Available online at www.sciencedirect.com
SCIENCE DIRECT®

JOURNAL OF Econometrics

ELSEVIER

Journal of Econometrics 122 (2004) 349–384

www.elsevier.com/locate/econbase

Analytical evaluation of the power of tests for the absence of cointegration

Elena Pesavento

Department of Economics, Emory University, Atlanta, GA 30322-2240, USA

Accepted 23 October 2003

Abstract

This paper proposes a theoretical explanation for the common empirical results in which different tests for cointegration give different answers. Using local to unity parametrization, this paper analytically computes the power of four 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 (EC) term, and Boswijk (1994) Wald test. The test statistics are shown to converge under a local alternative to random variables whose distributions are functions of Brownian Motions and Ornstein–Uhlenbeck processes and of a single nuisance parameter. The nuisance parameter is determined by the correlation at frequency zero of the errors in the cointegration relation with the shocks of the right-hand variables. I show that, when this correlation is high, system approaches, like the Johansen maximum eigenvalue or tests of the EC model, can exploit this correlation and significantly outperform single equation tests. Many of the varying results from applying different tests can be attributed to different values of this nuisance parameter.

JEL classification: C32

Keywords: Cointegration; Error correction; Local asymptotic power

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

E-mail address: epesave@emory.edu (E. Pesavento).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2003.10.025

that is stationary, then the variables are said to be cointegrated. Cointegrated variables will never move too far apart, and will be attracted 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 a variety of 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 (EC) Model introduced by Hendry (1987), Boswijk's (1994) Wald test, tests for the number of unit roots (Johansen, 1988; Stock and Watson, 1988; Horvath and Watson, 1995; Saikkonen, 1992), and tests on null hypotheses on the cointegrating vector (Saikkonen, 1992; Johansen, 1995; Elliott, 1998) among others (see Watson (1994) for a review). Currently, there is no consensus as to which is the best test for cointegration, and the general empirical approach is to report the results for a variety of tests.

This paper focuses on the class of tests that have no cointegration as the null hypothesis.[1] 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's (1988) test, tests on the coefficient of the EC terms among others), while a second group looks at single-equation regressions involving the variables that are potentially cointegrated (Engle and Granger's (1987) Augmented Dickey Fuller test, Phillips and Ouliaris' (1990) tests). Such tests are non-standard and their asymptotic distributions are non-normal (functions of Brownian motions). In this environment, there is no theory that suggests which test should have higher power, as the data do not give rise to locally asymptotic normal likelihood and the standard asymptotic theory is inapplicable.

Using local-to-unity parameterization, I compute the analytical asymptotic power for the most commonly used tests for the null of no cointegration to show which features of the model are important for power. In particular, I compare the Augmented Dickey Fuller test applied to the residuals from the cointegrating regression, Johansen's maximum eigenvalue test and three tests in the EC Model: the t-test on the EC 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). These results generalize and extend some of the previous work using a general and coherent framework that allows a comparison of the asymptotic power. The results show analytically that the asymptotic distributions of these tests depend on a single nuisance parameter under the local alternative. This parameter is a function of the long-run correlation of the errors in the cointegration relation to the shocks to the set of 'X' variables. Intuition suggests, when this correlation is high, we expect a full system approach to perform better by exploiting this correlation. Evaluations of the analytical power functions confirm the intuition and show that the tests have

[1] The complementary paper for the case of tests that take cointegration as the null hypothesis is Jansson and Haldrup (2002).

significantly different performances for different values of the nuisance parameter. The results suggest which test is best to use for the particular application at hand.

There are a number of practical reasons for being interested in the asymptotic distribution of these tests under a local alternative. First, although the asymptotic local power for some of the tests for no cointegration has been computed before (Banerjee et al., 1986; Kremers et al., 1992; Johansen, 1995; Zivot, 2000), different tests have been mostly compared on the basis of Monte Carlo analysis for a particular Data Generating Process (Haug, 1996; Gonzalo and Lee, 1998; Bewley and Yang, 1998; Boswijk and Frances, 1992; Kremers et al., 1992; Ericsson and Mackinnon, 1999). The problem with Monte Carlo simulations is that the results of the experiments are dependent on the particular design and no general conclusions are available. In general, these experiments are unable to find any ranking of the tests or to find important parameters that would allow such a ranking. The analytical derivation of the asymptotic distribution under a local alternative allows the isolation of a single parameter that is important for power, relates understanding to the results of previous Monte Carlo analysis, and assists in designing a more informative set of experiments. Second, classical tests for cointegration and no-cointegration can exhibit severe size distortions in small samples. One solution proposed in the recent literature is to apply bootstrap methods to such tests (Li and Maddala, 1997; Berkowitz and Kilian, 2000; Inoue and Kilian, 2002; Chang et al., 2002 among others). Evaluation of bootstrap methods in a non-stationary environment shows that bootstrapping provides substantial improvements to the finite sample size of tests. Although controlling the size of a test is extremely important, it is also relevant to evaluate whether the refinement obtained with bootstrap algorithms does not come at the cost of lower power. This paper presents a benchmark to evaluate the local power of bootstrap tests that address the relevant question of comparing the bootstrap power with the asymptotic power of classical tests (see also Swensen, 2003).

The next section introduces the model and discusses some of the tests that have been used in the literature to test for the absence of cointegration. In the third section, the asymptotic distribution of the tests under a local alternative is derived. The tests are then compared in terms of their local asymptotic power and their size-adjusted power in small samples.

2. The model

The model can be written as $z_t = \mu + \tau t + \zeta_t$, a sum of a deterministic and a stochastic part where t = 1, ..., T. The stochastic component can be written as

$$\Phi(L)[I(1-L) - (\rho - 1)\alpha\beta' L]\zeta_t = \varepsilon_t, \tag{2.1}$$

where L is the lag operator, $0 \leqslant \rho \leqslant 1$, $\alpha'\beta = 1$, $\Phi(L)$ is a polynomial of possibly infinite order with first element equal to the identity matrix (the usual normalization), and $|\Phi(r)| = 0$ has roots outside the unit circle.

The random variable z_t can be partitioned into $z_t' = [x_t' \ y_t]$, where x_t is an $n_1 \times 1$ vector, and y_t is a scalar. The error term $\varepsilon_t = [\varepsilon_{1t}' \ \varepsilon_{2t}]'$ is defined implicitly by Assumption 1, which is valid under a wide range of different assumptions on the process (see Phillips and Solo (1992) or Wooldridge (1994)).

Assumption 1. $\{\varepsilon_t\}$ satisfies a FCLT, i.e. $T^{-1/2}\sum_{t=1}^{[T\cdot]}\varepsilon_t \Rightarrow \Sigma^{1/2}W(\cdot)$ where $W(\lambda)$ is a standard $(n_1+1)\times 1$ Brownian motion and $\Rightarrow$ denotes weak convergence.

Partition $\mu=[\mu_1' \;\; \mu_2]'$ and $\tau=[\tau_1' \;\; \tau_2]'$. Without loss of generality, I can normalize α and β such that $\beta=[-\gamma' \;\; 1]'$, and $\alpha=[\alpha_1' \;\; \alpha_2]'$. Furthermore, I assume that the researcher knows that x_t individually have unit roots and are not cointegrated. As in Elliott et al. (2002), I will consider the rotation of the model through pre-multiplication of the matrix $K=\begin{bmatrix} I & 0 \\ -\gamma' & 1\end{bmatrix}$:

$$\Delta\begin{bmatrix} x_t \\ \beta' z_t\end{bmatrix}=\begin{bmatrix}\tau_1\\ \beta'\tau\end{bmatrix}-(\rho-1)\begin{bmatrix}\alpha_1\\ 1\end{bmatrix}\beta'(\mu+\tau(t-1))$$
$$+(\rho-1)\begin{bmatrix}\alpha_1\\ 1\end{bmatrix}\beta' z_{t-1}+\begin{bmatrix}v_{1t}\\ v_{2t}\end{bmatrix}, \tag{2.2}$$

where I have used the fact that $\alpha'\beta=1$. $v_t=[v_{1t}' \;\; v_{2t}]'=K\Phi(L)^{-1}\varepsilon_t$ is partitioned conformably with z_t.

Elliott et al. (2002) show that the assumption that x_t are not mutually cointegrated and have roots that are known a priori to be equal to one, corresponds to the assumption that $\alpha_1=0$.[2] Imposing the restriction that $\alpha_1=0$, along with $\alpha'\beta=1$, implies that $\alpha_2=1$. Model (2.2) can then be written as

$$\Delta x_t=\tau_1+v_{1t},$$
$$y_t=(\mu_2-\gamma'\mu_1)+(\tau_2-\gamma'\tau_1)t+\gamma' x_t+u_t,$$
$$u_t=\rho u_{t-1}+v_{2t}, \tag{2.3}$$

where $u_t=\beta'\zeta_t$. Model (2.3) is in the form of Phillips triangular model and has an intuitive interpretation. The first equation represents the dynamics of x_t, while the second equation represents the cointegrating relation. The third equation allows an easy interpretation of the role of ρ: when $\rho<1$, u_t is stationary, y_t and x_t are cointegrated and system (2.3) contains n_1 unit roots; when $\rho=1$, the two variables are not cointegrated and there are $n=n_1+1$ unit roots in the system. Thus, a test for no cointegration is testing $H_0:\rho=1$ vs. $H_a:\rho<1$.

Assumption 1 along with the assumptions on $\Phi(L)$ means that $T^{-1/2}\sum_{t=1}^{[T\cdot]}v_t \Rightarrow \Omega^{1/2}W(\cdot)$, where $\Omega=K\Phi(1)^{-1}\Sigma\Phi(1)^{-1'}K'$ is the spectral density at frequency zero of v_t scaled by 2π, and Σ is the variance covariance matrix of ε_t. Partition Ω and $\Phi(L)$ as

[2] Elliott et al. (2002) show this as just a ‘directional’ assumption that restricts the attention to tests for the set of relevant alternatives in which the cointegration vector is mean reverting and that fully utilize the information that x_t is $I(1)$.

$\Omega = \begin{bmatrix} \Omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{bmatrix}$ and $\Phi(L) = \begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix}$, and define $R^2 = \delta'\delta$, where $\delta = \Omega_{11}^{-1/2}\omega_{12}\omega_{22}^{-1/2}$ is a vector containing the bivariate zero frequency correlations of each element of v_{1t} with v_{2t}. R^2 lies between zero and one, and represents the contribution of the right-hand variables in the second equation of (2.3), as it is zero when these variables are not correlated in the long run with the errors from the cointegration regression. Under the assumption that x_t are not individually cointegrated, Ω_{11} is non-singular.

The triangular model (2.3) is general and does not assume that the x_t variables are weakly exogenous for the cointegration parameter γ. In fact, as Elliott et al. (2002) show, the 'directional' assumption $\alpha_1 = 0$ and weak exogeneity coincide only if $\Phi_{12}(1) = 0$. This point is only relevant under the alternative hypothesis where the variables are indeed cointegrated and γ is identified.

I will consider three combinations of restrictions for the deterministic component: (i) $\mu_2 - \gamma'\mu_1 = 0$ and $\tau = 0$, (ii) $\tau = 0$, (iii) $\tau_2 - \gamma'\tau_1 = 0$, or there are no restrictions. In all cases, μ_1 is allowed to be nonzero. The first case corresponds to no deterministic terms. In the second case, there is no drift or trend in Δx_t, but a constant in the cointegrating vector. Case three has x_t with a unit root and drift, and a mean, or a mean and trend, in the cointegrating vector. For case three, I assume that a constant and a time trend are included in the estimated regressions to obtain tests that are similar under the null. In all cases, the number of lags is approximated by a finite number k, chosen with some criteria that can be data dependent.

2.1. Tests for no cointegration

A variety of tests for no cointegration exist. This section briefly introduces the tests that will be analyzed in the next section.

To test the hypothesis of no cointegration, Engle and Granger (1987) first suggested the application of unit root tests to least-squares residuals from the cointegration regression

$$y_t = d_t + \gamma' x_t + u_t \tag{2.4}$$

for various possible choices of a deterministic d_t. Rejection of a unit root in the residuals from (2.4) 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) is to use the t-ratio test[3] on φ_0 in the Augmented Dickey-Fuller (ADF) regression:

$$\Delta\hat{u}_t = \varphi_0 \hat{u}_{t-1} + \sum_{i=1}^{k} \pi_i \Delta\hat{u}_{t-i} + \xi_{tk}, \tag{2.5}$$

[3] The t-test is not the only test choice. Although it may be interesting to compare the difference in power between different unit root tests on the residuals, this paper only looks at the most commonly used ADF t-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 (2001).

where $\hat{u}_t$ are the residuals from the LS estimation of the cointegration regression (2.4) estimated, respectively, with no mean for case (i), mean only for case (ii) and mean and trend for case (iii).[4]

As the Granger Representation theorem shows, a necessary and sufficient condition for cointegration is that the cointegrated series can be represented by a Vector Error Correction Model (VECM), where the EC term is present under cointegration. Testing whether the coefficient of the EC term is significant can be used as a test for the null of no cointegration. With a small amount of rearrangement and the use of Beveridge and Nelson (1981) decomposition, model (2.1) can be written in its EC form:

$$\Delta z_t = d_t + (\rho - 1)\Phi(1)\alpha\beta' z_{t-1} + \Phi^*(L)\Delta z_{t-1} + \varepsilon_t, \tag{2.6}$$

where d_t are the deterministic terms and $\Phi^*(L)$ is a polynomial of order k.[5] Implicit in Assumption 1 is the summability condition $\sum_{j=0}^{\infty} j\|\Phi_j\| < \infty$, which also implies $\sum_{j=0}^{\infty} \|\Phi_j^*\| < \infty$, where $\|\cdot\|$ is the standard Euclidian norm. From the EC model, the conditional model for y_t given x_t can be written as

$$\Delta y_t = d_{2t} + \varphi_0 u_{t-1} + \sum_{i=0}^{k} \pi_i' \Delta x_{ct-i} + \xi_{2tk}, \tag{2.7}$$

where $\pi_i' = [\pi_{1i}' \;\; \pi_{2i}]$, $\pi_{20} = 0$, $x_{ct} = [x_t' \;\; y_t]'$, $\varphi_0 = (\rho - 1)\theta$ with θ a function of $\Phi(1)$ and Ω.[6] d_{2t} is the second element of d_t, and it is such that it is zero when there is no drift in the 'X' variable (case (i) and (ii)), and it is constant for case (iii).[7] Under the alternative, x_t are not weakly exogenous for γ and the full system (2.6) should be used for inference on γ. At the same time, under the null of $\rho = 1$, $\beta' z_{t-1}$ does not enter the marginal equation for x_t, and a test for the stability of the conditional EC model (for $\varphi_0 = 0$) can be used to test whether y_t and x_t are cointegrated (Boswijk, 1994). This suggests that, if γ is known, a t-test on the parameter φ_0 estimated from the single equation (2.7) can be used to test the null of no cointegration.[8]

The assumption of known cointegration vector is a restrictive assumption. Unless the econometrician finds himself in the unlikely case of perfect knowledge of the cointegrating vector, the EC test presented in the previous section is not feasible. Banerjee et al. (1986, 1993, 1998) suggest adding a redundant regressor to avoid

[4] Hansen (1992) shows that the limiting distribution is sensitive to the actual trend and the detrending procedure. In the case in which only a mean is present in the cointegrating vector, using the unrestricted model with a trend has slightly less power than the restricted model with a constant only. On the other hand, if the cointegration regression is run with a constant only, the critical value is dependent on whether the true trend is zero or not. This paper only considers the widely used case in which the regression is run with a trend when the variables have a drift.

