Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers

Cowles Foundation

6-1-1987

Asymptotic Properties of Residual Based Tests for Cointegration

Peter C.B. Phillips

Sam Ouliaris

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

Part of the Economics Commons

Recommended Citation

Phillips, Peter C.B. and Ouliaris, Sam, "Asymptotic Properties of Residual Based Tests for Cointegration" (1987). *Cowles Foundation Discussion Papers*. 1090.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/1090

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station

New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 847-R

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

"ASYMPTOTIC PROPERTIES OF RESIDUAL BASED TESTS FOR COINTEGRATION"

PETER C.B. PHILLIPS

and

S. OULIARIS

July 1988

ASYMPTOTIC PROPERTIES OF RESIDUAL BASED TESTS FOR COINTEGRATION

by

P. C. B. Phillips and S. Ouliaris*

Cowles Foundation for Research in Economics
Yale University

0. ABSTRACT

This paper develops an asymptotic theory for residual based tests for cointegration. These tests involve procedures that are designed to detect the presence of a unit root in the residuals of (cointegrating) regressions among the levels of economic time series. Attention is given to the augmented Dickey-Fuller (ADF) test that is recommended by Engle-Granger (1987) and the Z_α and Z_t unit root tests recently proposed by Phillips (1987). Two new tests are also introduced, one of which is invariant to the normalization of the cointegrating regression. All of these tests are shown to be asymptotically similar and simple representations of their limiting distributions are given in terms of standard Brownian motion. The ADF and Z_t tests are asymptotically equivalent. Power properties of the tests are also studied. The analysis shows that all the tests are consistent if suitably constructed but that the ADF and Z_t tests have slower rates of divergence under cointegration than the other tests. This indicates that, at least in large samples, the Z_α test should have superior power properties.

The paper concludes by addressing the larger issue of test formulation. Some major pitfalls are discovered in procedures that are designed to test a null of cointegration (rather than no cointegration). These defects provide strong arguments against the indiscriminate use of such test formulations and support the continuing use of residual based unit root tests.

First version: June 1987
Revision: July 1988

---

*We are grateful to Lars Peter Hansen and two referees for helpful comments on the original version of this paper. Our thanks also go to Glena Ames for her skill and effort in keyboarding the manuscript of this paper and to the NSF for research support under grant number SES 8519595. Sam Ouliaris is now at the University of Maryland and is presently visiting the National University of Singapore.

1. INTRODUCTION

The purpose of this paper is to provide an asymptotic analysis of residual based tests for the presence of cointegration in multiple time series. Residual based tests rely on the residuals calculated from regressions among the levels (or log levels) of economic time series. They are designed to test the null hypothesis of no cointegration by testing the null that there is a unit root in the residuals against the alternative that the root is less than unity. If the null of a unit root is rejected then the null of no cointegration is also rejected. The tests might therefore be more aptly named residual based unit root tests. Some of the tests we shall study involve standard procedures applied to the residuals of the cointegrating regression to detect the presence of a unit root. Two of the procedures we shall look at are new to this paper. They all fall within the framework of residual based unit root tests.

Approaches other than residual based tests for cointegration are also available. A likelihood ratio test has been considered by Johansen (1988) in the context of vector autoregresssions (VAR's). A common stochastic trends test has been proposed by Stock and Watson (1986). A bounds test has been suggested by the authors in earlier work (1988). None of these tests rely on the residuals of cointegrating regressions. However, it is the residual based procedures which have attracted the attention of empirical researchers. This is partly because of the recommendations of Engle and Granger (1987), partly because the tests are so easy and convenient to apply and partly because it is clear intuitively what they set out to test.

Unfortunately, little is known from existing work about the properties of residual based unit root tests. Engle and Granger (1987) provide some experimental evidence on the basis of which they recommend the use of the augmented Dickey-Fuller (ADF) t-ratio test. They also show that this test and many others are similar tests when the data follows a vector random walk driven by iid normal innovations. They conjecture that the ADF procedure is asymptotically similar in more general time series settings. The present paper confirms that conjecture. We also study the asymptotic behavior of various other tests, including the Z_α and Z_t tests recently suggested in Phillips (1987). Our asymptotic theory covers both the null of no cointegration and the alternative of a cointegrated system. It is shown that the power properties of many of the tests depend critically on their method of construction. In particular, test consistency relies on whether residuals or first differences are used in serial correlation corrections that are designed to eliminate nuisance parameters under the null. Our analysis of power also indicates some major differences between the tests. In particular, t-ratio procedures such as the ADF and the Z_t test diverge under the alternative at a slower rate than direct coefficient tests such as the Z_α test and the new variance ratio tests that are developed in the paper. This indicates that coefficient and variance ratio tests should have superior power properties over t-ratio tests at least in large samples.

One of the new tests developed in the paper is invariant to the normalization of the cointegrating regression. This is in contrast to other residual based tests, such as the ADF, which are numerically dependent on the precise formulation of the cointegrating regression. Invariance is a useful property, since it removes conflicts that can arise in empirical work where the test outcome depends on the normalization selected.

The general question of how to formulate tests for the presence of cointegration is also addressed. In particular, we examine the potential of certain procedures which seek to test a null of cointegration against an alternative of no cointegration, rather than vice versa. This question of formulation is important. It arises frequently in seminar and conference discussions (for example Engle (1987)) where it is often argued that a null of cointegration is the more appealing. But the question has not to our knowledge been formally addressed in the literature until now. Our analysis points to some major pitfalls in the alternate approach. The source of the difficulties lies in the failure of conventional asymptotic theory under a null of cointegration. This is not just a matter of nonstandard limit theory. In fact, no general limit theory applies in this case to certain statistics (like long run variance estimates) that are most relevant to the null. Moreover, if tests based on a specific distribution theory are used they turn out to be inconsistent. These difficulties provide good arguments for the continuing use of tests that are based on the composite null of no cointegration.

The plan of the paper is as follows. Section 2 provides some preliminary theory, including a theorem that is likely to be very useful on invariance principles for processes which are linear filters of other time series. This theory is needed for an asymptotic analysis of the ADF. In Section 3 we review a class of residual based tests for cointegration and develop two new procedures: a variance ratio test and a multivariate trace test. Both tests have interesting interpretations and the second has the invariance property mentioned earlier. An asymptotic theory for the tests is developed in Section 4 and it is shown that the Z_α, Z_t and ADF tests all have limiting distributions which can be simply expressed as stochastic integrals.

The ADF and Z_t tests are asymptotically equivalent. Section 5 studies test consistency and the asymptotic behavior of the tests under the alternative of cointegration. Issues of test formulation are considered in Section 6 and some conclusions are drawn in Section 7. Proofs are given in the Appendix A.

In matters of notation we use the symbol " $\Rightarrow$ " to signify weak convergence, the symbol " $\equiv$ " to signify equality in distribution and the inequality " > 0 " to signify positive definite when applied to matrices. Continuous stochastic processes such as the Brownian motion $B(r)$ on $[0,1]$ are written as B to achieve notational economy. Similarly, we write integrals with respect to Lebesgue measure such as $\int_0^1 B(s)ds$ more simply as $\int_0^1 B$.

2. PRELIMINARY THEORY

Let $\{z_t\}_0^\infty$ be an m-vector integrated process whose generating mechanism is

$$(1) \qquad z_t = z_{t-1} + \xi_t , \quad t = 1, 2 \dots .$$

Our results do not depend on the initialization of (1) and we therefore allow z_0 to be any random variable including, of course, a constant. The random sequence $\{\xi_t\}_1^\infty$ is defined on a probability space $(X, \mathcal{F}, P)$ and is assumed to be strictly stationary and ergodic with zero mean, finite variance and spectral density matrix $f_{\xi\xi}(\lambda)$. We also require the partial sum process constructed from $\{\xi_t\}$ to satisfy a multivariate invariance principle. More specifically, for $r \in [0,1]$ and as $T \to \infty$ we require

$$\text{(C1)} \qquad X_T(r) = T^{-1/2}\Sigma_1^{[Tr]}\xi_t \Rightarrow B(r) \quad \text{(R-mixing)}$$

where $B(r)$ is m-vector Brownian motion with covariance matrix

$$\text{(2)} \qquad \Omega = \lim_{T\to\infty} T^{-1}E((\Sigma_1^T\xi_t)(\Sigma_1^T\xi_t')) = 2\pi f_{\xi\xi}(0) \ .$$

Writing

$$\Omega_0 = E(\xi_0\xi_0') \ , \quad \Omega_1 = \Sigma_{k=1}^{\infty}E(\xi_0\xi_k')$$

we have

$$\Omega = \Omega_0 + \Omega_1 + \Omega_1' \ .$$

The convergence condition (C1) is Reyni-mixing (R-mixing). This requires the random element $X_T(r)$ to be asymptotically independent of each event $E \in \mathcal{F}$ i.e.

$$P(\{X_T \in \cdot\} \cap E) \to P(B \ \in \cdot)P(E) \ , \quad T \to \infty \ .$$

In this sense, the random element X_T may be thought of as escaping from its own probability space when R-mixing applies. The reader is referred to Hall and Heyde (1980, p. 57) for further discussion. Functional limit theorems under R-mixing such as (C1) are known to apply in very general situations. For example, the theorems of McLeish (1975) that were used in the paper by Phillips (1987) are all R-mixing limit theorems. Extensions to multiple time series follow as in Phillips and Durlauf (1986).

It will be convenient for much of this paper to take ξ_t to be the linear process generated by

$$\xi_t = \Sigma_{j=-\infty}^{\infty} C_j \varepsilon_{t-j} \ , \quad \Sigma_{j=-\infty}^{\infty} \|C_j\| < \infty \ , \quad C(1) = \Sigma_{j=-\infty}^{\infty} C_j \tag{3}$$

where the sequence of random vectors $\{\varepsilon_t\}$ is $\mathrm{iid}(0,\Sigma)$ with $\Sigma > 0$ and $\|C_j\| = \max_k\{\Sigma_\ell |c_{jk\ell}|\}$ where $C_j = (c_{jk\ell})$. This includes all stationary ARMA processes and is therefore of wide applicability. The process ξ_t has a continuous spectral density matrix given by

$$f_{\xi\xi}(\lambda) = (1/2\pi)(\Sigma_j C_j e^{ij\lambda})\Sigma(\Sigma_j C_j e^{ij\lambda})^* \ .$$

In addition to the absolute summability of $\{C_j\}$ in (3) we will use the following condition (based on (5.37) of Hall and Heyde (1980))

$$\Sigma_{k=1}^{\infty} [\|\Sigma_{j=k}^{\infty} C_j\| + \|\Sigma_{j=k}^{\infty} C_{-j}\|] < \infty \tag{C2}$$

which is again satisfied by all stationary ARMA models. Note that (C2) holds for all sequences $\{C_j\}$ that are 1-summable in the sense of Brillinger (1981, equation 2.7.14).

