

Federal Reserve Bank of Minneapolis
Research Department Staff Report 75

October 1981

FORMULATING AND ESTIMATING
CONTINUOUS TIME RATIONAL EXPECTATIONS MODELS¹

Lars Peter Hansen

Carnegie-Mellon University

Thomas J. Sargent

Federal Reserve Bank of Minneapolis
and University of Minnesota

ABSTRACT

This paper proposes a method for estimating the parameters of continuous time, stochastic rational expectations models from discrete time observations. The method is important since various heuristic procedures for deducing the implications for discrete time data of continuous time models, such as replacing derivatives with first differences, can sometimes give rise to very misleading conclusions about parameters. Our proposal is to express the restrictions imposed by the rational expectations model on the continuous time process generating the observable variables. Then the likelihood function of a discrete time sample of observations from this process is obtained. Parameter estimates are computed by maximizing the likelihood function with respect to the free parameters of the continuous time model.

The views expressed herein are solely those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1.INTRODUCTION

This paper describes how to estimate a class of continuous time stochastic rational expectations models from discrete time data. A continuous time model within this class consists of a collection of stochastic differential equations that are linear in the variables but nonlinear in the deep parameters that are to be estimated. These nonlinearities arise from the extensive and complicated cross-equation restrictions that are the hallmark of rational expectations models.

The basic idea behind our estimation method is simply to maximize the Gaussian likelihood function of a sample of discrete time observations with respect to the free parameters of the continuous time model. However, to make this idea operational and practical, two problems have to be overcome. The first problem, which is solved in section two, is to obtain analytically tractable formulas for the equilibrium of the model which impose the cross-equation restrictions in a practical way. Researchers who have worked with discrete time rational expectations models are aware of a variety of more or less complicated procedures for characterizing the cross-equation restrictions. (See Muth [29], Lucas [24], Saracoglu and Sargent [36], Hansen and Sargent [11], Futia [5] and Whiteman [47] for a menu of such procedures.) The method used here is in effect a nontrivial adaptation and extension to continuous time systems of the method used by Futia [6] and Hansen and Sargent [11] in discrete time systems. The

method reported in section two comes as close as is logically possible to giving the solution of the model as an analytic expression of the deep parameters of the model. This is an important virtue since it makes the method computationally much quicker than are alternative methods of computing the solution, which are iterative. Speed in calculating the solution of the model is an essential ingredient in making practical the nonlinear maximum likelihood estimation strategy. Furthermore, the prediction formulas that we present are interesting in themselves because they can be used to support continuous time versions of theoretical work like Futia's [6].

The second problem is to find a reliable and computationally practical way of deducing the restrictions that the continuous time model places on the theoretical values of the discrete time second moments. This problem is solved in section three, where we supply analytic formulas for computing the matrix covariogram and spectral density of the discrete time data as a function of the parameters of the continuous time model. Using these formulas, it is straightforward to form the normal likelihood function of the discrete time data, and various alternative approximations to it.

With these basic problems solved, the next three sections indicate procedures for handling several practical problems that arise in applying the method. Section four describes how to handle the situation in which some of the discrete time data are point-in-time observations, while the remaining discrete time observations are unit averages of the continuous observations. Section five describes two alternative models of nonstationarity

for which it is easily possible to transform nonstationary raw data series into the stationary stochastic processes being modeled here. Section six briefly indicates how, say, weekly and quarterly data can be "pooled" in estimation. Section seven presents a numerical example that illustrates the computational costs involved in using our method.

This paper has a variety of antecedents. The basic philosophy behind estimating stochastic rational expectations models of the general class that we consider is explained by Lucas and Sargent [27], Hansen and Sargent [11], and Sargent [38]. A case for formulating and estimating models in continuous time is made in the work of Sims [39], Geweke [7], and P.C.B. Phillips [32, 33, 34]. Identification of particular versions of the present model is discussed by Hansen and Sargent [14, 15]. Hansen and Sargent [14] describe the role of the cross-equation rational expectations restrictions in solving the "aliasing" identification problem involved in moving from the second moments of the discrete time sampled data to the second moments of the continuous time process. (See Sims [39], Geweke [7], and P.C.B. Phillips [33] for alternative manifestations of the aliasing problem).

Because derivations of various of the results are technically involved, we have relegated the details of several of them to a technical working paper [17], which is available from the authors upon request.

2. COMPUTING THE SOLUTION

a. The Model

A representative member of the class of models that we study is given by the covariance stationary solution of the following system of stochastic differential equations:

$$(1) \quad H(D)y(t) = E_t \{ J(-D)^{-1} [x_1(t) + x_2(t)] \}$$

$$(2) \quad x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} \psi(D) \\ \theta(D) \end{bmatrix} w(t) = \phi(D)w(t)$$

where E_t is the linear least squares forecast operator conditioned on $\{x(v) : v \leq t\}$, and D is the time derivative operator. Here $y(t)$ is an $(n \times 1)$ vector of endogenous variables; $x_1(t)$ is an $(n \times 1)$ vector of variables observed by agents but not by the econometrician; $x_2(t)$ is an $(n \times 1)$ vector of forcing variables observed by both agents and the econometrician; $x_3(t)$ is a $(p-2n \times 1)$ vector of variables observed both by agents and the econometrician and which help predict $x_2(t)$.² Throughout this paper, primes denote transposition but not complex conjugation. The vector $w(t)$ in (2) is a $(p \times 1)$ vector white noise with a generalized covariogram

$$E w(t)w(t-u)' = \delta(t-u)V$$

where V is a positive definite matrix and δ is the Dirac delta generalized function.³ The matrix polynomials in the time

derivative operator D are given by

$$\begin{aligned} J(-D) &= J_0 + J_1(-D) + \dots + J_k(-D)^k \\ H(D) &= H_0 + H_1D + \dots + H_kD^k \\ \psi(D) &= \psi_0 + \psi_1D + \dots + \psi_{m-1}D^{m-1} \\ \theta(D) &= \theta_0 + \theta_1D + \dots + \theta_mD^m \end{aligned}$$

where the J_j 's and H_j 's are $(n \times n)$ matrices, the ψ_j 's are $(p \times p)$ matrices, and the θ_j 's are scalars. Let $\det A$ denote the determinant of the matrix A . We assume that the zeroes of $\det H(s)$ have real parts that are less than zero and that the zeroes of $\det J(-s)$ have real parts that are greater than zero.⁴ Also, we assume that the zeroes of $\theta(s)$ have real parts that are less than zero and that the zeroes of $\det \psi(s)$ have real parts that are less than or equal to zero. An implication of this last assumption is that the error in forecasting $x(t+u)$ from a linear function of $\{x(v) : v \leq t\}$ can be expressed as an integral of $\{w(t+v) : 0 \leq v \leq u\}$. We also note that while w is not a physically realizable process, the restrictions placed on ψ and θ imply that x is physically realizable as long as $\theta_m \neq 0$.

b. Some Economic Examples

We mention several examples of models from the class formed by (1) - (2).

(i) Interrelated factor demand problems.

These are linear quadratic stochastic versions of the models of Mortensen [28] and Treadway [44]. Assume that a competitive

firm maximizes over linear contingency plans for $D^k y(t)$ the expected undiscounted present value criterion:

$$(3) \quad \lim_{T \rightarrow \infty} E_0 \int_0^T \{ [x_1(t) + x_2(t)]' y(t) - y(t)' F_1 y(t) - [G(D)y(t)]' F_2 [G(D)y(t)] \} dt$$

subject to $x(t) = \phi(D)w(t)$, and given $y(0)$, $Dy(0)$, ..., $D^{k-1}y(0)$. Here $y(t)$ is an $(n \times 1)$ vector of stocks of factors of production; $x_1(t)$ is an $(n \times 1)$ vector of productivity shocks; $x_2(t)$ is an $(n \times 1)$ vector of real factor rentals; $G(D) = G_0 + G_1 D + \dots + G^k D^k$ is a generalized "cost of adjustment" matrix polynomial; and F_1 and F_2 are positive definite matrices. To match this setup with (1) and (2), we first note that the characteristic polynomial of the Euler equation⁵ for the certainty equivalent problem associated with (3) is $[F_1 + G(-s)' F_2 G(s)]$. It is known that there exists a factorization of the polynomial $[F_1 + G(-s)' F_2 G(s)] = C(-s)' C(s)$ where the zeroes of $\det C(s)$ are less than zero in real part.⁶ The factorization is unique up to premultiplication of $C(s)$ by an orthogonal matrix. Then the solution of the firm's maximum problem is given by (1) with $H(D) = C(D)$ and $J(-D)' = C(-D)'$.

