TESTS OF THE LIFE CYCLE-PERMANENT INCOME HYPOTHESIS IN THE PRESENCE OF RANDOM WALKS: ASYMPTOTIC THEORY AND SMALL-SAMPLE INTERPRETATIONS

Anindya Banerjee *
Juan Dolado **

* Jesus College. University of Oxford
** Bank of Spain. Research Department

We wish to thank Manuel Arellano, Alvaro Escribano, John Galbraith and David Hendry for their help. Helpful comments by three referees and participants in the Econometrics Workshop (Oxford, 1986), the ESRC Econometrics Conference (York, 1987) and the European Meeting of the Econometric Society (Copenhagen, 1987) are also gratefully acknowledged. Any remaining errors are attributable solely to us.

Banco de España. Servicio de Estudios
Documento de Trabajo n.º 8707

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: 84-505-7103-0
ISSN: 0213 - 2710
Depósito legal: M. 42883 - 1987

Imprenta del Banco de España

«We shall not cease from our exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.»

from «Little Gidding» by
T. S. Eliot,
with apologies to
Adrian Pagan.

Abstract:

The recent literature on cointegration and unit roots has focused attention on the distribution of test statistics frequently used to test efficiency in rational expectations models. In this paper we concentrate on the permanent income hypothesis of real consumption. We illustrate, by using the proper asymptotic theory and small-sample approximations, the cases in which tests of such a hypothesis are biased towards rejection and cases where they have the correct sizes. Our results serve to interpret numerous Monte Carlo studies in the literature on this issue. Special emphasis is placed on the distinction between "weak" and "semi-strong" rationality tests.

1. Introduction.

"…. Optimization on the part of consumers is shown to imply that the marginal utility of consumption evolves according to a random walk with trend. To a reasonable approximation, consumption itself should evolve in the same way. In particular, no variable apart from current consumption should be of any value in predicting future consumption. This implication is tested with time-series data for the post-war United States. It is confirmed for real disposable income... The paper concludes that the evidence supports a modified version of the permanent income hypothesis..."

R. E. Hall: J.P.E 1978

The recent literature on cointegration and unit roots has helped to offer insight to applied econometricians on the special consequences of econometric modelling with variables that are nonstationary or only borderline stationary. The literature, in part, has focussed attention on the distributions of test statistics frequently used in testing efficiency in models where a rational use, by economic agents, of available information is assumed. The test of the random-walk hypothesis of consumption, implied by the rational expectations version of the permanent income hypothesis, is perhaps the most popular application and is therefore the subject of this paper.

In the above context, the following question has been asked, in various forms:

"Do rationality test have the correct sizes if innovations in the explanatory series are correlated with the regressand and the explanatory series are substantially autocorrelated or have a unit root?"

Put slightly differently, the question relates to whether tests of rationality are biased in favour of rejection, if standard critical values are used.

The answers to the question posed have been somewhat mixed and therefore prone to confusion. The seminal work of Dickey and Fuller (1979), inter alia, pointed out the non-standard character of tests for unit roots and the lack of invariance of the critical regions to the existence of incidental parameters, say, a constant or trend, displayed by such tests. Their message has been re-emphasised in recent work by Mankiw and Shapiro (1985, 1986) (henceforth referred to as MS), who present evidence demonstrating the over-rejection of the permanent income hypothesis of consumption in its rational expectations version. This property of over-rejection applies to a restricted form of the test, both in the unit-root and borderline cases. However, emphasising the role of incidental parameters, West (1987) provides theorems on the possibility of achieving asymptotic normality of the test when integrated regressors have drifts. Stock and West (1987) Sims, Stock and West (1987) and Banerjee, Dolado and Galbraith (1987(a)) also provide evidence to support the claim that a detailed consideration of more general data generating processes and models lead to different versions of the tests having the correct sizes.

This paper is an attempt to explain this slightly bewildering catalogue of mixed results. We aim to provide an explanation of the opposing positions, based on which the literature relating to the critical values of rationality tests can be assessed.

Part of this work of assessment appears in Banerjee and Dolado (1987) where MS's findings on the borderline stationary are explained using finite sample expansions. This paper concentrates, for the most part, on the pure unit root case.

The basic point we make in this paper is that the mixed results can be quite easily understood once a distinction, between "weak-efficiency" and "semi-strong efficiency" tests of rationality, is made. The standard test of, say, the permanent income hypothesis of consumption in its rational expectations version consists of checking whether consumption is excessively sensitive to income, once the role of current income in signalling changes in permanent income has been taken into account. This is what we call the simple "weak-efficiency test" due to Flavin (1981) and the intuition for the resulting incorrect inference is clear. If income follows a random walk, the model, in which the test is carried out, is mis-specified since the two sides of the equation do not have the same order of integration.[1] Hence the regression is not consistent and the diagnostics do not, in general, follow the traditional distributions. The exception occurs when both the model and the process generating income have the same incidental parameters, for example, a constant or trend.

However, when testing for consumption being a random walk, another possibility is to use a "semi-strong efficiency test" as in Hall (1978), Davidson and Hendry (1981) and Muellbauer (1983), among others. Here

---

[1] A variable x_t (t=1,...T) is said to be integrated of order d, denoted I(d), if $\Delta^d x_t$ admits a stationary ARMA representation. An NX1 vector of I(d) variables is said to be cointegrated of order b, denoted C(d, b) where d > b, if there exists a set of r linear combinations denoted by α, $r \leq N-1$, such that $\alpha' x_t$ is integrated of order b.

more than one lagged regressor appear as explanatory variables. If the regressors are cointegrated then both sides of the regression equation have the same order of integration and the test statistics follow standard distributions. The inferences based on the ordinary t-tables are now correct. The permanent income hypothesis provides a convenient pair of cointegrated variables namely consumption and disposable income. It is of considerable interest to examine the consequences of extending the tests of rationality to such tests of "semi-strong efficiency".

In sum, this paper has a three-fold purpose. First, drawing upon recent results by Phillips (1986, 1987(a)) on the asymptotic theory for integrated processes, we derive the asymptotic results for the "weak-efficiency test" when income follows a random walk. The main result here is that the statistic used to test over-sensitivity does not follow the standard norml distribution asymptotically. This sheds considerable light on the Monte Carlo results reported by MS. Second, given that the correct distribution has unknown characteristics, we use small-sample approximations, based on Nagar expansions of the continuous normalisation of the bias, to approximate the expected values of these random variables and thereby get a feeling for why the Monte Carlo critical values are so disparate. These approximations are not available in the literature (see White (1961) and references therein for other expansions) and we find them useful in explaining the important role played by non-centrality, relative to other features of the distribution, in the Monte Carlo results. The asymptotic and small-sample theory results in this part of the paper explain the poor performance of weak-efficiency of the "weak-efficiency tests". We also comment on the extension of the previous results to cases where income follows a borderline stationary process and on the role of incidental parameters.

The analysis in the second part of the paper moves on to an explanation of "semi-strong efficiency tests". This is the third purpose of the paper. We analyse the asymptotics of the "semi-strong efficiency test" using again the asymptotic theory for integrated processes. The main result in this case is that the individual coefficients and their t-ratios are normally distributed but an F-test of their joint significance does not follow a standard distribution. Monte Carlo simulations are used to demonstrate the theoretical proposition.

The paper is organised as follows. Section 2 presents a canonical model, used by MS, of the permanent income hypothesis in order to discuss the properties of the tests. In the light of this canonical model, several combinations of the data generation process (DGP) and models related to the simple weak-efficiency tests are analysed. Section 3 provides a set of asymptotic theory results for the weak-efficiency test where income follows a random walk. Section 4 demonstrates the corresponding small-sample results obtained through Nagar expansions and uses these approximations to compute critical values and explain the MS results for unit roots. Section 5 contains the asymptotic theory results for the semi-strong efficiency tests. The analysis is linked with the work of Stock, West and Sims (1987) and Banerjee, Dolado and Galbraith (1987(a)). Features of results appearing in the above papers are explained. Particular attention is focused on the misbehaviour of the F-test of joint significance of the regression coefficients and an explanation is proposed in terms of the invariance of these models to linear transformations. This is a surprising result which has as its background the asymptotic singularity of the variance-covariance matrix for cointegrated processes (see Hendry (1986)). Section 6 concludes the paper. An appendix contains some of the algebra of the Nagar expansions.

2. A Canonical Model of the Permanent Income Hypothesis.

We begin by reviewing and revising the canonical model of the permanent income hypothesis (PIH) used by MS in order to illustrate the subsequent form of the tests. The model, in its simplest form, entails constancy of the interest rate and no taste shifts or transitory consumption.

Suppose that consumption C_t is set according to the PIH:

$$C_t = (1-\gamma)[W_t + H_t] \qquad (1)$$

where,

r = real interest rate

$\gamma = (1 + r)^{-1}$

W_t = Non-human wealth at the end of period t

H_t = Human wealth

Thus,

$$H_t = \sum_{s=0}^{\infty} \gamma^s \cdot \underset{t}{E}\, Y_{t+s} \qquad (2)$$

Y_t is real labour income received in period t.

Non-human wealth evolves as,

$$W_{t+1} = \gamma^{-1}[W_t + Y_t - C_t] \qquad (3)$$

Suppose also that Y_t follows an autoregressive (possibly infinite) process with drift:

$$\lambda(L)Y_t = \mu + \varepsilon_t \qquad (4)$$

where ε_t is white-noise with variance σ^2.

Then, substituting (1) in (3), we obtain,

$$\Delta W_{t+1} = \gamma^{-1}[Y_t - (1-\gamma)H_t] \qquad (5)$$

Differencing (1) at (t+1) and substituting (5) back into the corresponding expression yields,

$$\Delta C_{t+1} = (1-\gamma)[H_{t+1} - \gamma^{-1}(H_t - Y_t)] \qquad (6)$$

In order to obtain the expression in square brackets in (6) we make use of the Wiener-Kolmogorov prediction formula which yields,

$$H_t = \sum_{s=0}^{\infty} \underset{t}{E}\, \gamma^s Y_{t+s} = \frac{[1-\gamma L^{-1}\lambda(L)\lambda(\gamma)^{-1}]}{[1-\gamma L^{-1}]} \cdot Y_t + \frac{\gamma\mu}{[1-\gamma][\lambda(\gamma)]} \qquad (7)$$

where,

$$L^{-r}Y_t = \underset{t}{E}\, Y_{t+r}$$

Applying the corresponding Koyck transformation to both sides of (7) we obtain,

$$H_t - \gamma \cdot \underset{t}{E}\, H_{t+1} = Y_t - \gamma\lambda(\gamma)^{-1}L^{-1}[\mu + \varepsilon_t] + \gamma\mu\lambda(\gamma)^{-1} \qquad (8)$$

or,

$$[H_{t+1} - \gamma^{-1}(H_t - Y_t)] = H_{t+1} - \underset{t}{E}\, H_{t+1} = \lambda(\gamma)^{-1}\varepsilon_{t+1} \qquad (9)$$

since $L^{-1}\varepsilon_t = 0$

Substituting (9) back into (6), we obtain the process which governs the evolution of C_t, that is,

$$\Delta C_t = (1-\gamma)\lambda(\gamma)^{-1}\varepsilon_t \qquad (10)$$

If income follows a random walk (with or without drift), then

$$\lambda(L) = (1-L)$$

and consequently,

$$\Delta C_t = \varepsilon_t \qquad (11)$$

that is, permanent income equals current income and also consumption. Notice however that if there is a drift in the income process, still the consumption process will be driftless. We elaborate on this feature in Section 5.

Consider now the traditional manner, developed by Flavin (1981), in which excess sensitivity is measured. Basically, it consists of specifying a structural equation relating the change in consumption to the contemporaneous revision in permanent income as in (10) and the current change in income. The coefficient of the change in income is a measure of the behavioural marginal propensity to consume out of current income, since the role of current income in signalling changes in permanent income has been explicitly modelled. The PIH with rational expectations can then be tested by testing whether the marginal propensity to consume is significantly different from zero. In terms of (4) and (10) in the AR(1) case, the just identified enlarged model is the following:

$$Y_t = \mu + \lambda Y_{t-1} + \varepsilon_t \qquad (12)$$

$$\Delta C_t = (1-\gamma)(1-\lambda\gamma)^{-1}\varepsilon_t + \beta\Delta Y_t$$

$$= \beta\mu + \beta(\lambda-1)Y_{t-1} + (\beta + (1-\gamma)(1-\lambda\gamma)^{-1})\varepsilon_t \qquad (13)$$

which in an unrestricted reduced form can be re-written as,

$$\Delta C_t = \alpha + \pi Y_{t-1} + u_t \qquad (14)$$

The test for excess sensitivity is thus just the t-test for $\pi = 0$ in (14). Note that if $|\lambda| < 1$, both sides of (14) are stationary and, asymptotically, the t-ratio of Y_{t-1} will follow a standardized normal distribution. Nevertheless, the asymptotic distribution might be a bad approximation for finite sample sizes as MS claim. We shall come on to discuss this claim in the next section.

In the realistic case in which income follows a random walk, say, with drift, then it can be expressed as,

$$Y_t = \mu t + Y_t^* \qquad (15)$$

where

$$Y_t^* = \sum_{s=1}^{t} \varepsilon_s \text{ and is a driftless random walk. } Y_0^* = 0$$

Inserting (15) in (14) and augmenting with a trend, we get,

$$\Delta C_t = \alpha + \beta t + \pi[\mu(t-1) + Y_{t-1}^*] + u_t$$

$$= (\alpha - \pi\mu) + (\beta + \pi\mu)t + \pi Y_{t-1}^* + u_t$$

$$= k + dt + \pi Y_{t-1}^* + u_t \qquad (16)$$

This equation mimics the standard practice, followed in the literature, of detrending the data (Flavin (1981)) in the belief that this procedure eliminates non-stationary features of the regressors and then carrying out the test in model (14). By the Frisch-Waugh Theorem this procedure is equivalent to running the regression equation (16).

Once the trend is included in the model, the results are invariant to the true value of μ. We therefore set $Y_{t-1}{}^* = Y_{t-1}$ in (16). Similarly, since the results on the t-statistic of π derived in the paper are invariant to the variances of the innovations in consumption and income, we normalise them to unity. Also, to allow for the possibility of transitory consumption or capital gains in (3), we adopt the simplifying procedure of allowing for imperfect correlation between the innovations of income and consumption.[2]

In order to simplify interpretation of the results derived in Sections 3, 4 and 5 of the paper, we re-write a canonical version of both the model in which the test is carried out and the data generation process (DGP). According to the permanent income hypothesis, consumption and income are generated as follows:

DGP **MODEL**

$$Y_t = \lambda Y_{t-1} + \varepsilon_t \quad (17) \qquad \Delta C_t = k + dt + \pi Y_{t-1} + u_t \quad (18)$$

$$\Delta C_t = e_t$$

$$E(\varepsilon_t e_t) = \delta_{ts}\rho$$

$$\sigma^2_e = \sigma^2_\varepsilon = 1$$

This DGP and model constitute the object of the Monte-Carlo study undertaken by MS (1985, 1986). They study the behaviour of the t-statistic of $\hat{\pi}$ when $\lambda = 1$ and when $\lambda < 1$ but barely below unity (borderline cases). Their main result is that the t-statistic of the coefficient of Y_{t-1} is biased towards rejection if reliance is placed on

---

[2] The likely existence of stationary autocorrelation in e_t is disregarded on the grounds that it does not alter any of the main implications of the theorems. However, in the presence of such autocorrelation the empirical implementation of the tests requires the use of non-parametric conditions as suggested by Phillips (1987(a)).

the standard asymptotic distribution. The bias is particularly severe when the data are detrended but remains even if the incidental parameters (k and/or d) do not appear in the model. Note that the process generating Y_t has been assumed driftless, that is $\mu = 0$.

The results, although ignored for some time in the empirical macro-econometric literature, are not surprising. The intuition runs as follows. When $\lambda = 1$ in (17), it is immediately noticeable that the LHS of (18) is stationary whereas Y_{t-1} is non-stationary. The non-stationarity of the income process has profound implications irrespective of the presence or absence of a trend. The left-hand and the right-hand sides of equation (18) have different orders of integration. It is in this sense that the model is mis-specified and hence "inconsistent". When $\rho = 1$, the test corresponds to testing for a unit root, a case where non-standard distributional features were shown to be present, analytically by White (1959) and through simulation analyses by Dickey (1975). When λ belongs to the borderline region, the finite sample distribution of the t-statistic is known to be a smooth function of λ and again the MS results are not particularly surprising.