Let $\{a_j\}$ be a scalar sequence that is absolutely summable and define the new process

$$\xi_t^* = \Sigma_{j=-\infty}^{\infty} a_j \xi_{t-j} \ , \quad \Sigma_{-\infty}^{\infty} |j|^s |a_j| < \infty \ , \quad s > 1 \tag{4}$$

and the associated random element

$$X_T^*(r) = T^{-1/2} \Sigma_1^{[Tr]} \xi_t^* \ . \tag{5}$$

Let $a(1) = \Sigma_{-\infty}^{\infty} a_j$. We shall make use of the following important lemma describing the asymptotic behavior of $X_T^*(r)$.

<u>LEMMA 2.1</u>. *If (C2) holds then as* $T \to \infty$

$$\sup_{0 \le r \le 1} |X_T^*(r) - a(1)X_T(r)| \xrightarrow[P]{} 0 \tag{6}$$

and

$$X_T^*(r) \Rightarrow B^*(r) = a(1)B(r) \tag{7}$$

or vector Brownian motion with covariance matrix $\Omega^* = a(1)^2\Omega$.

We now partition $z_t = (y_t, x_t')'$ into the scalar variate y_t and the n-vector x_t $(m = n+1)$ with the following conformable partitions of Ω and $B(r)$:

$$\Omega = \begin{array}{cc} \begin{array}{cc} 1 & n \end{array} & \\ \begin{bmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{bmatrix} & \begin{array}{c} 1 \\ n \end{array} \end{array} , \quad B(r) = \begin{bmatrix} B_1(r) \\ B_2(r) \end{bmatrix} \begin{array}{c} 1 \\ n \end{array} .$$

We shall assume $\Omega_{22} > 0$ and use the block triangular decomposition of Ω

$$\Omega = L'L , \quad L = \begin{bmatrix} \ell_{11} & 0 \\ \ell_{21} & L_{22} \end{bmatrix} \tag{8}$$

with

$$\ell_{11} = \left(\omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21}\right)^{1/2} , \quad \ell_{21} = \Omega_{22}^{-1/2}\omega_{21} , \quad L_{22} = \Omega_{22}^{1/2} . \tag{9}$$

Let $W(r)$ be m-vector standard Brownian motion and define:

$$A = \int_0^1 BB' = \begin{bmatrix} a_{11} & a_{21}' \\ a_{21} & A_{22} \end{bmatrix} ,$$

$$F = \int_0^1 WW' = \begin{bmatrix} f_{11} & f_{21}' \\ f_{21} & F_{22} \end{bmatrix} ,$$

$$\eta' = (1, -a_{21}' A_{22}^{-1}) \; , \quad \kappa = (1, -f_{21}' F_{22}^{-1}) \; ,$$

$$Q(r) = W_1(r) - \left(\int_0^1 W_1 W_2'\right)\left(\int_0^1 W_2 W_2'\right)^{-1} W_2(r) \; .$$

LEMMA 2.2.

(a) $B(r) = L'W(r)$;

(b) $L\eta = \ell_{11}\kappa$, $\eta'\Omega\eta = \omega_{11\cdot 2}\kappa'\kappa$;

(c) $\eta' B(r) = \ell_{11} Q(r)$;

(d) $\eta'\int_0^1 BdB'\eta = \omega_{11\cdot 2}\int_0^1 QdQ$;

(e) $\eta' A\eta = a_{11\cdot 2} = \omega_{11\cdot 2}\int_0^1 Q^2$;

where

$$\omega_{11\cdot 2} = \omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21} = \ell_{11}^2$$

and

$$a_{11\cdot 2} = a_{11} - a_{21}' A_{11} a_{21} \; .$$

REMARKS

(a) This lemma shows how to reformulate some simple linear and quadratic functionals of the Brownian motion $B(r)$ into distributionally equivalent functionals of standard Brownian motion. These representations turn out to be very helpful in identifying key parameter dependencies in the original expressions. As is clear from (b)-(e) the conditional variance $\omega_{11\cdot 2}$ is the sole carrier of these dependencies in (b)-(e).

(b) Note that $\det \Omega = \omega_{11\cdot 2} \det \Omega_{22}$ and is zero iff $\omega_{11\cdot 2} = 0$ (given $\Omega_{22} > 0$). Note also that we may write

$$\omega_{11\cdot 2} = \omega_{11}(1 - \rho^2) \ , \quad \rho^2 = \omega_{21}' \Omega_{22}^{-1} \omega_{21} / \omega_{11}$$

where ρ^2 is a squared correlation coefficient. When $\omega_{11\cdot 2} = 0$ ($\rho^2 = 1$) then Ω is singular and y_t and x_t are cointegrated, as pointed out in Phillips (1986). At the other extreme when there is no correlation between the innovations of y_t and x_t we have $\rho^2 = 0$, Ω nonsingular and a regression of y_t on x_t is spurious in the sense of Granger and Newbold (1974).

(c) Consider the Hilbert space $L_2[0,1]$ of square integrable functions on the interval [0,1] with inner product $\int_0^1 fg$ for $f, g \in L_2(0,1)$. In this space, Q is the projection of W_1 on the orthogonal complement of the space spanned by the elements of W_2 .

3. RESIDUAL BASED TESTS OF COINTEGRATION

We consider the linear cointegrating regressions

$$y_t = \hat{\beta}' x_t + \hat{u}_t \ . \tag{10}$$

Residual based tests seek to test a null hypothesis of no cointegration using scalar unit root tests applied to the residuals of (10).

This null may be formulated in terms of the conditional variance parameter $\omega_{11\cdot 2}$ as the composite hypothesis

$$H_0 : \ \omega_{11\cdot 2} \neq 0 \ .$$

The alternative is simply

$$H_1 : \omega_{11\cdot 2} = 0$$

leading to $\rho^2 = 1$ and cointegration, as pointed out above. Engle and Granger (1987) discuss the procedure and suggest various tests. Their main recommendation is to use the augmented Dickey-Fuller (ADF) test and they provide some critical values obtained by Monte Carlo methods for the case $m = 2$.

We shall consider the asymptotic properties of the following residual based tests:

(i) Augmented Dickey Fuller:

ADF = $t_{\hat{\alpha}_*}$ in the regression $\Delta\hat{u}_t = \hat{\alpha}_*\hat{u}_{t-1} + \Sigma_{i=1}^p\hat{\varphi}_i\Delta\hat{u}_{t-i} + \hat{v}_{tp}$;

(ii) Phillips' (1987) Z_α test:

regress $\hat{u}_t = \hat{\alpha}\hat{u}_{t-1} + \hat{k}_t$ and compute

$$\hat{Z}_\alpha = T(\hat{\alpha}-1) - (1/2)(s_{T\ell}^2 - s_k^2)(T^{-2}\Sigma_2^T\hat{u}_{t-1}^2)^{-1}$$

where

$$s_k^2 = T^{-1}\Sigma_1^T\hat{k}_t^2 ,$$

(11) $$s_{T\ell}^2 = T^{-1}\Sigma_1^T\hat{k}_t^2 + 2T^{-1}\Sigma_{s=1}^{\ell}w_{s\ell}\Sigma_{t=s+1}^{T}\hat{k}_t\hat{k}_{t-s}$$

for some choice of lag window such as $w_{s\ell} = 1 - s/(\ell+1)$;

(iii) Phillips' (1987) Z_t test:

regress $\hat{u}_t = \hat{\alpha}\hat{u}_{t-1} + \hat{k}_t$ and compute

$$\hat{Z}_t = (\Sigma_2^T \hat{u}_{t-1}^2)^{1/2}(\hat{\alpha}-1)/s_{T\ell} - (1/2)(s_{T\ell}^2 - s_k^2)\left[s_{T\ell}(T^{-2}\Sigma_2^T \hat{u}_{t-1}^2)^{1/2}\right]^{-1}$$

with s_k^2 and $s_{T\ell}^2$ as in (ii);

(iv) Variance ratio test:

$$\hat{P}_u = T\hat{\omega}_{11\cdot 2}/(T^{-1}\Sigma_1^T \hat{u}_t^2) \tag{12}$$

where $\hat{\omega}_{11\cdot 2} = \hat{\omega}_{11} - \hat{\omega}_{21}'\hat{\Omega}_{22}^{-1}\hat{\omega}_{21}$ and

$$\hat{\Omega} = T^{-1}\Sigma_1^T \hat{\xi}_t \hat{\xi}_t + T^{-1}\Sigma_{s=1}^{\ell} w_{s\ell}\Sigma_{t=s+1}^T (\hat{\xi}_t \hat{\xi}_{t-s}' + \hat{\xi}_{t-s}\hat{\xi}_t') \tag{13}$$

for some choice of lag window such as $w_{s\ell} = 1 - s/(\ell+1)$ (see Phillips and Durlauf (1986), Newey and West (1987)) and where $\{\hat{\xi}_t\}$ are the residuals from the least squares regression:

$$z_t = \hat{\Pi} z_{t-1} + \hat{\xi}_t \; ; \tag{14}$$

(v) A multivariate trace statistic:

$$\hat{P}_z = T\ \mathrm{tr}(\hat{\Omega}\, M_{zz}^{-1}) \ , \quad M_{zz} = T^{-1}\Sigma_1^T z_t z_t'$$

where $\hat{\Omega}$ is as in (13).

REMARKS

(a) Note that $\hat{Z}_\alpha$ and $\hat{Z}_t$ are constructed using an estimate $s^2_{T\ell}$ that is based on the residuals $\hat{k}_t$ from the autoregression of $\hat{u}_t$ on $\hat{u}_{t-1}$. When the estimate $s^2_{T\ell}$ is based on first differences $\Delta\hat{u}_t$ in place of $\hat{k}_t$ (as suggested by the null of no cointegration) we shall denote the resulting tests by Z_α and Z_t. The distinction is important since these tests have very different properties under the alternative hypothesis of cointegration, as we see below.

(b) In a similar way, $\hat{P}_u$ and $\hat{P}_z$ are constructed using the covariance matrix estimate $\hat{\Omega}$ that is based on the residuals $\hat{\xi}_t$ from the first order vector autoregression (14). When the estimate $\hat{\Omega}$ is based on first differences $\xi_t = \Delta z_t$ we denote the resulting tests by P_u and P_z. Again the distinction is important since $\hat{P}_u$ and $\hat{P}_z$ have different properties under the alternative from those of P_u and P_z.

(c) The variance ratio test $\hat{P}_u$ is new. Its construction is intuitively appealing. $\hat{P}_u$ measures the size of the residual variance from the cointegrating regression of y_t on x_t, viz. $T^{-1}\Sigma_1^T\hat{u}_t^2$, against that of a direct estimate of the population conditional variance of y_t given x_t, viz. $t\hat{\omega}_{11\cdot2}$. If the model (1) is correct and has no degeneracies (i.e. Ω nonsingular) then the variance ratio should stabilize asymptotically. If there is a degeneracy in the model, then this will be picked up by the cointegrating regression and the variance ratio should diverge.

