persistent (or integrated) with a small (or zero) δ, sample estimates of δ will tend to be further from zero in an ADF regression than in a covariate regression.

6 Conclusion

This paper has analyzed the distribution theory for tests of unit roots in regression models with covariates. We have found that the distributions of the standard test statistics is a non-trivial function of the included covariates. We have also found that this complication implies a major benefit: the power functions are dramatically improved as well. Our Monte

Carlo study suggested that the finite sample size of the regression t-statistics is adequate, and the power against stationarity excellent. We applied these tests to long time series, and found strong evidence to support the contentions that real per capita GNP and the unemployment rate are stationary but highly persistent, and that Industrial Production is $I(1)$.

References

[1] Andrews, D.W.K. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817-858.

[2] Andrews, D.W.K. and J.C. Monahan (1992), "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica*, 60, 953-966.

[3] Berk, K.N. (1974), "Consistent autoregressive spectral estimates," *Annals of Statistics*, 2, 489-502.

[4] Chan, N.H. and C.Z. Wei (1987), "Asymptotic inference for nearly nonstationary AR(1) processes," *The Annals of Statistics*, 15, 1050-1063.

[5] 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.

[6] Elliott, G., T.J. Rothenberg, and J.H. Stock (1992), "Efficient tests for an autoregressive unit root," manuscript, Harvard University.

[7] Hansen, B.E. (1992a), "Convergence to stochastic integrals for dependent heterogeneous processes," *Econometric Theory,* 8, 489-500.

[8] Hansen, B.E. (1992b), "Consistent covariance matrix estimation for dependent heterogeneous processes," *Econometrica*, 60, 967-972.

[9] Herrndorf, N. (1984), "A functional central limit theorem for weakly dependent sequences of random variables," *The Annals of Probability*, 12, 141-153.

[10] Horvath, M.T. and M.W. Watson (1993), "Testing for cointegration when some of the cointegrating vectors are known," Federal Reserve Bank of Chicago, #WP-93-15.

[11] Johansen, S. and K. Juselius (1992), "Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for the UK," *Journal of Econometrics*, 53, 211-244.

[12] Kremers, J.J.M., N.R. Ericsson, and J.J. Dolado, "The power of cointegration tests," *Oxford Bulletin of Economics and Statistics*, 54, 325-348.

[13] Nelson, C.R. and C.I. Plosser (1982), "Trends and random walks in macroeconomic time series," *Journal of Monetary Economics*, 10, 139-162.

[14] Phillips, P.C.B. (1987a), "Time series regression with a unit root," *Econometrica*, 55, 277-302.

[15] Phillips, P.C.B. (1987b), "Toward a unified asymptotic theory for autoregression," *Biometrika*, 74, 535-547.

[16] Phillips, P.C.B. and M. Loretan (1991), "Estimating long run economic equilibria," *Review of Economic Studies*, 58, 165-193.

[17] Phillips, P.C.B. and P. Perron (1988), "Testing for unit roots in time series regression," *Biometrika*, 74, 335-346.

[18] Said, S.E. and D.A. Dickey (1984), "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, 71, 599-608.

[19] Schotman, P.C. and H.K. van Dijk (1991), "On Bayesian routes to unit roots," *Journal of Applied Econometrics*, 6, 387-401.

[20] Stock, J.H. and M. Watson (1993), "A simple estimator of cointegrating vectors in higher order integrated systems," *Econometrica,* 61, 783-820.

Appendix A. Mathematical Proofs

Proof of Lemma 1.

Part 1: Note that

$$a(L) = a(1) + a^*(L)(1 - L) \tag{22}$$

where $a^*(L)$ has all roots outside the unit circle. Let $\xi_t = a^*(L)\Delta S_t$ which satisfies

$$\sup_{t \leq T} |\xi_t| = o_p\left(\sqrt{T}\right).$$

Using (8), and (22) we find that (5) can be rewritten as

$$\Delta S_{at} = -\frac{c}{T}S_{at-1} + \frac{c}{T}\xi_{t-1} + v_t,$$

where $S_{at} = a(L)S_t$. Hence

$$\frac{1}{\sqrt{T}} \sup_{t \leq T} |S_{at} - S_{at}^*| \to_p 0 \tag{23}$$

where S_{at}^* is generated by

$$\Delta S_{at}^* = -\frac{c}{T}S_{at-1}^* + v_t.$$

By Theorem 4.4 (a) of Hansen (1992a) we know that

$$\frac{1}{\sqrt{T}}S_{a[Tr]}^* \Rightarrow \sigma_v W_1^c(r). \tag{24}$$

(23) and (24) yield

$$\frac{1}{\sqrt{T}}S_{a[Tr]} \Rightarrow \sigma_v W_1^c(r). \tag{25}$$

Finally, let $k(L) = a(L)^{-1}$. We have $k(L) = k(1) + k^*(L)(1 - L)$ where $k^*(L)$ has all roots outside the unit circle since $a(L)$ does. By (25),

