Econometric Theory, **11**, 1995, 1148–1171. Printed in the United States of America.

RETHINKING THE UNIVARIATE APPROACH TO UNIT ROOT TESTING

Using Covariates to Increase Power

Bruce E. Hansen
Boston College

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 estimates of the largest autoregressive root and its t-statistic. The asymptotic distribution is not the conventional Dickey–Fuller distribution, but a convex combination of the Dickey–Fuller distribution 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.

Consider the AR(1) model $\Delta y_t = \delta y_{t-1} + u_t$, where u_t is i.i.d. $(0,\sigma_u^2)$. The hypothesis of a unit root $H_0: \delta = 0$ is typically tested by the ordinary least-squares (OLS) t-statistic for δ, as it is widely believed to be the best classical procedure in this context. It is rare, however, that we observe the time series y_t in isolation. More typically, we observe at least one related time series, say x_t. Suppose that x_t is I(1), so that Δx_t is I(0). For simplicity, assume that $(\Delta x_t, u_t)$ is an i.i.d. and zero mean. Set $\sigma_{xu} = E(\Delta x_t u_t)$, $\sigma_x^2 =$

I thank three referees and the editor for helpful comments and the National Science Foundation and the Sloan Foundation for research support. Address correspondence to: Professor Bruce E. Hansen, Department of Economics, Boston College, Chestnut Hill, MA 02167-3806, USA.

  0266-4666/95 $9.00 + .10

$E(\Delta x_t)^2$, $b = \sigma_{xu}/\sigma_x^2$, and $e_t = u_t - \Delta x_t b$, so that the process for y_t can alternatively be written as

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

Under the assumptions, the parameter δ retains the same meaning as in the AR(1) model. An important difference, however, is that the new error variance, $\sigma_e^2 = \sigma_u^2 - \sigma_{xu}^2/\sigma_x^2$, will be smaller than σ_u^2 (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 (1) rather than the AR(1) model. Thus, 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 (1)? 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 later that this belief is incorrect.

Our results relate to some previous results in the literature. Kremers, Ericsson, and Dolado (1992) discussed the model

$$\begin{aligned} \Delta z_{1t} &= a_0 \Delta z_{2t} + a_1 (z_{1t-1} - z_{2t-1}) + \epsilon_{1t} \\ \Delta z_{2t} &= \epsilon_{2t} \end{aligned} \tag{2}$$

with ϵ_{1t} and ϵ_{2t} i.i.d. uncorrelated normal random variables. We can see that (2) falls in class (1) by setting $y_t = z_{1t} - z_{2t}$, $x_t = z_{2t}$, $\delta = a_1$, and $b = a_0 - 1$. Our results are thus generalizations of theirs, although it should be emphasized that Kremers et al. were discussing tests for cointegration, not for univariate unit roots.

Horvath and Watson (1993) recently proposed tests for cointegration when the cointegrating vector is known a priori. Although 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 obtained are different from those analyzed here.

A final interesting possibility was mentioned by Johansen and Juselius (1992). Conditioning on a known cointegrating rank of the data, they proposed testing 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. It appears that our tests are complementary to those of Johansen and Juselius.

Section 2 introduces a generalized version of (1), 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 the AR(1) model. The asymptotic distributions of the OLS estimates of (1) are also found under local alternatives to a unit root. Section 3 investigates the t-statistic for $\delta = 0$. Its asymptotic distribution is 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. 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 and that industrial production is I(1). The Appendix contains the mathematical proofs. A GAUSS procedure that calculates the test statistics and critical values is available from the author upon request.

2. REGRESSION FRAMEWORK

2.1. Model and Assumptions

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

$$y_t = d_t + S_t, \tag{3}$$

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 as

$$a(L)\Delta s_t = \delta S_{t-1} + v_t, \tag{4}$$

where $a(L) = 1 - a_1 L - a_2 L^2 - \cdots - a_p L^p$ is a pth order polynomial in the lag operator and

$$v_t = b(L)'(\Delta x_t - \mu_x) + e_t. \tag{5}$$

In (5), Δx_t is an m-vector, $\mu_x = E(\Delta x_t)$, and $b(L) = b_{q2}L^{-q2} + \cdots + b_{q1}L^{q1}$ is a lag polynomial allowing for (but not requiring) both leads and lags of Δx_t to enter the equation for v_t. In many applications (such as a standard VAR in y_t and Δx_t, where (5) is one equation from the VAR), only lagged values of Δx_t will enter the regression, so $b(L)$ will take the form $b_1 L + \cdots + b_{q1}L^{q1}$.

Define the 2×2 long-run covariance matrix

$$\Omega = \sum_{k=-\infty}^{\infty} E\left(\begin{pmatrix} v_t \\ e_t \end{pmatrix} (v_{t-k} \quad e_{t-k})\right) = \begin{pmatrix} \sigma_v^2 & \sigma_{ve} \\ \sigma_{ve} & \sigma_e^2 \end{pmatrix}, \tag{6}$$

and define

$$\rho^2 = \frac{\sigma_{ve}^2}{\sigma_e^2 \sigma_v^2},$$

the long-run (zero-frequency) squared correlation between v_t and e_t. One interpretation of ρ^2 is that it measures the relative contribution of Δx_t to v_t

at the zero frequency. One extreme is obtained when $b(L) = 0$, for then $v_t = e_t$ and $\rho^2 = 1$. The other extreme is obtained when the regressors Δx_t explain nearly all the zero-frequency movement in v_t, in which case $\rho^2 \approx 0$, which we exclude for technical reasons. We also define the variance ratio

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

To help improve our understanding of what these parameters mean, consider the important special case where e_t is uncorrelated with Δx_{t-k} for all k, which holds in a well-specified dynamic regression (or when both q_1 and q_2 are sufficiently large). In this case, $\sigma_{ve} = \sigma_e^2$ and $\rho^2 = R^2$. Furthermore, when $\delta = 0$ the long-run variance of Δy_t may be determined from (4) as $\sigma_{\Delta y}^2 = \sigma_v^2/a(1)^2$. Thus, $\rho^2 = a(1)^2\sigma_e^2/\sigma_{\Delta y}^2$. As another special case, suppose that Δx_t is serially uncorrelated with covariance matrix Σ. Then, $\rho^2 = \sigma_e^2/(\sigma_e^2 + b(1)'\Sigma b(1))$.

