

Estimating Long-run Economic Equilibria

PETER C. B. PHILLIPS and MICO LORETAN

Cowles Foundation for Research in Economics, Yale University

First version received November 1989; final version accepted February 1991 (Eds.)

Our subject is estimation and inference concerning long-run economic equilibria in models with stochastic trends. An asymptotic theory is provided to analyze a menu of currently existing estimators of cointegrated systems. We study in detail the single-equation ECM (SEECM) approach of Hendry. Our theoretical results lead to prescriptions for empirical work, such as specifying SEECM's nonlinearly and including lagged equilibrium relationships rather than lagged differences of the dependent variable as covariates. Simulations support these prescriptions, and point to problems of overfitting not encountered in the semiparametric approach of Phillips and Hansen (1990).

1. INTRODUCTION

An early stimulus for work on error-correction modelling was the research of A. W. Phillips (1954, 1956, 1958, 1962) on trade cycle and cyclical growth models. The error-correction models (ECM's) used by Phillips were formulated in continuous time and raised problems of econometric estimation that were dealt with only slowly over the next two decades. Bergstrom (1988) provides an historical review of this line of research. A more direct stimulus came from the work of Sargan (1964) on discrete time ECM's and their application to aggregate wage and price series in the U.K. Sargan argued for the presence of levels in regressions that were specified in differences and that were designed to model adjustment processes. During the 1970's these ideas became key elements in the methodology advocated by David Hendry. They served to distinguish this methodology from the Box-Jenkins techniques that were of growing popularity in statistics and from the astructural modelling methods that were beginning to emerge in North America. The fact that the levels of nonstationary economic variables like wages and prices appear in regressions that are formulated in differences requires that certain linear combinations of the levels must be stationary if the regression is to have stationary residuals. Otherwise the regression would have nonstationary residuals and would be spurious in the sense of Granger and Newbold (1974). For this reason ECM's were recognized at an early stage to carry some interesting statistical implications for nonstationary series—see the discussion by Hendry (1976) for example. Certainly the empirical success of these models, especially the consumption equation of Davidson *et al.* (1978), motivated analytic research and was in large part responsible for the birth of the concept of cointegration in Granger (1981).

In the last few years cointegration has become an enormously active research area with a wide range of participants. But, in spite of this enormous flurry of intellectual activity, there is still no agreement about the prescriptions for applied econometric research. This is unfortunate because, as we shall argue in the present paper, analytical research provides some very clear pointers for empirical work. Moreover, in light of the

empirical success of ECM modelling, empirical work itself provides some important guidelines. Finally, since ECM modelling is the immediate precedent of cointegration, it is enlightening to study the statistical properties of the single-equation empirical ECM's that gave rise to the concept of cointegration in the first place. In this sense the present article extends the analysis that was begun in Phillips (1988*a*).

Let us start by looking at the presently available prescriptions. The principal of these are:

- (i) Unrestricted VAR and Bayesian VAR methods, such as those advocated by Sims (1980) and Doan, Litterman and Sims (1984).
- (ii) Two-step methods, such as those advocated by Engle and Granger (1987), in which long-run equilibrium relationships are first estimated and subsequently used in the fitting of short-run dynamics. Engle and Yoo (1990) have recently suggested a three-step method in which the estimates of the long-run parameters are revised in the third stage.
- (iii) Single-equation ECM (SEECM) methods advocated by Hendry (1987) which seek a tentatively adequate data-characterization that encompasses rival models, displays parameter constancy, has martingale difference errors with respect to a selected information set and parsimoniously orthonormalizes the regressors.
- (iv) Systems maximum-likelihood methods proposed independently by Johansen (1988) for VAR specifications and by Phillips (1991) in a general time-series setting.
- (v) Nonparametric spectral-regression methods proposed by Phillips (1988*b*) which permit the direct estimation of long-run equilibrium relationships in the frequency domain.
- (vi) Single-equation semiparametric least squares and instrumental-variables methods proposed by Phillips and Hansen (1990). These, like (v), permit direct estimation of the long-run relationship but involve a two-step method whereby the data is filtered in the first step using a nonparametric correction.

With all of these differing prescriptions it is difficult for the applied researcher to decide what is appropriate in practice. The diversity of the choices obscures the fact that statistical theory is quite clear about optimal inference procedures. As discussed in Phillips (1991) this depends on whether unit roots are incorporated in the specification, as they are in systems ECM methods (iv), or whether they are implicitly estimated, as they are in unrestricted VAR formulations (i). Of course, choices also depend on issues of modelling methodology. This is particularly so in the case of (iii) where judgmental exercises in model selection are an integral part of the procedure.

The main purpose of the present article is to discuss the issues that arise in the context of the choices described above. We shall review what statistical theory has to say about the properties of the various approaches and look into the question of whether it has any guidance to offer empirical researchers. Since much of the empirical work in the field has involved SEECM methods like (iii), this will form our central focus of attention.

The plan of the paper is as follows. Section 2 describes various time-series representations of cointegrated systems. These include structural formulations, the triangular system ECM representation, the autoregressive ECM representation, unrestricted VAR representations, a latent variable representation and SEECM specifications. Section 3 provides the asymptotic theory for a wide menu of estimator choices based on single-equation and systems approaches and utilizing different prior information about the presence of unit

roots and the nature of short-run dynamic adjustment. SEECM's are studied in detail and the use of lagged long-run equilibrium relations in dynamic specification is found to have certain asymptotic advantages. Section 4 reports the results of a simulation study that seeks to evaluate the sampling properties of single equation approaches, again with an emphasis on the ECM methodology. Conclusions are given in Section 5.

For convenience of the reader we list the notation and acronyms used in the paper in Table 1.

TABLE 1
List of symbols and acronyms

OLS	ordinary least squares	FM	fully modified OLS
GLS	generalized least squares	iid	independently and identically distributed
2SLS	two stage least squares	wd	weakly dependent
3SLS	three stage least squares	\Rightarrow	converges weakly to
NLS	nonlinear least squares	lrvar	long-run variance
MLE	maximum likelihood estimator	mg	martingale
LIML	limited information maximum likelihood	mds	martingale difference sequence
FIML	full information maximum likelihood	BM (Ω)	Brownian motion with covariance matrix Ω
ECM	error correction model	a.s.	almost surely
SEECM	single equation error correction model	L	lag operator, $L^k y_t = y_{t-k}$
SEM	simultaneous equations model	Δ	first difference operator $\Delta y_t = (1 - L)y_t = y_t - y_{t-1}$
VAR	vector autoregression	$I(k)$	integrated process of order k
		$f_{uu}(\lambda)$	spectral density matrix of u_t

2. COINTEGRATED SYSTEMS AND TIME-SERIES REPRESENTATIONS

2.1. A typical cointegrated system

Let y_t be an n -vector $I(1)$ process and u_t be an n -vector stationary time series. We partition these vectors as follows

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix}_m^1, \quad u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}_m^1, \quad n = m + 1 \quad (1)$$

and assume that the generating mechanism for y_t is the cointegrated system

$$y_{1t} = \beta' y_{2t} + u_{1t} \quad (2)$$

$$\Delta y_{2t} = u_{2t}. \quad (3)$$

This system is in structural equation format and (2) may be regarded as a stochastic version of the partial equilibrium relationship $y_{1t} = \beta' y_{2t}$, with u_{1t} representing stationary deviations from equilibrium. Equation (3) is a reduced form which specifies y_{2t} as a general integrated process, the outcome of superimposed shocks u_{2s} ($s \leq t$) that influence the process period after period. The equation system (2) and (3) is typical of more general models, such as those studied in Phillips (1991), where the cointegrating relationship is multidimensional and where deterministic trends may coexist with the stochastic trend in (3). We shall work with the single-equation relationship (2) because our attention will later concentrate on SEECM methods.

In the general case u_t is wd ($0, f_{uu}(\lambda)$) and absorbs all stationary short-run dynamic adjustments towards the partial equilibrium. An important subcase is the prototypical

system where $u_t = \text{mds}(0, \Sigma)$ or even $\text{iid}(0, \Sigma)$. In both cases we shall assume that $\Sigma > 0$. This is a useful simplification. Indeed, the prototypical case may be treated as a pseudo-model for the general case of stationary errors. This is possible because we can construct a martingale difference sequence $u_t = \text{mds}(0, \Sigma)$, with $\Sigma = 2\pi f_{uu}(0)$, that approximates the actual time series $u_t = \text{wd}(0, f_{uu}(\lambda))$ in a well-defined sense—see Hall and Heyde (1980, Chapter V). Subsequently it will be useful to work with Σ in partitioned format as

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma'_{21} \\ \sigma_{21} & \Sigma_{22} \end{bmatrix}$$

with the partition conformable with (1).

The prototypical system is an important aid to intuition. Note that if we replace (3) by

$$y_{2t} = Ay_{2t-1} + u_{2t} \quad (3)'$$

and require the coefficient matrix A to have stable roots, then the new system (2) and (3)' is a conventional SEM. More than that, it is a triangular system because (3)' is a reduced form. This means that when Σ is block diagonal (i.e. $\sigma_{21} = 0$) the Gaussian MLE of β in (2) and (3) is derived by OLS on (2). When Σ is not block-diagonal and is unrestricted, then GLS and feasible GLS procedures on (2) and (3)' are asymptotically equivalent to the Gaussian MLE (see Lahiri and Schmidt (1978)).

These implications of a triangular structure continue to apply in the nonstationary cointegrated system (2) and (3). Thus, when $u_t \equiv \text{iid } N(0, \Sigma)$ and Σ is block-diagonal the OLS estimate of β in (2) is the MLE. When Σ is not block-diagonal, the MLE is obtained by using OLS on the augmented regression equation

$$y_{1t} = \beta' y_{2t} + \gamma' \Delta y_{2t} + u_{1 \cdot 2t} \quad (4)$$

Here $u_{1 \cdot 2t} = u_{1t} - \sigma_{12} \Sigma_{22}^{-1} u_{2t}$ and $\gamma = \Sigma_{22}^{-1} \sigma_{21}$.

Observe that (4) is a regression in levels that is augmented by the differences Δy_{2t} . In this respect (4) is related to typical ECM formulations, which are specified in differences but include levels amongst the regressors. Indeed, subtracting y_{1t-1} from (4) we obtain the equation

$$\Delta y_{1t} = \delta(y_{1t-1} - \beta' y_{2t-1}) + (\gamma + \beta)' \Delta y_{2t} + u_{1 \cdot 2t} \quad (4)'$$

where $\delta = -1$. Equation (4)' is now explicitly in ECM format.

2.2. The triangular system ECM representation

Taking differences of (2) we obtain a direct $I(1)$ representation for the system viz.

$$\Delta y_{1t} = \beta' u_{2t} + \Delta u_{1t} \quad (5)$$

$$\Delta y_{2t} = u_{2t} \quad (6)$$

in which the errors are stationary. This can be written as

$$\Delta y_{1t} = -u_{1t-1} + (1, \beta') u_t$$

$$\Delta y_{2t} = u_{2t}$$

or in systems format as

$$\Delta y_t = -e\alpha' y_{t-1} + v_t \quad (7)$$

where

$$e = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \alpha' = (1, \beta'), \quad v_t = \begin{bmatrix} 1 & \beta' \\ 0 & I \end{bmatrix} u_t.$$

Equation (7) is a triangular system ECM representation. The model explains the differences Δy_t in terms of the lagged levels y_{t-1} and stationary errors v_t that are a simple transform of the original errors u_t (2) and (3).

Triangular system representations like (7) are discussed in detail in Phillips (1991). They have several natural advantages which are worth emphasizing here:

1. The model (7) is linear in the unknown parameter β of the long-run equilibrium relationship;
2. The coefficient vector e is specified and is not estimated. This means that identification is achieved from the specification of e rather than the specification of martingale difference errors on an equation with explicit dynamics. Further, since v_t is stationary, the usual interpretation of a coefficient like the first component of e in (7) as an adjustment coefficient loses its normal meaning because it is no longer natural to think of the equation as explaining the extent of the adjustment towards equilibrium each period. In place of this interpretation it is more natural to think of the equation as explaining the stationary deviations about the equilibrium level that persist period upon period. The same interpretation was given to (2), and (7) is simply an algebraic reformulation of that original system.
3. All of the transient dynamics are absorbed into the residual process v_t in (7). Note that if u_t in (2) and (3) is a finite-order ARMA process then so is v_t .
4. Model (7) retains the triangular structure of (2) and (3). This simplifies issues of inference. And, as discussed above, many ideas from structural SEM theory continue to apply, such as the reduction of the Gaussian MLE to OLS on a suitably augmented single equation.

