This is a postprint version of the following published document:

EXPLORING EQUILIBRIUM RELATIONSHIPS IN ECONOMETRICS THROUGH STATIC MODELS: SOME MONTE CARLO EVIDENCE*

Anindya Banerjee, Juan J. Dolado, David F. Hendry, Gregor W. Smith

INTRODUCTION

This paper investigates the properties of estimators of claimed long-run relationships between integrated processes based on static models. The relationships involved are asserted to have the long-run or equilibrium property that deviations from them are bounded (in a statistical sense). This assertion is tested, and the associated coefficients are estimated, by fitting static regressions. Our concern is to assess the relevance for econometric practice of recent asymptotic theory in this area.

This task is a topical one. Hypothesized long-run relationships such as the quantity theory and the Fisher effect are familiar in macro-economics. Lucas (1980), Whiteman (1984), Summers (1983, 1984), McCallum (1984) and others have debated the usefulness of estimating these relationships using static models. The theory of cointegration developed by Granger and Engle (1985) shows how tests for the existence of equilibrium relationships can be constructed using these models. Our interest is primarily in their suggestion that static ordinary least squares regressions in some cases also may be used to parameterize such relationships. Thus we mainly consider a simple data generation process (we shall use the abbreviation DGP), discussed by Granger and Engle, in which the null hypothesis is that two time series are cointe-grated. Using Monte Carlo simulation and asymptotic approximations we find that the biases in static estimates of the equilibria built into such DGPs may be large and may decline slowly as the sample size increases. As a by-product of the Monte Carlo experiments we examine the performance of several tests for cointegration. We use the asymptotic distributional theory developed by Anderson (1959), Evans and Savin (1981, 1984), White (1959) and Phillips (1985a) to simplify the task of interpreting the results. In addition, one of Phillips's findings is that the limiting distributions of various statistics used in regression analysis are analytically non-normal when there is a unit root, a result which was known through Monte Carlo experiments (see White (1958), Dickey

1

and Fuller (1979) and Evans and Savin). This finding also helps reinterpret some studies of the consumption function.

The paper is organized as follows. In Section I we begin by setting out the theory to which these findings relate, regarding cointegration theory as a link between the methods of econometricians and of time-series analysts. In Section II the simple DGP proposed by Granger and Engle is examined. Simulation evidence of the bias in static estimates of the cointegrating vector is summarized in a response surface. Section III offers similar evidence for data generated by an error correction mechanism (denoted ECM). Here we also study various tests for cointegration. Section IV briefly discusses the applicability of the approach to tests of the random-walk model of aggregate consumption. In general we find that drawing inference about parameters from static models may be hazardous, a result which we elaborate in a brief concluding section. The appendix considers some asymptotic distribution theory in the presence of unit roots which helps to interpret the Monte Carlo results.

I. SPURIOUS REGRESSION, DYNAMIC ECONOMETRICS, AND COINTEGRATION THEORY

It is a well-known empirical fact that many macroeconomic time series are typically non-stationary, as indicated by the high serial correlation between successive observations particularly when the sampling interval is small. Observing that changes in the series are small compared to the levels confirms this fact. At the same time, though, casual observation tells us that some of these variables tend to trend together. Most of the traditional statistical tests that are used in inference have been developed for stationary, ergodic stochastic processes and the absence of a formal statistical procedure for testing whether the existence of similar trends implies a bounded linear relationship in the levels of several series has induced two different approaches to modelling economic time series.

On the one hand some econometricians have traditionally disregarded this issue and run static models in levels. This practice has been criticised by time-series analysts as being inconsistent with most data and giving rise to spurious inferences (see Yule (1926), Granger and Newbold (1974), Granger (1981)). The simple idea behind this criticism is that if no bounded combination of the levels exists then the error term in the regression must be non-stationary under the null hypothesis so that known distributional results do not apply. Given that Yule had proved that the R^2 of the regression in unrelated, non-stationary series tends to unity nothing could be learned by examining that statistic in this environment. Granger and Newbold pointed out that the interest should therefore centre on the Durbin–Watson statistic which would tend to zero since the residuals also would be non-stationary. They

2

suggested the rule $R^* > dw$ to define and discriminate against spurious regressions. The possibility of introducing dynamics through automatic residual serial correlation procedures turned out to provide no escape; the usual t-ratios still reject the null too often, thus again suggesting that non-normal distribution theory might be relevant to series containing a unit root.

On the other hand, in the light of the likely spurious inferences discussed above, time-series analysts trained in the Box–Jenkins approach advocated differencing and prewhitening the series prior to estimating multivariate models. By being extremely faithful to the properties of data their models could be used to describe only relationships between changes in variables. This approach disregards the potentially important, long-run relationships among the levels of the series to which the hypotheses of economic theory are usually taken to apply.

ECM models provide a way of combining the advantages of these two approaches. In this type of model the dynamics both of short-run (changes) and long-run (levels) adjustment processes are modelled simultaneously. In particular, the idea of incorporating the dynamic adjustment to steady-state targets in the form of error correction (i.e. a parsimonious parameterisation of the bounded, linear combination of the levels of the variables) suggested by Sargan (1964) and developed by Davidson et al. (1978) seems to have introduced a useful approach to modelling dynamics, with successful empirical applications. One complication in this approach must be noted. Take a simple, canonical ECM model containing an error correction term along with the regressand and remaining regressors in stationary changes. If no bounded, linear combination exists among the levels one would expect the coefficient of the error correction term to converge rapidly to zero. However, under this null hypothesis the ordinary t-ratio for that coefficient is no longer valid, as Theorem 4 in the appendix shows. As shown below, at the 5 per cent level, the power of the usual t-ratio is satisfactory in Monte Carlo experiments but this finding seems to be peculiar to that critical value, the only one for which the asymptotic normal distribution and the true distribution give similar results. For any other significance level inference based on the usual t-ratio would have been wrong. Therefore, it is essential either to correctly calibrate the t-test on the ECM, or to check for the existence of a bounded relationship before estimating the dynamics in an ECM model.

Granger (1983) in the theory of cointegration established a unified framework for the analysis of ECMs and of time series in which the variables stochastically trend together. Granger and Weiss (1983) and Granger and Engle (1985) developed the theory further. We summarize it below.

Consider a vector time series $\{x_t\}$ $t = 1, \ldots, T$ of observations on economic variables. An element $z_t = x_{it}$ often can be described by an

ARIMA model of the form:

$$\alpha(L)\Delta_j^k z_t = \tau + \beta(L)\epsilon_t \tag{1}$$

where $\{\epsilon_t\}$ is (empirically) white noise with a finite (though rarely homoskedastic) variance σ_{it}, $\Delta_j = (1 - L^j)$, $L^n z_t = z_{t-n}$, τ is the constant 'drift', $k \geqslant 0$ and $\alpha(L)$, $\beta(L)$ are finite-order polynomials in L with roots outside the unit circle and no roots in common. Granger has called such a variable integrated of order k, denoted by $I(k)$, so that differencing k times is required to produce a series with finite variance. Generally, k is zero, one or two. If a sample from a time series has a reasonably large value of T (e.g. 80 or more) and $k \geqslant 1$, then the observed variance of $\{z_t\}$ will be extremely large relative to the variance of $\Delta^k z_t$, which (as mentioned above) is a well-known phenomenon in economics: for example, Working (1934) argued for the random walk process being a useful benchmark for empirical economic time series comparisons.