$$\frac{1}{\sqrt{T}}S_{[Tr]} = \frac{1}{\sqrt{T}}k(L)S_{a[Tr]} = k(1)\frac{1}{\sqrt{T}}S_{a[Tr]} + o_p(1) \Rightarrow k(1)^{-1}\sigma_v W_1^c(r).$$

The proof is completed by noting that $k(1) = a(1)^{-1}$.

Part 2: Follows from Part 1 by the continuous mapping theorem.

Part 3: Theorem 4.4 of Hansen (1992a), Part 1, and (7) yields

$$\frac{1}{T}\sum_{t=1}^{T} S_{t-1}e_t \Rightarrow \int_0^1 a(1)^{-1}\sigma_v W_1^c \sigma_e d\left(\rho W_1 + \left(1-\rho^2\right)^{1/2} W_2\right),$$

which is the stated result. □

Proof of Theorem 1.

Rearranging terms in (9), we find that

$$LR = a(1)^2 \frac{(\bar{c}^2 - 2\bar{c}c)}{\sigma_v^2 T^2}\sum_{t=2}^{T} S_{t-1}^2 + a(1)\frac{2\bar{c}}{\sigma_v^2 T}\sum_{t=2}^{T} S_{t-1}e_t.$$

An application of Lemma 3 yields

$$LR \Rightarrow \left(\bar{c}^2 - 2\bar{c}c\right)\int_0^1 (W_1^c)^2 + 2\bar{c}\frac{\sigma_e}{\sigma_v}\left(\rho\int_0^1 W_1^c dW_1 + \left(1-\rho^2\right)^{1/2}\int_0^1 W_1^c dW_2\right).$$

The stated result follows by the definition $R = \sigma_e/\sigma_v$.

Proof of Theorem 2.

We prove (13). The extension to the cases with a mean or trend correction is standard and omitted.

Let $\phi_t = (\Delta y_{t-1}, ..., \Delta y_{t-p}, x'_{t+q_2}, ..., x'_{t-q_1})'$. Since ϕ_t is covariance stationary, ergodic, and strong mixing, and $E(\phi_t e_t) = 0$ under Assumption 1, we know that

$$\frac{1}{T}\sum_{t=2}^{T} S_{t-1}\phi'_t = O_p(1) \tag{26}$$

$$\frac{1}{T}\sum_{t=2}^{T} \phi_t\phi'_t \rightarrow_p F > 0 \tag{27}$$

and

$$\frac{1}{T}\sum_{t=2}^{T} \phi_t e_t \rightarrow_p 0. \tag{28}$$

Thus

$$\begin{aligned} T\left(\hat{\delta} - \delta\right) &= \frac{\frac{1}{T}\sum_{t=2}^{T} S_{t-1}e_t - \frac{1}{T}\sum_{t=2}^{T} S_{t-1}\phi'_t\left(\frac{1}{T}\sum_{t=2}^{T}\phi_t\phi'_t\right)\frac{1}{T}\sum_{t=2}^{T}\phi_t e_t}{\frac{1}{T^2}\sum_{t=2}^{T} S_{t-1}^2 - \frac{1}{T^2}\sum_{t=2}^{T} S_{t-1}\phi'_t\left(\frac{1}{T}\sum_{t=2}^{T}\phi_t\phi'_t\right)\frac{1}{T}\sum_{t=2}^{T}\phi_t S_{t-1}} \\ &= \frac{\frac{1}{T}\sum_{t=2}^{T} S_{t-1}e_t}{\frac{1}{T^2}\sum_{t=2}^{T} S_{t-1}^2} + o_p(1). \end{aligned} \tag{29}$$

Hence by Lemma 3,

$$T\left(\hat{\delta}-\delta\right) \Rightarrow \frac{a(1)^{-1}\sigma_v\sigma_e \int_0^1 W_1^c\left(\rho dW_1+(1-\rho^2)^{1/2}\,dW_2\right)}{a(1)^{-2}\sigma_v^2\int_0^1 (W_1^c)^2}$$

$$= a(1)\frac{\sigma_e}{\sigma_v}\left(\rho\frac{\int_0^1 W_1^c dW_1}{\int_0^1 (W_1^c)^2}+\left(1-\rho^2\right)^{1/2}\frac{\int_0^1 W_1^c dW_2}{\int_0^1 (W_1^c)^2}\right).$$

The equality $R=\sigma_e/\sigma_v$ establishes the result. □

Proof of Theorem 3.

