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

By ANINDYA BANERJEE *and* JUAN DOLADO

“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.

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 co-integration 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 focused 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

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

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 regressor have drifts. Stock and West (1987), Sims, Stock and Watson (1988) and Galbraith, Dolado and Banerjee (1987) 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 in this paper is that the mixed results can be 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

¹ A variable x_t ($t = 1, \dots, T$) is said to be integrated of order d , denoted $I(d)$, if $\Delta^d x_t$ admits a stationary ARMA representation. An $N \times 1$ vector of $I(d)$ variables is said to be co-integrated 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 .

not *consistent*² 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 two or more lagged regressors appear as explanatory variables. If the regressors are co-integrated 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 co-integrated 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 normal distribution asymptotically. This illuminates 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 thus see 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”, again using the asymptotic theory for integrated processes. The main result here 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.

²The word is used here specifically with reference to the regression model and not to the probability limits of the coefficient estimates. This terminology was first used by Granger (1982).

The paper is organised as follows. Section 2 presents the MS canonical model of the permanent income hypothesis in order to discuss the properties of the tests. In the light of this, 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 Galbraith, Dolado and Banerjee (1987). 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 co-integrated processes (see Hendry (1986)). Section 6 concludes. 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.³ 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:

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

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,

$$H_t = \sum_{s=0}^{\infty} \gamma^s \cdot E_t Y_{t+s} \quad (2)$$

Y_t is real labour income received in period t . Non-human wealth evolves as,

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

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

³ By “canonical” we mean the familiar Flavin (1981) representation.

with drift:

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

where ε_t is white-noise with variance σ_ε^2 . Then, substituting (1) in (3), we obtain,

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

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

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

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

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

where,

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

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

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

or,

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

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

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

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

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

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

and consequently,

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

that is, permanent income equals current income and also consumption. Notice however that, for the case considered in our paper (namely an infinitely lived individual), if the income process displays drift (see Sargent (1977)), the consumption process will not. 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:

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

$$\begin{aligned} \Delta C_t &= (1 - \gamma)(1 - \lambda\gamma)^{-1}\varepsilon_t + \beta\Delta Y_t \\ &= \beta\mu + \beta(\lambda - 1)Y_{t-1} + (\beta + (1 - \gamma)(1 - \lambda\gamma)^{-1})\varepsilon_t \end{aligned} \quad (13)$$

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

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

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 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,

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

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,

$$\begin{aligned} \Delta C_t &= \alpha + \beta t + \pi[\mu(t - 1) + Y_{t-1}^*] + u_t \\ &= (\alpha - \pi\mu) + (\beta + \pi\mu)t + \pi Y_{t-1}^* + u_t \\ &= k + dt + \pi Y_{t-1}^* + u_t. \end{aligned} \quad (16)$$

This equation mimics the standard practice 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.⁴

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:

<p><i>DGP</i></p> $Y_t = \lambda Y_{t-1} + \varepsilon_t \quad (17)$ $\Delta C_t = e_t$ $E(\varepsilon_t e_t) = \delta_{ts} \rho$ $\sigma_\varepsilon^2 = \sigma_e^2 = 1$	<p><i>MODEL</i></p> $\Delta C_t = k + dt + \pi Y_{t-1} + u_t \quad (18)$
---	--

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 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 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, emphasizing their difference from the standard distributions. We also discuss the cases in which it is valid to use the ordinary critical values.

⁴ 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 suggested by Phillips (1987(a)).

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.

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

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

$$x_t = S_t + x_0 \quad (20)$$

We may define $S_0 = 0$ and set the initial condition with probability one. The distributional results of this section will use the following standardized sums,

$$X_T(t) = (T^{-\frac{1}{2}}\sigma_\varepsilon^{-1})S_{[Tt]} = (T^{-\frac{1}{2}}\sigma_\varepsilon^{-1})S_{j-1} \quad j - 1/T \leq t < j/T, \quad (j = 1, 2, \dots, T) \quad (21)$$

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

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,

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

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 $g(\cdot)$ is any continuous function on $C[0, 1]$, then $X_T(t) \rightarrow W(t)$ implies that

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

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

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

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

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 lemma due to Phillips (1986, 1987(a)) will be used extensively.

