BANCO DE ESPAÑA

ASYMPTOTIC DISTRIBUTION THEORY FOR ECONOMIC ESTIMATION WITH INTEGRATED PROCESSES: A GUIDE

Juan J. Dolado

SERVICIO DE ESTUDIOS
Documento de Trabajo nº 9103

ASYMPTOTIC DISTRIBUTION THEORY FOR ECONOMETRIC ESTIMATION WITH INTEGRATED PROCESSES: A GUIDE (*)

Juan J. Dolado

(*) This paper contains material used in a set of lectures on Unit Root Econometrics given at various places during the last two years. It has been written for discussion and teaching purposes. I am very grateful to Anindya Banerjee and John Galbraith for their various comments and corrections.

Banco de España. Servicio de Estudios
Documento de Trabajo n.º 9103

In publishing this series, the Banco de España seeks to diffuse worthwhile studies that help acquaint readers better with the Spanish economy.

The analyses, opinions and findings of these studies represent the views of their authors; they do not necessarily coincide with those of the Banco de España.

ISBN: 84-7793-076-7
Depósito legal: M. 2508 - 1991

Imprenta del Banco de España

1. Introduction

The recent burgeoning literature on unit roots and cointegration has helped to offer insight on the special consequences of econometric modelling with integrated variables. A random walk is a simple example of an integrated process, and this model has been extensively used to characterise the behaviour of many economic time series in both financial and commodity market theories. Moreover, following from the seminal work by Box and Jenkins (1976), ARIMA models have been widely used because of their ability to represent the behaviour of many time series. Thus, the treatment of integrated processes both in econometrics and statistics has expanded very rapidly, having developed diverse applications and a new methodology. Applications and theory have become increasingly diffused and fragmented, but they share a common feature, i.e. they are built upon an alternative asymptotic theory which takes into account the different statistical analysis underlying the behaviour in the limit of this type of non-stationary time series.

In the absence of a textbook which incorporates in a comprehensive form this increased diversity of results, it may be useful, from a pedagogic point of view, to take stock of the most important results in this field, interpreting such results and, also, comparing them to conventional central limit theory for stationary processes. This is the purpose of this paper and we believe that it could be useful to a substantial number of teachers for preparing material on this branch of the statistical literature for inclusion in econometrics and mathematical statistics courses. This, of course, does not preclude consultation of the original references, suitably quoted, where details and extensions of the results summarised in this paper, eschewed for greater brevity and simplicity, can be further analysed.

The statistical analysis to be presented below distinguishes between the unknown data generation process (DGP) and the assumed model. The problem to be considered is the behaviour of estimates and tests based on models which may not correspond to the DGP's. For pedagogic purposes we proceed through simple examples, first explaining the basic ideas and then introducing further "complications" which arise almost inevitably when analysing econometric time series.

This paper is organised as follows: Section 2 develops some preliminary notation and introduces the basic concepts of the appropiate limiting distribution for integrated processes of order one. In Section 3 we apply the previous theory to derive the distributions of several tests for the existence of unit roots. Section 4 examines results in multivariate regression models, including spurious regressions, detrending and cointegrating regressions, as well as issues related to causality tests in a framework of integrated variables. Finally, Section 5 extends some of the previous results to higher order integrated and near integrated time series.

2. Preliminary Theory

According to the definition by Engle and Granger (1987), an integrated process of order d, is a stochastic process which needs a d-th order differencing to achieve an invertible moving-average representation, also known as Wold representation. Drawing on the previous authors' we will denote these processes as I(d) processes. We will concentrate for most of this paper on the statistical properties which stem from the presence of a single unit root, i.e, on processes which are I(1), and only extend the results to more general I(d) processes in Section 5.

We will start by considering as the DGP, the following process

$$y_t = \rho_b\, y_{t-1} + \mu_b + u_t \; ; \; \rho_b = 1, \; y_0 = 0 \quad (t = 1, 2...) \tag{1}$$

or, after suitable integration

$$y_t = \mu_b\, t + y_t^* \; ; \; y_t^* = S_t = \sum_{i=1}^{t} u_i \; (t = 1, 2...) \tag{2}$$

where $\{u_t\}_1^\infty$ is a weakly stationary, zero mean innnovation sequence with spectral density $f_u(\lambda)$. In general, I(1) series such as y_t are linear functions of time (with a slope of zero if $\mu_b = 0$). The deviations from this function of time, denoted as y^*_t, are I(1) since they are the accumulation of past random shocks. Hence, in general any non-stationary series is the sum of a deterministic and a stochastic component. When discussing the properties of I(1) series we will generally refer to the latter.

Notice that the formulation (1) does not assume that u_t is a white noise disturbance, only its I(0) nature is presupposed. Therefore, to complete the specification of the DGP we need to impose some conditions on the sequence $\{u_t\}_1^\infty$. These restrictions are necessary if non degenerate limiting distributions of the statistics discussed below are to be derived. A weak set of conditions that achieve this aim are given in detail in Phillips (1987a) and can be summarised as follows:

<u>Assumption 1</u>: Let $\{u_t\}_0^\infty$ be a stochastic process such that

a) $E(u_t) = 0$ for all t

b) $\sup E|u_t|^\beta < \infty$ for some $\beta > 2$

c) $\omega^2 = \lim_{t \to \infty} T^{-1} E(S_T^2)$ exists and $\omega^2 > 0$ $(S_T = \sum_{i=1}^{T} u_i)$

where ω^2, denoted as the long-run variance, can also be written as

$$\omega^2 = \sigma^2 + 2\lambda$$

$$\text{where } \sigma^2 = E(u_1^2), \quad \lambda = \sum_{j=2}^{\infty} E(u_1 u_j)$$

d) u_t is strongly-mixing with mixing coefficients α_m such that

$$\sum_1^{\infty} \alpha_m^{(1-2/\beta)} < \infty$$

Condition (b) restrains the heterogeneity of the process, while (c) controls the normalisation at a rate which ensures non-degenerate limiting distributions. Condition (d) moderates the extent of temporal dependence in relation to the probability of outliers (see White (1984)).

The generality of the previous set of conditions implies that the expression in (1) encapsulates a wide variety of DGP's. These include virtually any ARMA model with a unit root and even ARMAX models with unit roots (see Andrews (1988)), where the exogenous variables are I(0). It is important to notice at this stage that only if we assume that the error term in (1) is $iid(0,\sigma^2)$, will $\omega^2=\sigma^2$. This restrictive case is, however, an interesting one since most of the limiting distributions that have been simulated are based on that assumption. Nevertheless, that will not be the case in most empirical applications and hence in general we will consider $\omega^2\neq\sigma^2$. Note that ω^2 has a very clear interpretation, as given in condition (c), if we look at the frequency domain, i.e. it is simply $2\pi\ f_u(0)$, where $f_u(0)$ is the spectral density at frequency zero. So, for example, if u_t is an MA(1) process, $u_t = \varepsilon_t - \theta\, \varepsilon_{t-1}$, then $\sigma^2 = \sigma_\varepsilon^2(1+\theta^2)$ whereas $\omega^2 = \sigma_\varepsilon^2(1+\theta)^2$.

As we mentioned above, the ordinary probability limits and central limit theorems (CLT) do not apply in the case of I(1) variables (neither in more general I(d), d > 1, cases). So, in order to derive proper limiting distributions, it is necessary, as in the stationary framework, to use a sequence of random variables, whose convergence is ensured by suitable transformations. Intuitively, when we are considering a time-series process which is dominated by a growing secular component, its evolution can be suitably smoothed by a choice of horizontal and vertical axis, which control for its explosivity and curvature respectively. More precisely, in the I(1) framework, we need to focus on the sequence $\{S_t\}_1^T$ which can be transformed so that each element of the sequence lies in the space D(0,1) of all real valued functions on the interval [0,1] that are right continous and have finite left limits. This is achieved by substituting the stochastic component, denoted by y^*_t, of the original series by the concentrated series.

$$y^*_T\,(r) = \frac{S_{[Tr]}}{T^{1/2}}\ ,\ r\ \varepsilon[0,1] \tag{3}$$

where [z] represents the integer part of any rational number z. In this way we are able to concentrate the original horizontal axis of 1 to T, to the closed interval [0,1], indexing the observations by r. For example, if T = 100, the original observation y_{50} will be indexed by $r\varepsilon[50,.51)$ and so on. The choice of the power of T in the denominator of (3) is such that the series y_t is neither explossive nor converges to zero. Since, for example, when u_t is iid$(0,\sigma^2)$, $\mathrm{var}(y^*_T) = \sigma^2\, T$, its standard deviation will be of order $O(T^{1/2})$ and this is precisely the power chosen to modify the ordinate axis.

Under Assumption 1, we have that as T tends to infinity

$$y^*_T(r) \rightarrow \omega B(r) \tag{4}$$

The symbol "->" here signifies weak convergence of the associated probability measure, while B(r) is a scalar Brownian motion with unit variance, also known as Wiener process, which lies in the space C[0,1] of all real valued functions continuous on the interval [0,1]. This is known as Donsker's Theorem and the interested reader is referred to Billingsley (1968) and Hall and Heyde (1980) for the details of the proof. Note that B(r) behaves like a random walk in continous time, so that for fixed r, $B(r) \equiv N(0,r)$ and has independent increments.