(d) The multivariate trace statistic $\hat{P}_z$ is also new. Its appeal is similar to that of $\hat{P}_u$. Thus, $t\hat{\Omega}$ is a direct estimate of the covariance matrix of z_t, while M_{zz} is simply the observed sample moment matrix. Any degeneracies in the model such as cointegration ultimately manifest themselves in the behavior of M_{zz} and, hence, that of the statistic $\hat{P}_z$. This behavior will be examined in detail below. Note that $\hat{P}_z$ is constructed in the form of Hotelling's T_0^2 statistic, which is a common statistic (see, e.g., Muirhead (1982, Chapter 10)) in multivariate analysis for tests of multivariate dispersion.

(e) Note that none of the tests (i)-(iv) are invariant to the formulation of the regression equation (10). Thus, for these tests, different outcomes will occur depending on the normalization of the equation. One way around this problem is to employ regression methods in fitting (10) which are invariant to normalization. The obvious candidate is orthogonal regression, leading to

$$\tilde{b}'z_t = \hat{u}_t \ , \tag{15}$$

where

$$\tilde{b} = \arg\min b'M_{zz}b \ , \quad b'b = 1 \ .$$

Here $\tilde{b}$ is the direction of smallest variation in the observed moment matrix M_{zz} and corresponds to the smallest principal component with

$$T^{-1}\Sigma_1^T\tilde{u}_t^2 = \tilde{b}'M_{zz}\tilde{b} = \lambda_{min}(M_{zz})$$

where $\lambda_{min}(M_{zz})$ is the smallest latent root of M_{zz}. Smallest latent root tests based on $\lambda_{min}(M_{zz})$ may be constructed. For example, an orthogonal regression version of the variance ratio statistic $\hat{P}_u$ would be:

$$\hat{P}_\lambda = T\,\lambda_{min}(\hat{\Omega})/\lambda_{min}(M_{zz}) \ .$$

Unfortunately, $\hat{P}_\lambda$ has a limiting distribution which depends on the nuisance parameter Ω. The multivariate trace statistic $\hat{P}_z$ offers a very convenient alternative. $\hat{P}_z$ is a normalization invariance analogue of $\hat{P}_u$ and has the same general appeal as statistics such as $\hat{P}_\lambda$; yet, as we see below, its asymptotic distribution is free of nuisance parameters.

(f) Each of the test statistics (i)-(iv) has been constructed using the residuals $\hat{u}_t$ of the least squares regression (10). These statistics may also be constructed using the residuals $\bar{u}_t$ of the least squares regression

$$y_t = \bar{\alpha} + \bar{\beta}'x_t + \bar{u}_t \tag{16}$$

with a fitted intercept. In a similar way, for test (v) the statistic $\hat{P}_z$ may be constructed using $\bar{M}_{zz} = T^{-1}\Sigma_1^T(z_t - \bar{z})(z_t - \bar{z})'$ and residuals $\hat{\bar{\xi}}_t$ from a VAR such as (14) with a fitted intercept. These modifications do not affect the interpretation of the tests but the alternate construction does have implications for the asymptotic critical values. These will be considered below.

4. ASYMPTOTIC THEORY

Our first concern is to develop a limiting distribution theory for the tests (i)-(v) under the null of no cointegration. In this case, the covariance matrix Ω is positive definite. The statistic that presents the main difficulty in this analysis is the ADF. We shall give the asymptotic theory for this test separately in the second result below.

THEOREM 4.1. *If* $\{z_t\}_0^\infty$ *is generated by* (1), *if* $\Omega > 0$ *and if* (C1) *holds then as* $T \to \infty$:

(a) $\hat{Z}_\alpha \Rightarrow \int_0^1 R dR$;

(b) $\hat{Z}_t \Rightarrow \int_0^1 R dS$;

(c) $\hat{P}_u \Rightarrow (\int_0^1 Q^2)^{-1}$;

(d) $\hat{P}_z \Rightarrow \mathrm{tr}((\int_0^1 WW')^{-1})$

where notations are the same as in Lemma 2.2 *and*

$$R(r) = Q(r)/(\int_0^1 Q^2)^{1/2} \quad ,$$

$$S(r) = Q(r)/(\kappa'\kappa)^{1/2} \quad .$$

We require the lag truncation parameter $\ell \to \infty$ *as* $T \to \infty$ *and* $\ell = o(T)$.

REMARKS

(a) Recall from Lemma 2.2 that

$$Q(r) = W_1(r) - \int_0^1 W_1 W_2' \left[\int_0^1 W_2 W_2'\right]^{-1} W_2(r)$$

whose distribution depends on a single parameter n = dimension of W_2 . Recall too that

$$\kappa' = \left[1, -\int_0^1 W_1 W_2' (\int_0^1 W_2 W_2')^{-1}\right]$$

whose distribution is also independent of nuisance parameters.

(b) We deduce from the preceding remark that the limiting distributions of $\hat{Z}_\alpha$, $\hat{Z}_t$, $\hat{P}_u$ and $\hat{P}_z$ are free of nuisance parameters and are dependent only on the known dimension number n (or $m = n+1$ in the case of $\hat{P}_z$). These statistics therefore lead to (asymptotically) similar tests. Critical values for these statistics have been computed by simulation and are reported in Appendix B. For the case of the $\hat{Z}_t$ statistic demeaned (i.e. computed from the regression (16) with a fitted intercept) the values in Table 2b in Appendix B correspond closely with those reported by Engle and Yoo (1987, Table 2, p. 157) for the Dickey Fuller t statistic. Differences occur only at the second decimal place and are likely to be the result of: (i) differences in the actual sample sizes used in the simulations ($T = 200$ in Engle and Yoo (1987); and $T = 500$ in ours); and (ii) sampling error.

(c) The $\hat{Z}_\alpha$ and $\hat{Z}_t$ tests have the same limiting distribution in the general case as the Dickey Fuller residual based α and t tests do in the highly restrictive case of $iid(0,\Omega)$ errors. This point is discussed further in the original version of the paper which is available as a technical report on request, Phillips and Ouliaris (1987). Thus, the $\hat{Z}_\alpha$ and $\hat{Z}_t$ tests have the same property in this context of cointegrating regressions for which they were originally designed in Phillips (1987) as scalar unit root tests, viz. that they eliminate nuisance parameters and lead to limit distributions which are the same as those possessed by the Dickey-Fuller tests in the iid error environment. However, the limiting distributions here, $\int_0^1 RdR$ and $\int_0^1 RdS$, _are_ different from the simple unit root case, and they _are_ dependent on the dimension number n .

(d) We remark that the limiting distributions of Z_α , Z_t , P_u , P_z (the statistics mentioned earlier which are based on first differences

$k_t = \Delta\hat{u}_t$ and $\xi_t = \Delta z_t$ rather than regression residuals $\hat{k}_t$ and $\hat{\xi}_t$) are the same as those of $\hat{Z}_\alpha$, $\hat{Z}_t$, $\hat{P}_u$, $\hat{P}_z$ given in Theorem 4.1. This follows in a straightforward way from the proof given in the Appendix.

(e) Note, finally, that if the statistics are based on the regression (16) with a fitted intercept then the limiting distributions of $\hat{Z}_\alpha$, $\hat{Z}_t$ and $\hat{P}_u$ have the same form as in (a)-(c) but now R , S and Q are functionals of the demeaned standard Brownian motion $\bar{W}(r) = W(r) - \int_0^1 W$. Observe that $\bar{W}$ is the projection in $L_2[0,1]$ of W onto the orthogonal complement of the constant function, thereby justifying this terminology. In a similar way, if the cointegrating regression involves fitted time trends the limiting distributions in (a)-(c) continue to retain their stated form but involve functionals of correspondingly detrended standard Brownian motion.

<u>THEOREM 4.2</u>. *Let* $\{z_t\}$ *be generated by* (1) *and suppose* $\{\xi_t\}$ *follows a stationary vector ARMA process. If* $\Omega > 0$ *and* (C1) *holds then as* $T \to \infty$

$$ADF \Rightarrow \int_0^1 R\,dS$$

provided the order of the autoregression in the ADF is such that $p \to \infty$ *as* $T \to \infty$ *and* $p = o(T^{1/3})$.

<u>REMARKS</u>

(a) Theorem 4.2 shows that ADF and $\hat{Z}_t$ have the same limiting distribution. This distribution is conveniently represented as a stochastic integral in terms of the continuous stochastic processes $(R(r), S(r))$. These processes are, in turn, continuous functionals of the m-vector standard Brownian motion $W(r)$. In accord with our earlier remarks concerning $\hat{Z}_t$, the limiting distribution of the ADF depends only on the dimension number n (the number of regressors in (10) or, equivalently, the system dimension m $(= n+1)$. Given m, the ADF is an asymptotically similar test.

(b) The proof of Theorem 4.2 depends critically on the fact that the order of the autoregression $p \to \infty$. While this behavior is also required for a general unit root test in the scalar case (see Said and Dickey (1984)) it is not required when the scalar process is driven by a finite order AR model with a unit root. It is important to emphasize that this is not the case when the ADF is used as a residual based test for cointegration. Thus, we still need $p \to \infty$ even when the vector process ξ_t is driven by a finite order VAR. This is because the residuals on which the ADF is based are (random) linear combinations of ξ_t. These linear combinations no longer follow simple AR processes. In general, they satisfy (conditional) ARMA models and we need $p \to \infty$ in order to mimic their behavior.

(c) We mention one special case where the requirement $p \to \infty$ is not needed. This occurs when the elements of ξ_t are driven by a diagonal AR process of finite order viz

$$b(L)\xi_t = \varepsilon_t \ , \quad b(L) = \Sigma_{i=0}^{p} b_i L^i \ , \quad b_i = 1 \ ; \quad \varepsilon_t \ \text{iid} \ (0,\Sigma) \ .$$

In this case

$$\Omega = (1/b(1))^2 \Sigma \ ,$$

$$\Omega_0 = \left(\int_{-\pi}^{\pi} |b(e^{i\lambda})|^{-2} d\lambda\right)\Sigma$$

and we observe that Ω is a scalar multiple of Ω_0 . The examples chosen by Engle and Granger (1987) for their simulation experiments (Tables II and III in their paper) both fall within this special case.

(d) Note that the ADF test is basically a t-test in a long autoregression involving the residuals $\hat{u}_t$. In this sense, the ADF is a simple extension of the Dickey-Fuller t test. Note that no such extension of the Dickey-Fuller α test is recommended by Said and Dickey (1984) since even as $p \to \infty$ the coefficient estimate $T\hat{\alpha}_*$ has a limiting distribution that is dependent on nuisance parameters (cf. Said and Dickey (1984, p. 605)) in the scalar unit root case. In contrast, the $\hat{Z}_\alpha$ statistic is an asymptotically similar test. Thus, the nonparametric correction of the Z_α test successfully eliminates nuisance parameters asymptotically even in the case of cointegrating regressions. This point will be of some importance later when we consider the power of these various tests.

