

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 928

NOTE: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

ESTIMATING LONG RUN ECONOMIC EQUILIBRIA

by

Peter C. B. Phillips and Mico Loretan

October 1989

ESTIMATING LONG RUN ECONOMIC EQUILIBRIA*

by

Peter C. B. Phillips and Mico Loretan

Cowles Foundation for Research
in Economics
Yale University

September 1989

* 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. Our thanks go to Glenna Ames for keyboarding the manuscript and to the NSF for research support under Grants SES 8519595 and SES 8821180.

0. ABSTRACT

Our subject is econometric estimation and inference concerning long-run economic equilibria in models with stochastic trends. Our interest is focussed on single equation specifications such as those employed in the error correction model (ECM) methodology of David Hendry (1987, 1989 *inter alia*) and the semiparametric modified least squares method of Phillips and Hansen (1989). We start by reviewing the prescriptions for empirical time series research that are presently available. We argue that the diversity of choices is confusing to practitioners and obscures the fact that statistical theory is clear about optimal inference procedures. Part of the difficulty arises from the many alternative time series representations of cointegrated systems. We present a detailed analysis of these various representations, the links between them, and the estimator choices to which they lead. An asymptotic theory is provided for a wide menu of econometric estimators and system specifications, accommodating different levels of prior information about the presence of unit roots and the nature of short-run dynamic adjustments. The single equation ECM approach is studied in detail and our results lead to certain recommendations. Weak exogeneity and data coherence are generally insufficient for valid conditioning on the regressors in this approach. Strong exogeneity and data coherency are sufficient to validate conditioning. But the requirement of strong exogeneity rules out most cases of interest because long-run economic equilibrium typically relates interdependent variables for which there is substantial time series feedback. One antidote for this problem in practice is the inclusion of leads as well as lags in the differences of the regressors. The simulations that we report, as well as the asymptotic theory support the use of this procedure in practice. Our results also support the use of dynamic specifications that involve lagged long-run equilibrium relations rather than lagged differences in the dependent variable. Finally, our simulations point to problems of overfitting in single equation ECM's. These appear to have important implications for empirical research in terms of size distortions that are produced in significance tests that utilize nominal critical values delivered by conventional asymptotic theory. In sum, our results indicate that the single equation ECM methodology has good potential for further development and improvement. But in comparison with the semi parametric modified least squares method of Phillips and Hansen (1989) the latter method seems superior for inferential purposes in most cases.

1. INTRODUCTION

One of the many intellectual gifts that the London School of Economics (LSE) has bequeathed to the world of econometrics is error correction modeling. An early stimulus for the work at LSE in this field 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 slowly but steadily over the next two decades. Bergstrom (1989) 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 UK. 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 modeling 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 from 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. It is exciting and unusual to see macroeconomists and econometricians, applied economists and theorists as well as probabilists and statisticians working together in the same field. It used to be said that the sun never sets on the British empire. Now the sun never sets on the empire of cointegration. The concept has even percolated its way through the popular press in Britain—see Gilbert (1989).

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 modeling empirical work itself provides some important guidelines. Finally, since ECM modeling 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 (1988a).

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.
- (iii) Single equation ECM 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 (1988b) in a general time series setting.
- (v) Nonparametric spectral regression methods proposed by Phillips (1988c) 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 (1989). 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 (1988b) 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 modeling 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 in Britain and at LSE has involved single equation ECM methods (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

representation, a latent variable representation and single equation ECM 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. Single equation ECM'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 about here

2. COINTEGRATED SYSTEMS AND TIME SERIES REPRESENTATIONS

2.1. General Issues

Much of the motivation for econometric research derives from the most obvious characteristics of economic data. In the case of time series the typical characteristics are: (i) joint dependence among the series; (ii) serial dependence in individual series; and (iii) nonstationarity. The central core of simultaneous equations theory emerged in large part because of (i) and attendant concerns of economists with problems of general equilibrium. Time series and structural time series methods developed in response to (ii). And modern unit root and integrated process theory has arisen in response to (iii).

In spite of these common characteristics, distinct modeling practices have developed which emphasize certain features of the data, sometimes at the expense of others. Thus, both ECM and VAR approaches treat serial dependence through the use of unrestricted

systems and the methodology of working from general towards restricted models. But in VAR models no structural elements are included, whereas in ECM modeling long-run structural characteristics are retained through the presence of partial equilibrium formulations that furnish particular solutions to the system when the equation errors are set at their expected value of zero. In contrast, simultaneous equations models (SEM's) emphasize structural elements and usually pay less attention to serial dependence. In fact the very presence of serial dependence puts the identifiability of structural relationships at risk. This deficiency was recognized long ago by Orcutt (1952), who likened simultaneous equations methodology to an attempt to learn about the internal workings of a radio by twiddling the knobs on the outside of the set. In Orcutt's words:

The problem thus is somewhat analogous to that of giving a radio to a physicist and asking him to determine the operating characteristic of each component part merely by observing the radio as it plays but without being able in any way to take it apart or rearrange its circuits or perform experiments on it (Orcutt (1952, p. 166)).

Cointegrated systems offer some new perspectives on these different models and approaches because they admit SEM, VAR and ECM representations. In doing so, they help to bring together apparently divergent methodologies within the same theoretical framework. But, most usefully, in cointegrated systems theory the old ideas of simultaneity and prior information gain a new significance. In particular, they have an operational statistical impact which helps to explain what is gained and what is lost in the individual SEM, VAR and ECM representations. This has been explored recently in Phillips (1988b). There it is shown that the use of prior information about nonstationarity (in the form of the presence of unit roots) and about joint dependence (in the form of long-run structural relations) has a major impact on estimation and inference, improving statistical efficiency, removing bias and bringing the asymptotics within the realm of conventional theory so that usual testing procedures apply.

2.2. 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

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

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

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

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

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 superposed 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 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 single equation ECM 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 $u_t = mds(0, \Sigma)$ or even $iid(0, \Sigma)$. In both cases we shall assume that $\Sigma > 0$. This is a useful simplification. Indeed, as shown in Phillips (1988b), 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 $u_t = mds(0, \Sigma)$, 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

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

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 delivered 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

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

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

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

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

2.3. The Triangular System ECM Representation

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

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

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

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

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

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 on (2) and (3).

Triangular system representations like (7) are discussed in detail in Phillips (1988b). 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, as in (2), to think of the equation as explaining the stationary deviations about the equilibrium level that persist period upon period.
3. All of the short run 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.

2.4. 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

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

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

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

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

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

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 non-trivial 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 an assumption that the autoregression is of known finite order.
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.

2.5. 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)):

$$(11) \quad C(L) = C(1) + (L-1)\tilde{C}(L)$$

$$\tilde{C}(L) = \sum_0^\infty \tilde{C}_j L^j, \quad \tilde{C}_j = \sum_{j+1}^\infty C_k.$$

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

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

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

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

where $Y_t = mg(\Omega)$.

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 (1988b). However, as the discussion of the prototypical model in §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 single equation ECM methodology which ensures that least squares is asymptotically equivalent to full systems Gaussian estimation of α .

2.6. Single Equation ECM Methodology

The vector autoregressive ECM representation given in (10) is quite different from the single equation empirical specifications that follows in the Davidson *et al* (1978) tradition for the consumption function. The single equation approach to ECM modeling 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.

1. *Data coherency*: the model fits the data up to an error that is an mds with respect to a given information set.
2. *Valid conditioning*: the model validly conditions on variables that are weakly exogenous with respect to the parameters of interest.
3. *Encompassing*: the selected equation encompasses rival models.
4. *Theory compatible*: the formulation is compatible with economic theory, particularly the long–run steady state properties of the central variables.
5. *Parsimonious, orthogonal decision variables*: the model is parsimonious in its number of regressors and these are selected to be nearly orthogonal variables that are interpretable in some economic sense.
6. *Parameter constancy*: the parameters of the model are constant over the sample period and in *ex post* prediction tests.

The practical steps in finding a single equation model that satisfies these prescriptions inevitably involve many judgmental elements. But the basic steps are as follows:

1. Fit a general unrestricted dynamic specification of the autoregressive, distributed lag type.
2. Test the residuals from this equation for serial correlation, heterogeneity and ARCH effects. The general unrestricted specification should pass the usual residual diagnostic checks.
3. Integrate long–run equilibrium relations into the specification by prudent parameterizations and perform temporal transformations on the regressors so that they are approximately orthogonal and parsimonious in number.
4. Reestimate the transformed and parsimonious equation. This will usually be specified in differences with levels as regressors to accommodate the long–run

equilibrium relationship. Perform diagnostic checks on the residuals and test for parameter constancy within sample and by *ex post* prediction tests.

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

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

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 pretesting 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

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

allowing for the fact that in practice some of these regressors may be eliminated and others may be higher differences. In ECM format this takes the form

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

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

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

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

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

The intended equation (17) would seem to satisfy the Hendry–Richard prescriptions. However, as we shall explore in §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 failure of valid conditioning (#2 in the prescriptions) in (15) and (17), in general, 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.

