Title

Analytical Evaluation of the Power of Tests for the Absence of Cointegration

Author

PESAVENTO, ELENA

Publication Date

2000-09-01

2000-24

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

ANALYTICAL EVALUATION OF THE POWER OF TESTS FOR THE ABSENCE OF COINTEGRATION

BY

ELENA PESAVENTO

DISCUSSION PAPER 2000-24
SEPTEMBER 2000

ANALYTICAL EVALUATION OF THE POWER OF TESTS FOR THE ABSENCE OF COINTEGRATION.

Elena Pesavento
Department of Economics
University of California, San Diego

This Draft September 2000.

Abstract

This paper proposes a theoretical explanation to the common empirical results in which different tests for cointegration give different answers. Using local to unity parametrization I compute the analytical power of some tests for the null of no cointegration: The ADF test on the residuals of the cointegration regression, Johansen's maximum eigenvalue test, the t-test on the Error Correction term and Boswijk (1994) Wald test. The tests are shown to be functions of Brownian Motions and Ornstein-Uhlenbeck processes and to depend on a single nuisance parameter, which is, in turn determined by the correlation at frequency zero of the independent variables with the errors of the cointegration regression. Monte Carlo experiments show that the tests can have significantly different performances for different values of the nuisance parameter. An application to the money demand equation is presented.

JEL classification: C32

Keywords: Unit root, cointegration, local alternative.

Corresponding author: Elena Pesavento, epesaven@emory.edu.

1. INTRODUCTION

Since its formal introduction by Granger (1983) and Engle and Granger (1987), the concept of cointegration has been widely used in empirical analysis to study the relationship between integrated variables. If a group of variables are individually integrated of order one and there exists at least one linear combination of these variables that is stationary, then the variables are said to be cointegrated. Cointegrated variables will never move too far apart, and will revert to their long-run relationship. For this reason the knowledge that some variables are cointegrated can have a significant impact on the analysis of the long and short run dynamics of economic variables. As usual, testing the assumptions of the model (testing for cointegration) has become an important step in any empirical analysis of economic data.

The current literature is prolific in all sort of different tests for cointegration: tests for the presence of cointegration (Park (1990, 1992), Phillips and Hansen (1990)), tests for the absence of cointegration (Engle and Granger (1987), Phillips and Ouliaris (1990), tests on the Error Correction Model introduced by Hendry (1987), Boswijk (1994) Wald test), test for the number of unit roots (Johansen (1988), Stock and Watson (1998), Saikonnen (1992)), and tests on the null hypothesis on the cointegrating vector (Saikonnen (1992), Johansen (1995) and Elliott (1998)) among others (see Watson(1994) for a review). Currently there is no consensus as to the best test for cointegration and the general empirical approach is to report the results for a variety of tests.

This paper looks at the class of tests that have no cointegration as the null hypothesis. There are two types of tests proposed in the literature to test for the absence of cointegration. One group of tests looks at the full system of equations in a VAR framework (Johansen's tests, Stock and Watson (1988) SW test, tests on the coefficient of the error correction terms among others) while a second group looks at single equation regressions involving the variables that are potentially cointegrated (Engle and Granger Augmented Dickey Fuller test, Phillips and Ouliaris

$\hat{Z}_a$ and $\hat{Z}_t$ tests). Such tests are non-standard and their asymptotic distributions are non-normal (functions of Brownian motions). In this non-standard environment, no uniformly most powerful (UMP) test exists so there is no theoretical "best" test. Until now, tests have been compared on the basis of Monte Carlo analysis for particular Data Generating Process (Haug (1996), Gonzalo and Lee (1998), Bewley and Yang (1998), Boswijk and Frances (1992), Kremers et al. (1992) and Ericsson and Mackinnon (1999) among others). The problem with Monte Carlo simulations is that the results of the experiments are dependent on the particular design run and no general conclusions are available.

Haug (1996), for example, used Monte Carlo analysis to study the small sample power of some of the most common tests for cointegration. In particular, he examined the case of a fixed alternative in which the root of the residuals of the cointegration regression is equal to 0.85. He found that, in general, single equation tests have smaller size distortions, but also have lower power than system-based tests. Although Haug (1996) shows that for larger samples the power increases considerably, differences in the performance of the two sets of tests persist. The paper is unable to find any ranking of the tests or to find important parameters that would allow such a ranking, and so concludes by recommending the application of both sets of tests in empirical exercises.

In this paper, I derive analytical results for power of the tests that unify these experiments and I show which features of the model are important for power. Using local-to-unity parametrization, I am able to analytically compute the power for some of the most commonly used tests for the null of no cointegration: the Augmented Dickey Fuller test (ADF) applied to the residuals from the cointegrating regression and Johansen's maximum eigenvalue test. Analytical power is also computed for 3 tests in the Error Correction Model: The t-test on the error correction term (that is unfeasible if the cointegration vector is not known), a feasible version of the t-test obtained by adding a redundant regressor, and the Wald test proposed by

Boswijk (1994). Once we have a full understanding of which parameters are important for power, the results of previous Monte Carlo analysis can be easily understood and a more informative set of Monte Carlo can be designed.

The tests are shown analytically to depend on a single nuisance parameter under the alternative. This parameter is a function of the long run correlation of shocks to the independent variable with the errors of the cointegration regression. As intuition suggests, when this correlation is high, we would expect a full system approach to perform better through exploiting this correlation. Evaluations of the analytical power functions confirm the intuition and show that the tests have significantly different performances for different values of the nuisance parameter. These results suggest which test could be better used for the particular application at hand, depending on which level of simultaneity is suspected.

The next two sections introduce the model and the tests for the absence of cointegration. The analytical asymptotic power functions are analyzed in section 3 while section 4 presents the results for the Monte Carlo simulations. Section 5 contains an empirical application to the money demand equation. All the proofs are in the Appendix.

2. THE MODEL

Consider the model:

$$\begin{aligned} \Delta x_t &= d_{1t} + v_{1t} \\ y_t &= d_{2t} + x_t^{'} b + u_t \qquad (1.1) \\ u_t &= r\, u_{t-1} + v_{2t} \end{aligned}$$

where t=1,….T ; x_t is a $n_1 \times 1$ vector; y_t is a scalar, $d_{1t} = G_1 z_{1t}$ and $d_{2t} = G_2 z_{2t}$. $v_t = \left[v_{1t}^{'} \quad v_{2t}\right]^{'}$, $\Phi(L) v_t = e_t$, $e_t = [e_{1t}^{'}\, e_{2t}]^{'}$ is a $n \times 1$ vector of martingale differences with positive definite variance covariance Σ such that the partial sum $\frac{1}{\sqrt{T}}\sum_{1}^{t} v_t$ satisfies a multivariate invariance principle[1]. $\Phi(L)$ is an invertible lag polynomial of known order partitioned conformably to v_t such that $\Phi(L) = \begin{bmatrix} f_{11}(L) & f_{12}(L) \\ f_{21}(L) & f_{22}(L) \end{bmatrix}$. The spectral density of v_t at frequency zero (scaled by $2p$) is $\Omega = \Phi(1)^{-1}\Sigma\Phi(1)^{-1'}$ where $\Phi(1) = \sum_i \Phi_i$. Ω can be partitioned as:

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{11}^{1/2} w_2 D \\ D' w_2 \Omega_{11}^{1/2} & w_2^2 \end{bmatrix}$$

where $D' = \left[d_1 d_2 \ldots d_{n_1}\right]$ contains the bivariate zero frequency correlations of each element of v_{1t} with v_{2t}. Ω_{11} is assumed to be non-singular (so that the elements of x_t are not individually cointegrated with each other). For the purpose of this paper, the following cases for the deterministic part of the model are considered: (i) $z_{1t} = 0$ and $z_{2t} = 0$, (ii) $z_{1t} = 0$, $z_{2t} = 1$ and a constant is included in the regression, (iii) $z_{1t} = 1$, $z_{2t} = 1$ and a constant and a time trend are included in the regressions.

When $r < 1$, y_t and x_t are cointegrated and the system (1.1) contains n_1 unit roots; when $r = 1$ the two variables are not cointegrated and there are n unit roots in the system. Thus a tests for no cointegration is testing $H_0: r = 1$ vs $H_a: r < 1$.

2.1 TESTS ON THE RESIDUALS

As mentioned a variety of tests for no cointegration exist. This section briefly introduces the residual based tests for cointegration. To tests the hypothesis of no cointegration, Engle and

---

[1] This is valid for general case of weakly dependent heterogeneous variables. For conditions under which the invariance principle holds see Phillips and Solo (1992) for the univariate case. For the multivariate case see Chan and Wei (1988) or Wooldridge (1994).

Granger (1987) first suggested the application of unit root tests to the residuals of the cointegrating regression

(1.2) $$y_t = m + x_t^{'} b + u_t .$$

Rejection of a unit root in the residuals from (1.2) is an indication of cointegration between the two variables. A variety of tests for autoregressive unit roots are available (Stock (1994) offers an exhaustive survey on the argument). One of the suggestions of Engle and Granger (1987) was to use the t_a ratio test in the Augmented Dickey-Fuller regression:

(1.3) $$\Delta\hat{u}_t = a\hat{u}_{t-1} + \sum_1^p p_i \Delta\hat{u}_{t-i} + x_t$$

Where $\hat{u}_t$ are the residuals from the LS estimation of the cointegration regression (1.2) run respectively with no mean, mean only and mean and trend for case (i), (ii) and (iii)[2]. The t-test is not the only choice of test: Phillips and Ouliaris (1990) suggest also the use of Phillips (1987) Z_a and Z_t tests, and propose a variance ratio test, P_u, and a trace statistic, P_z. As in unit roots tests, the $T(\hat{r}-1)$ test and the Z_a test may be expected to perform better in small samples. Although it may be interesting to compare the difference in power between different unit root tests on the residuals, in this paper I will only look at the most commonly used ADF test. Extensions of results in this paper to other unit root tests are applications of the theorems presented here and can be found in Pesavento (2000).

2.2 TESTS ON THE ERROR CORRECTION TERM

The ADF test on the residuals presented in the previous section is based on OLS estimates of a single equation for y_t. As the Granger Representation theorem shows, a necessary and sufficient condition for cointegration is for the cointegrated series to be represented by an Error Correction model:

$$(1.4)\qquad \begin{bmatrix}\Delta x_t \\ \Delta y_t\end{bmatrix} = P\Phi(1)\overline{G} + P\Phi(1)\, M \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ z_{2t-1}\end{bmatrix} + \Pi(L)\begin{bmatrix}\Delta x_{t-1} \\ \Delta y_{t-1}\end{bmatrix} + Pe_t$$

where $\overline{G}$ is zero if $z_{1t} = 0$ and $z_{2t} = 0$ or $z_{1t} = 0$ and $z_{2t} = 1$. For the case in which $z_{1t} = 1$ and $z_{2t} = 1$ as Boswijk(1994) also mentions, a time trend need to be included in the regression to have similar tests under the null. In this case then $\overline{G}$ contains the parameters representing the drift in the variables. Finally, $P = \begin{bmatrix} 1 & 0 \\ b & 1 \end{bmatrix}$ and $M = \begin{bmatrix} 0 & 0 & 0 \\ -b(r-1) & (r-1) & -m(r-1) \end{bmatrix}$.

As the theorem suggests, testing whether the error correction term is significant can be used as a test for the null of no cointegration. In fact, under the null $r = 1$, $M = 0$ and the error correction term does not enter the model. Under the alternative $r < 1$ and the coefficient on the error correction term is different from zero.

If the right hand variable of the cointegration equation (1.2) is weakly exogenous for b under the null, as it is in the case of the triangular model (1.1), all information about the cointegration vector is contained in the conditional equation for y_t and the analysis can be limited to the single equation[3]:

$$(1.5)\qquad \Delta y_t = a\, u_{t-1} + j\, \Delta x_t + \sum_1^p p_{1i} \Delta x_{t-i} + \sum_1^p p_{2i} \Delta y_{t-i} + x_{2t}$$

---

[2] For the case (iii) in which $z_{1t} = 1$ and $z_{2t} = 1$, Hansen(1992) shows that, if the cointegrating regression is run only with a constant, the distribution of the test will be different. In this paper I only consider the more widely used case in which the regression is run with a trend if the variables have a drift.

[3] See also Assumption 1’ in Boswijk (1994).

where $u_{t-1} = y_{t-1} - x'_{t-1}b - G_2 z_{2t-1}$ is the Error Correction term and $a = (r-1)q$ where q is functions of $\Phi(1)$ and Ω[4]. If no bounded linear combination among the levels of the variables exists, one would expect the coefficient on the error correction term to be zero. This suggests that, if b is known, a t-test on the parameter a estimated from the single equation (1.5) could be used to test the null of no cointegration[5].

The assumption of known cointegrating vector is a very restrictive assumption. Unless the econometrician finds himself in the unlikely case of perfect knowledge of the cointegrating vector, the error correction test presented in the previous section is not feasible. Banerjee et al. (1986, 1993, 1998) suggest adding a redundant regressor to avoid imposing a particular cointegrating vector. The equation to be estimated is then

$$(1.6) \qquad \Delta y_t = m + a\,(y_{t-1} - x_{t-1}) + f\,x_{t-1} + j\,\Delta x_t + \sum_1^p p_{1i}\Delta x_{t-i} + \sum_1^p p_{2i}\Delta y_{t-i} + x_{2t}$$

where a is the same as before while $f = a\,(1-b)$ and $m = -a\,G_2 z_{2t}$. If $z_{1t} = 1$ and $z_{2t} = 1$, the regression (1.6) is estimated with mean and trend to obtain similar tests under the null. A test of the hypothesis $a = 0$, based on the t-statistic $t_{\hat{a}}$ in (1.6) is still a valid test for the absence of cointegration.

In Section 4.2 I derive the asymptotic distribution of the Error Correction (EC) test in (1.5) and the Error Correction with Redundant Regressor (ECR) test in (1.6) under the local alternative $T(r-1) = c$.

In equation (1.6) under the null of no cointegration in which $r = 1$ both coefficients on the variables in level are zero. A F test can then be used to test for the joint significance of the coefficients. This test is equivalent to the Wald test proposed by Boswijk (1994). As Boswijk (1994) shows, the asymptotic distribution of this test under the null is a function of standard

---

[4] Specifically $q = \left[f_{11}(1) f_{22}(1) - f_{21}(1) f_{12}(1)' \,/\, f_{11}(1) + f_{12}(1)\Omega_{21}\Omega_{11}^{-1} \right]$

Brownian Motions. In the next sections the power of the Wald test is analyzed and compared with the power of the ADF and ECR tests.

2.3 MAXIMUM EIGENVALUE TEST.

Johansen (1988,1991) and Johansen and Juselius (1990) suggest the use of a full information maximum likelihood[6] approach to determine the number of cointegrating vectors in a n-dimensional vector.

The Error Correction model (1.4) can be written as:

(1.7) $$\Delta Y_t = G + \Pi Y^*_{t-1} + \Pi(L)\Delta Y_{t-1} + w_t$$

with $Y_t = [x_t \quad y_t]'$ and $Y^*_t = [x_t \quad y_t \quad z_{2t}]'$. The mean G[7] is zero in the case in which $z_{1t} = 0$ and $z_{2t} = 0$ and in the case in which $z_{1t} = 0$, $z_{2t} = 1$[8]. The matrix $\Pi = P\Phi(1)P^{-1}M$ gives information about possible cointegrating vectors among the variables in Y^*_t. If Π has full rank n then all the variables are stationary. If $rank(\Pi) = h < n$, there exist two $n \times h$ matrices a and g of rank h such that $\Pi = ag'$. In this case, although each element of Y_t is integrated of order one, the linear combination $g'Y^*_t$ is stationary, and h cointegrating vectors exist.

To test the hypothesis that there are h cointegrating vectors against the alternative that there are n cointegrating vectors (that is, Y_t is stationary), Johansen proposes the likelihood ratio trace test, defined as:

---

[5] The invertibility of $\Phi(L)$ implies that all the roots of the polynomial are outside the unit circle so that $\Phi(1) \neq 0$. A t-test for the significance of a corresponds to testing $r = 1$.

[6] Of course if the right hand variables are not weakly exogenous a full system approach is necessary. The triangular form (3.1) precludes this possibility as, under the null, x_t is always weakly exogenous for b.

[7] $G = P\Phi(1)\overline{G}$

[8] The Johansen procedure is presented for the case in which the constant is imposed equal to zero under the null as in model (3.1).

(1.8) $\quad l_{tr} = -T \sum_{i=h+1}^{n} \ln(1-\hat{l}_i)$

where $\hat{l}_i$ are the eigenvalues of the matrix $S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$ listed in descending order. S_{11} and S_{00} are the variance-covariance matrices of the residuals of a regression of Y_{t-p}^* and ΔY_t respectively, on $\Delta Y_{t-1},\ldots,\Delta Y_{t-p+1}$. S_{10} and S_{01} are the covariance matrices between the residuals of the two equations.

