BANCO DE ESPAÑA

COINTEGRATION AND UNIT ROOTS: A SURVEY

Juan J. Dolado, Tim Jenkinson and Simon Sosvilla-Rivero

SERVICIO DE ESTUDIOS
Documento de Trabajo nº 9005

COINTEGRATION AND UNIT ROOTS: A SURVEY (*)

Juan J. Dolado

Tim Jenkinson

Simon Sosvilla-Rivero

(*) We are grateful to A. Banerjee, P. Burridge, A. Escribano, A. Espasa, J. Galbraith, D. Hendry and A. Maravall for their comments on previous versions of this paper. The usual caveat applies.

Banco de España. Servicio de Estudios

Documento de Trabajo n.º 9005

El Banco de España al publicar esta serie pretende facilitar la difusión de estudios de interés que contribuyan al mejor conocimiento de la economía española.

Los análisis, opiniones y conclusiones de estas investigaciones representan las ideas de los autores, con las que no necesariamente coincide el Banco de España.

ISBN: 84-7793-044-9
Depósito legal: M. 8195 - 1990

Imprenta del Banco de España

ABSTRACT

This paper provides an updated survey of a burgeoning literature on testing, estimation and model specification in the presence of integrated variables. Integrated variables are a specific class of non-stationary variables which seem to characterise faithfully the properties of many macroeconomic time series. The analysis of cointegration develops out of the existence of unit roots and offers a generic route to test the validity of the equilibrium predictions of economic theories. Special emphasis is put on the empirical researcher's point of view.

Keywords: Unit Root, Cointegration, Trends, Error Correction Mechanisms.

1. INTRODUCTION

Economic theory generally deals with equilibrium relationships. Most empirical econometric studies are an attempt to evaluate such relationships by summarising economic time series using statistical analysis. To apply standard inference procedures in a dynamic time series model we need the various variables to be stationary, since the majority of econometric theory is built upon the assumption of stationarity. Until recently, this assumption was rarely questioned, and econometric analysis proceeded as if all the economic time series were stationary, at least around some deterministic trend function which could be appropiately removed. However, stationary series should at least have constant unconditonal mean and variance over time, a condition which appears rarely to be satisfied in economics. The importance of the stationarity assumption had been recognised for many years, but the important papers by Granger and Newbold (1974), Nelson and Kang (1981) and Nelson and Plosser (1982) allerted many to the econometric implications of non-stationarity. Integrated variables are a specific class of non-stationary variables with important economic and statistical properties. These are derived from the presence of unit roots which give rise to stochastic trends, as opposed to pure deterministic trends, with innovations to an integrated process being permanent instead of transient. For example, the presence of a large permanent component in aggregate output conficts with traditional formulations of both Keynesian and Classical macroeconomic theories in terms of countercyclical policies, implying, in addition, that the welfare costs and benefits of policy actions are far different than when output movements are seen as transitory deviations from a slowly growing natural level.

The presence of, at least, a unit root is implied by many economic models by the rational use of available information by economic agents. Standard applications include futures contracts, stock prices, yield curves, real interest rates, exchange rates, money velocity, hysteresis theories of unemployment, and, perhaps the most popular, the implications of the permanent income hypothesis for real consumption.

Statisticians were well aware of the existence of integrated series and, in fact, Box and Jenkins (1970) argue that a non-stationary series can be transformed into a stationary one by successive differencing of the series. Therefore, from their point of view, the differencing operation seemed to be a pre-requisite for econometric modelling both from an univariate and a multivariate perspective. However Sargan (1964), Hendry and Mizon (1978) and Davidson et al. (1978), among others, have criticised on a number of grounds the specification of dynamic models in terms of differenced variables only, especially because it is then impossible to infer the long-run steady state solution from the estimated model.

Granger (1983) and Granger and Weiss (1983), resting upon the previous ideas, point out that a vector of variables, all of which achieve stationarity after differencing, may have linear combinations which are stationary without differencing. Engle and Granger (1987) formalise the idea of variables sharing an equilibrium relationship in terms of cointegration between time series, providing us with tests and an estimation procedure to evaluate the existence of equilibrium relationships, as implied by economic theory, within a dynamic specification framework. Standard examples include the relationship between real wages and productivity, nominal exchange rates and relative prices, consumption and disposable income, long and short-term interest rates, money velocity and interest rates, production and sales, etc.

In view of this epidemic of martingales in economics, a voluminous literature on testing, estimation, prediction, control and model specification in the presence of integrated variables has developed in the last few years[1]. The purpose of this survey is to provide a useful guide through this increasingly technical literature, paying special attention to the point of view of the applied researcher, who being a non-specialist in this particular subject wants to get a unified coverage of the main techniques available in this field.

The paper is organised as follows. The concepts of cointegration and unit roots are introduced in Section 2. In Section 3 we survey several alternative tests for the existence of unit roots, including cases where seasonality is present. Section 4 deals with alternative definitions of integration. Section 5 examines the application of some of the previous tests to determine the existence of cointegrating relationships. Finally, Section 6 contains a review of some new test procedures for cointegration. Finally, brief conclusions follow in Section 7.

2. UNIT ROOTS AND COINTEGRATION

Wold's (1938) decomposition theorem states that a stationary time series process with no deterministic component has an infinite moving average (MA) representation. This, in turn, can be represented approximately by a finite autoregressive moving average (ARMA) process (see, e.g. Hannan, 1970).

However, as was mentioned in the Introduction, some time series need to be appropiately differenced in order to achieve stationarity. From this comes the definition of integration (as adopted by Engle and Granger, 1987): A variable y_t is said to be integrated of order d [or $y_t \sim I(d)$] if it has a stationary, invertible, non-deterministic ARMA representation after differencing d times. Thus, a time series integrated of order zero is stationary in levels, while for a time series integrated of order one, the first difference is stationary. A white noise series and a stable first-order autorregressive [AR(1)] process are examples of I(0) series, while a random walk process is a example of an I(1) series.

Granger (1986) and Engle and Granger (1987) discuss the main differences between processes that are I(0) and I(1). They point out that an I(0) series: (i) has finite variance which does not depend on time, (ii) has only a limited memory of its past behaviour (i.e. the effects of a particular random innovation are only transitory), (iii) tends to fluctuate around the mean (which may include a deterministic trend), and (iv) has autocorrelations that decline rapidly as the lag increases. For the case of an I(1) series, the main features are: (i) the variance depends upon time and goes to infinity as time goes to infinity, (ii) the process has an infinitely long memory (i.e. an innovation will permenently affect the process), (iii) it wanders widely, and (iv) the autocorrelations tend to one in magnitude for all time separations.

Consider now two time series y_t and x_t which are both I(d) (i.e. they have compatible long-run properties). In general, any linear combination of y_t and x_t will be also I(d). If, however, there exists a vector $(1,-\beta)'$, such that the combination

$$z_t = y_t - \beta\ x_t \tag{1}$$

is I(d-b), b>0, then Engle and Granger (1987) define y_t and x_t as cointegrated of order (d,b) [or $(y_t,x_t)' \sim CI(d,b)$], with $(1,-\beta)'$ called the cointegrating vector.

The concept of cointegration tries to mimic the existence of a long-run equilibrium to which an economic system converges over time. If, e.g., economic theory suggests the following long-run relationship between y_t and x_t

$$y_t = \alpha + \beta\ x_t \tag{2}$$

then z_t can be interpreted as the equilibrium error (i.e., the distance that the system is away from the equilibrium, given in this case by the constant α).

Engle and Granger also show that if y_t and x_t are cointegrated CI(1,1), then there must exist an error correction model (ECM) representation of the following form[2]

$$\Delta y_t = \theta_0 + \theta_1 z_{t-1} + \Sigma\, \theta_{2i} \Delta x_{t-i} + \Sigma\, \theta_{3i} \Delta y_{t-i} + \varepsilon_t \qquad (3)$$

where Δ denotes the first-order time difference (i.e $\Delta y_t = y_t - y_{t-1}$) and where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance σ^2_ε (i.e. $\varepsilon_t \sim iid(0,\sigma^2_\varepsilon)$). Furthermore, they prove the converse result that an ECM generates cointegrated series.

Note that the term z_{t-1} in equation (3) represents the extent of the disequilibrium between levels of y and x in the previous period. The ECM states that changes in y_t depend not only on changes in x_t, but also on the extent of disequilibrium between the levels of y and x. The appeal of the ECM formulation is that it combines flexibility in dynamic specification with desirable long-run properties: it could be seen as capturing the dynamics of the system whilst incorporating the equilibrium suggested by economic theory (see Hendry and Richard, 1983)[3].