2.7. The Unrestricted VAR Representation

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

$$y_{1t} = \beta' y_{2t-1} + (1, \beta') u_t$$

$$y_{2t} = A y_{2t-1} + u_{2t}, \quad A = I$$

or

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

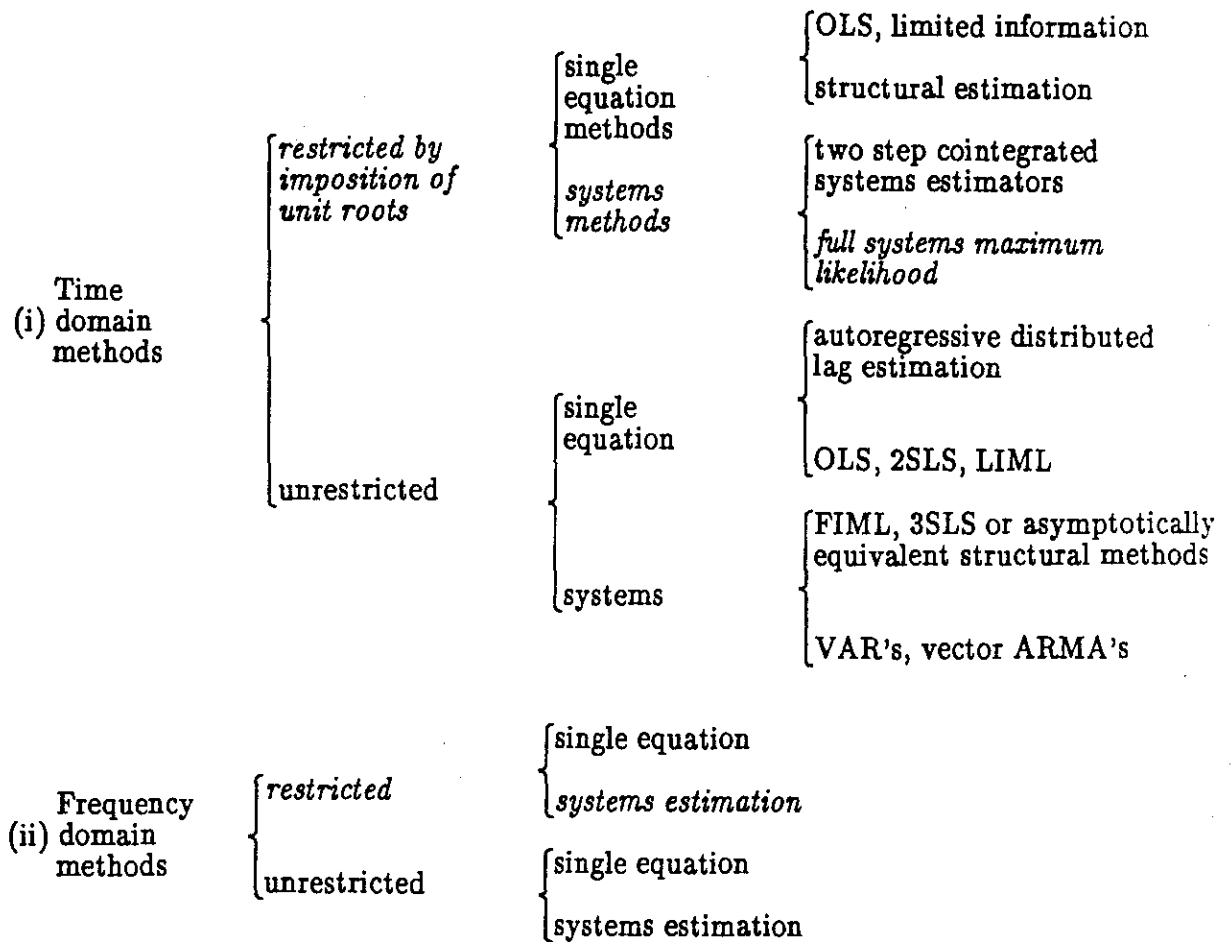
Then, if v_t has the autoregressive form $K(L)v_t = \epsilon_t$ with $\epsilon_t \equiv \text{iid}(0, \Sigma_\epsilon)$ and with the lag operator $K(L)$ possibly of infinite order we deduce the representation

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

The autoregressive polynomial $B(L)$ has $m = n-1$ unit roots, which are implicitly estimated in an unrestricted regression. As discussed in Phillips (1988b) this has nontrivial effects on the asymptotic theory of inference concerning β .

2.8. 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 (1988c). 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 §3.2 below.

Asymptotic theory, at least, provides some guidance concerning optimal choices in (i) and (ii). These are indicated by underlining 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 (1989) can produce optimal single equation techniques. As we shall see below in §3.4, the same is true of the Hendry approach with some appropriate modifications to the implementation procedure described above.

3. ASYMPTOTIC THEORY

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 §2.8(i). It will be helpful in presenting the outcomes to set $S(r) = \text{BM}(\Sigma)$ 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) and Phillips and Hansen (1989).

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

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

Using Lemma 3.1 of Phillips (1989) we write

$$S_1 = \sigma_{21}' \Sigma_{22}^{-1} S_2 + S_{1 \cdot 2}$$

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

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

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

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

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 exogenous.

2. 2SLS, LIML and FIML: Using y_{2t-1} as a 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:

$$(23) \quad \begin{aligned} T(\tilde{\beta} - \beta) &\Rightarrow \left[\int_0^1 S_2 S_2' \right]^{-1} \left(\int_0^1 S_2 dS_1 \right) \\ &= \left[\int_0^1 S_2 S_2' \right]^{-1} \left(\int_0^1 S_2 dS_{1.2} \right) + \left[\int_0^1 S_2 S_2' \right]^{-1} \left(\int_0^1 S_2 dS_2' \right) \Sigma_{22}^{-1} \sigma_{21} . \end{aligned}$$

Compared with (22) this limit variate eliminates one source of asymptotic bias—the third term of (22). 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 $(\Sigma_{22}, \sigma_{21})$ 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 also 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.

3. 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

$$(24) \quad -(T/2)\ln(\sigma_{11.2}) - (1/2 \sigma_{11.2})\Sigma_1^T \left[y_{1t} - \beta' y_{2t} - \sigma_{21}' \Sigma_{22}^{-1} u_{2t} \right]^2.$$

The marginal log likelihood is

$$-(T/2)\ln|\Sigma_{22}| - (1/2)\Sigma_1^T u_{2t}' \Sigma_{22}^{-1} u_{2t}$$

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^\dagger be the estimate of β obtained in this way. Then, in an obvious partitioned regression notation,

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

and the following asymptotics apply:

$$(26) \quad T(\beta^\dagger - \beta) \Rightarrow \left[J_0^1 S_2 S_2' \right]^{-1} (J_0^1 S_2 dS_{1.2}).$$

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

$$(27) \quad \int_{g>0} N(0, g\sigma_{11} \cdot 2\Sigma_{22}^{-1}) dP(g)$$

where $g = e' \left[\int_0^1 S_2 S_2' \right]^{-1} e$, $W_2 \equiv BM(I_m)$ and e 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

As argued in Phillips (1988b), these 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 in Johansen (1988) is included as a subcase. However, these remarks 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 $\left[\int_0^1 S_2 S_2' \right]^{-1} \left(\int_0^1 S_2 dS_2' \right)$. 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.

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 here. Model (2) and (3) is transformed to the frequency domain by taking discrete Fourier transforms (dft's) leading to

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

where, for example, $w_*(\lambda) = (2\omega T)^{-1/2} \Sigma_1^T y_{*t} e^{it\lambda}$ is the dft if $y_{*t} = (y_{1t}, \Delta y_{2t})$. 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:

$$\begin{aligned} \beta^{\#} &= \left[\Sigma_1^M e' \hat{f}_{uu}^{-1}(\omega_j) e \hat{f}_{22}(\omega_j)' \right]^{-1} \left[\Sigma_1^M \hat{f}_{2*}(\omega_j) \hat{f}_{uu}(\omega_j) e \right] \\ \beta_0^{\#} &= \hat{f}_{22}(0)^{-1} \hat{f}_{2*}(0) \hat{f}_{uu}^{-1}(0) e / e' \hat{f}_{uu}(0)^{-1} e. \end{aligned}$$

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

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

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

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 (1989), 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

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

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

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

With these corrections the new estimator β^+ has the same asymptotic behavior 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

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

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

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

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 (1989) and Phillips and Hansen (1989) 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

Suppose we are interested in building a single equation ECM to model the long-run cointegrating relationship (2) and the stationary deviations u_{1t} and the stationary deviations u_{2t} about it. As we have seen in §2.6, the Hendry methodology leads to 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

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

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

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

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 $(\epsilon_{t-1}, \epsilon_{t-2}, \dots)$, where we partition ϵ_t conformably with u_t . Under Gaussian assumptions we deduce that