The next section tackles these statistical issues, concentrating on the unit-root case. We compute the correct asymptotic distribution of estimators and test statistics in models of the type discussed above, emphasising their difference from the standard distributions. We also discuss the cases in which it is valid to use the ordinary critical values.

3. Asymptotic Theory Results.

This section, drawing on recent results by Phillips (1986, 1987(a)) and Phillips and Durlauf (1985), starts by providing the correct asymptotic statistical theory for the combination of the DGP and model specified in (18) with $\lambda = 1$. This enables us to interpret the results corresponding to the non-stationary case derived from Monte Carlo experiments. The principal distinguishing feature of integrated processes is that, suitably normalized, moments of the series do not converge to constants but to random variables.

Suppose that $\{x_t\}$ is a stochastic process generated by a random walk.

$$x_t = x_{t-1} + \varepsilon_t \qquad (19)$$

where $\varepsilon_t \sim$ n.i.d $(0, \sigma_\varepsilon^2)$

Alternatively, as in (15), x_t can be represented in terms of the partial sums S_t of the innovation sequence $\{\varepsilon_t\}$ and the initial condition, where,

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

and,

$$x_t = S_t + x_o \qquad (20)$$

We may define $S_o = 0$ and set the initial condition with probability one. The distributional results of this section will use the following standardized sums,

$$X_T(t) = (T^{-1/2}\sigma_\varepsilon^{-1})S_{[Tt]} = (T^{-1/2}\sigma_\varepsilon^{-1})S_{j-1} \qquad (21)$$

$$j\text{-}1/T \leq t < j/T, \quad (j=1,2, \ldots T)$$

$X_T(1) = (T^{-1/2}\sigma_\varepsilon^{-1})S_T$

where [b] denotes the integer part of b. $X_T(t)$ is a random element in the function space D[0,1], that is the space of all real-valued functions on [0,1] which are right-continuous at each point of [0,1] and have finite left limits. Under the conditions mentioned above, $X_T(t)$ can be shown to converge weakly to a limit process known as a Wiener process, which is denoted by W(t).

Symbolically,

$$X_T(t) \rightarrow W(t) \text{ as } T \rightarrow \infty \qquad (22)$$

where W(t) lies in C[0,1], the space of real-valued, continuous functions on [0,1]. W(t) is N(0, t) for fixed t and has independent increments. Moreover, an extension of the Slutsky theorem in conventional asymptotic theory also applies in this framework, in the sense that if g(.) is any continuous function on C[0,1], then $X_T(t) \rightarrow W(t)$ implies that

$$g[X_T(t)] \rightarrow g(W(t)) \qquad (23)$$

Now consider the pairs of random walks specified in (17) for the λ=1 case.

$$C_t = P_t = \sum_{i=1}^{t} e_i \qquad (24a)$$

$$Y_t = S_t = \sum_{i=1}^{t} \varepsilon_i \qquad (24b)$$

where the standardized sums of P_t and S_t, as defined in (20), converge to the Wiener processes V(t) and W(t) respectively.

In order to prove the main result of this section, the following set of lemmas due to Phillips (1986, 1987(a)) will be used extensively.

Lemmae: If the sequences $\{C_t\}$ and $\{Y_t\}$ are generated under the assumptions set in (17), then as $T \to \infty$,

(a) $$T^{-3/2} \sum_{1}^{T} Y_t \to \int_0^1 W(t)dt$$

(b) $$T^{-2} \sum_{1}^{T} Y_t^2 \to \int_0^1 W(t)^2 dt$$

(c) $$T^{-1} \sum_{1}^{T} Y_{t-1}e_t \to \frac{\rho}{2} [W(1)^2 -1] + N(0, \frac{1-\rho^2}{2})$$

(d) $$T^{-5/2} \sum_{1}^{T} tY_t \to \int_0^1 tW(t)dt$$

We now state the theorem that characterizes the asymptotic behaviour of the regression coefficients and diagnostics for the DGP-MODEL combination given by (17)-(18). The proof of the theorem is available on request from the authors .

Theorem 1: For the Model in (18), if (17) characterizes the DGP, then,

(I) $$T\hat{\pi} \to \frac{(1/12)A - BC}{(1/12)D - B^2}$$

(II) $$T^{3/2}\hat{d} \to \frac{CD - AB}{(1/12)D - B^2}$$

(III) $$T^{1/2}\hat{k} \to V(1) - 0.5T^{3/2}\hat{d} - T\hat{\pi} \cdot \int_0^1 W(t)dt$$

(IV) $t_{\pi=0} \rightarrow (12)^{1/2}(T\hat{\pi})((1/12)D - B^2)^{1/2}$

(V) $DW \rightarrow 2$

(VI) $R^2 \rightarrow 0$

where,

$$A = \frac{\rho(W(1)^2 - 1)}{2} + N(0, \frac{1-\rho^2}{2}) - \rho W(1)\int_0^1 W(t)dt$$

$$B = \int_0^1 tW(t)dt - 0.5\int_0^1 W(t)dt$$

$$C = \int_0^1 V(t)dt - 0.5V(1)$$

$$D = \int_0^1 W(t)^2dt - (\int_0^1 W(t)dt)^2$$

The theorem emphasises the importance, for determining coefficient consistency, of orders of magnitude of sampling variability. For the regressand, the sample moment $\Sigma\,(\Delta C_t^2)/T$ is $O_p(1)$. For the intercept vector, the corresponding moment is $O_p(1)$, whereas that of the trend is $O_p(T^2)$ and that of Y_{t-1} is $O_p(T)$. When the regressor sample variance is of higher order of magnitude than the dependent variable, the regression coefficients are consistent, converging to their true values at a higher speed than usual, as it is the case with the trend and Y_{t-1}. The value of the DW reflects the stationarity of the regressand while the converging value of the R^2 reflects the misspecification of the model.

The most striking result for our purpose is that reflected in part (IV) of the theorem, which relates to hypothesis testing. As discussed in the previous section, it shows that the statistic which tests the over-sensitivity of consumption to current income does not converge to a N(0, 1) distribution, as would be the case were the process stationary. The size of the test will be distorted if the ordinary N(0, 1) table is used. This is shown by MS in their papers.

It is also interesting to note that if $\mu = 0$ in the DGP and no trend or constant is included in the model, that is $k = d = 0$, then the t-statistics converges to,

$$\text{(i)}\quad t_{\pi=0} \rightarrow \tilde{T\pi}\left[\int_0^1 W(t)^2 dt\right]^{1/2}$$

If $k \neq 0$ and $d = 0$ in (18),

$$\text{(ii)}\quad t_{\pi=0} \rightarrow \tilde{T\pi} D^{1/2}$$

Hence the inclusion of the time-trend and the constant have only qualitative effects on the asymptotic distribution, explaining why the Monte Carlo results of MS show over-rejection in these instances.

However, it is important to point out that the previous analysis only goes through under the assumption that the number of incidental parameters in the model is always larger than in the DGP, except when there are no such parameters in either the model or the DGP.
West (1987) discussed cases in which the correct distribution of the test is again the standard asymptotic distribution. His analysis can be easily extended to our case by considering the following slight

variations in the DGP-Model combinations discussed in Section 2.

DGP		MODEL	
(1) $Y_t = \mu + \lambda Y_{t-1} + \varepsilon_t$	(27)	$\Delta C_t = k + \pi Y_{t-1} + u_t$	(28)
(2) $Y_t = \mu + \beta t + \lambda Y_{t-1} + \varepsilon_t$		$\Delta C_t = k + dt + \pi Y_{t-1} + u_t$	

The rest of the DGP is unchanged. In combination (1), when $\lambda = 1$, it is easy to show that using the normalisation matrix **diag** $(T^{1/2}, T^{3/2})$, the joint distribution of $(\hat{k} - \mu)$ and $\hat{\pi}$ is asymptotically normal.

In particular,

$$T^{3/2}\hat{\pi} \rightarrow N(0,\ 12/\mu^2) \text{ and}$$

$$t_\pi \rightarrow N(0,\ 1).$$

Similar results obtain, when $\lambda = 1$, in combination (2) [3], using the normalisation matrix **diag** $(T^{1/2}, T^{3/2}, T^{5/2})$, for $(\hat{k} - \mu)$, $(\hat{d} - \beta)$ and $\hat{\pi}$.

Here,

$$T^{5/2}\hat{\pi} \rightarrow N(0,\ 180/\beta^2) \text{ and}$$

$$t_\pi \rightarrow N(0,\ 1).$$

The intuition for these results is again clear. Take the case where there is only a constant in the DGP and model. Income depends therefore on a deterministic trend and a stochastic one. The sample variability of the deterministic trend is of $O_p(T^{3/2})$ which dominates the role of the sample variability of the stochastic trend which is of $O_p(T^2)$. But we know that the existence of a deterministic trend in a regression model does not affect the asymptotic normality of the standardised estimates, hence normality follows. Similar considerations

[3] This is however a rather unlikely case since trend plus unit root implies an ever increasing (decreasing) rate of growth of the variable.

apply to the constant plus trend case. The message is therefore clear. If we knew the incidental parameters in the process generating income, we could replicate them in the model, using the standard distributions to carry out the proposed tests. A possible strategy to test for the presence of these incidental parameters in the income process is given in Dolado and Jenkinson (1987) where use is made of the critical values contained in Table 8.5.2 of Fuller (1976).

It is necessary to comment briefly on the borderline case. When $\lambda < 1$, Anderson (1959), for example, showed that $T^{1/2}\pi$ has a normal distribution, emphasising the sharp discontinuity in the limiting distribution when $\lambda = 1$. However, for finite samples, the distibutions of the estimator and t-statistic are smooth functions of λ which do not exhibit discontinuities at $\lambda = 1$. The discontinuities are only the result of the limiting process. Cavanaugh (1986) and Phillips (1987(b)) have elaborated on this point by embedding the previous results in a more general framework. They consider a sequence of local alternatives for $\lambda = 1$, generated by,

$$\lambda = \exp(T^{-1}\delta), \quad \delta < 0 \tag{29}$$

where the rate of convergence to unity is controlled at $O(T^{-1})$, as shown in part (I) of Theorem 1. Defining, as in (21), a sequence of partial sums $X_T(t)$ where now,

$$S_t = \sum_{1}^{t-1} \lambda^j \varepsilon_{t-j} \tag{30}$$

A simple application of the Lindeberg-Levy Central Limit Theorem shows that this random variable is asymptotically distributed as $N(0, (1-\exp(2\delta t))/2\delta)$ for fixed t. Therefore it is necessary to find a limiting process, similar to W(t) in (22), which for fixed t has such a distribution. Phillips (1987(b), p.538, formula (7)) proves that such a process, known as a diffusion process, satisfies the following relationship:

$$D_{\delta}(t) = W(t) + \delta \int_{0}^{1} \exp(t-s)W(s)ds \qquad (31)$$

Hence, when $\delta = 0$, the diffusion process reduces to the Wiener process. Similar distributional results to those contained in Lemma 1 apply in this case substituting $W(t)$ by $D_{\delta}(t)$.

Under the sequence of local alternatives defined in (27) we have that $\lambda = \exp(T^{-1}\delta) \simeq 1 + T^{-1}\delta$. This allows the formulation of non-central asymptotic theory for this case. In particular, the following re-formulation of the coefficient and the t-statistic in parts (I) and (IV) of Theorem 1 reflect the likely lack of power of the test in such an instance.

$$T\hat{\pi} \rightarrow \beta\delta + \frac{(1/12)A' - B'C}{(1/12)D' - B'^{2}} \qquad (32)$$

$$t_{\pi=0} \rightarrow \beta\delta((1/12)D' - B')^{1/2} + \frac{(1/12)A' - B'C'}{[(1/12)D' - B'^{2}]^{1/2}} \qquad (33)$$

where A', B' and D' are expressions like A, B and D in Theorem 1 where $W(t)$ is substituted by $D_{\delta}(t)$.

The results in this section have important implications when interpreting the statistical meaning of MS's findings. We confirm that the presence of integrated or nearly integrated series in regression models like (18) can seriously damage the associated diagnostic tests. However, at a purely analytical level, very little can be said about the numerical values towards which the functional random variables converge. The only alternative is to simulate, by Monte Carlo, the actual distribution. The next section of this paper is devoted to overcoming, in part, this difficulty by using a finite sample analytical approximation to the central values of the distribution.

4. Nagar Expansions.

This section makes use of finite sample approximations to the t-ratio of the income coefficient in model (18). We have chosen, for illustrative purposes, a version of the model where the trend is absent. There is no loss in generality involved as the inclusion of a time trend has only qualitative effects on the asymptotic distribution when the process generating income has no drift. This has already been shown in expressions (25) - (26) of this paper. The approximation, defined below, is a Nagar expansion of the continuous normalisation of the OLS estimate of π in (18) (see Evans and Savin (1984)). The aim is to shed light on the pure Monte Carlo results and to emphasise the non-standard features, in finite samples, of the distributional results derived in the previous section. We focus on the unit root case, that is $\lambda = 1$, since similar approximations for the borderline case have been used successfully elsewhere (see Banerjee and Dolado (1987)). The key result in our paper on borderline stationarity was that the non-centrality of the distribution was a much more important feature than its skewness or kurtosis. This was especially true at the lower tail of the distribution, which is the one used in empirical implementations of the permanent income hypothesis of consumption.

In order to define a Nagar expansion, consider the following regression model where the variables represent deviations from the means or any other projection over a larger set of variables:

$$Y_t = \hat{\theta} x_t + \hat{u}_t \tag{34}$$

where,

$$\hat{\theta} = (\sum x_t y_t)(\sum x_t^2)^{-1} = H/F \tag{35}$$

Then,

$$E(\hat{\theta}) = E\left[\frac{H}{E(F)[1-(1-(F/E(F)))]}\right]$$

$$\simeq E\left[\frac{H}{E(F)} \cdot \left[2 - \frac{F}{E(F)}\right]\right] \tag{36}$$

since $(1+x)^m = 1 + mx + O(m(m-1)x^2)$ for small values of x. The error in the approximation has an order of magnitude of $[(F-E(F))^2 H]/(E(F))^2$ which can be shown to be $O_p(T^{-1})$. In order to simplify the approximation of the actual t-ratio, we use the continuous normalization of the bias. This is obtained by computing the ratio of (36) to $[E(s^2F^{-1})]^{1/2}$. s^2 is the residual variance given by $\hat{u}'\hat{u}/(T-2)$. The expansion proposed is of $O_p(T^{-5/2})$ and is currently not available in the literature.

Next, using lengthy but straightforward calculations with fourth moment products of random variables, Theorem 2 and its corollary follow.

Theorem 2: For the DGP-MODEL combination described in (17) and (18), where d = 0, and $\lambda = 1$, the Nagar expansion of the t-ratio of the π coefficient is given by,

$$E(t_{\pi=0}) = [E(1) - E(2) + E(3)]/E(4)$$

where,

$$E(1) = -(2\mathbf{i}'\mathbf{H}\Omega\ \mathbf{i})/\zeta^2(T-1)^2$$

$$E(2) = -(T-1)^2\zeta^{-4}[\mathbf{g} - 2(T-1)^{-1}\mathbf{i}'\mathbf{H}\Omega\ \mathbf{i}]$$

$$E(3) = (T-1)^{-1}\zeta^{-4}[(T-1)^{-2}(\mathrm{trace}(\Omega)\mathbf{i}'\mathbf{H}\ \mathbf{i} + 2\mathbf{i}'\Omega\mathbf{H}\ \mathbf{i}) - 3(T-1)^{-3}(\mathbf{i}'\mathbf{H}\ \mathbf{i})(\mathbf{i}'\Omega\ \mathbf{i})]$$

$$E(4) = [(T-1)\zeta^2]^{-1/2}$$

$$\zeta^2 = (T)(T-1)/2 - (\mathbf{i}'\Omega\ \mathbf{i})(T-1)^{-2}$$

where,

$$\mathbf{H} = E(\mathbf{Y}_{-1}\boldsymbol{\varepsilon}')$$

$$\Omega = E(\mathbf{Y}_{-1}\mathbf{Y}_{-1}')$$

$$\mathbf{g} = E(\boldsymbol{\varepsilon}'\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\mathbf{Y}_{-1})$$

and

$$\mathbf{Y}_{-1} = (Y_1, Y_2, \ldots, Y_{T-1}),\quad \boldsymbol{\varepsilon} = (\varepsilon_2, \varepsilon_3, \ldots \varepsilon_T).$$

Corollary: If both k = d = 0, the expansion reads as follows:

$$E(t_{\pi=0}) = E(5)/E(6)$$

where,

$$E(5) = -(T-1)^{-2}\sigma_Y^{-4}\cdot\mathbf{g}$$

$$E(6) = [(T-1)\sigma_Y^2]^{-1/2}$$

$$\sigma_Y^2 = (T(T-1))/2.$$

We use the exact expressions of the quadratic forms in Theorem 2 to generate the expected values of the diagnostic tests. The expressions are listed in the appendix and the proofs are avilable upon request. The central 'pseudo t-statistics' are computed for values in a subset of the parameter space ($T \times \lambda \times \rho$) considered by MS (1986) where T = {50, 200}, λ = 1 and ρ = {1.0, 0.8, 0.5}. The simulation study of MS (1985) for the unit root case contains a trend. A companion paper (see Banerjee, Dolado and Galbraith (1987(b))) deals with detrended data but for simplicity we have discarded the trend term here. Hence, in order to check the accuracy of our analytical approximation using simulation data, we carried out a Monte Carlo study with 500 replications for the above parameter space. The initial values Y_0 and C_0 are set equal to zero. The simulations allow us to estimate certain desired percentiles of the distribution of the statistic. In particular, we compute the true critical values required to carry out the test for significance levels α of one, five and ten per cent. Following MS, these critical values are defined as two-tailed, so that if τ is the test statistic, the α critical value is the number ψ such that under the null hypothesis Prob $(|\tau| \geq \psi) = \alpha$.