Moreover, an extension of the Slutsky Theorem in conventional asymptotic theory (see, e.g., White (1984)) also applies in this framework, in the sense that if g(.) is any continous function on C[0,1] then $y^*_T(r) \rightarrow \omega B(r)$ implies that

$$g[y^*_T(r)] \rightarrow g[\omega B(r)] \tag{5}$$

The previous results is known as the Continuous Mapping Theorem (CMT) (see Billingsley (1968)). The most striking difference between conventional and this new asymptotic theory is that whereas in the former the sample moments converge to constants, they converge to random variables in the latter. Similarly, as a result of the absence of stationarity and ergodicity, traditional CLT are substituted by Functional Central Limit Theorems (FCLT).

As an example of the previous remarks, let us take the sample mean of $\{y^*_t\}_1^T$ when $\alpha<1$ and $\alpha=1$ in (1). In the I(0) case, a simple application of the law of the Large Numbers (see White (1984)), will show that

$$\text{plim } T^{-1} \sum_{1}^{T} y_t^* = 0 \qquad (6)$$

since $E(y^*_t) = 0$

However, in the I(1) case, we will have that $\sum_{1}^{T} y_t^*$ can be written in terms of the corresponding Wiener process as follows

$$T^{-3/2} \sum_{1}^{T} y_t^* = T^{-1} \sum_{1}^{T} (T^{-1/2} \; y^*_{[tT/T]}) = \sum_{1}^{T} \int_{t-1/T}^{t/T} T^{-1/2} \; y^*_{[tT/T]} \; d\, r$$

$$= \int_0^1 \sum_{1}^{T} (T^{-1/2} \; \Sigma \; y^*_{[Tt/T]}) \; 1\!\!1_{t-1/T < r < t/T} \; d\, r \to \omega \int_0^1 B(r) \; d\, r \qquad (7)$$

by application of the CMT in (5) where $1\!\!1$ is an indicator function and g(.) is the integral function.

Similar techniques can be applied to show how the following standarised sample moments converge to functional of Wiener processes,

$$T^{-2} \sum_{1}^{T} y_t^{*2} \to \omega^2 \int_0^1 B^2(r) \; d\, r \qquad (8)$$

$$T^{-1} \; \Sigma \; y^*_{t-1} \; u_t \to \quad \frac{\omega^2}{2} \; [B(1)^2 - \frac{\sigma^2}{\omega^2}] \equiv \frac{\omega^2}{2} \; [B(1)^2 - 1] + \lambda \qquad (9)$$

$$T^{-5/2} \; \Sigma \; t \; y^*_t \to \omega \int_0^1 r \; B(r) \qquad (10)$$

Note that the difference between the orders of magnitude of these limiting distributions and the conventional stationary distributions, i.e. order of probability $O(T^{3/2})$ instead of $O(T)$ in (7), $O(T^2)$ instead of $O(T)$ in (8), $O(T)$ instead of $O(T^{1/2})$ in (9) and $O(T^{5/2})$ instead of $O(T^{3/2})$ in (10). These differences, for example, shed light on the non-conventional features and on coefficient consistency and limiting distributions when testing for unit roots. These will be analysed in the next section.

3. Unit Root Tests

Example 1: (Dickey-Fuller tests)

Let us suppose that y_t is generated by the DGP in (1), with $u_t \sim iid(0,\sigma^2)$, and we want to test the null hypothesis H_0: $\rho_c = 1$, $\gamma_c = 0$, in the model

$$y_t = \hat{\mu}_c + \hat{\gamma}_c\, t + \hat{\rho}_c\, y_{t-1} + \hat{u}_t \tag{11}$$

that is, the null hypothesis is that the series is a random walk with drift as in (1) and the alternative that it is stationary around a deterministic trend. Because of the unit root under the null hypothesis, it is convenient to use a transformation suggested by Sims, Stock and Watson (1990), so that under the null, (11) can be rewritten as

$$y_t = \hat{\theta}'\, z_{t-1} + \hat{u}_t \tag{12}$$

where $z_t=[z_t^1,\ z_t^2,\ z_t^3]$ and $\theta'=[\theta_1,\ \rho_c,\theta_3]$, where $z_t^1= 1$, $z_t^2= y_t^*=y_t-\mu_b t$, $z_t^3=t$ and θ_1 and θ_3 are a function of the parameters in (11). The transformed regressors are linear combinations of the original regressors with the linear combinations chosen to isolate the regressors with different stochastic properties: constant, integrated process with no time trend component and a linear time trend. Given the rates of convergence described in (7) - (10), the coefficients in $\hat{\theta}$ converge at different rates; so we need to define the scaling matrix Υ_T= diag $(T^{1/2},\ T,\ T^{3/2})$ partioned conformably with z_t and θ.

With these definitions, the OLS estimator of θ is given by

$$\hat{\theta} = (\sum_{2}^{T} z_{t-1} z'_{t-1})^{-1} (\sum_{2}^{T} z_{t-1} y_t) \tag{13}$$

Thus

$$\Upsilon_T[\hat{\theta} - \theta] = V_T^{-1} \phi_T \tag{14}$$

where $V_T = \Upsilon_T^{-1} \Sigma z_{t-1} z'_{t-1} \Upsilon_T^{-1}$ and $\phi_T = \Upsilon_T^{-1} \Sigma z_{t-1} u_t$

From (7) - (10) we can derive the limiting distribution of the six different elements in V_T and the three different elements in ϕ_T. This is done assuming that $\mu_b = 0$, without loss of generality since having included a trend in (11), the estimates $\hat{\theta}$ are invariant to the true value of μ_b. These elements are:

$V_{T,1,1} = T^{-1} \Sigma z_{1,t-1}^2 \rightarrow 1$

$V_{T,1,2} = T^{-3/2} \Sigma y_{t-1} \rightarrow \sigma \int_0^1 B(r)\, dr$

$V_{T,1,3} = T^{-2} \Sigma(t-1) \rightarrow 1/2$

$V_{T,2,2} = T^{-2} \Sigma y_{t-1}^{*2} \rightarrow \sigma^2 \int_0^1 B(r)^2\, dr$

$V_{T,2,3} = T^{-5/2} \Sigma y_{t-1}(t-1) \rightarrow \sigma \int_0^1 r\, B(r)\, dr$

$V_{T,3,3} = T^{-3} \Sigma(t-1)^2 \rightarrow 1/3$

$\phi_{T,1,1} \rightarrow N(0,\sigma^2) \equiv \sigma\, B(1)$

$\phi_{T,1,2} = T^{-1} \Sigma y^*_{t-1} u_t \rightarrow \sigma^2/2[B(1)^2 - 1]$

$\phi_{T,1,3} = T^{-3/2} \Sigma(t-1) u_t \rightarrow N(0,\sigma^2/3) \equiv \sigma \int_0^1 r\, d\, B(r)$

where the sums go from 1 to T.

If, as in the Dickey and Fuller test, we are particularly interested in the estimator of ρ_c and its t-ratio, $t\rho_c$, choosing the appropiate elements we would get

$$T(\hat{\rho}_c - 1) \rightarrow f\ (B) \tag{15}$$

and

$$t\rho_c = [\sigma^2\ V^{22}]^{-1/2}\ T(\hat{\rho}_c - 1) \rightarrow f(B) \tag{16}$$

where V^{22} is the second element on the diagonal of V^{-1}, and f(.), denote generically, an appropiate combination of the functionals of Wiener processes derived above. Henceforth, we will use the short notation f(B) to characterise different Wiener functionals. From (15) we note that $(\hat{\rho}_c - 1)$ converges at a rate $O(T^{-1})$ instead of the conventional $O(T^{-1/2})$, Similarly, from (16), the corresponding t-ratio has a non-degenerate distribution which is different from the standarised normal distribution which is used in conventional asymptotic theory.

There are analogous expressions for general Wald statistics for testing, e.g. joint hypothesis of the form $\rho_c = 1$, $\mu_c = 0$, $\gamma_c = 0$ or $\rho_c = 1$, $\mu_c = 0$ in (11). Suppose that the Wald statistic tests the q hypothesis $R\theta = r$ in (12). The test statistic is

$$F_T = (R\hat{\theta} - \theta)'\ [R(\sum_{1}^{T} z_{t-1}\ z'_{t-1})^{-1}\ R']^{-1}\ (R\hat{\theta} - \theta)/\hat{\sigma}^2 \tag{17}$$

Then the asymptotic behaviour of this test statistic is

$$F_T \rightarrow (R\theta - r)\ [R\ V^{-1}\ R']^{-1}\ (R\theta - r)/\sigma^2 \tag{18}$$

where V is the (3x3) matrix whose elements where derived above. The distributions of (15), (16) and (18) have been tabulated by numerical integration procedures by Dickey and Fuller (1979, 1981).