(ii) Continuous time versions of Lucas-Prescott equilibrium models of investment under uncertainty.

By combining the observations of the previous example with example (ii) in Hansen and Sargent [12], one obtains a multiple factor, continuous time stochastic version of Lucas and Prescott's [25] model.

(iii) Cagan's model of hyperinflation.

In (1) set $n=1$ and let $y(t)$ be the logarithm of the price level, $x_1(t)$ be a disturbance to portfolio balance, and $x_2(t)$ be the logarithm of the money supply. Then take $H(D) = 1$, and $J(-D) = (1 + \alpha D)$ where $\alpha < 0$ is the slope of the portfolio balance schedule with respect to expected inflation. Then (1) and (2) form a continuous time, rational expectations version of Cagan's [2] model. For further discussion of this example, see Hansen and Sargent [13].

(iv) Stochastic versions of Dornbusch's model of exchange rate "overshooting".

By referring to Wilson [48] and Dornbusch [4] and using example (iii), the reader can readily construct this example.

(v) Continuous time versions of Taylor's "staggered contract" models.

Models that come from stochastic linear-quadratic optimum problems have $J(D) = H(D)$ in (1), as in example (i). However, in a discrete time framework, Taylor [42, 43] has described a class of models that represent timing features of wage contracting processes in which there is no such symmetry between the analogues of the feedback function $H(D)$ and the feedforward function $J(-D)$. In continuous time, these functions of D will not be simple polynomials. Nonetheless, our methods can be modified in a straightforward manner to accommodate continuous time versions of Taylor's models.

Many other examples could be formed by appropriately modifying various of the discrete time linear rational expectations models described by Hansen and Sargent [13] and Sargent [37, 38].

c. The Solution

The solution to the forecasting problem on the right-hand-side of (1) is a complicated function of the parameters of $J(-D)$, $\theta(D)$, and $\psi(D)$. To deduce from (1) the implied second order properties of the y process, it is necessary to solve explicitly for an expression of the form

$$E_t J(-D)^{-1} [x_1(t) + x_2(t)] = \int_0^{\infty} L(\tau) w(t-\tau) d\tau$$

where $L(\tau)$ is a function of the parameters of $J(D)$, $\phi(D)$, and $\theta(D)$, and thereby embodies the rational expectations cross-equation restrictions. In effect, we now describe how to calculate $L(\tau)$, proceeding in two steps.

First, we obtain a matrix partial fractions representation of $J(-s)^{-1}$. We assume that the zeroes of $\det J(s)$ are distinct, so that $\det J(s) = \rho_0 (s-\rho_1)(s-\rho_2) \dots (s-\rho_k)$ where $k = \text{gn}$. Then we have

$$(4) \quad J(s)^{-1} = \frac{B_1}{s-\rho_1} + \dots + \frac{B_k}{s-\rho_k}$$

where

$$B_j = \frac{\text{adj} [J(\rho_j)]'}{\rho_0 \prod_{\substack{h=1 \\ h \neq j}}^k (\rho_j - \rho_h)}$$

and where $\text{adj}(A)$ denotes the adjoint of the matrix A .⁷

Since the zeroes of $\det J(-s)$ have been assumed to have positive real parts, the ρ_j 's, which are zeroes of $\det J(s)$, therefore have negative real parts. Thus, from (4) we have

$$(5) \quad J(-s)^{-1} = \frac{B_1}{-s-\rho_1} + \dots + \frac{B_k}{-s-\rho_k},$$

where

$$\text{Re}(\rho_j) < 0 \text{ for } j = 1, \dots, k$$

and where $\text{Re}(s)$ denotes the real part of the complex number s .

Now substitute (5) into (1) to obtain

$$(6) \quad H(D)y(t) = E_t \sum_{j=1}^k \frac{B_j}{-D-\rho_j} [x_1(t) + x_2(t)].$$

Recall that the Laplace transform of $\frac{1}{-s-\rho}$ for $\text{Re}(\rho) < 0$ is given by the function f where

$$f(u) = \begin{cases} e^{-\rho u} & u \leq 0 \\ 0 & u > 0. \end{cases}$$

Therefore (6) is equivalent with

$$(7) \quad H(D)y(t) = \sum_{j=1}^k B_j E_t \int_0^{\infty} e^{\rho_j u} [x_1(t+u) + x_2(t+u)] du.$$

Thus, the first step of using the partial fraction representation of $J(-D)$ leads to (7).

The second step in obtaining an operational solution for $y(t)$ is to determine expressions for the forecasts

$$E_t \int_0^{\infty} e^{\rho_j u} [x_1(t+u) + x_2(t+u)] du.$$

Hansen and Sargent [17] give the formula:⁸

$$(8) \quad E_t \int_0^{\infty} e^{\rho_j u} x(t+u) du = \left[\frac{-\phi(D) + \phi(-\rho_j)}{D + \rho_j} \right] w(t) .$$

It is easily verified that the rational function $\frac{-\phi(s) + \phi(-\rho_j)}{s + \rho_j}$

has a removable singularity at $-\rho_j$ and has lowest common denominator $\theta(s)$.

Now let M be an $(n \times p)$ matrix such that $Mx(t) = x_1(t) + x_2(t)$. Then substituting (8) into (7) gives the representation⁹

$$(9) \quad H(D)y(t) = \sum_{j=1}^k B_j M \left[\frac{-\phi(D) + \phi(-\rho_j)}{s + \rho_j} \right] w(t) .$$

With (9), we have completed our first task, that of solving the system of differential equation (1) and (2). Equations (9) and (2) completely characterize the cross-equation restrictions implied by rational expectations. Further, (10) is as close to being an analytical, closed form for the solution as it is possible to obtain. The only step that is not analytic, that of calculating the zeroes of $\det J(s)$ in forming the partial fractions representation (4), simply cannot be avoided.

The formula (8) is a closed form representation of the cross-equation restrictions. It extends the results of Futia [6] and Hansen and Sargent [11, 12] to continuous time systems. It is straightforward to evaluate, and is useful not only for estimation purposes but also for theoretical investigations such as Futia's [6].

d. A Model of the Error Term

As often happens in rational expectations models, the solution (9) expresses $y(t)$ as an exact function of current and past x 's. This implies that the spectral density matrix of the continuous time (y, x) process is singular. In order to induce a nonsingular spectral density matrix for the resulting discrete time process, it is necessary to resort to some device that restricts the information set of the econometrician relative to the information possessed by the agent who is being modeled.

At this point, in the interests of constructing a tractable model of the disturbance term, we further restrict the specification of the stochastic process governing x . We define

$$z(t) = \begin{bmatrix} x_2(t) \\ x_3(t) \end{bmatrix} \quad \text{and} \quad x_t = \begin{bmatrix} x_1(t) \\ z(t) \end{bmatrix} .$$

We partition $\theta(D)$, $\psi(D)$, and $w(t)$ conformably with the partitioning of $x(t)$, and assume that

$$(10) \quad w(t) = \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}$$

$$\phi(D) = \begin{bmatrix} \psi^1(D) & 0 \\ \frac{\psi^1(D)}{\theta^1(D)} & 0 \\ 0 & \frac{\psi^2(D)}{\theta^2(D)} \end{bmatrix} = \begin{bmatrix} \phi^1(D) & 0 \\ 0 & \phi^2(D) \end{bmatrix}$$

where $\psi^1(D)$ is an $(n \times n)$ operator, $\psi^2(D)$ is a $[(p-n) \times (p-n)]$ operator, and $\theta^1(D)$ and $\theta^2(D)$ are scalar operators. We assume that $\psi^1(s)$ is an $(m_1-1)^{\text{th}}$ order matrix polynomial, $\theta^1(s)$ is an m_1^{th} order scalar polynomial, $\psi^2(s)$ is an $(m_2-1)^{\text{th}}$ order matrix polynomial, and $\theta^2(s)$ is an m_2^{th} order scalar polynomial where $m_1 + m_2 = m$, and $\theta(s) = \theta^1(s)\theta^2(s)$. Equation (10) together with the conditions on the location of the poles and zeroes of $\det\phi(s)$ imply that $w_1(t)$ is contained in the space spanned by $\{x_1(u) : u \leq t\}$, and that $w_2(t)$ is contained in the space spanned by $\{z(u) : u \leq t\}$. Therefore, w_1 is fundamental for x_1 and w_2 is fundamental for z . It then follows from (10) that for all t and v , with $v \geq t$,

$$(11) \quad E[x_1(t+v) | x_1(u), z(u) \text{ for } u \leq t] = E[x_1(t+v) | x_1(u) \text{ for } u \leq t]$$

$$(12) \quad E[z(t+v) | x_1(u), z(u) \text{ for } u \leq t] = E[z(t+v) | z(u) \text{ for } u \leq t].$$

Equation (11) asserts that z fails for to Granger-cause x_1 , while (12) asserts that x_1 fails to Granger-cause z . Recall, however, that, $Ew(t)w(t-u)' = \delta(t-u)V$ where V is permitted to have non-zero off-diagonal elements, so that $x_1(t)$ and $z(t)$ can be correlated.