Based upon the concept of cointegration (and on its closely related concept of ECM representation), Engle and Granger suggest a 2-step estimation procedure for dynamic modelling which has become recently very popular in applied research. Let us assume that y_t and x_t are both I(1), then the procedure goes as follows:

(i) First, in order to test whether the series are cointegrated, the "cointegrating regression"

$$y_t = \alpha + \beta\, x_t + z_t \qquad (4)$$

is estimated by ordinary least squares (OLS) and it is tested whether the "cointegrating residuals" $\hat{z}_t = y_t - \hat{\alpha} - \hat{\beta}\, x_t$ are I(0). Stock (1987) has shown that if two I(1) series are cointegrated, then the OLS estimates from equation (4) provide "super-consistent" estimates of the cointegrating vector, in the sense that they converge to the true parameter at a rate proportional to the inverse sample size, T^{-1}, rather than at $T^{-1/2}$ as in the ordinary stationary case[4]. The intuition behind this remarkable result can be seen by analysing the behaviour of the OLS estimator of β in (4) (where the constant is eliminated for simplicity), when $z_t \sim iid(0,\sigma_z^2)$ and x_t follows a random walk:

$$\Delta\, x_t = \varepsilon_t \;\; ; \;\; (x_0 = 0,\; \varepsilon_t \sim iid(0,\sigma_\varepsilon^2)) \qquad (5)$$

Integrating (5) backwards we get

$$x_t = \sum_{i=1}^{t} \varepsilon_i = S_t$$

and therefore var $(x_t) = t\, \sigma_\varepsilon^2$, exploding as $T \uparrow \infty$[5]. Nevertheless, $T^{-2}\, \Sigma x_t^2$ converges to a random variable. Similarly the cross-moment $T^{-1/2}\, \Sigma\, x_t\, z_t$ will explode, in contrast to the stationary case, where it is asymptotically normally distributed. In the I(1) case $T^{-1}\, \Sigma\, x_t\, z_t$ converges also to a random variable. Both random variables are functionals of Brownian Motions or Wiener processes, which will be denoted henceforth, in general, as f(W) (see Phillips (1987), Phillips and Perron (1988) and Park and Phillips (1988) for a general discussion on convergence of the above mentioned distributional limits). Since the OLS estimator of β is given by

$$\hat{\beta} - \beta = \Sigma\, x_t\, z_t \,/\, \Sigma\, x_t^2$$

it follows from the previous discussion that $T(\hat{\beta} - \beta)$ is asymptotically the ratio of two non-degenerate random variables, and it is in general not normal. Thus, standard inference cannot be applied to β, even if it is "super-consistent", a question to which we will come back in Section 6.

ii) Finally, the residuals $\hat{z}_t$ are entered into the ECM. Now, all the variables in equation (3) are I(0) and conventional modelling strategies can be applied.

3. TESTING THE ORDER OF INTEGRATION OF THE RELEVANT VARIABLES

Once the relevant set of variables suggested by economic theory has been identified, the first stage in testing for cointegration between those variables is to determine the order of integration of the individual time series.

Several statistical for unit roots have been developed to test for stationarity in time series. Since many macroeconomic series have been found to be integrated of order one (see, e.g. Nelson and Plosser, 1982), we will only consider tests for a single unit root[6].

The search sequence is as follows: first test for the existence of a unit root in the levels of the series. If the unit root is not rejected by the data, the time difference of the series would then be tested for the presence of a second unit root and so on.

3.1. Tests of Unit Roots

(i) Dickey and Fuller (1979, 1981) present a class of test statistics, known as Dickey-Fuller (DF) statistics, generally used to test that a pure AR(1) process (with or without drift) has a unit root.

Let the time series y_t satisfy the following data generating process (DGP)

$$y_t = \beta_0 + \beta_1 t + \rho y_{t-1} + \varepsilon_t \qquad (6)$$

where $\varepsilon_t \sim iid(0,\sigma_\varepsilon^2)$, t is a time trend and the initial condition, y_0, is assumed to be a known constant (zero, without loss of generality). Equation (6) can also be written as

$$y_t = \beta_0 \sum_{j=1}^{t} \rho^{t-j} + \beta_1 \sum_{j=1}^{t} j \rho^{t-j} + \sum_{j=1}^{t} \varepsilon_j \rho^{t-j} \qquad (7)$$

while in the case that $\rho=1$

$$y_t = \beta_0 t + \beta_1 t(t+1)/2 + S_t \qquad (8)$$

where $S_t = \sum_{j=1}^{t} \varepsilon_j$

Dickey and Fuller (1979) consider the problem of testing the null hypothesis $H_0: \rho = 1$ versus $H_1: \rho < 1$, i.e. non-stationarity vs. stationarity around a deterministic trend, suggesting OLS estimation of a reparameterised version of (6), i.e.

$$\Delta y_t = \beta_0 + \beta_1 t + \gamma y_{t-1} + \varepsilon_t \qquad (9)$$

where $H_0: \rho=1$ is equivalent to $H_0: \gamma=0$ (since $\gamma = \rho-1$). The test is implemented though the usual t-statistic of $\hat{\gamma}$, denoted here as τ_τ. In addition, Dickey and Fuller (1981) suggest and F-statistic for the joint null hypothesis $\beta_0 = \beta_1 = \gamma = 0$ and $\beta_1 = \gamma = 0$, denoted as Φ_2 and Φ_3 respectively. Note that under the null hypothesis τ_τ, Φ_2 and Φ_3 will not have the standard t and F distributions, instead they are functions of Brownian motion; we must use the asymptotic distributions tabulated in Fuller (1973, p. 373) and in Dickey and Fuller (1981, p. 1.063) respectively. If $\beta_1 = 0$ ($\beta_0 = 0$) in (9), the t and F statistics, corresponding to H_0: $\gamma = 0$ and H_1: $\beta_0 = \gamma = 0$, are denoted $\tau_\mu(\tau)$ and Φ_1 respectively and the corresponding critical values are also given in the previous references. In all cases the critical values given there crucially depend upon the sample size. It should also be noted that the critical values depend upon the "nuisance" parameters contained in the model and in the DGP. To discuss this more formally, consider the sample variance of y_t when it is generated by (8) (i.e. $\rho=1$)

$$T^{-1} \Sigma y_t^2 = T^{-1}\Sigma[(\beta_0+\beta_1/2)^2 t^2 + (\beta_1/2)^2 t^4 + S_t^2 +(\beta_0+\beta_1/2)\beta_1 t^3$$

$$+ 2 (\beta_0+\beta_1/2) t S_t + \beta_1 t^2 S_t] \tag{10}$$

From the distibutional results in Park and Phillips (1988), it is known that $T^{-2}\Sigma S_t^2$, $T^{-5/2} \Sigma t S_t$ and $T^{-7/2} \Sigma t^2 S_t$ tend to f(W), hence, by taking probability limits in (10), we get

$$T^{-1} \Sigma y_t^2 => \beta_1^2/20 \; O(T^4) + (\beta_0+\beta_1/2)\beta_1/4 \; O(T^3) + \beta \; f(W) \; O(T^{5/2}) \tag{11}$$

$$+ (\beta_0+\beta_1/2)^2/3 \; O (T^2) + 2(\beta_0+\beta_1/2) \; f(W) \; O(T^{3/2}) + f(W) \; O(T)$$

whereby it is seen that

$$T^{-5} \Sigma y_t^2 => \beta_1^2/20 \qquad \text{if } \beta_1 \neq 0$$

$$T^{-3} \Sigma y_t^2 => \beta_0^2/3 \qquad \text{if } \beta_0 \neq 0, \; \beta_1 = 0$$

$$T^{-2} \Sigma y_t^2 => f(W) \qquad \text{if } \beta_0 = \beta_1 = 0$$

That is, if the unit root process contains a linear trend or a drift, its variability will be dominated by a quadratic or a linear trend which, appropiately normalised, converges to constants. It is only when $\beta_0=\beta_1=0$ that it converges to a non-standard distribution. This means that for example, if $\beta_1\neq 0$ in (8) and the model is estimated as in (9), the quadratic term in (10) will dominate the integrated process and normality of the τ_τ will follow. Similarly if $\beta_0\neq 0$ in (8) (with $\beta_1=0$) and only a constant appears in (9), the linear trend will dominate and normality of τ_μ will also follow[7]. It is only when $\beta_0=\beta_1=0$, both in the DGP and in the model, that the non-standard distribution will dominate. This implies that in order to use the DF critical values if a linear trend is included in the maintained hypothesis (9), the relevant null hypothesis should be a random walk with drift ($\beta_0\neq 0$), whilst if only a drift is included in (9), the relevant null hypothesis is a random walk without drift[8].

From the previous discussion we consider that the following testing strategy is most appropiate. First, start by the most unrestricted model (9), ($\beta_0\neq 0,\beta_1\neq 0$) if it is suspected that the differenced series has a drift. Then use τ_τ to test for the null hypothesis. If it is rejected there is no need to go further. If it is not rejected, test for the significance of the trend under the null. If it is significant, then test again for a unit root using the standarised normal. If the trend is not significant in the mantained model, estimate (9) without trend ($\beta_1=0$). Test again for the unit root using τ_μ. If the null hypothesis is rejected, again there is no need to go further. If it is not rejected, test for the significance of the trend under the null hypothesis and so on.