2.3. The autoregressive ECM representation

This works from an explicit version of the $I(1)$ representation (5) and (6) which we write in systems form as

$$\Delta y_t = w_t; \quad w_t = C(L)\varepsilon_t, \quad C(L) = \sum_{j=0}^{\infty} C_j L^j, \quad C_0 = I. \quad (8)$$

Here w_t is a linear process and ε_t is mds $(0, \Sigma)$ with $\Sigma > 0$. Under the summability condition

$$\sum_1^{\infty} j^{1/2} \|C_j\| < \infty \quad (9)$$

and the assumption that y_t is cointegrated with cointegrating dimension one we may extract an autoregressive form of the ECM

$$J(L) \Delta y_t = \gamma \alpha' y_{t-1} + \varepsilon_t. \quad (10)$$

A recent demonstration of this representation and the converse is given by Solo (1989). Systems estimation of (10) is recommended and discussed by Johansen (1988).

In (10) $J(L)$ is usually an infinite autoregressive lag polynomial although in practice finite-order approximations are employed. We observe:

1. The cointegrating coefficient α and subvector β appear nonlinearly in (10).
2. The coefficient vector γ is unknown and must be estimated. Identification in (10) is achieved by the reduction of the regression error to the mds ε_t . This requires

that an autoregressive operator $J(L)$ of infinite order be accommodated. In practice consistent estimation methods necessitate the use of consistent order selection routines, such as those that are incorporated in the Hannan and Kavalieris (1984) recursions. For vector systems these involve some nontrivial complications—see Hannan and Deistler (1988, pp. 246, 292ff) for details and discussion. Note that these problems are bypassed in Johansen's (1988) approach by the assumption that the autoregression in (10) is of known finite order and that this model is the data-generating mechanism.

3. The representation (10) is effectively a nonlinear in parameters reduced form. Since γ is unknown the model loses the simplicity of the triangular structure of (7) and its linear in parameters feature. This is the price that is paid for the explicit autoregressive representation.
4. Unlike (7), all short-run dynamics are now embodied in $J(L)$. Full system estimation requires that these coefficients be estimated simultaneously with γ and α . As remarked above, since the order of $J(L)$ is unknown in practice, this will necessitate order-selection methods and the use of nonlinear methods.
5. An advantage of (10) is that the model may be used to test the rank of the cointegration space, i.e. the row rank of α' . This means that the Johansen (1988) methodology permits estimation of the number of cointegrating vectors as well as α itself.

2.4. The latent variable representation

Working from (8) we replace w_t by a martingale difference sequence approximation. In the present case this may be achieved by the so-called Beveridge–Nelson decomposition (see Beveridge and Nelson (1981) and Solo (1989)):

$$\begin{aligned} C(L) &= C(1) + (L-1)\tilde{C}(L) \\ \tilde{C}(L) &= \sum_0^\infty \tilde{C}_j L^j, \quad \tilde{C}_j = \sum_{j+1}^\infty C_k. \end{aligned} \quad (11)$$

Under (9), the series $\tilde{C}(L)$ is square summable, i.e. $\sum_0^\infty \|\tilde{C}_j\|^2 < \infty$, and then we can write

$$\begin{aligned} w_t &= C(L)\varepsilon_t \\ &= C(1)\varepsilon_t + \tilde{w}_{t-1} - \tilde{w}_t \end{aligned} \quad (12)$$

where $\tilde{w}_t = \tilde{C}(L)\varepsilon_t$ is covariance stationary. Now, writing $W_t = C(1)\varepsilon_t = \text{mds}(0, \Omega)$ where $\Omega = C(1)\Sigma C(1)'$, we have the decomposition

$$w_t = W_t + \tilde{w}_{t-1} - \tilde{w}_t$$

where W_t is an mds approximation to w_t . This yields the mg approximation to y_t from

$$y_t = Y_t + \tilde{w}_0 - \tilde{w}_t, \quad Y_t = \sum_1^t W_j.$$

The mg Y_t may be considered as a latent variable for y_t , differing from it by the stationary process $\tilde{w}_0 - \tilde{w}_t$. This formalization was mentioned recently by Deistler and Anderson (1989).

Taking the argument further, we next observe that y_t is cointegrated with cointegrating vector α or $\alpha'y_t = I(0)$ iff $\alpha'W_t = 0$ a.s., iff $\alpha'Y_t = 0$ a.s., iff $\alpha'\Omega = 0$, iff $\alpha'C(1) = 0$. This leads us directly to the latent variable representation

$$\alpha'Y_t = 0 \quad (13)$$

where $Y_t = \text{mg}(\Omega)$, i.e. an mg whose differences have covariance matrix Ω . Suppose α is normalized so that $\alpha'\alpha = 1$ and let H_2 be its orthogonal complement. Then we have $Y_t = H_2 Y_t^c$ a.s. where $Y_t^c = \text{mg}(H_2'\Omega H_2)$ is the "common trend" of Y_t . The mg approximation to y_t is then

$$y_t = H_2 Y_t^c + \tilde{w}_0 - \tilde{w}_t$$

which is a common trends representation of y_t . Stock and Watson (1988) provide a detailed study of this type of representation.

From the practical standpoint it would appear that the representation (13) has few uses. However, under Gaussian assumptions and zero initial conditions we find that $Y_t \equiv N(0, t\Omega)$. Thus (13) makes it apparent that asymptotically efficient estimation of the cointegrating vector α will rely on consistent estimation of Ω —the long-run covariance matrix of w_t . From this standpoint (13) is useful because it points to the fact that the entire matrix Ω must be estimated for efficient estimation of α .

The last observation is suggestive. It indicates that, in general, systems estimation is required for the efficient estimation of α . This was the conclusion reached earlier in Phillips (1991). However, as the discussion of the prototypical model in section 2.2 and the augmented regression equation (4) indicates, there are cases where single-equation least squares is asymptotically efficient. Further examples are given in the methodological discussion in Phillips (1988a). One of our objectives here is to extend that treatment to the general case of stationary errors. In particular, we shall develop a modification to the SEECM methodology which ensures that least squares is asymptotically equivalent to full systems Gaussian estimation of α .

2.5. Single-equation ECM (SEECM) methodology

The vector autoregressive ECM representation given in (10) is quite different from the single-equation empirical specifications that follow in the Davidson *et al.* (1978) tradition for the consumption function. The single-equation approach to ECM modelling is explained in detail in two papers by Hendry and Richard (1982, 1983) and in recent years has become known as the Hendry methodology for empirical research. It is also supported by a suite of computer software written by Hendry (1989) and designed to assist in its implementation. The approach is to seek out a tentatively adequate single-equation specification that meets the following criteria, which we refer to as the Hendry-Richard prescriptions: (i) data coherency; (ii) valid conditioning; (iii) encompassing; (iv) theory compatibility; (v) parsimonious, orthogonal decision variables; (vi) parameter constancy. The empirical determination of a model that satisfies these prescriptions inevitably involves judgmental elements. These are illustrated in the Hendry-Richard papers, in Hendry (1989) and in many of the references cited therein. We shall not go into more detail about the general methodology here.

For a model such as (2) and (3) above, the starting point in the single-equation approach to modelling y_{1t} is a general unrestricted regression of the form:

$$y_{1t} = \hat{a}'y_{2t} + \hat{f}'x_t + \hat{w}_t. \quad (14)$$

Here x_t is a p -vector of autoregressive and distributed lag components. Typically it will contain lagged values of Δy_{1t} and present and lagged values of Δy_{2t} . Temporal transformations and pre-testing may reduce the number of elements in this vector and replace some differences by higher-order differences to achieve a parsimonious partial orthogonalization of the regressors. In the Hendry approach what appears to be important is that the chosen regressors adequately represent the information set embodied in the past history of Δy_{1t} .

and the present and past history of Δy_{2t} . We could write (14) explicitly as

$$y_{1t} = \hat{a}' y_{2t} + \sum_{k=1}^p \hat{f}_{1k} \Delta y_{1t-k} + \sum_{k=0}^p \hat{f}_{2k}' \Delta y_{2t-k} + \hat{w}_t \quad (15)$$

allowing for the fact that in practice some of these regressors may be eliminated and others may be higher differences. An alternative specification is to use an SEECM format that takes the form

$$\Delta y_{1t} = \tilde{f}_0'(y_{1t-1} - \tilde{a}' y_{2t-1}) + \sum_{k=1}^p \tilde{f}_{1k} \Delta y_{1t-k} + \sum_{k=0}^p \tilde{f}_{2k}' \Delta y_{2t-k} + \tilde{w}_t. \quad (16)$$

Of course, (16) involves a change in the regressand as well as a change in the regressors, so it is not suggested that (15) and (16) are equivalent. However, both types of specification are employed in empirical searches conducted under the Hendry methodology, a good example being the searches described in Mizon and Hendry (1980) for a quarterly consumption function for Canadian consumer durable expenditures. Moreover, to the extent that the transient dynamics are properly modelled, we might expect estimates of the long run parameters (viz. \hat{a} and \tilde{a}) derived from (15) and (16) to be asymptotically equivalent.

To this end and in order that the past history be fully captured in the information set, we would expect to have to allow $p \rightarrow \infty$ in both (15) and (16) as the sample size $T \rightarrow \infty$. The intended equation on which the empirical regression (15) is based would then have the form

$$y_{1t} = a' y_{2t} + \sum_{k=1}^{\infty} f_{1k} \Delta y_{1t-k} + \sum_{k=0}^{\infty} f_{2k}' \Delta y_{2t-k} + \eta_t. \quad (17)$$

The error on (17), η_t , is a martingale difference sequence with respect to the filtration $\mathcal{F}_{t-1} = \sigma\{\Delta y_{1t-1}, \Delta y_{1t-2}, \dots, \Delta y_{2t}, \Delta y_{2t-1}, \dots\}$, i.e. $E(\eta_t | \mathcal{F}_{t-1}) = 0$. In fact, from (2) we have $a = \beta$ and

$$\eta_t = u_{1t} - E(u_{1t} | \mathcal{F}_{t-1}). \quad (18)$$

The intended equation (17) would seem to satisfy the Hendry-Richard prescriptions. However, as we shall explore in Section 3.4 below, regression specifications such as (15) encounter difficulties in general because the truncation error is non-negligible due to shock persistence. Moreover, as pointed out earlier in Phillips (1988a) there is a general failure of valid conditioning (No. (ii) in the prescriptions) in (15) and (17), due to the presence of feedback from u_{1t} to u_{2t} . We shall systematically explore the effects of this failure below and suggest an antidote for the general case.

2.6. The unrestricted VAR representation

We can write (2) and (3) in the form

$$\begin{aligned} y_{1t} &= \beta' y_{2t-1} + (1, \beta') u_t \\ y_{2t} &= A y_{2t-1} + u_{2t}, \quad A = I \end{aligned}$$

or

$$y_t = G y_{t-1} + v_t, \quad G = \begin{bmatrix} 0 & \beta' \\ 0 & A \end{bmatrix}.$$

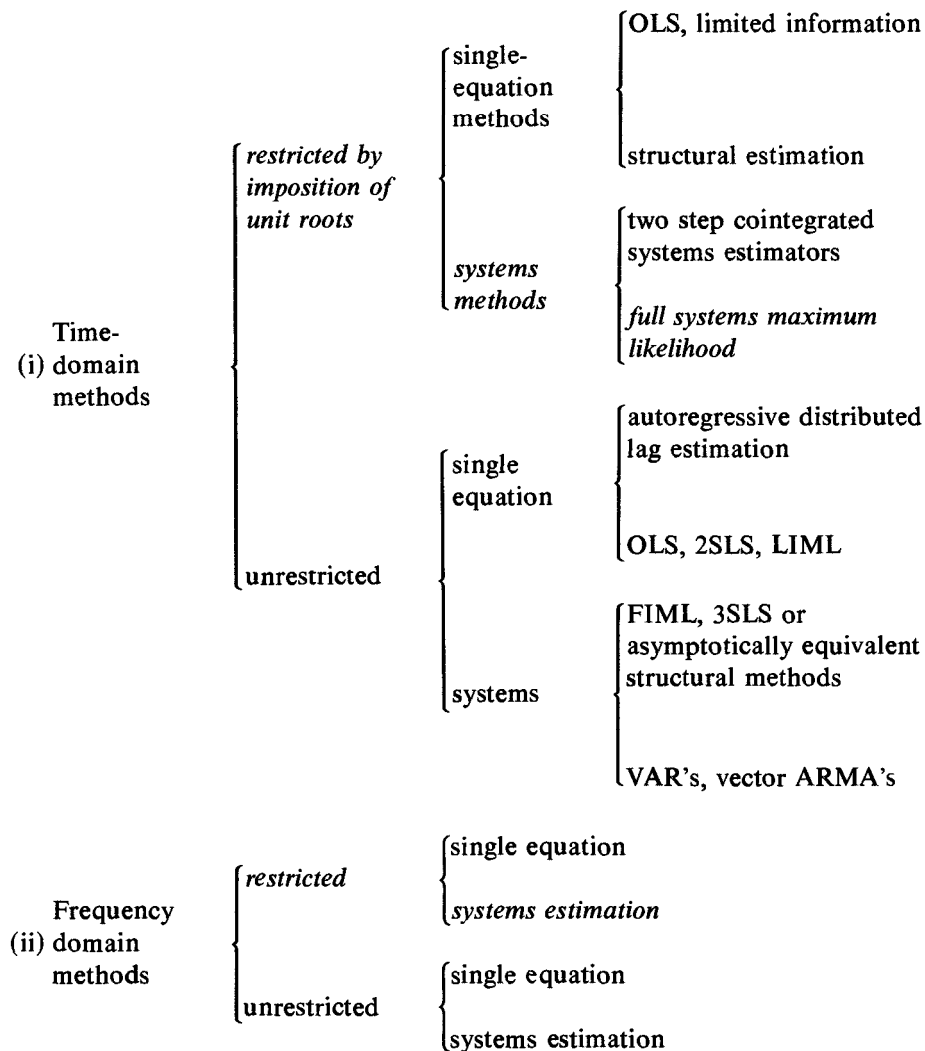
Then, if v_t has the autoregressive form $K(L)v_t = \varepsilon_t$ with $\varepsilon_t \equiv \text{iid}(0, \Sigma_\varepsilon)$ and with the lag operator $K(L)$ possible of infinite order we deduce the representation

$$B(L)y_t = \varepsilon_t, \quad B(L) = K(L)(1 - GL). \quad (19)$$

The autoregressive polynomial $B(L)$ has $m = n - 1$ unit roots, which are implicitly estimated in an unrestricted regression. As discussed in Phillips (1991), this has nontrivial effects on the asymptotic theory of inference concerning β . Park and Phillips (1989) and Sims, Stock and Watson (1990) provide a detailed study of the asymptotics in this case.