Consider now two series $\{x_{it}\}$, $\{x_{1t}\}$ which are respectively $I(k_i)$ and $I(k_1)$. If $k_i \neq k_1 \geqslant 1$ then no linear combination of x_{it} and x_{1t} can have finite variance since $(x_{it} \pm \lambda x_{1t})$ is at least $I(|k_i - k_1|)$. Thus, $\forall \lambda$, x_{it} and x_{1t} differ by a 'disturbance' with unbounded variability so that no relationship can be said to exist between them. Of course, if $k_i = 2$, $k_1 = 1$ (say), then $\Delta_1 x_{it}$ and x_{1t} *may* be related, but x_{it} and x_{1t} cannot be. Granger defines[1] two series to be cointegrated if they are both $I(d)$, $d \geqslant 1$ and there exists a linear combination of the series which is $I(0)$. If such a linear combination exists, in the bivariate case it must be unique (up to a scale factor) since adding or subtracting any proportion of either x_{it} or x_{1t} to the error will make it $I(d)$. Conversely, the existence of cointegrating vectors (i.e. vectors of real numbers linking cointegrated series) must be rare since the same relationship is required between x_{it} and x_{1t} for prolonged periods.

In fact, interesting parallels exist between the concept of cointegrating vectors and (static) equilibrium relationships. If it is asserted in an economic theory that $x_i = \lambda x_{1t}$ then $(1 : -\lambda)$ should be a cointegrating vector. For stationary series this is not a very dramatic insight but if, historically, $\{x_{it}\}$ and $\{x_{1t}\}$ are both $I(d)$, $d \geqslant 1$ then the theory is meaningful only if $\{x_{it} - \lambda x_{1t}\}$ has bounded variability. This is a powerful restriction and has testable implications as shown by Granger and Engle (1985) and discussed below.

An important implication of the above is that Δx_{it} is asymptotically negligible relative to x_{it} when x_{it} is $I(d)$, $d \geqslant 1$. Thus, say a complex dynamic relationship links the $I(1)$ series $\{x_{it}\}$ and $\{x_{1s}\}$ ($s \leqslant t$) but these are cointegrated. Granger and Engle provide in a general theorem the various representations of cointegrated processes. Consider the

---

[1] Actually, his definition is more general, but the special case suffices for our exposition.

simple static regression of x_{it} on x_{1t} (or *vice versa*) with no drift factors ($\tau = 0$ in (1)):

$$x_{it} = \lambda x_{1t} + v_t \tag{2}$$

Now v_t contains all the omitted dynamics, but these can be reparameterized purely in terms of Δx_{it-j}, Δx_{1t-m}, and $(x_{it-r} - \lambda x_{1t-r})$ which are all $I(0)$ if cointegrability holds. Thus λ is *consistently* estimated by the regression in (2) despite the complete omission of the dynamics. In fact:

$$\hat{\lambda} = (\Sigma x_{1t}^2)^{-1} \Sigma x_{it} x_{1t}$$
$$= \lambda + (\Sigma x_{1t}^2)^{-1} (\Sigma x_{1t} v_t) \tag{3}$$

Since $\{v_t\}$ is $I(0)$ under cointegrability but $\{x_{1t}\}$ is $I(1)$, if (2) holds (with $\tau_i = 0$)

$$\frac{1}{T} \Sigma x_{1t}^2 = m_{11} \text{ is } 0_p(T)$$

whereas

$$\frac{1}{T} \Sigma x_{1t} v_t \text{ is } 0_p(1) \tag{4}$$

Thus $\hat{\lambda} - \lambda = 0_p(T^{-1})$ and hence $\hat{\lambda}$ converges to λ at a rate of $0_p(T^{-1})$ and *not* at the usual rate of $0_p(T^{-1/2})$. Thus, convergence is very rapid, although the conventional least squares formula for estimating var $(\hat{\lambda})$ is invalid due to the autocorrelation in $\{v_t\}$. The logic generalizes, with more complicated algebra (and even faster convergence), for $\tau_i \neq 0$, and a constant in (2). Stock (1984) gives this super-consistency result and shows that the bias in $\hat{\lambda}$ will be of order $0_p(T^{-1})$. Phillips and Durlauf (1985) provide similar results in a more general framework.

We conclude that if variables are $I(d)$, $d \geqslant 1$ and are cointegrated it should be rather easy to establish the value of λ and hence to numerically parameterize equilibrium economic theories. Indeed, the derived long-run solutions from dynamic models for $I(1)$ variates should essentially coincide with the estimates obtained from the initial static regressions; this provides a check on the validity of the dynamic modelling. Of course, if $(x_{it-r} - \hat{\lambda} x_{1t-r})$ is used as the error correction term then no cross check is possible, but modelling might be easier.

One caveat must be noted. The equilibrium may relate $I(1)$ and $I(0)$ variables if a cointegrating vector exists for a subset of the observables: e.g. if x_{mt} is $I(0)$ but via δx_{mt} (say) influences $(x_{it} - \lambda x_{1t})$ (which is also $I(0)$) then δ *cannot* be consistently estimated from the prior static regression – all of the conventional econometric problems reappear for $I(0)$ variables. However this is not the case, for example, when the

equilibrium inter-relates $I(2)$ variables which cointegrate to $I(1)$ with other $I(1)$ variables.

To test an equilibrium assertion, Granger and Engle propose testing the null that $\{v_t\}$ has a unit root against the alternative that it has a root less than unity. Recent results by Fuller (1976), Dickey and Fuller (1979, 1981), Evans and Savin (1981, 1984), Sargan and Bhargava (1983), and Bhargava (1983) can be used. The first five papers consider the t-ratio of a levels term in a regression in which the remaining variables are in changes. The last two papers consider using the Durbin–Watson statistic, having derived its distribution on the null of a unit root, and provide tables for critical values. Strictly these are relevant to λ *known* (or x_{1t} non-stochastic) and various technicalities remain to be resolved. In the simulations which follow we use the lower bounds for the Durbin–Watson statistic provided by Sargan and Bhargava, confirming that nominal and actual sizes coincide by means of preliminary simulations. Our preference for this statistic (as opposed to the t-test) stems from the invariance of its distribution to nuisance parameters such as a constant.

The idea of cointegration tends to make the ECM model generic in the sense that the validity of a static equilibrium implies the validity of this dynamic representation. In theory the error correction term can be parameterized through a simple, auxiliary, static regression. The extent to which the bias in this first stage may be important and whether it can be summarized by the other main statistic of the static model — the long-neglected R^2 — are discussed below.

II. THE GRANGER-ENGLE DATA GENERATION PROCESS

A first set of experiments involves the data generation process used by Granger and Engle (1985).[2] This DGP is

$$x_t + y_t = v_t \qquad v_t(1 - \rho_1 L) = \epsilon_{1t} \qquad (5)$$

$$y_t + 2x_t = u_t \qquad u_t(1 - \rho_2 L) = \epsilon_{2t} \qquad (6)$$

$$\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \sim IN\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right]$$

where ϵ_{1t} and ϵ_{2t} are independent, pseudo-normal variates. The reduced form consists of the following two equations:

$$y_t = \frac{2\epsilon_{1t}}{1 - \rho_1 L} - \frac{\epsilon_{2t}}{1 - \rho_2 L} \qquad (7)$$

[2] All simulations were conducted using NAIVE which is part of the AUTOREG library. The stochastic processes begin at zero and no observations are discarded.