The results are contained in Tables 1A and 1B. The topmost entries in each box show the computed central values using the Nagar expansion. In all cases we observe that the expected t-ratios are centered around negative values. As ρ decreases, the expected values converge to the correct ones. The computed expressions are proportional to ρ. This reduces much of the computational burden as, for example, the computed central values for ρ = 0.5 are simply those for ρ = 1 multiplied by two. The results on the expected t-values alone indicate, quite clearly, that there is a strong bias towards over-rejection of the null if conventional asymptotics are used, since the latter distributions are centered around zero. There are other non-standard features of the distribution, like skewness and kurtosis, which cannot be analysed without carrying out a complete simulation analysis to tabulate the actual distribution or engaging in even messier algebra to compute the Nagar expansions of higher order sample moments. However, the interesting issue is whether we can say anything about the deviations from asymptotic normality of the small-sample distributions just by using information on the central values of the t-statistics. The middle entries of each box contain the Monte Carlo critical values for the three significance levels chosen. Note that for ρ = 1 the test is equivalent to testing for a unit root in a univariate framework and the critical values should be similar to those in Table 8.5.2 of Fuller (1976) for the distribution of the test statistic τ_{μ}. The critical values from our Monte Carlo study are based on a smaller number of replications but are very close to the Fuller values. This lends credibility to our small scale study. In order to complete the comparison between the numerical and simulation critical values we need to work out a set of critical values from our computations of the expected t-ratios. The critical values correspond to two-tailed symmetric cut-off points but, as properly noted by MS (1986, footnote 3), these are not equivalent to two-tailed critical regions. We saw that the distribution is not symmetric around the origin. Indeed, given that the central values are so negative, it would not be too incorrect to infer that all the rejections occur in the lower tail of the distribution; that is, for large negative values of the t- ratio the test is not far from being a one-tailed test, at least for the significance levels considered in the study. This intuition is confirmed in the Monte Carlo rejection distribution since the number of rejections in the upper tail is negligibly small. We elaborated upon this idea, for the borderline cases, in Banerjee and Dolado (1987), suggesting a mapping from the central computed t-statistics to the corresponding critical values. We extend this to the unit root case in thos paper. The approximation works by adding the one-tailed α per cent critical values, corresponding to the Student's t-distribution, to the central values obtained by the Nagar expansions. This emphasises the important role of non-centrality relative to the other features of the distribution. The use of the t-distribution instead of the asymptotic standardized normal stems from a small finite sample correction factor.[4] The results are tabulated in the bottom entries of Tables 1A and 1B. As in the borderline case, when we compare our critical values with those obtained in the Monte Carlo study we observe a close similarity, especially at one and five per cent significance levels where the likelihood of rejections in the upper tail is minimal. This shows that for larger significance levels the approximation does not work so well since the neglected rejections in the upper tail tend to occur more frequently. For example, at α = 0.95, the Monte Carlo critical values

[4] The critical values of the t-distribution for the two sample sizes are as follow:

-2.41(1%)	-1.68(5%)	-1.30(10%)	Degrees of freedom = 47
-2.33	-1.65	-1.28	Degrees of freedom = 197

for ρ = 1 are -0.01 and -0.04 (for T=50, 200 respectively) whereas the corresponding approximations are 0.47 and 0.49. We do not claim that the complete finite sample distribution can be recovered simply by shifting the asymptotic standard distribution. Rather, the claim is that for the lower tail tail of the distribution, which is the one normally used when carrying out the test, the shifting mapping works extremely accurately. We thereby offer a numerical explanation of the over-rejection problem as put forward by MS in the simple weak-efficiency version of the excess-sensitivity test.

5. Semi-strong Efficiency Tests.

In this final section of the paper, we extend the analysis to the case where the set of regressors in the regression model (18) is expanded. After all, a test of rationality of the kind considered in the paper is a test of the orthogonality of an expectational error (e_t) with respect to the information set. In the simple version of the weak efficiency test only the first lag of income is considered. This makes such tests rather restricted versions of the test considered by Hall (1978) who used lags of income, consumption and stock prices in the extended set of regressors. Similar considerations apply, among the most well known applications of the "semi-strong efficiency forms" of the rationality test, to the papers by Davidson and Hendry (1981) and Muellbauer (1983). This extension is important because by augmenting the size of the set of elements in the information set, we allow for the possibility of consistency of the orders of integration on both sides of the regression equation. That is, if the regressors, each individually

I(1), were cointegrated C(1,0) both the RHS and LHS of the equation would have the same order of integration, I(0).

It is interesting to note that the permanent income hypothesis of consumption provides a convenient pair of convenient variables, consumption and disposable income. To verify this property, first pointed out by Campbell (1987), let us start by defining savings S_t in period t as,

$$S_t = Y_t^d - C_t = Y_t + (1-\gamma)W_t - C_t \tag{37}$$

where Y_t^d is disposable income defined as the sum of labour income and the annuity on non-human wealth. Substituting (1) into (37) and making use of the definition of H_t, yields,

$$S_t = -(\gamma \cdot \underset{t}{E}\, \Delta Y_{t+1})(1-\gamma L^{-1})^{-1} = -\gamma \cdot \sum_{s=0}^{\infty} \gamma^s \underset{t}{E}\, \Delta Y_{t+s+1} \tag{38}$$

whereby saving is the discounted present value of expected future declines in income. This underlies the notion of "saving for a rainy day" in Campbell's nomenclature. The importance of (38) derives from the fact that even if labour income has a unit root S_t will be stationary, with or without drift, depending on the incidental parameters appearing in the income process. In fact if the process is like (12), then saving is identically zero given that labour income and disposable income are identical. In general, this will not be the case and if, say, labour income follows an ARIMA(1, 1, 0) process, as it seems to follow with US data (see Campbell and Deaton (1987)), then

$$s_t = \left[\frac{\lambda\gamma(1-\lambda L)^{-1}\varepsilon_t}{(1-\lambda\gamma)} - \frac{\gamma\mu}{(1-\lambda)(1-\gamma)} \right] \qquad (39)$$

which is I(0), apart from a constant. Therefore disposable income and consumption are cointegrated with cointegrating vector (1, -1). hence if lags of consumption and disposable income are used as regressors, the RHS of the model will be I(0). To analyse the implications of such an important property for undertaking "semi-strong rationality tests" we use the following simple variation of the previous DGP-MODEL combination.

DGP **MODEL**

$$Y_t = C_t + s_t \qquad \Delta C_t = k + dt + \pi_1 C_{t-1} + \pi_2 Y_{t-1} + u_t \quad (41)$$

$$\Delta C_t = e_t \qquad (40)$$

$$E(s_t, e_s) = \delta_{ts}\rho$$

$$\sigma^2_s = \sigma^2_e = 1$$

The disturbance term s_t plays the role of saving. Note that although s_t will, in general, be serially correlated, we have assumed it to be white noise. One can justify this assumption on the familiar grounds that more general structures on the error term will have only qualitative effects on the asymptotic distributions derived below without altering any of the main conclusions (see the conditions stated by Phillips (1987a) for the disturbance sequences). Similarly we have used a one-lag version of the "semi-strong efficiency test", without loss of generality. Y_t now refers to disposable income and not labour income.

As in Section 3, the fact that a trend is included in the model makes the presence of constants in the DGP irrelevant.

We now state the theorem that characterises the asymptotic behaviour of the diagnostic tests of the null of the permanent income hypothesis of consumption, either jointly [H_0: $\pi_1 = \pi_2 = 0$] or separately [H_0': $\pi_1 = 0$; H_0'': $\pi_2 = 0$]. The proofs are available upon request.

Theorem 3: For Model (41), if (40) characterises the DGP, then the Wald tests of the hypothesis H_0: $\pi_1 = \pi_2 = 0$, H_0': $\pi_1 = 0$, H_0'': $\pi_2 = 0$ are asymptotically distributed as follows:

(I') $F_{\pi_1 = \pi_2 = 0} \to 1 + (1/12)((1/12)G - F)^{-1} \cdot E$

(II') $t_{\pi_1 = 0} \to N(0, 1)$

(III') $t_{\pi_2 = 0} \to N(0, 1)$

where

$$E = (1/2) \cdot [V(1)^2 - 1] - V(1) \cdot \int_0^1 V(t)dt$$

$$G = \int_0^1 V(t)^2 \, dt - (\int_0^1 V(t)dt)^2$$

$$F = \int_0^1 tV(t)dt - (1/2) \cdot \int_0^1 V(t)dt$$

The theorem points out that the individual t-ratios are normally distributed but that the Wald test for the joint null hypothesis is a functional of a Wiener process and therefore its distribution is non-standard.[5] It is also easy to prove that if the model is reparameterised as,

$$\Delta C_t = k + dt + \gamma_1 C_{t-1} + \gamma_2 s_{t-1} + u_t \qquad (42)$$

where $\gamma_1 = \pi_1 + \pi_2$, $\gamma_2 = \pi_2$, then $t_{\gamma 1=0}$ is again a functional of a Wiener process and $t_{\gamma 2=0}$ is asymptotically normally distributed. The Wald test of H_0: $\gamma_1 = \gamma_2 = 0$ is, naturally, non-standard. Similar results obtain when the model is parameterised in terms of s_{t-1} and Y_{t-1}. The parameterisation in (42) has been used by Stock and West(1987) to illustrate by Monte Carlo simulations the correct rejection frequency for the t-statistic of the error correction term s_{t-1} Using West's (1986) results, as discussed in the previous section, if only a constant is included in the model and consumption has a drift, then the t-ratio of the coefficient of C_{t-1} will also be normally distributed. However, as we mentioned in Section 2, in the simplified version of the permanent income hypothesis of consumption $E(\Delta c_t) = 0$ and under the null hypothesis the assumption of a random walk with drift does not stand up well to scrutiny. Nevertheless it does remain valid for income in the alternative parameterisation with Y_{t-1} and s_{t-1}.

It is important to remark that the parameterisation chosen in Theorem 3 points out an interesting and unnoticed result: when testing for a random walk the individual t-statistics of the coefficients of the levels of cointegrated variables follow standard asymptotic distributions, even if their joint distribution is non-standard.

---

[5] It can also be proved that $T^{1/2}\hat{k}$ and $T^{3/2}\hat{d}$ converge to functionals of the Wiener process as in Theorem 1.

The previous results also extend easily to less restricted versions of the weak efficiency version of the rationality tests, where, say, two lags of income, Y_{t-1} and Y_{t-2}, are used. Similarly they extend to univariate tests of a random walk where two lags of consumption, C_{t-1} and C_{t-2} are included. In the reparameterisation of the set of regressors given by C_{t-1} and ΔC_{t-1}, we are back to the augmented version (ADF) of the Dickey-Fuller test (1981) with critical values for the coefficient of the first regressor being non-standard.

As in Theorem 1, if income and, consequently, savings have no drift, then the role of the constant and the trend is only qualitative and the order of convergence to the asymptotic distribution of the test is unchanged. This result which has also been left unnoticed in the literature (see, for example, Sims, Stock and Watson (1986) who assume the presence of a trend in the model when consumption and income have zero drift). The key result is that it is possible to find a cointegrating vector, for the variables on the RHS of a model like (41), with no zero elements, irrespective of the presence of incidental parameters provided these same incidental parameters do not appear in the DGP. The $\{\hat{\pi}_i\}$ coefficients are asymptotically normally distributed with the ordinary normalisation rate of $T^{1/2}$. In the alternative parameterisation of the regressors, for example, the one in (42), this is not the case since C_{t-1} is I(1) and s_{t-1} is I(0) and the only cointegrating vector is (0,1). In this case, the coefficient $\hat{\gamma}_1$ of C_{t-1} converges to a Wiener functional at rate T rather than $T^{1/2}$.

There is however an apparent puzzle in the discussion. It is well known that the linear regression model is invariant to linear transformations. So, how can one explain the fact that simple reparameterisations alter the distribution of coefficients? The answer to the question is to be found in the different rates at which the coefficients tend to non-degenerate distributions. Take $\hat{\gamma}_1$ in (42) which tends to a non-degenerate distribution at rate T. $T^{1/2}\hat{\gamma}_2$ is asymptotically normally distributed. Then,

$$T^{1/2}\hat{\pi}_1 = T^{1/2}\hat{\gamma}_1 - T^{1/2}\hat{\gamma}_2 = o_p(1) + O_p(1) \tag{43}$$

Hence,

$$T^{1/2}\hat{\pi}_1 \rightarrow -T^{1/2}\hat{\gamma}_2 \tag{44}$$

$$T^{1/2}\hat{\pi}_2 \rightarrow T^{1/2}\hat{\gamma}_2 \tag{45}$$

since $\hat{\pi}_2$ and $\hat{\gamma}_2$ are identical.

Take now,

$$T^{1/2}\hat{\gamma}_1 = T^{1/2}\hat{\pi}_1 + T^{1/2}\hat{\pi}_2 \tag{46}$$

Although $(T^{1/2}\hat{\pi}_i)$ are asymptotically normally distributed, their variance covariance matrix is singular, as can easily be seen from (44) and (45). This makes the distribution of $T^{1/2}\hat{\gamma}_1$ degenerate. The distribution of $T\hat{\gamma}_1$ is not degenerate since $\hat{\pi}_1$ and $\hat{\pi}_2$ can be written as (proof available upon request)

$$\hat{\pi}_1 = (f_3(W))^{-1}\cdot(f_1(W) - f_2(W)) \tag{47a}$$

$$\hat{\pi}_2 = (f_3(W))^{-1}\cdot(f_4(W) - f_1(W)) \tag{47b}$$

where $f_i(W)$ $(i = 1,2,..4)$ are Wiener functionals such that $f_1(W)$ is $O_p(T^{5/2})$, $f_2(W)$ is $O_p(T^2)$, $f_3(W)$ is $O_p(T^3)$ and $f_4(W)$ is $O_p(T^2)$. Therefore,

$$\hat{\gamma}_1 = (f_3(W))^{-1}\cdot(f_4(W) - f_2(W)) = O_p(T^{-1}) \tag{48}$$

In order to illustrate the finite sample performance of the previous asymptotic results, we close this section by reporting the results of a small-scale Monte Carlo study which concentrates on simplified versions of models (41) and (42). We have set the trend equal to zero on the grounds that it only affects qualitatively the asymptotic distributions of the Wald tests. The simulations were carried out using 500 replications in the parameter space T = (120) and ρ = (0, 1) for the DGP described in (40). The initial observations, Y_0 and C_0, have been set equal to zero. Tables 2A and 2B present the actual rejection frequencies at the chosen significance levels. They confirm the standard distribution followed by the $(\hat{\pi}_i)$ and $\hat{\gamma}_2$ coefficients and the non-standard character of the t-ratio of the $\hat{\gamma}_1$ coefficient. Finally, the F test over-rejects at both significance levels, confirming its non-standard distribution. It is interesting to note that this last result explains the unsatisfactory behaviour of the F asymptotic approximation in the Monte Carlo analysis of Banerjee, Dolado and Galbraith (1987(a)) and Stock and West (1987), in the weak-efficiency framework, where only two lags of income are included in the set of regressors. The latter study also verifies the result that when the DGP and Model share the same incidental parameters then all the previous tests, including the t-ratio on $\hat{\gamma}_1$ and the F test, converge to asymptotic standard distributions. This confirms West's (1987) results as explained in Section 3. It is also worth noting that the results do not depend on ρ, in strong contrast to the single regressor case analysed in Sections 3 and 4. This may be seen from expressions I', II' and III' of Theorem 3.

