COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 882R

ERROR CORRECTION AND LONG RUN EQUILIBRIA IN CONTINUOUS TIME

by

Peter C. B. Phillips

July 1989

ERROR CORRECTION AND LONG RUN EQUILIBRIA IN CONTINUOUS TIME*

by

P. C. B. Phillips

Cowles Foundation for Research in Economics
Yale University

First draft: June 1988

Revised: July 1989

*My thanks go to Rex Bergstrom and three referees for helpful comments on the first draft of this paper. I also thank Glena Ames for her skill and effort in keyboarding the manuscript of this paper and the NSF for research support under Grant No. SES 8519595 and No. SES 8821180. The paper was first drafted in June 1988 and revised while the author was visiting STICERD at the London School of Economics in June 1989. The hospitality of STICERD and the LSE is gratefully acknowledged.

0. ABSTRACT

This paper deals with error correction models (ECM's) and cointegrated systems that are formulated in continuous time. Problems of representation, identification, estimation and time aggregation are discussed. It is shown that every ECM in continuous time has a discrete time equivalent model in ECM format. Moreover, both models may be written as triangular systems with stationary errors. This formulation simplifies both the continuous and the discrete time ECM representations and it helps to motivate a class of optimal inference procedures. It is further shown that long run equilibria in the continuous system are always identified in the discrete time reduced form, so that there is no aliasing problem for these coefficients. Frequency domain procedures are outlined for estimation and inference. These methods are asymptotically optimal under Gaussian assumptions and they have the advantages of simplicity of computation and generality of specification, thereby avoiding some methodological problems of dynamic specification. In addition, they facilitate the treatment of data irregularities such as mixed stock and flow data and temporally aggregated partial equilibrium formulations Models with restricted cointegrating matrices are also considered.

The properties of Gaussian estimators of the traditional discrete approximation and exact discrete models are studied. It is shown that estimators from both models are consistent but that the specification error in the discrete approximation induces a second order bias effect in the limit distribution. Both procedures suffer inferential difficulties through nuisance parameters and nonstandard limit theory when there is untreated serial dependence.

Key Words: Aliasing; Error correction; Long run equilibria; Spectral regression; Stochastic differential equations; Triangular system; Temporal aggregation.

1. INTRODUCTION

During the 1980's there has been a steady growth of interest in econometric modeling in continuous time. The growing research activity in the field covers a broad range of topics from theoretical work on computational and inferential issues through to major empirical modeling projects. An important development in this renewed activity is the introduction of algorithms which help to relieve the computational burden of working with continuous time systems. The Kalman filter, for instance, simplifies the handling of mixed stock and flow data, missing observations and other data irregularities in the computation of likelihoods (see Harvey and Stock (1985) and Zadrozny (1988)). Progress has also been made on the treatment of higher order systems (Bergstrom (1983, 1984), Hansen and Sargent (1981)) and on the problem of aliasing (Robinson (1976, 1978), Hansen and Sargent (1983)). Readers who are less familiar with this field and its recent developments will find the historical review and commentary by Bergstrom (1988) particularly helpful.

Econometric methods for estimating continuous time models fall into two categories. The first approach is to work from discrete approximations to the underlying continuous system. The discrete approximations may be constructed in either the time domain or the frequency domain. They are then estimated by traditional time domain or Fourier methods. Time domain discrete approximations and estimation are explored in Bergstrom (1966), Wymer (1972) and Sargan (1974). Frequency domain discrete approximations and estimation were developed by Robinson (1976a, 1976b, 1976c, 1977, 1978).

The second approach is to work from the exact discrete model that is induced by the continuous system. This approach was explored in Phillips (1972, 1974) and has been the subject of recent research by Bergstrom (1983, 1985, 1986), Harvey and Stock (1985), Zadrozny (1988) and Agbeyegbe (1986, 1987). Much of the applied work in the field now uses this approach and follows the paradigm laid out by Bergstrom and Wymer (1976) in their model of the UK economy.

For the purpose of statistical inference, all of the articles cited above assume that the variables in the system are either stationary or stationary about deterministic trends. This assumption aids the development of an asymptotic distribution theory along traditional lines. As in discrete time models, however, the assumption of stationarity is an important one and empirical evidence suggests that it is unlikely to be satisfied either with economy wide data or financial data. For example, Bergstrom and Wymer (1976) found evidence of a statistically significant unstable root in their empirical model of the UK.

The purpose of the present paper is to investigate what happens when the assumption of stationarity is relaxed. We focus our attention on estimation and inference rather than computation and on long run equilibria rather than dynamic adjustment mechanisms. Our models are specified as simple first order stochastic differential equations systems but we allow these systems to be driven by general stationary errors. Stochastic trends are introduced by permitting the system to have some unstable latent roots at the origin. Such models may always be written in a triangular system error correction model (ECM) format in continuous time. Moreover, the equivalent discrete time system that is satisfied by equispaced observations of the continuous system can also be written in an analogous ECM format. This has major implications for identification and estimation.

First, since the long run equilibria also appear in the discrete time ECM, the corresponding coefficients are always identified. In effect, there is no aliasing problem in the estimation of long run equilibria. This conclusion strengthens earlier work by the author (1973) on the use of structural prior information in a continuous system to aid in the identification of coefficients.

Second, the triangular system format of the discrete time ECM opens the way to simple estimation methods. It is well known that maximum likelihood reduces to generalized least squares in triangular simultaneous equations models with serially uncorrelated and hence, under Gaussian assumptions, iid errors. However, in the present case the model is driven by stationary not iid errors. To deal with this complication we suggest that generalized least squares in the frequency domain not the time domain be used for estimation. This leads us to the class of spectral regression procedures due to Hannan (1963a), which have been explored in a related context in other recent work by the author (1988e) —hereafter simply (1988e).

An important aspect of these spectral methods is their nonparametric treatment of the regression errors. In effect, the methods allow us to proceed under general assumptions of stationary errors. At this level of generality, it is immaterial whether the model is estimated using instantaneously observed data, flow data or a mixture of the two. In addition, the partial equilibrium formulations that are implicit in the ECM may themselves be specified in terms of time averages. Thus, problems of temporal aggregation which present major impediments to computation and inference in traditional approaches simply do not arise in the present context..

Other recent work on continuous time models with stochastic trends has been done by Harvey and Stock (1988, 1989). The models considered by these authors involve stochastic trends that are represented by Brownian motion processes and stationary components that evolve according to an autoregressive system of stochastic differential equations. They develop algorithms based on the Kalman filter for the construction of the Gaussian likelihood and show how to deal with both stock and flow data. Their concerns are therefore computational and quite complementary to the present paper since they do not address the issues of identification, optimal inference and specification bias of discrete approximations that are dealt with here. In addition, the strategy for estimation and inference that we put forward involves frequency domain not time domain procedures.

The following notation is used in the paper. $D = d/dt$ represents the mean square differential operator with respect to continuous time and Δ the first difference operator in discrete time. We use $\text{vec}(A)$ to stack rows of the matrix A into a column vector and $\bar{A}$, A^* to represent the complex conjugate and complex conjugate transpose of A and $\|A\|$ to signify the matrix norm $(\text{tr}(A'A))^{1/2}$. The symbol "$\Rightarrow$" signifies weak convergence of associated probability measures, the symbol " $\equiv$ " signifies equality in distribution and the inequality " >0 " signifies positive definite when applied to matrices. Stochastic processes such as the standard Brownian motion $W(r)$ on [0,1] are often written as W to achieve notational economy. Similarly, we write integrals with respect to Lebesgue measure such as $\int_0^1 W(s)ds$ more simply as $\int_0^1 W$. Vector Brownian motion with covariance matrix Ω is written " $BM(\Omega)$ ". We use $[x]$ to denote the smallest integer $\leq x$. All limits given in the paper are as the sample size $T \to \infty$ unless otherwise stated.

The symbolism " $I(0)$ " will be taken to mean all covariance stationary processes in continuous time with bounded spectral densities $f(\lambda)$ for which $f(0) \neq 0$. This will be taken to include some generalized random processes such as the derivative of standard Brownian motion i.e. $\zeta(t) = DW(t)$ whose spectrum is the constant function $1/2\pi$ on $(-\infty,\infty)$. Since $DW(t)$ does not exist in the mean square sense an alternative here would be to write $dW(t) = \zeta(dt) = I(0)$, meaning that increments in $W(t)$ are stationary. Note that $D^2W(t)$, as a generalized process, is not $I(0)$ according to this definition because of the bounded spectrum requirement. The continuous time process $y(t)$ is said to be integrated of order one and we write $y(t) = I(1)$ if $Dy(t) = I(0)$. A vector process will be $I(0)$ or $I(1)$ if all of its elements are $I(0)$ or $I(1)$ respectively. However, in the vector case we may have $Dy(t) = u(t) = I(0)$, so that each element $u_i(t)$ has bounded spectrum $f_i(\lambda)$ with $f_i(0) \neq 0$, yet the spectral density matrix $f(\lambda)$ of $u(t)$ may be singular at the origin. In this case the elements of $y(t)$ are cointegrated.

2. REPRESENTATION THEORY AND IDENTIFICATION

Let $y(t)$ be an n–vector $I(1)$ process in continuous time and $u(t)$ be an n–vector stationary time series. The process $u(t)$ will be used to represent a stationary continuous time residual. It is not necessary at this stage to be more explicit about its properties. Indeed, as we shall soon see, it is advantageous to our approach to preserve generality in the specification of the residual. However, for the development of an asymptotic theory of inference it will be necessary to make conditions explicit to ensure the validity of the limit theory. This will be done later in the paper.

We partition the vectors $y(t)$ and $u(t)$ into subvectors as follows:

$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} \begin{matrix} n_1 \\ n_2 \end{matrix}, \quad u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} \begin{matrix} n_1 \\ n_2 \end{matrix} \tag{1}$$

We shall assume that the generating mechanism for $y(t)$ is the cointegrated system

$$y_1(t) = By_2(t) + u_1(t) \tag{2}$$

$$Dy_2(t) = u_2(t) . \tag{3}$$

As in discrete time formulations, the idea here is that (2) embodies a long run equilibrium relationship between the variables. This relationship is sufficiently strong that it is perturbed only by stationary deviations which are represented by $u_1(t)$. In particular, all high frequency perturbations from equilibrium are absorbed by the process $u_1(t)$. This will often be realistic in practice as we now discuss.