Lemma: If the sequences $\{C_t\}$ and $\{Y_t\}$ are generated under the assumptions set in (17), with $\lambda = 1$, then as $T \rightarrow \infty$,

$$(a) \quad T^{-\frac{1}{2}} \sum_1^T Y_t \rightarrow \int_0^1 W(t) dt$$

$$(b) \quad T^{-2} \sum_1^T Y_t^2 \rightarrow \int_0^1 W(t)^2 dt$$

$$(c) \quad T^{-1} \sum_1^T Y_{t-1} e_t \rightarrow \frac{\rho}{2} [W(1)^2 - 1] + N\left(0, \frac{1 - \rho^2}{2}\right)$$

$$(d) \quad T^{-\frac{1}{2}} \sum_1^T t Y_t \rightarrow \int_0^1 t W(t) dt$$

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) and $\lambda = 1$. The proof of the theorem is available on request from the authors.⁵

Theorem 1: For the Model in (18), if (17) characterizes the *DGP*, and $\lambda = 1$ then,

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

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

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

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

$$(V) \quad DW \rightarrow 2$$

$$(VI) \quad R^2 \rightarrow 0$$

⁵The theorem presented here is a special case of the general theory developed by Phillips (1986, 1987(a)).

where,

$$\begin{aligned}
A &= \frac{1}{2}\rho(W(1)^2 - 1) + N(0, \frac{1}{2}(1 - \rho^2)) - \rho W(1) \int_0^1 W(t) dt \\
B &= \int_0^1 tW(t) dt - 0.5 \int_0^1 W(t) dt \\
C &= \int_0^1 V(t) dt - 0.5V(1) \\
D &= \int_0^1 W(t)^2 dt - \left(\int_0^1 W(t) dt \right)^2
\end{aligned}$$

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 relates to part (IV) of the theorem, on hypothesis testing. 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 it would 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,

$$(i) \quad t_{\pi=0} \rightarrow T\hat{\pi} \left[\int_0^1 W(t)^2 dt \right]^{\frac{1}{2}} \quad (25)$$

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

$$(ii) \quad t_{\pi=0} \rightarrow T\hat{\pi}D^{\frac{1}{2}} \quad (26)$$

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.

<i>DGP</i>	<i>MODEL</i>
(1) $Y_t = \mu + \lambda Y_{t-1} + \varepsilon_t$	(27) $\Delta C_t = k + \pi Y_{t-1} + u_t$
(2) $Y_t = \mu + \xi b t + \lambda Y_{t-1} + \varepsilon_t$	(28) $\Delta C_t = k + dt + \pi Y_{t-1} + u_t$

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

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

and

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

Similar results obtain, when $\lambda = 1$, in combination of (2),⁶ using the normalisation matrix $\text{diag}(T^{\frac{1}{2}}, T^{\frac{3}{2}}, T^{\frac{5}{2}})$, for $(\hat{k} - \mu)$, $(\hat{d} - \xi)$ and $\hat{\pi}$. Here,

$$T^{\frac{5}{2}}\hat{\pi} \rightarrow N(0, 180/\xi^2)$$

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 $O_P(T^{\frac{1}{2}})$ 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. So normality follows. Similar considerations 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).

The main result of this section, given by Theorem 1, relates to the case where income has a unit root. In the borderline case, λ is close to⁷ (but strictly less than) 1 in absolute value. When $|\lambda| < 1$, Anderson (1959), for

⁶ This is, however, a rather unlikely case since trend plus unit root implies an ever increasing (decreasing) rate of growth of the variable.