Part 1 follows from Theorem 2 and the fact that $T\delta=-ca(1)$ under (8). Since

$$t(\hat{\delta}) = \hat{\sigma}_e^{-1}\left(\sum_{t=2}^T S_{t-1}^2-\sum_{t=2}^T S_{t-1}\phi_t'\left(\sum_{t=2}^T\phi_t\phi_t'\right)\sum_{t=2}^T\phi_t S_{t-1}\right)^{1/2}\hat{\delta}$$

$$= \hat{\sigma}_e^{-1}\left(\frac{1}{T^2}\sum_{t=2}^T S_{t-1}^2\right)^{1/2} T\hat{\delta}+o_p(1),$$

we can see that

$$t(\hat{\delta}) \Rightarrow \sigma_e^{-1}\left(a(1)^{-2}\sigma_v^2\int_0^1 (W_1^c)^2\right)^{1/2}\left[-ca(1)+a(1)R\left(\rho\frac{\int_0^1 W_1^c dW_1}{\int_0^1 W_c^2}+\left(1-\rho^2\right)^{1/2}\frac{\int_0^1 W_1^c dW_2}{\int_0^1 W_c^2}\right)\right]$$

$$= -\frac{c}{R}\left(\int_0^1 (W_1^c)^2\right)^{1/2}+\rho\frac{\int_0^1 W_1^c dW_1}{\left(\int_0^1 (W_1^c)^2\right)^{1/2}}+\left(1-\rho^2\right)^{1/2}\frac{\int_0^1 W_1^c dW_2}{\left(\int_0^1 (W_1^c)^2\right)^{1/2}}.$$

Since W_2 is independent of W_1 and W_1^c, $\int_0^1 W_c dW_2/\left(\int_0^1 (W_1^c)^2\right)^{1/2}$ is distributed $N(0,1)$, and is independent of W_1^c. This completes the proof. □