6. Conclusions.

In this paper we have provided asymptotic and small-sample explanations of the over-rejection phenomenon which affects tests of the permanent income hypothesis of consumption in its rational expectations versions. We have confirmed our results using Monte Carlo simulations. The test in which over-rejection occurs is a rather restricted version, proposed by Flavin (1981), of a "weak efficiency test" . Most empirical studies of the Hall hypothesis adopt more general frameworks in the spirit of semi-strong rationality tests. In particular, we have illustrated how the presence of more than one regressor in the orthogonality conditions, such as our regression equations (41) and (42), allows the use of standard critical values for a subset of regressors. We have also shown the interesting and previously unnoticed result that in a particular representation of the regression model all the individual t-statistics are asymptotically normally distributed, a result which is confirmed by Monte Carlo experimentation. However, the F-test is non-standard. Summarising, one can consider the content of this paper as part of a general programme devoted to drawing implications from regression models with integrated variables. In particular, when considering rationality tests we believe that the extent of over-rejection may have been overstated.

References

Anderson, T.W. (1959): "On Asymptotic Distributions of Estimates of Parameters of Stochastic Difference Equations", Annals. Math. Statist., Vol. 30, pp. 676-87.

Banerjee, A. and Dolado, J. (1987): "Do We Reject Rational Expectations Models too Often? Interpreting Evidence using Nagar Expansions", Economics Letters, Vol. 24, pp. 27-32.

Banerjee, A., Dolado, J. and Galbraith, J.W. (1987a): "Rejections of Orthogonality in Rational Expectations Models: Further Monte Carlo Results for an Extended Set of Regressors", Economics Letters (forthcoming).

-- (1987b): "Orthogonality Tests Using Detrended Data: Interpreting Monte Carlo Results Using Nagar Expansions", (mimeo).

Campbell, J.Y. (1985): "Does Saving Anticipate Declining Labour Income? An Alternative Test of the Permanent Income Hypothesis", Econometrica (forthcoming).

------ and Deaton, A.S. (1987): "Is Consumption too Smooth?", Princeton University, Woodrow Wilson School Discussion

Paper No. 122.

Cavanaugh, C.L. (1986): "Roots Local to Unity", Harvard Institute of Economic Research Discussion Paper No. 1259.

Davidson, J.E.H and Hendry, D.F. (1981): "Interpreting Econometric Evidence: The Behaviour of Consumers' Expenditure in the U.K.", European Economic Review, Vol. 16, pp. 177-92.

Dickey, D.A. (1975): "Estimation and Hypothesis Testing for Nonstationary Time Series", Ph.D Thesis, Iowa State University.

------ and Fuller, W.A. (1979): "Distributions of the Estimates for Autoregressive Time Series with Unit Root", Journal of the American Statistical Association, Vol. 74, pp. 427-31.

----------------------- (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", Econometrica, Vol. 49, pp. 1657-72.

Dolado, J. and Jenkinson, T. (1987): "A Recent Survey on Cointegration and Unit Roots", Economic Perspectives (forthcoming).

Evans, G.B.A and Savin, N.E. (1981): "Testing for Unit Roots: 1" Econometrica, Vol. 49, pp. 1057-72.

---------------------------- (1984): "Testing for Unit Roots: 2", Econometrica, Vol. 52, pp. 1241-69.

Flavin, M. (1981): "The Adjustment of Consumption to Change in Expectations about Future Incomes", Journal of Political Economy, Vol. 89, pp. 974-1009.

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

Hall, R.E. (1978): "Stochastic Implications of the Life-Cycle Permanent Income Hypothesis: Theory and Evidence", Journal of Political Economy, Vol. 86, pp. 971-87.

Hendry, D.F. (1986): "Econometric Modelling with Cointegrated Variables: An Overview", Oxford Bulletin of Economics and Statistics (Special Issue on Econometric Modelling with Cointegrated Variables), Vol. 48, pp. 201-12.

Mankiw, N.G. and Shapiro, M.D. (1985): "Trends, Random Walks and Test of the Permanent Income Hypothesis", Journal of Monetary Economics, Vol. 16, pp. 165-174.

------------------------------ (1986): "Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models", Economics Letters, Vol. 20, pp. 139-145.

Muellbauer, J. M. (1983): "Surprises in the Consumption Function", Economic Journal (Supplement) pp. 34-50.

Nagar, A.L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations", Econometrica, Vol. 27, pp. 575-595.

Phillips, P.C.B (1986): "Understanding Spurious Regressions in Econometrics", Journal of Econometrics (forthcoming).

---------------- (1987a): "Time Series Regression with a Unit Root", Econometrica, Vol. 55, pp. 277-302.

---------------- (1987b): "Towards a Unified Asymptotic Theory for Autoregression", Biometrika, Vol. 70, pp. 535-47.

and Durlauf, S.N. (1985): " Multiple Time Series Regression with Integrated Processes", Review of Economic Studies, Vol. 53, pp. 473-496.

Sims, C., Stock, J.H. and Watson, M. (1986): "Inference in Linear Time Series with Some Unit Roots", (manuscript).

Stock, J.H. and West, K.D. (1987): "Integrated Regressors and Tests of the Permanent Income Hypothesis", Princeton

University, Woodrow Wilson School Discussion Paper No. 127.

West, K.D. (1987): "Asymptotic Normality when Regressors have a Unit Root", Princeton University, Woodrow Wilson School Discussion Paper No. 110.

White, J.S. (1959): "The Limiting Distribution of the Serial Correlation Coefficients in the Explosive Case", Annals. Math. Statist, Vol. 29, pp. 1188-97.

----------- (1961): "Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient", Biometrika, Vol. 48, pp. 85-94.

Table 1A.

Mean t-ratios and Critical Values
T=50

ρ →	1.000	0.800	0.500
α=0.01	-1.206 -3.590 -3.620	-0.965 -3.360 -3.380	-0.603 -2.970 -3.010
α=0.05	-1.206 -2.910 -2.890	-0.965 -2.680 -2.650	-0.603 -2.280 -2.280
α=0.10	-1.206 -2.580 -2.510	-0.965 -2.360 -2.270	-0.603 -1.920 -1.900

Table 1B.

T=200

ρ →	1.000	0.800	0.500
α=0.01	-1.150 -3.530 -3.480	-0.920 -3.260 -3.250	-0.575 -2.920 -2.900
α=0.05	-1.150 -2.860 -2.800	-0.920 -2.610 -2.570	-0.575 -2.240 -2.220
α=0.10	-1.150 -2.530 -2.430	-0.920 -2.320 -2.200	-0.575 -1.890 -1.860

Note: The topmost entry in each box, of Tables 1A and 1B, shows the central t-ratios computed from the Nagar expansions; the middle entry reproduces the Monte Carlo critical values; the bottom entry shows the critical values obtained, by using the rule described in the text, from the Nagar expansion t- ratios. The Monte Carlo simulations have been carried out using 500 replications.

Table 2A.

Rejection frequencies of Wald Tests in the Semi-Strong Efficiency Model

Model: $\Delta C_t = k + \pi_1 C_{t-1} + \pi_2 Y_{t-1} + u_t$

		$t(\pi_1=0)$	$t(\pi_2=0)$	$F(\pi_1=\pi_2=0)$
$\rho=0$	$\alpha=0.05$	0.055	0.033	0.123
	$\alpha=0.10$	0.118	0.122	0.232
$\rho=1$	$\alpha=0.05$	0.050	0.042	0.130
	$\alpha=0.10$	0.121	0.130	0.264

Table 2B.

Model: $\Delta C_t = k + \gamma_1 C_{t-1} + \gamma_2 (Y_{t-1} - C_{t-1}) + u_t$

		$t(\gamma_1=0)$	$t(\gamma_2=0)$	$F(\gamma_1=\gamma_2=0)$
$\rho=0$	$\alpha=0.05$	0.285	0.035	0.126
	$\alpha=0.10$	0.452	0.120	0.241
$\rho=1$	$\alpha=0.05$	0.260	0.050	0.133
	$\alpha=0.10$	0.436	0.133	0.264

Note: The Monte Carlo simulations have been carried out using 500 replications. The two models are non-detrended versions of models (41) and (42) in the main text.

Appendix

In this appendix, we present a summary of the formulae used to compute the Nagar expansions described in Section 4. All these expressions were written into a simple TSP programme which then produced the central 't' values for any required combinations of (T, λ, ρ). Both the program and the detailed proofs are available upon request.

Lemma 1: Let $\boldsymbol{\varepsilon}$ and $\mathbf{Y}_{-1}$ denote, respectively, column vectors of dimension (T-1) of the form

$\varepsilon = (\varepsilon_2, \varepsilon_3, \ldots, \varepsilon_T)'$, $\mathbf{Y}_{-1} = (Y_1, Y_2, \ldots, Y_{T-1})'$.

Let $\mathbf{i}$ be a conformable unit vector.

Define the following expectations:

$\mathbf{H} = E(\mathbf{Y}_{-1}\boldsymbol{\varepsilon}')$

$\boldsymbol{\Omega} = E(\mathbf{Y}_{-1}\mathbf{Y}_{-1}')$

$\mathbf{g} = E(\boldsymbol{\varepsilon}'\mathbf{Y}_{-1}\mathbf{Y}_{-1}'\mathbf{Y}_{-1})$

Then if (17) characterizes the true DGP with $\lambda = 1$, with $Y_0 = 0$, we have that,

A) $\text{trace}(\boldsymbol{\Omega}) = T(T-1)/2$

B) $$\mathbf{i}'\boldsymbol{\Omega}\,\mathbf{i} = \left[\frac{T^2(T-1)}{2} - \frac{1}{12}(T(T-1)(2T-1)) - \frac{1}{4}(T(T-1)\right]$$

C) $$\mathbf{i}'\mathbf{H}\,\mathbf{i} = \frac{\rho(T-1)(T-2)}{2}$$

D) $$\mathbf{i}'\mathbf{H}\boldsymbol{\Omega}\,\mathbf{i} = \left[\frac{7}{24}(T-1)^2T^2 - \frac{T^2}{12}(T-1)(2T-1) - \frac{T(T-1)^2}{12}\right]$$

$$\text{E)}\quad \mathbf{i}'\boldsymbol{\Omega}\mathbf{H}\,\mathbf{i} = \rho\left[\frac{(T-2)(T-1)T}{3}\left[T - \frac{1}{2}\right]\right.$$

$$- \frac{T}{2}\left[\frac{(T-1)(2T-1)T}{6} - 1\right]$$

$$+ \frac{(3T-1)}{6}\left[\frac{T(T-1)}{2} - 1\right]$$

$$\left. + \frac{1}{6}\left[\frac{T^2(T-1)^2}{4} - 1\right]\right]$$

$$\text{F)}\quad \mathbf{g} = 2\rho\left[T\left[\frac{(T+1)}{2} - 1\right] - T(T-1) - \left[\frac{T(T+1)(2T+1)}{6} - 1\right]\right.$$

$$\left. + \left[\frac{T(T+1)}{2} - 1\right]\right]$$

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**: Spanish industrial unemployment: some explanatory factors *(versión inglés)*. El desempleo en el sector industrial español: algunos factores explicativos *(versión español)*.

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**: Parsimony and omitted factors: The airline model and the census X-11 assumptions *(versión inglés)*. Parquedad en la parametrización y omisiones de factores: el modelo de las líneas aéreas y las hipótesis del census X-11 *(versión español)*.

8517 **Ignacio Mauleón**: A stability test for simultaneous equation models.

8518 **José Viñals**: ¿Aumenta la apertura financiera exterior las fluctuaciones del tipo de cambio? *(versión español)*. Does financial openness increase exchange rate fluctuations? *(versión inglés)*.

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.

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.

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.

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 y Juan Dolado**: Tests of the Life Cycle-Permanent Income Hypothesis in the Presence of Random Walks: Asymptotic Theory and Small-Sample Interpretations.

---

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

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

COINTEGRATION: A SURVEY OF RECENT DEVELOPMENTS

Juan J. Dolado
Tim Jenkinson

Banco de España. Servicio de Estudios
Documento de Trabajo n.º 8708

COINTEGRATION: A SURVEY OF RECENT DEVELOPMENTS

Juan J. Dolado *
Tim Jenkinson **

* Bank of Spain. Research Department
** Oxford University. Keble College

This paper has been prepared for the 1987 volume of *Economic Perspectives.* Harwood Academic Press. New York. We are grateful to M. Arellano, A. Banerjee, S. Burgess, A. Espasa, J. Galbraith, D. Hendry, A. Maravall, S. Nickell, N. Rau and H. Stamler for their comments. The usual caveat applies.

Banco de España. Servicio de Estudios
Documento de Trabajo n.º 8708

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: 84-505-7104-9
ISSN: 0213-2710
Depósito legal: M. 42884 - 1987

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. Their statistical properties and implications for the interpretation of regression models are covered in a unified way.

Key words: Unit root, Cointegration, Trends, Error Correction Mechanisms.

I. INTRODUCTION.

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 a deterministic trend. Stationary series should, however, at least have constant unconditional 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) 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 stochastic trends, as opposed to deterministic trends, with innovations to an integrated process being permanent instead of transient. For example, in terms of welfare costs, this implies that the costs of expectational errors produced by, say, policy shifts are far more serious than in the case where the shocks were purely transient.

In particular the presence of a unit root is implied in 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, hysteresis theories of unemployment, and, perhaps the most popular, the implications of the permanent income hypothesis for real consumption. In view of this epidemic of martingales in economics a voluminous literature on testing, estimation, and model specification in the presence of integrated variables has developed in the last few years, and the purpose of this survey is to provide a guide through this increasingly technical literature.

The analysis of cointegration developed out of the work on testing for, and implications of, unit roots in economic time-series. This survey covers both literatures in a unified way. Cointegration considers the conditions under which the use of standard regression analysis, when the individual series under consideration are integrated, is valid. Some of the properties of these 'cointegrating regressions' are extremely surprising, and suggest new ways to incorporate 'long-run' information (and constraints imposed by theory) into the statistical model. In addition, the concept of cointegration is, in many ways, a statistical definition of equilibrium. As such, cointegration offers a generic route to test the validity of the equilibrium predictions of economic theories.

The analysis of non-stationary variables requires a different statistical framework from the standard stationary case, and in Section II this framework is introduced, and testing procedures for unit roots are discussed. The proper treatment of integrated processes in regression analysis is then analysed using a variety of examples. This Section includes a number of more technical parts (denoted by an asterix) which could be avoided by those readers wishing to proceed quickly to the discussion of cointegration. Section III introduces the concept of cointegration, and discusses the implications for economic modelling and estimation, and the use of cointegration to discriminate between economic theories. Section V concludes.

It should be stressed that the concept of cointegration is relatively new, and that further developments, applications and Monte Carlo studies are appearing extremely rapidly. As a result, this survey is selective, and "best-practice" methods may well change in the near future.

II. INTEGRATION AND UNIT ROOTS.

A weakly stationary series should have a mean and variance that are time-invariant. However, many economic time-series certainly do not satisfy this condition, having first and second moments that appear to be increasing over time (see Escribano (1987) for precise definitions of integration in the i^{th} moment of a stochastic process). Such series are non-stationary, and may require differencing to induce stationarity[2]. A series requiring differencing d times to induce stationarity is denoted I(d), or "integrated of order d" (see Granger (1983)). A simple example of an I(1) series is the random walk:

$$\Delta y_t = \epsilon_t \qquad y_o = 0$$

where, for instance, ϵ_t is distributed $IN(0,\sigma_\epsilon^2)$. If, however, y were an autoregressive series such as

$$\Delta y_t = -\alpha y_{t-1} + \epsilon_t \qquad |\alpha| < 1 \qquad y_o \sim N(0,[1 - (1 - \alpha)^2]^{-1}\sigma_\epsilon^2)$$

then y would be stationary, or I(0). In this Section some of the important and far-reaching implications of the existence of unit roots ($\alpha = 0$) in economic time series are discussed.

