ERROR CORRECTION AND LONG-RUN EQUILIBRIUM IN CONTINUOUS TIME

By P. C. B. Phillips[1]

This paper deals with error correction models (ECM's) and cointegrated systems that are formulated in continuous time. Long-run equilibrium coefficients in the continuous system are always identified in the discrete time reduced form, so that there is no aliasing problem for these parameters. The long-run relationships are also preserved under quite general data filtering. 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.

KEYWORDS: Aliasing, error correction, long run equilibrium, 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. Readers are referred to Bergstrom (1988) for a history and a review of recent developments in the field. 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, which may be constructed in either the time domain or the frequency domain. The second approach is to work from the exact discrete model that is induced by the continuous system. 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, it has been conventional to 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. Second, the triangular system format of the discrete time ECM opens the way to simple frequency domain estimation methods of the type discussed by the author (1988b) elsewhere. An important aspect of these methods is their nonparametric treatment of the regression 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. Thus, problems of temporal aggregation which present impediments to computation and inference in more traditional approaches simply do not arise in the present context.

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. Vector Brownian motion with covariance matrix Ω is written "$BM(\Omega)$". We use $[x]$ to denote the smallest integer $\leqslant x$. All limits given in the paper are as the sample size $T \to \infty$.

The symbolism "$I(0)$" will be taken to mean all covariance stationary processes in continuous time with bounded continuous spectra $f(\lambda)$ for which $f(0) > 0$. 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 $f_i(0) > 0$, 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, IDENTIFICATION, AND FILTERING

2.1. *Representation*

Let $y(t)$ be an m-vector $I(1)$ process in continuous time and $u(t)$ be an m-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. We partition the vectors $y(t)$ and $u(t)$ into the subvectors

$$(1) \qquad y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} \begin{matrix} m_1 \\ m_2 \end{matrix}, \qquad u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} \begin{matrix} m_1 \\ m_2 \end{matrix}$$

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

$$(2) \qquad y_1(t) = By_2(t) + u_1(t),$$

$$(3) \qquad Dy_2(t) = u_2(t).$$

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, high frequency perturbations that are unimportant to the long-run relationship are absorbed by the process $u_1(t)$. Solving (3) with initial conditions at $t = 0$ we have

$$(4) \qquad y_2(t) = \int_0^t u_2(s)\, ds + y_2(0)$$

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.

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

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

and combining this equation with (3) to give the system

$$(5) \qquad Dy(t) = -EAy(t) + w(t),$$

where

$$E = \begin{bmatrix} I \\ 0 \end{bmatrix} \begin{matrix} m_1 \\ m_2 \end{matrix}, \qquad A = [I, -B], \qquad w(t) = \begin{bmatrix} u_1(t) + Bu_2(t) + Du_1(t) \\ u_2(t) \end{bmatrix}.$$

The coefficient matrix E in (5) 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 (5) is a continuous time analogue of the ECM representation given in Phillips (1991) for discrete time models. It has the same advantage as the discrete time counterpart that it is in triangular system format. There is a converse version of this relationship. Suppose we start with the stochastic differential equation system

$$(6) \qquad Dy(t) = Fy(t) + w(t),$$

where $w(t) = I(0)$ and the coefficient matrix F is known to have m_1 latent roots in the left half plane and m_2 zero latent roots. We simply write F as $F = HG$, where H $(m \times m_1)$ and G $(m_1 \times m)$ both have rank m_1, and then (6) can be recast as

$$(7) \qquad Dy(t) = EGy(t) + (H - E)Gy(t) + w(t) = EGy(t) + v(t),$$

where $v(t) = I(0)$, putting the system into the same error correction format as (5). Discrete time series generated by (5) also satisfy the exact discrete model

$$(8) \qquad y(n) = \exp\{-EA\}y(n-1) + \varepsilon(n),$$

$$\varepsilon(n) = \int_0^1 \exp\{-sEA\}w(n-s)\, ds.$$

Indeed, from the series representation

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

and the fact that $AE = I$ we have $\exp(-EA) = I - fEA$ where $f = (e-1)/e$. It follows that (8) may be rewritten as

$$(9) \qquad \Delta y(n) = -EAy(n-1) + x(n), \qquad \text{with}$$

$$x(n) = \varepsilon(n) + (1/e)EAy(n-1),$$

where $x(n) = I(0)$ since both $\varepsilon(n)$ and $\Delta y(n)$ are stationary. Model (9) is now in triangular system ECM format for discrete time models. Such models have been studied in earlier work by the author (1991, 1988b) and results in those papers will be drawn on below.

