QED

Queen's Economics Department Working Paper No. 914

Dynamic Specification and Testing for Unit Roots and Co-Integration

Anindya Banerjee

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

11-1994

Discussion Paper #914

Dynamic Specification and Testing for Unit Roots and Co-integration

by

Anindya Banerjee
Queen's University
and
Wadham College and IES

November 1994

DYNAMIC SPECIFICATION AND TESTING FOR UNIT ROOTS AND CO-INTEGRATION[1]

Anindya Banerjee
Wadham College and IES
Oxford OX1 3PN
UNITED KINGDOM

December 1994

JEL Classification: C 10, C 13, C 51, C 52

Keywords: Dynamic specification, co-integration, unit roots, exogeneity

---

1 This paper is part of a programme of research with which I have been associated for several years. During this period I have benefited greatly from discussions with my colleagues Juan Dolado, John Galbraith, and David Hendry. While absolving them of any responsibility for the views expressed in this paper, it is a pleasure to acknowledge their help. I also wish to thank Kevin Hoover, Steve Sheffrin and Eric Zivot for helpful comments and the Department of Economics at Queen's University, Kingston, where the final version of this paper was written while on sabbatical, for their hospitality. The financial assistance of the Economic and Social Research Council (UK) under grant R231184 is also gratefully acknowledged.

Abstract

This paper provides an interpretation of the vast literature on testing for unit roots and estimating co-integrating relations. Emphasis is placed on identifying the particular ways in which methods of dynamic specification need to be modified in order to take account of the possible presence of unit roots in the time series being modelled. The discussion is undertaken in the settings of both single-equation and systems methods and attention is paid to problems of estimation and inference. It is argued that the importance of issues such as exogeneity and rich dynamic specification, developed in the context of stationary time series, carry over to a very large extent when dealing with non-stationary series.

1. Introduction

In the field of modelling economic time series, the eighties might easily be described as the "decade of co-integration". During this decade, theoretical and applied econometricians alike have invested a great deal of effort in dealing with the theoretical and empirical implications of Nelson and Plosser's (1982) central observation that time series of important economic variables such as consumption and per-capita GNP may have statistical properties quite distinct from those which would warrant the use of standard tools , such as normal, *t*-, and *F*-tables, of inference and estimation.

The results of this research have given rise to an unmanageably vast literature on almost every aspect of estimation and inference in the presence of non-stationary series. It is therefore impossible, in the space available, to provide a complete account of this field. The case for writing a formal survey is in any case rather limited, given the several surveys which have already appeared (see, for example, Stock and Watson (1988), Dolado, Jenkinson and Rivero (1990), Campbell and Perron (1991)). The purpose of this paper is instead to focus on specific issues which, in my view, are important in any evaluation of this literature. In particular, I will attempt to address, at least partially, the issue of the extent to which our notions of what constitutes good or appropriate modelling practice have changed as a result of the research on unit roots.

The mathematical and statistical tools on which the econometrics literature on unit roots depends, date back at least to the 1920's, 1930's and 1940's, notably to the work of Wiener, Lévy, Doob and many others. Thus no claim for mathematical or statistical originality can be made, *per se*, on behalf of this literature. Rather, I will argue that econometricians have brought the highly developed theory on stochastic processes into the realm of every day econometric modelling and, by applying it to problems particular to econometrics, have added considerably to our store of knowledge of the very special properties of non-stationary series and the implications of these properties for estimation and inference.

However, I will also argue that while a greater realization of how things can go wrong when dealing with these series has had an important effect on our econometric consciousness, this does not lead us necessarily to a fundamental re-evaluation of modelling practice, in particular of dynamic modelling. For example, it will be shown that some of the inferential problems which arise may be overcome by suitable transformations and appropriate augmentation of equations. These methods will often allow us to return to using standard tables for inference.

In the next section, two examples are presented which illustrate the fundamental differences which could arise between the treatments of unit root and non-unit root processes. The differences emerge particularly because the critical values of standard tests, such as t- or F- tests, are affected by the presence or absence of unit roots. In Sections 3.1 and 3.2 it is shown how these differences can, in some circumstances, be eliminated by a proper reformulation of the model. However there are cases where such reformulations are not possible and it therefore becomes very important, before proceeding to the formal task of econometric modelling, to classify the variables of interest by their orders of integration. This is a task which is by no means an easy one to accomplish, even with our fairly advanced understanding of the asymptotic theory, given the low powers of most available tests for unit roots. In Section 3.4 I propose a conservative testing strategy to allow for the possibility of incorrect classification. Sections 3.3 and 3.4 also deal with the issue of estimating co-integrating relationships in single equations, dealing with, in particular, the two-step method proposed by Engle and Granger (1987) and suggesting a simple alternative. This is linked with the issue of *testing* for co-integration and unit roots, an area which has generated considerable interest in the literature. Section 4 considers single-equation *vs.* systems methods of estimation and shows that the choice between these two methods can be made and understood within the familiar concepts of exogeneity. Section 5 concludes. Since frequent reference is made in the text to the concepts of weak and strong exogeneity, an appendix contains a discussion, based on Engle, Hendry and Richard (1983), of these concepts.

2. Spurious and Inconsistent Regressions

2.1 Spurious regressions

While Nelson and Plosser's paper provided one of the early surprising insights in the literature on unit roots, Yule (1926) had already alerted the profession to the potential dangers of undertaking stationary inference in an environment with non-stationary variables. He termed the phenomenon "nonsense" regressions and showed how regressing one variable which followed a random walk on another totally unrelated random walk led to findings of significant correlations between the two series. Granger and Newbold (1974) returned to the Yule example and their formulation of the problem forms the starting point for our analysis. Granger and Newbold called such regressions "spurious" and this has come to be regarded as the more commonly accepted terminology.

Granger and Newbold considered the following data generation process (DGP) for the data series $\{x_t\}_{t=1}^{T}$, $\{y_t\}_{t=1}^{T}$:

$$y_t = y_{t-1} + u_t, \; u_t \sim \text{IID}(0, \sigma_u^2), \qquad (1)$$

$$x_t = x_{t-1} + v_t, \; v_t \sim \text{IID}(0, \sigma_v^2), \qquad (2)$$

$$E(u_t v_s) = 0 \; \forall \; t \neq s; \; E(u_t u_{t-k}) = E(v_t v_{t-k}) = 0 \; \forall \; k \neq 0.$$

Thus x_t and y_t are uncorrelated random walks. Before proceeding to the formal details of the example some important features of the data generation process may be noted. The process generates two variables, each with a unit root. This terminology can be understood more readily by rewriting the DGP in lag polynomial operator form:

$$(1 - \rho_1 L) y_t = u_t, \qquad (1')$$

$$(1 - \rho_2 L) x_t = v_t, \qquad (2')$$

where the processes generating $\{u_t\}$ and $\{v_t\}$ remain unchanged and L is the lag operator such that $L^j x_t = x_{t-j}$, $L^j y_t = y_{t-j}$.

The "unit root" in the $\{x_t\}$ and $\{y_t\}$ processes refers to the value of unity for the coefficients ρ_1 and ρ_2. Values of ρ_1 and ρ_2 such that $|\rho_1| < 1$, and $|\rho_2| < 1$ correspond to "stationary roots". In the terminology of Engle and Granger (1987), (1) and (2) are called processes "integrated of order 1", denoted I(1), *i.e.* they need to be differenced once to achieve stationarity.[2]

It is interesting to compare the properties of the series $\{x_t\}$ and $\{y_t\}$ when $\rho_1 = \rho_2 = 1$ with the properties of the series generated by values of $|\rho_1| < 1$ and $|\rho_2| < 1$. In the former case both series have unconditional variances which grow with time (at rate t) while the series have time-invariant finite variances in the latter. The autocorrelation function $r_i = E(z_t z_{t-i})/[var(z_t)var(z_{t-i})]^{1/2}$, $z = y, x$, is an exponentially declining function of i when $|\rho_1| < 1$, $|\rho_2| < 1$, *i.e.* the past of the series becomes increasingly less important. When the processes have unit roots however, the correlations persist at significantly large values even when the observations are substantially far apart.

It is this property of persistence which drives many of the properties of spurious regressions. Each time series is growing but for entirely different reasons and by increments that are uncorrelated. Hence a correlation, induced simply by persistent but *independent* growth, cannot be interpreted in the way that it could be if it arose among stationary series.

Granger and Newbold showed that if standard normal tables were used to conduct tests of significance on the *t*-ratio, $t_{\beta_1=0}$, in the regression

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \qquad (3)$$

the tests would reject the null of $\beta_1 = 0$, on average, between 50% and 70% of the time (at a nominal significance level of 5%)! Thus the use of standard tables would be grossly misleading in the presence of integrated processes.