6

$$x_t = \frac{\epsilon_{2t}}{1 - \rho_2 L} - \frac{\epsilon_{1t}}{1 - \rho_1 L} \tag{8}$$

Our interest is in the ordinary least squares (OLS) regression of y_t on x_t. From (7) and (8) the OLS-estimate $\hat{\lambda}$ of λ, the potentially cointegrating coefficient, in this static model has plim as follows:

$$\text{plim}(\hat{\lambda}) = -1 - \frac{1}{1 + \dfrac{\sigma_2^2(1 - \rho_1)^2}{\sigma_1^2(1 - \rho_2)^2}} \tag{9}$$

Consider three cases of the DGP given by (5) and (6) above. In Case A, $|\rho_1| < 1$, $|\rho_2| < 1$ so that both x and y are $I(0)$ variables. In Case B, $\rho_1 = 1$, $\rho_2 = 1$ so that both variables are $I(1)$ and are not cointegrated. In Case C, $\rho_1 = 1$, $|\rho_2| < 1$ so that the variables are still $I(1)$ but are now cointegrated. In this last case, the parameters of the model are identified and the cointegrating coefficient has $\text{plim}(\hat{\lambda}) = \lambda = -2$.[3]

For Case B, we conducted experiments with $N = 200$ and $T = 99$, varying σ_1/σ_2, the noise-signal ratio widely. The null hypothesis that there is a unit root in the residuals of the static model holds. Using the Sargan–Bhargava critical values variation of the noise-signal ratio had no effect on the size of the test, which in all cases fell within the relevant binomial confidence interval.

For Case C, the null hypothesis of a unit root in the error dynamics in (6) is false; the two series are cointegrated. Here we are interested in exploring the usefulness of the estimate of the cointegrating parameter. In this case we simulated with 200 replications in the parameter space $s x T x \rho_2$ where $s = \sigma_1/\sigma_2 = (16\ 8\ 4\ 2\ 1\ \frac{1}{2})$, $T = (33\ 66\ 99\ 150\ 199)$, and $\rho_2 = (0.6\ 0.8\ 0.9)$ giving rise to 90 experiments. The range of the noise-signal ratio is very large here; obviously it will be difficult to distinguish between σ_u^2 and σ_v^2 if σ_1 and T are small and σ_2 and ρ_2 are large; for large values of s OLS essentially picks up equation (5) instead of (6).

This problem accounts for our interest in the DGP used by Granger and Engle. Asymptotic theory holds that the static model can be used to parameterize the long-run relationship and that the bias in the estimate of λ from this model should be $0_p(T^{-1})$. How relevant is this result to practice? In the present DGP if T is small and ρ_2 and s are large, then the two equations will be difficult to distinguish. Thus there is a finite-sample problem which is illustrated in the Monte Carlo study. For example if $T = 33$, $s = 2$, and $\rho_2 = 0.9$ the average bias is 0.589. If we double the sample size to 66 the bias is 0.468; clearly it does *not* decline at the theoretical rate. This outcome seems at odds

[3] In (9) as $\rho_2 \to 1$, $\text{plim}(\hat{\lambda}) = -2$ and as $\rho_1 \to 1$, $\text{plim}(\hat{\lambda}) = -1$.

with the theory. To resolve this puzzle we undertake analysis (in the appendix) and Monte Carlo study across a wide range of parameter values.

We calculate the bias in the estimate of λ from the static regression along with its estimated standard error. Also, we continue to apply the Sargan–Bhargava test to reject the null of non-cointegration (which is indeed false in Case C). The test is biased at high values of the noise-signal ratio. Except at such values the power is non-increasing in ρ_2 as one would expect. Figure 1 presents some of the same experimental results, projected in two dimensions. The sample size for the least squares estimation is on the horizontal axis, the bias is on the vertical axis, and the curves are drawn for noise-signal ratios declining down the page from 2.0–0.5. For each ratio and sample size the bias is given by the solid curve for $\rho_2 = 0.9$, by the dashed curve for $\rho_2 = 0.8$, and by the dotted curve for $\rho_2 = 0.6$. The wide, solid line is $10/T$ which is included for comparison with the asymptotic-theoretical rate of decline. The diagram shows that the bias is higher the higher ρ_2. The significance of this characteristic can be tested by means of response surfaces, which we discuss below.

First, we note the slow rate at which the bias vanishes in Figure 1. This suggests limited value for the static model as a source of pre-tests since often when cointegration obtains we do not find it and often when we find it our estimate of λ is very inaccurate. The apparent conflict between the simulation results in the figure and Stock's asymptotic result leads us to examine the asymptotic theory of the least-squares regression estimator where the regression relates random walks. In Theorem 1 of the appendix we find the limiting distribution of the bias and of $(1 - R^2)$ and use this theory to construct a response surface.[4] In approximating the conditional expectation $E(\hat{\lambda} - \lambda \,|\, s, T, \rho_2)$ our main interest lies in exploring the rate at which this bias vanishes.[5] We stress that although this data generation process may not be very interesting economically it is a simple one in which to test the rate at which the bias declines; moreover the properties of tests can be compared with those found by Granger and Engle.

The experimental results in Figure 1 can be summarized in a response surface (see Hendry (1984)) which approximates the conditional expectation of the bias in a simple expression. Our response surface also survives a number of misspecification tests; hence we have some warrant in using it for inference. For example, we can reproduce the asymptotic result that there is no constant term in the response surface so that the bias is $0_p(T^{-1})$. We can also make inexpensive predictions of experimental outcomes elsewhere in the parameter space.

---

[4] We also calculate a Nagar-type expansion for the estimator. This gives rise to the same first term in the response surface as the analysis in the appendix.

[5] See Rothenberg (1984) on the interpretation of moments of approximations.
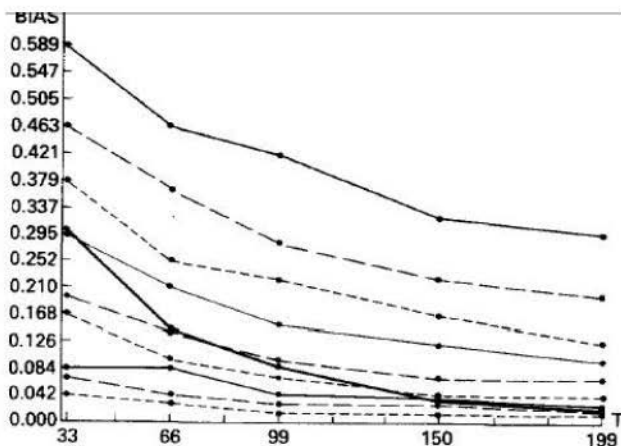
Fig. 1. Plot of bias versus T

The response surface estimated with the data is as follows:

$$\xi(E\hat{\lambda} - \lambda) = 3.99 \cdot \xi(1 - R^2) + 3.04 \cdot s^2 \cdot \xi(1 - \rho_2^2) \cdot (1 - R^2)^2 \qquad (10)$$
$$\phantom{\xi(E\hat{\lambda} - \lambda) =} (0.062) (0.6)$$

$$SE = 0.926 \quad n = 90 \quad R^2 = 0.98 \quad dw = 2.00$$

$$z_1(2,86) = 0.283 \quad z_2(9,79) = 0.299 \quad z_3(3) = 12.2 \quad z_4(2) = 1.6$$

Heteroskedasticity-consistent standard errors are given in parentheses. The functional form is suggested by the analysis of the appendix. The bias and regressors have been standardized by the Monte Carlo-estimated standard deviation of $\hat{\lambda}$, denoted ξ, as a heteroskedasticity transform. Consequently the equation standard error should be approximately unity if the response surface captures the salient features of the underlying conditional expectation.

The stability of (10) can be examined by splitting the sample into halves with $\rho_2 = 0.6$ in one half, $\rho_2 = 0.9$ in the other, and the data generated with $\rho_2 = 0.8$ divided between the two. An ANOVA test gives z_1 where $z_1 \sim F(2,86)$ under the null. A Chow test using randomly selected nine observations gives z_2. In addition, White's (1980) heteroskedasticity test gives z_3 while the Jarque–Bera test for non-normality in residuals gives z_4. The test z_3 suggests that the heteroskedasticity transform has not been completely successful.

