RESIDUALS BASED TESTS FOR COINTEGRATION: AN ANALYTICAL COMPARISON.

Elena Pesavento*
Department of Economics
Emory University

5 March, 2002

*Department of Economics, Emory University, phone: (404) 712-9297.
RESIDUALS BASED TESTS FOR COINTEGRATION: AN ANALYTICAL COMPARISON.

Address correspondence to:

Elena Pesavento  
Department of Economics  
Emory University  
Atlanta, GA 30322;

e-mail: epesave@emory.edu;  
phone: (404) 712 9297;  
fax: (404) 727 4639.
Abstract

This paper compares five residuals-based tests for the null of no cointegration to identify which unit root test should be used when testing for cointegration. The tests are compared in terms of power and size distortions. The asymptotic distribution of the tests under the local alternative is shown to be a function of Brownian Motions and Ornstein-Uhlenbeck processes, depending on a single nuisance parameter, which is, determined by the correlation at frequency zero of the independent variables with the errors of the cointegration regression. It is shown that no significant improvement can be achieved by using different unit root tests than the t-test originally proposed by Engle and Granger (1987).

JEL classification: C32

Keywords: cointegration, residual tests, unit root, power.

This paper is based on Chapter 1 of my dissertation. I thank Robert DeJong, Hashem Dezhbakhsh, Graham Elliott, Clive Granger, Jeff Wooldrige, and the participants to the 2001 Australasian Econometrics Society meeting for valuable comments. All errors remain my responsibility.
1. INTRODUCTION

The purpose of this paper is to provide a careful analysis of residuals-based tests for the absence of cointegration. In their seminal paper on cointegration, Engle and Granger (1987) suggest a two-step analysis to test for cointegration: estimate the cointegrating regression by OLS in the first step and then test for a unit root in the residuals from the cointegrating regression in the second step. If the null of a unit root in the residuals is rejected, then there is evidence of cointegration. Because the cointegrating vector is estimated in the first stage, the asymptotic distributions of tests on the residuals will be different than the non-estimated unit root case. The distribution generally shifts to the right. Engle and Yoo (1987) compute the asymptotic distribution of the Dickey-Fuller (DF) and the Augmented Dickey-Fuller (ADF) tests under the null hypothesis that the series are independent random walks. Phillips and Ouliaris (1990) relax the independence assumption and compute the asymptotic null distribution of the Dickey-Fuller tests and other residuals-based tests.

The asymptotic distribution of residuals-based tests are functions of Brownian Motion and are similar with respect to the cointegration vector and the variance covariance matrix of the innovations. Although Engle and Granger recommend using the ADF test, in practice, different unit root tests have been applied to test for cointegration. As in the case of unit roots, we expect different tests to have different asymptotic distributions and to perform differently in term of size and power. In particular, since statistics that perform better than the ADF test have been found in unit root testing, it is conceivable that the same tests will be superior when testing cointegration.
There is a vast literature that examines various integration tests in terms of their asymptotic power, size distortions, and performances in small samples\(^1\). Elliott et al. (1996), for example, compute the asymptotic power envelope for unit root tests and show that, although no feasible test will lie exactly on the power envelope, it is possible to find tests that are asymptotically close to the power envelope and also have very good small sample properties. They show that GLS (or quasi-difference) detrending leads to estimates of the mean and trend that are asymptotically more efficient than OLS estimates, and show that the power for the Dickey-Fuller test with GLS detrending is almost indistinguishable from the power envelope. Since the improvement in the test is mostly given by the different detrending procedure, the gain from using the DF-GLS test or the PT test of Elliott et al. (1996) is large in the case when either mean or trend are present. Because in cointegration tests the unit root tests are applied to residuals from an OLS regression, which have by definition mean zero, there is no gain in applying Elliott et al. (1996) tests directly. However, little work has been done to compare the power of residuals-based tests for cointegration. Pesavento (2000) shows that different tests for the null of no cointegration can have significantly different power. In fact, in this non-standard environment, all the tests have non-normal asymptotic distribution and no uniformly most powerful test exists. It is therefore not clear which test is better when testing cointegration. It therefore merits to examining how different residual-based tests of cointegration behave.

Currently, there is no analytical solution for the asymptotic power of residual tests for cointegration and the relative performance of the tests has been studied using Monte Carlo experiments (Haug (1993, 1996), Gonzalo and Lee (1998), Kremers et al. (1992))

\(^1\) See Stock (1994, 1999) and Phillips and Xiao (1998) for a comprehensive survey of tests for integration
However, with Monte Carlo simulations, the results of the experiments depend on the particular design run and conclusions are tentative. In this paper, I analytically derive results for the power of residuals-based tests for cointegration, and I show which features of the model are important for power. Once we fully understand which parameters are important for power, a complete set of Monte Carlo simulations can be designed. A study of the performance of different tests in terms of power, performance in small samples, and size distortions can determine if there is any gain in using a different test than the ADF test originally recommended by Engle and Granger.

Since all tests are consistent, they all have power equal to one asymptotically and thus the asymptotic power for fixed alternatives cannot be used to rank the tests. I therefore derive the asymptotic distribution of these tests under a local alternative parameterization. Having an analytical formulation for the asymptotic distribution of the tests, I can approximate the asymptotic power directly. I show that, unlike for unit root tests, there is little gain in using different residual tests. Small sample properties and size issues are also examined.

Section 2 gives a brief description of the different tests analyzed in this paper, the analytical asymptotic power functions are analyzed in section 3, and section 4 compares the local power of these tests in large samples. Small sample properties and size distortions are presented in section 5. Section 6 concludes.

2. RESIDUALS BASED TESTS

Consider the model:
\[
\Delta x_t = d_{1t} + v_{1t}, \\
y_t = d_{2t} + x_t'\beta + u_t, \\
u_t = \rho u_{t-1} + v_{2t},
\]

(2.1)

where \( t = 1, \ldots, T \); \( x_t \) is a \( n_1 \times 1 \) vector, \( y_t \) is a scalar, and \( d_{1t} = G_1z_{1t} \) and \( d_{2t} = G_2z_{2t} \) are deterministic terms. \( v_t = [v_{1t}' v_{2t}'] \) are serially correlated errors with \( \Phi(L)v_t = \varepsilon_t \). \( \varepsilon_t = [\varepsilon_{1t}' \varepsilon_{2t}'] \) is a \( n \times 1 \) vector of martingale differences with positive definite variance covariance \( \Sigma \) such that the partial sum \( \frac{1}{\sqrt{T}} \sum_i v_i \) satisfies a multivariate invariance principle\(^2\). \( \Phi(L) \) is an invertible lag polynomial of known order partitioned conformably to \( v_t \) such that \( \Phi(L) = \begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) \\ \phi_{21}(L) & \phi_{22}(L) \end{bmatrix} \). The spectral density of \( v_t \) at frequency zero (scaled by \( 2\pi \)) is \( \Omega = \Phi(1)^{-1}\Sigma(1)^{-1}' \) where \( \Phi(1) = \sum_i \Phi_i \). \( \Phi(1) = \sum_i \Phi_i \). \( \Omega \) can be partitioned as:

\[
\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}
\]