We now display the form of the solution (9) that incorporates the special assumptions (10) about $\theta(D)$ and $\psi(D)$. Let \tilde{M} be an $[n \times (p-n)]$ matrix such that $\tilde{M}z(t) = x_2(t)$. Then the solution for $y(t)$ can be written

$$(13) \quad H(D)y(t) = \sum_{j=1}^k B_j \tilde{M} \left[\frac{-\phi^2(D) + \phi^2(-\rho_j)}{s + \rho_j} \right] w_z(t) \\ + \sum_{j=1}^k B_j \left[\frac{-\phi^1(D) + \phi^1(-\rho_j)}{s + \rho_j} \right] w_1(t) .$$

The stochastic differential equation (13) and the equation for the observable forcing variables

$$z(t) = \phi^2(D) w_z(t)$$

form a system of linear stochastic differential equations in (y, z) that is driven by the vector white noise (w_1, w_z) . The system can be written compactly as

$$(14) \quad \begin{bmatrix} K_1(D) & 0 \\ 0 & K_2(D) \end{bmatrix} \begin{bmatrix} y(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} Q_{11}(D) & Q_{12}(D) \\ 0 & Q_{22}(D) \end{bmatrix} \begin{bmatrix} w_1(t) \\ w_z(t) \end{bmatrix}$$

or

$$K(D) \begin{bmatrix} y(t) \\ x(t) \end{bmatrix} = Q(D) \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}$$

where

$$K_1(D) = H(D)$$

$$K_2(D) = I$$

$$Q_{11}(D) = \sum_{j=1}^k B_j \left[\frac{-\phi^1(D) + \phi^1(\rho_j)}{D + \rho_j} \right]$$

$$(15) \quad Q_{12}(D) = \sum_{j=1}^k B_j \tilde{M} \left[\frac{-\phi^2(D) + \phi^2(\rho_j)}{D + \rho_j} \right]$$

$$Q_{22}(D) = \phi^2(D) = \frac{\psi^2(D)}{\theta^2(D)}$$

$$\phi^1(D) = \frac{\psi^1(D)}{\theta^1(D)}$$

$$B_j = \frac{\text{adj}[J(\rho_j)]'}{\rho_{o_{h-1}}^k (\rho_h - \rho_j)} \quad h \neq j$$

$$\det J(D) = \rho_o (s - \rho_1) \dots (s - \rho_k) .$$

Note that $K_1(s)$ is a finite order matrix polynomial in s whereas $Q_{11}(s)$, $Q_{12}(s)$ and $Q_{22}(s)$ are rational matrix functions of s . The matrix function $Q_{11}(s)$ has poles at the zeroes of $\theta^1(s)$ and the matrix functions $Q_{12}(s)$ and $Q_{22}(s)$ have poles at the zeroes of $\theta^2(s)$.

The restrictions (15) on the equation system (14) are across equations as manifested by the link between Q_{12} and Q_{22} . The free parameters of the model are the free parameters of the polynomials $H(s)$, $J(s)$, $\theta^1(s)$, $\theta^2(s)$, $\psi^1(s)$, $\psi^2(s)$ as well as the free parameters of the intensity matrix V . The model is linear in the variables $[y(t)']$, $[z(t)']$ that are observable to the econometrician, but is highly nonlinear in the free parameters of $H(D)$, $J(D)$, $\theta(D)$, and $\psi(D)$. For reasons indicated by Lucas and Sargent [27] and Hansen and Sargent [11], it is necessary to

estimate these deep parameters in order to overcome Lucas's [24] critique of procedures for econometric policy evaluation.

3. DEDUCING IMPLICATIONS FOR DISCRETE TIME DATA

We assume that the econometrician has observations on (y, z) sampled at the integers, but does not possess observations on $(D^j y, D^j z)$ for any positive j at any point in time. Since the continuous time model for (y, z) characterized by (14) and (15) involves derivatives and convolution integrals of (y, z) , the econometrician faces a massive problem of systematically missing data. The estimation procedure that we propose simply involves maximizing the likelihood function of a record of point-in-time sampled discrete data (y_t, z_t) for $t = 1, \dots, T$, where the maximization is carried out with respect to the free parameters of the continuous time model in $H(D)$, $J(D)$, $\theta^1(D)$, $\theta^2(D)$, $\psi^1(D)$, $\psi^2(D)$ and V . From the examples analyzed by Hansen and Sargent [14], there are reasons to hope that the continuous time parameters are identified from the likelihood function of the discrete time data.¹⁰

Define the covariogram of the (y, z) process to be the $(p \times p)$ matrix function

$$R(\tau) = E \begin{bmatrix} y(t) \\ z(t) \end{bmatrix} \begin{bmatrix} y(t-\tau) \\ z(t-\tau) \end{bmatrix}',$$

where τ is a real number. The spectral density of the (y, z) process is defined as the Fourier transform of R , namely,

$$S(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau) d\tau.$$

To solve for the spectral density matrix function, we first obtain a moving average representation for the (y, z) process by applying $K(D)^{-1}$ to both sides of (14). More specifically, write

$$H(D)^{-1} = \frac{\text{adj}H(D)}{\text{det}H(D)}.$$

Then

$$(16) \begin{bmatrix} y(t) \\ z(t) \end{bmatrix} = P(D) \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}$$

where

$$P(D) = \begin{bmatrix} P_{11}(D) & P_{12}(D) \\ 0 & P_{22}(D) \end{bmatrix}$$

$$P_{11}(D) = \frac{\text{adj}H(D)}{\text{det}H(D)} Q_{11}(D)$$

$$P_{12}(D) = \frac{\text{adj}H(D)}{\text{det}H(D)} Q_{12}(D)$$

$$P_{22}(D) = Q_{22}(D).$$

Note that $P_{11}(s)$ has poles at the zeroes of $\text{det}H(s)$ and at the zeroes of $\theta^1(s)$; P_{12} has poles at the zeroes of $\text{det}H(s)$ and at the zeroes of $\theta^2(s)$; and $P_{22}(s)$ has poles at the zeroes of $\theta^2(s)$. We assume that the product $\theta^1(s)\theta^2(s)\text{det}H(s)$ can be factored

$$\theta^1(s)\theta^2(s)\text{det}H(s) = \lambda_0(s-\lambda_1)(s-\lambda_1) \dots (s-\lambda_r)$$

where the λ_j 's are distinct for $j = 1, \dots, r$. We note that for many problems, such as those in which (1) is an Euler equation, the zeroes of $\det H(s)$ have to be calculated in evaluating Q_{11} and Q_{12} . So sometimes calculating the zeroes of $\det H(s)$ is not much of an additional computational burden, over and above that involved in computing the solution (15).

The moving average representation (16) can be used to obtain a formula of the spectral density matrix of the observables. It can be shown (see Phillips [30] or Kwakernaak and Sivan [21]) that S is given by

$$(17) S(\omega) = P(i\omega)VP(-i\omega)' .$$

Equation (16) provides a convenient expression for calculating the spectral density of the continuous time process (y, z) . To obtain the likelihood function of the discrete sampled data, in effect we require an expression for the spectral density of the discrete time data. The discrete time, point-in-time sampled data has spectral density S^d , which is related to S by the "folding formula":

$$(18) S^d(\omega) = \sum_{j=-\infty}^{\infty} S(\omega+2\pi j), \quad -\pi \leq \omega \leq \pi .$$

(For example, see Koopmans [20]). For the econometric applications that we have in mind, creating S^d numerically by using (18) is perhaps feasible, but is more expensive and less reliable than the following procedure, which builds on the results of A. W. Phillips [31].