Example 2 (Augmented Dickey-Fuller tests)

In this case we asume that the DGP is similar to (1) but where the DGP is an AR(p) process with a unit root. The corresponding model can be appropiately parameterised as follows

$$y_t = \hat{\mu}_c + \hat{\gamma}_c t + \hat{\rho}_c y_{t-1} + \hat{\beta}(L) y_{t-1} + \hat{u}_t \tag{19}$$

where $\beta(L)$ is a lag-polynomial of order (p-1). Under the null hypothesis H_0: $\gamma_c = 0$, $\rho_c = 1$, the DGP corresponds to the AR(p) generalisation of (1) so that we can use again the transformation

$$y_t = \hat{\theta}' z_{t-1} + \hat{u}_t$$

where now $\theta' = (\beta', \theta_2, \rho_c, \theta_4)$ and $z'_t = (z_t^{1'}, z_t^2, z_t^3, z_t^4)$ with $z_t^{1'} = (\Delta y^*_t, \ldots, \Delta y^*_{t-p+1})$ with $\Delta y^*_{t-i} = \Delta y_{t-i} - \bar{\mu}_b$, $z_t^2 = 1$, $z_t^3 = y^*_t = y_t - \bar{\mu}_b t$, $z_t^4 = t$,

where $\bar{\mu}_b = E \Delta y_t = (1-\beta(1))^{-1} \mu_b$, i.e, the unconditional mean under the null. Defining the scaling matrix $\Upsilon_T = \mathrm{diag}(T^{1/2} I_p, T^{1/2}, T, T^{3/2})$ where I_p is an identity matrix and Ω_p the var-cov matrix of $\Delta y^*_t \ldots \Delta y^*_{t-p+1}$, so that $E(z_{1t} z'_{1t}) = \Omega_p$. The elements of the V_T and $\emptyset_T$ matrices are the same of before for the corresponding blocks, except for the following elements, appropriately defined

$$V_{T1,1} = T^{-1} \Sigma z_{1t-1} z'_{1t-1} \rightarrow \Omega_p$$

$$V_{T1,2} \rightarrow 0$$

$V_T(1,3) \to 0$

$V_T(1,4) \to 0$

$\emptyset_{T,1,1} \to N(0,\sigma^2\ \Omega_p)$

Therefore, V is block diagonal and the estimator of the nuisance parameters β are asymptotically normal and do not affect the asymptotic distribution of the Dickey-Fuller statistics. Thus the same tables of critical values can be used as above.

Example 3 (Non-Parametric Tests)

In Example 2, u_t was assumed to be $iid(0,\sigma^2)$, whereas in Example 3 this assumption could only achieved after filtering by an AR(p) process. In general if u_t is any ARMA model such that assumption 1 is satisfied, then the AR(p) approximation can be a poor choice. Phillips and Perron (1988) have suggested carrying out the test "as if" (11) was the mantained hypothesis and then modifying the corresponding test statistic by a non-parametric correction, so that the tables for the AR(1) can be still used to obtain critical values.

To illustrate the distributional properties of this approach, we will choose a simple particular case of (11) where $\mu_c=\gamma_c=0$. The extension to the more general cases is straightforward. Therefore the DGP is

$$y_t = y_{t-1} + u_t \tag{20}$$

where u_t satisfies assumption 1, whilst the model is

$$y_t = \hat{\rho}\ y_{t-1} + \hat{u}_t \tag{21}$$

Using the results in (7) - (10) the estimator $\hat{\rho}$ and its t-ratio t_ρ have the following limiting distributions

$$T(\hat{\rho}-1) = T(\Sigma\, y_{t-1}^2)^{-1}\,(\Sigma\, y_{t-1}\, u_t) \to (\int B(r)^2\, dr)^{-1}[\tfrac{1}{2}\,[B(1)^2-1] + \frac{\lambda}{\omega^2}] \quad (22)$$

$$t_\rho = (\hat{\sigma}\,\Sigma\, y_{t-1}^2)^{1/2}\;\; T(\hat{\rho}-1) \to \frac{\omega}{\sigma}\;\frac{1/2[B(1)^2 - 1] + \lambda/\omega^2}{[\int B^2(r)\, dr]^{1/2}} \quad (23)$$

Notice that if u_t were iid$(0,\sigma^2)$ then $\lambda=0$ and $\sigma^2=\omega^2$, which correspond to the case of the distributions of the two statistics simulated by Dickey and Fuller. Note that they have been isolated in the first terms of the RHS of the limiting distributions in (22) and (23). Following this procedure Phillips and Perron, suggest transforming the statistics (22) and (23), by computing the consistent sample counterparts of

$$Z(\hat{\rho}) = T(\hat{\rho}-1) - \frac{\lambda}{\omega^2}\,(\int B(r)^2\, dr)^{-1} \quad (24)$$

and

$$Z(t\rho) = \frac{\sigma}{\omega}\, t\rho - \frac{\lambda}{\omega^2}\,(\int B(r)^2\, dr)^{-1/2} \quad (25)$$

To implement the correction factors, we need consistent estimators of λ, ω^2 and the functional of the Wiener process; we can get consistent estimates of σ^2 and ω^2 from the residuals of (21) by means of the variance of the residuals, $\hat{\sigma}^2$, and the estimator of the long-run variance suggested by Newey and West (1987)

$$\hat{\omega}^2 = T^{-1}[\Sigma\, \hat{u}_t^2 + 2\sum_{j=1}^{\ell} w_\ell(j) \sum_{t=j+1}^{T} \hat{u}_t\, \hat{u}_{t-j}]$$

where $w_\ell(j) = 1 - j/1 + \ell$. This estimator is consistent when $\ell = O(T^{1/4})$ and condition b) in assumption 1 is strengthed to sup $E|u_t|^{2\beta}<\infty$ for some $\beta>2$. The Wiener functional, according to (8) can be estimated by $T^{-2}\;\Sigma\, y_{t-1}^2$.

Similar arguments can be used to obtain $Z(\rho_i)$ and $Z(t\rho_i)$ (i=b,c) as given in Phillips and Perron (1988).

Example 4 (Asymptotic normality of unit root tests)

An interesting case which has been emphasised by West (1988) is when carrying out a unit root test in a model which contains a constant or a constant and a trend, the same nuisance parameters appear also under the null hypothesis of a unit root.

For purposes of illustration, let us assume that y_t is generated by

$$y_t = \mu_b + y_{t-1} + u_t \tag{26}$$

with $u_t \sim iid(0,\sigma^2)$ and the mantained hypothesis is

$$y_t = \hat{\mu}_b + \hat{\rho}_b y_{t-1} + \hat{u}_t \tag{27}$$

where

$$T^{3/2} (\hat{\rho}_b - 1) = \frac{T^{-3/2}(\Sigma y_{t-1} u_t) - (T^{-2} \Sigma y_{t-1})(T^{-1/2} \Sigma u_t)}{T^{-3} \Sigma y_{t-1}^2 - T^{-4}(\Sigma y_{t-1}^2)} \tag{28}$$

Under the null, $y_t = \mu_b t + y^*_t$, and therefore

$$T^{-3} \Sigma y_{t-1}^2 \to \mu_b^2/3$$

$$T^{-2} \Sigma y_{t-1} \to \mu_b/2$$

$$T^{-3/2} \Sigma y_{t-1} u_t \to N(0, \sigma^2 \mu_b^2/3)$$

Subsbtituting the previous expressions into (28) it is easy to show that

$$T^{-3/2}\ (\rho_b-1)\ \to\ N(0,12\ \sigma^2/\mu^2) \tag{29}$$

and consequently

$$t\rho_b\ \to\ N(0,1) \tag{30}$$

Thus, both statistics are asymptotically normal. The intuition behind this result is that if the DGP is (26), the integrated series depends on a deterministic trend and a stochastic trend. The sample variability of the deterministic trend is $O(T^2)$ which dominates the sample variability of the stochastic trend which is $O(T)$. But it is well known that the existence of a deterministic trend in a regression model does not affect the asymptotic normality of the standarised estimator, hence normality follows.

The same result obtains when both the model and the DGP contain a drift and a trend. In this case it can be shown that $T^{5/2}(\hat{\rho}_c-1)\to N(0,180\ \sigma^2/\mu^2)$ and $t\rho_c\to N(0,1)$

<u>Example 5</u> (Unit root tests with general deterministic trends)

Following the methodology in Ouliaris, Park and Phillips (1988), we extend the results in Example 1, by letting y_t have the following DGP

$$y_t = \mu_d + \sum_{1}^{p-1} \gamma_k\ t^k + y_{t-1} + u_t \tag{31}$$

Where $u_t \sim iid(0,\sigma^2)$. The model corresponding to the maintained hypothesis is

$$y_t = \hat{\mu}_d + \sum_{1}^{p} \hat{\gamma}_k\ t^k + \hat{\rho}_d\ y_{t-1} + \hat{u}_t \tag{32}$$

The null hypotheses of interest are therefore $\rho_d=1$ and $\beta_p=0$.