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

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

$$-(T/2) \ln \sigma_{11.2} - (1/2 \sigma_{11.2}) \Sigma_1^T \left[\epsilon_{1t} - \sigma'_{21} \Sigma_{22}^{-1} \epsilon_{2t} \right]^2.$$

Maximizing this is equivalent to minimizing the sum of squares

$$\Sigma_1^T \left\{ (b_{11}(L), b_{12}(L)) u_t - \sigma'_{21} \Sigma_{22}^{-1} (b_{21}(L), B_{22}(L)) u_t \right\}^2$$

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

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

where

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

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

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

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

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

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

Then (17) is

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

where $f_{1k} = d_{11}$ ($\forall k$), $f_{20} = d_{20}$, $f_{21} = d_{21} - d_{11}\beta$, $f_{2k} = 0$, $k > 1$. 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 the 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 usual ECM formulation that is based on (38). For specificity assume that terms in (38) are truncated at a single lag. The fitted regression is

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

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.

$$(40) \quad B(L) = \begin{bmatrix} b_{11}(L) & 0 \\ 0 & B_{22}(L) \end{bmatrix}, \quad \Sigma = \sigma^2 \mathbf{1}_n.$$

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

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

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 $\eta_t = \epsilon_t = \text{iid } N(0, \sigma^2)$ 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

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

(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

$$L_0 = [\beta, I, I], \quad L_1 = [\Delta_1' \beta, \Delta_0', \Delta_1']$$

$$\Delta_0 = \sum_{k=0}^{\infty} E(u_{20} u_{2k}') = V_{22} (I - D_{22})^{-1}$$

$$\Delta_1 = \sum_{k=-1}^{\infty} E(u_{20} u_{2k}') = D_{22} V_{22} + V_{22} (I - D_{22})^{-1}$$

$$Q = \begin{bmatrix} \beta' V_{22} \beta + 2(1 - d_{11}) v_{11} & \beta' D_{22} V_{22} & \beta' V_{22} \\ V_{22} D_{22}' \beta & V_{22} & D_{22} V_{22} \\ V_{22} \beta & V_{22} D_{22}' & V_{22} \end{bmatrix}$$

$$q' = [(1 - d_{11}) v_{11}, 0, 0], \quad v_{11} = E(u_{1t}^2),$$

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 §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 ECM 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 §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

$$(42) \quad \tilde{E}(u_{1t} | (u_2)_{-\infty}^t) = \tilde{E}(u_{1t} | (u_2)_{-\infty}^t, (u_2)_{t+1}^{\infty}).$$

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:

$$\Sigma_1^T \left\{ y_{1t} - \hat{\beta} y_{2t} - \hat{d}_1(L)(y_{1t} - \hat{\beta} y_{2t}) - \hat{d}_2(L) \Delta y_{2t} \right\} (y_{2t} - \hat{d}_1(L) y_{2t}) = 0$$

which we may rewrite in the form

$$(43) \quad \Sigma_1^T \left\{ [d_1(L) - \hat{d}_1(L)](y_{1t} - \beta' y_{2t}) + (\beta - \hat{\beta})'(1 - \hat{d}_1(L))y_{2t} \right. \\ \left. + [d_2(L) - \hat{d}_2(L)]\Delta y_{2t} + \eta_t \right\} [1 - \hat{d}_1(L)]y_{2t} = 0.$$

Since $\hat{d}_1(L) - d_1(L)$, $\hat{d}_2(L) - d_2(L) = o_p(1)$ and since the sample moments $\{T^{-2} \Sigma_1^T y_{2t-k} y_{2t}'\}$ 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

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

where

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

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

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

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

$$(46) \quad (1 - d_1(1))^{-1} \left[\int_0^1 S_2 S_2' \right]^{-1} \left\{ \int_0^1 S_2 dS_{\eta \cdot 2} + \int_0^1 S_2 dS_2' \Phi_{22}^{-1} \varphi_{21} \right\}.$$

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 $(\Delta y_{2t})_{-\infty}^{\infty}$. The revised specification has the form

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

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

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

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 $\mathcal{M}_{t-1} = \sigma(u_{1t-1}, u_{1t-2}, \dots; (u_{2t})_{-\infty}^{\infty})$.

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)':

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

Now observe that

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

with covariance matrix

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

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 $(u_{1s})_{s=-\infty}^{t-1}$ and $(u_{2s})_{s=-\infty}^{\infty}$. 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 \underline{v}_t where $\underline{v}_t = d_1(L)\underline{v}_t + \eta_t$. But this is equivalent to the long run variance of u_{1t} given $(u_{2s})_{s=-\infty}^{\infty}$, which is $\omega_{11.2}$. Hence

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

so that

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

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 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 ECM modelling, weak exogeneity of the regressors is not sufficient for valid conditioning. With feedback from u_1 to u_2 , *leads* of Δy_{2t} must be included in the ECM 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 ECM modelling there is an asymptotic advantage to the use of lagged equilibrium relationships in the regression and thereby the use of non linear 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 favors the use of NLS on non linear in parameters ECM's rather than simply OLS on linear ECM 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.

4. SIMULATIONS

The simulations reported here are intended to address the small sample performance of the following procedures: OLS, fully modified ("FM") OLS, linear ECM or Hendry ("H") estimators based on (17), and our new estimator ("PL") based on the nonlinear regression equations (35) and (35)'. 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 (1989). The

model is

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

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

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

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 (1989) 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) & (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 the Hendry style regression equations, we employ models with the following combinations of lags and leads of the covariates Δy_{1t}

and $\Delta y_{2t} : (2,0), (4,0), (2,1)$ and $(4,2)$. The same combinations of lags and leads are used for our nonlinear estimator (PL). Thus, 10 estimators were calculated at each replication. (The simulations reported in Phillips and Hansen (1989) 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:

$$(51) \quad y_{1t} = \hat{\alpha} + \hat{\beta}y_{2t} + \hat{d}_{11}(y_{1t-1} - \hat{\alpha} - \hat{\beta}y_{2t-1}) \\ + \hat{d}_{12}(y_{1t-2} - \hat{\alpha} - \hat{\beta}y_{2t-2}) + \hat{d}_{20}\Delta y_{2t} \\ + \hat{d}_{21}\Delta y_{2t-1} + \hat{d}_{22}\Delta y_{2t-2} + \hat{d}_{31}\Delta y_{2t+1} + \hat{\nu}_t.$$

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 easily obtained 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 simulations so we chose a different solution. At each iteration, we compute the Hessian: if it is positive definite we compute the updated parameter vector by the ridge walking method; if not, we calculate the new parameters by iterated OLS, using the parameters obtained in the preceding step as starting values. (Initial starting 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 leads to rapid convergence to an optimum for most replications, in usually 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.

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.

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. Figures 1–16 graph the densities of $\hat{\beta} - \beta$ and $\tau(\hat{\beta})$ for selected parameter configurations. Our main results are:

(1) *Sampling Distribution of the bias term $(\hat{\beta} - \beta)$*

- (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 ECM 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 ECM's provide essentially unbiased estimates of β , and are better centered than the FM distribution. This supports the theoretical result that $\theta_{21} = 0$ is an essential ingredient of valid asymptotic conditioning in conventional ECM 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.
- (v) The nonlinear EMS's do not, in general, reduce the bias of the estimates when compared to the corresponding linear specifications.

 Figures 1-8 about here

- (vi) Figures 1 to 8 depict the densities of some of the estimators for selected values of θ_{21} and σ_{21} . Figures 1 to 4 are based on $(\theta_{21} = 0.8, \sigma_{21} = -0.85)$, and show how biased and skewed the OLS estimator is compared to the other estimators. The density of the FM is the least biased and skewed; adding additional lags and leads in the ECM's reduces the bias and skewness of this method. Figures 5 and 6 show how the densities of various ECM estimators are essentially indistinguishable when $\theta_{21} = 0$. Figure 7 illustrates the skewness problems that the FM

estimator has for $\theta_{21} < 0$ and $\sigma_{21} > 0$. Figure 8 demonstrates the relative performance of the Hendry and PL estimators for these parameter values, which seem to be broadly equivalent in terms of location characteristics.

(2) *The sampling distribution of the t-ratio statistics $\tau(\hat{\beta})$*

- (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 ECM's is strongly evident. The nonlinear estimators improve the performance of the 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 overfitting that occurs in the H models when including many lags and leads of covariates: 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 ECM's. On the other hand, FM's standard deviation is less than that of the ECM's, leaving it with a superior mean square error.
- (iv) When $\sigma_{21} > 0$ neither the linear nor the nonlinear ECM'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 ECM estimators.

 Figures 9–16 about here

- (v) Figures 9 to 16 graph the densities of the t -ratios for selected estimators and combinations of θ_{21} and σ_{21} . When $(\theta_{21} > 0, \sigma_{21} < 0)$ the density of the FM estimator comes remarkably close to its limiting distribution, given the moderate size of the sample. The improvement due to the nonlinear specification is evident in the comparison of $H(2,0)$ vs. $PL(2,0)$ and $H(2,1)$ vs. $PL(2,1)$ in Figure 10.