Define the matrices:

$$(19) W_j = \lim_{s \rightarrow \lambda_j} P(s)VP(-s)'(s-\lambda_j)$$

for $j = 1, 2, \dots, r$. A partial fractions representation of S in (17) is then

$$(20) S(\omega) = \sum_{j=1}^r \left[\frac{W_j}{(i\omega - \lambda_j)} + \frac{W_j'}{(-i\omega - \lambda_j)} \right].$$

Since the real part of λ_j is negative, the inverse Fourier transform of $W_j^*/(i\omega - \lambda_j)$ is given by the function

$$\begin{cases} W_j e^{\lambda_j \tau} & \text{for } \tau \geq 0 \\ 0 & \text{for } \tau < 0. \end{cases}$$

Therefore, taking the inverse Fourier transform of each side of (20) gives

$$(21) R(\tau) = \begin{cases} \sum_{j=1}^r W_j e^{\lambda_j \tau} & \text{for } \tau \geq 0 \\ \sum_{j=1}^r W_j' e^{-\lambda_j \tau} & \text{for } \tau < 0. \end{cases}$$

By sampling (21) at integer τ , we obtain the covariogram of the discrete time process $\{(y_t, z_t) : t = 0, \pm 1, \pm 2, \dots\}$. The essential element in writing down the likelihood function of the discrete time data as a function of the model's free parameters is

the ability to represent R sampled at the integers as a function of those free parameters. Expression (21) and the steps leading up to it accomplish that task.

It will also prove useful to have another expression for the spectrum of the sampled (y_t, z_t) process. To derive it, let $\alpha_j = e^{\lambda_j}$ and write (21) sampled at the integers as

$$(22) R(\tau) = \begin{cases} \sum_{j=1}^r W_j (\alpha_j)^\tau & \text{for } \tau \geq 0 \\ \sum_{j=1}^r W_j (\alpha_j)^{-\tau} & \text{for } \tau < 0. \end{cases}$$

Define the covariance generating function

$$(23) g(\zeta) = \sum_{\tau=-\infty}^{+\infty} R(\tau) \zeta^\tau .$$

Using (23), we readily obtain

$$(24) g(\zeta) = \sum_{j=1}^r W_j \frac{1}{1 - \alpha_j \zeta} + \sum_{j=1}^r W_j \frac{\zeta^{-1} \alpha_j}{1 - \alpha_j \zeta^{-1}}$$

Evaluating (24) at $\zeta = e^{-i\omega}$ gives the spectral density of the integer sampled process (y_t, z_t) , $S^d(\omega) = g(e^{-i\omega})$.

With these results in hand, we can now indicate how to construct the likelihood function for a set of observations on (y_t, z_t) , $t = 1, \dots, T$, assuming that w is a Gaussian process. Define the stacked vector of observations on (y_t, z_t) , $t = 1, \dots, T$ as

$$(25) \bar{y}_T = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_T \end{bmatrix} \quad \text{and} \quad \bar{z}_T = \begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ z_T \end{bmatrix}$$

Define the covariance matrix of $[\bar{y}_T', \bar{z}_T']$.

$$(26) \Gamma_T = E \begin{bmatrix} \bar{y}_T \\ \bar{z}_T \end{bmatrix} [\bar{y}_T' \quad \bar{z}_T']$$

The $(pT \times pT)$ covariance matrix Γ_T can be computed as a function of the free parameters of the continuous time model $H(D)$, $J(D)$, $\theta^1(D)$, $\theta^2(D)$, $\psi^1(D)$, $\psi^2(D)$ and V by using (25) and (26) and the results leading up to (22). The normal log likelihood for $[\bar{y}_T', \bar{z}_T']$ is given by

$$(27) L_T = -\frac{1}{2} Tp \log 2\pi - \frac{1}{2} \log \det \Gamma_T - \frac{1}{2} [\bar{y}_T' \quad \bar{z}_T'] (\Gamma_T)^{-1} \begin{bmatrix} \bar{y}_T \\ \bar{z}_T \end{bmatrix}.$$

The log likelihood function (27) is to be maximized with respect to the free parameters of $H(D)$, $J(D)$, $\phi^1(D)$, $\phi^2(D)$, and V of the continuous time model. These parameters make their appearance in (27) through the covariance matrix Γ_T .

The maximization of (27) must be achieved by the application of numerical procedures, such as the "acceptable gradient" methods described by Bard [1]. From the standpoint of these iterative hill-climbing procedures, (27) is a formidable function because the $(pT \times pT)$ matrix Γ_T must be inverted each time (27) is evaluated for different points in the parameter space. Since the

matrix Γ_T is liable to be very large, this difficulty has lead researchers such as Hannan [10] and Phadke and Kadem [30] to propose frequency domain approximations to the normal likelihood function that economize on computations. Let the periodogram of the (y_t, z_t) process at frequency $\omega_j = 2\pi j/T$, $j=1, \dots, T$ be the $(p \times p)$ matrix $I(\omega_j)$. The approximations used are then

$$\begin{aligned} [\bar{y}'_T \quad \bar{z}'_T] (\Gamma_T)^{-1} \begin{bmatrix} \bar{y}_T \\ \bar{z}_T \end{bmatrix} &\approx \sum_{j=1}^T \text{trace} [S^d(\omega_j)^{-1} I(\omega_j)] \\ \log \det \Gamma_T &\approx \sum_{j=1}^T \log \det S^d(\omega_j) . \end{aligned}$$

Substituting these approximations into (27) gives the approximate log likelihood function

$$(28) \quad \begin{aligned} L_T^* &= -\frac{1}{2} T p \log 2\pi - \frac{1}{2} \sum_{j=1}^T \log \det S^d(\omega_j) \\ &\quad - \sum_{j=1}^T \text{trace} [S^d(\omega_j)^{-1} I(\omega_j)] . \end{aligned}$$

Equation (24) evaluated at $\tau = e^{-i\omega_j}$ and the results leading up to it express $S^d(\omega_j)$ as a function of the free parameters of $H(D)$, $J(D)$, $\theta(D)$, $\psi(D)$, and V of the continuous time model. By maximizing (28) instead of (27) the analyst avoids the need to invert the $(pT \times pT)$ matrix Γ_T at each function evaluation, and instead has to invert the $(p \times p)$ matrix $S^d(\omega_j)$. Expression (28) is a good approximation to (27) in the sense that maximizing it delivers estimators that are asymptotically equivalent to those obtained by maximizing (27).

We now very briefly describe an alternative time domain approximation of the likelihood function. By pursuing the calculations described by A. W. Phillips [31], it is possible analytically to deduce from representation (14) the vector mixed autoregressive, moving average process that governs the discrete time sampled process (y, z) , say

$$\tilde{K}(L) \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \tilde{Q}(L) \varepsilon_t$$

where ε_t is a discrete time vector white noise, and $\tilde{K}(L)$ and $\tilde{Q}(L)$ are each finite order matrix polynomials that are one-sided nonnegative powers of L . From $K(D)$ and $Q(D)$ and (14) it is possible analytically to calculate $\tilde{K}(L)$ and $\tilde{Q}(L)$, by following the suggestions of Phillips [31]. In most circumstances, $\tilde{K}(L)$ and $\tilde{Q}(L)$ can be chosen so that they are invertible in which case the suggestions in Hansen and Sargent [11, pp 28-30] can be used to approximate the likelihood function (27) by a likelihood function conditional on the same numbers of initial values of y and z and of ε as there are powers of L in $\tilde{K}(L)$ and $\tilde{Q}(L)$, respectively. Given $\tilde{K}(L)$ and $\tilde{Q}(L)$, it is straightforward to evaluate this approximate likelihood function. Once again, the approximate likelihood function is to be maximized with respect to the free parameters of $H(D)$, $J(-D)$, $\theta(D)$, $\psi(D)$ and V of the continuous time model.

The approximations described above are known to deteriorate when either $\det\tilde{K}(z)$ or $\det\tilde{Q}(z)$ have zeroes that are close to the unit circle. Hillmer and Tiao [19] have deduced a representation

of the exact likelihood function (27) in which inversion of the $(pT \times pT)$ matrix Γ_T is circumvented. Associated with their representation is a time domain approximation that permits the $\det \tilde{Q}(\zeta)$ to have zeroes that are on the unit circle.