To facilitate the derivation of the asymptotic distributions of the test statistics it is convenient to define $B_k(r)$ to be the stochastic process on [0,1] such that $B_k(r)$ is the projection residual of the Wiener process B(r) on the subspace generated by the polynomial functions 1, r,..., r^k in the Hilbert space of square integrable functions on [0,1]. It is also defined r_p to be the projection of r^p on the space spanned by the polynomials $1,r\ldots r^{p-1}$. Denoting by y^p_{t-1}, the projection of y_{t-1} on $1,t,\ldots,t^p$ we have that

$$T^{-2}\ \Sigma\ y^{p2}_{t-1} \to \sigma^2 \int_0^1 B^2_p(r)\ dr$$

$$T^{-1}\ \Sigma\ y^{p}_{t-1}\ u_t \to \sigma^2 \int_0^1 B_p(r)\ d\ B(r)$$

$$T^{-(2p+1)/2}\ \Sigma\ t^p\ u_t \to \sigma \int_0^1 r_p\ d\ B(r)$$

and therefore

$$T(\hat{\rho}_d-1) = [\Sigma\ y^{p2}_{t-1}]^{-1}\ [\Sigma\ y^{p}_{t-1}\ u_t] \to (\int_0^1 B_p(r)\ d\ B(r))^{-1}\ (\int_0^1 B^2(r)dr) \tag{33}$$

$$t\ \rho_d \to (\int_0^1 B_p(r)\ d\ B(r))\ (\int_0^1 B^2_p(r))^{-1/2} \tag{34}$$

$$F(\rho_d,\gamma_p) \to [(\int_0^1 B_p(r)\ d\ B(r))^2 (\int_0^1 B^2_p(r)dr)^{-1} + (\int_0^1 r_p\ d\ B(r))(\int_0^1 r^2_p)^{-1}] \tag{35}$$

Notice that (15), (16) and (17) are particular cases of (33), (34) and (35) for p=1. The corresponding distributions of (33) - (35) have been tabulated by Ouliaris, Park and Phillips (1988) up to p=5.

<u>Example 6</u> (Recursive and Sequential Statistics)

Following the analysis of Banerjee, Lumsdaine and Stock (1989), we can extend the tests for unit roots considered in examples 1 and 2 to two types of statistics: recursive and sequential statistics. The sequence of recursive statistics is computed recursively over subsamples of length k, for $k = k_0 \ldots T$ where k_0 is a startup value; the sequential statistics are computed using the full sample, where the statistics in the sequence vary by incrementing the date of the hypothetical break (or shift point). To illustrate the working of this approach we will consider a DGP as in (1), while the alternative hypothesis corresponding to the computation recursive and sequential statistics are given respectively by

$$y_t = \hat{\mu}_b + \hat{\rho}_b \, y_{t-1} + \hat{u}_t \tag{36}$$

and

$$y_t = \tilde{\mu}_b + \tilde{\mu}_{1b} \, \Upsilon_{1t} + \tilde{\rho}_b \, y_{t-1} + \tilde{u}_t \tag{37}$$

where $\Upsilon_{1t} = \mathbb{1}(t \geq k_0)$, i.e. a shift in the drift. Choosing a linear combination with zero mean regressors as in (12), we can write the recursive and sequential OLS estimator corresponding to (36) and (37) as

$$\hat{\theta}\,(\delta) = \Big(\sum_{1}^{[T\delta]} z_{t-1}\, z'_{t-1}\Big)^{-1} \Big(\sum_{1}^{[T\delta]} z_{t-1}\, y_t\Big) \tag{38}$$

$$\tilde{\theta}\,(\delta) = \Big(\sum_{1}^{T} z_{t-1}[T\delta]\; z'_{t-1}\,[T\delta]\Big)^{-1} \Big(\sum_{1}^{T} z_{t-1}\,[T\delta]\; y_t\Big) \tag{39}$$

where $T\delta = k$ $(k = k_0 \ldots T)$ in (38) and $k = k_0 \ldots T-k_0$ in (39), $z_t = (1,\, y_{t-1})$ in (38) and $z_t = (1,\, \Upsilon_{1t},\, y_{t-1})$ in (39).

Choosing scaling matrices Υ_T = diag $(T^{1/2},T)$ and Υ_T=diag $(T^{1/2},T^{1/2},T)$ respectively we have that

$$\Upsilon_T(\hat{\theta}(\delta)-\theta) \to V^R(\delta)^{-1}\emptyset^R(\delta) \tag{40}$$

and

$$\Upsilon_T(\tilde{\theta}(\delta)-\theta) \to V^s(\delta)^{-1}\emptyset^s(\delta) \tag{41}$$

where

$V^R_{11}(\delta) \to \delta$; $\emptyset^R_{11}(\delta) \to N(0, \sigma^2 . \delta)$

$V^R_{12}(\delta) \to \sigma \int_0^\delta B(r)\,dr$; $\emptyset^R_{11}(\delta) \to \frac{\sigma^2}{2}[B^2(\delta) - \delta]$

$V^R_{22}(\delta) \to \sigma^2 \int_0^\delta B^2(r)\,dr$

and

$V^s_{11}(\delta)\to 1$; $V^s_{22}(\delta)\to 1-\delta$; $\emptyset^s_{11}(\delta)\to N(0,s^2)$

$V^s_{12}(\delta)\to 1-\delta$; $V^s_{23}(\delta)\to\sigma \int_\delta^1 B(r)dr$; $\emptyset^s_{12}(\delta)\to N(0,\sigma^2\delta)$

$V^s_{13}(\delta)\to\sigma \int_0^1 B(r)dr$; $V^s_{33}(\delta)\to\sigma^2 \int_0^1 B^2(r)dr$; $\emptyset^s_{13}(\delta)=\frac{\sigma^2}{2}[B(1)^2-1]$

Several remarks serve to highlight different features of the previous results. First, the asymptotic representations (40) apply for $\delta\geq\delta_0>0$, i.e. it accounts for start up observations, while those in (45) apply for $0<\delta_0\leq\delta\leq(1-\delta_0)<1$, i.e. requires a "trimming"value. Second, the results apply uniformly in δ, i.e, the "marginals at any fixed δ are simply those that would be obtained using conventional arguments. Third, the results can be extended to other test-statistics like t-ratios, F-tests, etc. as well as to other hypothetical breaks, e.g. shift and/or jump in trends, etc. Critical values have been computed using simulation procedures by Perron (1989), Banerjee, Lumsdaine and Stock (1989) and Banerjee, Dolado and Galbraith (1990).

4. Multivariate Regression Models

We now extend the previous analysis to regression models containing several integrated variables which may be cointegrated, including time trends.

Example 7 (Spurious detrending)

This case, analysed by Durlauf and Phillips (1986), deals with the issue of innapropiate de-trending of integrated processes, under the traditional belief that conventional asymptotic theory could be applied to de-trended time series. Let $\{y_t\}_1^\infty$ have the DGP given in (1) and consider the model

$$y_t = \hat{\mu} + \hat{\gamma}t + \hat{e}_t \tag{42}$$

Then, from (1), since

$$y_t = \mu_b\, t + y^*_t \quad (y^*_t = S_t)$$

(42) can be rewritten as

$$y^*_t = \hat{\mu} + (\hat{\gamma} - \mu_b)t + \hat{e}_t \tag{43}$$

Therefore

$$\hat{\mu} = \frac{(\Sigma\, t^2)(\Sigma y_t^*) - (\Sigma t)(\Sigma t\ y_t^*)}{\psi_T}$$

$$(\hat{\gamma} - \mu_b) = \frac{T(\Sigma t\ y_t^*) - (\Sigma t)\ (\Sigma\ y_t^*)}{\psi_T}$$

where $\psi_T = T\ \Sigma\ t^2 - (\Sigma t)^2$

Using the limiting distributions in (7) - (10), it is easy to show that

$$T^{-1/2}\ \hat{\mu} \rightarrow 4\ \sigma \int_0^1 B(r)\ dr - 6\ \sigma \int_0^1 r\ B(r)\ dr \tag{44}$$

$$T^{1/2}(\hat{\gamma}-\mu_b) \rightarrow 12\ \sigma[\int_0^1 r\ B(r)dr - 1/2 \int_0^1 B(r)dr] \tag{45}$$

Similarly it can be shown that

$$T^{-1/2}\ t_{\gamma} \rightarrow f(B)$$

$$T^{-1/2}\ t_{\mu} \rightarrow f(B)$$

$$T^{-1}\ \hat{\sigma}_e^2 \rightarrow f(B)$$

$$R^2 \rightarrow f(B)$$

$$TDW \rightarrow f(B)$$

where f (B) are generic functionals of Wiener processes as in (44) and (45).

From the previous results, we observe that the $\hat{\gamma}$ estimator is consistent, converging to its true value μ_b at a rate $O(T^{-1/2})$. However, its t-ratio diverges to infinity, confirming the Monte Carlo results of Nelson and Kang (1981). Both the drift and its t-ratio diverge. The estimated variance of the residuals ($\hat{\sigma}_e^2$) also diverges, reflecting the fact that the residuals of the model are integrated around the trend. The coefficient of multiple correlation (R^2) converges to a non-degenerate limiting distribution. The results for the Durbin-Watson statistic appear quite promising, confirming its powerful role as a misspecification diagnostic (see Sargan and Bhargava (1983)).