A question that might well be asked is whether the misleading inferences arise simply because the two integrated series are unrelated with each other and if we considered two integrated but related series whether the problem would disappear. Unfortunately, this too happens to be untrue. In a regression such as (3) the finite and asymptotic distributions of both β_0 and β_1 are non-normal even when y_t and x_t are linked by some hypothesized equlibrium relationship (say, y is income and x is consumption). The tools of inference are, in general, non-standard. However, as we shall show in a later section, it is possible to reparameterize and extend (3) such that at least some of the inferences can be undertaken using standard tables.

A detailed theoretical analysis of the spurious regression problem was undertaken by Phillips (1986). He showed that the *t*-statistic diverged asymptotically. Thus the inferential problem would become worse as the sample size increased and therefore, in the limit, to avoid making spurious inferences, infinitely large *t*-values would be needed to reject the null $\beta_1 = 0$. The asymptotic distribution of the R^2 of the regression would also have substantial weight at the ends of its support (-1 and 1) and values well away from zero would therefore be very likely. For stationary series however, none of these problem would arise. In particular, β_1 would tend in probability to zero.

It is possible to provide a good intuition for some of the rather dramatic results described above. In (3), both the null hypothesis $\beta_1 = 0$ and alternative $\beta_1 \neq 0$ lead to false models, since the true DGP is not nested within (3). It is therefore not surprising that the null hypothesis, implying that $y_t = \varepsilon_t$, in other words that $\{y_t\}$ is a white-noise process, is rejected: the persistence in $\{y_t\}$ is projected onto $\{x_t\}$, also a random walk and therefore also highly persistent, and spurious correlations arise. Further, the use of the normal table asymptotically is based on the assumption that $\{\varepsilon_t\}$ is a white-noise process under H_0. This is clearly false and it follows then that the *t*-statistic is not asymptotically normally distributed.

Phillips (1986) also demonstrated an important feature of the

Durbin-Watson statistic (DW) calculated from the residuals of (3). When the regression is spurious, DW $\rightarrow$ 0 in probability. This is a consequence of precisely the property discussed in the previous paragraph. Under H_0, $\{\varepsilon_t\}$ far from being a white-noise process is instead highly correlated and this is revealed in a low value of DW. When the two series are genuinely related, the DW statistic converges to a non-zero value and the behaviour of the DW statistic therefore provides a way of discriminating between genuine and spurious regressions. We will return to this issue in a later section.

2.2 Inconsistent Regressions

Another example of the dangers involved in using standard distributions for inference when there are non-stationary variables present was highlighted by Mankiw and Shapiro (1985, 1986) in their discussion of what have come to be called "inconsistent" or "unbalanced" regressions.

In this terminology, a regression is said to inconsistent (unbalanced) if the regressand is not of the same order of integration as the regressors, or any linear combination of the regressors. The inconsistency (imbalance) refers to the disparity in the orders of integration of the variables on the two sides of the regression. Thus the problem occurs if, say, the regressand is an I(0) variable while the regressors, individually *and* in combination, are I(1). The problem also appears if the regressors are only near-integrated, *i.e* ρ_1 is close to, but not equal to, one in absolute value. Mankiw and Shapiro's analysis which we describe below concentrates on these near-integrated cases but their results apply equally well to integrated series.

The difficulty with these inconsistent regressions can be understood in the context of spurious regressions. An I(0) variable cannot be related in any meaningful sense to an I(1) variable (where the I(1) variable may be taken to be a linear combination or composite of several I(1) variables), given their very different statistical properties and behaviour over time. Thus an inconsistent or near-inconsistent (if only near-integrated variables appear) regression may be regarded as a special kind of spurious regression. The use of standard tables will therefore lead to misleading inferences on the significance of estimates of parameters.

This discussion has considerable economic interest since tests for rational expectations, in consumption (Flavin, 1981) or the stock market (Fama and French, 1989), typically give rise to such regressions. For example, Flavin's (1981) test of Hall's (1978) random walk hypothesis for consumption takes the form of regressing differenced consumption ($(1-L)c_t = \Delta c_t$) on lagged income (y_{t-1}). If both consumption and income are I(1) variables this regression is inconsistent. Under the null hypothesis, that consumption follows a random walk, the coefficient on the lagged income term should not be significantly different from zero. However, given the form of the regression, the *t*-statistic for the coefficient estimate on lagged income does not have the standard *t*-distribution and use of the *t*-table will lead to spurious findings of significance and hence rejections of the random walk hypothesis.

Mankiw and Shapiro consider the following DGP for hypothetical $\{c_t\}$ and $\{y_t\}$ series:

$$\Delta c_t = \nu_t,$$
$$y_t = \theta y_{t-1} + \varepsilon_t,$$
$$E_{t-1}(\nu_t) = E_{t-1}(\varepsilon_t) = 0, \quad (4)$$
$$\text{corr}(\nu_t, \varepsilon_t) = \rho,$$
$$\text{corr}(\varepsilon_{t+j}, \nu_t) = 0 \ \forall \ j \neq 0.$$

E_{t-1} is the expectation, conditional on information available at time t-1, of the value of variables dated in the future. The model is given by:

$$\Delta c_t = d_1 + d_2 y_{t-1} + u_t. \quad (5)$$

The null hypothesis is given by H_0: $d_2 = 0$ and Mankiw and Shapiro use Monte Carlo simulations to tabulate the *actual* rejection frequencies of H_0: $d_2 = 0$, when standard *t*-critical values are used, for a range of specified values for θ, ρ, and T (the size of the samples generated). Table I below gives some of their results for model (5) and also for a model with a linear time trend,

$$\Delta c_t = d_1 + d_2 y_{t-1} + d_3 t + u_t. \quad (6)$$ [3]

Table I Percentage rejection frequencies in standard *t*-tests at nominal 5% level[4]

DGP: (4); Sample size = T; No. of replications = 1000

$\theta \setminus \rho \rightarrow$ ↓	Model (5)					Model (6)				
	1.0	0.9	0.8	0.5	0.0	1.0	0.9	0.8	0.5	0.0
(a) T = 50										
0.999	30	24	20	11	7	60	45	36	16	6
0.99	26	20	15	10	7	54	40	33	15	6
0.98	22	17	15	8	7	50	37	30	14	5
0.95	17	12	10	7	6	38	30	25	12	6
0.90	12	9	8	6	6	28	22	19	10	6
0.00	5	6	6	5	5	6	7	7	5	6
(b) T = 200										
0.999	29	23	20	10	5	61	48	38	18	5
0.99	18	15	13	8	4	41	32	27	13	5
0.98	13	10	9	7	5	29	24	20	11	6
0.95	9	7	7	6	5	17	14	12	7	6
0.90	7	6	6	6	6	10	9	8	6	7
0.00	5	4	4	5	5	5	5	4	5	5

Table I has several interesting features. First, the size distortions are an increasing function of the value of the autoregressive parameter. Critical values given by Mankiw and Shapiro show that this arises from a leftward (from zero) shift of the *t*-density for values of θ close to 1. The closer θ is to 1 the greater the shift and the more likely the occurrence of values in excess of -2, leading to a higher probability of rejection of the true null and hence giving rise to greater size distortions. Because of the nature of the alternative hypothesis H_A: $d_2 < 0$, rejections of the null hypothesis all take place in the lower tail. In fact, because $d_2 > 0$ implies an explosive (or exponentially growing) process, two-sided critical values are not significantly different from one-sided critical values, giving the scarcity of rejections in the upper tail of the density. For values of θ well within the unit circle, the size distortions disappear.

Second, the distortions are a decreasing function of T although this only holds when the series are near-integrated. For $\theta = 1$ there will be no reductions in size distortions when the sample size increases.

Finally, the distortions are an increasing function of the number of *nuisance* parameters (such as constant or trend) estimated. Thus the distortions in model (6) are higher than those in model (5). This is again a characteristic feature of such densities.

In summary, the results in this section again show the dangers of conducting inference using standard tables when unit roots are present. This leads naturally to the conclusion that tests of economic hypotheses of interest, such as the excess sensitivity of consumption to income, crucially depend on a pre-classification of the variables of interest by their orders of integration, as this determines what critical values should be used for inference. Incorrect pre-classifications will lead to an inappropriate choice of critical values and thus lead to incorrect inferences.

For most of the next section we duck this issue of possible incorrect classification and proceed conditionally upon a classification of the orders of integration of the variables. We ask the question whether, given that some of the variables have been classified as I(1), it is possible under any circumstances to return to using standard tools of inference and whether such a return has any other practical benefits such as unbiased or efficient estimation. We then return briefly to the issue of ameliorating the consequences of incorrect classification by being conservative in our testing strategy.

3. Dynamic Regressions

3.1 Overview

The fundamental point, which is the first major theme of this survey, emerging from the discussion above is the striking difference which may arise between the critical values required to conduct inference in a stationary environment and those required when unit roots are present, and, as a corollary, the mistakes which can occur if incorrect critical values are used.

The classification of time series by their integration properties, say, into I(0) and I(1), is an area fraught with difficulty. There is a vast literature just on *testing* for unit roots, with a wide variety of tests proposed. Each of these tests may have satisfactory power properties against a given set of alternative hypotheses but may be powerless against a range of other alternatives.