\( R^2 = \delta \delta' \) where \( \delta = \Omega_{11}^{-1/2}\omega_{12}\Omega_{22}^{-1/2} \) is the bivariate zero frequency correlations of each element of \( v_{1t} \) with \( v_{2t} \). \( R^2 \) represents the contribution of the right hand variables as it is zero when these variables are not correlated in the long run with the errors from the cointegration regression. To assure that \( x_t \) are not individually cointegrated, \( \Omega_{11} \) is assumed non-singular. For the purpose of this paper, I consider the cases where (i)

\(^2\) This is valid for general case of weakly dependent heterogeneous variables. For conditions under which the multivariate invariance principle holds see Phillips and Solo (1992), or Wooldridge (1994) for a review.
\[ z_{1t} = 0 \text{ and } z_{2t} = 0, \quad (ii) \quad z_{1t} = 0, \quad z_{2t} = 1 \text{ and a constant is included in the regression, (iii) } \]
\[ z_{1t} = 1, \quad z_{2t} = 1 \text{ and a constant and a time trend are included in the regressions.} \]

When \( \rho < 1 \), the linear combination \( y_t - d_{2t} - x_t' \beta \) is stationary, \( y_t \) and \( x_t \) are cointegrated and the system (2.1) contains \( n_1 \) unit roots; when \( \rho = 1 \) the two variables are not cointegrated and there are \( n \) unit roots in the system.

Residual tests for cointegration are based on the OLS residuals from the cointegration regression:
\[ y_t = d_{2t} + x_t' \beta + u_t \quad (2.2) \]

Since \( \hat{\beta}_{OLS} \) is a consistent estimator for the true \( \beta \), testing stationarity of the residuals is a good proxy for testing that the linear combination above is stationary. Rejection of a unit root in the residuals from (2.2) is an indication of cointegration between the two variables.

As Stock (1999) shows, most unit root tests can be classified within one unique class. Following his and Phillips and Ouliaris (1990) I will consider the asymptotic properties of the following residual based tests:

1) The Augmented Dickey-Fuller \( t_u \) and \( \hat{\alpha}_T \) tests in the augmented regression
\[ \Delta \hat{u}_t = \alpha \hat{u}_{t-1} + \sum_{i=1}^p \pi_i \Delta \hat{u}_{t-i} + \xi_t. \]

2) The Phillips (1987) \( Z_t \) and \( Z_\alpha \) tests from the regression \( \hat{u}_t = \hat{\rho} \hat{u}_{t-1} + \hat{k} \), where
\[ \hat{Z}_t = \frac{(\hat{\rho} - 1)(\sum \hat{u}_{t-1}^2)^{1/2}}{s_{TT}^2} - \frac{1}{2} \left( s_{TT}^2 - s_k^2 \right)^{1/2} \quad \text{and} \quad \hat{Z}_\alpha = T (\hat{\rho} - 1) - \frac{1}{2} \left( s_{TT}^2 - s_k^2 \right) \]

where
\[ s_k^2 = T^{-1} \sum_{t=1}^{T} \hat{k}_t^2 \text{ and} \]
\[ s_{lT}^2 = T^{-1} \sum_{t=1}^{T} \hat{k}_t^2 + 2T^{-1} \sum_{s=1}^{l} w_{s} \sum_{t=s+1}^{T} \hat{k}_t \hat{k}_{t-s} \] with \( w_s = 1 - s/(l+1) \). \( s_{lT}^2 \) is the sum of covariances estimator of the spectral density at frequency zero of \( \hat{k}_t \).

3) A modified version of the Sargan and Bhargava (1983) test:
\[ MSB = \frac{\left( T - 2 \sum \hat{u}_t^2 \right)^{1/2}}{\hat{\omega}_{AR}} \] where \( \hat{\omega}_{AR} \) is a parametric estimator of the spectral density at frequency zero of \( \hat{k}_t \) that can be easily obtained from the augmented regression in 1.

For each test, \( \hat{u}_t \) are the residuals from the LS estimation of the cointegration regression (2.2) estimated with no mean when \( z_{1t} = 0 \) and \( z_{2t} = 0 \), or mean only when \( z_{1t} = 0, \ z_{2t} = 1 \). When \( z_{1t} = 1, \ z_{2t} = 1 \), mean and trend are included in the cointegrating regression. Hansen (1992) shows that, if the variables have a drift, and the regression (2.2) is run only with a mean, the above tests will have different distributions, with critical values that depend on whether the drift is different than zero or not. Although Hansen (1992) shows that in such case the tests have higher power, we may still want to estimate the regression with a trend when we are not sure if any of the elements of \( x_t \) have a drift. This paper only considers the case when a mean and trend are included in the regression (2.2) for the above reason and also because it is in general more widely used.

The following conditions on the expansion rate of the lag length \( p \) and the lag truncation parameter \( l \) are assumed throughout the paper:

\[ \frac{\sum_1^p \phi_p}{\sum_1^l} < \frac{1}{\sqrt{T}} \]
ASSUMPTION A1: \( p \to \infty \) as \( T \to \infty \) and \( p = o(T^{1/3}) \).

ASSUMPTION A2: \( l \to \infty \) as \( T \to \infty \) and \( l = o(T^{1/4}) \).

3. ASYMPTOTIC POWER

Given that the traditional optimality theory cannot be applied to the case of tests for the absence of cointegration, there is, in general, no reason to expect one test to perform uniformly better. The literature comparing tests for cointegration uses Monte Carlo experiments with various possible combinations of values for the parameters of the data generating process. To be able to compare the performance of the different unit root tests when testing for cointegration, I need to compute the analytical power of the tests. The knowledge of which nuisance parameter enters the local power will help us design Monte Carlo experiments that are informative in terms of the direction of the parameter space that we need to examine.

As Phillips and Ouliaris (1990) show, the residuals tests from the previous sections are consistent tests, so they cannot be compared on the basis of their asymptotic power for fixed alternatives as it equals to one for any large sample. As it is commonly done, we can look at a sequence of local alternatives. The model (2.1) naturally suggests the use of the parametrization \( \rho = 1 + c/T \). When \( c \) is equal to zero the errors \( u_t \) are integrated of order one. For \( c \) negative, the variables in equation (2.2) are cointegrated. Using the results of Phillips (1988) for local to unit processes I compute the analytical power for the three residuals-based tests previously presented.
THEOREM 1: When the model is generated according to (2.1) and Assumption A1 and A2 are valid, then, as $T \to \infty$:

(1) $\hat{Z}_t, \hat{\alpha}_T \Rightarrow c \left( \begin{bmatrix} \eta_c^d \ A^d \eta_c^d \\ \eta_c^d \ D \eta_c^d \end{bmatrix} \right)^{-1/2} \int W^d_c d\hat{\tilde{W}} \eta_c^d$

(2) $\hat{Z}_t, \hat{\alpha}_T \Rightarrow c + \frac{\eta_c^d \ A^d \eta_c^d}{\eta_c^d \ A^d \eta_c^d}$

(3) $MSB \Rightarrow \left( \begin{bmatrix} \eta_c^d \ A^d \eta_c^d \\ \eta_c^d \ D \eta_c^d \end{bmatrix} \right)^{-1/2}$

where:

$\eta_c^d = \left( \int W^d_{12} J^d_{12c} \left( \int W^d_{12} W^d_{12} \right)^{-1} \right)$, $W^d_c = \left[ W^d_{12} \ J^d_{12c} \right]$, $A^d_c = \int W^d_c W^d_c$.