4. TIME AVERAGED DATA

The procedures of the preceding sections assume that the discrete time data are point-in-time observations on the underlying continuous time data. Often, however, one or more of the available series consist of unit averaged data, which correspond to integrals of continuous flows over a month or a quarter, for example. Observations on GNP, sales, and manhours are usually recorded in this way. In this section, we tell how the preceding method can be modified to accommodate the situation in which some or all of the data are unit averages.

We can completely indicate the modifications required by supposing that y and z are both scalar processes, so that R is a (2×2) matrix function of τ with

$$R(\tau) = \begin{bmatrix} r_{11}(\tau) & r_{12}(\tau) \\ r_{21}(\tau) & r_{22}(\tau) \end{bmatrix} = E \begin{bmatrix} y(t) \\ z(t) \end{bmatrix} \begin{bmatrix} y(t-\tau) \\ z(t-\tau) \end{bmatrix}' .$$

Recall representation (21) for R ,

$$(21) \quad R(\tau) = \sum_{j=1}^r W_j e^{\lambda_j \tau} , \quad \tau \geq 0$$

where the W_j 's are functions of the deep parameters of the model. Suppose that the discrete data on z are point-in-time while those on y are unit averaged, namely,

$$\bar{y}(t) = \int_0^1 y(t-s) ds \quad .$$

Define the cross-covariogram of the joint (\bar{y}, z) process as

$$\begin{aligned} \bar{R}(\tau) &= E \begin{bmatrix} \bar{y}(t) \\ z(t) \end{bmatrix} \begin{bmatrix} \bar{y}(t-\tau) \\ z(t-\tau) \end{bmatrix}' \quad \text{for integer } \tau \\ &= \begin{bmatrix} \bar{r}_{11}(\tau) & \bar{r}_{12}(\tau) \\ \bar{r}_{21}(\tau) & \bar{r}_{22}(\tau) \end{bmatrix} . \end{aligned}$$

Evaluation of the terms in \bar{R} will indicate how to compute the discrete time autocovariogram and spectral density of a general $(n \times 1)$ vector process, some of whose members correspond to point-in-time observations, while others are unit averaged.

Hansen and Sargent [17] derive the following formulas¹¹

$$\bar{r}_{11}(0) = \sum_{j=1}^r \frac{2}{\lambda_j} \left[\frac{\alpha_j^{-1}}{\lambda_j} - 1 \right] W_j^{11}$$

(29)

$$\bar{r}_{11}(\tau) = \sum_{j=1}^r V_j^{11} (\alpha_j)^\tau, \quad \tau > 0$$

where

$$V_j^{11} = W_j^{11} \begin{bmatrix} \frac{\alpha_j^{-1}}{\lambda_j} \\ \frac{(-\alpha_j)^{-1+1}}{\lambda_j} \end{bmatrix}$$

$$\alpha_j = e^{\lambda_j}$$

$$W_j = \begin{bmatrix} W_j^{11} & W_j^{12} \\ W_j^{21} & W_j^{22} \end{bmatrix} .$$

The z transform of the sequence $\{\bar{r}_{11}(\tau)\}_{\tau=-\infty}^{+\infty}$ is given by

$$(30) \quad \bar{g}_{11}(z) = \sum_{j=1}^r \frac{V_j^{11} \alpha_j z}{(1-\alpha_j z)} + \sum_{j=1}^r \frac{V_j^{11} \alpha_j z^{-1}}{(1-\alpha_j z^{-1})} \\ + \sum_{j=1}^r \frac{2}{\lambda_j} \left[\frac{\alpha_j^{-1}}{\lambda_j} - 1 \right] W_j^{11} .$$

Substituting $e^{-i\omega}$ for z in (30) gives the discrete time spectral density for the unit averaged process \bar{y} .

Hansen and Sargent [17] also show that

$$(31) \quad \bar{r}_{21}(\tau) = \begin{cases} \sum_{j=1}^r W_j^{21} (\alpha_j)^\tau \frac{(\alpha_j^{-1})}{\lambda_j} & \tau \geq 0 \\ \sum_{j=1}^r W_j^{12} (\alpha_j)^{-\tau} \frac{(\alpha_j^{-1})}{\lambda_j} & \tau < 0 . \end{cases}$$

The z transform of $\{\bar{r}_{21}(\tau)\}_{\tau=-\infty}^{+\infty}$ is

$$(32) \quad \bar{g}_{21}(z) = \sum_{j=1}^r \frac{W_j^{21} \alpha_j z}{(1-\alpha_j z)} \frac{(\alpha_j^{-1})}{\lambda_j} + \sum_{j=1}^r \frac{W_j^{12} \alpha_j z^{-1} (\alpha_j^{-1})}{(1-\alpha_j z^{-1}) \lambda_j^{-1}} \\ + \sum_{j=1}^r W_j^{21} \frac{(\alpha_j^{-1})}{\lambda_j} .$$

Substituting $e^{-i\omega}$ for z in (32) gives the discrete time cross-spectral density between the point-in-time data z and the unit averaged data \bar{y} .

Equations (29) and (30) are readily generalized to calculate the cross-covariogram and cross-spectral density between two variables, each of which is a unit average. Equations (31) and (32) give the typical form for the discrete time cross-covariogram and cross-spectrum between two series, the first of which is unit averaged while the second is point-in-time.

With formulas (29) - (32) in hand, the estimation strategy advocated in section three can be executed for a discrete time data set which is any arbitrary mixture of series, some of which are unit averages while others are point-in-time observations.

5. MODELS OF NONSTATIONARITY

We briefly describe two alternative interpretations of the model formed by (1) and (2) or (14) in which the $[y(t), x(t)]$ variables are interpreted as transformations of the underlying variables of interest, which are themselves nonstationary. Turning to the first model of nonstationarity, we reproduce (1), (2), and (14) for convenience:

$$(1) \quad H(D)y(t) = E_t J(-D)^{-1} [x_1(t) + x_2(t)]$$

$$(2) \quad \theta(D) \begin{bmatrix} x_1(t) \\ z(t) \end{bmatrix} = \psi(D)w(t)$$

$$(14) \quad K(D) \begin{bmatrix} y(t) \\ z(t) \end{bmatrix} = Q(D)w(t)$$

where $z(t)' = [x_2(t)', x_3(t)']$.

Our first model of nonstationarity assumes that $y(t)$, $x(t)$ and $w(t)$ are themselves the following transformations of the continuous time records of the underlying raw data series $\hat{y}(t)$, $\hat{x}(t)$, and the white noise $\hat{w}(t)$:

$$y(t) = e^{-\frac{\gamma}{2}t} \hat{y}(t)$$

$$(33) \quad x(t) = e^{-\frac{\gamma}{2}t} \hat{x}(t) ,$$

$$w(t) = e^{-\frac{\gamma}{2}t} \hat{w}(t) , \quad \gamma > 0.$$

From (33) it follows that \hat{w} is a vector white noise with generalized covariogram

$$E \hat{w}(\tau) \hat{w}(t-s)' = V e^{\gamma t} \delta(t-s)$$

The continuous time model in terms of the original variables can be represented either as ¹²

$$(1') \quad H(D - \frac{\gamma}{2}) \hat{y}(t) = E_t J(-D - \frac{\gamma}{2})^{-1} [\hat{x}_1(t) + \hat{x}_2(t)]$$

$$(2') \quad \theta(D - \frac{\gamma}{2}) \hat{x}(t) = \psi(D - \frac{\gamma}{2}) \hat{w}(t)$$

or as

$$(14') \quad K(D - \frac{\gamma}{2}) \begin{bmatrix} \hat{y}(t) \\ \hat{x}(t) \end{bmatrix} = Q(D - \frac{\gamma}{2}) \hat{w}(t).$$

With this model of nonstationarity, the idea is simply to transform the raw data $[\hat{y}(t), \hat{x}(t)]$ according to (33) at the integer points in time, and then to proceed with estimation as described in section three.