⁷ That is λ is in the neighbourhood of 1, where “neighbourhood” can be rigorously defined.

example, showed that $T^{\frac{1}{2}}\hat{\pi}$ has a normal distribution, emphasising the sharp discontinuity in the limiting distribution when $\lambda = 1$. However, for finite samples, the distributions of the estimator and t -statistic are smooth functions of λ , with no 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,

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

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,

$$S_t = \sum_1^{t-1} \lambda^j \varepsilon_{t-j} \quad (30)$$

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:

$$D_\delta(t) = W(t) + \delta \int_0^1 \exp(t-s)W(s) ds \quad (31)$$

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 $\lambda = \exp(T^{-1}\delta) \approx 1 + T^{-1}\delta$. This allows the formulation of non-central asymptotic theory for this case. In particular, the following reformulation 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.

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

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

where A' , B' and D' are expressions like A , B and D in Theorem 1 where $W(t)$ is substituted by $D_\delta(t)$. β is the marginal propensity to consume as in (13). These formulae highlight the non-centrality of the distribution of the t -statistic and of the co-efficient estimate in the borderline case. Note that when $\delta = 0$, we back to the formulae given by Theorem 1.

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). To illustrate, we choose a version of the model where the trend is absent. There is no loss of generality, 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). 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 Section 3. 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.

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:

$$y_t = \hat{\theta}x_t + \hat{u}_t \quad (34)$$

where,

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

Then,

$$\begin{aligned} E(\hat{\theta}) &= E \left[\frac{H}{E(F)[1 - (F/E(F))]} \right] \\ &\approx E \left[\frac{H}{E(F)} \cdot \left[2 - \frac{F}{E(F)} \right] \right] \end{aligned} \quad (36)$$

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 $[\{F - E(F)\}^2 H] / \{E(F)\}^3$ 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 $[E(s^2 F^{-1})]^{\frac{1}{2}}$. s^2 is the residual variance given by $u'u / (T - 2)$. The expansion proposed is of $O_p(T^{-2})$ and is currently not available in the literature.

Next, using 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,

$$\begin{aligned} E(1) &= -(2i'H\Omega i) / \zeta^2(T-1)^2 \\ E(2) &= -(T-1)^2 \zeta^{-4} [g - 2(T-1)^{-1} i'H\Omega i] \\ E(3) &= (T-1)^{-1} \zeta^{-4} [(T-1)^{-2} \{\text{trace}(\Omega) i'H i + 2i'\Omega H i\} \\ &\quad - 3(T-1)^{-3} (i'Hi)(i'\Omega i)] \\ E(4) &= [(T-1)\zeta^2]^{-\frac{1}{2}} \\ \zeta^2 &= (T)(T-1)/2 - (i'\Omega i)(T-1)^{-2} \end{aligned}$$

where,

$$\begin{aligned} H &= E(Y_{-1}\varepsilon') \\ \Omega &= E(Y_{-1}Y_{-1}') \\ g &= E(\varepsilon'Y_{-1}Y_{-1}'Y_{-1}) \end{aligned}$$

and

$$Y_{-1} = (Y_1, Y_2, \dots, Y_{T-1}), \quad \varepsilon = (\varepsilon_2, \varepsilon_3, \dots, \varepsilon_T).$$

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

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

where,

$$\begin{aligned} E(5) &= -(T-1)^{-2} \sigma_Y^{-4} \cdot g \\ E(6) &= [(T-1)\sigma_Y^2]^{-\frac{1}{2}} \\ \sigma_Y^2 &= (T(T-1))/2. \end{aligned}$$

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; proofs are available 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 $T = \{50, 200\}$, $\lambda = 1$ and

$\rho = \{1.0, 0.8, 0.5\}$. The simulation study of MS (1985) for the unit root case contains a trend. A companion paper (see Banerjee, Dolado and Galbraith (1987)) 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 $\text{Prob}(|\tau| \geq \psi) = \alpha$.