2.2. Identification

The continuous time error correction model (5) generates equispaced data that satisfy the analogous discrete time model (9). 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 discrete time cointegrated system induced by (9) is obtained from (2) simply by sampling at integer intervals. The problem of identifying the coefficient matrix B in the cointegrating equation (2) using only discrete data then reduces to the problem of identifying the submatrix B of A in (9). In discrete time these are issues that have been fully explored elsewhere—see Phillips and Durlauf (1986) and Stock (1987a) for details.

It is worth observing how this result on identification bears on the usual aliasing problem. Let $f_{ww}^c(\lambda)$ be the spectral density matrix of the residual process $w(t)$ in (5). Suppose $f_{ww}^c(\lambda)$ is continuous and bounded over the interval $(-\infty, \infty)$. The spectral matrix of $y(t)$ is

$$(10) \qquad f_{yy}^c(\lambda) = (-i\lambda I - EA)^{-1} f_{ww}^c(\lambda)(i\lambda I - A'E')^{-1}$$

and the spectrum of the discrete sequence $\{y(n)\}_1^\infty$ is given by the folding formula $f_{yy}^d(\lambda) = \sum_{j=-\infty}^\infty f_{yy}^c(\lambda + 2\pi j)$. (Note that spectra such as (10), which represent nonstationary $I(1)$ 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 $f_{yy}^c(\lambda + 2\pi j)$ is bounded for all $j \neq 0$ in the vicinity of $\lambda = 0$ whereas when $j = 0$ we have $f_{yy}^c(\lambda) = O(1/\lambda^2)$ as $\lambda \to 0$. Thus, the behavior of the discrete spectrum $f_{yy}^d(\lambda)$ at the origin is prescribed by that of $f_{yy}^c(\lambda)$ as $\lambda \to 0$. This means that we can identify the long-run components that dominate the behavior of $f_{yy}^c(\lambda)$ from the discrete spectrum $f_{yy}^d(\lambda)$. 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 $(1 + \lambda^2)^{-1}Af_{ww}^c(\lambda)A'$. Clearly, A is identified from $f_{yy}^c(\lambda)$ as the linear transformation of $y(t)$ that annihilates the pole at the origin in $f_{yy}^c(\lambda)$. The matrix A is then unique up to normalization. However, since the behavior of $f_{yy}^d(\lambda)$ mirrors the behavior of $f_{yy}^c(\lambda)$ at the origin we may equivalently identify A from $f_{yy}^d(\lambda)$. This eliminates the aliasing problem for the long-run equilibrium parameters in continuous time.

2.3. Filtered Series and Temporal Aggregation

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 spectral matrix of the filtered series $Y(n)$ is $f_{YY}^c(\lambda) = G(\lambda)f_{yy}^c(\lambda)G(\lambda)^*$. From the above analysis we know that $f_{yy}^c(\lambda) = O(1/\lambda^2)$ as $\lambda \to 0$ and that the matrix A for which $Af_{yy}^c(0)A' = Af_{ww}^c(0)A' < \infty$ is uniquely defined up to normalization. The spectral matrix of $AY(n)$ at the origin is $AG(0)f_{yy}^c(0)G(0)'A'$ and this matrix is bounded iff the rows of the matrix $AG(0)$ are spanned by the rows of A, i.e. iff $G(0)'A' \in \mathcal{R}(A')$. This leads us to the following result.

THEOREM: *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*

(11) $G(0)'A' \in \mathscr{R}(A')$.