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 $\sum_{m=1}^{\infty} \alpha_m^{1/r-1/p} < \infty$;
2. $\sup_t E[|\Delta x_t|^p + |e_t|^p] < \infty$;
3. $E(\Delta x_{t-k} e_t) = 0$ for $q_1 \le k \le q_2$;
4. $E(e_t e_{t-k}) = 0$ for all $k \ge 1$;
5. the roots of $a(L)$ all lie outside the unit circle;
6. $E(\phi_t \phi_t') > 0$, where $\phi_t = (\Delta y_{t-1}, \dots, \Delta y_{t-p}, \Delta x_{t+q2}' - \mu_x, \dots, \Delta x_{t-q_1}' - \mu_x)'$;
7. $\sigma_v^2 > 0$ and $\rho^2 > 0$.

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 that grows with sample size, following the technique of Berk (1974) and Said and Dickey (1984). The assumption that $\sigma_v^2 > 0$ ensures that y_t is I(1) when $\delta = 0$, which is necessary for our interpretation of tests of $\delta = 0$ as tests for a unit root in y_t.[1]

Under Assumption 1, the errors (v_t, e_t) satisfy the conditions of Herrndorf (1984), so their partial sums converge weakly to a Brownian motion with covariance matrix Ω. Furthermore, using definition (6), we can write this limit as

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

where W_1 and W_2 are independent standard Brownian motions and $\Rightarrow$ denotes weak convergence with respect to the uniform metric.

Our asymptotic theory will be based on "local-to-unity asymptotics," following the technique of Phillips (1987) and Chan and Wei (1987). Models (3)–(5) contain a unit root under the null hypothesis $H_0: \delta = 0$. 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

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

2.2. Power Envelope

The Gaussian power envelope for the unit root testing problem in model (5) can be easily derived. Assume that the nuisance parameters $a(L)$, $b(L)$, μ, μ_x, σ_e^2, and θ are known, the error e_t is i.i.d. $N(0, \sigma_e^2)$ and is independent of Δx_t at all leads and lags, and the initial condition S_0 is fixed. The likelihood ratio test for a unit root ($\delta = 0$) against a fixed alternative $\bar{\delta} < 0$ rejects for small values of

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

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

THEOREM 1.

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

Note that the limiting distribution of the likelihood ratio statistic depends on the parameters $(c, \bar{c}, R^2, \rho^2)$. The point optimal likelihood ratio statistic sets $\bar{c} = c$, 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 c) 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 is an 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[2] 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.

2.3. Least-Squares Estimation

When $d_t = 0$ and $\mu_x = 0$, 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

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

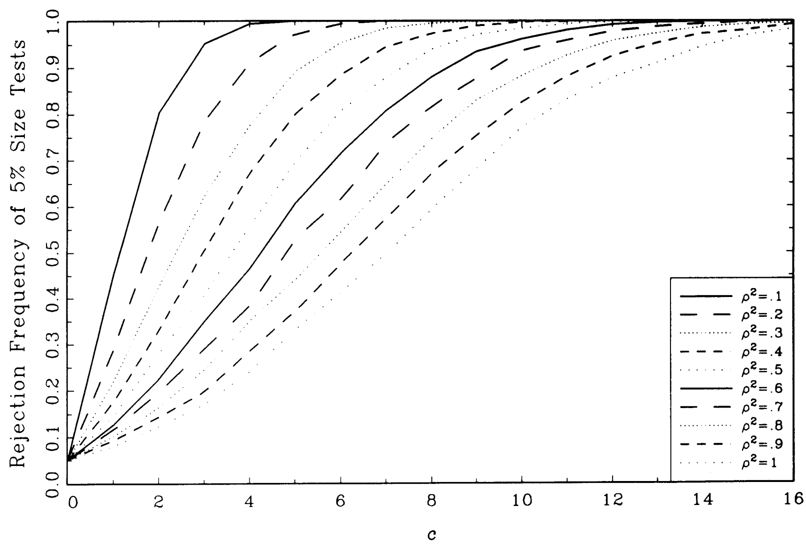

FIGURE 1. Asymptotic local power envelope, $R^2 = \rho^2$.

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 $\theta^* = -\delta\theta$. Equations (10), (11), or (12) can be estimated by OLS. Let $\hat{\delta}$, $\hat{\delta}^\mu$, and $\hat{\delta}^\tau$ denote the respective OLS estimates of δ.

THEOREM 2.

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

The asymptotic distributions for $T(\hat{\delta}^\mu - \delta)$ *and* $T(\hat{\delta}^\tau - \delta)$ *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.*

3. TESTING FOR A UNIT ROOT

3.1. Test Statistics

The natural test statistic for the hypothesis of a unit root $H_0: \delta = 0$ in models (10)–(12) is the t-statistic $t(\hat{\delta}) = \hat{\delta}/s(\hat{\delta})$, where $s(\hat{\delta})$ is the OLS standard error for $\hat{\delta}$. For models (11) and (12), we denote the statistics by $t(\hat{\delta}^\mu)$ and $t(\hat{\delta}^\tau)$, respectively. We will refer to $t(\hat{\delta})$ as the CADF(p, q_1, q_2) statistic, where CADF stands for "covariate augmented Dickey–Fuller" and (p, q_1, q_2) stands for the orders of the polynomials $a(L)$ and $b(L)$, as specified in (4) and (5).