$\hat{\tilde{W}} = \left[ \begin{bmatrix} W^d_{12} \ W^d_{12} \end{bmatrix} \right]^\prime$, $W_{12} = \sqrt{\frac{R^2}{1 - R^2} + W_{12}}$, $D = \left[ \begin{bmatrix} I \ \hat{\delta} \\ \hat{\delta} \ 1 + \hat{\delta} \hat{\delta} \end{bmatrix} \right]$ and $R^2 = \hat{\delta} \hat{\delta}$. $J_{12c}$ is a scaled Ornstein Uhlenbeck process such that $J_{12c}(\lambda) = W_{12}(\lambda) + c \int_0^\lambda e^{(\lambda - s)} W_{12}(s) ds$ with

(i) $W^d_c = W^d_c$ and $J^d_{12c} = J^d_{12c}$ if $z_{1t} = 0$ and $z_{2t} = 0$.

(ii) $W^d_c = W^d_c - \int W^d_c$ and $J^d_{12c} = J^d_{12c} - \int J^d_{12c}$ if $z_{1t} = 0$, $z_{2t} = 1$ and a mean is included in the regression.

(iii) $W^d_c = W^d_c - (4 - 6\lambda) \int W^d_c - (12\lambda - 6) \int s W^d_c$ and $J^d_{12c} = J^d_{12c} - (4 - 6\lambda) \int J^d_{12c} - (12\lambda - 6) \int s J^d_{12c}$ if $z_{1t} = 1$, $z_{2t} = 1$ or $z_{1t} = 1$, $z_{2t} = (1, t)$ and mean and trend are included in the regressions.
Theorem 1 shows that the asymptotic distributions of each of the tests are different; thus they have different asymptotic local power functions. Although the power is invariant to the cointegrating vector (i.e. the value of \( \beta \)), the variances of the error terms and the parameters of the serial correlation, it depends additionally on a few parameters. It depends on the alternative \( c \) through the first block of the tests and the Ornstein-Uhlenbeck processes, and on the dimension of \( x_t \). More importantly, although asymptotically the tests do not depend on any nuisance parameter under the null as shown in previous literature (Phillips and Ouliaris (1990)), under the local alternative the power contains the nuisance parameter \( R^2 \) in the functionals \( J_{12c}^d \) and \( W_{12} \). Since \( R^2 \) enters the asymptotic distribution under the alternative, we expect the power of the tests to vary with \( R^2 \).

When \( c = 0 \), \( J_{12c} = W_{12} \) so the asymptotic distribution of the tests under the null is a function of standard Brownian motions and depends only on the dimension of \( x_t \) (see also Phillips and Ouliaris (1990) and Stock (1999)).

4. LARGE SAMPLE RESULTS

The analytical limit distributions of the tests under the local alternative that was derived in the previous section give us a better idea as to how the power is affected by various parameters and how the tests behave in large sample. For this experiment I consider \( c = 0, -1, -5, -10, -15 \) and \( R^2 = 0, 0.5, 0.7 \). Each Brownian Motions is approximated by step functions using Gaussian random walks with \( T = 1000 \) observations; the first 100 observations are discarded to eliminate initial condition effect. Monte Carlo
simulations with 5000 replications of the functionals of Theorem 1 are used to compute the rejections probabilities for each $c$.

Figures 1-3 compare the asymptotic power of the tests in Theorem 1 for the case with no mean, mean only and mean and trend in the cointegration vector. When there is no mean in the cointegration and no constant is included in the OLS regression, all the tests perform very similarly: The ADF $\hat{\alpha}_r$ test has marginally higher asymptotic power especially when we are further away from the null. Generally, no significant increase in power results from using a different test than the most commonly used $t$-test. It is interesting to note that, in the case with a constant, the rankings of the residual tests is the same as that in the unit root case: MSB has slightly better power than $\hat{\alpha}_r$ which performs better than ADF. The power gain from using a test other than $\hat{t}_a$, however, is not as big as in the unit root case. For example, Stock (1994) reports that for $c=-10$ the power of the ADF test when used to test for integration is 35% while the power of MSB test is 55%. The difference in power between the same two tests when used to test for cointegration is only between 0.02% and 0.06% depending on the value of $R^2$. When a mean and a trend are included in the regression, there is no unique ranking between the tests anymore. The $\hat{t}_a$ test performs better in a neighborhood of the null, while $\hat{\alpha}_r$ performs better for alternatives further away. For most of the cases, when a drift in the variable is present, the MSB has the lowest asymptotic power.

As I show in Theorem 1, the asymptotic power of the tests is different for different values of the nuisance parameter $R^2$. As $R^2$ increases, i.e. as the right hand variables in (2.2) are more correlated with the errors from the cointegration regression, I expect the power of the tests to be decreasing. In fact, residual based tests for
cointegration draw on a single equation, (2.2), which does not fully exploit this correlation. Figure 4 shows the asymptotic power of the ADF $\hat{t}_a$ test for different values of $R^2$. As $R^2$ increases, the power of the ADF test shifts to the right. In fact the reduction in power for large value of $R^2$ can be very significant. For example, for the demeaned case, when $c$ is $-20$, the probability of rejecting a false hypothesis is 30% lower when $R^2$ is 0.7 than when $R^2$ is zero. For the detrended case this difference is even larger\(^3\).

To see whether the asymptotic distribution is a good approximation for practical purpose we need to look at the small samples properties of the tests: the power in small samples and the size properties of the tests will be studied in the next sections.

5. SMALL SAMPLES PROPERTIES

Starting with the work of Schwert (1989) a vast literature has been developed to study the small sample performance of unit root tests both in term of size and power. In general, the findings confirm that finite size distortions for unit root tests vary widely across tests depending on the way the spectral density estimator is constructed and the lag length involved\(^4\). Haug (1993, 1996) conducts some Monte Carlo simulations to compare small sample distortions of unit root tests applied to residuals from a cointegration regression. He finds results similar to the unit root tests, in that the tests present significant size distortions especially in the presence of large negative moving average roots. The model

\(^3\) Pesavento (2000) shows that other cointegration tests based on a full system approach can exploit this correlation and have power higher than residual based tests when $R^2$ is large.

\(^4\) Stock (1994) lists many such studies.
derived in Section 2 allows us to compare the small sample power and size distortions in a more general setting than Haug’s (1993).