2.7. Implications

With so many ways of representing and working with cointegrated systems it is hardly surprising that there is no consensus about how best to proceed in empirical research. We can classify the many possibilities as follows:



Frequency-domain methods were suggested in Phillips (1988*b*). Since these methods were originally designed for regression models with strictly exogenous regressors, they raise questions concerning the treatment of endogeneity of the regressors (e.g. in (2)) and issues about the choice of spectral bands to be used in the regression. More will be said about the alternative provided by these methods in Section 3.2. below.

Asymptotic theory, at least, provides some guidance concerning optimal choices in (i) and (ii). These are shown in *italic* in the various branches of the tree above and they involve the use of systems methods on the model in restricted form so that all unit roots are explicitly included. On the face of it, this optimal decision path would seem to exclude single-equation methods. However, it is known that semiparametric corrections, as in Phillips and Hansen (1990) can produce optimal single-equation techniques. As we shall see below in Section 3.4, the same is true of the Hendry approach with some appropriate modifications to the implementation procedure described above.

3. ASYMPTOTIC THEORY

Throughout this section we assume that (2) is correctly specified in that there is a cointegrated relationship to estimate. If this were not so then u_t would be $I(1)$ and (2) would be a spurious regression. The asymptotic theory for that case is given in Phillips (1986). If there is more than one cointegrating relationship then it follows from Theorem 1 of Phillips (1991) that asymptotically efficient estimates of a single cointegrating relationship may be obtained even if the other cointegrating relations are ignored provided there are no across-relationship restrictions on the parameters of the long-run equilibria. The latter case does require joint estimation and is explored in the cited paper.

3.1. *The prototypical case*

This is the simple cointegrated system (2) and (3) with $u_t \equiv \text{iid}(0, \Sigma)$. We shall select typical entries in the tree of time-domain estimation possibilities shown in Section 2.7(i). It will be helpful in presenting the outcomes to set $S(r) = BM(\Sigma)$ and to partition S conformably with u_t in (1) as $S' = (S_1, S_2')$. The asymptotic results given below for this prototypical case are easily derived from those given in Phillips (1988b, 1991) and Phillips and Hansen (1990).

(i) *OLS*: Let β^* be the single equation OLS estimator of β in (1). Then

$$T(\beta^* - \beta) \Rightarrow \left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2' dS_1 + \sigma_{21} \right). \quad (20)$$

Using Lemma 3.1 of Phillips (1989) we write

$$S_1 = \sigma_{21}' \Sigma_{22}^{-1} S_2 + S_{1.2}$$

where $S_{1.2}(r) = BM(\sigma_{11.2})$ and $\sigma_{11.2} = \sigma_{11} - \sigma_{21}' \Sigma_{22}^{-1} \sigma_{21}$ is the conditional variance of u_{1t} given u_{2t} . Then the limit variate given in (20) may be written

$$\left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2 dS_{1.2} \right) + \left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2 dS_2' \right) \Sigma_{22}^{-1} \sigma_{21} + \left(\int_0^1 S_2 S_2' \right)^{-1} \sigma_{21}. \quad (21)$$

The first term of (21) is a Gaussian mixture of the form

$$\int_{G>0} N(0, \sigma_{11.2} G) dP(G), \quad G = \left(\int_0^1 S_2 S_2' \right)^{-1}. \quad (22)$$

The second term is a matrix unit root distribution, arising from the m unit roots in y_{2t} . The third term is a bias term arising from the contemporaneous correlation of u_{1t} and u_{2t} and thereby the endogeneity of the regressor y_{2t} .

Observe that in general (i.e. when $\sigma_{21} \neq 0$) the second and third terms of (21) induce asymptotic bias, asymmetry and nuisance parameters (i.e. Σ_{22} , σ_{21}) into the limit distribution. These effects make the OLS estimator β^* a poor candidate for inference, even though it is $O(1/T)$ consistent.

When $\sigma_{21} = 0$, the second and third terms of (21) are null and the limit theory comes from the first term alone and is therefore the Gaussian mixture (22). This component is, in fact, the same as the limit theory for the Gaussian MLE of β . Indeed β^* is the Gaussian MLE in this case because the system is completely triangular (i.e. has a diagonal error covariance matrix as well as a triangular structural matrix). Note also that when $\sigma_{21} = 0$ the regressor y_{2t} in (2) is strictly exogeneous.

(ii) *2SLS, LIML and FIML*: Using y_{2t-1} as the vector of instruments from the reduced form (3) we obtain the 2SLS estimator $\tilde{\beta}$ in (2). The asymptotics for $\tilde{\beta}$ are given by:

$$\begin{aligned} T(\tilde{\beta} - \beta) &\Rightarrow \left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2 dS_1 \right) \\ &= \left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2 dS_{1,2} \right) + \left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2 dS_2' \right) \Sigma_{22}^{-1} \sigma_{21}. \end{aligned} \quad (23)$$

Compared with (21) this limit variate eliminates one source of asymptotic bias—the third term of (21). But it still involves a matrix unit root distribution. This is because the unit roots in (3) are implicitly estimated in the first stage of the 2SLS regression. Also, the nuisance parameters (Σ_{22} , σ_{21}) still figure in the limit except when $\sigma_{21} = 0$. Thus, although single-equation structural methods like 2SLS seem to provide some improvement over single-equation least squares, they are still poor candidates for inference.

Observe that since (2) and (3) are triangular and (3) is in reduced form, FIML on the system of equations is the same as LIML on (2). But LIML and 2SLS are asymptotically equivalent (as in the stationary regressor case). Thus, (23) gives the asymptotics for the FIML estimator of β as well as 2SLS.

(iii) *Full system MLE (restricted by unit roots)*: In this case the Gaussian likelihood is the product of the conditional and marginal densities of u_{1t} and u_{2t} :

$$\prod_1^T \text{pdf}(u_{1t} | u_{2t}) \text{pdf}(u_{2t}).$$

The conditional log-likelihood is

$$-(T/2) \ln(\sigma_{11,2}) - (1/2\sigma_{11,2}) \sum_1^T [y_{1t} - \beta' y_{2t} - \sigma_{21}' \Sigma_{22}^{-1} u_{2t}]^2. \quad (24)$$

The marginal log-likelihood is

$$-(T/2) \ln |\Sigma_{22}| - (1/2) \sum_1^T u_{2t}' \Sigma_{22}^{-1} u_{2t}$$

and does not involve β . Thus the full system MLE of β is obtained by maximizing (24). Equivalently, this is just OLS on the augmented linear model (4), given earlier. Let β^+ be the estimate of β obtained in this way. Then, in an obvious partitioned regression notation,

$$\beta^+ = (Y_2' Q_\Delta Y_2)^{-1} (Y_2' Q_\Delta y_1) \quad (25)$$

and the following asymptotics apply:

$$T(\beta^+ - \beta) \Rightarrow \left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2 dS_{1,2} \right). \quad (26)$$

This is equivalent to the first term of (21) and thus the limit distribution is the Gaussian mixture given in (22). Full system MLE with prior information about the unit roots in (3) is therefore consistent and asymptotically median unbiased. As shown in Theorem 3.2 of Phillips (1989) the limit distribution (22) involves only scale parameters. It can be expressed in the equivalent form

$$\int_{g>0} N(0, g\sigma_{11}^{-1} \Sigma_{22}^{-1}) dP(g) \quad (27)$$

where $g = j'[\int_0^1 W_2 W_2']^{-1}j$, $W_2 \equiv BM(I_m)$ and j is any unit vector.

A major effect of (27) is that the scale nuisance parameters are readily eliminated by the use of conventional test statistics. In particular, asymptotic chi-squared criteria are obtained from the usual construction of Wald, Lagrange multiplier and likelihood-ratio tests. Thus, full system maximum likelihood in the presence of prior information about unit roots offers major advantages for statistical inference.

3.2. Models with weakly-dependent errors

(i) *MLE*: As argued in Phillips (1991), models with dependent errors offer no new complications as far as the asymptotic theory is concerned. The prototypical model considered above is then simply an approximation to the true model in which the weakly-dependent error is replaced by its mds approximation. The only change in the theory given in the last section for full system maximum likelihood (restricted by the use of prior information on the number of unit roots) is that the covariance matrix Σ is now the long-run covariance matrix $2\pi f_{uu}(0)$ where $f_{uu}(\lambda)$ is the spectral density matrix of the true error sequence u_t . The above remarks apply equally well to the case of finite-order autoregressive ECM representations such as (10), with the order of the matrix polynomial $J(L)$ finite. Thus, the asymptotic theory of estimation in Johansen (1988) is included, although we note again that Johansen also deals with the asymptotics of testing the rank of the cointegration space, which we do not treat here.

The remarks in the preceding paragraph do not apply to unrestricted VAR's in levels, where unit roots are not imposed but are implicitly estimated by the unrestricted regression. In such cases the limit distribution theory for the estimated vector β is of the same form as (23) and carries the matrix unit root component $[\int_0^1 S_2 S_2']^{-1}(\int_0^1 S_2 dS_2')$. By contrast, ECM formulations have the natural advantage that they work to model stationary deviations about the long-run equilibria and eliminate unit roots by construction.

(ii) *Spectral Regression*: The time-series case may be handled most simply by employing spectral regression methods. These are the subject of Phillips (1988c) to which the reader is referred. We shall mention only the main ideas and results of that paper here but we shall suggest a new single-step method of estimation. Model (2) and (3) is transformed to the frequency domain by taking discrete Fourier transforms (dft's) leading to

$$w_*(\lambda) = e\beta'w_2(\lambda) + w_u(\lambda), \quad e' = (1, 0) \quad (28)$$

where, for example, $w_*(\lambda) = (2\pi T)^{-1/2} \sum_1^T y_{*t} e^{it\lambda}$ is the dft of $y_{*t} = (y_{1t}, \Delta y_{2t})'$. Efficient and band spectral estimators of β are obtained by applying weighted least squares to (28), using estimates of the spectra of u_t as the weights. Because (28) is linear in β the estimators have the simple form:

$$\begin{aligned} \beta^\# &= [\sum_1^M e' \hat{f}_{uu}^{-1}(\omega_j) e \hat{f}_{22}(\omega_j)']^{-1} [\sum_1^M \hat{f}_{2*}(\omega_j) \hat{f}_{uu}(\omega_j) e] \\ \beta_0^\# &= \hat{f}_{22}(0)^{-1} \hat{f}_{2*}(0) \hat{f}_{uu}^{-1}(0) e / e' \hat{f}_{uu}(0)^{-1} e. \end{aligned}$$

Here $\omega_j = \pi j/M$ ($j = -M+1, \dots, M$), $\hat{f}_{ab}(\cdot)$ denotes any consistent estimate of the cross spectrum of y_{at} and y_{bt} and M is a bandwidth parameter satisfying $M \rightarrow \infty$ and $M = o(T^{1/2})$.

The asymptotics for $\beta^\#$ and $\beta_0^\#$ are identical with that of the full system MLE given in (26). They have the advantage of allowing for general stationary errors and they avoid methodological issues of specifying finite parameter error schemes and model selection issues with respect to short-run dynamics. Both estimates $\beta^\#$ and $\beta_0^\#$ arise from linear estimating equations and amount to the use of feasible GLS in the frequency domain on (28) over full and restricted frequency bands respectively.