The results are contained in Tables 1A and 1B. The top 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 $\rho = 0.5$ are simply those for $\rho = 1$ divided 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 $\rho = 1$ 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

TABLE 1A
Mean t -ratios and Critical Values $T = 50$

$\rho \rightarrow$	1.000	0.800	0.500
$\alpha = 0.01$	-1.206	-0.965	-0.603
	-3.590	-3.360	-2.970
	-3.620	-3.380	-3.010
$\alpha = 0.05$	-1.206	-0.965	-0.603
	-2.910	-2.680	-2.280
	-2.890	-2.650	-2.280
$\alpha = 0.10$	-1.206	-0.965	-0.603
	-2.580	-2.360	-1.920
	-2.510	-2.270	-1.900

TABLE 1B
 $T = 200$

$\rho \rightarrow$	1.000	0.800	0.500
$\alpha = 0.01$	-1.150	-0.920	-0.575
	-3.530	-3.260	-2.920
	-3.480	-3.250	-2.900
$\alpha = 0.05$	-1.150	-0.920	-0.575
	-2.860	-2.610	-2.240
	-2.800	-2.570	-2.220
$\alpha = 0.10$	-1.150	-0.920	-0.575
	-2.530	-2.320	-1.890
	-2.430	-2.200	-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.

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 this 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.⁸ 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 $\alpha = 0.95$, the Monte Carlo critical values for $\rho = 1$ 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 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

We now 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 best 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 co-integrated $C(1, 0)$ both the RHS and LHS of the equation would have the same order of integration, $I(0)$.

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

⁸ The critical values of the t -distribution for the two sample sizes are as follows:

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

savings S_t in period t as,

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

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,

$$S_t = -\left(\gamma \cdot \mathbb{E}_t \Delta Y_{t+1}\right)(1 - \gamma L^{-1})^{-1} = -\gamma \cdot \sum_{s=0}^{\infty} \gamma^s \mathbb{E}_t \Delta Y_{t+s+1} \quad (38)$$

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 with US data (see Campbell and Deaton (1987)), then

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

which is $I(0)$, apart from a constant. Therefore disposable income and consumption are co-integrated with co-integrating 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.

<i>DGP</i>	<i>MODEL</i>
$Y_t = C_t + s_t$	$\Delta C_t = k + dt + \pi_1 C_{t-1} + \pi_2 Y_{t-1} + u_t \quad (41)$
$\Delta C_t = e_t$	
$E(s_t, e_t) = \delta_s \rho$	
$\sigma_e^2 = \sigma_s^2 = 1$	

(40)

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:

$$\begin{aligned} \text{(I')} \quad & F_{\pi_1=\pi_2=0} \rightarrow 1 + (G - 12f)^{-1} \cdot E \\ \text{(II')} \quad & t_{\pi_1=0} \rightarrow N(0, 1) \\ \text{(III')} \quad & t_{\pi_2=0} \rightarrow N(0, 1) \end{aligned}$$

where

$$\begin{aligned} E &= (1/2) \cdot [V(1)^2 - 1] - V(1) \cdot \int_0^1 V(t) dt \\ G &= \int_0^1 V(t)^2 dt - \left(\int_0^1 V(t) dt \right)^2 \\ F &= \int_0^1 tV(t) dt - (1/2) \cdot \int_0^1 V(t) dt \end{aligned}$$

The theorem shows 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.⁹ It is also easy to prove that if the model is reparameterised as,

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

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

⁹ It can also be proved that $T^{-1/2}\hat{k}$ and $T^{-1/2}\hat{d}$ converge to functionals of the Wiener process, as in Theorem 1.

$E(\Delta C_t) = 0$ 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} .

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

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 (1988) 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 co-integrating 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 $T^{\frac{1}{2}}$. 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 co-integrating 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 $T^{\frac{1}{2}}$.