Using the DGP of equation (2.1) I randomly generate the errors from a bivariate Normal with mean zero and variance-covariance matrix

\[ \Omega = \begin{bmatrix} \omega_1^2 & \omega_1 \omega_2 \delta \\ \omega_1 \omega_2 \delta & \omega_2^2 \end{bmatrix} \]

For \( R^2 = \delta^2 \), I consider \( R^2 = 0, 0.3, 0.7, \) and \( c = 0, -1, -5, -10, -15 \) that corresponds to \( \rho = 1, 0.99, 0.95, 0.9, 0.85 \) for \( T=100 \). The tests are all invariant to \( \beta \) and the variance of the errors so I can choose any number. Table 1 shows the rejection probabilities for the case in which there is no serial correlation in the error terms. Since no serial correlation is assumed all the regressions are estimated without lags. For now, I am only interested in examining the quality of the local-to-unity asymptotic approximations so Tables 1 presents the size-adjusted power. Given that the regressions are estimated without lags and no serial correlation is present, the estimated spectral density in \( \hat{Z}_t \) and \( \hat{Z}_a \) is only marginally different so the results are presented only for the \( \hat{t}_a \), \( \hat{\alpha}_T \) and MSB tests.

Similarly to the unit root case, \( \hat{t}_a \) does not perform as well as the other tests in small samples. Although none of the tests has very high power in a neighbor of the null, the difference in performance is more evident for small values of \( R^2 \). When \( R^2 \) is large, the gain from using MSB over the \( \hat{t}_a \) test is less than 5% in term of power.

More interesting is the case where is some serial correlation in the error terms. It is known that in this case tests for integration and cointegration may have very severe size distortions. Since a test with very good power, but very bad size may not be the best choice, it is important to evaluate the size properties of the tests. For this experiment I
look at case (ii) in which there in no drift in the variables but a mean is present in the cointegration regression. The data are generated as in model (2.1) with 

\[(1-\Phi L)v_t = (1+\Theta L)\epsilon_t\] and \[\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}, \quad \Theta = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix}\] and \(T=100\) observations.

All the regressions are estimated with a mean and the lag length is chosen by BIC with a maximum of four lags\(^5\). The same lag length is used to estimate the spectral density at frequency zero for \(\hat{Z}_t\) and \(\hat{Z}_a\). Table 2 presents the results for different combinations of values for the autoregressive and moving average components.

The size properties of the tests are very similar to what has been previously reported in the literature for univariate unit root tests. For autoregressive roots, the ADF test has more stable size than other tests. For \(\Phi\) diagonal both \(\hat{Z}_t\) and \(\hat{Z}_a\) have empirical size significantly below the nominal level of 5% while the other tests are slightly above. Because the main difference between the two sets of tests is the method of estimating the spectral density, these results emphasize the role of the type of estimator chosen in determining the small sample properties of the tests. When \(\Phi\) is not diagonal the ADF \(\hat{t}_a\) test has empirical size closer to 5% while \(\hat{Z}_t\) and \(\hat{Z}_a\) have empirical size above 10%.

Similar results can be found when looking at positive MA roots in the errors.

It is well known that, in the presence of large negative moving average roots, unit root tests have severe size distortions. The last two rows of Table 2 show that the probability of rejecting the true null of no cointegration when it is true can be above 50%.

---

\(^5\) When \(\theta_{11} = 0\) and \(\theta_{22} \neq 0\), the model simulated in this paper is equivalent to Haug’s (1993). While Haug (1993) estimates the model with no lags, I consider the case in which an information criteria chooses the number of lags in the estimation, a situation closer to what is usually done in empirical applications.
and up to 99% for $\hat{Z}_t$ and $\hat{Z}_a$. Although the size distortions are extremely severe for all the tests in this case, Table 2 suggests that $\hat{t}_a$ and MSB still are to be preferred to other tests. These results emphasize the well-known trade-off between size and power in unit root tests; although the $\hat{t}_a$ test has power lower than the other tests, its size distortions are less evident across different specifications for the error terms.

6. CONCLUSIONS

Over the past years, residual tests for cointegration have been widely used in testing for the null of no cointegration. Until now, evaluation of the tests in term of their power of size distortions has been mainly done with Monte Carlo simulations. However, it is hard to design the correct experiment without the knowledge of which parameters enter the asymptotic distribution of the tests under the alternative. This paper computes the analytical local power of five unit root tests applied to the residuals of a cointegration regression. I show that there is no particular gain in using different unit root tests from the ADF t-test originally suggested by Engle and Granger (1987). The difference in power between the tests is less significant than in the unit root tests. The paper also shows that the power depends on a single nuisance parameter, which is, determined by the correlation at frequency zero of the independent variables with the errors of the cointegration regression. Simulations show that for large values of the nuisance parameter, the power of residuals based tests is very low, leaving room for improvement and future work.
REFERENCES:


APPENDIX

LEMMA A1: When the model is generated according to (2.1) with $T \rho - 1 = c$, then, as $T \to \infty$:

\begin{enumerate}
\item $\Omega_{11}^{-1/2} T^{-2} \Sigma x_{i,t}^d x_t^d \Omega_{11}^{-1/2} \Rightarrow \int W_t^d W_t^d$,
\item $\Omega_{11}^{-1/2} \omega_{2,1}^{-1/2} T^{-2} \Sigma x_t^d u_t^d \Rightarrow \int W_t^d J_{12c}^d$,
\item $\omega_{2,1}^{-1/2} T^{-2} \Sigma u_t^d \Rightarrow \int J_{12c}^d$,
\item $\omega_{2,1}^{-1/2} T^{-2} \Sigma u_t^d \Rightarrow \int J_{12c}^d dW$,
\item $\Omega_{11}^{-1/2} T^{-2} \Sigma x_{i,t}^d \Rightarrow \int W_t^d dW$
\end{enumerate}

where $i = 1, 2$, the summation goes from 1 to $T$ and $\Rightarrow$ denotes weak convergence, \[ \begin{bmatrix} \cdot \end{bmatrix} \] denotes the $ii$ element of a matrix. $W = \begin{bmatrix} W_1' & W_2' \end{bmatrix}$ is a $n \times 1$ vector of standard independent Brownian motions partitioned conformably to $v_{1t}$ and $v_{2t}$. $J_{12c}$ is a scaled Ornstein Uhlenbeck process such that: $J_{12c} (\lambda) = W_{12} (\lambda) + c \int_0^\lambda e^{(\lambda-s)c} W_{12} (s) ds$ with

\[ W_{12} = \sqrt{\frac{R^2}{1 - R^2}} W_1 + W_2 \]

\begin{enumerate}
\item if $z_{1t} = 0$ and $z_{2t} = 0$, $x_t^d = x_t$ and $W^d = W$, $J^d = J$
\item if $z_{1t} = 0$ and $z_{2t} = 1$, $x_t^d = x_t - \bar{x}$ and $W^d = W - \int W$, $J^d = J - \int J$
\item if $z_{1t} = 1$ and $z_{2t} = 1$, or $z_{1t} = 1$, $z_{2t} = (1, t)$, $x_t^d$ is $x_t$ detrended by OLS and $W^d = W - (4 - 6\lambda) \int W - (12\lambda - 6) \int s W$
\end{enumerate}

\[ \lambda \] follow the usual convention and suppress the $\lambda$ from the Brownian motion terms. Unless specified otherwise, all the integrals are intended to be between 0 and 1.
\[ J^d = J - (4 - 6\lambda) \int J - (12\lambda - 6) \int sJ \]