Figure 1: Asymptotic Local Power Envelope, $R^2=\rho^2$

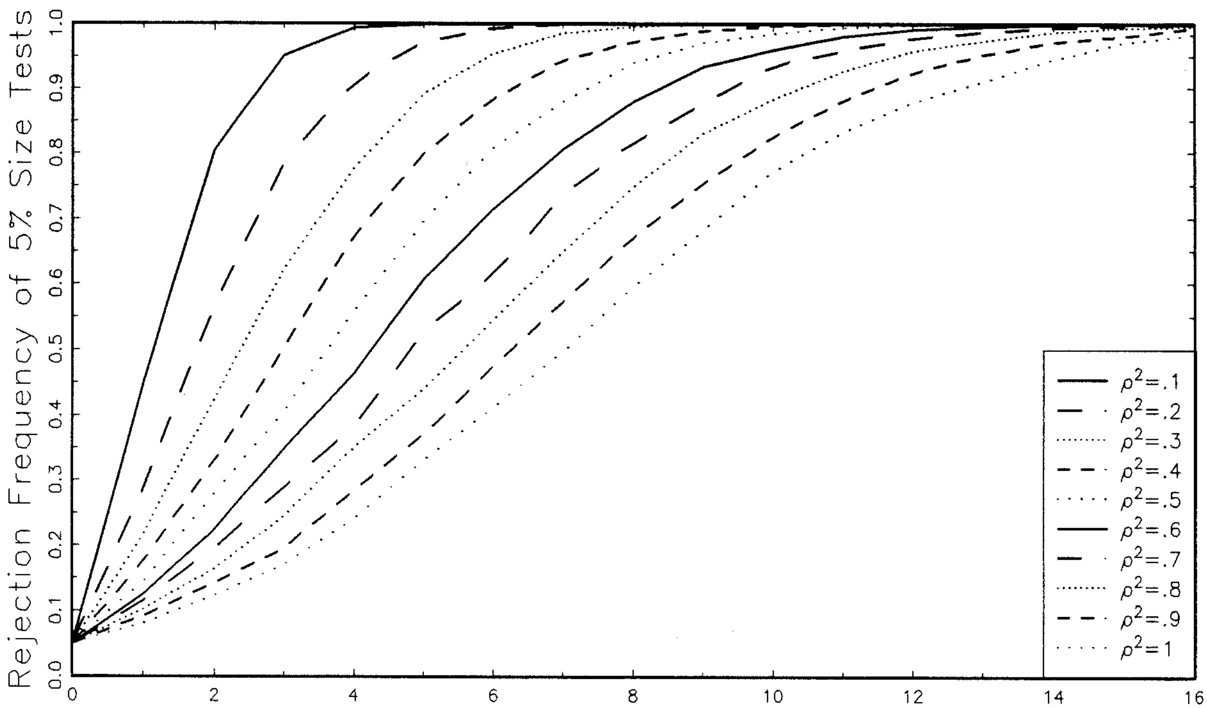

Figure 2: Comparison of Power Envelope
and Covariate Tests with No Mean Correction, $R^2=\rho^2$