A similar approach can be used to test the hypothesis of h cointegrating vectors against the alternative of $h+1$ cointegrating vectors. In this case, the likelihood ratio test becomes:

(1.9) $\quad l_{tr} = -T\ln(1-\hat{l}_{h+1})$

In particular, when the test (1.9) is used to test the hypothesis of no cointegrating vectors against the alternative of 1 cointegrating vector, the test is called the maximum eigenvalue test:

(1.10) $\quad l_{\max} = -T\ln(1-\hat{l}_1)$

Notice that the trace test (1.8) is testing for h cointegrating vectors, or that there are $n-h$ unit roots in Y_t^* against the hypothesis that there are no unit roots, so all the elements of Y_t^* are stationary. This is a different set of hypotheses compared to the one tested by the Engle and Granger procedure. In the ADF test (1.3), in fact, under the null hypothesis, $r=1$ and there are n unit roots in the system while under the alternative of $r<1$ Y_t contains $n-1$ unit roots. This is exactly the hypothesis tested by the maximum eigenvalue test (1.10). For the purpose of this paper, only the $l_{\max}$ test is directly comparable to the ADF and the Error Correction tests introduced in the previous sections.

3. ASYMPTOTIC POWER FUNCTIONS

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. Almost all 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 understand the previous studies we need to know how the nuisance parameters affect the power. One way of capturing this is to compute the analytical power of the tests. The knowledge of which nuisance parameter enters the local power will help us design the correct experiment and will suggest in which direction of the parameter space to look.

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. The usual approach is to examine a sequence of local alternatives (e.g. Van der Vault p.194). In this case the tests are all diverging at rate T so we need a sequence of alternatives that collapses on the null at rate T, leading to the use of the alternative $r = 1 + c/T$. When c is equal to zero the errors u_t are integrated of order one. For c negative, the variables in equation (1.1) are cointegrated. Using this parameterization and the results of Philips (1988) I evaluate the power of the tests for the absence of cointegration previously presented. This section is one of the main contributions of this paper in that, in my knowledge, the analytical power of the ECRR, the Wald and the residuals based tests for cointegration have never been computed before. For this version of the paper the asymptotic power functions are computed for the univariate case only.

3.1 ADF TEST

In the case of the ADF test, Phillips and Ouliaris (1990) show that, under the null, t_a has a non-standard distribution that is a function of standard Brownian motions, and they compute the critical values. The authors also show that residuals-based tests are consistent provided that first differences are used as dependent variables in the residual regression. This paper goes one step further and computes the local analytical power.

LEMMA 1: *When the model is generated according to* (1.1) *with* $n_1 = 1$*, then, as* $T \to \infty$[9],

$$(2.1) \qquad \left(\hat{b}_{ols} - b\right)' \Rightarrow \left(\int B_1^{d^2}\right)^{-1}\left(\int B_1^d \, K_{2c}^d\right)$$

where $\hat{b}_{ols}$ *is the LS estimator in the cointegration regression* (1.2) *and* $B' = \left[B_1 \ B_2\right]$ *is a* $n \times 1$ *vector Brownian motion partitioned conformably to* v_{1t} *and* v_{2t} *with covariance matrix* Ω. K_{2c} *is a scaled Ornstein Uhlenbeck process such that:*

$$K_{2c}(l) = B_2(l) + c\int_0^l e^{(l-s)c} B_2(s)\, ds \quad \text{and}$$

(i) $B^d = B$ *and* $K_{2c}^d = K_{2c}$ *if* $z_{1t} = 0$ *and* $z_{2t} = 0$

(ii) $B^d = B - \int B$ *and* $K_{2c}^d = K_{2c} - \int K_{2c}$ *if* $z_{1t} = 0$ *and* $z_{2t} = 1$ *and a mean is included in the regressions.*

(iii) $B^d = B - (4 - 6l)\int B - (12l - 6)\int sB$ *if* $z_{1t} = 1$ *and* $z_{2t} = 1$ *and mean and trend are included in the regressions.*

If $D' = 0$ and $c = 0$, then $K_{2c}^d \equiv B_2^d$ and (2.1) coincides with the usual distribution for spurious regressions as defined by Granger and Newbold (1974). In the original definition of spurious regression, there is no cointegration and indeed no relationship between the two variables. As Phillips (1986) shows, the same result is valid in the more general case in which $D \neq 0$. In this case, even though there is a relationship between the two variables, this relationship is not consistently estimated; the estimated beta does not converge in probability to the true value and the asymptotic distribution of Lemma 1 is the same as in Phillips and Ouliaris (1990).

THEOREM 1: *When the model is generated according to* (1.1) *with* $n_1 = 1$*, then, as* $T \to \infty$*:*

[9] From now on I adopt the conventional simplified notation B to denote $B(l)$. All the integrals are intended to be between 0 and 1, unless otherwise specified.

$$(2.2)\quad \hat{t}_a^{ADF} \Rightarrow c\frac{\left[\int W_1^{d\,2}\int J_{12c}^{d\ 2}-\left(\int W_1^{d}J_{12c}^{d}\right)^2\right]^{1/2}}{\left[\int W_1^{d\,2}+\left(\int W_1^{d}J_{12c}^{d}\right)^2\left(\int W_1^{d\,2}\right)^{-1}-2d\int W_1^{d}J_{12c}^{d}\right]^{1/2}}+$$

$$+\frac{\left[\int W_1^{d}J_{12c}^{d}\left(\int W_1^{d\,2}\right)^{-1}-d\right]\left[\int W_1^{d}J_{12c}^{d}\int W_1^{d}dW_1-\int J_{12c}^{d}dW_1\left(\int W_1^{d\,2}\right)\right]+\left(1-d^2\right)^{1/2}\left[\int W_1^{d\,2}\int J_{12c}^{d}dW_2-\int W_1^{d}J_{12c}^{d}\int W_1^{d}dW_2\right]}{\left[\int W_1^{d\,2}\int J_{12c}^{d\ 2}-\left(\int W_1^{d}J_{12c}^{d}\right)^2\right]^{1/2}\left[\int W_1^{d\,2}+\left(\int W_1^{d}J_{12c}^{d}\right)^2\left(\int W_1^{d\,2}\right)^{-1}-2d\int W_1^{d}J_{12c}^{d}\right]^{1/2}}$$

where $\hat{t}_a^{ADF}$ *is the t ratio test from the Augmented Dickey Fuller regression* (1.3) *and* $J_{12c}=d\,J_{1c}+\left(1-d^2\right)^{1/2}J_{2c}$. J_{ic} *are standard Ornstein-Uhlenbeck process defined as* $J_{ic}(l)=W_i(l)+c\int_0^l e^{(l-s)c}W_i(s)\,ds$ *for* $i=1,2$, $W_i(l)$ *are standard Brownian motions and*

(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* $J_{12c}^d=J_{12c}-\int J_{12c}$ *if* $z_{1t}=0$ *and* $z_{2t}=1$

(iii) $W^d=W-(4-6l)\int W-(12l-6)\int sW$ *and*

$J_{12c}^d=J_{12c}-(4-6l)\int J_{12c}-(12l-6)\int sJ_{12c}$ *if* $z_{1t}=1$ *and* $z_{2t}=1$.

The result of Theorem 1 shows that the asymptotic distribution of the ADF test on the residuals is a function of few parameters. It depends on the alternative c thorough the first block of the test and the Ornstein-Uhlenbeck process. The test also contains a nuisance parameter d in the second block of the test and in the functional J_{12c}^d. Finally, although the proof is presented for the univariate case, the test will depend on the dimension of x_t.

When $c=0$, $J_{12c}=d\,W_1+(1-d^2)^{1/2}W_2$ so the asymptotic distribution of the t statistic under the null is a function of standard Brownian motions and depends only on the dimension of

x_t (see also theorem 4.2 in Phillips and Ouliaris (1990)). So, while under the null hypothesis the limiting distribution of the test is free of nuisance parameters, when $c \neq 0$ the asymptotic power depends on c and on the value of d, the correlation between error terms. Notice that the power does not depend on either the true b or the variances of the errors.

3.2 EC, ECR AND WALD TESTS

Various papers have been written on the properties of the EC test. Assuming b known, Banerjee et al (1986) and Kremer et al. (1992) compute the asymptotic distribution of the EC test in (1.5) for the case in which the correlation between the two error terms in (1.1) (d) is zero. They also compute the power of the test under fixed and local alternative. As the mentioned paper shows, the asymptotic distribution of the EC test is non-standard and it is a function of Brownian Motions. The following theorem extends their results to the case in which D is different from zero[10].

THEOREM 2: *When the model is generated according to* (1.1) *with* $n_1 = 1$, *then as* $T \to \infty$:

$$\hat{t}_a^{EC} \Rightarrow \frac{c}{\left(1-d^2\right)^{1/2}\left(\int J_{12c}^2\right)^{-1/2}} + \left(\int J_{12c}^2\right)^{-1/2}\left(\int J_{12c}\,dW_2\right) \qquad (2.3)$$

where $\hat{t}_a^{EC}$ *is the t ratio test on the Error Correction term in* (1.5), *and* J_{12c} *and* W_2 *are defined as in Theorem 1.*

As Zivot (2000) also shows the EC test has the same asymptotic distribution as Hansen's (1995) (Theorem 3) unit root test on u_t when Δx_t is used as the stationary covariate. When

[10] The results of this paper assume that the common factor restriction imposed by the ADF regression is valid. For a study on the effects of violation of this restriction on the DF test when the cointegration vector

$d = 0$, as for example in the case in which x_t are strictly exogenous, the result of Theorem 2 agrees with of Kremer et al. (1992) in the particular case in which $a = 1$ (i.e. what they call " a common factor restriction " is valid).

When $c = 0$, $J_{12c} = d\, W_1 + (1-d)^{1/2} W_2$ and the asymptotic distribution of $\hat{t}_a^{EC}$ coincides with the result theorem 3 in Hansen (1995). Under the null the test is not invariant with respect to a nuisance parameter under the null. For this reason a unique set of critical values for the tests cannot be obtained.

As we showed, when the EC equation (1.5) is extended by adding a redundant regressor, a test for no cointegration can be performed by looking at the significance of the difference of the variables in level without the knowledge of the cointegrating vector. This transformation also renders the studentized test statistic invariant with respect to d under the null. Banerjee et al. (1998) compute the distribution of this test under the null for the case in which $d = 0$. Theorem 3 presents the general asymptotic distribution of the test under the local alternative $T(r-1)$.

THEOREM 3: *When the model is generated according to* (1.1), *then as* $T \to \infty$*:*

$$(2.4) \qquad \hat{t}_a^{ECR} \Rightarrow c\frac{\left[\int W_1^{d^2} \int J_{12c}^{d\;2} - \left(\int W_1^d J_{12c}^d\right)^2\right]^{1/2}}{(1-d^2)^{1/2}\left[\int W_1^{d^2}\right]^{1/2}} + $$

$$+ \frac{\int W_1^{d^2} \int J_{12c}^d dW_2 - \int W_1^d J_{12c}^d \int W_1^d dW_2}{\left[\int W_1^{d^2} \int J_{12c}^{d\;2} - \left(\int W_1^d J_{12c}^d\right)^2\right]^{1/2}\left[\int W_1^{d^2}\right]^{1/2}}$$

is imposed, see Kremers et al. (1992). Ostermark and Hoglund (1998) generalize Kremers et al.(1992) results to the case in which the cointegration vector is estimated.

where $\hat{t}_a^{ECR}$ *is the t ratio test in equation* (1.6)*, and* J_{12c}^d *and* W_1^d *are defined as in Theorem 1.*

Under the local alternative, the ECR test depends not only on the particular alternative c, but also on the value of the nuisance parameter d. In contrast with the results for the EC test, under the null the asymptotic distribution of the ECR is free of the nuisance parameter. In fact when $c = 0$,

$$\hat{t}_a^{ECR} \Rightarrow \frac{\int W_1^d \int W_2^d dW_2 - \int W_1^d W_2^d \int W_1^d dW_2}{\left[\int W_1^{d\,2} \int W_2^{d\,2} - \left(\int W_1^d W_2^d\right)^2\right]^{1/2} \left[\int W_1^{d\,2}\right]^{1/2}}$$

so in this case the asymptotic distribution of the test depends only on the dimension of x_t and it is invariant of any other parameter of the DGP. Zivot (2000) also computes the local power for the ECR test in a slightly different model.

Under the null hypothesis of no cointegration, both a and f in (1.6) are zero so, as Boswijk (1994) suggests, the null hypothesis can be tested using a joint test on both parameters. Theorem 4 computes the local power of Boswijk (1994) Wald test.

THEOREM 4: *When the model is generated according to* (1.1)*, then as* $T \to \infty$*:*

$$\text{(2.5)} \qquad \hat{F} \Rightarrow c^2\left(1-d^2\right)^{-1} \int J_{12c}^{d\;2} + 2c\left(1-d^2\right)^{-1/2} \int J_{12c}^d dW_2 + \\ + \frac{\int W_1^{d\,2}\left(\int J_{12c}^d dW_2\right)^2 - 2\int W_1^d J_{12c}^d \int W_1^d dW_2 \int J_{12c}^d dW_2 + \int J_{12c}^{d\;2}\left(\int W_1^d dW_2\right)^2}{\int W_1^{d\,2} \int J_{12c}^{d\;2} - \left(\int W_1^d J_{12c}^d\right)^2}$$

where $\hat{F}$ *is the Wald test for the joint significance of* a *and* f *in* (1.6) *and* J_{12c}^d *and* W_i^d *are defined as in Theorem 1.*

In this case too the only parameter entering the asymptotic distribution of the test are the particular alternative c and the correlation term d. The test does not depend on the variances of the error term or the true value of the cointegrating vector.

For $c = 0$ the asymptotic distribution of the test does not depend on any nuisance parameter and coincides with the result of Theorem 1 of Boswijk (1994):

$$\hat{F} \Rightarrow \frac{\int W_1^{d\,2}\left(\int W_2^d dW_2\right)^2 + \int W_2^{d\,2}\left(\int W_1^d dW_2\right)^2 - 2\left(\int W_1^d W_2^d\right)\left(\int W_1^d dW_2\right)\left(\int W_2^d dW_2\right)}{\int W_1^{d\,2}\int W_2^{d\,2} - \left(\int W_1^d W_2^d\right)^2}$$

3.3 JOHANSEN TEST

The ECM representation of the simple model (1.1) is

(2.6) $\Delta Y_t = G + \Pi Y_{t-1}^* + \Pi(L)\Delta Y_{t-1} + w_t$

where $\Pi = a\,g'$ with $a_1' = [f_{12}(1)(r-1) \quad [f_{22}(1) + bf_{12}(1)](r-1)]$ and $g' = [-b \quad 1 \quad -m_1]$. While under the null of no cointegration $r = 1$ and $rank(\Pi) = 0$, when y_t and x_t are cointegrated $r < 1$ and $rank(\Pi) = 1$.

The local alternative for the rank test suggested by Johansen (1995) is of the form:

$$H_a:\ \Pi = a g' + \frac{a_1 g_1'}{T}$$

where a_1 and g_1 are $n \times s$ matrices. Under the local alternatives the process has s extra cointegrating vectors, g_1, that enter the process with very small adjustment coefficients $T^{-1}a_1$.

In the case of the $l_{\max}$ test the relevant local alternative is:

$$H_a:\ \Pi = \frac{a_1 g_1'}{T}$$

where $a' = c[f_{12}(1) \quad f_{22}(1) + bf_{12}(1)]$ and $g_1' = [-b \quad 1 \quad -m_1]$.

Johansen (1995) discusses the power of the rank test under this general local alternative. Since the maximum eigenvalue test is just a special case of the rank test, it can be shown that:

THEOREM 5: *When the model is generated according to* (1.1)*, then as* $T \to \infty$*:*

(2.7) $$\hat{l}_{\max} \Rightarrow \max eig\Big\{\left(\int W_c W_c^{'}\right)^{-1}\int W_c\, dW^{'}\int dW W_c^{'} + \left(\int W_c W_c^{'}\right)^{-1}\int W_c\, dW^{'} A_c^{'} + + \left(\int W_c W_c^{'}\right)^{-1} A_c \left(\int dW W_c^{'}\right)^{\prime} + \left(\int W_c W_c^{'}\right)^{-1} A_c A_c^{'}\Big\}$$

where $\max eig\{M\}$ *denotes the maximum eigenvalue of the matrix M, ,* $W = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}$,

$A'_c = \begin{bmatrix} 0 \\ c(1-d^2)^{-1/2} \end{bmatrix} \int W_c \,' J_{12c}$ *and*

(i) $W_c = \begin{bmatrix} W_1 \\ J_{12c} \end{bmatrix}$ *if* $z_{1t} = 0$ *and* $z_{2t} = 0$ *and no mean is included in the VAR*

(ii) $W_c = \begin{bmatrix} W_1 \\ J_{12c} \\ 1 \end{bmatrix}$ *if* $z_{1t} = 0$ *and* $z_{2t} = 1$ *and the mean is restricted to enter the Error Correction term.*

(iii) $W_c = \begin{bmatrix} W_1^d \\ J_{12c}^d \\ 1 \end{bmatrix}$ *if* $z_{1t} = 1$ *and* $z_{2t} = 1$*, a mean is restricted to enter the Error Correction term and the VAR is estimated with a mean.* W_1^d *and* J_{12c}^d *are demeaned Brownian Motions.*

If the mean is not restricted to enter the Error Correction term then $W_c = \begin{bmatrix} W_1^d \\ J_{12c}^d \end{bmatrix}$ and the Brownian Motions are demeaned for case (ii) and detrended for case (iii). See also Johansen

(1995) and Saikonnen and Lutkepohl (1999). Again the only parameter entering the power function is the correlation coefficient d. Under the null hypothesis, $W_c = W$ and the distribution of the test is free of nuisance parameters as in Johansen and Juselius (1990).