The response surface in (10) represents the experimental results fairly well. It also is important to emphasize its consistency with the asymptotic theory of the appendix. The first term follows from the

9

application of Theorem 2 in (A4) in the Appendix with $(1 - R^2)$ being an indicator of the extent of the bias. According to (A4) its coefficient has an expected value of four given the values assigned to the parameters in the DGP. The second term, which is $O(T^{-2})$, follows from the form of the variance of the corresponding $O(T^{-1})$ term in (A4). Both terms contribute to the explanation of large biases whereas the second is important in explaining why the bias does not decline at $O(T^{-1})$: note that $(1 - R^2)$ (which is of $O(T^{-1})$) plays a pivotal role in determining the magnitude of the biases in this DGP. This can be seen most easily from the summary (*) below of (10).

$$(*) \quad (E\hat{\lambda} - \lambda) = 4 \cdot (1 - R^2)\{1 + 0.75 \cdot (1 - \rho_2^2)(1 - R^2)\}$$

As a final exercise with the Granger–Engle DGP we set $\epsilon_{1t} = \rho_3 \epsilon_{1t-1} + e_t$ where $e_t \sim IN(0, \sigma_3^2)$ and then let $\rho_3 \to 1$. This allows us to examine the test for cointegration and the static estimates of relationships between $I(2)$ variables. Some results are given in the table below.

This table can be compared with Figure 1. Estimates of λ from the static model are obviously vastly more reliable for $I(2)$ than for $I(1)$ variables, but the test for a unit root still lacks power and again the rate of convergence to zero of the bias is less than the theoretical rate.

Thus we can conclude this section with a general warning that the biases in estimates of equilibrium relationships found from static models may be large and may decline slowly. However, since $(1 - R^2)$ provides a rough index of the bias in this simple DGP it clearly offers guidance as to the potential unreliability of estimates from static models. Nevertheless, we stress that in multivariate models a high R^2 does *not* imply that each element in the cointegrating vector is estimated with negligible bias (see the applications by Hall and Jenkinson elsewhere in this issue). Finally, the power of the DW test to reject the null of non-cointegration against alternatives close to the unit circle is low. Consequently, we should not be over hasty in discarding claims

TABLE 1
Cointegration Tests with I(2) data
$\rho_2 = 0.9 \quad s = 1 \quad \rho_3 = 0.99$

Bias	ESE	T	Rejection frequency
0.00109	0.00147	33	0.11
0.00002	0.00074	66	0.19
0.00002	0.00074	99	0.43
0.00053	0.00026	150	0.37
0.00004	0.00023	199	0.39

to long-run relationships on this basis alone. Indeed, we must consider the alternative route to modelling the long-run relationship using error correction models and we devote the next section to this aim.

III. AN ERROR CORRECTION MECHANISM

We next examine a more general DGP. We contrast inferences drawn in the strategy of estimating static models with those based on dynamic modelling. Thus a second set of experiments was conducted with the following data generation process:

$$\Delta x_t = \epsilon_{1t} \tag{11}$$

$$\Delta y_t = \alpha_1 \Delta x_t - \alpha_2 (y - x)_{t-1} + \epsilon_{2t} \tag{12}$$

Again the errors are mutually and serially uncorrelated pseudo-normal variates. Here Δy and Δx are always $I(0)$ and they are cointegrated if $\alpha_2 \neq 0$.

Simulations were conducted with 200 replications at sample sizes of 66, 150, and 199. In all cases x_t is $I(1)$ and for α_2 non-zero the endogenous series x_t and y_t are cointegrated. In the simulations T, α_1, α_2, and $s = \sigma_1/\sigma_2$ were varied as shown in the 72 experiments listed in the tables below. Note that the values adopted for α_1 and α_2 give two strongly differing lag distributions in the response of y to x. With $\alpha_1 = 0.5$ and $\alpha_2 = 0.1$ the mean lag is five periods and the median lag is zero while in the opposite case the values are 1.8 and 0.9 respectively.

For each DGP we estimated the following models:

(A) $y_t = a \cdot x_t + v_{1t}$

(B) $\Delta y_t = b \cdot \Delta x_t - c \cdot (y - x)_{t-1} + d \cdot x_{t-1} + v_{2t}$

(C) $\Delta y_t = e \cdot \Delta x_t - f \cdot (y - x)_{t-1} + v_{3t}$

In comparing the three models we are particularly interested in inferences about cointegration based on the DW statistic in model A and on the t-tests of the error correction coefficients c and f respectively in models B and C. We also are interested in the biases in estimates of the cointegrating vector. In B the test $H: c = 0$ is a test for cointegration despite the presence of the unrestricted term x_{t-1}. The modeller knows that both $\{x\}$ and $\{y\}$ are $I(1)$ series; hence model B makes no sense with $c = 0$ and $d \neq 0$. Table 3 shows the powers of the three tests. Table 2 shows the biases in the static model and the unrestricted ECM model under the null of no cointegration ($\alpha_2 = 0$) and under the alternative. Model C is excluded from the comparison since it is correctly specified under the null. We are particularly interested in the first case given the results of Section 2.

11

TABLE 2
Test Rejection Frequencies in ECMs

S		3/1			1/3		
T		66	150	199	66	150	199
(a) $\alpha_2 = 0$ $\alpha_1 = 0.1$							
	A	0.050	0.050	0.050	0.050	0.050	0.050
	B	0.155	0.195	0.215	0.180	0.195	0.135
	C	0.055	0.060	0.055	0.050	0.070	0.055
(b) $\alpha_2 = 0$ $\alpha_1 = 0.5$							
	A	0.050	0.050	0.050	0.050	0.050	0.050
	B	0.205	0.180	0.195	0.160	0.175	0.135
	C	0.050	0.065	0.045	0.060	0.035	0.060
(c) $\alpha_2 = 0.1$ $\alpha_1 = 0.5$							
	A	0.110	0.405	0.435	0.165	0.535	0.575
	B	0.820	0.980	1.000	0.560	0.825	1.000
	C	0.925	1.000	1.000	0.455	0.960	0.995
(d) $\alpha_2 = 0.5$ $\alpha_1 = 0.1$							
	A	1.000	1.000	1.000	1.000	1.000	1.000
	B	1.000	1.000	1.000	1.000	1.000	1.000
	C	1.000	1.000	1.000	1.000	1.000	1.000

Note: (a) and (b) refer to the sizes of the tests. The size of the DW test has been set to 0.05 by interpolation of the relevant Sargan–Bhargava critical values. (c) and (d) refer to the powers of the tests.

TABLE 3
Biases in the Long-Run Coefficient

S		3/1			1/3		
T		66	150	199	66	150	199
(a) $\alpha_2 = 0$ $\alpha_1 = 0.1$							
	A	0.0800	0.0850	0.0930	−0.1399	0.0590	0.0299
	B	0.1305	0.1013	0.1024	−0.1585	0.0841	0.0650
(b) $\alpha_2 = 0$ $\alpha_1 = 0.5$							
	A	0.4956	0.4897	0.4913	0.2562	0.4713	0.4775
	B	0.4384	0.5089	0.5005	0.3022	0.5603	0.6918
(c) $\alpha_2 = 0.1$ $\alpha_1 = 0.5$							
	A	−0.0953	−0.0490	−0.0352	−0.1107	−0.0409	−0.0301
	B	−0.0002	−0.0003	−0.0002	−0.0124	−0.0050	−0.0040
(d) $\alpha_2 = 0.5$ $\alpha_1 = 0.1$							
	A	−0.1842	−0.1001	−0.0821	−0.1848	−0.1060	−0.1164
	B	−0.0346	−0.0076	−0.0065	−0.0222	−0.0271	−0.0304

Note: The biases were computed as $(\hat{a} - \lambda)$ in model A and $(1 + \bar{d}/\hat{c} - \lambda)$ in model B where $\lambda = 0$ for case (a) and (b) and unity for cases (c) and (d).