(ii) In the analysis of the DF tests, we have assumed that the DGP is a pure AR(1) process. If instead, the DGP is AR(p)

$$y_t = \beta_0 + \beta_1 t + \sum_{i=1}^{p} \rho_i\, y_{t-i} + \varepsilon_t \qquad (12)$$

let

$$\lambda^p - \sum_{i=1}^{p} \rho_i \lambda^{p-i} = 0 \qquad (13)$$

be the characteristic equation of the time series, where λ_i (i=1..p) are the eigenvalues of the process. Dickey and Fuller (1979, 1981) consider the problem of testing the null hypothesis H_0: $\lambda_1=1$ and $|\lambda_2|<1$ for i=2,.. p, suggesting OLS estimation of the reparameterised regression model

$$\Delta y_t = \beta_0 + \beta_1 t + \gamma_1 y_{t-1} + \sum_{i=1}^{p-1} \gamma_{2i} \Delta y_{t-i} + \varepsilon_t \qquad (14)$$

where p is large enough to ensure that the residual series ε_t is white noise. The tests are based on the t-ratio on $\hat{\gamma}_1$ and are known as "Augmented Dickey Fuller" (ADF) statistics. The critical values are the same as those discussed for the DF statistics, since the $\hat{\gamma}_{2i}$ (i=1... p-1) estimates converge to their true values at a rate $O(T^{-1/2})$, being asymptotically dominated by the distribution of $\hat{\gamma}_1$ which, as we mentioned in (4), is $O(T^{-1})$. The same testing strategy discussed above, applies in this case[9].

The sample distribution of the ADF statistics critically depend on the assumption that the time series y_t is generated by a pure AR process. However, since there is evidence that many macroeconomic series contain moving average (MA) components (see Schwert, 1987), we would want to consider also the possibility of an MA component in the DGP, so that the null hypothesis would be that the data are generated by a mixed autoregressive integrated moving average (ARIMA) process.

Said and Dickey (1984) extend the ADF test by exploiting the fact that an ARIMA (p,1,q) process can be adequately approximated by a high-order autoregressive process, AR(ℓ), where $\ell=O(T^{1/3})$ as $T\uparrow\infty$. In practice the test proceeds as before with p in (12) and (14) equal to ℓ. This approach permits to test the null hypothesis of the presence of a unit root without knowing the orders of p and q. However, it involves the estimation of additional nuisance parameters which reduces the effective number of observations due to the need for extra initial conditions.

When p and q are known, Said and Dickey (1985) present a test for the hypothesis that the process is ARIMA (p,1,q), i.e.

$$H_0: \emptyset(L)\ \Delta\ y_t = \Theta(L)\ \varepsilon_t$$

where $\emptyset(.)$ and $\Theta(.)$ are pth and qth order polynomials in the lag operator L, versus the alternative hypothesis that it is ARIMA (p,0,q)

$$H_1: \emptyset(L)\ (1-\rho L)y_t = \Theta(L)\ \varepsilon_t$$

To perform the test of $\rho=1$, we specify initial estimates of the parameters that are consistent under the null and alternative hypothesis. We next perform a one-step of the Gauss-Newton numerical estimation procedure (see, e.g., Harvey, 1981 p. 17). The t-statistic associated with ρ, after applying the iteration has the limiting distribution of τ, tabulated by Fuller (1976, p. 373). Similarly if the series mean y is substracted from each observation of y_t prior to analysis, the t-statistic has the limiting distribution of τ_μ.

(iii) An alternative approach, based upon the DF procedure has been presented by Phillips (1987) and Phillips and Perron (1988), While the ADF statistics are based upon the assumption that the disturbance term ε_t is identically and independently distributed, they suggest amending these statistics to allow for weak dependence and heterogeneity in ε_t. Under such general conditions, a wide class of DGP's for ε_t, such as most finite order ARIMA (p,o,q) models, can be allowed. The procedure consists of computing the DF statistics and then use some non-parametric adjustment of τ_μ and τ_τ in order to eliminate the dependence of their limiting distributions on additional nuisance parameters stemming from the ARIMA process followed by the error terms. Their adjusted counterparts are denoted $Z(\tau_\mu)$ and $Z(\tau_\tau)$, respectively.

For the regression model (9), with $\beta_1=0$, Phillips and Perron (PP) define

$$Z(\tau_\mu)=(\hat{s}/\hat{s}_{Tm})\tau_\mu-0.5(\hat{s}^2_{Tm}-\hat{s}^2)\ T\ \{s^2_{Tm}\sum_{2}^{T}(y_t-\bar{y}_{-1})^2\}^{-1/2} \tag{15}$$

where T is the sample size and m is the number of estimated autocorrelations; $\bar{y}_{-1}=(T-1)^{-1}\sum_{2}^{T}y_{t-1}$, $\hat{s}^2$ and τ_μ are, respectively, the sample variance of the residuals and the t-statistic associated with $\hat{Y}$ from the regression (9) (with $\beta_1=0$); and s^2_{Tm} is the long-run variance estimated as[10]

$$\hat{s}^2_{Tm}=T^{-1}\sum_{t=1}^{T}\varepsilon_t^2+2\ T^{-1}\sum_{s=1}^{\ell}w_{sm}\sum_{t=s+1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-s} \tag{16}$$

where $\hat{\varepsilon}$ are the residuals from the regression (9) and where the triangular kernel

$$w_{sm}=[1-s(m+1)],\ s=1\ldots m \tag{17}$$

is used to ensure that the estimate of the variance $\hat{s}^2_{Tm}$ is positive (see Newey and West, 1987)

When $\beta_1\neq 0$ in (9), the corresponding statistic is

$$Z(\tau_\tau) = (\tilde{s}/\tilde{s}_{Tm})\tau_\tau - (\tilde{s}^2_{Tm} - \tilde{s}^2)T^3\{4\ \tilde{s}_{Tm}[3\ D_{xx}]^{1/2}\}^{-1} \tag{18}$$

where $\tilde{s}$ and $\tilde{s}_{Tm}$ are defined as above, but with the residual e obtained from the estimation of (9) with $\beta_1 \neq 0$. D_{xx} is the determinant of the regressor cross-product matrix, given by

$$D_{xx} = [T^2(T^2-1)/12]\ \Sigma\ y^2_{t-1} - T(\Sigma\ t\ y_{t-1})^2$$

$$+ T(T+1)\ \Sigma\ t\ y_{t-1}\ \Sigma\ y_{t-1} - [T(T+1)(2T+1)/6](\Sigma\ y_{t-1})^2$$

The Phillips and Perron statistics have the same limiting distributions as the corresponding DF and ADF statistics, provided that $m\uparrow\infty$ as $T\uparrow\infty$, such that $m/T^{1/4}\uparrow 0$.

(iv) Simulation evidence in Molinas (1986) and Schwert (1986, 1989), shows that the tests proposed by Dickey and Fuller and by Phillips and Perron are affected by the process generating the data in large finite samples. In particular, when the underlying process is ARIMA (0,1,1) with a MA parameter close to one, the ADF and PP statistics have critical values that are far below the Dickey-Fuller distributions (i.e. these tests will lead to the conclusion that economic data are stationary too frequently). The intuition behind this result is that if the DGP of y_t is

$$\Delta\ y_t = (1-\Theta L)\varepsilon_t \tag{19}$$

if Θ is close to one, (1-L) will tend to cancel on both sides of (19), giving the impression that y_t behaves like a white noise. However, the Said and Dickey (1984) high-order autoregressive t test for the unit root, with a suitable choice of ℓ, has size close to its nominal level for all values of the MA parameter. Schwert suggests searching for the correct specification of the ARIMA process before testing for the presence of a unit root in the AR polynomial and provides the relevant critical values for the Said and Dickey (1984), Phillips (1987) and Phillips and Perron (1988) tests based on Monte Carlo experiments.

(v) Hall (1989) proposes a new approach to testing for a unit root in a time series with a moving average component based on an instrumental variable (IV) estimator

Let y_t be generated by the DGP

$$y_t = \beta_0 + \beta_1 t + \rho y_{t-1} + \Theta(L) \varepsilon_t \tag{20}$$

when $\Theta(L)$ is again a q-th order lag polynomial. Then, the IV estimator for model (20) is defined as follows

$$[\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\rho}_1]^{IV'} = (\sum_k Z_{1t} X'_{1t})^{-1} (\sum_k^T Z_{1t} y_t) \tag{21}$$

where $Z_{1t}=(1,t,y_{t-k})$ and $X_{1t}=(1,t,y_{t-1})$, $k=(q+1)$ (see Dolado (1989) for the choice of optimal IV in this framework). For the model (20) when $\beta_1=0$, the IV estimator is given by

$$[\hat{\beta}_0, \hat{\rho}_2]^{IV'} = (\sum_k^T Z_{2t} X'_{2t})^{-1} (\sum_k^T Z_{2t} y_t) \tag{22}$$

where now $Z_{2t} = (1, y_{t-k})$, $X_{1t} = (1, y_{t-1})$

Let $\tilde{t}(\tilde{\rho}_1^{IV})$ and $\hat{t}(\hat{\rho}_2^{IV})$ be the t-statistics associated with the null hypothesis $\rho=1$ in (19) (with and without trend), then Hall proves that

$$\tilde{\tau}_{IV} = s_\varepsilon \tilde{\tau}(\tilde{\rho}_1^{IV}) / \tilde{s} \Rightarrow \tau_\tau \tag{23}$$

and

$$\hat{\tau}_{IV} = s_{\varepsilon}\ \hat{t}(\hat{\rho}_2^{\ IV})\ /\ \hat{s} \Rightarrow \tau_{\mu} \qquad (24)$$

where s_{ε}^2, $\tilde{s}^2$ and $\hat{s}^2$ are consistent estimators of the variances of ε and the long-run variance $e(=\Theta(L)\varepsilon)$ obtained as in (16) and (17).