4. MONTE CARLO RESULTS

Previous Monte Carlo comparisons of cointegration tests have shown that different tests can perform very differently depending on the particular design. Haug (1996), for example, compares 9 different tests for cointegration on the basis of power and size distortions due to the presence of moving average component for endogenous or exogenous x_t. Haug found that, in general, the single equation tests have smaller size distortions, but also have lower power than system-based tests. The paper concludes by recommending the application of both sets of tests in empirical exercises.

As Haug (1996) also points out “A theory that gives the direction in which to experiment would be necessary but this theory is not available at the moment”. The analytical power computed in the previous section tells us exactly which parameters are important for power and gives a precise indication of which direction we need to look in the Monte Carlo analysis: The asymptotic distribution is a function of a unique nuisance parameter, the number of unit roots in the system and the local alternative. Because of the lack of asymptotic normality and the fact that the test is not invariant on the particular alternative, an uniformly most powerful test for model (1.1) cannot be computed. There is no reason then why we should expect one of the tests to behave uniformly better than the others. At the same time, if the right hand variables in (1.2) were highly correlated with the errors from the cointegration regression (i.e. delta is large), we would expect a full system approach to exploit this correlation, to have smaller standard errors, and to perform better.

Using a Monte Carlo experiments, this section compares the power of the tests for the absence of cointegration presented in the previous sections.

4.1 LARGE SAMPLE

The power functions are computed as the probability that the tests are less than some critical value. Since we have an expression for the limit distribution of all the tests we can approximate the asymptotic power by simulating the distributions presented in the previous sections. For this experiment I consider $c = 0, -1, -5, -10, -15$ and $d = 0, 0.2, 0.3, 0.5, 0.7, 0.9$, negative values of delta give similar results. Each Brownian Motions piece in the asymptotic distribution is approximated by step functions using Gaussian random walk with T=1000 observations. To eliminate initial condition effects, in practice I generated 1100 observations and discarded the first 100. For example, $T^{-2}\Sigma u_{ct}^2$ is used to approximate $\int J_{1c}^2$, where $u_{ct} = (1 + c/T)u_{ct-1} + e_{1t}$, and e_{1t} is randomly generated form a standard normal. 5000 replications for are used to compute the critical values and the rejection probabilities for each c and delta..

Since we showed that the local power for all the tests depends on a nuisance parameter, the power functions of the tests are compared for different values of delta. The large sample results in Figure 1 confirm the intuition. For d small, in a close neighbor of one, the ECR test has slightly higher power than the ADF and Wald test. For d equal to zero using a full system approach is inefficient and the ADF, Wald and ECR tests perform better than the $l_{\max}$ test. As d increases, all the tests with the exception of the ADF test perform better. As expected, given that they are both based on the conditional error correction equation on a rotated model, Wald and ECR tests have in general similar large sample power.

For large values of delta, we expect the single equation approach of the ADF test on the residuals to have very low power. In fact when d is higher than 0.5, this test performs badly when compared to the other tests. Although the power of the maximum eigenvalue test is significantly higher than the power of the ADF test, the power of both the ECR and the Wald is even higher than the Johansen's test. For d equal to 0.9, the difference between the ADF and the other tests is even more significant. All other tests have very similar power.

Figure 2 and 3 show the local power of the tests for the demeaned and detrended cases. For the Wald and the ECR I consider the case in which a mean enters the cointegration regression but the mean or the trend in equation (3.6) are left unconstrained. Although the tests have generally lower power when I include a deterministic, as is the general case for unit root tests, the rankings between the tests is unchanged

4.2 SMALL SAMPLE

To examine the usefulness of these asymptotic approximation in practice we need to study the small sample behavior of the tests. Using the DGP of equation (1.1), I randomly generate the errors from a bivariate Normal with mean zero and variance-covariance matrix

$$\Omega = \begin{bmatrix} w_1^2 & w_1 w_2 d \\ w_1 w_2 d & w_2^2 \end{bmatrix}$$

I consider $d = 0, 0.3, 0.5, 0.7, 0.9$ and $c = 0, -1, -5, -10, -15$ that corresponds for T=100 to values for $r = 1,\ 0.99,\ 0.95,\ 0.9,\ 0.85$. The tests are all invariant to b and the variances so I can choose any number. I use $w_1 = w_2 = 0.25$ and $b = 1$. Table 1 presents the size adjusted rejection rates for the case in which there is no serial correlation in the error terms. Since no serial correlation is assumed all the regression are estimated with no lags.

As Table 1 shows, all tests have a low power when the root is close to one. For d small, the ECR and the Wald tests have in general power higher than all the other tests for any value of r. The relative ranking of the tests is the same as in the large sample case. When delta is 0.9, the difference in the small sample power of the tests is significant even when ρ is large.

More interesting is the case in which there 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 looked 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 (1.1) with $(1-\Phi L)v_t = (1+\Theta L)e_t$ and $\Phi = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}$, $\Theta = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$ 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. Table 4 presents the results for different combinations of values for the autoregressive and moving average components. In general if only an autoregressive component is present in the error terms the BIC performs really well in choosing the appropriate number of lags. In this case the Wald test has the worse performance in term of size distortions with an empirical size of 10%. As it is well known, large size distortions occur when large negative roots are present in the moving average components. In this case all tests present large empirical sizes with a remarkable performance of the Wald test that shows a size of less than 20% when all the other tests are well above 50%.

As a general conclusion, table 1-3 and table 4 seam to indicate that the ECR and the Wald test, although the less commonly used of the tests presented in this paper, not only perform better than the other tests in term of power in large and small samples, but they are also not worse or better in term of size distortions.

5. EMPIRICAL APPLICATION

As an empirical application I analyze whether there exists a stable relation between money balances and the determinants of money demand. There is a vast literature testing for cointegration in the general money demand equation $m_t - p_t = a + b\, y_t + g\, i_t$ or its various restricted forms. The results are in general inconclusive; the rejection of the hypothesis of no cointegration is very sensitive to the sample size and the particular data used. Miller (1991), for example, considers 5 alternatives for the monetary aggregates, two alternatives for the interest rate and three different specifications for the money demand equation. The data are quarterly data for the period from 1959 to 1987. The ADF test on the residuals of the 15 cointegration regressions, finds cointegration only in few cases. The results change with the choice of the interest rate and the lag length used in the test.

A full system approach has also been used in various cases: Hoffman and Rasche (1991), Friedman and Kuttner (1992), Hafer and Jansen (1991), among others, studied the money demand relation in a Vector Error Correction Model (VECM) model and used the Johansen tests to test for cointegration. Some evidence of cointegration can be found in data up to the beginning of 1990 although the results are again highly sensitive to the choice of the monetary variable and the interest rate.

Miyao (1996) argues that the evidence of cointegration found in previous studies is not strong even though the data show sign of cointegration in samples up to 1990. When the sample is increased to include data up to 1993, the hypothesis of no cointegration cannot be rejected. Only Johansen's maximum eigenvalue test finds cointegration in few cases. Miyao argues that this is due to size distortions in Johansen's test and concludes that the rejections of the maximum eigenvalue test are highly questionable.

Following our analysis above, in cases in which the right hand variables in the cointegration vector and the errors in the cointegration equation are highly correlated, the power of the ADF test is so low that a rejection is highly unlikely. The money demand equation is a clear example of a situation in which we would expect the correlation between the residuals to be large and a full system approach to have better power.

I test for cointegration in the three version of the money demand equation: one with unrestricted coefficients for both the income and the interest rate, one where the coefficient on income is imposed equal to one and finally one in which the interest rate is not included. The data considered are from the Federal Reserve Economic Database (FRED). I use quarterly data from the first quarter of 1959 to the fourth quarter of 1997. The variables considered are the logarithm of GDP in 1992 dollars, the logarithm of M2 and real GDP deflator. For the interest rate I follow Miyao (1996) and use the three-month Treasury bill rate not in logarithm transformation. For M2 and the interest rate quarterly data are obtained by averaging monthly observations over the months within the quarter.

Preliminary unit roots tests are conducted on all variables and are available upon request. We then test for cointegration between $m_t - p_t$ and y_t and i_t, and between $m_t - p_t - y_t$ and i_t. Since it is arguable that the interest rate does not contain a unit root, we also test for cointegration only between $m_t - p_t$ and y_t. For all cases a constant is included in the regressions and for both the ECR and the Wald test the constant is also jointly tested to be equal to zero. The lag length is chosen using BIC in the residual regression for the ADF test and on a VAR on differences and levels for all other tests.

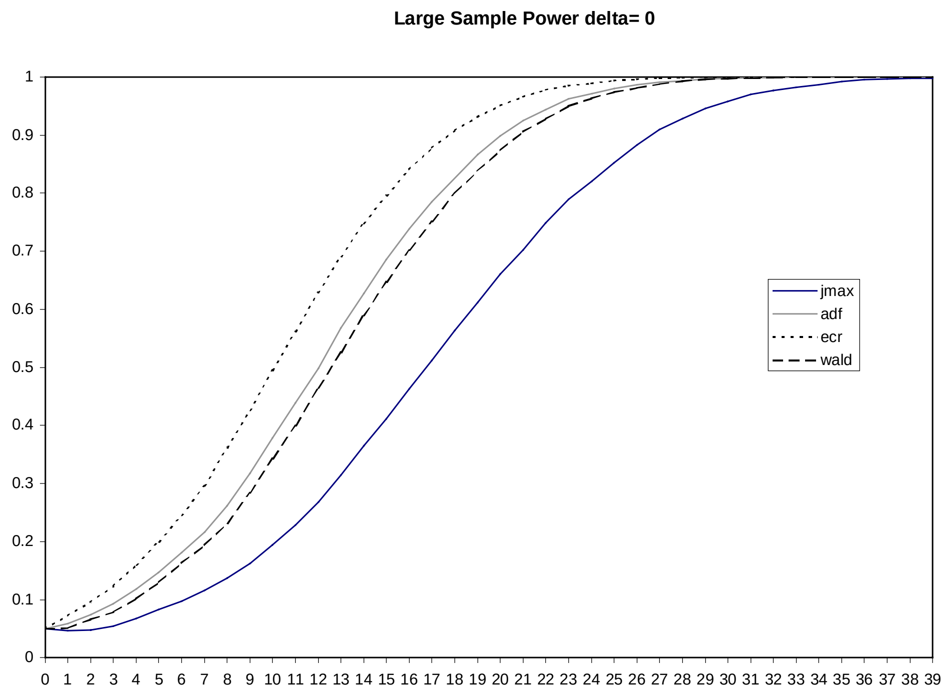
Table 5 presents the results for all the tests considered where the regressions are estimated with mean and trend to allow for a drift in the variables. As we can see from the table, the ADF test never rejects the null of no cointegration. The maximum eigenvalue strongly rejects in all three models while the Wald test rejects in most cases. Since we can expect that there is a lot of simultaneity in the money demand equation, table 5 also presents an estimate for D. The estimated value for the correlation at frequency zero are between 0.3 and 0.8. In light of my theoretical results it is not surprising that the ADF test is not able to reject the null of no cointegration. On the other hand, system based tests reject the null more frequently suggesting the existence of cointegration in the money demand equation. Obviously these results are only preliminary and a more in deep analysis of the money demand problem is necessary. At the same time, Table 5 and the theoretical analysis of this paper can help in understanding the puzzling empirical results that different tests for cointegration can lead to contradictory conclusions.

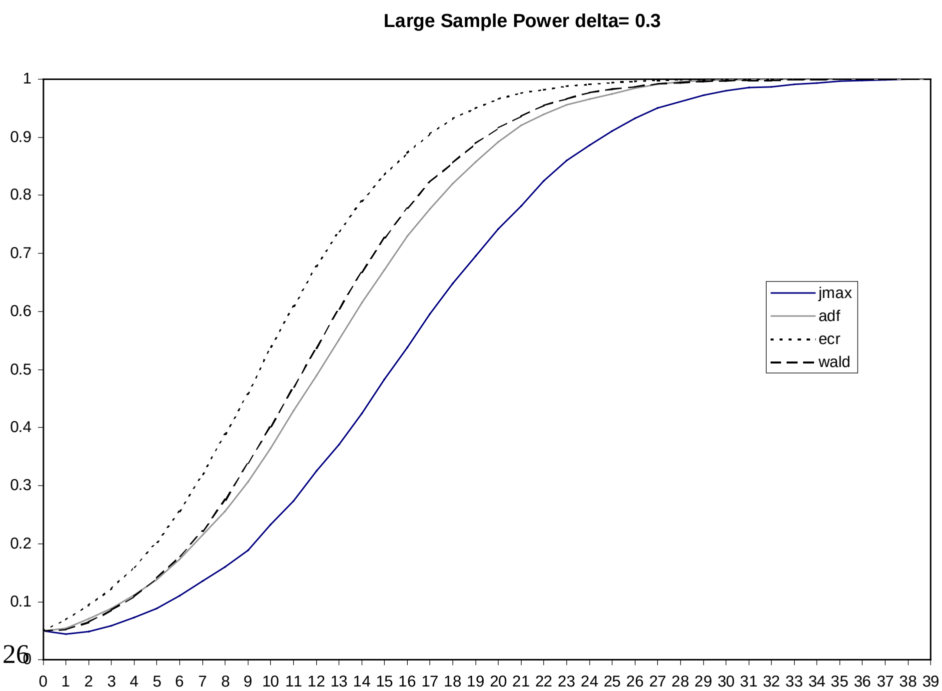
6. CONCLUSIONS

Over the past years testing for cointegration has become a very important step in any empirical analysis. This paper illustrates how the analytical analysis of the local power of the tests can help identifying what nuisance parameters are relevant. A complete theoretical understanding of current methods can help in designing the appropriate Monte Carlo experiment to evaluate the relative performance of the tests. In particular, this paper look at the class of set of tests for the absence of cointegration and show that a very important role is played by the correlation of the independent variable with the errors of the cointegration regression. As the intuition suggests, when this correlation is very high, system approaches like the Johansen maximum eigenvalue or tests of the Error Correction model can exploit this correlation and significantly outperform single equation tests. An empirical application to the money demand equation explains why the Augmented Dickey Fuller test on the residuals of the cointegration regression and other tests can give contradictory results.

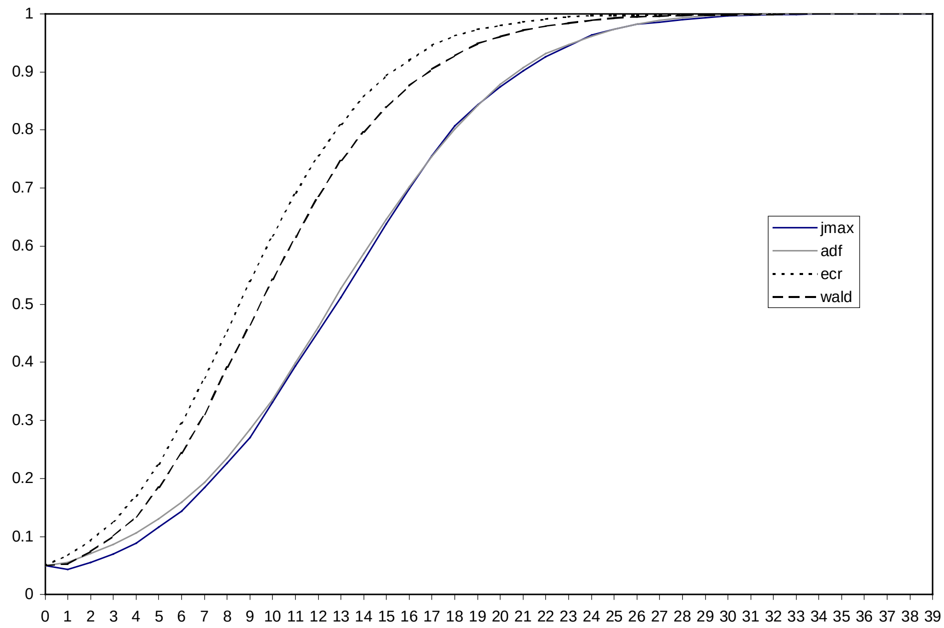
Further research still needs to be done to fully understand how all the tests for cointegration work. The absence of asymptotic normality and the fact that under the null the cointegration vector is not identified, destroy the usual optimality theory. This new understanding of the asymptotic power functions of tests for the absence of cointegration is a good stepping-stone for future research.