One disadvantage of $\beta^\#$ and $\beta_0^\#$ is that they both rely on an initial estimate of the residual spectrum $\hat{f}_{uu}(\cdot)$. It is possible to avoid this two-step procedure by using an augmented regression in the frequency domain that is analogous to (4) in the time domain. Indeed, extending (4) to the frequency domain we have

$$w_1(\lambda) = \beta' w_2(\lambda) + \sigma_{21}' \Sigma_{22}^{-1} w_{\Delta y_2}(\lambda) + w_{1.2}(\lambda)$$

with $w_{1.2}(\lambda) = w_{u_1}(\lambda) - \sigma_{21}' \Sigma_{22}^{-1} w_{u_2}(\lambda)$ and $\Sigma = 2\pi f_{uu}(0)$. Now if $\lambda \rightarrow 0$ as $T \rightarrow \infty$ we have

$$w(\lambda) = \begin{bmatrix} w_{u_1}(\lambda) \\ w_{u_2}(\lambda) \end{bmatrix} \Rightarrow N(0, (1/2\pi)\Sigma)$$

so that $w_{1.2}(\lambda)$ is asymptotically independent of $w_{u_2}(\lambda) = w_{\Delta y_2}(\Delta)$. The system of equations (28) is now

$$w_1(\lambda) = \beta' w_2(\lambda) + \sigma_{21}' \Sigma_{22}^{-1} w_{\Delta y_2}(\lambda) + w_{1.2}(\lambda) \quad (29)$$

$$w_{\Delta y_2}(\lambda) = w_{u_2}(\lambda). \quad (30)$$

When $\lambda \rightarrow 0$ as $T \rightarrow \infty$ we have

$$\begin{bmatrix} w_{1.2}(\lambda) \\ w_{u_2}(\lambda) \end{bmatrix} \Rightarrow N\left(0, \frac{1}{2\pi} \begin{bmatrix} \sigma_{11.2} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}\right).$$

Thus, the system (29) and (30) has block-independent errors asymptotically and at the same time retains the triangular structure of (2) and (3). It is therefore apparent that OLS applied to (29) using frequencies in the vicinity of the origin is asymptotically equivalent to the systems estimators $\beta_0^\#$ and $\beta^\#$.

Hence the arguments that lead to the use of OLS in the time domain on the augmented regression (4) apply equally well to the frequency domain augmented regression (29).

3.3. Fully modified OLS

This is an optimal single-equation method based on the use of OLS on (2) with semiparametric corrections for serial correlation and endogeneity. The method is developed in Phillips and Hansen (1990), to which the reader is referred for a detailed discussion. The heuristic idea in the procedure is to modify the OLS estimator $\beta^* = (Y_2' Y_2)^{-1} Y_2' y_1$ so that the limit distribution given in (21) involves only the leading term. The third term may be eliminated by employing a serial correlation correction to remove σ_{21} . Recall that in the time-series case

$$\sigma_{21} = \sum_{k=0}^{\infty} E(u_{20} u_{1k}).$$

If $\hat{\sigma}_{21}$ is a consistent estimator of σ_{21} then we have a modified OLS estimator

$$\beta^{**} = (Y_2' Y_2)^{-1} (Y_2' y_1 - T \hat{\sigma}_{21})$$

whose asymptotics are

$$T(\beta^{**} - \beta) \Rightarrow \left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2 dS_1 \right), \quad (31)$$

thereby eliminating the third term of (21). The further modification for the endogeneity of y_{2t} in (2) is required to remove the correlation between the Brownian motions S_1 and S_2 in (31). This is achieved by constructing

$$y_{1t}^+ = y_{1t} - \hat{\sigma}_{21}' \hat{\Sigma}_{22}^{-1} \Delta y_{2t},$$

$$\hat{\delta}^+ = \hat{\Delta} \begin{bmatrix} 1 \\ -\hat{\Sigma}_{22}^{-1} \hat{\sigma}_{21} \end{bmatrix}$$

where $\hat{\Delta}$ is a consistent estimate of $\Delta = \sum_{k=0}^{\infty} E(u_{20} u_k')$ and $\hat{\Sigma}$ is consistent for Σ . These estimates can be constructed from the residuals of a preliminary OLS regression on (2).

The fully modified OLS estimator employs both the serial correlation and endogeneity corrections and is given by

$$\beta^+ = (Y_2' Y_2)^{-1} (Y_2' y_1^+ - T \hat{\delta}^+). \quad (32)$$

With these corrections the new estimator β^+ has the same asymptotic behaviour as the full systems MLE. Observe that it is a two-step estimator, however, and relies on the preliminary construction of y_{1t}^+ and $\hat{\delta}^+$.

Fully modified test statistics that are based on β^+ may be constructed in the usual way. Thus, for t -ratios we define

$$t_i^+ = (\beta_i^+ - \beta_i) / s_i^+$$

where $(s_i^+)^2 = \hat{\sigma}_{11.2} [(Y_2' Y_2)^{-1}]_{ii}$. Here $\hat{\sigma}_{11.2} = \hat{\sigma}_{11} - \hat{\sigma}_{21}' \hat{\Sigma}_{22}^{-1} \hat{\sigma}_{21}$ and is based on components of $\hat{\Sigma}$ which is again an estimate of the long-run covariance matrix $\Sigma = 2\pi f_{uu}(0)$. With this construction s_i^+ is a fully modified standard error and we have

$$t_i^+ \Rightarrow N(0, 1)$$

so that conventional asymptotic t -tests apply. General tests of analytic restrictions such as

$$H_0: h(\beta) = 0 \quad \text{with } H(\beta) = \partial h / \partial \beta' \text{ of full rank } q$$

may also be conducted in the usual manner. Thus, the Wald test has the form

$$W_T = h(\beta^+) [H^+ V_T H^+]^{-1} h(\beta^+)$$

with $V_T = \hat{\omega}_{11.2} (Y_2' Y_2)^{-1}$ and $H^+ = H(\beta^+)$. Since

$$W_T \Rightarrow \chi_q^2$$

under H_0 , conventional chi-squared critical values apply.

Simulations reported in Hansen and Phillips (1990) and Phillips and Hansen (1990) indicate that these tests perform adequately in samples of size $T = 50$, at least for small scale models with only two or three variables in the long-run relationship.

3.4. Single-equation ECM's (SEECM's)

Suppose we are interested in building a SEECM to model the long-run cointegrating relationship (2) and the stationary deviations u_{1t} about it. As we have seen in Section 2.5, the Hendry methodology works with an equation of the form (15) or (16), allowing for the fact that some of the regressors may be eliminated due to insignificant coefficients and others may be replaced by higher-order differences for reasons of explanatory power, parsimony or economic interpretability. As argued in that section, these equations are

empirical versions of (17), whose error η_t is an mds with respect to the filtration $\mathcal{F}_{t-1} = \sigma\{\Delta y_{1t-1}, \Delta y_{1t-2}, \dots, \Delta y_{2t}, \Delta y_{2t-1}, \dots\}$.

Suppose the generating mechanism for u_t in (2) and (3) is the linear process

$$u_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}; \quad A_0 = I, \quad \sum_0^{\infty} j^{1/2} \|A_j\| < \infty \quad (33)$$

where $(\varepsilon_t) = \text{iid } N(0, \Sigma)$. Assume also that (33) can be inverted and written in autoregressive form as

$$B(L)u_t = \varepsilon_t, \quad B(L) = \sum_{j=0}^{\infty} B_j L^j, \quad B_0 = I. \quad (34)$$

This will be useful in what follows. We shall now try to build a single-equation model from this explicit representation for the error.

Observe that if η_t is orthogonal to Δy_{2t} and the past history $(\Delta y_{t-1}, \Delta y_{t-2}, \dots)$ then η_t is orthogonal to ε_{2t} and the past history $(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$, where we partition ε_t conformably with u_t . Under Gaussian assumptions we deduce that

$$\eta_t = \varepsilon_{1t} - \sigma'_{21} \Sigma_{22}^{-1} \varepsilon_{2t} = \text{iid } N(0, \sigma_{11.2}).$$

Thus, η_t is just ε_{1t} centred about its conditional mean given ε_{2t} . The conditional log-likelihood of $(\varepsilon_{1t})_1^T$ given $(\varepsilon_{2t})_1^T$ is:

$$-(T/2) \ln \sigma_{11.2} - (1/2\sigma_{11.2}) \sum_1^T (\varepsilon_{1t} - \sigma'_{21} \Sigma_{22}^{-1} \varepsilon_{2t})^2.$$

Maximizing this is equivalent to minimizing the sum of squares

$$\sum_1^T \{(b_{11}(L), b_{12}(L))u_t - \sigma'_{21} \Sigma_{22}^{-1} (b_{21}(L), B_{22}(L))u_t\}^2$$

where $B(L)$ in (34) is partitioned conformably with u_t . This is the same as running least squares on the equation

$$y_{1t} = \beta' y_{2t} + d_1(L)(y_{1t} - \beta' y_{2t}) + d_2(L)' \Delta y_{2t} + \eta_t \quad (35)$$

where

$$d_1(L) = b_{11}(L) - \sigma'_{21} \Sigma_{22}^{-1} b_{21}(L) - 1 = \sum_{j=1}^{\infty} d_{1j} L^j, \quad \text{say}$$

$$d_2(L) = b_{12}(L) - \sigma'_{21} \Sigma_{22}^{-1} B_{22}(L) = \sum_{j=0}^{\infty} d_{2j} L^j, \quad \text{say.}$$

Models, such as (35), that involve lagged equilibria as regressors have been employed in SEECM empirical work, e.g. Hendry and von Ungern-Sternberg (1981).

In spite of their apparent similarity, equations (35) and (17) can have very different implications in practice. Suppose, for example, that the polynomials $d_1(L)$ and $d_2(L)$ in (35) are of finite degree with $d_1(L) = d_{11}L$ and $d_2(L) = d_{20} + d_{21}L$. Then (35) is simply

$$y_{1t} = \beta' y_{2t} + d_{11}(y_{1t-1} - \beta' y_{2t-1}) + d_{20} \Delta y_{2t} + d_{21} \Delta y_{2t-1} + \eta_t. \quad (36)$$

Estimation of β in (36) involves a simple nonlinear least-squares regression.

The corresponding equation (17) can be deduced from (36) by taking initial conditions at $t = 0$ setting $y_0 = 0$ and writing

$$y_{1t-1} = \sum_{k=1}^{t-1} \Delta y_{1t-k}, \quad y_{2t-1} = \sum_{k=1}^{t-1} \Delta y_{2t-k}. \quad (37)$$

Then (17) is

$$y_{1t} = \beta' y_{2t} + \sum_{k=1}^{t-1} f_{1k} \Delta y_{1t-k} + \sum_{k=0}^{t-1} f'_{2k} \Delta y_{2t-k} + \eta_t \quad (38)$$

where $f_{1k} = d_{11}(\nabla k)$, $f_{20} = d_{20}$, $f_{21} = d_{21} - d_{11}\beta$, $f_{2k} = 0$, $k > 1$. As is apparent from these formulae, the coefficients in (38) and by implication the coefficients in (17), in general, do not decay as the lag increases. This is because the representation relies on the partial sums (37) where the weight of individual innovations is unity. Thus, in order to model short-run dynamics using the variables Δy_{1t-k} and Δy_{2t-k} it is necessary, in general, to include all lags because of shock persistence. Of course, this is quite impractical in applications and cannot be justified in theory where truncation arguments are needed to develop the asymptotics. These arguments strongly support the use of nonlinear formulation (35) over linear models like (17).

It is illuminating to develop the asymptotics in the special case (36) and examine the practical impact of truncation in the formulation that is based on (38). For specificity assume that terms in (38) are truncated at a single lag. The fitted regression is

$$y_{1t} = \hat{\beta}' y_{2t} + \hat{f}_{11} \Delta y_{1t-1} + \hat{f}_{20} \Delta y_{2t} + \hat{f}_{21} \Delta y_{2t-1} + \hat{\eta}_t. \quad (39)$$

This is equivalent to replacing the composite variable $y_{1t-1} - \beta' y_{2t-1}$ in (36) by the regressor Δy_{1t-1} . Next suppose that the true generating mechanism for the error u_t in (2) and (3) is (34) with a diagonal autoregressive operator and a scalar error covariance matrix, i.e.

$$B(L) = \begin{bmatrix} b_{11}(L) & 0 \\ 0 & B_{22}(L) \end{bmatrix}, \quad \Sigma = \sigma^2 I_n. \quad (40)$$

Let the degrees of the block diagonal elements of $B(L)$ all be unity. Then we have

$$d_1(L) = b_{11}(L) - 1 = d_{11}L, \quad B_{22}(L) = I - D_{22}L, \quad \text{say}$$

and

$$d_2(L) = b_{12}(L) - \sigma_{21}' \Sigma_{22}^{-1} B_{22}(L) = 0. \quad (41)$$

Equation (36) is correctly specified but has surplus variables Δy_{2t} and Δy_{2t-1} since $d_{20} = d_{21} = 0$ in view of (41). The error in (36) is simply $\eta_t = \varepsilon_t = \text{iid } N(0, \sigma^2)$ and is independent of u_{2s} for all t and s . Some standard asymptotic calculations now lead to the following results.