Condition (11) is necessary and sufficient for the invariance of the cointegrating relationship (1) under the action of the linear filter with response function $G(\lambda)$. Three filters that are of interest in applications are the following.

(i) *Flow data.* Here, $Y(n) = \int_{n-1}^{n} y(t)\,dt$ and $G(\lambda) = h(\lambda)I_m$ with $h(\lambda) = (e^{i\lambda} - 1)/i\lambda$. Observe that $G(0) = I_m$ and (11) holds trivially.

(ii) *Mixed stock and flow data.* Write $y(t)$ in partitioned format as

(12) $y(t) = Q\begin{bmatrix} y^s(t) \\ y^f(t) \end{bmatrix} = Q_s y^s(t) + Q_f y^f(t)$.

Here, the affixes "s" and "f" signify that the associated components are measured as stocks or as flows, respectively. Let there be m_s and m_f components in each category. Q is a permutation matrix which reorders the elements to conform with the earlier format given in (1). Define the filtered series $Y(n) = Q_s y^s(n) + Q_f Y^f(n)$, where $Y^f(n) = \int_{n-1}^{n} y^f(t)\,dt$. Then the response function of the filter is

(13) $G(\lambda) = Q\begin{bmatrix} I_{m_s} & 0 \\ 0 & h(\lambda)I_{m_f} \end{bmatrix} Q'$.

Again, $G(0) = I_m$ and (11) holds trivially, so that cointegrating relationships are preserved under the action of this filter. Stock (1987b) made a similar observation and gave an example for the case $m_s = m_f = 1$.

(iii) *General time averaging filters.* Here, the length (δ) of the time averaging filter may be distinct from the sampling interval of the econometrician, which is set to unity. Weights may also be assigned in averaging the data. In general, we might have $Y_\delta^f(n) = \int_0^\delta w(s)y^f(t-s)\,ds$ with $\int_0^\delta w(s)\,ds = 1$. In place of (13), the response function of the filter that accommodates instantaneous observations (i.e. $y^s(n)$) and time averaged data such as $Y_\delta^f(n)$ is

$G_\delta(\lambda) = Q\begin{bmatrix} I_{m_s} & 0 \\ 0 & h_\delta(\lambda)I_{m_f} \end{bmatrix} Q'$,

where $h_\delta(\lambda) = \int_0^\delta e^{i\lambda s} w(s)\,ds$. When $\lambda = 0$, we obtain $G(0) = I_m$ again and (11) continues to hold.

General time averaging filters like $Y_\delta^f(n)$ arise in models not only because of the manner in which data are actually collected. They can also arise because partial equilibrium formulations in continuous time models may distinguish between the decision making intervals of agents and the time unit that is used in

the model. The cointegrating long-run relationship between the variables is then invariant to the replacement of $Y_\delta^f(n)$ by $y(n)$ or $Y^f(n)$. Ideally, one would like the asymptotic properties of the estimation procedure that is employed to be invariant to this type of replacement. This would be especially helpful in cases where the filter was itself unknown. The methods now to be discussed have this property.

3 ESTIMATION, INFERENCE, AND ASYMPTOTICS

3.1. *Unrestricted Cointegrating Matrices*

In this subsection we shall work with the linear model (9) where the cointegrating matrix A is unrestricted other than by normalization, i.e. the submatrix B is unrestricted. The approach we suggest is to use the discrete time ECM formulation (9) rather than to attempt to estimate the differential equation system (5) directly. Since (9) has the triangular system format, a number of different estimation methods are available including instrumental variables (Phillips and Hansen (1990)), maximum likelihood (Phillips (1991)) and spectral regression (Phillips (1988b).) 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, thereby facilitating the treatment of data irregularities such as the presence of mixed stock and flow data. Spectral 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 (9). We assume that the residual process $x(n)$ in (9) is stationary with spectral matrix $f_{xx}^d(\lambda) > 0$ that is continuous at the origin $\lambda = 0$. We set $\Omega = 2\pi f_{xx}^d(0)$ 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