[5] In particular, $d_t = \Phi(1)\tau - (\rho - 1)\Phi(1)\alpha\beta'(\mu - \tau) - (\rho - 1)\Phi(1)\alpha\beta'\tau t$.

[6] Specifically $\theta = \Phi_{22}(1) - [\Phi_{21}(1) + \Phi_{22}(1)\gamma' + \Phi_{22}(1)\omega_{21}\Omega_{11}^{-1}][\Phi_{11}(1) + \Phi_{12}(1)\gamma + \Phi_{12}(1)\omega_{21}\Omega_{11}^{-1}]\Phi_{12}(1)$.

[7] I rule out the possibility of a trend in Eq. (2.7), as this would imply a mean and trend in Δx_t which is unlikely in practice.

[8] 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 φ_0 corresponds to testing $\rho = 1$.

imposing a particular cointegrating vector. The equation to be estimated is then:

$$\Delta y_t = d_{2t} + \varphi_0(y_{t-1} - i'_{n_1}x_{t-1}) + \varphi'_1 x_{t-1} + \sum_{i=0}^{k} \pi'_i \Delta x_{ct-i} + \xi_{2tk}. \tag{2.8}$$

φ_0 is unchanged, $\varphi'_1 = \varphi_0(i'_{n_1} - \gamma')$ and i_{n_1} is an $n_1 \times 1$ vector of ones. In case (iii), the partial system (2.8) needs to be estimated with a mean and a trend to obtain similar tests under the null.[9] A test of the hypothesis $\varphi_0 = 0$, based on the t-statistic on φ_0 in (2.8), is still a valid test for the absence of cointegration. In Section 4.2, I derive the asymptotic distribution of the EC test in (2.7) and the EC with redundant regressor (ECR) test in (2.8) under the local alternative $T(\rho - 1) = c$.

In Eq. (2.8), under the null hypothesis in which $\rho = 1$, both coefficients on the variables in levels are zero. Thus, an F test can be used to test for the joint significance of φ_0 and φ_1. This test is equivalent to the Wald test proposed by Boswijk (1994). For the case examined in this paper, in which the right-hand variables are not mutually cointegrated and there is at most one cointegration vector, this test is also equivalent to the test proposed by Harbo et al. (1998). Boswijk (1994) shows that the null asymptotic distribution of this test is a function of standard Brownian Motions. In the next section, the asymptotic power of the Wald test is analytically computed and compared with the power of the ADF and ECR tests.

One of the implications of the model is that the mean in Eq. (2.8) is also zero under the null. This restriction can be jointly tested in the ECR and Wald tests.

The fourth test this paper considers is the Johansen and Juselius' (1990) maximum eigenvalue test for the hypothesis of no cointegrating vectors based on the statistic (2.9):

$$\lambda_{\max} = -T\ln(1 - \hat{\lambda}_1). \tag{2.9}$$

Note that the general trace test proposed by Johansen (1988), to test the hypothesis of h cointegrating vectors against the hypothesis that there are no unit roots, is used for a different set of hypotheses than the one tested in the conditional EC model or in Engle and Granger's procedure. In the ADF test, under the null hypothesis, $\rho = 1$ and there are unit roots in the system, while, under the alternative of $\rho < 1$, z_t contains n_1 unit roots. This is the hypothesis tested by the maximum eigenvalue statistics (2.9). Although the $\lambda_{\max}$ test is not as general as the trace test, the analysis is limited to the case in which there is only one cointegration vector. For the purpose of this paper, only the $\lambda_{\max}$ test is directly comparable to the ADF and the EC tests.

3. Asymptotic power functions

Given that 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 than the others. A majority of the literature comparing tests for cointegration uses Monte Carlo experiments with various combinations of values for

[9] See also Boswijk (1994) and Harbo et al. (1998).

the parameters of the data-generating process. To understand the previous studies, we need to know how the nuisance parameters affect power. This can be accomplished by computing the analytical power of the tests. The knowledge of which nuisance parameter enters the asymptotic distributions helps us in designing the correct experiment, and suggests in which direction of the parameter space to look.

Since all tests are consistent, they all have power equal to one asymptotically and the asymptotic power for fixed alternatives cannot be used to rank the tests. The usual approach is to examine a sequence of local alternatives of the type $\rho = 1 + c/T$. When c is equal to zero, the errors u_t are integrated of order one. For c negative, the variables in Eq. (2.1) are cointegrated. Using this parameterization and the results of Phillips (1988), I evaluate and compare the power of the tests for the absence of cointegration presented in the previous section. The derivation of the analytical power is one of the main contributions of this paper, as it generalizes some of the previous work done in this area using a consistent framework. This section also extends some of the previous results as, to my knowledge, the local power of the Wald and the residuals-based tests for cointegration has never been computed before.

The following condition on the expansion rate of the truncation lag k is assumed throughout the paper:

Assumption 2. $T^{-1/3}k \to 0$ and $k \to \infty$ as $T \to \infty$.

The condition in Assumption 2 specifies an upper bound for the rate at which the value k is allowed to tend to infinity with the sample size. Ng and Perron (1995) show that conventional model selection criteria like AIC and BIC yield $k = O_p(\log T)$, which satisfies Assumption 2.

Often, a second condition is also assumed to impose a lower bound on k. The lower bound condition is only necessary to obtain $\sqrt{T}$ consistency of the parameters on the stationary variables, and it is sufficient but not necessary to prove the limiting distribution of the relevant test statistics (Ng and Perron, 1995; Lutkephol and Saikkonen, 1999). Since I am only interested in the t-ratio statistics, Assumption 2 is necessary and sufficient to prove the asymptotic distribution of the tests.

3.1. ADF test

In the case of the ADF test, Phillips and Ouliaris (1990) show that, under the null, the t-statistic has a non-standard distribution that is a function of standard Brownian motions. This paper goes one step further and analytically computes the local power.

Lemma 1. *When the model is generated according to* (2.1) *and Assumption* 1 *is valid, then, as* $T \to \infty$,[10]

$$(\hat{\gamma}_{ols} - \gamma) \Rightarrow \omega_{2.1}^{1/2} \Omega_{11}^{-1/2} \left(\int W_1^d W_1^{d\prime} \right)^{-1} \left(\int W_1^d J_{12c}^d \right), \tag{3.1}$$

[10] All the integrals are intended to be between 0 and 1, unless otherwise specified.

where $\hat{\gamma}_{ols}$ *is the LS estimator in the cointegration regression* (2.4), $\tilde{W}_1$ *is a univariate standard Brownian motion,* $W(\lambda)'=[W_1(\lambda)' \;\; W_2(\lambda)]$ *is an* $(n_1+1)\times 1$ *vector of standard independent Brownian motions partitioned conformably to* v_{1t} *and* v_{2t}, $\omega_{2.1}=\omega_{22}-\omega_{21}\Omega_{11}^{-1}\omega_{12}$, $\delta=\Omega_{11}^{-1/2}\omega_{12}\omega_{22}^{-1/2}$ *and* $R^2=\delta'\delta$. $J_{12c}(\lambda)$ *is an Ornstein–Uhlenbeck process such that*: $J_{12c}(\lambda)=W_{12}(\lambda)+c\int_0^\lambda \mathrm{e}^{(\lambda-s)c}W_{12}(s)\,\mathrm{d}s$ *with* $W_{12}(\lambda)=\sqrt{R^2/(1-R^2)}\tilde{W}_1(\lambda)+W_2(\lambda)$, *and*

(i) $W^d(\lambda)=W(\lambda)$ *and* $J_{12c}^d(\lambda)=J_{12c}(\lambda)$ *if* $\mu_2-\gamma'\mu_1=0$, $\tau=0$, *and no deterministic terms are included in the regressions.*
(ii) $W^d(\lambda)=W(\lambda)-\int W(s)\,\mathrm{d}s$ *and* $J_{12c}^d(\lambda)=J_{12c}(\lambda)-\int J_{12c}(s)\,\mathrm{d}s$ *if* $\tau=0$, *and a constant is included in the regression.*
(iii) $W^d(\lambda)=W(\lambda)-(4-6\lambda)\int W(s)\,\mathrm{d}s-(12\lambda-6)\int sW(s)\,\mathrm{d}s$ *and* $J_{12c}^d(\lambda)=J_{12c}(\lambda)-(4-6\lambda)\int J_{12c}(s)\,\mathrm{d}s-(12\lambda-6)\int sJ_{12c}(s)\,\mathrm{d}s$ *if* $\tau_2-\gamma'\tau_1=0$ *or no restrictions, and the mean and trend are included in the regression.*

If $R^2=0$ and $c=0$, then $J_{12c}^d(\lambda)\equiv W_2^d(\lambda)$ and (3.1) coincides with the usual distribution in the spurious regression, 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 sets of variables. As Phillips (1986) shows, the same results are valid in the more general case in which $R^2\neq 0$. In this case, even though there is a relationship between the two sets of variables, this relationship is not consistently estimated 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* (2.1) *and Assumptions* 1 *and* 2 *are valid, then, as* $T\to\infty$:

$$\hat{t}_{\hat{\varphi}_0}^{\mathrm{ADF}} \Rightarrow c\,\frac{[\eta_c^{d\prime}A_c^d\eta_c^d]^{1/2}}{[\eta_c^{d\prime}D\eta_c^d]^{1/2}}+\frac{\eta_c^{d\prime}\int W_c^d d\tilde{W}'\eta_c^d}{[\eta_c^{d\prime}D\eta_c^d]^{1/2}[\eta_c^{d\prime}A_c^d\eta_c^d]^{1/2}}, \tag{3.2}$$

where $\hat{t}_{\hat{\varphi}_0}^{\mathrm{ADF}}$ *is the t-statistic on* φ_0 *from the Augmented Dickey Fuller regression* (2.5).

$$\eta_c^{d\prime}=\left[-\left(\int W_1^{d\prime}J_{12c}^d\right)\left(\int W_1^dW_1^{d\prime}\right)^{-1} \quad 1\right],$$

$$W_c^d(\lambda)=\left[\,W_1^{d\prime}(\lambda) \quad J_{12c}^d(\lambda)\,\right]', \quad A_c^d=\int W_c^dW_c^{d\prime},$$

$$\tilde{W}(\lambda)=[W_1'(\lambda) \quad W_{12}(\lambda)]', \quad W_{12}(\lambda)=\sqrt{R^2/(1-R^2)}\tilde{W}_1(\lambda)+W_2(\lambda),$$

$$D=\begin{bmatrix} I & \bar{\delta} \\ \bar{\delta}' & 1+\bar{\delta}'\bar{\delta}\end{bmatrix},$$

$\bar{\delta}'=\omega_{2.1}^{-1/2}\omega_{21}\Omega_{11}^{-1/2}$, $R^2=\delta'\delta$, $\tilde{W}_1$ *is a univariate standard Brownian motion, and* $J_{12c}(\lambda)$ *is an Ornstein–Uhlenbeck process such that* $J_{12c}(\lambda)=W_{12}(\lambda)+c\int_0^\lambda \mathrm{e}^{(\lambda-s)c}W_{12}(s)\,\mathrm{d}s$. $W^d(\lambda)$ *and* $J_{12c}^d(\lambda)$ *are defined in Lemma* 1, (i)–(iii).

The result of Theorem 1 shows that the asymptotic distribution of the ADF test on the residuals is a function of few parameters. The test depends: (1) on the alternative c, (2) on the dimension of x_t through the dimension of the Brownian motion $W_1(\lambda)$, (3) and more importantly, on a single nuisance parameter R^2.

When $c=0$, $J_{12c}=W_{12}^d$ and 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). Under the null hypothesis, the limiting distribution is free of nuisance parameters; however, when $c \neq 0$, the local asymptotic power depends on c and on the value of R^2, but is asymptotically similar with respect to γ and the variance of the errors.

3.2. EC, ECR and WALD tests

Various papers have been written on the properties of the EC test. Assuming γ known, Banerjee et al. (1986) and Kremers et al. (1992) compute the asymptotic distribution of the EC test in (2.7) for the case in which the correlation between the two error terms in (2.3) (δ) is zero. They also compute the power of the test under fixed and local alternatives. The following theorem extends their results to the case in which δ is different from zero and to infinite order VAR with a data-dependent choice of lag length.[11]

Theorem 2. *When the model is generated according to* (2.1) *and Assumptions* 1 *and* 2 *are valid, then, as* $T \to \infty$:

$$\hat{t}^{\mathrm{EC}}_{\hat{\varphi}_0} \Rightarrow \frac{c}{\left(\int J_{12c}^{d2}\right)^{-1/2}} + \left(\int J_{12c}^{d2}\right)^{-1/2}\left(\int J_{12c}^{d}\,\mathrm{d}W_2\right) \tag{3.3}$$

where $\hat{t}^{\mathrm{EC}}_{\hat{\varphi}_0}$ *is the t-statistic on the EC term in* (2.7), $J_{12c}(\lambda)$ *and* $W_2(\lambda)$ *are defined as in Lemma* 1. $J_{12c}^d(\lambda)=J_{12c}(\lambda)$ *for cases* (i) *and* (ii) *and*, $J_{12c}^d(\lambda)=J_{12c}(\lambda)-\int J_{12c}(s)\,\mathrm{d}s$ *for case* (iii).

As Zivot (2000) also shows, the EC test statistic has the same asymptotic distribution as Hansen's (1995) unit root test on $\hat{u}_t$, when Δx_t is used as the stationary covariate. When $R^2=0$, as in the case in which x_t is strictly exogenous, the result of Theorem 2 agrees with Kremers et al. (1992).

When $c=0$, $J_{12c}(\lambda)=W_{12}(\lambda)$ and the asymptotic distribution of $\hat{t}^{\mathrm{EC}}_{\alpha}$ coincides with the result of Theorem 3 in Hansen (1995). The test is not invariant with respect to a nuisance parameter under the null, and a unique set of critical values cannot be obtained.