We close by remarking that our simulations suggest that one area in which the performance of the ECM'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 orthogonalize the remaining regressors in the model would reduce 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 ECM's and employing them as the new 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 ECM'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 ECM'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 cointegrated 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 semi-parametric corrections, as a benchmark in evaluating the parametric estimation methodology of Hendry. On the basis of asymptotic theory we propose a nonlinear ECM 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 ECM 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 ECM estimators are both substantially better than OLS. The performance of the ECM 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 ECM 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 ECM models seem to provide a sound basis for estimation.

However, our simulations show that there is size distortion in inferences that are based on ECM estimates. This distortion is due to a tendency in ECM regressions to overfit and hence bias downwards the error sum of squares in the regression. Of the many ECM specifications considered here we found that the nonlinear in parameters ECM 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 in 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 VAR approach. The authors hope to pursue this idea in future research.

TABLE 1: List of Symbols, Abbreviations and Acronyms

(i) Estimation Methods, Models

OLS	ordinary least squares
GLS	generalized least squares
2SLS	two stage least squares
3SLS	three stage least squares
NLS	nonlinear least squares
MLE	maximum likelihood estimation
LIML	limited information maximum likelihood
FIML	full information maximum likelihood
ECM	error correction model
SEM	simultaneous equations model
AR(p)	autoregressive model of order p
MA(q)	moving average model of order q
VAR	vector autoregression
FM	fully modified OLS
PL	Phillips–Loretan nonlinear ECM

(ii) Symbols from Mathematical Statistics

E	mathematical expectation
Σ	covariance matrix
$f_{uu}(\lambda)$	spectral density matrix at frequency λ
Ω	long run covariance matrix = $2\pi f_{uu}(0)$
iid	independent and identically distributed
wd	weakly dependent
\Rightarrow	converges weakly to
mg	martingale
mds	martingale difference sequence
BM	Brownian motion
\mathcal{F}_t	natural filtration, the σ field generated by the past history of X_t , i.e. $\sigma(X_t, X_{t-1}, \dots)$

(iii) Matrices, Functions of Matrices, Operators

I	identity matrix
B_{ij}	element (i,j) of matrix B
$ A $	determinant of matrix A
$\ A\ $	norm of matrix A, = $\{\text{tr}(A'A)\}^{1/2}$
L	lag operator, $L^k y_t = y_{t-k}$
Δ	first difference operator, $\Delta y_t = (1-L)y_t = y_t - y_{t-1}$

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	-.1466 (.124)	-.0957 (.092)	-.0564 (.060)	-.0204 (.030)
FM(5)	-.0092 (.122)	-.0278 (.077)	-.0252 (.050)	-.0038 (.022)
H(2,0)	-.0573 (.099)	-.0188 (.060)	-.0021 (.036)	.0073 (.017)
H(4,0)	-.0430 (.103)	-.0146 (.065)	-.0016 (.041)	.0069 (.021)
H(2,1)	-.0358 (.100)	-.0095 (.063)	-.0015 (.038)	-.0059 (.017)
H(4,2)	-.0189 (.113)	-.0067 (.075)	-.0017 (.048)	.0065 (.023)
PL(2,0)	-.0789 (.118)	-.0249 (.064)	-.0036 (.035)	.0057 (.015)
PL(4,0)	-.0498 (.121)	-.0144 (.072)	-.0022 (.042)	.0049 (.018)
PL(2,1)	-.0284 (.108)	-.0101 (.066)	-.0030 (.038)	-.0049 (.016)
PL(4,2)	-.0006 (.128)	-.0025 (.083)	-.0028 (.049)	.0035 (.020)
$\sigma_{21} = -0.5$	$\theta_{21} = 0.8$	$\theta_{21} = 0.4$	$\theta_{21} = 0.0$	$\theta_{21} = -0.8$
OLS	-.0782 (.089)	-.0611 (.077)	-.0433 (.061)	-.0144 (.033)
FM(5)	-.0435 (.101)	-.0262 (.082)	.0170 (.062)	.0067 (.033)
H(2,0)	-.0552 (.092)	-.0283 (.075)	-.0060 (.057)	.0180 (.034)
H(4,0)	-.0466 (.098)	-.0227 (.083)	-.0024 (.065)	.0190 (.040)
H(2,1)	-.0165 (.088)	-.0069 (.078)	-.0056 (.061)	-.0086 (.032)
H(4,2)	-.0100 (.104)	-.0086 (.094)	-.0033 (.077)	.0073 (.041)
PL(2,0)	-.0806 (.117)	-.0337 (.082)	-.0066 (.058)	.0111 (.030)
PL(4,0)	-.0657 (.133)	-.0253 (.093)	-.0026 (.067)	.0121 (.036)
PL(2,1)	-.0079 (.096)	-.0013 (.084)	-.0061 (.062)	-.0096 (.031)
PL(4,2)	-.0005 (.123)	-.0044 (.103)	-.0033 (.078)	.0033 (.037)
$\sigma_{21} = +0.5$	$\theta_{21} = 0.8$	$\theta_{21} = 0.4$	$\theta_{21} = 0.0$	$\theta_{21} = -0.8$
OLS	-.0219 (.036)	-.0173 (.042)	-.0071 (.048)	.0296 (.058)
FM(5)	-.0213 (.044)	-.0105 (.048)	.0063 (.056)	.0697 (.094)
H(2,0)	-.0193 (.041)	-.0164 (.048)	-.0090 (.057)	.0270 (.068)
H(4,0)	-.0155 (.046)	-.0116 (.055)	-.0042 (.064)	.0323 (.079)
H(2,1)	.0070 (.040)	.0025 (.049)	-.0083 (.061)	-.0125 (.068)
H(4,2)	-.0069 (.048)	-.0061 (.062)	-.0043 (.075)	-.0048 (.082)
PL(2,0)	-.0223 (.048)	-.0185 (.053)	-.0098 (.059)	.0320 (.070)
PL(4,0)	-.0163 (.048)	-.0122 (.057)	-.0042 (.067)	.0375 (.086)
PL(2,1)	.0145 (.048)	.0066 (.054)	-.0085 (.063)	-.0115 (.067)
PL(4,2)	-.0087 (.054)	-.0066 (.066)	-.0042 (.080)	-.0032 (.083)

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)