(i) *OLS on equation (2)*

$$T(\beta^* - \beta) \Rightarrow \left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2 dS_1 \right) \equiv \int_{G>0} N(0, \omega_{11}G) dP(G)$$

where $G = \left(\int_0^1 S_2 S_2' \right)^{-1}$ and

$$\begin{bmatrix} S_1 \\ S_2 \end{bmatrix} = BM(\Omega), \quad \Omega = \begin{bmatrix} \omega_{11} & 0 \\ 0 & \Omega_{22} \end{bmatrix} = \sigma^2 \begin{bmatrix} b_{11}(1)^{-2} & 0 \\ 0 & B_{22}(1)^{-1} B_{22}'(1)^{-1} \end{bmatrix}$$

(ii) *NLS on equation (36)*

$$T(\tilde{\beta} - \beta) \Rightarrow \left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2 dS_1 \right) \equiv \int_{G>0} N(0, \omega_{11}G) dP(G).$$

(iii) *OLS on equation (39)*

$$T(\hat{\beta} - \beta) \Rightarrow \left(\int_0^1 S_2 S_2' \right)^{-1} \left\{ \int_0^1 S_2 dS_1' - \left[\int_0^1 S_2 dS_2' L_0 + L_1 \right] Q^{-1} q \right\}$$

where

$$\begin{aligned}
 L_0 &= [\beta, I, I], & L_1 &= [\Delta'_1 \beta, \Delta'_0, \Delta'_1] \\
 \Delta_0 &= \sum_{k=0}^{\infty} E(u_{20} u'_{2k}) = V_{22} (I - D_{22})^{-1} \\
 \Delta_1 &= \sum_{k=-1}^{\infty} E(u_{20} u'_{2k}) = D_{22} V_{22} + V_{22} (I - D_{22})^{-1} \\
 Q &= \begin{bmatrix} \beta' V_{22} \beta + 2(1 - d_{11}) v_{11} & \beta' D_{22} V_{22} & \beta' V_{22} \\ V_{22} D'_{22} \beta & V_{22} & D_{22} V_{22} \\ V_{22} \beta & V_{22} D'_{22} & V_{22} \end{bmatrix} \\
 q' &= [(1 - d_{11}) v_{11}, 0, 0], & v_{11} &= E(u_{1t}^2).
 \end{aligned}$$

These results show that for this special autoregressive cointegrated system, β^* and $\tilde{\beta}$ are asymptotically equivalent and the limit distribution given in (i) and (ii) is the same as that of the full system MLE and spectral regression estimators described in Sections 3.1 and 3.2. Observe that OLS in (2) ignores the dynamic error structure of u_{1t} . Yet it is still asymptotically equivalent to the MLE which jointly estimates the error dynamics. This outcome is indicative of a general result: viz. that efficient estimation of β can be achieved without fully estimating the error dynamics of equation (2). Note that NLS on equation (36) does fit the error dynamics on (2) by the inclusion of the autoregressive term. Yet this does not improve the asymptotic efficiency of the estimator. On the other hand, there is a clear advantage to the use of NLS on (36) in practice, even in this simple case. This is that the error η_t is an mds and inference can proceed in the usual way with asymptotic normal t -ratios and asymptotic chi-squared criteria constructed in the usual fashion. This is *not* true of OLS on (2). The usual test statistics need to be modified in the latter case to allow for the serial dependence in u_{1t} . This is achieved by the use of long-run variance estimates in the construction of the statistics rather than the usual error sum of squares. The situation is identical to that described in the study of Phillips and Park (1988) on the asymptotic equivalence of OLS and GLS in cointegrating regressions with exogenous regressors.

The outcome for case (iii) shows the disadvantages of working with equation (39). In fact, Δy_{1t-1} is not an adequate proxy for the lagged cointegrating relationship $y_{1t-1} - \beta' y_{2t-1}$ that appears in (36). As a result, the limit distribution of $\hat{\beta}$ in (39) involves second-order bias effects, is asymmetric and involves non-scale nuisance parameters. These factors suggest that, at least on the basis of asymptotic theory, this type of specification will be a poor choice for inference compared with models like (36). The extent of these problems in finite samples will be explored in Section 4.

The structure imposed by (40) on the error dynamics is very special because it ensures that u_{1t} and u_{2t} are incoherent time series, i.e. they are uncorrelated at all lags and the spectral density matrix of u_t is block diagonal, just like the matrices in (40). This ensures that the regressor y_{2t} in (2) is *strongly* exogenous and we have the following equivalence of the wide sense conditional expectations

$$\tilde{E}(u_{1t} | (u_2)'_{-\infty}) = \tilde{E}(u_{1t} | (u_2)'_{-\infty}, (u_2)_{t+1}^{\infty}). \quad (42)$$

Of course u_{1t} and u_{2t} will generally be coherent time series and (42) will not apply. Some simple examples are studied in Phillips (1988a). When this happens there is a failure of valid conditioning in models like (2) and (35). The consequences are easy to explore in the general case. We shall work with the general ECM specification (35) and assume that the lag polynomials $d_1(L)$ and $d_2(L)$ are either of finite degree p or that they

are of infinite degree and in the regression p is permitted to move to infinity as $T \rightarrow \infty$ but at the controlled rate $p = o(T^{1/3})$, so that conventional asymptotic arguments, as in Berk (1974) and Said and Dickey (1984), may be employed to accommodate the infinite-dimensional case. The relevant first-order conditions for NLS are:

$$\sum_1^T \{y_{1t} - \hat{\beta}y_{2t} - \hat{d}_1(L)(y_{1t} - \hat{\beta}y_{2t}) - \hat{d}_2(L)\Delta y_{2t}\}(y_{2t} - \hat{d}_1(L)y_{2t}) = 0$$

which we may rewrite in the form

$$\begin{aligned} \sum_1^T \{[d_1(L) - \hat{d}_1(L)](y_{1t} - \beta'y_{2t}) + (\beta - \hat{\beta})'(1 - \hat{d}_1(L))y_{2t} \\ + [d_2(L) - \hat{d}_2(L)]\Delta y_{2t} + \eta_t\}[1 - \hat{d}_1(L)]y_{2t} = 0. \end{aligned} \quad (43)$$

Since $\hat{d}_1(L) - d_1(L)$, $\hat{d}_2(L) - d_2(L) = o_p(1)$ and since the sample moments $\{T^{-2} \sum_1^T y_{2t-k}y'_{2t}\}$ are asymptotically collinear for all fixed k (the case of infinite operators $d_i(L)$ can be dealt with by truncation arguments) we find that (43) gives rise to the asymptotics

$$T(\tilde{\beta} - \beta) \Rightarrow \left\{ (1 - d_1(1))^2 \int_0^1 S_2 S_2' \right\}^{-1} \left\{ (1 - d_1(1)) \int_0^1 S_2 dS_\eta \right\} \quad (44)$$

where

$$\begin{bmatrix} S_\eta \\ S_2 \end{bmatrix} = \text{BM}(\Phi). \quad (45)$$

We define $\zeta_t = (\eta_t, u'_{2t})'$ and then

$$\zeta_t = \begin{bmatrix} 1 & -\sigma'_{21}\Sigma_{22}^{-1} \\ a_{21}(L) & A_{22}(L) \end{bmatrix} \varepsilon_t = G(L)\varepsilon_t, \text{ say.}$$

The limit covariance matrix in (45) is therefore

$$\Phi = \omega f'_{\zeta\zeta}(0) = G(1)\Sigma G(1)'.$$

The limit distribution represented by (44) can alternatively be decomposed as

$$(1 - d_1(1))^{-1} \left(\int_0^1 S_2 S_2' \right)^{-1} \left\{ \int_0^1 S_2 dS_\eta + \int_0^1 S_2 dS_2' \Phi_{22}^{-1} \varphi_{21} \right\}. \quad (46)$$

As is apparent from this formula bias, asymmetry and non-scale nuisance parameters are a feature of the limit distribution in the general case. These features are due to the non-diagonal nature of Φ arising from the feedback from η_t to u_{2t} .

The antidote to the failure of (35) to produce an asymptotically efficient estimator of β in the general case is the elimination of this feedback. This can be achieved by including leads of Δy_{2t} in the regression so that in the limit η_t is orthogonal to the entire history $(\Delta y_{2t})_{-\infty}^{\infty}$.¹ The revised specification has the form

$$y_{1t} = \beta'y_{2t} + d_1(L)(y_{1t} - \beta'y_{2t}) + d_2(L)\Delta y_{2t} + d_3(L^{-1})'\Delta y_{2t} + v_t \quad (35)'$$

where

$$d_3(L^{-1}) = \sum_{k=1}^{\infty} d_{3k}L^{-k}$$

and $d_1(L)$ and $d_2(L)$ are as before in (35). The coefficients of $d_3(L^{-1})$ are delivered from the linear least-squares projection

$$\tilde{E}(\eta_t | (u_{2s})_{t+1}^{\infty}) = \sum_{k=1}^{\infty} d_{3k}u_{2t+k}.$$

1. The inclusion of leads of Δy_{2t} as a means of achieving valid conditioning has also been advanced by Saikkonen (1991) and Stock and Watson (1991). We became aware of the existence of these papers after our own work was completed.

The new error on (35)' is

$$\nu_t = \eta_t - d_3(L^{-1})' u_{2t}$$

and is an mds with respect to the filtration $\mathcal{M}_{t-1} = \sigma(u_{1t-1}, u_{1t-2}, \dots; (u_{2t})_{-\infty}^{\infty})$.

Observe that (35)' has the alternate SEECM format

$$\Delta y_{1t} = \bar{d}_1(L)(y_{1t} - \beta' y_{2t}) + \bar{d}_2(L)\Delta y_{2t} + d_3(L^{-1})'\Delta y_{2t} + \nu_t \quad (35)''$$

where

$$\bar{d}_1(L) = d_1(L) - L, \quad \bar{d}_2(L) = d_2(L) + L.$$

Estimates of β obtained from (35)' and (35)'' using NLS will be asymptotically equivalent. We shall therefore focus on (35)' in the following discussion.

Following the same line of argument as that above based on first-order conditions we deduce the following asymptotics for the NLS estimator of β in (35)':

$$T(\tilde{\beta} - \beta) \Rightarrow (1 - d_1(1))^{-1} \left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2 dS_\nu \right). \quad (47)$$

Now observe that

$$\begin{bmatrix} S_\nu \\ S_2 \end{bmatrix} = BM(\Psi)$$

with covariance matrix

$$\Psi = \begin{bmatrix} (1 - d_1(1))\omega_{11.2} & 0 \\ 0 & \Omega_{22} \end{bmatrix}. \quad (48)$$

To verify the form of the matrix given in (48) we note the following:

- (i) Ψ is block-diagonal because if $\zeta_t = (\nu_t, u_{2t}')'$ then by virtue of the construction of ν_t we have a block-diagonal autocovariance function $\gamma_\zeta(h) = E(\zeta_t \zeta_{t+h}')$. Thus the spectral density matrix

$$f_{\zeta\zeta}(\lambda) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma_\zeta(h) e^{ih\lambda}$$

is also block-diagonal. This includes the value of $f_{\zeta\zeta}(\lambda)$ at the origin and, hence, Ψ .

- (ii) The error ν_t is that part of u_{1t} that is orthogonal to $(u_{1s})_{-\infty}^{t-1}$ and $(u_{2s})_{-\infty}^{\infty}$. The long-run variance of $(1 - d_1(1))^{-1}\nu_t$ is the same as the long-run variance of $(1 - d_1(L))^{-1}\nu_t$ or $\bar{\nu}_t$ where $\bar{\nu}_t = d_1(L)\nu_t + \nu_t$. But this is equivalent to the long-run variance of u_{1t} given $(u_{2s})_{-\infty}^{\infty}$, which is $\omega_{11.2}$. Hence

$$lr \text{ var } (1 - d_1(1))^{-1}\nu_t = \omega_{11.2}$$

so that

$$lr \text{ var } (\nu_t) = (1 - d_1(1))^2 \omega_{11.2}.$$

Next we observe that

$$(1 - d_1(1))^{-1}S_\nu = S_{1.2}$$

and thus (47) may also be written as

$$T(\tilde{\beta} - \beta) \Rightarrow \left(\int_0^1 S_2 S_2' \right)^{-1} \left(\int_0^1 S_2 dS_{1.2} \right).$$

Hence the NLS estimator of β from (35)' is asymptotically equivalent to the full system MLE and spectral regression estimator.

To sum up, the results of this section are as follows:

- (1) Asymptotically efficient estimation of long-run equilibrium relationships can be achieved by a variety of methods. In particular, the following methods are fully efficient and asymptotically median unbiased:
 - full systems MLE (restricted by the imposition of unit roots)
 - fully modified OLS (with semiparametric serial correlation and endogeneity corrections)
 - systems estimation in the frequency domain (restricted by the imposition of unit roots).
 - single-equation band spectral estimation (with the dft of Δy_{2t} as an additional regressor)
 - nonlinear ECM's (with lagged equilibrium relations and both lags and leads of Δy_{2t} as regressors)

Each of these methods achieves full efficiency in the limit by working to estimate (and eliminate) the effects of long-run feedback between the errors on the long-run equilibrium relationship (i.e. u_{1t}) and the errors that drive the regressors (i.e. u_{2t}). The methods are asymptotically equivalent for the estimation of long-run economic equilibria and lead to conventional chi-squared criteria for inferential purposes with respect to these coefficients.