THEOREM 3.

$$t(\hat{\delta}) \Rightarrow -\frac{c}{R}\left(\int_0^1 (W_1^c)^2\right)^{1/2} + \rho \frac{\int_0^1 W_1^c\, dW_1}{\left(\int_0^1 (W_1^c)^2\right)^{1/2}} + (1-\rho^2)^{1/2} N(0,1), \tag{14}$$

where the $N(0,1)$ *variable is independent of* W_1. *Under the null hypothesis* $\delta = 0$,

$$t(\hat{\delta}) \Rightarrow \rho \frac{\int_0^1 W_1\, dW_1}{\left(\int_0^1 W_1^2\right)^{1/2}} + (1-\rho^2)^{1/2} N(0,1). \tag{15}$$

The asymptotic distributions for $t(\hat{\delta}^\mu)$ *and* $t(\hat{\delta}^\tau)$ *are similar, except that* W_1^c *is replaced by* $W_1^{c\mu}$ *and* $W_1^{c\tau}$, *respectively.*

While the local asymptotic distribution of (14) depends on the two nuisance parameters ρ^2 and R^2, the null distribution of (15) only depends on ρ^2. The latter distribution is a convex mixture of the standard normal and the Dickey–Fuller distribution, with the weights determined by ρ^2. As $\rho^2 \to 1$ we find the Dickey–Fuller, and as $\rho^2 \to 0$ we obtain the normal distribution. Estimated[3] asymptotic 1, 5, and 10% critical values for the CADF statistic are given in Table 1, for values for ρ^2 in steps of .1. For intermediate values of ρ^2, critical values could be selected by interpolation.

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 Theorem 3. We can see from the form of the asymptotic distributions that the conventional asymptotic critical values are conservative, implying that tests mistakenly based on the Dickey–Fuller critical values will have reduced power.

To use the correct critical values from Table 1, we need a consistent estimate of ρ^2. A nonparametric choice is

$$\hat{\rho}^2 = \frac{\hat{\sigma}_{ve}^2}{\hat{\sigma}_v^2 \hat{\sigma}_e^2}, \tag{16}$$

where

$$\hat{\Omega} = \begin{pmatrix} \hat{\sigma}_v^2 & \hat{\sigma}_{ve} \\ \hat{\sigma}_{ve}^2 & \hat{\sigma}_e^2 \end{pmatrix} = \sum_{k=-M}^{M} w(k/M) \frac{1}{T} \sum_t \hat{\eta}_{t-k} \hat{\eta}_t' \tag{17}$$

and $\hat{\eta}_t = (\hat{v}_t \hat{e}_t)'$ are least-squares estimates of $\eta_t = (v_t e_t)'$ from the appropriate regression model.[4] The estimated $\hat{\rho}^2$ from (16) is then used to select the

TABLE 1. Asymptotic critical values for CADF t-statistics

	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
.9	−2.57	−1.94	−1.61	−3.39	−2.81	−2.50	−3.88	−3.33	−3.04
.8	−2.57	−1.94	−1.60	−3.36	−2.75	−2.46	−3.83	−3.27	−2.97
.7	−2.55	−1.93	−1.59	−3.30	−2.72	−2.41	−3.76	−3.18	−2.87
.6	−2.55	−1.90	−1.56	−3.24	−2.64	−2.32	−3.68	−3.10	−2.78
.5	−2.55	−1.89	−1.54	−3.19	−2.58	−2.25	−3.60	−2.99	−2.67
.4	−2.55	−1.89	−1.53	−3.14	−2.51	−2.17	−3.49	−2.87	−2.53
.3	−2.52	−1.85	−1.51	−3.06	−2.40	−2.06	−3.37	−2.73	−2.38
.2	−2.49	−1.82	−1.46	−2.91	−2.28	−1.92	−3.19	−2.55	−2.20
.1	−2.46	−1.78	−1.42	−2.78	−2.12	−1.75	−2.97	−2.31	−1.95

appropriate row from Table 1. The function $w(\cdot)$ in (17) may be any kernel weight function that produces positive semidefinite covariance matrices, such as the Bartlett or Parzen kernels, and M is a bandwidth selected to grow slowly with sample size. Conditions under which these estimates are consistent are given in Hansen (1992a), and selection rules for M that minimize asymptotic mean squared error are given in Andrews (1991) and Newey and West (1994).

3.2. Asymptotic Local Power Functions

Theorem 3 gives the asymptotic distribution of the CADF t-statistic under the local-to-unity alternative in (8). The expressions show that the asymptotic distribution, and hence the asymptotic local power, depends on ρ^2 and R^2. Figure 2 displays the asymptotic local power functions[5] for the t-test when $d_t = 0$ for $\rho^2 = 1.0$, .7, and .4, and .1, setting $R^2 = \rho^2$. The power envelope is displayed as well for contrast. We can see that the power is quite close to the power envelope, especially for small c or large ρ^2.