When the conditional EC equation (2.7) is extended by adding a redundant regressor, a test for no cointegration can be performed by looking at the significance of the

[11] 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 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.

coefficient on the difference of the variables in level without the knowledge of the cointegration vector. This transformation also renders the studentized test statistic invariant with respect to R^2 under the null. Banerjee et al. (1998) compute the distribution of this test under the null when $R^2 = 0$. Theorem 3 presents the asymptotic distribution of the test under the local alternative $T(\rho - 1)$ under more general conditions:

Theorem 3. *When the model is generated according to* (2.1) *and Assumptions* 1 *and* 2 *are valid, then, as* $T \to \infty$:

$$\hat{t}_{\hat{\varphi}_0}^{\mathrm{ECR}} \Rightarrow c\left[\int J_{12c}^{d2} - \int W_1^{d\prime} J_{12c}^d \left(\int W_1^d W_1^{d\prime}\right)^{-1} \int W_1^d J_{12c}^d\right]^{1/2} + \frac{\left[\int J_{12c}^d \,\mathrm{d}W_2 - \int W_1^{d\prime} J_{12c}^d \left(\int W_1^d W_1^{d\prime}\right)^{-1} \int W_1^d \,\mathrm{d}W_2\right]}{\left[\int J_{12c}^{d2} - \int W_1^{d\prime} J_{12c}^d \left(\int W_1^d W_1^{d\prime}\right)^{-1} \int W_1^d J_{12c}^d\right]^{1/2}}, \tag{3.4}$$

where $\hat{t}_{\hat{\varphi}_0}^{\mathrm{ECR}}$ *is the t-ratio test in Eq.* (2.8), *and* $J_{12c}^d(\lambda)$ *and* $W_1^d(\lambda)$ *are defined as in Lemma* 1.

The local asymptotic distribution of $\hat{t}_{\hat{\varphi}_0}^{\mathrm{ECR}}$ is identical to the local distribution of EC t-test with γ pre-specified presented in Zivot (2000). Under the local alternative, the ECR test depends not only on the particular alternative c, but also on the value of the nuisance parameter R^2. In contrast with the results for the EC test, under the null, the asymptotic distribution of $\hat{t}_{\hat{\varphi}_0}^{\mathrm{ECR}}$ is free of the nuisance parameter. When $c = 0$,

$$\hat{t}_{\hat{\varphi}_0}^{\mathrm{ECR}} \Rightarrow \frac{\left[\int W_2^d \,\mathrm{d}W_2 - \int W_1^{d\prime} W_2^d \left(\int W_1^d W_1^{d\prime}\right)^{-1} \int W_1^d \,\mathrm{d}W_2\right]}{\left[\int W_2^{d2} - \int W_1^{d\prime} W_2^d \left(\int W_1^d W_1^{d\prime}\right)^{-1} \int W_1^d W_1^d\right]^{1/2}}$$

and the asymptotic distribution of the statistic depends only on the dimension of x_t, and is invariant to any other parameter of the DGP.

Under the null hypothesis of no cointegration, both φ_0 and φ_1 in (2.8) are zero. As Boswijk (1994) suggests, the null hypothesis can be tested using a joint test on both parameters. Theorem 4 computes the local asymptotic distribution of Boswijk's (1994) Wald test statistic.

Theorem 4. *When the model is generated according to* (2.1) *and Assumptions* 1 *and* 2 *are valid, then, as* $T \to \infty$:

$$\hat{F} \Rightarrow c^2 \int J_{12c}^{d2} + 2c \int J_{12c}^d \,\mathrm{d}W_2 + \int W_c^{d\prime} \,\mathrm{d}W_2 \left[\int W_c^d W_c^{d\prime}\right]^{-1} \int W_c^d \,\mathrm{d}W_2, \tag{3.5}$$

where $\hat{F}$ *is the Wald test for the joint significance of* φ_0 *and* φ_1 *in* (2.8) *and* $J_{12c}^d(\lambda)$ *and* $W_c^d(\lambda)$ *are defined as in Lemma* 1 *for case* (i)–(iii). *For the case in which the mean is also restricted to be equal to zero,* $W_c^d(\lambda)' = [1 \;\; W_1(\lambda)' \;\; J_{12c}(\lambda)]$.

In this case also, the only parameters entering the asymptotic distribution are the alternative c and R^2. The test does not depend on the variances of the error terms or the true value of the cointegration vector. For $c=0$, the asymptotic distribution of the test is invariant to any nuisance parameter and coincides with the result of Theorem 1 in Boswijk (1994).

3.3. Johansen test

The ECM representation (2.6) can be written as

$$\Delta z_t = d_t + \Pi z_{t-1} + \sum_{i=1}^{k} \Pi_i \Delta z_{t-i} + \varepsilon_t, \tag{3.6}$$

where $\Pi = \bar{\alpha}\beta'$, $\bar{\alpha} = (\rho-1)\Phi(1)\alpha = [(\rho-1)\Phi_{12}(1)'\ (\rho-1)\Phi_{22}(1)]'$ and the number of lags k in the model is chosen by some criteria that can be data dependent. While, under the null of no cointegration, $\rho = 1$ and $rank(\Pi) = 0$, when y_t and x_t are cointegrated $\rho < 1$ and $rank(\Pi) = 1$. The local alternative for the rank test suggested by Johansen (1995) is of the form $H_a: \Pi = \bar{\alpha}\beta' + \bar{\alpha}_1\beta_1'/T$, where $\bar{\alpha}_1$ and β_1 are $(n_1+1)\times 1$ matrices. Under the local alternative, the process has one extra cointegrating vector, β_1, that enters the process with small adjustment coefficients $T^{-1}\bar{\alpha}_1$. In the $\lambda_{\max}$ test, the relevant local alternative is $H_a: \Pi = \bar{\alpha}_1\beta_1'/T$, where $\bar{\alpha}_1 = c[\Phi_{12}(1)'\ \Phi_{22}(1)]'$ and $\beta_1 = [-\gamma'\ 1]'$. Since model (2.1) implies restrictions on the deterministic terms under the null, often the model is written in the form

$$\Delta z_t = \tilde{d}_t + \Pi^* z_{t-1}^* + \sum_{i=1}^{k} \Pi_i \Delta z_{t-i} + \varepsilon_t, \tag{3.7}$$

where $z_t^{*\prime} = [x_t'\ y_t\ 1]$ and $\Pi^* = \bar{\alpha}\beta^{*\prime}$, where $\beta^* = [-\gamma'\ 1\ (\mu_2 - \gamma'\mu_1)]'$.

Johansen (1995) discusses the power of the rank test under this general local alternative. Lutkephol and Saikkonen (2000) and Saikkonen and Lutkephol (1999, 2000) discuss the restrictions imposed on the deterministic component by cointegration, and provide a general framework for deriving the local power properties of likelihood ratio tests for the cointegrating rank under various assumptions on the deterministic term. 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* (2.1) *and Assumptions* 1 *and* 2 *are valid, then, as* $T \to \infty$:

$$\hat{\lambda}_{\max} \Rightarrow \max \operatorname{eig} \left\{ \left(\int W_c W_c' \right)^{-1} \int W_c \, \mathrm{d}W' \int \mathrm{d}W W_c' \right.$$
$$+ \left(\int W_c W_c' \right)^{-1} \int W_c \, \mathrm{d}W' G_c'$$

$$+\left(\int W_c W_c'\right)^{-1} G_c \left(\int W_c \, \mathrm{d}W'\right)'$$

$$+\left(\int W_c W_c'\right)^{-1} G_c G_c' \Bigg\}, \tag{3.8}$$

where $\max \operatorname{eig}\{M\}$ *denotes the maximum eigenvalue of matrix* M, $W(\lambda) = \begin{bmatrix} W_1(\lambda) \\ W_2(\lambda) \end{bmatrix}$, $G_c' = \begin{bmatrix} 0 \\ c \end{bmatrix} \int W_c' J_{12c}$ *and*

(i) $W_c(\lambda) = \begin{bmatrix} W_1(\lambda) \\ J_{12c}(\lambda) \end{bmatrix}$ *if* $\mu_2 - \gamma' \mu_1 = 0$ *and* $\tau = 0$ *and no mean is included in the VAR.*

(ii) $W_c^d(\lambda) = \begin{bmatrix} W_1^d(\lambda) \\ J_{12c}^d(\lambda) \end{bmatrix}$ *if* $\tau = 0$ *and the mean is unrestricted.*

(iii) $W_c(\lambda) = \begin{bmatrix} W_1(\lambda) \\ J_{12c}(\lambda) \\ 1 \end{bmatrix}$ *if* $\tau = 0$ *and the mean is restricted to enter the EC term.*

(iv) $W_c(\lambda) = \begin{bmatrix} W_1^d(\lambda) \\ J_{12c}^d(\lambda) \\ 1 \end{bmatrix}$ *if* $\tau_2 - \gamma' \tau_1 = 0$ *or no restrictions, mean and trend are restricted to enter the EC term and the VAR is estimated with a mean.*

$W_1^d(\lambda)$ *and* $J_{12c}^d(\lambda)$ *are demeaned Brownian Motions.*

If the mean is not restricted to enter the EC term, then $W_c(\lambda) = \begin{bmatrix} W_1^d(\lambda) \\ J_{12c}^d(\lambda) \end{bmatrix}$, and the Brownian Motions are demeaned for case (ii) and detrended for case (iii). See also Johansen (1995) and Saikkonen and Lutkephol (1999). Again, the only parameter entering the power function is the correlation coefficient R^2. 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 differently depending on the particular design. Haug (1996), for example, compares nine different tests for cointegration on the basis of power and size distortions induced by the presence of moving average components. Haug (1996) finds that, in general, single-equation tests have smaller size distortions, but also 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 local asymptotic distributions computed in the previous section tell us exactly which parameters are important for power, and give a precise indication of which direction we need to look in the Monte Carlo analysis. For each test statistic, 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 tests are not invariant to the particular alternative, a uniformly most powerful test for model (2.3) cannot be computed. There is no reason to expect one of the tests to have uniformly higher power than the others. If R^2 is large, we would expect a full system approach that exploits this correlation to have smaller standard errors and to perform better. This section compares the power of the tests for the absence of cointegration presented in the previous sections using Monte Carlo experiments.

4.1. Large sample

The power functions are computed as the probability that the tests are less than some critical value. Given the expression for the limit distribution of all the tests, the asymptotic local power can be approximated by simulating the distributions presented in the previous sections. This experiment considers $c = 0, -5, -10, -15, -20$ and $R^2 = 0, 0.3, 0.7$. Each Brownian Motion's piece in the asymptotic distribution is approximated by step functions using Gaussian random walk with $T = 1000$ observations. 5000 replications are used to compute the critical values and the rejection probabilities for each c and R^2.

Since the local power for all the tests depends solely on one nuisance parameter, the power functions of the tests are compared for different values of R^2. Note that the asymptotic distributions computed under the local alternatives do not depend on the particular estimator used to estimate other nuisance parameters such as $\omega_{2.1}$. As pointed out by a referee, the finite-size sample properties of tests for no cointegration can be sensitive to the choice of estimation method for the nuisance parameters. In this respect, the local asymptotic curves presented in this section should be interpreted as approximations to the finite sample size-adjusted power curves of the corresponding tests.

Table 1 reports the large sample results for the univariate case. For small R^2, in a close neighborhood of zero, the ECR test has slightly higher power than the ADF and Wald test. For R^2 equal to zero, using a full system approach is inefficient and the ADF, Wald, and ECR tests perform better than the $\lambda_{\max}$ test. As R^2 increases, all the tests, with the exception of the ADF test, perform better. As expected, given that they are both based on the conditional EC equation on a rotated model, Wald and ECR tests have similar large sample power.

The ranking between the tests is even clearer in Fig. 1,[12] where the asymptotic power for the local alternative is presented. For large values of R^2, we expect the single-equation approach of the ADF test on the residuals to have low power. When R^2

[12] Only the results for the demeaned case are presented here. The results for the no mean and demeaned and detrended cases are similar: Although the tests have generally lower power when a deterministic part is included, as is the general case for unit root tests, the ranking between the tests is unchanged.

Table 1
Large sample power

$-c$		0	5	10	15	20
$R^2 \setminus \rho$		1	0.995	0.99	0.985	0.98
Case (i)						
0	ADF	0.050	0.147	0.379	0.686	0.899
	ECR	0.050	0.200	0.495	0.796	0.951
	Wald	0.050	0.130	0.343	0.647	0.874
	$\lambda_{\max}$	0.050	0.083	0.195	0.412	0.660
0.3	ADF	0.050	0.127	0.328	0.638	0.873
	ECR	0.050	0.235	0.647	0.908	0.984
	Wald	0.050	0.207	0.591	0.867	0.970
	$\lambda_{\max}$	0.050	0.126	0.374	0.694	0.906
0.7	ADF	0.050	0.089	0.266	0.556	0.824
	ECR	0.050	0.425	0.905	0.992	0.999
	Wald	0.050	0.607	0.941	0.995	0.999
	$\lambda_{\max}$	0.050	0.433	0.893	0.992	1.000
Case (ii)						
0	ADF	0.050	0.094	0.221	0.431	0.679
	ECR	0.050	0.117	0.269	0.498	0.746
	Wald	0.050	0.093	0.209	0.411	0.656
	$\lambda_{\max}$	0.050	0.070	0.129	0.252	0.442
0.3	ADF	0.050	0.081	0.177	0.358	0.598
	ECR	0.050	0.129	0.355	0.674	0.894
	Wald	0.050	0.131	0.356	0.663	0.872
	$\lambda_{\max}$	0.050	0.092	0.227	0.467	0.736
0.7	ADF	0.050	0.046	0.104	0.242	0.468
	ECR	0.050	0.204	0.690	0.947	0.995
	Wald	0.050	0.337	0.806	0.968	0.996
	$\lambda_{\max}$	0.050	0.227	0.712	0.961	0.998
Case (iii)						
0	ADF	0.050	0.086	0.162	0.319	0.527
	ECR	0.050	0.083	0.171	0.330	0.542
	Wald	0.050	0.081	0.162	0.305	0.501
	$\lambda_{\max}$	0.050	0.062	0.093	0.172	0.301
0.3	ADF	0.050	0.072	0.127	0.234	0.411
	ECR	0.050	0.083	0.206	0.429	0.705
	Wald	0.050	0.098	0.246	0.495	0.751
	$\lambda_{\max}$	0.050	0.069	0.141	0.309	0.542
0.7	ADF	0.050	0.037	0.049	0.100	0.221
	ECR	0.050	0.082	0.374	0.773	0.957
	Wald	0.050	0.198	0.647	0.925	0.989
	$\lambda_{\max}$	0.050	0.132	0.490	0.873	0.983

The power is computed with $T = 1000$ and 5000 replications. Case (i) corresponds to the no mean case, cases (ii) and (iii) are for the demeaned and demeaned and detrended cases, respectively.