$$(14) \qquad T^{-1/2} P_{[Tr]} \Rightarrow S(r) \equiv BM(\Omega)$$

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

$$(15) \qquad T^{-1} \Sigma_1^{[Tr]} P_n x(n)' \Rightarrow \int_0^r S\,dS' + r\Delta,$$

where the first term on the right side of (15) is a matrix stochastic integral with respect to the Brownian motion $S(r)$. Results (14) and (15) 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 (e.g., see Phillips and Durlauf (1986)) and are reviewed by the author in (1988a). They certainly apply when $x(n)$ is a linear process of the type

$$(16) \qquad x(n) = \sum_{j=-\infty}^{\infty} C_j \varepsilon_{n-j}; \qquad \{\varepsilon_t\} \equiv \mathrm{iid}(0, \Sigma_\varepsilon), \qquad \sum_{-\infty}^{\infty} j^{1/2} \|C_j\| < \infty.$$

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

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

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 (9). These are simply the matrix extensions of the spectral regression estimators developed for the single equation ECM setting in the (1988b) paper. It will be helpful, therefore, to use the same notation.

Specifically, we define

$$y_*(n)' = (y_1(n)', \Delta y_2(n)'), \qquad w_*(\lambda) = (2\pi T)^{-1/2} \sum_{n=1}^{T} y_*(n)e^{in\lambda},$$

$$w_2(\lambda) = (2\pi T)^{-1/2} \sum_{n=1}^{T} y_2(n-1)e^{in\lambda} \qquad \text{for} \qquad \lambda \in [-\pi, \pi].$$

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 m_1 equations of (9). 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

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

where the summation is over $\lambda_s \in \mathscr{B}_j = (\omega_j - \pi/2M < \lambda \leqslant \omega_j + \pi/2M)$, a frequency band of width π/M centered on $\omega_j = \pi j/M$, $j = -M+1, \ldots, M$ for M integer. Setting $l = [T/M]$, we are effectively averaging l 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 (1987a)) we find that when $\omega_j \to \omega$ we have $\hat{f}_{xx}(\omega_j) \to_p f_{xx}(\omega)$ as $T \to \infty$. This follows because $f_{xx}(\omega)$ is continuous in view of (16), 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:

$$(17) \qquad \text{vec}(\tilde{B}) = \left[\frac{1}{2M} \sum_{j=-M+1}^{M} E'\hat{f}_{xx}(\omega_j)^{-1} E \otimes \hat{f}_{22}'(\omega_j) \right]^{-1}$$
$$\cdot \left[\frac{1}{2M} \sum_{j=-M+1}^{M} \left(E'\hat{f}_{xx}(\omega_j)^{-1} \otimes I \right) \text{vec}\left(\hat{f}_{*2}(\omega_j) \right) \right],$$

where $\hat{f}_{22}(\omega_j) = l^{-1} \sum_{\mathscr{B}_j} w_2(\lambda_2) w_2(\lambda_s)^*$ and $\hat{f}_{*2}(\omega_j) = l^{-1} \sum_{\mathscr{B}_j} w_*(\lambda_2) w_2(\lambda_s)^*$. The band spectral estimator $\tilde{B}_0$ is similarly defined but is based only on spectral estimates at the origin.

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 (9) or the exact discrete model (8), more especially when there are data irregularities to accommodate. Note that the latter methods also require explicit modeling of the error process and, hence, the short run dynamics of the model with the attendant difficulties of model selection.

Both $\tilde{B}$ and $\tilde{B}_0$ rely on an initial estimate of B such as least-squares in order to construct the residual spectral estimate $\hat{f}_{xx}$. $\tilde{B}$ and $\tilde{B}_0$ are therefore two-step estimators. Further iterations are possible and may lead to some improvement in finite sample performance because of the second order bias in first stage estimates like least squares (Phillips and Durlauf (1986), Stock (1987a)). 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 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.