However, given that the classification has been properly made, series integrated, say, of order one and *related* to each other seem to offer a special advantage to the applied econometrician. The asymptotic property which confers this advantage is called "super-consistency" and discussion of this property takes us to the second important theme of the literature on I(1) processes, namely the modelling of long-run economic relationships by means of static regressions.

Engle and Granger (1987) introduced the notion of "co-integration" to the mainstream of applied and theoretical econometric research. The idea is simple yet very powerful and can be understood by looking at a simple DGP (taken from Engle and Granger):

$$y_t + \beta x_t = u_t \tag{7a}$$

$$y_t - \alpha x_t = e_t \tag{7b}$$

$$u_t = u_{t-1} + \varepsilon_{1t} \tag{7c}$$

$$e_t = \rho e_{t-1} + \varepsilon_{2t}, \text{ with } |\rho| < 1; \tag{7d}$$

$(\varepsilon_{1t}, \varepsilon_{2t})'$ is distributed identically and independently as a bivariate normal with zero means, finite variances, and zero covariance.

From (7a) - (7b), x_t and y_t can be expressed as linear combinations of the error processes u_t and e_t (where the weights in the linear combinations are functions of α and β). Thus, both x_t and y_t are weighted sums of an I(1) variable and an I(0) variable and are therefore both I(1). Yet a linear combination of x_t and y_t, given by (7b), is I(0). In the terminology of Engle and Granger the two series x_t and y_t are said to be *co-integrated* with each other. In slightly more formal terms, two I(d) series are said to be co-integrated with each other if a linear combination of these two series is integrated of order d - b, where b > 0; *i.e.* linearly combining the two series leads to a series of a lower order of integration. In the example above x_t and y_t are said to be co-integrated of order 1 and 1, denoted CI(1, 1), where the first 1 gives the order of the component series and the second the reduction in the order of integration. Thus two series are CI(d, b) if each series is individually I(d) while the linear combination is I(d-b).

For the most part in this chapter we will focus on reductions from I(1) to I(0). The concept of co-integration applies to multivariate (greater than two) systems of variables where the non-uniqueness of the co-integrating relationship becomes an important issue.

The concept of co-integration gains importance from the fact that the statistical properties of the composite variable (which is I(0)) are so dramatically different from the properties of the component series (both of which are I(1)). Thus series growing stochastically over time are said to be linked together in the long-run or co-integrated if a linear combination of these series remains bounded in a statistical sense.[5] Co-integration captures the notion of long-run relationships in economics, such as consumption being a fraction of permanent income or purchasing power parity, which are testable on data sets albeit with non-standard tools. Co-integration allows for possibly extensive divergences in the short-run but because a stable relationship among variables cannot be meaningfully said to exist if these divergences persist in the long-run, it uses the criterion of stationarity in a series such as $\{e_t\}$ in (7b), or its estimated counterpart $\{\hat{e}_t\}$, to define the existence of a long-run or "equilibrium" relationship. Tests for co-integration therefore reduce to testing for unit roots in the estimated series $\{\hat{e}_t\}$, introducing all the difficulties in testing for unit roots into the literature on testing for co-integration.

In an important sense, the concept of co-integration is the natural opposite to the concept of spurious regressions. If, say, a bivariate static regression is spurious, there does not exist a stationary linear combination of the variables. If a stationary linear combination does exist the regression is the co-integrating regression.

Looking more closely at the asymptotic properties of a co-integrating regression, the difficulty of non-standard distributions remains. However, the non-standardness has several important implications.

First, consider estimating (7b) as the co-integrating regression. Then,

$$(\hat{\alpha} - \alpha) = \left[\sum_{t=1}^{T} x_t^2\right]^{-1} \left[\sum_{t=1}^{T} x_t e_t\right]. \qquad (8)$$

Earlier in the chapter we noted that the variance of x_t, if x_t is generated by a unit root process, grows with t. From this, and from the properties of e_t, it is possible to see after some tedious manipulation that the variance of the numerator of (8) is of order T^2 while the denominator has variance of order T^4. Thus to prevent explosive (or degenerate) behaviour of both the numerator and the denominator, they need to be scaled by T and T^2 respectively.[6] Simplifying, it is then clear that $T(\hat{\alpha} - \alpha)$ has a non-degenerate distribution.

The scaling is one important way in which the presence of unit roots alters the asymptotic theory of the distributions of estimators even in models as simple as (8). $\hat{\alpha}$ tends in probability to α, denoted $\hat{\alpha} \xrightarrow{p} \alpha$ at rate T, instead of the usual $T^{1/2}$ [7], the rate of convergence of consistent estimators in stationary asymptotic theory. This is known as "super-" or "T-" consistency, as distinct from $T^{1/2}$-consistency.

The integratedness of the series feeds into the distribution of the estimator in another important respect. The denominator of (8) does not tend in probability to a finite limit but has an asymptotic distribution. Similarly, the distribution of the numerator of (8) is not asymptotically normal. Standard central limit theorems do not apply because of the non-stationarity of x_t.

Finally, consider taking the Cochrane-Orcutt transform of (7b). This yields

$$y_t = \alpha x_t + \rho(y_{t-1} - \alpha x_{t-1}) + \varepsilon_{2t}. \tag{9}$$

Equation (8) differs from (9) only in the former's omission of an I(0) term given by $(y_{t-1} - \alpha x_{t-1})$ (I(0) because y_t and x_t are co-integrated), which thereby enters the residual of the regression in (8). Nevertheless, a remarkable consequence of T-consistency is that the omission of I(0) terms in the model does not affect the consistency of the coefficient estimator on an I(1) variable. Thus, with a long data series, a static regression such as (8) is enough to obtain a consistent estimate of the long-run relationship between the variables. However, the finite-sample and asymptotic distributions of the coefficient estimator is affected by the presence of un-modelled terms, an observation which is of some importance in our later discussion. This is also a point taken up by Zivot in his comment which follows this chapter.

Engle and Granger's paper emphasized the value of static regressions in an environment where the processes are integrated of order 1 and recommended modelling such integrated processes in two stages. In the first stage, the long-run relationship is estimated *via* static regressions. At the second stage, the dynamics of the model, all parameterized as I(0) variables, are estimated. Thus in an I(1) environment, static regressions formed, according to this line of argument, an important part of good modelling practice. How much of this recommendation still holds true, in the light of subsequent research, is a matter of some debate. Several interlinked issues are important in any evaluation of this point. These include, in particular, issues concerning the distributions of the coefficient estimators, biases in these estimates and tests for co-integration based on these estimates. We address all these issues in turn.

3.2 Asymptotic Theory

The non-standardness of the distributions, both in spurious and non-spurious regressions, arises, in the main, from the coefficient of interest being a coefficient on a variable integrated of order 1 (or higher). From this follows a central observation, due to Sims, Stock and Watson (1990) *inter alia* (for related analysis, see Banerjee and Dolado (1988) and Stock and West (1988)) that if the coefficients of interest can be written as coefficients on I(0) variables, by means of suitable linear transformations of the variables in the original regression equation, then standard asymptotic tools can be used to conduct inference. In essence this implies that the regression equation is rich enough to allow transformations such as differencing of the variables and linearly combining lagged levels of the variables - *i.e.* a sufficiently rich dynamic specification is required.

Consider a test of the permanent income hypothesis of the form given by (5) but augmented to include lags of y_t and c_t, where c_t is consumption in period t and y_t is disposable income. If the permanent income hypothesis is correct, y_t is co-integrated with c_t and since c_t has a unit root, so does y_t. Further, in the regression equation

$$\Delta c_t = \beta c_{t-1} + \pi_1 y_{t-1} + \pi_2 y_{t-2} + \ldots + \pi_p y_{t-p} + u_t \qquad (10),$$

the permanent income hypothesis, which is taken to be the null hypothesis implies, first, that $\beta = 0$ and, second, that $\pi_i = 0$ (i=1 ...p). Take the latter as the main hypothesis of interest, and to simplify matters further, suppose that β is correctly imposed at its true value of 0. Thus, as in (5), we are interested in testing for the excess sensitivity of consumption to income. Then a rewriting of (10) yields

$$\Delta c_t = (\pi_1 + \pi_2 + \ldots + \pi_p)k + (\pi_1 + \pi_2 + \ldots + \pi_p)c_{t-1}$$
$$+ \pi_1(y_{t-1} - c_{t-1} - k) + \ldots + \pi_p(y_{t-p} - c_{t-p} - k) + u_t,$$
$$= m + \phi c_{t-1} + \pi_1(y_{t-1} - c_{t-1} - k)$$
$$+ \ldots + \pi_p(y_{t-p} - c_{t-p} - k) + u_t,$$

where k is the intercept of the long-run consumption function, possibly equal to 0, $m = k\Sigma_{i=1}^{p}\pi_i$, and $\phi = \Sigma_{i=1}^{p}\pi_i$. In this rewriting, the coefficients on all the disposable income variables, $\pi_1,\ldots,\pi_p$, have been expressed as coefficients on I(0) variables (given that y_t and c_t are co-integrated). Because it is possible to achieve this rewriting, the distributions of the coefficient estimators of $\pi_1,\ldots,\pi_p$, denoted $\hat{\pi}_1,\ldots,\hat{\pi}_p$, are individually and jointly asymptotically normally distributed. Thus standard normal tables can be used to test for significance of the individual $\hat{\pi}_i$s while *F*-tables can be used to conduct tests of joint significance (asymptotically). This therefore represents a considerable simplification of the process of inference although it is based, importantly, on a proper classification of the integration and co-integration properties of the variables concerned. This analysis alone is enough to reinstate the use of dynamic regressions and should lead to a focus away from static regressions. But there are at least two other important reasons, the first concerning biases in estimates of the co-integrating relation and the second linked with the issue of testing for co-integration.