<u>Example 8</u> (Spurious Regression)

To illustrate the consequences of running regression models where variables are spuriously related (as discussed by Yule (1926) and Granger and Newbold (1974)), we apply the previous limiting distribution to the following case (see Phillips (1987)). Let $\{y_t\}_1^\infty$ and $\{x_t\}_1^\infty$ be generated by the following DGP.

$$y_t = y_{t-1} + u_t \tag{46}$$

$$x_t = x_{t-1} + \varepsilon_t \tag{47}$$

where $u_t \sim iid(0,\sigma_u^2)$, $\varepsilon_t \sim iid(0,\sigma_\varepsilon^2)$ and $E(u_t\ \varepsilon_s) = 0\ \forall t,s$

The regression model is

$$y_t = \hat{\mu} + \hat{\beta}\ x_t + \hat{e}_t \tag{48}$$

To facilitate the derivation of the asymptotic distribution of the estimators and test-statistics in (48) it is convenient to define $\sigma_u\ B_u(r)$ and $\sigma_\varepsilon\ B_\varepsilon(r)$ to be the Wiener processes on [0,1] obtained from the disturbances in (46) and (47). The limiting distributions in (7) - (10) can be applied to this case plus the following cross-moment limiting distribution derived in a similar fashion

$$T^{-2} \Sigma x_t y_t \to \sigma_u \sigma_\varepsilon \int_0^1 B_u(r) B_\varepsilon(r) dr \quad (49)$$

Therefore, the OLS estimator of μ and β in (48) are such that

$$T^{-1/2} \hat{\mu} \to [\int_0^1 B_\varepsilon^2 - (\int_0^1 B_\varepsilon)^2]^{-1} \sigma_u [(\int_0^1 B_\varepsilon^2)(\int_0^1 B_u - (\int_0^1 B_\varepsilon)(\int_0^1 B_\varepsilon B_u)] \quad (50)$$

$$\hat{\beta} \to [\int_0^1 B_\varepsilon^2 - (\int_0^1 B_\varepsilon)^2]^{-1} (\sigma_u/\sigma_\varepsilon)[\int_0^1 B_u B_\varepsilon - \int_0^1 B_u \int_0^1 B_\varepsilon] \quad (51)$$

where the differentials in the notation of the corresponding integrals have been ommited for simplicity in the notation.

Similarly it is easy to show that

$$T^{-1} \hat{\sigma}_\varepsilon^2 \to f(B) \; ; \; T^{-1/2} t_\mu \to f(B) \; ; \; T^{-1/2} t_\beta \to f(B)$$

$$R^2 \to f(B) \; ; \; TDW \to f(B)$$

where again f(B) denote generic functionals of the Wiener processes corresponding to the first and second moments of y_t and x_t.

This case interprets the familiar Monte-Carlo results of Granger and Newbold (1974), reinforcing analytically the divergence of t_β despite the fact that $\hat{\beta}$ and R^2 have non-degenerate distributions. Again, as in Example 6 the DW statistic detect misspecification of the model. Finally, it is important to notice that if we detrended the variables entering the regression model, the coefficient of the trend is $O(T^{-1/2})$ and is therefore consistent. The orders of magnitude of the remaining estimators are the same.

Example 9 (Cochrane-Orcutt procedure)

Let now $\{y_t\}_1^\infty$ and $\{x_t\}_1^\infty$ be generated as in Example 8, except that, without loss of generality, $E(u_t \varepsilon_t) \neq 0$. This implies the following DGP

$$\Delta y_t = \beta \, \Delta x_t + u_t \tag{52}$$

$$x_t = x_{t-1} + \varepsilon_t \tag{53}$$

The regression model, abstracting from the constant term for simplicity, is

$$y_t = \beta \, x_t + e_t \tag{54}$$

$$e_t = \rho \, e_{t-1} + u_t \tag{55}$$

In a classic paper, Cochrane and Orcutt (1949) suggested estimation of (β, ρ) by quasi-differencing the data:

$$y^*_t = y_t - \hat{\rho} \, y_{t-1}$$

$$x^*_t = x_t - \hat{\rho} \, x_{t-1}$$

where $\rho = (\Sigma \, \hat{e}_t \, \hat{e}_{t-1}) \, / \, \Sigma \, \hat{e}^2_{t-1}$, $\hat{e}_t$ being the OLS residuals in (54) obtained from the OLS estimator $\hat{\beta}$ in (54). Then apply OLS to the transformed equation

$$y^*_t = \tilde{\beta} \, x^*_t + \tilde{e}_t \tag{56}$$

where under the DGP, $\Delta e_t = u_t$. To obtain the asymptotic distribution of $\tilde{\beta}$, it is necessary to obtain first the asymptotic distribution of $\hat{\beta}$ and $\hat{\rho}$. Since (54) is a "spurious regression", we know that

$$\hat{\beta} - \beta \rightarrow f(B) \tag{57}$$

as in (51), where now $\beta \neq 0$

Similarly, under the DGP

$$T(\hat{\rho} - 1) \rightarrow f(B) \tag{58}$$

since

$$T^{-2} \Sigma \hat{e}^2_{t-1} = T^{-2} \Sigma[e_{t-1} - (\hat{\beta}-\beta) x_{t-1}]^2 = T^{-2}(\hat{\beta}-\beta) \Sigma x_t e_t + o(1) \to f(B)$$

given (49) and (57) (note that x_t and e_t are I(1)), and

$$T^{-1} \Sigma \Delta \hat{e}_t \hat{e}_{t-1} = T^{-1} \Sigma u_t e_{t-1} - (\hat{\beta}-\beta) T^{-1} [\Sigma u_t x_{t-1} + \Sigma \varepsilon_t e_{t-1}]$$

$$+ (\hat{\beta}-\beta)^2 T^{-1} \Sigma \varepsilon_t x_{t-1} \to f(B)$$

Therefore, from (57) and (58), it can be shown that $\tilde{\beta}$ in (56) tends to β, a constant, not to a random variable as does $\hat{\beta}$. This is so since

$$\tilde{\beta} = (\Sigma y*_t y*_t) / (\Sigma x*_t)^2$$

and

$$T^{-1} \Sigma x*^2_t = T^{-1} \Sigma[\Delta x_t-(\hat{\rho}-1) x_{t-1}]^2 = T^{-1} \Sigma \Delta x^2_t - 2(\hat{\rho}-1) T^{-1}\Sigma \varepsilon_t x_{t-1}+$$

$$+(\hat{\rho}-1)^2 T^{-1} \Sigma x^2_{t-1} \to E(\varepsilon^2_t) = \sigma^2_\varepsilon$$

$$T^{-1}\Sigma x^*_t y^*_t = T^{-1} \Sigma \Delta y_t \Delta x_t - (\hat{\rho}-1) T^{-1} [\Sigma \Delta y_t x_{t-1} + \Sigma \Delta x_t y_{t-1}]$$

$$+(\hat{\rho}-1)^2 T^{-1} \Sigma y_{t-1}x_{t-1} \to E(\Delta y_t \Delta x_t) = \beta \sigma^2_\varepsilon$$

Hansen (1990), relying upon the previous results, has suggested obtaining the second-stage estimator of ρ, by computing

$$\tilde{\rho} = \Sigma\, \tilde{e}_t\, \tilde{e}_{t-1} / (\Sigma\, \tilde{e}_{t-1})^2$$

where $\tilde{e}_t$ are the second-stage residuals, as defined in (56).

Given that

$$T^{-2}\, \Sigma\, \tilde{e}^2_{t-1} = T^{-2}\, \Sigma[e_{t-1} - (\tilde{\beta}-\beta)\, x_{t-1}]^2 = T^{-2}\, \Sigma\, e^2_{t-1} - 2(\tilde{\beta}-\beta) T^{-2} \Sigma\, e_{t-1} x_{t-1}$$

$$+ (\tilde{\beta}-\beta)\, T^{-2}\, \Sigma\, x^2_{t-1} \to \sigma^2_u \int_0^1 B^2_u$$

since $(\tilde{\beta}-\beta)$ is o(1), and

$$T^{-1}\, \Sigma\, \Delta\, \tilde{e}_t\, e_{t-1} = T^{-1}[\Delta\, e_t - (\tilde{\beta}-\beta)\, \Delta\, x_t]\, [e_{t-1} - (\tilde{\beta}-\beta)\, x_{t-1}] =$$

$$= T^{-1}\, \Sigma\, u_t\, e_{t-1} - (\tilde{\beta}-\beta)\, T^{-1}[\Sigma\, u_t\, x_{t-1} + \Sigma\, \varepsilon_t\, e_{t-1}] + (\tilde{\beta}-\beta)^2\, T^{-1}\, \Sigma\, \varepsilon_t x_{t-1}$$

$$\to 1/2\, \sigma^2_u [B^2_u(1) - 1]$$

Therefore

$$T(\tilde{\rho}-1) \to 1/2\, [B^2_u(1) - 1] \,/\, \int_0^1 B^2_u \qquad (59)$$

which is identical to the distribution of the univariate Dickey-Fuller unit root test. The advantage of the test based upon (59) over the cointegration test discussed in Examples 7 and 8, is that it does not depend on the number of variables included in the regression model (constrained for simplicity to a single regressor in these examples). This independence of dimensionality is important since the critical values presented in Sargan and Bhargava (1983) or in Engle and Yoo (1987) reveal that the asymptotic distributions of the test statistics shift away from the origin as the dimensionality increases, and this is expected to reduce power.