Figures 3 and 4 display the same set of power functions for the cases $d_t = \mu$ and $d_t = \mu + \theta t$, respectively. As expected, 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 $\rho^2 = 1$. These power curves lie uniformly beneath the power envelopes, which, to retain clarity, are not displayed on

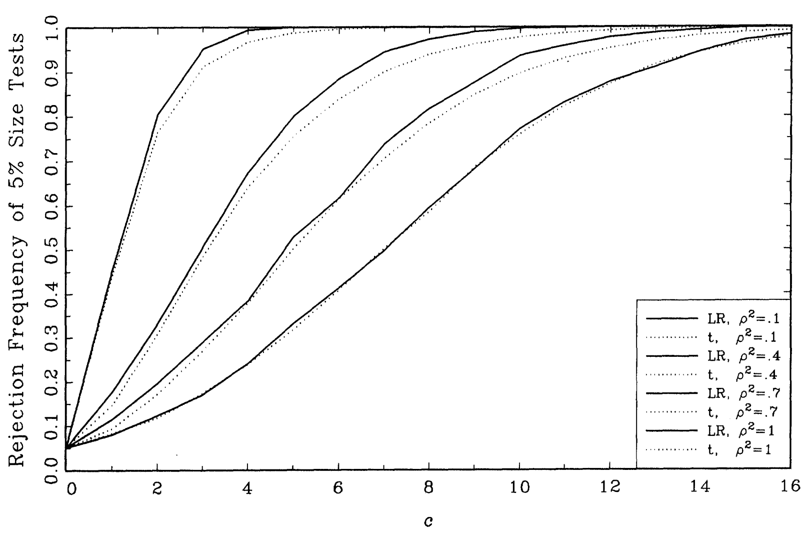

FIGURE 2. Comparison of power under no mean correction with power envelope, $R^2 = \rho^2$.

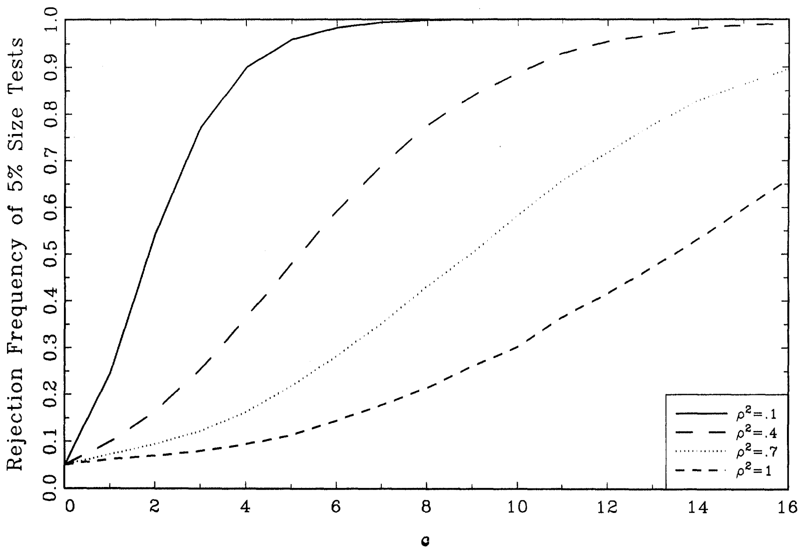

FIGURE 3. Power under mean correction, $R^2 = \rho^2$.

these graphs. Just as in the conventional case, the estimation of the mean and/or trend coefficient implies a significant loss in power.

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

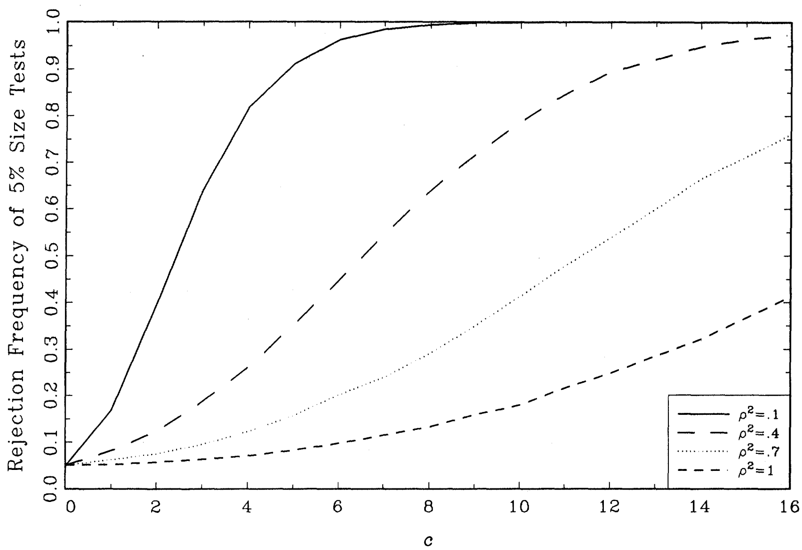

FIGURE 4. Power under trend correction, $R^2 = \rho^2$.

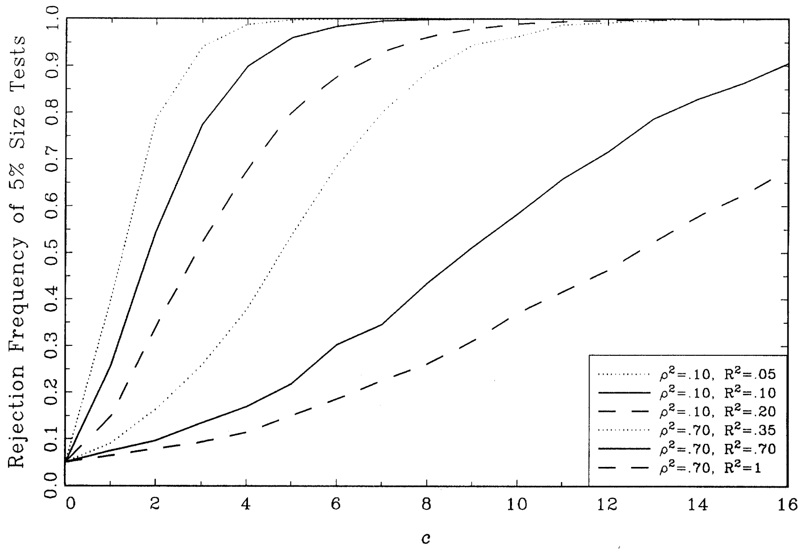

FIGURE 5. Power under mean correction, $R^2 \neq \rho^2$.