5. TEST CONSISTENCY

Our next concern is to consider the behavior of the tests based on $\hat{Z}_\alpha$, $\hat{Z}_t$, ADF , $\hat{P}_u$ and $\hat{P}_z$ under the alternative of cointegration. To be specific we define z_t to be cointegrated if there exists a vector h on the unit sphere $(h'h = 1)$ for which $q_t = h'z_t$ is stationary with continuous spectral density $f_{qq}(\lambda)$. This ensures that the action of the cointegrating vector h reduces the integrated process z_t to a stationary time series with properties broadly in agreement with those of the innovations ξ_t in (1). The spectral density of $h'(z_t - z_{t-1})$ satisfies

$$h'f_{\xi\xi}(\lambda)h = f_{qq}(\lambda)|1 - e^{i\lambda}|^2 \quad .$$

from which we deduce that

$$h' f_{\xi\xi}(\lambda)h = f_{qq}(0)\lambda^2 + o(\lambda^2) \ , \quad \lambda \to 0 \ .$$

This implies that $h'\Omega h = 0$ so that Ω is singular.

<u>THEOREM 5.1</u>. *If* $\{z_t\}_0^\infty$ *is generated by* (1) *and is a cointegrated system with cointegrating vector* h *and* $\Omega_{22} > 0$, *if* $q_t = h'z_t$ *and* $f_{qq}(0) > 0$ *then*

(a) $\hat{Z}_\alpha = O_p(T)$

(b) $\hat{Z}_t = O_p(T^{1/2})$

(c) $ADF = O_p(T^{1/2})$

so that each of these tests is consistent.

REMARKS

(a) We see from Theorem 5.1 that $\hat{Z}_\alpha$ diverges faster as $T \to \infty$ under the alternative of cointegration than do either of the statistics $\hat{Z}_t$ and ADF. This suggests that $\hat{Z}_\alpha$ is likely to have higher power than $\hat{Z}_t$ and ADF in samples of moderate size. It also suggests that the null distribution of $\hat{Z}_\alpha$ in finite samples is likely to be more sensitive than $\hat{Z}_t$ and ADF to changes in parameters which move the null closer to the alternative (i.e. as $\omega_{11\cdot 2} \to 0$ or $\rho^2 \to 1$).

(b) As is clear from the proof of Theorem 5.1 the requirement that $f_{qq}(0) > 0$ is needed for results (b) and (c). It is not needed for result (a). Moreover, when $f_{qq}(0) = 0$ we show in the proof that $\hat{Z}_t = O_p(T)$. In this case the cointegrating vector does more than reduce z_t to the stationary process $q_t = h'z_t$. It actually annihilates all spectral power at the origin. When this happens, there should be more evidence in the data for cointegration and, correspondingly, the $\hat{Z}_t$ statistic diverges at the faster rate $O_p(T)$.

(c) Note further that the $\hat{Z}_\alpha$ statistic, which is based directly on the coefficient estimate $\hat{\alpha}$ in the residual based regression, does not involve an estimate of the standard error of regression like $\hat{Z}_t$. Its rate of convergence is $O_p(T)$ under the alternative irrespective of the value of $f_{qq}(0)$.

(d) The requirement that $\Omega_{22} > 0$ in Theorem 5.1 is not essential to the validity of results (a)-(c) provided y_t and x_t are still cointegrated. However, when it is relaxed and we allow Ω_{22} to be singular then we need to allow for cointegrated regressors x_t in (10). In such cases we have $\hat{b} = b + O_p(T^{-1/2})$ in place of $\hat{b} = b + O_p(T^{-1})$

(see Park and Phillips (1987, Section 5.2) for details) and the proofs become more complicated.

In order to develop our next theorem let us continue to assume that $\Omega_{22} > 0$. Define an orthogonal matrix $H = [H_1, h]$ and the process:

$$(17) \qquad w_t = \begin{bmatrix} H_1'\xi_t \\ h'z_t \end{bmatrix} = \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix} \begin{matrix} n \\ 1 \end{matrix} , \quad \text{say.}$$

This is zero mean, stationary with spectrum $f_{ww}(\lambda)$, say, under the alternative of cointegration.

<u>THEOREM 5.2</u>: *If* $\{z_t\}_0^\infty$ *is generated by* (1) *and is a cointegrated system with* $h > 0$, $\Omega_{22} > 0$ *and if* $f_{ww}(0) > 0$ *then as* $T \to \infty$:

(a) $\hat{P}_u = O_p(T)$

(b) $\hat{P}_z = O_p(T)$

so that each of these tests is consistent.

We observed earlier that the Z and P tests may be constructed using first differences rather than residuals. Thus, we denoted by Z_α and Z_t the statistics which utilize the first differences $\Delta\hat{u}_t$ rather than $\hat{k}_t$ in the estimators s_k^2 and $s_{T\ell}^2$. Similarly, we denoted by P_u and P_z the statistics which utilize the first differences $\Delta z_t = \xi_t$ in place of the residuals $\hat{\xi}_t$ from the VAR (14). These modified tests have very different properties under the alternative as the following result shows.

THEOREM 5.3. *If (1) is a cointegrated system with cointegrating vector* h *and* $\Omega_{22} > 0$ *then as* $T \to \infty$

$$Z_{\alpha},\ Z_{t},\ P_{u},\ P_{z} = O_{p}(1)$$

and each of these tests is inconsistent.

The use of residuals rather than first differences in the construction of these tests has a big impact on their asymptotic behavior under the alternative of cointegration. Clearly, the formulation in terms of residuals leading to $\hat{Z}_{\alpha}$, $\hat{Z}_{t}$, $\hat{P}_{u}$, $\hat{P}_{z}$ is preferable. Phillips and Durlauf (1986) reached a similar conclusion in a related context, dealing with multivariate unit root tests.

6. COMMON CONCEPTUAL PITFALLS

Since it is the hypothesis of cointegration that is of primary interest rather than the hypothesis of no cointegration it is often argued that cointegration would be the better choice of the null hypothesis. For example, in a recent survey Engle (1987) concludes that a "null hypothesis of cointegration would be far more useful in empirical research than the natural null of non cointegration." In spite of such commonly expressed views, no residual based statistical test of cointegration proceeds along these lines.

A major source of difficulty lies in the estimation of Ω under (the null of) cointegration. In order to assess whether a multiple time series is cointegrated residual based tests seek, in effect, to determine whether there exists a linear combination of the series whose variance is an order of magnitude (in T) smaller than that of the individual series. Equivalently, one can work directly with the covariance matrix Ω and seek to determine whether its smallest latent root is zero and Ω is singular. Let us assume that the supposed cointegrating linear combination h were known. In such a case, we would seek to test

$$H_0' : h'\Omega h = 0 .$$

In order to test H_0' it would seem appropriate to estimate Ω by $\hat{\Omega}$ using the first differences $\xi_t = \Delta z_t$ of the data and base some test statistic on $h'\hat{\Omega}h$. Since $h'\hat{\Omega}h \xrightarrow[p]{} 0$ under the null H_0, but not under the alternative $(h'\Omega h > 0)$, we might expect a suitably rescaled version of $h'\hat{\Omega}h$ to provide good discriminatory power. Of course, this approach relies on critical values for the statistic that is based on $h'\hat{\Omega}h$ being worked out. Likewise, if h were not known, we would seek to test

$$H_0'' : \Omega \text{ singular} .$$

If $\Omega_{22} > 0$ then the obvious approach would be to base some test statistic on the estimated conditional variance

$$\hat{\omega}_{11\cdot 2} = \hat{\omega}_{11} - \hat{\omega}_{21}'\hat{\Omega}_{22}^{-1}\hat{\omega}_{21} .$$

Since $\hat{\omega}_{11\cdot 2} \xrightarrow[p]{} 0$ under the null H_0'' similar considerations apply.

The following lemma indicates the pitfalls inherent in this approach. It will be convenient for the proof to employ the smoothed periodogram estimate of Ω :

$$(18) \qquad \hat{\Omega} = \frac{2\pi}{2\ell+1} \Sigma_{s=-\ell}^{\ell} I_{\xi\xi}(2\pi s/T)$$

where $I_{xx}(\lambda) = w_x(\lambda)w_x(\lambda)^*$ denotes the periodogram and $w_x(\lambda) = (2\pi T)^{-1/2}\Sigma_1^T x_t e^{i\lambda t}$ the finite Fourier transform of a (multiple) time series (x_t). In (18) the bandwidth parameter ℓ plays a similar role to that of the lag truncation parameter in the weighted covariance estimator (14). We shall assume that $\ell = o(T^{1/2})$. We have:

<u>LEMMA 6.1</u>

(a) $h'\hat{\Omega}h = T^{-1}(q_T - q_0)^2 + o_p(T^{-1})$;

(b) $\hat{\omega}_{11\cdot 2} = h'\hat{\Omega}h + o_p(T^{-1})$.

<u>REMARKS</u>

(a) We see from Lemma 6.1 that both $Th'\hat{\Omega}h$ and $T\hat{\omega}_{11\cdot 2}$ are $O_p(1)$. Indeed, both of these statistics are weakly convergent in a trivial way viz

$$Th'\hat{\Omega}h,\ T\hat{\omega}_{11\cdot 2} \Rightarrow (q_\infty - q_0)^2 \tag{19}$$

where q_∞ is a random variable signifying the (weak) limit of the stationary sequence (q_T) as $T \to \infty$. Tests that are based on these statistics therefore result in inconsistent tests. Note also that the limiting distribution given by (19) is dependent on that of the (stationary) sequence q_t, which in turn depends on that of the data z_t. Thus, no central limit theory is applicable in this context. And any statistical tests that are based on $h'\hat{\Omega}h$ or $\hat{\omega}_{11\cdot 2}$ under the null of cointegration would need critical values tailored to the distribution of the data. Such specificity is highly undesirable.

(b) The above results suggest that classical procedures designed to test a null of cointegration can have serious defects. Statistics that are based on $\hat{\Omega}$ or $\hat{\omega}_{11\cdot 2}$ are not to be recommended. An alternative approach that is inspired by principal components theory is not to test H_0^* directly but to examine whether any of the latent roots of $\hat{\Omega}$ are small enough to be deemed negligible. This approach proceeds under the hypothesis that $\Omega > 0$ (no cointegration) and is well established in multivariate analysis (e.g. Anderson (1984)). It has been explored in the present context by Phillips and Ouliaris (1988).

7. ADDITIONAL ISSUES

The results of this paper are all asymptotic. They are broadly consistent with the simulation findings reported in Engle and Granger (1987) for the ADF and in Phillips and Ouliaris (1988) for the Z_α , Z_t and ADF tests. However, it is certain that there are parameter sensitivities that are likely to affect the finite sample properties of these tests in important ways. This is because as we approach the alternative hypothesis of cointegration, the model undergoes a fundamental degeneracy. This seems destined to manifest itself in the finite sample behavior of the tests in differing degrees, depending on their construction.

Some guidance on this issue is given by the performance of the Z_α , Z_t and ADF tests in simple tests for the presence of a unit root in raw time series (rather than regression residuals). Simulation findings in this context have been reported by Schwert (1987) and Phillips and Perron (1988). These studies indicate the power advantages of the Z_α test that we have established by asymptotic arguments in this paper. But they also show that size distortions can be substantial for all of the tests in models with parameters approaching the stability region. It seems likely that similar conclusions will hold for residual based unit root tests. However, the issues deserve to be explored systematically in simulation experiments.