There is an apparent puzzle here. 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 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 . $T^{\frac{1}{2}}\hat{\gamma}_2$ is asymptotically normally distributed. Then,

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

Hence,

$$T^{\frac{1}{2}}\hat{\pi}_1 \rightarrow -T^{\frac{1}{2}}\hat{\gamma}_2 \quad (44)$$

$$T^{\frac{1}{2}}\hat{\pi}_2 \rightarrow T^{\frac{1}{2}}\hat{\gamma}_2 \quad (45)$$

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

$$T^{\frac{1}{2}}\hat{\gamma}_1 = T^{\frac{1}{2}}\hat{\pi}_1 + T^{\frac{1}{2}}\hat{\pi}_2 \quad (46)$$

Although $\{T^{\frac{1}{2}}\hat{\pi}_i\}$ are asymptotically normally distributed, their variance covariance matrix is singular, as can easily be seen from (44) and (45). This makes the distribution of $T^{\frac{1}{2}}\hat{\gamma}_1$ 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)

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

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

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

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

To illustrate the finite sample performance of the previous asymptotic results, we report the results of a small-scale Monte Carlo study which concentrates on simplified versions of models (41) and (42). The trend is set at zero since 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 $\rho = \{0, 1\}$ 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 $\hat{\gamma}_1$

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.

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 Galbraith, Dolado and Banerjee (1987) 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. Monte Carlo simulations confirm our results. 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 confirmed by Monte Carlo experimentation. However, the F -test is non-standard. In particular, when considering rationality tests we believe that the extent of over-rejection may have been overstated.

*Jesus College, Oxford
Bank of Spain*

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 ε and Y_{-1} denote, respectively, column vectors of dimension $(T-1)$ of the form

$$\varepsilon = (\varepsilon_2, \varepsilon_3, \dots, \varepsilon_T)', \quad Y_{-1} = (Y_1, Y_2, \dots, Y_{T-1})'.$$

Let i be a conformable unit vector.

Define the following expectations:

$$H = E(Y_{-1}\epsilon')$$

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

$$g = E(\epsilon'Y_{-1}Y'_{-1}Y_{-1})$$

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

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

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

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

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

E) $i'\Omega Hi = \rho \left[\frac{(T-2)(T-1)T}{3} \left[T - \frac{1}{2} \right] \right. \\ \left. - \frac{T}{2} \left[\frac{(T-1)(2T-1)T}{6} - 1 \right] \right. \\ \left. + \frac{(3T-1)}{6} \left[\frac{T(T-1)}{2} - 1 \right] \right. \\ \left. + \frac{1}{6} \left[\frac{T^2(T-1)^2}{4} - 1 \right] \right]$

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

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. (1987): "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*, Vol. 55, pp. 1249–1273.
- 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): “Co-integration: A Survey of Recent Developments”, Applied Economics Discussion Paper Series (Oxford), D.P. No: 39 and forthcoming in *Economic Perspectives*.
- 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.
- GALBRAITH, J. W., DOLADO, and BANERJEE, A. (1987): “Rejections of Orthogonality in Rational Expectations Models: Further Monte Carlo Results for an Extended Set of Regressors”, *Economics Letters*, Vol. 25, pp. 243–247.
- GRANGER, C. W. J. (1981): “Some Properties of Time Series Data and Their Use in Econometric Model Specification”, *Journal of Econometrics*, Vol. 16, pp. 121–130.
- 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 Co-integrated Variables: An Overview”, *Oxford Bulletin of Economics and Statistics* (Special Issue on Econometric Modelling with Co-integrated 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*, Vol. 33, pp. 311–340.
- — (1987(a)): “Time Series Regression with a Unit Root”, *Econometrica*, Vol. 55, pp. 277–302.
- (1987(b)): “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.
- SARGENT, T. J. (1977): “Observations on Improper Methods of Simulating and Teaching Friedman’s Time Series Consumption Model”, *International Economic Review*, Vol. 18, 445–462.
- SIMS, C., STOCK, J. H. and WATSON, M. (1988): “Inference in Linear Time Series with Some Unit Roots”, (typescript)
- 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.