(vi) As it might have been noticed, one important limitation of all of the previous testing procedures is that they are not independent of the nuisance parameters contained in the deterministic component of the time-series process. The disappearence of this limitation has produced an alternative strand in the literature on testing. In this respect, Bhargava (1986) has developed most powerful invariant (MPI) tests for the null hypothesis corresponding to DGP (9) (with and without trend). These tests are valid in small samples and are independent of the nuisance parameters, but only valid for the AR(1) case. They are based upon transformations of Von Neumann type ratios, as for example the Durbin-Watson appproach emphasised by Sargan and Bhargava (1983) in a different context, as discussed below. The statistics proposed to test H_0: $\rho=1$, when $\beta_1=0$ and $\beta_1\neq 0$, are given by

$$R_1 = \sum_2^T (\Delta\ y_t)^2\ /\ \sum_2^T (y_t-\bar{y})^2 \qquad (25)$$

and

$$R_2 = [\sum_2^T (\Delta\ y_t)^2-(T-1)^{-1} \sum_1^T \Delta\ y_i]/D \qquad (26)$$

where $D=(T-1)^{-2} \sum_1^T [(T-1)y_t - (t-1)y_T - (T-t)y_1 - (T-1)(\bar{y}-0.5(y_1+y_T))]^2$

The corresponding critical values are given by Bhargava (1986, p. 378). The test is found to have slightly greater power than the tests proposed by Dickey and Fuller, when the data are generated by an AR(1) process.

(vii) Another limitation of all the previous testing procedures is that the distributions of the corresponding statistics are non-standard and hence a different set of critical values has to be used in each case. This problem has originated a new strand of research (see Phillips and Ouliaris, 1988), which exploits the fact that differencing a stationary series induces a unit root in the moving average representation. This fact provides a diagnostic for testing whether the series is I(0) or I(1), by using the long-run variance of the first difference of the time series y_t. To clarify the interpretation of the test, let us assume that y_t is generated by

$$\Delta y_t = \Theta(L)\varepsilon_t \;;\; \Theta(L) = (1-\Theta_1 L)\Theta'(L)$$

Then the long-run variance of Δy_t is $\sigma^2 = \sigma_\varepsilon^2\,\Theta(1)^2$. If $\Theta_1 \neq 1$ and $\Theta'(1) \neq 0$, then σ^2 is finite, whilst if $\Theta_1 = 1$, σ^2 is zero. In other words, if the time series y_t is I(0), Δy_t will have $\sigma^2 = 0$, whereas if it is I(1), $\sigma^2 \neq 0$. Therefore the null hypothesis is H_0: $\sigma^2 \neq 0$ or $H_0: \tau^2 = \sigma^2/\sigma_\varepsilon^2 \neq 0$, getting rid of the units of measurement. Obtaining an estimate of σ^2 as in (16), Phillips and Ouliaris prove that

$$m^{1/2}(\hat{\tau}^2 - \tau^2)/\tau^2 \sim N(0,1) \tag{27}$$

Since only the alternative hypothesis is a simple hypothesis, i.e. H_1: $\tau^2 = 0$, Phillips and Ouliaris propose a bounds procedure based upon the corresponding confidence interval in (27), yielding

$$\hat{\tau}^2/[1+(z_\alpha/m^{1/2})] \leq \tau^2 \leq \hat{\tau}^2/[1-(z_\alpha/m^{1/2})] \tag{28}$$

where z_α is the (1–α) percentage point of the standard normal distribution. According to the bounds test, H_0 is rejected if the upper limit of τ^2 in (28) is sufficiently small. Similarly H_0 is non rejected if the lower bound is sufficiently large. Phillips and Ouliaris recommend using 0.10 as the rejection point for the upper and lower bound. Simulation results show, however, that the suggested value can be very conservative in some instances. For example if the DGP is ARIMA (1,1,1) with parameter values in the interval (–0.6, 0.6), the average upper bound is 0.45 wheress the value of the lower bound is close to 0.10.

A very nice implication of this type of tests is that, given their asymptotic normality, they can be applied to deal with very general trend–cycle models (e.g piecewise linear functions of time, any type of impulse or step dummy). All that is needed is to perform the previous test on the differenced residuals of the regression of y_t on the general trend function.

3.2. Integration and Seasonality

Due to the fact that many economic time series contain important seasonal components, there have been several developments in the concept of seasonal integration.

Osborn et al. (1988) amend the Engle and Granger (1987) definition of integration to account for seasonality: a variable y_t, is said to be integrated of order (d,D) [or $y_t \sim I(d,D)$], if it has a stationary, invertible, non–deterministic ARMA representation after one–period differencing d times and seasonally differencing D times.

Following Pierce (1976), let us assume that seasonality has both deterministic and stochastic components, then a seasonal observed series y_t can be seen as the sum of a purely stochastic process x_t and a purely deterministic seasonal component μ_t

$$y_t = x_t + \mu_t \tag{29}$$

where

$$\mu_t = \beta_0 + \beta_1 t + \sum_{j=1}^{q-1} \beta_{2j} S_{jt} \tag{30}$$

where S_{jt} are zero/one seasonal dummies, and q=12 for monthly data, q=4 for quarterly data and so on.

By regressing y_t on μ_t, we can remove the deterministic seasonality, using the residuals from that regression as if they were the true x_t. Then, the following tests can be applied for testing I(d,D) integration, where we present the case for q=4 (i.e. we are dealing with quaterly data).

(i) Dickey et al. (1984) present a test for the presence of a single unit root at a seasonal lag. The null hypothesis is H_0: I(0,1) and the alternative is H_1: I(0,0). The test is a 3-step procedure as follows

1. The regression equation

$$\Delta_4 x_t = \theta_0 + \sum_{i=1}^{p} \theta_i \Delta_4 x_{t-i} + \varepsilon_t$$

is estimated by OLS, where $\Delta_4 x_t = x_t - x_{t-4}$

2. Using the estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_p$ define

$$z_t = \hat{\theta}(L) x_t = (1-\hat{\theta}_1 L-\ldots -\hat{\theta}_p L^p) x_t$$

3. Run the regression

$$\Delta_4 x_t = \xi_0 z_{t-1} + \sum_{i=1}^{p} \xi_i \Delta_4 x_{t-i} + \varepsilon_t$$

and compute the t–ratio on $\hat{\xi}_0$. This sample statistic, denoted $\tau_{\mu 4}$ is compared to the tabulated critical values given in Table 7 of Dickey, Hasza and Fuller (1984, p. 362).

(ii) Dickey et al. (1986, Appendix B) show that the limiting distribution of the unit root statistics is not affected by removal of seasonal means from autoregressive series. Therefore, We can use the ADF statistic from the regression equation

$$\Delta x_t = \gamma_1 \Delta x_{t-1} + \sum_{i=1}^{p} \gamma_{2i} \Delta x_{t-i} + \varepsilon_t \qquad (31)$$

to test the null hypothesis H_0: $x_t \sim I(1,0)$ versus the alternative H_1: $x_t \sim I(0,0)$. The relevant critical values are given by Fuller (1976, p. 373) for τ_μ.

(iii) Engle et al. (1987) present the following 3–step procedure to test for seasonal unit roots in the possible presence of a zero frequency unit root.

1. Compute $\hat{\theta}(L)$ as for the Dickey et al. (1984) statistic

2. Compute

$$z_{1t} = \hat{\theta}(L) \; (1+L+L^2+L^3) \; x_t$$

$$z_{2t} = -\hat{\theta}(L) \; (1-L+L^2-L^3) \; x_t$$

$$z_{3t} = -\hat{\theta}(L)\ (1-L^2)\ x_t$$

3. Run the regression

$$\Delta_4\ x_t = \pi_1\ z_{1t-1} + \pi_2\ z_{2t-1} + \pi_3\ z_{3t-2} + \sum_{i=1}^{p} \pi_{4i}\ \Delta_4\ x_{t-i} + \varepsilon_t$$

and compute the values of the t-ratios on $\hat{\pi}_1$, $\hat{\pi}_2$ and $\hat{\pi}_3$. The critical values are given in Table 2.1 of Engle et al. (1987, p. 14). If $x_t \sim I(0,0)$, then all theree of these statistics should be significant. If the test statistic for $\pi_1=0$ not significant, then $x_t \sim I(1,0)$. If either of the test statistics for $\pi_2=0$ or $\pi_3=0$ is not significant, then $x_t \sim I(0,1)$

(iv) Osborn et al. (1988) present an alternative 3-step test procedure:

1. Run the regression

$$\Delta\Delta_4\ x_t = \Psi_0 + \sum_{i=1}^{p} \Psi_i\ \Delta\Delta_4\ x_{t-i} + \varepsilon_t$$

and compute $\hat{\Psi}_1, \hat{\Psi}_2, \ldots, \hat{\Psi}_p$

2. Compute $z_{4t} = \hat{\Psi}(L)\ \Delta_4\ x_t$ and $z_{5t} = \hat{\Psi}(L)\ \Delta\ x_t$, where $\hat{\Psi}(L)= = (1 - \hat{\Psi}_1\ L - \ldots - \hat{\Psi}_p\ L^p)$

3. Run the regression

$$\Delta\Delta_4\ x_t = \Phi_1\ z_{4t-1} + \Phi_2\ z_{5t-4} + \sum_{i=1}^{p} \Phi_{3i}\ \Delta\Delta_4\ x_{t-i} + \varepsilon_t$$

and compute the F-statistics for the null hypothesis H_0: $\Phi_1 = \Phi_2 = 0$, and the t-ratio on $\hat{\Phi}_1$ and $\hat{\Phi}_2$. The null hypothesis for both type of statistics is H_0: $x_t \sim I(1,1)$ with alternative hypothesis H_1: $x_t \sim I(0,0)$ or H_2: $x_t \sim I(0,1)$. The critical values of these statistics are given in Table A.1 of Osborn et al. (1988, p. 376).