3.3 *Biases*

The discussion in this section is based on the important, but clearly not always realistic, assumption that single-equation methods are statistically valid (and efficient) for estimating the parameters of interest. The parameters of interest include those describing short-run behaviour and also those giving the long-run relationships among the variables. The claim in this section is that, even when single-equation estimation methods suffice,[8] both the long-run and the short-run parameters are better estimated using dynamic methods. Thus the important notion, that the errors should form a martingale difference sequence[9] in a well-specified model, continues to hold true when dealing with non-stationary series. Therefore, in finite samples, the biases in the estimates of the long-run parameters, introduced by not explicitly modelling the dynamic I(0) terms, which thus enter the residuals of a static regression, are considerable.[10]

We illustrate the argument with an example taken from Banerjee, Dolado, Galbraith and Hendry (1993) (referred to henceforth as Banerjee *et al.* (1993)). While an important charge that can be made against any example is that the DGP is too special, the results here are representative of a large number of Monte Carlo results (Hendry and Neale (1987), Stock (1987), Phillips and Hansen (1990)).

Suppose the series $\{y_t\}$ and $\{x_t\}$ are generated by the following process:

$$y_t = \gamma_1 y_{t-1} + \gamma_2 x_t + \gamma_3 x_{t-1} + \varepsilon_{1t} \qquad (11a)$$

$$x_t = x_{t-1} + \varepsilon_{2t}; \qquad (11b)$$

$$\varepsilon_{1t} \sim NID(0, \sigma_1^2), \ \varepsilon_{2t} \sim NID(0, \sigma_2^2), \ cov(\varepsilon_{1t}, \varepsilon_{2s}) = 0, \ \forall\ t, s;$$

$$\gamma_1 + \gamma_2 + \gamma_3 = 1.$$

Thus, both $\{y_t\}$ and $\{x_t\}$ are I(1) series and are CI(1, 1), with a long-run multiplier of 1 linking the two series.[11] Further, $\{x_t\}$ is strongly exogenous for the regression parameters.

Consider now estimating the long-run by means of the static regression:

$$y_t = \alpha x_t + u_t. \tag{12}$$

The omitted dynamic I(0) terms are given by $(y - x)_{t-1}$ and Δx_t[12] and are included in the residual u_t which, in general, is therefore serially correlated. The data are generated using the specification in (11) above. The strong exogeneity of x_t is ensured by drawing the $\{\varepsilon_{1t}\}$ and $\{\varepsilon_{2t}\}$ series from uncorrelated pseudo-normal distributions. The γ_is are chosen to preserve homogeneity and sample sizes ranging from 25 to 400 are considered. The ratio of the standard deviations of ε_{1t} and ε_{2t} are also varied and each parameter configuration is run 5000 times and the results noted. Table II below, taken from Banerjee *et al.* (1993) summarises the estimated mean biases in the static model:

Table II Biases in Static Models[a,b]

DGP: (11a) + (11b); 5000 replications

	Sample size (T)				
	25	50	100	200	400
$\gamma_1 = 0.9$, $\gamma_2 = 0.5$, $\sigma_1/\sigma_2 = 3$	-0.39	-0.25	-0.15	-0.07	-0.04
$\gamma_1 = 0.9$, $\gamma_2 = 0.5$, $\sigma_1/\sigma_2 = 1$	-0.32	-0.22	-0.14	-0.08	-0.04
$\gamma_1 = 0.5$, $\gamma_2 = 0.1$, $\sigma_1/\sigma_2 = 3$	-0.23	-0.13	-0.07	-0.03	-0.02
$\gamma_1 = 0.5$, $\gamma_2 = 0.1$, $\sigma_1/\sigma_2 = 1$	-0.21	-0.12	-0.06	-0.03	-0.02

Source: Banerjee *et al.* (1993), Table 7.3.

[a]Standard errors of these estimates vary widely, but the estimated biases are in almost all cases significantly different from zero. The biases do not decline at rate T but they do decline more quickly than $T^{1/2}$. The simulations used GAUSS.

[b]The mean biases are computed as $\left[(5000)^{-1}\sum_{i=1}^{5000}(\hat{\alpha}_i - 1)\right]$ where i is the index for the replications for each parameter configuration.

The significance of the results reported in Table II become clear when these are compared with the biases arising from estimating a dynamic model corresponding to (12), say, regressing y_t not only on x_t but also on y_{t-1} and x_{t-1} which, expressed in error-correction form[13], is given as (13) below:

$$\Delta y_t = b\Delta x_t + c(y_{t-1} - x_{t-1}) + dx_{t-1} + \nu_t. \qquad (13)$$

The extra lagged term is included to avoid imposing homogeneity on the relation between y and x. Although homogeneity would be a valid restriction in this case, *i.e.* d = 0, the extra term allows for the possible ignorance of the investigator and, as results reported in Banerjee, Galbraith and Dolado (1990) show, does not affect the estimate of the short-run adjustment coefficient c. The estimate of the long-run multiplier is deduced from (13) as $(1 - \hat{d}/\hat{c})$.[14] For the same configuration of parameters given in Table II, the biases in the estimate of the long-run parameter derived from (13) are all insignificantly different from zero.

It is possible to extend this set of experiments to allow for weakly exogenous x_t. The results carry through in this case. However, as we show in a later section, if weak exogeneity fails to hold, the usefulness of estimates from dynamic single equations is reduced substantially and the comparison between static and dynamic estimates becomes ambiguous. Systems estimation is the main route to consider in the absence of weak exogeneity and Monte Carlo evidence here (see, for example, Gonzalo (1994)) again supports the claim of substantial inferiority, in general, of estimates derived from static regressions.

Finally, it is important to note that it is possible to derive analytical expressions which explain the difference in the accuracy of the estimates using alternative methods. From these analytical expressions, it is possible to derive the parameter configurations or DGP specifications for which, say, static regressions are likely to out-perform (or perform as well as) dynamic regressions or *vice versa* (see Kremers, Ericsson and Dolado (1992)).

3.4 Testing for Co-integration

This paper has taken as its main theme the importance of dynamic regressions in a non-stationary environment. This theme has been emphasized *via* several sub-themes discussing the particular benefits of conducting estimation and testing in a dynamic setting. This section discusses the third sub-theme of this line of argument, testing for co-integration and discusses, first, using a specific simple example and then more generally, tests of co-integration in dynamic models.

A large class of tests for co-integration (Phillips and Ouliaris (1990)) takes as its starting point a static regression such as (12) and tests for unit roots in the estimated residual series $\hat{u}_t$. A popular test in this category is known as the Augmented Dickey-Fuller test and consists of estimating the regression

$$\Delta\hat{u}_t = (c) + \beta(t) + \rho\hat{u}_{t-1} + \sum_{i=1}^{\ell} \delta_i \Delta\hat{u}_{t-i} + \omega_t, \qquad (14)$$

and testing for the significance of the estimate of ρ. A significant $\hat{\rho}$, according to appropriate critical values, constitutes a rejection of the null of a unit root in the residual series and hence provides *prima facie* evidence of the existence of co-integration.

The critical values used for this test are essentially modified Dickey-Fuller critical values. The unmodified critical values, given in Fuller (1976), apply to testing for unit roots in raw series. These critical values have to be adjusted for size when the series to be tested for a unit root is constructed or derived from a regression such as (12). Naturally this implies that the critical values are sensitive to the number of variables in the co-integrating regression and to the existence of a constant or a trend in the regression.[15] An extensive set of tables is provided by MacKinnon (1991), while Granger and Engle (1987) and Engle and Yoo (1987) provide a more limited set.

Return now to the experiment described in (11). The parameter configuration given by $\gamma_1 = 1$, $\gamma_2 = \gamma_3 = 0$ generates the case where y_t and x_t are not co-integrated. Equation (11a) can be written in error correction form as

$$\Delta y_t = \gamma_2 \Delta x_t + (\gamma_1 - 1)(y_{t-1} - x_{t-1}) + \varepsilon_{1t}. \quad (15)$$

Thus a test for co-integration can be based on the $\underline{t}$-statistic for $\hat{c}$ in (13) because under the usual null hypothesis of "no co-integration", c = 0 from (15). The distribution of this test statistic is non-standard because under H_0, $(y_{t-1} - x_{t-1})$ is I(1) while both differenced terms are I(0) and the regression, in the terminology discussed above, is inconsistent. However the test is straightforward and would be useful if it had good power properties.

