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

by

Bruce E. Hansen

Working Paper No. 300
May 1995

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

Bruce E. Hansen*
Department of Economics
Boston College
(617) 552-3678

May 1994
Revised: May 1995

*I thank three referees and the editor for helpful comments, and the National Science Foundation and the Sloan Foundation for research support.

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 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.

Dolado (1992) discuss the model

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

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

with ϵ_{1t} and ϵ_{2t} iid uncorrelated normal random variables. We can see that (2) falls in the 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) 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 here.

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. It appears that our tests are complementary to those of Johansen-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 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 which 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 $\rho^2 = \sigma_e^2/\left(\sigma_e^2 + b(1)'\Sigma b(1)\right)$.

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. *$\sup_t E\left[|\Delta x_t|^p + |e_t|^p\right] < \infty$;*
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 $\phi_t = \left(\Delta y_{t-1}, \ldots, \Delta y_{t-p}, \Delta x'_{t+q_2} - \mu_x, \ldots, \Delta x'_{t-q_1} - \mu_x\right)'$;*
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

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 iid $N(0,\sigma_e^2)$ 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 $\overline{\delta} < 0$ rejects for small values of

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

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

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 $(c, \overline{c}, R^2, \rho^2)$. 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 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

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

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.

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

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

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

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

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 (14) depends on the two nuisance parameters ρ^2 and R^2, the null distribution (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 of ρ^2 in steps of 0.1. 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

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

3.2 Asymptotic Local Power Functions

Theorem 3 gives the asymptotic distribution of the CADF t-statistic under the local-to-unity alternative (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 $\rho^2 = 1.0$, 0.7, 0.4, and 0.1, 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 are not displayed on these graphs to retain clarity. 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 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. 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 upon 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 $I(1)$.

[5] 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 Table 1.

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

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 non-zero 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.)

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 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.

[6]Rejection frequencies were calculated from 5000 samples each of length 1000, 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 the paper also considered tests based on the normalized coefficient $T(\hat{\delta}-1)$. It turns out that the size of these test statistics are highly senstive 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.

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

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 $a_{11} = 0$, since 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 (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 $\{-.3,\ 0,\ .3,\ .6\}$.

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. 10 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 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.

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

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

[9] All experiments used 5000 replications.

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

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 similar and excluded to conserve space. The ADF tests have power which is roughly independent of the design, ranging from 17-22%. 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.

Table 7: Power of Asymptotic 5% Tests, MA Model

	ADF Statistic	CADF Statistics		
		$\rho^2 = .8$	$\rho^2 = .4$	$\rho^2 = .1$
$\theta = -0.8$	25	37	74	99
$\theta = -0.7$	24	35	74	99
$\theta = -0.6$	20	32	70	99
$\theta = -0.5$	17	29	67	99
$\theta = 0.5$	14	23	60	97
$\theta = 0.6$	14	21	57	98
$\theta = 0.7$	17	27	57	97
$\theta = 0.8$	16	24	60	98

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.

Table 9: Unit Root Tests for IP, Using UR as Covariate

	ADF	CADF 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
$\hat{\rho}^2$		.21	.16	.17	.15
* Significant at the asymptotic 5% level					
** Significant at the asymptotic 1% level					

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, while 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, while 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 indicates 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 IP 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 is $I(0)$. The conflict between the tests makes an definitive conclusion difficult.

One might question the wisdom of this final test, since 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, the stated asymptotic distributions are

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

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

[5] Engle, R.F. and C.W.J. Granger (1987), "Co-integration and error correction: Representation, estimation and testing," *Econometrica*, 55, 251-276.

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

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

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

[9] 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.

[10] 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.

[11] Kremers, J.J.M., N.R. Ericsson, and J.J. Dolado (1992), "The power of cointegration tests," *Oxford Bulletin of Economics and Statistics*, 54, 325-348.

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

[13] Newey, W.K. and K.D. West (1994), "Automatic lag selection in covariance matrix estimation," *Review of Economic Studies*, 61, 631-653.