4. OTHER FORMS OF INTEGRATION

In this section, we review alternative forms of integration based upon the possibility that the model parameters are allowed to vary (periodic integration) or the possibility of using non-integer differencing orders to achieve stationarity in the data (fractional integration). Both ideas have received recent attention in the literature

4.1. Periodic Integration

Osborn et al. (1988), building upon the framework developed by Tiao and Grupe (1980), investigate the use of a periodic model (whose parameters are allowed to vary according to the time at which observations are made) as an alternative to the conventional approaches to modelling for seasonal data

The non-deterministic periodic AR(1) process is given by the following expression

$$y_t = \sum_{j=1}^{q} \omega_j S_{jt} y_{t-1} + \varepsilon_t \tag{32}$$

or

$$y_t = \omega_j y_{t-1} + \varepsilon_t \tag{33}$$

when t falls in season j. As in equation (30), S_{jt} are seasonal dummy variables corresponding to season j(j=1...q). Equation (33) states that y_t is seasonal, seasonality arising not from any direct dependence of y_t on y_{t-q}, but from the annual variation in the autoregressive coefficients ω_j. This dependence can arise, for example, if the allocation of expenditure over the year reflects seasonal tastes and hence seasonality in the underlying utility function (see Osborn, 1988).

Osborn et al. (1988) define periodic integration as follows: A variable y_t is periodically integrated of order one [or $y_t \sim PI(1)$] if y_t is non-stationary and $\delta_j\, y_t$ is stationary, where the generalised difference operator δ_j is defined as

$$\delta_j\, y_t = y_t - \omega_j\, y_{t-1} \tag{34}$$

the product $\omega_1\, \omega_2 \ldots \omega_q$ being equal to one.

Osborn et al. (1988) propose two ways of testing for periodic integration:

(i) After regressing y_t on μ_t (as defined in (30)) to remove conventional deterministic seasonality, a non-deterministic periodic AR(1) process (as defined in (33)) is fitted to the residuals x_t. This case is referred to as the removed deterministic seasonality case.

(ii) The case of included deterministic seasonality is given by fitting the following periodic AR(1) process to the original observations y_t

$$y_t = \Omega_j + \omega_j\, y_{t-i} + \varepsilon_t \quad (j=1, \ldots, q)$$

To allow for the possibility of a periodic disturbance variance, they suggest a 2-step estimation procedure for both cases. In the first step, the appropiate equation is estimated by OLS applied to observations on each of the q seasonal realisations. (i.e. four for quaterly data); then the equation is transformed by dividing each variable by the appropiate seasonal residual standard deviation estimated in this first stage regression. Using the transformed data, in the second step the periodic AR(1) model is estimated in its two versions (i.e. removed and included deterministic seasonality), with imposition of the restriction $\omega_1 \omega_2 \ldots \omega_q = 1$.

Finally, the tests (i) to (iii) in Section 3.2 are applied to the residuals of the periodic AR(1) model.

4.2. Fractional Integration

As was seen in Section 2, one of the main characteristics of the existence of unit roots in the Wold representation of a time series is that they have "long memory" (i.e. shocks have a permanent effect on the level of the series). In general it is known that the coefficient on ε_{t-j} in the MA representation of any I(d) process has a leading term j^{d-1} (for example, the coefficient in a random walk is unity, since d=1). This implies that the variance of the original series is $O(t^{2d-1})$. So, all that is needed to have "long-memory", in the sense that the variance explodes as $t\uparrow\infty$, is a degree of differencing $|d|>0.5$. Thus, it is clear that a wide range of dynamic behaviour is ruled out a priori if d is restricted to integer values.

Granger and Joyeux (1980) and more recently Diebold and Rudebusch (1989) have proposed a new family of "long-memory" processes, denoted by ARFIMA (autoregressive fractionally integrated moving-average processes), of which the ARIMA processes are particular cases: A variable y_t is fractionally integrated of order d [or $y_t \sim FI(d)$] if y_t is non-stationary and Δ^d is stationary, where the operation Δ^d, using a binomial expansion, is as follows

$$(1-L)^d = 1 - dL + \frac{d(d-1)}{2!} L^2 - \frac{d(d-1)(d-2)}{3!} L^3 + \ldots \qquad (35)$$

where d belongs to the rational set of numbers and d>0.5.

Note that these processes can always be constrained to belong to the open interval (0.5,1.0) by substracting the integer part of the differencing order. So if the degree of diferencing is, for example, 1.7, we can always redefine the degree of differencing as d-1 (0.7 in this case).

Diebold and Rudebusch (198) propose the following method of testing and estimation for fractional integration:

(i) First difference the relevant series denoted $\tilde{y}_t=(1-L)y_t$. As d of the level series equals $1+\tilde{d}$, a value of $\tilde{d}$ equal to zero corresponds to a unit root in y_t. Thus, we wish to estimate $\tilde{d}$ in the model

$$(1-L)^{\tilde{d}} \; y_t = \Theta(L)\varepsilon_t \qquad (36)$$

(ii) Estimate by OLS the following regression

$$\ln[I(\lambda_j)] = \beta_0 - \beta_1 \ln\{4 \sin^2(\lambda_j/2)\} + \eta_j, \; j=1,\ldots, T^{1/2} \qquad (37)$$

where $\lambda_j=2\pi_j/T$ (j=0...T-1) denote the harmonic ordinates of the sample and $I(\lambda_j)$ denote the periodogram at ordinate j (see Harvey, 1981, p. 66). Geweke and Porter-Hudak (1983) prove that $\hat{\beta}_1$ is a consistent and asymptotically normal estimate of d. Furthermore, the variance of the estimate of β_1 is given by the usual OLS estimator, which can be used to test the null hypothesis H_0: d=0[i.e. $y_t \sim I(1)$]. Moreover they show that the variance of the disturbance η_j is know to be equal to $\pi^2/6$, which can be imposed to increase efficiency.

(iii) Given an estimate of $\tilde{d}$ we transform the series y_t by the "long-memory" filter (35), truncated at each point to the available sample. The transformed series is then modelled as in (36) (or in the ARMA representation) following the traditional Box and Jenkins (1970) procedure.

5. TESTING FOR STATIONARITY IN THE COINTEGRATING RESIDUALS

In the two previous sections we have discussed procedures to test for the order of integration of individual time series. This is, as we mentioned in Section 2, a first stage in the estimation and testing of cointegrating relationship. The reason is a matter of "integration or growth accounting" in the words of Brusch and Pagan (1989) (i.e. the left and right hand sides of an equation, such as (4) must be of the same order of integration, otherwise, the residual will not be stationary). If for example, the dependent variable is I(1), the independent variables need to be I(1) and not cointegrate among themselves to an I(0) variable or, perhaps, be I(2) and cointegrate among themselves to an I(1) variable.

In order to illustrate testing for cointegration, we will consider a bivariate case where say, y_t and x_t have been found to contain a single unit root at the regular frequency (i.e. both are I(1)). Then, the following part of the cointegration test is to estimate the cointegrating regression (4) and test whether the "cointegrating residuals" ($\hat{z}_t = y_t - \hat{\alpha} - \hat{\beta}\, x_t$) are I(0).

Engle and Granger (1987) suggest seven alternative tests for determining if $\hat{z}_t$ is stationary. Here we will consider only two of their suggested tests, namely the Durbin-Watson statistic for the cointegration equation (CRDW) and the ADF statistic for the cointegrating residuals (CRADF).

The DW statistic for equation (4) will approach zero if the cointegrating residuals contain an autoregressive unit root, and thus the test rejects the null hypothesis of non-cointegration if the CRDW is significantly greater than zero. The intuition underlying this test can be understood by means of a simple example. Suppose that z_t is assumed to follow an AR(1) process with coefficient ρ. Then the null hypothesis of non-cointegration is H_0: $\rho=1$. Since it can be shown that the DW statistic is such that DW $\simeq 2(1-\rho)$ (see, e.g. Harvey, 1981, p. 20), the previous null hypothesis can be translated into H_0: DW=0 versus the alternative H_1: DW>0. Engle and Granger (1987, p. 269) present the critical values of this test for 100 observations.