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 $R^2 = .7$, whereas 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. In summary, we find that enormous power gains can be achieved by inclusion of appropriate covariates Δx_t.

3.3. Under-Differenced Regressors

The theory of the previous sections is based on the strong assumption that the series Δx_t is stationary. This assumption is important for the results. We consider in this section the consequences of Δx_t being I(1).

Consider the simple model

$$\Delta y_t = \delta y_{t-1} + \Delta x_t b + e_t \tag{18}$$

with e_t i.i.d. and Δx_t I(1). The null hypothesis that y_t is I(1) holds either when $\delta = 0$ and $b = 0$ or when y_t and Δx_t are cointegrated.

When y_t and Δx_t are I(1) but not cointegrated, the CADF statistic will not have the asymptotic distribution specified in (15); instead, it will be of the form found by Phillips and Ouliaris (1990) for the Engle and Granger

(1987) two-step test for no cointegration. The tables in Phillips and Ouliaris (1990) show that this asymptotic distribution is more biased away from the normal distribution than is the standard Dickey–Fuller distribution. Thus, inferences based on the distribution theory of Section 3.1 when Δx_t is I(1) and not cointegrated with y_t will be considerably biased.

Alternatively, if we allow y_t and Δx_t to be cointegrated, then y_t will be I(1) even when $\delta \neq 0$, rending a test of the restriction $\delta = 0$ meaningless.

It appears that the assumption that Δx_t is stationary is not innocuous. Violations of this condition will invalidate the theory derived in the previous sections. A sensible conclusion is that applications should use first-differenced regressors, hence the notation Δx_t. A caveat should be noted that some variables, such as the price level, may be I(2), in which case first differences will not be sufficient to induce stationarity.

3.4. Over-Differenced Regressors

To avoid the problems mentioned in the previous section, we recommended taking first differences before including a highly serially correlated variable as a regression covariate. This of course raises the possibility that Δx_t could be "over-differenced," or I(−1). When Δx_t is I(−1), $\rho^2 = 1$, so the asymptotic critical values and power of the CADF test is equivalent to that of the ADF test.

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, let us assume that x_t is near-integrated. Specifically, assume that x_t satisfies

$$\Delta x_t = -\frac{g}{T} x_{t-1} + u_t, \tag{19}$$

with u_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 asymptotic critical values and power function for the CADF test should equal that of the ADF test.

The model is essentially the same as before, except that there is one more parameter, g. The asymptotic distribution for the regression test will therefore depend on ρ^2, R^2, and g. A nonzero g will induce bias in both the asymptotic size and power of the tests. Asymptotic size was calculated[6] for a variety of values of g and ρ^2, setting $R^2 = \rho^2$. Results for nominal 5% size tests with a fitted intercept are reported in Table 2. (The size distortion for the case without a fitted intercept was minimal, and that in the case of a fitted intercept and time trend was similar to the case reported in Table 2.)

TABLE 2. Asymptotic size under over-differencing

	$\rho^2 = .1$	$\rho^2 = .3$	$\rho^2 = .5$	$\rho^2 = .7$	$\rho^2 = .9$
$g = -10$	.2	.3	.4	.4	.6
$g = -6$	.6	.6	.6	.4	.8
$g = -2$	3.5	3.3	2.9	3.2	4.4
$g = 2$	5.6	6.1	5.4	5.8	5.4
$g = 6$	6.3	6.8	5.8	5.5	5.4
$g = 10$	7.3	7.8	6.2	5.5	4.8

Our major concern is with the over-differenced case ($g > 0$). Table 2 indicates only mild size distortion, with the t-test rejecting slightly too frequently relative to the asymptotic distribution. The distortion increases as ρ^2 falls. In the empirically less relevant case of a locally explosive root ($g < 0$), we find substantial underrejection.[7]

We are also interested in the effect of over-differencing on power. Figure 6 plots the power functions for $\rho^2 = .1$ and g equal to 0, 2, 4, 8, 16, and ∞. The qualitative impact of letting $g > 0$ is as anticipated: increasing g decreases the power function, eventually reaching the power function of the Dickey–Fuller test. What is somewhat surprising (at least to the author) is the quan-

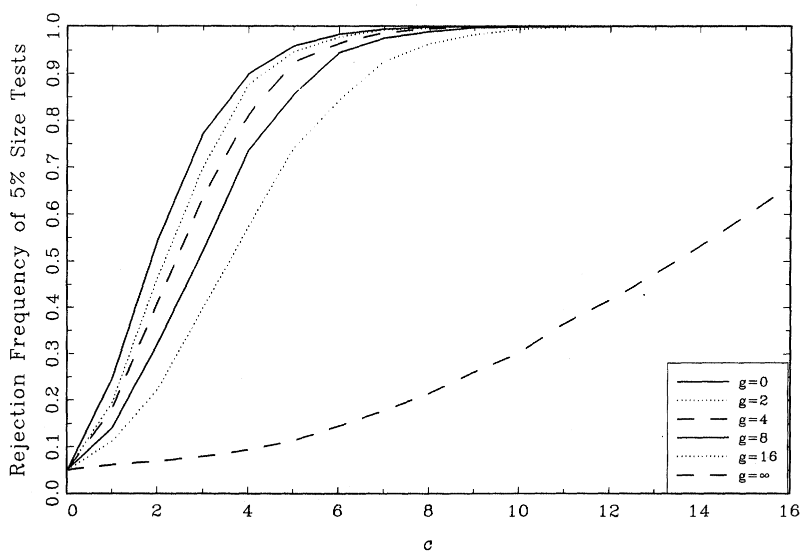

FIGURE 6. Power with over-differenced regressors.

titative finding that the magnitude of the power loss is fairly mild. Even setting $g = 16$ does not lead to a major loss of power.