It is interesting to remark that under the alternative hypothesis of co-integration, $\underline{t}_{c=0}$ is asymptotically normally distributed. This follows from the property that in this case c is a coefficient on an I(0) variable and the arguments discussed in section *3.2* apply.

The $\underline{t}_{c=0}$ test is a simple example of what Boswijk and Franses (1992) call a Wald test for co-integration and is based on a dynamic regression model. It is instructive to compare the power properties of this test with an ADF(1)[16] test based on the residuals of the static model (12). The critical values for both these tests are derived by simulating the model under the null (*i.e.* $\gamma_1 = 1$, $\gamma_2 = \gamma_3 = 0$, $s = \sigma_1/\sigma_2 = 1$ in (11a)-(11b)) for 5000 replications. These critical values are then used for deriving the test powers, when the null hypothesis of no co-integration is false, of the $\underline{t}_{c=0}$ test and the

ADF(1) test. Critical values are reported in Table IIIa below while powers of the tests are given in Table IIIb. Both tables are taken from Banerjee *et al.* (1993) and the simulations used PC-NAIVE (Hendry, Neale and Ericsson (1990)).[17]

Table IIIa Fractiles of $\underline{t}_{c=0}$ in (13) and ADF(1) in (14)
DGP: (11a) + (11b) with $\gamma_1 = 1$, $\gamma_2 = \gamma_3 = 0$, $s = \sigma_1/\sigma_2 = 1$
5000 replications

	Fractiles of $\underline{t}_{c=0}$			Fractiles of ADF(1)		
T	0.10	0.05	0.01	0.10	0.05	0.01
25	-2.99	-3.42	-4.22	-3.15	-3.51	-4.30
50	-2.95	-3.33	-4.06	-3.10	-3.41	-4.08
100	-2.93	-3.28	-3.95	-3.09	-3.39	-4.00

Table IIIb Test rejection frequencies in ECMs and ADFs
DGP: (11a) + (11b); Model: (13), (14)
5000 replications

	Estimated power at given fractile		
	0.10	0.05	0.01
	$t_{c=0}$/ADF	$t_{c=0}$/ADF	$t_{c=0}$/ADF
(a) $\gamma_1 = 0.9$, $\gamma_2 = 0.5$, $\sigma_1/\sigma_2 = 3$			
T = 25	0.13/0.13	0.06/0.06	0.01/0.01
50	0.21/0.17	0.10/0.10	0.02/0.02
100	0.44/0.31	0.26/0.20	0.07/0.05
(b) $\gamma_1 = 0.9$, $\gamma_2 = 0.5$, $\sigma_1/\sigma_2 = 1$			
T = 25	0.14/0.11	0.06/0.05	0.01/0.01
50	0.21/0.15	0.10/0.09	0.02/0.02
100	0.49/0.30	0.30/0.19	0.08/0.04
(c) $\gamma_1 = 0.9$, $\gamma_2 = 0.5$, $\sigma_1/\sigma_2 = 1/3$			
T = 25	0.13/0.10	0.07/0.05	0.02/0.01
50	0.24/0.13	0.12/0.07	0.03/0.01
100	0.59/0.24	0.40/0.14	0.13/0.03
(d) $\gamma_1 = 0.5$, $\gamma_2 = 0.1$, $\sigma_1/\sigma_2 = 3$			
T = 25	0.66/0.35	0.45/0.20	0.16/0.05
50	0.99/0.84	0.97/0.72	0.78/0.34
100	1.00/1.00	1.00/1.00	1.00/0.97
(e) $\gamma_1 = 0.5$, $\gamma_2 = 0.1$, $\sigma_1/\sigma_2 = 1$			
T = 25	0.79/0.31	0.66/0.18	0.29/0.04
50	1.00/0.80	1.00/0.67	0.94/0.28
100	1.00/1.00	1.00/1.00	1.00/0.96
(f) $\gamma_1 = 0.5$, $\gamma_2 = 0.1$, $\sigma_1/\sigma_2 = 1/3$			
T = 25	0.94/0.23	0.87/0.12	0.64/0.03
50	1.00/0.75	1.00/0.60	1.00/0.22
100	1.00/1.00	1.00/1.00	1.00/0.94

Several issues may now be noted. First, the most significant divergences in the powers of the two tests appear in the last three blocks of Table IIIb and may be understood in terms of the common-factor restriction imposed by an ADF(1) type of test. The ADF(1) test involves testing $\gamma_1 = 1$ in the regression

$$\Delta(y_t - \alpha x_t) = (\gamma_1 - 1)(y_{t-1} - \alpha x_{t-1}) + \delta\Delta(y_{t-1} - \alpha x_{t-1}) + \omega_t, \quad (16)$$

where α, which here has a true value of unity, may be replaced by its estimate $\hat{\alpha}$ from the first-step regression of y_t on x_t. In (16) therefore the dynamics of the model implicitly impose the short-run elasticity to equal its long-run value of 1. In (13) however, the coefficient on Δx_t is unrestricted. When the common-factor restriction is far from being satisfied (γ_2 is very different from 1), as in the last three blocks of the table, the performance of the ADF statistic, relative to $\underline{t}_{c=0}$, is very poor.

Second, calculations in Banerjee *et al.* (1993) provide estimates of the non-centrality of both test statistics for fixed alternatives. For T = 25, these are given in Table IIIc below and show that the non-centralities of the ECM test are substantially greater, in most cases, than the corresponding non-centralities for the ADF(1) test.

Table IIIc Non-centrality (NC) of $\underline{t}_{c=0}$ and ADF(1)
DGP: (11a) + (11b), 5000 replications

Case	(a)	(b)	(c)	(d)	(e)	(f)
NC_{ADF}	-1.15	-1.15	-1.15	-2.89	-2.89	-2.89
NC_{ECM}	-1.19	-1.28	-1.52	-3.25	-3.88	-5.32

The relative magnitudes of the non-centralities help further to explain the performances of the two tests and again provide reasons for not modelling dynamics restrictively, both in general and where common factor restrictions are likely to be violated.

Third, to illustrate the issue of a conservative testing strategy, consider the choice of critical values to test for significance of $\hat{c}$ in (13) using the ECM test. In Table IIIa we require *t*-statistics greater than, roughly, 3.5 in absolute magnitude to reject the null of no co-integration at the 5% confidence level. Let us call these the I(1) critical values. If the variables in the regression had been integrated of order 0, *t*-values in excess of 2.0 in absolute value would have sufficed. An I(0) critical value is therefore given by 2.0. Four possible situations are evident, labelled S1 to S4, when H_0: c = 0 is true:

Table IV

	Order of integration	Critical values used	Rejection of H_0(%)
S1	I(1)	I(1)	5
S2	I(1)	I(0)	> 5
S3	I(0)	I(0)	5
S4	I(0)	I(1)	< 5

A conservative strategy would be one which sets the *maximum* of the rejections probability of the true null equal to the nominal confidence level of the test (here 5%). Looking at Table IV, this requires the use of the I(1) critical values, although this may entail a loss of power (if the variables are I(0)).

The conservative strategy may be adopted as a crude device to minimise the consequences of *incorrect* classifications. However it is not a strategy which attacks the pre-classification issue directly. Papers by Elliott and Stock (1992), Stock (1992) and Phillips and Ploberger (1991) propose Bayesian methods of classification and inference in models when the orders of integration of the variables are unknown. A description of these methods is beyond the scope of this paper.

Finally, the logic of the argument in favour of the ECM test generalises to cases where the regressors are only weakly exogenous for the parameters of interest. Boswijk and Franses (1992) propose a Wald test for co-integration identical in spirit to the one developed above. In the simplest case, they consider the case where there is at most one co-integrating vector and the vector $\underset{\sim}{\mathbf{x}}_t$ is weakly exogenous for the co-integration parameters. Boswijk and Franses' test involves writing a conditional error correction model for the dependent variable y_t given $\underset{\sim}{\mathbf{x}}_t$ (where linearity of the conditional model follows from the assumption of normality of the joint distribution of $\underset{\sim}{\mathbf{z}}_t = (y_t:\underset{\sim}{\mathbf{x}}_t')'$):

$$\begin{aligned}\Delta y_t &= c + \underset{\sim}{\phi}_0'\Delta\underset{\sim}{\mathbf{x}}_t + \lambda(y_{t-1} - \underset{\sim}{\theta}'\underset{\sim}{\mathbf{x}}_{t-1}) + \sum_{j=1}^{p-1}(\psi_j\Delta y_{t-j} + \underset{\sim}{\phi}_j'\Delta\underset{\sim}{\mathbf{x}}_{t-j}) + \eta_t \\ &= c + \underset{\sim}{\phi}_0'\Delta\underset{\sim}{\mathbf{x}}_t + \underset{\sim}{\pi}'\underset{\sim}{\mathbf{z}}_{t-1} + \sum_{j=1}^{p-1}(\psi_j\Delta y_{t-j} + \underset{\sim}{\phi}_j'\Delta\underset{\sim}{\mathbf{x}}_{t-j}) + \eta_t, \qquad (17)\end{aligned}$$