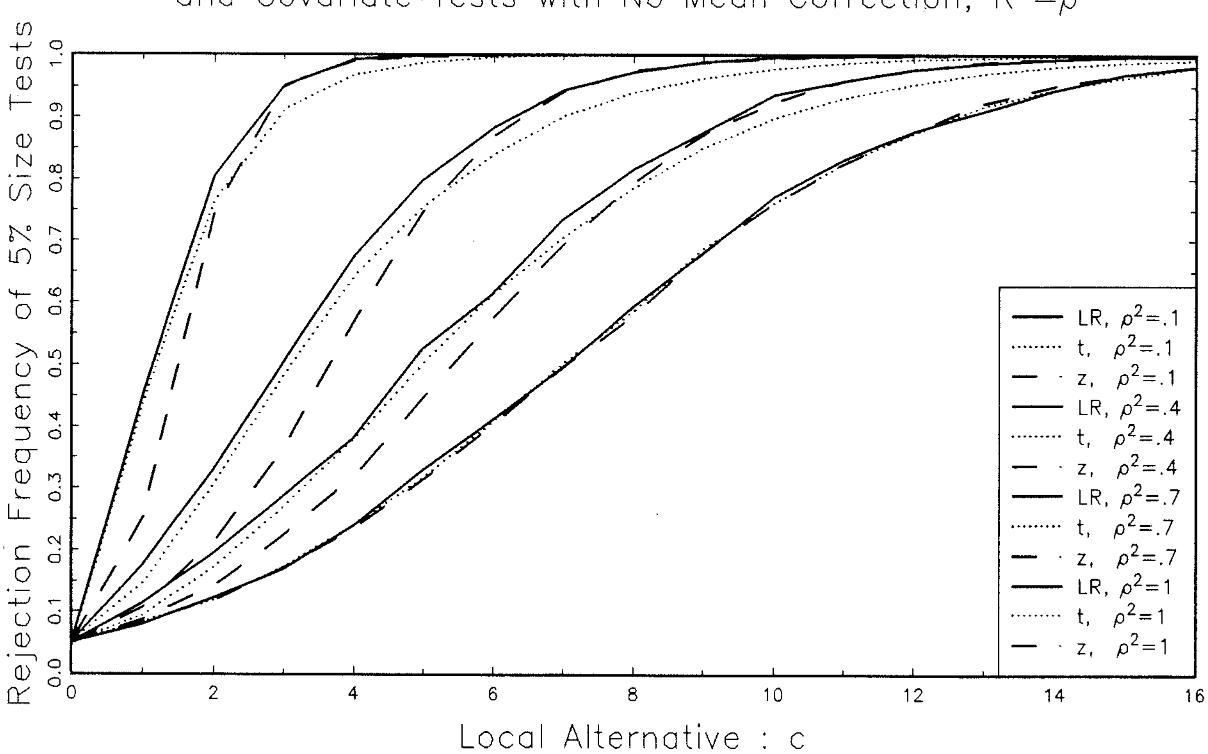

Figure 3: Power of Covariate tests
With Mean Correction, $R^2 = \rho^2$

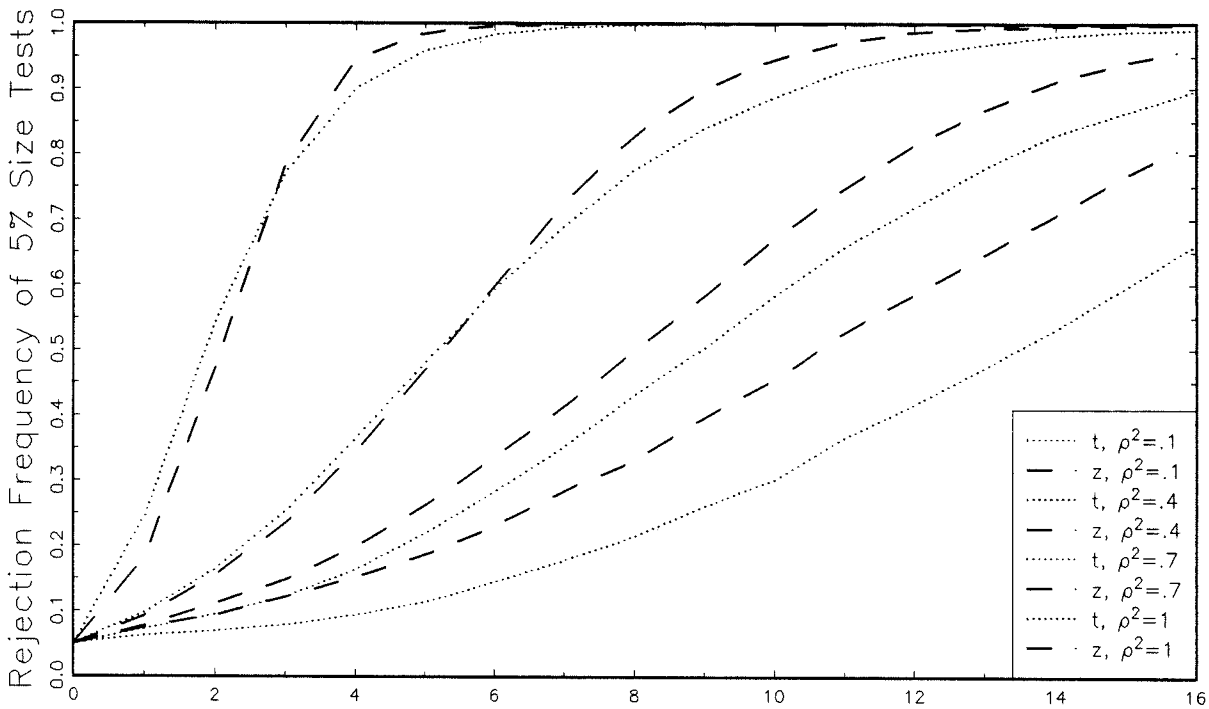

Figure 4: Power of Covariate tests