Note that the variables in (2) are formulated as instantaneous functions of time not as time averages. Thus, if $y_1(t)$ represents aggregate expenditure it measures the rate at which expenditure is taking place at the instant t in time; and correspondingly if $y_2(t)$ represents income it measures the rate at which income is being generated. Naturally, there will be substantial high frequency activity in these processes. Aggregate expenditure and income will fluctuate between peak and off peak times during the course of any fixed period like a day or a week. Much of this high frequency variation is unimportant to the long run relationship between income and expenditure and it will therefore become absorbed in the residual $u_1(t)$.

Other formulations of the long run relationship (2) are possible. Some of these will involve temporal aggregation to allow explicitly for flow data. In other cases stock and flow data will occur together in the same relationship. Alternative versions of (2) such as these will be considered later in Section 4.

Solving (3) with initial conditions at $t = 0$ we have

$$y_2(t) = \int_0^t u_2(s)ds + y_2(0) \tag{4}$$

so that $y_2(t)$ is the outcome of accumulated innovations over the interval $[0,t]$. If $u_2(t) = I(0)$ then $y_2(t)$ is an $I(1)$ process and the system (2) and (3) may be regarded as being driven by the superposition of accumulated innovations over time and stationary deviations from long run equilibrium. According to our usage of the symbolism $I(0)$, $u_2(s)$ may be a generalized process such as continuous time white noise. In this case it is more usual to write the first member on the right side of (4) as a stochastic integral of the form $\int_0^t \zeta(ds)$ where $\zeta(\cdot)$ is a process of orthogonal increments.

Now suppose that both $u_1(t)$ and $Du_1(t)$ are $I(0)$. This would certainly be true if $u_1(t)$ were mean square differentiable. Then, differentiating (2) we have

$$Dy_1(t) = Bu_2(t) + Du_1(t) \tag{5}$$

and

$$Dy_2(t) = u_2(t)$$

which we write collectively as

$$Dy(t) = \underline{u}(t) \tag{6}$$

where $\underline{u}(t) = I(0)$. According to this representation, the elements of $y(t)$ are individually $I(1)$ processes. But (6) is degenerate in the sense that it has only n_2 independent stochastic trends as determined by (4). Note that, as in the discrete time treatment of cointegration, the spectral matrix of $\underline{u}(t)$ is singular at the origin (cf. Phillips (1986)) and the rows of $[I,-B]$ are in its null space.

We also have an ECM representation of (2) and (3). This is obtained by writing (5) in the form

$$Dy_1(t) = -[I, -B]y(t) + u_1(t) + Bu_2(t) + Du_1(t) .$$

Combining this with (3) we obtain

$$(7) \qquad Dy(t) = -EAy(t) + w(t)$$

where

$$E = \begin{bmatrix} I \\ 0 \end{bmatrix} \begin{matrix} n_1 \\ n_2 \end{matrix}, \quad A = [I, -B], \quad w(t) = \begin{bmatrix} u_1(t) + Bu_2(t) + Du_1(t) \\ u_2(t) \end{bmatrix} .$$

The coefficient matrix E in (7) is known. Only the cointegrating matrix A is to be estimated and this matrix is normalized to accord with the normalization of the long–run equilibrium formulation (2). Thus, it is the submatrix B that is the focus of interest in what follows.

The representation (7) is a continuous time analogue of the ECM representation given in Phillips (1988a) for discrete time models. It has the same advantage as the discrete time counterpart that it is in triangular system format. This is made possible by the fact that we have proceeded under general conditions on the residuals $u(t)$. Thus, all short run fluctuations about equilibrium are absorbed into $u_1(t)$ in (2) and $u_2(t)$ measures the shocks that accumulate over time to produce $y_2(t)$. Correspondingly, in (7) the short run dynamics of the adjustment process are absorbed into $w(t)$. In both cases, the residuals u(t) and w(t) are treated as general I(0) processes and in what follows we shall refer to (7) as the *triangular system ECM format.*

There is a converse version of the above relationship. Suppose we start with the stochastic differential equation system

$$Dy(t) = Fy(t) + w(t) \tag{8}$$

where $w(t) = I(0)$ and suppose the coefficient matrix F in (8) is known to have n_1 latent roots in the left half plane and n_2 zero latent roots. We may then write F in the form $F = HG$ where H $(n \times n_1)$ and G $(n_1 \times n)$ both have rank n_1. If necessary, we can reorder variables so that the leading $n_1 \times n_1$ submatrix of G has full rank n_1.

It is easy to see that $Gy(t) = I(0)$ since this process is determined by the system

$$D(Gy) = GH(Gy) + Gw$$

and the latent roots of GH have negative real parts, being identical with the non zero latent roots of HG .

We now write (8) in the form

$$\begin{aligned} Dy(t) &= EGy(t) + (F-E)Gy(t) + w(t) \\ &= EGy(t) + v(t) \qquad (9) \end{aligned}$$

where $v(t) = I(0)$. This puts (8) in the same error correction format as (7). Indeed, upon renormalization of G in the decomposition $F = HG$ the system (10) is precisely in triangular system format.

The following proposition summarizes the above results.

PROPOSITION 1

(a) *Every continuous time cointegrated system such as* (2) *and* (3) *has a triangular system ECM format of the type given in* (7). *The converse is also true.*

(b) *Every stochastic differential equation system of the form* (8) *whose coefficient matrix* F *has latent roots in the left half plane and some latent roots at the origin has a triangular system ECM representation.*

We shall be concerned with the situation where continuous time recording is not possible and only discrete time observations are available for the purpose of estimation. Instantaneous data recording at equispaced points in time leads to the time series $\{y(n)\}_{n=1}^{T}$. It is well known (e.g. Bergstrom (1983)) that discrete time series of this type that are generated by (7) also satisfy the model

$$y(n) = \exp\{-EA\}y(n-1) + \epsilon(n)\ , \quad \epsilon(n) = \int_0^1 \exp\{-sEA\}w(n-s)ds\ . \tag{10}$$

This system is known as the *exact discrete model* corresponding to (7) and the discrete time series $\{y(n)\}_1^N$ is known to satisfy (10) almost surely. Since $w(t) = I(0)$, $\epsilon(n)$ is stationary and when $w(t)$ is continuous time white noise $\epsilon(n)$ is an orthogonal sequence.

The exact discrete model (10) has an equivalent formulation in terms of a discrete time ECM. Indeed, from the series representation

$$\exp(-EA) = I - EA + \frac{1}{2!}(EA)^2 - \frac{1}{3!}(EA)^3 + \cdots$$

and the fact that $AE = I$ we find that $\exp(-EA) = I - fEA$ where $f = (e-1)/e$. It follows that (10) may be rewritten as

$$\Delta y(n) = -fEAy(n-1) + \epsilon(n)$$

which is in the form of a discrete time ECM. This may be further simplified by writing

$$\Delta y(n) = -EAy(n-1) + (1/e)EAy(n-1) + \epsilon(n)$$

or

(11) $$\Delta y(n) = -EAy(n-1) + x(n) \text{ , say}$$

where $x(n) = I(0)$ since both $\epsilon(n)$ and $Ay(n)$ are stationary. The model (11) is now in triangular system ECM format for discrete time models. Such models have been studied in earlier work by the author (1988a, 1988e) and results in those papers will be drawn on below.

We now have:

PROPOSITION 2

Every continuous time ECM system of the form (7) *generates an exact discrete model that may equivalently be written in the discrete time triangular system ECM format* (11). *This discrete time ECM may itself be written in cointegrated system format as*

(12) $$y_1(n) = By_2(n) + \nu_1(n)$$

(13) $$\Delta y_2(n) = \nu_2(n)$$

in which the cointegrating matrix $A = [I, -B]$ *is identical to that of the continuous time analogue* (2). *The error process* $\nu(n) = (\nu_1(n)', \nu_2'(n))'$ *in* (12) *and* (13) *is stationary and is determined from the continuous time residual process* $u(t)$ *by the action of the linear filter*

(14) $$\nu_1(n) = u_1(n) \text{ , } \nu_2(n) = \int_{n-1}^{n} u_2(s)ds \text{ .}$$

REMARK (a) Continuous time error correction models such as (7) generate equispaced data which satisfy analogous discrete time models. Moreover, the long run equilibrium coefficients in the two models are the same. This leads to the conclusion that the long run parameters of a continuous time model may be estimated directly from discrete data by formulating and estimating the corresponding discrete time ECM. This shows that at least for these parameters there is no aliasing or identification problem.

Note that the cointegrated system (12) and (13) is simply obtained from (2) and (3) by integration over a unit time interval. The problem of identifying the coefficient matrix B in the cointegrating equation (2) using only discrete data then reduces to the problem of identifying B in (12). Since (12) is a structural relation, identification would normally require further conditions such as the exogeneity of y_2. However, y_2 is an I(1) process and, although it is in general contemporaneously and serially dependent on ν_1 in (12), the signal that it imparts is sufficiently strong relative to the stationary error ν_1 that B is identifiable and may be consistently estimated by a variety of methods. In discrete time these are issues that have been fully explored elsewhere—see Phillips and Durlauf (1986) and Stock (1987) for details. Similar remarks apply to quite general stochastic differential equation systems such as (8) with some zero roots. Note that (8) has the ECM formulation (10) and in (10) the cointegrating matrix G is identified up to normalization. This means that long run equilibrium relations are always identified in continuous time models irrespective of the short run dynamic behavior of the system and provided only that the residuals in the system are I(0). It seems likely that this conclusion can be strengthened to allow for some degree of long memory in the residuals but this will not be explored here.