where $\underset{\sim}{\pi}' = \lambda(1,-\underset{\sim}{\theta}')$. (17) is a generalisation of the test given in (15). As before, $\lambda = 0$ implies that there is no co-integration while if $\lambda \neq 0$, (17) can be reparameterised as an autoregressive distributed lag model of order p with y_t and $\underset{\sim}{\mathbf{x}}_t$ co-integrated. If $\underset{\sim}{\theta}$ were known, a test for co-integration could be based on a $\underline{t}$-test of $\lambda = 0$. In general, this test is not implementable because $\underset{\sim}{\theta}$ is not known but must be estimated. Thus the regression must either be reparameterised as

$$\begin{aligned}\Delta y_t = c + \underset{\sim}{\phi}_0'\Delta\underset{\sim}{\mathbf{x}}_t + \lambda(y_{t-1} - \underset{\sim}{\iota}'\underset{\sim}{\mathbf{x}}_{t-1}) + \underset{\sim}{\zeta}'\underset{\sim}{\mathbf{x}}_{t-1} \\ + \sum_{j=1}^{p-1}(\psi_j\Delta y_{t-j} + \underset{\sim}{\phi}_j'\Delta\underset{\sim}{\mathbf{x}}_{t-j}) + \eta_t, \qquad (18)\end{aligned}$$

where $\underset{\sim}{\iota}$ is a vector (of dimension equal to the dimension of $\underset{\sim}{\mathbf{x}}_t$) whose elements are all equal to 1 and $\underset{\sim}{\zeta}' = \lambda(\underset{\sim}{\iota}' - \underset{\sim}{\theta}')$, and the $\underline{t}$-test now based on the $\underline{t}$-statistic for $\hat{\lambda}$ in (18), or, using the result that $\lambda = 0$ implies $\pi = 0$, based on a Wald-type statistic of the form

$$\text{Wald} = \hat{\pi}'[\hat{V}(\hat{\pi})]^{-1}\hat{\pi}, \qquad (19)$$

where $\hat{\pi}$ is the OLS estimator of π in (17), with estimated variance matrix $\hat{V}(\hat{\pi})$. Boswijk and Franses' results suggest that a test of the form (19) is likely to have better statistical properties than the one given by (18).

Boswijk and Franses present power calculations comparing the power of the Wald test with an ADF test based on the residuals of the static regression. In this framework, selection of lag length becomes an issue of some importance. For underparameterised models, relative to the DGP, the Wald test may have incorrect size (too many rejections of the true null hypothesis at any nominal confidence level) while over-parameterisation leads to loss in power of the test. Using various "optimal" lag length selection criteria, Boswijk and Franses show that the Wald test is superior to the ADF test (and to a test based on Johansen's (1988) procedure discussed below). Also, because both λ and θ can be retrieved from the estimate of π, after suitable transformation, the evidence from the simpler DGP (11(a) - (11b)) suggests that the long-run is also likely to be better estimated in a dynamic model such as (17).

4. Systems Estimation

It is natural to ask, given that the discussion above has focused exclusively on single-equation estimation techniques, how the arguments generalise to estimation in systems. In essence, this is equivalent to asking what happens when the regressors are not weakly exogenous for the co-integrating parameters. Based on our discussion so far, it should not be surprising to observe that the long-run relation is poorly estimated not only in static regressions but that single-equation dynamic models in general do not perform much better.

There are at least four interrelated issues in single-equation estimation which are worth highlighting. First, the presence of unit roots introduces non-standard distributions of the coefficient estimates. Second, the errors may be processes which are autocorrelated. Thirdly, in a multivariate setting, there can exist several co-integrating relations and there is no longer a natural ordering of the variables (which dependent and which independent) in a static regression.[18] Finally, the explanatory variables in the single equation may not be weakly exogenous for the parameters of interest.

In some of the discussion above, we have provided examples of how dynamic regressions can overcome some of the problems posed by the first two effects. However, for most purposes of empirical modelling the question ultimately boils down to a discussion of the circumstances in which only systems-estimation provides efficient and unbiased estimates.

The answer in this non-stationary world is not dramatically different from the answer one would have given, say, fifteen years ago - when weak-exogeneity is violated, estimating only the conditional model is suboptimal, leading to biased, inconsistent and inefficient estimates. Thus the debate which raged in the seventies on the need to have *at least* weakly exogenous regressors to conduct estimation and inference in single-equation models remains an issue of considerable importance.

A discussion of estimating co-integrated systems requires us first to provide a formal statement of a co-integrated system:

Definition (Engle and Granger 1987). The components of the vector $\tilde{x}_t$ are said to be co-integrated of order d, b, denoted $\tilde{x}_t \sim CI(d, b)$ if (i) each component of $\tilde{x}_t$ is I(d) and (ii) there exists a non-zero vector $\tilde{\alpha}$ such that $\tilde{\alpha}'\tilde{x}_t \sim I(d-b)$, $d > b \geq 0$. The vector $\tilde{\alpha}$ is called the co-integrating vector.

As we have stated before, if $\tilde{x}_t$ has $n > 2$ components then it is possible for r linearly independent (equilibrium) relationships to govern the evolution of the variables. It may be shown that $0 \leq r \leq n - 1$. The r linearly independent co-integrating vectors can be gathered into an $n \times r$ matrix $\tilde{\alpha}$ with rank r.

The Granger Representation Theorem (Engle and Granger (1987)) shows that a co-integrated system can be written in several equivalent forms (see Engle and Granger (1987), Banerjee _et al._ (1993)). We focus on one of these forms, the ECM form, to illustrate the importance of estimation in systems. In ECM form, the system can be written as

$$\tilde{A}(L)(1 - L)\tilde{x}_t = -\tilde{\gamma}\tilde{z}_{t-1} + \tilde{\omega}_t, \qquad (20)$$

where $\tilde{\omega}_t$ is a stationary multivariate disturbance, $\tilde{A}(L)$ is a stationary lag-polynomial matrix, with $\tilde{A}(0) = \tilde{I}_n$ and $\tilde{A}(1)$ finite, $\tilde{z}_t = \tilde{\alpha}'\tilde{x}_t$ and $\tilde{\gamma}$ is a non-zero matrix with rank r.

Thus, because $\tilde{\alpha}$ is also a reduced-rank matrix of rank r, the reduced rank (r) of $\tilde{\pi} = \tilde{\gamma}\tilde{\alpha}'$ is the restriction implied by co-integration in the system given by (20).

Consider the case where n = 2 and r = 1, let $\underset{\sim}{\alpha}' = (1, -\kappa)$, and let the system be given by

$$x_{1t} - \kappa x_{2t} = u_{1t},$$
$$\Delta x_{2t} = u_{2t}, \qquad (21)$$

where $(u_{1t}, u_{2t})'$ follow a jointly stationary process. Then, in ECM form, the system can be written as

$$\Delta x_{1t} = d_{12}(L)\Delta x_{2t} + d_{11}(L)\Delta x_{1t-1} + \gamma_1(x_{1t-k} - \kappa x_{2t-k}) + e_{1t},$$
$$\Delta x_{2t} = d_{21}(L)\Delta x_{1t} + d_{22}(L)\Delta x_{2t-1} + \gamma_2(x_{1t-k} - \kappa x_{2t-k}) + e_{2t}. \qquad (22)$$

Weak exogeneity will be violated in (22) if, say, $\gamma_1\gamma_2 \neq 0$[19] and thus the error-correction term (which Engle and Granger proposed estimating at the first step by a static regression) enters both equations. The x_2 process therefore contains information about the process generating x_1 and therefore, in the absence of *a priori* information of the form, say, $\gamma_1\gamma_2 = 0$,[20] systems-based methods of estimation are essential. Hendry (1992) provides several examples of how violations of weak-exogeneity leads to problems of bias in single-equation based estimators.

It is impossible, in the space available, to provide an account of the various methods proposed of estimating co-integrated systems. Two main methods of estimation have been proposed, one due to Phillips (1991) which uses the triangular or canonical form of the system given by (21) and the other due to Johansen (1988) which estimates the model in the ECM form given by (22)[21]. Estimation focuses on several issues, especially the co-integrating rank given by the rank of $\underset{\sim}{\pi}$ (or equivalently the number of co-integrating vectors), the economic interpretability of the linear equilibrium relationships and the testing of any special economic hypotheses of interest, where the latter are given by restrictions on the parameters of the equlibrium relationships. As in the usual analysis of simultaneous equation systems, it is also necessary to take account of restrictions which serve to identify the system.

5. Conclusion

We have taken as our main theme of this paper the observation that the fundamental methods *per se* of econometric modelling, for specification and estimation, remain unaltered in the new world of non-stationary econometrics. Some of the dramatic simplifications which appeared possible, in terms of, say, modelling the long-run and short-run separately, or focusing primarily on the long-run properties of and inter-relationships between series, seem eventually to be fraught with difficulties of tests having low powers and estimators having large biases.