In summary, over-differencing is in principle a cause for concern because the asymptotic null distribution depends on the unknown (and nonestimable) parameter g (and is, thus, nonsimilar). One response is that the CADF test exploits known prior information concerning the order of integration of the regressors Δx_t. Another response is that neither the size distortion nor the power loss is severe, so no major inferential error is likely. A final argument is that the hypothesis is really a joint null hypothesis that both y_t and x_t are I(1), and a "rejection" of the joint hypothesis should be interpreted with caution.

4. SMALL SAMPLE DISTRIBUTIONS

4.1. VAR Design

To demonstrate the performance of the CADF test in small samples, I performed two simulation experiments. For the first, data were generated from the VAR model

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

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

where

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

Each replication discarded the first 100 observations to eliminate start-up effects. Sample sizes of 50, 100, and 250 were examined; only those for $T = 100$ are reported here to conserve space.

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

The test statistics considered are ADF(2), CADF(2,0,0), CADF(2,1,0), CADF(2,0,1), and CADF(2,1,1). All regressions included a fitted intercept. To implement the CADF tests, asymptotic critical values were taken from Table 1 for each sample using a sample estimate of ρ^2, calculated using a Parzen kernel and Andrews' (1991) automatic bandwidth estimator.

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. Ten samples of length 10,000 were generated from each parameterization, and the average estimated $\hat{\rho}^2$ and $\hat{R}^2$ are reported[8] in Table 3. The results are quite interesting. We can see that it is possible for R^2 to exceed 1 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 meant that the ρ^2 and R^2 measures will be low. It will depend on the nature of the correlation.

Finite sample size for tests of nominal size 5% is reported[9] in Table 4. We find a substantial range of size behavior, with some parameter designs producing over-rejection, and others producing under-rejection. In general, the CADF tests have more size distortion than the ADF test.

Power against the alternatives $c = 4$, 8, and 15 was examined. To eliminate size distortion, the power calculations were done with finite sample critical values, obtained from the simulated data generated under the null hypothesis. Table 5 reports the results for $c = 8$, the other results being sim-

TABLE 3. Simulation design, VAR model

	Parameters			CADF(2,0,0)		CADF(2,1,0)		CADF(2,0,1)		CADF(2,1,1)	
Design	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

Table 4. Size of 5% tests, VAR model

Design	ADF(2)	CADF(2,0,0)	CADF(2,1,0)	CADF(2,0,1)	CADF(2,1,1)
1	5	5	5	5	5
2	5	10	6	7	5
3	5	10	6	10	6
4	6	10	6	12	6
5	6	11	6	12	6
6	5	5	5	5	5
7	5	5	5	5	5
8	5	5	5	5	5
9	5	5	5	4	4
10	5	2	5	5	6
11	5	2	5	2	5
12	6	2	5	1	4
13	6	3	5	1	4
14	4	1	4	5	6
15	5	1	5	1	5
16	6	2	5	0	5
17	7	2	6	0	3

Table 5. Power of 5% tests, VAR model

Design	ADF(2)	CADF(2,0,0)	CADF(2,1,0)	CADF(2,0,1)	CADF(2,1,1)
1	19	19	19	18	19
2	19	14	18	17	22
3	20	20	20	20	20
4	20	28	23	28	25
5	21	35	22	40	34
6	20	20	20	20	19
7	19	28	27	28	27
8	20	38	37	45	43
9	20	52	48	65	63
10	19	23	26	20	26
11	20	37	44	34	42
12	17	51	68	67	72
13	19	69	86	88	88
14	21	26	38	18	35
15	20	39	64	35	63
16	19	59	88	82	88
17	18	81	99	99	99

ilar and excluded to conserve space. The ADF tests have power that is roughly independent of the design, ranging from 17 to 22%. 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. Note that it is important to get the "correct" covariates, for major losses in power can be obtained by inclusion or exclusion of covariates.

4.2. Moving Average Design

Most Monte Carlo investigations of unit root tests have used moving average specifications. We report here a simple multivariate extension. The data generating process is

$$\Delta y_t = -(c/T)y_{t-1} + v_t + \theta v_{t-1}$$

$$v_t = b\Delta x_t + e_t$$

with Δx_t and e_t i.i.d. mutually independent, $N(0,1)$ variables. The unconditional process for Δy_t is an ARMA(1,1), which nests the conventional simulation design and is not a function of b. The "optimal" regression model contains all past lags of Δy_t and the current value of Δx_t, and the parameters b and ρ^2 are related as $\rho^2 = 1/(1 + b^2)$.

In our implementation of the tests, the sample size was set at 100, and five AR lags[10] were included in each regression, as well as an intercept. For the CADF tests, the current value of Δx_t was also included, and rejection was based on estimated critical values calculated from Table 1 and an estimate $\hat{\rho}^2$. Note that the finite sample distribution of the ADF test is independent of b.

Table 6 reports null rejection frequencies based on the asymptotic 5% critical values. As is well known, the ADF test over-rejects for large negative values of θ. Similarly, the CADF test over-rejects for large negative θ, but the size distortion is much more severe and is accentuated for small ρ^2. The size distortion largely disappears for $\theta \geq -.5$.

Table 7 reports finite sample power against the alternative $c = 8$. Finite sample critical values were used to eliminate size distortion. The power of the tests is found to be dramatically influenced by the choice of ρ^2 and is much less affected by the value of θ. The Monte Carlo results make clear that the power of the CADF test can be substantially greater than the ADF test.

5. APPLICATION TO U.S. TIME SERIES

Nelson and Plosser (1982) challenged the prevailing wisdom of the 1970's by showing that the ADF test did not reject the hypothesis of a unit root for many U.S. annual macroeconomic time series. The original Nelson–Plosser

TABLE 6. Size of asymptotic 5% tests, MA model

	ADF statistic	CADF statistics $\rho^2 = .8$	$\rho^2 = .4$	$\rho^2 = .1$
$\theta = -.8$	19	22	34	66
$\theta = -.7$	9	10	12	24
$\theta = -.6$	6	6	7	9
$\theta = -.5$	6	5	6	6
$\theta = .5$	6	6	5	5
$\theta = .6$	6	7	6	5
$\theta = .7$	6	6	7	6
$\theta = .8$	7	7	7	5

data ran up to 1970 but were 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. Because 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.