With Trend Correction, $R^2 = \rho^2$

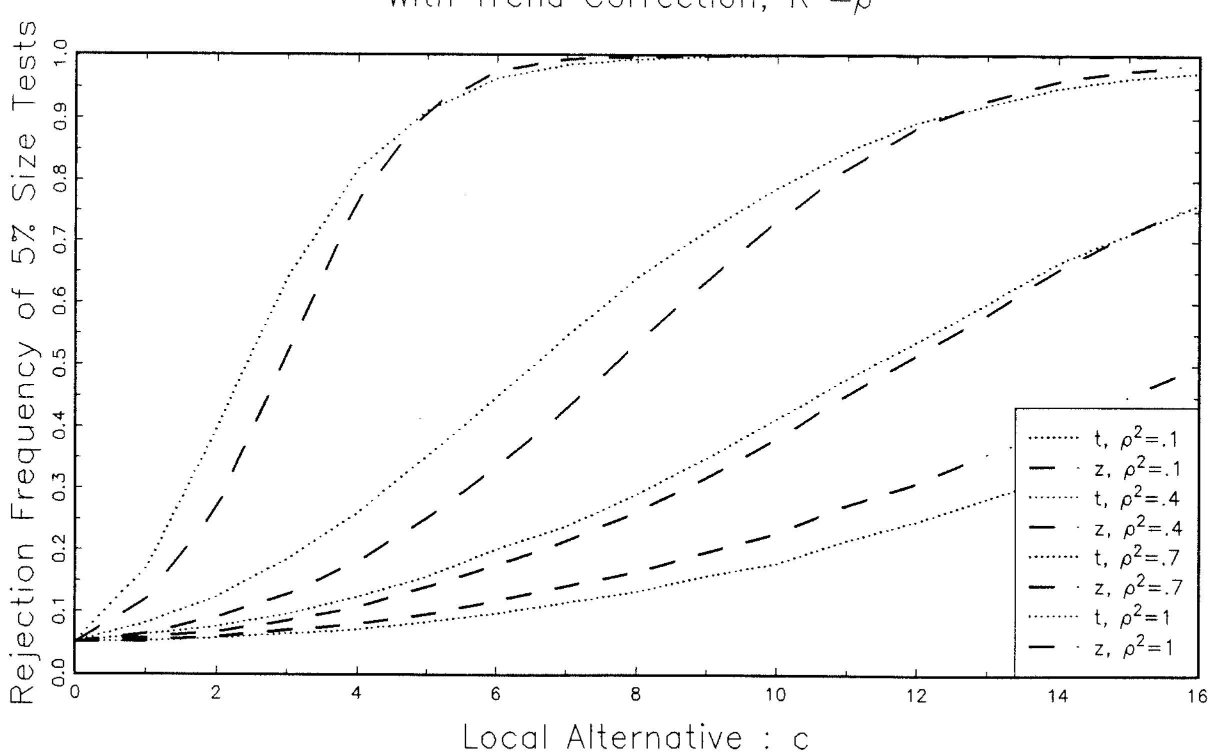

Figure 5: Power of t-Test With Mean Correction

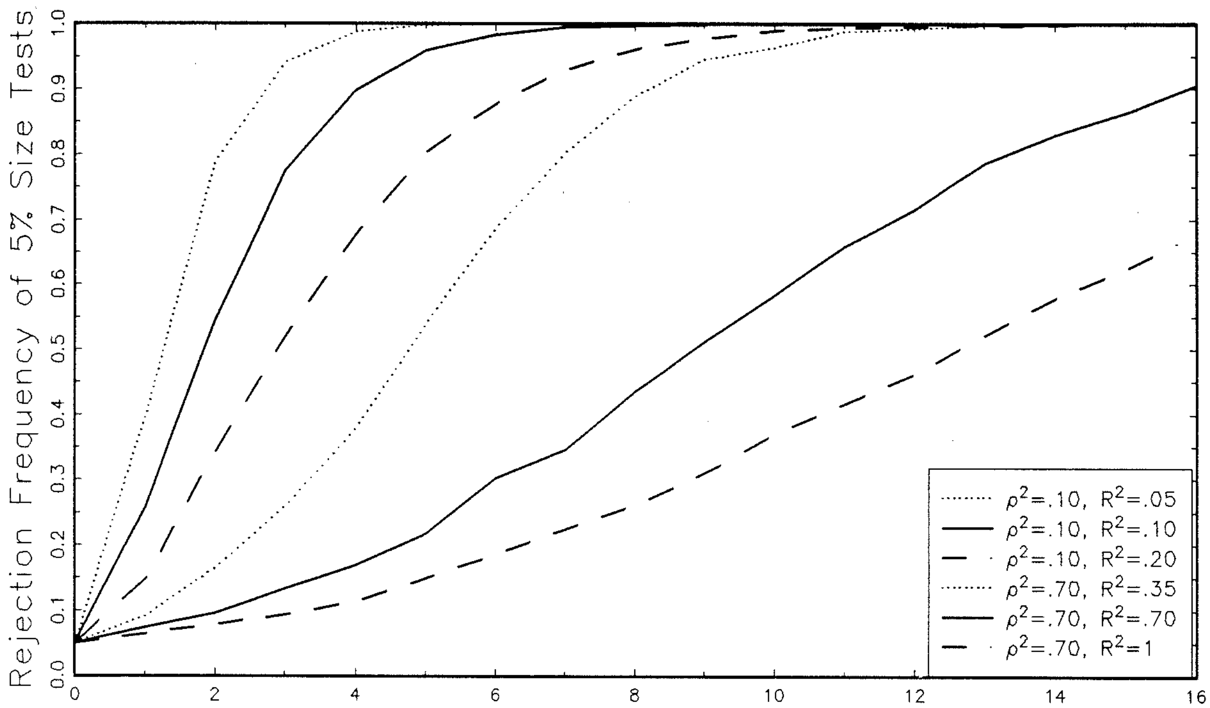

Figure 6: Power with Over-Differenced Regressors