It is worth observing how these results on identification bear on the usual aliasing problem. Let $f^c_{ww}(\lambda)$ be the spectral density matrix of the residual process w(t) in (7). Suppose $f^c_{ww}(\lambda)$ is continuous and bounded over the interval $(-\infty,\infty)$. The spectral matrix of y(t) is

$$f^c_{yy}(\lambda) = (-i\lambda I - EA)^{-1} f^c_{ww}(\lambda)(i\lambda I - A'E')^{-1} . \tag{15}$$

and the spectrum of the discrete sequence $\{y(n)\}_1^\infty$ is given by the folding formula $f^d_{yy}(\lambda) = \Sigma_{j=-\infty}^{\infty} f^c_{yy}(\lambda+2\pi j)$. (Note that spectra such as (15), which represent non stationary I(1) processes and have a singularity at the origin, may be defined as the pointwise limit of the expectation of the periodogram—see Solo (1987)). Now $f^c_{yy}(\lambda+2\pi j)$ is bounded for all $j \neq 0$ in the vicinity of $\lambda = 0$ whereas when $j = 0$ we have $f^c_{yy}(\lambda) = O(1/\lambda^2)$ as $\lambda \to 0$. Thus, the behavior of the discrete spectrum $f^d_{yy}(\lambda)$ at the origin is prescribed by that of $f^c_{yy}(\lambda)$ as $\lambda \to 0$. This means that we can identify the long run components that dominate the behavior of $f^c_{yy}(\lambda)$ from the discrete spectrum $f^d_{yy}(\lambda)$. Next observe that $A(i\lambda I + EA)^{-1} = (1+i\lambda)^{-1}A$ and therefore the spectrum of $Ay(t)$ is the continuous and bounded function $(1 + \lambda^2)^{-1} A f^c_{ww}(\lambda)A'$. Clearly, A is identified from $f^c_{yy}(\lambda)$ as the linear transformation of $y(t)$ that annihilates the pole at the origin in $f^c_{yy}(\lambda)$. The matrix A is then unique up to normalization. However, since the behavior of $f^d_{yy}(\lambda)$ mirrors the behavior of $f^c_{yy}(\lambda)$ at the origin we may equivalently identify A from $f^d_{yy}(\lambda)$. This eliminates the aliasing problem for the long–run equilibrium parameters in continuous time.

REMARK (b) Some of the inherent difficulties of estimating differential equation systems from discrete data are avoided by working with the discrete time ECM (11). That is, instead of attempting to estimate the differential equation system (7) directly, it is sufficient to concentrate attention on the discrete time triangular system ECM formulation (11). This system may be estimated by a variety of time domain and frequency domain procedures which are discussed in other work by the author (1988a, 1988e). The frequency domain procedures considered in (1988e) have the advantage of generality by virtue of their nonparametric treatment of the regression errors. They also involve only linear estimating equations and thereby avoid some of the complexities of conventional continuous time estimation methods—see e.g. Phillips (1972, 1974) and Bergstrom (1983). The application of these methods in the present context will be briefly described in Section 3.1 below.

REMARK (c) The proposition brings into question the traditional interpretation of the exact discrete model of a continuous time system as a reduced form. This has been emphasized in some earlier work (e.g. Bergstrom (1984), p. 1170). As seen above, the exact discrete model (10) is equivalent to the cointegrated system given by (12) and (13) and (12) is simply a discrete time sampled version of (2). Thus, in contrast to the conventional view of the exact discrete model as a reduced form, the discrete model carries the same structural information as the underlying continuous system. Moreover, this information is carried in the same ECM format in both discrete and continuous time models.

REMARK (d) The proposition also has a converse in the sense that for every discrete time cointegrated system such as (12) and (13) there is an underlying continuous system such as (2) and (3) which gives rise to it. In fact, (12) is obtained by sampling (2) at the integers $\{t = n : n = 0, 1, 2, \ldots\}$ and (13) is obtained by integrating (3) over the unit time interval $[n-1, n]$. The residuals in (12) and (13) are then determined by (14). The required continuous system is obtained by selecting a continuous time residual process $u(t)$ that leads to (12) and (13) upon the action of the linear filter (14).

3. ESTIMATION, INFERENCE AND ASYMPTOTICS

3.1. Unrestricted Cointegrating Matrices

In this subsection we shall work with the linear model (11) where the cointegrating matrix A is unrestricted other than by normalization, i.e. the submatrix B is unrestricted. Our approach is to use the discrete time ECM formulation (11) rather than to attempt to estimate the differential equation system (7) directly. Since (11) has the triangular system ECM format a number of different estimation methods are available including instrumental variables (Phillips and Hansen (1988)), maximum likelihood (Johansen (1988), Phillips (1988a)) and spectral regression (Phillips (1988e)). Of these, spectral regression procedures seem desirable in the present context because of the generality they permit with regard to the regression errors. Generality is important here since the only conditions on the regression errors that have been used in our discussion of representation and identification are stationarity and the existence of a continuous spectral density matrix. In view of their nonparametric treatment of residuals, spectral regression methods allow us to proceed at a comparable level of generality. These methods also have the advantage of computational simplicity since at least when B is unrestricted they avoid the nonlinear optimization problems of other approaches.

Before we detail formulae for our estimators we shall make explicit the conditions that we require on (11). We assume that the residual process $x(n)$ in (11) is stationary with spectral matrix $f^d_{xx}(\lambda) > 0$ that is continuous at the origin $\lambda = 0$. We set $\Omega = 2\pi f^d_{xx}(0)$ and decompose this long run covariance matrix as follows: $\Omega = \Sigma + \Lambda + \Lambda'$ where $\Sigma = E(x(0)x(0)')$, $\Lambda = \Sigma_{k=1}^{\infty} E(x(0)x(k)')$ and we define $\Delta = \Sigma + \Lambda$. We further assume that the partial sum process $P_t = \Sigma_{n=1}^{t} x(n)$ satisfies the invariance principle

$$T^{-1/2} P_{[Tr]} \Rightarrow S(r) \equiv BM(\Omega) \tag{16}$$

and the sample covariance matrix between P_n and $x(n)$ converges weakly as follows:

$$T^{-1} \Sigma_1^{[Tr]} P_n x(n)' \Rightarrow \int_0^r S dS' + r\Delta \tag{17}$$

where the first term on the right side of (17) is a matrix stochastic integral with respect to the Brownian motion $S(r)$. Results (16) and (17) are known to hold under quite mild moment and weak dependence assumptions on the residual process $x(n)$. These conditions and results are discussed in earlier work (see Phillips and Durlauf (1986) and Phillips (1988b, 1988c)) and are reviewed recently in Phillips (1988d). They certainly apply when $x(n)$ is a linear process of the type

$$x(n) = \Sigma_{j=-\infty}^{\infty} C_j \epsilon_{n-j} ; \; \{\epsilon_t\} \equiv iid(0, \Sigma_\epsilon) , \; \Sigma_{-\infty}^{\infty} j^{1/2} \|C_j\| < \infty . \tag{18}$$

This also accommodates discrete sampling of a wide class of continuous processes such as stable ARMA systems in continuous time.

It is convenient for the statement of our results to partition the limit Brownian motion S and the matrices Ω, Σ, Λ and Δ conformably with the partition of y(t) given in (1). For example, we shall write

$$S = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{21}' \\ \Omega_{21} & \Omega_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{21}' \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

and so on. We also define

$$\Omega_{11\cdot 2} = \Omega_{11} - \Omega_{21}'\Omega_{22}^{-1}\Omega_{21} .$$

The estimators we suggest are the Hannan efficient and band spectral estimators (see Hannan (1963a, 1963b)) of the regression coefficient matrix B in (11). These are simply the matrix extensions of the spectral regression estimators developed for the single equation ECM setting in (1988e). It will be helpful, therefore, to use the same notation as that of the (1988e) paper.

Specifically, we define $y_*(n)' = (y_1(n), \Delta y_2(n)')$, $w_*(\lambda) = (2\pi T)^{-1/2}\Sigma_{n=1}^{T} y_*(n)e^{in\lambda}$, $w_2(\lambda) = (2\pi T)^{-1/2}\Sigma_{n=1}^{T} y_2(n-1)e^{in\lambda}$ for $\lambda \in [-\pi,\pi]$. Some natural economies in the computation of the discrete Fourier transforms (dft's) w_* and w_2 can be achieved e.g. by the use of $(e^{-i\lambda} - 1)w_2(\lambda)$ for the dft of $\Delta y_2(n)$. We also need an estimate of the spectral matrix $f_{xx}(\lambda)$ and this may be based on the residuals of an initial least squares regression on the first n_1 equations of (11). We write $\hat{x}(n) = \Delta y(n) + E\hat{A}y(n-1) = y_*(n) - E\hat{B}y_2(n-1)$ and may then compute the smoothed periodogram estimate

$$\hat{f}_{xx}(\omega_j) = \frac{M}{T}\sum_{B_j}[w_*(\lambda_s) - E\hat{B}w_2(\lambda_s)][w_*(\lambda_s) - E\hat{B}w_2(\lambda_s)]^*$$

where the summation is over $\lambda_s \in B_j = (\omega_j - \pi/2M < \lambda \le \omega_j + \pi/2M)$, a frequency band of width π/M centered on

$$\omega_j = \frac{\pi j}{M}, \quad j = -M+1, \ldots, M$$

for M integer. Setting $m = [T/M]$ we are effectively averaging m neighboring periodogram ordinates around the frequency ω_j to obtain the estimate $\hat{f}_{xx}(\omega_j)$. We require $M \to \infty$ so that the band shrinks as $T \to \infty$ but in such a way that $M = o(T^{1/2})$. Since the least squares estimator $\hat{B}$ is consistent (see Phillips and Durlauf (1986) and Stock (1987)) we find that when $\omega_j \to \omega$ we have $\hat{f}_{xx}(\omega_j) \xrightarrow[p]{} f_{xx}(\omega)$ as $T \to \infty$. This follows because $f_{xx}(\omega)$ is continuous in view of (18), although we make use of the consistency of $\hat{f}_{xx}(\omega_j)$ only for sequences $\omega_j \to 0$.