Tables 8–10 present the results. For GNP and industrial production the first difference of the unlogged unemployment rate was used as Δx_t, and

TABLE 7. Power of asymptotic 5% tests, MA model

	ADF statistic	CADF statistics $\rho^2 = .8$	$\rho^2 = .4$	$\rho^2 = .1$
$\theta = -.8$	25	37	74	99
$\theta = -.7$	24	35	74	99
$\theta = -.6$	20	32	70	99
$\theta = -.5$	17	29	67	99
$\theta = .5$	14	23	60	97
$\theta = .6$	14	21	57	98
$\theta = .7$	17	27	57	97
$\theta = .8$	16	24	60	98

TABLE 8. Unit root tests for GNP, using unemployment rate as covariate

		CADF tests			
		$k_2 = 0$		$k_2 = 2$	
	ADF	$k_1 = 0$	$k_1 = 2$	$k_1 = 0$	$k_1 = 2$
$\hat{\delta}$	−.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*
$\hat{\rho}^2$		.06	.08	.07	.08

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

for the unemployment rate the growth rate of industrial production was used as Δx_t. The OLS estimate of $\hat{\delta}$, its standard error, and $t(\hat{\delta})$ are presented for all cases and the estimated $\hat{\rho}^2$ for the covariate regressions.

First examine GNP (Table 8). The ADF t-test is not significant. The point estimate for the coefficient on lagged GNP is about −.20, with large standard errors, suggesting that the univariate series is uninformative. The CADF tests are more revealing. For each lag specification, three statistics are significant at the asymptotic 5% level and one at the asymptotic 1% level. The estimated $\hat{\rho}^2$ are extremely low, ranging from .06 to .08. This indicates that the estimates should be quite precisely estimated and the power considerably higher than for the ADF tests. Interestingly, whereas the t-statistics show that δ is statistically significantly different from 0, the point estimates (about −.08) are much closer to 0 than the ADF estimates. Hence, although real per capita GNP appears to be I(0), it is highly persistent.

Second, examine the industrial production series (Table 9). The ADF t-test statistic lies quite close to the asymptotic 5% critical value, and the point estimate and standard error of δ indicate considerable uncertainty. The CADF tests, however, strongly support the unit root hypothesis. The point estimates of δ are about −.06, with insignificant t-statistics. We conclude that the industrial production series is I(1).

Third, turn to the unemployment rate (Table 10). The ADF test suggests that the series is I(0), but the CADF tests suggest that the series is I(1). For the specification with the lowest $\hat{\rho}^2$, we find $\hat{\delta} = -.11$, suggesting considerable persistence, even if the data are I(0). The conflict between the tests makes a definitive conclusion difficult.

One might question the wisdom of this final test, because it flips the definition of y_t and x_t from the previous test. The procedure is justified if we think of the joint hypothesis that all both series are I(1). Under this joint null,

TABLE 9. Unit root tests for industrial production, using unemployment rate as covariate

		CADF tests			
		$k_2 = 0$		$k_2 = 2$	
	ADF	$k_1 = 0$	$k_1 = 2$	$k_1 = 0$	$k_1 = 2$
$\hat{\delta}$	−.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
$\hat{\rho}^2$		.21	.16	.17	.15

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

the stated asymptotic distributions are appropriate. It is true that the power of the tests will be a function of the order of integration of the included covariate, but it is not (asymptotically) less than the power of the ADF test.

6. CONCLUSION

This paper has analyzed the distribution theory for tests of unit roots in regression models with covariates. We have found that the asymptotic distribution of the t-statistic for a unit root is a convex combination of the standard normal and the Dickey–Fuller distribution. We have also found that this

TABLE 10. Unit root tests for unemployment rate, using industrial production as covariate

		CADF tests			
		$k_2 = 0$		$k_2 = 2$	
	ADF	$k_1 = 0$	$k_1 = 2$	$k_1 = 0$	$k_1 = 2$
$\hat{\delta}$	−.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
$\hat{\rho}^2$		.53	.20	.58	.39

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

complication implies a major benefit: power is dramatically improved. Our Monte Carlo study suggests that the finite sample size of the CADF t-statistic is adequate and the power against stationarity excellent. We applied these tests to long time series and found evidence to support the contentions that real per capita GNP is stationary but highly persistent, and that industrial production is I(1).

NOTES

1. Assumption 1.7 does exclude certain multivariate processes. For example, if the model is

$$\begin{pmatrix} y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ a & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \Delta x_{t-1} \end{pmatrix} + \epsilon_t$$

$$A = \begin{bmatrix} 1 & -1 \\ a & 0 \end{bmatrix},$$

then equations (4) and (5) are satisfied with $\delta = 0$. In this model, if $0 < a < \frac{1}{4}$, the long-run variance of y_t is $2/a^2$, so y_t is stationary, and, hence, a test of $\delta = 0$ cannot be interpreted as a test for integration in y_t. Assumption 1.7 excludes this class of models from consideration because σ_v^2 equals the long-run variance of Δy_t, which is 0 when $a > 0$.

2. The power envelope was calculated for each value of ρ^2 shown and at $c = 1,2,\ldots,16$. The distributions were approximated by calculations from samples of size 1,000 with i.i.d. 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.

3. The critical values were calculated from 60,000 draws generated from samples of size 1,000 with i.i.d. Gaussian innovations.