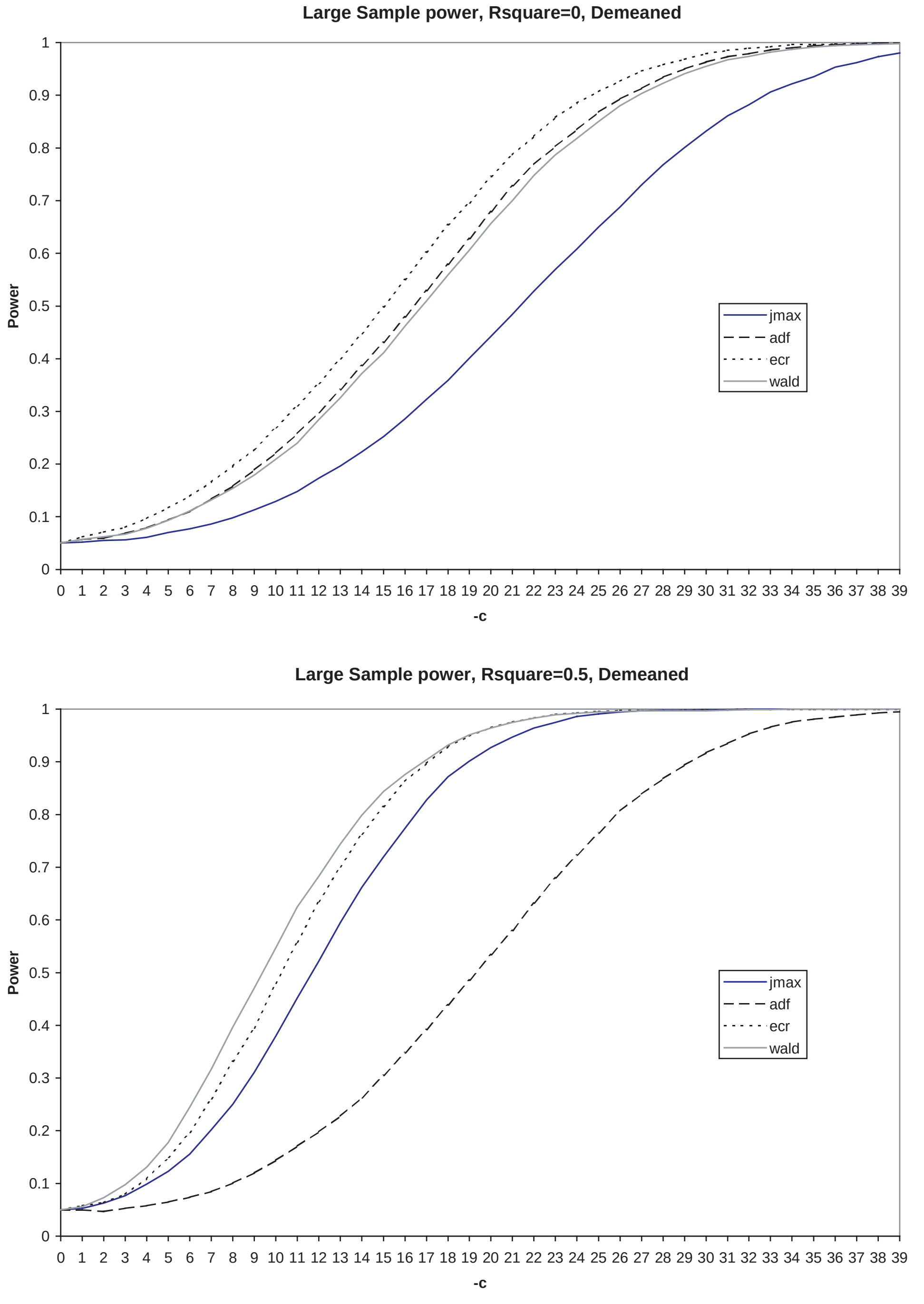
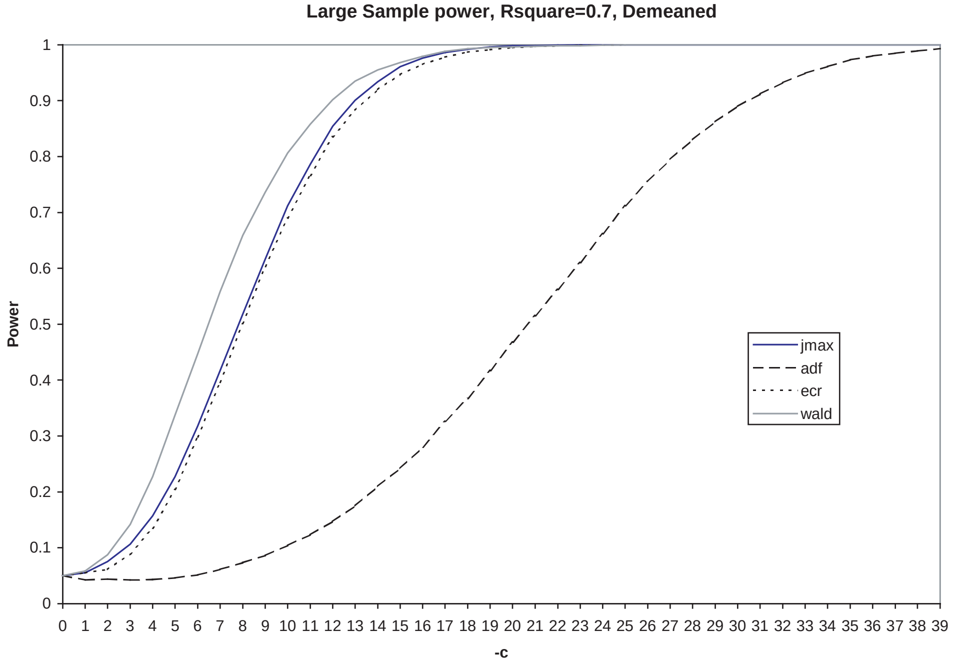


Fig. 1. Large sample power, demeaned.



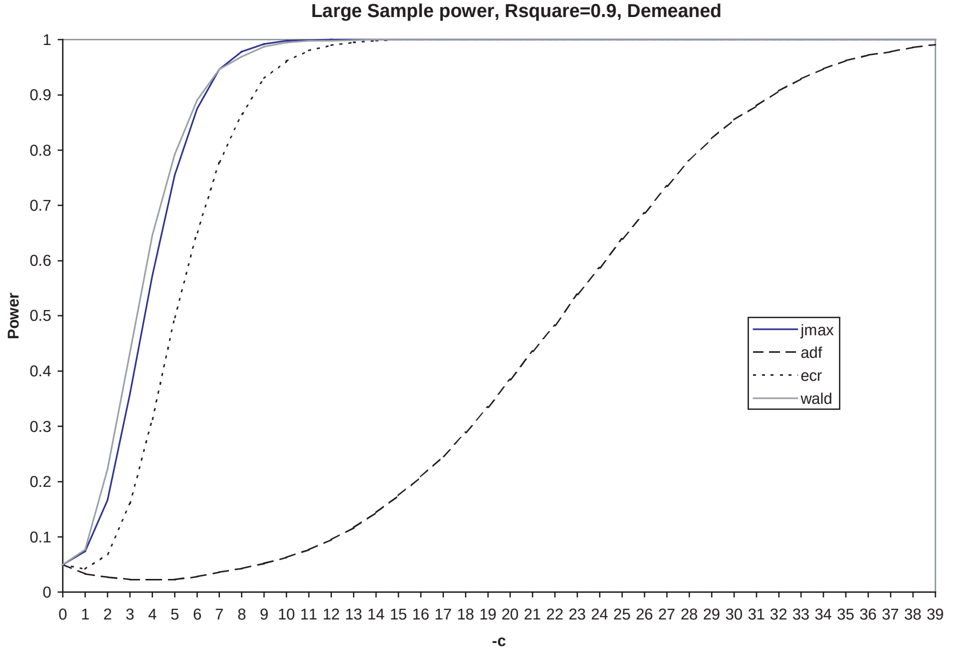


Fig. 1. *continued.*

is higher than 0.3, which corresponds in the univariate case to a correlation coefficient of 0.5, the ADF test performs poorly when compared to the other tests. For $R^2 = 0.7$ ($\delta = 0.83$), the power of the ECR test is two to six times larger than the power of the unit root test on the residuals. Although the power of the maximum eigenvalue test is significantly higher than the power of the ADF test, the power of both ECR and Wald tests is slightly higher than Johansen's test. For the extreme case of $R^2 = 0.9$, the difference between the ADF and the other tests is even more pronounced. All other tests have similar power.

Fig. 2 shows the rejection rates for different values of c as a function of R^2. For values of R^2 larger than approximately 0.2, the system tests generally perform better. The difference in the power functions is more significant as R^2 increases and as we move away from the null. The results for the multivariate case are qualitatively equivalent.

4.2. Small sample

To examine the usefulness of these asymptotic approximations in practice, we need to study the small sample behavior of the tests. Using the DGP of Eq. (2.3), errors are randomly generated from a bivariate Normal with mean zero and variance–covariance matrix

$$\Omega = \begin{bmatrix} \omega_1^2 & \omega_1\omega_2\delta \\ \omega_1\omega_2\delta & \omega_2^2 \end{bmatrix}.$$

For $R^2 = \delta^2$, I consider $R^2 = 0, 0.3, 0.7$ and $c = 0, -5, -10, -15, -20$ that correspond for $T = 100$ to values for $\rho = 1, 0.95, 0.9, 0.85$. The tests are all invariant to γ and the variance of the errors so any reasonable number can be chosen. All the tests are computed as described in Section 2. Table 2 presents the size-adjusted rejection rates for the case in which there is no serial correlation in the error terms, and the regressions are estimated with no lags.

As Table 2 shows, all tests have low power when the root is close to one. For small R^2, the ECR and the Wald tests have, in general, higher power than all the other tests for any value of ρ. The small sample results are consistent with the predictions of the asymptotic theory: The relative ranking of the tests is the same as in the large-sample case. When R^2 is 0.7, the difference in the small sample power of the tests is significant even when ρ is large.

It is more interesting when there is some serial correlation in the error terms, as in this case tests for integration and cointegration may have severe size distortions. Since a test with good power but bad size may not be the best choice, it is important to evaluate the size properties of the tests. This experiment considers case (ii) in which there is no drift in the variables but a mean is present in the cointegration regression. The data are generated as in model (2.3), with $(1 - \Phi L)v_t = (1 + \Theta L)\varepsilon_t$, $\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$, $\Theta = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix}$ and $T = 100$ observations. All the regressions are estimated with a mean, and the lag length is chosen by BIC with a maximum of 12 lags. Table 3

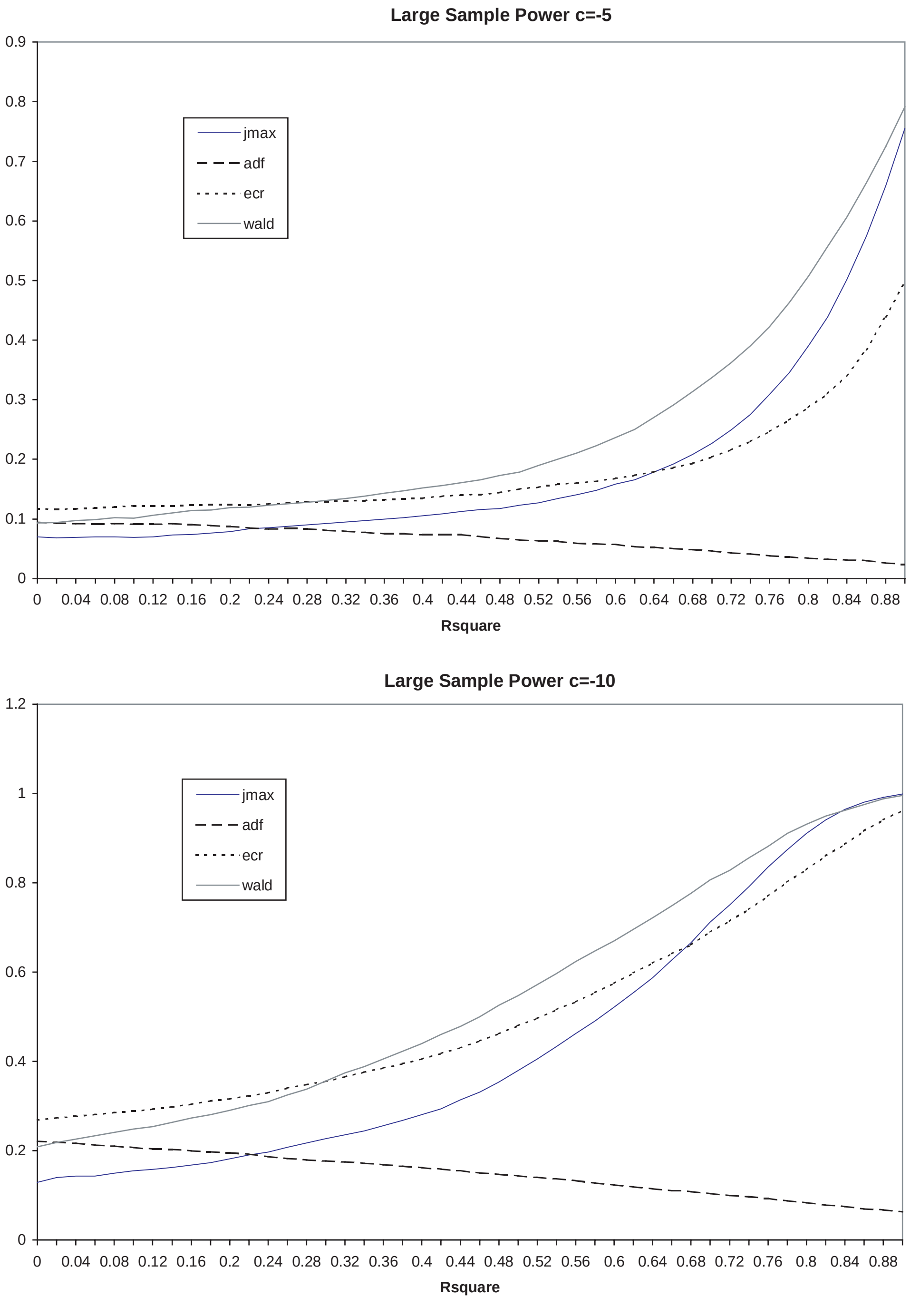


Fig. 2. Large sample power as a function of R^2, demeaned case.