Figure 1

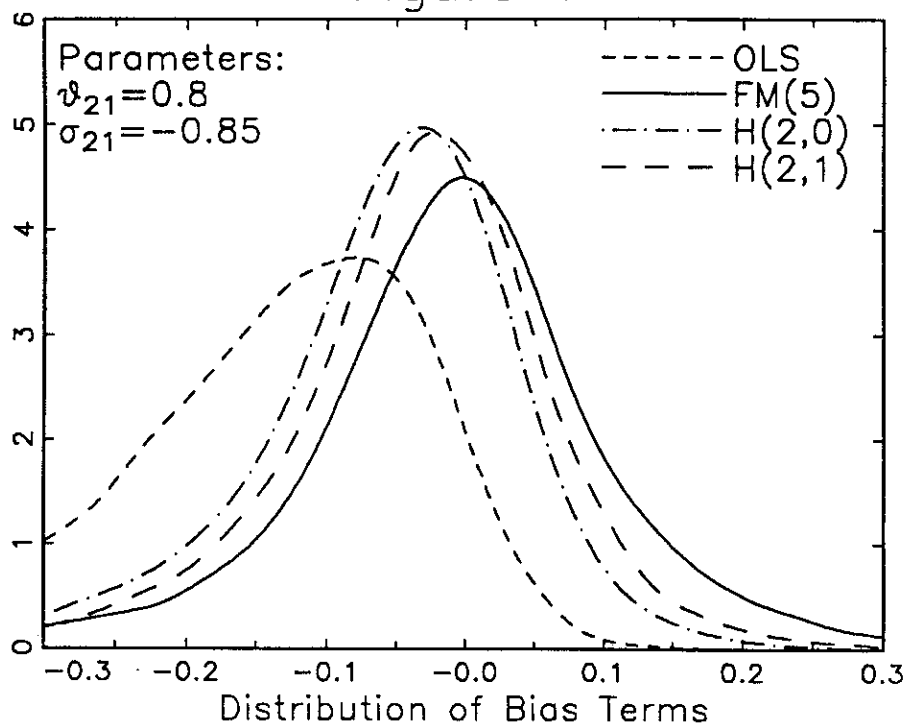


Figure 2

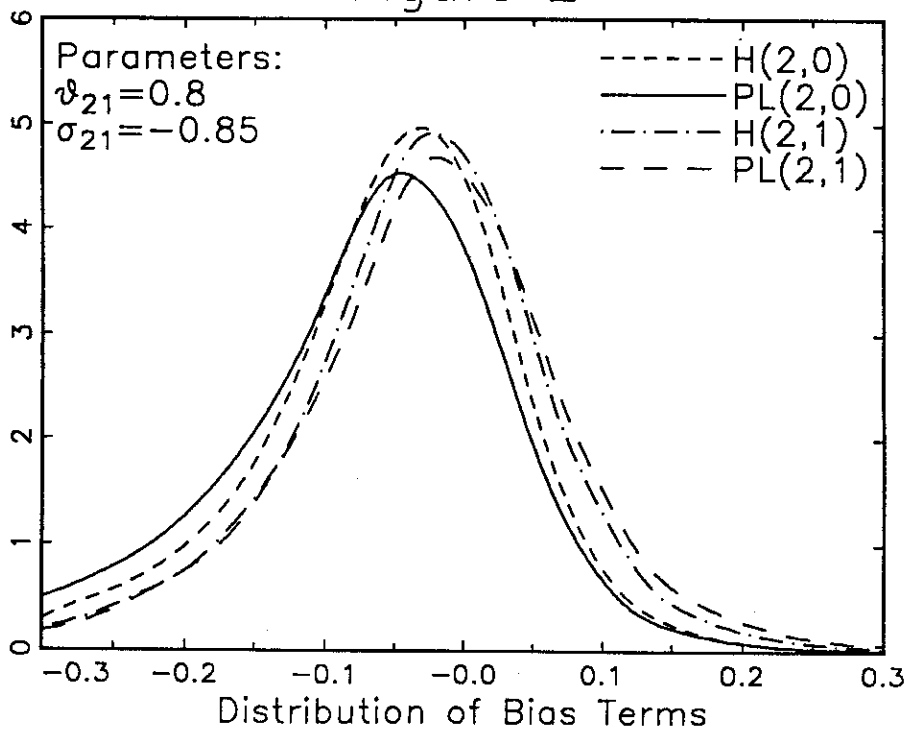


Figure 3

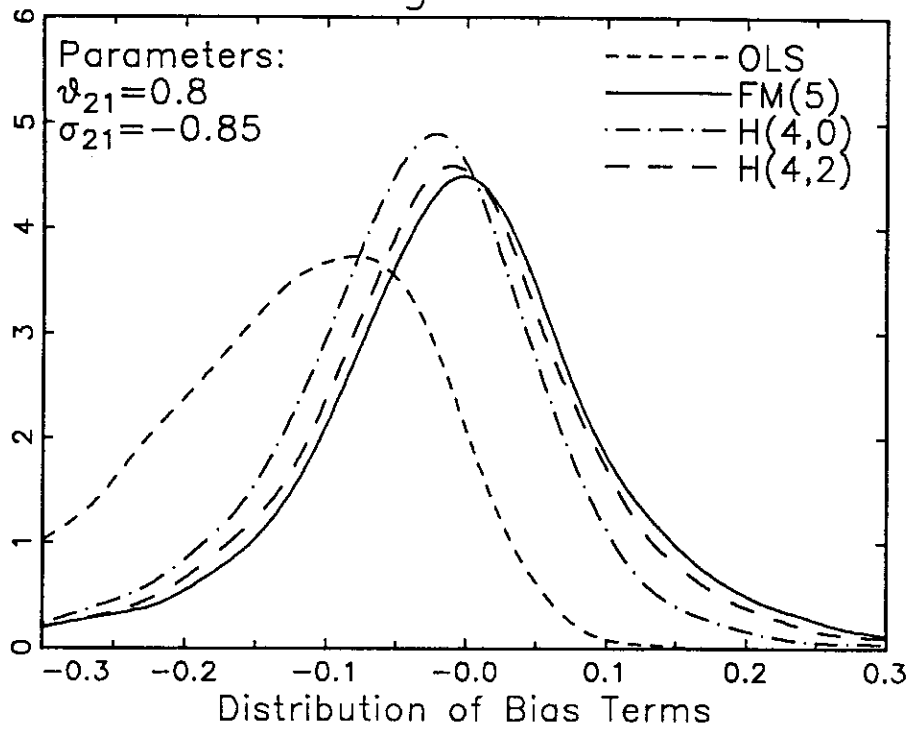


Figure 4

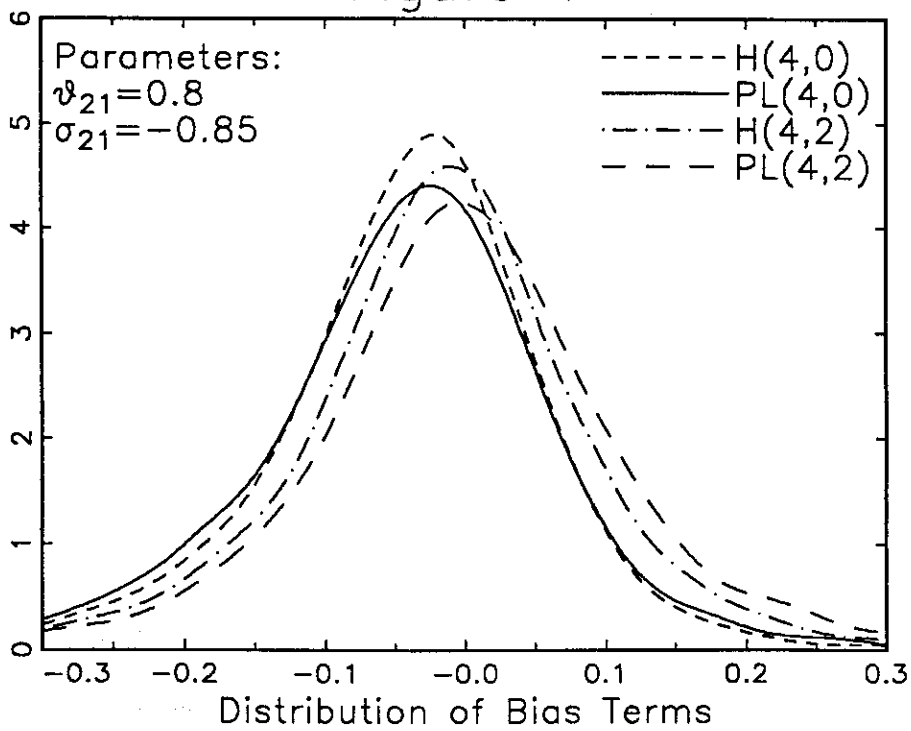


Figure 5

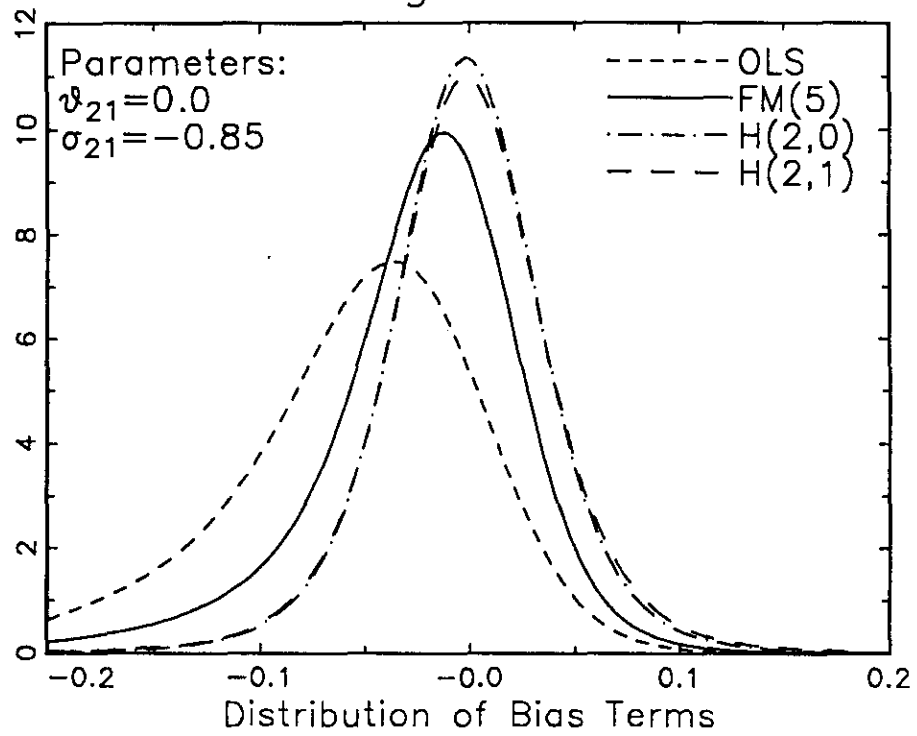


Figure 6

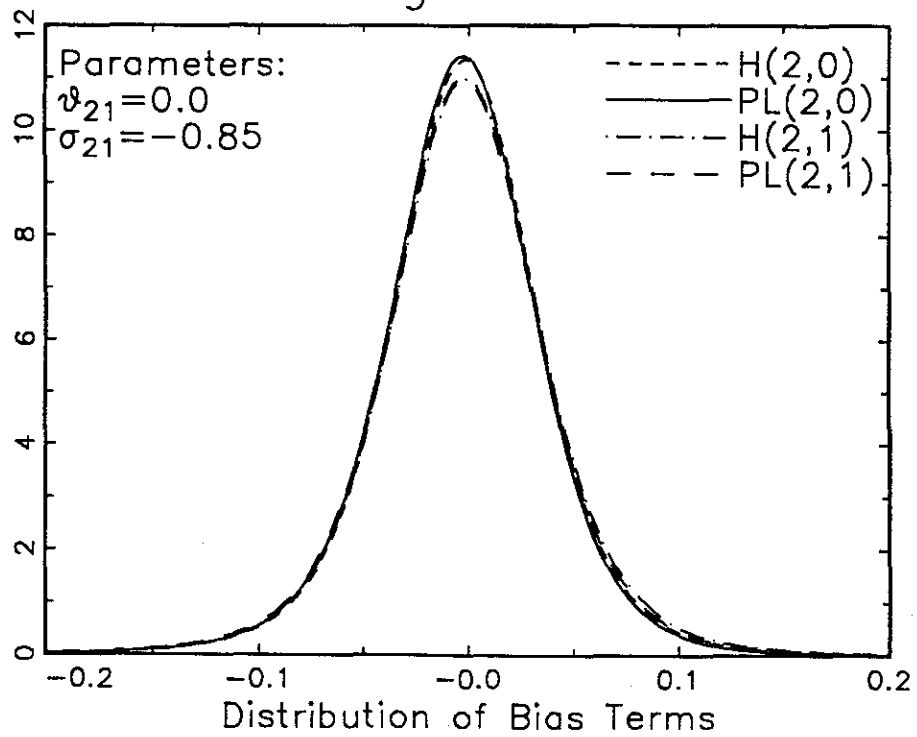


Figure 7

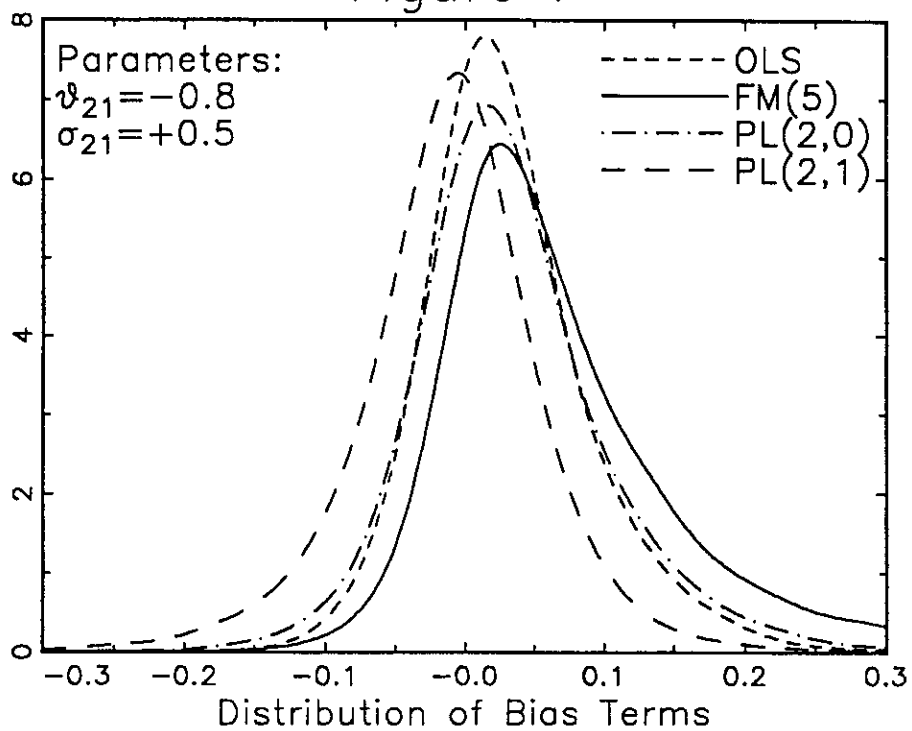


Figure 8

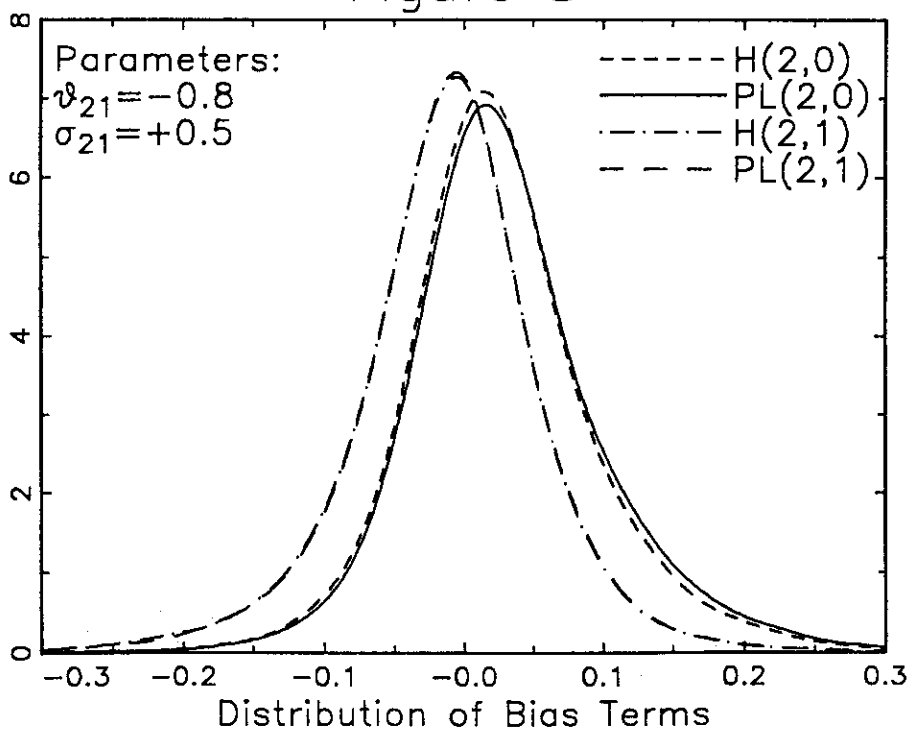


Figure 9

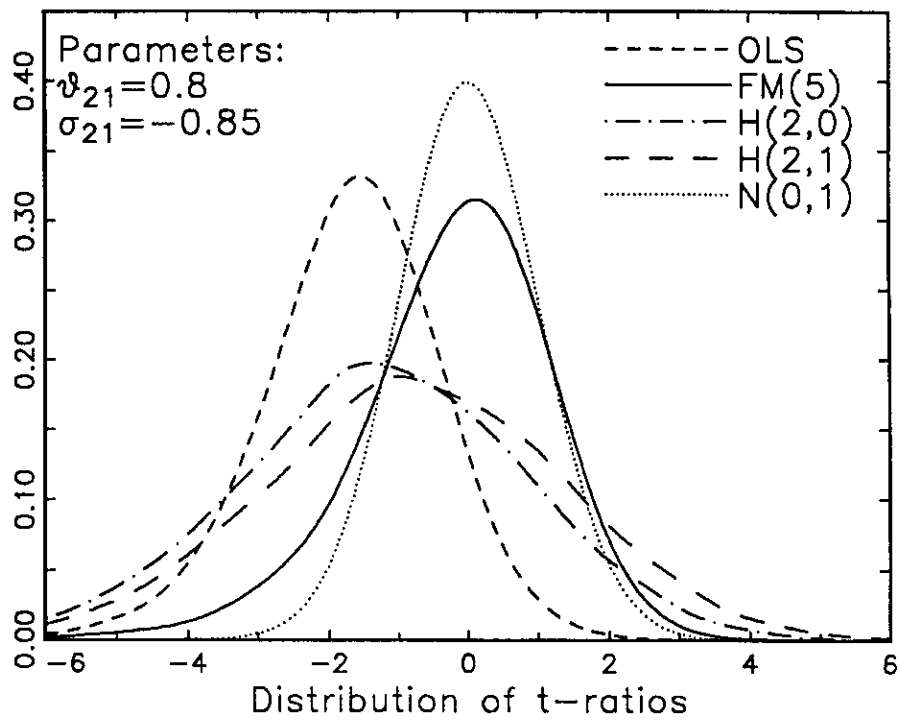


Figure 10

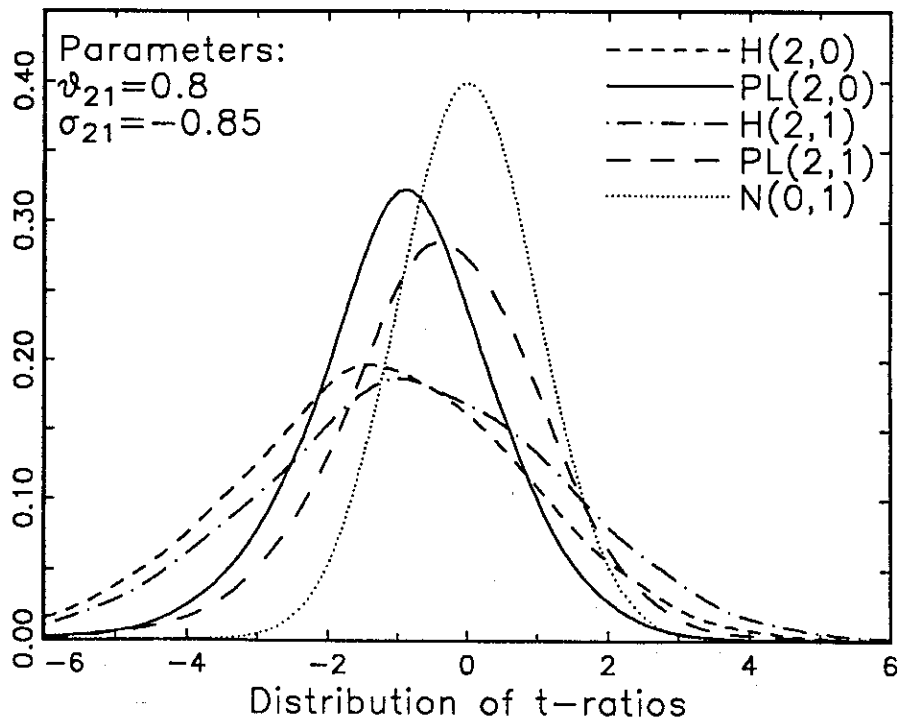


Figure 11

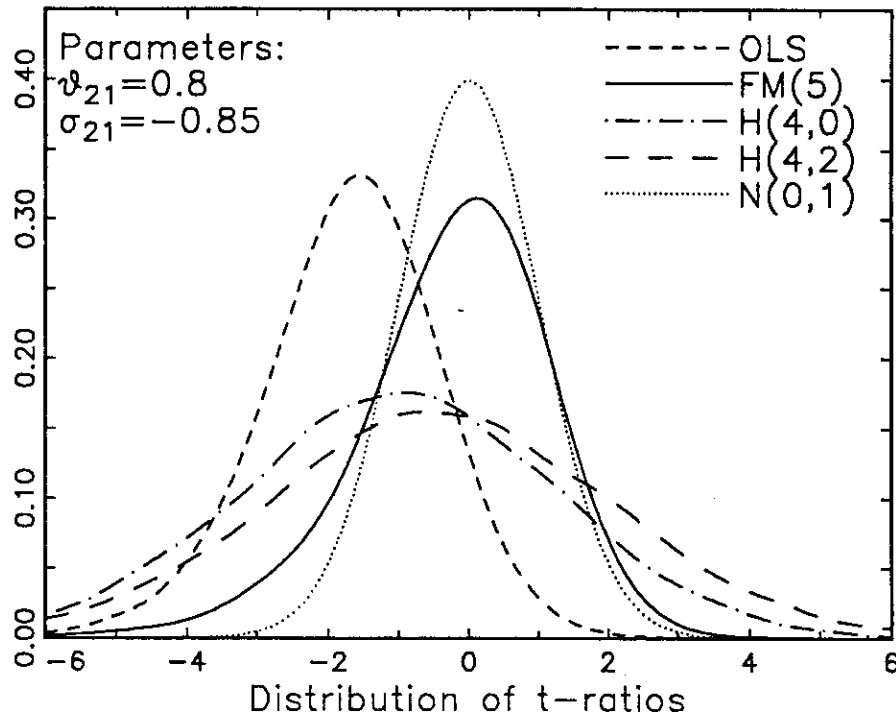


Figure 12

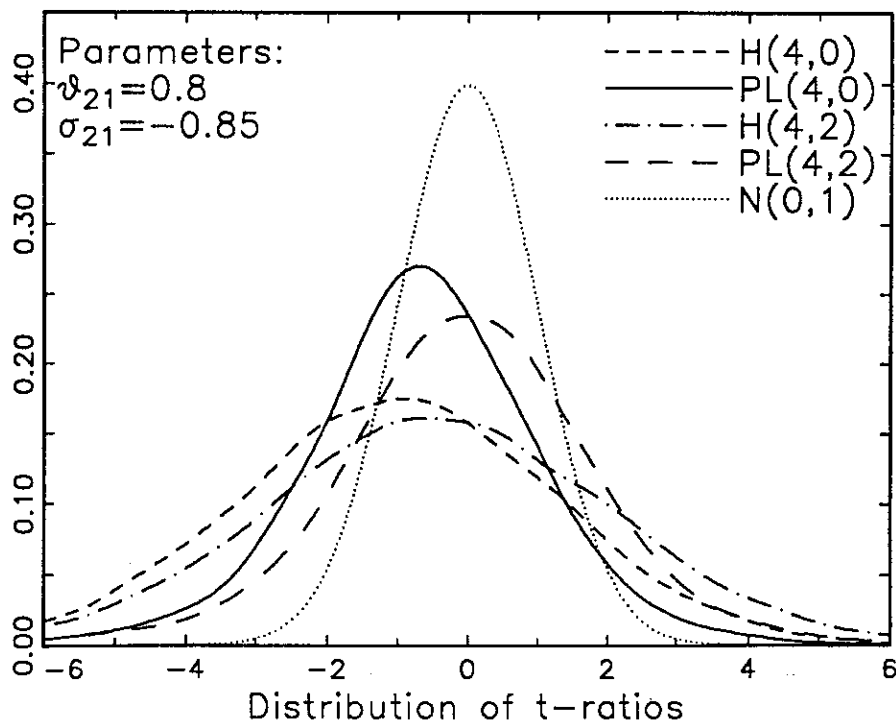


Figure 13

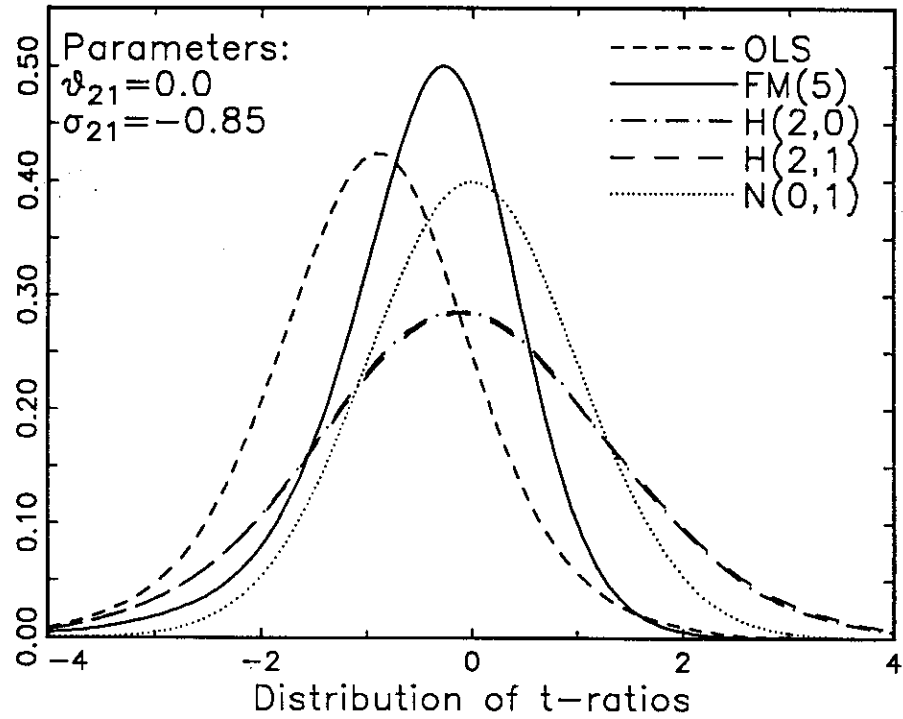


Figure 14

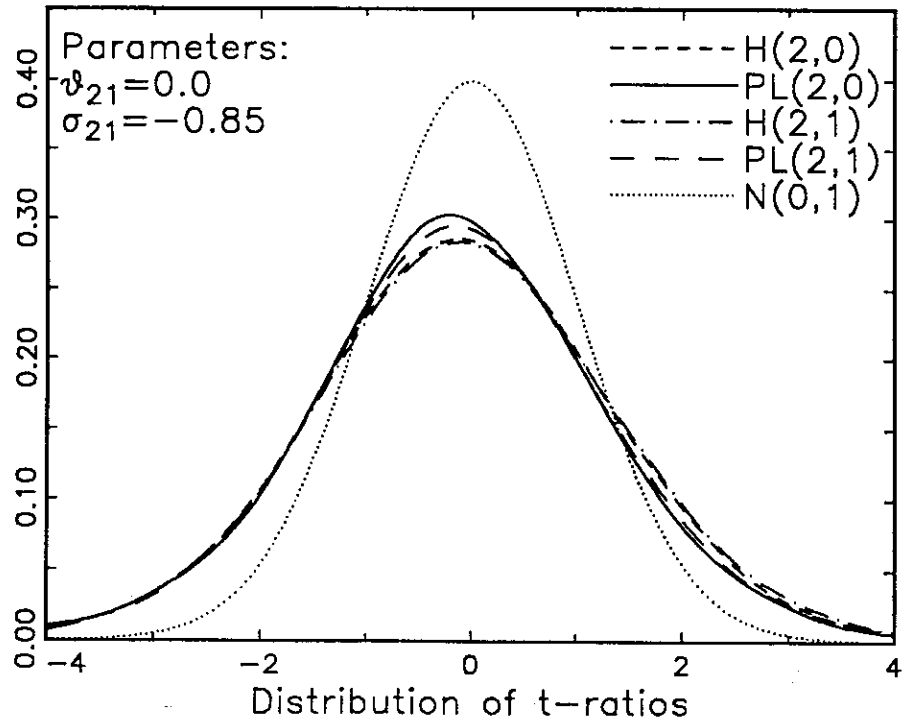


Figure 15

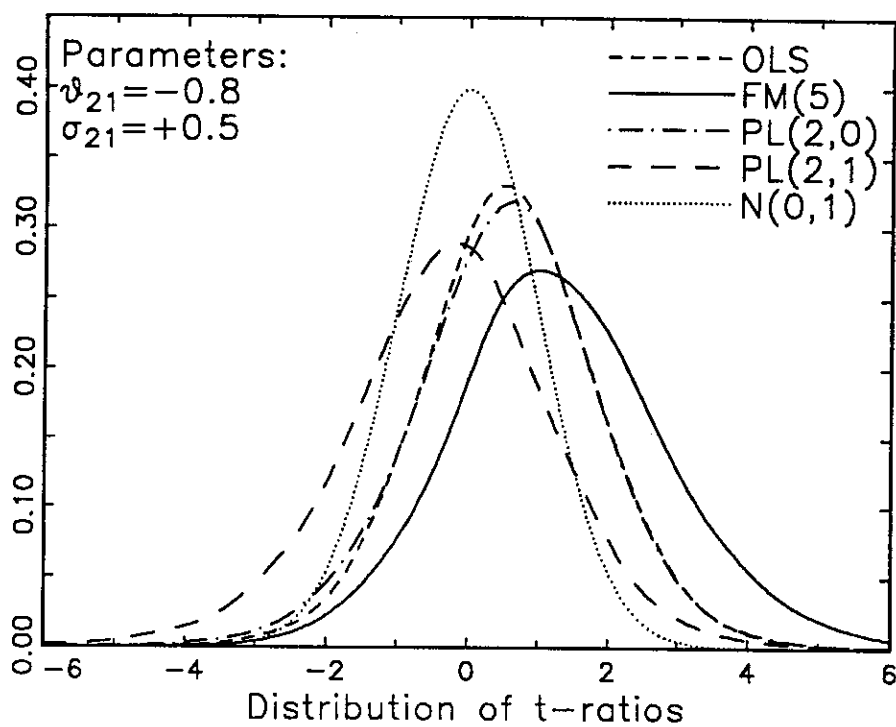
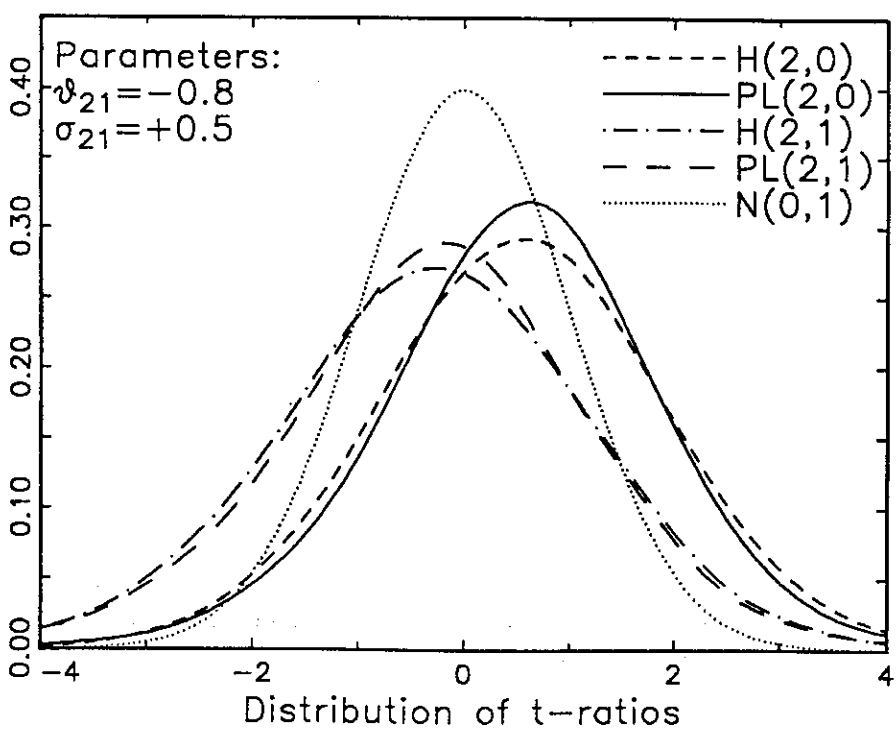


Figure 16



REFERENCES

- Banerjee, A., J. J. Dolado, D. F. Hendry and G. W. Smith (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 C. R. Nelson (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., D. F. Hendry, F. Srba and S. Yeo (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 B. D. O. Anderson (1989), "Linear Dynamic Errors-in-Variables Models: Some Structure Theory," *Journal of Econometrics* 41, 39–63.