APPENDIX A

Proof of Lemma 2.1. The first part of the argument relies on the construction of a sequence $\{Y_t\}$ of stationary and ergodic martingale differences which are representative of the sequence $\{\xi_t\}$. Under (C2) this construction may be performed as in the proof of Theorem 5.5 of Hall and Heyde (1980, pp. 141-142). We then have

$$\text{(A1)} \qquad \xi_t = Y_t + Z_t - Z_{t+1}$$

where $\{Y_t = C(1)\epsilon_t\}$ is the required martingale difference sequence and Z_t is strictly stationary. Note that with this construction $E(Y_0Y_0') = \Omega$ and Z_t is square integrable. Now

$$\xi_t^* = \Sigma_{-\infty}^{\infty} a_j Y_{t-j} + \Sigma_{-\infty}^{\infty} a_j (Z_{t-j} - Z_{t-j+1})$$

and

$$X_T^*(r) = T^{-1/2}\Sigma_{t=1}^{[Tr]}\Sigma_{-\infty}^{\infty} a_j Y_{t-j} + T^{-1/2}\Sigma_{-\infty}^{\infty} a_j (Z_{1-j} - Z_{[Tr]-j+1}) .$$

As in Hall and Hyde (1980, p. 143) it follows that

$$\text{(A2)} \qquad \sup_{0\le r\le 1} |X_T^*(r) - T^{-1/2}\Sigma_1^{[Tr]} Y_t^*| \xrightarrow[p]{} 0$$

where $Y_t^* = \Sigma_{-\infty}^{\infty} a_j Y_{t-j}$.

The new sequence $\{Y_t^*\}$ is strictly stationary, ergodic and square integrable with spectral density matrix

$$f_{Y*Y*}(\lambda) = (1/2\pi)(\Sigma_{-\infty}^{\infty} a_j e^{ij\lambda})\Omega(\Sigma_{-\infty}^{\infty} a_j e^{ij\lambda})^* = |\Sigma_{-\infty}^{\infty} a_j e^{ij\lambda}|^2 \Omega .$$

It follows that

$$T^{-1}E\{(\Sigma_1^T Y_t^*)(\Sigma_1^T Y_t^*)'\} \to 2\pi f_{Y^*Y^*}(0) = a(1)^2\Omega$$

(see Ibragimov and Linnik (1971, Theorem 18.2.1)). Under (4) we may now obtain a martingale representation of Y_t^* analogous to (A1) for ξ_t viz

$$\text{(A3)} \qquad Y_t^* = Q_t + Z_t^* - Z_{t+1}^*$$

where $Q_t = a(1)Y_t$ is a stationary ergodic sequence of martingale differences with covariance matrix $a(1)^2\Omega$. We deduce from (A3) that

$$T^{-1/2}\Sigma_1^{[Tr]}Y_t^* = T^{-1/2}\Sigma_1^{[Tr]}Q_t + T^{-1/2}(Z_1^* - Z_{[Tr]+1})$$

and as before

$$\text{(A4)} \qquad \sup_{0\le r\le 1} |T^{-1/2}\Sigma_1^{[Tr]}(Y_t^* - Q_t)| \xrightarrow[p]{} 0 .$$

Similarly,

$$\text{(A5)} \qquad \sup_{0\le r\le 1} |T^{-1/2}\Sigma_1^{[Tr]}(\xi_t - Y_t)| \xrightarrow[p]{} 0 .$$

Noting that $X_T(r) = T^{-1/2}\Sigma_1^{[Tr]}\xi_t$, we now obtain from (A2), (A4) and (A5)

$$\sup_{0\le r\le 1} |X_T^*(r) - a(1)X_T(r))| \xrightarrow[p]{} 0$$

as required for (6). (7) follows directly.

Proof of Lemma 2.2. Part (a) is immediate from (8). To prove (b) note that

$$L\eta = \begin{bmatrix} \ell_{11} \\ \ell_{21} - L_{22}A_{22}^{-1}a_{21} \end{bmatrix} = \begin{bmatrix} \ell_{11} \\ -\ell_{11}(\int_0^1 w_2w_2')^{-1}\int_0^1 w_2w_1 \end{bmatrix} = \ell_{11}\kappa$$

giving the results required. Next

$$\text{(A6)} \qquad \eta'B(r) = \eta'L'W(r) = \ell_{11}\kappa'W(r) = \ell_{11}Q(r)$$

and (c)-(e) follow directly.

Proof of Theorem 4.1. We first observe that

$$\text{(A7)} \qquad \hat{b} \Rightarrow \eta \ , \quad \text{where} \quad \hat{b}' = (1, -\hat{\beta}')$$

(Phillips (1986)) and then

$$\text{(A8)} \qquad T^{-2}\Sigma_1^T\hat{u}_{t-1}^2 = \hat{b}'(T^{-2}\Sigma_1^T z_{t-1}z_{t-1}')\hat{b} \Rightarrow \eta'A\eta = a_{11\cdot 2} \ .$$

To prove (a) we write

$$\begin{aligned} \hat{Z}_\alpha &= T(\hat{\alpha}-1) - (1/2)(s_{T\ell}^2 - s_k^2)/(T^{-2}\Sigma_1^T\hat{u}_{t-1}^2) \\ &= (T^{-1}\Sigma_1^T\hat{u}_{t-1}\Delta\hat{u}_t - (1/2)(s_{T\ell}^2 - s_k^2))/(T^{-2}\Sigma_1^T\hat{u}_{t-1}^2) \ . \end{aligned}$$

Now

$$(1/2)(s_{T\ell}^2 - s_k^2) = T^{-1}\Sigma_{s=1}^{\ell} w_{s\ell}\Sigma_{t=s+1}^T \hat{k}_t\hat{k}_{t-s}$$

and $\hat{k}_t = \hat{u}_t - \hat{\alpha}\hat{u}_{t-1} = \Delta\hat{u}_t - (\hat{\alpha}-1)\hat{u}_{t-1} = \hat{b}'(\xi_t - (\hat{\alpha}-1)z_{t-1})$ so that

(A9) $$\hat{Z}_\alpha = \hat{b}'(T^{-1}\Sigma_1^T z_{t-1}\xi_t - T^{-1}\Sigma_{s=1}^{\ell} w_{s\ell}\Sigma_{t=s+1}^T[\xi_{t-s} + (1-\hat{\alpha})z_{t-s-1}]$$
$$[\xi_t + (1-\hat{\alpha})z_{t-1}]')\hat{b}/(T^{-2}\Sigma_1^T\hat{u}_{t-1}^2) \ .$$

Since ξ_t is strictly stationary with continuous spectral density matrix $f_{\xi\xi}(\lambda)$ we have

$$T^{-1}\Sigma_{s=1}^{\ell} w_{s\ell}\Sigma_{t=s+1}^T \xi_{t-s}\xi_t' \xrightarrow[p]{} \Omega_1$$

provided $\ell \to \infty$ as $T \to \infty$ with $\ell = o(T)$. Moreover, as in the proof of Theorem 2.6 of Phillips (1988), we find

$$T^{-1}\Sigma_1^T z_{t-1}\xi_t - T^{-1}\Sigma_{s=1}^{\ell} w_{s\ell}\Sigma_{t=s+1}^T \xi_{t-s}\xi_t' \Rightarrow \int_0^1 BdB' \ .$$

Finally, since $1-\hat{\alpha} = O_p(T^{-1})$ we deduce from (A8) and (A9) that

$$\hat{Z}_\alpha \Rightarrow \eta'\int_0^1 BdB'\eta/a_{11\cdot 2} = \int_0^1 RdR$$

with the final equivalence following from parts (d) and (e) of Lemma 2.2. Note that the distribution of

$$R = Q/(\int_0^1 Q^2)^{1/2}$$

depends only on n , the number of regressors in (10). The proof of part (b) follows in the same manner. To prove (c) and (d) we observe that $\hat{\xi}_t = \xi_t + O_p(T^{-1})$ from (14) and hence

$$\hat{\Omega} \xrightarrow[p]{} \Omega$$

as $T \to \infty$ provided $\ell \to \infty$ and $\ell = o(T)$. We deduce that

$$\hat{\omega}_{11\cdot 2} \xrightarrow[p]{} \omega_{11\cdot 2}$$

and using (A8) and Lemma 2.2(e) we obtain

$$\hat{P}_u \Rightarrow 1/\int_0^1 Q^2$$

as required for (c). Part (d) follows by noting that

$$\hat{P}_z \Rightarrow \mathrm{tr}(\Omega(\int_0^1 BB')^{-1}) = \mathrm{tr}(\Omega L^{-1}(\int_0^1 WW')^{-1}L'^{-1})$$

$$= \mathrm{tr}(\int_0^1 WW')^{-1}$$

as required.

Proof of Theorem 4.2. The ADF test statistic is the usual t-ratio for $\hat{\alpha}_*$ in the regression

$$\text{(A10)} \qquad \Delta\hat{u}_t = \hat{\alpha}_*\hat{u}_{t-1} + \Sigma_{i=1}^{p}\hat{\varphi}_i\Delta\hat{u}_{t-i} + \hat{v}_{tp} .$$

In conventional regression notation this statistic takes the form

$$\mathrm{ADF} = (u_{-1}'Q_{X_p}u_{-1})^{1/2}\hat{\alpha}_*/s_v$$

where X_p is the matrix of observations on the p regressors $(\Delta\hat{u}_{t-1}, \ldots, \Delta\hat{u}_{t-p})$, u_{-1} is the vector of observations of $\hat{u}_{t-1}$, $Q_{X_p} = I - X_p(X_p'X_p)^{-1}X_p'$ and $s_v^2 = T^{-1}\Sigma_1^T\hat{v}_{tp}^2$. Now

$$\text{(A11)} \qquad (u_{-1}'Q_{X_p}u_{-1})^{1/2}\hat{\alpha}_* = (T^{-2}u_{-1}'Q_{X_p}u_{-1})^{-1/2}(T^{-1}u_{-1}'Q_{X_p}\Delta\hat{u})$$

and

$$T^{-2}u_{-1}'Q_{X_p}u_{-1} = T^{-2}u_{-1}'u_{-1} + o_p(1)$$

(A12) $$\Rightarrow \eta'\int_0^1 BB'\eta = \omega_{11\cdot 2}\int_0^1 Q^2 \ .$$

Moreover under (C1) we have

$$\hat{b} \Rightarrow \eta \quad \text{(R-mixing)}$$

so that the limit variate η is asymptotically independent of each event $E \in F = \bigcup_{i=-\infty}^{\infty} F_i$ where $F_i = \sigma(\xi_j,\ j \le i)$. This independence makes it possible to condition on η without affecting the probability of events $E \in F$. Note that

(A13) $$\Delta\hat{u}_t = \hat{b}'\xi_t \Rightarrow \eta'\xi_t = \zeta_t \ , \quad \text{say.}$$

Since ξ_t is a stationary (vector) ARMA process by assumption, it is clear that the new scalar process $\zeta_t = \eta'\xi_t$, given η , is also a stationary ARMA process (see, for instance, Lutkepohl (1984)). We write its AR representation as