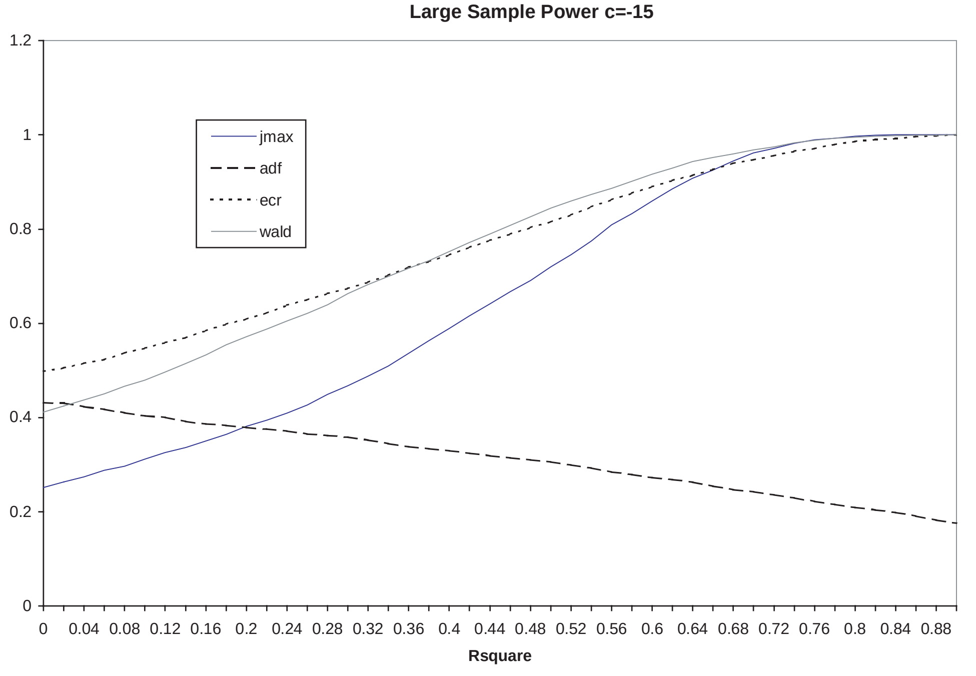


Fig. 2. *continued.*

presents the results for different combinations of values for the autoregressive and moving-average components. As expected, when only an autoregressive component is present in the error terms, the BIC performs well in choosing the appropriate number of lags. In this case, the Wald test has the worst performance in terms of size distortions with an empirical size around 10%. As it is well known, large size distortions occur when large negative roots are present in the moving average components. All tests present large empirical sizes well above 50%, with the exception of the remarkable performance of the Wald test that shows a size of less than 20%.

As a general conclusion, Tables 1–3 indicate that the ECR and the Wald test, although the least commonly used of the tests presented in this paper, not only perform better than other tests in terms of power in large and small samples, but are also not worse or better in terms of size distortions.

5. Conclusions

Over the past years, testing for cointegration has become an important step in any empirical analysis. A complete theoretical understanding of current methods is relevant in designing the appropriate Monte Carlo experiment to evaluate the relative performance of the tests. This paper illustrates how the analytical analysis of the local power of the tests can help in identifying which nuisance parameters are important. In

Table 2
Size-adjusted small sample power

$-c$		0	5	10	15	20
$R^2 \setminus \rho$		1	0.95	0.90	0.85	0.80
Case (i)						
0	ADF	0.050	0.154	0.379	0.683	0.903
	ECR	0.050	0.207	0.498	0.793	0.950
	Wald	0.050	0.128	0.336	0.638	0.868
	$\lambda_{\max}$	0.050	0.090	0.215	0.440	0.701
0.3	ADF	0.050	0.123	0.343	0.652	0.887
	ECR	0.050	0.256	0.661	0.916	0.987
	Wald	0.050	0.205	0.565	0.855	0.971
	$\lambda_{\max}$	0.050	0.127	0.392	0.715	0.920
0.7	ADF	0.050	0.103	0.306	0.600	0.860
	ECR	0.050	0.509	0.936	0.955	1.000
	Wald	0.050	0.599	0.938	0.996	1.000
	$\lambda_{\max}$	0.050	0.458	0.898	0.993	1.000
Case (ii)						
0	ADF	0.050	0.095	0.218	0.447	0.707
	ECR	0.050	0.111	0.248	0.492	0.748
	Wald	0.050	0.093	0.200	0.411	0.668
	$\lambda_{\max}$	0.050	0.060	0.117	0.242	0.436
0.3	ADF	0.050	0.077	0.173	0.376	0.624
	ECR	0.050	0.109	0.338	0.667	0.896
	Wald	0.050	0.113	0.342	0.661	0.881
	$\lambda_{\max}$	0.050	0.078	0.196	0.435	0.718
0.7	ADF	0.050	0.050	0.109	0.257	0.508
	ECR	0.050	0.190	0.681	0.952	0.995
	Wald	0.050	0.328	0.797	0.973	0.997
	$\lambda_{\max}$	0.050	0.201	0.668	0.947	0.996
Case (iii)						
0	ADF	0.050	0.074	0.144	0.286	0.516
	ECR	0.050	0.080	0.160	0.315	0.542
	Wald	0.050	0.077	0.143	0.279	0.492
	$\lambda_{\max}$	0.050	0.062	0.096	0.169	0.311
0.3	ADF	0.050	0.062	0.109	0.213	0.398
	ECR	0.050	0.079	0.188	0.423	0.717
	Wald	0.050	0.086	0.216	0.466	0.747
	$\lambda_{\max}$	0.050	0.065	0.137	0.295	0.546
0.7	ADF	0.050	0.037	0.049	0.101	0.222
	ECR	0.050	0.081	0.358	0.784	0.964
	Wald	0.050	0.173	0.618	0.919	0.990
	$\lambda_{\max}$	0.050	0.118	0.466	0.854	0.982

The power is computed with $T = 100$ and 5000 replications. Case (i) corresponds to the no mean case, Cases (ii) and (iii) are for the demeaned and demeaned and detrended cases respectively.

Table 3
Size distortions

AR errors	MA errors	ADF	ECR	Wald	$\lambda_{\max}$
$\phi_{11}=\phi_{22}=0$	$\vartheta_{11}=\vartheta_{22}=0$	0.062 (1.1)	0.070 (1.6)	0.128 (1.6)	0.060 (1.2)
$\phi_{11}=\phi_{22}=0.2$	$\vartheta_{11}=\vartheta_{22}=0$	0.063 (1.1)	0.073 (1.6)	0.132 (1.6)	0.066 (1.8)
$\phi_{11}=\phi_{22}=0.8$	$\vartheta_{11}=\vartheta_{22}=0$	0.067 (1.1)	0.100 (1.7)	0.172 (1.7)	0.120 (2.2)
$\phi_{11}=\phi_{22}=0.2$ $\phi_{12}=\phi_{21}=0.5$	$\vartheta_{11}=\vartheta_{22}=0$	0.039 (1.3)	0.058 (1.6)	0.133 (1.6)	0.077 (2.2)
$\phi_{11}=\phi_{22}=0$	$\vartheta_{11}=\vartheta_{22}=0.2$	0.072 (1.1)	0.080 (1.6)	0.131 (1.6)	0.065 (1.8)
$\phi_{11}=\phi_{22}=0$	$\vartheta_{11}=\vartheta_{22}=0.8$	0.081 (3.0)	0.097 (4.2)	0.222 (4.2)	0.106 (5.0)
$\phi_{11}=\phi_{22}=0$	$\vartheta_{11}=\vartheta_{22}=0.2$ $\vartheta_{12}=\vartheta_{21}=0.5$	0.094 (1.1)	0.088 (2.3)	0.156 (2.3)	0.080 (3.3)
$\phi_{11}=\phi_{22}=0$	$\vartheta_{11}=\vartheta_{22}=-0.8$	0.676 (1.6)	0.588 (2.3)	0.158 (2.3)	0.633 (2.5)
$\phi_{11}=\phi_{22}=0$	$\vartheta_{11}=\vartheta_{22}=-0.8$ $\vartheta_{12}=\vartheta_{21}=0.5$	0.627 (1.5)	0.279 (1.9)	0.166 (1.9)	0.543 (2.3)

The size distortions are computed with $T=100$ and 5000 replications. Lags in each regression are chosen using BIC with a maximum of 12 lags: the number in parenthesis represents the average number of lags chosen by BIC.

particular, this paper looks at the class of tests for the absence of cointegration and shows that an important role is played by the correlation of the ‘X’ variables with the errors of the cointegration regression. As intuition suggests, when this correlation is high, system approaches like the Johansen maximum eigenvalue or tests on the EC model can exploit this correlation and significantly outperform single-equation tests.

Further research still needs to be done to fully understand how all the tests for cointegration work. Because of the lack of locally asymptotic normal likelihood, standard asymptotic theory is inapplicable. In fact, no upper bound on the achievable local power of tests for no cointegration with unknown cointegration vector is available at present. The analytical results of this paper allow the computation of an asymptotic approximation of the power function of tests for no cointegration, which can be of help in the construction of tests with nearly optimal local power. Additionally, the local asymptotic power can be used as a benchmark to compare the power of bootstrap methods in testing for no cointegration.

Acknowledgements

I thank Graham Elliott, Silvia Goncalves, Scott Goode, Clive Granger, Bruce Hansen, Michael Jansson, Barbara Rossi, Peter Schmidt, Jeff Wooldrige, for valuable comments. I would also like to thank the editor and two referees for insightful comments. All errors remain my responsibility.

Appendix A

For the proofs, 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. $\Rightarrow$ denotes weak convergence.

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

$$(1)\quad \Omega_{11}^{-1/2}T^{-2}\sum_{t=1}^{T}x_t^d x_t^{d\prime}\Omega_{11}^{-1/2}\Rightarrow\int W_1^d W_1^{d\prime},$$

$$(2)\quad \Omega_{11}^{-1/2}\omega_{2.1}^{-1/2}T^{-2}\sum_{t=1}^{T}x_t^d u_t^d\Rightarrow\int W_1^d J_{12c}^d,$$

$$(3)\quad \omega_{2.1}^{-1}T^{-2}\sum_{t=1}^{T}u_t^{d2}\Rightarrow\int J_{12c}^{d2},$$

$$(4)\quad \omega_{2.1}^{-1/2}T^{-1}\sum_{t=1}^{T}u_{t-1}^d(\Sigma^{-1/2}\varepsilon_t)\Rightarrow\int J_{12c}^d\,\mathrm{d}W,$$

$$(5)\quad \Omega_{11}^{-1/2}T^{-1}\sum_{t=1}^{T}x_{t-1}^d(\Sigma^{1/2}\varepsilon_t)\Rightarrow\int W_1^d\,\mathrm{d}W,$$

where $[\ ]_{ii}$ *denotes the ii element of a matrix.* $W'=[W_1'\ W_2]$ *is an* $n\times 1$ *vector of standard independent Brownian motions partitioned conformably to* v_{1t} *and* v_{2t}. J_{12c} *is a scaled Ornstein–Uhlenbeck process such that*: $J_{12c}(\lambda)=W_{12}(\lambda)+c\int_0^\lambda \mathrm{e}^{(\lambda-s)c}W_{12}(s)\,\mathrm{d}s$ *with* $W_{12}=\sqrt{R^2/(1-R^2)}\tilde{W}_1+W_2$, $\tilde{W}_1$ *is an univariate standard Brownian motion independent of* W_2, *and*

(i) *if* $\mu_2-\gamma'\mu_1=0$ *and* $\tau=0$, $x_t^d=x_t$ *and* $W^d=W$, $J^d=J$
(ii) *if* $\tau=0$, $x_t^d=x_t-\bar{x}$ *and* $W^d=W-\int W$, $J^d=J-\int J$

(iii) *if* $\tau_2 - \gamma'\tau_1 = 0$ *or no restrictions,* x_t^d *is* x_t *detrended by OLS and* $W^d = W - (4-6\lambda)\int W - (12\lambda - 6)\int sW$,

$$J^d = J - (4-6\lambda)\int J - (12\lambda - 6)\int sJ.$$

Proof. Assumption 1 along with the assumptions on $\Phi(L)$ implies $T^{-1/2}\sum_{s=1}^{[T\lambda]} v_s \Rightarrow \Omega^{1/2}W$ for $\lambda \in [0,1]$ where, $\Omega^{1/2} = \begin{bmatrix} \Omega_{11}^{1/2} & 0 \\ \omega_{21}\Omega_{11}^{-1/2} & \omega_{2.1}^{1/2} \end{bmatrix}$, $\omega_{2.1} = \omega_{22} - \omega_{21}\Omega_{11}^{-1}\omega_{12}$, and $W' = [W_1' \;\; W_2]'$. Using this notation $T^{-1/2}\sum_{s=1}^{[T\lambda]} v_{2s} \Rightarrow \omega_{21}\Omega_{11}^{-1/2}W_1 + \omega_{2.1}^{1/2}W_2$. Define $\bar{\delta}' = \omega_{2.1}^{-1/2}\omega_{21}\Omega_{11}^{-1/2}$ so that $\bar{\delta}'\bar{\delta} = R^2/(1-R^2)$; then $T^{-1/2}\sum_{s=1}^{[T\lambda]} \bar{\delta}'\Omega_{11}^{-1/2}v_{1s} \Rightarrow \bar{\delta}'W_1 = \sqrt{R^2/(1-R^2)}\tilde{W}_1$, where $\tilde{W}_1$ is a univariate standard Brownian Motion independent of W_2. From Phillips (1987b) and the multivariate FCLT $\omega_{2.1}^{-1/2}T^{-1/2}u_{[T\lambda]} \Rightarrow \sqrt{R^2/(1-R^2)}\tilde{W}_{1c} + W_{2c} = J_{12c}$. Results (1)–(3) follow from the Continuous Mapping Theorem (CMT thereafter), while (4) and (5) follow directly from Chan and Wei (1988) or Phillips (1987b). □

Proof of Lemma 1. $\hat{\gamma}$ can be estimated by regressing y_t^d on x_t^d. For case (i) $x_t^d = x_t$ and $y_t^d = y_t$, for case (ii) both variables are demeaned, and for case (iii) both variables are demeaned and detrended. Since $(\hat{\gamma} - \gamma) = \left(T^{-2}\sum_{t=1}^{T} x_t^d x_t^{d\prime}\right)^{-1}\left(T^{-2}\sum_{t=1}^{T} x_t^d u_t^d\right)$, from Lemma A.1 and the CMT $(\hat{\gamma} - \gamma) \Rightarrow \omega_{2.1}^{1/2}\Omega_{11}^{-1/2}\left(\int W_1^d W_1^{d\prime}\right)^{-1}\left(\int W_1^d J_{12c}^d\right)$. □

Before proving Theorem 1, I will introduce some auxiliary results.