<u>Example 10</u> (Cointegrating Regression)

Let now $\{y_t\}_1^\infty$ and $\{x_t\}_1^\infty$ be generated by

$$y_t = \beta x_t + u_t \tag{60}$$

$$\Delta x_t = \varepsilon_t \tag{61}$$

with $u_t \sim iid(0,\sigma_u^2)$, $\varepsilon_t \sim iid(0,\sigma_\varepsilon^2)$, $E(u_t \varepsilon_s) = \delta_{ts} \sigma_{u\varepsilon}$, where δ_{ts} is Kroenecker's delta. From (61) we can see that x_t is I(1). Substituting (61) in (60) and differencing we get

$$\Delta y_t = \beta \varepsilon_t + \Delta u_t$$

i.e. y_t is IMA(1,1) and therefore I(1) as well. There is however a linear combination of y_t and x_t given by (60) which is I(0). The estimator of β in the regression model

$$y_t = \hat{\beta} x_t + \hat{u}_t$$

is given by

$$\hat{\beta} = [\Sigma x_t^2]^{-1} [\Sigma x_t y_t] = \beta + [\Sigma x_t^2]^{-1} [\Sigma x_t u_t] \tag{62}$$

The limiting distribution of $T^{-2} \Sigma x_t^2$ is given by the corresponding expression in (8), whilst to get the limiting distribution of $\Sigma x_t u_t$, it is convenient to condition u_t on ε_t such that

$$u_t = \gamma \varepsilon_t + v_t \; ; \; \gamma = \sigma_{u\varepsilon}/\sigma_\varepsilon^2; \; \sigma_v^2 = \sigma_u^2 - \sigma_{u\varepsilon}^2/\sigma_\varepsilon^2$$

where $E(\varepsilon_t v_s) = 0$

Then it is possible to show that

$$T^{-1} \Sigma x_t u_t = T^{-1} \Sigma x_t(\gamma \varepsilon_t + v_t) = T^{-1} \Sigma (x_{t-1} + \varepsilon_t)(\gamma \varepsilon_t + v_t)$$

$$= T^{-1} [\gamma \Sigma x_{t-1} \varepsilon_t + \gamma \Sigma \varepsilon_t^2 + \Sigma x_{t-1} v_t] + O(T^{-1/2})$$

$$\to \gamma \sigma_\varepsilon^2/2 \, [B_\varepsilon^2(1) + 1] + \sigma_\varepsilon \quad \sigma_v \int_0^1 B_\varepsilon \, d B_v \qquad (63)$$

Since ε_t and v_t are independent by construction, it can also be shown using similar arguments to the Mann and Wald (1943) Theorem that conditional on the σ-algebra $\mathcal{F} = \sigma(B_v(r) \quad 0<r<1)$.

$$\int_0^1 B_\varepsilon \, d B_v \to N(0, \int_0^1 B_\varepsilon^2) \qquad (64)$$

and therefore

$$[\int_0^1 B_\varepsilon^2]^{-1/2} \int_0^1 B_\varepsilon \, d B_v \to N(0,1) \qquad (65)$$

The previous asymptotic distributions are known as "mixture of normals" (henceforth MN) (see Billingsley, 1968).

Substituting (8) and (63) into (60) we get

$$T(\hat\beta-\beta) \to [\gamma \sigma_\varepsilon^2/2 \, [B_\varepsilon^2(1) + 1] + \sigma_\varepsilon \sigma_v \int_0^1 B_\varepsilon \, d B_v] \, (\sigma_\varepsilon^2 \int_0^1 B_\varepsilon^2)^{-1} \qquad (66)$$

and

$$t_\beta \to \frac{\gamma/2 \, [B_\varepsilon^2(1) + 1]}{\sigma_u [\int_0^1 B_\varepsilon^2]^{1/2}} + N(0,1) \qquad (67)$$

So, in general, $\hat\beta$ is a "super-consistent" estimator of β (see Stock (1987) but its t-ratio will not have a standard distribution unless $\gamma = 0$, i.e. x_t is exogenous (weakly and strongly in this example). In fact when $\gamma \neq 0$ the first term in

(66) gives rise to the so called "second-order" or "endogeneity" bias (see Phillips and Hansen (1988) and Gonzalo (1989)) which, though asymptotically neglegible, can be important in finite samples, as emphasised by Banerjee et al. (1986).

Similar arguments can be used to show that

$$T(1 - R^2) \rightarrow f(B)$$
$$DW \rightarrow 2$$

The latter result obtains because u_t has been assumed to be iid. If it were correlated then $DW \rightarrow 2(1-\rho_1)$, where ρ_1 is the first order autocorrelation. If a constant term, μ, is included in the model, then $T^{1/2}\,\hat{\mu} \rightarrow f(B)$.

The existence of nuisance parameters is also important in this case. Let us suppose that x_t is generated by

$$\Delta x_t = \mu + \varepsilon_t$$

Then x_t is dominated by a linear trend and hence

$$T^{3/2}(\hat{\beta}-\beta) \rightarrow N(0,\ 3\,\sigma_u^2\,\sigma_\varepsilon^2/2) \tag{68}$$

The reason why (68) obtains is that the linear combination (1,-β) not only cointegrates y_t and x_t but also their respective trends. This is seen by noticing that substituting (67) in (60) and differencing we get

$$\Delta y_t = \beta\,\mu + \beta\,\varepsilon_t + \Delta u_t$$

i.e. y_t has a linear trend with slope β.

If a linear trend is included in the model, then the asymptotic normality in (68) disappears, since the equation can be reparameterised as a regresssion model with a linear trend and zero mean regressors, and results similar to (66) hold.

Example 11 (Fully Modified Estimator)

From (66) we have seen that the elements of the cointegrating vector converge to their true values super-consistently, i.e. at a rate $O(T^{-1})$, but the asymptotic distribution of their test-statistics is non-standard, unless the regressor satisfies certain exogeneity properties. Phillips and Hansen (1988) in a similar vein to the non-parametric approach examined in Example 3, have proposed a non-parametric correction which converts the distributions of the transformed estimators into "mixture of normals".

To illustrate the nature of the approach, consider the DGP given in (60) and (61). The conditional distribution of y_t on x_t can be written as

$$y_t = \beta\, x_t + \gamma\, \Delta\, x_t + v_t \tag{69}$$

so that estimation by OLS of β and γ in this model is equivalent to estimating (60) and (61) by full-information maximum likelihood. It is easy to show that in this case

$$T\,(\hat{\beta}-\beta) \rightarrow [\textstyle\int_0^1 B_\varepsilon^2]^{-1}\, \sigma_v/\sigma_u\, [\int_0^1 B_\varepsilon\, d\, B_v] \equiv MN$$

and

$$T^{1/2}(\hat{\gamma}-\gamma) \rightarrow N(0,\ \sigma_v^2/\sigma_\varepsilon^2)$$

Since the asymptotic variance-cov matrix of x_t and $\Delta\, x_t$ is diagonal, Phillips and Hansen propose estimating γ from the regression of the residuals in (60) on $\Delta\, x_t$ and then estimating β in a second step in the model

$$y_t - \hat{\gamma}\,\Delta x_t = \beta\, x_t + u_t \tag{70}$$

which is asymptotically equivalent to the FIML estimator.

The previous procedure can be extended to more general cases as the following example illustrates. Let $\{y_t\}_1^\infty$, $\{x_t\}_1^\infty$ have the DGP given by (60) and (61) where now $(u_t,\ \varepsilon_t)$ are assumed to have the following "long-run" variance

$$\Omega = \begin{pmatrix} \omega_u^2 & \omega_{u\varepsilon} \\ \cdot & \omega_\varepsilon^2 \end{pmatrix} = 2\,\pi\, f(0) \tag{71}$$

Then, since $\omega_u\ B_u(r)=\gamma\ \ \omega_\varepsilon\ \ B_\varepsilon(r)+\ \ \omega_v\ \ B_v(r)$ with $\gamma= \omega_{u\varepsilon}/\omega_\varepsilon^2$, we have that

$$T^{-1}\Sigma\, x_t\ v_t \;\text{->}\; \omega_\varepsilon\ \omega_u \int_0^1 B_\varepsilon\ d\,B_u + \lambda - \gamma[\omega_\varepsilon^2 \int_0^1 B_\varepsilon\ d\ B_\varepsilon + \omega_\varepsilon^2] \tag{72}$$

where λ is now the non contemporaneous long run covariance

$$\lambda = \sum_{k=1}^{\infty} E(\varepsilon_0\ u_k)$$

Then, subtituting $\omega_u\ d\ B_u(r) = \gamma\ \omega_\varepsilon\ d\ B_\varepsilon(r) + \omega_v\ d\ B_v(r)$ in (72) we get

$$T^{-1}\ \Sigma\, x_t\ v_t - (\lambda - \gamma\,\omega_\varepsilon^2) \;\text{->}\; \omega_\varepsilon\ \omega_v \int_0^1 B_\varepsilon\ d\ B_v \tag{73}$$

which is a "mixture of normals" according to (64).