---

[2]This condition is really too strong. In fact all that is needed is absence of trend in variance after suitable mean transformations (see, for example, Dickey *et al* (1986) and Escribano (1987)).

2.1 Statistical Properties of Integrated Series*

We will concentrate, in this section, on the statistical properties which stem from the presence of a single unit root, and start by considering the following data generation process (DGP) for the canonical stochastic integrated process $\{y_t\}_0^\infty$

$$\Delta y_t = -\alpha y_{t-1} + \mu + \varepsilon_t \qquad \alpha = 0 \qquad y_0 = 0 \qquad \ldots.(1)$$

$$\text{or} \quad y_t = \mu t + S_t \qquad S_t = \sum_1^t \varepsilon_j \qquad \ldots.(2)$$

where as a particularly interesting case we consider the driftless version of (1) with $\mu = 0$. In general, integrated series such as y_t are linear functions of time (with a slope of zero if $\mu = 0$). The deviations from this function of time are non-stationary, as they are the accumulation of past random shocks, giving rise to the concept of an integrated series.

To complete the specification of the DGP we need to impose some conditions on the innovation sequence $\{\varepsilon_t\}_1^\infty$. These restrictions are necessary if non-degenerate limiting distributions of the statistics discussed below are to be derived. The weakest set of conditons that achieve this aim is defined in detail in Phillips (1987a), and can be summarised as follows:

(a) $E(\varepsilon_t) = 0$ for all t

(b) $\sup_t E|\varepsilon_t|^{2\beta} < \infty$ for some $\beta>2$

(c) $\sigma^2 = \lim E(T^{-1}S_T^2)$ exists and $\sigma^2>0$

(d) ε_t is strong-mixing with mixing coefficients α_m such that $\Sigma\alpha_m^{(1-2/\beta)}<$ infinity

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 model (1) encapsulates a wide variety of DGP's. These include virtually any auto-regressive moving average (ARMA) model with a unit root, and even ARMAX models with unit roots and non-evolutionary exogenous processes. It is important to notice at this stage that only if we assume that the errors are iid$(0,\sigma^2)$ will $\sigma^2 = \sigma_\epsilon^2$. This restrictive case is an interesting one since most limiting distributions that have been numerically tabulated have been based on this assumption. However, this will not be the case in most empirical applications and hence in general $\sigma^2 \neq \sigma_\epsilon^2$.[3]

In order to derive the aforementioned limiting distributions, it is necessary, as in the stationary framework, to use a sequence of random variables, whose convergence is ensured by suitable transformation. More precisely, in the non-stationary framework, we need to focus on the sequence of partial sums $\{S_t\}_1^T$ which has to be transformed so that each element lies in the space D(0,1) of all real valued functions on the interval [0,1] that are right continuous and have finite left limits. This is achieved by defining the functions

$$X_T(r) = T^{-1/2}S_{j-1} \qquad \frac{j-1}{T} \leq r < \frac{j}{T} \qquad (j=1,\ldots T)$$
$$X_T(1) = T^{-1/2}S_T$$

Under the previous assumptions on the sequence $\{\epsilon_t\}$ we have that as T

---

[3]As an example, if ϵ_t follows an MA(1) process then $\epsilon_t = e_t - \theta e_{t-1}$ where e_t is iid$(0,\sigma_e^2)$. Then $\sigma^2/\sigma_\epsilon^2 = (1 + \theta)^2/(1 + \theta^2)$.

tends to infinity, $X_T(r) \rightarrow W(r)$, where $\rightarrow$ denotes weak convergence in probability. That is, $X_T(r)$ converges to a Wiener process.[4] Notice that $W(r)$ behaves like a random walk in continuous time such that for fixed r it is $N(0,r)$ and has independent increments.[5]

The most striking difference between the 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 Central Limit Theorems are substituted by Functional Limit Theorems (see, for example, Billingsley (1968)).

As an example of the previous remarks, the following standardised sample moments converge to Wiener functionals:

$$\text{(i)} \quad T^{-2}\Sigma y_t^2 \quad \rightarrow \quad \sigma^2 \int_0^1 W(t)^2\, dt \qquad \text{....(3)}$$

$$\text{(ii)} \quad T^{-3/2}\Sigma y_t \quad \rightarrow \quad \sigma \int_0^1 W(t)\, dt \qquad \text{....(4)}$$

$$\text{(iii)} \quad T^{-1}\Sigma y_{t-1}\epsilon_t \quad \rightarrow \quad \frac{\sigma^2}{2} [W(1)^2 - \sigma_\epsilon^2/\sigma^2] \qquad \text{....(5)}$$

Note the divergences in the orders of magnitude of these limiting distributions with the conventional stationary distributions, i.e. order in probability T^2, denoted $O_p(T^2)$, instead of $O_p(T)$ in (3); $O_p(T^{3/2})$ instead of $O_p(T)$ in (4); and $O_p(T)$ instead of $O_p(T^{1/2})$ in (5). These

---

[4]This Wiener process will lie in the space C[0,1] of all real-valued functions continuous on the interval [0,1].

[5]Moreover, an extension of the Slutsky Theorem in conventional asymptotic theory also applies in this framework, in the sense that if g(.) is any continuous function on C[0,1] then $X_T(r) \rightarrow W(r)$ implies that $g[X_T(r)] \rightarrow g[W(r)]$.

differences shed light on the non-conventional features of the coefficient consistency and limiting distributions when testing for unit roots, and will be important in the discussion of cointegration in Section III.

If, for instance, OLS is applied to (1) , it is easy to show that, using the sample variability results summarised in (3) - (5), the slope $\hat{\alpha}$ and its t-ratio converge to the following distributions, in the case when $\mu = 0$:

$$T\hat{\alpha} \rightarrow \frac{\frac{1}{2}\left[W(1)^2 - \sigma_\epsilon^2/\sigma^2\right]}{\int_0^1 W(t)^2\,dt} \qquad \ldots.(6)$$

$$t_{\alpha=0} \rightarrow \frac{\frac{1}{2}\sigma\left[W(1)^2 - \sigma_\epsilon^2/\sigma^2\right]}{\sigma_\epsilon\left[\int_0^1 W(t)^2\,dt\right]^{1/2}} \qquad \ldots.(7)$$

From (6) we note that $\hat{\alpha}$ converges to its true value zero at a rate of $O_p(T^{-1})$ instead of the conventional $O_p(T^{-1/2})$. Similarly, from (7), the corresponding t-ratio has a non-degenerate distribution which is different from the standardised normal distribution which is used in conventional asymptotic theory.

2.2 Testing for Unit Roots.

The previous statistical implications of the unit root hypothesis in the time-series representation of univariate models underscore the need to have reliable procedures to test formally this hypothesis. Investigations by Dickey (1976), Dickey and Fuller (1979, 1981) and Fuller (1976) have constructed by numerical simulations the corresponding critical values of the limiting distributions expressed in (6) and (7). Table 1 collects the exact null and alternative hypotheses under which these simulations were performed. The unrestricted model contains both a constant and a trend as regressors, plus an error term subject to a first order autoregressive representation. Three interesting cases follow. Case 1 in that in which both the drift and the trend are zero, such that under the null we find a pure driftless random walk. Case 2 describes the case in which there is a drift but no trend, and consequently the model under the null is again the driftless random walk. Finally, Case 3 relates to the most general case in which both constant and trend are different from zero, and hence the model under the null hypothesis is random walk with drift. Notice that in all cases the error terms under H_0 are assumed to be $iid(0,\sigma^2)$ for simulation purposes.

Table 1: Use of Tables to Test for a Unit Root in Univariate Models.

	H_0 $(\alpha = 0)$		H_1 $(y_t = \mu + \beta t + u_t)$, $\Delta u_t = -\alpha u_{t-1} + \nu_t$
Case 1 $(\mu = \beta = 0)$	$\Delta y_t = \epsilon_t$	$\longleftarrow B_1 \longrightarrow$	$\Delta y_t = -\alpha y_{t-1} + \epsilon_t$
Case 2 $(\mu \neq 0, \beta = 0)$	$\Delta y_t = \epsilon_t$	$\longleftarrow B_2 \longrightarrow$	$\Delta y_t = -\alpha y_{t-1} + \alpha\mu + \epsilon_t$
	$\swarrow$	T_0	$\nearrow$ (Case 2 H_1)
Case 3 $(\mu \neq 0, \beta \neq 0)$	$\Delta y_t = \beta + \epsilon_t$	$\longleftarrow B_3 \longrightarrow$	$\Delta y_t = -\alpha y_{t-1} + \beta\alpha t + [\mu\alpha + \beta(1-\alpha)] + \epsilon_t$

Note: B_i (i=1,2,3) denote 1^{st}, 2^{nd} and 3^{rd} blocks of Table 8.5.2 in Fuller (1976), T_0 denotes Table for critical values for the standardised Normal distribution.

From (6) and (7) two basic statistics can be derived to test the null hypothesis of a unit root. The first test refers to the scaled regression coefficient $T\hat{\alpha}$ while the second concentrates on the t-ratio $t\hat{\alpha}$. Critical values for both asymptotic distributions are found in Fuller (1976)[6]. The arrow scheme in Table 1 explains the proper use of these tables, depending on the choice of the model representing the unrestricted hypothesis. If we start with Case 1, then we should use the first block (denoted B_1) of critical values in Tables 8.5.1 and 8.5.2. Similarly, the choice of model with constant and constant plus trend implies the use of the second and third blocks (denoted B_2 and B_3) respectively. A very interesting case, of which some practitioners are unaware, is that, when choosing the models with a constant, if that nuisance parameter is significant under the null (checked by simply regressing Δy_t on a constant) then the right critical value for the t-ratio will be found in the standardized normal distribution table (denoted T_0), rather than in the Dickey-Fuller tables (see West (1986))[7].

The same peculiar result obtains when, after using the most general model, the constant and the trend are significant under the null (checked

---

[6] Chapter 8, Tables 8.5.1 and 8.5.2.

[7] In general for Case 2 it can be shown that $T^{3/2}\hat{\alpha} \sim N(0,\ 12\sigma^2/\mu^2)$ and $t_{\hat{\alpha}} \sim N(0,\ \sigma^2/\sigma_\epsilon^2)$.

by regressing Δy_t on a constant and trend)[8]. In both instances, the interesting outcome of looking at the wrong tables is enlightened when we find $t_{\hat{\alpha}}$,say at the 5% level, larger than 1.96 but smaller than the corresponding critical values in the D-F tables. Upon these conditions, we should be rejecting the null hypothesis instead of accepting it. The intuition behind this peculiar result is that if there is a unit root and, say, a constant, the integrated series depends on a deterministic trend and a stochastic one. Moreover, the sample variability of the deterministic trend is of $O_p(T^3)$ which dominates the order of the sample variability of the stochastic trend which is of $O_p(T^2)$. But we know that the existence of a deterministic trend in a regression model does not affect the asymptotic normality of the standardised estimates, hence normality follows.

It is clear from the previous discussion and the derivations of the conventional and unconventional statistics shown in (6) & (7) that if the error sequence $\{\epsilon_t\}$ is correlated, the distributions will depend on the nuisance parameter $\sigma^2/\sigma_\epsilon^2$. In such a case there is a need to either change the estimation method (that is, adopt another regression model), or modify the statistics described above. Dickey and Fuller (1981) favour the first approach by enlarging the regression model by adding in a lag polynomial of Δy_t such that these terms capture the serial correlation in any of the unrestricted models contained in Table 1. It can be shown,

---

[8]The case where the DGP contains a unit root and a trend does not seem to be too realistic a *priori* since, in logarithmic form, it implies an ever increasing (or decreasing) rate of change. In general for Case 3 it can be shown that $T^{5/2}\hat{\alpha} \sim N(0,\ 180\sigma^2/\beta^2)$ and $t_{\hat{\alpha}} \sim N(0,\ \sigma^2/\sigma_\epsilon^2)$.

that under the null hypothesis, $t_{\hat{\alpha}}$ in the enlarged model has the same limiting distribution as when the errors are iid, giving rise to the so-called Augmented Dickey-Fuller tests. Note, however, that it is no longer legitimate to use $T\hat{\alpha}$ as the basis of a test in any of the variants, since they are not invariant to the true population value of the parameters of the distributed lag in Δy_t.

Nevertheless, this solution introduces the problem that we might need a large number of lags of Δy_t in order to obtain uncorrelated residuals. Recently, Said and Dickey (1984) have shown that if ϵ_t contains moving average terms, the number of extra regressors needs to increase with the sample size at a rate $(T^{1/3})$. Given that the majority of the macroeconomic variables studied in the seminal paper by Nelson and Plosser (1982) were adequately represented by an IMA(1) process, this seems a quite likely situation. Schwert (1985), using Monte Carlo simulations, has recently shown that the exact size of the test may be far from the nominal size if the order of the autoregressive correction is not increased as the sample size increases. Accordingly, it would be desirable to have an approach for the test which takes into consideration the structure of the residuals in a non-parametric way under the assumptions (a) - (d) above. This is the approach developed by Phillips and Perron (1986) and Perron (1987), and described briefly in the Appendix.

Next, we briefly discuss a testing strategy based on the choice of the appropriate initial unrestricted model in Table 1, as well as on the choice of data sample. With respect to the first issue, we advocate estimating the most unrestricted model initially, as in Case 3. Then use the test statistic (c) in Table 2 of the Appendix to test for a unit root, using the critical values contained in B_3 of Table 8.5.2. If the null hypothesis of a unit root is rejected there is no need to go further. If it is not rejected, test for the significance of the trend (a rather implausible case, as discussed earlier) using the test statistic in row (d) of Table 2. If it is significant, then test for its significance under the null using the ordinary tables. Its significance under the null would imply that the ordinary tables, instead of Table 8.5.2 should have been used to test for the unit root. If the trend is not significant under the alternative, estimate the unrestricted model in Case 2 in Table 1. Test again for the unit root using the test statistic (b) in Table 2, looking at B_2 in Table 8.5.2. 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 constant under the alternative using the test statistic shown in row (e) of Table 2. If the procedure reaches the most restrictive alternative model, as in Case 1, then the unit root should be tested with the critical values contained in B_1 of Table 8.5.2. Failure to follow this strategy may lead to serious misinterpretations.

An alternative strand to the literature on testing for unit roots is that suggested by Sargan and Bhargava (1983). They advocate the use of the conventional Durbin-Watson (DW) statistic from the simple OLS regression of the variable under consideration on a constant, that is

$$y_t = c + u_t$$

$$\Delta u_t = -\alpha u_{t-1} + \epsilon_t \qquad \epsilon_t \text{ distributed } IN(0,\sigma_\epsilon^2)$$

Then the null hypothesis of α=0 is tested against the alternative that the errors follow a stationary first order autoregressive process. A unit root for the error process is equivalent to the structural element following a random walk. The value of the DW statistic will obviously tend to be very low when the root in the error process tends towards unity, since DW $\simeq 2(\alpha)$. The test can be performed using the standard DW statistic generated by most statistical programs along with the table of critical values presented by Sargan and Bhargava (1983) under the unit root null hypothesis. This test can be shown to be the uniformly most powerful invariant test against the alternative of a stationary first order autoregressive error process. An important feature of the test is its invariance to whether a trend enters into the true model, unlike the other tests considered above. However, the test is only powerful in discriminating between the simple random walk and stationary first order autoregressive processes, and thus lacks generality.

Having discussed the main tests that have been proposed for unit roots, an important qualification should be noted. In practice, economic time series emerge from this testing procedure as appearing to be I(1). However, in the context of, for example, the Sargan and Bhargava approach, the estimated degree of autoregression in the residuals is often in excess of 0.95. In other words, a value of 0.1 for the DW statistic is fairly typical in the static regression (given that DW $\simeq 2\alpha$). However, as Sargan and Bhargava note, the power of the test for a unit root against such highly autoregressive alternatives is exceedingly low. This is hardly suprising, since discrimination between a 0.95 autoregressive process and a random walk is extremely difficult in the relatively short samples typically used in economics. The practical implications are, however, important when we consider the powerful cointegration results that depend upon the individual time series possessing unit roots.

It should be noted that the definitions and properties introduced up to now for scalar random variables extend to multivariate cases (see Phillips and Durlauf (1986b)) by applying the properties to each element of the vector. This extension immediately raises the question of having components with different degrees of integration, or the possibility of finding linear transformations of those components with a different order of integration to the order of the individual elements of the vectors. Both these issues are raised in the next two Sections.