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

(1) $\omega_{2.1}^{-1}T^{-2}\sum_{t=k+2}^{T} \hat{u}_t^{d2} \Rightarrow \eta_c^{d\prime}A_c^d\eta_c^d$,
(2) $\omega_{2.1}^{-1}T^{-1}\sum_{t=k+2}^{T} \hat{u}_{t-1}^d\Delta\hat{u}_t^d \Rightarrow cd(1)\eta_c^{d\prime}A_c^d\eta_c^d + d(1)\eta_c^{d\prime}\int W_c^{d\prime}d\tilde{W}'\eta_c^d$,
(3) $s_{\xi_k}^2 \Rightarrow d(1)^2\omega_{2.1}\eta_c^{d\prime}D\eta_c^d$,

where $\hat{u}_t$ *are the residuals from the cointegration regression* (2.4) *and* $s_{\xi_k}^2 = T^{-1}\sum_{t=k+2}^{T} \hat{\xi}_{tk}^2$ *is the estimated variance of the residuals of the ADF regression* (2.5).

$$\eta_c^{d\prime} = \left[-\left(\int W_1^{d\prime}J_{12c}^d\right)\left(\int W_1^d W_1^{d\prime}\right)^{-1} \quad 1\right],$$

$$A_c^d = \int W_c W_c' = \begin{bmatrix} \int W_1^d W_1^{d\prime} & \int W_1^d J_{12c}^d \\ \int W_1^{d\prime}J_{12c}^d & \int J_{12c}^{d2} \end{bmatrix},$$

$$W_c^d = [W_1^{d\prime}\ J_{12c}^d]', \quad \tilde{W} = \begin{bmatrix} W_1 \\ \sqrt{\dfrac{R^2}{1-R^2}}\tilde{W}_1 + W_2 \end{bmatrix} \text{ and } D = \begin{bmatrix} I & \bar{\delta} \\ \bar{\delta}' & 1+\bar{\delta}'\bar{\delta} \end{bmatrix}.$$

Proof of Corollary A.1. (1) By OLS projections, $\hat{u}_t = \hat{u}_t^d = u_t^d - (\hat{\gamma}-\gamma)'x_t^d$. By Lemma A.1 and the CMT, $\omega_{2.1}^{-1/2}T^{-1/2}\hat{u}_{[T.]}^d \Rightarrow \eta_c^{d\prime}W_c^d$ and $\omega_{2.1}^{-1}T^{-2}\sum_{t=k+2}^T \hat{u}_t^{d2} \Rightarrow \eta_c^{d\prime}\int W_c^d W_c^{d\prime}\eta_c^d = \eta_c^{d\prime}A_c^d\eta_c^d$ when $k/T \to 0$ which is true under Assumption 2.

(2) Follows from the same exact argument of Phillips and Ouliaris (1990), with the additional piece $(\rho-1)u_{t-1}$. The theory used in this paper is within the framework of Phillips (1987b, 1988). It can be shown that $\Delta\hat{u}_t^d = \Delta u_t^d - (\hat{\gamma}-\gamma)'\Delta x_t^d$, so that $\Delta\hat{u}_t^d = (\rho-1)\hat{u}_{t-1}^d + v_{2t}^d - (\hat{\gamma}-\gamma)'v_{1t}^d + (\hat{\gamma}-\gamma)'(\rho-1)x_{t-1}^d$.

Write $v_{2t}^d - (\hat{\gamma}-\gamma)'v_{1t}^d = \hat{b}'v_t^d.\hat{b}'v_t^d \Rightarrow \eta_c^{d\prime}\Gamma v_t^d = \varsigma_t^d$ where $\Gamma = \begin{bmatrix} \Omega_{11}^{-1/2}\omega_{2.1}^{1/2} & 0 \\ 0 & 1 \end{bmatrix}$.

Following Phillips and Ouliaris (1990, p. 183), $\varpi_t^d = d(L)\varsigma_t^d$ is an absolute summable sequence. The weak convergence in Lemma A.1 is R-Mixing, so that η_c^d is asymptotically independent of v_t and we can condition on η_c^d without affecting the probability of events within the sigma-algebra of v_t. The variance of the orthogonal sequence ϖ_t^d given η_c^d can then be written as $d(1)^2\eta_c^{d\prime}\Gamma\Omega\Gamma'\eta_c^d$. The ADF procedure requires the lag order in the augmented residuals regression to be large enough to capture the correlation structure of the errors. As Said and Dickey (1984) originally showed, in the context of unit root tests, it is required that $k = o(T^{1/3})$. A similar argument can be applied in the local to unity case (Xiao and Phillips, 1998). If Assumption 2 is valid, it can then be shown that, conditionally on η_c^d,

$\Delta\hat{u}_t^d = d(1)(\rho-1)\hat{u}_{t-1}^d + (\hat{\gamma}-\gamma)'(\rho-1)d(L)x_{t-1}^d + \bar{d}(L)\Delta\hat{u}_t^d + \varpi_t^d$, where $\bar{d}(L)$ is a polynomial of order k and

$$\begin{aligned}
&(\rho-1)\omega_{2.1}^{-1}T^{-1}\sum_{t=k+2}^{T}\hat{u}_{t-1}^d x_{t-1}^d \\
&=\omega_{2.1}^{-1}cT^{-2}\sum_{t=k+2}^{T}\hat{u}_{t-1}^d x_{t-1}^d \\
&=\omega_{2.1}^{-1}cT^{-2}\sum_{t=k+2}^{T}u_{t-1}^d x_{t-1}^d - \omega_{2.1}^{-1}cT^{-2}\sum_{t=k+2}^{T}x_{t-1}^d x_{t-1}^{d}{}'(\hat{\gamma}-\gamma) \\
&\Rightarrow \omega_{2.1}^{-1/2}\Omega_{11}^{1/2}c\int W_1^d J_{12c}^d \\
&\quad -\omega_{2.1}^{-1/2}\Omega_{11}^{1/2}c\left(\int W_1^d W_1^{d\prime}\right)\left(\int W_1^d W_1^{d\prime}\right)^{-1}\int W_1^d J_{12c}^d = 0.
\end{aligned}$$

If $\frac{k}{T} \to 0$, $(\hat{\gamma}-\gamma)'(\rho-1)\Omega_{11}^{-1/2}\omega_{2.1}^{-1/2}T^{-1}\sum_{t=k+2}^T \hat{u}_{t-1}^d x_{t-1}^d \Rightarrow 0$.

Using the above results, we have that

$$\omega_{2.1}^{-1}T^{-1}\sum_{t=k+2}^{T}\hat{u}_{t-1}^{d}\Delta\hat{u}_{t}^{d}=cd(1)\omega_{2.1}^{-1}T^{-2}\sum_{t=k+2}^{T}\hat{u}_{t-1}^{d2}+\omega_{2.1}^{-1}T^{-1}\sum_{t=k+2}^{T}\hat{u}_{t-1}^{d}\varpi_{t}^{d}+\mathrm{o}_{p}(1).$$

The first piece converges to $cd(1)\eta_c^{d\prime}A_c^d\eta_c^d$ by (i). For the second piece, note that the variance of ϖ_t given η_c^d can be written as $d(1)^2\eta_c^{d\prime}\Gamma\Omega\Gamma'\eta_c^d$, where $\Gamma\Omega\Gamma'$ is the variance covariance matrix of $\omega_{2.1}^{1/2}\begin{bmatrix} W_1 \\ \bar{\delta}' W_1 + W_2\end{bmatrix}$. Under Assumption 1 and since $k/T\to 0$ under Assumption 2,

$$\omega_{2.1}^{-1}T^{-1}\sum_{t=k+2}^{T}\hat{u}_{t-1}^{d}\varpi_{t}^{d\prime}\Rightarrow d(1)^{2}\eta_{c}^{d\prime}\int W_{c}^{d}\,d\tilde{W}\eta_{c}^{d},$$

where

$$\tilde{W}'=[W_1'\ \bar{\delta}'W_1+W_2]'=\left[W_1'\quad \sqrt{\frac{R^2}{1-R^2}}\,\tilde{W}_1+W_2\right]'.$$

(3) Under Assumption 2, the estimates of the ADF regression are consistent. $\hat{\alpha}\to\rho-1$ and

$$s_{\xi_k}^{2}=T^{-1}\sum_{t=k+2}^{T}\hat{\xi}_{tk}^{2}=T^{-1}\sum_{t=k+2}^{T}\varpi_{t}^{d2}+\mathrm{o}_{p}(1)\Rightarrow d(1)^{2}\eta_{c}^{d\prime}\Gamma\Omega\Gamma'\eta_{c}^{d}$$

$$=d(1)^{2}\omega_{2.1}\eta_{c}^{d\prime}\begin{bmatrix} I & \bar{\delta} \\ \bar{\delta}' & 1+\bar{\delta}'\bar{\delta}\end{bmatrix}\eta_{c}^{d}. \qquad \square$$

Proof of Theorem 1. The ADF test statistic is the usual t ratio test in the augmented regression $\Delta\hat{u}_t=\alpha\hat{u}_{t-1}+\sum_1^k\pi_i\Delta\hat{u}_{t-i}+\xi_{tk}$, or $t_\alpha=T^{-1}\hat{u}'_{-1}M_U\Delta\hat{u}/s_v[T^{-2}\hat{u}'_{-1}M_U\hat{u}_{-1}]^{1/2}$, where $\hat{u}_{-1}$ and $\Delta\hat{u}$ are the vectors of observations $\hat{u}_{t-1}^d$ and $\Delta\hat{u}_t^d$, U is the matrix of observation on the k regressors $[\Delta\hat{u}_{t-1}^d\ \Delta\hat{u}_{t-2}^d\cdots\Delta\hat{u}_{t-k}^d]$, $M_U=I-U(U'U)^{-1}U'$ and $s_{\xi_k}^2=T^{-1}\sum_{t=k+2}^{T}\xi_t^2$. Under Assumptions 1 and 2, and the conditions for Lemma A.1, and Lemma 1,

$$t_\alpha=\frac{\omega_{2.1}^{-1}T^{-1}\sum_{t=k+2}^{T}\hat{u}_{t-1}^{d}\Delta\hat{u}_{t}^{d}}{\omega_{2.1}^{-1/2}s_{\xi_k}\left[\omega_{2.1}^{-1}T^{-2}\sum_{t=k+2}^{T}\hat{u}_{t-1}^{d2}\right]^{1/2}}+o_p(1).$$

By Corollary A.1,

$$t_\alpha\Rightarrow c\,\frac{d(1)[\eta_c^{d\prime}A_c^d\eta_c^d]^{1/2}}{d(1)[\eta_c^{d\prime}D\eta_c^d]^{1/2}}+\frac{d(1)\eta_c^{d\prime}\int W_c^d\,d\tilde{W}'\eta_c^d}{[d(1)^2\eta_c^{d\prime}D\eta_c^d]^{1/2}[\eta_c^{d\prime}A_c^d\eta_c^d]^{1/2}}. \qquad \square$$

To prove Theorem 2, I shall first provide some auxiliary results. For easiness of notation, I will consider the case in which $d_{2t}=0$ (Cases (i) and (ii) of this paper). The extension to the case of a mean in Eq. (2.7) is straightforward. Following the same methodology of Sims et al. (1990) rewrite Eq. (2.7) as

$$\Delta y_t = \pi' w_{tk} + \xi_{2tk}, \quad \xi_{2tk} = \xi_{2t} + \sum_{j>k} \pi_j' \Delta x_{ct-j},$$

where $\pi_j' = [\pi_{1j}' \;\; \pi_{2j}]$, $\pi_{20}=0$, $x_{ct}=[x_t' \;\; y_t]'$, $w_{tk}=[u_{t-1} \;\; \Delta X']'$, $\Delta X' = [\Delta x_{ct}' \;\; \Delta x_{ct-1}' \;\; \cdots \;\; \Delta x_{ct-k}']$ and $\pi' = [\varphi_0 \;\; \pi_1' \;\; \cdots \;\; \pi_k']$.

The proof closely follows Berk (1974), Said and Dickey (1984) and Saikkonen (1991). As Berk (1974), I use the standard Euclidean norm $\|x\| = (x'x)^{1/2}$ of a column vector x to define a matrix norm $\|B\|$ such that $\|B\| = \sup\{\|Bx\|\colon x<1\}$. Note that $\|B\|^2 \leqslant \sum_{ij} b_{ij}$ and that $\|B\|$ is dominated by the largest modulus of the eigenvalues of B.

Let Υ denote the diagonal matrix of dimension $(1+n_1)(1+k)$

$$\Upsilon = \begin{bmatrix} T-k-1 & 0 & \cdots & 0 \\ 0 & (T-k-1)^{1/2} I_{n_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (T-k-1)^{1/2} \end{bmatrix}$$

and $\hat{R} = \Upsilon^{-1}\left(\sum_{t=k+2}^{T} w_{tk} w_{tk}'\right)\Upsilon^{-1}$. We are interested in the difference between $\hat{R}$ and $R = \operatorname{diag}\left[(T-k-1)^{-2}\sum_{t=k+2}^{T} u_{t-1}^2 \;\; \Gamma_{\Delta X}\right]$, where $\Gamma_{\Delta X} = \mathrm{E}[\Delta X \;\; \Delta X']$.

Lemma A.2. $\|\hat{R}-R\| = \mathrm{O}_p(k^2/T)$.

Proof. Denote $Q=[q_{ij}]=\hat{R}-R$. By definition, $q_{11}=0$. When $i>1, j>1$, Dickey and Fuller (1984) show that $(T-k-1)\mathrm{E}[q_{ij}^2] \leqslant C$ for some $0<C<\infty$, and it is independent of i, j and T. Since Q has dimension $(1+n_1)(1+k)$, $\mathrm{E}(\|Q\|^2) \leqslant (k+1)^2 (1+n_1)^2/(T-k-1)$. So, if $\frac{k^2}{T}\to 0$, $\|Q\|$ converges in probability to zero. □

Lemma A.3. $\|R^{-1}\| = \mathrm{O}_p(1)$.

Proof. Since R^{-1} is block diagonal, $\|R^{-1}\|$ is bounded by the sum of the norms of the diagonal blocks. Under Lemma A.1 and if $k/T\to 0$, $(T-k-1)^{-2}\sum_{t=k+2}^{T} u_{t-1}^2 \Rightarrow \omega_{2.1}\int J_{12c}^2$, while the lower right corner of R^{-1} is $\Gamma_{\Delta X}^{-1}$, which is bounded since ΔX is stationary. □

Lemma A.4. $\|\hat{R}^{-1}-R^{-1}\| = \mathrm{O}_p(k/T^{1/2})$

Proof. The proof follows directly from Said and Dickey (1984). □

Denote $e_t = \sum_{j>k} \pi_j' \Delta x_{ct-j}$ so that $\xi_{2tk} = \xi_{2t} + e_t$. Note that $\mathrm{E}\|e_t\|^2 \leqslant c \left(\sum_{j=k+2}^{\infty} \|\pi_j\| \right)^2$.