Figure 1: Large Sample Power

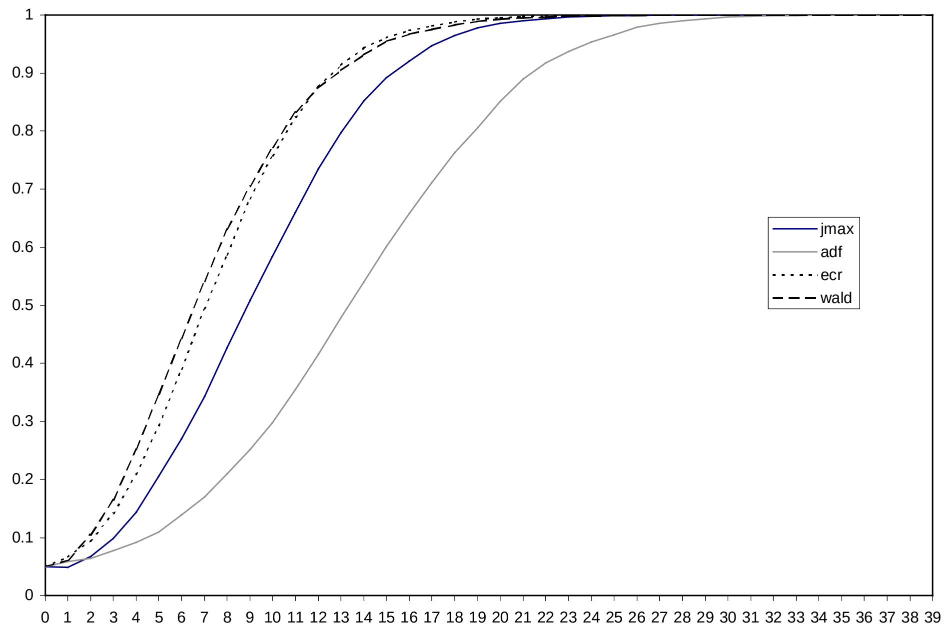




Large Sample Power delta= 0.5



Large Sample Power delta= 0.7



Large Sample Power delta= 0.9

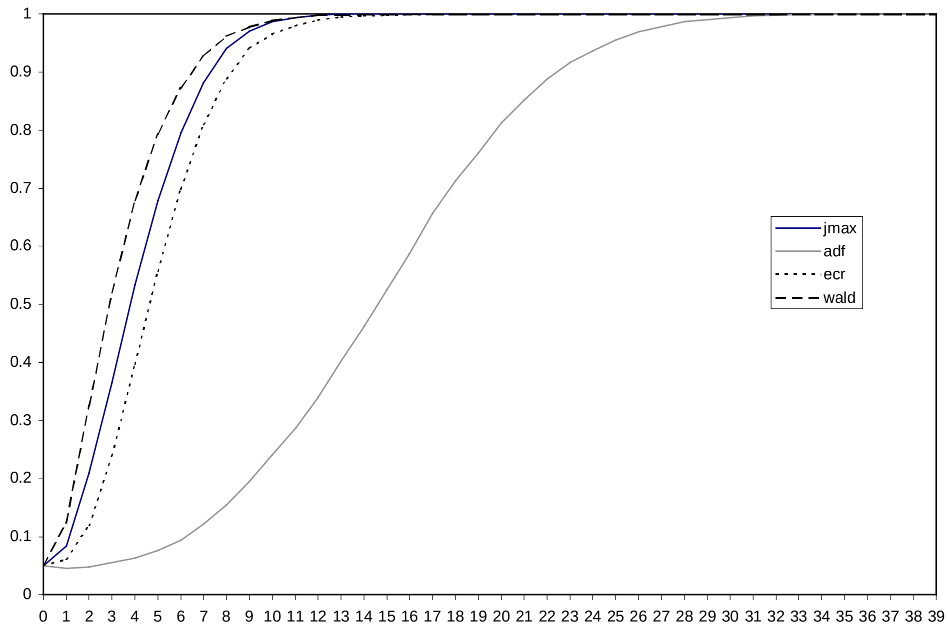
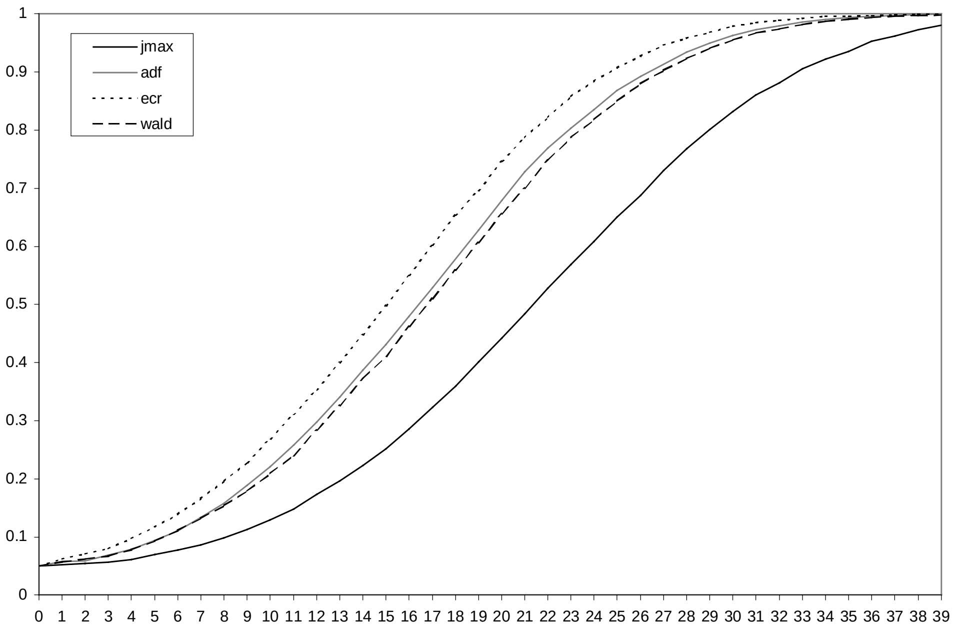
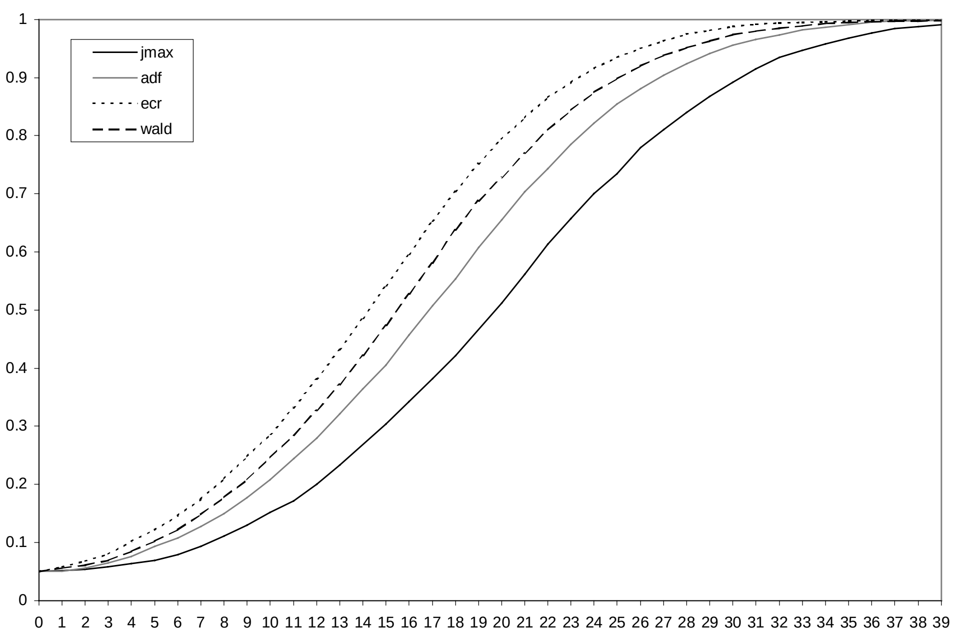


Figure 2: Large Sample Power, Demeaned case.

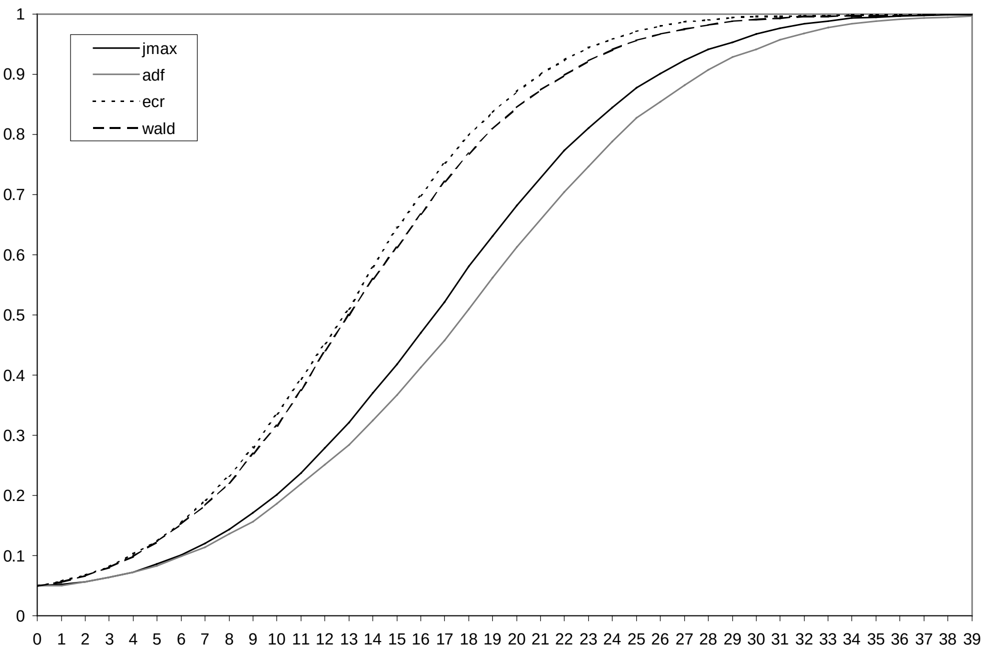
Large Sample Power delta=0 , demeaned



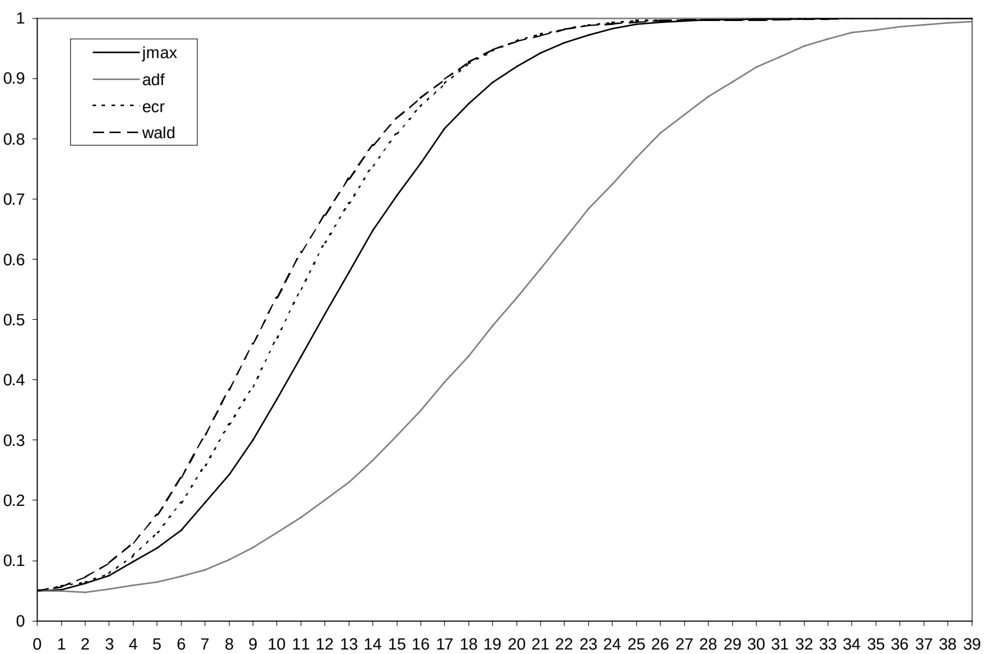
Large Sample Power delta=0.3 , demeaned



Large Sample Power delta=0.5 , demeaned



Large Sample Power delta=0.7 ,demeaned



Large Sample Power delta=0.9 , demeaned

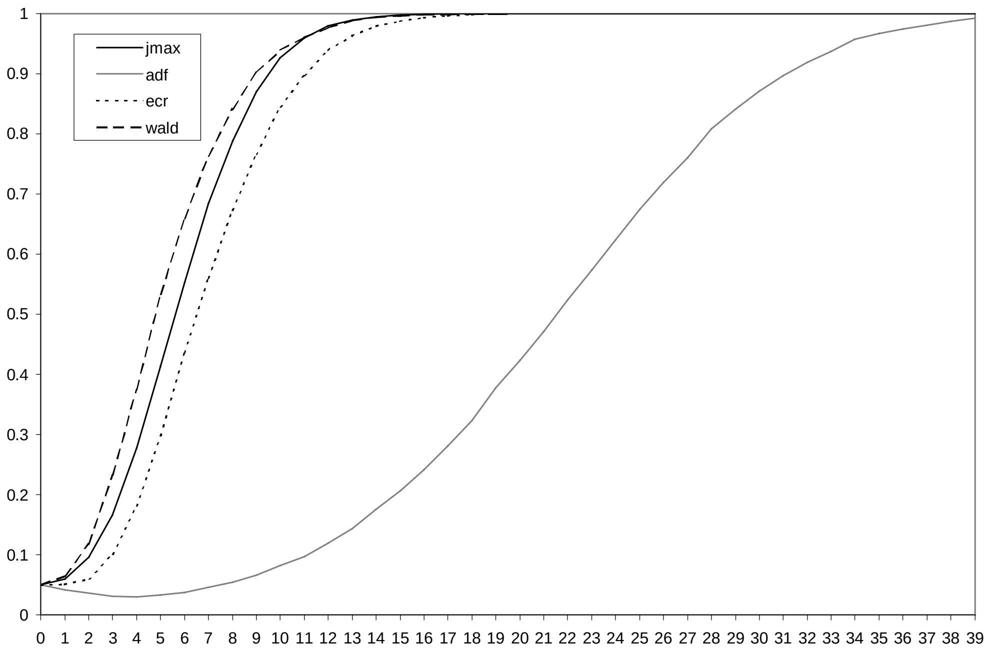
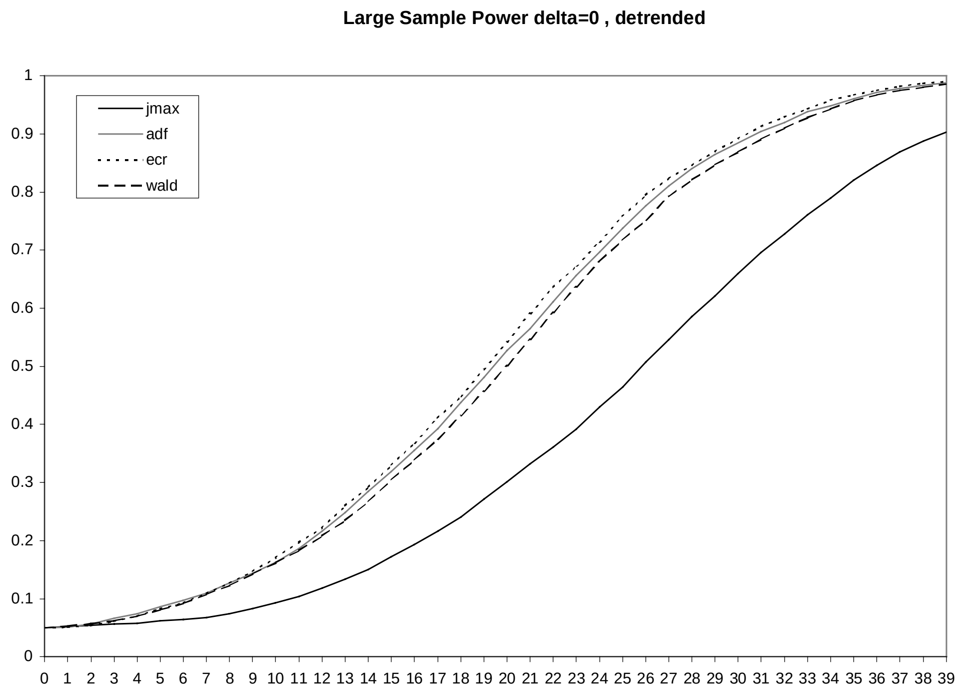
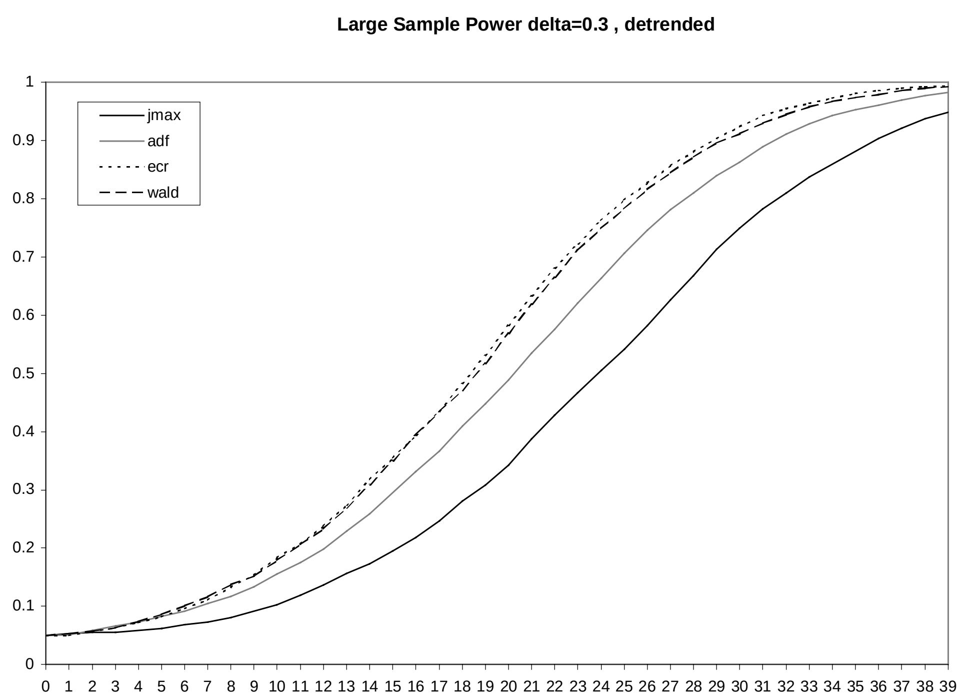
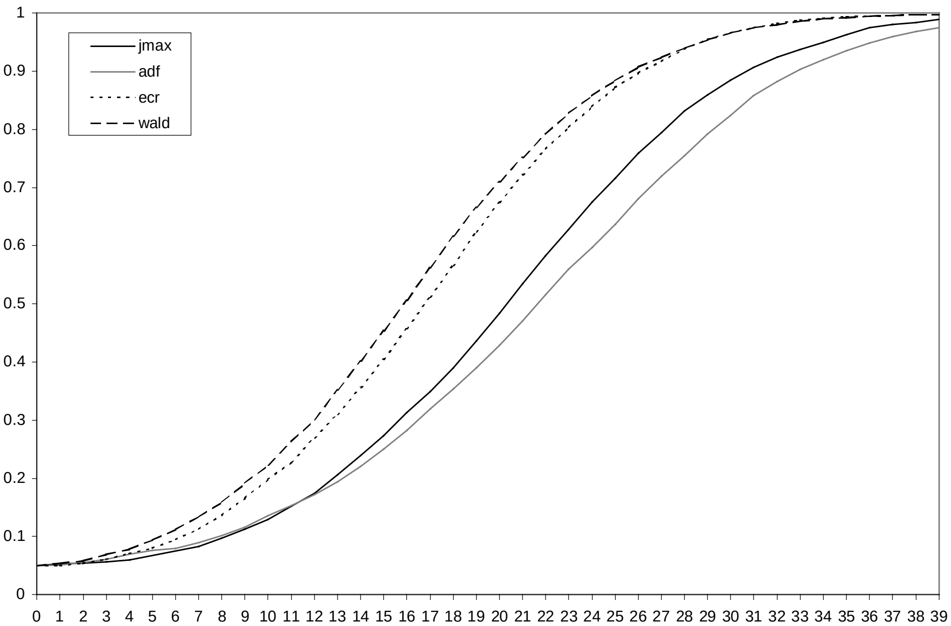