Finally, as far as the choice of data sample is concerned the main result concerns the trade-off between span and sampling interval (see Shiller and Perron (1985)). For a given span, more observations lead to higher power of the previous tests. Similarly, a longer span for a given number of observations leads to higher power. Of course, this intuitive result had to be mediated by the relevant alternative. So, for example, since for macroeconomic series, the natural alternative is mean reversion over a period similar to the length of business cycles, a long span of annual data should be preferred to a shorter span with, say, quarterly or monthly data.

2.3 Asymptotic Theory and Monte Carlo Results*

Having examined the important statistical implications of integrated processes, we proceed to use this theory to interpret a number of results concerning the treatment of integrated series in regression analysis. An explicit analytical solution to the asymptotic behaviour of parameter estimates and regression statistics permits a unification of the disparate Monte Carlo studies that presently exist in the literature. We present a summary of results on analyses which range from inappropriate detrending of integrated series, to efficiency tests, including the familiar spurious regression results. Most of the results derive from the work of Phillips or Phillips and Durlauf in a recent long sequence of papers that are referenced at the beginning of each case.

In order to unify as much as possible the treatment of different cases, the following description procedure is adopted. Each case will be characterised by a DGP and an estimated model (denoted simply MODEL). The distributional results, which happen to be functionals of Wiener processes, will be denoted generically by f(W), whose precise expressions are given in the appropriate references. At the end of each case we offer an intuitive explanation of the analytical results, together with some remarks about the use of certain regression statistics, which prove to be useful to detect misspecifications in the estimated models.

<u>De-trending</u> (Phillips and Durlauf (1986a))

$$\text{DGP} \quad \Delta y_t = \epsilon_t$$
$$\text{MODEL} \quad y_t = \hat{\mu} + \hat{\beta} t + \hat{u}_t \qquad \ldots.(8)$$

Summary of Results:

$T^{-1/2}\hat{\mu} \rightarrow f(W)$ $\quad T^{1/2}\hat{\beta} \rightarrow f(W)$ $\quad T^{-1/2} t_{\mu=0} \rightarrow f(W)$

$T^{-1/2} t_{\beta=0} \rightarrow f(W)$ $\quad T^{-1} s^2 \rightarrow f(W)$ $\quad T.DW \rightarrow f(W)$

$R^2 \rightarrow f(W)$

This case tackles the issue of inappropriate de-trending of integrated processes, under the traditional belief that conventional asymptotic theory could be applied to detrended series. We observe that the $\hat{\beta}$ coefficient is consistent, converging to its true value of zero. 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 (s^2) also diverges reflecting the fact that the residuals of the model are non-stationary around the trend. The coefficient of multiple correlation (R^2) converges to a non-degenerate limiting distribution. The results for the Durbin-Watson statistic (DW) appear quite promising, confirming its powerful role as a misspecification diagnostic (see Sargan and Bhargava (1983)). The intuition behind all these disparate results stems from the different orders of magnitude of the sampling variability of the regressors and regressand in the model, i.e. $O_p(T) = \hat{\mu}\, O_p(1) + \hat{\beta}\, O_p(T^2)$. The divergence of the order of magnitude highlights the fact that $\hat{\beta}$ converges while $\hat{\mu}$ diverges, according to when the sample variances of their corresponding regressors are larger or smaller than the sample variance of the regressand.

<u>Encompassing Tests</u> (Phillips and Durlauf (1986a))

$$\text{DGP} \qquad \Delta y_t = \varepsilon_t$$

$$\text{MODEL} \qquad \Delta y_t = \hat{\mu} + \hat{\beta} t - \hat{\alpha} y_{t-1} + \hat{u}_t \qquad \ldots.(9)$$

Summary of Results:

$T^{3/2}\,\hat{\beta} \rightarrow f(W)$ $\qquad$ $T\,\hat{\alpha} \rightarrow f(W)$ $\qquad$ $T^{-1/2}\, t_{\alpha=0} \rightarrow f(W)$

$s^2 \rightarrow \sigma^2_{\varepsilon}$ $\qquad$ $F_{\beta=0,\alpha=0} \rightarrow f(W)$

This case interprets the unit root test in Case 3 of Table 1, where the issue is to discriminate between trends and integrated processes. The model embodies both alternatives, and uses the F test to discriminate between the alternatives. The encompassing test works as follows: $H_A(\beta=0,\alpha=0)$ corresponds to the integrated process, whereas $H_B(\alpha=1)$ corresponds to the deterministic trend. Denoting rejection of a hypothesis by $\tilde{}H$, the following combinations of rejections and non-rejections would operate the encompassing tests: $(H_A, \tilde{}H_B)$ supports the Random Walk; $(\tilde{}H_A, H_B)$ supports the Deterministic Trend. In view of the divergence of $t_{\alpha=0}$, we would conclude that H_B is always rejected for a sufficiently large sample. The F-test for H_A converges to a non-degenerate distribution, which differs from the ordinary F distribution, and hence requires the Dickey-Fuller critical values as explained above. The disparate sample variability of regressand and regressors is given by $O_p(1) = \hat{\mu}\, O_p(1) + \hat{\beta}\, O_p(T^2) - \hat{\alpha}\, O_p(T)$.

<u>Non de-trended Spurious Regression.</u> (Phillips (1987b))

$$\text{DGP} \quad \Delta y_t = \epsilon_t \;;\; \Delta x_t = \nu_t \;;\; E(\epsilon_t \nu_t) = \rho\sigma_\epsilon\sigma_\nu\delta_{ts} \qquad \ldots.(10)$$

$$\text{MODEL} \quad y_t = \hat{\mu} + \hat{\alpha} x_t + \hat{u}_t$$

Summary of Results:

$\hat{\alpha} \rightarrow f(W)$ $\qquad T^{-1/2}\,\hat{\mu} \rightarrow f(W)$ $\qquad T^{-1/2}\, t_{\alpha=0} \rightarrow f(W)$

$T^{-1}\, s^2 \rightarrow f(W)$ $\qquad T.DW \rightarrow f(W)$ $\qquad R^2 \rightarrow f(W)$

This case interprets the familiar Monte-Carlo results of Granger and Newbold (1974), reinforcing analytically the divergence of $t_{\alpha=0}$ despite the fact that $\hat{\alpha}$ and R^2 possess non-degenerate distributions. Again, as in the de-trending case, the DW statistic detects misspecification of the model, although GLS corrections fail to provide the right answer. The orders of magnitude of the sampling variability in the model are: $O_p(T) = \hat{\mu}\ O_p(1) + \hat{\beta}\ O_p(T)$. Notice that the equality of the orders of magnitude between y_t and x_t provides the possibility of finding certain combinations of both variables such that the residuals are stationary, despite the non-stationarity nature of the variables themselves.

<u>De-trended Spurious Regression.</u> (Phillips and Durlauf (1986a))

$$\text{DGP} \quad \Delta y_t = \epsilon_t \ ; \quad \Delta x_t = \nu_t \ ; \quad E(\epsilon_t \nu_s) = \rho\sigma_\epsilon\sigma_\nu\delta_{ts} \qquad \ldots.(11)$$

$$\text{MODEL} \quad y_t = \hat{\mu} + \hat{\beta}t + \hat{\alpha}x_t + \hat{u}_t$$

Summary of Results:

$$\hat{\alpha} \to f(W) \qquad T^{-1/2}\hat{\mu} \to f(W) \qquad T^{1/2}\hat{\beta} \to f(W)$$

$$T^{-1/2}\, t_{\alpha=0} \to f(W) \qquad T^{-1}s^2 \to f(W) \qquad T.DW \to f(W)$$

This case interprets results from spurious regression models where y_t and x_t are de-trended, with the aim of inducing stationarity in the variables prior to the regression. The results are similar to the previous case with the addition that $\hat{\beta}$ is consistent, since the order of magnitude of the sample variance of the trend is $O_p(T^2)$. Notice that the presence of a trend in the regression only has qualitative effects on the asymptotic distribution.

<u>Efficiency Tests</u>. (Banerjee and Dolado (1987))

$$\text{DGP} \quad \Delta y_t = \epsilon_t \; ; \quad \Delta x_t = \nu_t \; ; \quad E(\epsilon_t \nu_t) = \rho\sigma_\epsilon \sigma_\nu \delta_{ts} \qquad \ldots.(12)$$

$$\text{MODEL} \quad \epsilon_t = \hat{\mu} + \hat{\beta}t + \hat{\alpha}x_{t-1} + \hat{u}_t$$

Summary of Results:

$T\hat{\alpha} \rightarrow f(W)$ $\qquad T^{1/2}\hat{\mu} \rightarrow f(W)$ $\qquad T^{3/2}\hat{\beta} \rightarrow f(W)$

$t_{\alpha=0} \rightarrow f(W)$ $\qquad T.R^2 \rightarrow f(W)$

This case interprets recent Monte-Carlo results by Mankiw and Shapiro (1985) on the over-rejection of the orthogonality condition which characterises rational expectations models. In this case, the three estimated parameters and R^2 are consistent, but $t_{\alpha=0}$, the basis of the previous test (see Flavin (1981)) does converge to a non-degenerate distribution which differs from the standardised normal. The orders of magnitude of the sample variances are $O_p(1) = \hat{\mu}\ O_p(1) + \hat{\beta}\ O_p(T^2) + \hat{\alpha}\ O_p(T)$.

III. COINTEGRATION.

Whereas the analysis and implications of unit roots in individual time series excited mainly the econometrician, far more general economic interest has developed in the concept of cointegration, which analyses groups of integrated variables. The major reason for this is the possibility of estimating, and testing the existence of, long run economic relationships suggested by theory. As was explained in the previous sections, many individual economic time series appear to be non-stationary, requiring differencing at least once to induce stationarity. Yet economic theory rarely suggests equilibria that are not stationary functions of the variables involved. This would imply that there may exist fundamental economic forces that, over time, make variables move stochastically together. In other words, whereas the individual economic variables involved in a theory may all be non-stationary, the deviations from a given equilibrium may be bounded.

For many years the problems associated with static regressions between time-series have been known (for an interesting historical account see Hendry (1986)). The problem of 'spurious' regressions discussed earlier led many economists to adopt the Box-Jenkins (1970) methodology of transforming all the variables to stationary series prior to regression, so that, for the most part, differenced variables were considered. This, of course, resulted in models that disregarded the low frequencies of the variables, and so did not allow for any of the long run relationships which economic theory normally suggested. These features made the models difficult to interpret.

One response to such problems was the use of error-correction mechanisms (ECM) in econometric models. Models including ECMs have been widely used since Sargan (1964), and have the advantage of retaining information about the levels of variables, and hence any long-run relationships between such variables, within the model (see, for example Davidson *et al* (1978), Currie (1981) and Salmon (1982)). In an important paper, Granger (1983) establishes the equivalence between cointegration and error-correction. That is, ECM's produce cointegrated sets of variables, and, if a cointegrated set of variables is found, it must have an ECM representation. To a great extent, cointegration provides formal statistical support for the use of error-correcting models, and suggests additional procedures to test model specification in a ***static*** sense, and proposes ways to parameterise the error-correcting mechanism.

A vector of variables x_t is said to be cointegrated if

(i) each element of x_t is I(d)

and (ii) there exists a vector α such that $\alpha' x_t$ is I(d-b), where $\alpha \neq 0$ and $b > 0$.

For example, in the case of d=b=1, if x_t is cointegrated, each variable in x_t would each be I(1), but some linear combination of them would be I(0). If such a linear combination can be found, α is called the cointegrating vector.

The relationship between cointegration and equilibrium now becomes clearer. One natural way to characterise equilibrium between a set of variables is to define equilibrium to occur when a linear constraint is satisfied, such as

$$\alpha' x_t = 0 \qquad \text{....(13)}$$

For example, if we believe that a proportion λ of any increase in labour productivity is eventually passed on in the form of real wages then, in equilibrium, $w = c + \lambda Q$ where w and Q denote real wages and productivity respectively, and c is a constant. Therefore, if

$$w - c - \lambda Q = 0 \qquad \text{....(14)}$$

in any time period, then the labour market would be in equilibrium. Of course, real wages may take some time to respond to changes in productivity, and the process by which equilibrium tends to be restored may be complex, in which case the scalar $z_t = w_t - c - \lambda Q_t$ would measure the deviation from equilibrium, or disequilibrium, in period t. If w and Q are cointegrated, then, by the above definition, the deviations from equilibrium will be bounded. An obvious way of testing the theory is then to determine the order of integration of z_t. If it is not possible to reject the null hypothesis of a unit root for z_t then there will be no tendency for the real wage to move towards the putative equilibrium, in which case the estimated equilibrium would be misleading and irrelevant.

In the case of testing for cointegration between two variables x_1 and x_2, if a cointegrating vector exists, it must be unique. To see this, suppose x_1 and x_2 are both I(1) variables and $z_t = x_{1t} + \mu x_{2t}$ is I(0). Then μ must be unique, since any other linear combination would add or subtract a term in x_{2t}, which would be I(1), which would result in z_t also being I(1). However, when x has more than two components, if a cointegrating vector exists, it need not be unique. In general, if x has N components, there may be r linearly independent cointegrating vectors, where $r \leq N-1$.

To illustrate the possible outcomes, consider the following example, taken from Granger and Engle (1987). Suppose y_t and x_t are jointly distributed according to the following data generation process:

$$\begin{aligned} y_t + \beta x_t &= v_t & \qquad v_t &= \rho_1 v_{t-1} + \varepsilon_{1t} \\ y_t + \alpha x_t &= u_t & \qquad u_t &= \rho_2 u_{t-1} + \varepsilon_{2t} \end{aligned} \qquad \ldots.(15)$$

where ε_1 and ε_2 are distributed independently N(0,1). Four possible cases exist:

(i) $\rho_1 = 1$, $\rho_2 < 1$ which implies that x_t and y_t are I(1) and the cointegrating vector is $(1,-\alpha)$.

(ii) $\rho_1 < 1$, $\rho_2 = 1$ which implies that x_t and y_t are I(1) and the cointegrating vector is $(1,-\beta)$.

(iii) $\rho_1 = \rho_2 = 1$ which implies that x_t and y_t are I(1) but there does not exist a cointegrating vector.

(iv) $\rho_1 < 1$, $\rho_2 < 1$ which implies that x_t and y_t are I(0) and so any linear combination of x and y will be I(0).

The last case introduces some interesting issues. The test for cointegration is actually a conditional test: conditional on x_t and y_t being I(1), the discovery of an I(0) linear combination would imply that the variables are cointegrated. However, as noted above, when ρ_1 and ρ_2 are unknown, the power of tests for unit roots against alternatives of roots close to the unit circle is often exceedingly low. In such situations type II errors, that is the acceptance of a unit root rather than a root of, say, 0.95, are likely to occur. Jenkinson (1986b) presents some Monte Carlo evidence on the hazards of inference when some, or all, of the variables under consideration are, in fact, highly autoregressive rather than I(1). The intuition is that we need extremely long time-series in order to distinguish borderline from unit root cases. Banerjee et al (1987a,b) present easily computable approximations to the correct critical values in general cases.

3.1 Estimation

An obvious issue is the question of estimating, and testing for the existence of, cointegrating vectors. Consider again the problem of estimating $\boldsymbol{\alpha}$ and testing for the stationarity of z in the model

$$\boldsymbol{\alpha}'\mathbf{x}_t = z_t \qquad \ldots.(16)$$

If all the variables in x are I(1), then in general a linear combination of these variables, and hence z_t , will be I(1). Therefore, almost all the α vectors will produce a series z with asymptotically infinite variance. The exceptions to this will be any cointegrating vectors. Now since Ordinary Least Squares estimation minimises the residual variance of z_t , the estimated α vector derived from an OLS regression of the simple model (16) where all variables are in levels and no dynamics are included, should yield an excellent approximation to a true cointegrating vector, if one exists.

This result is one important reason why interest in cointegration has itself exploded like a non-stationary series. It implies that to parameterise a long-run equilibrium relationship between a set of variables, all that is needed is a simple static OLS regression between those variables. This simple regression can even be performed at the first stage of a research program, as is advocated by the Engle & Granger (1987) 'two-step estimator' discussed in Section 3.2 below. In any event, such an initial check may indicate to what extent the equilibrium predictions of the economic theory are consonant with the data, and, to put the argument at its strongest, whether it is fruitful to expend resources to model the short term dynamics around the equilibrium. To make things even easier, at least one econometric software package, PC-GIVE, automatically provides basic tests to determine the order of integration of the variables in the model as an addition to such summary measures as the means and standard deviations of the variables!