$$(18) \qquad A(\alpha) = A(\alpha^0) \qquad \text{implies} \qquad \alpha = \alpha^0,$$

where α^0 is the true value of α. The (nonlinear) efficient spectral regression estimator of α is

$$\tilde{\alpha} = \text{argmin}_\alpha \sum_{\mathscr{B}} \left[w_*(\lambda_s) - EB(\alpha) w_2(\lambda_s) \right]^*$$
$$\cdot \Phi(\lambda_s) \left[w_*(\lambda_s) - EB(\alpha) w_2(\lambda_s) \right],$$

where $\Phi(\lambda_s) = \hat{f}_{xx}(\omega_j)^{-1}$, $\lambda_s = 2\pi s/T \in \mathscr{B}_j$ and $\mathscr{B} = \cup_{-M+1}^{M} \mathscr{B}_j$. Both $\tilde{\alpha}$ and the corresponding band spectral estimator 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

$$(19) \qquad w_1(\lambda_s) = Bw_2(\lambda_s) + Cw_{\Delta y_2}(\lambda_s) + w_{1 \cdot 2}(\lambda_s).$$

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 $C = \Omega_{12}\Omega_{22}^{-1}$, respectively. Band spectral regression on (19) for $\lambda_s \in \mathscr{B}_0$ is simply multivariate least-squares. When there are no restrictions on B in (19) 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 $\lambda_s \in \mathscr{B}_0$. In effect, the efficiency of least squares on (19) is just a frequency domain version of the result (from Phillips (1991)) that OLS in the time domain on

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

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

The subsystem band spectral estimator of B in (19) has the form

$$(20) \qquad B_0^+ = \left[\hat{f}_{12}(0) - \hat{f}_{1\Delta}(0)\hat{f}_{\Delta\Delta}(0)^{-1}\hat{f}_{\Delta 2}(0) \right]$$

$$\cdot \left[\hat{f}_{22}(0) - \hat{f}_{2\Delta}(0)\hat{f}_{\Delta\Delta}(0)^{-1}\hat{f}_{\Delta 2}(0) \right]^{-1},$$

where we use $\hat{f}_{\Delta\Delta}(\cdot)$ to denote the estimated spectral matrix of $\Delta y_2(n)$ 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 estimates could equally well be employed.

3.4. Asymptotic Theory

Following the approach set out in the (1988b) paper, with some modifications to deal with the multivariate character of the regressions, it is straightforward to derive the asymptotic distribution of $\tilde{B}$, $\tilde{B}_0$, and B_0^+. The estimators are asymptotically equivalent, have the following limit theory, and are optimal under Gaussian assumptions:

$$(21) \qquad T(\tilde{B} - B), T(\tilde{B}_0 - B), T(B_0^+ - B) \Rightarrow \left(\int_0^1 dS_{1 \cdot 2} S_2' \right)\left(\int_0^1 S_2 S_2' \right)^{-1},$$

where

$$\begin{bmatrix} S_{1 \cdot 2} \\ S_2 \end{bmatrix}\begin{matrix} m_1 \\ m_2 \end{matrix} \equiv BM\left(\begin{bmatrix} \Omega_{11 \cdot 2} & 0 \\ 0 & \Omega_{22} \end{bmatrix} \right).$$

The common limit distribution may alternatively be represented in the mixed normal form

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

where $G = (\int_0^1 S_2 S_2')^{-1}$, $g = \int_0^1 (Q_2 W_1)^2$, and $Q_2 W_1 = W_1 - (\int_0^1 W_1 W_2')$ $(\int_0^1 W_2 W_2')^{-1} W_2'$ 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

$$\begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \begin{matrix} 1 \\ m_2 - 1 \end{matrix} \equiv BM(I_{m_2}).$$

As discussed in the (1988b) paper, the nuisance parameters of the limit distribution (22) 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

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

where $h(\cdot)$ 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

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

Similar considerations apply to $\tilde{B}_0$ and B_0^+.

The case of restricted cointegrating matrices $A = A(\alpha)$ may be handled in much the same way. Setting $b = b(\alpha) = \text{vec}(B(\alpha))$, we have

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

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

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

$$\equiv \int_{G>0} N\left(0, \left[J'(\Omega_{11 \cdot 2}^{-1} \otimes G) J \right]^{-1}\right) dP(G).$$

Again, hypothesis testing about α^0 may be conducted using asymptotic chi-squared criteria constructed in the usual way.