The second model of nonstationarity assumes, for example, that the stationary model (1) - (2) or (14) applies to the j th time derivatives of the raw series in continuous time. Thus, let the raw series in continuous time be $[Y(t), Z(t)]$ so that

$$\begin{aligned} y(t) &= D^j Y(t) \\ z(t) &= D^j Z(t). \end{aligned}$$

We suppose that we have integer-sampled, point-in-time data on the raw series $\{(Y_t, Z_t): t=1, \dots, T\}$. The continuous time autocovariogram of the j th derivative series $\{y(t), z(t)\}$ is given by representation (21).

$$(21) \quad R(\tau) = \sum_{j=1}^r W_j e^{\lambda_j \tau}, \quad \tau \geq 0.$$

From (21) and the sort of calculations in the previous section, we can deduce how the continuous time model places restrictions on the discrete time covariance stationary process defined by the j th differenced series $(\Delta^j Y, \Delta^j Z)$ where Δ is the unit finite difference operator defined by $\Delta Y(t) = Y(t) - Y(t-1)$. The main idea can be illustrated for the case in which $j=1$, so that (y, z) correspond to first derivatives of the original data. Our assumptions imply that (y, z) is a covariance stationary process, and that $(DY, DZ) = (y, z)$. It follows, for example, that

$$Y(t) = \int_0^t y(s) ds + Y(0),$$

for any real t . For positive integers t , the above equality can be expressed

$$Y(t) = \sum_{j=1}^t \int_{j-1}^j y(s) ds + Y(0) .$$

Subtracting the analogous equation for $Y(t-1)$ from the above equality gives

$$Y(t) - Y(t-1) = \int_0^1 y(t-s) ds .$$

Thus, the first differences of the original data Y are unit averages of the first derivative series y . Since the unit average of a covariance stationary series is covariance stationary, it follows that ΔY is covariance stationary.

Formulas (29) and (30) can be used to deduce the spectral density matrix and cross-covariogram for the first-differenced raw series $(\Delta Y, \Delta Z)$. Estimation can then proceed along the lines indicated in the preceding section.

This idea generalizes to the case in which $(y, z) = (D^j Y, D^j Z)$ for $j > 1$. For example, where $j=2$, the second difference of the raw data Y are covariance stationary and obey

$$\Delta^j Y(t) = \int_0^2 b(s) y(t-s) ds$$

where

$$b(s) = \begin{cases} s, & 0 \leq s < 1 \\ 2-s, & 1 \leq s \leq 2. \end{cases}$$

Using calculations along the lines of those summarized in section four and reported by Hansen and Sargent [17], the discrete cross-covariogram and spectral density of the j th-differenced series can be deduced from (21).

6. POOLING DATA SAMPLED AT DIFFERENT INTERVALS

Suppose that we have available discrete point-in-time observations of the following kind: There are T_1 monthly observations on $v(t) = [y(t)', z(t)']'$ followed by T_2 weekly observations on $v(t)$. For convenience, assume that there are 13/3 weeks per month. We briefly describe how to construct and to approximate the likelihood of the pooled sample of monthly and weekly data as a function of the parameters of the continuous time model. One of the advantages of formulating and estimating models in continuous time is the availability of these natural procedures for pooling observations at different sampling intervals.

Define the stacked vector of the T_1 monthly observations

$$v_m = \begin{bmatrix} v(\frac{13}{3}) \\ v(\frac{26}{3}) \\ \vdots \\ v(\frac{13}{3}T_1) \end{bmatrix} .$$

The stacked vector of weekly observations is defined by

$$V_w = \begin{bmatrix} v(\frac{13}{3} T_1 + 1) \\ v(\frac{13}{3} T_1 + 2) \\ \vdots \\ v(\frac{13}{3} T_1 + T_2) \end{bmatrix} .$$

The stacked vector of pooled monthly and weekly observations is then

$$\bar{v} = \begin{bmatrix} v_m \\ v_w \end{bmatrix} .$$

Define the covariance matrix of \bar{v} as $\Gamma = E\bar{v}\bar{v}'$. The matrix Γ is dimensioned $[(T_1 + T_2)p] \times [(T_1 + T_2)p]$. Its elements are functions of the deep parameters of the model, and can be filled in by using representation (21). The normal log likelihood function for vector of pooled monthly and weekly observations is given by the obvious counterpart to (27).

Various approximations to the log likelihood function can be constructed along the lines described in section three. For example, from (24) the theoretical spectral density of the weekly data is

$$(34) S_w(\omega) = \sum_{j=1}^r W_j \frac{1}{1 - \alpha_j e^{-i\omega}} + \sum_{j=1}^r W_j \frac{\alpha_j e^{+i\omega}}{1 - \alpha_j e^{+i\omega}} .$$

Using a similar logic to that which led to (34), it is straightforward to show that the spectral density of the monthly data is then

$$(35) S_m(\omega) = \sum_{j=1}^r W_j \frac{1}{1 - \delta_j e^{-i\omega}} + \sum_{j=1}^r W_j \frac{\delta_j e^{+i\omega}}{1 - \delta_j e^{+i\omega}}$$

where $\delta_j = e^{\frac{13\lambda_j}{3}} = \alpha_j^{\frac{13}{3}}$. Now let $I_m(\omega_j)$ be the periodogram of the monthly data for $\omega_j = 2\pi j/T_1$, $j = 1, \dots, T_1$, and let $I_w(\omega_j)$ be the periodogram of the weekly data for $\omega_j = 2\pi j/T_2$, $j = 1, \dots, T_2$. Then an approximation to the log likelihood function is

$$(36) L^{**} = -p(T_1 + T_2) \log 2\pi - \frac{1}{2} \sum_{j=1}^{T_1} \log \det S_m(\omega_j) - \frac{1}{2} \sum_{j=1}^{T_2} \log \det S_w(\omega_j) - \frac{1}{2} \sum_{j=1}^{T_1} \text{trace} [S_m(\omega_j)^{-1} I_m(\omega_j)] - \frac{1}{2} \sum_{j=1}^{T_2} \text{trace} [S_w(\omega_j)^{-1} I_w(\omega_j)] .$$

An advantage of (36) is that it can be evaluated rapidly. A disadvantage is that it ignores information contained in the sample in the form of covariance between some of the weekly and monthly data. Presumably, an approximate likelihood function in the time domain, along the lines suggested in section three, could be constructed that utilizes that information.

7. AN ILLUSTRATION

To illustrate the computational feasibility and cost of the method of sections two and three, we report here the results of estimating a synthetic example along the lines of example (i) of section two. The example is a model of a firm or industry that maximizes (3). We set $n=1$, $p=2$, so that $y(t)$, $x_1(t)$ and $x_2(t)$ are each univariate. We set $G(D) = \sqrt{1/2} + D$, with F_1 and F_2 as positive scalars, and

$$\phi(D) = \begin{bmatrix} \psi_1^1 D & 0 \\ 0 & \frac{\psi_0^2}{\theta_0^2 + D} \end{bmatrix}$$

where we set $\theta_0^2 = 2$, $\psi_1^1 = 2.83$, and $\psi_0^2 = 2$. For this optimization problem, the characteristic polynomial of the Euler equation is

$$-F_2 s^2 + (F_1 + \frac{1}{2} F_2) = -F_2 (s - (\frac{F_1}{F_2} + \frac{1}{2})^{\frac{1}{2}}) (s + (\frac{F_1}{F_2} + \frac{1}{2})^{\frac{1}{2}}).$$

Therefore, the solution of this optimization problem is (1) with

$H(D) = D + \left(\frac{F_1}{F_1} + \frac{1}{2}\right)^{\frac{1}{2}}$ and $J(-D) = -\frac{1}{F_2}(-D + \left(\frac{F_1}{F_2} + \frac{1}{2}\right)^{\frac{1}{2}})$, which can be written as

$$\left[D + \left(\frac{F_1}{F_2} + \frac{1}{2}\right)^{\frac{1}{2}}\right]y(t) = -\frac{1}{F_2} \int_0^{\infty} \exp[-u\left(\frac{F_1}{F_2} + \frac{1}{2}\right)^{\frac{1}{2}}] E_t[x_1(t+u) + x_2(t+u)] du.$$

The econometrician observes $y(t)$ and $x_2(t)$, but not $x_1(t)$. For this particular model, the section three calculations imply that the joint (y, x) process in continuous time has a representation of the particular first-order vector autoregressive form,

$$\begin{bmatrix} D + K_{11} & K_{12} \\ 0 & \theta_0^2 + D \end{bmatrix} \begin{bmatrix} y(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} Q_{11}^0 & 0 \\ 0 & \psi_0^2 \end{bmatrix} \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}.$$

The parameters K_{11} , K_{12} , and Q_{11} are functions of the deep parameters F_1 , F_2 , θ_0^2 , ψ_1^1 , ψ_0^2 . The model is just identified.