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

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 (23),

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

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

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

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

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

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

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

Hence by the Lemma,

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

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

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

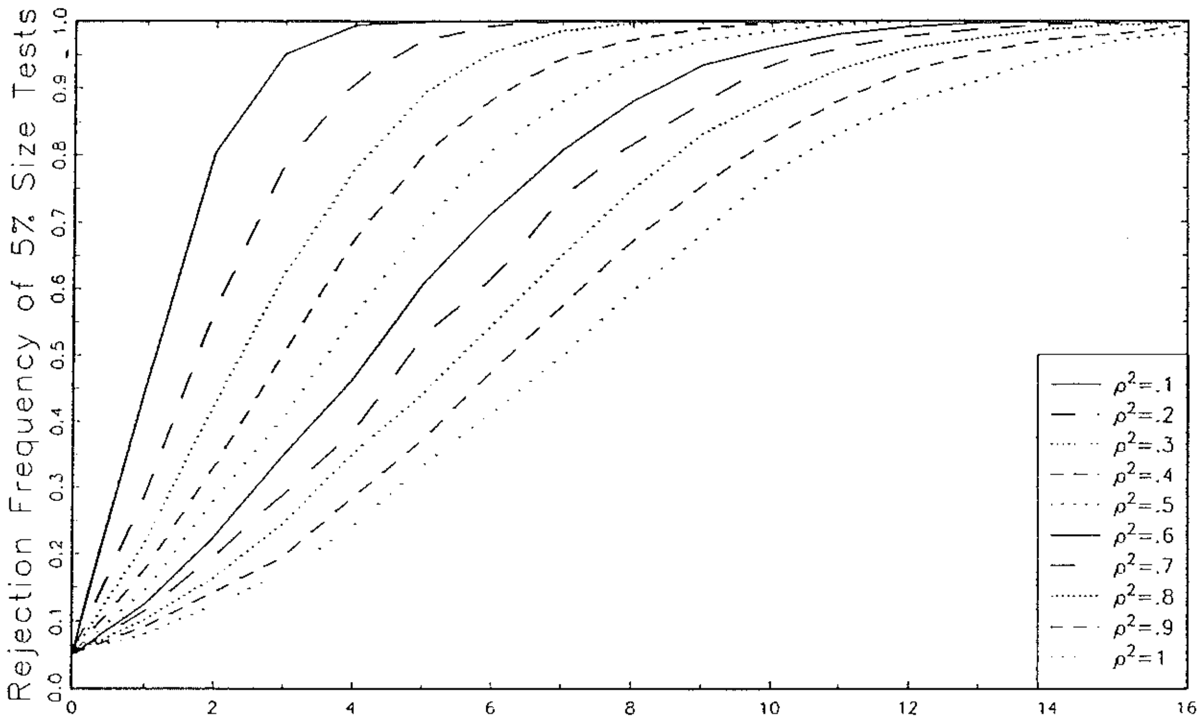

Figure 2: Comparison of Power Under No Mean Correction with Power Envelope, $R^2=\rho^2$