**Proof of Lemma A1:**

By the multivariate FCLT and Phillips (1987) \( T^{-1/2} \sum_{t=1}^{\ell} v_t = \Omega_1^{1/2} T^{-1/2} \sum_{s=1}^{\delta} \eta_s \Rightarrow \Omega^{1/2} W \)

where, \( \omega_{2,1} = \omega_{2,2} - \omega_{2,1} \Omega^{-1}_{11}\omega_{1,2} \), \( W' = \begin{bmatrix} W_1' & W_2' \end{bmatrix} \). Using this notation I can say that

\[ T^{-1/2} \sum_{t=1}^{\ell} v_t \Rightarrow \omega_{2,1} \Omega^{-1/2}_{11} W_1 + \omega_{1,2} W_2. \]

Define \( \delta = \omega_{2,1}^{-1} \omega_{2,2} \Omega^{-1/2}_{11} \) so that \( \delta' \delta = \frac{R^2}{1 - R^2} \);

then \( T^{-1/2} \sum_{t=1}^{\ell} \delta' \eta_s \Rightarrow \delta' W_1 = \sqrt{\frac{R^2}{1 - R^2} W_1} \) where \( W_1 \) is a univariate standard Brownian Motion independent of \( W_2 \). From Phillips (1987) and the Multivariate FCLT we have that \( \omega_{2,1}^{-1/2} T^{-1/2} u_{[r]} \Rightarrow \sqrt{\frac{R^2}{1 - R^2} W_1} + W_2 = J_{12c} \). Results (1)-(3) follows from the Continuous Mapping Theorem (CMT) while (4) and (5) follows directly from Chan and Wei (1988) or Phillips(1987).

**COROLLARY A2:** *When the model is generated according to (2.1) then, as \( T \to \infty \),

\[ (A1) \quad (\hat{\beta}_{ols} - \beta) \Rightarrow \omega_{2,1}^{1/2} \Omega_{11}^{-1/2} \left( \int W_1^d W_1^d \right)^{-1} \left( \int W_1^d J_{12c}^d \right) \]

where \( \hat{\beta}_{ols} \) is the LS estimator in the cointegration regression (2.2), and \( W' = \begin{bmatrix} W_1' & W_2' \end{bmatrix} \)

is a \( n \times 1 \) vector of standard independent Brownian motions partitioned conformably to \( v_\mu \) and \( v_\nu \), \( \omega_{2,1} = \omega_{2,2} - \omega_{2,1} \Omega^{-1}_{11}\omega_{1,2} \), \( \delta = \Omega_{11}^{-1/2} \omega_{1,2} \omega_{2,2}^{-1/2} \) and \( R^2 = \delta' \delta \). \( J_{12c} \) is a scaled
Ornstein Uhlenbeck process such that: \( J_{12c} (\lambda) = W_{12} (\lambda) + c \int_0^\lambda e^{(\lambda-r)s} W_{12} (s) \, ds \) with

\[
W_{12} = \sqrt{\frac{R^2}{1 - R^2}} W_1 + W_2
\]

(i) \( W^d = W \) and \( J_{12c}^d = J_{12c} \) if \( z_{1t} = 0 \) and \( z_{2t} = 0 \).

(ii) \( W^d = W - \int W \) and \( J_{12c}^d = \int_0^\lambda J_{12c} \) if \( z_{1t} = 0, \ z_{2t} = 1 \) and a mean is included in the regression.

(iii) \( W^d = W - (4 - 6\lambda) \int W - (12\lambda - 6) \int sW \) and

\[
J_{12c}^d = J_{12c} - (4 - 6\lambda) \int J_{12c} - (12\lambda - 6) \int sJ_{12c}
\]

if \( z_{1t} = 1, \ z_{2t} = 1 \) or \( z_{1t} = 1, \ z_{2t} = (1,1) \) and mean and trend are included in the regressions.

**Proof of COROLLARY A2:**

\( \hat{\beta} \) can be estimated by regressing \( y^d_i \) on \( x^d_i \). If \( z_{2t} = 0 \) then \( x^d_i = x_i \) and \( y^d_i = y_i \), for case (ii) both variable are demeaned and for case (iii) both variables are demeaned and detrended. In the general case \( (\hat{\beta} - \beta)' \left( T^{-2} \Sigma x^d_i x^d_i \right)^{-1} \left( T^{-2} \Sigma x^d_i u^d_i \right) \). From LEMMA A1 and the CMT we have that \( (\hat{\beta} - \beta)' \Rightarrow \omega_{21}^{1/2} \Omega_{11}^{-1/2} \left( \int W^d_i W^d_i \right)^{-1} \left( \int W^d_i J_{12c}^d \right) \).

**COROLLARY A3:** When the model is generated according to (2.1) with \( T(p-1) = c \), then, as \( T \to \infty \):

(i) \( \omega_{21}^{-1} T^{-2} \Sigma u^d_i u^d_i \Rightarrow \eta^d_i A^4 \eta^d_i \)
(ii) $\omega_{2,1}^{-1} T^{-1} \Delta \hat{u}_{t-1} d \Rightarrow cd(1) \eta_{c}^{d} A_{c}^{d} \eta_{c}^{d} + d(1) \eta_{c}^{d} \int W_{c}^{d} d \tilde{W} \eta_{c}^{d}$

(iii) $s_{\xi}^{2} \Rightarrow d(1)^{2} \omega_{2,1} \eta_{c}^{d} D \eta_{c}^{d}$

(iv) $\omega_{AR} \Rightarrow \omega_{2,1} \eta_{c}^{d} D \eta_{c}^{d}$

where $\hat{u}_{t}$ are the residuals from the cointegration regression (2.2) and $s_{\xi}^{2} = T^{-1} \Sigma \hat{\xi}_{t}^{2}$ is the estimated variance of the residuals of the ADF regression.

\[
\eta_{c}^{d} = \left[ -\left( \int W_{1,d}^{d} J_{12c}^{d} \right)^{-1} \int W_{1,d}^{d} W_{1,c}^{d} \right]^{1}, \quad A_{c}^{d} = \int W_{1,c}^{d} W_{1,c}^{d} \left[ \int W_{1,d}^{d} W_{1,d}^{d} \right],
\]

\[
W_{c}^{d} = \left[ W_{1,d}^{d} J_{12c}^{d} \right], \quad \tilde{W} = \left[ \frac{W_{1}}{\sqrt{1-R^{2}}} \right] \quad \text{and} \quad D = \left[ I_{n} \quad \frac{\delta}{\delta} \right].
\]

Proof of Corollary A3:

(i) By OLS projections $\hat{u}_{t} = \hat{u}_{t}^{d} = u_{t}^{d} - (\hat{\beta} - \beta) x_{t}$. By Lemma A1 and CMT

\[
\omega_{2,1}^{-1/2} A_{t-1}^{d} T^{-1/2} \Rightarrow \eta_{c}^{d} W_{c}^{d} \quad \text{and} \quad \omega_{2,1}^{-1} T^{-3/2} \Sigma \hat{\xi}_{t}^{2} \Rightarrow \eta_{c}^{d} \int W_{c}^{d} W_{c}^{d} \eta_{c}^{d} = \eta_{c}^{d} A_{c}^{d} \eta_{c}^{d}
\]

(ii) This proof follows from the same exact argument of Phillips and Ouliaris (1990) with the only difference is that now there is the extra piece $(\rho - 1) u_{t-1}$. In fact it can be shown that

\[
\Delta \hat{u}_{t}^{d} = (\rho - 1) \hat{u}_{t-1}^{d} + v_{2,t} - (\hat{\beta} - \beta) v_{1,t} - \hat{\gamma}_{1} + (\hat{\beta} - \beta) (\hat{\mu}_{1} - \mu_{1}) + (\hat{\beta} - \beta) (\rho - 1) x_{t-1},
\]

where $\hat{\gamma}_{1}$ and $\hat{\mu}_{1}$ are the OLS estimates from regressing $u_{t}$ and $x_{t}$ on a mean and trend (or mean only for $z_{t}$) with $\hat{\gamma}_{1}$ converging at rate $\sqrt{T}$ and $\Delta x_{t}^{d} = \Delta x_{t} - \hat{\mu}_{1}$ if detrended.
Write \( v_{2t} - (\hat{\beta} - \beta) v_{t} = \hat{b}' v_{t}. \) \( b' v_{t} \Rightarrow \eta_{c}^d \Gamma_{v_{t}} = \zeta, \) where \( \Gamma = \begin{bmatrix} \Omega_{11}^{-1/2} \omega_{21}^{1/2} & 0 \\ 0 & 1 \end{bmatrix}. \) Following Phillips and Ouliaris (1990) p. 183, I can write \( w_{t} = d(L) \zeta \), that is an absolute summable sequence. The variance of the orthogonal sequence \( w_{t} \) given \( \eta_{c}^d \) can then be written as \( d(1)^{2} \eta_{c}^d \Gamma \Omega \Gamma \eta_{c}^d. \) The ADF procedure requires the lag order in the augmented residuals regression to be large enough to capture the correlation structure of the errors. As Said and Dickey (1984) originally showed, in the context of unit root tests it is required that \( p = o(T^{1/3}) \). Similar argument can be applied in the local to unity case (Xiao and Phillips (1998)). If Assumption A1 is valid, it can then be shown that, conditionally on \( \eta_{c}^d \),

\[
\Delta \hat{\mu}_{t} = d(1)(\rho - 1) \hat{\mu}_{t} - d(1) \hat{\gamma}_{1} \Rightarrow O_p(1), \quad d(1) T^{-1} \Sigma \hat{\mu}_{t-1} \hat{\gamma}_{1} = d(1) T^{-3/2} \Sigma \hat{\mu}_{t-1} (T^{1/2} \hat{\gamma}_{1} ) \Rightarrow 0 \quad \text{and} \quad \omega_{21}^{-1/2} T^{-3/2} \hat{\mu}_{t} \Rightarrow \int J_{12c}(d(1) d_{11} d_{12c} \int W_{11} d_{12c} )^{-1/2} W_{11} d_{12c} = 0 \quad \text{since} \quad \int J_{12c} = 0 \quad \text{and} \quad \int W_{11} d_{12c} = 0.
\]

Similarly \( d(1) \left( \hat{\beta} - \beta \right) T^{-1} \Sigma \hat{u}_{t-1} (\hat{\mu}_{1} - \mu_{1}) \Rightarrow 0 \)

And \( \Omega_{11}^{-1/2} \omega_{21}^{-1/2} T^{-2} \Sigma \hat{u}_{t-1} x_{t-1} \Rightarrow 0 \) so \( \left( \hat{\beta} - \beta \right) (\rho - 1) \Omega_{11}^{-1/2} \omega_{21}^{-1/2} T^{-1} \Sigma \hat{u}_{t-1} x_{t-1} \Rightarrow 0 \)

Using the above results it easy to see that:

\[
\omega_{21}^{-1/2} T^{-1} \Sigma \hat{u}_{t-1} \Delta \hat{u}_{t} = c d(1) \omega_{21}^{-1/2} T^{-2} \Sigma \hat{u}_{t-1}^{2} + \omega_{21}^{-1/2} T^{-1} \Sigma \hat{u}_{t-1} w_{t} + o_p(1). \quad \text{The first piece converges to} \quad c d(1) \eta_{c}^{d} A^d_{\eta_{c}^d} \text{ by (i). For the second piece, notice that the variance of } w_{t} \text{ given } \eta_{c}^{d} \text{ can be written as} \quad d(1)^{2} \eta_{c}^{d} \Gamma \Omega \Gamma \eta_{c}^{d} \quad \text{where } \Gamma \Omega \Gamma \text{is the variance covariance matrix of} \quad \omega_{21}^{1/2} \begin{bmatrix} W_{1} \\ \delta W_{1} + W_{2} \end{bmatrix} \quad \text{so} \quad \omega_{21}^{-1/2} T^{-1} \Sigma \hat{u}_{t-1} w_{t} \Rightarrow d(1)^{2} \eta_{c}^{d} \int W_{t} d \bar{W} \eta_{c}^{d} \quad \text{where} \quad \nu
\[ \bar{W}' = \left[ W_1' \delta W_1 + W_2' \right] = \left[ W_1' \sqrt{\frac{R^2}{1 - R^2} \bar{W} + W_2} \right]. \]

(iii) Under Assumption A1, the estimates of the ADF regression will be consistent. 
\( (\hat{\alpha} \to \rho - 1). \)

\[ s_i^2 = T^{-1} \Sigma_{Z_i}^2 = T^{-1} \Sigma_{W_i}^2 + o_p(1) \Rightarrow d(1)\hat{\eta}_c^d \Gamma \Omega \Gamma \hat{\eta}_c^d = d(1)\omega_{2,1} \hat{\eta}_c^d \left[ \frac{I}{\delta'} + \frac{\delta}{1 + \delta \delta} \right] \hat{\eta}_c^d. \]

(iv) Following the argument in (ii) and (iii) it is easy to show that \( \hat{k}_i = \Delta \hat{u}_i - (\hat{\rho} - 1) \hat{u}_{i-1}. \)

If assumption A1 is valid, the autoregressive estimate of the spectral density of \( \hat{k}_i \) is a consistent estimate of the spectral density at frequency zero of \( \hat{v}_i \) which, as shown in (ii), converges to \( \hat{\eta}_c^d \Gamma \Omega \Gamma \hat{\eta}_c^d = \omega_{2,1} \hat{\eta}_c^d D \hat{\eta}_c^d. \)

**Proof of Theorem 1:**

(1i) The ADF test statistic is the usual \( t \) ratio test in the augmented regression of the residuals:

\[ t_\alpha = \frac{T^{-1} \hat{u}_{-1}' M_u \Delta \hat{u}_i}{s_i \left[ T^{-2} \hat{u}_{-1}' M_u \Delta \hat{u}_i \right]^{1/2}} \]