- (2) OLS, conventional SEM methods, and unrestricted VAR's lead in general to estimators that are asymptotically biased and whose distributions involve unit root asymptotics and nontrivial nuisance parameters.
- (3) In SEECM modelling, valid conditioning on the regressors generally fails. With feedback from u_1 to u_2 , *leads* of Δy_{2t} must be included in the SEECM specifications to obtain errors that form an mds sequence with respect to the past history of u_1 and the full history of u_2 . This is important for estimator efficiency, unbiasedness and for inference.
- (4) In SEECM modelling there is an asymptotic advantage to the use of lagged equilibrium relationships in the regression and thereby the use of nonlinear least squares (NLS). This is because lags of Δy_1 are not in general an adequate proxy for the past history of u_1 , because of the persistence in the effects of the innovations that arises from the presence of unit roots in the system. Thus, asymptotic theory favours the use of NLS on non-linear-in-parameters SEECM's rather than simply OLS on linear SEECM models formulated with lags (and possibly leads) of differences in all variables in the system. In effect, the requisite information set for valid conditioning is better modelled by employing lagged equilibria than it is by the use of lagged differences in the dependent variable. We note that such formulations have indeed been used in past empirical work—see Hendry and von Ungern-Sternberg (1981).

4. SIMULATIONS

The simulations reported here are intended to address the small-sample performance of the following procedures: OLS, full modified (“FM”) OLS, linear ECM or Hendry (“H”) estimators based on (17) and (35), and our new estimator (“PL”) based on the nonlinear regression equations (35)' that includes leads of Δy_{2t} .¹ The data-generating process we use is related to that of Banerjee *et al.* (1986) and is identical to the one in Phillips and

Hansen (1990). The model is

$$y_{1t} = \alpha + \beta y_{2t} + u_{1t} \quad (49)$$

$$y_{2t} = y_{2t-1} + u_{2t}, \quad t = 1, \dots, T \quad (50)$$

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = u_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \equiv \text{iid } N(0, \Sigma).$$

We set

$$\alpha = 0, \quad \beta = 2, \quad T = 50, \\ \theta = \begin{bmatrix} 0.3 & 0.4 \\ \theta_{21} & 0.6 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & \sigma_{21} \\ \sigma_{21} & 1 \end{bmatrix}$$

and allow θ_{21} and σ_{21} to vary. Earlier work in Phillips and Hansen (1990) established these as the critical parameters. We consider values of $\{0.8, 0.4, 0.0, -0.8\}$ for θ_{21} and $\{-0.85, -0.5, 0.5\}$ for σ_{21} . Our simulation model is thus a special case of the general cointegrated system (2) and (3) studied earlier. Here y_{2t} is a scalar regressor and the time-series dynamics are generated by an MA(1). The number of replications for all simulations is 10,000.

To make the OLS t -statistic comparable to those of the other estimators, we use an estimate of the long-run variance $\hat{\omega}_{11}$ rather than the standard variance estimator, thus following the recommended procedure in Phillips and Park (1988) for models with exogenous regressors. The nuisance parameter estimates that enter the fully-modified procedure are obtained by using a triangular Bartlett window of lag length 5, the OLS residuals \hat{u}_{1t} and $u_{2t} = \Delta y_{2t}$ to calculate $\hat{\Omega}_{21}$ and $\hat{\Delta}_{21}$. In order to distinguish the separate effects of including additional lags and adding leads to Hendry-style regression equations based on (15), we employ models with the following combinations of lags and leads of the covariates Δy_{1t} and Δy_{2t} : (2, 0), (4, 0), (2, 1) and (4, 2). The same combinations of lags and leads are used for estimators based on (35) and (35)', these being designated as PL. Thus, 10 estimators were calculated at each replication. (The simulations reported in Phillips and Hansen (1990) studied the first three (OLS, FM(5), and H(2, 0)) of these estimators.)

For example, the nonlinear regression equation for the PL(2, 1) model is:

$$\begin{aligned} y_{1t} = & \hat{\alpha} + \hat{\beta} y_{2t} + \hat{d}_{11}(y_{1t-1} - \hat{\alpha} - \hat{\beta} y_{2t-1}) \\ & + \hat{d}_{12}(y_{1t-2} - \hat{\alpha} - \hat{\beta} y_{2t-2}) + \hat{d}_{20} \Delta y_{2t} \\ & + \hat{d}_{21} \Delta y_{2t-1} + \hat{d}_{22} \Delta y_{2t-2} + \hat{d}_{31} \Delta y_{2t+1} + \hat{v}_t, \end{aligned} \quad (51)$$

i.e. two lags and one lead in (35)'. Since (51) is bilinear in the parameters $\{\alpha, \beta\}$ and $\{d_{11}, d_{12}, d_{20}, d_{21}, d_{22}, d_{31}\}$, nonlinear least-squares estimation requires an iterative method to minimize the residual sum of squares. Rather than use a general nonlinear optimization method, we found that substantial gains were possible from using an algorithm that takes advantage of the bilinearity of the model. One such algorithm, proposed by Stone (1954) in a related context, is iterated OLS. However a drawback of this method is that its rate of convergence is usually very slow. A superior method described in Deaton (1973) is the "ridge-walking algorithm", a variant of the Newton-Raphson method that fully exploits the bilinear structure of the model by solving one subset of the first-order conditions conditional on the other subset of parameters. It relies

on explicit calculation of the gradients and the Hessian, both of which are obtained analytically for our model. The ridge-walking algorithm converges more quickly and is computationally simpler than the generic Newton–Raphson method in the bilinear case. But the ridge-walking algorithm, like the Newton–Raphson, has the drawback that the Hessian is not necessarily positive definite for a given set of parameters (especially when they are not yet close to the optimum), leading to an unacceptable step. Deaton (1973) therefore used a modified ridge-walking algorithm (pp. 240–242), essentially using the quadratic hill-climbing method (Marquardt (1963)). We found that this modification was unsuccessful in many of our replications so we chose a different solution. At each iteration, we computed the Hessian: if it was positive-definite we computed the updated parameter vector by the ridge-walking method; if not, we calculated the new parameters by iterated OLS, using the parameters obtained in the preceding step as starting values. (Initial parameter values were also obtained by iterated OLS.) One advantage of this combined method is that the objective function need not be computed at each step, since the method is guaranteed to produce a valid step. Furthermore, it is very simple to implement. The good choice of starting values combined with the efficiency of the ridge-walking method led to rapid convergence to an optimum for most replications, usually in not more than four steps. This is in contrast to all the other methods mentioned above, which encountered convergence difficulties for a small percentage (5–10%) of the replications.

Monte Carlo means and standard deviations of the bias term ($\hat{\beta} - \beta$) and the t -ratio statistic $\tau(\hat{\beta})$ for the 10 estimators and the 12 combinations of $\{\theta_{21}, \sigma_{21}\}$ are given in Tables 2 and 3, respectively. The densities of the estimators and their t -ratios were calculated using nonparametric density estimates computed with a standard normal kernel and bandwidths of 0.02 for the bias terms and 0.4 for the t -ratios. These choices were made after some experimentation with other values. Figures 1–6 graph the densities of $\tau(\hat{\beta})$ for selected parameter configurations. Graphs of the densities of $\hat{\beta} - \beta$ are given in the original version of the paper, Phillips and Loretan (1989), which is available on request.

Our main results are:

(1) *The bias term ($\hat{\beta} - \beta$): see Table 2*

- (i) When $(\theta_{21} > 0, \sigma_{21} < 0)$ OLS is usually the most biased estimator. The FM procedure reduces the bias substantially. Adding additional lags—e.g. going from $H(2, 0)$ to $H(4, 0)$ —or including leads—e.g., going from $H(2, 0)$ to $H(2, 1)$ —of the covariates reduces the bias of the linear SEECM estimators. These results strongly endorse the asymptotic theory of Section 3.
- (ii) When $\theta_{21} \neq 0$, adding leads is more effective than increasing the number of lags in terms of bias reduction. Conversely, adding leads is less useful when $\theta_{21} = 0$.
- (iii) When $\theta_{21} = 0$ (i.e. when there is no feedback from u_1 to u_2) all SEECM's provide essentially unbiased estimates of β , and are better centred than the FM distribution. This supports the theoretical result that $\theta_{21} = 0$ is an essential ingredient of valid asymptotic conditioning in conventional SEECM methodology.
- (iv) The FM method is more biased than OLS when $(\theta_{21} < 0, \sigma_{21} > 0)$. The failure of the semiparametric corrections procedure in this case is analogous to the size distortions of the unit root tests of Phillips (1987) and Phillips and Perron (1988) in the presence of negative serial correlation of the errors. The simulations in Schwert (1987) reveal these effects for unit root tests under MA(1) errors with negative serial correlation.

TABLE 2

Means and standard deviations of the bias term $(\hat{\beta} - \beta)$, for various values of σ_{21} and θ_{21}

$\sigma_{21} = -0.85$	$\theta_{21} = 0.8$	$\theta_{21} = 0.4$	$\theta_{21} = 0.0$	$\theta_{21} = -0.8$
OLS	-0.1466 (0.124)	-0.0957 (0.092)	-0.0564 (0.060)	-0.0204 (0.030)
FM (5)	-0.0092 (0.122)	-0.0278 (0.077)	-0.0252 (0.050)	-0.0038 (0.022)
H (2, 0)	-0.0573 (0.099)	-0.0188 (0.060)	-0.0021 (0.036)	0.0073 (0.017)
H (4, 0)	-0.0430 (0.103)	-0.0146 (0.065)	-0.0016 (0.041)	0.0069 (0.021)
H (2, 1)	-0.0358 (0.100)	-0.0095 (0.063)	-0.0015 (0.038)	-0.0059 (0.017)
H (4, 2)	-0.0189 (0.113)	-0.0067 (0.075)	-0.0017 (0.048)	0.0065 (0.023)
PL (2, 0)	-0.0789 (0.118)	-0.0249 (0.064)	-0.0036 (0.035)	0.0057 (0.015)
PL (4, 0)	-0.0498 (0.121)	-0.0144 (0.072)	-0.0022 (0.042)	0.0049 (0.018)
PL (2, 1)	-0.0284 (0.108)	-0.0101 (0.066)	-0.0030 (0.038)	-0.0049 (0.016)
PL (4, 2)	-0.0006 (0.128)	-0.0025 (0.083)	-0.0028 (0.049)	0.0035 (0.020)
$\sigma_{21} = -0.5$	$\theta_{21} = 0.8$	$\theta_{21} = 0.4$	$\theta_{21} = 0.0$	$\theta_{21} = -0.8$
OLS	-0.0782 (0.089)	-0.0611 (0.077)	-0.0433 (0.061)	-0.0144 (0.033)
FM (5)	-0.0435 (0.101)	-0.0262 (0.082)	0.0170 (0.062)	0.0067 (0.033)
H (2, 0)	-0.0552 (0.092)	-0.0283 (0.075)	-0.0060 (0.057)	0.0180 (0.034)
H (4, 0)	-0.0466 (0.098)	-0.0227 (0.083)	-0.0024 (0.065)	0.0190 (0.040)
H (2, 1)	-0.0165 (0.088)	-0.0069 (0.078)	-0.0056 (0.061)	-0.0086 (0.032)
H (4, 2)	-0.0100 (0.104)	-0.0086 (0.094)	-0.0033 (0.077)	0.0073 (0.041)
PL (2, 0)	-0.0806 (0.117)	-0.0337 (0.082)	-0.0066 (0.058)	0.0111 (0.030)
PL (4, 0)	-0.0657 (0.133)	-0.0253 (0.093)	-0.0026 (0.067)	0.0121 (0.036)
PL (2, 1)	-0.0079 (0.096)	-0.0013 (0.084)	-0.0061 (0.062)	-0.0096 (0.031)
PL (4, 2)	-0.0005 (0.123)	-0.0044 (0.103)	-0.0033 (0.078)	0.0033 (0.037)
$\sigma_{21} = +0.5$	$\theta_{21} = 0.8$	$\theta_{21} = 0.4$	$\theta_{21} = 0.0$	$\theta_{21} = -0.8$
OLS	-0.0219 (0.036)	-0.0173 (0.042)	-0.0071 (0.048)	0.0296 (0.058)
FM (5)	-0.0213 (0.044)	-0.0105 (0.048)	0.0063 (0.056)	0.0697 (0.094)
H (2, 0)	-0.0193 (0.041)	-0.0164 (0.048)	-0.0090 (0.057)	0.0270 (0.068)
H (4, 0)	-0.0155 (0.046)	-0.0116 (0.055)	-0.0042 (0.064)	0.0323 (0.079)
H (2, 1)	0.0070 (0.040)	0.0025 (0.049)	-0.0083 (0.061)	-0.0125 (0.068)
H (4, 2)	-0.0069 (0.048)	-0.0061 (0.062)	-0.0043 (0.075)	-0.0048 (0.082)
PL (2, 0)	-0.0223 (0.048)	-0.0185 (0.053)	-0.0098 (0.059)	0.0320 (0.070)
PL (4, 0)	-0.0163 (0.048)	-0.0122 (0.057)	-0.0042 (0.067)	0.0375 (0.086)
PL (2, 1)	0.0145 (0.048)	0.0066 (0.054)	-0.0085 (0.063)	-0.0115 (0.067)
PL (4, 2)	-0.0087 (0.054)	-0.0066 (0.066)	-0.0042 (0.080)	-0.0032 (0.083)

(v) The nonlinear models do not, in general, reduce the bias of the estimates compared to the corresponding linear specifications.