4. For example, under model (11), $\hat{e}_t = \hat{a}^\mu(L)\Delta y_t - \hat{\delta}^\mu y_{t-1} - \hat{\mu}^* - \hat{b}(L)'\Delta x_t$ and $\hat{v}_t = \hat{b}(L)'(\Delta x_t - \overline{\Delta x}) + \hat{e}_t$.

5. The power functions were calculated for each ρ^2 and $c = 1,2,\ldots,16$ from 20,000 samples of size 1,000 with i.i.d. Gaussian innovations, using the asymptotic critical values from Table 1.

6. Rejection frequencies were calculated from 5,000 samples each of length 1,000, and the asymptotic 5% critical values from Table 1 were used. The choice of row from Table 1 was determined from a sample estimate of ρ^2. Similar results were found at the 10 and 1% levels, samples of size 100, and imposing the true value of ρ^2.

7. An earlier version of this paper also considered tests based on the normalized coefficient $T(\hat{\delta} - 1)$. It turns out that the size of these test statistics is highly sensitive to over-differencing, reducing the appeal of the normalized coefficient tests relative to the t-tests. I thank a referee for drawing my attention to this point.

8. Standard errors (not shown) indicate that the estimates are quite precise.

9. All experiments used 5,000 replications.

10. Fewer lags were also tried but resulted in larger size distortions.

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Berk, K.N. (1974) Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.

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

Elliott, G., T.J. Rothenberg, & J.H. Stock (1992) Efficient Tests for an Autoregressive Unit Root. Manuscript. Harvard University.
Engle, R.F. & C.W.J. Granger (1987) Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55, 251–276.
Hansen, B.E. (1992a) Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* 60, 967–972.
Hansen, B.E. (1992b) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.
Herrndorf, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables. *Annals of Probability* 12, 141–153.
Horvath, M.T. & M.W. Watson (1993) Testing for Cointegration When Some of the Cointegrating Vectors Are Known. Working paper 93-15, Federal Reserve Bank of Chicago.
Johansen, S. & 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.
Kremers, J.J.M., N.R. Ericsson, & J.J. Dolado (1992) The power of cointegration tests. *Oxford Bulletin of Economics and Statistics* 54, 325–348.
Nelson, C.R. & C.I. Plosser (1982) Trends and random walks in macroeconomic time series. *Journal of Monetary Economics* 10, 139–162.
Newey, W.K. & K.D. West (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–653.
Phillips, P.C.B. (1987) Toward a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.
Phillips, P.C.B. & S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–193.
Said, S.E. & D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–608.
Schotman, P.C. & H.K. van Dijk (1991) On Bayesian routes to unit roots. *Journal of Applied Econometrics* 6, 387–401.

APPENDIX: MATHEMATICAL PROOFS

LEMMA.

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

Proof. We prove part 1 later. Part 2 follows by the continuous mapping theorem, and part 3 by Theorem 4.4 of Hansen (1992b) and (7).

Note that

$$a(L) = a(1) + a^*(L)(1-L), \tag{A.1}$$

where $a^*(L)$ has all roots outside the unit circle. Let $\xi_t = a^*(L)\Delta S_t$, which satisfies $\sup_{t\le T}|\xi_t| = o_p(\sqrt{T})$. Using (8) and (A.1), we find that (5) can be rewritten as $\Delta S_{at} = -(c/T)S_{at-1} + (c/T)\xi_{t-1} + v_t$, where $S_{at} = a(L)S_t$. Hence,

$$\frac{1}{\sqrt{T}} \sup_{t\le T} |S_{at} - S^*_{at}| \to_p 0, \tag{A.2}$$

where S^*_{at} is generated by $\Delta S^*_{at} = -(c/T)S^*_{at-1} + v_t$. By Theorem 4.4(a) of Hansen (1992b), we know that

$$\frac{1}{\sqrt{T}} S^*_{a[Tr]} \Rightarrow \sigma_v W_1^c(r). \tag{A.3}$$

Equations (A.2) and (A.3) yield

$$\frac{1}{\sqrt{T}} S_{a[Tr]} \Rightarrow \sigma_v W_1^c(r). \tag{A.4}$$

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 because $a^*(L)$ does. By (A.4),

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

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

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 the lemma yields

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

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 ϕ_t be defined as before Assumption 1. Because ϕ_t is covariance stationary, ergodic, and strong mixing, and $E(\phi_t e_t) = 0$ under Assumption 1, we know that $(1/T)\sum_{t=2}^T S_{t-1}\phi_t' = O_p(1)$, $(1/T)\sum_{t=2}^T \phi_t\phi_t' \to_p E(\phi_t\phi_t') > 0$, and $(1/T)\sum_{t=2}^T \phi_t e_t \to_p 0$. Thus,

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

$$= \frac{\frac{1}{T}\sum_{t=2}^T S_{t-1}e_t}{\frac{1}{T^2}\sum_{t=2}^T S_{t-1}^2} + o_p(1).$$

Hence, by the lemma,

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

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

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

Proof of Theorem 3. Because

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

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

and $T\delta = -ca(1)$ under (8), Theorem 2 yields

$$t(\hat{\delta}) \Rightarrow \sigma_e^{-1}\left(a(1)^{-2}\sigma_v^2 \int_0^1 (W_1^c)^2\right)^{1/2}$$

$$\times \left[-ca(1) + a(1)R\left(\rho \frac{\int_0^1 W_1^c\, dW_1}{\int_0^1 W_c^2} + (1-\rho^2)^{1/2}\frac{\int_0^1 W_1^c\, dW_2}{\int_0^1 W_c^2}\right)\right]$$

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

Finally, the fact that W_2 is independent of W_1 and W_1^c implies that $\int_0^1 W_1^c\, dW_2/(\int_0^1 (W_1^c)^2)^{1/2}$ is distributed $N(0,1)$ and is independent of W_1^c. This completes the proof. ■