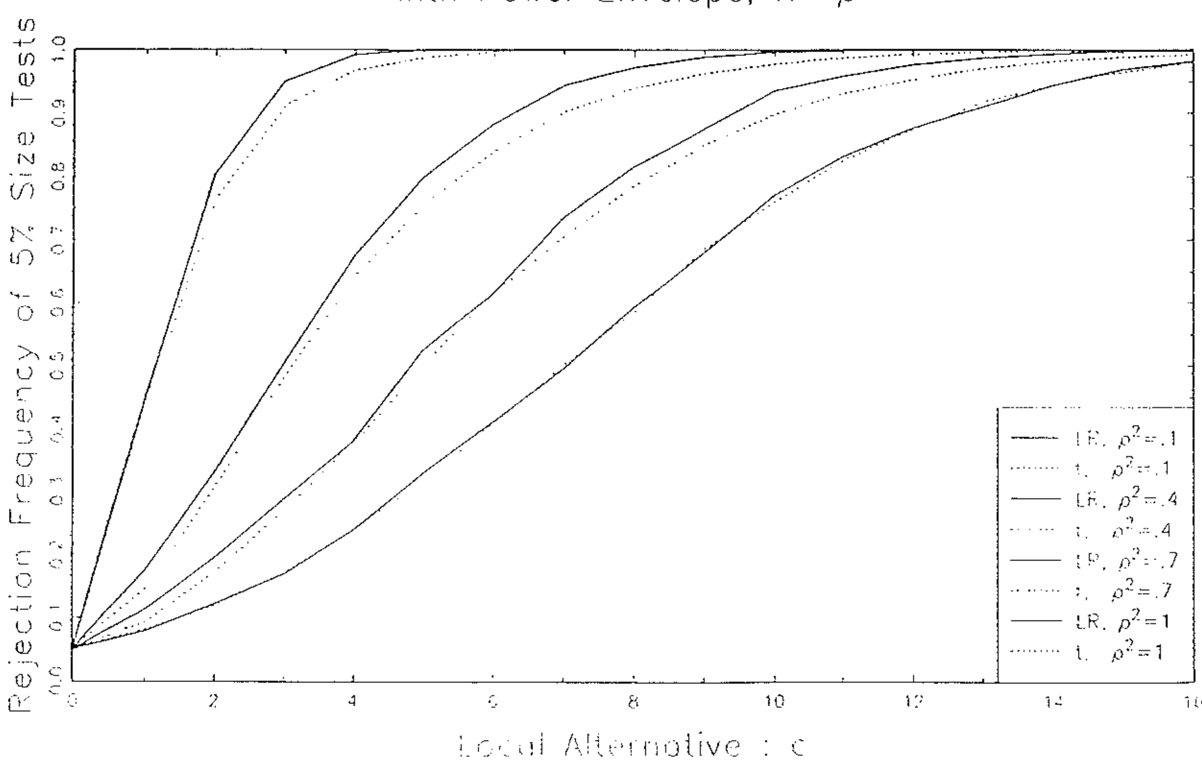

Figure 5: Power Under Mean Correction, $R^2 \neq \rho^2$

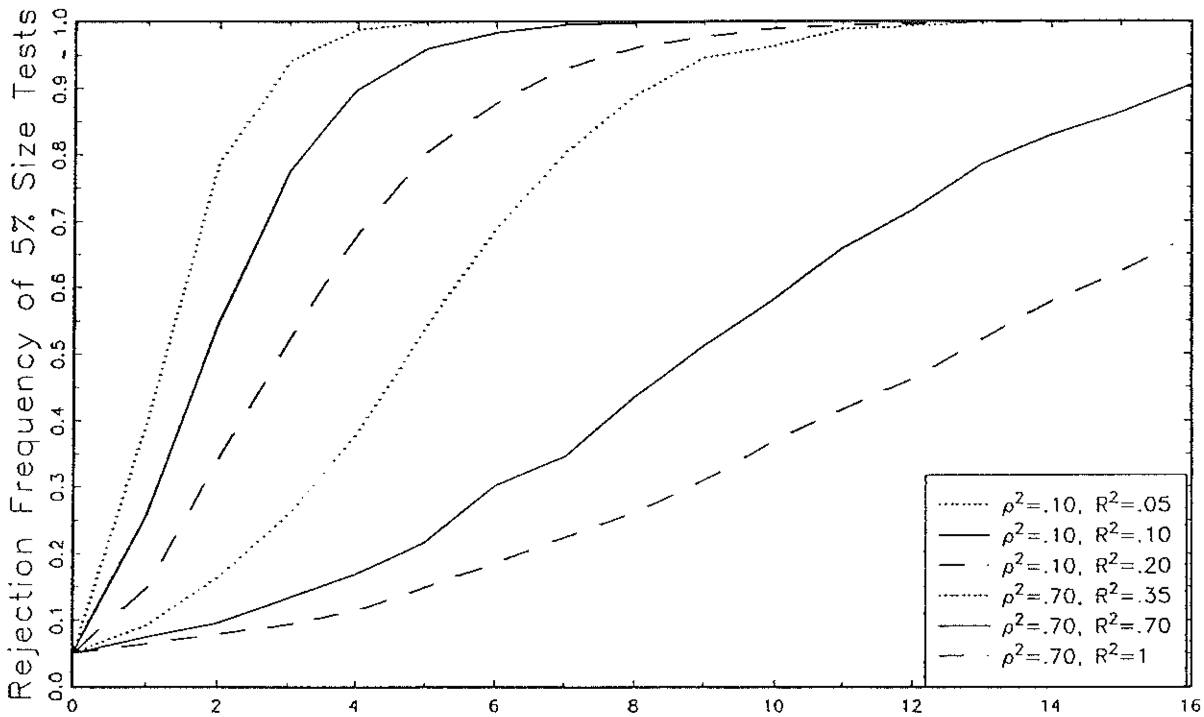

Figure 6: Power with Over-Differenced Regressors