The Hannan efficient estimator of B here takes the form:

$$\text{(19)} \qquad \text{vec}(\tilde{B}) = \left[\frac{1}{2M}\sum_{j=-M+1}^{M} E'\hat{f}_{xx}(\omega_j)^{-1}E \otimes \hat{f}_{22}'(\omega_j)\right]^{-1} \cdot \left[\frac{1}{2M}\sum_{j=-M+1}^{M} (E'\hat{f}_{xx}(\omega_j)^{-1} \otimes I)\text{vec}(\hat{f}_{*2}(\omega_j))\right]$$

where $\hat{f}_{22}(\omega_j) = m^{-1}\Sigma_{\mathcal{B}_j} w_2(\lambda_s)w_2(\lambda_s)^*$ and $\hat{f}_{*2}(\omega_j) = m^{-1}\Sigma_{\mathcal{B}_j} w_*(\lambda_s)w_2(\lambda_s)^*$. The band spectral estimator, which is based on spectral estimates at the origin, correspondingly has the form:

$$\text{vec}(\tilde{B}_0) = \left[E'\hat{f}_{xx}(0)^{-1}E \otimes \hat{f}_{22}(0)\right]^{-1}\left[(E'\hat{f}_{xx}(0)^{-1} \otimes I)\text{vec}(\hat{f}_{*2}(0)\right]$$

or more directly

$$\text{(20)} \qquad \tilde{B}_0 = \left[E'\hat{f}_{xx}(0)^{-1}E\right]^{-1} E'\hat{f}_{xx}(0)^{-1}\hat{f}_{*2}(0)f_{22}(0)^{-1}.$$

The computational requirements of the two estimators $\tilde{B}$ and $\tilde{B}_0$ are small, particularly in comparison with direct maximum likelihood methods applied to (11) or the exact discrete model (10). Note also that the latter methods require explicit modeling of the error process and, hence, the short run dynamics of the model with the attendant difficulties of model selection. These problems are bypassed by the use of $\tilde{B}$ or $\tilde{B}_0$ at the cost of dealing with the short run dynamics by indirect nonparametric methods.

Both (19) and (20) rely on an initial estimate of B such at least squares in order to construct the residual spectral estimate $\hat{f}_{xx}$. $\tilde{B}$ and $\tilde{B}_0$ may therefore be regarded as two step estimators. Further iterations are possible and may lead to some improvement in finite sample performance because of the second order bias in the first stage estimates like least squares (Phillips and Durlauf (1986), Stock (1987). Further iterations will not, of course, influence the asymptotics. Finally, we observe that many alternative choices of spectral estimates for $\hat{f}_{xx}$, $\hat{f}_{22}$ and $\hat{f}_{*2}$ other than the smoothed periodogram estimates may be used in the estimation formulae (19) and (20) without affecting asymptotic behavior.

3.2. Restricted Cointegrating Matrices

Now suppose that $B = B(\alpha)$ where α is a p–vector of underlying parameters. Suppose also that $\alpha \in \Phi$, a compact set in R^p, that $A(\alpha)$ is a continuously differentiable matrix function and that the usual identification condition holds, viz

$$A(\alpha) = A(\alpha^0) \text{ implies } \alpha = \alpha^0 \tag{21}$$

where α^0 is the true value of α.

The (nonlinear) efficient spectral regression estimator of α is now

$$\tilde{\alpha} = \operatorname{argmin}_{\alpha} \sum_{B} [w_*(\lambda_s) - EB(\alpha)w_2(\lambda_s)]^* \Phi(\lambda_s))[w_*(\lambda_s) - EB(\alpha)w_2(\lambda_s)]$$

with

$$\Phi(\lambda_s) = \hat{f}_{xx}(\omega_j)^{-1}, \quad \lambda_s = 2\pi s/T \in B_j$$

and $\mathcal{B} = \cup_{-M+1}^{M}\mathcal{B}_j$. The corresponding band spectral estimator is

$$\tilde{\alpha}_0 = \mathrm{argmin}_{\alpha} \sum_{\mathcal{B}_0} [w_*(\lambda_s) - EB(\alpha)w_2(\lambda_s)]^* \hat{f}_{xx}(0)^{-1}[w_*(\lambda_s) - EB(\alpha)w_2(\lambda_s)] .$$

Both these estimators are special cases of the nonlinear spectral regression estimators studied in Robinson (1972). Again all that is new here is that they are being applied in a context where the regressors are nonstationary and coherent with the equation errors.

3.3. Subsystem One Step Estimation

The above estimators rely on a preliminary regression in order to construct the weighting matrix $\hat{f}_{xx}(\omega_j)^{-1}$. This can be avoided at least for the band spectral estimator when there are no restrictions on B by working with the equation

$$(22) \qquad w_1(\lambda_s) = Bw_2(\lambda_s) + Cw_{\Delta y_2}(\lambda_s) + w_{1\cdot 2}(\lambda_s) .$$

Here $w_1(\lambda)$, $w_{\Delta y_2}(\lambda)$ and $w_{1\cdot 2}(\lambda)$ denote the dft's of $y_1(n)$, $\Delta y_2(n)$ and $x_{1\cdot 2}(n) = x_1(n) - Cx_2(n) = x_1(n) - C\Delta y_2(n)$ where $C = \Omega_{12}\Omega_{22}^{-1}$. Band spectral regression on (22) for $\lambda_s \in \mathcal{B}_0$ is simply multivariate least squares. When there are no restrictions on B in (22) this produces an asymptotically efficient estimation procedure, the reason being that $w_{1\cdot 2}(\lambda_s)$ is asymptotically independent of $w_{\Delta y_2}(\lambda_s)$ for $\lambda_s \in \mathcal{B}_0$. In effect, the efficiency of least squares on (22) is just a frequency domain version of the result (from Phillips (1988a) that OLS on the following equation in the time domain

$$y_1(n) = By_2(n) + C\Delta y_2(n) = u_{1\cdot 2}(n)$$

is optimal when $\{u(n)\}$ is iid $N(0,\Omega)$.

The subsystem band spectral estimator of B in (22) has the following form

$$(23) \qquad B_0^+ = [\hat{f}_{12}(0) - \hat{f}_{1\Delta}(0)\hat{f}_{\Delta\Delta}(0)^{-1}\hat{f}_{\Delta 2}(0)][\hat{f}_{22}(0) - \hat{f}_{2\Delta}(0)\hat{f}_{\Delta\Delta}(0)^{-1}\hat{f}_{\Delta 2}(0)]^{-1}$$

and the full band estimator is given by the corresponding weighted regression formula

$$(24)\ \mathrm{vec}(B^{+}) = \Big[\Sigma_j \hat{f}_{1\cdot 2}^{-1}(\omega_j) \otimes \hat{f}_{22}(-\omega_j) - (\Sigma_j \hat{f}_{1\cdot 2}^{-1}(\omega_j) \otimes \hat{f}_{2\Delta}(-\omega_j))$$

$$\cdot \Big[\Sigma_j \hat{f}_{1\cdot 2}^{-1}(\omega_j) \otimes \hat{f}_{\Delta\Delta}(-\omega_j)\Big]^{-1} (\Sigma_j \hat{f}_{1\cdot 2}^{-1}(\omega_j) \otimes \hat{f}_{\Delta 2}(-\omega_j)\Big]^{-1}$$

$$\cdot \Big[\Sigma_j (\hat{f}_{1\cdot 2}^{-1}(\omega_j) \otimes I)\mathrm{vec}(\hat{f}_{12}(\omega_j)) - (\Sigma_j \hat{f}_{1\cdot 2}^{-1}(\omega_j) \otimes \hat{f}_{2\Delta}(-\omega_j)) \cdot \Big[\Sigma_j \hat{f}_{1\cdot 2}^{-1}(\omega_j) \otimes \hat{f}_{\Delta\Delta}(-\omega_j)\Big]^{-1}$$

$$\cdot (\Sigma_j (\hat{f}_{1\cdot 2}^{-1}(\omega_j) \otimes I))\mathrm{vec}(\hat{f}_{1\Delta}(\omega_j))\Big] .$$

There may be some advantage to the use of B^{+} in small samples. However, unlike B_0^{+}, B^{+} relies on a preliminary estimate of the spectral matrix $f_{1\cdot 2}(\omega)$ of the residuals in (22). It is therefore operational only as a two step procedure, whereas B_0^{+} is directly computable from the data.

In formulae (23) and (24) we use $\hat{f}_{\Delta\Delta}(\cdot)$ and $\hat{f}_{1\cdot 2}(\cdot)$ to denote the estimated spectral matrices of $\Delta y_2(n)$ and $u_{1\cdot 2}(n)$ and $\hat{f}_{12}(\cdot)$, $\hat{f}_{1\Delta}(\cdot)$, $\hat{f}_{2\Delta}(\cdot)$ to denote the estimated cross spectral matrices of $(y_1(n), y_2(n))$, $(y_1(n), \Delta y_2(n))$ and $(y_2(n), \Delta y_2(n))$, respectively. Again smoothed periodogram estimates underlie the stated formulae but other types of spectral estimate could equally well be employed.

3.4. Asymptotic Theory

The asymptotic distributions of $\tilde{B}$ and $\tilde{B}_0$ are given in the following theorem. The approach is the same as that of (1988e) with straightforward modifications to deal with the multivariate character of the regressions. Details of the derivations are therefore omitted.

THEOREM 1

(a) $T(\tilde{B}-B) \Rightarrow (\int_0^1 dS_{1\cdot 2}S_2')\left[\int_0^1 S_2S_2'\right]^{-1}$

(b) $T(\tilde{B}_0 - B) \Rightarrow (\int_0^1 dS_{1\cdot 2}S_2')\left[\int_0^1 S_2S_2'\right]^{-1}$

where

$$(25)\qquad \begin{bmatrix} S_{1\cdot 2} \\ S_2 \end{bmatrix} \begin{matrix} n_1 \\ n_2 \end{matrix} \equiv BM\left(\begin{bmatrix} \Omega_{11\cdot 2} & 0 \\ 0 & \Omega_{22} \end{bmatrix} \right) .$$

The common limit distribution in (a) *and* (b) *may alternatively be represented in the mixed normal form*

$$(26)\qquad \int_{G>0} N(0, \Omega_{11\cdot 2} \otimes G)dP(G) \equiv \int_{g>0} N\left[0, g\Omega_{11\cdot 2} \otimes \Omega_{22}^{-1}\right] dP(g)$$

where $G = \left[\int_0^1 S_2 S_2'\right]^{-1}$, $g = \int_0^1 (Q_2 W_1)^2$ *and* $Q_2 W_1 = W_1 - (\int_0^1 W_1 W_2')\left[\int_0^1 W_1 W_2'\right] W_2'$ *is the Hilbert projection in* $L_2[0,1]$ *of* W_1 *on the orthogonal complement of the space spanned by the elements of* W_2 . *Here*

$$\begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \begin{matrix} 1 \\ n_2 - 1 \end{matrix} \equiv BM(I_{n_2}) .$$

REMARK (a) The limit distribution (26) is a normal mixture for which the mixing variate may be either a matrix (G) or a scalar (g) . The limit theory belongs to the LAMN family of Jeganathan (1980, 1982), LeCam (1986) and Davies (1986). The estimators $\tilde{B}$ and $\tilde{B}_0$ are, in fact, asymptotically equivalent to the maximum likelihood estimator of B in (11). They have, correspondingly, all of the advantages of the latter including asymptotic median unbiasedness and optimality (see LeCam (1986) and Sweeting (1983)).