Figure 3: Large Sample Power, demeaned and detrended case

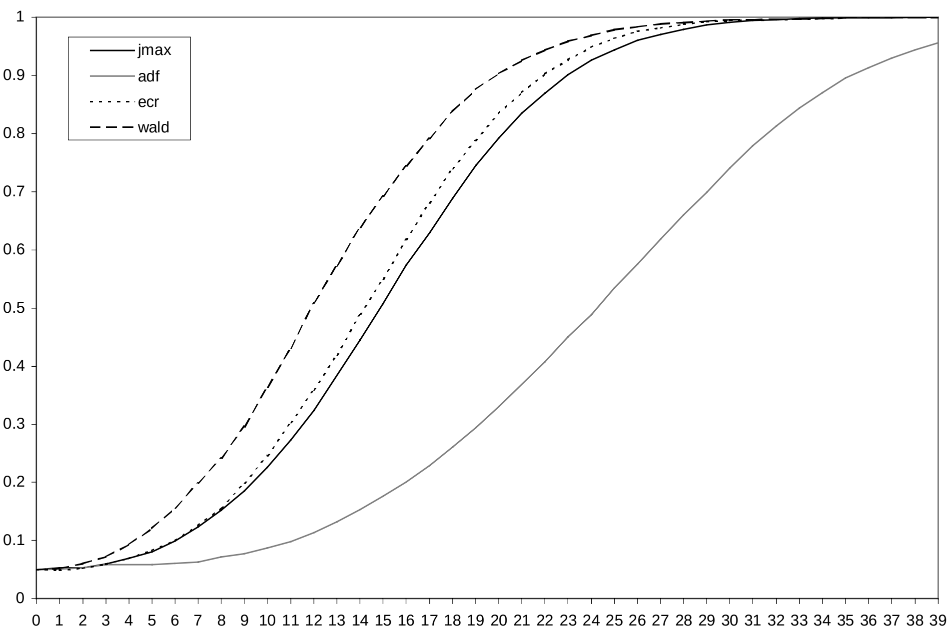




Large Sample Power delta=0.5 , detrended



Large Sample Power delta=0.7 , detrended



Large Sample Power delta=0.9 , detrended

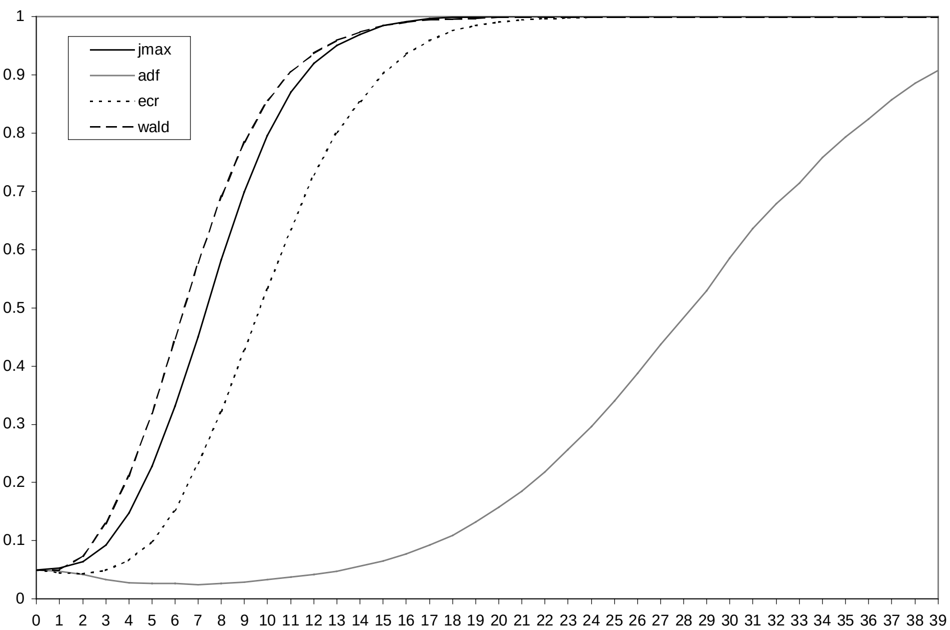


Table 1: Size adjusted small sample power, no constant, no serial correlation

-c		0	5	10	15	20
$\delta \setminus \rho$		1	0.95	0.90	0.85	0.80
	ADF	0.050	0.141	0.386	0.695	0.912
0	ECR	0.050	0.208	0.493	0.802	0.955
	Wald	0.050	0.117	0.326	0.615	0.864
	λ_{max}	0.050	0.083	0.202	0.422	0.679
0.3	ADF	0.050	0.141	0.373	0.687	0.913
	ECR	0.050	0.231	0.542	0.841	0.971
	Wald	0.050	0.138	0.385	0.701	0.908
	λ_{max}	0.050	0.090	0.239	0.492	0.761
0.5	ADF	0.050	0.134	0.353	0.670	0.905
	ECR	0.050	0.267	0.645	0.907	0.986
	Wald	0.050	0.190	0.516	0.818	0.965
	λ_{max}	0.050	0.121	0.338	0.643	0.884
0.7	ADF	0.050	0.129	0.324	0.646	0.893
	ECR	0.050	0.358	0.645	0.907	0.986
	Wald	0.050	0.330	0.757	0.951	0.995
	λ_{max}	0.050	0.215	0.586	0.896	0.990
0.9	ADF	0.050	0.106	0.294	0.604	0.870
	ECR	0.050	0.700	0.984	1.000	1.000
	Wald	0.050	0.788	0.988	1.000	1.000
	λ_{max}	0.050	0.684	0.986	1.000	1.000

Note: The power is computed with T=100 and 5000 replications.

Table 2: Size adjusted small sample power, demeaned, no serial correlation

-c		0	5	10	15	20
δ \ ρ		1	0.95	0.90	0.85	0.80
	ADF	0.050	0.098	0.227	0.454	0.702
0	ECR	0.050	0.113	0.268	0.514	0.758
	Wald	0.050	0.102	0.227	0.452	0.695
	λ_{max}	0.050	0.063	0.118	0.250	0.451
0.3	ADF	0.050	0.093	0.213	0.431	0.691
	ECR	0.050	0.115	0.298	0.560	0.806
	Wald	0.050	0.107	0.267	0.515	0.764
	λ_{max}	0.050	0.068	0.138	0.299	0.521
0.5	ADF	0.050	0.084	0.191	0.387	0.650
	ECR	0.050	0.123	0.349	0.655	0.883
	Wald	0.050	0.132	0.348	0.652	0.868
	λ_{max}	0.050	0.075	0.193	0.411	0.683
0.7	ADF	0.050	0.069	0.154	0.321	0.588
	ECR	0.050	0.145	0.487	0.814	0.973
	Wald	0.050	0.198	0.565	0.853	0.976
	λ_{max}	0.050	0.112	0.352	0.693	0.916
0.9	ADF	0.050	0.041	0.086	0.217	0.462
	ECR	0.050	0.309	0.848	0.993	1.000
	Wald	0.050	0.550	0.944	0.998	1.000
	λ_{max}	0.050	0.382	0.908	0.998	1.000

Note: The power is computed with T=100 and 5000 replications.

Table 3: Size adjusted small sample power, demeaned and detrended , no serial correlation

-c δ \ ρ		0 1	5 0.95	10 0.90	15 0.85	20 0.80
	ADF	0.050	0.072	0.155	0.317	0.535
	ECR	0.050	0.075	0.162	0.327	0.538
0	Wald	0.050	0.075	0.153	0.300	0.508
	λ_{max}	0.050	0.059	0.100	0.180	0.319
	ADF	0.050	0.069	0.140	0.294	0.502
	ECR	0.050	0.076	0.166	0.350	0.587
0.3	Wald	0.050	0.077	0.168	0.348	0.574
	λ_{max}	0.050	0.059	0.106	0.206	0.378
	ADF	0.050	0.065	0.124	0.253	0.438
	ECR	0.050	0.074	0.185	0.407	0.681
0.5	Wald	0.050	0.085	0.218	0.453	0.709
	λ_{max}	0.050	0.062	0.131	0.286	0.512
				0.087		
	ADF	0.050	0.050	0.243	0.178	0.344
	ECR	0.050	0.080	0.361	0.561	0.839
0.7	Wald	0.050	0.114	0.228	0.691	0.905
	λ_{max}	0.050	0.075		0.507	0.797
	ADF	0.050	0.024	0.032	0.072	0.174
	ECR	0.050	0.094	0.535	0.906	0.994
0.9	Wald	0.050	0.314	0.844	0.988	1.000
	λ_{max}	0.050	0.218	0.769	0.985	1.000

Note: The power is computed with T=100 and 5000 replications.

Table 4: Size distortions

AR errors	MA errors	ADF	ECR	Wald	λ_{max}
$f_{11} = f_{22} = 0$	$J_{11} = J_{22} = 0$	0.059 (1.1)	0.057 (1.3)	0.102 (0.3)	0.058 (1.1)
$f_{11} = f_{22} = 0.2$	$J_{11} = J_{22} = 0$	0.060 (1.1)	0.059 (1.3)	0.102 (1.3)	0.063 (1.7)
$f_{11} = f_{22} = 0.8$	$J_{11} = J_{22} = 0$	0.061 (1.1)	0.084 (1.3)	0.151 (1.3)	0.112 (2.1)
$f_{11} = f_{22} = 0.2$ $f_{12} = f_{21} = 0.5$	$J_{11} = J_{22} = 0$	0.036 (1.3)	0.045 (1.3)	0.116 (1.3)	0.075 (2.1)
$f_{11} = f_{22} = 0$	$J_{11} = J_{22} = 0.2$	0.068 (1.1)	0.064 (1.3)	0.098 (1.3)	0.061 (1.7)
$f_{11} = f_{22} = 0$	$J_{11} = J_{22} = 0.8$	0.069 (2.7)	0.065 (3.0)	0.122 (3.0)	0.073 (3.7)
$f_{11} = f_{22} = 0$	$J_{11} = J_{22} = 0.2$ $J_{11} = J_{22} = 0.5$	0.090 (1.1)	0.065 (1.8)	0.115 (1.8)	0.067 (3.0)
$f_{11} = f_{22} = 0$	$J_{11} = J_{22} = -0.8$	0.673 (1.6)	0.592 (1.9)	0.194 (1.9)	0.629 (2.4)
$f_{11} = f_{22} = 0$	$J_{11} = J_{22} = -0.8$ $J_{11} = J_{22} = 0.5$	0.629 (1.5)	0.290 (1.5)	0.159 (1.5)	0.537 (2.2)

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.

Table 5: Cointegration tests for the money demand equation.

Model	ADF	$l_{\max}$	ECR	WALD	$\hat{D}$
$m_t - p_t = c + b\, y_t + g\, i_t$	-3.04 (1)	78.90 (3)**	0.506 (3)	14.58(3)**	0.39, -0.30
$m_t - p_t = c + b\, y_t$	-2.95(1)	61.97 (1)**	-1.43 (1)	2.22 (1)	0.56
$m_t - p_t - y_t = c + g\, i_t$	-3.08 (4)	55.98 (2)**	-3.226 (2)	15.03 (2)**	-0.76

Note: The number in parenthesis represent the number of lags selected by BIC. The maximum possible number of lags was set to 4. Critical values are obtained by simulations, asterisk denotes rejection at 5%.

APPENDIX

LEMMA A1: *When the model is generated according to (3.1) with* $T(r-1)=c$ *, then, as* $T\to\infty$ [11]*:*

(i) $T^{-2}\Sigma x_t^{d} x_t^{d\,'} \Rightarrow \int B_1^d B_1^{d\,'}$

(ii) $T^{-2}\Sigma x_t^{d} u_t \Rightarrow \int B_1^d K_{2c}$

(iii) $T^{-2}\Sigma u_t^2 \Rightarrow \int K_{2c}^2$

(iv) $T^{-1}\Sigma u_{t-1} e_t \Rightarrow \Sigma^{1/2}\int K_{2c}\, dW$

(v) $T^{-1}\Sigma x_{t-1}^{d} e_t \Rightarrow \Sigma^{1/2}\int B_1^d\, dW$

where $i=1,2$ *, the summation goes from* 1 *to* T *and* $\Rightarrow$ *denotes weak convergence,* $[\;]_{ii}$ *denotes the ii element of a matrix.* $B^{'}=\left[B_1^{'}\;\; B_2\right]$ *is a* $n\times 1$ *vector Brownian motion partitioned conformably to* v_{1t} *and* v_{2t}*, with covariance matrix* Ω*.* K_{2c} *is a scaled Ornstein Uhlenbeck process such that:* $K_{2c}(l)=B_2(l)+c\int_0^l e^{(l-s)c}B_2(s)ds$ *and*

(1) if $z_{1t}=0$ *and* $z_{2t}=0$*,* $x_t^d=x_t$ *and* $B^d=B$*.*

(2) if $z_{1t}=0$ *and* $z_{2t}=1$*,* $x_t^d=x_t-\overline{x}$ *and* $B^d=B-\int B$*.*

(3) if $z_{1t}=1$ *and* $z_{2t}=1$*,* x_t^d *is* x_t *detrended by OLS and* $B^d=B-(4-6l)\int B-(12l-6)\int sB$*.*

Proof of Lemma A1:

By the multivariate FCLT and Phillips(1987) we have that $T^{1/2}\sum_{s=1}^{t} e_s \Rightarrow B(l)$ and $T^{-1/2}u_t \Rightarrow K_{2c}(l)$. Results (i)-(iii) follows from the Continuos Mapping Theorem (CMT) while (iv) and (v) follows directly from Chan and Wei(1988) or Phillips(1987).

Proof of Lemma 1:

$\hat{b}$ can be estimated by regressing y_t^d on x_t^d. If $z_{2t}=0$ then $x_t^d=x_t$ and $y_t^d=y_t$, for case (ii) both variable are demeaned and for case (iii) both variables are demeaned and

[11] I follow the usual convention and suppress the (λ) from the Brownian motion terms. Unless specified otherwise, all the integrals are intended to be between 0 and 1.

detrended. In the general case then $\hat{b} - b = \left(T^{-2}\Sigma\left(x_t^d\right)^2\right)^{-1}\left(T^{-2}\Sigma\, x_t^d u_t^d\right)$. From LEMMA A1 and the CMT we have that $\hat{b} - b \Rightarrow \left(\int B_1^{d\,2}\right)^{-1}\left(\int B_1^d K_{2c}^d\right)$.