Substituting (69) into (73) we get the so called "fully modified estimator", given by

$$\hat{\beta}^+ = (\Sigma\, x_t^2)^{-1}\ [\Sigma\, x_t\ (y_t - \hat{\gamma}\ \Delta\ x_t) - T(\hat{\lambda} - \hat{\gamma}\ \ \hat{\sigma}_\varepsilon^2) \tag{74}$$

where γ, λ, and ω_{ε}^{2} can be consistently estimated from the residuals in (60) and (61) using a truncating lag $\ell = O(T^{1/4})$.

The corresponding t-ratio is given by

$$t_{\beta}= (\hat{\beta}^{+}-\beta)(\Sigma\, x_t^2)^{1/2} / \omega_v \rightarrow N(0,1)$$

where notice that in the computation of conventional "standard errors" calculated by statistical packages $\hat{\sigma}_u$ is to be substituted by $\hat{\omega}_v$.

<u>Example 12</u> (Causality Tests)

By means of a simple example, we can use the theoretical arguments developed in Examples 10 and 11, to analyse the consequences of having I(1) variables when examining two common tests of linear restrictions in applied work: a test for the number of lags with which a variable should enter a regression equation and a "causality" or predictability test that contemporaneous or lagged values of one variables do not enter the equation for a second variable. To simplify the discussion without loss of generality we abstract from drifts. A general treatment can be found in Sims, Stock and Watson (1990) and an application to efficiency tests in Banerjee and Dolado (1988).

Let $\{y_t\}_1^{\infty}$ and $\{x_t\}_1^{\infty}$ have the DGP in (46) and (47), where the regression model is

$$y_t = \hat{\alpha}_0\, x_t + \hat{\alpha}_1\, x_{t-1} + \hat{\alpha}_2\, y_{t-1} + e_t \qquad (75)$$

A test of the null hypothesis "x does not Granger-cause y" is $H_0^1: \alpha_0=\alpha_1=0$ or $H_0^2: \alpha_0+\alpha_1=0$ whereas a test of the appropiate lag length is $H_0^3: \alpha_0=0$ $H_0^4: \alpha_1=0$, $H_0^5: \alpha_2=0$ or $H_0^6: \alpha_1=\alpha_2=0$.

Because under the DGP, y_t follow a random walk it is convenient to rewrite (75) as

$$\Delta y_t = \hat{\theta}_0 \ \Delta \ x_t + \hat{\theta}_1 \ x_{t-1} + \hat{\theta}_2 \ y_{t-1} + \hat{e}_t = \hat{\theta}' z_t + \hat{e}_t \tag{76}$$

where $z_t = (\Delta x_t, x_{t-1}, y_{t-1})'$ and $\theta' = (\theta_0, \theta_1, \theta_2)$ with $\theta_0 = \alpha_0$, $\theta_1 = \alpha_0 + \alpha_1$ and $\theta_2 = \alpha_2 - 1$.

Choosing a scaling matrix $\Upsilon_T = \text{diag}\ (T^{1/2}, T, T)$, we have that

$$\Upsilon_T [\hat{\theta} - \theta] = V_T^{-1} \ \emptyset_T$$

where $V_T = \Upsilon_T^{-1} \ \Sigma \ z_t \ z'_t \ \Upsilon_T^{-1}$ and $\emptyset_T = \Upsilon_T^{-1} \ \Sigma \ z_t \ u_t$

The limiting distributions of the different elements in V_T and $\emptyset_T$ are

$V_{T,1,1} \to \sigma_\varepsilon^2$

$V_{T,1,2} \to 0$

$V_{T,1,3} \to 0$

$V_{T,2,2} \to \sigma_\varepsilon^2 \int_0^1 B_\varepsilon^2(r)\ dr$

$V_{T,2,3} \to \sigma_u \ \sigma_\varepsilon \int_0^1 B_u(r)\ B_\varepsilon(r)\ dr$

$V_{T,3,3} \to \sigma_u^2 \int_0^1 B_u^2(r)\ dr$

$\phi_{T,1,1} \to N(0, \ \sigma_u^2 \ \sigma_\varepsilon^2)$

$\phi_{T,1,2} \to \sigma_u \ \sigma_\varepsilon \int B_u(r)\ d\ B_\varepsilon(r) \equiv \sigma_u \ \sigma_\varepsilon \ N(0, \int B_u^2)$

$$\phi_{T,1,3} \rightarrow \frac{\sigma_u^2}{2} [B_u(1)^2 - 1]$$

Therefore V is block diagonal with respecto to Δx_t and the estimator of $\theta_0(=\alpha_1)$ is asymptotically normal. This implies that a test like H_0^4 in (65) follows asymptotically the standarised normal distribution. Similarly since the reparametrisation in (66) is identical to that having Δx_t and x_t as regressors, now with $\theta_0 = -\alpha_1$ and $\theta_1 = \alpha_1 + \alpha_2$, the same argument applies to H_0^3. However the test for H_0^1, H_0^2, H_0^5, H_0^6, are functionals of the Wiener processes given above and the corresponding t and F test-statistics do not follow standard distributions.

Let us assume now that $\varepsilon_t = u_{t-1}$ in (47), then by substraction of (46) from (47), we find that y_t and x_t are cointegrated with cointegrating vector (1,-1); i.e,

$$y_t = x_t + \varepsilon_t \tag{77}$$

In this case the limiting distributions of $V_{T,2,2}$, $V_{T,3,3}$ are identical and the corresponding submatrix does not have full rank, reflecting the asymptotic perfect collinearity between x_{t-1} and y_{t-1} given in (77). However in this case (75) can be reparameterised as follows

$$\Delta y_t = \hat{\theta}'_0 \, \Delta x_t + \hat{\theta}'_1(y_{t-1} - x_{t-1}) + \hat{\theta}'_2 \, y_{t-1} + \hat{e}_t \tag{78}$$

where $\theta_0 = \alpha_0$, $\theta_1 = (\alpha_0 + \alpha_1)$, $\theta_2 = (\alpha_0 + \alpha_1 + \alpha_2 - 1)$

By similar arguments as before it can be shown that by choosing the scaling matrix $\Upsilon_T = \mathrm{diag}(T^{1/2}, T^{1/2}, T)$, the corresponding V_T matrix is a diagonal matrix an that the joint distribution of the estimator of θ'_2 and θ'_1 is asymptotically normal, hence H_0^1 or H_0^2 in (75) can also be tested using the standard distributions. This result can be generalised to any cointegrating vector in (78) and in general can be stated as follows: Parameters that can be rewritten as coefficients on mean zero, non-integrated regresors will be asymptotically normally distributed, while any other coefficients will have non-normal asymptotic distributions.

5. Extensions to Higher Order and Near-Integrated Variables

It should be noted that the results obtained in all the examples examined above can be generalised to any degree of differencing d>1 as follows.

Let $\{y^*_t\}_1^\infty$ be an stochastic process with the following Wold representation

$$(1-L)^d \; y^*_t = u_t, \; y^*_0 = \ldots = y^*_{-d} = 0 \tag{79}$$

where $\{u_t\}_d^\infty$ has the same properties as in (1). Then expanding $(1-L)^{-d}$ around L=0 we get

$$y_t = \sum_{j=d}^{t} \theta(t-j) \, u_j \tag{80}$$

with $\theta(t-j)=\Gamma(t-j+d)/\Gamma(d)(t-j)!$, where $\Gamma(.)$ is the gamma function such that $\Gamma(d)=(d-1)!$. Then it is possible to substitute for the original y^*_t with the concentrated series

$$y^*_T(r) = \frac{S_{[Tr]}}{T^{d-1/2}} = \frac{\sum_{j=d}^{[Tr]} \theta([Tr]-j) \, u_j}{T^{d-1/2}} \tag{81}$$

Under Assumption 1, we have that as $T\uparrow\infty$

$$y^*_T(r) \to \omega\, B_d(r) = \frac{\omega}{\Gamma(d)} \int_0^r (r-d)^{d-1}\, d\, B_d(s) \qquad (82)$$

where $B_d(r)$ is a scalar d^{th} - order Wiener process (see Billingsley (1968) and Gourieroux, Maurel and Monfort (1988)). Integrating by parts, it is easy to derive the following particular cases

$$d = 1 \to B_d(r) = B(r)$$

$$d = 2 \to B_d(r) = \int_0^r B(r)\, dr$$

etc.