REMARK (b) As discussed in (1988e) the nuisance parameters of the limit distribution (25) involve only scale effects and hypothesis testing may be conducted in the usual way with conventional asymptotic chi–squared criteria. Thus, if we wish to test

$$H_0 :\ h(b) = 0 ,\ b = \text{vec } B$$

where h( ) is a twice continuously differentiable q–vector function of restrictions on b , the Wald statistic is constructed from $\tilde{b} = \text{vec}\,\tilde{B}$ in the usual way, viz. $M = h(\tilde{b})'[\tilde{H}V_T\tilde{H}']^{-1}h(\tilde{b})$ where $\tilde{H} = \partial h(\tilde{b})/\partial b'$ and

$$V_T = \frac{1}{T}\left[\frac{1}{2M}\sum_{j=-M+1}^{M} E'\hat{f}_{xx}(\omega_j)^{-1}E \otimes \hat{f}_{22}'(\omega_j)\right]^{-1}.$$

Similarly, setting $\tilde{b}_0 = \text{vec}\;\tilde{B}_0$ we have the statistic $M_0 = h(\tilde{b}_0)[\tilde{H}_0V_{T0}\tilde{H}_0']^{-1}h(\tilde{b}_0)$ where

$$V_{TO} = \frac{1}{T}\left[E'\hat{f}_{xx}(0)^{-1}E \otimes \hat{f}_{22}(0)\right]^{-1}.$$

Under the assumption that $H = \partial h(b)/\partial b'$ has full rank q we find as in (1988e) that

(27) $$M, M_0 \Rightarrow \chi^2_q$$

leading to conventional chi–squared tests of H_0.

REMARK (c) Identical asymptotic results apply to the subsystem estimators B_0^+ and B^+ given in Section 3.3. In particular, the one step estimator B_0^+ is asymptotically equivalent to $\tilde{B}$. Moreover, using the same approach to asymptotic tests as that of Remark (b) we have the test statistic $M_0^+ = h(b_0^+)[H_0^+V_{T0}^+H_0^{+\prime}]^{-1}h(b_0^+)$ where $b_0^+ = \text{vec}(B_0^+)$, $H_0^+ = \partial h(b_0^+)/\partial b'$ and $V_{T0}^+ = T^{-1}[\hat{f}_{1\cdot 2}(0) \otimes \hat{f}_{22}(0)^{-1}]$. As in (27) we find $M_0^+ \Rightarrow \chi^2_q$ under H_0.

REMARK (d) The case of restricted cointegrating matrices $A = A(\alpha)$ may be handled in a similar way. Setting $b = b(\alpha) = \text{vec}(B(\alpha))$ we have

$$b(\tilde{\alpha}) - b(\alpha^0) = J(\alpha^*)(\tilde{\alpha} - \alpha^0)$$

where $J(\,) = \partial b/\partial\alpha'$ and α^* is on the line segment that connects $\tilde{\alpha}$ and α^0. Since $b(\,)$ is continuously differentiable the asymptotic theory for $T(\tilde{\alpha} - \alpha^0)$ follows by conventional arguments, leading to the following extension of Theorem 1.

THEOREM 2

$$T(\tilde{\alpha}-\alpha^0) \Rightarrow \left[J'(\Omega_{11\cdot 2}^{-1} \otimes \int_0^1 S_2 S_2')J\right]^{-1}\left[J'(\Omega_{11\cdot 2}^{-1} \otimes I)\int_0^1 dS_{1\cdot 2} \otimes S_2\right]$$
$$\equiv \int_{G>0} N\left[0, \left[J'(\Omega_{11\cdot 2}^{-1} \otimes G)J\right]^{-1}\right] dP(G)$$

where $G = \int_0^1 S_2 S_2'$.

Hypothesis testing about α^0 may be conducted as in Remark (b) using asymptotic chi–squared criteria constructed in the usual way.

4. FILTERED SERIES AND TEMPORAL AGGREGATION

If all of the components of $y(t)$ in (1) are of the flow variable type then (10) is integrated over a unit time period to conform with the observable data. The discrete time ECM transforms to

(28) $$\Delta Y_n = -EAY_{n-1} + X_n \, .$$

where

(29) $$Y_n = \int_{n-1}^{n} y(t)dt \, , \; X_n = \int_{n-1}^{n} x(t)dt = \int_0^1 x(n-t)dt \, .$$

Treated as a continuous time process X_n has spectral matrix $f_{XX}^c(\lambda) = (h(\lambda))^2 f_{xx}^c(\lambda)$, where $h(\lambda)$ is the frequency response function of the temporal averaging filter (29) and is given by $h(\lambda) = (e^{i\lambda}-1)/i\lambda$. The spectrum of the discrete time sequence $\{X_n\}_{n=1}^{\infty}$ is then

$$f_{XX}^d(\lambda) = \Sigma_{j=-\infty}^{\infty} |h(\lambda+2\pi j)|^2 f_{xx}^c(\lambda+2\pi j) \, .$$

Since the estimation procedures given in Section 3 operate under general assumptions of stationary residuals it is just as valid to apply them to the transformed ECM (28) as it is to apply them to the original system (11). In fact, the use of time averaged data such as $\{Y_n\}$ does not in any way affect our methods. No modifications to our estimators or to our inferential procedures are needed to accommodate flow data. The only change in our results is the replacement of the long run covariance matrix Ω and its submatrices in Theorem 1 with $\underline{\Omega} = 2\pi f^d_{XX}(0) = 2\pi f^c_{xx}(0)$. (Note that the final equality holds because $h(2\pi j) = 0 \ (j \neq 0)$, $1 \ (j = 0)$.)

The above conclusion is rather simple in comparison to existing procedures, for which temporally aggregated data present substantial complications in terms of both the algorithms of estimation and the asymptotic theory of inference—see Phillips (1978) and Bergstrom (1983, 1984) for details. These complications increase in severity when there are stock and flow data simultaneously present in a continuous time system. Recent work in this latter context has been done by Harvey and Stock (1985), Agbeyegbe (1986), Zadrozny (1988) and Stock (1987b). Most recently Harvey and Stock (1988, 1989) have shown how to extend earlier Kalman filter type algorithms for the computation of the Gaussian likelihood to models with common stochastic trends.

Interestingly and in contrast to the above mentioned studies, the simultaneous presence of stock and flow data causes no difficulties in computation or in inference when the frequency domain methods of Section 3 are applied. We shall write $y(t)$ in partitioned format as

$$y(t) = Q\begin{bmatrix} y^s(t) \\ y^f(t) \end{bmatrix} = Q_s y^s(t) + Q_f y^f(t), \text{ say} \tag{30}$$

where the superscripts " s" and " f " signify that the associated components are measured as stocks or as flows respectively. Q is a permutation matrix which reorders the elements to conform with the earlier format given in (1). We define

$$\underline{Y}_n = Q\begin{bmatrix} y^s(n) \\ Y_n^f \end{bmatrix}$$

where

$$Y_n^f = \int_{n-1}^n y^f(t)dt .$$

Then

(31) $$Y_n - \underline{Y}_n = Q\begin{bmatrix} \int_{n-1}^n y^s(t)dt - y^s(n) \\ 0 \end{bmatrix} = Q_s(Y_n^s - y^s(n)) .$$

Note that $Z_n = Y_n - \underline{Y}_n$, when treated as a continuous time process, has spectral matrix

$$f_{zz}^c(\lambda) = |h(\lambda) - 1|^2 Q_s f_{y^s y^s}^c(\lambda) Q_s' .$$

Since $f_{y^s y^s}^c(\lambda)$ is continuous everywhere but the origin where its behavior is characterized by a pole of $O(1/\lambda^2)$ as $\lambda \to 0$ and since $|h(\lambda) - 1|^2 = O(\lambda^2)$ as $\lambda \to 0$ we deduce that $f_{zz}^c(\lambda)$ is continuous throughout $(-\infty,\infty)$. It follows that $Z_n = I(0)$. This applies also to the discrete time process $\{Z_n\}_{n=1}^\infty$ whose spectrum is given by

$$f_{zz}^d(\lambda) = \sum_{j=-\infty}^{\infty} |h(\lambda+2\pi j) - 1|^2 Q_s f_{y^s y^s}^c(\lambda+2\pi j) Q_s' .$$

From (26) and (31) we deduce that

(32) $$\Delta \underline{Y}_n = -EA\underline{Y}_{n-1} - EAZ_{n-1} - \Delta Z_n + X_n = -EA\underline{Y}_{n-1} + \underline{X}_n$$

where $\underline{X}_n$ is stationary. The ECM (32) can be used for estimation purposes with the observable sequence $\{\underline{Y}_n\}_{n=1}^\infty$ in just the same way as (11). Again, the inferential procedures outlined in Section 3 are not affected. As for the case of flow data, all that changes is the covariance matrix of the limit Brownian motion that appears in Theorem 1. The limit distribution given by (26) remains valid in form but now the long–run covariance matrix is given by $\underset{\sim}{\Omega} = 2\pi f^{d}_{\underline{XX}}(0)$. Thus, mixed stock and flow data are no more complicated to handle in estimation and inference than simple instantaneous data sets.