(vi) Densities of $\hat{\beta} - \beta$ are graphed in Phillips and Loretan (1989). These show the bias and skewness of the OLS estimator and the effect of adding additional lags and leads of the covariates in the SEECM's in reducing these deficiencies. When $\theta_{21} = 0$ the densities of the various SEECM estimators are essentially indistinguishable.

(2) The t -ratio statistic $\tau(\hat{\beta})$: see Table 3 and Figures 1-6.

(i) When $(\theta_{21} > 0, \sigma_{21} < 0)$, OLS again yields the greatest bias for the t -ratios. The FM estimator performs very well in these cases.

(ii) The advantage of the nonlinear over the linear specification of the SEECM's is strongly evident. The nonlinear estimators improve the performance of the

TABLE 3
Means and standard deviations of the t -ratios $\tau(\hat{\beta})$ for various values of σ_{21} and θ_{21}

$\sigma_{21} = -0.85$	$\theta_{21} = 0.8$	$\theta_{21} = 0.4$	$\theta_{21} = 0.0$	$\theta_{21} = -0.8$
OLS	-1.679 (1.22)	-1.279 (1.08)	-0.938 (0.95)	-0.504 (0.83)
FM (5)	-0.216 (1.35)	-0.426 (1.02)	-0.454 (0.82)	-0.137 (0.61)
H (2, 0)	-1.266 (2.07)	-0.532 (1.71)	-0.079 (1.37)	0.383 (0.89)
H (4, 0)	-1.015 (2.30)	-0.433 (1.86)	-0.056 (1.44)	0.336 (0.99)
H (2, 1)	-0.790 (2.18)	-0.251 (1.76)	-0.051 (1.40)	0.342 (0.99)
H (4, 2)	-0.421 (2.51)	-0.191 (1.95)	-0.053 (1.51)	0.351 (1.18)
PL (2, 0)	-0.921 (1.33)	-0.510 (1.36)	-0.137 (1.37)	0.422 (1.24)
PL (4, 0)	-0.646 (1.68)	-0.317 (1.69)	-0.076 (1.65)	0.359 (1.54)
PL (2, 1)	-0.394 (1.45)	-0.222 (1.41)	-0.108 (1.41)	-0.377 (1.32)
PL (4, 2)	-0.029 (1.86)	-0.079 (1.79)	-0.086 (1.75)	0.244 (1.72)
$\sigma_{21} = -0.5$	$\theta_{21} = 0.8$	$\theta_{21} = 0.4$	$\theta_{21} = 0.0$	$\theta_{21} = -0.8$
OLS	-1.258 (1.31)	-1.020 (1.23)	-0.788 (1.12)	-0.353 (0.92)
FM (5)	-0.724 (1.49)	-0.492 (1.33)	0.360 (1.14)	0.102 (0.82)
H (2, 0)	-1.140 (1.77)	-0.604 (1.61)	-0.138 (1.39)	0.577 (1.00)
H (4, 0)	-0.942 (1.85)	-0.472 (1.69)	-0.057 (1.47)	0.558 (1.09)
H (2, 1)	-0.384 (2.01)	-0.149 (1.69)	-0.118 (1.42)	-0.302 (1.14)
H (4, 2)	-0.229 (2.19)	-0.180 (1.83)	-0.067 (1.53)	0.246 (1.31)
PL (2, 0)	-0.907 (1.25)	-0.496 (1.34)	-0.131 (1.40)	0.464 (1.30)
PL (4, 0)	-0.731 (1.54)	-0.374 (1.62)	-0.055 (1.67)	0.476 (1.54)
PL (2, 1)	-0.125 (1.44)	-0.031 (1.42)	-0.110 (1.43)	-0.397 (1.37)
PL (4, 2)	-0.022 (1.79)	-0.084 (1.75)	-0.066 (1.77)	0.127 (1.72)
$\sigma_{21} = +0.5$	$\theta_{21} = 0.8$	$\theta_{21} = 0.4$	$\theta_{21} = 0.0$	$\theta_{21} = -0.8$
OLS	-0.676 (1.14)	-0.477 (1.20)	-0.172 (1.22)	0.596 (1.20)
FM (5)	-0.583 (1.22)	-0.238 (1.28)	0.174 (1.34)	1.334 (1.51)
H (2, 0)	-0.622 (1.26)	-0.458 (1.34)	-0.208 (1.37)	0.545 (1.33)
H (4, 0)	-0.470 (1.35)	-0.308 (1.41)	-0.088 (1.46)	0.627 (1.43)
H (2, 1)	-0.252 (1.50)	-0.064 (1.42)	-0.184 (1.39)	-0.269 (1.46)
H (4, 2)	-0.260 (1.72)	-0.168 (1.55)	-0.081 (1.52)	-0.093 (1.61)
PL (2, 0)	-0.503 (1.17)	-0.394 (1.21)	-0.204 (1.27)	0.548 (1.29)
PL (4, 0)	-0.463 (1.67)	-0.283 (1.62)	-0.084 (1.60)	0.597 (1.51)
PL (2, 1)	0.341 (1.30)	-0.135 (1.31)	-0.168 (1.30)	-0.262 (1.46)
PL (4, 2)	-0.226 (1.79)	-0.154 (1.75)	-0.080 (1.70)	-0.097 (1.77)

t -ratios: both the bias and the standard deviations are lower in the PL models than in the corresponding H models. This illustrates the consequences of over-fitting that occurs in the H models when many lags and leads of covariates are included: the fit obtained is often "too good" and the estimated residual variance therefore too low, leading to t -ratios with excessive dispersion.

- (iii) The case of $\theta_{21} = 0$ shows that although FM is less biased than OLS, it is outperformed by SEECM's. On the other hand, FM's standard deviation is less than that of the SEECM's, leaving it with a superior mean square error.
- (iv) When $\sigma_{21} > 0$ neither the linear nor the nonlinear SEECM's have a clear advantage over the other. As was the case for the FM bias term, when $\theta_{21} < 0$, the FM t -ratio suffers from bias, increased variance, and skewness. Adding lags and leads of the covariates strongly reduces the bias of the SEECM estimators.

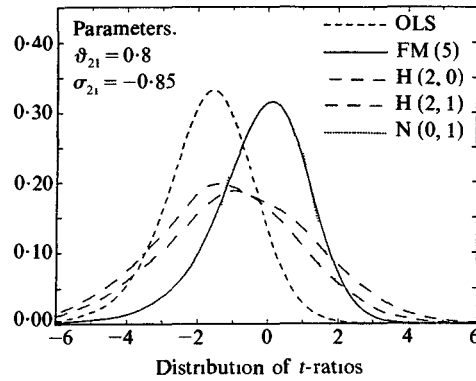


FIGURE 1

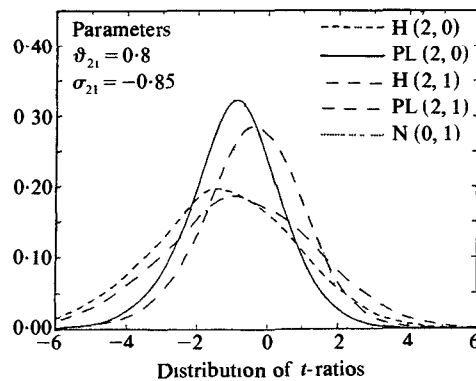


FIGURE 2

- (v) We present three pairs of figures in which we graph the densities of the t -ratios for selected estimators and combinations of θ_{21} and σ_{21} . The standard normal density is included in all figures as the “benchmark” asymptotic distribution. The first two figures are based on $(\theta_{21} = 0.8, \sigma_{21} = -0.85)$, a case of strong positive feedback from y_{1t} to y_{2t} , and show how biased and skewed the distribution of the OLS estimator is in comparison to the other estimators. The density of the FM t -ratio is the least biased and skewed, and comes reasonably close to its limiting distribution, given the moderate size of the sample ($T = 50$). In comparison to fully modified OLS, the linear SEECM’s perform less well (Figure 1), even when a lead term is included to capture the feedback effect. Figure 2 compares the linear and nonlinear SEECM’s, and makes clear the substantial improvement achieved by the nonlinear specification of the model: both PL (2, 0) and PL (2, 1) have densities which are much closer to the standard normal density than those of the linear models H (2, 0) and H (2, 1). In Figures 3 and 4 we graph the densities of the same estimators as for Figures 1 and 2, but now for $(\theta_{21} = 0, \sigma_{21} = -0.85)$. The density of the t -ratio of the OLS estimator is again negatively biased, and the FM (5) estimator has good location and size properties. Since $\theta_{21} = 0$ represents the case of no-feedback from the dependent variable to the independent variable, so that valid conditioning on \mathcal{F}_{t-1} applies,

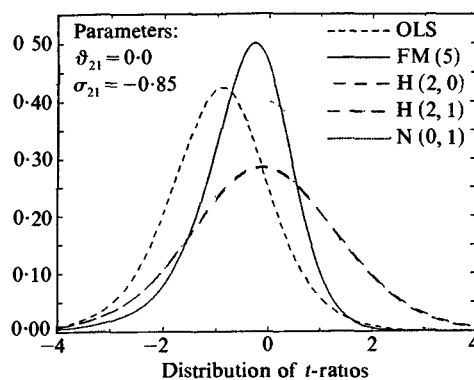


FIGURE 3

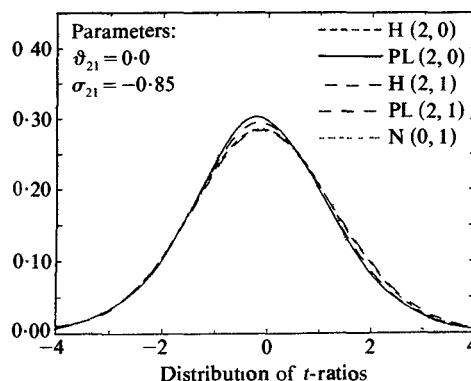


FIGURE 4

we would not expect inclusion of leads of Δy_{2t} to improve the properties of the H model, which is confirmed (Figure 3). The case of no-feedback is also one where the nonlinear (PL) specification is not noticeably superior to the linear (H) model (Figure 4). Finally, densities of the t -ratios in the case of negative feedback are graphed in Figures 5 and 6, for $(\theta_{21} = -0.8, \sigma_{21} = 0.5)$. Figure 5 illustrates the skewness problems the FM estimator has for θ_{21} close to -1 and $\sigma_{21} > 0$. The nonlinear SEECM's perform somewhat better than their linear counterparts (Figure 6), and inclusion of a lead term of Δy_{2t} helps to better centre the distributions about the origin.

We close by remarking that our simulations suggest that one area in which the performance of SEECM's might be improved substantially is in further reducing the dispersion of the t -ratios. We expect that methods which reduce the number of parameters to be estimated and which orthogonalize the remaining regressors in the model would attenuate the overfitting problem without sacrificing much in terms of bias reduction. The process of successive elimination of insignificant regressors and variable orthogonalization is, in practice, as much an art as a science. For our simulations, we tried a few methods that are intended to mimic the variable reduction stage of the Hendry-Richard modelling strategy. For example, we tried retaining only the first few leading principal components of the covariates in the linear SEECM's and employing them as the new

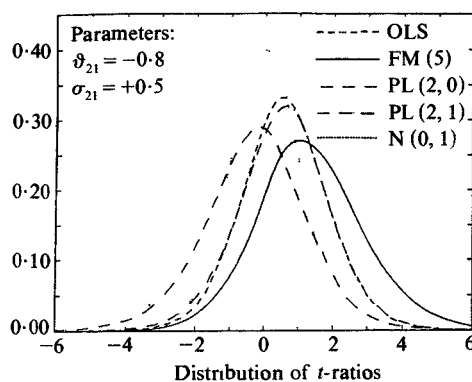


FIGURE 5

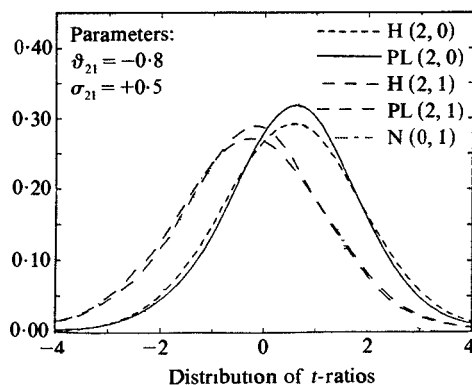


FIGURE 6

regressors. However, this mechanical approach did not produce good results. The reduction of the standard deviations was minimal, whereas the bias problem became immediately apparent. The shift to the nonlinear specification of the SEECM's, on the other hand, brought the desired effect of simultaneous reductions of bias and variance in the t -ratios, at least for negative values of σ_{21} . But the t -ratio statistics of the nonlinear SEECM's are still more dispersed than those of the FM method, suggesting that they can be improved even further.