The implications of these experiments can be summarized as follows.[6] First, when the two series are cointegrated the biases in the estimates of the cointegrating parameter are larger in the static than in the dynamic model. Again, the bias in the static model never vanishes as rapidly as T^{-1}. However, the usual practice in dynamic modelling of adding and testing the presence of an extra term in levels using a conventional t-test as in Case B, is invalid if the series are cointegrated with the true cointegrating parameter contained in the error correction term. Obviously, under the null hypothesis of cointegration, every term in Case B is $I(0)$ except the level term, hence the ordinary t-ratio does not apply. We shall have the case of a variable whose coefficient tends to zero at $0(T^{-1})$ but has a t-test with a non-degenerate distribution different from the ordinary one. The t-tests in the dynamic models are more powerful than the Durbin–Watson test in the static models; this is especially so when the mean lag is large, otherwise it is difficult to get any power comparison as the shorter mean lag case shows.

Second, when the two series are not cointegrated the t-test in the unrestricted model has a larger size than the 5 per cent level used to set the critical values of the test. However, it is remarkable to see that in the restricted model the t-test has about the correct size at the 5 per cent level, although the results in Evans and Savin (1981) suggest this is not true of other levels. Hence, when testing for differencing as a restriction, as opposed to a filter, it seems appropriate to test the significance of error correction terms at a 5 per cent level. Overall, the Durbin–Watson test in the static model usually rejects less often than in both dynamic models, hence the unrestricted version of the dynamic model may sometimes misleadingly suggest that cointegration holds.[7] Note also that under the null of no cointegration the biases are large and quite similar in magnitude to the value of α_1 emphasizing the fact that we might need larger critical values than the ordinary ones to test the null hypothesis H_0: $c = d = 0$ in Model B.

Finally, we note that the use of the leading term in the response surface for the static regression, which may be derived from the results in Theorem 2 in the appendix, again gives a quite accurate summary of the simulation results. This is true even when only three sample sizes are used for each experiment. For example, taking the case $\alpha_1 = 0.1; \alpha_2 = 0.5; s = \frac{1}{3}$, the coefficient of $(1 - R^2)$ according to the formula is -0.07 while that obtained from averaging across the three sample sizes is -0.158. Similarly in the case $\alpha_1 = 0.5$, $\alpha_2 = 0.1$, $s = 3.0$, the values are 1.32 and 1.63 respectively.

---

[6] We draw these inferences after calculating binomial confidence intervals for the rejection frequencies.

[7] Molinas (1986) finds that the power of the DW test can be very low under the null of no cointegration when the series are generated by IMA(1,1) processes.

We devote this section to pointing out briefly some implications of the existence of integrated processes for testing Hall's (1978) hypothesis about the life-cycle/permanent-income theory of consumption. He demonstrated that if consumers are assumed to measure permanent income using rational expectations of future income rather than a distributed lag of past income, then a consequence of life-cycle theory (under various well-known assumptions) is that consumption will follow a random walk. The standard test of the model is to regress the first difference of consumption on lagged information (e.g. lagged income) and test the null hypothesis of a zero coefficient for that variable. However, Mankiw and Shapiro (1985a,b) in two recent papers have argued that if income follows a random walk, as it seems to in the US, then the standard testing procedure is invalid since it is greatly biased towards rejecting 'too often' the null hypothesis. They also offer Monte Carlo evidence to substantiate their claims.

In order to understand the DGP that will be analysed below, it is helpful to explain a rather simple version of the LC/PI theory of consumption by means of the following example.

Consider the simple consumption function,

$$c_t^P = (1 - \delta)\, i_t^P : 0 < \delta \leqslant 1 \tag{13}$$

where c_t^P is consumption, and i_t^P is permanent income. Permanent income is measured using forward-rational expectations as,

$$y_t^P = \sum_0^\infty \delta^k \cdot E_t\, y_{t+k} = (1 - \delta)^{-1} y_t$$

if $\Delta i_t = \epsilon_t$. Assume also that $\Delta c_t = \Delta c_t^P + u_t$.

Then, lagging (13), and subtracting, we obtain

$$c_t = c_{t-1} + (\epsilon_t + u_t) \tag{14}$$

and hence consumption itself follows a random walk whose disturbance is correlated with that of the income process (in fact if $u_t \equiv 0$, the correlation is one).

With two slight modifications, i.e. $\alpha_1 = 0$ and $\mathrm{corr}(\epsilon_{1t}\epsilon_{2t}) = \rho \neq 0$, the bivariate process generating consumption and income corresponds to that analysed in Theorems 3 and 4 in the appendix where now, following the notation there, the following regression is performed.

$$\Delta y_t = \gamma x_{t-1} + \eta_t \tag{15}$$

For this regression, after transforming the system, $\epsilon_{2t} = \rho \sigma_2 \sigma_1^{-1} \epsilon_{1t} + \nu_t$ where $\sigma_\nu^2 = \sigma_2^2(1 - \rho^2)$, we can show that

$$T\hat{\gamma} \rightarrow \frac{\{\rho/2s[W(1)^2-1] + N(0, 1-\rho^2/2s^2)\}}{\int_0^1 W(t)^2 \, dt}$$

and

$$t_{\hat{\gamma}} \rightarrow \frac{\{\rho/2[W(1)^2-1] + N(0, (1-\rho^2)/2)\}}{\int_0^1 W(t)^2 \, dt}$$

Thus we simply confirm that the t-statistic does not have the usual distribution and, therefore, inference based on the usual tables is invalid. Also, $R^2 \rightarrow 0$ and DW $\rightarrow 2$. Note that even if ρ equals one and the normal-distribution term disappears, the t-ratio follows a non-degenerate distribution which differs from the usual standardized normal (which approximates the t-distribution in large samples). Another feature of interest is the dependence of the obtained distribution on ρ and its independence of the variances of the error terms, both results in agreement with the Monte Carlo results reported by Mankiw and Shapiro who compute the corresponding critical values. The intuition behind all this is very similar to that discussed in Theorem 4 in the appendix. Even if there is correlation between the disturbances of the random walk, the null hypothesis of no cointegration stands and hence the estimate of γ tends to zero quite rapidly to make equation (15) consistent.

Even if income and consumption were cointegrated as in the DGP examined in Section 3, the coefficient and the t-ratio in a regression like (15) would have the same order of magnitude as before, although with a different non-degenerate distribution, emphasizing again that only under the null can the error term in (15) be treated as stationary. In fact by taking E_{t-1} in the second equation of that DGP, and substituting from the first equation, we can get an expression similar to (14). This points out the close or exact observational equivalence of consumption models generated by both types of DGPs when only income conveys relevant information about the future, a point already acknowledged by Hall (1981) in his comment on Davidson and Hendry (1981) who propose the DGP in Theorem 2 as realistic.

However there is an issue which merits further research. Until now we have been dealing solely with the case of driftless correlated random walks and no constant or trend in (15). Some ongoing research (see Banerjee et al. (1986)) shows that in this case the t-ratio is centred around a value less than one, so that by using the standard critical

15

values one may be under-rejecting instead of over-rejecting as claimed by Mankiw and Shapiro. Since the whole issue depends on the use of raw or detrended data and the presence of constant terms, it may be conjectured that their results apply only to certain cases and are therefore not robust to the presence of nuisance parameters in the testing regression.