The CRADF statistic is based upon the OLS estimation of

$$\Delta \hat{z}_t = \gamma_1 \hat{z}_{t-1} + \sum_{i=1}^{p} \gamma_{2i} \Delta \hat{z}_{t-i} + \varepsilon_t \tag{38}$$

where again p is selected on the basis of being sufficiently large to ensure that ε_t is a close approximation to white noise. The t-ratio statistic on $\hat{\gamma}_1$ is the CRADF statistic. We cannot use the critical values tabulated by Fuller (1976) to test for a unit root in the cointegrating residuals. Intuitively, since OLS estimation of the cointegrating regression equation I(0) chooses $\hat{\alpha}$ and $\hat{\beta}$ to minimise the residual variance, we might expect to reject the null hypothesis H_0: $z_t \sim I(1)$ rather more often than suggested by the nominal test size, so that the critical values have to be raised in order to correct the test bias. Engle and Granger (1987, p. 269) present the critical values for the CRADF statistic generated from Monte Carlo simulations of 100 observations.

Note that the critical values for both CRDW and CRADF statistics are for the bivariate case (i.e., for one dependent and one independent variable in the cointegrating regression), and for 100 observations. Engle and Yoo (1987) produce expanded critical values for CRDW and CRADF statistics for 50, 100 and 200 observations, and for systems of up to five variables.

6. SOME NEW DEVELOPMENTS IN COINTEGRATION

In this section we survey some new test procedures for cointegration that have recently been proposed in the literature. Most of these procedures extend the testing and estimation approach introduced in Section 2 to a multivariate context where there may exist more than a single cointegrating relationships among a set of n variables. For example, among nominal wages, prices employment and productivity, there may exist two relationships, one determining employment and another determining wages (see, inter alia, Hall, 1986, and Jenkinson, 1986).

In general, if X_t represents a vector of n I(1) variables whose Wold representation is

$$\Delta X_t = C(L)\ \varepsilon_t \tag{39}$$

where now $\varepsilon_t \sim nid(0,\Sigma)$, being Σ the covariance matrix of ε_t and C(L) in an invertible matrix of polynomial lags. If there exists a cointegrating vector α, then, permultiplying (39) by α', we obtain

$$\alpha'\ \Delta X_t = \alpha'[C(1) + C^*(L)\ (1-L)]\varepsilon_t \tag{40}$$

where C(L) has been expanded around L=1 and $C^*(L)$ can be shown to be invertible (see Engle and Granger, 1987). If the linear combination $\alpha' X_t$ is stationary, then $\alpha' C(1)=0$ and then (1–L) would cancel out on both sides of (40). If (39) is represented in AR form, we have that

$$A(L)\ C(L) = (1-L)I \tag{41}$$

where I is an identity matrix, and hence

$$A(1)\ C(1) = 0 \tag{42}$$

This implies that A(1) can be written as $A(1)=\Upsilon\ \alpha'$. If there were r cointegrating vector (r<n-1), then $A(1)=B\Gamma'$, where B and Γ are (nxr) matrices which collect the r different Υ and α vectors. Testing the rank of A(1) or C(1) constitutes the basis of the following procedures.

(i) Johansen (1988) and Johansen and Juselius (1988) develop a maximum likelihood estimation procedure that has several advantages on the 2-step regression procedure suggested by Engle and Granger. It relaxes the assumption that the cointegrating vector is unique and it takes into account the error structure of the underlying process.

Johansen considers the p-th order autoregressive representation of X_t

$$X_t = \Pi_1\ X_{t-1} + \Pi_2\ X_{t-2} + \ldots + \Pi_p\ X_{t-p} + \varepsilon_t \tag{43}$$

which, following a similar procedure to the ADF test, can be reparameterised as

$$\Delta\ X_t = \tilde{\Pi}'_1\ \Delta\ X_{t-1} + \ldots + \tilde{\Pi}'_{p-1}\ \Delta\ X_{t-p+1} + \tilde{\Pi}'_p\ X_{t-p} + \varepsilon_t$$

where $\tilde{\Pi}'_p = -\ \Pi(1)\quad (= -\ (\Pi_1 + \ldots + \Pi_p))$. To estimate $\tilde{\Pi}'_p$ maximum-likelihood, we estimate by OLS the following regressions

$$\Delta X_t = \Gamma_{01} \Delta X_{t-1} + \ldots + \Gamma_{0k-1} \Delta X_{t-k+1} + e_{0t}$$

and

$$X_{t-p} = \Gamma_{11} \Delta X_{t-1} + \ldots + \Gamma_{1k-1} \Delta X_{t-k+1} + e_{1t}$$

and compute the product moment matrices of the residuals

$$\hat{S}_{ij} = T^{-1} \sum_{t=1}^{T} \hat{e}_{it} \hat{e}'_{jt} \; ; \; i, j = 0, 1$$

The likelihood ratio test statistic of the null hypothesis H_0: $\Pi'_p = B\Gamma'$, i.e. there are at most r cointegrating vectors is

$$-2 \ln(Q) = - T \sum_{i=r+1}^{p} (1-\hat{\lambda}_i) \tag{45}$$

where $\hat{\lambda}_{r+1} \ldots \hat{\lambda}_p$ are the p–r smallest eigenvalues of $\hat{S}_{10} \hat{S}_{00} \hat{S}_{01}$ with respect to S_{11}, obtained from the determinant

$$|\hat{\lambda} \hat{S}_{11} - \hat{S}_{10} \hat{S}_{00} \hat{S}_{01}| = 0$$

Under the hypothesis that there are at most r cointegrating vectors, Johansen (1988) shows that the likelihood ratio test (45) is asymptotically distributed as a functional f(W). Johansen (1988, p. 239) provides a table with various quantiles of the distribution of the likelihood ratio test for r=1,2,...5. He also shows that these quantiles can be obtained by approximating the distribution by c $\chi^2(f)$ where c=0.85–0.85/f, and $\chi^2(f)$ is a central chi–square distribution with $f=2(p-r)^2$ degrees of freedom.

(ii) Stock and Watson (1988) focus on testing for the rank of C(1) in (40) and denote their approach as a "common trends" approach, by noticing that if there exists r cointegrating vectors in (40), then there exist a representation such that[11]

$$x_t = C(1)\ \Phi\ \tau_t + C^*(L)\ \varepsilon_t$$

where Φ is an nx(n–r) matrix and τ_t is an n–r vector random walk. In other words, x_t can be written as the sum of n–r common trends and an I(0) component. Estimating (39) as a multivariate ARMA (1,q) model, the null hypothesis that there are r cointegrating vectors is equivalent to the null hypothesis that there are n–r "common trends". This implies that, under the null hypothesis, the first (n–r) eigenvalues of the autoregressive matrix should be unity and the remaining eigenvalues should be smaller than one. The test is based on $T(\hat{\lambda}_{n-r+1}-1)$ and the critical values can be found in Stock and Watson (1988, p. 1104).

Phillips and Ouliaris (1988) have also proposed a multivariate extension of their unit root test, as discussed in Section 3, based upon the eigenvalues of the long–run variance of the differenced multivariate series.

(iii) As discussed in Section 2, when concentrating on a single equation estimator in the case of a single cointegrating C(1,1) relationship, the OLS estimator of the slope in the static regression (4) is "super–consistent" but its distribution is, in general, non–normal and in finite samples is biased (see Banerjee et al., 1986 and Gonzalo, 1989).

This bias and non–normality stem from the I(1) character of the regressor and its possible correlation with the I(0) disturbance z_t. Phillips (1988) has shown that in the case where x_t and z_t are independent at all leads and lags, that distribution is a "mixture of normals" and, hence, the distribution of the t–statistic on $\hat{\beta}$ is asymptotically normal. Phillips and Hansen (1988) have developed an estimation procedure, equivalent to FIML, which corrects for the bias and yields asymptotic normality in the case where such correlation exists. The procedure, denoted as a "fully modified estimator" (FME), is based upon a "non–parametric" correction by which the error term z_t is conditioned on the process followed by Δx_t and, hence, orthogonality between regressors and disturbance is achieved by construction. The FME estimators of α and β in (4) are given by

$$(\hat{\alpha},\hat{\beta})^{+'} = \left(\sum_{t=1}^{T} X_t X_t'\right)^{-1}\left[\sum_{t=1}^{T} X_t(y_t - \hat{\sigma}_{\surd}^2 \Delta x_t) - e_2 T(\Delta_{21} - \hat{\sigma}_{\surd}^2 \hat{\sigma}_{\Delta x}^2)\right]$$

where $X_t = (1, x_t)$, $e_2 = (0,1)'$ and

$$\hat{\Delta}_{21} = T^{-1} \sum_{k=0}^{\ell} \sum_{t=k+1}^{T} \Delta x_{t-k} \hat{z}_t$$

$$\hat{\sigma}_{\surd}^2 = \hat{\sigma}_{\Delta x, z} / \hat{\sigma}_{\Delta x}^2$$

the long–run variances obtained from the first–stage residuals $\hat{z}_t$, as in Engle and Granger (1987). Notice that when $\sigma_{\surd}^2 = \Delta_{21} = 0$ the FME estimators coincide with OLS for the static regression (4).

It is interesting to notice that the FME procedure coincides with the Hendry–Sargan approach, as summarised in (3), through the ECM representation of dynamic single equation models, except when z_t or Δx_t contain a moving–average disturbance in their respective representations. Even in that case it is possible to modify slightly equations like (3) by including leads of Δx_t in the regression model (see Saikkonen, 1989).