Under the condition for Corollary A1, A2, and Assumption A1,

\[ t_\alpha = \frac{T^{-1} \Sigma_{\hat{u}_{-1}} \Delta \hat{u}_i}{\omega_{2,1} \left[ T^{-2} \Sigma_{\hat{u}_{-1}} \right]^{1/2} + o_p(1)} \]

By Corollary A3, \( t_\alpha \Rightarrow c \frac{d(1)\hat{\eta}_c^d A^d \hat{\eta}_c^d \gamma^{1/2}}{d(1)\hat{\eta}_c^d D \hat{\eta}_c^d} + \frac{d(1)\hat{\eta}_c^d \int W_i^d d\bar{W} \hat{\eta}_c^d}{d(1)\hat{\eta}_c^d D \hat{\eta}_c^d} + \frac{d(1)\hat{\eta}_c^d \int W_i^d d\bar{W} \hat{\eta}_c^d}{d(1)\hat{\eta}_c^d D \hat{\eta}_c^d}. \]
(1ii) Recall that $\hat{k}_t = \Delta \hat{u}_t - (\hat{\rho} - 1) \hat{u}_{t-1}$. As proved in Corollary A3, part (ii)

$$\Delta \hat{u}_{t}^d = \Delta \hat{u}_t = (\rho - 1) \hat{u}_{t-1} + v_{2t} - (\hat{\beta} - \beta) \hat{v}_{1t} - \hat{\gamma}_1 + (\hat{\beta} - \beta) \hat{\delta}_1 + (\hat{\beta} - \beta)(\rho - 1)x_{t-1}.$$ 

Under assumption A2

$$\frac{1}{2}(s_t^2 - s_x^2) = T^{-1} \sum_{t=1}^{T} w_s T \sum_{t=1}^{T} \hat{k}_t \Delta \hat{u}_{t-1} = T^{-1} \sum_{t=1}^{T} w_s T \sum_{t=1}^{T} \hat{b}' \hat{v}_{t-1} \hat{b} + o_p(1).$$

$$T^{-1} \sum_{t=1}^{T} w_s T \sum_{t=1}^{T} \hat{b}' \hat{v}_{t-1} \hat{b} \Rightarrow \omega_2 \eta_c^d \Gamma \Gamma_i \Gamma' \eta_c^d = \omega_2 \Gamma \Gamma \Gamma \eta_c^d = \eta_c^d \Omega \Gamma \eta_c^d = \eta_c^d D \eta_c^d$$ which is the spectral density at frequency zero of $\hat{b}' \hat{v}_t$ and

$$T^{-1} \sum_{t=1}^{T} \hat{u}_{t-1} \Delta \hat{u}_t = c T^{-1} \sum_{t=1}^{T} \hat{u}_{t-1} \hat{u}_{t-1} + \hat{b}' T^{-1} \sum_{t=1}^{T} \hat{u}_{t-1} \hat{v}_t + o_p(1).$$ (See also proof of Corollary A3, part (ii)). Because $v_t$ is a dependent process, following Park and Phillips (1988) we have that

$$\omega_2^{-1} T^{-1} \sum_{t=1}^{T} \hat{u}_{t-1} \Delta \hat{u}_t \Rightarrow c \eta_c^d A \eta_c^d + \eta_c^d \int W_c d \hat{W}_c + \eta_c^d \Gamma \Gamma_i \eta_c^d.$$ 

Noting that

$$T(\hat{\rho} - 1) = \left( T^{-2} \sum_{t=1}^{T} \hat{u}_{t-1}^d \right)^{-1} \left( T^{-1} \sum_{t=1}^{T} \hat{u}_{t-1} \Delta \hat{u}_t \right)$$ 
and substituting each piece in the expression for $\hat{Z}_r$, I obtain that

$$\hat{Z}_r = \frac{c \eta_c^d A \eta_c^d + \eta_c^d \int W_c d \hat{W}_c + \eta_c^d \Gamma \Gamma_i \eta_c^d - \eta_c^d \Gamma \Gamma_i \eta_c^d}{\left[ \eta_c^d A \eta_c^d \right]^{1/2} \left[ \eta_c^d D \eta_c^d \right]^{1/2}}$$

which simplifies to the expression in Theorem 1.

(2i) We can write $\hat{\alpha}_T = \frac{T(\hat{\rho} - 1)\hat{\omega}_{AR}}{s_T^2}$. 

vii
\[
T(\hat{\rho} - 1) = \frac{T^{-1} \sum \hat{u}_{t-1} \Delta \hat{u}_t + o_p(1)}{T^{-2} \sum \hat{u}_{t-1}^2 + o_p(1)} \Rightarrow \frac{cd(1)\eta^d_c A^d \eta^d_c + d(1)\eta^d_c \int W^d_c d\hat{W} \eta^d_c}{\eta^d_c A^d \eta^d_c} \quad \text{from Corollary A3.}
\]

Substituting in the limits for the variances we have

\[
\hat{\alpha}_T \Rightarrow c \frac{d(1)\eta^d_c A^d \eta^d_c}{d(1)\left[\eta^d_c A^d \eta^d_c\right]} + \frac{d(1)\eta^d_c \int W^d_c d\hat{W} \eta^d_c}{d(1)\left[\eta^d_c A^d \eta^d_c\right]} = c + \frac{\eta^d_c \int W^d_c d\hat{W} \eta^d_c}{\eta^d_c A^d \eta^d_c}
\]

(2ii) The proof for \( \hat{Z}_\alpha \) follows from (1ii):

\[
\hat{Z}_\alpha \Rightarrow \frac{c \eta^d_c A^d \eta^d_c + \eta^d_c \int W^d_c d\hat{W} \eta^d_c + \eta^d_c \Gamma \eta^d_c - \eta^d_c \Gamma \eta^d_c}{\eta^d_c A^d \eta^d_c} = c + \frac{\eta^d_c \int W^d_c d\hat{W} \eta^d_c}{\eta^d_c A^d \eta^d_c}
\]

(3) \( MSB = \frac{\left(T^{-2} \sum \hat{u}_t^2\right)^{1/2}}{\hat{\omega}_{AR}} \). Substituting all the terms from Corollary A3:

\[
MSB \Rightarrow \frac{\omega_{T1}^{1/2} \left[\eta^d_c A^d \eta^d_c\right]^{1/2}}{\omega_{T2}^{1/2} \left[\eta^d_c D \eta^d_c\right]^{1/2}} = \left[\eta^d_c A^d \eta^d_c\right]^{1/2} / \left[\eta^d_c D \eta^d_c\right]^{1/2}
\]
Figure 1: Large Sample Power, no constant case.
Figure 2: Large Sample Power, Demeaned