COROLLARY A1: *When the model is generated according to (3.1) with* $T(r-1) = c$, *then, as* $T \to \infty$ *:*

(i) $T^{-2}\Sigma\,\hat{u}_t^{d\,2} \Rightarrow h_c^{d'} A_c^d h_c^d$

(ii) $T^{-1}\Sigma\,\hat{u}_{t-1}^d \Delta\hat{u}_t^d \Rightarrow c h_c^{d'} A_c^d h_c^d + d(1) h_c^{d'} \int B_c^d dB' h_c^d$

(iii) $s_x^2 \Rightarrow d(1)^2 h_c^d{}' \Omega h_c^d$

where $\hat{u}_t$ *are the residuals from the cointegration regression (3.2) and* $s_x^2 = T^{-1}\Sigma\hat{x}_t^2$ *is the estimated variance of the residuals of the ADF regression.* $h_c^d{}' = \begin{bmatrix} -\left(\int B_1^{d\,2}\right)^{-1}\left(\int B_1^d K_{2c}^d\right) & 1 \end{bmatrix}$, $A_c^d = \begin{bmatrix} \int B_1^{d\,2} & \int B_1^d K_{2c}^d \\ \int B_1^d K_{2c}^d & \int K_{2c}^{d\,2} \end{bmatrix}$, $B_c^d = \begin{bmatrix} B_1^d \\ K_{2c}^d \end{bmatrix}$ *and* $B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$.

Proof of Corollary A1:

(i) By OLS projections, we can write $\hat{u}_t = \hat{u}_t^d = y_t^d - x_t^d{}'\hat{b} = u_t^d - \left(\hat{b} - b\right)x_t^d{}'$

By Lemma A1 and CMT $T^{1/2}\hat{u}_t \Rightarrow K_{2c}^d - \left(\int B_1^{d\,2}\right)^{-1}\left(\int B_1^d K_{2c}^d\right)B_1^d{}'$ so

$T^{-2}\Sigma\,\hat{u}_t^2 \Rightarrow \int K_{2c}^{d\,2} - \left(\int B_1^{d\,2}\right)^{-1}\left(\int B_1^d K_{2c}^d\right)^2 = h_c^{d'} A_c^d h_c^d$

(ii) Follows from the same exact argument of Phillips and Ouliaris (1990).
The only difference is that now there is the extra piece $(r-1)u_{t-1}$. In fact we can write $\Delta\hat{u}_t = \Delta\hat{u}_t^d = \Delta u_t^d - \left(\hat{b} - b\right)\Delta x_t^d$. $u_t^d = u_t - \hat{g}_0 - \hat{g}_1 t$ and $x_t^d = x_t - \hat{m}_0 - \hat{m}_1 t$ where $\hat{g}_i$ and $\hat{m}_i$ are the OLS estimates from regressing u_t and x_t on a mean and trend (or mean only for $z_{1t} = 0$). Then $\Delta u_t^d = \Delta u_t - \hat{g}_1$ with $\hat{g}_1$ converging at rate $\sqrt{T}$ and $\Delta x_t^d = \Delta x_t - \hat{m}_1$ if detrended.

Substituting we have that

$$\Delta\hat{u}_t^d = \Delta\hat{u}_t = (r-1)u_{t-1} + v_{2t} - \left(\hat{b}-b\right)v_{1t} - \hat{g}_1 + \left(\hat{b}-b\right)\hat{m}_1 - \left(\hat{b}-b\right)m_1 .$$

Also $u_{t-1} = \hat{u}_{t-1} + \left(\hat{b}-b\right)x_{t-1}$ so

$$\Delta\hat{u}_t^d = (r-1)\hat{u}_{t-1} + v_{2t} - \left(\hat{b}-b\right)v_{1t} - \hat{g}_1 + \left(\hat{b}-b\right)(\hat{m}_1 - m_1) + \left(\hat{b}-b\right)(r-1)x_{t-1}$$

Write $v_{2t} - \left(\hat{b}-b\right)v_{1t} = b'v_t$. $\hat{b}'v_t \Rightarrow h'_c v_t = z_t$ says. Following the same exact argument in Phillips and Ouliaris (1990) p 183, I can write $w_t = d(L)z_t$ that is an absolute summable sequence. The variance of the orthogonal sequence x_t can the be written as $d(1)^2 h'_c \Omega h_c$. If the lag order in (3.3) is large enough to capture the correlation structure of the errors it can be shown that, conditionally on h_c,

$$\Delta\hat{u}_t^d = d(1)(r-1)\hat{u}_{t-1} - d(1)\hat{g}_1 + \left(\hat{b}-b\right)d(1)(\hat{m}_1 - m_1) + \left(\hat{b}-b\right)(r-1)d(L)x_{t-1} + \bar{d}(L)\Delta\hat{u}_t + w_t$$

Note that:

$T^{-1}\Sigma\Delta\hat{u}_t\,\Delta\hat{u}_{t-k}$ is $o_p(1)$, $d(1)T^{-1}\Sigma\hat{u}_{t-1}\hat{g}_1 = d(1)T^{-3/2}\Sigma\hat{u}_{t-1}(T^{1/2}\hat{g}_1) \Rightarrow 0$ since $T^{1/2}\hat{g}_1$ is $O_p(1)$.

$T^{-3/2}\Sigma\hat{u}_{t-1} \Rightarrow \int K_{2c}^d - \left(\int B_1^{d2}\right)^{-1}\left(\int B_1^d K_{2c}^d\right)\int B_1^{d}{}' = 0$ since $\int K_{2c}^d = 0$ and $\int B_1^{d}{}' = 0$.

Similarly $d(1)\left(\hat{b}-b\right)T^{-1}\Sigma\hat{u}_{t-1}(\hat{m}_1 - m_1) \Rightarrow 0$ since $\sqrt{T}(\hat{m}_1 - m_1)$ is $O_p(1)$.

Also $T^{-2}\Sigma\hat{u}_{t-1}x_{t-1} \Rightarrow \int B_1^d K_{2c}^d - \left(\int B_1^{d2}\right)^{-1}\left(\int B_1^d K_{2c}^d\right)\int B_1^{d2} = 0$ so

$$\left(\hat{b}-b\right)(r-1)T^{-1}\Sigma\hat{u}_{t-1}x_{t-1} = \left(\hat{b}-b\right)cT^{-2}\Sigma\hat{u}_{t-1}x_{t-1} \Rightarrow 0$$

Using the above results it easy to see that :

$$T^{-1}\Sigma\hat{u}_{t-1}^d\Delta\hat{u}_t^d = cT^{-2}\Sigma\hat{u}_{t-1}^2 + \hat{b}'T^{-1}\Sigma z_{t-1}^d w_t + o_p(1) \Rightarrow ch_c^{d'}A_c^d h_c^d + d(1)h_c^{d'}\int B_c^d dB' h_c^d .$$

(iii) If the log order p in the autoregressive part of the ADF test is large enough to capture the correlation structure of the errors (see Phillips and Ouliaris (1990) theorem 4.2), the estimates of the ADF regression will be consistent. ($\hat{a} \rightarrow r-1$).

$$\hat{x}_t = \Delta\hat{u}_t - \hat{a}\,\hat{u}_{t-1} - lags = -d(1)\hat{g}_1 + \left(\hat{b}-b\right)d(1)(\hat{m}_1 - m_1) + \left(\hat{b}-b\right)(r-1)d(L)x_{t-1} + w_t$$

$$s_x^2 = T^{-1}\Sigma\hat{x}_t^2 = T^{-1}\Sigma w_t^2 + o_p(1) \Rightarrow d(1)^2 h_c^{d}{}'\Omega h_c^d$$

Proof of Theorem 1:

The ADF test statistic is the usual t ratio test for the regression (3.3):

$$t_a = \frac{T^{-1}\hat{u}_{-1}^{'} M_u \Delta\hat{u}}{s_v \left[T^{-2}\hat{u}_{-1}^{'} M_u \hat{u}_{-1}\right]^{1/2}}$$

Under the condition for Lemma 1 and Lemma 2 ,

$$t_a = \frac{T^{-1}\Sigma\,\hat{u}_{t-1}\Delta\hat{u}_t + o_p(1)}{s_x\left[T^{-2}\Sigma\hat{u}_{t-1}^2\right]^{1/2} + o_p(1)}$$

By Corollary A1,

$$t_a \Rightarrow c\frac{d(1)\left[h_c^d{}' A_c^d h_c^d\right]^{1/2}}{d(1)\left[h_c^d{}'\Omega h_c^d\right]^{1/2}} + \frac{d(1) h_c^d{}' \int B_c^d dB' h_c^d}{\left[d(1)^2 h_c^d{}'\Omega h_c^d\right]^{1/2}\left[h_c^d{}' A_c^d h_c^d\right]^{1/2}}$$

Define $\Gamma = \begin{bmatrix} \Omega_{11}^{1/2} & 0 \\ 0 & \Omega_{22}^{1/2} \end{bmatrix}$. Notice that $B_c = \Gamma W_c$ and $B = \Gamma W_d$ and where $W_c = \begin{bmatrix} W_1 \\ J_{12c} \end{bmatrix}$ and $W_d = \begin{bmatrix} W_1 \\ W_{12} \end{bmatrix}$.

$$h_c^d{}' A_c^d h_c^d = h_c^d{}' \int B_c^d B_c^d{}' h_c^d = h_c^d{}'\Gamma \int W_c^d W_c^d{}' \Gamma' h_c^d = \Omega_{22} h^d{}' \int W_c^d W_c^d{}' h^d$$

with $h^{d'} = \left[-\left(\int W_1^{d\,2}\right)^{-1}\left(\int W_1^d J_{12c}^d\right) \quad 1 \right]$

Similarly $h_c^d{}' \int B_c^d dB' h_c^d = \Omega_{22} h^d{}' \int W_c^d dW' h^d$.

Finally $\left(h_c^d{}'\Omega h_c^d\right)^{1/2} = \left(h_c^d{}'\Omega^{1/2}\Omega^{1/2'} h_c^d\right)$

Since $h_c^d{}'\Omega^{1/2} = \left[-\left(\int B_1^{d\,2}\right)^{-1}\left(\int B_1^d K_{2c}^d\right) \quad 1 \right] \begin{bmatrix} \Omega_{11}^{1/2} & 0 \\ \Omega_{21}\Omega_{11}^{-1/2} & \Omega_{2.1}^{1/2} \end{bmatrix} =$

$$\Omega_{22}^{1/2}\left[-\left(\int W_1^{d\,2}\right)^{-1}\left(\int W_1^d J_{12c}^d\right) + d \qquad \left(1-d^2\right)^{1/2} \right]$$

Substituting these expressions in the tests, all the terms involving the variances simplifies, and only d appears in the asymptotic power as stated in Theorem 1.

Note that when $c = 0$, B_c and A_c in our notation coincide with B and A matrices in Phillips and Ouliaris (1990) and the asymptotic distribution of the tests coincides with Theorem 4.2 in Phillips and Ouliaris.

Proof of Theorem 2:

Following the same methodology of Sims, Stock and Watson (1990) rewrite equation (3.5) as

$\Delta y_t = z_t' g + e_{2t}$ where $z'_t = \begin{bmatrix} u_{t-1} & \Delta x_t & \Delta x_{t-1} & \dots & \Delta x_{t-p} & \Delta y_{t-1} & \dots & \Delta y_{t-p} \end{bmatrix}$ and $g = \begin{bmatrix} a & j & p_{11} & \cdots & p_{1p} & p_{21} & \cdots & p_{2p} \end{bmatrix}$.

Let Υ denotes the $(2p+2)\times(2p+2)$ diagonal matrix

$$\Upsilon = \begin{bmatrix} T & 0 & 0 & 0 \\ 0 & \sqrt{T} & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{T} \end{bmatrix}$$

With this notation it is easy to show the asymptotic distribution of the LS estimator of equation (3.5) $\hat{g}$.

We can write

(C.1) $\Upsilon(\hat{g} - g) = \left(\Upsilon^{-1} \Sigma z_t z_t' \Upsilon^{-1}\right)^{-1} \left(\Upsilon^{-1} \Sigma z_t x_{2t}\right)$ where

$$\left(\Upsilon^{-1} \Sigma z_t z_t' \Upsilon^{-1}\right) = \begin{bmatrix} T^{-2}\Sigma u_{t-1}^2 & T^{-3/2}\Sigma u_{t-1}\Delta x_t & T^{-3/2}\Sigma u_{t-1}\Delta x_{t-1} & \cdots & T^{-3/2}\Sigma u_{t-1}\Delta y_{t-p} \\ T^{-3/2}\Sigma u_{t-1}\Delta x_t & T^{-1}\Sigma\Delta x_t^2 & T^{-1}\Sigma\Delta x_t\Delta x_{t-1} & \cdots & T^{-1}\Sigma\Delta x_t\Delta y_{t-p} \\ T^{-3/2}\Sigma u_{t-1}\Delta x_{t-1} & T^{-1}\Sigma\Delta x_t\Delta x_{t-1} & T^{-1}\Sigma\Delta x_{t-1}^2 & \cdots & T^{-1}\Sigma\Delta x_{t-1}\Delta y_{t-p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T^{-3/2}\Sigma u_{t-1}\Delta y_{t-p} & T^{-1}\Sigma\Delta x_t\Delta y_{t-p} & T^{-1}\Sigma\Delta x_{t-1}\Delta y_{t-p} & \cdots & T^{-1}\Sigma\Delta y_{t-p}^2 \end{bmatrix}$$

The convergence of this matrix can be analyzed piece by piece.

$T^{-2}\Sigma u_{t-1}^2 \Rightarrow \int K_{2c}^2$ by Lemma 1.

$T^{-3/2}\Sigma u_{t-1}\Delta x_t \to 0$ since $T^{-1}\Sigma u_{t-1}\Delta x_t = T^{-1}\Sigma u_{t-1} v_{1t} \Rightarrow \int K_{2c} dB_1 + \Sigma\Gamma_{1,i}$ by Phillips(1988) where $\Gamma_{1,i} = E\left(e_{11}e_{1i}\right)$. Similarly we can show that $T^{-3/2}\Sigma u_{t-1}\Delta y_{t-p} \to 0$ so that we have the usual block diagonality between stationary and non-stationary variables.

For the stationary block we have that $T^{-1}\Sigma\Delta x_{t-1}^2 = T^{-1}\Sigma v_{2,t-1}^2 \to \Omega_{11}$

$T^{-1}\Sigma\Delta x_t\Delta x_{t-1} = T^{-1}\Sigma v_{1t} v_{1t-1} \to \Gamma_{1,1}$

$T^{-1}\Sigma\Delta x_t\Delta y_{t-p} = T^{-1}\Sigma v_{1t} v_{2t-p} \to \Gamma_{12,p}$ where $\Gamma_{12,i} = E\left(v_{11}v_{2i}\right)$

$T^{-1}\Sigma\Delta y_{t-p}^2 \to \Omega_{22}$

and $T^{-1}\Sigma\Delta x_{t-1}\Delta y_{t-p} \to \Gamma_{12,p-1}$ so that

$$\left(\Upsilon^{-1}\,\Sigma\, z_t z_t{}'\,\Upsilon^{-1}\right)\Rightarrow\begin{bmatrix}\int K_{2c}^2 & 0 & 0 & \cdots & 0\\ 0 & \Omega_{11} & \Gamma_{1,1} & \cdots & \Gamma_{12,p}\\ 0 & \Gamma_{1,1} & \Omega_{11} & \cdots & \Gamma_{12,p-1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & \Gamma_{12,p} & \Gamma_{12,p-1} & \cdots & \Omega_{22}\end{bmatrix}$$

Finally we have that

$$\left(\Upsilon^{-1}\,\Sigma\, z_t e_{2t}\right)=\begin{bmatrix}T^{-1}\Sigma u_{t-1}x_{2t}\\ T^{-1}\Sigma\Delta x_t x_{2t}\\ T^{-1}\Sigma\Delta x_{t-1}x_{2t}\\ \vdots\\ T^{-1}\Sigma\Delta y_{t-p}x_{2t}\end{bmatrix}\Rightarrow\begin{bmatrix}q\,\Omega_{2.1}^{1/2}\int K_{2c}dW_2\\ Z_i\\ Z_{11}\\ \vdots\\ Z_{2p}\end{bmatrix}$$

The convergence of the first element comes directly from Lemma1, while the convergence of the other elements of the vector to (multivariate) normal distributions comes form the stationarity of Δx_t and Δy_t.