V. CONCLUSIONS

Monte Carlo methods could be used to study a number of further issues in the exploration of equilibrium relationships through static models. Data generation processes with drift and trend and with multiple (in the multivariate case) or time-varying cointegrating vectors would be of interest. In this paper we have begun to assess what can be learned from static models about bivariate relationships. We have found that tests for a unit root in the residuals often lead to the erroneous conclusion that the regression is spurious. Even where a relationship is found to exist the estimated cointegrating coefficient may have substantial bias. A high R^2 in the static model can serve as an index of the value of the estimate, but estimates from dynamic models are more valuable still.

We conclude that care should be exercised in trying to parameterize long-run relationships using static regressions. This warning has some force precisely because we have examined the bivariate case, the case in which inference should be easiest. In the multivariate case a high R^2 in the static model does not imply that the estimated coefficient on each variable is close to its true value.[8] For example, R^2 may be high in the regression of money on prices, real income, and an interest rate but low in a regression of velocity on the interest rate; it is not invariant to transformations of the model.

Nuffield College, Oxford

REFERENCES

Anderson, T. W. (1959). 'On the Asymptotic Distribution of Estimates of Parameters of Stochastic Difference Equations', *Annals of Mathematical Statistics*, Vol. 30, pp. 676-98.
Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, New York, Wiley.

[8] That is, some partial correlations may be low.

16

Banerjee, A., Dolado, J., Hendry, D. F. and Smith, G. W. (1986). 'Tests of the LC/PI Hypothesis in the Presence of Random Walks', (Nuffield College) Manuscript.

Bhargava, A. (1983). 'On the Theory of Testing for Unit Roots in Observed Time Series', LSE: ICERD No. 67.

Davidson, J., Hendry, D. F., Srba, F. and Yeo, S. (1978). 'Econometric Modelling of the Aggregate Time-Series Relationship Between Consumer's Expenditure and Income in the United Kingdom', *Economic Journal*, Vol. 88, pp. 661-92.

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

Dickey, D. A. and Fuller, W. A. (1979). 'Distributions of the Estimators for Autoregressive Time Series with a Unit Root', *Journal of American Statistical Association*, Vol. 74, pp. 427-31.

Dickey, D. A. and Fuller, W. A. (1981). 'Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root', *Econometrica*, Vol. 49, pp. 1057-72.

Evans, G. B. A. and Savin, N. E. (1981). 'Testing for Unit Roots: 1', *Econometrica*, Vol. 49, pp. 753-77.

Evans, G. B. A. and Savin, N. E. (1984). 'Testing for Unit Roots: 2', *Econometrica*, Vol. 52, pp. 1241-69.

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

Granger, C. W. J. (1981). 'Some Properties of Time Series Data and Their Use in Econometric Model Specification', *Journal of Econometrics*, Vol. 16, pp. 121-30.

Granger, C. W. J. (1983). 'Cointegrated Variables and Error Correcting Models', UCSD Discussion Paper 83-13a.

Granger, C. W. J. and Newbold, P. (1974). 'Spurious Regressions in Econometrics', *Journal of Econometrics*, Vol. 2, pp. 111-20.

Granger, C. W. J. and Weiss, A. A. (1983). 'Time Series Analysis of Error Correcting Models', UCSD Discussion Paper 82-28.

Granger, C. W. J. and Engle, R. F. (1985). 'Dynamic Model Specification with Equilibrium Constraints: Cointegration and Error Correction', UCSD mimeo.

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.

Hall, R. E. (1981). 'Comments: "Interpreting Econometric Evidence" by Davidson and Hendry', *European Econometric Review*, Vol. 16, pp. 193-94.

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

Hendry, D. F. and Mizon, G. (1978). 'Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England', *Economic Journal*, Vol. 88, pp. 549-63.

Hendry, D. F. (1983). 'An Expository Note on Cointegrability and Testing the Existence of Equilibrium Relationships', Nuffield College, Oxford, mimeo.

Hendry, D. F. (1984). 'Monte Carlo Experimentation in Econometrics', Chapter 16, pp. 937-76 in Griliches, Z. and Intriligator, M. D. (eds.), *Handbook of Econometrics*, Vol. II, Amsterdam, North-Holland.

Lucas, R. E., Jr. (1980). 'Two Illustrations of the Quantity Theory of Money', *American Economic Review*, Vol. 70, pp. 1005-14.

Mankiw, N. G. and Shapiro, M. D. (1985a). 'Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models', Harvard Institute

of Economic Research Discussion Paper No. 1191 and *Economics Letters*, Vol. 20, 1986 in press.

Mankiw, N. G. and Shapiro, M. D. (1985b). 'Trends, Random Walks and Tests of the Permanent Income Hypothesis', *Journal of Monetary Economics*, Vol. 16, pp. 165-74.

McCallum, B. T. (1984). 'On Low-Frequency Estimates of Long-Run Relationships in Macroeconomics', *Journal of Monetary Economics*, Vol. 14, pp. 3-14.

Molinas, C. (1986). 'A Note on Spurious Regressions with Integrated Moving Average Errors', (this issue).

Phillips, P. C. B. (1985a). 'Time Series Regression with Unit Roots', Cowles Foundation D.P. 740.

Phillips, P. C. B. (1985b). 'Understanding Spurious Regressions in Econometrics', Cowles Foundation D.P. 757.

Phillips, P. C. B. (1985c). 'Asymptotic Expansions in Nonstationary Vector Autoregressions', Cowles Foundation D.P. 765.

Phillips, P. C. B. and Durlauf, S. N. (1985). 'Multiple Time Series Regression with Integrated Processes', Cowles Foundation D.P. 768.

Rothenberg, T. J. (1984). 'Approximating the Distributions of Econometric Estimators and Test Statistics', Chapter 15, pp. 882-933 in Griliches Z. and Intriligator, M. D. (eds.), *Handbook of Econometrics*, Vol. II, Amsterdam, North-Holland.

Sargan, J. D. (1964). 'Wages and Prices in the United Kingdom: A Study in Econometric Methodology', in Hart, P. E., Mills, G. and Whittaker, J. N. (eds.), *Econometric Analysis for National Economic Planning*, London, Butterworths.

Sargan, J. D. and Bhargava, A. (1983). 'Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk', *Econometrica*, Vol. 51, pp. 153-74.

Stock, J. H. (1984). 'Asymptotic Properties of a Least Squares Estimator of Cointegrating Vectors', Harvard University, mimeo.

Summers, L. H. (1983). 'The Nonadjustment of Nominal Interest Rates: A Study of the Fisher Effect', pp. 201-41, in Tobin, J. (ed.), *Macroeconomics, Prices, and Quantities*, Oxford, Basil Blackwell.

Summers, L. H. (1984). 'Estimating the Long-Run Relationship Between Interest Rates and Inflation: A Response to McCallum', Harvard Institute of Economic Research D. P. 1097.

White, H. (1980). 'A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity', *Econometrica*, Vol. 48, pp. 817-38.

White, H. (1984). *Asymptotic Theory for Econometricians*, London, Academic Press.

White, J. S. (1958). 'The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case', *Annals of Mathematical Statistics*, Vol. 29, pp. 1188-97.

White, J. S. (1959). 'The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case II', *Annals of Mathematical Statistics*, Vol. 30, pp. 831-34.

Whiteman, C. H. (1984). 'Lucas on the Quantity Theory: Hypothesis Testing without Theory', *American Economic Review*, Vol. 74, pp. 742-49.

Working, H. (1934). 'A Random Difference Series for Use in the Analysis of Time Series', *JASA*, Vol. 29, pp. 11-24.

Yule, G. U. (1926). 'Why Do We Sometimes Get Nonsense Correlations Between Time Series? A Study in Sampling and the Nature of Time Series', *Journal of the Royal Statistical Society*, Vol. 84, pp. 1-64.