(A14) $$v_t = \Sigma_{j=0}^{\infty} d_j \zeta_{t-j} = d(L)\zeta_t$$

where L is the backshift operator. Note that the sequence (d_j) is majorized by geometrically declining weights and is therefore absolutely summable. Moreover, given η , v_t is an orthogonal $(0,\ \sigma^2(\eta))$ sequence with

(A15) $$\sigma^2(\eta) = d(1)^2\eta'\Omega\eta \ .$$

We now note that the ADF procedure requires the lag order p in the autoregression (A10) to be large enough to capture the correlation structure of the errors. Even if ξ_t is itself driven by a finite order vector AR model, the scalar process ζ_t will follow an ARMA model with a non zero MA component. It is therefore always necessary to let $p \to \infty$ in (A10) in order to capture the time series behavior of ζ_t. The only exception occurs when ξ_t is itself an orthogonal sequence. Formally, in the context of unit root tests, Said and Dickey (1984) require p to increase with T in such a way that $p = o(T^{1/3})$. When this happens, noting that $\hat{\alpha}_* = O_p(T^{-1})$, we see that (A10) converges to (A14), conditional on η. In particular, we have

$$
\begin{aligned}
T^{-1}u'_{-1}Q_{X_p}\Delta\hat{u} &= T^{-1}u'_{-1}v + o_p(1) \\
&= \hat{b}'T^{-1}\Sigma_1^T z_{t-1}v_t + o_p(1) \\
\text{(A16)} \qquad &= \eta' T^{-1}\Sigma_1^T z_{t-1}v_t + o_p(1)
\end{aligned}
$$

Now write

$$v_t = d(L)\zeta_t = d(L)\xi_t'\eta$$

and note that by Lemma 2.1

$$\text{(A17)} \qquad T^{-1/2}\Sigma_1^{[Tr]}v_t = d(1)(T^{-1/2}\Sigma_1^{[Tr]}\xi_t')\eta + o_p(1)$$

uniformly in r (from (6)). Also

$$\text{(A18)} \qquad T^{-1/2}\Sigma_1^{[Tr]}\xi_t \Rightarrow B(r)$$

and we obtain from (A16)-(A18)

(A19) $$\eta' T^{-1}\Sigma_1^T z_{t-1} v_t \Rightarrow d(1)\eta'\int_0^1 BdB'\eta \ .$$

We deduce from (A11), (A12), (A15) and (A19) that

$$\begin{aligned} ADF \Rightarrow \frac{d(1)\eta'\int_0^1 BdB'\eta}{(\eta'\int_0^1 BB'\eta)^{1/2}\sigma(\eta)} &= \frac{\eta'\int_0^1 BdB'\eta}{(\eta'\int_0^1 BB'\eta)^{1/2}(\eta'\Omega\eta)^{1/2}} \\ &= \frac{\int_0^1 QdQ}{(\int_0^1 Q^2)^{1/2}(\kappa'\kappa)^{1/2}} \\ &= \int_0^1 RdS \end{aligned}$$

as required.

<u>Proof of Theorem 5.1</u>. First observe that since the system is cointegrated we have:

$$\hat{b} = \begin{bmatrix} 1 \\ -\hat{\beta} \end{bmatrix} \xrightarrow[p]{} \begin{bmatrix} 1 \\ -\beta \end{bmatrix} = b \ , \quad \text{say}$$

where

$$b = ch \ , \quad c = (b'b)^{1/2}$$

so that h and b are collinear. We have

$$\hat{b} = b + O_p(T^{-1})$$

from Phillips and Durlauf (1986, Theorem 4.1) and thus

$$\hat{u}_t = \hat{b}'z_t = b'z_t + O_p(T^{-1/2}) = cq_t + O_p(T^{-1/2}) \ .$$

In the residual based regression

$$\hat{u}_t = \hat{\alpha}\hat{u}_{t-1} + \hat{k}_t$$

we now obtain

$$\hat{\alpha} \xrightarrow[p]{} \alpha = \gamma_q(1)/\gamma_q(0) \ , \quad \gamma_q(r) = E(q_t q_{t-r})$$

with $|\alpha| < 1$, and

$$\hat{k}_t = c(q_t - \alpha q_{t-1}) + O_p(T^{-1/2}) = k_t + O_p(T^{-1/2}) \ , \quad \text{say}$$

where k_t is stationary with continuous spectral density

$$f_{kk}(\lambda) = c^2|1 - \alpha e^{i\lambda}|^2 f_{qq}(\lambda) \ .$$

It follows that

$$s^2_{T\ell} \xrightarrow[p]{} 2\pi f_{kk}(0) = 2\pi c^2(1-\alpha)^2 f_{qq}(0) > 0 \tag{A20}$$

and

$$\begin{aligned} s^2_k \xrightarrow[p]{} \mathrm{var}(k_t) &= c^2((1 + \alpha^2)\gamma_q(0) - 2\alpha\gamma_q(1)) \\ &= c^2(\gamma_q(0)^2 - \gamma_q(1)^2)\gamma_q(0) \ . \end{aligned}$$

Now

$$T^{-1}\Sigma_1^T\hat{u}^2_{t-1} \xrightarrow[p]{} c^2\gamma_q(0) \ , \quad T^{-1}\Sigma_1^T\hat{u}_t\hat{u}_{t-1} \xrightarrow[p]{} c^2\gamma_q(1) \tag{A21}$$

and then

$$T^{-1}\Sigma_1^T\hat{u}_t\hat{u}_{t-1} - T^{-1}\Sigma_1^T\hat{u}_{t-1}^2 - (1/2)(s_{T\ell}^2 - s_k^2)$$

(A22) $$\xrightarrow[p]{} c^2(\gamma_q(1) - (1/2)\gamma_q(0) - \pi(1-\alpha)^2 f_{qq}(0) - (1/2)\gamma_q(1)^2/\gamma_q(0)) \ .$$

It follows that

$$\hat{Z}_\alpha = T(T^{-1}\Sigma_2^T\hat{u}_t\hat{u}_{t-1} - T^{-1}\Sigma_2^T\hat{u}_{t-1}^2 - (1/2)(s_{T\ell}^2 - s_k^2))/(T^{-1}\Sigma_2^T\hat{u}_{t-1}^2)$$
$$= O_p(T)$$

as required for part (a). Similarly, we find that

$$\hat{Z}_t = T(T^{-1}\Sigma_1^T\hat{u}_t\hat{u}_{t-1} - T^{-}\Sigma_1^T\hat{u}_{t-1}^2 - (1/2)(s_{T\ell}^2 - s_k^2))(T^{1/2}s_{T\ell}(T^{-1}\Sigma_2^T\hat{u}_{t-1}^2)^{1/2})^{-1}$$
$$= O_p(T^{1/2})$$

in view of (A20)-(A22). Note, however, that if $f_{qq}(0) = 0$ (so that q_t has an MA unit root) we have, as in Lemma 6.1,

$$Ts_{T\ell}^2 = O_p(1)$$

and in this case

$$\hat{Z}_t = O_p(T)$$

as for $\hat{Z}_\alpha$. This proves part (b). To prove (c) we observe that when $f_{qq}(0) > 0$, q_t has an AR representation,

(A23) $$\Sigma_{j=0}^{\infty} a_j q_{t-j} = e_t \ , \quad a_0 = 1$$

where $\{e_t\}$ is an orthogonal $(0, \sigma_e^2)$. We take $\{a_j\}$ to be absolutely summable and then, following Fuller (1976, p. 374) we write (A23) in alternate form as:

$$\text{(A24)} \qquad \Delta q_t = (\theta_1 - 1)q_{t-1} + \Sigma_{k=1}^{\infty}\theta_{k+1}\Delta q_{t-k} + e_t$$

where $\theta_i = \Sigma_{j=i}^{\infty}a_j$ $(i = 2, 3, \ldots)$ and $\theta_1 = -\Sigma_{j=1}^{\infty}a_j$. Since q_t is stationary we know that $\theta_1 \neq 1$. In the ADF regression (A10) (as $p \to \infty$) we find that

$$s_v^2 \xrightarrow[p]{} \sigma_e^2$$

$$\hat{\alpha}_* \xrightarrow[p]{} (\theta_1 - 1) \neq 0$$

$$T^{-1}u'_{-1}Q_{X_p}u_{-1} \xrightarrow[p]{} (1 + (\Sigma_1^{\infty}a_j)^2)\sigma_e^2$$

and hence

$$ADF = O_p(T^{1/2})$$

as required for (c).

<u>Proof of Theorem 5.2</u>. Note that both $\hat{P}_u$ and $\hat{P}_z$ rely on the covariance matrix estimate $\hat{\Omega}$ given by (13). This estimate relies on the residuals $\hat{\xi}_t$ from the VAR (14) i.e.

$$z_t = \hat{\Pi}z_{t-1} + \hat{\xi}_t \ .$$

As $T \to \infty$ we have (Park and Phillips (1988))

$$\hat{\Pi} \to H\begin{bmatrix} I_{n-1} & \vdots & 0 \\ 0 & \vdots & \end{bmatrix}H' = \bar{\Pi} \ , \quad \text{say}$$

where

$$g = E(w_t w_{2t-1})/E(w_{2t}^2) \ .$$

We may write

$$\begin{aligned}
\hat{\xi}_t &= \xi_t + (I-\bar{\Pi})z_{t-1} + O_p(T^{-1/2}) \\
&= HH'\xi_t + H\left\{\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & \vdots & g \end{bmatrix}\right\}H'z_{t-1} + O_p(T^{-1/2}) \\
&= H\left\{\begin{bmatrix} w_{1t} \\ h'\xi_t \end{bmatrix} + \begin{bmatrix} 0 \\ w_{2t-1} \end{bmatrix} - gw_{2t-1}\right\} + O_p(T^{-1/2}) \\
&= H\left\{\begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix} - gw_{2t-1}\right\} + O_p(T^{-1/2}) \\
&= H(I - [0 \ \vdots \ g])w_t + O_p(T^{-1/2}) \\
&= \bar{\xi}_t + O_p(T^{-1/2}) \quad \text{say.}
\end{aligned}$$

Now $\bar{\xi}_t$ is zero mean, stationary with spectral density matrix

$$f_{\bar{\xi}\bar{\xi}}(\lambda) = H(I - [0,g]e^{i\lambda})f_{ww}(\lambda)(I - [0,g]e^{i\lambda})^{*}H' \ .$$

Observe that

$$\bar{\Omega} = f_{\bar{\xi}\bar{\xi}}(0) = H(I - [0,g])f_{ww}(0)(I - [0,g])'H' \tag{A25}$$

is positive definite, since $f_{ww}(0) > 0$ and $1 - g_n \neq 0$ where

$$g_n = E(w_{2t}w_{2t-1})/E(w_{2t}^2) \ .$$

We now obtain

(A26) $$\hat{\Omega} \xrightarrow[p]{} \bar{\Omega} > 0$$

as $T \to \infty$. It follows that

(A27) $$\hat{\omega}_{11\cdot 2} = \hat{\omega}_{11} - \hat{\omega}_{21}'\hat{\Omega}_{22}\hat{\omega}_{21} \xrightarrow[p]{} \bar{\omega}_{11\cdot 2} = \bar{\omega}_{11} - \bar{\omega}_{21}'\bar{\Omega}_{22}\bar{\omega}_{21}$$

where we partition $\bar{\Omega}$ conformably with Ω . Hence, using (A21) and (A27) we find that

$$\hat{P}_u = O_p(T)$$

as required for part (a).