Table 1 reports estimates that were obtained by using the frequency domain estimator of section three, as well as the cost of obtaining the estimates on the Cyber computer at the University of Minnesota. We report the parameter estimates together with 95% confidence bands calculated using the normal distribution together with the asymptotic covariance matrix calculated from the approximate information matrix. The first three experiments explore the effects and costs of increasing the sample size T .

Relative to experiment 1, experiments 4 and 5 study the effect of increasing the weight of the "feedforward" part of the decision rule F_2^{-1} , while leaving fixed the integrating factor

$\exp[-u(F_2^{-1} + \frac{1}{2})^{\frac{1}{2}}]$. Experiment 6 studies the effect of driving the root of the $\theta^2(s)$ polynomial toward zero. All experiments, with the exception of the one cited in footnote 2 of table 1, used the true parameters as the initial values for the nonlinear optimization. The nonlinear optimization procedure first used DFP of Goldfeld and Quandt [8] until convergence, and then switched to GRADX to assure convergence.

We have also estimated higher order systems. An example is the overidentified system in which the true parameters are $F_1 = .5$, $F_2 = 1$, $\theta^1(D) = 1$, $\psi^1(D) = 2.83$, $\psi^2(D) = 2$, $\theta^2(D) = 3 + 3.5D + D^2$. For $T = 120$, we obtained the following estimates: $F_1 = .41 \pm .49$, $F_2 = .41 \pm .49$, $\psi^1(D) = 1.30 \pm 1.47$, $\psi^2(D) = 1.23 \pm .42$, $\theta^2(D) = [2.32 \pm .73] + [2.12 \pm 1.01] D + D^2$. This required 44.6 c.p.u. seconds and cost \$3.02.

These computations indicate that the estimation procedures that we have suggested are economical, at least for systems with few parameters.

Table 1

Experiment number	F_1	F_2	θ_0^2	ψ_0^1	ψ_0^2	T	C.P.U.(seconds)	Cost ¹
Truth	1/2	1	2	2.83	2			
1	.44±.63 ²	.43±.60	1.83±.93	1.35±1.85	1.72±.43	120	21.6 ³	\$ 1.46
2	.49±.52	.72±.68	1.94±.56	2.09±1.97	2.06±.30	400	52.3	3.54
3	.38±.27	.97±.62	2.00±.30	2.75±1.75	1.96±.15	1600	162.4	11.01
Truth	1/4	1/2	2	2.83	2			
4	.22±.31	.21±.30	1.83±.93	1.35±1.85	1.72±.43	120	21.2	1.44
Truth	.0025	.005	2	2.83	2			
5	.0022±.0019	.0021±.0032	1.83±.86	1.35±.028	1.72±.41	120	22.6	1.53
Truth	1/2	1	.1	2.83	2			
6	.67±.23	.53±.35	.14±.09	1.76±.95	1.95±.22	120	17.7	1.20

1. At 6.78¢ per c.p.u. second.
2. Approximate 95% confidence band.
3. When all starting value were doubled, c.p.u. seconds = 36.9, cost = \$2.50.

8. CONCLUSIONS

This paper describes formulas that explicitly characterize the cross-equation rational expectations restrictions in continuous time linear models. Like the corresponding formulas for discrete time systems developed by Futia [6] and Hansen and Sargent [11], these formulas are interesting in their own right for theoretical work, such as Futia's [6]. The formulas become all the more valuable when the goal is empirical estimation, since they are quick to evaluate and therefore potentially compatible with maximum likelihood and other iterative nonlinear estimation strategies.

The paper also describes an exact and computationally feasible procedure for deducing the discrete time covariogram and spectral density as a function of the continuous time parameters. This is an essential tool in acquiring the ability to employ both exact and various approximate maximum likelihood procedures.

The theoretical and estimation framework described in this paper has implications in a number of directions, only some of which we have explored above. Theorizing and estimating in terms of a continuous time model has a number of advantages. As we have seen, such a model provides a relatively automatic answer to the question of how to utilize discrete data, some of which are point-in-time, while others are unit averages. It also provides relatively automatic answers to the problem of optimally pooling observations at systematically different time intervals, as analyzed in section six. Similarly, although we did not draw them out here, our model naturally implies a set of distinct procedures

for dealing with the problem of "optimal interpolation by related series".¹³ For the purpose of estimating models like ours, with their extensive cross-equation restrictions, it would not generally be an optimal procedure to create a synthetic series by interpolating from related series. Instead, the appropriate procedure would be to choose free parameters to maximize the exact or approximate likelihood of the available sample, including in the sample whatever data are available on the missing variables. Such procedures could be described, using methods similar to those of section six.

However, perhaps the biggest advantage of theorizing and estimating in continuous time is its potential for protecting us against the kinds of errors of aggregation over time studied by Sims [38, 39, 40] and Geweke [7]. While the nature of such approximation errors has yet to be studied in the specific context of rational expectations models, the results of Sims [39, 40] provide clear enough indication that for certain specifications of θ , ψ , H , and J of (1) and (2), estimating the parameters of an "analogous" discrete time system could lead to serious errors in interpretation and policy advice.

FOOTNOTES

1. Helpful comments on an earlier draft were made by Christopher A. Sims. The example in section seven was calculated by Lawrence Christiano. This research was supported in part by NSF Grant SES-8007016.
2. Thus, $x_3(t)$ is an "information variable" that will appear in the solution to (1) and (2) only because it "Granger-causes" $x_2(t)$ in continuous time. The device of withholding data on $x_1^2(t)$ from the econometrician in order to rationalize an error term in the econometric model was extensively studied by Hansen and Sargent [11] in a discrete time setting.
3. See Kwakernaak and Sivan [21] for a definition of a continuous time vector white noise, and for a description of the operational properties of the Dirac delta generalized function.
4. With some straightforward modification of interpretations, the results obtained by Whiteman [47, ch. 4] for discrete time systems could be used to characterize the existence and uniqueness of covariance stationary solutions to models of the form (1) and (2).
5. See Hansen and Sargent [17] for a proof and a thorough discussion of this example.
6. Hansen and Sargent [17] describe in detail an algorithm for factoring $[F_1 + G(-D)'F_2G(D)]$ in the desired way. An algorithm for factoring more general non-symmetric polynomials could be constructed by pursuing the results of Whiteman [47, ch. 4].
7. Hansen and Sargent [17] describe a recursive algorithm for calculating the B_j 's of (4).
8. Equation (8) is a solution to the prediction problem in terms of current and past values of the white noise w . Hansen and Sargent [17] display alternative representations of this solution. For instance, if x has a continuous time autoregressive representation, then the solution can be represented in terms of current and past values of x and/or derivatives of x . Hansen and Sargent [17] also show that certain representations of this prediction formula remain valid when the x process has explosive autoregressive roots.
9. For models in which $J(-D) = H(-D)$ and which come from linear quadratic optimal control problems, equations of the form (1) are the Euler first-order necessary conditions for the certainty equivalent problem. Solutions of the form (9) for such problems are typically calculated by solving the algebraic matrix Riccati equation for an associated optimal linear regulator problem. The algebraic Riccati equation is

necessarily solved by iterative methods, e.g. by Vaughan's [45] eigenvalue method, which are typically much more expensive than using (9). For some illustrative cost calculations for a related discrete time problem, see Hansen and Sargent [12].

10. Hansen and Sargent theoretically establish identification only for two special cases of the model (1) - (2). However, in any actual applied context, identification can easily be checked numerically in the course of maximizing the approximate likelihood function. In particular, to check that identification is secure against the aliasing problem, the procedures of Hansen and Sargent [17, appendix C] can be used to construct an "aliased" version of the driving process z at a tentative set of parameter estimates. The likelihood can be evaluated with the driving process parameters at these "aliased" values, and compared with the likelihood at the original tentative estimates.
11. The formula for $\bar{r}_{11}(\tau)$, $\tau > 0$ is calculated by observing that
- $$\bar{r}_{11}(\tau) = E \int_0^{\tau} \int_0^{\tau-u} x_1(t-s) x_1(t-\tau-u) ds du = \int_0^{\tau} \int_0^{\tau-u} r_{11}(\tau+u-s) ds du.$$
- Substituting for $r_{11}(\tau+u-s)$ from (21), and evaluating the resulting integral gives the second line of (29). The remaining of formulas (20) and (31) are derived in a similar way.
12. Notice that since the zeroes of $\det H(s)$ have been assumed less than zero in real part, it follows that the zeroes of $\det H(s-\gamma/2)$ are less than $\gamma/2$ in real part. Similarly, the zeroes of $\det \theta(s-\gamma/2)$ are less than $\gamma/2$ in real part, while the zeroes of $\det J(-s-\gamma/2)$ are greater than $-\gamma/2$ in real part. From these last observations, it follows that the convolution integrals
- $$E_t J(-D-\gamma/2)^{-1} [x_1(t) + x_2(t)]$$
- converge.
13. See Milton Friedman [5] and Gregory Chow and An-loh Lin [3].