Some Asymptotic Theory Results for Non-Stationary Processes

In order to sharpen understanding of some of the Monte Carlo results derived in Sections 2 and 3 of this study, we will use some results recently provided by Phillips (1985a, 1985b, 1985c). These results deal with the proper asymptotic theory for the least-squares regression estimator and its associated statistics in the case where regression relates independent or correlated random walks. As is well known, the limiting distribution theory that is usually applied in econometrics for stationary time series is no longer valid for non-stationary cases. The distribution theory for the latter belongs to a general class of functional limit theory on metric spaces rather than the central limit theorems in Euclidean spaces conventionally used in econometrics. In order to duplicate the properties of the DGPs used in the Monte Carlo studies, we will develop all the theoretical results for the special case in which the innovations are n.i.d. $(0, \sigma^2)$. Nevertheless, these results hold with much weaker conditions on the innovation processes.[9] The proofs of the theorems below are available from the authors.

Let us suppose that $\{x_t\}$ is a stochastic process generated by a random walk,

$$x_t = x_{t-1} + \epsilon_t \tag{A1}$$

where $\epsilon_t \sim$ n.i.d. $(0, \sigma^2)$. From (A1) we have the alternative representation for x_t in terms of the partial sums $S_t = \Sigma_{i=1}^t \epsilon_i$ of the innovation sequence $\{\epsilon_t\}$ and the initial condition x_0:

$$x_t = S_t + x_0 \tag{A1'}$$

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

$$X_T(t) = (T^{-1/2} \cdot \sigma^{-1}) \cdot S_{[Tt]} = (T^{-1/2} \cdot \sigma^{-1}) \cdot S_{j-1};$$

$$(j-1)/T \leqslant t < j/T, (j = 1 \ldots T) \tag{A2}$$

$$X_T(1) = (T^{-1/2} \cdot \sigma^{-1}) \cdot S_T \tag{A2'}$$

where $[b]$ denotes the integer part of b. $X_T(t)$ is a random element in the function space $D[0,1]$ i.e. in the space of all real-valued functions on $[0,1]$ which are right-continuous at each point of $[0,1]$ and have

---

[9] They allow for a wider class of weakly dependent heterogeneously distributed processes (see e.g. lemmae 2.1 and 2.2 in Phillips (1985a). These results cover a large class of α and ρ mixing processes as defined in White (1984), in particular applying to many stationary finite order ARMA processes with moderate degrees of heteroskedasticity, extending the notion of $I(1)$ processes in the Granger–Engle sense to a wider class, i.e. that class of processes that after first differencing becomes 'evolutionary' in the sense mentioned before.

finite left limits. Under the conditions mentioned above, $X_T(t)$ can be shown to converge weakly to a limit process known as the Wiener process which is denoted by $W(t)$. Symbolically,

$$X_T(t) \rightarrow W(t) \text{ as } T \rightarrow \infty \qquad \text{(A3)}$$

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

$$g(X_T(t)) \rightarrow g(W(t))$$

Consider the following pairs of random walks,

$$x_t = S_t + x_0; y_t = P_t + y_0; \ \epsilon_{it} \sim \text{n.i.d}(0, \sigma_i^2) \ (i = 1,2); E(\epsilon_{1t}\epsilon_{2s}) = 0 \ \forall \ t, s$$

S_t and P_t denote respective partial sums of ϵ_{1t} and ϵ_{2t}.

The following lemma (see Phillips, 1985b) will be used extensively.

Lemma: If the sequences $\{y_t\}$ and $\{x_t\}$ are generated by the independent random walks defined as above, then as $T \rightarrow \infty$,

(a) $T^{-2} \sum_1^T y_t^2 \rightarrow \sigma_2^2 \cdot \int_0^1 V(t)^2 \, dt$

(b) $T^{-2} \sum_1^T x_t^2 \rightarrow \sigma_1^2 \cdot \int_0^1 W(t)^2 \, dt$

(c) $T^{-2} \sum_1^T x_t y_t \rightarrow \sigma_1 \sigma_2 \cdot \int_0^1 V(t)W(t) \, dt$

(d) $T^{-1} \sum_1^T y_{t-1} \epsilon_{2t} \rightarrow (\sigma_2^2/2) \cdot (V(1)^2 - 1)$

(e) $T^{-1} \sum_1^T x_{t-1} \epsilon_{1t} \rightarrow (\sigma_1^2/2) \cdot (W(1)^2 - 1)$

(f) $T^{-1} \sum_1^T y_t \epsilon_{1t} \rightarrow N(0, \sigma_1^2 \sigma_2^2/2)$

20

(g) $T^{-1} \sum_{1}^{T} x_t \epsilon_{2t} \to N(0, \sigma_1^2 \sigma_2^2/2)$

where $V(t)$ and $W(t)$ are independent Wiener processes on $C[0,1]$.

We have stated the basic distributional results and will proceed to develop applications related to the DGPs analysed in the Monte Carlo study.

Theorem 1: If the stochastic processes $\{y_t\}$ and $\{x_t\}$ are generated by the following DGP:

$$y_t + \beta x_t = u_t; \quad u_t = u_{t-1} + \epsilon_{1t}$$

$$y_t + \alpha x_t = v_t; \quad v_t = \rho v_{t-1} + \epsilon_{2t}$$

$$(\epsilon_{1t}, \epsilon_{2t})' \sim \text{n.i.d} \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right]$$

Define

$$S_t = \sum_{j=1}^{t} \epsilon_{1j}; \quad s = \sigma_1/\sigma_2$$

and y_t is regressed on x_t giving the least squares regression

$$y_t = \hat{\gamma} x_t + \hat{\eta}_t$$

then as $T \to \infty$,

(a) $T(\hat{\gamma} + \alpha) \to (\beta - \alpha) \dfrac{[N(0, \frac{1}{2}s^2(1 - \rho^2)) - (1/s^2(1 - \rho^2))]}{\displaystyle\int_0^1 W(t)^2 \, dt} = \xi_1$

(b) $T(1 - R^2) \to \dfrac{\left[\dfrac{(\beta - \alpha)^2}{\alpha^2 s^2 (1 - \rho^2)} \right]}{\displaystyle\int_0^1 W(t)^2 \, dt}$

(c) $t_{\hat{\gamma}} \to \dfrac{\xi_1 s(1 - \rho^2)^{1/2}}{(\beta - \alpha)} \left[\int_0^1 W(t)^2 \, dt \right]^{1/2}$

(d) $DW \to 2(1 - \rho)$

where $W(t)$ is a Wiener process on $C[0,1]$.

Theorem 1 relates to the DGP developed by Granger and Engle (1985). This DGP is the object of the Monte Carlo study in the section 2. Further, (a) shows that the equilibrium regression is in fact the second equation of the DGP as expected and that any bias in $\hat{\gamma}$ for α converges to zero at order T^{-1} thereby reproducing Stock's (1984) results. The intuition behind Stock's analysis was discussed in Section 1. The same rate of convergence obtains for $(1 - R^2)$ while the t-ratio and the DW are of order one. Note however that the t-ratio does not converge in probability to a constant.

The results (a) and (b) above can be used to construct a response surface for the behaviour of the bias in the equilibrium regression with the Granger–Engle DGP. In other words, from the limiting distribution, (a) and (b) can be used to construct the leading terms in the response surface of the bias by substituting for the unknown functionals of the Wiener process by a known statistic of the static regression. The obvious candidate for such a statistic is $(1 - R^2)$ since it has the same order of convergence as the bias. More importantly, the role of this statistic in the bivariate equilibrium regression as an indicator of the bias is clearly revealed. Further implications of this observation are discussed in the text.