7. BRIEF CONCLUSION

The considerable gap between the economic theorist, who has much to say about equilibrium but relatively little to say about dynamics, and the econometrician, whose models concentrate on dynamic adjustment processes, has, to some extent, been bridged by the concept of cointegration. In addition to allowing the data to determine the dynamics of the model, cointegration suggests that models can be significantly improved by introducing, and allowing the data to parameterise, equilibrium conditions suggested by economic theory. Furthermore, the generic existence of such long-run relationship can, and should, be tested, using the battery of tests for unit roots discussed in this paper.

NOTES

1. There is an early survey by two of us (see Dolado and Jenkinson, 1987) and a more recent one by one of us (see Sosvilla-Rivero, 1989), and some excellent overviews by Granger (1986), Hendry (1986), Gilbert (1986), Stock and Watson (1987), Diebold and Nerlove (1988), Pagan and Wickens (1989) and Haldrup and Hylleberg (1989).

2. Even though cointegration implies at least one causal direction, it does not imply any explicit causal relationship. Here we have assumed that the causal relation suggested by the theory (i.e. x_t causes y_t) is the correct one. See Granger (1988) for a study of cointegration and causality.

3. Nickell (1985) shows that the ECM is also consistent with optimising behaviour on the part of economic agents.

4. Alternatively we would say that a "super"consistent" estimator is such that $\hat{\beta}-\beta$ has probabilistic order of magnitude $O(T^{-1})$.

5. The explosivity of the variance characterises the "integration in variance". Integration can also be applied to other higher moments (see Escribano (1987) and Hansen (1988)).

6. See the Appendix for a description for $k(k \geq 2)$ unit roots.

7. This result has been noticed by West (1988) and it is applicable also to regression models like (2) where x_t has a unit root with drift. However, Hylleberg and Mizon (1989) have noted in simulation studies that the drift has to be quite large for the deterministic trend to dominate the integrated component. If there are two I(1) regressors with drift in the model, a trend should also be included to avoid asymptotic perfect collinearity.

8. Ouliaris et al. (1988) compute critical values when in the mantained hypothesis there is up to a quintic trend. Similarly, Perron (1987) computes critical values when there is a piecewise linear trend under the mantained hypothesis.

9. Sims et al. (1986) and Banerjee and Dolado (1988), have show that the estimates of coefficients on I(0) variables in regression models with I(1) variables are $O(T^{1/2})$ and asymptotically normally distributed.

10. In the frequency domain notation, the long-run variance is equal to $2\pi f_{\varepsilon}(0)$, where $f_{\varepsilon}(0)$ in the spectrum of ε_t evalutated at frequency zero.

11. The size of C(1) in a univariate context, has been called the "size the unit root", giving rise to a literature (see Cochrane (1988) and references therein) which deals with the relative importance of the trend and cyclical componentes in the decomposition of a time-series.

APPENDIX: TESTING FOR K UNIT ROOTS

Dickey and Pantula (1987) suggest a sequence of tests for unit roots, starting with the largest number of roots under consideration (k) and decreasing by one each time the null hypothesis is rejected, stopping the procedure when the null hypothesis is accepted.

They illustrate their sequential procedure for the case k=3. It is as follows:

1. Run the regression

$$\Delta^3 y_t = \xi_0 + \xi_1 \Delta^2 y_{t-1} + \varepsilon_t$$

(where Δ^3 denotes third difference), and compute the "pseudo t-statistic" $t^*_{3,n}$ (3) (i.e. the t-statistic on $\hat{\xi}_1$). Reject the hypothesis H_3 of three unit roots and go to step 2 if $t^*_{3,n}$ (3) $\leqslant \tau_\mu$, where τ_μ is given by Fuller ((1976), p. 373).

2. Run the regression

$$\Delta^3 y_t = \xi'_0 + \xi'_1 \Delta^2 y_{t-1} + \xi'_2 \Delta y_{t-1} + \varepsilon_t$$

and compute $t^*_{3,n}$ (3) and $t^*_{2,n}$ (3). Reject the hypothesis H_2 of exactly two unit roots and go to step 3 if in addition to $t^*_{3,n}$ (3) $\leqslant \tau_\mu$ it is also found that $t^*_{2,n}$ (3) $\leqslant \tau_\mu$

3. Run the regression

$$\Delta^3 y_t = \xi^*_0 + \xi^*_1 \Delta^2 y_{t-1} + \xi^*_2 \Delta y_{t-1} + \xi^*_3 y_{t-1} + \varepsilon_t$$

and compute $t^*_{3,n}(3)$, $t^*_{2,n}(3)$ and $t^*_{1,n}(3)$. Reject the hypothesis H_1 of exactly one unit root in favour of the hypothesis H_0 if $t^*_{i,n}$ (3) $\leqslant \tau_\mu$ (i=1,2,3).

REFERENCES

Banerjee, A. and Dolado, J. (1988) Tests of the life cycle permanent income hypothesis in the presence of random walks: Asymptotic theory and small-sample intepretations. Oxford Economic Papers 40, 610–633.

Banerjee, A., Dolado, J., Hendry, D. and Smith G. (1986) Exploring equilibrium relationships in econometrics trough static models: Some Monte-Carlo evidence. Oxford Bulletin of Economics and Statistics 48, 253–277.

Bhargava, A. (1986) On the theory of testing for unit roots in observed time series. Review of Economic Studies 53, 369–384.

Box, G. and Jenkins, G. (1970) Time Series Analysis, Forecasting and Control San Francisco: Holden-Day.

Cochrane, J. (1988) How big is the random walk in GNP? Journal of Political Economy 96, 893–920.

Davidson, J., Hendry, D., Srba, F. and Yeo S. (1978) Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. Economic Journal 88, 349–363.

Dickey, D. and Fuller, W. (1979) Distribution of the estimators for autoregressive time-series with a unit root. Journal of the American Statistical association 74, 427–431.

– (1981) Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49, 1057–1072.

Dickey, D. and Pantula, S. (1987) Determining the order of differencing in autoregressive processes. Journal of Business and Economic Statistics 15, 455–461.

Dickey, D., Bell, W. and Miller, R. (1986) Unit roots in time series models: Tests and implications. The American Statistician 40, 12–26.

Dickey, D., Harza, D. and Fuller, W. (1984) Testing for unit roots in seasonal time series. Journal of the American Statistical Association 79, 355–367.

Diebold, F. and Nerlove, M. (1988) Unit roots in economic time series: A selective survey. Discussion Paper No., 49. Federal Reserve Board. Washington D.C.

Diebold, F. and Rudebusch, G. (1989) Long memory and persistence in aggregate output. Journal of Monetary Economics 24, 189–209.

Dolado, J. (1989) Optimal instrumental variable estimator of the AR parameter of an ARMA (1,1), mimeo, (forthcoming in Econometric Theory).

Dolado, J. and T. Jenkinson (1987) Cointegration: A survey of recent developments. Applied Economics Discussion Paper No., 39 University of Oxford (forthcoming in Economic Perpectives).

Engle, R. and Granger, G. (1987) Cointegration and error correction: Representation, estimation and testing. Econometrica 55, 251–276.

Engle, R., Granger, C., Hylleberg, S. and Yoo, S. (1987) Seasonal integration and cointegration. Discussion Paper No., 88–32. University of S. Diego.

Engle, R. and Yoo, S. (1987) Forecasting and testing in cointegrated systems <u>Journal of Econometrics</u> 35, 143–159.

Escribano, A. (1987) Cointegration, time co-trends and error correction systems: An alternative approach. CORE Discussion Paper No., 8715. University of Louvain.

Fuller, W. (1976) <u>Introduction to Statistical Time Series</u>. New York: John Wiley and Sons.

Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. <u>Journal of Time Series Analysis</u> 4, 221–238.

Gilbert, C. (1986) Professor Hendry's econometric methodology. <u>Oxford Bulletin of Economics and Statistics</u> 48, 283–307.

Gonzalo, J. (1989) Comparison of five alternative methods of estimating long-run equilibrium relationships. Discussion Paper 89–55. University of California, San Diego.

Granger, C. (1981). Some properties of time series data and their use in econometric model specification. <u>Journal of Econometrics</u> 28, 121–130.

– (1986) Developments in the study of cointegrated economic variables <u>Oxford Bulletin of Economics and Statistics</u> 48, 213–228.

– (1988) Some recent developments in a concept of causality. <u>Journal of Econometrics</u> 39, 199–211.

Granger, C. and Newbold, P. (1974) Spurious regressions in econometrics. <u>Journal of Econometrics</u> 26, 1045–1066.

Granger, C. and Joyeux, R. (1980) An introduction to long memory time series models and fractional integration. Journal of Time Series Analysis, 1, 15–39.

Granger, C. and Weiss, A. (1983) Time series analysis of error correction models, in Karlin, S. Amemiya, T. and Goodman, L (eds), Studies in Economic Time Series and Multivariate Statistics, New York: Academic Press.

Haldrup, N. and Hylleberg, S. (1989) Unit roots and deterministic trends, with yet another comment on the existence and interpretation of a unit root in U.S. GNP, mimeo. University of Aarhus.

Hall, A. (1989). Testing for a unit root in the presence of moving average errors. Biometrika 76, 49–56.

Hall, S. (1986) An applicaton of the Engle and Granger two-step estimation procedure to U.K. aggregate wage data. Oxford Bulletin of Economics and Statistics 48, 229–240.