Indeed, the OLS estimate of any cointegrating vector should converge to the true value extremely quickly. To see this, consider the following case characterised as in the taxonomy of Section 2.3.

$$\begin{aligned} \text{DGP:} \quad \Delta x_t &= \varepsilon_t \\ y_t &= \alpha x_t + e_t \qquad e_t = (1 - \rho L)\xi_t \qquad \ldots.(17) \\ \text{MODEL:} \quad y_t &= \hat{\mu} + \hat{\alpha} x_t + \hat{u}_t \end{aligned}$$

Summary of Results:

$T(\hat{\alpha} - \alpha) \rightarrow f(W)$ $\qquad T(1 - R^2) \rightarrow f(W)$ $\qquad T^{1/2}\hat{\mu} \rightarrow f(W)$

$T.DW \rightarrow 2(1 - \rho)$ $\qquad t_{\hat{\alpha}=\alpha} \rightarrow f(W)$

The interpretation of the results illustrates very clearly the previous informal discussion. The slope in the static regression converges to its true value α at a rate of $O_p(T^{-1})$ instead of the ordinary rate of $O_p(T^{-1/2})$. The intuition is again clear: $\hat{\alpha}$ is computed using the ratio of a covariance, which is of $O_p(T)$, by a variance, which is of $O_p(T^2)$, given that both x_t and y_t are I(1). Therefore, any bias in $\hat{\alpha}$ is of $O_p(T^{-1})$. However, in spite of this super-consistency of $\hat{\alpha}$, its distribution is not asymptotically normal, and therefore the computed standard errors of the coefficients lack meaning. Since both x_t and y_t are driftless processes, $\hat{\mu}$ converges consistently to zero, although at a slower speed than $\hat{\alpha}$. The coefficient of multiple correlation R^2 is also $O_p(T)$ consistent to unity, reflecting the fact that in the bivariate case, under cointegration, the product of the slope and the inverse slope is unity. This feature will be exploited in the discussion below. Finally, the DW statistic converges to the standard result under the assumption that e_t follows an AR(1) process.

An important associated result relates to the existence of simultaneity biases and errors in variables. Such biases in parameter estimates normally derive from the correlation between the regressors and the errors, which is ordinarily assumed to be of $O_p(T)$. However, given the fact that in cointegrating regressions $\Sigma x_t e_t$ will be of a lower order of magnitude than Σx_t^2, such biases are asymptotically negligible. This implies that issues of endogeneity and exogeneity are not, in general, relevant in static cointegrating regressions.

The most important result of the previous discussion relates to the super-consistency of $\hat{\alpha}$. However, biases in $\hat{\alpha}$, despite being $O_p(T^{-1})$, can still be large in small samples. In a Monte Carlo study, Banerjee et al (1986) discovered large biases in $\hat{\alpha}$ derived from bivariate cointegrating regressions. In addition, $\hat{\alpha}$ did not converge rapidly to α. Given that the R^2 of the regression converges at the same rate as the bias, they propose $(1 - R^2)$ as a proxy for the latter. In fact, for the canonical model discussed previously, the linear relationship between both statistics turns out to be:

$$\hat{\alpha} - \alpha = \alpha^2(1 - R^2) + O_p(T^{-1}) \qquad \text{....(18)}$$

which suggests rather strongly that cointegrating regressions without R^2 very close to unity should be viewed with caution (see, for example, Campbell and Shiller (1986)). However, in the context of a multiple regression, the R^2 of an equation cannot fall when an additional variable is added, and this implies that a high R^2 is not sufficient to guarantee that each included variable is germane to the model, nor that the estimated coefficients closely approximate their true values. This issue of functional forms is discussed in more detail below.

Another important implication is that in contrast to normal regressions where multicollinearity amongst the regressors is often considered a problem, in the context of a cointegrating static regression such multicollinearity is *essential*: if variables do not follow similar trends over time then no linear combination of the (individually non-stationary) time-series will be stationary. Indeed, in terms of estimation of the cointegrating vector α, the multicollinearity amongst the regressors will produce a nearly-singular (X'X) matrix corresponding to the cointegrating vector. In this sense multicollinearity is a positive advantage!

The effect of running the static regression (16) to estimate α is to push all the dynamic adjustment terms into the residual $\hat{u}_t$. These dynamic terms can all be parameterised in terms of I(0) series of the form Δy_{t-i}, Δx_{t-j}, and $(y - \beta x)_{t-k}$ where the values of i,j, and k will depend upon the nature of the ARMA processes generating x and y. To illustrate this, consider a simple model in which the true dynamic relationship is given by:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 x_t + \alpha_3 x_{t-1} + u_t \qquad \text{....(19)}$$

where y and x are I(1) and CI(0). Suppose that in the long run the homogeneity restriction $\alpha_1 + \alpha_2 + \alpha_3 = 1$ holds. This is equivalent to saying that in the long run y and x move together. Now equation (9) can be rewritten as

$$y_t = \alpha_1 y_{t-1} + \alpha_2 \Delta x_t + (1-\alpha_1) x_{t-1} + u_t \qquad \text{....(20)}$$

or as

$$y_t = x_t + \alpha_1 (y - x)_{t-1} + (\alpha_2 - 1) \Delta x_t + u_t \qquad \text{....(21)}$$

Now (y - x) and Δx must both be I(0) under our assumptions, as will u_t. Hence, by estimating the static regression

$$y_t = \alpha x_t + \epsilon_t \qquad \text{....(22)}$$

these dynamic terms are all contained in the residual ϵ_t. The fact that the OLS estimate of α is super-consistent in such circumstances is truly remarkable.

3.2 Two-Step Estimators.

Once the cointegrating regression has been performed, this essentially parameterises the long-run relationship between the variables. Engle & Granger (1987) then suggest that the lagged value of z_t, the derived estimate of disequilibrium in any period, should be included in the general dynamic model. If cointegration holds, z_t will be I(0), and the dynamic modelling problem is to transform the individually I(1) variables into reasonably orthogonal I(0) regressors. The lagged value of z_t is completely analogous to an error-correction term in the equation. Recent applications of this methodology include Hall (1986), Jenkinson (1986a), Campbell (1986) and Campbell and Shiller (1986).

The alternative approach to estimating the cointegrating vector α is to include an error-correction mechanism in the dynamic model, since, as noted above, error correction and cointegration are equivalent concepts. This is clear from equation (20) above, which can be transformed into the following dynamic model

$$\Delta y_t = \alpha_2 \Delta x_t - (1 - \alpha_1)(y - x)_{t-1} + u_t \qquad \ldots.(23)$$

where the second regressor is the ECM. Unless y and x are cointegrated, the ECM will be I(1), and hence, since Δy_t and Δx_t are both assumed to be I(0), will have an estimated coefficient tending rapidly to zero. In other words, rather than use the static regression as a kind of pre-test of the model, the full dynamic model is formulated and estimated, with the estimate of any cointegrating vector only being derived once a satisfactory representation of the DGP has been found. Of course, the specification of dynamic adjustment processes in economic models is to a considerable extent a matter of discretion, even when broadly agreed rules are being followed. As such, the ECM approach lacks the conceptual and practical simplicity of the static regression approach, but may be considerably more robust. In fact, Banerjee et al (1986) find that the biases in the cointegrating vector are much smaller when the short-run dynamics are jointly modelled with the long-run relationship, providing some support for the dynamic ECM modelling strategy.

3.3 Testing.

Testing for cointegration between a set of time series is simply a test for the existence of a unit root. The analysis of Section 2 follows through entirely, except that instead of searching for unit roots in the individual time series, the tests are for the existence of a unit roots in the residuals, z_t, from the static cointegrating regression. Because the tests are based on constructed regressors, the critical values obtained for the previous case have to be adjusted upwards, otherwise the test will reject the null too often. If we cannot reject the null hypothesis of a unit root in the residuals, then these 'equilibrium errors' are themselves non-stationary, and cannot be relied upon to move the system systematically back towards equilibrium. In these circumstances cointegration could not be established and hence considerable statistical doubt would be cast upon the theoretical equilibrium.

In actual applications, the Cointegrating Regression Durbin-Watson (CRDW) test, suggested by Sargan and Bhargava (1983), and discussed briefly in section 2.2, has proved extremely popular with researchers. That is,

$$\mathrm{CRDW} = \frac{\Sigma(\hat{u}_t - \hat{u}_{t-1})^2}{\Sigma \hat{u}_t^2} \qquad \ldots.(24)$$

where $\hat{u}_t$ denote the OLS residuals from the cointegrating regression. However, special problems exist with this test. Firstly, whereas in testing for integration (when there are no regressors other than a constant) the critical values of the test, as reported in Sargan and Bhargava (1983), are exact, the test statistics for cointegration depend upon the number of regressors in the cointegrating equation, and only bounds on the critical values are available. This is because, as in the case of the standard D-W test (which is based upon a null hypothesis of white noise residuals), the exact critical value for the D-W statistic is itself a function of the data generation process. The bounds on the test provide a benchmark, and can be used to accept the null hypothesis of no cointegration, but they become rather wide apart as the number of regressors is increased. Without the addition of, for example, the Imhof Routine (1961) to standard regression software, inference will be impossible whenever the value of the DW statistic falls between the bounds. It is possible to compute exact critical values using Monte Carlo methods for a given DGP, an example of which are those reported in Engle and Granger (1987) for a white noise DGP, but these values are not generally applicable to other experiments, and should be very carefully interpreted as the basis of cointegration tests.

An alternative approach suggested earlier is to test for cointegraton using the long run solution in the autoregressive distributed lag model. If the error correction term is restricted as in (23) for theoretical reasons (e.g. log consumption and log income should have a cointegrating slope of unity) then the t-ratio of the coefficient of this term is a useful statistic. Banerjee et al (1986) show that this t-test has about the correct size at the 5% level, although the results of Evans and Savin (1981) suggest that this is not true at other levels. When the level terms are left unrestricted, non-parametric tests, based on deviations of the computed long-run solution, seem a fruitful approach. Some Monte Carlo evidence in Banerjee et al (1986) suggests that the power of these tests is higher than the power of the test based on the static regression. One explanation for this may be the smaller biases obtained using the dynamic modelling approach.

All the other tests described in Section 2 can be used to test for the stationarity of the residuals from the cointegrating regression, but in practise the Dickey-Fuller and Augmented Dickey-Fuller tests have proved most popular. Of course, the choice of the lag structure in ADF tests is still to a great extent *ad hoc*, and different results can be obtained by changing the length of the autoregression, which suggests that greater use should be made of the non-parametric tests described in the Appendix.

At this stage it is important to emphasize that, given the fragility of the tests for cointegration, simple auxiliary tests may be interesting. Granger and Weiss (1983) suggest increasing (or decreasing) the coefficients of the cointegrating vector by, say, 10% and then examine whether the corresponding sum of squares is much larger than for the chosen cointegrating vector. The intuition of this additional check is clear, since only using the latter should the variance be finite, and so it should be easily distinguishable from other cases.

3.4 Functional Forms.

The actual functional form of the cointegrating regression, or ECM, is normally dictated by economic theory. However, what inferences are valid on completion of a set of cointegration tests? Consider first what the inability to find a cointegrating vector, or significant ECM, might imply. It may, of course, simply be that the theoretical equilibrium is without statistical foundation. On the other hand, it may be that a crucial I(1) variable has been omitted from the analysis, which if added to the model would generate I(0) residuals. The temptation of the researcher is to continue adding variables until stationarity of the residuals is achieved. Whilst such general models may be necessary to establish cointegration, the parsimony of the relationship should then be questioned. The t-ratios of the variables in the cointegrating regression will be badly biased, given the degree of autocorrelation of the residuals, but if the autocorrelation is positive, we know that t-ratios are biased upwards. On what criterion, then, should variables be included or excluded from the equilibrium, since it is perfectly possible that some subset of the variables is cointegrated? A low t-ratio will be suggestive, but standard tables cannot be used. There is in fact little alternative to testing all subsets of the variables for cointegration, and only if all these tests are rejected can the researcher be sure that each variable is germane to the relationship. Thus, the discovery of a cointegrating vector should signal the start of a further series of tests.

The choice of functional form is also, in practice, an important consideration. It can be proved that

(i) cointegration implies Granger-Causality

(ii) cointegration in levels implies cointegration in logs

and (iii) cointegration in logs does not imply cointegration in levels. If theory is used to select functional form (rather than, for example, Box-Jenkins techniques), then the use of cointegration methods to evaluate the validity of equilibrium predictions of theories can yield inconclusive results (see, for example, the debate between Jenkinson (1987) and Nickell (1987) regarding the existence of NAIRUs). Intuition suggests that there will always exist some *non-linear* combination of I(1) series that will be I(0), which, if it were true, would have important implications for the use of cointegration analysis in modelling, although much of the simplicity and strength of the original linear case would be lost (see, for example, Escribano (1986)).

3.5 Common Trends.

An alternative way to approach the existence of cointegration is based upon the idea of common trends (see, for example, Stock and Watson (1986), Phillips and Ouliaris (1986), King et al (1987), Aoki (1987)). Suppose that we have a DGP as considered in Section 3.1 equations (17), with $\rho = 0$ without loss of generality. If we add the following process

$$w_t = \gamma x_t + \zeta_t \qquad \ldots.(25)$$

then the vector $(1, \gamma^{-1}\alpha)$ in the regression of y_t on z_t is a cointegrating vector, since it eliminates the presence of x_t. We then say that y_t and w_t have a common trend of x_t. If a third process of the same form is added, there are at most two linearly independent cointegrating vectors. Thus, in general, with N series and r common trends, there are at most (N - r) cointegrating vectors. It is also possible to test for uniqueness of the cointegrating vector. The test basically consists of checking that no subset of regressors is cointegrated in the cointegrating relationship (see Gourieroux et al (1985)).

This approach suggests that multivariate autoregressions of the form:

$$Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + \epsilon_t \qquad E(\epsilon_t) = 0, \quad E(\epsilon_t \epsilon_t') = \Omega \qquad \ldots.(26)$$

should be considered, where Y denotes an n-vector of random variables. This can be rewritten as:

$$\Delta Y_t = -BY_{t-1} - \sum_{j=1}^{p-1} C_j \Delta Y_{t-j} + \epsilon_t \qquad \ldots.(27)$$

where $B = (I - \sum_{i=1}^{p} A_i)$ and $C_j = \sum_{s=1}^{p-j} A_{j+s}$

Diagonalising B such that $P^{-1}BP = \Lambda$, the transformed system can be written

$$\Delta Y_t^* = -\Lambda Y_{t-1}^* - \sum_{j=1}^{p-1} C_j \Delta Y_{t-j}^* + \epsilon_t^* \qquad \ldots.(28)$$

where $Y_t^* = P^{-1}Y_t$ and $\epsilon_t = P^{-1}\epsilon_t$. Then we can test for the number of common trends by testing how close the largest eigenvalue of Λ is to zero, followed by the next largest, and so forth. Dickey and Fountis (1987) show that tests of the form $T\hat{\lambda}$ can be compared with the critical values in Table 8.5.1 of Dickey and Fuller. The natural corollary to the existence of common trends is that there are linear combinations of the regression coefficients in (26) which are $O_p(\sqrt{T})$ consistent, and are asymptotically normally distributed, a result which was first conjectured by Sims (1978) and later formally proved by Phillips and Ouliaris (1986).

The implications of this result are very interesting. Take, for instance, the case of testing whether consumption follows a random walk, when income and consumption are I(1). The test for parameter exclusion in the regression of the change in consumption on the lagged level of income and consumption will have the ordinary F distribution, if they are cointegrated, but will have a non-normal limiting distribution otherwise (see, for example, Mankiw and Shapiro (1985) and Banerjee et al (1987b)).

Finally, taking advantage of the framework used to interpret the existence of common trends, we will briefly discuss the notion of cointegration in trends and in variance (see Escribano (1987)). Consider, for example, the following DGP

$$\begin{aligned} y_t &= \mu + \varphi t + \beta x_t + e_t \\ w_t &= \mu' + \varphi' t + \gamma x_t + v_t \end{aligned} \qquad \ldots.(29)$$

where $\Delta x_t = \epsilon_t$. Then, as before, the vector $(1, \gamma^{-1}\beta)$ provides cointegration in variance, since it eliminates x_t which has a trend in variance. But unless $(\gamma\mu - \beta\mu') = 0$ and $(\gamma\varphi - \beta\varphi') = 0$ there will not be cointegration in means (trends). Therefore, the relevant concept of cointegration, in the framework so far discussed, is that of variance. So in order to decide if a set of series is cointegrated in variance, we first have to detrend in mean, or include a drift and/or trend in the cointegrating relationship. Alternatively these nuisance parameters could be excluded from the cointegrating regression and test for their presence jointly with integrated variance in the corresponding residuals. If the first test, i.e. the exclusion of the constant and trend, is accepted (rejected) and the second test, i.e. the existence of a unit root, is accepted (rejected) there will be cointegration in trends but not in variance (no cointegration in trends but cointegration in variance). The use of non-parametric tests, corrected for the existence of generated regressors, seems another fruitful testing approach which needs to be developed further.