Lemma A.5.

$$\left\| \Upsilon^{-1} \sum_{t=k+2}^{T} w_{tk} e_t \right\| = \mathrm{O}_p(k^{1/2}).$$

Proof. $\mathrm{E}\|\Upsilon^{-1} \sum_{t=k+2}^{T} w_{tk} e_t\|^2 = \mathrm{E}\|(T-k-1)^{-1} \sum_{t=k+2}^{T} u_{t-1} e_t\|^2 + \mathrm{E}\|(T-k-1)^{-1/2} \sum_{t=k+2}^{T} \Delta X e_t\|^2$ $\mathrm{E}\|(T-k-1)^{-1} \sum_{t=k+2}^{T} u_{t-1} e_t\|^2 = \mathrm{E}((T-k-1)^{-1} \sum_{t=k+2}^{T} u_{t-1} e_t)^2$, where $e_t = \sum_{j>k} \pi_j' \Delta x_{ct-j} = \sum_{j>k} (\pi_{1j}' + \pi_{2j}\gamma') v_{1t-j} + (\rho-1) \sum_{j>k} \pi_{2j} v_{1t-j} + \sum_{j>k} \pi_{2j} v_{2t-j}$. Since $(\rho-1) = c/T$, under Lemma A.1, if $k/T \to 0$, $(T-k-1)^{-1} \sum_{t=k+2}^{T} u_{t-1} e_t = \mathrm{O}_p(1)$ and $\mathrm{E}\|(T-k-1)^{-1} \sum_{t=k+2}^{T} u_{t-1} e_t\|^2 = \mathrm{O}_p(1)$. Additionally,

$$\begin{aligned}
\mathrm{E} &\left\| (T-k-1)^{-1/2} \sum_{t=k+2}^{T} \Delta x_{ct-j} e_t \right\|^2 \\
&\leqslant (T-k-1)^{-1} \sum_{t=k+2}^{T} \mathrm{E}\|\Delta X e_t\|^2 \\
&\leqslant (T-k-1)^{-1} \sum_{t=k+2}^{T} \mathrm{E}\|\Delta X\|^2 E\|e_t\|^2 \\
&\leqslant c(T-k-1)^{-1} \sum_{t=k+2}^{T} \mathrm{tr}(\Gamma_{\Delta X}) \left(\sum_{j>k} \|\pi_j\| \right)^2 \\
&\leqslant c(k+1)\, \mathrm{tr}(\Gamma_{\Delta x_c}) \sum_{j>k} \|\pi_j\|^2.
\end{aligned}$$

Under Assumption 1, $\sum_{j>k} \|\pi_j\|^2$ is bounded, $\mathrm{E}\|(T-k-1)^{-1/2} \sum_{t=k+2}^{T} \Delta x_{ct-j} e_t\|^2$ is $\mathrm{O}_p(k)$, $E\|\Upsilon^{-1} \sum_{t=k+1}^{T} w_{tk} e_t\|^2 = \mathrm{O}_p(k)$ and $\|\Upsilon^{-1} \sum_{t=k+2}^{T} w_{tk} e_t\| = \mathrm{O}_p(k^{1/2})$. □

Lemma A.6.

$$\left\| \Upsilon^{-1} \sum_{t=k+2}^{T} w_{tk} \xi_{2t} \right\| = \mathrm{O}_p(k^{1/2}).$$

Proof.

$$\mathrm{E} \left\| \Upsilon^{-1} \sum_{t=k+2}^{T} w_{tk} \xi_{2t} \right\|^2 = \mathrm{E} \left\| (T-k-1)^{-1} \sum_{t=k+2}^{T} u_{t-1} \xi_{2t} \right\|^2$$

$$+\mathrm{E}\left\|(T-k-1)^{-1/2}\sum_{t=k+2}^{T}\Delta X\xi_{2t}\right\|^2.$$

$$\mathrm{E}\left\|(T-k-1)^{-1}\sum_{t=k+2}^{T}u_{t-1}\xi_{2t}\right\|=\mathrm{E}\left((T-k-1)^{-1}\sum_{t=k+2}^{T}u_{t-1}\xi_{2t}\right)^2=\mathrm{O}_p(1)$$

under Lemma A.1 and given that $k/T\to 0$. Since Δx_t and Δy_t are stationary and independent of ξ_{2t} for $j>0$,

$$\begin{aligned}&\mathrm{E}\left\|(T-k-1)^{-1/2}\sum_{t=k+2}^{T}\Delta x_{ct-j}\xi_{2t}\right\|^2\\ &\quad\leqslant (T-k-1)^{-1}\sum_{t=k+2}^{T}\mathrm{E}\|\Delta X\|^2\mathrm{E}\|\xi_{2t}\|^2\\ &\quad=(k+1)\,\mathrm{tr}(\Gamma_{\Delta x_c})\sigma^2=\mathrm{O}_p(k);\end{aligned}$$

so $\mathrm{E}\|\Upsilon^{-1}\sum_{t=k+2}^{T}w_{tk}\xi_{2t}\|=\mathrm{O}_p(k^{1/2})$, by Markov's inequality. $\square$

Proof of Theorem 2. Write

$$\begin{aligned}\Upsilon(\hat{\pi}-\pi)&=\hat{R}^{-1}\Upsilon^{-1}\sum_{t=k+2}^{T}w_{tk}\xi_{2tk}=(\hat{R}^{-1}-R^{-1})\Upsilon^{-1}\sum_{t=k+2}^{T}w_{tk}\xi_{2tk}\\ &\quad+R^{-1}\Upsilon^{-1}\sum_{t=k+2}^{T}w_{tk}\xi_{2tk}\\ &=(\hat{R}^{-1}-R^{-1})\Upsilon^{-1}\sum_{t=k+2}^{T}w_{tk}\xi_{2t}-(\hat{R}^{-1}-R^{-1})\Upsilon^{-1}\sum_{t=k+2}^{T}w_{tk}e_t\\ &\quad+R^{-1}\Upsilon^{-1}\sum_{t=k+2}^{T}w_{tk}\xi_{2tk}\\ &=E_1+E_2+E_3.\end{aligned}$$

By Lemmas A.4–A.6, if $k^3/T\to 0$, both $\|E_1\|$ and $\|E_2\|$ are of order $\mathrm{o}_p(1)$. Since R^{-1} is block diagonal,

$$\begin{aligned}(T-k-1)(\hat{\varphi}_0-\varphi)&=\left(\frac{1}{T-k-1}\sum_{t=k+2}^{T}u_{t-1}^2\right)^{-1}\left(\frac{1}{T-k-1}\sum_{t=k+2}^{T}u_{t-1}\xi_{2tk}\right)\\ &\quad+\mathrm{o}_p(1)\end{aligned}$$

with

$$\left(\frac{1}{T-k-1}\sum_{t=k+2}^{T} u_{t-1}\xi_{2tk}\right) = \left(\frac{1}{T-k-1}\sum_{t=k+2}^{T} u_{t-1}\xi_{2t}\right) + o_p(1).$$

(See also Ng and Perron (1995, p. 278)). Under Assumption 1, by Chan and Wei (1988), Phillips (1987a, b) and Lemma A.1, if $k/T \to 0$,

$$\frac{1}{T-k-1}\sum_{t=k+2}^{T} u_{t-1}\xi_{2t} \Rightarrow \theta\omega_{2.1}\int J_{12c}^2\,\mathrm{d}W_2$$

and

$$(T-k-1)(\hat{\varphi}_0 - \varphi_0) \Rightarrow (\omega_{2.1}\int J_{12c}^2)^{-1}\left(\theta\omega_{2.1}\int J_{12c}^2\,\mathrm{d}W_2\right).$$

Define $\hat{s}^2_{\xi_{2tk}} = (T-k-1)^{-1}\sum_{t=k+2}^{T}\hat{\xi}^2_{2tk}$. By a similar argument (see also Said and Dickey (1984)), $\hat{s}_{\xi_{2tk}}$ converges in probability to $\sigma = \theta\omega_{2.1}^{1/2}$; so $(T-k-1)\mathrm{SE}(\hat{\varphi}_0) \Rightarrow \theta\left[\int J_{12c}^2\right]^{-1/2}$. Since

$$\begin{aligned} t^{\mathrm{EC}}_{\hat{\varphi}_0} &= \frac{(T-k-1)\varphi_0}{(T-k-1)\mathrm{SE}(\hat{\varphi}_0)} + \frac{(T-k-1)(\hat{\varphi}_0-\varphi_0)}{(T-k-1)\mathrm{SE}(\hat{\varphi}_0)} \\ &= \frac{(T-k-1)(\rho-1)\theta}{(T-k-1)\mathrm{SE}(\hat{\varphi}_0)} + \frac{(T-k-1)(\hat{\varphi}_0-\varphi_0)}{(T-k-1)\mathrm{SE}(\hat{\varphi}_0)} \\ &= \frac{T(\rho-1)\theta}{(T-k-1)\mathrm{SE}(\hat{\varphi}_0)} + \frac{(T-k-1)(\hat{\varphi}_0-\varphi_0)}{(T-k-1)\mathrm{SE}(\hat{\varphi}_0)} - \frac{k(c/T)\theta}{(T-k-1)\mathrm{SE}(\hat{\varphi}_0)}, \end{aligned}$$

$$t^{\mathrm{EC}}_{\hat{\varphi}_0} \Rightarrow \frac{c}{\left[\int J_{12c}^2\right]^{-1/2}} + \frac{\left(\int J_{12c}^2\right)^{-1}\left(\int J_{12c}\,\mathrm{d}W_2\right)}{\left[\int J_{12c}^2\right]^{-1/2}} \quad \text{as long as } \frac{k}{T} \to 0.$$

If there are no restrictions in the deterministic terms, d_{2t} is not zero and the same proof can be applied to demeaned variables. □

Proof of Theorem 3. To compute the local asymptotic power of the ECR test, we can follow the same approach used to prove Theorem 2. Eq. (2.6) can be written as

$$\Delta y_t = \mu^* + \psi_0' x_{t-1} + \varphi_0 u_{t-1} + \sum_{j>k}\pi_j'\Delta x_{ct-j} + \xi_{2tk}, \tag{A.1}$$

so that the model is again in the same framework of Sims et al. (1990), with n_1+1 non-stationary and n_1k+1 stationary variables. μ^* is zero for case (i), contains a mean for case (ii), and a mean and trend for case (iii). φ_0 is defined as in (2.7). The coefficients ψ_0 on the redundant regressor are truly zero. This proof is very similar to the proof of Theorem 3 in Zivot (2000), but I generalize his results to a VAR of possibly infinite order and data-dependent choice of the truncation lag.

The model can be written in compact form as $\Delta y_t = \mu^* + w_{tk}'\pi + \xi_{2tk}$. As for Theorem 1, π can be estimated by first detrending both the right and left sides of the equation; so $\Delta y_t^d = w_{tk}^{d}{}'\pi + \xi_{2tk}^d$ where $w_{tk}^d = [h_{t-1}^{d}{}' \;\; \Delta X']'$, $h_{t-1}^{d}{}' = [x_{t-1}^{d}{}' \;\; u_{t-1}^d]$. As before, $\Delta X' = [\Delta x_{ct}' \;\; \Delta x_{ct-1}' \cdots \;\; \Delta x_{ct-k}']$, $\pi' = [\psi' \;\; \pi_1' \;\; \cdots \;\; \pi_k']$, $\psi' = [\psi_0' \;\; \varphi_0]$, and $x_{ct} = [x_t' \;\; y_t]'$.

Υ is now a diagonal matrix of dimension $(1+n_1)(1+p)+1$:

$$\Upsilon = \begin{bmatrix} (T-k-1)I_{n_1} & 0 & 0 & \cdots & 0 \\ 0 & (T-k-1) & 0 & \cdots & 0 \\ 0 & 0 & (T-k-1)^{1/2}I_{n_1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & (T-k-1)^{1/2} \end{bmatrix}.$$

Denote $\hat{R} = \Upsilon^{-1}\left(\sum_{t=k+2}^{T} w_{tk}w_{tk}'\right)\Upsilon^{-1}$ and $R = \mathrm{diag}[(T-k-1)^{-2}\sum_{t=k+2}^{T} h_{t-1}^d h_{t-1}^{d}{}' \;\; \Gamma_{\Delta X}]$, where $\Gamma_{\Delta X} = \mathrm{E}[\Delta X \;\; \Delta X']$. Lemmas A.2–A.7 still apply.

Since R^{-1} is block diagonal,

$$(T-k-1)(\hat{\psi}-\psi) = \left((T-k-1)^{-2}\sum_{t=k+2}^{T} h_{t-1}^d h_{t-1}^{d}{}'\right)^{-1} \times \left((T-k-1)^{-1}\sum_{t=k+2}^{T} h_{t-1}^d \xi_{2t}\right) + \mathrm{o}_p.\ (1)$$

By Lemma A.1 and the CMT,

$$(T-k-1)^{-2}\sum_{t=k+2}^{T} h_{t-1}^d h_{t-1}^{d}{}' \Rightarrow \begin{bmatrix} \Omega_{11}^{1/2}\left(\int W_1^d W_1^{d}{}'\right)\Omega_{11}^{1/2} & \Omega_{11}^{1/2}\omega_{2.1}^{1/2}\left(\int W_1^d J_{12c}^d\right) \\ \left(\int W_1^{d}{}' J_{12c}^d\right)\omega_{2.1}^{1/2}\Omega_{11}^{1/2} & \omega_{2.1}\int J_{12c}^{d2} \end{bmatrix},$$

while

$$(T-k-1)^{-1}\sum_{t=k+2}^{T} h_{t-1}^d \xi_{2t} \Rightarrow \begin{bmatrix} \theta\Omega_{11}^{1/2}\omega_{2.1}^{1/2}\int W_1^d\,\mathrm{d}W_2 \\ \theta\omega_{2.1}\int J_{12c}^d\,\mathrm{d}W_2 \end{bmatrix}.$$

By the inverted partitioned formula,

$$\left((T-k-1)^{-2}\sum_{t=k+2}^{T} h_{t-1}^{d} h_{t-1}^{d}{}'\right)^{-1}$$

$$\Rightarrow \frac{1}{\Delta}\begin{bmatrix} * & -\Omega_{11}^{-1/2}\omega_{2.1}^{1/2}\left(\int W_1^d W_1^{d}{}'\right)^{-1} \\ & \times \int W_1^d J_{12c}^d \\ -\int W_1^{d}{}' J_{12c}^d \left(\int W_1^d W_1^{d}{}'\right)^{-1} & \\ \times \Omega_{11}^{-1/2}\omega_{2.1}^{1/2} & 1 \end{bmatrix},$$

where

$$\Delta = \omega_{2.1}\left(\int J_{12c}^{d2} - \int W_1^{d}{}' J_{12c}^d \left(\int W_1^d W_1^{d}{}'\right)^{-1} \int W_1^d J_{12c}^d\right)$$

and

$$* = \Delta\left[\Omega_{11}^{-1/2}\left(\int W_1^d W_1^{d}{}'\right)^{-1} + \omega_{2.1}\left(\int W_1^d W_1^{d}{}'\right)^{-1}\left(\int W_1^d J_{12c}^d\right)\left(\int W_1^{d}{}' J_{12c}^d\right)\right.$$

$$\left.\times\left(\int W_1^d W_1^{d}{}'\right)^{-1}\Omega_{11}^{-1/2}\right].$$