To prove part (b) we first note that by Proposition 5.4 of Park and Phillips (1987)

(A28) $$\begin{aligned} M_{22}^{-1} &= h(h'M_{zz}h)^{-1}h' + o_p(T^{-1}) \\ &= hh'/m_{qq} + o_p(T^{-1}) \end{aligned}$$

where

$$m_{qq} = T^{-1}\Sigma_1^T w_{2t}^2 = T^{-1}\Sigma_1^T q_t^2 .$$

It follows that

(A29) $$\begin{aligned} \hat{P}_z &= T\{h'\hat{\Omega}h/m_{qq} + o_p(T^{-1})\} \\ &= O_p(T) \end{aligned}$$

as required for (b).

Proof of Theorem 5.3. The Z_α and Z_t tests use

$$k_t = \hat{u}_t - \hat{u}_{t-1} = \hat{b}'\xi_t = b'\xi_t + o_p(T^{-1})$$

$$= c(q_t - q_{t-1}) + o_p(T^{-1}) \ .$$

We therefore have

$$s^2_{T\ell} \xrightarrow[p]{} 0$$

and, as in Lemma 6.1,

$$Ts^2_{T\ell} = O_p(1) \ .$$

Now

$$T^{-1}\Sigma\hat{u}_{t-1}(\hat{u}_t - \hat{u}_{t-1}) + (1/2)s^2_k$$

$$= c^2T^{-1}\Sigma q_{t-1}(q_t - q_{t-1}) + (1/2)c^2T^{-1}\Sigma_1^T(q_t - q_{t-1})^2 + o_p(T^{-1})$$

$$= o_p(T^{-1})$$

so that

$$Z_\alpha = T\{T^{-1}\Sigma_2^T\hat{u}_{t-1}(\hat{u}_t - \hat{u}_{t-1}) - (1/2)(s^2_{T\ell} - s^2_k)\}/(T^{-1}\Sigma_2^T\hat{u}^2_{t-1})$$

$$= O_p(1)$$

as $T \to \infty$. The result for Z_t follows in the same way.

In the case of P_z it is easy to see from (A27) and (A28) that

$$P_z = T\{(h'\hat{\Omega}h/m_{qq}) + o_p(T^{-1})\}$$

where $\hat{\Omega}$ is constructed using the first differences $\Delta z_t = \xi_t$. However, since the system is cointegrated the limit matrix Ω of $\hat{\Omega}$ is singular. Indeed $h'\Omega h = 0$ and, further, since $h'\xi_t = q_t - q_{t-1}$ we find

$$Th'\hat{\Omega}h = O_p(1) \quad . \tag{A30}$$

We deduce that

$$P_z = O_p(1)$$

and the test is inconsistent, as stated. In the case of P_u we observe that

$$\begin{aligned} \hat{\omega}_{11\cdot 2} &= \det \hat{\Omega}/\det \hat{\Omega}_{22} \\ &= \det(H'\hat{\Omega}H)/\det \hat{\Omega}_{22} \\ &= (h'\hat{\Omega}h - h'\hat{\Omega}H_1(H_1'\hat{\Omega}H_1)^{-1}H_1'\hat{\Omega}h)\det(H_1'\hat{\Omega}H_1)/\det \hat{\Omega}_{22} \\ &= O_p(T^{-1}) \end{aligned} \tag{A31}$$

in view of (A30) and the fact that

$$h'\hat{\Omega}H_1 = o_p(T^{-1/2})$$

(see Lemma 6.1). We deduce that

$$\hat{P}_u = O_p(1)$$

as stated.

Proof of Lemma 6.1. $\hat{\Omega}$ is a consistent estimator of Ω based on ξ_t. $h'\hat{\Omega}h$ is a consistent estimator of $h'\Omega h = 0$ based on $r_t = h'\xi_t = q_t - q_{t-1}$. Consider the following smoothed periodogram estimates

$$\hat{\Omega} = \frac{2\pi}{2\ell+1} \Sigma_{s=-\ell}^{\ell} I_{\xi\xi}(2\pi s/T) ,$$

$$h'\hat{\Omega}h = \frac{2\pi}{2\ell+1} \Sigma_{s=-\ell}^{\ell} I_{rr}(2\pi s/T) .$$

Note that for $s = -\ell, -\ell+1, \ldots, \ell$ we have

$$\begin{aligned} w_r(2\pi s/T) &= (2\pi T)^{-1/2}\Sigma_1^T(q_t - q_{t-1})e^{i2\pi st/T} \\ &= (2\pi T)^{-1/2}\Sigma_1^T q_t e^{i2\pi st/T} - (2\pi T)^{-1/2}\Sigma_1^T q_{t-1} e^{i2\pi s((t-1)+1)/T} \\ &= (2\pi T)^{-1/2}(q_T e^{i2\pi s} - q_0) + O_p\left(\frac{s}{T}\right) \\ &= (2\pi T^{-1/2})(q_T - q_0) + O_p\left(\frac{\ell}{T}\right) \end{aligned}$$

and, thus, for $\ell = o(T^{-1/2})$ we deduce that

(A32) $$w_r(2\pi s/T) = (2\pi T^{-1/2})(q_T - q_0) + o_p(T^{-1/2})$$

uniformly in s. Hence,

$$h'\hat{\Omega}h = T^{-1}(q_T - q_0)^2 + o_p(T^{-1}) .$$

It follows that

$$Th'\hat{\Omega}h = O_p(1)$$

as required for part (a). The result continues to hold for other choices of spectral estimator. To prove (b) note from (A31) that

$$\hat{\omega}_{11\cdot 2} \; (h'\hat{\Omega}h - h'\hat{\Omega}H_1(H_1'\hat{\Omega}H_1)^{-1}H_1'\hat{\Omega}h)\det(H_1'\hat{\Omega}H_1)/(\det \hat{\Omega}_{22}) \; .$$

Now $\hat{\Omega}_{22} \xrightarrow[p]{} \Omega_{22} > 0$, $H_1'\hat{\Omega}H_1 \xrightarrow[p]{} H_1'\Omega H_1 > 0$ and

$$h'\hat{\Omega}H_1 = \frac{2\pi}{2\ell+1} \Sigma_{s=-\ell}^{\ell} I_{rw_1}(2\pi s/T) \; .$$

Now $w_r(2\pi s/T) = O_p(T^{-1/2})$ uniformly in s and

$$(2\ell+1)^{-1}\Sigma_{s=-\ell}^{\ell} w_{w_1}(2\pi s/T) = O_p(1)$$

so that

$$h'\hat{\Omega}H_1 = O_p(T^{-1/2}) \; .$$

We deduce that

$$\hat{\omega}_{11\cdot 2} = h'\hat{\Omega}h + O_p(T^{-1})$$

as required for part (b).

APPENDIX B

Tables 1-4 present estimates of the critical values for the $\hat{Z}_\alpha$, $\hat{Z}_t$, $\hat{P}_u$, and $\hat{P}_z$ statistics. The tables allow for cointegrating regressions with up to five explanatory variables ($n \leq 5$). Critical values are provided for Models (10) and (16) and for cointegrating regressions with a constant term and trend.

The critical values were generated using the monte-carlo method with 10000 iterations and 500 observations. All the computations were performed on an IBM/AT using the GAUSS programming language. The random innovations were drawn from the standard normal random number generator in GAUSS (i.e "RNDNS"). Thus $\Omega = I$ and $\rho^2 = 0$ for the generated data, thereby simplifying the computation of the statistics.

Approximate 95% confidence intervals for the critical values were computed using the method described in Rohatgi(1984, pp 496-500). In order to provide some indication of the degree of precision in the estimates, we present the approximate 95% confidence intervals for $n = 1$ (refer to the rows labelled Δ_1). Confidence intervals for $n \geq 2$ are available from the authors on request.

<u>Usage</u>:

For Tables 1 and 2 ($\hat{Z}_\alpha$ and $\hat{Z}_t$):

Reject the null hypothesis of no cointegration if the computed value of the statistic is <u>smaller</u> than the appropriate critical value. For example, for a regression with a constant term and one explanatory variable (i.e. $n = 1$), we reject at the 5% level if the computed value of $\hat{Z}_\alpha$ is less than -20.4935 or the computed value of $\hat{Z}_t$ is less than -3.3654.

For Tables 3 and 4 ($\hat{P}_u$ and $\hat{P}_z$):

Reject the null hypothesis of no cointegration if the computed value of the statistic is <u>greater</u> than the appropriate critical value. For example, for a regression with two explanatory variables (i.e. $n = 2$) but no constant term, we reject at the 5% level if the computed value of $\hat{P}_u$ is greater than 32.9392 or the computed value of $\hat{P}_z$ is greater than 71.2751.

Table 1a

Critical Values for the $\hat{Z}_{\alpha}$ statistic (standard)

Size

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	-10.7444	-11.5653	-12.5438	-13.8123	-15.6377	-18.8833	-22.8291
2	-16.0164	-17.0148	-18.1785	-19.6142	-21.4833	-25.2101	-29.2688
3	-21.5353	-22.6211	-23.9225	-25.5236	-27.8526	-31.5432	-36.1619
4	-26.1698	-27.3952	-28.8540	-30.9288	-33.4784	-37.4769	-42.8724
5	-30.9022	-32.2654	-33.7984	-35.5142	-38.0934	-42.5473	-48.5240
Δ_1	(-0.2009)	(-0.1866)	(-0.2210)	(-0.2863)	(-0.5282)	(-0.5053)	(-0.8794)
	(+0.2283)	(+0.2338)	(+0.2941)	(+0.3163)	(+0.3899)	(+0.6036)	(+0.6801)

Table 1b

Critical Values for the $\hat{Z}_{\alpha}$ statistic (demeaned)

Size

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	-14.9135	-15.9292	-17.0390	-18.4836	-20.4935	-23.8084	-28.3218
2	-19.9461	-21.0371	-22.1948	-23.8739	-26.0943	-29.7354	-34.1686
3	-25.0537	-26.2262	-27.5846	-29.5083	-32.0615	-35.7116	-41.1348
4	-29.8765	-31.1512	-32.7382	-34.7110	-37.1508	-41.6431	-47.5118
5	-34.1972	-35.4801	-37.0074	-39.1100	-41.9388	-46.5344	-52.1723
Δ_1	(-0.2646)	(-0.2664)	(-0.3035)	(-0.2660)	(-0.4174)	(-0.6163)	(-0.9824)
	(+0.1834)	(+0.3011)	(+0.3329)	(+0.3348)	(+0.4319)	(+0.4834)	(+1.1440)

Table 1c

Critical Values for the $\hat{Z}_{\alpha}$ statistic (demeaned and detrended)