IV. CONCLUSIONS.

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 (in the spirit of Hendry; see, for example, Hendry (1986)) cointegration suggests that models can be significantly improved by introducing, and allowing the data to parameterise, equilibrium conditions suggested by economic theory. Furthermore, the putative existence of such long-run equilibrium relationships can, and should, be tested, using the tests for unit roots discussed in this paper.

Appendix: Non-Parametric Tests for Unit Roots.

The basic idea of this non-parametric approach is quite appealing. The derivation of the statistics such as (6) and (7) highlights the way in which the ratio $\sigma^2/\sigma_\epsilon^2$ affects the shape of the distribution. It is then possible to find an affine transformation of the various statistics which eliminate the dependence of the limiting distribution on the nuisance parameter $\sigma^2/\sigma_\epsilon^2$, accomplished in such a way that the transformed statistics converge to the same random variable as do the untransformed statistics when the errors are iid, i.e. when $\sigma^2/\sigma_\epsilon^2 = 1$. This implies that the critical values of the transformed statistics are the same as those tabulated by Dickey and Fuller.

A simple example will help to understand the procedure. From (6), we look for a transformation such that

$$AT(\hat{\alpha}) + B = \frac{1/2[W(1)^2 - 1]}{\int_0^1 W(t)^2\, dt} \qquad \text{....(A1)}$$

that is, in this case

$$A = 1 \quad \text{and} \quad B = -\frac{1/2[\sigma^2 - \sigma_\epsilon^2]}{\sigma^2 \int_0^1 W(\,)^2\, dt} \qquad \text{....(A2)}$$

Using (A2) and consistent estimates of σ^2 and σ_ϵ^2 , which will be discussed later, we find a consistent estimate of B, that is

$$\hat{B} = -\frac{1/2[\hat{\sigma}^2 - \hat{\sigma}_\epsilon^2]}{T^{-2}\ \Sigma y_{t-1}^2} \qquad \text{....(A3)}$$

such that $T\hat{\alpha} + \hat{B}$ has the same asymptotic critical values as those tabulated by Dickey and Fuller. Similar arguments follow for those tests which are based upon the t-ratio of $\hat{\alpha}$. Since the latter have proved to be more powerful tests than the former, we will concentrate on testing through t-statistic from now on. Rows (a),(b) and (c) in Table 2 present the corresponding transformed t-statistics for the three unrestricted models shown in Table 1. Therefore, these statistics provide a relatively easy way to implement tests of hypotheses of a unit root with possibly heterogeneously and dependently distributed data. However, an important caveat to bear in mind is that the previous equivalence is asymptotic in Tables 8.5.1 and 8.5.2 in Fuller (1976), whereas the finite sample counterparts are not the same. This implies that when dealing with relatively small samples the transformations are not adequate, and unless there is strong evidence of a moving average error term, we advise the extended regression and the Augmented Dickey-Fuller test.

The next step in the test implementation consists of discussing the choice of consistent estimates for σ^2 and σ_ϵ^2. The residual variance $\hat{\sigma}_\epsilon^2$ in the unrestricted models provide consistent estimates of σ_ϵ^2 except in the case where the unrestricted model does not contain a drift, and the true DGP is a random walk with drift. To consistently estimate σ^2, it is important to notice that it is equivalent to $2\Pi s(0)$, s(0) being the spectral density function at zero frequency. Newey and West (1987) have proposed a simple estimate which uses a triangular smoothing window. The estimate is

$$\hat{\sigma}^2 = T^{-1}\{\sum_1^T \hat{\epsilon}_t^2 + 2\sum_{k=1}^m [1 - \frac{k}{m+1}] \sum_{t=k+1}^T \hat{\epsilon}_t \hat{\epsilon}_{t-k}\}$$

The choice of the truncation lag k, i.e. the suspected number of non-zero autocorrelations, is sometimes suggested by the framework in which the test is carried out (see, for example, Corbae and Oularis (1986) for the case of unit roots in spot and forward exchange rates). In general we suggest k ranging from 1 to 8 for quarterly data, and 1 to 24 for monthly data.

Table 2: Summary of Test Statistics.

H_0	Test Statistic ($At_\alpha + B$) A	B
a) α=0 in Case 1 Fuller (1976) Table 8.5.2 (B1)	$\hat{\sigma}_\epsilon/\hat{\sigma}$	$-1/2\ (\hat{\sigma}^2 - \hat{\sigma}^2_\epsilon)[\hat{\sigma}^2 T^{-2}\Sigma y^2_{t-1}]^{-1/2}$
b) α=0 in Case 2 Fuller (1976) Table 8.5.2 (B2)	$\hat{\sigma}'_\epsilon/\hat{\sigma}'$	$-1/2\ (\hat{\sigma}^{2'} - \hat{\sigma}^{2'}_\epsilon)[\hat{\sigma}^{2'} T^{-2}\Sigma \tilde{y}^2_{t-1}]^{-1/2}$
c) α=0 in Case 3 Fuller (1976) Table 8.5.2 (B3)	$\hat{\sigma}''_\epsilon/\hat{\sigma}''$	$-1/2\ (\hat{\sigma}^{2''} - \hat{\sigma}^{2''}_\epsilon)\{\hat{\sigma}^{2''}[(T^{-2}\Sigma\tilde{y}^2_{t-1})(1/12) - T^{-5/2}\Sigma\tilde{y}_{t-1}\tilde{t}]\}^{-1/2}$
d) β=0 in Case 3 Dickey and Fuller (1981)	$\hat{\sigma}''_\epsilon/\hat{\sigma}''$	$1/2\ [1 - (\hat{\sigma}^{2''}/\hat{\sigma}^{2''}_\epsilon)][T^{-5/2}\Sigma\tilde{y}^2_{t-1}\tilde{t}]\ [T^{-4}(\Sigma\tilde{y}_{t-1})(1/12) - T^{-7}(\Sigma\tilde{y}_{t-1}\tilde{t})^2(\Sigma\tilde{y}^2_{t-1})]^{-1/2}$
e) μ=0 in Case 2 Dickey and Fuller (1981)	$\hat{\sigma}'_\epsilon/\hat{\sigma}'$	$1/2\ (\hat{\sigma}^{2'} - \hat{\sigma}^{2'}_\epsilon)[T^{-3/2}\ \Sigma y_{t-1}][T^{-4}\Sigma\tilde{y}^2_{t-1}\Sigma y^2_{t-1}]^{-1/2}$

Note: (ˆ) denotes estimates based on residuals from the unrestricted model in Case 1; (ˆ') denotes estimates based on residuals from the unrestricted model in Case 2; (ˆ'') denotes estimates based on residuals from the unrestricted model in Case 3; (˜) denotes deviations with respect to sample means.

References.

Aoki, M. (1987), "Evidences of Unit Roots and Cointegration in the Time Series for U.S. GNP, M1 and CPI", Working Paper, University of California, Los Angeles.

Banerjee, A.; Dolado, J.; Hendry, D.F. and Smith, G. (1986), "Exploring Equilibrium Relationships in Econometrics Through Static Models: Some Monte Carlo Evidence", Oxford Bulletin of Economics and Statistics 48, pp. 253-277.

Banerjee, A. and Dolado, J. (1987), "Tests of the Life Cycle-Permanent Income Hypothesis in the Presence of Random Walks: An Interpretation Using Nagar Expansions", Nuffield College Discussion Paper 13, Oxford University.

Banerjee, A.; Dolado, J. and Galbraith, J. (1987a), "Orthogonality Tests Using Detrended Data: Interpreting Monte Carlo Results Using Nagar Expansions", *mimeo*, Oxford University.

Banerjee, A.; Dolado, J. and Galbraith, J. (1987b), "Rejecting Orthogonality in Rational Expectations Models: Further Monte Carlo Results for Extended Sets of Regressors", *mimeo*, Oxford University.

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

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

Campbell, J.N. (1986), "Does Savings Anticipate Declining Labour Income? An Alternative Test of the Permanent Income Hypothesis", NBER Working Paper 1805.

Campbell, J.N. and Shiller, R.J. (1986), "Cointegration and Tests of Present Value Models", NBER Working Paper 1885.

Corbae, D. and Ouliaris, S. (1986), "Robust Tests for Unit Roots in Foreign Exchange Market", Economic Letters 22, pp.375-380.

Currie, D. (1981), "Some Long-Run Features of Dynamic Time-Series Models", Economic Journal 91, pp. 704-715.

Davidson, J.; Hendry, D.F.; Srba, F. and Yeo, S. (1978), "Econometric Modelling of the Aggregate Time-Series Relationships Between Consumer's Expenditure and Income in the United Kingdom", Economic Journal 88, pp. 661-692.

Dickey, D.A. (1976), "Estimation and Hypothesis Testing for Nonstationary Time Series", Ph.D Dissertation, Iowa State University.

Dickey, D.A.; Bell, W.R. and Miller, R.B. (1986), "Unit Roots in Time Series Models: Tests and Implications", American Statistician 40, pp. 12-26.

Dickey, D.A. and Fountis, N.G. (1987), "Testing for a Unit Root Nonstationarity in Multivariate Autoregressive Time Series", *mimeo*.

Dickey, D.A. and Fuller, W.A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", Journal of the American Statistical Association 74, pp. 427-431.

Dickey, D.A. and Fuller, W.A. (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", Econometrica 49, pp. 1057-1072.

Engle, R.F. and Granger C.W.J. (1987), "Co-integration and Error Correction: Representation, Estimation, and Testing", Econometrica 55, pp. 251-276.

Escribano, A. (1986), "Identification and Modelling of Economic Relationships in a Growing Economy", Ph.D Dissertation. University of California, San Diego.

Escribano, A. (1987), "Cointegration, Time Co-trends and Error Correction Systems: An Alternative Approach", CORE Discussion Paper 8715, University of Louvain.

Evans, G.B.A. and Savin, N.E. (1981), "Testing for Unit Roots: 1", Econometrica 49, pp. 753-777.

Flavin, M. (1981), "The Adjustment of Consumption to Changing Expectations about Future Income", Journal of Political Economy 89, pp. 974-1009.

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

Gourieriox, G.; Monfort, A. and Renault, E. (1985), "Testing Unknown Linear Restrictions on Parameter Functions", Discussion Paper 8516, CEPREMAP.

Granger, C.W.J. (1983), "Co-Integrated Variables and Error-Correcting Models", University of California San Diego Discussion Paper 83-13.

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

Granger, C.W.J. and Weiss, A.A. (1983), "Time Series Analysis of Error Correcting Models" in Karlin, S.; Amemiya, T. and Goodman, L.A. (eds), Studies in Econometric Time-Series and Multivariate Statistics. New York: Academic Press.

Hall, S.G. (1986), "An Application of the Granger & Engle Two-Step Estimation Procedure to United Kingdom Aggregate Wage Data", Oxford Bulletin of Economics and Statistics 48, pp. 229-240.

Hendry, D.F. (1986), "Econometric Modelling with Cointegrated Variables: An Overview", Oxford Bulletin of Economics and Statistics 48, pp. 201-212.

Imhof, P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variates", Biometrika 48, pp. 419-426.

Jenkinson, T.J. (1986a), "Testing Neo-Classical Theories of Labour Demand: An Application of Cointegration Techniques", Oxford Bulletin of Economics and Statistics 48, pp.241-251.

Jenkinson, T.J. (1986b), "A Note on Alternative Tests for Cointegration when Individual Series are Highly Autoregressive rather than Non-Stationary", *mimeo*, Oxford University.

Jenkinson, T.J. (1987), "The NAIRU: Statistical Fact or Theoretical Straitjacket?", in Cross, R. (ed), Unemployment, Hysteresis and the Natural Rate Hypothesis. Oxford: Blackwell.

King, R.; Plosser, C.; Stock, J.; and Watson, M. (1987), "Stochastic Trends and Economic Fluctuations", Working Paper, University of Rochester.

Mankiw, N.G. and Shapiro, M.D. (1985), "Trends, Random Walks and Tests of the Permanent Income Hypothesis", Journal of Monetary Economics 16, pp. 165-174

Newey, W.K. and West, K.D. (1987), "A Simple Positive Semi-Definite Hetroskedasticity and Autocorrelation Consistent Covariance Matrix", Econometrica 55, pp. 703-708.

Nelson, C.R. and Kang, H. (1981), "Spurious Periodicity in Inappropriately Detrended Time Series", Econometrica 49, pp. 741-751.

Nelson, C.R. and Plosser, G.J. (1982), "Trends and Random Walks in Macroeconomic Time Series", Journal of Monetary Economics 10, pp. 139-162.

Nickell, S.J. (1987), "The NAIRU: Some Theory and Facts", in Cross, R. (ed), Unemployment, Hysteresis and the Natural Rate Hypothesis. Oxford: Blackwell.

Perron, P. (1987), "Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach", Cahier 8650, University of Montreal.

Phillips, P.C.B. (1987a), "Time Series Regression with a Unit Root", Econometrica 55, pp. 277-301.

Phillips, P.C.B. (1987b), "Understanding Spurious Regressions in Econometrics", *forthcoming*, Journal of Econometrics.

Phillips, P.C.B. and Durlauf, S.N. (1986a), "Trends vs. Random Walks in Time Series Analysis", Cowles Foundation Discussion Paper 788, Yale University.

Phillips, P.C.B. and Durlauf, S.N. (1986b), "Multiple Time Series with Integrated Processes", Review of Economic Studies 53, pp.473-495.

Phillips, P.C.B. and Ouliaris, S. (1986), "Testing for Cointegration", Cowles Foundation Discussion Paper 809, Yale University.

Phillips, P.C.B. and Perron, P. (1986), "Testing for Unit Roots in Time Series Regression", Cahier 2186, University of Montreal.

Said, S.E. and Dickey, D.A. (1984), "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order", Biometrika 71, pp. 599-607.

Salmon, M. (1982), "Error Correction Mechanisms", Economic Journal 92, pp. 615-629.

Sargan, J.D. (1964), "Wages and Prices in the United Kingdom: a Study in Econometric Methodology", in P.E. Hart *et al*, Econometric Analysis for National Economic Planning. London: Butterworth.

Sargan, J.D. and Bhargava, A. (1983), "Testing Residuals from Least Squares Regression for Being Generated by the Guassian Random Walk", Econometrica 51, pp. 153-174.

Schwert, G.W. (1985), "Effects of Model Specification on Tests for Unit Roots", *mimeo*.

Shiller, R.J. and Perron, P. (1985), "Testing the Random Walk Hypothesis: Power vs. Frequency of Observation", Economic Letters 18, pp. 381-386.

Stock, J.H. (1984), "Asymptotic Properties of a Least Squares Estimator of Co-integrating Vectors", *mimeo*, Harvard University.

Stock, J.H. and Watson, M. (1986), "Testing for Common Trends", Discussion Paper 1222, Harvard University.

West, K.D. (1986), "Asymptotic Normality when Regressors have a Unit Root", Princeton University Discussion Paper No. 110.

White, H. (1984), Aymptotic Theory for Econometricians. London: Academic Press.

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**: Spanish industrial unemployment: some explanatory factors *(versión inglés)*. El desempleo en el sector industrial español: algunos factores explicativos *(versión español)*.
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**: Parsimony and omitted factors: The airline model and the census X-11 assumptions *(versión inglés)*. Parquedad en la parametrización y omisiones de factores: el modelo de las líneas aéreas y las hipótesis del census X-11 *(versión español)*.
8517 **Ignacio Mauleón**: A stability test for simultaneous equation models.
8518 **José Viñals**: ¿Aumenta la apertura financiera exterior las fluctuaciones del tipo de cambio? *(versión español)*. Does financial openness increase exchange rate fluctuations? *(versión inglés)*.
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.
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.
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.
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 y 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 y Tim Jenkinson**: Cointegration: A survey of recent developments.

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

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