5. CONCLUSIONS

We have studied various methods of estimating long-run economic equilibria or co-integrated relations. At present, the multitude of available methods for estimation and inference in cointegrated systems is potentially confusing and disconcerting to applied researchers. We have argued that asymptotic theory actually provides some clear guidelines on the most suitable choice of methods. However, there are still a wide range of asymptotically equivalent choices ranging from full systems to single-equation methods (see our summary at the end of Section 3). Our focus in this paper has been on the performance of single-equation methods. We use the method of fully-modified OLS, with its semiparametric corrections, as a benchmark in evaluating the parametric estimation

methodology of Hendry. On the basis of asymptotic theory we propose a nonlinear SEECM specification augmented by leads as well as lags in the differences of the regressors. For such specifications, the use of (parametric) nonlinear least squares on the SEECM is asymptotically equivalent to (semiparametric) fully-modified OLS and to full systems maximum likelihood on the entire system of equations (restricted by the imposition of unit roots).

Our simulations with a small-scale cointegrated system show that FM and SEECM estimators are both substantially better than OLS. The performance of the SEECM estimator is itself substantially improved by: (i) adding more lags, (ii) including leads of differences in the regressor variables when these are not strongly exogenous, and (iii) most importantly, by formulating the SEECM nonlinearly in the parameters through the explicit use of lagged equilibria as regressors. These (nonlinear in parameters) variables serve better to model the information set that is needed for valid conditioning than lagged differences in the dependent variable. With these modifications, parametric SEECM models seem to provide a sound basis for estimation.

However, our simulations show that there is size-distortion in inferences that are based on SEECM estimates. This distortion is due to a tendency in SEECM regressions to overfit and hence bias downwards the error sum of squares in the regression. Of the many SEECM specifications considered here we found that the nonlinear in parameters SEECM leads to the best performance in this respect; yet the size-distortion in this case is still appreciable in simple asymptotic t -tests. We conclude that there is room for further improvement in the methodology to deal with the potential problem of regression overfitting. In practice, this problem may be partially eliminated by careful residual diagnostic checks that include some post-sample predictive tests, as indeed is recommended in the Hendry approach.

It would be interesting to extend our simulations to systems methods of estimation. In particular, it would be of interest to determine whether parametric methods, such as the Johansen VAR approach, suffer from similar difficulties of overfitting and size distortion as the single-equation ECM methods considered in this paper. In this respect it is important to recognize that the asymptotic theory makes it clear that it is *not* necessary to fully estimate (and a fortiori, efficiently estimate) the generating mechanism of u_t in order to efficiently estimate long-run equilibria. All that is needed for the latter is a consistent estimate of the contribution from the short-run dynamics to the long-run variance, *viz.* the value of the long-run covariance matrix of u_t . One advantage of the semiparametric approach is that this feature of the theory of efficient estimation is explicitly recognized and used in the construction of the estimator.

Since short-run dynamics are important in other aspects of modelling, especially prediction, it seems likely that the best operational methods in practice will always involve some parametric elements. One way of proceeding in systems estimation that takes advantage of the strong points of both approaches would be the following: (i) to estimate efficiently the long-run equilibrium relationships by systems methods (or asymptotically equivalent semiparametric single equation methods) and (ii) utilize the estimates from stage (i) in the construction of parsimoniously parameterized ECM's for each of the dependent variables. This approach would lead to efficient estimates of long-run equilibria and the short-run dynamics would be individually modelled on an equation-by-equation basis following the present Hendry methodology. An additional advantage of this proposal is that it would *not* be necessary to include leads in the ECM regressions since the long-run equilibria are already efficiently estimated and incorporated in the regression. Furthermore, the equation-by-equation approach would enable the investigator to employ

many of the judgemental aspects in parsimoniously selecting and orthogonalizing the regressor set that are presently an integral part of this methodology. Such a possibility is not available in complete systems methods like the Johansen (1988) VAR approach. The authors hope to pursue this idea in future research.

Acknowledgement. This paper is based on an inaugural lecture delivered by the first author at the Institute of Advanced Studies, Vienna in May 1989 and on an invited lecture at the Conference on the Econometrics of Financial Markets at the London School of Economics in June 1989. The paper was first circulated in September 1989 and subsequently revised in July 1990. The second author is now at the Department of Economics of the University of Wisconsin at Madison. Our thanks go to the referees for helpful comments, to Gena Ames for wordprocessing and to the NSF for research support under Grants SES 8519595 and SES 8821180.

REFERENCES

- BANERJEE, A., DOLADO, J. J., HENDRY, D. F. and SMITH, G. W. (1986), "Exploring Equilibrium Relationships in Econometrics through Statistical Models: Some Monte Carlo Evidence", *Oxford Bulletin of Economics and Statistics*, **48**, 253-277.
- BERGSTROM, A. R. (1988), "The History of Continuous-Time Econometric Models", *Econometric Theory*, **4**, 365-383.
- BERK, K. N. (1974), "Consistent Autoregressive Spectral Estimates", *Annals of Statistics*, **2**, 489-502.
- BEVERIDGE, S. and NELSON, C. R. (1981), "A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'", *Journal of Monetary Economics*, **7**, 151-174.
- DAVIDSON, J. E., 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.
- DEATON, A. (1973), *Models and Projections of Demand in Post-War Britain* (London: Chapman Hall).
- DEISTLER, M. and ANDERSON, B. D. O. (1989), "Linear Dynamic Errors-in-Variables Models: Some Structure Theory", *Journal of Econometrics*, **41**, 39-63.
- DOAN, T., LITTERMAN, R. B. and SIMS, C. A. (1984), "Forecasting and Conditional Projections Using Realistic Prior Information", *Econometric Reviews*, **3**, 1-100.
- ENGLE, R. F. and GRANGER, C. W. J. (1987), "Cointegration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, **55**, 251-276.
- ENGLE, R. F. and YOO, B. S. (1990), "Cointegrated Economic Time Series: A Survey with New Results" (mimeographed, UC San Diego).
- GRANGER, C. W. J. (1981), "Some Properties of Time Series Data and their Use in Econometric Model Specification", *Journal of Econometrics*, **16**, 121-130.
- GRANGER, C. W. J. and NEWBOLD, P. (1974), "Spurious Regressions in Econometrics", *Journal of Econometrics*, **2**, 111-120.
- HALL, P. and HEYDE, C. C. (1980) *Martingale Limit Theory and Its Application* (New York: Academic Press).
- HANNAN, E. J. and DEISTLER, M. (1988) *The Statistical Theory of Linear Systems* (New York: Wiley).
- HANNAN, E. J. and KAVALLIERIS, L. (1984), "A Method for Autoregressive-Moving Average Estimation", *Biometrika*, **71**, 273-280.
- HANSEN, B. E. and PHILLIPS, P. C. B. (1990), "Estimation and Inference in Models of Cointegration: A Simulation Study", in Fomby, T. B. and Rhodes, G. F., Jr. (Eds.), *Advances in Econometrics*, Vol. 8: *Co-Integration, Spurious Regressions, and Unit Roots* (Greenwich, CT: JAI Press), 225-248.
- HENDRY, D. F. (1976), "The Structure of Simultaneous Equations Estimators", *Journal of Econometrics*, **4**, 51-88.
- HENDRY, D. F. (1987), "Econometric Methodology: A Personal Perspective", in Bewley, T. (Ed.), *Advances in Econometrics*, Vol. 2. (Cambridge: Cambridge University Press).
- HENDRY, D. F. (1989) *PC-GIVE: An Interactive Econometric Modelling System. Version 6.01*.
- HENDRY, D. F. and RICHARD, J. F. (1982), "On the Formulation of Empirical Models in Dynamic Econometrics", *Journal of Econometrics*, **20**, 3-33.
- HENDRY, D. F. and RICHARD, J. F. (1983), "The Econometric Analysis of Economic Time Series (with discussion)", *International Statistical Review*, **51**, 111-163.
- HENDRY, D. F. and UNGERN-STERNBERG, T. von (1981), "Liquidity and Inflation Effects on Consumers' Expenditure", in Deaton, A., (Ed.), *Essays in the Theory and Measurement of Consumer Behaviour in Honour of Sir Richard Stone*. (Cambridge: Cambridge University Press), 237-260.
- JOHANSEN, S. (1988), "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, **12**, 231-254.
- LAHIRI, K. and SCHMIDT, P. (1978), "On the Estimation of Triangular Structural Systems", *Econometrica*, **45**, 1217-1223.

- MARQUARDT, D. W. (1963), "An Algorithm for Least Squares Estimation of Nonlinear Parameters", *Journal of the Society for Industrial and Applied Mathematics*, 11, 431-441.
- MIZON, G. E. and HENDRY, D. F. (1980), "An Empirical Application and Monte Carlo Analysis of Tests of Dynamic Specification", *Review of Economic Studies*, 47, 21-46.
- PARK, J. Y. and PHILLIPS, P. C. B. (1989), "Statistical Inference in Regressions with Integrated Processes: Part 2", *Econometric Theory*, 5, 95-131.
- PHILLIPS, A. W. (1954), "Stabilization Policy in a Closed Economy", *Economic Journal*, 64, 290-323.
- PHILLIPS, A. W. (1956), "Some Notes on the Estimation of Time-Forms of Reaction in Interdependent Dynamic Systems", *Economica*, 23, 99-113.
- PHILLIPS, A. W. (1958), "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom", *Economica*, 25, 283-299.
- PHILLIPS, A. W. (1962), "Employment, Inflation and Growth", *Economica*, 29, 1-16.
- PHILLIPS, P. C. B. (1986), "Understanding Spurious Regressions in Econometrics", *Journal of Econometrics*, 33, 311-340.
- PHILLIPS, P. C. B. (1987), "Time Series Regression with a Unit Root", *Econometrica*, 55, 277-301.
- PHILLIPS, P. C. B. (1988a), "Reflections on Econometric Methodology", *Economic Record*, 64, 344-359.
- PHILLIPS, P. C. B. (1988b), "Spectral Regression for Cointegrated Time Series" (Cowles Foundation Discussion Paper No. 872, to appear in Barnett, W. (Ed.), *Nonparametric and Semiparametric Methods in Economics and Statistics*. (Cambridge: Cambridge University Press, 1991).
- PHILLIPS, P. C. B. (1989), "Partially Identified Econometric Models", *Econometric Theory*, 5, 181-240.
- PHILLIPS, P. C. B. (1991), "Optimal Inference in Cointegrated Systems", *Econometrica*, 59, 283-306.
- PHILLIPS, P. C. B. and HANSEN, B. E. (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes", *Review of Economic Studies*, 57, 99-125.
- PHILLIPS, P. C. B. and PARK, J. Y. (1988), "Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions with Integrated Regressors", *Journal of the American Statistical Association*, 83, 111-115.
- PHILLIPS, P. C. B. and PERRON, P. (1988), "Testing for a Unit Root in Time Series Regression", *Biometrika*, 75, 335-346.
- PHILLIPS, P. C. B. and LORETAN, M. (1989), "Estimating Long-run Economic Equilibria" (Cowles Foundation Discussion Paper No. 928).
- SAID, S. E. and DICKEY, D. A. (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order", *Biometrika*, 71, 599-608.
- SAIKKONEN, P. (1991), "Asymptotically Efficient Estimation of Cointegration Regressions", *Econometrica Theory*, 7, 1-21.
- SARGAN, J. D. (1964), "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", in Hart, P. E., Mills, G., and Whitaker, J. K. (Eds.), *Econometric Analysis for National Planning*. (London: Butterworths). (Reprinted in Hendry, D. F. and Wallis, K. F. (Eds.), *Econometrics and Quantitative Economics*. (Oxford: Basil Blackwell (1984)), 275-314).
- SCHWERT, G. W. (1987), "The Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data", *Journal of Monetary Economics*, 20, 73-103.
- SIMS, C. (1980), "Macroeconomics and Reality", *Econometrica*, 48, 1-48.
- SIMS, C. A., STOCK, J. H. and WATSON, M. W. (1990), "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica*, 58, 113-144.
- SOLO, V. (1989), "A Short Proof of the Engle-Granger Representation Theorem" (mimeo, Johns Hopkins University).
- STOCK, J. H. and WATSON, M. W. (1988), "Testing for Common Trends", *Journal of the American Statistical Association*, 83, 1097-1107.
- STOCK, J. H. and WATSON, M. W. (1991), "A Simple MLE of Cointegrating Vectors in Higher Order Integrated Systems" (mimeo, Dept. of Economics, Northwestern University).
- STONE, J. R. N. (1954), "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand", *Economic Journal*, 64, 511-527.