Hannan, E. (1970) Multiple Time Series. New York: Wiley.

Hansen, B. (1988) A model of hetoskedastic cointegration, mimeo. Yale University.

Hansen, B. and Phillips, P. (1988) Estimation and inference in models of cointegration: A simulation study. Cowles Foundation Discussion Paper No. 881. Yale University.

Harvey, A. (1981) Time Series Models Oxford: Phillip Allan.

Hendry, D. (1986) Econometric modelling with cointegrated variables: An overview Oxford Bulletin of Economics and Statistics 48, 201–212.

Hendry, D. and Mizon, G. (1978) Serial correlation as a convenient simplification not a nuisance: A comment on a study of the demand for money by the Bank of England. Economic Journal 88, 349-363.

Hendry, D. and Richard, J.F. (1983) The econometric analysis of economic time series. International Statistical Review 51, 111-163.

Hylleberg, S. and Mizon, G. (1989) A note on the ditribution of the least squares estimation of a random walk with drift. Economics Letters 29, 225-230.

Jenkinson, T. (1986) Testing neo-classical theories of labour demand: An application of cointegration techniques. Oxford Bulletin of Economics Statistics 48, 241-251.

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

Johansen, S. and Juselius, K. (1988) Hypothesis testing for cointegration vectors with an application to the demand for money in Denmark and Finland. Working Paper No., 88-05. University of Copenhaguen.

Molinas, C. (1986) A note on spurious regressions with integrated moving average errors. Oxford Bulletin of Economics and Statistics 48.

Nelson, C. and Kang, H. (1981) Spurious periodicity in inappropiately detrended time series. Econometrica 49, 741-751.

Nelson, C. and Plosser, C. (1982) Trends an random walks in macroeconomic time series. Journal of Monetary Economics 10, 139-162.

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

Nickell, S. (1985) Error correction, partial adjustment and all that: An expository note. Oxford Bulletin of Economics and Statistics 47, 119–129.

Osborn, D. (1988) Seasonality and habit persistence in a life cycle model of consumption. Journal of Applied Econometrics 3, 255–266.

Osborn, D., Chui, A., Smith, J. and Birchenhall, C. (1988) Seasonality and the order of integration for consumption. Oxford Bulletin of Economics and Statistics 50, 361–377.

Ouliaris, S., Park, J. and Phillips, P. (1988) Testing for a unit root in the presence of a mantained trend (forthcoming in B. Raj (ed) Advances in Econometrics and Modelling. Needham: Kluwer Academic Press).

Pagan, A. and Wickens, M. (1989) Econometrics: A survey, mimeo, (forthcoming in Economic Journal).

Park, J. and Phillips, P. (1988) Statistical inference in regression with integrated processes: Part 1. Econometric Theory 4, 468–497.

Perron, P. (1987) The great crash, the oil price shock and the unit root hypothesis. Cahier 8749. University of Montreal (forthcoming in Econometrica).

Phillips, P (1986) Undestanding spurious regression in econometrics. Journal of Econometrics 33, 311–340.

– (1987) Time series regression with a unit root. Econometrica 55, 277–301.

– (1988) Optimal inference in cointegrated systems. Cowles Foundation Discussion Paper No., 866. Yale University.

Phillips, P. and Ouliaris, S. (1988) Testing for cointegration using principal components methods. Journal of Economic Dynamics and Control 12, 205–230.

Phillips, P. and Perron, P. (1988) Testing for a unit root in time series. Biometrika 75, 335–346.

Pierce, D. (1976) Seasonality adjustment when both deterministic and stochastic seasonality are present, in A. Zellner (ed) Seasonal Anlaysis of Economic Time Series. Washington: Bureau of the Census.

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

– (1985) Hypothesis testing in ARIMA (p,q,q) models. Journal of the American Statistical Association 80, 369–374.

Saikkonen, P. (1989) Asymptotically efficient estimation of cointegration regressions. Research Report No. 72, University of Helsinki.

Sargan, J. (1964) Wages and prices in the UK: A study in econometric methodology, in Hart, P., Mills, G. and Whittaker, J. (eds.) Econometric Analysis for National Planning, London: Butterworths.

Schwert, G. (1987) Effects of model specification on tests for unit roots in macroeconomic data. Journal of Monetary Economics 20, 73–103.

– (1989) Tests for unit roots. A Montecarlo investigation. Journal of Business and Economic Statistics 7, 147–159.

Sims, C., Stock, J. and Watson, M. (1986) Inference in linear time-series models with some unit roots, mimeo (forthcoming in Econometrica).

Sosvilla-Rivero, S. (1989) Cointegration and unit roots: A survey Discussion Paper 89-5, University of Birmingham.

Stock, J. (1987) Asymptotic properties of least squares estimators of cointegrating vectors. Econometrica 55, 381–386.

Stock, J. and Watson, M. (1987) Variable trends in economic time series. Journal of Economic Perspectives 2, 147–174.

– (1988) Testing for common trends. Journal of the American Statistical Association. 83, 1094–1107.

Tiao, G. and Grupe, M. (1980) Hidden periodic autoregressive-moving average models in time-series data. Biometrika 67, 365–373.

West, K. (1988) Asymptotic normality when regressors have a unit root. Econometrica 56, 1397–1417.

Wold, H. (1938) A study in the Analysis of Stationary Time-Series. Stockholm: Almguist and Wiknell.

DOCUMENTOS DE TRABAJO (1):

8501 **Agustín Maravall**: Predicción con modelos de series temporales.
8502 **Agustín Maravall**: On structural time series models and the characterization of components.
8503 **Ignacio Mauleón**: Predicción multivariante de los tipos interbancarios.
8504 **José Viñals**: El déficit público y sus efectos macroeconómicos: algunas reconsideraciones.
8505 **José Luis Malo de Molina y Eloísa Ortega**: Estructuras de ponderación y de precios relativos entre los deflactores de la Contabilidad Nacional.
8506 **José Viñals**: Gasto público, estructura impositiva y actividad macroeconómica en una economía abierta.
8507 **Ignacio Mauleón**: Una función de exportaciones para la economía española.
8508 **J. J. Dolado, J. L. Malo de Molina y A. Zabalza**: El desempleo en el sector industrial español: algunos factores explicativos. (Publicada una edición en inglés con el mismo número).
8509 **Ignacio Mauleón**: Stability testing in regression models.
8510 **Ascensión Molina y Ricardo Sanz**: Un indicador mensual del consumo de energía eléctrica para usos industriales, 1976-1984.
8511 **J. J. Dolado y J. L. Malo de Molina**: An expectational model of labour demand in Spanish industry.
8512 **J. Albarracín y A. Yago**: Agregación de la Encuesta Industrial en los 15 sectores de la Contabilidad Nacional de 1970.
8513 **Juan J. Dolado, José Luis Malo de Molina y Eloísa Ortega**: Respuestas en el deflactor del valor añadido en la industria ante variaciones en los costes laborales unitarios.
8514 **Ricardo Sanz**: Trimestralización del PIB por ramas de actividad, 1964-1984.
8515 **Ignacio Mauleón**: La inversión en bienes de equipo: determinantes y estabilidad.
8516 **A. Espasa y R. Galián**: Parquedad en la parametrización y omisiones de factores: el modelo de las líneas aéreas y las hipótesis del census X-11. (Publicada una edición en inglés con el mismo número).
8517 **Ignacio Mauleón**: A stability test for simultaneous equation models.
8518 **José Viñals**: ¿Aumenta la apertura financiera exterior las fluctuaciones del tipo de cambio? (Publicada una edición en inglés con el mismo número).
8519 **José Viñals**: Deuda exterior y objetivos de balanza de pagos en España: Un análisis de largo plazo.
8520 **José Marín Arcas**: Algunos índices de progresividad de la imposición estatal sobre la renta en España y otros países de la OCDE.
8601 **Agustín Maravall**: Revisions in ARIMA signal extraction.
8602 **Agustín Maravall y David A. Pierce**: A prototypical seasonal adjustment model.
8603 **Agustín Maravall**: On minimum mean squared error estimation of the noise in unobserved component models.
8604 **Ignacio Mauleón**: Testing the rational expectations model.
8605 **Ricardo Sanz**: Efectos de variaciones en los precios energéticos sobre los precios sectoriales y de la demanda final de nuestra economía.
8606 **F. Martín Bourgón**: Indices anuales de valor unitario de las exportaciones: 1972-1980.
8607 **José Viñals**: La política fiscal y la restricción exterior. (Publicada una edición en inglés con el mismo número).
8608 **José Viñals y John Cuddington**: Fiscal policy and the current account: what do capital controls do?
8609 **Gonzalo Gil**: Política agrícola de la Comunidad Económica Europea y montantes compensatorios monetarios.
8610 **José Viñals**: ¿Hacia una menor flexibilidad de los tipos de cambio en el sistema monetario internacional?
8701 **Agustín Maravall**: The use of ARIMA models in unobserved components estimation: an application to spanish monetary control.

(1) Los Documentos de Trabajo anteriores a 1985 figuran en el catálogo de publicaciones del Banco de España.

Información: Banco de España
Sección de Publicaciones. Negociado de Distribución y Gestión
Teléfono: 446 90 55, ext. 2180
Alcalá, 50. 28014 Madrid