Given the block diagonality the asymptotic distribution of the parameter of interest is

$$T(\hat{a}-a)\Rightarrow q\,\Omega_{2.1}^{1/2}\left(\int K_{2c}^2\right)^{-1}\left(\int K_{2c}dW_2\right)$$

If enough lags are included in the regression to get rid of the serial correlation (p is assumed to be known) the standard error for $\hat{a}$ converges to $q\,\Omega_{2.1}^{1/2}\left(\int K_{2c}^2\right)^{-1/2}$ and the local asymptotic distribution of the t-stat for a is:

$$t_{\hat{a}}=\frac{T(r-1)q}{q\,\Omega_{2.1}^{1/2}\left(\int K_{2c}^2\right)^{-1/2}}+\frac{q\,\Omega_{2.1}^{1/2}\left(\int K_{2c}^2\right)^{-1}\left(\int K_{2c}dW_2\right)}{q\,\Omega_{2.1}^{1/2}\left(\int K_{2c}^2\right)^{-1/2}}$$

Since how shown in Lemma2, $K_{2c}=\Omega_{22}^{1/2}J_{12c}$ and $B_1=\Omega_{11}^{1/2}W_1$ and $\Omega_{2.1}^{1/2}=\Omega_{22}^{1/2}\left(1-d^2\right)^{1/2}$

the t-stat simplifies to $t_{\hat{a}}=\dfrac{c}{\left(1-d^2\right)^{1/2}\left(\int J_{12c}^2\right)^{-1/2}}+\dfrac{\left(\int J_{12c}^2\right)^{-1}\left(\int J_{12c}dW_2\right)}{\left(\int J_{12c}^2\right)^{-1/2}}$

Proof of Theorem 3:

To compute the local asymptotic power of the ECR test we can follow the same approach used to prove Theorem 1 and Theorem 2. Equation (3.6) can be written as

(C.2) $\quad \Delta y_t=m^*+a\,u_{t-1}+y\,x_{t-1}+j\,\Delta x_t+\sum_i^p p_{1i}\Delta x_{t-i}+\sum_i^p p_{2i}\Delta y_{t-i}+x_{2t}$

so that the model is again in the same framework of Sims, Stock and Watson(1990) with 2 non-stationary and 2p+1 stationary variables. a is defined as in (3.5) such that with q defined in note 1. The coefficient y on the redundant regressor is truly zero. Since the mean has been moved back in the error correction term, the true value for m^* is also zero.

The model can be written in compact form as $\Delta y_t = m^* + z_t' g + x_{2t}$. As for Theorem 1 $\hat{g}$ can be estimated by first detrending both the right and the left side of the equation. $\Delta y_t^d = z_t^d{}' g + x_{2t}^d$ where

$z'^d_t = \begin{bmatrix} u_{t-1}^d & x_{t-1}^d & \Delta x_t^d & \Delta x_{t-1}^d & \dots & \Delta x_{t-1}^d & \Delta y_{t-1}^d & \dots & \Delta y_{t-1}^d \end{bmatrix}$,

$g' = \begin{bmatrix} a & y & p_{11} & \cdots & p_{1p} & p_{21} & \cdots & p_{2p} \end{bmatrix}$ and

Υ is now a $(2p+4)\times(2p+4)$ diagonal matrix

$$\Upsilon = \begin{bmatrix} T & 0 & 0 & 0 & 0 \\ 0 & T & 0 & 0 & 0 \\ 0 & 0 & \sqrt{T} & 0 & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{T} \end{bmatrix}$$

$$\left(\Upsilon^{-1}\Sigma z_t^d z_t^d{}'\Upsilon^{-1}\right) = \begin{bmatrix} T^{-2}\Sigma u_{t-1}^{d^2} & T^{-2}\Sigma u_{t-1}^d x_{t-1}^d & T^{-3/2}\Sigma u_{t-1}^d \Delta x_t^d & \cdots & T^{-3/2}\Sigma u_{t-1}^d \Delta y_{t-p}^d \\ T^{-2}\Sigma u_{t-1}^d x_{t-1}^d & T^{-2}\Sigma x_{t-1}^{d^2} & T^{-3/2}\Sigma x_{t-1}^d \Delta x_t^d & \cdots & T^{-1}\Sigma x_{t-1}^d \Delta y_{t-p}^d \\ T^{-3/2}\Sigma u_{t-1}^d \Delta x_t^d & T^{-3/2}\Sigma x_{t-1}^d \Delta x_t^d & T^{-1}\Sigma \Delta x_t^{d^2} & \cdots & T^{-1}\Sigma \Delta x_t^d \Delta y_{t-p}^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T^{-3/2}\Sigma u_{t-1}^d \Delta y_{t-p}^d & T^{-1}\Sigma x_{t-1}^d \Delta y_{t-p}^d & T^{-1}\Sigma \Delta x_t^d \Delta y_{t-p}^d & \cdots & T^{-1}\Sigma \Delta y_{t-p}^2 \end{bmatrix}$$

Following exactly the same idea of Theorem 2 we can look at the asymptotic distribution of the LS estimates piece by piece.
The only extra pieces are now:
$T^{-2}\Sigma u_{t-1}^d x_{t-1}^d \Rightarrow \int B_1^d K_{2c}^d$, $T^{-2}\Sigma x_{t-1}^{d^2} \Rightarrow \int B_1^{d2}$.

Moreover, $T^{-3/2}\Sigma x_{t-1}^d \Delta x_t^d = T^{-3/2}\Sigma x_{t-1}^d v_{1t} - T^{-3/2}\Sigma x_{t-1}^d \hat{m}_1 = o_p(1)$ since $T^{-1}\Sigma x_{t-1}^d v_{1t} \Rightarrow \int B_1^d dB_1 + \Sigma\Gamma_{1,i}$ and $T^{-1}\Sigma\Delta y_{t-i}$ and $T^{-1}\Sigma\Delta x_{t-i}$ (for $i = 0,1,\dots,p$) are $o_p(1)$ since are all stationary variables. All other pieces behave as in Theorem 1 with the exception that now are all detrended BM.

From model (3.1) we can write $v_t = \Omega^{1/2} h_t$ and $e_t = \Sigma^{1/2} h_t$ where h_t is a vector of uncorrelated martingale differences errors with unit variances. Given that $\Omega = \Phi(1)^{-1}\Sigma\Phi(1)^{-1}$, I can write $\Sigma^{1/2} = \Phi(1)\Omega^{1/2}$. Because of the way the errors are orthogonalized in equation (3.5), it can be shown that $x_{2t} = q\,\Omega_{2.1}^{1/2} h_{2t}$. By Chan and Wei (1988) and Phillips (1987) then follows that $T^{-1}\Sigma\, x_{t-1}^d x_{2t} \Rightarrow q\,\Omega_{2.1}^{1/2}\int B_1^d\, dW_2$, $T^{-1}\Sigma\, u_{t-1}^d x_{2t} \Rightarrow q\,\Omega_{2.1}^{1/2}\int K_{2c}^d dW_2$.

The matrix of cross product then weakly converges to:

$$\left(\Upsilon^{-1}\Sigma\, z_t^d z_t^{d}{}'\Upsilon^{-1}\right) \Rightarrow \begin{bmatrix} \int K_{2c}^{d\,2} & \int B_1^d K_{2c}^d & 0 & 0 & \cdots & 0 \\ \int B_1^d K_{2c}^d & \int B_1^{d\,2} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \Omega_{11} & \Gamma_{1,1} & \cdots & \Gamma_{12,p} \\ 0 & 0 & \Gamma_{1,1} & \Omega_{11} & \cdots & \Gamma_{12,p-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \Gamma_{12,p} & \Gamma_{12,p-1} & \cdots & \Omega_{22} \end{bmatrix}$$

Similarly

$$\left(\Upsilon^{-1}\Sigma\, z_t^d x_{2t}\right) = \begin{bmatrix} T^{-1}\Sigma u_{t-1}^d x_{2t} \\ T^{-1}\Sigma\, x_{t-1}^d x_{2t} \\ T^{-1/2}\Sigma\Delta x_t^d x_{2t} \\ T^{-1/2}\Sigma\Delta x_{t-1}^d x_{2t} \\ \vdots \\ T^{-1/2}\Sigma\Delta y_{t-p}^d e_{2t} \end{bmatrix} \Rightarrow \begin{bmatrix} q\,\Omega_{2.1}^{1/2}\int K_{2c}^d dW_2 \\ q\,\Omega_{2.1}^{1/2}\int B_1^d\, dW_2 \\ Z \\ Z_{11} \\ \vdots \\ Z_{2p} \end{bmatrix}$$

The inverse of the (2×2) non-stationary block is (by the inverted partitioned formula)

$$\frac{1}{\Delta^d}\begin{bmatrix} \int B_1^{d\,2} & -\int K_{2c}^d B_1^d \\ -\int K_{2c}^d B_1^d & \int K_{2c}^{d\,2} \end{bmatrix} \text{ where } \Delta^d = \int K_{2c}^{d\,2}\int B_1^{d\,2} - \left(\int B_1^d K_{2c}^d\right)^2$$

The asymptotic distribution of $\hat{a}$ is:

$$T(\hat{a} - a) \Rightarrow \frac{q\,\Omega_{2.1}^{1/2}\left[\int B_1^{d\,2}\int K_{2c}^d dW_2 - \int K_{2c}^d B_1^d \int B_1^d dW_2\right]}{\int B_1^{d\,2}\int K_{2c}^{d\,2} - \left(\int B_1^d K_{2c}^d\right)^2}$$

Now $SE(\hat{a}) \Rightarrow \dfrac{q\,\Omega_{2.1}^{1/2}\left[\int B_1^d\right]^{1/2}}{\left[\int B_1^{d\,2}\int K_{2c}^{d\,2} - \left(\int B_1^d K_{2c}^d\right)^2\right]^{1/2}}$

Notice that $t_{\hat{a}} = \dfrac{Ta}{SE(\hat{a})} + \dfrac{T(\hat{a}-a)}{SE(\hat{a})}$ with $Ta = T\left(\dfrac{c}{T}\right)q = cq$

So

$$t_{\hat{a}} \Rightarrow c\frac{q\left[\int B_1^{d\,2}\int K_{2c}^{d\,2} - \left(\int B_1^d K_{2c}^d\right)^2\right]^{1/2}}{q\,\Omega_{2.1}^{1/2}\left[\int B_1^d\right]^{1/2}} + \frac{q\,\Omega_{2.1}^{1/2}\left[\int B_1^{d\,2}\int K_{2c}^d dW_2 - \int K_{2c}^d B_1^d \int B_1^d dW_2\right]}{q\,\Omega_{2.1}^{1/2}\left[\int B_1^{d\,2}\int K_{2c}^{d\,2} - \left(\int B_1^d K_{2c}^d\right)^2\right]^{1/2}\left[\int B_1^d\right]^{1/2}}$$

Since how shown in Lemma2, $K_{2c} = \Omega_{22}^{1/2} J_{12c}$ and $B_1 = \Omega_{11}^{1/2} W_1$ and $\Omega_{2.1}^{1/2} = \Omega_{22}^{1/2}\left(1-d^2\right)^{1/2}$ the test simplifies to

$$t_{\hat{a}} \Rightarrow c\frac{\left[\int W_1^{d\,2}\int J_{12c}^{d\,2} - \left(\int W_1^d J_{12c}^d\right)^2\right]^{1/2}}{\left(1-d^2\right)^{1/2}\left[\int W_1^{d\,2}\right]^{1/2}} + \frac{\left[\int W_1^{d\,2}\int J_{12c}^d dW_2 - \int J_{12c}^d W_1^d \int W_1^d dW_2\right]}{\left[\int W_1^{d\,2}\int J_{12c}^{d\,2} - \left(\int W_1 J_{12c}^d\right)^2\right]^{1/2}\left[\int W_1^{d\,2}\right]^{1/2}}$$

Proof of Theorem 4:

Using the same parametrization of Theorem 3, the model can be written in compact form as $\Delta y_t^d = z_t^{d\,\prime} g + x_{2t}^d$ where z_t^d and g are defined previously.

The Wald test for the joint significance of a and f is equivalent to testing for the joint significance of a and y in (C.2) and it can be written as

$$F = \left(R\Upsilon\hat{g}\right)'\left[R\hat{V}R'\right]^{-1}\left(R\Upsilon\hat{g}\right)$$

where q is the number of restrictions, R is $2\times(3+2p)$ matrix defines as $R = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \end{bmatrix}$, Υ is the same as Theorem 3 and $\hat{V} = \hat{s}_{x_2}^2\left(\Upsilon^{-1}\Sigma z_t^d z_t^{d\,\prime}\Upsilon^{-1}\right)^{-1}$ is the standard error of the LS estimates. For the purpose of this proof the F test can be decomposed as:

(C.3) $$\left(R\Upsilon\hat{g}\right)'\left[R\hat{V}R'\right]^{-1}\left(R\Upsilon\hat{g}\right)=\left(R\Upsilon\left(\hat{g}-g\right)\right)'\left[R\hat{V}R'\right]^{-1}\left(R\Upsilon\left(\hat{g}-g\right)\right)+$$

$$+\left(R\Upsilon g\right)'\left[R\hat{V}R'\right]^{-1}\left(R\Upsilon\left(\hat{g}-g\right)\right)+\left(R\Upsilon\left(\hat{g}-g\right)\right)'\left[R\hat{V}R'\right]^{-1}\left(R\Upsilon g\right)+$$
$$+\left(R\Upsilon g\right)'\left[R\hat{V}R'\right]^{-1}\left(R\Upsilon g\right)$$

Looking at the asymptotic behavior piece by piece (recall that $\hat{s}_{x_t}=q\,\Omega_{2.1}^{1/2}$):

$$\left[R\hat{V}R'\right]^{-1}\Rightarrow q^{-2}\,\Omega_{2.1}^{-1}\Delta^{d}\begin{bmatrix}\int B_1^{d\,2} & -\int K_{2c}^{d}B_1^{d}\\ -\int K_{2c}^{d}B_1^{d} & \int K_{2c}^{d\,2}\end{bmatrix}^{-1}=q^{-2}\,\Omega_{2.1}^{-1}\begin{bmatrix}\int K_{2c}^{d\,2} & \int K_{2c}^{d}B_1^{d}\\ \int K_{2c}^{d}B_1^{d} & \int B_1^{d\,2}\end{bmatrix}$$

$$R\Upsilon g=\begin{bmatrix}Ta\\ Ty\end{bmatrix}=\begin{bmatrix}cq\\ 0\end{bmatrix}$$

$$R\Upsilon\left(\hat{g}-g\right)\Rightarrow\frac{q\,\Omega_{2.1}^{1/2}}{\Delta^{d}}\begin{bmatrix}\int B_1^{d\,2}\int K_{2c}^{d}dW_2-\int K_{2c}^{d}B_1^{d}\int B_1^{d}dW_2\\ \int K_{2c}^{d\,2}\int B_1^{d}dW_2-\int K_{2c}^{d}B_1^{d}\int K_{2c}^{d}dW_2\end{bmatrix}$$

where, as defined earlier, $\Delta^{d}=\int K_{2c}^{d\,2}\int B_1^{d\,2}-\left(\int B_1^{d}K_{2c}^{d}\right)^{2}$ is the determinant of the submatrix corresponding to the non-stationary block.

For the first element of the (C.3) note that

$$\left[R\hat{V}R'\right]^{-1}\left(R\Upsilon\left(\hat{g}-g\right)\right)\Rightarrow\frac{q^{-2}\,\Omega_{2.1}^{-1}q\,\Omega_{2.1}^{1/2}}{\Delta^{d}}\begin{bmatrix}\int K_{2c}^{d\,2}\int B_1^{d\,2}\int K_{2c}^{d}dW_2-\left(\int K_{2c}^{d}B_1^{d}\right)^{2}\int K_{2c}^{d}dW_2\\ -\left(\int K_{2c}^{d}B_1^{d}\right)^{2}\int B_1^{d}dW_2+\int K_{2c}^{d\,2}\int B_1^{d\,2}\int B_1^{d}dW_2\end{bmatrix}=$$
$$=q^{-1}\Omega_{2.1}^{-1/2}\begin{bmatrix}\int K_{2c}^{d}dW_2\\ \int B_1^{d}dW_2\end{bmatrix}$$

So

$$\left(R\Upsilon\left(\hat{g}-g\right)\right)'\left[R\hat{V}R'\right]^{-1}\left(R\Upsilon\left(\hat{g}-g\right)\right)\Rightarrow$$

$$\Rightarrow \frac{q\,\Omega_{2.1}^{1/2}}{\Delta^d}\left[\int B_1^{d\,2}\int K_{2c}^d dW_2 - \int K_{2c}^d B_1^d \int B_1^d dW_2 \quad \int K_{2c}^{d\,2}\int B_1^d dW_2 - \int K_{2c}^d B_1^d \int K_{2c}^d dW_2\right] q^{-1}\,\Omega_{2.1}^{-1/2}\begin{bmatrix}\int K_{2c} dW_2\\ \int B_1 dW_2\end{bmatrix} =$$

$$= \frac{\int B_1^{d\,2}\left(\int K_{2c}^d dW_2\right)^2 - 2\int B_1^d K_{2c}^d \int B_1^d dW_2 \int K_{2c}^d dW_2 + \int K_{2c}^{d\,2}\left(\int B_1^d dW_2\right)^2}{\int B_1^{d\,2}\int K_{2c}^{d\,2} - \left(\int B_1^d K_{2c}^d\right)^2}$$

$$\left(R\Upsilon g\right)'\left[R\hat{V}R'\right]^{-1}\left(R\Upsilon(\hat{g}-g)\right) \Rightarrow$$

$$\Rightarrow \left[cq \quad 0\right] q^{-1}\,\Omega_{2.1}^{-1/2}\begin{bmatrix}\int K_{2c}^d dW_2\\ \int B_1^d dW_2\end{bmatrix} = c\,\Omega_{2.1}^{-1/2}\int K_{2c}^d dW_2$$

The third element is the transpose of the previous one and finally the fourth element on the right hand side of (C.3) converges to $c^2\Omega_{2.1}^{-1}\int K_{2c}^2$.