**Large Sample Power for Rsquare=0, Demeaned**

**Large Sample Power for Rsquare=0.5, Demeaned**

**Large Sample Power for Rsquare=0.7, Demeaned**
Figure 3: Large Sample Power, Demeaned and Detrended.
**Figure 4**: Asymptotic Power for ADF t-test for different values of Rsquare.
<table>
<thead>
<tr>
<th>( -c )</th>
<th>( R^2 )</th>
<th>( \rho )</th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF</td>
<td>0.050</td>
<td>0.141</td>
<td>0.386</td>
<td>0.695</td>
<td>0.912</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\alpha}_r )</td>
<td>0.050</td>
<td>0.181</td>
<td>0.485</td>
<td>0.806</td>
<td>0.964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSB</td>
<td>0.050</td>
<td>0.182</td>
<td>0.501</td>
<td>0.822</td>
<td>0.971</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>ADF</td>
<td>0.050</td>
<td>0.128</td>
<td>0.348</td>
<td>0.667</td>
<td>0.904</td>
<td></td>
</tr>
<tr>
<td>( \hat{\alpha}_r )</td>
<td>0.050</td>
<td>0.167</td>
<td>0.444</td>
<td>0.778</td>
<td>0.954</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSB</td>
<td>0.050</td>
<td>0.166</td>
<td>0.458</td>
<td>0.794</td>
<td>0.961</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>ADF</td>
<td>0.050</td>
<td>0.113</td>
<td>0.305</td>
<td>0.623</td>
<td>0.881</td>
<td></td>
</tr>
<tr>
<td>( \hat{\alpha}_r )</td>
<td>0.050</td>
<td>0.143</td>
<td>0.396</td>
<td>0.738</td>
<td>0.939</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSB</td>
<td>0.050</td>
<td>0.140</td>
<td>0.406</td>
<td>0.749</td>
<td>0.950</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>ADF</td>
<td>0.050</td>
<td>0.098</td>
<td>0.227</td>
<td>0.454</td>
<td>0.702</td>
<td></td>
</tr>
<tr>
<td>( \hat{\alpha}_r )</td>
<td>0.050</td>
<td>0.119</td>
<td>0.287</td>
<td>0.546</td>
<td>0.798</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSB</td>
<td>0.050</td>
<td>0.129</td>
<td>0.307</td>
<td>0.584</td>
<td>0.834</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>ADF</td>
<td>0.050</td>
<td>0.054</td>
<td>0.109</td>
<td>0.251</td>
<td>0.505</td>
<td></td>
</tr>
<tr>
<td>( \hat{\alpha}_r )</td>
<td>0.050</td>
<td>0.056</td>
<td>0.140</td>
<td>0.321</td>
<td>0.611</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSB</td>
<td>0.050</td>
<td>0.062</td>
<td>0.154</td>
<td>0.355</td>
<td>0.645</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Case (iii)

<table>
<thead>
<tr>
<th>( -c )</th>
<th>( R^2 )</th>
<th>( \rho )</th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF</td>
<td>0.050</td>
<td>0.072</td>
<td>0.155</td>
<td>0.317</td>
<td>0.535</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\alpha}_r )</td>
<td>0.050</td>
<td>0.078</td>
<td>0.179</td>
<td>0.362</td>
<td>0.595</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSB</td>
<td>0.050</td>
<td>0.081</td>
<td>0.191</td>
<td>0.379</td>
<td>0.618</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>ADF</td>
<td>0.050</td>
<td>0.063</td>
<td>0.116</td>
<td>0.237</td>
<td>0.419</td>
<td></td>
</tr>
<tr>
<td>( \hat{\alpha}_r )</td>
<td>0.050</td>
<td>0.065</td>
<td>0.131</td>
<td>0.271</td>
<td>0.472</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSB</td>
<td>0.050</td>
<td>0.066</td>
<td>0.138</td>
<td>0.285</td>
<td>0.496</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>ADF</td>
<td>0.050</td>
<td>0.038</td>
<td>0.051</td>
<td>0.106</td>
<td>0.240</td>
<td></td>
</tr>
<tr>
<td>( \hat{\alpha}_r )</td>
<td>0.050</td>
<td>0.034</td>
<td>0.051</td>
<td>0.118</td>
<td>0.276</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSB</td>
<td>0.050</td>
<td>0.036</td>
<td>0.055</td>
<td>0.128</td>
<td>0.294</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The power is computed with \( T=100 \) and 5000 replications.
<table>
<thead>
<tr>
<th>AR errors</th>
<th>MA errors</th>
<th>Lags</th>
<th>ADF</th>
<th>α_τ</th>
<th>MSB</th>
<th>Ũ_τ</th>
<th>Ũ_α</th>
</tr>
</thead>
<tbody>
<tr>
<td>ϕ_{11} = ϕ_{22} = 0</td>
<td>θ_{11} = θ_{22} = 0</td>
<td>1.1</td>
<td>0.059</td>
<td>0.064</td>
<td>0.097</td>
<td>0.060</td>
<td>0.040</td>
</tr>
<tr>
<td>ϕ_{11} = ϕ_{22} = 0.8</td>
<td>θ_{11} = θ_{22} = 0</td>
<td>1.1</td>
<td>0.061</td>
<td>0.076</td>
<td>0.174</td>
<td>0.026</td>
<td>0.000</td>
</tr>
<tr>
<td>ϕ_{11} = ϕ_{22} = 0.2</td>
<td>θ_{11} = θ_{22} = 0</td>
<td>1.5</td>
<td>0.052</td>
<td>0.063</td>
<td>0.076</td>
<td>0.170</td>
<td>0.129</td>
</tr>
<tr>
<td>ϕ_{11} = ϕ_{22} = 0.5</td>
<td>θ_{11} = θ_{22} = 0</td>
<td>1.5</td>
<td>0.084</td>
<td>0.109</td>
<td>0.174</td>
<td>0.023</td>
<td>0.009</td>
</tr>
<tr>
<td>ϕ_{11} = ϕ_{22} = 0</td>
<td>θ_{11} = θ_{22} = 0.5</td>
<td>2.7</td>
<td>0.069</td>
<td>0.129</td>
<td>0.203</td>
<td>0.022</td>
<td>0.010</td>
</tr>
<tr>
<td>ϕ_{11} = ϕ_{22} = 0</td>
<td>θ_{11} = θ_{22} = 0.2</td>
<td>1.1</td>
<td>0.090</td>
<td>0.109</td>
<td>0.130</td>
<td>0.203</td>
<td>0.170</td>
</tr>
<tr>
<td>ϕ_{11} = ϕ_{22} = 0</td>
<td>θ_{11} = θ_{22} = 0.5</td>
<td>1.6</td>
<td>0.673</td>
<td>0.727</td>
<td>0.650</td>
<td>0.992</td>
<td>0.987</td>
</tr>
<tr>
<td>ϕ_{11} = ϕ_{22} = 0.5</td>
<td>θ_{11} = θ_{22} = -0.8</td>
<td>1.2</td>
<td>0.428</td>
<td>0.491</td>
<td>0.537</td>
<td>0.651</td>
<td>0.625</td>
</tr>
</tbody>
</table>

**Note:** The size distortions are computed with T=100 and 5000 replications. Lags in each regression are chosen using BIC with a maximum of 4 lags: the number in parenthesis represents the average number of lags chosen by BIC.