- Doan, T., R. B. Litterman and C. Sims (1984), "Forecasting and Conditional Projections Using Realistic Prior Information," *Econometric Reviews* 3, 1–100.
- Engle, R. F. and C. W. J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation, and Testing," *Econometrica* 55, 251–276.
- Gilbert, C. (1989), "Cointegration May Be More Useful Than Chaos," Letter to the Editor of the Financial Times, May 23.
- 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 P. Newbold (1974), "Spurious Regressions in Econometrics," *Journal of Econometrics* 2, 111–120.
- Hall, P. and C. C. Heyde (1980), *Martingale Limit Theory and Its Application*. New York: Academic Press.
- Hannan, E. J. and M. Deistler (1988), *The Statistical Theory of Linear Systems*. New York: Wiley.
- Hannan, E. J. and L. Kavalieris (1984), "A Method for Autoregressive-Moving Average Estimation," *Biometrika* 71, 273–280.
- Hansen, B. E. and P. C. B. Phillips (1989), "Estimation and Inference in Models of Cointegration: A Simulation Study," in *Advances in Econometrics* (to appear).

- 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: T. Bewley (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 J. F. Richard (1982), "On the Formulation of Empirical Models in Dynamic Econometrics," *Journal of Econometrics* 20, 3–33.
- Hendry, D. F. and J. F. Richard (1983), "The Econometric Analysis of Economic Time Series (with discussion)," *International Statistical Review* 51, 111–163.
- Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control* 12, 231–254.
- Lahiri, K. and P. Schmidt (1978), "On the Estimation of Triangular Structural Systems," *Econometrica* 45, 1217–1223.
- Orcutt, G. H. (1952), "A Review of T. C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*," *American Economic Review* 42, 165–169.
- 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. (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), "Optimal Inference in Cointegrated Systems," Cowles Foundation Discussion Paper No. 866.
- Phillips, P. C. B. (1988c), "Spectral Regression for Cointegrated Time Series," Cowles Foundation Discussion Paper No. 872.
- Phillips, P. C. B. (1989), "Partially Identified Econometric Models," *Econometric Theory* 5, 181–240.
- Phillips, P. C. B. and B. E. Hansen, "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies* (forthcoming).

- Phillips, P. C. B. and J. Y. Park (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 P. Perron (1988), "Testing for a Unit Root in Time Series Regression," *Biometrika* 75, 335–346.
- Said, S. E. and D. A. Dickey (1984), "Testing for Unit Roots in Autoregressive–Moving Average Models of Unknown Order," *Biometrika* 71, 599–608.
- Sims, C. (1980), "Macroeconomics and Reality," *Econometrica* 48, 1–48.
- Sargan, J. D. (1964), "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in: P. E. Hart, G. Mills and J. K. Whitaker (eds.), *Econometric Analysis for National Planning*. London: Butterworths. Reprinted in: D. F. Hendry and K. F. Wallis (eds.), *Econometrics and Quantitative Economics*. Oxford: Basil Blackwell, 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.
- Solo, V. (1989), "A Short Proof of the Engle–Granger Representation Theorem," mimeo, Johns Hopkins 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.