The above conclusion is also robust to certain plausible alternative formulations of the long run relationship (2). In empirical work the relationship (2) is often based on the idea of a partial equilibrium representation of the value of one variable in terms of others in the system. For instance, if y_1 represents consumption and y_2 income, then (2) indicates that there will be only stationary fluctuations in consumption expenditures about its partial equilibrium level $\hat{y}_1(t) = \beta y_2(t)$ determined as a constant fraction of income y_2 . When rewritten in ECM format the model then represents the adjustment process of consumption expenditure towards the partial equilibrium level $\hat{y}_1$. The use of a partial equilibrium value in this way in discrete time modeling seems justifiable and is well accepted. See Davidson *et al.* (1978), Hendry (1983, 1984), and Hendry and Richard (1982) for numerous examples. However, when time is continuous the formulation of a partial equilibrium level of one variable in terms of the levels of other variables at each instant in time seems less satisfactory in view of the inevitable high frequency oscillations that can be expected. Variables like the instantaneous rate of aggregate expenditure and aggregate income generation, for instance, sustain substantial fluctuations between peak and off peak periods (like day time and night time). In such cases, a formulation of partial equilibrium levels in terms of time averages seems much more satisfactory. For example if δ represents the decision making interval of a representative consumer with respect to expenditure (δ could be a week or a month and is most likely to correspond broadly with the interval between income receipts) then we might write in place of (33)

$$(34) \qquad \hat{y}_1(s) = (\beta/\delta)\int_{t-\delta}^{t} y_2(r)dr$$

for all values of $s \in (t-\delta, t]$. Partial equilibrium formulations of this type were used in

Bailey, Hall and Phillips (1987). The cointegrating relationship (2) would then be derived as follows

$$y_1(t) = \hat{y}_1(s) + u_1(t)\ ,\quad t-\delta \le s \le t$$

$$= \beta y_2(t) + \beta\{(1/\delta)\int_0^\delta y_2(t-r)dr - y_2(t)\} + u_1(t)$$

or

(35) $$y_1(t) = \beta y_2(t) + \underline{u}_1(t)\ .$$

Observe that $\underline{u}_1(t)$ is stationary with spectrum

$$f^c_{\underline{u}_1\underline{u}_1}(\lambda) = \left[1,\ \beta(1-h_\delta(\lambda))/i\lambda\delta\right] f^c_{uu}(\lambda) \begin{bmatrix} 1 \\ \beta(h_\delta(\lambda)-1)/i\lambda\delta \end{bmatrix}$$

where $h_\delta(\lambda) = (e^{i\delta\lambda}-1)/i\lambda\delta$. Note that $(1-h_\delta(\lambda))/i\lambda\delta = O(1)$ as $\lambda \to 0$ and thus $f^c_{\underline{u}_1\underline{u}_1}(\lambda)$ is continuous throughout $(-\infty,\infty)$.

It follows that in order to estimate the coefficient β in the partial equilibrium formulation (34) it is enough to work with the equivalent cointegrating relationship (35). This brings us back to our original framework and the methods and results of Sections 3 apply without change.

In each of the above examples the cointegrating relationship is undisturbed by the temporal averaging that occurs in the data and in the model specification. We might ask whether the same is true of more general linear transformations. Suppose $Y(n)$ is obtained from the original series $y(t)$ by the action of a filter of the form $Y(n) = \int_a^b g(s)y(n-s)ds$, $\int_a^b |g(s)|ds < \infty$. This filter may be interpreted as a linear operator on the space of random functions where $y(t)$ is defined. Its frequency response function is $G(\lambda) = \int_a^b g(s)e^{i\lambda s}ds$. The following theorem gives necessary and sufficient conditions for the invariance of the cointegrating relationship (1) under the action of this filter.

THEOREM 3. *Suppose* $Y(n)$ *is obtained from* $y(t)$ *by a linear filter whose response function is* $G(\lambda)$. *Then* $AY(n) = I(0)$ *and the cointegrating relationship is preserved under the filter iff*

(36) $$G(0)'A' \in \mathcal{R}(A') .$$

REMARK. In the examples given earlier we have the following response functions:

(i) $G(\lambda) = h(\lambda)I_n$ (n–variable flow data)

(ii) $G(\lambda) = Q\begin{bmatrix} I_{n^s} & 0 \\ 0 & h(\lambda)I_{n^f} \end{bmatrix}Q'$ (n^s stock variables and n^f flow variables)

(iii) $G(\lambda) = \begin{bmatrix} 1 & 0 \\ 0 & h_\delta(\lambda) \end{bmatrix}$ (smoothed partial equilibrium formulation)

where $h_\delta(\lambda) = (e^{i\delta\lambda} - 1)/i\lambda\delta$, $h(\lambda) = (e^{i\lambda} - 1)/i\lambda$.

In each of these examples $G(0)$ is the identity and (36) holds trivially. Thus, the cointegrating relationship is preserved under the action of each of these filters.

5. GAUSSIAN ESTIMATION AND TRADITIONAL DISCRETE APPROXIMATIONS

Traditional methods of estimation of continuous systems such as (7) rely on either the exact discrete model (11) (see Phillips (1972, 1974) and Bergstrom (1983)) or direct discrete approximations to (7) (see Bergstrom (1966) and Sargan (1973)). These methods have been developed for models where the properties of the residual processes are made more explicit than we have done in the present paper. For example, suppose that $w(t)$ in (7) is a pure noise in continuous time with spectral matrix $(1/2\pi)\Sigma_w$. This is an extreme case because, given the form of (7), it implies that the speeds of adjustment towards partial equilibrium are the same for each component of $y_1(t)$. However, it is still a very useful example: there are no short run dynamic adjustment coefficients to estimate, no need for extra identifiability conditions and attention is naturally focussed on estimates of the unnormalized long run coefficients in the matrix B. Moreover, as we shall discuss, results for this polar case extend at least in qualitative terms to models where there are variable speeds of adjustment across equations. We shall examine the properties of Gaussian maximum likelihood estimates of B that are obtained from both the exact discrete model and the discrete approximation to (7).

We start with the exact discrete model (11). This may be written as

$$y(n) = (I - fEA)y(n-1) + \epsilon(n) \quad . \tag{37}$$

When $w(t)$ in (7) is pure noise, the discrete sequence $\{\epsilon(n)\}_{n=1}^{\infty}$ is iid(0,V) with $V = \int_0^1 \exp(-sEA)\Sigma_w \exp(-sA'E')ds$. Upon further reduction (37) becomes

$$y_+(n) = fEBy_2(n-1) + \epsilon(n) = EB\underline{y}_2(n-1) + \epsilon(n) \tag{38}$$

where $y_+(n)' = [\Delta y_1(n)' + fy_1(n-1)', \Delta y_2(n)']$.

Note that $f = (e-1)/e$ is known, is conveniently absorbed into $\underline{y}_2(n-1)$ and thus (38) is linear in the unknown coefficients B. The Gaussian estimator of B is easily expressed in instrumental variables form as:

$$\hat{B} = \left[E'\hat{V}^{-1}E\right]^{-1}(E'\hat{V}^{-1})(\Sigma_{n=1}^{T}y_+(n)\underline{y}_2(n-1)')\left[\Sigma_{n=1}^{T}\underline{y}_2(n-1)\underline{y}_2(n-1)'\right]^{-1}$$

where $\hat{V} = T^{-1}\Sigma_{n=1}^{T}(y_+(n) - E\hat{B}\underline{y}_2(n-1))(y_+(n) - E\hat{B}\underline{y}_2(n-1))'$. Its asymptotic distribution is as follows:

THEOREM 4

$$\text{(39)} \qquad T(\hat{B}-B) \Rightarrow (1/f)\left(\int_0^1 dR_{1\cdot 2}R_2'\right)\left[\int_0^1 R_2R_2'\right]^{-1}$$

where

$$\begin{bmatrix} R_{1\cdot 2} \\ R_2 \end{bmatrix} \equiv BM\left(\begin{bmatrix} V_{11\cdot 2} & 0 \\ 0 & V_{22} \end{bmatrix}\right), \; V_{11\cdot 2} = V_{11} - V_{21}'V_{22}^{-1}V_{21}$$

and V *is partitioned conformably with* y(t) *in* (1).

REMARK (a) The estimator $\hat{B}$ makes explicit use of the fact that the discrete model (37) has serially independent errors. In this respect it uses much more information than the spectral estimators studied in Section 3. Yet the limit distributions of the two types of estimators, given in Theorems 1 and 4, are the same. To see this we need only relate the long run variance matrix Ω, on which (25) depends, to V. From (11) we have $x(n) = \epsilon(n) + (1/e)EAy(n-1)$ and by elementary calculations we find $\Omega = 2\pi f_{xx}(0) = V + (e-1)^{-1}EAV + (e-1)^{-1}VA'E' + (e-1)^{-2}EAVA'E'$ and $\Omega_{11\cdot 2} = (e/(e-1))^2V_{11\cdot 2}$. Since $V_{22} = \Omega_{22}$ and $f^2 = ((e-1)/e)^2$ we deduce that the limit distribution in (39) is equivalent to $(\int_0^1 dS_{1\cdot 2}S_2')(\int_0^1 S_2S_2')^{-1}$, as in Theorem 1. The spectral estimators $\tilde{B}$ and $\tilde{B}_0$ are therefore truly adaptive and lose no asymptotic efficiency against Gaussian estimators like $\hat{B}$ that utilize much more specific information about the residuals. Although there is no loss of efficiency in infinite samples we would certainly expect the spectral estimators to be inefficient relative to $\hat{B}$ in finite samples.

REMARK (b) The limit theory (39) applies when the discrete sequence $\epsilon(n)$ is orthogonal with variance matrix V. If $\epsilon(n)$ is known only to be stationary, however, $\hat{B}$ is still consistent but it has no other useful properties. Its limit distribution no longer belongs to the LAMN family, it involves nuisance parameters and it has second order bias effects. In this case, all of the main advantages of working with the exact discrete model are gone.

REMARK (c) Since the limit distribution in (39) is mixed normal conventional methods of asymptotic testing remain valid. In fact, conventional formulae are still appropriate and, like the Wald tests studied in Section 3, these lead to criteria which are asymptotically chi–squared. Again, these advantages are lost when the assumption of serially uncorrelated errors is relaxed.