Size

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	-20.7931	-21.8068	-23.1915	-24.7530	-27.0866	-30.8451	-35.4185
2	-25.2884	-26.4865	-27.7803	-29.7331	-32.2231	-36.1121	-40.3427
3	-30.2547	-31.6712	-33.1637	-34.9951	-37.7304	-42.5998	-47.3590
4	-34.6336	-36.0288	-37.7368	-39.7286	-42.4593	-47.1068	-53.6142
5	-38.9959	-40.5939	-42.3231	-44.5074	-47.3830	-52.4874	-58.1615
Δ_1	(-0.2514)	(-0.3946)	(-0.3466)	(-0.3908)	(-0.5445)	(-0.6850)	(-0.9235)
	(+0.2771)	(+0.3020)	(+0.3044)	(+0.4081)	(+0.5049)	(+0.6158)	(+0.8219)

Table 2a

Critical Values for the $\hat{Z}_t$ and ADF statistics (standard)

Size

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	-2.2584	-2.3533	-2.4505	-2.5822	-2.7619	-3.0547	-3.3865
2	-2.7936	-2.8797	-2.9873	-3.1105	-3.2667	-3.5484	-3.8395
3	-3.2639	-3.3529	-3.4446	-3.5716	-3.7371	-3.9895	-4.3038
4	-3.6108	-3.7063	-3.8068	-3.9482	-4.1261	-4.3798	-4.6720
5	-3.9438	-4.0352	-4.1416	-4.2521	-4.3999	-4.6676	-4.9897
Δ_1	(-0.0232)	(-0.0247)	(-0.0269)	(-0.0328)	(-0.0439)	(-0.0382)	(-0.0600)
	(+0.0211)	(+0.0228)	(+0.0218)	(+0.0347)	(+0.0318)	(+0.0601)	(+0.0755)

Table 2b

Critical Values for the $\hat{Z}_t$ and ADF statistics (demeaned)

Size

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	-2.8639	-2.9571	-3.0657	-3.1982	-3.3654	-3.6420	-3.9618
2	-3.2646	-3.3513	-3.4494	-3.5846	-3.7675	-4.0217	-4.3078
3	-3.6464	-3.7306	-3.8329	-3.9560	-4.1121	-4.3747	-4.7325
4	-3.9593	-4.0528	-4.1565	-4.2883	-4.4542	-4.7075	-5.0728
5	-4.2355	-4.3288	-4.4309	-4.5553	-4.7101	-4.9809	-5.2812
Δ_1	(-0.0290)	(-0.0261)	(-0.0232)	(-0.0296)	(-0.0424)	(-0.0389)	(-0.0582)
	(+0.0186)	(+0.0263)	(+0.0317)	(+0.0380)	(+0.0304)	(+0.0415)	(+0.0501)

Table 2c

Critical Values for the $\hat{Z}_t$ and ADF statistics (demeaned and detrended)

Size

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	-3.3283	-3.4207	-3.5184	-3.6467	-3.8000	-4.0722	-4.3628
2	-3.6613	-3.7400	-3.8429	-3.9754	-4.1567	-4.3854	-4.6451
3	-3.9976	-4.0808	-4.1950	-4.3198	-4.4895	-4.7699	-5.0433
4	-4.2751	-4.3587	-4.4625	-4.5837	-4.7423	-5.0180	-5.3576
5	-4.5455	-4.6248	-4.7311	-4.8695	-5.0282	-5.3056	-5.5849
Δ_1	(-0.0259)	(-0.0246)	(-0.0244)	(-0.0259)	(-0.0350)	(-0.0469)	(-0.0629)
	(+0.0246)	(+0.0281)	(+0.0205)	(+0.0301)	(+0.0288)	(+0.0507)	(+0.0722)

Table 3a

Critical Values for the $\hat{P}_u$ statistic (standard)

Size

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	17.2146	18.6785	20.3933	22.7588	25.9711	31.8337	38.3413
2	22.9102	24.6299	26.7022	29.4114	32.9392	39.2236	46.4097
3	28.9811	31.0664	33.5359	36.5407	40.1220	46.3395	55.7341
4	34.5226	36.4575	39.2826	41.8969	46.2691	53.3683	63.2149
5	39.7187	41.7669	44.3725	47.6970	51.8614	59.6040	69.4939
Δ_1	(-0.4356)	(-0.3845)	(-0.3833)	(-0.5797)	(-0.6274)	(-1.3218)	(-1.4320)
	(+0.3777)	(+0.3842)	(+0.4706)	(+0.5793)	(+0.6159)	(+0.8630)	(+1.4875)

Table 3b

Critical Values for the $\hat{P}_u$ statistic (demeaned)

Size

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	24.1833	25.8456	27.8536	30.3123	33.7130	39.9288	48.0021
2	29.3836	31.4238	33.6955	36.4757	40.5252	46.6707	53.8731
3	35.1077	37.4543	39.6949	42.8111	46.7281	53.9710	63.4128
4	40.5469	42.5683	45.3308	48.6675	53.2502	61.2555	71.5214
5	45.3177	47.6684	50.3537	53.5654	57.7855	65.8230	76.7705
Δ_1	(-0.3913)	(-0.4662)	(-0.5310)	(-0.5323)	(-0.5064)	(-1.0507)	(-1.6209)
	(+0.4424)	(+0.5441)	(+0.4507)	(+0.7081)	(+0.8326)	(+1.3312)	(+2.1805)

Table 3c

Critical Values for the $\hat{P}_u$ statistic (demeaned and detrended)

Size

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	36.9055	38.8150	41.2488	44.2416	48.8439	56.0886	65.1714
2	41.2115	43.4320	46.1061	49.3671	53.8300	60.8745	69.2629
3	46.9643	49.2906	52.0015	55.4625	60.2384	68.4051	78.3470
4	51.9689	54.3205	57.3667	60.8175	65.8706	74.4712	84.5480
5	56.0522	58.6310	61.6155	65.3514	70.7416	79.0043	91.0392
Δ_1	(-0.5294)	(-0.5724)	(-0.6764)	(-0.7143)	(-1.0116)	(-1.2024)	(-2.1849)
	(+0.5171)	(+0.5187)	(+0.6762)	(+0.7989)	(+0.8773)	(+1.2936)	(+2.2679)

Table 4a

Critical Values for the $\hat{P}_z$ statistic (standard)

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	30.0137	31.7517	33.9267	36.6646	40.8217	47.2452	55.1911
2	56.7679	59.1613	62.1436	65.6162	71.2751	79.5177	89.6679
3	92.7621	95.7974	99.2664	103.8454	109.7426	119.3793	131.5716
4	135.2724	138.9636	143.0775	148.4109	155.8019	166.3516	180.4845
5	186.4277	190.6337	195.6202	201.9621	210.2910	224.0976	237.7723
Δ_1	(-0.4804)	(-0.4493)	(-0.5646)	(-0.7120)	(-0.8406)	(-1.1622)	(-1.2202)
	(+0.4633)	(+0.5042)	(+0.6770)	(+0.9202)	(+0.7319)	(+1.5961)	(+1.7356)

Table 4b

Critical Values for the $\hat{P}_z$ statistic (demeaned)

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	42.5452	44.8266	47.5877	50.7511	55.2202	61.4556	71.9273
2	74.1493	76.8850	80.2034	84.4027	89.7619	97.8734	109.4525
3	113.5617	116.6933	120.3035	125.4579	132.2207	142.5992	153.4504
4	160.8156	164.7394	168.8572	174.2575	182.0749	194.7555	209.8054
5	215.2089	219.5757	225.2303	232.4652	241.3316	255.5091	270.5018
Δ_1	(-0.4629)	(-0.7383)	(-0.6744)	(-0.6903)	(-1.0214)	(-0.7998)	(-1.8177)
	(+0.4873)	(+0.7355)	(+0.5972)	(+0.6662)	(+0.7440)	(+2.0530)	(+2.4081)

Table 4c

Critical Values for the $\hat{P}_z$ statistic (demeaned and detrended)

n	0.1500	0.1250	0.1000	0.0750	0.0500	0.0250	0.0100
1	66.2417	68.8271	71.9586	75.7349	81.3812	90.2944	102.0167
2	106.6198	109.9751	113.4929	118.3710	124.3933	133.6963	145.8644
3	154.8402	158.6619	163.1050	168.7736	175.9902	188.1265	201.0905
4	210.3150	214.3858	219.5098	225.6645	234.2865	247.3640	264.4988
5	273.3064	277.9294	284.0100	291.2705	301.0949	315.7299	335.9054
Δ_1	(-0.5433)	(-0.7346)	(-0.8305)	(-0.6905)	(-0.8651)	(-1.6500)	(-2.3915)
	(+0.6819)	(+0.5862)	(+0.7373)	(+1.1280)	(+1.4149)	(+1.8572)	(+2.1024)

REFERENCES

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis* (2nd Edition). New York: Wiley.

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

Engle, R. F. and B. S. Yoo (1987). "Forecasting and testing in cointegrated systems," *Journal of Econometrics*, 35, 143-159.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. New York: Wiley.

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

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Applications*. New York: Academic Press.

Ibragimov, I. A. and Y. V. Linnik (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

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

Lutkepohl, H. (1984). "Linear transformations of vector ARMA processes," *Journal of Econometrics*, 26, 283-294.

McLeish, D. L. (1975). "Invariance principles for dependent variables," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32, 165-178.

Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley.

Newey, W. K. and K. D. West (1987). "A simple positive definite heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703-708.

Park, J. Y. and P. C. Phillips (1988). "Statistical inference in regressions with integrated processes: Part 1," Cowles Foundation Discussion Paper No. 811R, forthcoming in *Econometric Theory*.

__________ (1987). "Statistical inference in regressions with integrated processes: Part 2," Cowles Foundation Discussion Paper No. 819R, forthcoming in *Econometric Theory*.

Phillips, P. C. B. (1986). "Understanding spurious regressions in econometrics," *Journal of Econometrics*, 33, 311-340.

__________ (1987). "Time series regression with a unit root," *Econometrica*, 55, 277-301.

__________ (1988). "Weak convergence to the matrix stochastic integral $\int_0^1 BdB'$," *Journal of Multivariate Analysis*, 24, 252-264.

__________ and S. N. Durlauf (1986). "Multiple time series regression with integrated processes," *Review of Economic Studies*, 53, 473-496.

Phillips, P. C. B. and S. Ouliaris (1988). "Testing for cointegration using principal components methods," *Journal of Economic Dynamics and Control*, 12, ...

Phillips, P. C. B. and P. Perron, "Testing for a unit root in time series regression," *Biometrika*, 75, 335-346.

Rohatgi, V. K. (1984). *Statistical Inference*. Wiley: New York.

Said, S. E. and D. A. Dickey (1984). "Testing for unit roots in autoregressive moving average models of unknown order," *Biometrika*, 71, 599-607.

Schwert, G. W. (1986). "Effects of model specification on tests for unit roots," University of Rochester Working Paper.

Stock, J. H. and M. W. Watson (1986). "Testing for common trends," Harvard Institute of Economic Research Discussion Paper No. 1222.