Substituting (b) into (a) we obtain,

$$T[(\hat{\gamma} - \alpha) - \alpha^2(\alpha - \beta)^{-1}(1 - R^2)] \to N\left[0 \frac{s^2 \cdot \alpha^4(1 - \rho^2)\,T^2(1 - R^2)^2}{2(\beta - \alpha)^2}\right]$$

(A4)

The intuition behind this result is quite clear. The coefficient obtained by regressing x_t on y_t in the equilibrium regression converges to α^{-1} and since the R^2 is just the product of the slopes, the closer it is to unity, the smaller will be the bias.

Theorem 2: Suppose now that the stochastic process generating $\{y_t\}$ and $\{x_t\}$ is,

$$\Delta x_t = \epsilon_{1t}$$

$$\Delta y_t = \alpha_1 \Delta x_t - \alpha_2(y - x)_{t-1} + \epsilon_{2t}$$

where

$$(\epsilon_{1t}, \epsilon_{2t})' \sim \text{n.i.d}(0, \Omega)$$

where the Ω matrix is unchanged from the previous DGP. Define S_t and s as above.

y_t is regressed on x_t and the regression coefficient is denoted again by $\hat{\gamma}$. Then as $T \to \infty$,

22

(a) $T(\hat{\gamma} - 1) \rightarrow \alpha_2^{-1}(\tfrac{1}{2}(\alpha_1 - 1)(W(1)^2 + 1) + N\left(0, \dfrac{s^2}{2}\right) / \displaystyle\int_0^1 W(t)^2 \, dt = \xi_2$

(b) $T(1 - R^2) \rightarrow ((\alpha_1 - 1)^2 + s^{-2})(1 - (1 - \alpha_2)^2)^{-1} / \displaystyle\int_0^1 W(t)^2 \, dt$

(c) $t_{\hat{\gamma}} \rightarrow \xi_2((1 - (1 - \alpha_2)^2))^{1/2}((\alpha_1 - 1)^2 + s^{-2})^{-1/2}\left(\displaystyle\int_0^1 W(t)^2 \, dt\right)^{1/2}$

(d) $DW \rightarrow 2\alpha_2$

Theorem 2 relates to the DGP developed in the Monte Carlo experiments in Section 3. We investigate and compare the powers of the DW statistic in the equilibrium regression and the t-ratio test of the error correction term in a restricted specification of the ECM model under the null hypothesis of cointegration. In particular, we deal with the equilibrium regression results under the assumption of cointegration. Since the ECM just represents a different parameterization of the bivariate representation analysed in Theorem 1 (for another example) the results turn out to be the same with both the bias and $(1 - R^2)$ again converging to zero at the rate of T^{-1} and so on. Also from the relationship between the bias and $(1 - R^2)$ we can identify the leading term in the response surface for the second Monte Carlo study, i.e.

$$(\hat{\gamma} - 1) = (\alpha_1 - 1)\frac{[1 - (1 - \alpha_2)^2](1 - R^2)}{2\alpha_2[(\alpha_1 - 1)^2 + s^{-2}]} + 0(1/T)$$

The distribution results under the alternative of stationarity both for the restricted and the unrestricted ECM specifications could be easily derived from the conventional asymptotic theory for $I(0)$ processes and therefore will not be quoted here although we shall be interested in the power of the test statistics.

Theorem 3: If the stochastic processes are generated by the DGP analysed in Theorem 2 with $\alpha_2 = 0$ and y_t regressed on x_t giving the least squares regression of y_t on x_t, then as $T \rightarrow \infty$,

(a) $\hat{\gamma} \rightarrow \alpha_1 + s^{-1}\left[\displaystyle\int_0^1 V(t)W(t) \, dt\right]\left[\displaystyle\int_0^1 W(t)^2 \, dt\right] = \xi_3$

23

(b) $R^2 \to \left[\xi_3{}^2 \cdot \int_0^1 W(t)^2 \, dt \right] \left[\nu + \xi_3{}^2 \int_0^1 W(t)^2 \, dt \right]^{-1}$

(c) $T^{-1/2} t_{\hat{\gamma}} \to \left[\xi_3{}^2 \int_0^1 W(t)^2 \, dt \right]^{1/2} / \nu$

(d) $T \cdot DW \to [(\alpha_1 - \xi_3)^2 + s^{-2}] / \nu$

where

$$\nu = (\alpha_1{}^2 - \xi_3{}^2) \int_0^1 W(t)^2 \, dt + s^{-2} \int_0^1 V(t)^2 \, dt + 2\alpha_1 s^{-1} \int_0^1 V(t)W(t) \, dt$$

and $V(t)$ and $W(t)$ are independent Wiener processes on $C[0,1]$.

Theorem 4: If the stochastic processes $\{y_t\}$ and $\{x_t\}$ are generated by the DGP analysed in Theorem 2 with $\alpha_2 = 0$ and the following regression is performed

$$\Delta y_t = \hat{\gamma}_1 \Delta x_t + \hat{\gamma}_2 (y - x)_{t-1} + \hat{\eta}_t$$

then as $T \to \infty$,

(a) $T\hat{\gamma}_2 \to \left[\dfrac{1}{2s^2} \left\{ V(1)^2 - 1 \right\} + \dfrac{\alpha_1^2}{2} \left\{ W(1)^2 - 1 \right\} - (\alpha_1 - 1) N\left(0, \dfrac{1}{2s^2} \right) \right]$

$\times \left[(\alpha_1 - 1)^2 \int_0^1 W(t)^2 \, dt + s^{-2} \int_0^1 V(t)^2 \, dt \right.$

$\left. + 2s^{-1}(\alpha_1 - 1) \int_0^1 V(t)W(t) \, dt \right]^{-1} = \xi_4$

(b) $t_{\hat{\gamma}} \to \xi_4 \left[(\alpha_1 - 1)^2 s^2 \int_0^1 W(t)^2 \, dt + \int_0^1 V(t)^2 \, dt \right.$

$\left. + 2(\alpha_1 - 1) s \int_0^1 V(t)W(t) \, dt \right].$

Theorems 2, 3, and 4 are valuable in analysing the Monte Carlo results in Section 3 of the paper. We have analysed theoretically the properties

of the bias and DW statistic in the static regression and the t-ratio test of the error correction term under the null hypothesis of (a) no cointegration and (b) cointegration. The corresponding simulation results are presented in Section 3.

Theorem 3 reproduces, but for a slightly different DGP, the results obtained by Phillips (1985b) in the spurious regression case. In the first place (a) and (b) show that the least squares coefficient and the R^2 have non-degenerate limiting distributions while the distribution of the corresponding t-ratio diverges according to (c). This result is in total agreement with the Monte Carlo results reported by Granger and Newbold (1974). In (d) it is shown that the DW statistic converges to zero again reproducing a typical feature of spurious regressions with data generated by random walks. The limiting distribution of $T \cdot DW$ may be used to construct an asymptotic critical value as suggested by Sargan and Bhargava (1983) and Granger and Engle.

By means of Theorem 4, we analyse the case in which an ECM is estimated. The results are, as expected, quite different from the equilibrium regression case. First, the coefficient of the error correction term goes to zero as $T \rightarrow \infty$ while in the previous case the slope did not converge to a constant. The intuition is again quite important in order to explain this different behaviour, i.e. in Theorem 1 the null hypothesis $\gamma = 0$ meant that η could not be stationary, while in Theorem 4, the null $\gamma_2 = 0$ implies stationarity of the residuals and so its estimate should converge quickly to zero, as it indeed does. Nevertheless the corresponding t-ratio has a non-degenerate distribution different from the standardized normal distribution that is used in conventional asymptotic theory. The corresponding critical values may be constructed by simulation as is done by Dickey and Fuller (1979) and Granger and Engle. This observation has interesting implications for testing models like Hall's (1978) random-walk-consumption model as in Section 4.