We have also argued that, even for modelling the long run or testing for co-integration, dynamic models provide the most effective way of obtaining information. Where weak exogeneity is violated, single-equation estimation techniques are, in general inadequate and systems-based methods are optimal. Thus the importance of this literature lies rather in the new awareness we have of the properties of common time series and of the consequent need to take account of these properties in estimation and to modify, wherever necessary, the tools required to conduct inference.

References

BANERJEE, A. and DOLADO, J. (1988), "Tests of the Life-Cycle-Permanent Income Hypothesis in the Presence of Random Walks: Asymptotic Theory and Small Sample Interpretations", *Oxford Economic Papers*, 40: 610-633.

—— —— GALBRAITH, J.W. and HENDRY, D.F. (1993), *Co-integration, Error Correction and the Econometric Analysis of Non-stationary Data*, Oxford University Press.

—— GALBRAITH, J.W. and DOLADO, J. (1990), "Dynamic Specification with the General Error-Correction Form", *Oxford Bulletin of Economics and Statistics*, 52: 95-104.

—— and HENDRY, D.F. (1992), "Testing Integration and Co-integration", *Oxford Bulletin of Economics and Statistics*, 54: 225-235.

—— DOLADO, J. and MESTRE, R. (1993), "On Some Tests for Co-integration: The Cost of Simplicity", typescript.

BOSWIJK, H.P. and FRANSES, P.H. (1992), "Dynamic Specification and Cointegration", *Oxford Bulletin of Economics and Statistics*, 54, 369-381.

CAMPBELL, J.Y. and PERRON, P. (1991), "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots", in O.J. Blanchard and S.Fischer (eds.), *NBER Economics Annual 1991*, MIT Press.

DAVIDSON, J.E.H., HENDRY, D.F., SRBA, F., and YEO, S. (1978), "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom", *Economic Journal*, 88: 661-692.

DOLADO, J., JENKINSON, T.J. and RIVERO S-S. (1990), "Cointegration and Unit Roots", *Journal of Economic Surveys*, 4: 249-273.

ELLIOTT, G. and STOCK, J.H. (1992), "Inference in Time Series Regression when the Order of Integration of a Regressor is Unknown", *National Bureau of Economic Research*, Technical Working Paper No. 122.

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

—— HENDRY, D.F. and RICHARD, J-F. (1983) "Exogeneity", *Econometrica*, 55: 277-304.

—— and YOO, B.S. (1987), "Forecasting and Testing in Cointegrated Systems", *Journal of Econometrics*, 35: 143-159.

FAMA, E.F. and FRENCH, K.R. (1989), "Business Conditions and Expected Returns on Stocks and Bonds", *Journal of Financial Economics*, 25: 23-49.

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

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

GONZALO, J. (1994), " Five Alternative Methods of Estimating Long-Run Equilibrium Relationships", *Journal of Econometrics*, 60: 203-233.

GRANGER, C.W.J. and NEWBOLD, P. (1974), "Spurious Regressions in Econometrics", *Journal of Econometrics*, 2: 111-120.

HALL, R.E. (1978), "Stochastic Implications of the Life-Cycle Permanent Income Hypothesis", *Journal of Political Economy*, 86: 971-87.

HANSEN, B.E. (1992), "A Powerful Simple Test for Cointegration Using Cochrane-Orcutt", University of Rochester, Department of Economics Working Paper No. 230.

HENDRY, D.F. and ANDERSON, G.J. (1977), "Testing Dynamic Specification in Small Simultaneous Models: An Application to a Model of Building Society Behaviour in the United Kingdom", ch. 8c in M.D. Inrilligator (ed.), *Frontiers of Quantitative Economics*, iii(a), North-Holland, Amsterdam, 361-383.

—— and NEALE, A.J. (1987), "Monte Carlo Experimentation using PC-NAIVE", in T. Fomby and G. Rhodes (eds.), *A Century of Economics*, Vol. 6, JAI Press, Greenwich, Conn., 91-125.

—— —— and ERICSSON, N.R. (1990), *PC-NAIVE: An Interactive Program for Monte Carlo Experimentation in Econometrics*, Institute of Economics and Statistics, Oxford University, Oxford.

HENDRY, D.F. (1992), "On the Interactions of Unit Roots and Exogeneity", typescript, Nuffield College, Oxford.

JOHANSEN, S. (1988), "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12: 231-254.

KREMERS, J.J.M., ERICSSON, N.R. and DOLADO, J. (1992), "The Power of Co-integration Tests", *Oxford Bulletin of Economics and Statistics*, 54: 325-348.

MACKINNON, J.G. (1991), "Critical Values for Co-integration Tests", in R.F. Engle and C.W.J. Granger (eds.), *Long-Run Economic Relationships*, Oxford University Press, 267-276.

MANKIW, N.G. and SHAPIRO, M.D. (1985), "Trends, Random Walks and Tests of the Permanent Income Hypothesis", *Journal of Monetary Economics*, 16: 165-174.

—— —— (1986) "Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models", *Economics Letters*, 20:

139-145.

NELSON, C.R. and PLOSSER, C.I. (1982), "Trends and Random Walks in Macro-Economic Time Series: Some Evidence and Implications", *Journal of Monetary Economics*, 10: 139-162.

PHILLIPS, P.C.B. (1986), "Understanding Spurious Regressions in Econometrics", *Journal of Econometrics*, 33: 311-340.

—— (1991), "Optimal Inference in Co-integrated Systems", *Econometrica*, 59: 282-306.

—— and HANSEN B.E. (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes", *Review of Economic Studies*, 57: 99-125.

—— and LORETAN, M. (1991), "Estimating Long-Run Economic Equilibria", *Review of Economic Studies*, 58: 407-436.

—— and OULIARIS, S. (1990), "Asymptotic Properties of Residual Based Tests for Cointegration", *Econometrica*, 58: 165-193.

—— and PLOBERGER, W. (1991), "Time Series Modelling with a Bayesian Frame of Reference: I. Concepts and Illustrations", manuscript, Yale University.

SARGAN, J.D. (1964), "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", in P.E. Hart, G. Mills and J.K. Whitaker (eds.), *Econometric Analysis for National Economic Planning*, Butterworth, London; reprinted in D.F. Hendry and K.F. Wallis (eds.), *Econometrics and Quantitative Economics*, Basil Blackwell, Oxford, 1984.

SIMS, C.A., STOCK, J.H. and WATSON, M.W. (1990), "Inference in Linear Time Series with Some Unit Roots", *Econometrica*, 58: 113-144.

STOCK, J.H. (1987), "Asymptotic Properties of Least-Squares Estimators of Co-integrating Vectors", *Econometrica*, 55: 1035-1056.

—— —— (1992), "Deciding between I(0) and I(1)", *National Bureau of Economic Research*, Technical Working Paper No. 121.

—— and WATSON, M.W. (1988), "Variable Trends in Economic Time Series", *Journal of Economic Perspectives*, 2: 147-174.

—— and WEST, K.D. (1988), "Integrated Regressors and Tests of the Permanent Income Hypothesis", *Journal of Monetary Economics*, 21: 85-96.

WEST, K.D. (1988), "Asymptotic Normality when Regressors have a Unit Root", *Econometrica*, 56: 1397-1418.

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*, 89: 1-64.

Appendix: Concepts of Exogeneity

Econometric analysis often proceeds by using a single-equation model of a process of interest. In basing inference on single- equation methods, we implicitly assume that knowledge of the processes generating the explanatory variables carry no information relevant to the parameters of the process of interest. Engle, Hendry and Richard (1983) provide conditions (concepts of exogeneity) relating to the circumstances in which this assumption is valid. Rather than refer to particular variables as exogenous in general, however, Engle *et al.* refer to a variable as exogenous *with respect to a particular parameter* if knowledge of the process generating the exogenous variable contains no information about that parameter.

Three different concepts are introduced by Engle *et al.* and correspond to three different ways in which a parameter estimate may be used: inference, forecasting based on forecasts of the exogenous variables, and policy analysis. These different uses require that increasingly stringent conditions be met for exogeneity (that is, for the irrelevance of the regressor process to a parameter of interest). These conditions can be examined with the following definitions.

Let $\underset{\sim}{x}_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}$ be generated by the process with conditional density function $D(\underset{\sim}{x}_t|\underset{\sim}{X}_{t-1},\underset{\sim}{\lambda})$, where $\underset{\sim}{X}_{t-1}$ denotes the history of the variable X: $\underset{\sim}{X}_{t-1} = (\underset{\sim}{x}_{t-1}, \underset{\sim}{x}_{t-2}, \ldots \underset{\sim}{x}_0)$. Let the set of parameters $\underset{\sim}{\lambda}$ be partitioned into $(\underset{\sim}{\lambda}_1,\underset{\sim}{\lambda}_2)$ such that

$$D(\underset{\sim}{x}_t|\underset{\sim}{X}_{t-1},\underset{\sim}{\lambda}) = D(y_t|z_t,\underset{\sim}{X}_{t-1},\underset{\sim}{\lambda}_1)D(z_t|\underset{\sim}{X}_{t-1},\underset{\sim}{\lambda}_2).$$

The first element of the product on the right hand side of the equality is known as the *conditional* density (or model) while the second element is the marginal density. In the simple case where $\underset{\sim}{x}_t$ is bivariate normal, the conditional model leads to a bivariate regression with y_t as the dependent variable and z_t the regressor.