REMARK (d) We may allow for variable speeds of adjustment towards partial equilibrium in the above analysis by working from (8) rather than (7). Suppose we set $F = HA$ with $A = [I, -B]$ as before and $H = E\Lambda$ where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n_1})$ and the λ_i are adjustment coefficients (with negative real parts) that vary across equations. Then in place of (38) we have the system

$$\Delta y(n) = ECAy(n-1) + \epsilon(n)$$

where $C = e^{\Lambda} - I$. Gaussian estimation leads to $\hat{B}$ and $\hat{C}$ and, if the elements of C are identifiable (which typically requires additional conditions—see Phillips (1973)), we find the following asymptotics in place of (39)

$$T(\hat{B}-B) \Rightarrow C^{-1}(\int_0^1 dR_{1.2}R_2')\left[\int_0^1 R_2R_2'\right]^{-1} .$$

The only change is therefore the replacement of the scale effect $(1/f)$ in (39) by the matrix C^{-1}. None of the substantive results or comparisons with the spectral estimators are affected by this change.

REMARK (e) Finally, we observe that the above analysis may be extended as in Section 3.2 to the case of restricted cointegrating matrices of the form $B = B(\alpha)$. For empirical applications this type of restricted formulation may well be more relevant. The observations made under Remarks (a)–(d) continue to apply in this case.

We now turn to the discrete approximations of (7) given by

$$(40)\qquad \Delta y(n) = -\tfrac{1}{2}EA\{y(n) + y(n-1)\} + v(n)\,,$$

$$v(n) = \textstyle\int_{n-1}^{n} w(t)dt - EA\{\int_{n-1}^{n} y(t)dt - (1/2)(y(n) + y(n-1))\}\,.$$

Model (40) is a nonrecursive discrete approximation to (7) of the type originally explored in Bergstrom (1966). It has reduced form

$$(41)\qquad y(n) = \left[I + \tfrac{1}{2}EA\right]^{-1}\left[I - \tfrac{1}{2}EA\right]y(n-1) + \eta(n) = \left[I - \tfrac{2}{3}EA\right]y(n-1) + \eta(n)\,.$$

This model approximates (37), the only difference being in the two scale factors $f = 0.63212$ in (37) and 2/3 in (41). We write (40) as

$$(42)\qquad y_{\#}(n) = EB\underset{\sim}{y}_2(n-1) + \eta(n)$$

where $y_{\#}(n)' = (y_1(n)' - (1/3)y_1(n-1)', \Delta y_2(n)')$ and $\underset{\sim}{y}_2(n-1) = (2/3)y_2(n-1)$. The Gaussian estimator of B in (42) is

$$B_{\#} = \left[E'\hat{\Sigma}_{\eta}^{-1}E\right]^{-1}E'\hat{\Sigma}_{\eta}^{-1}\left(\Sigma_{n=1}^{T} y_{\#}(n)\underset{\sim}{y}_2(n-1)'\right)\left[\Sigma_{n=1}^{T}\underset{\sim}{y}_2(n-1)\underset{\sim}{y}_2(n-1)'\right]^{-1}$$

where $\hat{\Sigma}_{\eta} = T^{-1}\Sigma_{n=1}^{T}(y_{\#}(n) - EB_{\#}\underset{\sim}{y}_2(n-1))(y_{\#}(n) - EB_{\#}\underset{\sim}{y}_2(n-1))'$. Using (38) the residual $\eta(n)$ in (42) is calculated as $\eta(n) = \epsilon(\eta) + (2/3 - f)EAy(n-1)$. This process is stationary but is serially dependent, unlike $\epsilon(n)$. Its discrete time spectral matrix is

$$f_{\eta\eta}^{d}(\lambda) = (1/2\pi)V + \frac{(2/3 - f)^2}{2\pi}\, d(\lambda)EA\Sigma_{w}A'E'$$

where $d(\lambda) = \Sigma_{j=-\infty}^{\infty}\left[1 + (\lambda+2\pi j)^2\right]^{-1}$. We set $\Omega_{\eta} = 2\pi f_{\eta\eta}^{d}(0)$ and decompose this matrix as $\Omega_{\eta} = \Sigma_{\eta} + \Lambda_{\eta} + \Lambda_{\eta}'$ where $\Sigma_{\eta} = E(\eta(0)\eta(0)')$ and $\Lambda_{\eta} = \Sigma_{k=1}^{\infty}E(\eta(0)\eta(k)')$. Next we define $N(r) = BM(\Omega_{\eta})$ and partition this process conformably with $y(t)$ as $N(r)' = (N_1(r)', N_2(r)')$. Note that from the construction of $\eta(n)$ we have the equivalence $N_2(r) \equiv R_2(r)$.

The asymptotic distribution of the estimator $B_{\#}$ is given in the following result. Since the derivation follows lines that are similar to earlier work of Park and Phillips (1988, 1989) the details are omitted.

THEOREM 5

$$T(B_{\#} - B) \Rightarrow (3/2)\left(\int_0^1 dN_E N_2' + \Lambda_E' E_*\right)\left[\int_0^1 N_2 N_2'\right]^{-1} \tag{43}$$

where $N_E(r) = \left[E'\Sigma_\eta^{-1}E\right]^{-1} E'\Sigma_\eta^{-1}N(r)$ *and* $\Lambda_E' = \left[E'\Sigma_\eta^{-1}E\right]^{-1} E'\Sigma_\eta^{-1}\Lambda_n'$.

REMARK (a) $B_{\#}$ is consistent for B so that, in this model at least, estimates from the discrete approximation (40) do not carry an asymptotic bias, in contrast to traditional theory for stationary systems (Bergstrom (1966), Sargan (1973)).

REMARK (b) Although $B_{\#}$ is consistent, its asymptotic distribution is miscentered and there is what is referred to elsewhere (see Phillips (1988a)) as a second order bias in the limit distribution (43). This has two causes: the term $\Lambda_E' E_*$ in the formula on the right side of (43); and the fact that the Brownian motions $N_E(r)$ and $N_2(r)$ are dependent. The first arises from the serial dependence in the process $\eta(n)$, which is not taken into account in the estimator $B_{\#}$. The second is the result of the misspecification bias that arises in the use of the discrete approximate model (40) rather than the exact model (37).

REMARK (c) The limit distribution (43) is not a normal mixture. It involves nuisance parameters arising from the serial correlation effects in $\Lambda_E' E_*$ and from the correlation between N_E and N_2 . These nuisance parameters are an obstacle to the use of (43) for inference and they lead to size distortions and invalid inferences if traditional asymptotic chi–squared tests are used for inferential purposes with $B_{\#}$. In this latter respect, our results correspond with the earlier theory in Bergstrom (1966) and Sargan (1973) concerning the use of discrete approximation in stationary continuous systems.

6. SOME CONCLUDING REMARKS

The model on which our attention has focused is the first order stochastic differential equation system (7). This model includes a wider class of continuous systems than may be apparent from its simple form. For example, $w(t)$ could be generated by a stable ARMA(p,q) system in continuous time of the form

$$A(D)w(t) = B(D)\zeta(t) \tag{44}$$

where $A(D) = \Sigma_0^p A_i D^i$ and $B(D) = \Sigma_{j=0}^q B_j D^j$ $(q \leq p)$ are matrix polynomials in D and $\zeta(t)$ is a pure noise vector with constant spectral matrix $(1/2\pi)\Sigma_\zeta > 0$. The system (7) is then the higher order model

$$A(D)(DI + EA)y(t) = B(D)\zeta(t) \tag{7$'$}$$

in which the coefficients of $A(D)$ and $B(D)$ embody the short run dynamics.

The system (7) also accommodates stationary exogenous inputs $z(t)$. These may be absorbed into the generating process for the residuals by writing

$$A(D)\underline{w}(t) = A(D)Cz(t) + B(D)\zeta(t)$$

in place of (44) so that the complete model for $y(t)$ is

$$Dy(t) = -EAy(t) + Cz(t) + w(t) = -EAy(t) + \underline{w}(t) \tag{7$''$}$$

where C is some constant matrix of coefficients. Interestingly, for the purposes of inference, no (asymptotic) efficiency is lost by absorbing the stationary process $z(t)$ into the residual in this way. This is so in spite of the fact that discrete observations of $z(t)$ are available. The reason is that the effects of $z(t)$ are already accounted for in our estimation procedure. By systems estimation we are, in effect, adjusting for the conditional mean in (the first n_1 equations of) (7)$''$ and this adjustment deals in a nonparametric way with the input $z(t)$. One might expect, however, that for correctly specified models the explicit use of observable exogenous series like $z(t)$ would lead to some finite sample gains. This is something that can be explored in Monte Carlo work.

Our model may also be extended to allow for deterministic as well as stochastic trends. All that is needed is to replace (7) by the system

$$(7)''' \qquad Dy(t) = k(t) - EAy(t) + w(t)$$

where $k(t) = \Sigma_{i=0}^{m} k_i t^i$ and $Ak_i = 0$ $(i = 0, ..., m)$. The latter condition ensures that the cointegrating relationship (2) persists and the matrix A annihilates the deterministic as well as the stochastic trends. Note from (7)‴ that $Ay(t)$ then satisfies the stable system

$$D(Ay(t)) = -Ay(t) + Aw(t) .$$

The discrete time equivalent of (7)‴ is

$$(11)' \qquad \Delta y(n) = k_*(n) - EAy(n-1) + x(n)$$

which replaces our earlier equation (11). In (11)′ we find after a small calculation that

$$k_*(n) = \Sigma_{i=0}^{m} k_i \Sigma_{j=0}^{i} \begin{bmatrix} i \\ j \end{bmatrix} \frac{(-1)^{j-i}}{j-i+1} n^j = \Sigma_{i=0}^{m} k_{*i} n^i , \text{ say.}$$

Now we need only remove the deterministic trend $k_*(n)$ from the data by regression before applying the methods of Section 3 to estimate A. The asymptotic results are unchanged except that suitably detrended Brownian motions appear in the limit distributions of Theorem 1; and inferential procedures are unaffected.