Similarly, the limiting distributions in (8) - (10), will gèneralise as follows

$$T^{-2d} \Sigma\, y^{*2}_t \to \omega^2 \int_0^1 B^2_d(r)\, dr \qquad (83)$$

$$T^{-d} \Sigma\, y^*_{t-1}\, u_t \to \omega^2 \int_0^1 B_d(r) dB(r) \qquad (84)$$

$$T^{-(d+3/2)} \Sigma\, t\, y^*_t \to \omega \int_0^1 r\, B_d(r)\, dr \qquad (85)$$

Finally, it should be remarked that several simulation exercises have shown that the discriminatory power of test statistics for the presence of unit roots is low against the alternative hypothesis of roots which are close to unity. This is explained because although we have shown that there is a discontinuity between the distribution theory applicable to stationary and integrated cases, that discontinuity only holds in the limit, and for finite samples the distributions are much more similar. Phillips (1987b, 1988), has developed an asymptotic theory for near integrated variables, which helps to bring together the apparently divergent theories mentioned above. The following example, related to testing for a unit root, tries to clarifly the issue.

<u>Example 13</u> (Near-Integrated Variables)

Let $\{y_t\}_1^{\infty}$ be generated by the following DGP

$$y_t = \rho\, y_{t-1} + u_t \tag{86}$$

where ρ = exp (c/T), where c is a fixed number and u_t satisfies assumption 1. Note that when c=0, y_t is an I(1) process and when $c \neq 0$, (6) represents a local alternative to H_0: c = 0.

To derive the limiting distribution of the test statistic for H_0, it is convenient to define the following functional, also known as Ornstein-Uhlenbeck or diffusion process

$$K_c(r) = B(r) + c \int_0^r \exp\,(c(r-s))\, B(s) \tag{87}$$

where B(r) is a unit variance scalar Brownian motion, and $K_c(r)$ is a Gaussian process, so that for fixed r, $K_c(r) \equiv N(0,(\int \exp(c(r-s))^2)$.

Using similar arguments as in Section 1, it is possible to prove that as $T\uparrow\infty$,

$$y^*_T(r) \to \omega\, K_c(r) \tag{88}$$

and

$$T^{-2}\, \Sigma\, y^{*2}_t \to \omega^2 \int_0^1 K_c(r)^2 \tag{89}$$

$$T^{-1}\, \Sigma\, y^*_{t-1}\, u_t \to \omega^2/2\; [K_c(1)^2 - 1] + \lambda \tag{90}$$

Since ρ = 1 + c/T +o(1), it is easy to prove that the OLS estimator of $\hat{\rho}$ in (86) is such that

$$T(\hat{\rho}-1) \to c + (1/2(K_c(1)^2 - 1) + \lambda/\omega^2)(\int K_c(r)^2)^{-1} \tag{91}$$

When the non-centrality parameter $c=0$, $K_c(r) \equiv B(r)$, and we recover the main distributional result of the Dickey-Fuller test statistic. From (91), it can be observed that the effect of near-integration entails a shift in the location as well as in the shape of the limiting distribution, though the convergence rate is identical, i.e $O(T^{-1})$.

REFERENCES

Andrews, D.W.K. (1987) "Least squares regression with integrated or dynamic regressors under weak error assumptions" Econometric Theory, 3, 98-116.

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.

Banerjee, A., Lumsdaine, R. and Stock, J.H. (1989) "Recursive and sequential tests for a unit root: Theory and international evidence". University of Florida (mimeo).

Banerjee, A., Dolado, J. and J. Galbraith (1990) "Recursive and Sequential Tests for unit roots and structural breaks in long annual GNP series" Documento de Trabajo 9010. Banco de España.

Billingsley, P. (1968) Convergence of Probability Measures, New York, Wiley.

Box, G. and G. Jenkins (1976) Time Series Analysis: Forecasting and Control. San Francisco, Holden-Day.

Cochrane, D. and G. Orcutt (1949) "Application of least squares regression to relationships containing autocorrelated error terms" Journal of the American Statistical Association 44, 32-61.

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.

Durlauf, S. and P.C.B. Phillips (1986) "Trends versus random walks in time series analysis" (mimeo, forthcoming in Econometrica).

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

Engle, R. and S. Yoo (1987) "Forecasting and Testing in Cointegrated Systems" Journal of Econometrics 35, 143-159.

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

Gourieroux, C., Maurel, F. and Monfort, A. (1987) "Regression and non-stationarity" Document de Travail 8708. Institut National de la Statistique et des Etudes Economiques.

Granger, C. and Newbold, P. (1974) "Spurious regressions in econometrics". Journal of Econometrics 26, 1045-1066.

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

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

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

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).

Perron, P. (1989) "The great crash, the oil price shock and the unit root hypothesis". Econometrica 57, 1361-1402.

Phillips, P (1986) "Understanding spurious regression in econometrics". Journal of Econometrics 33, 311-340.

Phillips, P. (1987a) "Time series regression with a unit root" Econometrica 55, 207-231.

Phillips, P, (1987b) "Towards a unified asymptotic theory for autoregression" Biometrika 74, 535-547.

Phillips, P. (1988) "Regression theory for near-integrated time series" Econometrica 56, 1021-1044.

Phillips, P. and Durlauf, S. (1986) "Multiple time series with integrated variables" Review of Economic Studies, 53, 473-496.

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

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

Sargan, J.D. and Bhargava, A. (1983) "Testing the residuals from least squares regression for being generated by the gaussian random walk" Econometrica, 51, 153-174.

Sims, C., Stock, J. and Watson, M. (1990) "Inference in linear time-series models with some unit roots". Econometrica 58, 113-145.

Stock, J. (1987) "Asymptotic properties of least-squares estimators of cointegrating vectors" Econometrica 55, 1035-1056.

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

White, H. (1984) Asymptotic theory for econometricians. London Academic Press.

Yule, G. (1926) "Why do we sometimes get nonsense correlations between time series? A study in sampling and the nature of time series" Jounal of the Royal Statistical Society 84, 1-64.

DOCUMENTOS DE TRABAJO (1):

8601 **Agustín Maravall:** Revisions in ARIMA signal extraction.
8602 **Agustín Maravall and 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 economia.
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 and 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.
8702 **Agustín Maravall:** Descomposición de series temporales: especificación, estimación e inferencia (Con una aplicación a la oferta monetaria en España).
8703 **José Viñals y Lorenzo Domingo:** La peseta y el sistema monetario europeo: un modelo de tipo de cambio peseta-marco.
8704 **Gonzalo Gil:** The functions of the Bank of Spain.
8705 **Agustín Maravall:** Descomposición de series temporales, con una aplicación a la oferta monetaria en España: Comentarios y contestación.
8706 **P. L'Hotellerie y J. Viñals:** Tendencias del comercio exterior español. Apéndice estadístico.
8707 **Anindya Banerjee and Juan Dolado:** Tests of the Life Cycle-Permanent Income Hypothesis in the Presence of Random Walks: Asymptotic Theory and Small-Sample Interpretations.
8708 **Juan J. Dolado and Tim Jenkinson:** Cointegration: A survey of recent developments.
8709 **Ignacio Mauleón:** La demanda de dinero reconsiderada.
8801 **Agustín Maravall:** Two papers on arima signal extraction.
8802 **Juan José Camio y José Rodríguez de Pablo:** El consumo de alimentos no elaborados en España: Análisis de la información de Mercasa.
8803 **Agustín Maravall and Daniel Peña:** Missing observations in time series and the «dual» autocorrelation function.
8804 **José Viñals:** El Sistema Monetario Europeo. España y la política macroeconómica. (Publicada una edición en inglés con el mismo número).
8805 **Antoni Espasa:** Métodos cuantitativos y análisis de la coyuntura económica.
8806 **Antoni Espasa:** El perfil de crecimiento de un fenómeno económico.
8807 **Pablo Martín Aceña:** Una estimación de los principales agregados monetarios en España: 1940-1962.
8808 **Rafael Repullo:** Los efectos económicos de los coeficientes bancarios: un análisis teórico.
8901 **M.ª de los Llanos Matea Rosa:** Funciones de transferencia simultáneas del índice de precios al consumo de bienes elaborados no energéticos.
8902 **Juan J. Dolado:** Cointegración: una panorámica.
8903 **Agustín Maravall:** La extracción de señales y el análisis de coyuntura.
8904 **E. Morales, A. Espasa y M. L. Rojo:** Métodos cuantitativos para el análisis de la actividad industrial española. (Publicada una edición en inglés con el mismo número).
9001 **Jesús Albarracín y Concha Artola:** El crecimiento de los salarios y el deslizamiento salarial en el periodo 1981 a 1988.
9002 **Antoni Espasa, Rosa Gómez-Churruca y Javier Jareño:** Un análisis econométrico de los ingresos por turismo en la economía española.
9003 **Antoni Espasa:** Metodologia para realizar el análisis de la coyuntura de un fenómeno económico. (Publicada una edición en inglés con el mismo número).
9004 **Paloma Gómez Pastor y José Luis Pellicer Miret:** Información y documentación de las Comunidades Europeas.

(1) Los Documentos de Trabajo anteriores a 1986 figuran en el catálogo de publicaciones del Banco de España.

Información: Banco de España
Sección de Publicaciones. Negociado de Distribución y Gestión
Teléfono: 338 51 80
Alcalá, 50. 28014 Madrid