References

- [1] Bard, Y.: Nonlinear Parameter Estimation, New York: Academic Press, 1974.
- [2] Cagan, P.: "The Monetary Dynamics of Hyperinflation," in: M. Friedman, ed., Studies in the Quantity Theory of Money, Chicago: University of Chicago Press, 1976.
- [3] Chow, G. and A. Lin: "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series," Review of Economics and Statistics, 53, (1971), 372 - 375.
- [4] Dornbusch, R., "Expectations and Exchange Rate Dynamics," Journal of Political Economy, 84, (1976), 1116 - 1176.
- [5] Friedman, M., "The Interpolation of Time Series by Related Series," Journal of the American Statistical Association, 57 (1962), 729 - 757.
- [6] Futia, C., A., "Rational Expectations in Stationary Linear Models," Econometrica, 49, (1981), 171 - 192.
- [7] Geweke, J.B.: "Temporal Aggregation in the Multivariate Regression Model," Econometrica, 46, (1978), 643 - 662.
- [8] Goldfeld, S. M., and R. E. Quandt: Nonlinear Methods in Econometrics, Amsterdam: North Holland, 1972.
- [9] Granger, C. W. J.: "Investigating Causal Relations by Econometric Models and Cross-spectral Methods", Econometrica, 37, (1969), 424-438.
- [10] Hannan, E. J.: Multiple Time Series, New York: Wiley, 1970.
- [11] Hansen, L. P., and T. J. Sargent: "Formulating and Estimating Dynamic Linear Rational Expectations Models," Journal of Economic Dynamics and Control, 2, (1980), 7-46.
- [12] Hansen, L. P., and T. J. Sargent: "Linear Rational Expectations Models for Dynamically Interrelated Variables," in: R. E. Lucas, Jr., and T. J. Sargent, eds., Rational Expectations and Econometric Practice, Minneapolis: University of Minnesota Press, 1981.
- [13] Hansen, L. P. and T. J. Sargent: "Aggregation Over Time and the Inverse Optimal Predictor Problem for Adaptive Expectations in Continuous Time," Working Paper 157, Federal Reserve Bank of Minneapolis, 1980.
- [14] Hansen, L. P., and T. J. Sargent:, "Identification of Continuous Time Rational Expectations Models from Discrete Time Data," manuscript, 1981.

- [15] Hansen, L. P. and T. J. Sargent: "The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities," manuscript, 1981.
- [16] Hansen, L. P. and T. J. Sargent: "A Note on Wiener - Kolmogorov Prediction Formulas for Rational Expectations Models," manuscript, 1981.
- [17] Hansen, L. P. and T. J. Sargent: "Methods for Estimating Continuous Time Rational Expectations Models from Discrete Time Data," Staff Report 59, Federal Reserve Bank of Minneapolis, 1980.
- [18] Hansen, L. P. and T. J. Sargent: "Exact Linear Rational Expectations Models: Specification and Estimation," manuscript, 1981.
- [19] Hillmer, S. C. and G. C. Tiao: "Likelihood Function for Stationary Multiple Autoregressive Moving Average Models," Journal of the American Statistical Association, 74, (1979), 652-660.
- [20] Koopmans, L. H.: The Spectral Analysis of Time Series, New York: Academic press, 1974.
- [21] Kwakernaak, H. and R. Sivan: Linear Optimal Control Systems, New York, Wiley, 1972.
- [22] Lucas, R. E., Jr.: "Distributed Lags and Optimal Investment Policy," manuscript, 1965.
- [23] Lucas, R. E., Jr.: "Optimal Investment With Rational Expectations," manuscript, 1966.
- [24] Lucas, R. E., Jr.: "Econometric Testing of the Natural Rate Hypothesis," in O. Eckstein, ed., The Econometrics of Price Determination: Conference October 30-31, 1970, Washington, D.C., Washington: Board of Governors of the Federal Reserve System, 1972.
- [25] Lucas, R. E., Jr.: "Econometric Policy Evaluation: A Critique", in K. Brunner and A. H. Meltzer, eds., The Phillips Curve and Labor Markets, Carnegie-Rochester Conferences on Public Policy 1, Amsterdam: North Holland, 1976.
- [26] Lucas, R. E., Jr., and E. Prescott: "Investment Under Uncertainty", Econometrica, 39, (1971), 659 - 681.
- [27] Lucas, R. E., Jr. and T. J. Sargent: Introductory Essay to Rational Expectations and Econometric Practice, Minneapolis: University of Minesota Press, 1981.

- [28] Mortensen, D. T.: "Generalized Costs of Adjustment and Dynamic Factor Demand Theory," Econometrica, 41, (1973), 657-665.
- [29] Muth, J. F.: "Rational Expectations and the Theory of Price Movements," Econometrica, 29, (1961), 315 - 335.
- [30] Phadke, M. S., and G. Kedeem: "Computation of the Exact Likelihood Function of Multivariate Moving Average Models," Biometrika, 65, (1973), 511- 519.
- [31] Phillips, A. W.: "The Estimation of Parameters in Systems of Stochastic Differential Equations," Biometrika, 59, (1959), 67 - 76.
- [32] Phillips, P. C. B.: "The Structural Estimation of a Stochastic Differential Equation System", Econometrica, 40, (1972), 1021 - 1041.
- [33] Phillips, P. C. B.: "The Problem of Identification in Finite Parameter Continuous Time Models," Journal of Econometrics, 1, (1973), 351 - 362.
- [34] Phillips, P. C. B.: "The Estimation of Some Continuous Time Models," Econometrica, 42, (1973) 803 - 824.
- [35] Rozanov, Y. A.: Stationary Random Processes, San Francisco: Holden-Day, 1967.
- [36] Saracoglu, R. and T. J. Sargent: "Seasonality and Portfolio Balance Under Rational Expectations," Journal of Monetary Economics, 4, (1978), 435 - 458.
- [37] Sargent, T. J., Macroeconomic Theory, New York: Academic Press, 1979.
- [38] Sargent, T. J., "Interpreting Economic Time Series," Journal of Political Economy, 89, (1981), 213 - 248.
- [39] Sims, C. A., "Discrete Approximations to Continuous Time Lag Distributions in Econometrics," Econometrica, 39, (1971), 545 - 564.
- [40] Sims, C. A.: "Optimal Stable Policies for Unstable Instruments," Annals of Economic and Social Measurement, 3, (1974), 257-266.
- [41] Sims, C. A.: "Approximate Specification in Distributed Lag Models," manuscript, 1971. Paper Prepared for the 38th Session of the International Statistical Institute, Washington, D.C., August, 1971.
- [42] Taylor, J. B.: "Staggered Wage Setting in a Macro Model," American Economic Review, Papers and Proceedings, (1979), 108 - 118.

- [43] Taylor, J. B.: "Output and Price Stability: An International Comparison," Journal of Economic Dynamics and Control, 2, (1980).
- [44] Treadway, A. B.: "On Rational Entrepreneur Behavior and the Demand for Investment," Review of Economic Studies, 36, (1969), 227 - 240.
- [45] Treadway, A. B.: "Adjustment Costs and Variable Inputs in the Theory of the Competitive Firm," Journal of Economic Theory, 2, (1970), 329 - 347.
- [46] Vaughan, D. R.: "A Negative-exponential Solution to the Matrix Riccati Equation", IEEE Transactions on Automatic Control, 14, (1969), 72 - 75.
- [47] Whiteman, C. H., "Moving Average Representations in Rational Expectations Models," unpublished Ph.D. dissertation, University of Minnesota, 1981.
- [48] Wilson, C. A.: "Anticipated Shocks and Exchange Rate Dynamics," Journal of Political Economy, 87, (1979), 639 - 647.
- [49] Wymer, C. R.: "Econometric Estimation of Stochastic Differential Equation Systems," Econometrica, 40, (1972), 565 - 577.