We end this paper with some brief comments that bear on the interpretation of empirical results and Monte Carlo experiments in this field. As shown in Sections 3 and 5, estimates of the long–run equilibrium coefficients converge at the rate $O_p(T^{-1})$. Correspondingly, conventional standard error estimates are also $O_p(T^{-1})$. This suggests that standard errors in models of the type we have studied will tend to be smaller than is usual in applied econometric work with stationary series of a comparable length. This is borne out by some of the empirical results in the area, where long run equilibrium coefficients seem to be estimated very precisely—see, for example, Table 1 of Bergstrom and Wymer (1976). It also seems likely that the results in Section 5 can be used to help explain some of the empirical differences between the use of the discrete approximation and the exact discrete model. For instance, estimated standard errors obtained from the exact discrete model seem on the whole to be much smaller than those from the discrete approximation (see Phillips (1972) and Bergstrom and Wymer (1976)). In part this may be a simple consequence of the optimality of Gaussian estimates of the exact discrete model (see Theorem 4 and the following Remark (a)). But it may also be explained by the fact that the specification error that is inherent in the discrete approximation induces a serial dependence in the residual which in many cases will lead to an increase in residual variance (i.e. Σ_η and Ω_η of Section 5). These are matters that can be further explored in appropriately designed sampling experiments.

APPENDIX

Proof of Theorem 3. The spectral matrix of $Y(n)$ is $f_{YY}(\lambda) = G(\lambda)f_{yy}(\lambda)G(\lambda)^*$ and $f_{yy}(\lambda) = O(1/\lambda^2)$ for $\lambda \sim 0$. From Remark (a) of Section 2 we know that $Af_{yy}(0)A' = Af^c_{ww}(0)A' < \infty$ and the matrix A is unique up to normalization. Now the spectral matrix of $AY(n)$ at the origin is $AG(0)f_{yy}(0)G(0)'A'$, so that this matrix is bounded iff $G(0)'A' \in \mathfrak{R}(A')$, as required.

Proof of Theorem 4. Observe that $\hat{B} = \text{argmin}_B D_T(B)$ where $D_T(B) = \ln\det\left\{T^{-1}\Sigma_1^T(y_+(n) - EB\underline{y}_2(n-1))(y_+(n) - EB\underline{y}_2(n-1)'\right\}$. At the true value of B we have $D_T \to_{as} \ln\det V$, whereas in any close set not containing B D_T diverges. It follows by Wu's (1981) Theorem that $\hat{B}$ and $\hat{V}$ are consistent. Then

$$\begin{aligned}
T(\hat{B}-B) &= \left[E'\hat{V}^{-1}E\right]^{-1}E'\hat{V}^{-1}(T^{-1}\Sigma\epsilon(n)\underline{y}_2(n-1)')\left[T^{-2}\Sigma\underline{y}_2(n-1)\underline{y}_2(n-1)'\right]^{-1} \\
&= \left[E'V^{-1}E\right]^{-1}E'V^{-1}(T^{-1}\Sigma\epsilon(n)\underline{y}_2(n-1)')\left[T^{-2}\Sigma\underline{y}_2(n-1)\underline{y}_2(n-1)'\right]^{-1} + o_p(1) \\
&\Rightarrow (E'VE)^{-1}E'V^{-1}(f\textstyle\int_0^1 dRR_2')\left[f^2\int_0^1 R_2R_2'\right]^{-1}
\end{aligned}$$

where $R = (R_1', R_2')' = BM(V)$. Noting that $E'V^{-1}R = BM(E'V^{-1}E)$ and $(E'VE)^{-1} = V_{11\cdot 2}$ we get the desired result.

REFERENCES

Agbeyegbe, T. D. (1984). "The exact discrete analog to a closed linear mixed-order system," *Journal of Economic Dynamics and Control*, 7, 363–375.

_______ (1987). "An exact discrete analog to a closed linear first order continuous time system with mixed sample," *Econometric Theory*, 3, 143–149.

Bailey, R. W., V. B. Hall and P. C. B. Phillips (1987). "A model of output, employment capital formation and inflation," in G. Gandolfo and F. Marzano (eds.), *Keynesian Theory Planning Models and Quantitative Economics*, Giuffrè–Editore.

Bergstrom, A. R. (1966). "Nonrecursive models as discrete approximations to systems of stochastic differential equations," *Econometrica*, 34, 173–182.

_______ (1983). "Gaussian estimation of structural parameters in higher order continuous time dynamic models," *Econometrica*, 51, 117–152.

_______ (1984). "Continuous time stochastic models and issues of aggregation over time," in Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 11. North Holland: Amsterdam.

_______ (1985). "The estimation of parameters in nonstationary higher order continuous time dynamic models," *Econometric Theory*, 1, 369–386.

_______ (1986). "The estimation of open higher order continuous time dynamic models with mixed stock and flow data," *Econometric Theory*, 2, 350–373.

_______ (1988). "The history of continuous time econometric models," *Econometric Theory* (forthcoming).

_______ and C. R. Wymer (1976). "A model of disequilibrium neoclassical growth and its application to the United Kingdom," in A. R. Bergstrom (eds)., *Statistical Inference in Continuous Time Economic Models*. North Holland: Amsterdam.

Davidson, J. E. H., D. F. Hendry, F. Srba, and J. S. Yeo (1978). "Econometric modeling of the aggregate time series relationship between consumers expenditure and income in the United Kingdom," *Economic Journal*, 88, 661–692.

Davies, R. B. (1986). "Asymptotic inference when the amount of information is random," in L. M. LeCam and R. A. Olshen (eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Vol. II. Wadsworth Inc.

Hannan, E. J. (1963a). "Regression for time series," in M. Rosenblatt (ed.), *Time Series Analysis*. New York: Wiley.

_______ (1963b). "Regression for time series with errors of measurement," *Biometrika*, 50, 293–302.

Hansen, L. P. and T. J. Sargent (1981). "Methods for estimating continuous time rational expectations models from discrete time data," Federal Reserve Bank, Minneapolis.

_______ (1983). "The dimensionality of the aliasing problem in models with rational spectral densities," *Econometrica*, 51, 377–387.

Harvey, A. C. and J. H. Stock (1985). "The estimation of higher order continuous time autoregressive models," *Econometric Theory*, 1, 97–112.

_______ (1988). "Continuous time autoregressive models with common stochastic trends," *Journal of Economic Dynamics and Control*, 12, 365–384.

_______ (1989). "Estimating higher order continuous time autoregressions with an application to money–income causality," *Journal of Econometrics*, 42, .

Hendry, D. F. (1983). "Econometric modeling: The consumption function of retrospect," *Scottish Journal of Political Economy*, 30, 193–220.

_______ and J. F. Richard (1983). "On the formulation of empirical models in dynamic econometrics," *Journal of Econometrics*, 20, 3–33.

Jeganathan, P. (1980). "An extension of a result of L. LeCam concerning asymptotic normality," *Sankhya Series A*, 42, 146–160.

_______ (1982). "On the asymptotic theory of estimation when the limit of the log–likelihood ratios is mixed normal," *Sankhya Series A*, 44, 172–212.

LeCam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* New York: Springer.

Park, J. Y. and P. C. B. Phillips (1988). "Statistical inference in regressions with integrated processes: Part 1," Cowles Foundation Discussion Paper No. 811R, Yale University.

_______ (1989). "Statistical inference in regressions with integrated processes: Part 2," Cowles Foundation Discussion Paper No. 819R.

Phillips, P. C. B. (1972). "The structural estimation of a stochastic differential equation system," *Econometrica*, 40, 1021–1041.

_______ (1973). "The problem of identification in finite parameter continuous time models," *Journal of Econometrics*, 1, 351–362.

_______ (1974). "The estimation of some continuous time models," *Econometrica*, 42, 803–824.

_______ (1978). "The treatment of flow data in the estimation of continuous time systems," in A. R. Bergstrom, A. J. L. Catts, M. H. Peston and B. D. J. Silverstone (eds.), *Stability and Inflation.* Wiley: New York.

_______ (1986). "Understanding spurious regressions in econometrics," *Journal of Econometrics*, 33, 311–340.

_______ (1988a). "Optimal inference in cointegrated systems," Cowles Foundation Discussion Paper No. 866, Yale University.

_______ (1988b). "Weak convergence to the matrix stochastic integral $\int_0^1 BdB'$," *Journal of Multivariate Analysis*, 24, 252–264.

_______ (1988c). "Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations," *Econometric Theory*, 4, 528–533.

_______ (1988d). "Multiple regression with integrated processes," in N. U. Prabhu (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics*, 80, 79–106.

_______ (1988e). "Spectral regression for cointegrated time series," Cowles Foundation Discussion Paper No. 872, Yale University. Forthcoming in W. Barnett (ed.), *Nonparametric and Semiparametric Methods in Economics and Statistics*, CUP, 1990.

Phillips, P. C. B. and S. N. Durlauf (1986). "Multiple time series with integrated variables," *Review of Economic Studies*, 53, 473–496.

Phillips, P. C. B. and B. E. Hansen (1989). "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies* (forthcoming).

Robinson, P. M. (1976a). "The estimation of linear differential equations with constant coefficients," *Econometrica*, 44, 751–764.

_______ (1976b). "Instrumental variables estimation of differential equations," *Econometrica*, 44, 765–776.

_______ (1976c). "The aliasing problem in differential equation estimation," *Journal of the Royal Statistical Society*, Series B, 38, 180–188.

_______ (1977). "The construction and estimation of continuous time models and discrete approximations in econometrics," *Journal of Econometrics*, 6, 173–197.

_______ (1979). "Continuous model fitting from discrete data," in D. R. Brillinger and G. C. Tiao (eds.), *Directions in Time Series: Proceedings of the IMS Special Topics Meeting on Time Series Analysis May 1–3, 1978.* Iowa State University.

Sargan, J. D. (1974). "Some discrete approximations to continuous time stochastic models," *Journal of the Royal Statistical Society*, Series B, 36, 74–90.

Solo, V. (1987). "Intrinsic random functions and the paradox of 1/f noise," Johns Hopkins (mimoegraphed).

Stock, J. H. (1987a). "Asymptotic properties of least squares estimators of cointegrating vectors," *Econometrica*, 55, 1035–1056.

_______ (1987b). "Temporal aggregation and structural inference in macroeconomics: A comment," *Carnegie–Rochester Conference Series of Public Policy.*

Sweeting, T. J. (1983). "On estimator efficiency in stochastic processes," *Stochastic Processes and Applications*, 15, 93–98.

Wymer, C. R. (1972). "Econometric estimation of stochastic differential equation systems," *Econometrica*, 40, 565–577.

Zadrozny, P. (1988). "Gaussian likelihood of continuous time ARMAX models when data are stocks and flows at different frequencies," *Econometric Theory*, 4, 108–124.