The asymptotic distribution of $\hat{\varphi}_0$ is then given by the second element of the vector $\left((T-k-1)^{-2}\sum_{t=k+2}^{T} h_{t-1}^{d} h_{t-1}^{d}{}'\right)^{-1}\left((T-k-1)^{-1}\sum_{t=k+2}^{T} h_{t-1}^{d}\xi_{2t}\right)$, so that

$$(T-k-1)(\hat{\varphi}_0-\varphi_0) \Rightarrow \frac{\theta\left[\int J_{12c}^d dW_2 - \int W_1^{d}{}' J_{12c}^d \left(\int W_1^d W_1^{d}{}'\right)^{-1}\int W_1^d \, \mathrm{d}W_2\right]}{\int J_{12c}^{d2} - \int W_1^{d}{}' J_{12c}^d \left(\int W_1^d W_1^{d}{}'\right)^{-1}\int W_1^d J_{12c}^d}$$

and

$$(T-k-1)\mathrm{SE}(\hat{\varphi}_0) \Rightarrow \frac{\theta}{\left[\int J_{12c}^{d2} - \int W_1^{d}{}' J_{12c}^d \left(\int W_1^d W_1^{d}{}'\right)^{-1}\int W_1^d J_{12c}^d\right]^{1/2}}.$$

As before, we can decompose the t-test as

$$t_{\hat{\alpha}}^{\mathrm{ECR}} = \frac{T(\rho-1)\theta}{(T-k-1)\mathrm{SE}(\hat{\varphi}_0)} + \frac{(T-k-1)(\hat{\varphi}_0-\varphi_0)}{(T-k-1)\mathrm{SE}(\hat{\varphi}_0)} - \frac{k(c/T)\theta}{(T-k-1)\mathrm{SE}(\hat{\varphi}_0)}.$$

Under Assumption 2, $k/T \to 0$ and

$$t_{\hat{\alpha}}^{\mathrm{ECR}} \Rightarrow c\left[\int J_{12c}^{d^2} - \int W_1^{d\prime} J_{12c}^d \left(\int W_1^d W_1^{d\prime}\right)^{-1} \int W_1^d J_{12c}^d\right]^{1/2}$$

$$+\frac{\left[\int J_{12c}^d \,\mathrm{d}W_2 - \int W_1^{d\prime} J_{12c}^d \left(\int W_1^d W_1^{d\prime}\right)^{-1} \int W_1^d \,\mathrm{d}W_2\right]}{\left[\int J_{12c}^{d2} - \int W_1^{d\prime} J_{12c}^d \left(\int W_1^d W_1^{d\prime}\right)^{-1} \int W_1^d J_{12c}^d\right]^{1/2}}. \qquad \square$$

Proof of Theorem 4. Using the same parametrization of Theorem 3, the model can be written in compact form as $\Delta y_t^d = w_{tk}^{d\prime}\pi + \xi_{2tk}^d$, where w_{tk}^d and π are as defined in Theorem 3. The Wald test for the joint significance of φ_0 and φ_1 is equivalent to testing for the joint significance of φ_0 and ψ_0 in (A.1) after demeaning or detrending, and it can be written as $F = (R\Upsilon\hat{\pi})'[R\hat{V}\ R']^{-1}(R\Upsilon\hat{\pi})$. R is a $2 \times (2(n_1 + p) + 1)$ matrix defined as $R = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \end{bmatrix}$. Υ is the same as Theorem 3 and $\hat{V} = \hat{s}_{\xi_{2tk}}^2 \left(\Upsilon^{-1} \sum_{t=k+2}^T w_{tk}^d w_{tk}^{d\prime} \Upsilon^{-1}\right)^{-1}$ is the estimated variance of the LS estimates. For the purpose of this proof, the F test can be decomposed as:

$$\begin{aligned}
&(R\Upsilon\hat{\pi}')'[R\hat{V}\ R']^{-1}(R\Upsilon\hat{\pi}') \\
&\quad= (R\Upsilon(\hat{\pi} - \pi)')'[R\hat{V}\ R']^{-1}(R\Upsilon(\hat{\pi} - \pi)') \\
&\qquad + (R\Upsilon\hat{\pi})'[R\hat{V}\ R']^{-1}(R\Upsilon(\hat{\pi} - \pi)') + (R\Upsilon(\hat{\pi} - \pi)')'[R\hat{V}\ R']^{-1}(R\Upsilon\pi') \\
&\qquad + (R\Upsilon\pi')'[R\hat{V}\ R']^{-1}(R\Upsilon\pi').
\end{aligned} \tag{A.2}$$

The limit distribution of each piece of the tests can be written using the same compact notation of Theorem 3 (recall that under Assumption 2, $\hat{s}_{\xi_{2tk}}$ converges in probability to $\sigma = \theta\omega_{2.1}^{1/2}$). So $[R\hat{V}\ R']^{-1} \Rightarrow \theta^{-2}\omega_{2.1}^{-1}\Gamma\left[\int W_c^d W_c^{d\prime}\right]\Gamma$, where $\Gamma = \begin{bmatrix} \Omega_{11}^{1/2} & 0 \\ 0 & \omega_{2.1}^{1/2} \end{bmatrix}$ and $W_c^{d\prime} = [W_1^{d\prime}\ \ J_{12c}^d]$, $R\Upsilon\pi' = \begin{bmatrix} (T-k-1)\psi_0 \\ (T-k-1)\varphi_0 \end{bmatrix} = \begin{bmatrix} 0 \\ c\theta \end{bmatrix} + \mathrm{o}_p(1)$, and $R\Upsilon(\hat{\pi} - \pi)' \Rightarrow \vartheta\omega_{2.1}^{1/2}\Gamma^{-1}\left[\int W_c^d W_c^{d\prime}\right]^{-1}\int W_c^d\,\mathrm{d}W_2$. For the first element of Eq. (A.1), note that

$$[R\hat{V}\ R']^{-1}(R\Upsilon(\hat{\pi} - \pi)') \Rightarrow \theta^{-1}\omega_{2.1}^{-1/2}\Gamma\int W_c^d\,\mathrm{d}W_2$$

that implies $(R\Upsilon(\hat{\pi} - \pi)')'[R\hat{V}\ R']^{-1}(R\Upsilon(\hat{\pi} - \pi)') \Rightarrow \int W_c^{d\prime}\,\mathrm{d}W_2[\int W_c^d W_c^{d\prime}]^{-1}\int W_c^d\,\mathrm{d}W_2$ and $(R\Upsilon\pi')'[R\hat{V}\ R']^{-1}(R\Upsilon(\hat{\pi}-\pi)') \Rightarrow [0\ c\theta]\theta^{-1}\omega_{2.1}^{-1/2}\Gamma\int W_c^d\,\mathrm{d}W_2 = c\int J_{12c}^d\,\mathrm{d}W_2$. The third element is the transpose of the previous one, and finally the fourth element on the right-hand side of Eq. (A.2) converges to $c^2 \int J_{12c}^{d2}$, so that

$$F \Rightarrow c^2 \int J_{12c}^{d^2} + 2c \int J_{12c}^{d} \, \mathrm{d}W_2 + \int W_c^{d\prime} \, \mathrm{d}W_2 \left[\int W_c^d W_c^{d\prime} \right]^{-1} \int W_c^d \, \mathrm{d}W_2. \qquad \square$$

Proof of Theorem 5. The $\lambda_{\max}$ test is just a special case of the trace test. So the proof of Theorem 5 follows directly from Johansen (1995), Lutkephol and Saikkonen (1999) or Saikkonen and Lutkephol (1999, 2000). Lutkephol and Saikkonen (1999) extend the results to data-dependent rules for choosing the truncation lag. $\square$

References

Banerjee, A., Dolado, J.J., Hendry, D.F., Smith, G.W., 1986. Exploring equilibrium relationship in econometrics through static models: some Monte Carlo evidence. Oxford Bulletin of Economics and Statistics 48, 253–277.

Banerjee, A., Dolado, J.J., Galbraith, J.W., Hendry, D.F., 1993. Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data. Oxford University Press, Oxford.

Banerjee, A., Dolado, J.J., Mestre, R., 1998. Error-correction mechanism tests for cointegration in a single-equation framework. Journal of Time Series Analysis 19, 267–283.

Berk, K.N., 1974. Consisten autoregressive spectral estimates. Annals of Statistics 2, 486–502.

Berkowitz, J., Kilian, L., 2000. Recent developments in bootstrapping time series. Econometric Reviews 19, 1–48.

Beveridge, S., Nelson, C.R., 1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ‘business cycle’. Journal of Monetary Economics 7, 151–174.

Bewley, R., Yang, M., 1998. On the size and power of system tests for cointegration. The Review of Economics and Statistics 80, 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., Frances, P.H., 1992. Dynamic specification and cointegration. Oxford Bulletin of Economics and Statistics 54, 369–381.

Chan, N.H., Wei, C.Z., 1988. Limiting distributions of the least squares estimates on unstable autoregressive processes. Annals of Statistics 16, 367–401.

Chang, Y., Park, J.Y., Song, K., 2002, Bootstrapping cointegrating regressions. Rice University, Department of Economics Working Paper #2002-04.

Dickey, E.S., Fuller, D.A., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71, 599–607.

Elliott, G., 1998. On the robustness of cointegration methods when regressors almost have a unit root. Econometrica 66, 149–158.

Elliott, G., Jansson, M., Pesavento, E., 2002, Optimal power for testing potential cointegrating vectors with known parameters for nonstationarity. Emory University Department of Economics, Working Paper #03-03.

Engle, R.F., Granger, C.W., 1987. Cointegration and error correction: representation, estimation and testing. Econometrica 55, 251–276.

Ericsson, N.R., Mackinnon, J.G., 1999. Distributions of error correction tests for cointegration. International Finance Discussion Paper (655), Board of Governors of the Federal Reserve System.

Gonzalo, J., Lee, T.-H., 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.W., Newbold, P., 1974. Spurious regressions in econometrics. Journal of Econometrics 2, 111–120.

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.

Harbo, I., Johansen, S., Nielsen, B., Rahbek, A., 1998. Asymptotic inference on cointegrating rank in partial systems. Journal of Business and Economic Statistic 16, 388–399.

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. Cambridge University Press, Cambridge (Chapter 10).

Horvath, W., Watson, M., 1995. Testing for cointegration when some of the cointegrating vector are prespecified. Econometric Theory 11, 984–1014.

Inoue, A., Kilian, L., 2002. Bootstrapping autoregressive processes with possible unit roots. Econometrica 70, 377–391.

Jansson, M., Haldrup, N., 2002. Regression theory for nearly cointegrated time series. Econometric Theory 18, 1309–1335.

Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231–254.

Johansen, S., 1995. Likelihood Based Inference in Cointegrated Autoregressive Models. Oxford University Press, Oxford.

Johansen, S., Juselius, K., 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., Ericsson, N.R., Dolado, J.J., 1992. The power of cointegration tests. Oxford Bulletin of Economics and Statistics 54, 325–348.

Li, H., Maddala, G.S., 1997. Bootstrapping cointegration regressions. Journal of Econometrics 80, 297–318.

Lutkephol, H., Saikkonen, P., 1999. Order selection in testing for the cointegrating rank of VAR process. In: Engle, R.F., White, H. (Eds.), Cointegration, Causality and Forecasting. Oxford University Press, Oxford, pp. 168–199.

Lutkephol, H., Saikkonen, P., 2000. Testing for the cointegrating rank of a VAR process with a time trend. Journal of Econometrics 95, 177–198.

Ng, S., Perron, P., 1995. Unit root tests in ARMA models with data-dependent methods for selection of the truncation lag. Journal of the American Statistical Association 90, 268–281.

Ostermark, R., Hoglund, R., 1998. Simulating Competing Cointegration Tests in Bivariate Systems. Abo Akademi University, Finland.

Park, J.Y., 1990. Testing for unit roots and cointegration by variable addition. In: Fomby, T., Rhodes, G.F. (Eds.), Advances in Econometrics, Vol. 8. JAI Press Inc., Greenwich, CT, USA, pp. 107–133.

Park, J.Y., 1992. Canonical cointegration regression. Econometrica 60, 119–144.

Pesavento, E., 2001. Residuals based tests for 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., Hansen, B.E., 1990. Statistical inference in instrumental variable regression with I(1) processes. Review of Economic Studies 57, 99–125.

Phillips, P.C.B., Ouliaris, S., 1990. Asymptotic properties of residuals based tests for cointegration. Econometrica 58, 165–193.

Phillips, P.C.B., Solo, V., 1992. Asymptotics of linear processes. The Annals of Statistics 20, 971–1001.

Said, E.D., Dickey, D.A., 1984. Testing for unit roots in autoregressive moving average models of unknown order. Biometrika 71, 599–607.

Saikkonen, P., 1991. Asymptotically efficient estimation of cointegration regression. Econometric Theory 7, 1–21.

Saikkonen, P., 1992. Estimation and testing of cointegrated systems by an autoregressive approximation. Econometric Theory 8, 1–27.

Saikkonen, P., Lutkephol, H., 1999. Local power of likelihood ratio tests for the cointegrating rank of a VAR process. Econometric Theory 15, 50–78.

Saikkonen, P., Lutkephol, H., 2000. Testing for the cointegrating rank of a VAR process with an intercept. Econometric Theory 16, 373–406.

Sims, C.A., Stock, J.H., Watson, M.W., 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: Mc Fadden, D., Engle, R.F. (Eds.), Handbook of Econometrics, Vol. 4. Elsevier, Amsterdam, pp. 2739–2841.

Stock, J.H., Watson, M.W., 1988. Testing for common trends. Journal of the America Statistical Association 83, 1097–1107.

Swensen, A.R., 2003. A note on the power of bootstrap unit root tests. Econometric Theory 19, 32–48.

Watson, M.W., 1994. Vector autoregression and cointegration. In: Mc Fadden, D., Engle, R.F. (Eds.), Handbook of Econometrics, Vol. 4. North-Holland, Amsterdam, pp. 2843–2915.

Wooldridge, J., 1994. Estimation and inference for dependent processes. In: Mc Fadden, D., Engle, R.F. (Eds.), Handbook of Econometrics, Vol. 4. North-Holland, Amsterdam, pp. 2639–2738.

Xiao, Z., Phillips, P.C., 1998. An ADF coefficient test for a unit root in ARMA models of unknown order with empirical applications to the US economy. Econometrics Journal 1, 27–43.

Zivot, E., 2000. The power of single equation tests for cointegration when the cointegrating vector is prespecified. Econometric Theory 16, 407–439.