4 SOME CONCLUDING REMARKS

The model on which our attention has focused is the first order stochastic differential equation system (5). 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

$$A(D)w(t) = B(D)\zeta(t),$$

where $A(D) = \Sigma_0^p A_i D^i$ and $B(D) = \Sigma_{j=0}^q B_j D^j$ $(q \le p)$ 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 (5) is then the higher order model

$$(5)^1 \qquad A(D)(DI + EA)y(t) = B(D)\zeta(t)$$

in which the coefficients of $A(D)$ and $B(D)$ embody the transient dynamics. The model also accommodates stationary exogenous inputs $z(t)$. These may be absorbed into the generating process for the residuals by writing the complete model for $y(t)$ as

$$(5)^2 \qquad Dy(t) = -EAy(t) + Cz(t) + w(t) = -EAy(t) + \underline{w}(t),$$

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 m_1 equations of) $(5)^2$ 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 could be explored in Monte Carlo work.

The model may also be extended to allow for deterministic as well as stochastic trends. All that is needed is to replace (5) by the system

$$(5)^3 \qquad Dy(t) = k(t) - EAy(t) + w(t),$$

where $k(t) = \Sigma_{i=0}^p k_i t^i$ and $Ak_i = 0$ $(i = 0, \ldots, p)$. 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 $(5)^3$ that $Ay(t)$ then satisfies the stable system

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

The discrete time equivalent of $(5)^3$ is

$$(9)^1 \qquad \Delta y(n) = k_*(n) - EAy(n-1) + x(n)$$

which replaces our earlier equation (9). In $(9)^1$ we find, after a small calculation, that

$$k_*(n) = \sum_{i=0}^p k_i \sum_{j=0}^i \binom{i}{j}\frac{(-1)^{j-i}}{j-i+1}n^j = \sum_{i=0}^p k_{*,i}n^i.$$

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 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. As shown in Section 3, 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 of maximum likelihood estimates 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). Asymptotics of the type discussed here can also be used to help explain some of the empirical differences between the use of the discrete approximation and the exact discrete model in the estimation of continuous time systems. 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. 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 will often lead to an increase in residual variance. These are matters that could be explored in appropriately designed sampling experiments.

Cowles Foundation for Research in Economics, Department of Economics, P.O. Box 2125 Yale Station, Yale University, New Haven, CT 06520-2125, U.S.A.

REFERENCES

BERGSTROM, A R (1988): "The History of Continuous Time Econometric Models," *Econometric Theory*, 4, 365–383

BERGSTROM, A. R., AND C. R. WYMER (1976): "A Model of Disequilibrium Neoclassical Growth and Its Application to the United Kingdom," in *Statistical Inference in Continuous Time Economic Models*, ed by A. R. Bergstrom. Amsterdam: North Holland.

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

——— (1963b): "Regression for Time Series with Errors of Measurement," *Biometrika*, 50, 293–302.

PHILLIPS, P. C. B. (1972): "The Structural Estimation of a Stochastic Differential Equation System," *Econometrica*, 40, 1021–1041.

——— (1988a): "Multiple Regression with Integrated Processes," in *Statistical Inference from Stochastic Processes, Contemporary Mathematics*, ed. by N. U. Prabhu. 80, pp. 79–106.

—— (1988b): "Spectral Regression for Cointegrated Time Series," Cowles Foundation Discussion Paper No. 872, Yale University. Forthcoming in *Nonparametric and Semiparametric Methods in Economics and Statistics*, ed. by W. Barnett. Cambridge: Cambridge University Press, 1991.

—— (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

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 (1990): "Statistical Inference in Instrumental Variables Regression with $I(1)$ Processes," *Review of Economic Studies*, 57, 99–125.

ROBINSON, P. M. (1972): "Nonlinear Vector Time Series Regression," *Journal of Applied Probability*, 9, 758–768.

SOLO, V. (1987): "Intrinsic Random Functions and the Paradox of $1/f$ Noise," Johns Hopkins (mimeographed).

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*.