Remember that $K_{2c} = \Omega_{22}^{1/2} J_{12c}$, $B_1 = \Omega_{11}^{1/2} W_1$ and $B_2 = \Omega_{22}^{1/2} W_{12}$ the local asymptotic power of the Wald test simplifies to:

$$F \Rightarrow c^2\left(1-d^2\right)^{-1}\int J_{12c}^{d\,2} + 2c\left(1-d^2\right)^{-1/2}\int J_{12c}^d dW_2 +$$

$$+ \frac{\int W_1^{d\,2}\left(\int J_{12c}^d dW_2\right)^2 - 2\int W_1^d J_{12c}^d \int W_1^d dW_2 \int J_{12c}^d dW_2 + \int J_{12c}^{d\,2}\left(\int W_1^d dW_2\right)^2}{\int W_1^{d\,2}\int J_{12c}^{d\,2} - \left(\int W_1^d J_{12c}^d\right)^2}$$

Proof of Theorem 5:

The proof of Theorem 5, follows Johansen (1995). I give here a less formal proof than Johansen's. The proof I present is only for the case in which there is a mean in the cointegrating vector. For a more general and formal proof for the asymptotic distribution and local power of the trace statistics of which, the $l_{\max}$ test is just a special case, see Johansen (1995) or Saikonnen and Lutkepohl (1999, 2000).

As I showed in Section 3.3, the $l_{\max}$ test is defined as $l_{\max} = -T\ln(1-\hat{l}_1)$. Johansen (1995) shows that $-T\ln(1-\hat{l}_1) = T\hat{l}_1 + o_P(1)$. Since $\hat{l}_1$ is the maximum eigenvalue of the matrix $S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$ the test can be written as $l_{\max} = \max eig\left\{TS_{11}^{-1}S_{10}\,S_{00}^{-1}S_{01}\right\} + o_P(1)$. The asymptotic distribution of the tests can be computed by looking at the limit of each single matrix. For the case in which $z_{1t}=0$ and $z_{2t}=1$, so that there is mean only in the cointegrating vector, the Error Correction model is:

$$\begin{bmatrix}\Delta x_t\\ \Delta y_t\end{bmatrix}=P\Phi(1)M\begin{bmatrix}x_{t-1}\\ y_{t-1}\\ z_{2t-1}\end{bmatrix}+\Pi(L)\begin{bmatrix}\Delta x_{t-1}\\ \Delta y_{t-1}\end{bmatrix}+Pe_t$$

For theoretical purposes I can re-parametrize the model as (see also Saikonnen and Lutkepohl (1999)).

$$\begin{bmatrix}\Delta x_t\\ \Delta y_t\end{bmatrix}=P\Phi(1)M^*\begin{bmatrix}x_{t-1}\\ y_{t-1}\\ \sqrt{T}\,z_{t-1}\end{bmatrix}+\Pi(L)\begin{bmatrix}\Delta x_{t-1}\\ \Delta y_{t-1}\end{bmatrix}+Pe_t$$

where $M^*=\begin{bmatrix}0 & 0 & 0\\ -b(r-1) & (r-1) & -m(r-1)T^{-1/2}\end{bmatrix}$ and $Y_t^*=\begin{bmatrix}x_t & y_t & \sqrt{T}\,z_t\end{bmatrix}'$

Of course this is an unfeasible regression but it such that the eigenvalues of the tests can also be obtained from this regression.
As Johansen (1995)
shows $S_{00}\xrightarrow{p}VAR\left[\Delta Y_t|Y_{t-1,}past\right]=VAR\left[Pe\right]=P\Sigma P'=P\Phi(1)\Omega\Phi(1)'P'$

$T^{-1}S_{11}\Rightarrow T^{-2}\Sigma\left(Y_{t-1}^*Y_{t-1}^{*'}\right)+o_p(1)\Rightarrow\int\tilde{B}_c\tilde{B}_c'$ where $\tilde{B}_c=\begin{bmatrix}B_1\\ bB_1+K_{2c}\\ 1\end{bmatrix}$.

Recall that $K_{2c}=\Omega_{22}^{1/2}J_{12c}$, $B_1=\Omega_{11}^{1/2}W_1$ and define $\Gamma=\begin{bmatrix}\Omega_{11}^{1/2} & 0 & 0\\ 0 & \Omega_{22}^{1/2} & 0\\ 0 & 0 & 1\end{bmatrix}$ and

$\overline{P}=\begin{bmatrix}1 & 0 & 0\\ b & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$ so $T^{-1}S_{11}$ can be written as $T^{-1}S_{11}\Rightarrow\overline{P}\Gamma\int W_cW_c'\Gamma'\,\overline{P}'$.

Finally $S_{10}=S_{11}\Pi_T'+S_{1\overline{e}}+o_p(1)$ where $S_{1e}=T^{-1}\Sigma\left(Y_{t-1}^*\overline{e}_t'\right)+o_p(1)$ where $\overline{e}_t=Pe_t$.
From Lemma 1 we have that
$S_{1e}\Rightarrow\int\tilde{B}_c d\overline{B}'$ where $\overline{B}=P\Sigma^{1/2}W$ and , as before, $\tilde{B}_c=\overline{P}\Gamma W_c$.
$S_{11}\Pi_T'=S_{11}M^{*'}\Phi(1)'P'$. Under the local alternative $(r-1)=c/T$ so

$$S_{11}\Pi_T'=cT^{-1}S_{11}\begin{bmatrix}0 & 0 & 0\\ -b & 1 & -m/\sqrt{T}\end{bmatrix}\Phi(1)'P'\Rightarrow c\overline{P}\Gamma\int W_cW_c'\Gamma'\,\overline{P}'\,\overline{M}^*\Phi(1)'P'$$

where $\bar{M}^* = \begin{bmatrix} 0 & 0 & 0 \\ -b & 1 & 0 \end{bmatrix}$. The limit of the tests can then be decomposed into the limit of 4 matrices:

$$TS_{11}^{-1}S_{10}S_{00}^{-1}S_{01} = TS_{11}^{-1}S_{11}\Pi_T^{'}S_{00}^{-1}S_{1e}^{'} + TS_{11}^{-1}S_{11}\Pi_T^{'}S_{00}^{-1}\Pi_T S_{11} + TS_{11}^{-1}S_{1e}S_{00}^{-1}S_{1e}^{'} + TS_{11}^{-1}S_{1e}S_{00}^{-1}\Pi_T S_{11}$$

For the first block:

$$TS_{11}^{-1}S_{11}\Pi_T^{'}S_{00}^{-1}S_{1e}^{'} \Rightarrow c\left(\bar{P}'\Gamma'\right)^{-1}\left[\int W_cW_c^{'}\right]^{-1}\Gamma^{-1}\bar{P}^{-1}\bar{P}\Gamma$$

$$\int W_cW_c^{'}\Gamma'\bar{P}'\bar{M}^*\Phi(1)'P'P'^{-1}\Phi(1)'^{-1}\Omega^{-1}\Phi(1)^{-1}P^{-1}\left[\int W_c dW\right]'\Gamma'\bar{P}'$$

Since $\Omega^{-1/2'} = \begin{bmatrix} \Omega_{11}^{1/2} & -\Omega_{2.1}^{-1/2}\Omega_{12}\Omega_{11}^{-1} \\ 0 & \Omega_{2.1}^{-1/2} \end{bmatrix}$

it is easy to show that $\Gamma'\bar{P}'\bar{M}^*\Omega^{-1/2'} = \begin{bmatrix} 0 & 0 \\ 0 & \Omega_{22}^{1/2}\Omega_{2.1}^{-1/2} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \left(1-d^2\right)^{-1/2} \\ 0 & 0 \end{bmatrix}$ and that

$$c\left[\int W_cW_c^{'}\right]\Gamma'\bar{P}'M^{*'}\Omega^{-1/2'} = c\left(1-d^2\right)^{-1/2}\begin{bmatrix} 0 & \int W_1 J_{12c} \\ 0 & \int J_{12c}^2 \\ 0 & \int J_{12c} \end{bmatrix} = A_c$$

The first block the simplifies to

$$S_{11}^{-1}S_{11}\Pi_T^{'}S_{00}^{-1}S_{1e}^{'} \Rightarrow \left(\bar{P}'\Gamma'\right)^{-1}\left[\int W_cW_c^{'}\right]^{-1}A_c\left[\int W_c dW'\right]'\Gamma'\bar{P}'$$

Similarly for the other blocks we obtain:

$$TS_{11}^{-1}S_{11}\Pi_T^{'}S_{00}^{-1}\Pi_T S_{11} \Rightarrow \left(\bar{P}'\Gamma'\right)^{-1}\left[\int W_cW_c^{'}\right]^{-1}A_cA_c^{'}\Gamma'\bar{P}'$$

$$TS_{11}^{-1}S_{1e}S_{00}^{-1}\Pi_T S_{11} \Rightarrow \left(\bar{P}'\Gamma'\right)^{-1}\left[\int W_cW_c^{'}\right]^{-1}\left[\int W_c dW'\right]A_c^{'}\Gamma'\bar{P}'$$

$$TS_{11}^{-1}S_{1e}S_{00}^{-1}S_{1e}^{'} \Rightarrow \left(\bar{P}'\Gamma'\right)^{-1}\left[\int W_cW_c^{'}\right]^{-1}\left[\int W_c dW'\right]\left[\int W_c dW'\right]'\Gamma'\bar{P}'$$

Finally, since $\max eig\{M\} = \max eig\left\{\left(\Gamma'\bar{P}'\right)M\left(\bar{P}'\Gamma'\right)^{-1}\right\}$ the theorem is proved.

The case with no mean can be proved similarly with

$M = \begin{bmatrix} 0 & 0 \\ -b(r-1) & (r-1) \end{bmatrix}$, $Y_t = [x_t \quad y_t]'$, $B_c = \begin{bmatrix} B_1 \\ bB_1 + K_{2c} \end{bmatrix}$, $\Gamma = \begin{bmatrix} \Omega_{11}^{1/2} & 0 \\ 0 & \Omega_{22}^{1/2} \end{bmatrix}$, and $\bar{P} = \begin{bmatrix} 1 & 0 \\ b & 1 \end{bmatrix}$.

The proof for the case with mean and trend follows exactly from Saikonnen and Lutkepohl (1999).

REFERNCES

Banerjee et al. (1986): " Exploring Equilibrium Relationship in Econometrics through static models: some Monte Carlo evidence", *Oxford Bulletin of Economics and Statistics,* 48, 3, 253-277.

Banerjee et al. (1993): *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data.* Oxford University Press.

Banerjee et al. (1998): " Error-Correction Mechanism Tests for Cointegration in a Single-Equation Framework", *Journal of Time Series Analysis,* 19, 3, 267:283.

Bewley R. and M.Yang (1998): "On the size and Power of system tests for cointegration", *The Review of Economics and Statistics*, 80(4), 675-679.

Boswijk, H. P. (1994): "Testing for an unstable root in conditional and structural error correction models", *Journal of Econometrics*, 63, 37-60.

Boswijk, H.P. and P.H. Frances (1992): "Dynamic specification and Cointegration", *Oxford Bulletin of Economics and Statistics*, 54, 3, 369:381.

Chan, N.H. and C.Z.Wei (1988): "Limiting distributions of the least Squares estimates on unstable auotregressive processes", *Annals of Statistics,* 16, 367-401.

Elliott G. (1998): "On the Robustness of Cointegration methods when regressors almost have a unit root", *Econometrica,* 66(1), 149-158.

Elliott, G., T.J.Rothemberg and J.H.Stock (1996): "Efficient tests for an autoregressive unti root", *Econometrica*, 64, 813-836.

Engle, R.F. and C.W.Granger (1987): "Cointegration and error correction: Representation, estimation and testing", *Econometrica,* 55,251-276.

Ericsson, N.R. and J.G. Mackinnon (1999): "Distributions of Error Correction tests for Cointegration", International Finance Discussion Paper (655), Board of Governors of the Federal Reserve System.

Friedman, B.M. and K.N. Kuttner (1992):"Money, Income, Prices and Interest Rates", *American Economic Review*, 82,472-492.

Gonzalo, J. and T-H Lee (1998): " Pitfalls in testing for long run relationships", *Journal of Econometrics*, 86, 129-154.

Granger, C. W. (1983): " Co-Integrated Variables and Error Correction Models", UCSD Discussion Paper 83-13.

Granger C. and P. Newbold (1974): " Spurious regressions in Econometrics", *Journal of Econometrics*, 2, 111-120.

Hafer, R.W. and D.W. Jansen (1991): "The demand for money in the United States: evidence from cointegration tests", *Journal of Money, Credit and Banking*, 23, 155-168.

Hamilton J. (1994): *Time Series Analysis*. Princeton University Press.

Hansen B. (1992): "Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends", *Journal of Econometrics*, 53, 87-121.

Hansen B. (1995): "Rethinking the Univariate Approach to Unit Root testing", *Econometric Theory*, 11, 1148-1171.

Haug, A.A. (1996): "Tests for cointegration. A Monte Carlo Comparison", *Journal of Econometrics*, 71, 89-115.

Hendry, D.F. (1987): "Econometric Methodology: A Personal Prospective", In Bewley, T. (ed.), *Advances in Econometrics*, vol.2. Cmabridge University Press.

Hoffman, D.L. and R.H. Rasche (1991): "Long-run Income and Interest elasticities on money demand in the United States", *Review of Economics and Statistics*, 73,665-674.

Johansen, S. (1988): "Statistical analysis of cointegration vectors", *Journal of Economic Dynamics and Control*, 12,231-254.

Johansen, S. (1991): "Estimation and hypothesis testing of cointegration vector in Gaussian vector autoregressive models", *Econometrica*, 59,1551-1580.

Johansen, S. (1995): *Likelihood Based Inference in Cointegrated autoregressive Models.* Oxford: Oxford University Press.

Johansen, S. and K. Juselius (1990): "Maximum likelihood estimation and inference on cointegration with application to the demand for money", *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

Kremers J.M. et al. (1992): “The power of cointegration tests”, *Oxford Bulletin of Economics and Statistics*, 54, 325-348.

Miller S.M. (1991):”Monetary Dynamic: an application of cointegration and error correction modeling”, *Journal of Money, Credit and Banking*, 23,139-154.

Miyao, R. (1996): “Does a cointegrating M2 demand relation really exist in the United States?”, *Journal of Money, Credit and Banking*, 28(3), 365-380.

Ostermark R. and R. Hoglund (1998) “Simulating competing cointegration tests in bivariate system”, Abo Akademi University WP.

Park J.Y. (1990): “Testing for unit roots and cointegration by variable addition”, *Advances in econometrics*, 8, 107-133.

Park J.Y. (1992): “Canonical Cointegration Regression”, *Econometrica*, 60(1), 119-144.

Pesavento E. (2000): “Residuals Based Tests fro Cointegration: An Analytical Comparison“, Mimeo.

Phillips, P.C.B. (1986): “ Understanding Spurious Regression in Econometrics”, *Journal of Econometrics*, 33, 311-340.

Phillips, P.C.B. (1987a): “ Toward and unified Asymptotic theory for autoregression”, *Biometrica*, 74, 535-547.

Phillips, P.C.B. (1987b): “Time Series Regression with a Unit Root”, *Econometrica*, 55, 277-301.

Phillips, P.C.B. (1988): “Regression Theory for Near-Integrated Time Series”, *Econometrica*, 56, 1021-1043.

Phillips, P.C.B. and B.E. Hansen (1990): “ Statistical Inference in Instrumental Variable Regression with I(1) Processes”, *Review of Economic Studies*.

Phillips, P.C.B. and S. Ouliaris (1990): “Asymptotic properties of residuals based tests for cointegration”, *Econometrica*, 58, 165-193.

Phillips, P.C.B. and V.Solo (1992): “Asymptotic for Linear Processes”, *The Annals Of Statistics*, 20(2), 971-1001.

Saikonnen P. (1992): “Estimation and testing of cointegrated systems by an autoregressive approximation”, *Econometric Theory*, 8, 1-27.

Saikonnen P. and Lutkephol H. (1999): “ Local Power of Likelihood Ratio Tests for the Cointegrating Rank of a VAR process”, *Econometric Theory*, 15, 50-78.

Sims C.A., Stock J.H. and M.W. Watson (1990): “Inference in Linear Time Series Models with some Unit Roots”, *Econometrica*, 58, 113-144.

Stock J.H. (1994): “Unit roots, structural breaks and trends”, in McFadden and R.F. Engle eds. Hanbook of Econometrics V4, North Holland, 2739-2841.

Stock J.H. and M.W.Watson (1988): “A simple estimator of cointegrating vectors in higher order integrated systems”, *Econometrica*, 61, 783-820.

Watson M.W. (1994): “ Vector Autoregression and Cointegration”, in *Handbook of Econometrics*, vol.4, ed. D. Mc Fadden and R.F. Engle. Amsterdam: North Holland, 2843-2915.

Wooldridge J. (1994): “ Estimation and Inference for Dependent Processes” in *Handbook of Econometrics*, vol.4, ed. D. Mc Fadden and R.F. Engle. Amsterdam: North Holland 2639-2738.

Zivot E. (2000): “ The Power of Single Equation Tests for Cointegration when the Cointegrating Vector is prespecified”, *Econometric Theory*, 16, 407-439.