Suppose our parameter of interest is given by ψ (in the conditional model). Weak exogeneity of z_t for ψ requires, (i) ψ is a function of $\underset{\sim}{\lambda}_1$ alone, and (ii) that there are no cross-restrictions between $\underset{\sim}{\lambda}_1$ and $\underset{\sim}{\lambda}_2$. The essential element of weak exogeneity is that $\underset{\sim}{\lambda}_2$ contains no information relevant to discovering $\underset{\sim}{\lambda}_1$. Inference concerning ψ can be made conditional on z_t with no loss of information relative to that which could be obtained using the joint density of y_t and z_t.

Strong exogeneity requires that z_t is weakly exogenous for ψ and

$$D(z_t | \underset{\sim}{X}_{t-1}, \underset{\sim}{\lambda}_2) = D(z_t | \underset{\sim}{Z}_{t-1}, y_0, \underset{\sim}{\lambda}_2),$$

so that Y does not Granger-cause Z (*i.e.* Y does not enter the process generating Z. In a simple regression model this implies that the equation generating z_t does not contains any lags of y_t). Strong exogeneity is necessary for forecasting which proceeds by forecasting future z's and then forecasting y's conditional on the predicted z's.

Finally, z_t is super-exogenous for ψ if and only if z_t is weakly exogenous for ψ and $d\underset{\sim}{\lambda}_1/d\underset{\sim}{\lambda}_2 = 0$. Super-exogeneity is necessary where models will be used for policy analysis, in which the investigator is interested in predicting the derivative of the dependent variable with respect to a change which can be made in an explanatory variable; a change made to the value of the explanatory variable by an external actor is in effect a change in the parameters of the process governing that variable's evolution.

[2] This definition can be extended to cover variables integrated of order d > 1. Roughly speaking, a series is said to be integrated of order d if it is stationary after differencing d times, but is not stationary after differencing only d - 1 times.

[3] An issue of some interest is which of model (5) or (6) is appropriate. The advantage of using a model such as (6) lies, generally, not in its plausibility but rather in making the critical values appropriate for testing the null H_0: $d_2 = 0$ invariant to the presence or absence of a constant in the process generating c_t. If this DGP has a constant, say, $\Delta c_t = d + \nu_t$, the critical values of the *t*-test in (5), used for testing H_0, are sensitive to the value of d whereas this is not the case in model (6). A related problem concerns what critical values should be used. West (1988) shows that if $d \neq 0$, $\underline{t}_{d_2=0}$ is asymptotically normally distributed in (5) while it has a non-standard distribution in (6). In the absence of *a priori* knowledge of whether or not there is a constant in the DGP, invariance of the critical values is a useful property and I would argue in favour of using the more general model, and hence non-standard critical values, given by (6).

[4] Source: Table 2 in Mankiw and Shapiro (1986). The standard errors of each of the entries a_{ij}, expressed as a fraction, is given by $[(a_{ij})(1 - a_{ij})/N]^{1/2}$ where N is the number of replications in the Monte Carlo (N = 1000 for this table).

[5] Work on non-linear co-integration, although mathematically more complex, shares this main idea of some function of the component series being bounded statistically

[6] By order T^k we mean that the variance grows at rate T^k. For example, for a random walk process such as (1), it may be shown that variance$(y_T) = T\sigma_u^2$ and thus k = 1 in the terminology given above. Thus if we define a variable $z_T = T^{-1/2}y_T$, variance$(z_T) = \sigma_u^2$ which is finite, bounded away from zero and remains constant with time. Scalings such as these are important in deriving the asymptotic theory because it ensures that the asymptotic distributions of the *scaled* expressions are neither non-degenerate (do not collapse on a single value) nor explosive (have infinitely large variances).

7 *i.e.* it is $T(\hat{\alpha} - \alpha)$ which has a non-degenerate, non-explosive distribution asymptotically rather than $T^{1/2}(\hat{\alpha} - \alpha)$. The latter is the scaling appropriate for estimators based on stationary processes.

8 It is important to make this qualification in order not to mislead the reader. There are many circumstances, some described below, especially under failure of weak-exogeneity, when single-equation dynamic models also provide badly biased estimates. However, this does not run counter to our main assertion here that static regressions are hardly ever desirable and thus two-step methods based on first estimating the long-run and then modelling the short-run have severe limitations. For estimation and inference, the two steps are best accomplished together.

9 A martingale difference sequence (MDS) generalises the concept of an identically and independently distributed sequence of random variables, by allowing for some amount of dynamic dependence among the elements of the sequence. An MDS is defined with respect to an information set $\mathcal{I}_{t-1}$ of data realised by time t-1. A sequence $\{y_t, t = 1, 2, \ldots\}$ is defined to be a MDS with respect to $\{\mathcal{I}_t, t = 1, 2, \ldots\}$ if $E\{|y_t|\} < \infty \; \forall \; t$ and that $E\{y_t | \mathcal{I}_{t-1}\} = 0 \; \forall \; t$. In words, we require that the expectation of the absolute value of y_t be finite and that the expectation of y_t *conditional on the past* be zero. For the purposes of the analysis which follows, the reader will not lose by thinking of the simpler special case of serially uncorrelated, mean zero, variables.

10 There is a literature on correcting, non-parametrically, the estimates derived from a static regression in order to account for the effects of the omitted dynamic terms (see, for example, Phillips and Hansen (1990)). Provided single equation estimation is valid, dynamic specification and non-parametric correction of the estimates from a static regression are asymptotically equivalent procedures. Comparisons in finite samples are usually ambiguous and depend on the particular specifications of the DGP.

11 The long-run multiplier is derived by setting $y_t = y^*$, $x_t = x^*$ $\forall$ t and ε_{1t} to its expected value of 0. Thus, $y^* = \gamma_1 y^* + (\gamma_2 + \gamma_3)x^*$, which implies that $(1 - \gamma_1)y^* = (\gamma_2 + \gamma_3)x^*$. Under the restriction that $\gamma_1 + \gamma_2 + \gamma_3 = 1$, it follows that $y^* = x^*$ in the long-run.

12 This may be seen by rewriting (11a), using the homogeneity restriction, as $y_t = x_t + \gamma_1(y - x)_{t-1} + (\gamma_2 - 1)\Delta x_t + \varepsilon_{1t}$.

[13] Error-correction models (ECMs) are the focus of an extensive literature, starting wih Sargan (1964), Hendry and Anderson (1977) and Davidson, Hendry, Srba and Yeo (1978). They are a way of capturing adjustments in a dependent variable which may depend not only on the level of some explanatory variable, but also on the extent of the deviation of the explanatory variable from an *equilibrium* relationship with the dependent variable. Thus, if the equilibrium relationship is given by $y^* = \theta x^*$, the error-correction term is given by $(y_t - \theta x_t)$. In (13) $\theta = 1$. The estimate of the coefficient on the error-correction term, given above by c, provides an estimate of the short-run adjustment to equilibrium.

[14] If it is necessary to have estimates of the standard error of this multiplier estimate, (13) can be estimated more conveniently in a linearly transformed form which gives the long-run multiplier directly. This is known as the Bewley transform and the estimates derived from this transform are numerically equivalent to those obtained from (13).

[15] See footnote 3.

[16] ADF(1) refers to an augmented Dickey-Fuller test with $\ell = 1$ in (14).

[17] Critical values for the $\underline{t}_{c=0}$ test for the case where (13) includes a constant and $\underset{\sim}{x}_t$ is a vector (at most of dimension 5) are given by Banerjee, Dolado and Mestre (1993). These critical values depend on the dimension of the system (number of regressors). Hansen (1992) provides an ECM test which is invariant to dimension but the test imposes a common factor restriction. Where this restriction is violated, the losses of power involved are important. Banerjee, Dolado and Mestre (1993) provide a full discussion of this issue.

[18] Techniques which rely on rotating the dependent variable in sequence are in general unsatisfactory.

[19] Naturally, this is only a *sufficient* condition for such a violation. To derive all the necessary conditions for weak exogeneity would require us to look at the properties of the error process $(e_{1t}, e_{2t})'$.

[20] To emphasize the point made in the earlier footnote, this is clearly only a necessary condition for weak exogeneity since the parameters of the two equations may be linked in other ways.

[21] For details of the first method, see Phillips and Hansen (1990), Phillips and Loretan (1991) and Phillips (1991). For the second method, in addition to the original account contained in Johansen (1988), also see Banerjee and Hendry (1992) and Banerjee *et al.* (1993).